Localization and Counting of Indoor Populations on a University Campus using Wi-Fi Connection Logs and Floor Plans

by

Lannois Carroll-Woolery

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Management Sciences

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

The localization and counting of persons in indoor spaces is an area of extensive research. Indoor population metrics can inform energy conservation, health and safety, security, resource optimization, and location-aware services such as marketing and navigation. Building utility is impacted by the number of persons in each space, and the management of person flows into and out of building spaces is a critical consideration of space design, and the COVID-19 pandemic elevated the need to accurately measure and monitor indoor populations.

Indoor populations' size, movement and location can be ascertained by a variety of automatic means, but scalability, repeatability and cost are limiting factors. One low-cost technique is the use of wireless logs from Wi-Fi-enabled devices, which provide precise counts but inaccurate locations due to Access Points' widely varying coverage areas. Population locations, as estimated by wireless logs, are usually defined at a floor, or building level.

In this paper, I propose a generalized technique for more precise identification of indoor populations' location, using wireless logs. It is based on the merging of connection logs with floor layout plans, to define floor zones, representing the general area(s) of wireless coverage provided by each wireless AP, including areas served by more than one AP. The combined information allows for more precise location and counting of indoor populations. This analysis could be useful across multiple functional domains, including sustainability management, resource optimization, and capacity monitoring. The technique can be implemented in any environment where there is an extensive wireless network, widespread usage of the network, and reliable data records. It is non-invasive and does not require the purchase or installation of new equipment.

As a case study, we applied the technique to data from a mid-sized university. Spatial and temporal population analyses were completed using wireless logs collected over a 6-week period prior to the COVID pandemic. The logs included unique User Ids and Device Ids; The floor layout plans included the installed locations of AP devices. Facilities management records included building, floor, and room metrics. Population analyses were completed by building, room types, work weeks, and duration of wireless connections.

The population estimations for size and location were compared to expected indoor populations, based on student class enrolments and employee work schedules, to gauge accuracy and utility. Linear Correlation Coefficients were calculated for measured vs. expected population counts.

The results indicated that the definition of Building Floor Zones provided more accurate indoor population location values than floor-level estimates, across a variety of building types and room types. Facilities management definitions for Building Floors allowed generic description of campus spaces that could be applied to any environment with varying building usage and occupant activity. The merged data allowed the estimation of indoor populations' size and location at various levels of aggregation: zones, floors, and buildings; and allows for comparisons of activity in similar environments in differing locations. Possible research and/or application areas include: the use of indoor spaces outside of business hours, occupancy/utility rates, and the measurement of indoor crowd densities.

Acknowledgements

First, I would like to thank my supervisor Prof. Lukasz Golab to whom I am truly grateful. He offered to be my supervisor for my research thesis when my previous supervisor moved to another university. He was also very patient through the various changes in research focus. Most of all, he communicated the exciting possibilities for doing research in Data Science, using data that is readily available but often overlooked.

I would like to thank the readers of this thesis, Prof. Stan Dimitrov and Prof. Oliver Schneider for their feedback and encouragement to continue the work.

I appreciate my staff colleagues in Information Systems and Technology and the Space Planning Office: Mike Patterson, Ryan Goggin, and Scott Nicoll, who shared the masked data and their expertise with me. I hope that staff, faculty, and students will work together more closely in the future on data science projects at the University of Waterloo.

My wife Anandi was very supportive for the very long time it took to finish my Master's program part-time. I look forward to taking her on a very special vacation to celebrate our accomplishment.

A special shout-out to my adult children: Jalen, Jasia, and Janna who I hope will be inspired to continue learning, whatever their ages.

Thanks also to my parents and siblings: Alva-Joye, Albert, Rontgen, Nadine, Marsha, Ava-Ann, and Annika, for their love and support.

Table of Contents

Listing of Figures

Listing of Tables

1.0 Introduction

The counting and localization of persons in indoor spaces is important for both short- and long-term facilities management, building utility, and occupants' wellness. Accurate metrics can inform health and safety, security, energy conservation, and location-aware services such as marketing and navigation [4,6,8,11]. Cleaning and maintenance schedules; room bookings; heating, ventilation, and air control (HVAC) settings; lighting; security patrolling and the comfort and health of tenants are all informed by population metrics. Resource optimization is another area of study that could benefit from the accurate and precise measuring of indoor crowds: which spaces are being used to their full capacity, and what areas are available for individual study and/or ad hoc congregating, are questions that can be answered using occupancy measures.

Indoor spaces are typically unique environments, designed and configured for a specific function/purpose. They vary considerably in size, layout, patterns of use, amenities, and furniture/equipment. Rooms may be designated for single or group use, and for a variety of work and social activities. Ventilation elements such as windows, doors and HVAC systems may be based on older designs or made with newer materials. The age of a building is often correlated with the types of materials used in its construction, floor layouts, and capacity. It is challenging to transfer research learnings from one type of space to another.

The measurement of indoor populations is an area of significant research focus. Crowds' sizes, locations, directions, timing, and purpose are important factors for space utilization and management, resource conservation, service provision, occupant health and safety. Manual counting is often cost-prohibitive, or possible only with small populations. Real-time or near-real-time data, especially in areas of high or fluctuating pedestrian traffic, require automated tools.

Automatic counting technologies (ACTs) include both mechanical and information & communication technology (ICT) tools. Occupancy estimates can be delivered by physical counters such as gates/turnstiles; temperature and humidity sensors which measure room atmospheric changes; chemical sensors that monitor CO_2 levels; electromagnetic sensors such as motion detectors, video cameras, and infra-red sensors can all be deployed to locate and count individuals within defined spaces [7,11,12].

The implementation and use of ACTs can be difficult due to several factors, including building and room layout, the size and location of inanimate objects, and privacy concerns. Video cameras require clear lines of sight; structural elements such as columns and walls may hide persons from view. Heat-, sound-, and radiation-emitting devices may create "noise" that confuse electromagnetic and audio sensors. Facial recognition software is computationally expensive and raises privacy concerns. Chemical sensors may take significant time to register the presence of individuals because of airflow and diffusion. Many ACTs involve long data-capture times, are slow to report results, or require the purchase and installation of specialized equipment, making them prohibitively difficult and/or expensive to implement [6,7,10,12,14]. ACTs differ widely in accuracy, reliability, complexity, timeliness, and cost.

The use of already-deployed technology and infrastructure for indoor population estimation is a popular alternative. Widely used technologies such as wireless networks can be used for automatic counting in circumstances where the uptake of digital devices is high and the network infrastructure is expansive or can be easily upgraded [4,26].

The widespread use of smartphones and other Wi-Fi-enabled personal devices has meant that the number of Wi-Fi devices detected in a space is considered a reasonable proxy for the number of persons present [1,2,14,22]. The Media Access Control (MAC) Address is a unique device identifier, and other information exchanged in the broadcast frames contain highly specific information about the device, or its user. Wi-Fi components have been extensively deployed in portable electronic devices such as smartphones and laptops. These portable devices have Wi-Fi enabled by default and the capability is rarely disabled. The number of smartphone users worldwide was estimated to be 3.8 billion in 2021 [8]. The number of devices that support the IEEE 802.11 standards was 22 billion in 2018 and is expected to exceed 38 billion by 2025 [1], including over a billion Wi-Fi Access Points (AP's) [26]. Individuals tend to keep their portable computers, especially their smartphones, close at hand for communication, entertainment, and/or productivity [4,10,15]. Free Wi-Fi in public spaces means that consumers are more likely to make use of Wi-Fi networks rather than paid networks, such as cellular. Studies have estimated that the correlation between scanned devices and actual person counts is high, varying from 0.78 to 0.95 [1,2,8,12,18,26].

The count of Wi-Fi-enabled devices as a proxy for population size is subject to several assumptions. First, that most persons are carrying a Wi-Fi enabled device- an individual may choose to use a different wireless network, or none. Second, that everyone has only one Wi-Fi-enabled device: some individuals carry multiple smartphones, or a laptop and a cellphone, on their person. Third, that the Wi-Fi sensor is not detecting devices beyond the area being monitored: depending on line-of-sight considerations, the presence/absence of walls and other absorbing/reflecting objects, the effective range of a Wi-Fi sensor may vary significantly. Fourth, that the individual carrying the device is present in the defined space for a reasonable amount of time, allowing for periodic connection to a nearby access point (AP). Fifth, that the Wi-Fi enabled device is related to an individual, and is not fixed equipment within a space, such as a printer, or company computer. Finally, that the device's signal is detectable, not absorbed/blocked by surrounding objects, or too low-powered to be detected [2,11].

University campuses are considered excellent environments for population estimates using Wi-Fi connections. Most campuses have clear physical boundaries, within which individuals move frequently between buildings and rooms. Large numbers of students and employees, and the higher usage rate of smartphones, tablets, and laptops among campus users, means that there will be many Wi-Fi enabled devices for detection and analysis. Campus Wi-Fi networks are extensive, designed to provide seamless access across the campus, as students and employees move between classrooms and buildings, with few/no "dead zones", where wireless connections are non-existent. The low cost of Wi-Fi relative to cellular networks, and its greater reliability indoors means that students are more likely to make use of the Wireless Local Area Networks (WLAN) for network connectivity. Many of the academic and social resources required by students and employees are on the campus network, such as work and class schedules, Learning Management Systems (LMS's) and Enterprise Resource Planning systems (ERPs). Finally, the campus wireless network is available to only authorized users, i.e., students and employees, permitting verification and validation of expected vs. actual users of the network.

Spatial and temporal analyses of wireless connections on a university campus are a strong proxy for analyses of populations' location, size, and movement. Wi-Fi-enabled devices automatically connect to available APs as students move around campus. Wi-Fi-enabled devices send out periodic probe requests, depending on the devices' settings and power levels, to determine the AP with which it can form a connection. The preferred AP may change depending on the number of wireless network users, physical proximity to an AP, or network configuration.

In this research project, Wi-Fi session logs are combined with AP architectural diagrams, floor layout plans, and facilities management building and room data. The session logs are used to calculate the number of connected users at each AP, for each period. The floor layouts are used to define AP zones within each floor, where zones are defined as the rooms in the vicinity of the AP. Facilities management data is used to categorize Building Floors by primary purpose/activity. The combined data is then used to estimate population levels for each zone, for each period, by type of space.

The contributions of this research to the locating and counting of dynamic indoor populations are as follows:

1. Defining zones within Building Floors to allow for more precise locating of indoor populations.
2. Locating APs within zones by combining architectural diagrams of APs with floor layouts.
3. Categorization of Building Floors on campus by the predominant types of spaces and/or rooms on each Building Floor.
4. Comparison of measured vs. expected traffic in each zone by incorporating data on employee office locations and student classroom schedules.
5. Applying the methodology above to all Academic and Academic Support buildings on the university campus, demonstrating its utility and applicability to a variety of building types, floor layouts, and occupancy scenarios. Academic and Academic Support spaces are those designed for teaching, research, and scholarship, and the administrative support functions that enable the academic mission.

Challenges in applying the methodology included data errors, missing records, and the merging of data from different administrative systems. The data from each system was at times inconsistent, or unreadable, or erroneous. Administrative systems of record are often selected and maintained for departmental priorities, which can limit their usefulness for cross-functional and institutional-level analyses.

The Wi-Fi data was collected prior to the school shutdown precipitated by the COVID-19 pandemic; it represented the then-normal traffic patterns of students and employees on the main campus.

Spatial and temporal analyses of the estimated population, as measured by connection logs, were conducted on the combined Wi-Fi logs, floor plans, facilities management data, class schedules, and employee office locations. The analyses demonstrated the potential of tracking indoor populations at a more precise level than Building Floor. The data also highlighted possibilities for population analyses outside of working hours, e.g., on evenings and weekends, to better understand campus activity.

Population estimates for each zone were compared to the expected occupancy, to gauge accuracy. The correlation analyses were limited to Academic and Academic Support buildings and floors, that is, campus areas where scheduled student and employee activities (e.g., classes, labs, office work) took place. This was done for two reasons: to limit the analysis to areas where corroborating data (in the form of class schedules and employee schedules) could be used to check the results, and because data for other types of campus buildings (e.g., residence floor plans) was not readily available.

2.0 Related Work

2.1 Introduction

The automatic counting of individuals by information & communication technology (ICT) means has been researched using video cameras, infrared sensors, motion detectors and wireless signals. Each methodology has differing strengths and weaknesses, including the (in)ability to detect objects behind objects and walls, (in)ability to distinguish two individuals in proximity, requirement for specialized equipment, computational processing power, and cost.

The use of wireless signals for automatic counting has been an area of research focus because of its unobtrusiveness, relatively low cost, and use of existing technology/infrastructure. It does not make individuals feel that they are being watched/monitored and does not require their co-operation or participation. Extensive Wi-Fi networks, the widespread use of wireless-enabled devices such as cellphones, tablets and cellphones, and the cheap costs of Wi-Fi adapters and related components have allowed both active and passive Wi-Fi automatic counting technologies (ACTs) to be implemented.

Research into the use of Wi-Fi logs to measure indoor populations' size, location and movement has generally been conducted in two situations: in specific buildings, floors and/or rooms, or over extensive areas with multiple buildings. In specific locations, the floor layout and occupancy patterns are known, and ground truth – the actual number of persons in a space- can be easily determined by manual counting or by use of security infrastructure, like access card readers. In extensive areas that span multiple buildings, the floor layouts, occupancy patterns and ground truth are unknown; population estimates are determined at a building and/or floor level by counting the number of wireless connections on each floor.

In this section, previous related research using Wi-Fi logs is recounted.

2.2 Wi-Fi Automatic Counting Technologies

Research into the use of Wi-Fi infrastructure and technology for automatic counting of persons can be divided into three areas: analysis of electromagnetic waves' signal interference, analysis of Wi-Fi packets broadcast by devices, and analysis of session logs.

Signal interference involves measuring the attenuation and reflection of radio waves caused by individuals in their path. This area of research involves the positioning of emitters and receivers that measure Received Signal Strength (RSS), and Channel State Information (CSI), i.e., changed electromagnetic properties of the radio waves [14,15]. After calibration in an empty space, the degrees of change caused by persons in the path of the waves may be used to determine the

number and positions of persons and can also classify types of motion. Recent research has focused on people counting, hand gestures, and activity classification [6]. This area of research does not require individuals to be carrying a Wi-Fi enabled device- the absorption and/or refraction of radio waves by their bodies provides the data for inferring their presence and activity. Signal interference can be used for fine-grained analysis, such as binary analysis (e.g., Are persons present in a space or not?) Macro analysis is more difficult because of the challenges in making sense of many individuals in a dynamic environment, interacting with each other and inanimate objects. The need to purchase and install transmitters and receivers is also a drawback for this research method.

The analysis of Wi-Fi packets involves the deployment of wireless receivers in defined spaces. It assumes that individuals are carrying a wireless-transmitting device. The research aims to determine the position, motion, and number of persons by the phase shift (Doppler effect), Time Difference of Arrival (TDOA) between sensors, Received Signal Strength (RSS) or Channel State Information (CSI) data associated with the periodic probe frames emitted by mobile devices [7,10]. The waves' properties measured at the receivers vary with distance, reflection, absorption, room layout, the presence of objects and people, and the frequency and strength of the radiation pulses emitted by the device. For these reasons, each space requires calibration for proper interpretation of the results [26]. Analysis of Wi-Fi probes is subject to errors caused by MAC address randomization, and incorrect deployment of receivers. Wi-Fi packet research also gives less accurate results when the sample population is small because the assumption of a one-to-one ratio between Wi-Fi devices and individuals may not hold true. Macro analysis is more accurate, because assumptions about the uptake and use of Wi-Fi-enabled portable devices are more likely to hold true for a larger group.

The analysis of session logs aims to determine persons' positions, counts, and activities by analyzing the history of wireless associations between devices and APs, given the spatial localization of the APs and other contextual data [21]. Depending on the APs' location and configuration, and the building and room layout, the timing and duration of the AP associations can be used to infer high-level estimates of occupancy and activity. The large amount of specific content in the probe requests, high degree of detail in the logs (e.g., timestamps, duration, MAC addresses, User Id, network) together provide ample information for Big Data analysis. Session logs' research techniques include time series analysis, association rules, clustering, machine learning, and forecasting. The session logs' data is often combined with related information such as building specifications, energy readings, or individuals' personal schedules, before analysis.

Privacy protection is also a significant concern for users of Wi-Fi data. To address user concerns about misuse of personal data, identifying information in Wi-Fi probe requests or session logs is usually hashed, or otherwise de-identified, before analysis.

2.2 Wi-Fi Session Logs Automatic Counting Technology Research

The use of session logs for crowd localization, counting and profiling has several advantages compared to analyses of Wi-Fi probe requests and/or signal interferences. It has lower start-up and implementation costs: probe request and signal interference analyses usually require the purchase, installation, and calibration/configuration of equipment, while session logs are provided by existing infrastructure [2]. Session logs are device-independent: Wi-Fi-enabled devices vary in their sleep-mode settings, power output, and frequency of network scans [12]. Session logs record User information as well as Device Ids (MACs), so unique users can be easily identified, even when they carry multiple Wi-Fi-enabled devices, or MAC addresses are changed after a software upgrade, or MAC addresses are randomized while scanning for nearby APs. Only completed/successful connections are recorded in the session logs.

Session logs' analysis is subject to several of the same assumptions and errors of other Wi-Fi-based crowd analyses. Wi-Fi signals from stationary (non-human) devices such as printers introduce errors. Individuals who do not carry Wi-Fi-enabled devices, or use other wireless connections such as cellular networks, also introduce estimation errors. Wi-Fi associations may be formed between AP and devices that are not in the same physical area, because of signal interference, load-balancing algorithms, or building floor layout. Conversely, a wireless device may continually switch association between nearby APs. Scaling and repeatability of the research conducted is challenging because of varied building, floor and rooms' layouts and usage.

Researchers have attempted to address these limitations in varied ways. Wi-Fi signals from stationary (i.e., non-human) sources have been scrubbed from session log datasets by noting which devices send signals during non-working hours, are active for more than 24 hours continuously or have a MAC address that does not correspond to a user device like a laptop or cellphone. Individuals who do not carry Wi-Fi-enabled devices are either treated as non-significant sources of errors, or are categorized/clustered, with specific analyses created for each category/cluster. Difficulties in localizing individuals in a particular area are addressed by confining analyses to specific building areas with clear boundaries, or summarizing results across an entire floor or building. In addressing the limitations, the researchers were able to draw useful conclusions and suggest areas for more in-depth research.

Ichifuji et al [33] traced foreign visitors' trip patterns by temporal analysis of session logs from specific tourist sites in several cities. Visitors were identified as Wi-Fi users whose total trip duration was seven days or less. The research assumed that most visitors made use of the free Wi-Fi at each site, and that visitors to the area did not stay longer than a week.

Office buildings provide a more controlled environment for crowd analysis, with more structured occupancy patterns, small numbers of users, and the ability to collect ground-truth data from optical cameras, access card readers, or manual counting of occupants. On the other hand, scalability and repeatability are challenging because of varying building layouts, wireless

architecture, and user behaviour. Rafsanjani and Ghahramani [34] studied the correlation between building occupancy and energy usage on a single floor of each of two office buildings. Each office area had a single AP, serving 11 and 16 employees, respectively. Identified sources of error included wireless connections to APs in adjacent offices, low spatial resolution, and sensitivity of the results to office workers turning off Wi-Fi. Wang et al [11] analysed session logs' durations from the third and fourth floors of an office building, housing up to 74 employees. The floors had seven and nine APs, respectively. They used machine learning algorithms to estimate the number of individuals in the offices, comparing the results to ground-truth data gathered from camera-based sensors. The major limitation identified was transferability to other spaces.

University campuses' session logs are much larger, and much more complex, to analyze. Wi-Fi users may be visitors, students, or employees. Building and floor layouts and usage vary significantly, as does the wireless architecture. The volume of session logs generated requires Big Data techniques. Hobson et al [8] analysed seven months of logs from a multi-story academic building with up to 565 occupants to forecast weekday and weekend usage. Kamińska and Graña [10] used geostatistical methods and the physical location of APs to create a 3D estimate of the number of persons in discretized areas on each of six floors of an academic building. Gao et al [13] categorized faculty vs. students by analyzing logs from 27 different campus buildings. Binthaisong et al [28] calculated the number of users present in each of 22 campus buildings. Zagatti et al [27] used session logs from the National University of Singapore, taken at 15-minute intervals, along with APs' geospatial information, to calculate near-real-time crowd sizes for each building and floor level on the campus.

In this project more specific crowd locations are proposed by defining "zones" that correspond to APs' floor locations. The number of wireless connections per zone is assumed to be a good estimate of the number of persons in each zone. The total counts per zone are compared to the total counts per floor, in each building, for differing Building Floor purposes. The total counts are also compared to the expected counts on each building and floor. The methodology is scalable and repeatable, assuming the complementary datasets of AP locations, floor layouts and room types are available for analysis.

3.0 Data

3.1 Introduction

University IT systems are owned and maintained by individual departments; they are designed and configured for specific operational and strategic functions, such as course registration, payroll processing, and scheduling. Enterprise resource planning (ERP) IT systems are used widely and have standardized definitions and data formats that are understood campus wide. Smaller administrative IT systems serve the purposes of individual units, and often have differing/unique data definitions, formats, and business rules.

In this research project, records from six (6) campus administrative systems were merged to draw insights on indoor populations. Each of the six datasets were filtered and transformed to facilitate merging: invalid/or and missing information was corrected/removed; data in image files were encoded; duplicate records were merged; building and floor names were standardized; datetime periods were harmonized.

The following sections give details on each dataset, and the Extract-Transform-Load processes applied to each.

3.2 Listing of Datasets

This research required the following datasets:

1. Wi-Fi session logs.
2. Building and floor layouts.
3. Wi-Fi Access Point Installations.
4. Space Metrics for Buildings and Floors.
5. Course Schedules.
6. Employee Office Locations.

Wi-Fi session logs record the User Id, Device Id, AP Name, Start Time, and End Time of each wireless connection on campus. Authorized Wi-Fi users include university employees, students, and guests. Staff and students from other universities can also connect to the wireless network using the EDUROAM roaming service.

The logs were provided for the first six weeks of 2020, after access was authorized. Prior to being shared, the Start Times and End Times were rounded to the nearest half-hour, and the User Ids

and Device Ids were Hashed for each 24-hr period, so that the IDs were distinct and consistent for each day, but not traceable over multiple days. Over 27 million Wi-Fi records were provided, for the first 6 weeks of 2020.

Building and floor layouts are architectural diagrams, made available online to authorized users. The floor layouts include structural elements and room names. Authorized users have access to the floor diagrams used by Academic and Academic Support employees for the university's teaching and research functions; floor layouts for off-limit areas such as residences, basements, and elevator bulkheads are unavailable. New and updated diagrams are made available periodically by facilities management staff, after construction and major renovations.

The building and floor layouts were accessed online. Over 600 pdf files are listed on the webpage, of which approximately 200 plans, for over 60 Academic and Academic Support buildings, were available for viewing.

AP installations are architectural diagrams showing the locations of APs relative to the building and floor layouts. The diagrams were provided as pdf files, with images of named APs superimposed on the floor layouts. Approximately 390 pdf files for 82 buildings were provided by the university's IT department.

Space metrics for buildings and floors include room names, room types, sizes, ownership, and designated use. Campus spaces include classrooms, offices, hallways, stairwells, washrooms, labs, athletic facilities, atria, and computer server rooms. Spaces are not necessarily defined by walls, may be multi-purpose, might be shared between departments, or be inactive/unusable. Space metrics are maintained by the Space Planning Office (SPO).

Space metrics were provided as comma-delimited extracts. Over 33,000 records for campus facilities in winter 2020 were delivered by the SPO.

Class schedules for winter 2020 included course names, course sections, course delivery mode, meeting dates and times, meeting locations, and enrolment. Class schedules are maintained by the Registrar's Office. The data was provided as comma-delimited files.

Class schedules were provided as comma-delimited extracts. Over 6,000 records were provided for the 2019/2020 academic year.

Full-time employees' office locations are available in an employee phone directory stored by the IT department. The data extract included building and room names. Information was provided for the Winter 2020 term.

The office locations were used to estimate the number of employees in each building zone for each period. Approximately 5,000 records were analyzed.

3.3 Wi-Fi Session Logs

The Security Operations Centre (SOC) is a subgroup of the Information Systems and Technology (IST) department's Information Security Services team. The SOC maintains logs of Wi-Fi network connections on the University of Waterloo campus. These logs can be used to monitor the performance, utility, and security vulnerabilities of the campus' wireless network.

The system of record for Wi-Fi connections at the time of the data collection was ONAlogs. The system logs included User Ids, Device Ids, APs, ssid (network information), Start Times and End Times for wireless connections. Employees, students, and guests generate approximately 600,000 wireless connections daily in campus buildings.

Prior to making the data available for analysis, data masking and data generalization was used to prevent any potential data breach. One-way hashing was used to mask confidential information, also Start and End Times were rounded. To preserve anonymity and privacy, the following edits were made to the data prior to delivery:

- All identifying information was scrubbed- User Ids and MAC addresses were replaced with new identifiers generated by a masking algorithm. The hashing function was reset every day, to prevent tracking of individuals across days. With the mapping, User Ids and Device Ids were consistently mapped within each 24-hour period.
- All Start Times and End Times were rounded to the nearest half-hour. That is, a Start Time after :45 and before :15 was mapped to :00, and a Start Time after :15 and before :45 was mapped to :30. The same transformation was applied to End Times. This meant that Start Times and End Times were the same whenever the connection to a particular AP lasted less than 15 minutes, also that all Start Times and End Times ended in :00 or 0:30.

The data transformations removed the possibility of tracking individuals across days and limited the scope of temporal analysis. Even with these limitations, the records provided useful insight into wireless connections during the first 6 weeks of the term.

A screen shot of the scrubbed data is shown in Figure 1. Each row represented one device's connection to a specific AP. The columns in the screenshot are, in order, **action**, **user**, **mac**, **start_time**, **end_time**, **ap** and **ssid**. For **action**, a value of LOGIN represented the first successful connection for the device for the 24-hr period, ROAM is used for all subsequent connections. The records are keyed by **user**, **mac**, and **ap**- that is, users with multiple devices connected to the network, and users with Wi-Fi connections to multiple APs, will have multiple rows in the data. The **Start Time** and **End Time** are logged with Eastern Standard Time (EST) values. The **ssid** values represent the service set identifier for the connection: *eduroam*, *guest*, and *uw-wifi-setup-no-encryption* are sample values shown.

action	user	mac	start_time	end_time	ap	ssid	u
LOGIN	uDO2xLJJGhQnnGOBOpsHfoTLbRQdsEHfCnZeSWuy	LcO4KY6HuGerIA2HyBFNtmYcCMMrTAJ6xb7eCwXm	2020-01-19 11:30:00+00	2020-01-20 02:00:00+00	RN-AP-CLVN-187-A	eduroam	7
LOGIN	SIEzzIV5h2LydI00lgQvFFPxtxuwK0PWdpujKS2i	bth2vHXnFUc1V7XTLNV4ONoIR4UW6C8o1UpCR1mv	2020-01-28 14:00:00+00	2020-01-28 14:00:00+00	WN-AP-HS-129-B	eduroam	b
ROAM	RgTtUK2fm1QuoAJP2F9D0V6BFUI4iURxiDRAPi5l	S0JDMMoCzFjo8yutLMKNFj9S99Bng81e6J94KP8Q	2020-01-19 15:30:00+00	2020-01-20 02:00:00+00	RN-AP-STPA-506-A	eduroam	8
ROAM	HuEwEd3efU4VeK73eJiSJIzPEmI0aQafi0paFFz9	1blIZtvb8lzuZCyN30poJxqWaCLuNxYG8O7BwdD27	2020-01-19 15:00:00+00	2020-01-20 02:00:00+00	RN-AP-CLVN-148-A	eduroam	1
LOGIN	3qwVc9Z0kpyKVUn5OHbIitSV0bbThEb37w1hO8se	gm0eQXKdsQR3YLoTz4XioI2hHXtSMTM5Z1v6HPP7	2020-01-19 16:00:00+00	2020-01-20 02:00:00+00	WN-AP-V1C-203-C	eduroam	f
ROAM	dXP3qhyvqDWH0nvKvBJRKnQah2Q4Hj0QT6UImcwD	v4ae1Jnep3fTjTk4w17TBYkR5Zzd29boAEB77CL8	2020-01-19 15:00:00+00	2020-01-20 02:00:00+00	RN-AP-REN-2709-A	eduroam	4
LOGIN	Umomu0BsNV5VyaY5cJOy2Hlw4V7SLVTTbfE90Ktc	S8OA1Tcg0bCRwaHa5NT6GPU0hRq6wXOEohBgmI7Y	2020-01-19 17:30:00+00	2020-01-20 02:00:00+00	RN-AP-STPA-255-A	eduroam	6
LOGIN	xFStbGRBwEdf6E5tNGD1sJRBpqVocyZoyIa6H4fx	E43ID8Okljh3GvsNnReWFotDe0dSxMe1aI3px9cY	2020-01-19 16:30:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-4013-A	eduroam	1
ROAM	pRWzzVMvYHHzPBK12FesTf5Zth7WDSGk0utQUR6g	vrP3GWkD4yXbOS1uKar5SWhFqncDtIN1DWJ2yPH7	2020-01-19 16:30:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-5018-A	eduroam	9
LOGIN	YpTNx8QUbzIRsJLEZIvOdICTIAINq20RNnatsNmO	kPY1ltpdf5hrNtn7b0HftCu68XEZ9ErXI3cEIcbg	2020-01-19 18:00:00+00	2020-01-20 02:00:00+00	RN-AP-CLVN-287-A	eduroam	3
LOGIN	mqMn0lAvESG66AZdfx9uP1jDJoWx1TgeTMoMSkXi	pxtj9z8evkloBWLooQt7Q1bScOM45i8rkIWgWYKA	2020-01-19 17:30:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-5509-A	eduroam	7
ROAM	lDmjIwQ0UiUwQN4CsoA3rn79uFm0Clvpa0Om1xxP	NmXpmJwCcRdfXs498AXAG3Gq5OERvyCPXbLq9eMJ	2020-01-19 19:30:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-4855-A	uw-wifi-setu	9
LOGIN	GQnBOoQMDO9fUjtkuia0mV5GKxIqaD0usx00viBh	yJP1DxP0tq0yY6obDb4upb5goWIVCHFqZHy4IuQl	2020-01-19 18:30:00+00	2020-01-20 02:00:00+00	RN-AP-STPA-512-A	eduroam	1
LOGIN	o98B0auecIJ5d4M9ZMSWc4vdr630GJeFKtsKEhiR	Ccdt5SY2oIvcZcaQ97SpJprZTr5pceahD4VXPU29	2020-01-19 20:00:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-4805-A	eduroam	a
ROAM	1XMPWA59THfR9YsqlS8V0sWnhLsoCTHPBivJBdlf	uBiL2pDdwIuz9N4pb8A6f0ezDHdUKDIs74V2iYrQ	2020-01-19 20:30:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-4506-A	eduroam	5
LOGIN	HabnpIkTt26Det3gov9J8q7LwBiGuCUT51en4A0D	ELhII1IjtgAzsEP03x97eowoWFUoIQkHJDuVH7Kh	2020-01-19 21:00:00+00	2020-01-20 02:00:00+00	WN-AP-STJRB-5016-A	eduroam	1
LOGIN	8X85u0NIE3F4ymCAunns1shLvwnzx1JkYjgVEcpx	5rKTOyogW6pOKGEUM2cCejSLReOiue0WyIJqb3xV	2020-01-10 13:00:00+00	2020-01-10 13:00:00+00	WN-AP-V1C-112-A	eduroam	6
LOGIN	bo8KWCL5LM5u2i5CXhluUbABsWJ5q2zd6RRtOxaX	2yXtSZBo5esrgMWXnJT4ekV0TfuVqdQvrwVtr5uz	2020-01-10 13:00:00+00	2020-01-10 13:00:00+00	WN-AP-E5-1908-A	eduroam	1
ROAM	bo8KWCL5LM5u2i5CXhluUbABsWJ5q2zd6RRtOxaX	LWjSSx3OuCXlPHww1tTTWAfKiD705rbb9tWnTujh	2020-01-10 13:00:00+00	2020-01-10 13:00:00+00	WN-AP-E5-1908-A	eduroam	f
LOGIN	vCq9Yu9XPQjEiQmecW5E4godi6Qk7Mkh9eiLS9zq	lORGVrN0R8SCZawKER4y7uVPMGhn3zTLAKMcXKVc	2020-01-19 21:30:00+00	2020-01-20 02:00:00+00	WN-AP-REN-201-A	eduroam	1
ROAM	K1B1HhSGHtDYCPcorHQe1NoPLcLa57XL40Gre0Kc	h4SBKIhc8jw0Zk0LALtDZi0emPXgs1M0ivVhpwQI	2020-01-19 21:30:00+00	2020-01-20 02:00:00+00	MATH-AP-M3-1--J	eduroam	b
LOGIN	lDmjIwQ0UiUwQN4CsoA3rn79uFm0Clvpa0Om1xxP	crZHg8r3L5LJGPeM5bIevBf2vsl4gif1alh4gUEU	2020-01-19 22:00:00+00	2020-01-20 02:00:00+00	AS-AP-SLC-2134-A	uw-guest	b
LOGIN	lDmjIwQ0UiUwQN4CsoA3rn79uFm0Clvpa0Om1xxP	0ywHpgiGQuyCdiy1Nsv0VUcL2g11h39pbqnAuUk8	2020-01-19 19:30:00+00	2020-01-20 02:00:00+00	WN-AP-DC-1505H-A	uw-guest	3
LOGIN	lDmjIwQ0UiUwQN4CsoA3rn79uFm0Clvpa0Om1xxP	LS5eAwwF3WUd10994dE8k8Ya6HAaUAn1E1G0w0Eo	2020-01-19 22:00:00+00	2020-01-20 02:00:00+00	WN-AP-DC-0823-A	eduroam	3

Figure 1: Sample Wi-Fi Connection Log data

Difficulties in data collection and storage, and a switch in the vendor systems, limited the useful Wi-Fi data to six weeks at the beginning of the Winter 2020 term.

Information Systems & Technology provided 27M rows of wireless logs for the first six weeks of the Winter 2020 term, from January to mid-February.

3.4 Building and Floor Layouts

Building and floor plans, including external walls, doors, rooms, stairwells, hallways, and elevators are available on the university's website, for authorized personnel. The architectural drawings display room numbers, hallways, and open areas. A section of a sample floor plan is shown in Figure 2. The image includes rooms, offices, hallways, stairwells, and washrooms. Complete drawings also display the building location relative to major thoroughfares on campus, measurement scale, and major structural elements.

The online floor plans were referenced for defining areas/zones in each building, and for mapping rooms and hallways to a specific zone.

Floor plans are published online for 67 Waterloo buildings, including Academic and Academic Support structures, residences, and places of assembly.

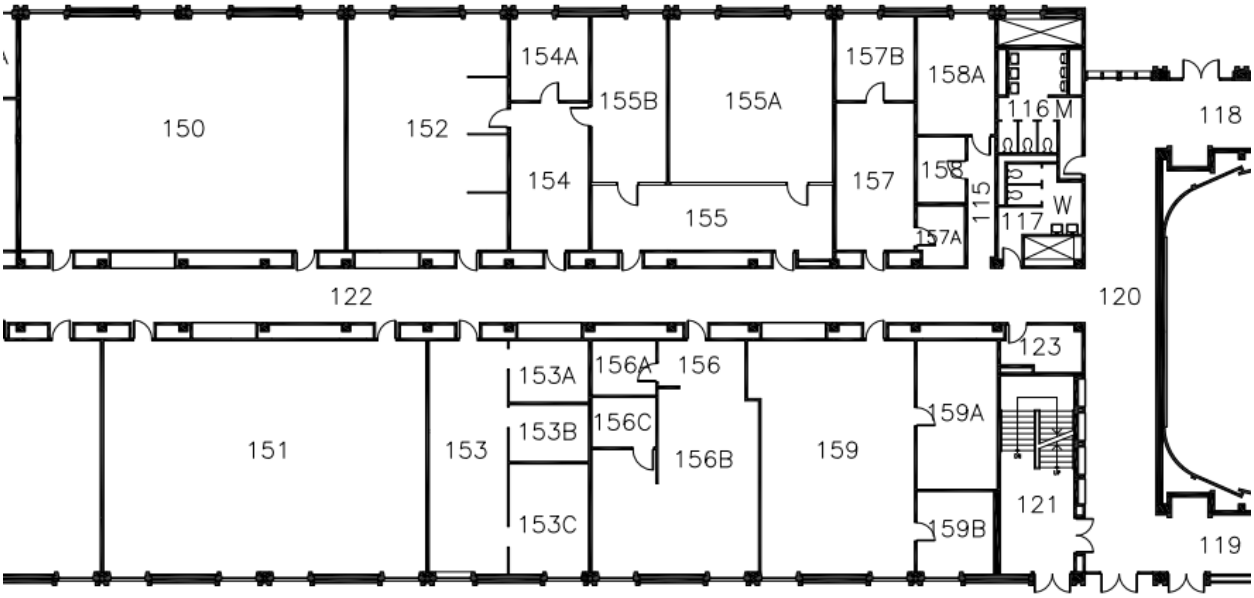

Figure 2: Sample floor layout, with room and hallway numbers

3.5 Wi-Fi Access Point Installations

Access Points' installation locations are documented as images in pdf floor layout files. The images include AP names and positions relative to rooms and hallways.

The APs are installed along the ceilings of hallways and rooms. Larger spaces have more APs installed. The precise locations are determined by IST, to provide extensive and overlapping coverage for users. Each floor of a building has numerous APs installed, with fewer devices in low-traffic and/or restricted areas.

The Access Points are named, in general, using the 3-character building code, the floor number, and the room/hallway in which it was installed. For rooms and/or hallways that have multiple APs, an additional suffix, e.g., "A", "B", "C" is appended to the AP name.

An example of a pdf is shown in Figure 3 for the 3rd floor of the Carl Pollock Hall (CPH) in the Faculty of Engineering. The drawing shows that multiple rooms can lie within the coverage area of a single AP, and that a single room may have multiple APs installed.

The APs provide overlapping coverage, and total connections to each AP are managed, for network load balancing. This means that persons in close physical proximity may be connected to different APs, also that a person may have their wireless connection automatically switched from one AP to another as they move about a building.

Figure 3: Example of AP installations in a campus building. AP Names have been redacted.

Information Systems & Technology provided 392 individual pdf files, one for each unique building and floor for buildings in which they install and maintain wireless networks.

The AP Installations pdf files were manually converted to comma-delimited files that defined the AP locations by room and hallway numbers. The new dataset listed the AP name, building, floor, and room/hallway number where the AP was installed. For APs located in hallways, atria, and other open spaces, the four closest Academic or Academic Support rooms to the AP were also listed. The four closest rooms were used in later processing to create a Centroid that more accurately described the exact location of the AP relative to rooms on the Building Floor.

Determining whether the AP was installed in a room/hallway can be difficult to determine visually due to the varied floor layouts in campus buildings. The Space Metrics dataset was used to define whether

an installed location was a “room” or a “hallway”: hallways are coded as COU Category 16.0 in the Space Metrics database (see Section 3.6).

3.6 Space Metrics for Buildings and Floors

Space Metrics include the name/number, size, capacity, ownership and purpose of each building, floor, and room in the university’s real estate portfolio. “Room” can refer both to enclosed or open areas, such as laboratories or hallways. Rooms do not necessarily have physical boundaries; a space may be subdivided if different sections are used for different purposes.

The administrative system of record for Waterloo’s facilities management is Archibus, a cloud-based application used by the Space Planning Office (SPO) and Plant Operations (Plant Ops) departments to manage the university’s owned and leased properties, and plan renovation projects.

The SPO provided a report from the Archibus system, for the Winter 2020 term, for this research. The report included over 33,000 records representing the space metrics at the institution in that period. Sample records are displayed in Table 1, for building *DWE*. The Floor Code and Room Code may include alphabetical characters. The Description of the space corresponds to the Room Type coding, documenting the purpose of the space.

Building Code	Floor Code	Room Code	Room Type	Description	Rm Standard Description	Seat Capacity	Room Area m²
DWE	2	2525	4.2	Research Office/Project Space	Project Space	4	18.85
DWE	2	2525A	4.2	Research Office/Project Space	Project Space	1	10.41
DWE	2	2526	2.1	Scheduled Class Laboratory Space	Laboratory - Wet	12	68.38
DWE	2	2526A	2.1	Scheduled Class Laboratory Space	Project Space	4	9.14
DWE	2	2526B	3.1	Research Laboratory Space	Wood/Metal/Electrical/Glass Shop:	0	26.73
DWE	2	2526C	2.1	Scheduled Class Laboratory Space	Project Space	4	11.86
DWE	2	2526D	16.2	Other Non-Assignable Areas	Corridor/Circulation Area	0	14.83
DWE	2	2527	1.2	Non-tiered Classrooms	Classroom	96	132.49
DWE	2	2528	14.1	Student Offices And Support Space	Office	6	26.65
DWE	2	2529	1.2	Non-tiered Classrooms	Classroom	90	134.66
DWE	2	2530	4.4	Departmental Administrative and Support Staff Offices	Office	0	19.36
DWE	2	2530A	4.4	Departmental Administrative and Support Staff Offices	Office	1	12.93
DWE	2	2530B	4.4	Departmental Administrative and Support Staff Offices	Office	1	16.15
DWE	2	2534	4.3	Graduate Student Offices	Student Office Facilities	16	66.64
DWE	2	2535	3.1	Research Laboratory Space	Project Space	0	92.06
DWE	2	2535A	4.3	Graduate Student Offices	Student Office Facilities	1	8.24
DWE	2	2535B	4.1	Academic Offices	Office	1	20.34
DWE	2	2535C	4.1	Academic Offices	Office	1	12.77
DWE	2	2558	16.2	Other Non-Assignable Areas	Stairs	0	13.67
DWE	2	2559	16.2	Other Non-Assignable Areas	Corridor/Circulation Area	0	175.43
DWE	2	2559A	16.2	Other Non-Assignable Areas	Corridor/Circulation Area	0	5.16
DWE	2	2559B	16.2	Other Non-Assignable Areas	Corridor/Circulation Area	0	1.45
DWE	2	2559C	16.2	Other Non-Assignable Areas	Corridor/Circulation Area	0	0.79

Table 1: Sample Space Metrics records from facilities management.

3.7 Course Schedules

Course Schedules are the listings of classes, locations, dates, times, and enrolment of all courses offered by a university. The schedules indicate course types (e.g., lecture, seminar, labs), method of delivery (e.g., online, in-person), building and room numbers.

The over 42,000 students that attend Waterloo may take any available combination of online and in-person classes, depending on their residence location, program of study and course load. A typical full-time student takes five courses (half-credits), each of which may have an associated tutorial and/or laboratory component.

Instructional courses are typically scheduled during working hours, 8:30 to 4:30, although some classes may be held during evening hours, e.g., 7:00 pm to 10:00 pm. In-person classes are usually taught in the rooms and buildings owned and managed by the relevant department, with a few exceptions. Courses with large enrolments may have multiple Sections, each with a fraction of the total Course Student Enrollment. In a typical Winter term, over 2,200 courses are taught, for thirteen weeks, to over 40,000 full-time and part-time students.

STRM	ACADEMIC TERM	CLASS	CLASS SECTION	PRIMARY COMPONENT IND	COMPONENT TYPE	ENROLLMENT	SCHEDULE DAYS	SCHEDULE START TIME	SCHEDULE END TIME	SCHEDULE START DATE	SCHEDULE END DATE	BUILDING NUMBER	BUILDING ABBREV	FLOOR	ROOM NUMBER
1201	Winter - 2020	BIOL239	104	Non Prim	Tutorial	44	R	9:30	10:20	6-Jan-20	3-Apr-20	1	DWE	3	3517
1201	Winter - 2020	BIOL239	105	Non Prim	Tutorial	46	R	9:30	10:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	106	Non Prim	Tutorial	45	R	10:30	11:20	6-Jan-20	3-Apr-20	1	DWE	3	3517
1201	Winter - 2020	BIOL239	107	Non Prim	Tutorial	47	M	12:30	1:20	6-Jan-20	3-Apr-20	1	DWE	3	3517
1201	Winter - 2020	BIOL239	108	Non Prim	Tutorial	46	W	3:30	4:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	109	Non Prim	Tutorial	43	M	12:30	1:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	110	Non Prim	Tutorial	45	M	2:30	3:20	6-Jan-20	3-Apr-20	17	MC	4	4058
1201	Winter - 2020	BIOL239	111	Non Prim	Tutorial	44	W	11:30	12:20	6-Jan-20	3-Apr-20	1	DWE	3	3522A
1201	Winter - 2020	BIOL239	112	Non Prim	Tutorial	44	T	9:30	10:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	113	Non Prim	Tutorial	40	T	9:30	10:20	6-Jan-20	3-Apr-20	1	DWE	3	3517
1201	Winter - 2020	BIOL239	114	Non Prim	Tutorial	46	R	11:30	12:20	6-Jan-20	3-Apr-20	1	DWE	3	3522A
1201	Winter - 2020	BIOL239	115	Non Prim	Tutorial	43	M	10:30	11:20	6-Jan-20	3-Apr-20	1	DWE	3	3522A
1201	Winter - 2020	BIOL239	116	Non Prim	Tutorial	45	T	1:30	2:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	117	Non Prim	Tutorial	46	F	2:30	3:20	6-Jan-20	3-Apr-20	57	EV3	3	3412
1201	Winter - 2020	BIOL239	118	Non Prim	Tutorial	45	T	2:30	3:20	6-Jan-20	3-Apr-20	1	DWE	1	1515
1201	Winter - 2020	BIOL239	119	Non Prim	Tutorial	44	M	12:30	1:20	6-Jan-20	3-Apr-20	1	DWE	3	3522A
1201	Winter - 2020	BIOL239	120	Non Prim	Tutorial	31	M	9:30	10:20	6-Jan-20	3-Apr-20	1	DWE	3	3517
1201	Winter - 2020	BIOL240	81	Primary	Online	71		NULL	NULL	6-Jan-20	3-Apr-20	999	INB	1	9999
1201	Winter - 2020	BIOL241	1	Primary	Lecture	291	TR	1:30	2:20	6-Jan-20	3-Apr-20	11	RCH	1	101
1201	Winter - 2020	BIOL241	2	Primary	Lecture	36		NULL	NULL	6-Jan-20	3-Apr-20	999	INB	1	9999
1201	Winter - 2020	BIOL241	101	Non Prim	Laboratory	29	T	2:30	5:20	6-Jan-20	3-Apr-20	999	INB	1	9999
1201	Winter - 2020	BIOL241	102	Non Prim	Laboratory	27	T	2:30	5:20	6-Jan-20	3-Apr-20	999	INB	1	9999

Table 2: Sample Class Schedule records

The class schedule records in Table 2 include the Academic Term, Class, Section, Component Type, Enrollment, Schedule Days, Start and End Dates, Start and End times, Building Number, Building Abbreviation, Floor Number, and Room number.

For the Winter 2020 term, the class schedule was represented by over 6,000 rows of data, listing both in-person and online courses. The information was provided in csv format by the Institutional Analysis and Planning department of the university.

3.8 Employee Office Locations

Employee Office Location records record the building and room where full-time permanent staff have their offices. The data is maintained by the university's Telephone Services department. Sample data included Names, Buildings, and Rooms.

Table 3 shows a subset of records from Winter 2020. 5,000 rows of data for both faculty and staff employees were provided, which included employees with multiple and/or shared office spaces. Over 4,100 office locations in 70 buildings were listed.

Name	Building/Room	Department
[redacted]	REN 3112	Renison University College
[redacted]	MC 6427	Applied Mathematics
[redacted]	HH 142	English Language & Literature
[redacted]	E2 2326	Civil and Environmental Engineering
[redacted]	EVO 1004	Waterloo Institute for Sustainable Energy
[redacted]	E2 2311	Civil and Environmental Engineering
[redacted]	MC 3050	Computing Facility (MFCF)
[redacted]	EC3 1017	Food Services
[redacted]	EIT 2034	Earth Sciences
[redacted]	E5 2109	Dean of Engineering Office-Student Design Centre
[redacted]	EC2 3122E	Info Systems and Technology
[redacted]	M3 2004	Math Business & Accounting
[redacted]	SCH 126e	Print & Retail Solutions - Waterloo Store
[redacted]	DC 3327	School of Computer Science
[redacted]	ESC 251	Science, Dean of
[redacted]	GSC 249	Plant Operations
[redacted]	E7 3422	Mechanical & Mechatronics Engineering
[redacted]	E2 3350A	Electrical & Computer Engineering
[redacted]	TCB 1F	Mathematics Faculty
[redacted]	MC 6427	Mathematics Faculty
[redacted]	SCH 014	Print & Retail Services
[redacted]	RAC 1103	Inst for Quantum Computing
[redacted]	CGR 1103	Conrad Grebel University College
[redacted]	DWE 1452	Civil and Environmental Engineering
[redacted]	DC 2621	Computer Science (CSCF)

Table 3: Sample employee office location records, with names redacted

4.0 Method

4.1 Introduction

The goal of this research project was to demonstrate a repeatable and scalable method for using wireless logs to estimate indoor populations' size and location, at a more precise level than Building Floor. Previous research work in this area was specific to a particular space (e.g., room) or type of space (e.g., office building), and so was non-repeatable. Other research provided general analyses that were transferable to other environments (e.g., Building Floor), but not precise in terms of locating indoor crowds.

The six datasets used in this research project are representative of the types of data available at many university campuses. The datasets are large (e.g., wireless records may be too big for conventional database systems), varied (e.g., structured and unstructured), have omissions and errors (i.e., "dirty data"), and contain implicit information that can be extracted using Big Data or Data Science techniques.

The methodology involved standardizing and normalizing the data, then combining the information to gauge insights not possible by analyzing any single dataset. The data processing was done using Python (Anaconda/Jupyter Notebook) running on a Linux server.

The following sections provide information on combining the varied datasets, the campus environment and the data processing.

4.2 Differences in Departmental Datasets

In this study, Wi-Fi data from the University of Waterloo, a medium-sized university in Ontario, Canada, with 42,000 students, was used to estimate the number of students present in Building Floor Zones during specific periods. The estimates were compared with the estimated number of occupants (employees and students) for each Building Floor Zone.

Waterloo's system administrators made available for investigation and analysis datasets from Wi-Fi logs, Building and Floor Layouts, Access Point Installations, Space Metrics for Buildings and Floors, Course Schedules, and Employee Office Locations. The datasets were cleaned and standardized, then combined and analysed for insights.

The six distinct datasets are maintained by different campus departments and/or units, for varying business purposes. As a result, combining the datasets was challenging. For example, staff may have multiple offices and phone records, and some offices have multiple employees, resulting in duplicate records in the Employee Office Locations. The Building and Floor Layouts define spaces based on structural elements, but the Space Metrics dataset defined spaces based on financial costing/chargeback rules, as a result there can be multiple spaces defined in the Space Metrics dataset for any defined space in the architectural diagrams.

The datasets are not usually combined for analysis, so merging the records was a challenging task. Some codes for some buildings, floors, and rooms were recorded differently in each administrative system of record. Also, some datasets had defined areas that were unmapped/unknown in the other administrative systems, such as basement and rooftop maintenance areas inaccessible to the general student and employee population. In addition, departmental updates to the administrative systems of record were not always synchronized in a timely fashion.

The datasets' contents were standardized to reflect consistent business definitions, and to prevent duplicate records that would skew the analysis. Master records were identified in cases where records were inconsistent, and their values used in the analysis.

The datasets were combined spatially after defining zones for each Building Floor. The Class Schedules and Employee Locations included room numbers, but the datasets for Building and Floor Layouts, AP Installations, and Space Metrics included spaces that were defined by rooms and open areas like hallways, foyers, and atria. Building Floor Zones were defined using the first two/three characters of the room/hallway/space name listed in the individual datasets.

The datasets were combined temporally at a level of detail that aligned with the highest-level time periods in the input files. The Wi-Fi logs' Start Times and End Times were rounded to 00:00 or 00:30 (hours and minutes) prior to being shared, so all time period calculations were modified to use half-hour periods.

4.3 University of Waterloo Main Campus

The University of Waterloo is a medium-sized research-intensive institution located in south-western Ontario, Canada. The university a main campus in the city of Waterloo, and three satellite campuses in the cities of Cambridge, Kitchener, and Stratford. The main campus dwarfs the other three by any measure of size, and is the hub of almost all teaching, learning and research activities. The main campus is also co-located with four affiliated university colleges: United College, Conrad Grebel, St. Jerome's and Renison, with which it shares resources and students. Waterloo also accommodates cross-registered students from Wilfrid Laurier University, located less than a mile away.

As with other tertiary institutions, the Waterloo main campus can be described as a small city. With over 42,000 students, including over 6,000 graduate students, Waterloo is one of the ten largest higher-ed institutions in Canada, for student population. Also, cross-registered students from the university colleges, and nearby Wilfrid Laurier University, significantly boost the number of students on the main campus. The university is open year-round, with over 29,000 students attending classes on campus from May to August, the spring/summer term.

Waterloo's main campus includes 92 buildings, occupied and managed by the six faculties and Academic Support units. Many of the buildings are located within a circular Ring Road (see Figure 4), within which vehicular traffic is limited. To the north, east and south of the campus are major public thoroughfares, which serve as gateways for students and employees to enter the campus. The eastern boundary includes a stop on the Region of Waterloo's light rail transit system, as well as a bus terminus.

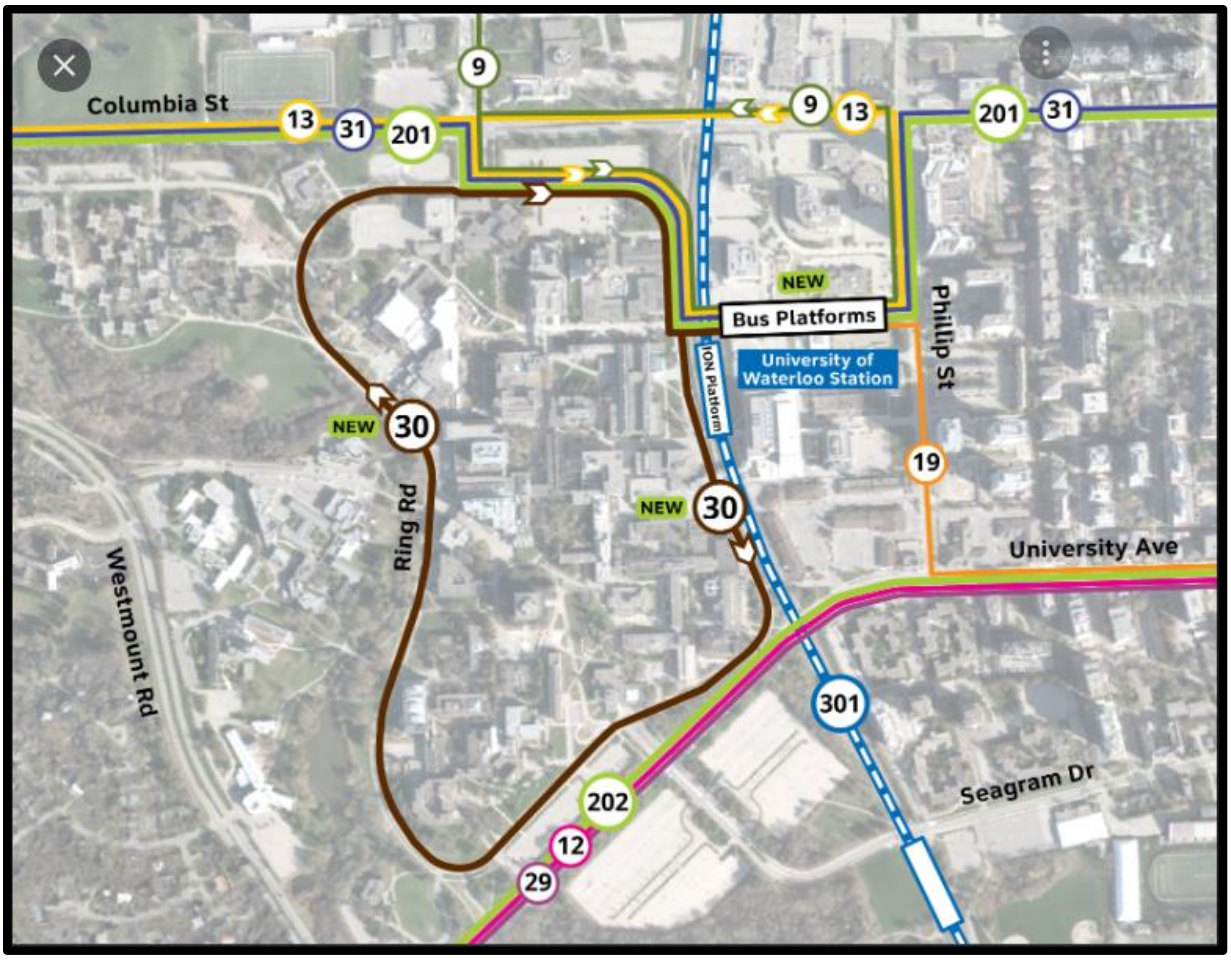

Figure 4: University of Waterloo's Main Campus, and public transit routes, courtesy GRT.

The buildings within the Ring Road are primarily Academic and Academic Support spaces. There are also spaces dedicated to social, health and wellness, athletic, and other student support activities. Additional academic and/or administrative buildings are situated outside of Ring Road, particularly on the eastern and northeastern side of the main campus. To the west and southeast of Ring Road are the university colleges and student residences.

University buildings are often referred to using two- and three-letter codes, which are typically abbreviations of the building names (see Figure 5). Many of the buildings are connected via tunnels and/or overpasses, facilitating pedestrian traffic.

Large numbers of pedestrians (students and employees) move about the Ring Road-enclosed buildings daily. Within Ring Road, vehicular traffic is limited and there is a high degree of pedestrian activity. Visitor, employee, and student pedestrian traffic enter the main campus from transit, or from student residences to the west and east of Ring Road (see Figure 4).

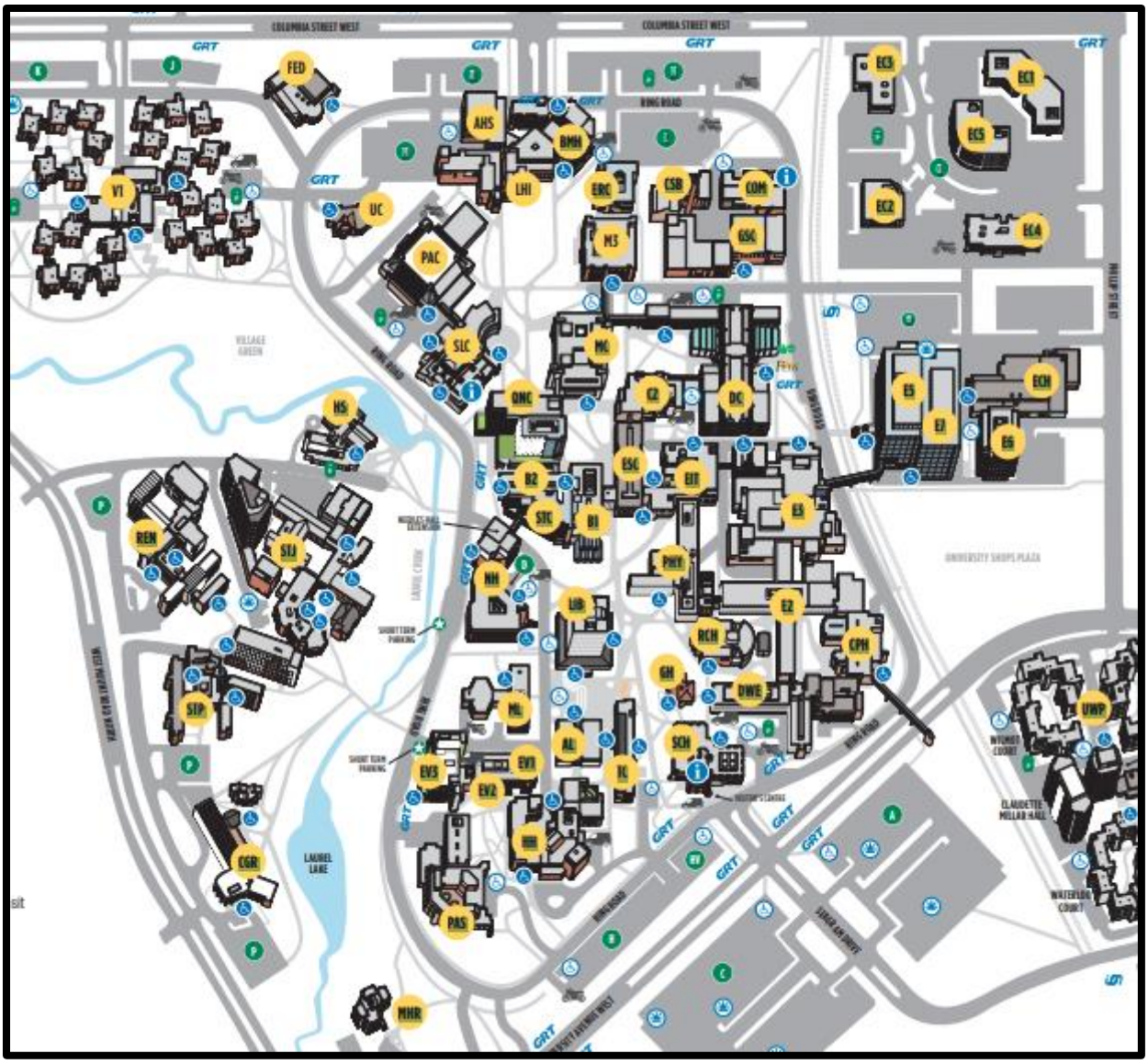

Figure 5: Building Layout within Ring Road, University of Waterloo, courtesy University of Waterloo.

Campus pedestrian activity is high even outside of working hours, because of large student residences that are located on-campus, or very close to campus: Waterloo boasts one of the highest numbers of on-campus student accommodation in the country, at over 8,000 beds. This makes the main campus a hub of activity during both operating and non-operating hours, as students use the campus buildings for both studying and socializing.

4.4 Data Processing

The six data sources were combined to create a superset for analysis. The data sources were from five different systems of record. The flowchart in Figure 5 shows how the six datasets were scrubbed and combined. The following sections describe the individual processing tasks shown in Figure 6.

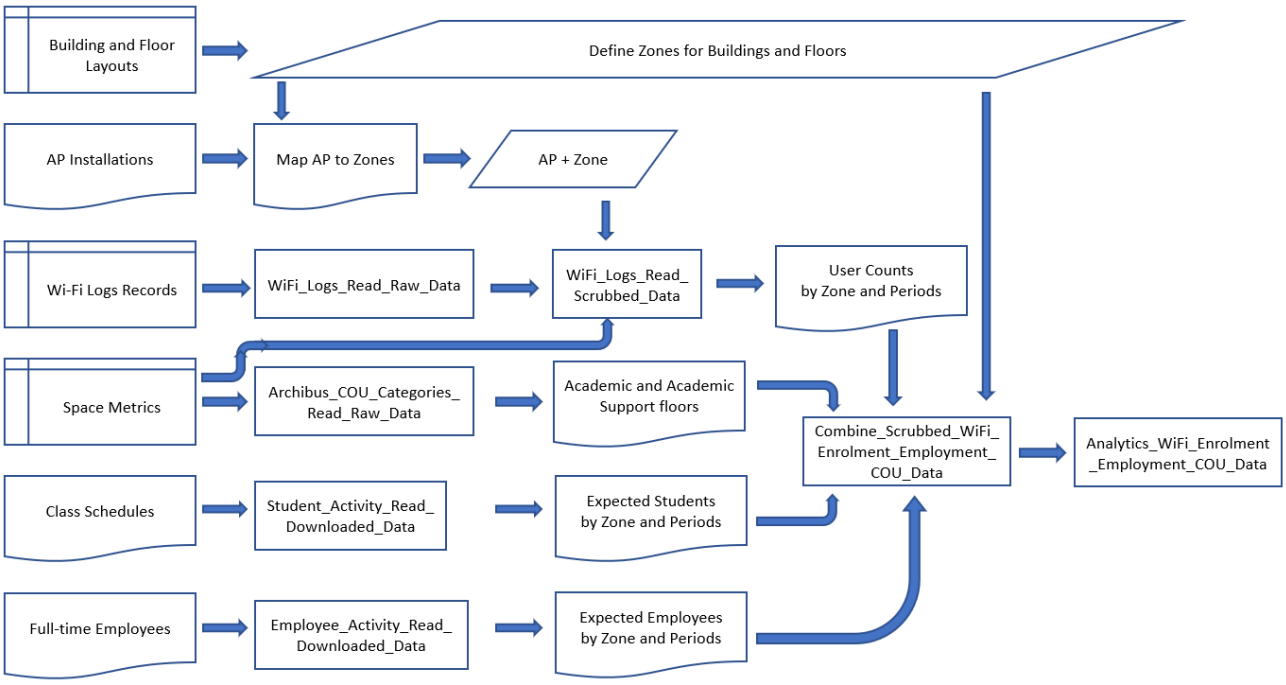

Figure 6: Data Processing flowchart,

4.4.1 Define Zones for Building and Floors

Estimating the number of persons present in an indoor space by analyzing the number of wireless connections is usually done for a specific room, or for an entire Building Floor. Varying building layouts, configuration of APs, and differing occupancy plans, make it difficult to accurately measure wireless users when building floor environments differ. Dedicated equipment in small spaces than can facilitate triangulation or signal-strength fingerprinting, or generalized estimates for larger spaces are the typical methods for calculating indoor crowd sizes and/or locations.

Wireless networks in interior spaces are architected for seamless integration with the wired network, overlapping coverage areas, minimal channel interference, and expected user loads. The locations, types, and count of installed APs vary with building type, floor layout and room usage. Multiple APs may be installed in one room, or one AP may provide network access to many rooms.

The closest AP to a wireless device is not necessarily the one with which a portable device forms a connection. Effective Wi-Fi range is affected by signal attenuation by walls, room objects, AP frequency (2.4 GHz or 5 GHz) and device power levels. Also, load balancers may distribute the network load across

multiple APs. An open floor layout may result in a direct “line of sight’ connection between a portable device and an AP that is a significant distance away.

Identifying the precise indoor location of a wireless user is challenging. Previous researchers have attempted to address the issue by fingerprinting (creating a calibrated map of device locations vs. signal strengths and/or established wireless connections), geospatial algorithms, or limiting analysis to a single room or small area. Another approach has been providing only macro-level analyses, such as the number of persons on a floor, or in a building.

A zone is a room, or collection of rooms, within a Building Floor. The zone represents the effective coverage area of an AP, as defined by the location(s) of the persons who have a wireless connection to that AP. Even though an AP may form a connection with a wireless device some distance away, in general connections are formed between APs and devices in the vicinity: that is, on average a wireless device will form a connection with an AP in the same space, or in an adjacent space, rather than an AP that’s on the far side of the same building.

In this project, a more granular spatial analysis of Wi-Fi signals was attempted by defining zones within floors. The zones were areas served by one/more APs. Where multiple APs were installed on a floor, the zones represented the rooms closest to each AP installation. Since APs are installed to provide network access, they are often positioned within or close to areas of high usage, such as classrooms and labs.

Zones were defined by combining the building abbreviation (e.g., *AL*) with the floor number (e.g., *02*) and the first two/three characters of the room codes/numbers (e.g., *234*, *236*, *250*). If the room code was three digits long, the first two digits were used; If the room code was four digits long, the first three characters were used. Zone definitions provide a greater location precision than Building Floor, without the challenges of identifying specific rooms, and without the need for additional installed equipment.

In most buildings, adjacent rooms have similar room codes/numbers, while rooms that are at opposite ends of a floor usually have room codes/numbers that differ significantly. Also, room codes/numbers on different floors usually begin with unique numbers- typically, that first digit is also the floor number. For example, rooms on the second floor of a building start with *2*, and rooms *2346A* and *2348* are more likely to be closer together than *2366* and *2500A*. At Waterloo room codes/numbers begin with the floor number. Also, room codes are three/four digits, sometimes with an alphabetical suffix (e.g., *AL 210*, *E2 2345A*).

Defining Building Floor Zones for the Class Schedules and Employee Locations datasets was straightforward, since the information in the datasets included a specific room number. All scheduled activities for students and employees take place in classrooms and offices, so the room number was adequate for defining Building Floor Zone.

4.4.2 Map AP to Zones

Defining Building Floor Zones for APs was challenging, since many APs are installed in hallways and/or open areas. The naming convention for hallway and/or open areas at the university differs significantly

from the naming convention for rooms; adjacent rooms and hallways do not share the first two/three characters. In addition, a hallway may extend a significant distance, providing access to dozens of rooms; multiple APs may be installed along the length of the hallway, each providing wireless access to a unique set of rooms.

At Waterloo the naming convention for APs includes the building and room/hallway location. For example, *WN-AP-DWE-2513-A* and *WN-AP-DWE-2513-B* are both located in hallway 2513 in the *DWE* building. In situations where the AP name did not follow the standard naming convention, the pdf drawings were used to locate it in a particular area, defined by a Centroid. If neither the AP's name, nor the pdf drawings, permitted identification, it was excluded from the processing.

The Centroid represented the room(s) in which the APs were installed, or the rooms immediately adjacent to the hallway in which the AP was installed. Figure 7 shows the locations and names of APs installed in specific rooms of a Building Floor. The AP names correspond to the room names: *3346-a*, *3347-a*, *3341-a*, etc. In Figure 7, the AP Centroids were the first two/three digits of the room numbers, e.g., *3346-a=>334*, *3347-a=>334*, *3339-a=>333*, etc.

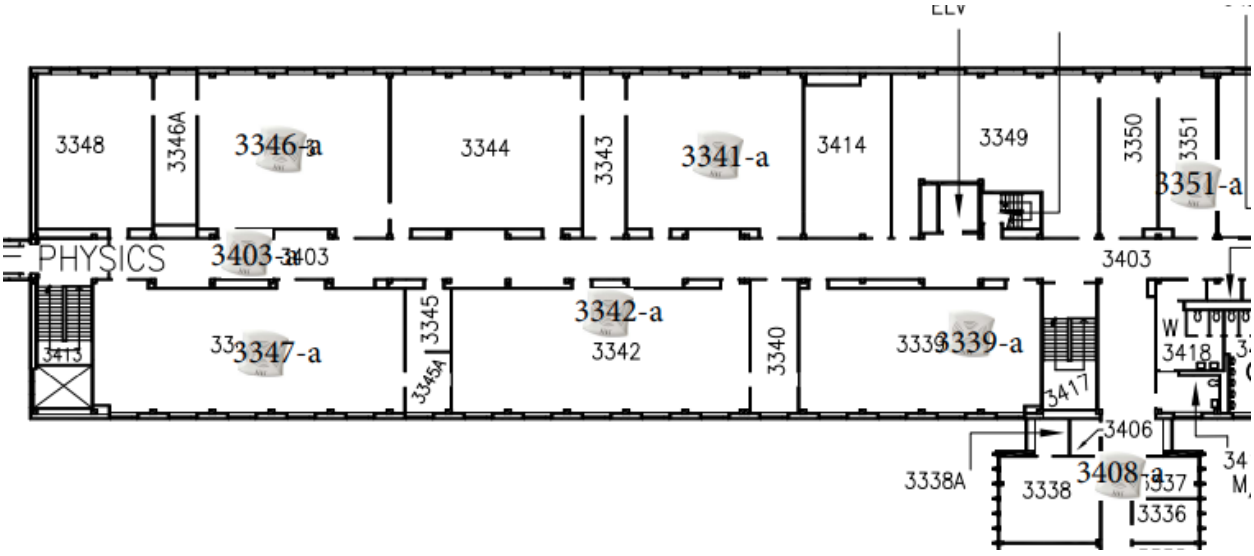

Figure 7: Examples of AP installation locations in rooms.

Hallways may extend the length of a Building Floor, connecting rooms that are far apart. Multiple APs installed in a corridor are assigned the same three/four digits, with unique suffixes. Figure 8 shows a different section of the 3rd floor of the same campus building shown in Figure 7, with five APs installed in a long corridor, *3408*. The APs are named *3408-a* through *3408-e*.

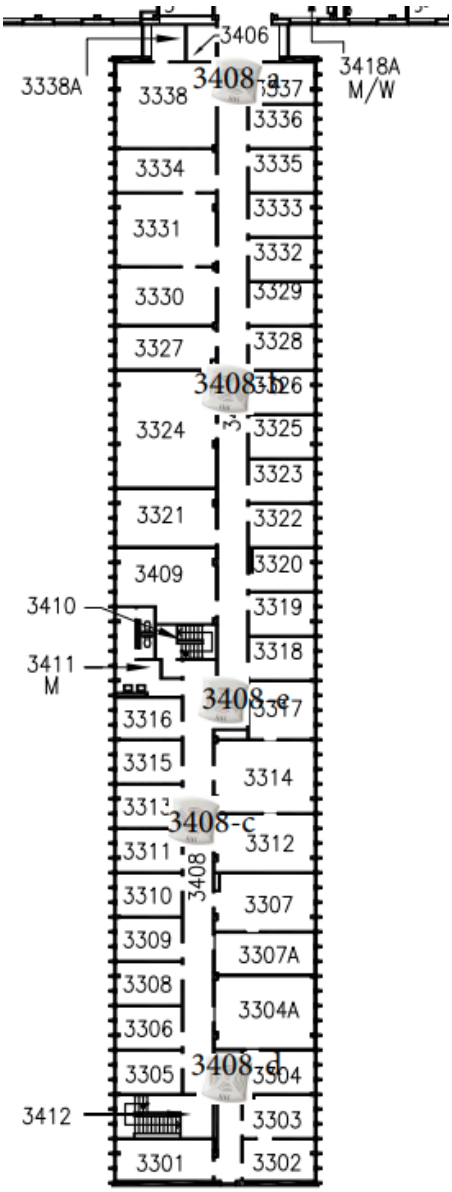

Figure 8: Examples of AP installations locations in hallways.

For APs in hallways, the Centroid was defined as the geographical "centre" of the four closest rooms to the AP. Centroids for APs in hallways were calculated by finding the arithmetic average of the four closest Academic or Academic Support rooms, after removing any suffixes from the room codes/names. The first two/three characters of the arithmetic average, after decimal places of the average were truncated, was defined to be the Centroid.

At Waterloo, special purpose rooms such electrical equipment rooms, elevators, and custodial rooms, do not use the same numbering scheme as Academic and Academic Support rooms. That is, special purpose rooms have significantly different room codes, even when they are adjacent to Academic (Support) rooms. The four rooms selected for determining the Centroid of a hallway AP must therefore all be Academic or

Academic Support spaces, since this is the only way that a Centroid value can be calculated that accurately represents the area of the Building Floor where the AP is being used for wireless connections.

In Figure 8, the zone served by each AP is assumed to be the area described by the average room codes of the closest Academic (Support) rooms. For example, AP *3408-c*, at the bottom of the image, is assigned a Centroid value "330", since room codes *3301*, *3303*, *3305*, and *3306* have an arithmetic average of *3303.75;* the truncated arithmetic average of *3303.75* is *3303*; and the first three characters are *330*. In a similar manner, the APs in the image above are assigned Centroid values of {*3408-a => 333*; *3408-b=> 332*; *3408-c => 331*; *3408-d => 330*; *3408-e => 331*}.

By assigning each AP to a centroid that represents by the cluster of rooms closest to it, it is assumed that wireless connections made to that AP are from devices in nearby rooms.

The data for used for determining each AP's zone was manually compiled by comparing and contrasting:

1. Access Point installations (pdfs)
2. Building and floor layouts (online files)
3. Space metrics for buildings and floors

The pdfs were used to identify the precise locations in rooms and hallways in which the APs were installed. The layout files were used to identify all nearby spaces (rooms and hallways). The Space Metrics dataset was used to select/filter only nearby spaces that had an Academic or Academic Support purpose; the purpose was determined by using the Room Type Descriptions and Room Standard Descriptions, as shown in Table 1.

The pdfs were occasionally out of date, compared to the floor layouts, which are updated regularly in the university's Space Metrics' system of record, during renovation and construction projects. In cases where there were discrepancies, the Wi-Fi logs, and floor layouts, were treated as correct.

	A	B	D	E	F	G	H	I	J	K
1	Building	Filename	Floor	AP_Name	Room_or_Hallway	Label	Triangulate1_Room	Triangulate2_Room	Triangulate3_Room	Triangulate4_Room
328	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2367-A	ROOM	2367				
329	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2385-A	HALLWAY	2385	2383	2385A	2385B	2384
330	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2386A-A	HALLWAY	2368A	2387	2398	2389	2384
331	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2395-A	HALLWAY	2395	2396A	2396B	2396C	2396D
332	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2387-A	ROOM	2387				
333	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2398-A	ROOM	2398				
334	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2368A-A	HALLWAY	2368A	2369A	2377	2367	2382A
335	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2369-B	HALLWAY	2369	2369D	2370	2369C	2369E
336	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2376J-A	ROOM	2376J				
337	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2368A-B	HALLWAY	2368A	2374A	2374B	2376G	2375
338	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2369-A	HALLWAY	2369	2373F	2373G	2373E	2373D
339	CPH	029CPH_02FLR-withAP.pdf	02	WN-AP-CPH-2376-A	HALLWAY	2376	2376B	2376C	2376D	2376E
2503										
2504										

Table 4: Sample records for precise locating of APs.

Over 2,500 records were manually created from the AP Installations pdfs to allow more precise locating of APs. Table 4 shows sample location records compiled for APs on the second floor of a university building. APs installed in hallways, as defined by the Space Metrics dataset, have four additional column values, representing the four closest Academic (Support) rooms to the AP location.

Figures 9, 10, and Table 5 below show respectively: the floor layout of the 2nd floor of a typical building; the AP installation locations in rooms and hallway; the triangulation of the hallway AP's location; and the calculation of each AP's zone (based on building name & AP_Centroid).

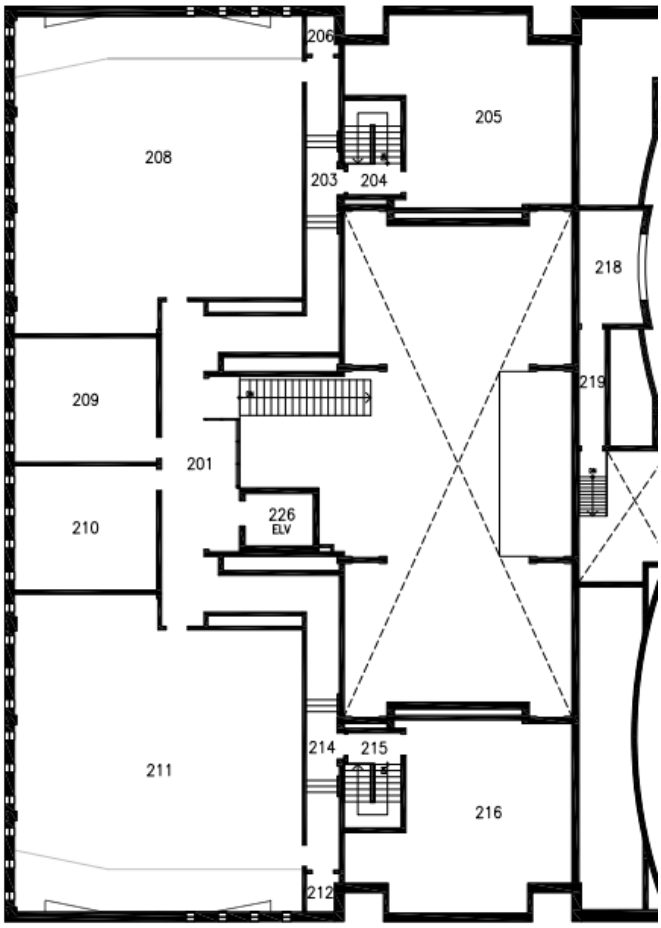

Figure 9: Sample Building Floor layout.

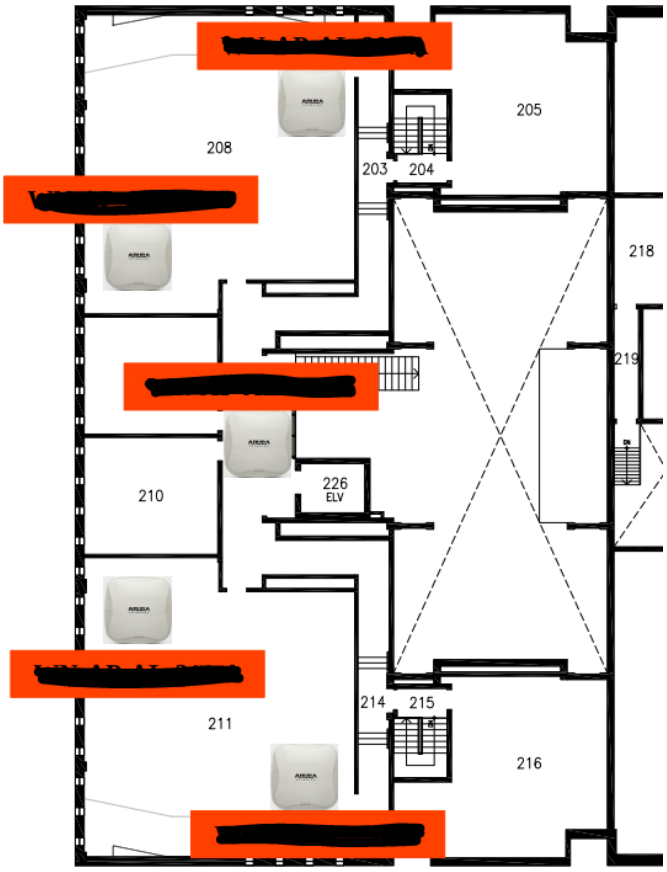

Figure 10: Sample Building Floor layout, with AP installations

ap	Room_or_Hallway	Building_Room	Triangulate1_Room	Triangulate2_Room	Triangulate3_Room	Triangulate4_Room	Triangulate_Avg	Corridor_Circulation	AP_Centroid
WN-AP-AL-208-A	ROOM	208	NaN	NaN	NaN	NaN	208	N	20
WN-AP-AL-208-B	ROOM	208	NaN	NaN	NaN	NaN	208	N	20
WN-AP-AL-201-A	HALLWAY	201	208	209	210	211	209	Y	20
WN-AP-AL-211-A	ROOM	211	NaN	NaN	NaN	NaN	211	N	21
WN-AP-AL-211-B	ROOM	211	NaN	NaN	NaN	NaN	211	N	21

Table 5: Sample location records for APs in Figure 9.

4.4.3 Wi-Fi Logs Read Raw Data

The campus wireless network is used by students, employees, and guests. Waterloo is part of the eduroam network, so visiting students and staff from other institutions can also seamlessly connect.

Wi-Fi connection logs were stored in a database maintained by the Information Systems and Technology (IST) department during the Winter term. The connection logs stored User Ids, MAC addresses, AP Names, Start Times, End Times, and other particulars of devices' successful connections to the campus wireless network (see Fig. 1). Each record represented a single wireless connection between a specific portable device and a specific AP. Start Times and End Times were rounded to the nearest half-hour.

The data provided was related to the first six weeks of the winter term (January to April) of 2020. In the six-week period 27,000,000 log records were collected from 5,373 Access Points across the campus. Table 6 shows a summary of the log records provided, from December 31, 2019 to February 7, 2020. The number of unique users is 1,046,332 because of the one-way hashing of User Ids each 24-hour period.

	action	ap	end_time	id	mac	ssid	start_time	user
count	27000000	27000000	27000000	27000000	27000000	26999470	27000000	27000000
unique	2	5373	1803	1000000	1643853	11	1767	1046332
top	LOGIN	WN-AP-LOT-D-B-OUTDOORTEST	2020-01-14 13:00:00-05:00	0	x2sdcOC0MVdUNX9ZiqKNKmWlEpcuUADo1sxn9nze	eduroam	2020-01-14 13:00:00-05:00	e5CP0g86wzm5JN9YEtG5NhT2WKMUD5u2zImtKHTp
freq	14497139	476461	60805	27	1012	26137579	65034	100816
first	NaN	NaN	2019-12-31 22:00:00-05:00	NaN	NaN	NaN	2019-12-31 22:00:00-05:00	NaN
last	NaN	NaN	2020-02-07 11:00:00-05:00	NaN	NaN	NaN	2020-02-06 17:00:00-05:00	NaN

Table 6: Summary of wireless logs provided for analysis

Figure 11 shows the number of unique Wi-Fi connections per day for the first 37 days of 2020. The number of successful connections on working days averaged close to 1,000,000.

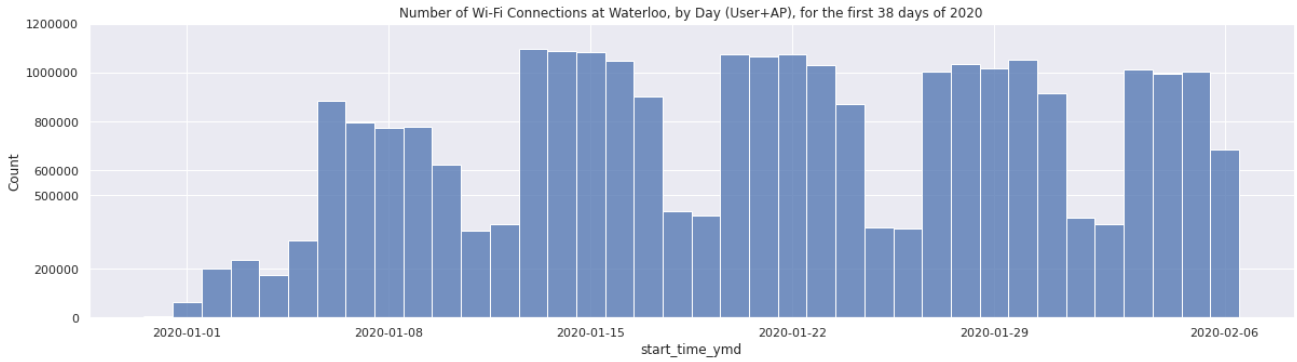

Figure 11: Count of unique Wi-Fi connections for the first 37 days of 2020

Figure 12 displays the number of unique User Ids, by day, for the first 37 days of 2020. The User Ids were scrambled daily to prevent tracking across days, so the total number of unique User Ids could not be determined, but the largest daily number of unique User Ids was 39,864 on January 14. The lowest number of unique wireless users occurred in the first week of the year, January 1 2020 (Wednesday) to January 5 2020 (Sunday) before classes started on Monday, January 6, 2020. Starting on January 6 2020, the number of unique daily Users of the wireless network averaged between 33,000 and 40,000 on

weekdays, with the lowest numbers on Fridays. Weekend records averaged between 13,000 and 15,000 unique User Ids.

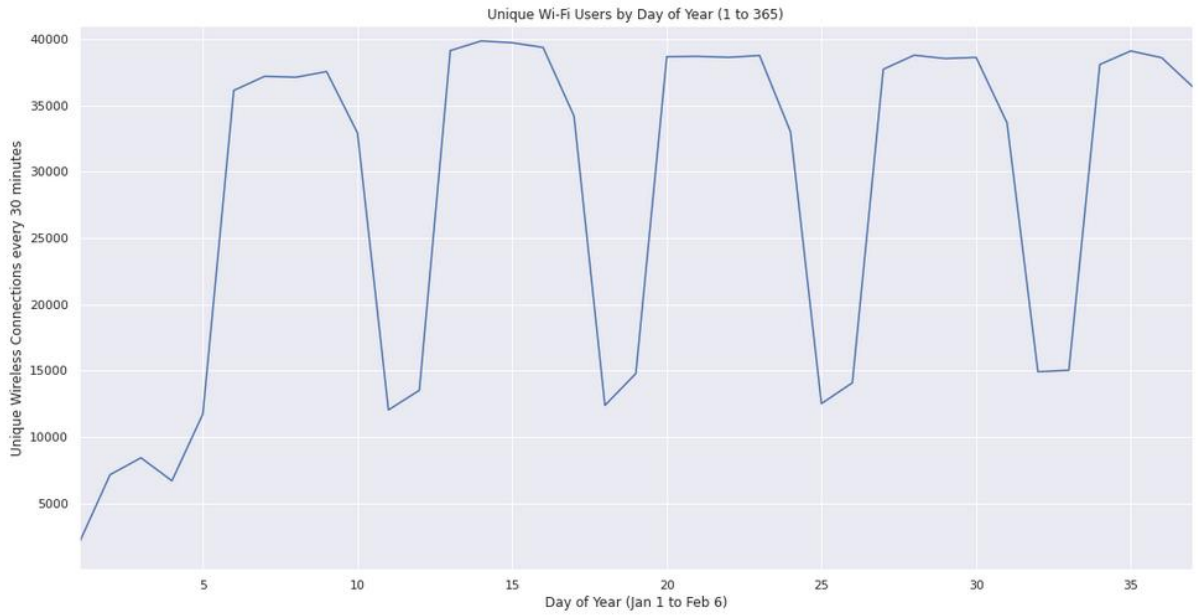

Figure 12: Count of unique Wi-Fi users for the first 37 days of 2020

In the Winter 2020 term over 31,300 graduate and undergraduate students were registered for classes; a further 468 students were registered as working on campus. Faculty and staff (full-time permanent employees) totaled 4,250. Contract and part-time research appointments totaled approximately 550. Not including casual staff, the total number of students and employees was 36,568.

The higher number of unique User Ids daily, relative to the expected number of 36,568 may have been due, in part, to casual employees , plus visiting students and employees from other universities- the University of Waterloo offers joint registrations and degrees with Wilfrid Laurier University, located less than a mile from the main campus.

The counts of unique Users at each AP, for every 30-minute period of the first six weeks of 2020, were analyzed. The number of unique Users at each AP, for each 30-minute period was assumed to be equivalent to the number of persons in the vicinity of the AP's location.

The dataset of 27M rows was first analysed for unusual/outlier values, which were removed. Outlier values in the wireless logs included users that had connected to more than 1,000 APs in a single 24-hour period. The scrubbed dataset included 24,284,118 rows after the outliers were removed.

The wireless logs dataset contained inadvertent duplicates, i.e., some devices were recorded as being connected to more than one AP, for a specific 30-minute period. This was caused by the Wi-Fi logs' Start Time and End Time being rounded to the closest half-hour. A device can only be successfully connected to a single AP at a time, however, as an individual moved around campus buildings and floors, the rounding of Start Times and End Times to the nearest half-hour meant that the Start Time and End Time for

connections to distinct APs would be identical if consecutive connections with multiple APs differed by less than 15 minutes.

The duplicate records that suggested that a single wireless device had been connected to multiple APs during a half-hour period were removed. The wireless log record corresponding to the maximum AP value for each user, for each 30-minute period, was selected as the record for analysis. Once duplicate records were removed, the wireless logs included 15,290,078 rows of data. Sample records are shown in Table 7:

	user	mac	start_time_pd	end_time_pd	ap
0	0000JSal7kUId0lyqYmyd9vbsSkboMplNRKGrxQs	oaAZg4a2s7WHblWS4O4YxJ3594kqsXo76lrZKeLD	2020-01-13 10:00:00-05:00	2020-01-13 10:30:00-05:00	ENG-AP-E6-3--A
1	0000JSal7kUId0lyqYmyd9vbsSkboMplNRKGrxQs	oaAZg4a2s7WHblWS4O4YxJ3594kqsXo76lrZKeLD	2020-01-13 12:00:00-05:00	2020-01-13 12:00:00-05:00	ENG-AP-E6-3--A
2	0000JSal7kUId0lyqYmyd9vbsSkboMplNRKGrxQs	oaAZg4a2s7WHblWS4O4YxJ3594kqsXo76lrZKeLD	2020-01-13 13:00:00-05:00	2020-01-13 13:30:00-05:00	ENG-AP-E6-3--A
3	0000JSal7kUId0lyqYmyd9vbsSkboMplNRKGrxQs	oaAZg4a2s7WHblWS4O4YxJ3594kqsXo76lrZKeLD	2020-01-13 13:30:00-05:00	2020-01-13 14:00:00-05:00	ENG-AP-E6-3--A
4	0000JSal7kUId0lyqYmyd9vbsSkboMplNRKGrxQs	oaAZg4a2s7WHblWS4O4YxJ3594kqsXo76lrZKeLD	2020-01-13 14:30:00-05:00	2020-01-13 15:00:00-05:00	ENG-AP-E6-3--A
...	...	...	...	...	...
15029073	zzzxtjYc0irpRsZQM4W7kkSK8aqRJC00FR5Xxz3I	YEcN0zzi3T20AyeuVG7000QNOT002YXJFxrgOwF1	2020-02-01 20:00:00-05:00	2020-02-01 20:00:00-05:00	RN-AP-NRB-4002-A
15029074	zzzxtjYc0irpRsZQM4W7kkSK8aqRJC00FR5Xxz3I	YEcN0zzi3T20AyeuVG7000QNOT002YXJFxrgOwF1	2020-02-01 20:00:00-05:00	2020-02-01 23:00:00-05:00	RN-AP-NRB-5202-A
15029075	zzzxtjYc0irpRsZQM4W7kkSK8aqRJC00FR5Xxz3I	YEcN0zzi3T20AyeuVG7000QNOT002YXJFxrgOwF1	2020-02-01 23:00:00-05:00	2020-02-01 23:00:00-05:00	RN-AP-NRB-5001-A
15029076	zzzxtjYc0irpRsZQM4W7kkSK8aqRJC00FR5Xxz3I	YEcN0zzi3T20AyeuVG7000QNOT002YXJFxrgOwF1	2020-02-01 23:00:00-05:00	2020-02-01 23:30:00-05:00	RN-AP-NRB-5206-A
15029077	zzzxtjYc0irpRsZQM4W7kkSK8aqRJC00FR5Xxz3I	YEcN0zzi3T20AyeuVG7000QNOT002YXJFxrgOwF1	2020-02-01 23:30:00-05:00	2020-02-02 00:00:00-05:00	RN-AP-NRB-5001-A

15029078 rows × 5 columns

Table 7: Sample wireless logs, after removal of duplicate user connections.

The Duration of each connection (End Time – Start Time) was calculated for each record. More than half the 15M records had Duration = 0, indicating that the Start Times and End Times of the wireless connection were identical. These “0 minute” Durations represented wireless connections that lasted a short time (i.e., from 0 to 30 minutes), but had had their Start Time and End Time rounded to the nearest half hour (e.g., 11:46 am to 12:14 pm would have been rounded to 12:00 pm to 12:00 pm).

To facilitate analyses of wireless connections that lasted for longer periods than 30 minutes, two additional datasets were then created by filtering the 15M rows to include records with Duration >= 30 minutes, and records with Duration >= 60 minutes. The assumption was wireless connections lasting a short time represented students and employees in motion, and that wireless connections lasting for longer periods represented students and employees taking part in a scheduled activity, e.g., students in a class or lab. Of the 15M rows, 7.06M had durations greater than 30 minutes, and 2.87M had durations greater than 60 minutes.

Limited conclusions could be drawn from analyzing differences between the three datasets. The rounding of Start Time and End Time meant that durations of 30 minutes included connections that lasted anywhere from 15 minutes to 60 minutes (e.g., 11:30 am to 11:45 am would have been rounded to 11:30 am to 12:00 pm; 11:45 am to 12:44 pm would have been rounded to 12:00 pm to 12:30 pm). Also, durations of 60 minutes could have represented connections as short as 30 minutes (e.g., 11:44 to 12:16 would have been rounded to 11:30 to 12:30).

The wireless logs data was transformed to calculate the number of unique Users connected to each AP, for each 30-minute period of the six weeks. The steps were as follows:

1. Since a device can only be connected to one AP at a time, calculate the maximum AP value for each set {User, MAC, Start Time, End Time} to create a new set {User, MAC, Start Time, End Time, AP}. This removed possible duplicate counts due to persons roaming between APs in the same 30-minute period.
2. From the new set {User, MAC, Start Time, End Time, AP} drop the MAC data column to create a new set {User, Start Time, End Time, AP} which may include duplicate values for specific combinations of Start Time, End Time, and AP for each User.
3. Select only unique values of {User, Start Time, End Time, AP}. This prevented the double-counting of Users who had multiple Wi-Fi-connected devices.
4. From the unique value set of {User, Start Time, End Time, AP}, create a list of unique APs.
5. For each AP, calculate the daily minimum Start Time and the daily maximum End Time. These times represented the earliest and latest times that the AP recorded a connection log.
6. For each AP, loop through the Minimum_Start_Time and Maximum_End_Time, in 30-minute increments.
7. For each 30-minute increment, check all unique connection logs {User, Start Time, End Time, AP} against the AP, Minimum_Start_Time, and Maximum_End_Time.
8. If the APs match, and the Start Times and End Times overlap, increment Connected_Users by 1
9. After checking all unique connection records, output the total connections for that ap and 30-minute period.
10. Increment the 30-min. bucket to the following period.
11. Return to Step 7.
12. After checking all 30-min. periods, from the Minimum_Start_Time to the Maximum_End_Time, retrieve the next AP value.
13. Return to Step 5.

In this manner, the total number of connections, for each AP, for each 30-minute period between January 1st and February 6th was calculated. In the end, the 27 million initial records yielded 15 million rows without duplicates, then 5.95 million rows of unique User counts by AP and 30-minute periods.

Note that, in calculating the maximum AP value in Step 1, we did not specify a particular type of wireless device. That is, it would have been more accurate to associate an individual who had multiple devices with their tablet or cellphone, rather than their laptop, since persons tend to always carry their cellphones on their persons. The algorithm did not attempt to identify the device for each User that had the widest/largest roaming profile each day. Even so, the algorithm only counted unique connections per user.

For more in-depth analysis, four counts were determined for each 30-minute period: number of devices (Devices), number of connected users (WiFi_Users), number of users connected for at least 30 minutes (Users30), and number of users connected for at least 60 minutes (Users60). Users30 and Users60 values are representative of Wi-Fi connections that lasted for a significant length of time, for example persons who attended a class. In Figure 10, the first record has values of 48, 46, 36, 16 for these Duration, Users, Users30, and Users60, respectively, for connections to AP *AS-AP-TC-1204-A* between 9:30 and 10;00 am on January 28, 2020.

Wi-Fi connection logs include both daytime and nighttime activity in buildings. A Boolean value for Working Hours was calculated as time spanning 9:00 am to 5:00 pm inclusive.

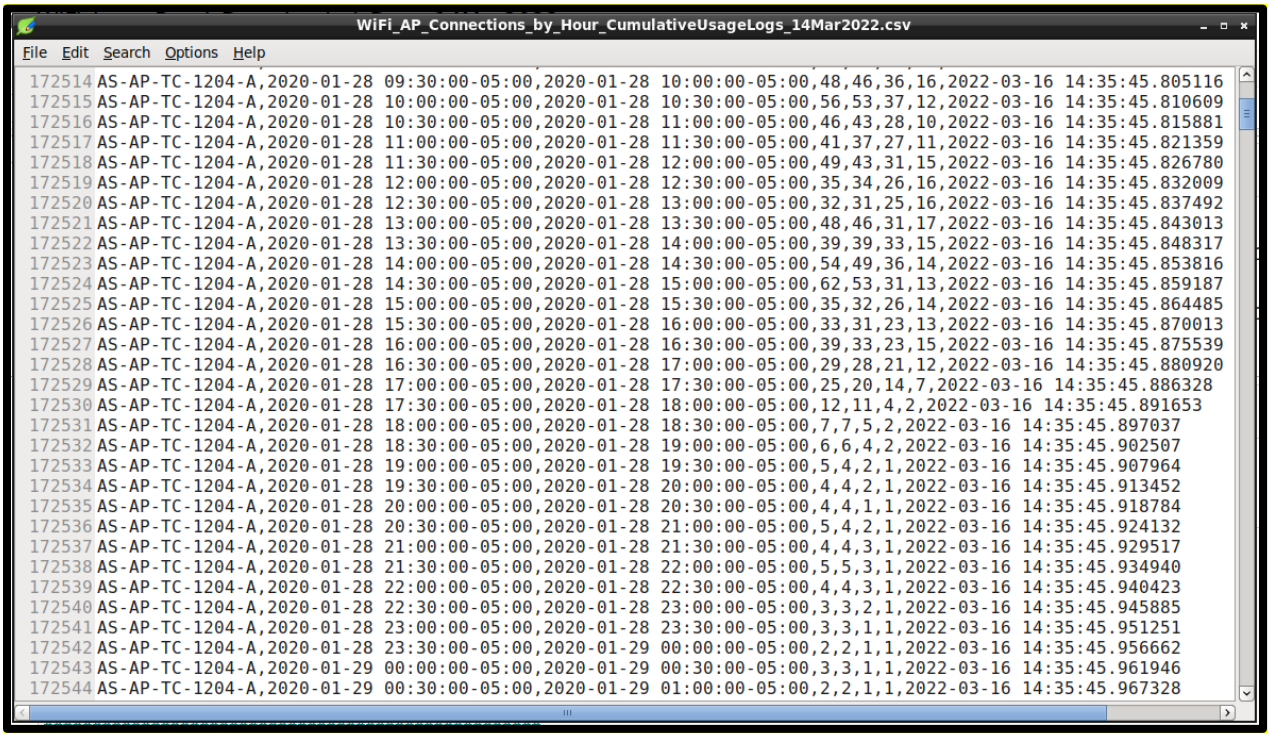

Figure 13: Counts of connected users by 30-minute period to AP AS-AP-TC-1204-A

Figure 13 displays a sample of the numbers of unique users, connected to AP *AS-AP-TC-1204-A*, during 30-minute periods on January 28th, 2020. The columns in the csv file are, respectively, AP, Start Period, End Period, Devices, Users, Users30, Users60, and DateTime recorded.

The final output file contained 5.95M rows, keyed by AP, Begin Period and End Period, where the Begin Period and End Period represented unique 30-minute periods between Dec 31st, 2019, and Feb 7th 2010, spanning the first 6 weeks of the Winter 2020 Term. There were 5,343 unique APs, with the number of unique wireless connections (Users), for each AP and 30-minute period, ranging from 0 to 654. The mean value of Users was 6.69, for each 30-minute period, for each AP. The maximum was 622 (see Table 8).

	AP	BeginPeriod	EndPeriod	Devices	Users	Users30	Users60
count	5946362	5946362	5946362	5946362.00000	5946362.00000	5946362.00000	5946362.00000
unique	5343	1804	1804	NaN	NaN	NaN	NaN
top	RN-AP-CLVN-237-A	2020-01-06 14:00:00-05:00	2020-01-06 14:30:00-05:00	NaN	NaN	NaN	NaN
freq	1804	5125	5125	NaN	NaN	NaN	NaN
first	NaN	2019-12-31 22:00:00-05:00	2019-12-31 22:30:00-05:00	NaN	NaN	NaN	NaN
last	NaN	2020-02-07 11:30:00-05:00	2020-02-07 12:00:00-05:00	NaN	NaN	NaN	NaN
mean	NaN	NaN	NaN	6.95714	6.68740	4.89223	2.97667
std	NaN	NaN	NaN	14.47715	13.65780	10.23090	6.72038
min	NaN	NaN	NaN	0.00000	0.00000	0.00000	0.00000
25%	NaN	NaN	NaN	1.00000	1.00000	1.00000	0.00000
50%	NaN	NaN	NaN	3.00000	3.00000	2.00000	1.00000
75%	NaN	NaN	NaN	7.00000	7.00000	5.00000	3.00000
max	NaN	NaN	NaN	654.00000	622.00000	345.00000	241.00000

Table 8: Statistical summary of AP logs for the first 6 weeks.

4.4.4 Wi-Fi Logs Read Scrubbed Data

The total users at each AP calculated in Section 4.4.3 can be used to determine the total users in each indoor space if the Building Floor location of each AP is known. The wireless logs must be merged with location information for each AP to calculate the number of wireless users present in an indoor area. The information required to identify the location of each AP is obtainable from four complementary sources: the AP Name, the AP Installations dataset, the Space Management dataset, and the Building and Floor layouts dataset. In this section, the information from all four datasets are combined and/or referenced to produce a report of wireless users in each indoor area by building, floor, zone, and AP.

The total Users at each AP for each 30-minute period, calculated in Section 4.4.3, represents the number of individuals in the area served by the AP. The area may be a single space (e.g., classroom), or a collection of rooms. The relationship between APs and spaces can be 1:1 (1 AP in one room), 1:many (1 AP serving many rooms), many:1 (multiple APs in one room), or many:many (multiple APs providing wireless connections to occupants in multiple rooms). To model the spaces with wireless connections, it is helpful to think of Building Floor Zones. A Building Floor Zone is a collection of spaces that contain one or more APs, and one or more rooms. It is assumed that all APs installed in the Building Floor Zone provide wireless connections to all devices present in the Building Floor Zone (many:many).

The concept of a Building Floor Zone addresses a number of issues related to localizing individuals within an indoor space using logs: a wireless network automatically switching a user connection between nearby APs; load balancing algorithms that connect users to APs that are not physically closest; the challenge of identifying which particular room on a Building Floor a wireless user is occupying; and providing a precise location of a user at a level of detail smaller than Building Floor.

The definition of Building Floor Zones within a Building Floor can be accomplished with various Global Positioning Systems (GPS) or other coordinate systems/frameworks; in this project room number prefixes

are used as a proxy for a coordinate system. In most buildings rooms are consecutively numbered along hallways: adjacent rooms have similar numbers, and rooms with widely differing numbers are unlikely to be in the same vicinity on a given Building Floor. The first two/three characters of a room number can be used to define a Building Floor Zone, allowing the grouping of rooms in close proximity and segmenting of areas in Building Floors.

To map an AP to a Building Floor Zone, the room where the AP installed must first be determined. The room can then be mapped to a Building Floor Zone. AP installation locations can be ascertained from a combination of one/more of the AP Name, AP Installations dataset, Space Management dataset and Building and Floor Layouts dataset.

At Waterloo, most AP Names incorporate the building abbreviation and room/hallway number where the AP is installed. If the installation location is a room, then the zone is calculated as the first n characters of the room number, where:

n = length(room number with suffix removed) – 1

e.g., for room 315E:

n = length(315) -1

n = 2

zone = 31

If the AP installation location is a hallway, the zone is calculated as the first n characters of the arithmetic average of the surrounding rooms' numbers.

n = length(integer(average(4 closest Academic Support room numbers with suffix removed)) – 1

e.g., for rooms 1351A, 1302, 1303, 1301A:

n = length(integer(average(1351,1302, 1303,1301))) – 1

n = length(1314) -1

n = 3

zone = 131

The surrounding rooms were determined using the AP Installations, Building and Floor Layout, and Space Management datasets. The AP Names, AP Installations, and Building and Floor Layouts were used to determine the building number, floor number and room/hallway number. The Space Management dataset was used to categorize the room/hallway number as a ROOM or HALLWAY based on the COU categorization of the space in the administrative system of record, Archibus. In Archibus, hallways, corridors and other circulation spaces are categorized as room type 16.0 (see Table 1). Confirming the installation locations as ROOM or HALLWAY was necessary because the varied layouts in buildings would

have required making a judgement about a particular installed location. All APs were identified as being installed in ROOMs or HALLWAYs, and additional information was added for the 966 hallway-installed APs.

For APs installed in hallways, the surrounding rooms were selected only from Academic (Support) rooms, determined using the Space Management dataset, i.e., no spaces categorized as room type 16.0 were used to precisely locate an AP. For example, APs *WN-AP-DC-1843-A* and *WN-AP-DC-1843-B* are installed in hallway *1843* of building *DC*. The set of the closest Academic (Support) rooms are {1314, 1315, 1316, 1317} and {1307, 1308, 1309, 1310} respectively (see Figure 14). AP *WN-AP-DC-1844-A* is installed in hallway 1844, between rooms {1336, 1340, 1339, 1343}.

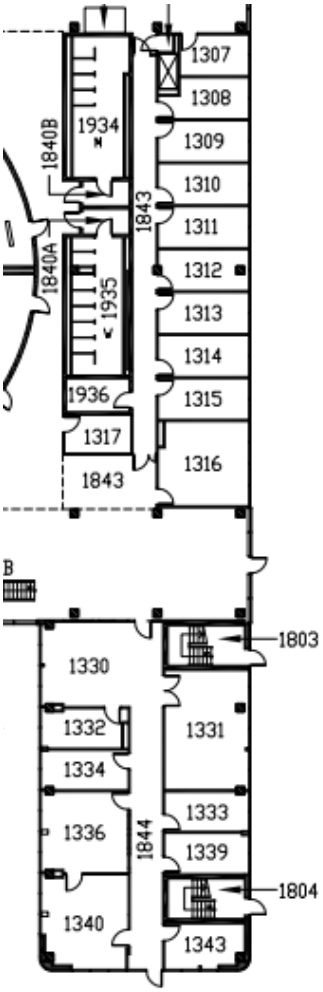

Figure 14: Sample floor layout showing room numbering for hallways, washrooms, and stairwells.

At Waterloo, spaces on Building Floors are numbered consecutively/similarly if they are Academic or Academic Support spaces, but elevators, hallways, utility rooms and other spaces not assigned/occupied to departments have different naming conventions (see Figure 12). Defining Building Floor Zones using room number prefixes gives a more precise location for each hallway-installed AP. Using Academic

(Support) spaces, like classrooms and offices, for room numbers closely associates the APs with the types of spaces being occupied by students and employees and is consistent with the zone definitions used for APs installed in rooms. Table 9 shows sample data from the triangulation method used for locating hallway APs:

AP_Name	Room_or_Hallway	Label	Triangulate1_Room	Triangulate2_Room	Triangulate3_Room	Triangulate4_Room
WN-AP-DC-1823-E	ROOM	1505F				
WN-AP-DC-1841-A	HALLWAY	1841	1351A	1302	1303	1301A
WN-AP-DC-1852-C	ROOM	1120				
WN-AP-DC-1852-B	ROOM	1120				
WN-AP-DC-1843-B	HALLWAY	1843	1307	1308	1309	1310
WN-AP-DC-1843-A	HALLWAY	1843	1316	1314	1317	1315
WN-AP-DC-1844-A	HALLWAY	1844	1336	1340	1339	1343

Table 9: Locating hallway APs using nearby Academic Support spaces.

The summarized file of 5.95M rows showing total Users per AP for each 30-minute period was further reduced by selecting only the records for APs installed in Academic or Academic Support buildings. Building and Floor Layout, and AP Installations datasets were not available for many buildings that were not Academic or Academic Support, e.g., residences, because of security and privacy concerns. The final number of AP records selected for further processing was 3.56M rows, including 2,192 unique APs installed in 1,207 rooms/hallways in 52 buildings.

Using the datasets, over 97% of the APs in Academic and Academic Support buildings had their locations precisely defined. The locations of 65 of the 2,192 APs (2.97%) could not be determined, due to missing data. 23 of the 52 buildings had APs that could not be precisely located, but most of the missing APs were in a small number of buildings: of the 23 buildings with missing APs, 14 were missing 1 or 2 APs. The 65 APs without a known location represented 2.92% (103,984 of 3,564,854) of the Wi-Fi connection logs. Tables 10 and 11 display the counts of known and unknown AP locations, and the corresponding counts of known and unknown wireless logs, by building. The complete listing of buildings with proportion of missing APs is in Appendix B.

	Building	N	Y
0	AL	NaN	21.00000
1	ARC	NaN	31.00000
2	B1	NaN	26.00000
3	B2	NaN	21.00000
4	BMH	1.00000	89.00000
5	C2	NaN	33.00000
6	CIF	2.00000	21.00000
7	COM	NaN	7.00000
8	CPH	1.00000	49.00000
9	CSB	1.00000	NaN
10	DC	4.00000	110.00000
11	DMS	1.00000	28.00000
12	DMS2	9.00000	NaN
13	DWE	NaN	54.00000
14	E10	3.00000	NaN
15	E2	1.00000	69.00000
16	E3	NaN	45.00000
17	E5	NaN	81.00000
18	E6	NaN	36.00000
19	E7	1.00000	199.00000
20	ECH	NaN	26.00000
21	EIT	1.00000	59.00000
22	ERC	1.00000	11.00000
23	ESC	1.00000	31.00000

Table 10: Sample counts of known and unknown AP locations, by Building.

	Building	N	Y
0	AL	0.00000	33454.00000
1	ARC	0.00000	52572.00000
2	B1	0.00000	44612.00000
3	B2	0.00000	35988.00000
4	BMH	1728.00000	152346.00000
5	C2	0.00000	57442.00000
6	CIF	3504.00000	35582.00000
7	COM	0.00000	12302.00000
8	CPH	1746.00000	82840.00000
9	CSB	1778.00000	0.00000
10	DC	6998.00000	180350.00000
11	DMS	1692.00000	47010.00000
12	DMS2	9462.00000	0.00000
13	DWE	0.00000	90186.00000
14	E10	5000.00000	0.00000
15	E2	1738.00000	118602.00000
16	E3	0.00000	77978.00000
17	E5	0.00000	138694.00000
18	E6	0.00000	63336.00000
19	E7	1704.00000	181156.00000
20	ECH	0.00000	44728.00000
21	EIT	1694.00000	100844.00000
22	ERC	1734.00000	19334.00000
23	ESC	1738.00000	52962.00000

Table 11: Sample counts of known and unknown AP connections, by Building.

Additional fields in the summarized Wi-Fi dataset included User30 and User60, the number of Users in each time period whose connection durations lasted at least 30 minutes, and at least 60 minutes, respectively. Sample records from the summarized file is shown in Table 12:

AP	Building	Floor_Code	Room_or_Hall	AP_Room_or_Hallway	BeginPeriod	EndPeriod	Users	Users30	Users60	Known_AP_Location
WN-AP-RCH-103-A	RCH	01	103	ROOM	2020-01-02 19:00:00-05:00	2020-01-02 19:30:00-05:00	1	1	0	Y
WN-AP-RCH-103-A	RCH	01	103	ROOM	2020-01-02 19:30:00-05:00	2020-01-02 20:00:00-05:00	1	1	0	Y
WN-AP-RCH-103-A	RCH	01	103	ROOM	2020-01-02 20:00:00-05:00	2020-01-02 20:30:00-05:00	1	1	0	Y
WN-AP-RCH-103-A	RCH	01	103	ROOM	2020-01-02 20:30:00-05:00	2020-01-02 21:00:00-05:00	1	1	0	Y
WN-AP-RCH-103-A	RCH	01	103	ROOM	2020-01-02 21:00:00-05:00	2020-01-02 21:30:00-05:00	1	1	0	Y
...	...	...	...	...	...	...	...	...	...	...
WN-AP-RCH-331-B	RCH	03	331	HALLWAY	2020-02-06 15:30:00-05:00	2020-02-06 16:00:00-05:00	64	36	22	Y
WN-AP-RCH-331-B	RCH	03	331	HALLWAY	2020-02-06 16:00:00-05:00	2020-02-06 16:30:00-05:00	58	35	19	Y
WN-AP-RCH-331-B	RCH	03	331	HALLWAY	2020-02-06 16:30:00-05:00	2020-02-06 17:00:00-05:00	37	25	12	Y
WN-AP-RCH-331-B	RCH	03	331	HALLWAY	2020-02-06 17:00:00-05:00	2020-02-06 17:30:00-05:00	32	19	11	Y
WN-AP-RCH-331-B	RCH	03	331	HALLWAY	2020-02-06 17:30:00-05:00	2020-02-06 18:00:00-05:00	9	9	7	Y

Table 12: Sample Wi-Fi records showing locations of APs and User counts.

4.4.5 Archibus COU Categories Read Raw Data

Wireless activity inside buildings is not only a function of building size, location, and layout; the types of activity that take place on each Building Floor may also determine the volume of activity. On a university campus, offices typically have lower counts of persons than classrooms, which themselves have fewer persons than food courts. Since similar spaces are often grouped together (e.g., offices), or an entire floor is dedicated to a specific teaching/research function (e.g., library) a Building Floor can often be generally described as having primarily classrooms, or offices, or athletic facilities.

The Council of Ontario Universities (COU), a collaborative forum for Ontario Universities, defines 20 high-level classifications for university space. The classifications are listed in Table 13. Each category has multiple sub-categories, e.g., Category 01 (*Classrooms*) include sub-categories *Tiered Classrooms* (1.1), *Non-Tiered Classrooms* (1.2), *Active Learning Classrooms* (1.3), and *Classroom Service and Exam Space* (1.4). Each interior space (e.g., lab, elevator, hallway, classroom, foyer) is measured and classified by the facilities management department. The complete listing of COU Space Categories is in Appendix C. Waterloo's spaces include most of the space categories defined by the COU.

Category	Category Description
1.0	CLASSROOM FACILITIES
2.0	LABORATORY - UNDERGRADUATE
3.0	RESEARCH LABORATORY SPACE
4.0	ACADEMIC DEPARTMENTAL OFFICE & RELATED
5.0	LIBRARY FACILITIES & LIBRARY STUDY SPACE
6.0	ATHLETIC / RECREATION SPACE
7.0	FOOD SERVICE
8.0	BOOKSTORE & OTHER MERCHANDISING FACILITIES
9.0	PLANT MAINTENANCE
10.0	CENTRAL ADMINISTRATIVE OFFICE AND RELATED
11.0	NON-LIBRARY STUDY SPACE
12.0	CENTRAL SERVICES
13.0	HEALTH SERVICE FACILITIES
14.0	COMMON USE & STUDENT ACTIVITY SPACE
15.0	ASSEMBLY & EXHIBITION FACILITIES
16.0	NON-ASSIGNABLE
17.0	RESIDENTIAL SPACE
18.0	ANIMAL SPACE
19.0	OTHER UNIVERSITY FACILITIES
20.0	HEALTH SCIENCE CLINICAL FACILITIES

Table 13: Council of Ontario Universities (COU) Space categories.

The Space Metrics dataset was used to group/categorize each Building Floor zone based on the primary activities taking place, to better describe the areas being served by each AP. Campus facilities for similar functions (e.g., teaching, research, administration) are usually co-located: there are clusters of like spaces found on each Building Floor. The COU category that has the largest percentage area of the Building Floor zone is assumed to describe the Building Floor's primary activity/purpose. Building Floor Zones were defined by the same formula used in section 4.4.4: Building Code + Floor Code + the first two/three characters of the Room Number.

Building Floor Zone = Building code & "_" & Floor code & "-" & Room Number prefix

The Space Metrics dataset for Winter 2020 included 33,637 rows of data on indoor areas, defined by building, floor, room/hallway, and owner. The dataset included spaces shared by multiple departments, which are shown as unique records in the dataset. The dataset included all buildings in the university's real estate portfolio. The records are maintained by facilities management staff and updated during renovation and construction projects.

The 33,844 defined spaces in the Space Metrics dataset were filtered to remove outdoor areas, stairwells, custodial rooms, mechanical and electrical equipment rooms, elevators, and other spaces that have minimal academic and/or administrative activity. Rooms that are shared spaces between multiple

departments were summarized to remove duplicates. The final row count of the Space Metrics dataset was 29,301 rows, representing 218 buildings and 684 floors.

The Space Metrics dataset was summarized by Building Floor Zone. 5,834 zones were calculated. The total area in each COU category for each Building Floor Zone was determined, as well as the percentage area for each COU category. Table 14 displays sample data for spaces within building *E2*, areas are measured in square metres.

Building_Alpha_Code	Floor	Building_Floor	Building_Floor_Zone	COU_01	COU_02	COU_03	COU_04	COU_01_Pct	COU_02_Pct	COU_03_Pct	COU_04_Pct
E2	1	E2_01	E2_01_171	0.00	125.28	0.00	26.93	0.00	0.33	0.00	0.07
E2	1	E2_01	E2_01_172	0.00	0.00	398.58	0.00	0.00	0.00	0.97	0.00
E2	1	E2_01	E2_01_173	243.73	0.00	194.19	0.00	0.56	0.00	0.44	0.00
E2	1	E2_01	E2_01_175	0.00	45.59	47.28	35.45	0.00	0.36	0.37	0.28
E2	1	E2_01	E2_01_176	0.00	193.99	0.00	0.00	0.00	1.00	0.00	0.00
E2	1	E2_01	E2_01_177	0.00	55.33	0.00	285.34	0.00	0.12	0.00	0.61
E2	1	E2_01	E2_01_178	0.00	78.45	0.00	111.37	0.00	0.19	0.00	0.27
E2	1	E2_01	E2_01_179	0.00	591.81	0.00	54.47	0.00	0.85	0.00	0.08
E2	1	E2_01	E2_01_180	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E2	1	E2_01	E2_01_181	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E2	1	E2_01	E2_01_182	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E2	1	E2_01	E2_01_183	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E2	1	E2_01	E2_01_184	0.00	9.45	0.00	0.00	0.00	0.05	0.00	0.00
E2	1	E2_01	E2_01_194	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
E2	2	E2_02	E2_02_230	0.00	0.00	0.00	116.70	0.00	0.00	0.00	1.00

Table 14: Sample Building Floor Zones showing total and percent areas for selected COU categories.

A K-Means algorithm was executed to group all Building Floors across the campus into similar categories. The 5,834 Building Floor Zones in the dataset were each classified based on the largest percentage area of their COU categories. The K-Means algorithm was seeded with an initial array of 20 records (see Figure 15) that were based on the twenty COU Categories shown in Table 13.

```
([[0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.500, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 1.000, 0.000, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.500, 0.000, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.500, 0.000, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.500, 0.000],
  [0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.500, 0.000, 0.000, 0.000, 0.500]])
```

Figure 15: Initial/starting array points for K-Means algorithm.

The cluster centroids are displayed in Table 15, with the largest values in each row (the primary COU Classification) highlighted. The different COU categories (Cat01, Cat02, etc.) are described in Table 13.

	Cat01	Cat02	Cat03	Cat04	Cat05	Cat06	Cat07	Cat08	Cat09	Cat10	Cat11	Cat12	Cat13	Cat14	Cat15	Cat16	Cat17	Cat18	Cat19	Cat20
0	0.61	0.02	0.03	0.13	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	-0.00
1	0.03	0.72	0.05	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.01	0.02	-0.00	0.00	0.00	0.01
2	0.00	0.02	0.75	0.19	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	-0.00	0.00	0.00	0.00
3	0.01	0.01	0.03	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	-0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
4	0.00	0.00	-0.00	0.01	0.89	-0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.01	0.00	0.00	0.00
5	-0.00	0.00	-0.00	-0.00	-0.00	0.96	-0.00	0.00	-0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.03	-0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.01	0.00	0.00	0.90	0.00	0.00	0.03	0.01	0.00	0.00	0.01	0.00	0.03	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.80	0.00	0.08	0.00	0.01	0.00	0.02	0.00	0.05	0.02	0.00	0.02	0.00
8	0.00	0.00	0.01	0.00	-0.00	0.00	0.01	0.00	0.88	0.02	0.00	0.01	0.00	0.01	0.00	0.06	0.02	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00
10	-0.00	0.00	0.00	0.01	0.00	-0.00	0.00	0.00	-0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	0.14	0.00	0.78	0.00	0.00	-0.00	0.04	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.02	0.00	0.00	0.93	0.00	-0.00	0.04	-0.00	0.00	0.01	0.00
13	0.02	0.01	0.00	0.04	-0.00	0.02	0.04	0.02	0.00	0.02	0.01	0.00	0.00	0.74	0.00	0.06	0.02	0.00	0.00	0.00
14	0.00	0.01	0.01	0.07	0.02	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.79	0.07	0.00	0.00	0.02	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.98	0.00	0.00	0.00	0.00
16	0.00	0.01	0.01	0.02	0.01	0.00	0.01	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.56	0.32	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.99	0.00	0.00	0.00
18	0.00	0.00	0.00	0.01	-0.00	-0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.01	-0.00	0.00	0.96	0.00
19	0.00	0.02	0.03	0.02	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.02	0.88

Table 15: K-Means centroids, with percentages of each COU category.

Each of the centroids was labelled according to the largest COU category by percentage. The 20 clusters for the 5,834 Building Floor Zones, the assigned cluster names and count of Building Floors are shown in Table 16.

Clusters	Cluster_Names	Building_Floor	Building_Floor_Zone
0	Classroom	87	154
1	UndergradLab	57	98
2	ResearchLab	86	251
3	AcademicOffice	121	602
4	Library	20	43
5	Athletics	4	44
6	FoodServices	21	56
7	Bookstore	4	9
8	PlantMaintenance	26	38
9	AdministrativeOffice	44	271
10	StudySpace	66	299
11	CentralServices	7	12
12	HealthServices	3	23
13	CommonUse	17	24
14	AssemblyExhibition	13	18
15	Circulation/OpenArea	400	1119
16	MixedUse	186	196
17	ResidenceRooms	384	2481
18	OtherSpace	29	78
19	ClinicalFacilities	4	15

Table 16: Building Floor Zone counts by K-Means cluster names.

The clustering algorithm demonstrated that the university has spaces in all COU categories except Cat18: Animal Space (see Table 15). By count, most of the 5,834 Building Floor Zones are in Residences or are Circulation/Open Areas, with large numbers of Building Floor Zones classified as Academic Offices, Study Spaces, and Administrative Offices.

The count of Building Floor Zones in each category correlates with the total space on the university campus in each COU category: Circulation/OpenArea, Residence Rooms, Research Labs, Academic Offices and Classrooms are the top five COU categories on the university campus when measured by floor area.

Identifying and categorizing clusters of Building Floor Zones was done to facilitate analysis of Wi-Fi records by type of space. Sample records for the first 4 Clusters (i.e., Classroom, UndergradLab, ResearchLab, AcademicOffice) are shown in Tables 17, 18, 19, 20.

Building_Floor	Building_Floor_Zone	Clusters	Cluster_Names	COU_01_Pct	COU_02_Pct	COU_03_Pct	COU_04_Pct	COU_16_Pct
AAR_01	AAR_01_100	0	Classroom	0.00	0.00	0.00	0.00	0.00
AL_00	AL_00_01	0	Classroom	0.57	0.00	0.00	0.00	0.29
AL_01	AL_01_11	0	Classroom	0.87	0.00	0.00	0.00	0.04
AL_01	AL_01_12	0	Classroom	0.77	0.00	0.00	0.00	0.10
AL_02	AL_02_20	0	Classroom	0.74	0.00	0.00	0.00	0.26
AL_02	AL_02_21	0	Classroom	0.90	0.00	0.00	0.00	0.10
ARC_01	ARC_01_110	0	Classroom	0.37	0.28	0.00	0.03	0.00
B1_02	B1_02_27	0	Classroom	0.40	0.00	0.00	0.27	0.34
BMH_01	BMH_01_100	0	Classroom	1.00	0.00	0.00	0.00	0.00
BMH_01	BMH_01_101	0	Classroom	0.51	0.00	0.00	0.49	0.00
BMH_01	BMH_01_104	0	Classroom	0.50	0.00	0.00	0.50	0.00

Table 17: Sample data for Building Floor Zones in cluster Classroom.

Building_Floor	Building_Floor_Zone	Clusters	Cluster_Names	COU_01_Pct	COU_02_Pct	COU_03_Pct	COU_04_Pct	COU_16_Pct
ACW_02	ACW_02_200	1	UndergradLab	0.00	1.00	0.00	0.00	0.00
ARC_01	ARC_01_100	1	UndergradLab	0.31	0.52	0.00	0.02	0.00
ARC_03	ARC_03_300	1	UndergradLab	0.00	0.59	0.00	0.39	0.00
ARC_03	ARC_03_310	1	UndergradLab	0.00	0.94	0.00	0.00	0.06
B1_03	B1_03_37	1	UndergradLab	0.00	0.52	0.23	0.21	0.00
B2_01	B2_01_14	1	UndergradLab	0.00	0.83	0.07	0.10	0.00
BMH_03	BMH_03_367	1	UndergradLab	0.00	0.71	0.00	0.00	0.00
BMH_03	BMH_03_368	1	UndergradLab	0.12	0.76	0.00	0.03	0.00
BMH_04	BMH_04_469	1	UndergradLab	0.00	1.00	0.00	0.00	0.00
C2_01	C2_01_16	1	UndergradLab	0.00	0.40	0.23	0.37	0.00
C2_02	C2_02_27	1	UndergradLab	0.07	0.71	0.00	0.22	0.00

Table 18: Sample data for Building Floor Zones in cluster UndergradLab.

Building_Floor	Building_Floor_Zone	Clusters	Cluster_Names	COU_01_Pct	COU_02_Pct	COU_03_Pct	COU_04_Pct	COU_16_Pct
AVR_01	AVR_01_100	2	ResearchLab	0.00	0.00	0.92	0.08	0.00
AVR_01	AVR_01_101	2	ResearchLab	0.00	0.00	1.00	0.00	0.00
B1_01	B1_01_16	2	ResearchLab	0.00	0.26	0.31	0.27	0.10
B1_01	B1_01_17	2	ResearchLab	0.00	0.03	0.82	0.10	0.05
B1_02	B1_02_28	2	ResearchLab	0.00	0.00	0.56	0.44	0.00
...	...	...	...	...	...	...	...	...
TJB_01	TJB_01_110	2	ResearchLab	0.00	0.00	1.00	0.00	0.00
TJB_01	TJB_01_114	2	ResearchLab	0.00	0.00	1.00	0.00	0.00
TJB_01	TJB_01_116	2	ResearchLab	0.00	0.00	1.00	0.00	0.00
TJB_01	TJB_01_117	2	ResearchLab	0.00	0.00	0.87	0.00	0.13
TJB_02	TJB_02_216	2	ResearchLab	0.00	0.00	0.47	0.53	0.00

Table 19: Sample data for Building Floor Zones in cluster ResearchLab.

Building_Floor	Building_Floor_Zone	Clusters	Cluster_Names	COU_01_Pct	COU_02_Pct	COU_03_Pct	COU_04_Pct	COU_16_Pct
ARC_02	ARC_02_200	3	AcademicOffice	0.00	0.38	0.00	0.62	0.00
ARC_02	ARC_02_201	3	AcademicOffice	0.00	0.25	0.00	0.75	0.00
ARC_02	ARC_02_202	3	AcademicOffice	0.33	0.00	0.00	0.67	0.00
ARC_03	ARC_03_301	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00
ARC_03	ARC_03_302	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00
...	...	...	...	...	...	...	...	...
TJB_02	TJB_02_226	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00
TJB_02	TJB_02_227	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00
TJB_02	TJB_02_230	3	AcademicOffice	0.00	0.00	0.00	0.96	0.04
TJB_02	TJB_02_231	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00
TJB_02	TJB_02_232	3	AcademicOffice	0.00	0.00	0.00	1.00	0.00

Table 20: Sample data for Building Floor Zones in cluster AcademicOffice.

For almost all Building Floor Zones on campus, the second largest COU classification for space is hallway/circulation: COU_16. The percentage of space given over to that category is also shown for the sample floors in Tables 12.

Additional contextual data was added to the Space Metrics dataset for later analysis. The building construction year and Gross Area to Net Assignable Ratio were included. The construction year was assumed to be a proxy for the types and quantities of materials used in constructing the buildings. The Gross to Net Area is a measure of the degree of openness of the floor layouts: buildings with a higher ratio have more open indoor areas, such as atria and wider hallways.

4.4.6 Student Activity Read Downloaded Data

The Class Schedule for Winter 2020 was transformed to calculate the number of expected students in each Building, Floor and Room, for each 30-minute period of the first 6 weeks of 2020. The 6,163 Class Schedule records were reduced to 3,611 in-person classes by removing class components with values such as “Online” and “Field Studies”, which had no valid building location or Start Date.

The number of students expected in each 30-minute timeslot in each room was determined by looping through each Course Schedule record, from the Course Start Date to the Course End Date, for the Course Schedule Days, from the Course Start Time to the Course End Time. For example, in Table 9 below, Section 1 of Class ACC607 had an enrolment of 53 persons and was scheduled on Mondays and Wednesdays from 14:30:00 to 15:50:00 EST in Room HH 1106. The algorithm assigned an expected enrolment value of 53 to space HH 1106, for all 30-minute periods:

- On Mondays and Wednesdays
- Between 14:30 and 15:50
- Between January 6th and April 3rd, inclusive

The algorithm was as follows:

1. Select only Class Schedule records that are taught/presented in-person (e.g., exclude online and distance education classes)
2. For each Class Schedule Record, determine the Start Date and End Date for each Class/Section.
3. For each Class Schedule Record, determine the Scheduled Days (e.g., MWF), Start Time and End Time, when the class is scheduled.
4. Create program loop. For each Class Schedule Record, starting from the first meeting date/time for the term, based on Start Date, Scheduled Days, and meeting dates/times, increment the period by 30 minutes, until the last meeting date/time for the term, based on the Class End Date and Scheduled Days.
5. If the Class Schedule Record’s Start Date and End Date, plus Schedule Days, plus Start Time and End Time, overlap with the current period calculated in the loop, increment the Expected_Student_Count by the Class Section Enrollment.
6. Output the total Expected_Student_Count for each Building + Room, for each 30-minute period between the Class Start Date/Time and End Date/Time.
7. Retrieve the next Class Schedule Record
8. Go To Step #2

CLASS	CLASS_SECTION	SCHEDULE_DAYS	SSCHEDULE_START_TIME	SSCHEDULE_END_TIME	COMPONENT_TYPE	ENROLLMENT	COURSE_START_DATETIME	COURSE_END_DATETIME	BUILDING_ABBREV	ROOM_NUMBER
ACC607	1	MW	14:30:00	15:50:00	Lecture	53	2020-01-06 14:30:00	2020-04-03 15:50:00	HH	1106
ACC607	2	TR	10:00:00	11:20:00	Lecture	50	2020-01-06 10:00:00	2020-04-03 11:20:00	HH	2107
ACC607	3	TR	11:30:00	12:50:00	Lecture	49	2020-01-06 11:30:00	2020-04-03 12:50:00	HH	2107
ACC607	4	TR	14:30:00	15:50:00	Lecture	51	2020-01-06 14:30:00	2020-04-03 15:50:00	HH	1106
ACC611	1	MW	11:30:00	12:50:00	Lecture	48	2020-01-06 11:30:00	2020-04-03 12:50:00	HH	1106

Table 21: Sample data for classes ACC607 and ACC611.

Figure 16 shows sample records from the final dataset shows enrollments of 53 persons in room HH 1106, every Monday and Wednesday, for every 30-minute period between 14:30:00 and 16:00:00 between January 6th and April 3rd, 2020:

```
STRM,Component_Type,Class,Class_Section,Weekday,Enrolment,StartDateTime,EndDateTime,Building,Floor,Room_Number,DateTimeAdded
1201,Lecture,ACC607,1,M,53,"01/06/2020, 14:30:00","01/06/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,M,53,"01/06/2020, 15:00:00","01/06/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,M,53,"01/06/2020, 15:30:00","01/06/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,W,53,"01/08/2020, 14:30:00","01/08/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,W,53,"01/08/2020, 15:00:00","01/08/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,W,53,"01/08/2020, 15:30:00","01/08/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:52"
1201,Lecture,ACC607,1,M,53,"01/13/2020, 14:30:00","01/13/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,M,53,"01/13/2020, 15:00:00","01/13/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,M,53,"01/13/2020, 15:30:00","01/13/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/15/2020, 14:30:00","01/15/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/15/2020, 15:00:00","01/15/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/15/2020, 15:30:00","01/15/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,M,53,"01/20/2020, 14:30:00","01/20/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,M,53,"01/20/2020, 15:00:00","01/20/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,M,53,"01/20/2020, 15:30:00","01/20/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/22/2020, 14:30:00","01/22/2020, 15:00:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/22/2020, 15:00:00","01/22/2020, 15:30:00",HH,1,1106,"09/23/2022, 21:07:53"
1201,Lecture,ACC607,1,W,53,"01/22/2020, 15:30:00","01/22/2020, 16:00:00",HH,1,1106,"09/23/2022, 21:07:53"
```

Figure 16: Expected occupancy in building HH room 1106, every 30 minutes, due to ACC607.

The final dataset included 151,225 rows representing 2,587 unique in-person Class Sections delivered in 30 buildings.

The classroom schedules display the total Course Section Enrollment: the number of students registered in the Classes and Class Sections. The Expected_Student_Count calculated from the enrollment likely represents a maximum value for the period, since students frequently choose not to attend classes in which they are registered. The algorithm did not consider pedestrian traffic into and out of classrooms at the start and end of classes.

4.4.7 Employee Activity Read Downloaded Data

The Employee Location records, sourced from the campus telephone directory, were transformed into a dataset of Expected Employee counts for each half-hour period, for each Building and Room Code, during working days in the first 6 weeks of 2020.

The telephone records are maintained to display the office location(s) and phone extensions of full-time employees. The 5,483 records included Name, Phone Extension, Department, Building and Room. 5,324 records for employees in 71 buildings, remained after removing invalid/unusual values.

The algorithm for transforming Employee Locations into Employee Count for each Building Floor Zone assumed that all employees worked between 8:00 am and 5:00 pm on weekdays, from Monday, January 6th to Friday, February 7th, to match the available Wi-Fi records. The algorithm was as follows:

1. Read the employee telephone directory, including Building and Room

2. Remove duplicate telephone records, for employees with multiple offices.
3. Create a Start Date Time of 8:00 am January 6, 2020, and an End Date Time of 17:00 on February 7th, 2020 (6 weeks), for each employee.
4. Loop through the six weeks in 30-minute increments, for each employee.
5. If the 30-minute period falls within 8:00 and 17:00 on MTWTF, assign value = 1 to the Expected_Employee_Count.
6. Output the Building, Room, Date, Time, Expected_Employee_Count, for each 30-minute period in the first six weeks.
7. Retrieve the next telephone directory record
8. Go To Step #4.

The algorithm did not include any logic related to lunch/break periods, or employee activity in any other area other than their home office location. The Expected_Employee_Count calculated did not reflect any non-standard working hours (e.g., evening staff), Out of Office days (e.g., vacation), or employees who worked in areas other than their offices (e.g., attending meetings elsewhere).

4.4.8 Combine Scrubbed Wi-Fi, Enrolment, and Employment Data

The summarized datasets created in the preceding sections were combined to create a superset of data for the purposes of comparing expected vs. actual User counts. Previous research into the use of wireless logs to estimate indoor crowds' size and location provided data summarized at a building level, or floor level. If estimates of indoor crowds' size and location at a Building Floor Zone level is consistently more accurate and more precise than at a Building Floor level, then the methodology can be expanded and/or adapted to provide better indoor crowd measurement using wireless logs.

The summarized Class Schedules data from section 4.4.6, and Employee Locations data from section 4.4.7 were combined by building, floor, room and 30-minute time periods. The merged dataset represented an estimate of expected wireless users (i.e., students + employees) in each room for each half-hour period in the first six weeks of the term. The count of expected Users is a rough estimate based on class enrolments and employee office locations- it makes no allowance for actual attendance, employee activities outside of home offices (e.g., meetings), or non-work and non-learning activities (e.g., lunch breaks).

The Wi-Fi counts summary data calculated in section 4.4.4 was merged with the dataset of expected wireless users to create a superset of actual vs. expected User counts, by building, floor, room, and time period.

The Space Metrics summary data calculated in section 4.4.5 was added to the combined dataset of expected vs actual wireless users. The Space Metrics' data fields included Clusters, building construction year, and Gross Area : Net Area ratio.

The four datasets required additional data transformations prior to merging. Building codes, floor codes and room codes were often inconsistent, e.g., *EC1* vs. *BB1* for building names, *00* vs *B1* for basement floors, *AL 06* vs. *AL 006* for room numbers. Data errors were also present, e.g., APs with missing names, and rooms that had existed at some previous date, before renovations changed the floor layouts. Any necessary data edits were done by cross-referencing the information in all source records, to identify the most accurate information.

The four datasets also had mismatched record keys: no one dataset had a master list of key values. This necessitated outer joins for all records, and additional data processing to fill in missing values. Data gaps included missing Wi-Fi logs for known APs, and Wi-Fi logs for which APs could not be located.

Sample data for the final combined records are shown in Table 22. Day_of_Year is the BeginPeriod of each half-hour, reformatted as the ordinal day number of the year. Day_and_Hour is the BeginPeriod of each half-hour, reformatted as a combination of weekday, hour, and minute. Enrolment is the number of expected students, Employment the number of expected Employees. Additional data fields calculated include Exp_Occ - the total of Enrolment and Employment. A new Boolean value for Working Hours was calculated for the merged dataset, defined as periods between 08:00 and 17:30, Mondays through Fridays.

Building_Alpha_Code	Room_or_Hall	Day_of_Year	Day_and_Hour	Work_Hrs_Bool	Users	Users30	Users60	Enrolment	Employment	Exp_Occ	Clusters	Construction_Year	Gross_to_NASM_x
C2	278	29.00	4.10.00	1	18.00	11.00	9.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	36.00	4.10.00	1	25.00	21.00	16.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	31.00	6.10.00	1	22.00	18.00	13.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	29.00	4.11.00	1	14.00	13.00	12.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	36.00	4.11.00	1	27.00	23.00	19.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	31.00	6.11.00	1	21.00	18.00	15.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	29.00	4.10.30	1	18.00	13.00	12.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	36.00	4.10.30	1	32.00	26.00	19.00	15.00	1.00	16.00	1	1971-01-01	1.70
C2	278	31.00	6.10.30	1	25.00	19.00	14.00	15.00	1.00	16.00	1	1971-01-01	1.70
E5	3052	27.00	2.13.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.13.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	27.00	2.14.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.14.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	27.00	2.15.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.15.00	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	27.00	2.12.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.12.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	27.00	2.13.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.13.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	27.00	2.14.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00
E5	3052	34.00	2.14.30	1	NaN	NaN	NaN	14.00	1.00	15.00	3	2010-07-01	2.00

Table 22: Sample data showing Wi-Fi Users and Expected Occupants, by building and room.

The final combined dataset included 5.08M rows, including 192 buildings, 582 floors, 5,147 zones, and 20 Clusters.

4.4.9 Analytics Wi-Fi Enrolment Employment Data

In this Python program the dataset produced in section 4.4.8 was summarized by Building Floor Zone and time period. Missing and/or inconsistent records were first removed from the dataset. Table 23 shows counts of connection logs for selected APs, by week of year.

Building	AP_Floor	AP	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Sparklines
BMH	2	WN-[redacted]	372	923	907	945	948	716	
BMH	2	WN-[redacted]-A	218	975	1,232	1,196	1,418	993	
BMH	2	WN-[redacted]-A	488	1,569	2,035	2,111	2,339	1,858	
BMH	2	WN-[redacted]-A	988	2,998	3,404	4,039	3,302	2,152	
BMH	2	WN-[redacted]	381	1,009	1,370	1,671	1,932	1,625	
BMH	2	WN-[redacted]	189	503	535	534	544	302	
BMH	2	WN-[redacted]	200	607	850	725	796	549	
BMH	2	WN-[redacted]	221	614	597	672	682	433	
BMH	2	WN-[redacted]	191	849	785	1,110	1,086	959	
BMH	2	WN-[redacted]	199	1,197	1,092	995	974	702	
BMH	2	WN-[redacted]	763	1,947	1,768	1,759	1,966	1,365	
BMH	2	WN-[redacted]-A	167	430	790	543	585	372	
BMH	2	WN-[redacted]-A	362	1,136	1,108	1,154	1,186	905	
BMH	2	WN-[redacted]-B	138	950	883	1,199	1,079	650	
BMH	2	WN-[redacted]	14	71	282	453	583	396	
BMH	2	WN-[redacted]	104	120	403	560	534	275	
BMH	2	WN-[redacted]	121	1,011	1,282	1,583	1,643	1,026	
BMH	2	WN-[redacted]	489	3,618	4,370	4,987	5,575	3,905	
BMH	2	WN-[redacted]	162	616	870	792	774	470	
BMH	2	WN-[redacted]	696	3,449	5,729	6,585	5,423	4,946	
BMH	2	WN-[redacted]	397	1,471	1,531	1,901	1,999	1,463	
BMH	2	WN-[redacted]	232	1,362	1,728	1,358	1,321	924	
BMH	3	WN-[redacted]	444	1,520	1,375	1,582	1,489	1,091	
BMH	3	WN-[redacted]	236	1,686	2,098	2,122	2,171	1,669	
BMH	3	WN-[redacted]	239	1,042	1,141	927	1,091	773	
BMH	3	WN-[redacted]	572	1,648	1,665	1,602	1,728	1,398	
BMH	3	WN-[redacted]	474	1,489	1,584	1,125	1,590	791	
BMH	3	WN-[redacted]	419	1,878	1,504	1,646	1,709	1,363	
BMH	3	WN-[redacted]	413	1,014	1,840	1,990	2,081	1,474	
BMH	3	WN-[redacted]	123	501	1,051	1,089	1,262	805	
BMH	3	WN-[redacted]	170	600	1,591	1,683	1,504	1,067	
BMH	3	WN-[redacted]	376	2,049	1,979	1,852	1,871	1,093	
BMH	3	WN-[redacted]	158	907	2,144	2,086	2,077	1,537	
BMH	3	WN-[redacted]	3	850	852	1,065	628	481	
BMH	3	WN-[redacted]	845	5,809	5,021	5,178	4,982	2,161	
BMH	3	WN-[redacted]	318	1,985	2,429	2,673	2,520	1,699	

Table 23: Sample counts of wireless connections by AP and week of year.

The sample information in Figure 13 was broadly representative of the dataset as a whole: there were relatively low numbers in week 1 and week 6 of the dataset, compared to weeks 2, 3, 4, and 5. The records

for weeks 1 and 6 were removed from the dataset before calculation of correlation coefficients between and expected occupants.

The calculated correlation ratio between actual vs. expected users was analysed by duration of wireless connection (i.e., User vs. User30 vs. User60), type of space (e.g., Clusters, building age, Gross Area : Net Area ratio, Building Floor), and time of week (e.g., working vs non-working hours).

The calculated correlation ratio between actual vs. expected users is expected to be greater than one (1) because students and employees present in any Building Floor Zone will not only be individuals with scheduled activities (e.g., classes) but persons who are present in the space because of any other activities, including being in-transit to another campus location.

5.0 Results

5.1 Introduction

The use of wireless connection logs to determine crowd size and location has been limited by scalability, repeatability, and localization issues. The unique AP installation/architecture required in each location (campus, building, floor) to service employees and students; the varied coverage provided by individual APs, including between Building Floors; and load-balancing algorithms make it challenging to repeat successes in different environments, or to localize persons within a floor.

This project proposes a scalable, repeatable, and more specific method of measuring crowd size and location using Wi-Fi connection logs by mapping each AP to a zone within a building and floor, and combining the connection logs with building data and occupant data to confirm and/or contextualize the crowd patterns suggested by the data. The additional information permits more detailed analysis/interpretation of the connection logs, even when it is de-identified and/or summarized.

Waterloo's extensive Wi-Fi network coverage, the widespread use of Wi-Fi enabled devices by students and employees, and the automatic collection of wireless connection data, provides a reliable means of counting individuals on the campus. Additional information on AP locations, and Building Floor plans, suggested more precise locations for APs within a Building Floor. High-level employee and student schedules by location and time permitted an overall measurement of precision.

The Wi-Fi logs were first analyzed from an overall campus perspective. Reasonability checks were done by comparing the number of logs per day were compared to registration records, looking at crowd traffic patterns, and other high-level checks. The results indicated that the records were useful for analyzing campus crowds during the six-week period.

The results were then analyzed from a Building Floor perspective. The 171 buildings on campus are used for a variety of purposes, both academic and non-academic, so identifying the general purpose/activities expected to take place on each Building Floor allows for more detailed analysis of the connection logs from each category of Building Floor. The categories of Building Floor were determined by cluster analysis, using the Council of Ontario Universities (COU) space utilization standards.

Finally, the results were analyzed from a Building Floor zone perspective. The Wi-Fi counts for each zone were calculated, then compared to the expected count of employees and students for the Zone and period. Strong correlations were identified between Wi-Fi counts and expected persons for zones that could be described as dedicated to Academic and/or Academic Support activities, such as classrooms, labs, and offices.

5.2 Wireless Activity on Campus

The summary data for the first 6 weeks of 2020 is shown in Table 24. The summary statistics show that 15,027,493 wireless network logs were collected over the first 37 days of 2020 (Day_of_Year), from 5,344 unique Access Points (ap). Most connections had a duration in minutes of 30 minutes or less, with the maximum duration being 1410 minutes (23:30 hours:mins). More than half the durations were 0 mins, likely reflecting the rounding of Start Time and End Time to the nearest half-hour. The summary data also shows over 1 million unique user and mac values- this is misleading and due to the true user and mac values being hashed daily before the data was released.

	user	mac	ap	duration_mins	Day_of_Year	Week_of_Year
count	15027493	15027493	15027493	15027493.00000	15027493.00000	15027493.00000
unique	1045787	1565460	5344	NaN	NaN	NaN
top	xmG4KFOR5rkm6aB0ZczniUF7KMdbd87Z41nZvfJW	y5gLAvE9ZjtE8vIkYCFVrzVn0cafgBzLTXVFZpMy	WN-AP-LOT-D-B-OUTDOORTEST	NaN	NaN	NaN
freq	533	94	135085	NaN	NaN	NaN
mean	NaN	NaN	NaN	31.71129	21.03217	3.83046
std	NaN	NaN	NaN	73.53416	9.69741	1.41157
min	NaN	NaN	NaN	0.00000	1.00000	1.00000
25%	NaN	NaN	NaN	0.00000	13.00000	3.00000
50%	NaN	NaN	NaN	0.00000	21.00000	4.00000
75%	NaN	NaN	NaN	30.00000	29.00000	5.00000
max	NaN	NaN	NaN	1410.00000	37.00000	6.00000

Table 24: : Summary statistics for Wi-Fi connection logs, for first 37 days of 2020.

A truer representation of users can be obtained by counting the number of unique users in each day. Figure 17 shows the number of unique hashed user ids for each of the 37 days in the connection logs.

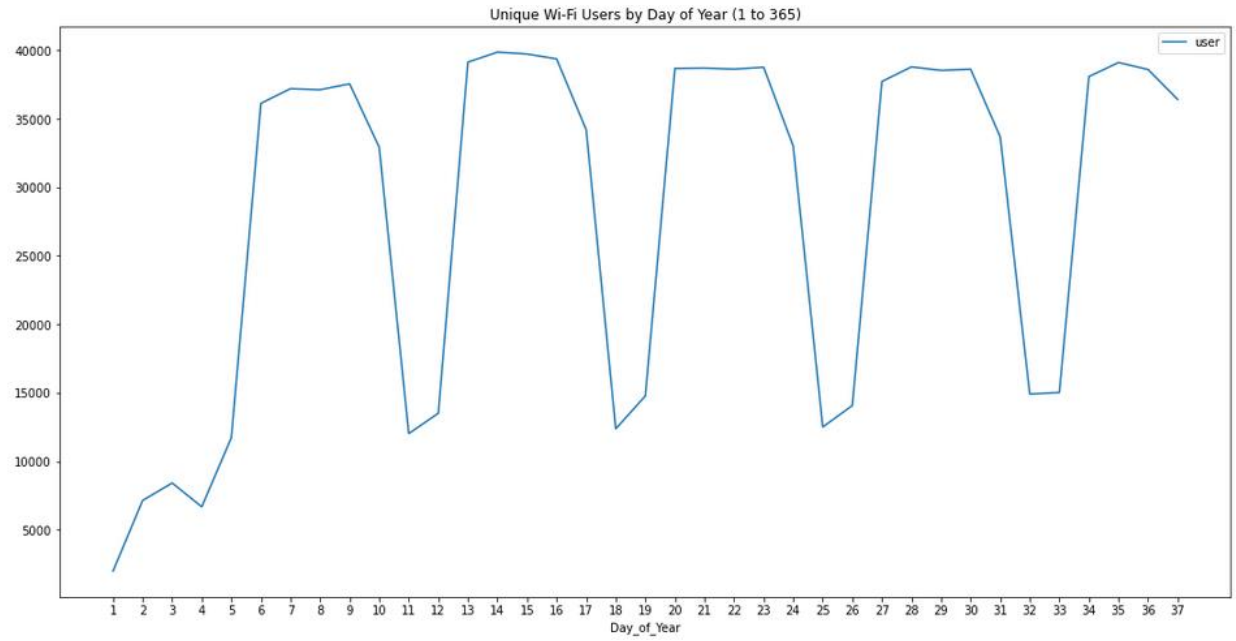

Figure 17: Number of unique users for the first 37 days of 2020.

The data in Figure 16 clearly shows a consistent number of daily users beginning on day *6* of 2020 (i.e., Monday, Jan 6, 2020) and continuing for most weekdays for the first 6 weeks. The number of campus wireless connections is slightly lower on Fridays (e.g., Days 10, 17, 24), and significantly lower on weekends (e.g., Days 11, 12, 18, 19). The number of unique wireless users on weekdays ranged between 36,000 and 40,000 on Mondays through Thursdays, dropping to approximately 33,000 on Fridays. This pattern of usage makes sense given that fewer academic activities are scheduled on Fridays, and none on Saturdays and Sundays. The 13,000 to 15,000 unique users on Saturdays and Sundays highlight that the campus is still a hub of activity on weekends, when no scheduled academic activity is taking place.

The log statistics are consistent with official records. The universities' official statistics show that there were 30,765 students registered for classes in Winter 2020 (https://uwaterloo.ca/institutional-analysis-planning/university-data-and-statistics/student-data/student-headcounts). This total did not include cross-registered students from the nearby Wilfrid Laurier University, or students on a Work Term who were present on the campus grounds. In January 2020, there were also 1,350 full-time faculty members and approximately 2,700 full-time permanent staff; the number of part-time and temporary staff in winter 2020 is unknown.

The significant levels of campus activity during non-working hours can also be demonstrated by the number of unique users each weekday before 8 am, and after 5 pm, plus weekends. Figure 18 shows the number of unique users for the first 37 days of 2020, outside of the university's core operating hours.

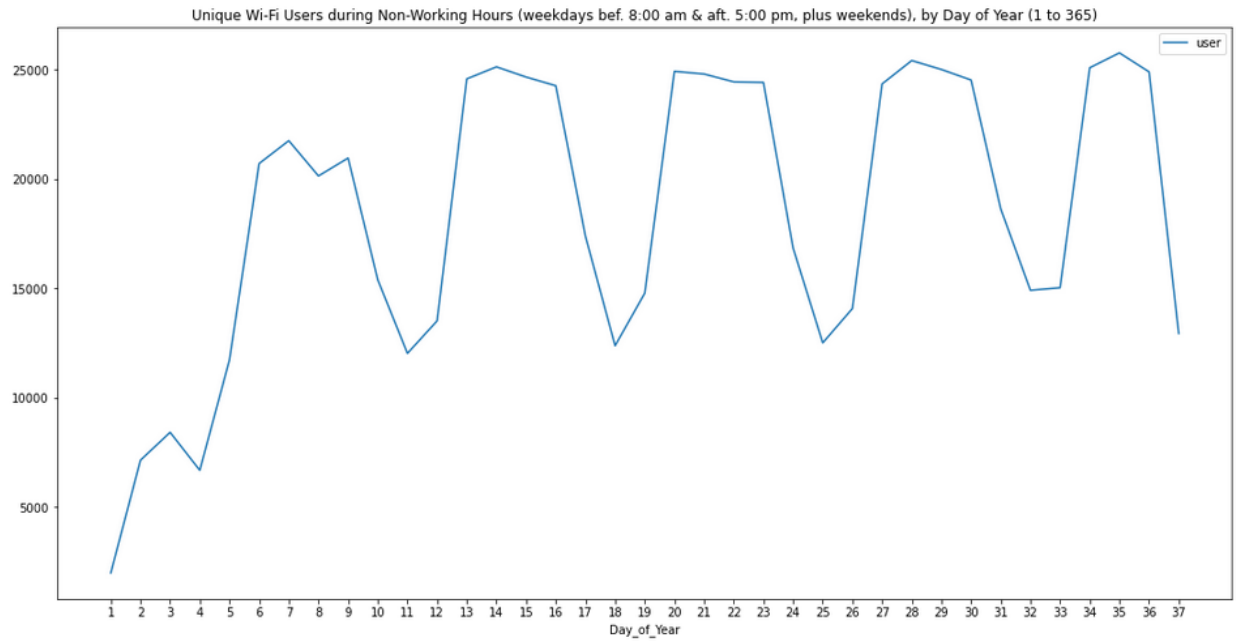

Figure 18: Number of unique users daily, during non-working hours.

The number of unique users daily, before 8:00 am and after 5:30 pm on weekdays and all day on weekends, peak close to 25,000 for the first 6 weeks of the semester. The pattern of users is consistent each week, with lower numbers on weekends. The data shows that student and employee activity on campus remains high outside of official hours, which could be interesting from a "student experience" perspective, if the university leadership chooses to investigate the scheduled and unscheduled activities (e.g., recreation activities, studying and homework) across campus.

5.3 Wireless Activity by Building

The number of wireless users by building varied with building size, as expected. The seven largest Academic and Academic Support buildings on the campus by gross area are, in order: DC, MC, QNC, E7, HH, BMH and PAS. Also, certain buildings (e.g., DC, BMH) serve as "gateways" to campus due to their proximity to public transit points (i.e., bus terminals, parking lots, major streets). The distribution of wireless connection counts by building code is displayed in Figure 19.

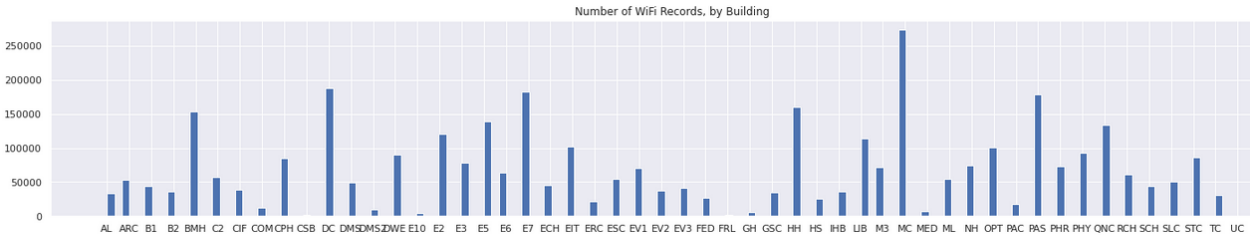

Figure 19: Number of Wi-Fi connection logs by Building Code.

The number of wireless connections in each building, during working hours in the period selected for analysis (weeks 2 through 5), varied widely for each building. Figure 20 shows the box and whisker plot for User counts per 30-minute period in the same buildings shown in Figure 19.

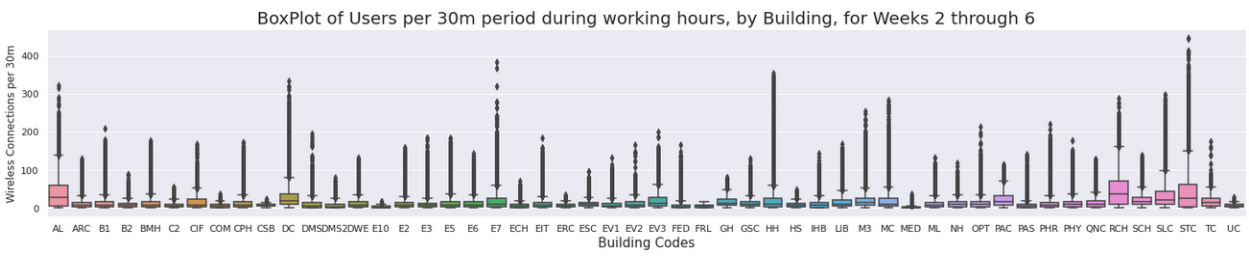

Figure 20: Box Plot of number of Wi-Fi Users per 30-minute period, by Building Code.

The buildings that were consistently occupied, as measured by the size of the interquartile range, were AL, RCH and STC. Other buildings had small interquartile ranges, relative to their overall range.

5.4 Wireless Activity by Building Floor

The buildings and floors with the 20 highest numbers of wireless connections over the 6-week period are shown in Table 25. The first floors of the Davis Centre (DC), Hagey Hall (HH), and Burt Matthews Hall (BMH) are ranked first through third; all are large buildings on the East, South and North boundaries of the campus. Nine of the entries on the list are Floor Code *01*. Fourteen (14) of the top twenty entries in the table are from six buildings: DC, E7, HH, MC, RCH, and STC.

Building_Alpha_Code	Floor_Code	Devices	Users	Users30	Users60
DC	01	1877091	1779724	1223965	668453
HH	01	1387121	1340745	1043643	579617
BMH	01	623870	604302	455279	264550
MC	04	631987	574147	431448	305292
STC	01	589218	549811	382966	221248
MC	02	588864	543366	431714	300526
AL	01	500042	480734	386564	230632
CIF	01	475921	469509	323798	129825
DC	02	455501	433920	307147	183927
STC	00	447702	419326	332323	219251
E7	04	434493	415078	334446	243842
MC	03	440008	411793	294492	182806
RCH	03	398347	381532	257133	173289
PHR	01	407887	366121	296218	87548
RCH	02	371635	348787	241577	126197
SLC	01	341022	328071	199177	61821
M3	01	335540	321634	254848	141796
HH	02	331163	318831	256934	166294
DC	03	323898	315107	235219	149416
E7	05	314608	301954	236457	168843

Table 25: Top twenty Building Floors for total number of wireless Users.

The number of unique wireless Users vary by Building Floor, as would be expected. For most campus buildings, the main points of entry/exit are on the ground floor (Floor *01*), so more wireless activity is

recorded there (Figure 21). Several buildings also have entrances and exits on the second floor, which results in higher traffic patterns.

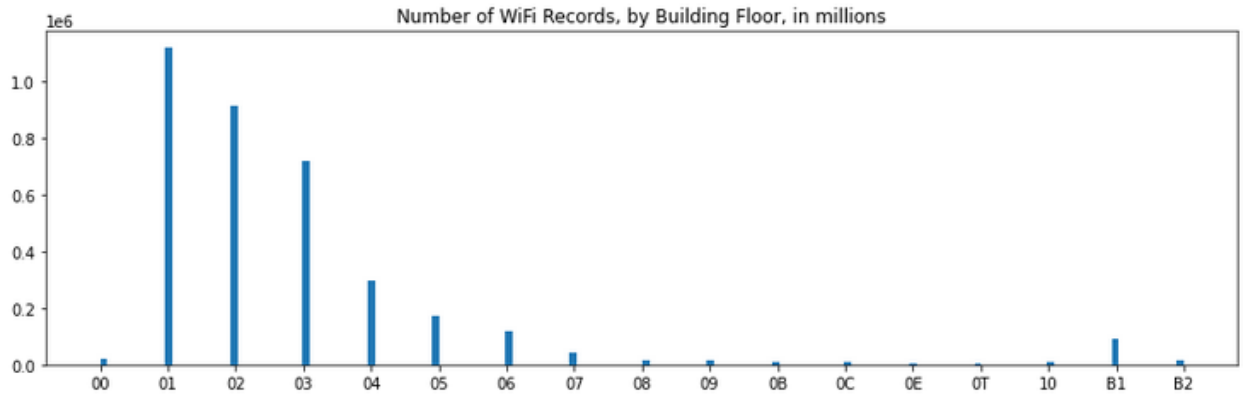

Figure 21: Number of Wi-Fi connection logs by Building Floor, in millions of records.

Many Building Floors had a wide variation in the number of Users during working hours. This can be displayed using a boxplot of the number of connections for each Building Floor. Records from the six (6) buildings that comprise the bulk of the top twenty list in Table 25 are displayed in Figure 22. Ground floors (Floor *01*) are excluded as many of the buildings are "gateway" structures through which many students and employees gain access to campus.

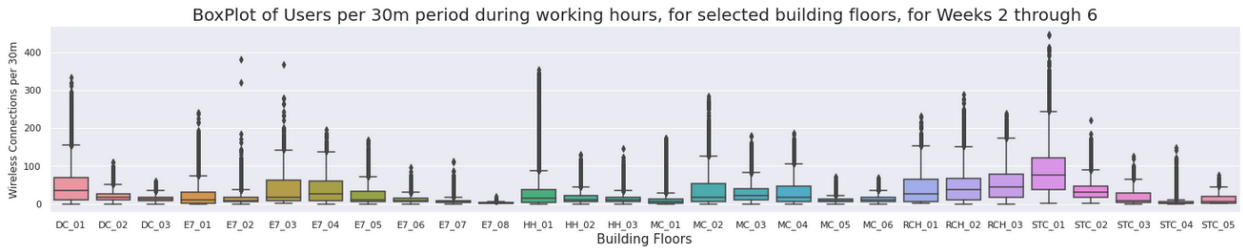

Figure 22: Boxplot of the number of Users by Building Floor, for Buildings DC, E7, HH, MC, RCH, STC during working hours.

An interesting question that is sometimes asked on university campuses is whether the physical assets (buildings) are being used to their full capacity, that is, can better use be made of the campus' space assets? One way to investigate this subject is to look at how the levels of occupancy in a building and floor vary over time. A Building Floor with a wide variation in occupancy levels may have periods when spaces are available for use.

5.5 Wireless Activity by Building Floor Zones

Wireless APs are deployed to provide overlapping, continual network connections for users. This means that often a user's Wi-Fi device may be within range of more than one AP. The AP chosen for a wireless connection depends on several factors, including signal strength, load balancing and interference. In certain circumstances, a device's connection may switch between APs while the individual remains stationary.

It is challenging to definitively determine a user's location on a floor, or in a room, based on the AP to which his/her/their wireless device is connected. APs may be installed in specific rooms on a Building Floor, where large numbers of students and/or employees are expected, such as lecture halls or laboratories; in larger rooms multiple APs may be present. APs may also be situated in hallways, each providing wireless coverage to multiple rooms.

Although single individuals using the wireless network are difficult to locate conclusively, it is possible to estimate a person's probable location with a high degree of confidence. In general, a wireless device will form a connection to an AP which provides a strong signal. The AP that provides the strongest signal is usually the closest AP, typically on the same floor and in the same vicinity as the user because of the attenuation of the wireless signal by occluding objects like walls.

While the probability of error for locating a single individual's floor location is non-zero, the probability of error for locating a larger group is smaller, per the Weak Law of Large Numbers ("*The weak law of large numbers essentially states that for any nonzero specified margin, no matter how small, there is a high probability that the average of a sufficiently large number of observations will be close to the expected value within the margin.*"). An indoor crowd's location, as suggested by connections to an Access Point, has a much higher probability of being in the AP's vicinity, than being in the vicinity of AP further away. That is, the odds of all/most of the individuals being somewhere other than in the AP's vicinity becomes vanishingly small, as the crowd size increases. This means that locating crowds in the vicinity of an AP is more precise than measuring individuals in the vicinity.

Since APs are deployed to provide overlapping coverage with limited interference, a zone for each AP (or, group of APs) can be defined spatially on a Building Floor. In the absence of GPS locations for each AP, a zone "proxy" can be defined using the typical/standard/average room numbers that are in the vicinity of the AP. Since nearby rooms have similar numbers, for example adjacent rooms are usually the next even or odd number in a sequence, and rooms opposite each other in a hallway are numbered consecutively, rooms within an APs coverage area usually begin with the same sequence of numbers. In many cases, rooms within an APs coverage area have the same room number, but different letter suffixes (see Figure 23).

The room and hallway locations of the APs on each floor were mapped to zones based on the room numbers, for APs installed in rooms, and the "average" room numbers of nearby rooms for APs installed in hallways. The total number of wireless Users for each Building Floor Zone, for each 30-minute period, was then calculated.

The classroom schedules and employee location datasets were used to calculate the number of expected occupants in each zone. Since all scheduled on-campus activity (e.g., lab) occurs in a room, mapping rooms with expected occupants to zones did not require determining an "average" room number.

For Building Floors that had four-digit room numbers, the zone was defined as the first three digits of the room numbers; for building with three-digit room numbers, only the first two digits were used. The algorithm for determining zone first stripped alphabetic suffixes from room numbers, then selected the first "x - 1" characters of the room numbers, where "x" was the length of the room number string, after alphabetic suffixes were removed.

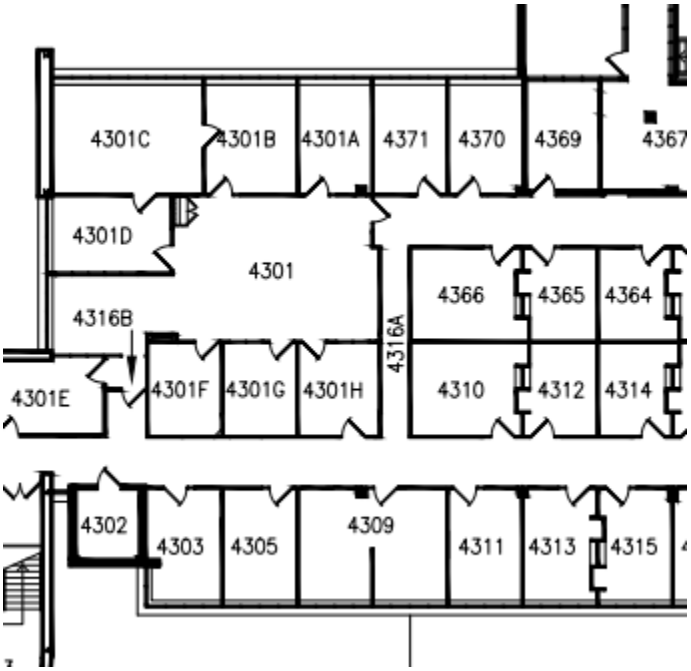

Figure 23: Sample floor layout, showing room numbers of adjacent rooms.

The number of wireless connections at an AP is a function of the number of users in the vicinity of that AP. Correlation Coefficients were calculated for the Users (wireless connections) vs expected occupants (students and employees) for each Building Floor, and for each Building Floor Zone, when expected occupants > 0.

Scheduled student and employee activity take place in Academic and Academic Support buildings, such as buildings containing offices, labs, and classrooms. Scheduled activities were defined by the classroom schedules and employee location datasets only; activities in areas such as athletics facilities, food courts, exhibition facilities were not captured.

Table 26 shows sample Correlation Coefficients for Users, Users30 and Users60 for various Building Floors for the period. Across campus, 115 floors in 42 buildings had scheduled activity. The number of data points for each measure is shown.

Building	Building_Floor	GrossNASMRatio	ConstructionYear	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparkli
AL	AL_00	2	12/1/2002	80	1.031	0.825	0.572	
AL	AL_01	2	9/1/1965	660	1.667	1.366	0.870	
AL	AL_02	2	9/1/1965	377	0.727	0.579	0.353	
ARC	ARC_01	1.5	9/1/1965	248	0.668	0.576	0.561	
ARC	ARC_02	1.5	9/1/2004	1212	1.389	1.210	1.068	
ARC	ARC_03	1.5	9/1/2004	504	1.194	0.884	0.568	
B1	B1_02	2.5	9/1/2004	984	1.373	0.970	0.391	
B1	B1_03	2.5	9/1/1964	168	1.147	0.953	0.826	
B2	B2_01	1.6	9/1/1964	96	0.833	0.614	0.346	
B2	B2_02	1.6	9/1/1967	380	8.324	5.855	4.145	
B2	B2_03	1.6	9/1/1967	404	1.361	1.089	0.679	
BMH	BMH_01	1.9	9/1/1967	1376	-0.940	-0.579	0.053	
BMH	BMH_02	1.9	9/1/1972	1500	1.933	1.543	0.808	
BMH	BMH_03	1.9	9/1/1972	508	0.234	0.354	0.234	
BMH	BMH_04	1.9	9/1/1972	380	14.012	11.382	7.772	
C2	C2_00	1.7	9/1/1972	380	7.053	4.584	2.995	
C2	C2_01	1.7	1/1/1971	380	6.776	4.926	2.455	
C2	C2_02	1.7	1/1/1971	760	1.068	0.981	0.839	
C2	C2_03	1.7	1/1/1971	377	10.729	8.024	4.926	
CIF	CIF_01	1.3	1/1/1971	380	9.389	7.963	4.934	
COM	COM_01	1.6	5/1/1987	380	1.443	1.296	0.659	
CPH	CPH_01	1.7	9/1/1966	878	1.863	1.466	0.902	
CPH	CPH_03	1.7	9/1/1972	440	1.105	0.943	0.693	
CPH	CPH_04	1.7	9/1/1972	398	1.740	1.540	1.315	
DC	DC_01	2	9/1/1972	1471	2.165	1.679	1.047	
DMS	DMS_01	1.8	3/1/1987	380	6.371	5.497	4.718	
DMS	DMS_02	1.8	12/1/2012	380	5.805	2.855	0.729	

Table 26: Correlation values for Users, Users30 and Users60 vs. Expected Users, for selected Building Floors.

The correlation values varied widely, likely reflecting the varying levels of activity on a Building Floor during operating hours. That is, students and employees are present on Building Floors during scheduled and unscheduled times- Building Floors that are high traffic areas, or have study areas for students, will have high crowd levels even when classes are not in session.

The Correlation values were calculated for User, User30 and User60 to determine if User60, a connection duration of at least 60 minutes, the typical length of a class period, was more closely correlated with Expected Users. The relative Correlation values for User, User30 and User60 are displayed in Table 26 using trend lines ("sparklines"). The Correlation values decrease for longer connection durations: The correlation for User60 was less than for User30, which was less than the correlation value for User.

The Correlation values were calculated for Building Floor Zones, for User, User30 and User60. Sample values are shown in Table 27. Table 27 shows the Trendlines for sample values for Correlation, displaying whether longer wireless connection durations correlate more closely with Expected Users.

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
AL	AL_00	AL_00_00	MixedUse	0.63	80	1.031	0.825	0.572	
AL	AL_01	AL_01_10	MixedUse	0.63	160	1.233	0.911	0.381	
AL	AL_01	AL_01_11	Classroom	0.04	276	1.566	1.263	0.808	
AL	AL_01	AL_01_12	Classroom	0.1	224	0.964	0.748	0.412	
AL	AL_02	AL_02_20	Classroom	0.26	102	-0.001	0.000	0.000	
AL	AL_02	AL_02_21	Classroom	0.1	275	1.176	0.936	0.572	
ARC	ARC_01	ARC_01_100	UndergradLab	0	96	0.318	-0.037	-0.528	
ARC	ARC_01	ARC_01_110	Classroom	0	152	0.590	0.394	0.677	
ARC	ARC_02	ARC_02_200	AcademicOffice	0	380	2.322	2.053	1.543	
ARC	ARC_02	ARC_02_201	AcademicOffice	0	380	4.324	3.368	2.355	
ARC	ARC_02	ARC_02_202	AcademicOffice	0	72	1.380	1.284	1.150	
ARC	ARC_02	ARC_02_210	Library	0	380	5.284	3.126	1.282	
ARC	ARC_03	ARC_03_300	UndergradLab	0	392	0.281	0.199	0.087	
ARC	ARC_03	ARC_03_310	UndergradLab	0.06	112	1.292	0.950	0.605	
B1	B1_02	B1_02_27	Classroom	0.34	224	1.326	0.942	0.395	
B1	B1_02	B1_02_28	ResearchLab	0	380	15.884	14.405	10.874	
B1	B1_02	B1_02_29	ResearchLab	0.19	380	8.808	7.261	4.682	
B1	B1_03	B1_03_37	UndergradLab	0	168	1.147	0.953	0.826	
B2	B2_01	B2_01_15	ResearchLab	0	96	0.833	0.614	0.346	
B2	B2_02	B2_02_25	ResearchLab	0	380	8.324	5.855	4.145	
B2	B2_03	B2_03_35	ResearchLab	0	404	1.361	1.089	0.679	
BMH	BMH_01	BMH_01_100	Classroom	0	104	2.073	1.591	0.946	
BMH	BMH_01	BMH_01_101	Classroom	0	232	1.553	1.425	1.322	
BMH	BMH_01	BMH_01_161	ResearchLab	0	380	8.016	3.739	1.266	
BMH	BMH_01	BMH_01_162	Classroom	0.1	124	1.824	1.527	1.085	
BMH	BMH_01	BMH_01_168	Classroom	0	380	112.403	88.097	47.384	
BMH	BMH_01	BMH_01_170	AcademicOffice	0	156	0.742	0.626	0.428	
BMH	BMH_02	BMH_02_230	AcademicOffice	0	380	10.068	6.282	4.100	

Table 27: Correlation values for Users, Users30 and Users60 vs. Expected Users, for selected Building Floor Zones.

The complete listing of values of Correlation, for all Building Floors and Building Floor Zones, are in Appendixes D and E. The trendlines indicate that longer periods of connection duration (e.g., User60 vs. User30, or User30 vs. User) create lower Correlation Coefficients for wireless connections vs. Expected Users.

Larger values for Coefficients of Correlation were found for Building Floor Zone values, when compared to Building Floor values. That is, most of the Coefficient values for Zones exceeded the Coefficient values for the entire floor. This suggests a stronger relationship between wireless connections and expected users was better explained by zones within a Building Floor, than for the entire floor.

Table 28 shows sample values for Coefficients of Correlation values for both Building Floor and Building Floor Zones. The full table of values is shown in Appendix F. 158 Zones had higher Coefficients of Correlation than the corresponding Floor, compared to 90 Zones that had a lower Coefficient. 45 Zones had the same Coefficient as the entire Floor- this occurred when the Floor had only one zone. Summary statistics are shown in Table 29.

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
E6	E6_02	2010	E6_02_202	AcademicOffice	256	256	0.380	0.380	-
E6	E6_04	2011	E6_04_402	Classroom	184	184	1.030	1.030	-
E7	E7_01	2011	E7_01_142	Classroom	380	380	44.210	44.210	-
E7	E7_02	2018	E7_02_230	Classroom	328	157	20.880	6.400	N
E7	E7_02	2018	E7_02_240	Classroom	328	171	20.880	25.840	Y
E7	E7_03	2018	E7_03_331	ResearchLab	1116	152	52.600	9.130	N
E7	E7_03	2018	E7_03_333	ResearchLab	1116	171	52.600	8.900	N
E7	E7_03	2018	E7_03_334	ResearchLab	1116	153	52.600	120.580	Y
E7	E7_03	2018	E7_03_335	ResearchLab	1116	155	52.600	184.090	Y
E7	E7_03	2018	E7_03_340	ResearchLab	1116	153	52.600	8.920	N
E7	E7_03	2018	E7_03_341	ResearchLab	1116	171	52.600	11.170	N
E7	E7_03	2018	E7_03_345	ResearchLab	1116	161	52.600	34.410	N
E7	E7_04	2018	E7_04_404	Classroom	1014	161	-17.080	177.280	Y
E7	E7_04	2018	E7_04_405	Classroom	1014	380	-17.080	81.710	Y
E7	E7_04	2018	E7_04_431	Classroom	1014	153	-17.080	9.070	Y
E7	E7_04	2018	E7_04_441	Classroom	1014	159	-17.080	65.320	Y
E7	E7_04	2018	E7_04_443	Classroom	1014	161	-17.080	62.210	Y
E7	E7_05	2018	E7_05_534	ResearchLab	1127	155	-29.910	167.090	Y
E7	E7_05	2018	E7_05_535	ResearchLab	1127	161	-29.910	78.570	Y
E7	E7_05	2018	E7_05_541	ResearchLab	1127	155	-29.910	7.530	Y
E7	E7_05	2018	E7_05_542	ResearchLab	1127	167	-29.910	13.310	Y
E7	E7_05	2018	E7_05_543	ResearchLab	1127	153	-29.910	7.180	Y
E7	E7_05	2018	E7_05_544	ResearchLab	1127	171	-29.910	3.610	Y
E7	E7_05	2018	E7_05_545	ResearchLab	1127	165	-29.910	9.380	Y
E7	E7_06	2018	E7_06_631	ResearchLab	805	169	10.380	8.680	N
E7	E7_06	2018	E7_06_640	ResearchLab	805	153	10.380	17.230	Y
E7	E7_06	2018	E7_06_641	ResearchLab	805	171	10.380	10.350	N
E7	E7_06	2018	E7_06_643	ResearchLab	805	159	10.380	4.790	N
E7	E7_06	2018	E7_06_644	ResearchLab	805	153	10.380	11.280	Y
E7	E7_07	2018	E7_07_731	AcademicOffice	1432	171	-3.760	3.020	Y
E7	E7_07	2018	E7_07_732	AcademicOffice	1432	152	-3.760	4.380	Y
E7	E7_07	2018	E7_07_733	AcademicOffice	1432	167	-3.760	9.570	Y
E7	E7_07	2018	E7_07_734	AcademicOffice	1432	153	-3.760	39.250	Y
E7	E7_07	2018	E7_07_740	AcademicOffice	1432	150	-3.760	4.620	Y

Table 28: Comparison of Coefficient of Correlation values for selected Building Floor and Building Floor Zones.

In general, larger values for Correlation Coefficients were found for Building Floor Zone values, when compared to Building Floor. Most of the Coefficient values for Zones exceeded the corresponding values for Building Floor (158 of 292). Higher values of Correlation Coefficients were found for almost all Cluster types, as shown in Table 29.

Many of the Building Floor Zones that had lower correlation coefficients (*N*) than the Building Floors they are situated in still have positive coefficients. Of the 90 Building Floor Zones (of 292) that had lower correlation coefficients, 81 had positive values. This suggests that 239 (158 + 81) zones had correlation coefficient values that suggest a strong relationship between expected occupants and wireless connections measured.

Coefficients of Correlation for Zones Compared to Floors				
Cluster Labels	Same	Lower	Higher	Total
AcademicOffice	10	15	24	49
AdministrativeOffice	4		7	11
Athletics	1			1
Bookstore		2		2
Classroom	12	18	24	54
CommonUse	1	2	6	9
FoodServices	1		4	5
HealthServices		2	3	5
Library	1	1	3	5
MixedUse	1	6	1	8
OtherSpace	1	1	6	8
PlantMaintenance			6	6
ResearchLab	6	17	34	57
StudySpace	1			1
UndergradLab	5	26	40	71
Totals	44	90	158	292

Table 29: Count of Building Floor Zones that had the same, lower, and higher Correlation Coefficients for User, by COU Cluster.

In Table 29, the number of Zones that have the same coefficients of correlation are areas on campus where only one zone was calculated on a Building Floor, i.e., all the rooms on the Building Floor have the same prefix.

The relative number of Building Floor Zones with higher Coefficients of Correlation, compared to Building Floors, did not change materially when the User30 and User60 duration periods were used as the basis for comparison (see Tables 30 and 30). This is likely because the User30 and User60 connections overlapped in the connection durations they represented, as discussed in section 4.4.3.

Coefficients of Correlation for Zones Compared to Floors (User30)				
ClusterNames	Same	Lower	Higher	Total
AcademicOffice	10	13	26	49
AdministrativeOffice	4	1	6	11
Athletics	1			1
Bookstore		2		2
Classroom	12	17	25	54
CommonUse	1	2	6	9
FoodServices	1	2	2	5
HealthServices		2	3	5
Library	1		4	5
MixedUse	1	4	3	8
OtherSpace	1	1	6	8
PlantMaintenance			6	6
ResearchLab	6	16	35	57
StudySpace	1			1
UndergradLab	5	32	34	71
Totals	44	92	156	292

Table 30: Count of Building Floor Zones that had the same, lower, and higher Correlation Coefficients for User30, by COU Cluster.

Coefficients of Correlation for Zones Compared to Floors (User60)				
ClusterNames	Same	Lower	Higher	Total
AcademicOffice	10	10	29	49
AdministrativeOffice	4	1	6	11
Athletics	1			1
Bookstore		2		2
Classroom	12	19	23	54
CommonUse	1	1	7	9
FoodServices	1	2	2	5
HealthServices		5		5
Library	1		4	5
MixedUse	1	4	3	8
OtherSpace	1	1	6	8
PlantMaintenance			6	6
ResearchLab	6	18	33	57
StudySpace	1			1
UndergradLab	5	34	32	71
Totals	44	97	151	292

Table 31: Count of Building Floor Zones that had the same, lower, and higher Correlation Coefficients for User60, by COU Cluster.

The higher correlation coefficients for Users vs expected occupants in Building Floor Zones, compared to Building Floor, is likely because the overall coefficient of correlation at the Building Floor level represents an averaging of the coefficients at each Building Floor Zone. Figure 24 shows the correlation between

expected occupants and Wireless Connections, for Building Floor *AL_01*. Three charts are displayed: one for each of Users, Users30, and Users60.

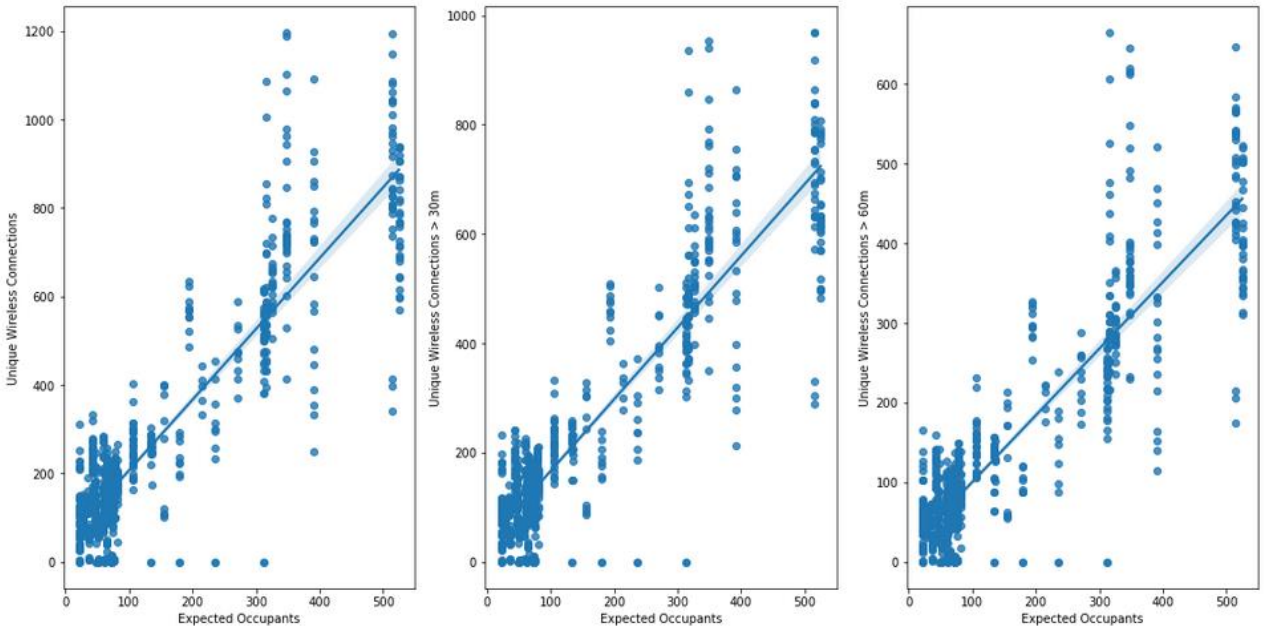

Figure 24: Correlation plots for Wi-Fi records (Users, Users30, Users60) vs. Expected Occupants from Building Floor AL_01.

The Coefficient of Correlation is positive for Users, Users30 and Users60, with highest values for Users vs. Expected Occupants.

Building Floor *AL_01* includes three zones, *AL_01_10*, *AL_01_11*, and *AL_01_12*. The correlation plot for each zone is shown in Figure 27. Three charts are displayed for each zone: one each for Users, Users30, Users60.

The charts in Figure 25 clearly display distinct activity in each of the three zones on the first floor of building AL. In Figure 24, the scatter plot for the entire floor is a summary of the activity occurring in all three zones. The zones provide more granular detail about the locations of persons on the first floor of AL.

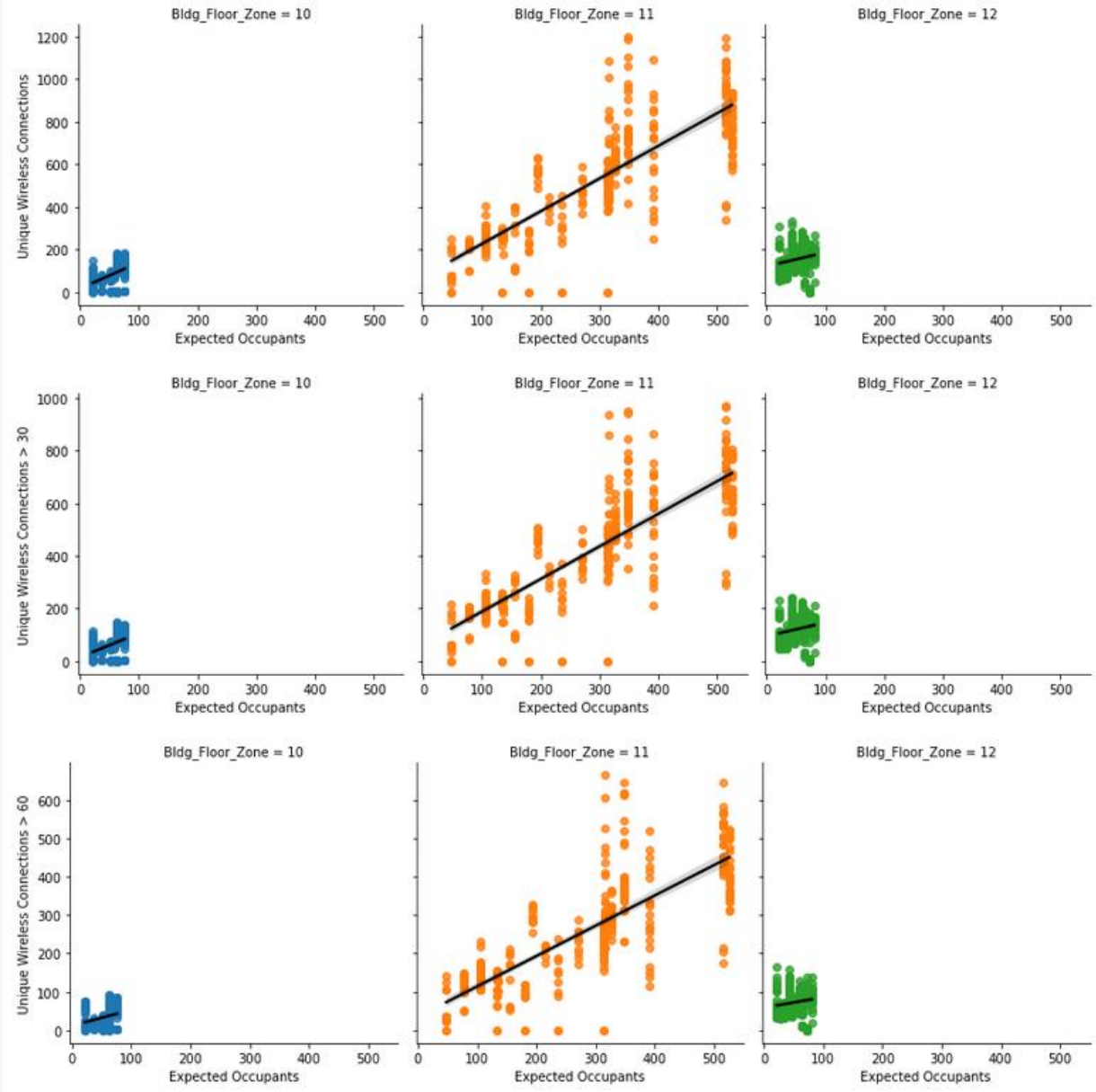

Figure 25: Correlation plots for Wi-Fi records (Users, Users30, Users60) vs. Expected Occupants from Building Floor AL_01, by individual Building Floor Zones 10, 11, 12.

5.6 Regression Analysis

Linear Regression analysis was performed for each Building_Zone to determine the most important feature variables for the class variable Users. The Users class variable represented the number of unique wireless network users for each 30-minute period, in each Building_Zone. The feature variables

considered were Expected_Employee_Count, Expected_Student_Count, Hour_and_Minute, and Day_of_Week.

1.715 million records, for 1250 Building_Zones, were used in the regression analysis. The value of Users for the 30-minute periods ranged from 0 to 2,006. Day_of_Week had a minimum of 2 (Monday) and maximum of 6 (Friday). The Hour_and_Minute variable ranged from 08.00 (8:00 am) to 19.50 (7:30 pm). Expected_Student_Count values spanned 0 to 668. Expected_Employee_Count ranged from 0 to 64. Sample records are displayed in Table 32.

Cluster_Names	Bldg_Zone	Day_of_Week	Hour_and_Min	Enrolment	Employment	Users
AssemblyExhibition	HH_01_15	3.00	9.00	31.00	0.00	62.00
AssemblyExhibition	HH_01_15	3.00	9.00	31.00	0.00	3.00
AssemblyExhibition	HH_01_15	3.00	9.00	31.00	0.00	64.00
AssemblyExhibition	HH_01_15	3.00	9.00	31.00	0.00	58.00
AssemblyExhibition	HH_01_15	5.00	9.00	31.00	0.00	69.00
AssemblyExhibition	HH_01_15	5.00	9.00	31.00	0.00	65.00
AssemblyExhibition	HH_01_15	5.00	9.00	31.00	0.00	3.00
AssemblyExhibition	HH_01_15	5.00	9.00	31.00	0.00	60.00
AssemblyExhibition	HH_01_15	2.00	10.00	28.00	0.00	32.00
AssemblyExhibition	HH_01_15	2.00	10.00	28.00	0.00	34.00
AssemblyExhibition	HH_01_15	2.00	10.00	28.00	0.00	43.00
AssemblyExhibition	HH_01_15	2.00	10.00	28.00	0.00	42.00
AssemblyExhibition	HH_01_15	3.00	10.00	23.00	0.00	70.00
AssemblyExhibition	HH_01_15	3.00	10.00	23.00	0.00	42.00
AssemblyExhibition	HH_01_15	3.00	10.00	23.00	0.00	99.00
AssemblyExhibition	HH_01_15	3.00	10.00	23.00	0.00	72.00
AssemblyExhibition	HH_01_15	4.00	10.00	28.00	0.00	35.00
AssemblyExhibition	HH_01_15	4.00	10.00	28.00	0.00	45.00
AssemblyExhibition	HH_01_15	4.00	10.00	28.00	0.00	42.00
AssemblyExhibition	HH_01_15	4.00	10.00	28.00	0.00	43.00
AssemblyExhibition	HH_01_15	5.00	10.00	23.00	0.00	76.00

Table 32: Sample records for Regression Analysis for Building_Zone HH_01_15

The relationship between the class variable User and each of the feature variables was non-linear. Figures 26, 27, and 28 display the pattern of User values in three distinct Building_Zones, for the weeks analyzed. Only working periods (hours and days) are shown.

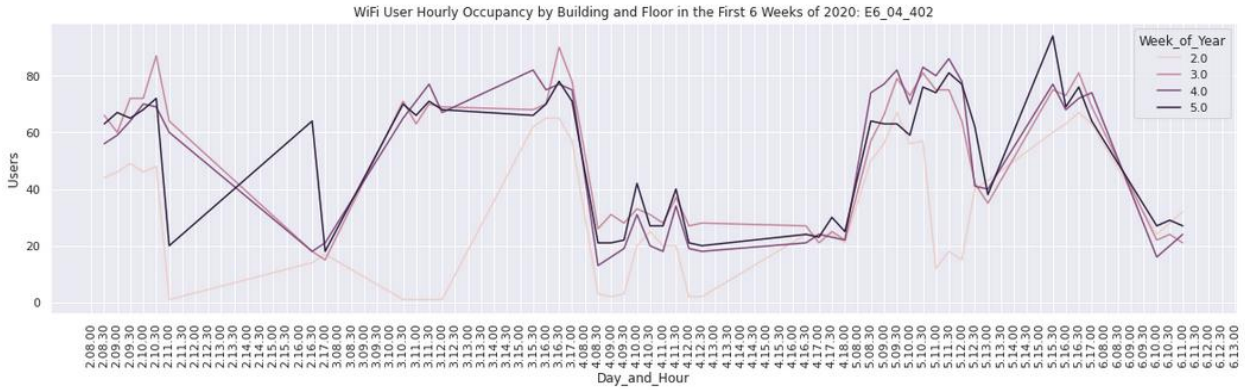

Figure 26: Wi-Fi Users by Day_and_Hour for Building Zone E6_04_402

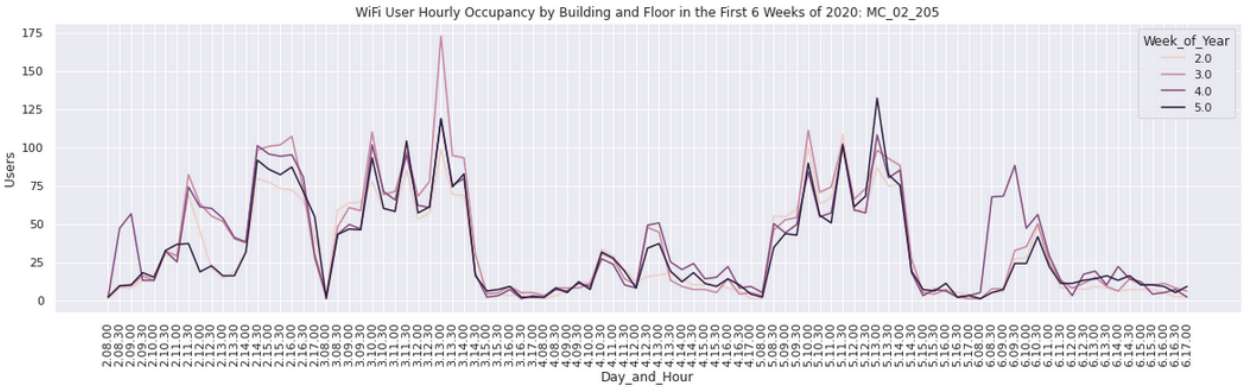

Figure 27: Wi-Fi Users by Day_and_Hour for Building Zone MC_02_205

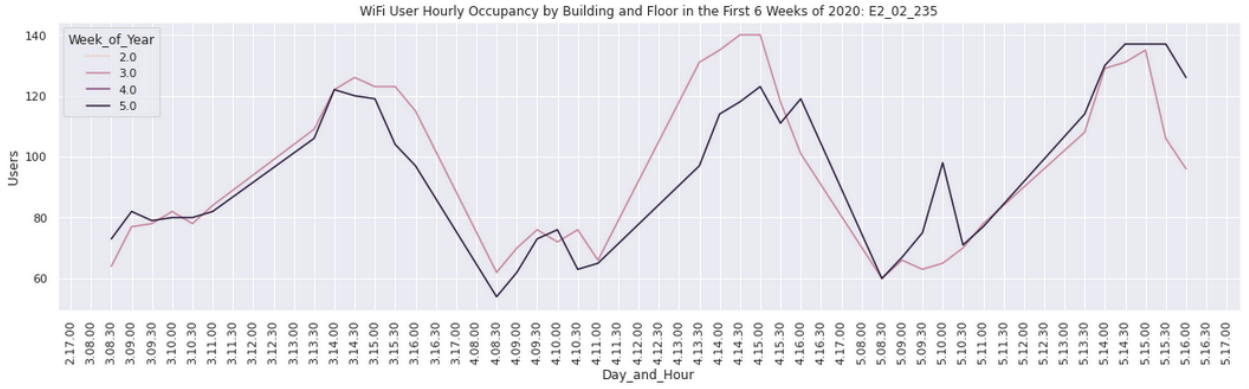

Figure 28: Wi-Fi Users by Day_and_Hour for Building Zone E2_02_235

The patterns of wireless usage for each Building_Zone were consistent week to week. The relative peaks and troughs occurred at approximately the same Day_and_Hour weekly.

Figures 29, 30, 31, and 32 display the pattern of Users in three distinct Building_Zones, by Expected_Student_Count or Expected_Employee_Count, for the weeks analyzed. Only working periods' values are shown:

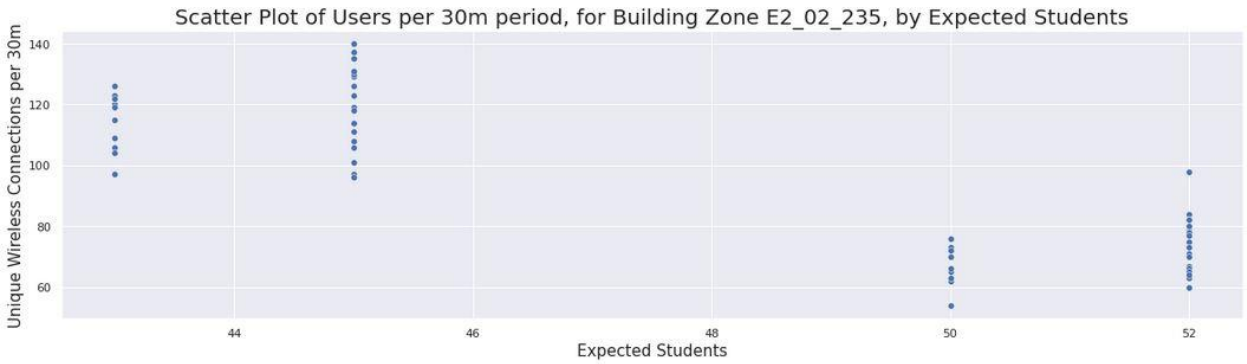

Figure 29: Wi-Fi Users by Expected Students for Building Zone E2_02_235

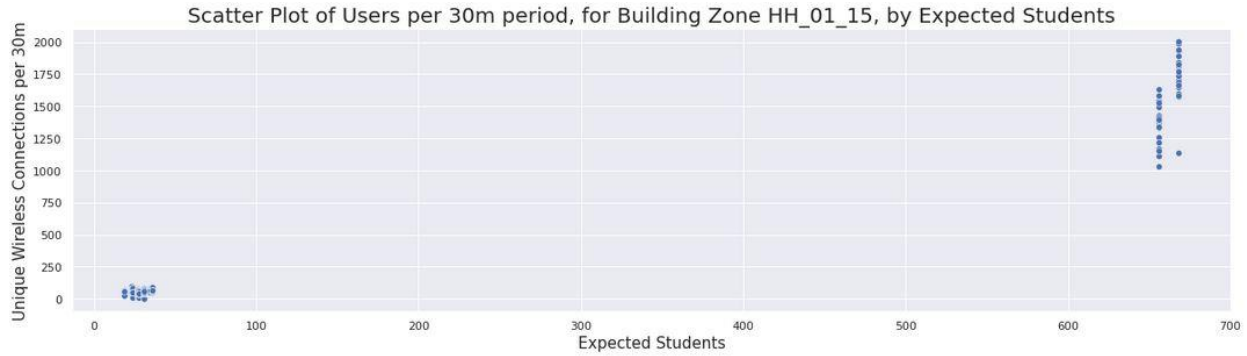

Figure 30: Wi-Fi Users by Expected Students for Building Zone HH_01_15

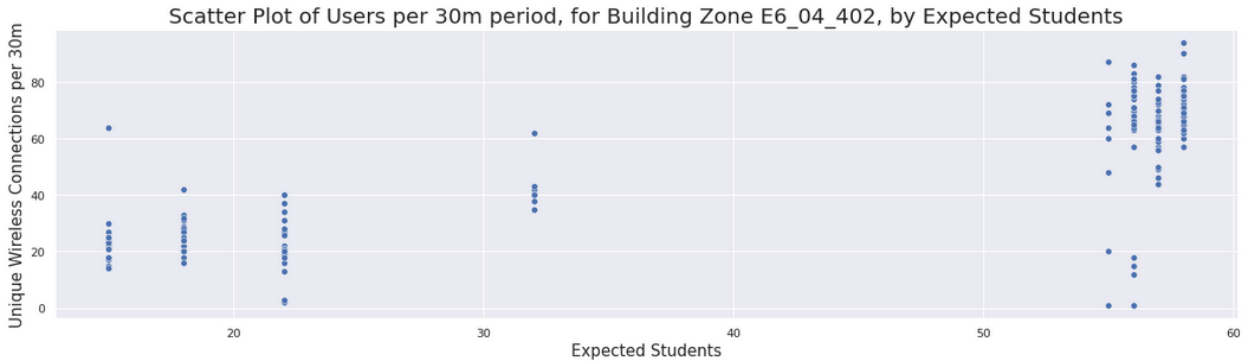

Figure 31: Wi-Fi Users by Expected Students for Building Zone E6_04_402

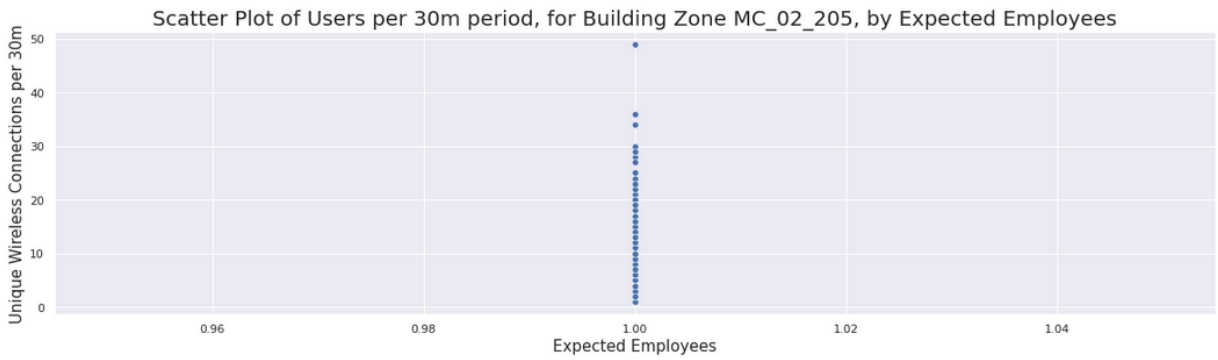

Figure 32: Wi-Fi Users by Expected Employees for Building Zone MC_02_205

The relationships between each feature variables and class variable differed for each Building_Zone, but was generally consistent week to week, in each Building_Zone. The value of Users was consistently high Mondays to Thursdays, and lower on Fridays. The values of Users during a single day tended to have higher values between 10:00 am and 2:00 pm. The values of Users had a positive correlation with Expected_Student_Count. There was also a positive correlation with Expected_Employee_Count.

The feature and class variables were normalized to values between 0 and 1, where 0 and 1 were the transformed minimum and maximum measured values for each variable. Sample rows are displayed in Table 33:

Cluster_Names	Bldg_Zone	Users_Norm	Day_of_Week_Norm	Hour_and_Min_Norm	Enrolment_Norm	Employment_Norm
AssemblyExhibition	HH_01_15	0.03	0.25	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.00	0.25	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.03	0.25	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.03	0.25	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.03	0.75	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.03	0.75	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.00	0.75	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.03	0.75	0.11	0.05	0.00
AssemblyExhibition	HH_01_15	0.02	0.00	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.00	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.00	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.00	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.03	0.25	0.22	0.03	0.00
AssemblyExhibition	HH_01_15	0.02	0.25	0.22	0.03	0.00
AssemblyExhibition	HH_01_15	0.05	0.25	0.22	0.03	0.00
AssemblyExhibition	HH_01_15	0.04	0.25	0.22	0.03	0.00
AssemblyExhibition	HH_01_15	0.02	0.50	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.50	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.50	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.02	0.50	0.22	0.04	0.00
AssemblyExhibition	HH_01_15	0.04	0.75	0.22	0.03	0.00

Table 33: Normalized Feature and Class variables for Building Zone HH_01_15

Multiple Linear Regression analysis was performed with the normalized feature variables: Day_of_Week_Norm, Hour_and_Min_Norm, Enrolment_Norm, Employment_Norm, and normalized class variable: Users_Norm. The regression analysis was conducted for each of the 1250 Building_Zone values. Records being analyzed were limited to data collected during university operating hours: Monday to Friday, 8:00 a.m. to 7:30 p.m. Python's Statsmodels Ordinary Least Squares (sm.OLS) functions were used in the analysis.

R-Squared was calculated for all possible combinations of the four feature variables. There were 15 combinations in total: 1 with all four variables, 4 with three variables, 6 with two variables; 4 with 1 variable. The varied combinations of feature variables were chosen to identify the most important factors influencing the value of the class variable, in each Building Zone.

The calculated values of R-Squared for each Building_Zone ranged from <undefined> to 0.975. Table 34 displays the R-Squared values of the first thirty Building_Zone values, for each of the fifteen combinations of feature variables. The feature variables are coded as D, H, S, E, where D = Day_of_Week_Norm, H = Hour_and_Minute_Norm, S = Expected_Enrolment_Norm, E = Expected_Employment_Norm.

COU_Category	Building_Zone	R2_DHSE	R2_DHS	R2_DHE	R2_DSE	R2_HSE	R2_DH	R2_DS	R2_DE	R2_HS	R2_HE	R2_SE	R2_D	R2_H	R2_S	R2_E
MixedUse	AL_00_00	0.294	0.294	0.194	0.268	0.243	0.194	0.268	0.088	0.243	0.036	0.134	0.088	0.036	0.134	0.000
MixedUse	AL_01_10	0.233	0.233	0.069	0.157	0.233	0.069	0.157	0.000	0.233	0.069	0.157	0.000	0.069	0.157	0.000
Classroom	AL_01_11	0.440	0.440	0.006	0.430	0.440	0.006	0.430	0.002	0.440	0.004	0.429	0.002	0.004	0.429	0.000
Classroom	AL_01_12	0.193	0.193	0.098	0.093	0.189	0.098	0.093	0.005	0.189	0.091	0.090	0.005	0.091	0.090	0.000
Classroom	AL_02_20	0.055	0.055	0.005	0.053	0.049	0.005	0.053	0.002	0.049	0.002	0.047	0.002	0.002	0.047	0.000
Classroom	AL_02_21	0.669	0.669	0.096	0.665	0.665	0.096	0.665	0.047	0.665	0.046	0.659	0.047	0.046	0.659	0.000
UndergradLab	ARC_01_100	0.936	0.931	0.936	0.936	0.936	0.012	0.930	0.936	0.930	0.936	0.936	0.005	0.007	0.929	0.936
Classroom	ARC_01_110	0.457	0.457	0.441	0.452	0.048	0.441	0.452	0.412	0.048	0.029	0.001	0.412	0.029	0.001	0.000
AcademicOffice	ARC_02_200	0.756	0.092	0.756	0.668	0.752	0.092	0.004	0.668	0.088	0.752	0.664	0.004	0.088	0.000	0.664
AcademicOffice	ARC_02_201	0.037	0.008	0.037	0.030	0.036	0.008	0.001	0.030	0.007	0.036	0.029	0.001	0.007	0.000	0.029
AcademicOffice	ARC_02_202	0.871	0.865	0.662	0.871	0.871	0.018	0.865	0.662	0.864	0.657	0.871	0.001	0.017	0.864	0.657
Library	ARC_02_210	0.240	0.008	0.240	0.234	0.238	0.008	0.002	0.234	0.006	0.238	0.232	0.002	0.006	0.000	0.232
UndergradLab	ARC_03_300	0.399	0.304	0.338	0.399	0.399	0.024	0.303	0.338	0.301	0.337	0.399	0.004	0.021	0.300	0.336
UndergradLab	ARC_03_310	0.186	0.186	0.029	0.168	0.175	0.029	0.168	0.011	0.175	0.018	0.157	0.011	0.018	0.157	0.000
Classroom	B1_02_27	0.884	0.884	0.785	0.884	0.884	0.004	0.884	0.784	0.884	0.785	0.884	0.002	0.001	0.884	0.784
ResearchLab	B1_02_28	0.028	0.028	0.028	0.000	0.027	0.028	0.000	0.000	0.027	0.027	0.000	0.000	0.027	0.000	0.000
ResearchLab	B1_02_29	0.019	0.019	0.019	0.003	0.016	0.019	0.003	0.003	0.016	0.016	0.000	0.003	0.016	0.000	0.000
UndergradLab	B1_03_37	0.825	0.775	0.500	0.822	0.825	0.058	0.764	0.494	0.774	0.497	0.822	0.001	0.056	0.763	0.491
ResearchLab	B2_01_15	0.656	0.656	0.656	0.655	0.655	0.116	0.655	0.655	0.655	0.655	0.655	0.019	0.101	0.654	0.654
ResearchLab	B2_02_25	0.004	0.004	0.004	0.000	0.004	0.004	0.000	0.000	0.004	0.004	0.000	0.000	0.004	0.000	0.000
ResearchLab	B2_03_35	0.444	0.441	0.148	0.444	0.444	0.021	0.440	0.136	0.440	0.146	0.444	0.002	0.020	0.440	0.135
Classroom	BMH_01_100	0.246	0.246	0.012	0.226	0.245	0.012	0.226	0.009	0.245	0.000	0.225	0.009	0.000	0.225	0.000
Classroom	BMH_01_101	0.874	0.865	0.818	0.874	0.874	0.002	0.865	0.818	0.865	0.818	0.874	0.002	0.000	0.865	0.818
ResearchLab	BMH_01_161	0.001	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.000	0.000
Classroom	BMH_01_162	0.855	0.855	0.696	0.855	0.855	0.034	0.854	0.692	0.855	0.691	0.855	0.018	0.016	0.854	0.687
Classroom	BMH_01_168	0.510	0.014	0.510	0.510	0.496	0.014	0.014	0.510	0.000	0.496	0.496	0.014	0.000	0.000	0.496
AcademicOffice	BMH_01_170	0.696	0.664	0.647	0.696	0.696	0.003	0.664	0.646	0.661	0.647	0.696	0.001	0.002	0.661	0.646
AcademicOffice	BMH_02_230	0.002	0.002	0.002	0.000	0.002	0.002	0.000	0.000	0.002	0.002	0.000	0.000	0.002	0.000	0.000
ResearchLab	BMH_02_240	0.664	0.663	0.584	0.663	0.660	0.017	0.663	0.578	0.659	0.578	0.659	0.009	0.009	0.658	0.572
AcademicOffice	BMH_02_268	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ResearchLab	BMH_02_269	0.057	0.057	0.057	0.000	0.057	0.057	0.000	0.000	0.057	0.057	0.000	0.000	0.057	0.000	0.000
Classroom	BMH_02_270	0.784	0.755	0.766	0.782	0.784	0.003	0.754	0.762	0.754	0.766	0.782	0.003	0.001	0.754	0.762
AcademicOffice	BMH_03_302	0.849	0.815	0.834	0.849	0.848	0.001	0.814	0.834	0.815	0.834	0.848	0.000	0.000	0.814	0.834

Table 34: Sample R-Squared values for varied combinations of Feature variables

Table 35 displays the top R-Squared values for all Building_Zone values. The table has highlighted values for the (joint) highest R-Squared values for each of the 4-feature, 3-feature, 2-feature, and 1-feature combinations of variables.

COU_Category	Building_Zone	R2_DHSE	R2_DHS	R2_DHE	R2_DSE	R2_HSE	R2_DH	R2_DS	R2_DE	R2_HS	R2_HE	R2_SE	R2_D	R2_H	R2_S	R2_E
AssemblyExhibition	HH_01_15	0.975	0.975	0.208	0.975	0.975	0.013	0.975	0.198	0.975	0.208	0.975	0.002	0.011	0.975	0.198
UndergradLab	ESC_03_34	0.969	0.967	0.954	0.969	0.969	0.055	0.967	0.954	0.967	0.954	0.969	0.000	0.055	0.967	0.954
AssemblyExhibition	ML_02_25	0.967	0.953	0.878	0.967	0.967	0.008	0.953	0.878	0.953	0.873	0.967	0.000	0.007	0.952	0.873
UndergradLab	E2_02_236	0.964	0.945	0.924	0.964	0.963	0.016	0.945	0.922	0.945	0.923	0.963	0.000	0.016	0.945	0.922
UndergradLab	ESC_01_14	0.943	0.835	0.942	0.943	0.943	0.075	0.835	0.942	0.833	0.942	0.943	0.025	0.054	0.833	0.942
AcademicOffice	E2_03_333	0.941	0.935	0.941	0.941	0.941	0.000	0.935	0.940	0.935	0.941	0.941	0.000	0.000	0.935	0.940
UndergradLab	ARC_01_100	0.936	0.931	0.936	0.936	0.936	0.012	0.930	0.936	0.930	0.936	0.936	0.005	0.007	0.929	0.936
Classroom	PAS_01_124	0.916	0.910	0.874	0.916	0.916	0.098	0.909	0.868	0.910	0.873	0.916	0.064	0.038	0.909	0.868
UndergradLab	ECH_01_123	0.913	0.913	0.348	0.832	0.910	0.348	0.832	0.204	0.910	0.213	0.822	0.204	0.213	0.822	0.000
ResearchLab	EIT_05_500	0.913	0.001	0.913	0.912	0.913	0.001	0.000	0.912	0.001	0.913	0.912	0.000	0.001	0.000	0.912
AcademicOffice	PAS_04_403	0.912	0.896	0.885	0.912	0.912	0.005	0.896	0.885	0.896	0.885	0.912	0.004	0.002	0.896	0.885
UndergradLab	EIT_01_100	0.912	0.912	0.306	0.911	0.912	0.005	0.911	0.305	0.911	0.304	0.911	0.002	0.003	0.911	0.303
AcademicOffice	PHY_03_35	0.906	0.906	0.906	0.906	0.906	0.002	0.906	0.906	0.906	0.906	0.906	0.001	0.000	0.906	0.906
AcademicOffice	HH_02_25	0.904	0.887	0.816	0.904	0.904	0.010	0.887	0.816	0.887	0.815	0.904	0.002	0.009	0.887	0.815
AcademicOffice	STC_02_203	0.903	0.001	0.903	0.903	0.902	0.001	0.001	0.903	0.000	0.902	0.902	0.001	0.000	0.000	0.902
UndergradLab	E2_03_335	0.903	0.874	0.894	0.902	0.903	0.006	0.872	0.893	0.874	0.894	0.901	0.000	0.006	0.871	0.893
Classroom	ML_03_34	0.897	0.834	0.833	0.897	0.896	0.008	0.834	0.833	0.834	0.832	0.896	0.003	0.005	0.834	0.832
AcademicOffice	EIT_02_205	0.894	0.781	0.853	0.894	0.894	0.016	0.781	0.851	0.781	0.853	0.894	0.000	0.016	0.781	0.851
AcademicOffice	E2_02_235	0.893	0.893	0.197	0.891	0.893	0.000	0.891	0.196	0.893	0.197	0.891	0.000	0.000	0.891	0.196
MixedUse	PHY_03_30	0.888	0.888	0.611	0.886	0.843	0.611	0.886	0.003	0.843	0.592	0.805	0.003	0.592	0.805	0.000
UndergradLab	ML_01_11	0.887	0.887	0.755	0.886	0.887	0.011	0.886	0.754	0.887	0.752	0.886	0.007	0.004	0.886	0.752
Classroom	OPT_03_34	0.886	0.880	0.885	0.885	0.885	0.007	0.880	0.884	0.879	0.884	0.884	0.007	0.000	0.879	0.883
AdministrativeOffice	GSC_01_110	0.885	0.004	0.885	0.883	0.884	0.004	0.001	0.883	0.003	0.884	0.881	0.001	0.003	0.000	0.881
Classroom	B1_02_27	0.884	0.884	0.785	0.884	0.884	0.004	0.884	0.784	0.884	0.785	0.884	0.002	0.001	0.884	0.784
Classroom	E5_03_310	0.883	0.882	0.387	0.883	0.882	0.090	0.882	0.352	0.881	0.387	0.882	0.001	0.089	0.881	0.352
Classroom	E6_04_402	0.882	0.881	0.700	0.880	0.881	0.001	0.878	0.700	0.880	0.700	0.879	0.001	0.000	0.878	0.699
Classroom	PHY_03_31	0.882	0.881	0.740	0.881	0.882	0.016	0.881	0.737	0.881	0.740	0.881	0.001	0.015	0.881	0.737
AcademicOffice	HH_01_11	0.882	0.874	0.817	0.882	0.881	0.009	0.874	0.816	0.874	0.816	0.881	0.005	0.004	0.874	0.815
Classroom	ML_02_24	0.881	0.880	0.811	0.881	0.881	0.015	0.880	0.809	0.880	0.811	0.881	0.006	0.010	0.880	0.809

Table 35: The Highest R-Squared values in the Regression Analysis, highlighting the most important Feature Variable combinations.

Table 35 shows that the highest R-Squared values for some Building_Zone values (e.g., HH_01_15, PHY_03_25) are constant for all feature variable combinations, suggesting that only one of the feature variables is important for the regression analysis (i.e., R2_S: Students). Some Building_Zone values (e.g., ECH_01_123, ML_03_34, EIT_02_205, PHY_03_30) show lower R-squared values with fewer feature variables, indicating that a stronger correlation results from a combination of feature variables.

All the Building_Zone values in Table 35 have their strongest correlation with only two of feature variables: Expected_Enrolment_Norm (i.e., Expected Students Normalized) and Expected_Employment_Norm (i.e., Expected Employees Normalized). This can be seen in the table where the highest values in are highlighted in green.

One Building_Zone, PHY_03_35, has equal R-Squared for both Expected_Enrolment_Norm and Expected_Employment_Norm: 0.906. Further investigation revealed that the equivalent values were coincidental.

The regression analysis indicated that, during university working hours, the most important variables for predicting the number of wireless users in each Building Zone were Expected_Enrolment_Norm (based on class enrollments) and Expected_Employment_Norm (based on faculty and staff office assignments).

No regression analysis was completed for non-working hours, which likely explains the low R-Squared values for variables Day_of_Week and Hour_and_Minute. Another likely explanation is the non-linear relationship between Users and the time variables Day_of_Week and Hour_and_Minute. Further investigations could include “one-hot encoding” for those two class variables, or the use of a transformation function to make the relationship between the variables linear.

6.0 Discussion

6.1 The University of Waterloo Context

The University of Waterloo is one of the largest universities in Canada, with over 42,000 students. It is situated in the 10th largest metropolitan area in the country, the Region of Waterloo, but most incoming students hail from outside the Region, primarily the Greater Toronto Area. The institution has a high percentage of international students. Waterloo also has many residence beds, relative to the student population size- first year students are guaranteed on-campus residence.

The large proportion of the student population that hails from outside the Region, plus the numerous on-campus residences, mean that the main campus sees significant activity, both during and after working hours. University buildings host both academic and social activities, and the campus is often described as a “small city”.

The university is open year-round, with upwards of 25,000 students registered during spring/summer terms. Over 60% of students are registered in co-op programs, where academic terms are interspersed with work terms during which students work full-time for employers to gain real-world experience. Most work locations are outside the Region, which means that many students change their city of residence multiple times during the year.

The year-round operation, highly mobile student population, high percentage of out-of-town students, and main campus as a hub of activity combine to contribute to a unique culture at the institution. Some full-time students may spend up to half of their undergraduate career outside the Region, and social connections to classmates may take longer to develop.

Understanding the culture of the university is important to administrators for marketing, retention, health and wellness, student and alumni engagement, and strategic planning, among other concerns.

The major elements of Culture are understood to be symbols, language, norms, values, artifacts. At the University of Waterloo, departments interested in understanding and influencing campus culture are continually investigating new ways of understanding student attitudes and behaviour. Tools used include surveys, social media analyses, and academic research projects.

One underutilized source of student activity data is Wi-Fi logs. The campus’ extensive Wi-Fi infrastructure and extensive use of Wi-Fi-enabled devices could be leveraged to understand population activities. If a population can be understood/defined in terms of {identities, count, location, activity, timing} [17,25] then Wi-Fi logs can be a rich source of research data to investigate student behaviour.

6.2 COVID-19 Considerations

The university's high percentage of international and out-of-town students may also be an issue for monitoring vaccinations and the potential impact of the virus. Students who move cities for co-op work opportunities also are part of Waterloo's highly mobile community.

Given the large numbers of students who travel out of the Region, it may be useful to understand student movement on campus, to limit the potential spread of illness. Potential hotspots could be identified based on crowd levels, or persons entering prohibited spaces could be automatically notified that an area has been declared off-limits for given periods.

Some researchers have suggested using Wi-Fi logs for contact tracing, but the utility of the data for that purpose is very limited. Wi-Fi logs data from typical AP deployments cannot be used to determine if two persons who had connections to the same AP were in proximity.

6.3 Data Processing Challenges

The six datasets used in this research project were generated from five administrative systems tracking the university's class schedules, building records, telephone extensions, Wi-Fi logs, and AP deployments. Differing definitions and value sets were used in each system for record-keeping.

Building codes and names differed in some of the systems. Some used three-letter building abbreviations while others used two. In some cases, buildings had been renamed, but had not been updated in all administrative systems. The definition of a "building" also differs between departments- some contiguous spaces are considered single buildings by one department, but separate named buildings by others.

Floor codes differed in some buildings- basements could be designated "B1" in one system, and "00" in another. Some floor levels, recorded in some administrative systems, were not present in others- usually because the rooms on those floors were for maintenance needs and not accessible to the general population.

For rooms, leading zeroes were an issue for room numbers, with some systems making use of them, and others not. In most circumstances this was not an issue, since the data was summarized at a level of Building Floor Zone, but occasionally lookup/matching functions failed because of inexact matches.

The separate administrative systems are not updated at the same time, so some systems have out of date information, where major renovations had taken place. For example, walls may have been moved, rooms renumbered, or areas of a building made inactive/unusable. In those circumstances, the SPO's administrative system was considered the system of record for building information.

Data errors were present in a few systems, including misspellings, duplicates, and missing data. Obvious errors were corrected manually using official university systems of record. Other errors were deleted from the dataset. Many of the errors were associated with older buildings, not within the Ring Road "circle". Since the research focused on Academic and Academic Support buildings on the main campus, most of which are located within Ring Road, these errors were not material to the analysis.

Half-hour rounding of the session logs' start and end times meant that matching blending datasets could be done for 30-minute periods only- correlating occupancy counts would be simpler with start and end times rounded to 5-minute periods, at most. Half-hour rounding also increased the degree of data processing, as it introduced duplicates in the data as individuals moved through a building and their connections to APs switched.

The most significant data processing challenge involved reading the millions of rows of Wi-Fi logs. On average, over 4 million log records were created weekly during the first six weeks of 2020. Reading and transformation of the Wi-Fi logs was done on an Engineering Computing Linux server used for long-running data science tasks.

6.4 Findings

Overall, the crowd patterns identified in the project were consistent with researcher expectations. Buildings considered "gateways" to campus, i.e., close to transportation hubs, had the most traffic. Also, spaces built for circulation and people traffic- foyers, courts, atria- also had high person counts.

Cluster analysis of the 800 unique values of Building Floor Zones space data suggested several distinct categories (19) very close to the number of space groupings defined by the government (20). This could be explained by the fact that two of the government categories (Health Science Clinical Facilities, Animal Space) are not well represented at the University of Waterloo; those two other areas (i.e., Food Services, Bookstores) are co-located with large circulation areas; and each Building Floor Zone generally consists of one type of space (e.g., Classrooms).

Wi-Fi activity in academic and academic-support buildings was strongly correlated with employee locations and class schedules during working hours, as expected. During working hours, the number of users measured by session logs far exceeded the expected number of students and employees. This likely meant that students were present in spaces during times they had no classes- perhaps lounging between classes or using empty classrooms for studying and socializing.

If only students and employees who remained in a Building Floor Zone space for at least an hour was counted, then there was a closer correlation with expected occupancy. This would be expected since class times are typically an hour in length.

There was significant activity in buildings after hours and on weekends.

Predictive analysis of person counts in Academic and Academic Support buildings had a low accuracy, based on the feature variables chosen. This was likely due to the presence of persons in a Building Floor Zone who did not have scheduled classes, i.e., additional persons who were using empty classrooms or remaining in circulation spaces between classes.

6.5 Limitations

The research findings were limited by the accuracy and completeness of the source datasets, and the difficulties in combining the different data files.

The Wi-Fi logs' Start Times and End Times were rounded to the nearest half-hour, which made it difficult to distinguish between users with short and long durations of connection to specific APs. Many records had identical Start Times and End Times, which meant that the true period of connection to an AP could have been anywhere from 1 to 29 minutes (e.g., duration 11:43 am to 11:44 am would have Start Time = 11:30 a.m. and End Time = 11:30 a.m.; duration 11:46 a.m. to 12:14 p.m. would have a Start Time = 12:00 p.m. and End Time = 12:00 p.m.); Start Times and End Times that differed by 30 minutes, could represent duration periods of anywhere from 2 to 58 minutes (e.g., duration 11:44 am to 11:46 am would have Start Time = 11:30 a.m. and End Time = 12:00 p.m.; duration 11:46 a.m. to 12:44 p.m. would have a Start Time = 12:00 p.m. and End Time = 12:30 p.m.). Difficulty is clearly distinguishing between connections of short and long durations limited the correlation analysis between the connection logs and class schedules, since classes are typically 1 hour long.

The rounding of Wi-Fi logs' Start Times and End Times also meant that some duplicate records were created when individuals connected to more than one AP within half-hour periods. An individual who connected, in turn, to nearby APs while traversing a space would have the same Start and End Times at each AP, if their total journey lasted less than 30 minutes. Duplicate records were removed from the wireless logs dataset before processing.

The Wi-Fi logs datasets were also incomplete. The recording of Wi-Fi connections was sometimes inconsistent, leading to missing data at certain APs. In some cases, very little data was collected from some buildings, or no data was collected from specific APs in buildings. The data was still sufficient to perform analyses, but there were limitations to the precision and accuracy measurements gleaned from the analysis.

Missing and/or incomplete data was also an issue in other datasets, e.g., employee office locations, AP installations, floor layouts. The reasons for the errors and omissions varied, including systems of records not being updated with the latest building changes, data entry typos, and illegible documents. In some cases, the missing/incomplete/incorrect data was corrected by cross-referencing with other official sources, or by manual verification (e.g., visiting a Building Floor to get the true physical location of an AP).

Finally, combining the six datasets involved making a few assumptions about the data. One important example involves students and/or employees that carry more than one Wi-Fi-enabled device: the Wi-Fi logs record both devices' connections, but there wasn't a simple way of distinguishing between a mobile device (e.g., a smartphone or tablet) or a laptop. In those cases, one device was chosen for everyone (the "maximum" value of their multiple Device Ids) to represent their activity across campus, which may not have correlated with the actual device on their person at any given time.

Another example of data assumptions involved shared rooms and spaces. Some building rooms on campus are used for multiple functions and are shared by two or three departments; their COU

Category classifications vary as well. In those cases, the categorization used by the largest user of the shared space was used to classify the room type.

The missing and/or incomplete data did not significantly impact the data analysis: the broad patterns and trends displayed in the analysis were still obvious, but the level of accuracy was lowered. The analysis of datasets from different sources, and the combining of those records, typically requires the use of techniques that are not overly sensitive to missing and/or incomplete data, i.e., methods that can embrace “messy” data and still produce meaningful results.

6.6 Future Analysis

Further research into the use of session logs could involve the use of more granular data, and additional datasets to provide supporting information.

Session logs with greater temporal granularity would also enable more accurate analyses. The rounding of Start Times and End Times to the nearest half hour limited the possible analysis- it meant that corroborating analyses of employee working hours and students’ class schedules could only be considered in 30-minute periods as well. For example, many university classes are scheduled for 50 minutes Monday, Wednesday, Friday; or for 1 hr. 15 mins. Tuesdays and Thursdays; or for 2 hrs. 30 mins. weekly.

The use of Building Floor zone as the smallest identifiable area of a building could be further refined using GPS coordinates, centered on AP locations. The use of the first two or three characters to define a Building Floor zone is inaccurate on Building Floors where the room numbers of adjacent rooms are not continuous, this happens occasionally, when perpendicular hallways meet.

Combining session logs with employee and student schedules, using hashed unique identifiers, could also help to determine the extent to which wireless connections span Building Floors. A mapping of AP connection to probable user locations could then be developed by machine learning algorithms.

The data analyzed was for a single six-week period from January to mid-February 2020. Further research could involve differing periods of low and high activity, e.g., holidays, exam periods, orientation week.

Further research could also be performed using sessions logs from non-academic buildings, such as athletic facilities, dining areas, student residences and health services. These analyses would help university administration understand the extent to which students make use of those spaces, to inform capacity planning, student engagement, and health and wellness.

This analysis of connection logs can be repeated with more detailed and up-to-date records for connection logs, Building Floor plans, employee, and student schedules. The method can be implemented in any environment with extensive wireless coverage, authenticated users and a large uptake/usage of wireless devices. Standard/consistent definitions of persons, spaces and equipment is necessary for minimizing the amount of data scrubbing required.

7.0 Conclusions

Previous research projects aimed at estimating crowd counts in internal spaces, using wireless data logs, were limited by Wi-Fi architecture and location-specific constraints. Load balancing across Access Points, signal attenuation by occluding objects such as walls, furniture, and equipment, overlapping AP coverage areas, open areas, user adoption rates and building layouts all impose calibration difficulties specific to the given research location. Person densities were either calculated at a building or floor level, reflecting the difficulty of providing a precise location for a person, given a Wi-Fi connection.

On a university campus, extensive Wi-Fi networks and high usage rates among students and employees make Wi-Fi connections a good proxy for person counts. The low cost of Wi-Fi, high data rates, ease of transferability, and that most new devices come with Wi-Fi components pre-installed, make uptake very high. Other, competing, wireless technologies have limited utility and higher costs.

In this project, Wi-Fi logs were combined with additional datasets to provide more specific information, and more detailed analyses. The datasets' levels of granularity, constructed to protect security and privacy, limited the possible analyses.

By using complementary datasets and Big Data techniques, a greater degree of confidence in user counts from Wi-Fi logs can be achieved. Building and floor layouts can suggest more specific areas of a building and floor, for a given AP. Class schedules and employee locations provide reference information which can be used to validate Wi-Fi logs data, for areas of campus where student and employee activity is scheduled, such as Academic and Academic Support buildings.

By combining datasets from multiple administrative systems, useful insights can be drawn. Facilities Management information on area, space usage, and building layout can explain patterns in Wi-Fi logs, or inform policies on space utilization and building management. The real-time person density in circulation areas can be calculated with a high degree of confidence if Wi-Fi logs are cross-referenced against student and employee schedules, and floor layouts. Common definitions and value sets across administrative systems make combining data much simpler.

The technique can be developed further by incorporating more granular wireless network logs, individual work/study schedules for employees and students, and building zones defined using GPS coordinates. The coverage areas for building APs can be determined by comparing wireless logs to employee and student schedules; GPS coordinates would provide more accurate zones than floor and room numbers. Finally, standardized encoding of space information in admin systems across departments, coordinated updates to administrative systems of record, and machine learning techniques applied to the combined data will streamline data processing and provide more useful results.

References

1. Khan A., Kabir A. (2016) COMPARISON AMONG SHORT RANGE WIRELESS NETWORKS: BLUETOOTH, ZIGBEE, & WI-FI, Advances in Computer Science and Engineering 4(2):19-28, Jan 2016
2. Grgurević I., Juršić K., Rajič V. Overview of Wi-Fi-Based Automatic Passenger Counting Solutions in Public Urban Transport. EAI/Springer Innovations in Communication and Computing, pp 181-196,
3. Shao, WH.; Luo, HY.; Zhao, F.; Tian, H.; Huang, JY.; Crivello, A. Floor Identification in Large-Scale Environments with Wi-Fi Autonomous Block Models. IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, 18(2): pp 847-858, FEB 2022
4. Vega-Barbas, M.; Alvarez-Campana, M.; Rivera, D.; Sanz, M.; Berrocal, J. AFOROS: A Low-Cost Wi-Fi-Based Monitoring System for Estimating Occupancy of Public Spaces. SENSORS, 21(11): pp -, JUN 2021
5. Patrikakis, CZ.; Kogias, DG.; Chatzigeorgiou, C.; Kalyvas, D.; Katsadouros, E.; Giannousis, C. A method for measuring urban space density of people and deliver notification, with respect to privacy. 2021 IEEE INTERNATIONAL CONFERENCE ON CONSUMER ELECTRONICS (ICCE), pp -, 2021
6. Khalili, A.; Soliman, AH.; Asaduzzaman, M.; Griffiths, A. Wi-Fi sensing: applications and challenges. JOURNAL OF ENGINEERING-JOE, 2020(3): pp 87-97, MAR 2020
7. Tang, XY.; Xiao, B.; Li, KL. Indoor Crowd Density Estimation Through Mobile Smartphone Wi-Fi Probes. IEEE TRANSACTIONS ON SYSTEMS MAN CYBERNETICS-SYSTEMS, 50(7): pp 2638-2649, JUL 2020
8. Hobson B.W., Burak Gunay H., Ashouri A., Newsham G.R. Wi-fi based occupancy clustering and motif identification: A case study. ASHRAE Transactions, 126: pp 256-264, 126
9. Statista (2019) Number of smartphone users worldwide from 2016 to 2021(in billions) [Internet]. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/
10. Kaminska-Chuchmala, A.; Grana, M. Indoor Crowd 3D Localization in Big Buildings from Wi-Fi Access Anonymous Data. SENSORS, 19(19): pp -, OCT 2019
11. Wang, Z.; Hong, T.; Piette, MA.; Pritoni, M. Inferring occupant counts from Wi-Fi data in buildings through machine learning. BUILDING AND ENVIRONMENT, 158: pp 281-294, JUL 2019
12. Wang, JQ.; Tse, NCF.; Chan, JYC. Wi-Fi based occupancy detection in a complex indoor space under discontinuous wireless communication: A robust filtering based on event-triggered updating. BUILDING AND ENVIRONMENT, 151: pp 228-239, MAR 15 2019
13. Gao, Y.; Tao, J.; Zeng, L.; Fang, XM.; Fang, Q.; Li, XY. User Profiling with campus Wi-Fi Access Trace and Network Traffic. 2019 IEEE International Conference on Multimedia and Expo (ICME), pp 922-927, 2019
14. Jarvis, N.; Hata, J.; Wayne, N.; Raychoudhury, V.; Gani, MO. MiamiMapper: Crowd Analysis using Active and Passive Indoor Localization through Wi-Fi Probe Monitoring. Q2SWINET'19: PROCEEDINGS OF THE 15TH ACM INTERNATIONAL SYMPOSIUM ON QOS AND SECURITY FOR WIRELESS AND MOBILE NETWORKS, pp 1-10, 2019

15. Pietropaoli, B.; Delaney, K.; Pesch, D.; Ploennigs, J. Probabilistic Occupancy Level Estimation Based on Opportunistic Passive Wi-Fi Localisation. PROCEEDINGS OF SAI INTELLIGENT SYSTEMS CONFERENCE (INTELLISYS) 2016, VOL 1, 15: pp 932-952, 2018
16. Kusakabe T., Yaginuma H., Fukuda D. Estimation of bus passengers' waiting time at a coach terminal with Wi-Fi MAC addresses. Transportation Research Procedia, 32: pp 62-68, 32
17. Schauer, L.; Linnhoff-Popien, C. Extracting Context Information from Wi-Fi Captures. 10TH ACM INTERNATIONAL CONFERENCE ON PERVASIVE TECHNOLOGIES RELATED TO ASSISTIVE ENVIRONMENTS (PETRA 2017), pp 123-130, 2017
18. Wang, W.; Chen, JY.; Song, XY. Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach. BUILDING AND ENVIRONMENT, 124: pp 130-142, NOV 1 2017
19. Wang, Y.; Shao, L. Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. BUILDING AND ENVIRONMENT, 114: pp 106-117, MAR 2017
20. Yaik, O.B., Wai, K.Z., Tan, I.K.T., Sheng, O.B. Measuring the accuracy of crowd counting using Wi-Fi probe-request-frame counting technique. Journal of Telecommunication, Electronic and Computer Engineering 8(2), pp. 79-81
21. Ciavarrini, G.; Marcelloni, F.; Vecchio, A. Improving Wi-Fi based localization using external constraints. 2015 9TH INTERNATIONAL CONFERENCE ON NEXT GENERATION MOBILE APPLICATIONS, SERVICES AND TECHNOLOGIES (NGMAST 2015), pp 126-131, 2015
22. Ma, WY.; Zhu, XN.; Huang, JF.; Shou, GC. Detecting Pedestrians Behavior in Building Based on Wi-Fi Signals. 2015 IEEE INTERNATIONAL CONFERENCE ON SMART CITY/SOCIALCOM/SUSTAINCOM (SMARTCITY), pp 1-8, 2015
23. Guo Y., Zhang S., Xiao D. Overview of Wi-Fi technology. Proceedings of the 2012 International Conference on Computer Application and System Modeling, ICCASM 2012, pp 1293-1296,
24. Del-Valle-Soto C., Valdivia L., Velazquez R., Rizo-Dominguez L., Lopez-Pimentel J.C. Smart Campus: An Experimental Performance Comparison of Collaborative and Cooperative Schemes for Wireless Sensor Network, pp, Aug 2019
25. Johnson, NE.; Mandiola, P.; Blankinship, C.; Bonczak, B.; Kontokosta, CE. Validating the Use of Wi-Fi Signals to Estimate Hyperlocal Urban Populations. 2019 IEEE INTERNATIONAL CONFERENCE ON BIG DATA (BIG DATA), pp 1650-1655, 2019
26. Pahlavan, K.; Krishnamurthy, P. Evolution and Impact of Wi-Fi Technology and Applications: A Historical Perspective. INTERNATIONAL JOURNAL OF WIRELESS INFORMATION NETWORKS, 28(1): pp 3-19, MAR 2021
27. Zagatti, GA.; Wu, TF.; Ng, SK.; Bressan, S. A Large-scale Disease Outbreak Analytics System based on Wi-Fi Session Logs. 2021 22ND IEEE INTERNATIONAL CONFERENCE ON MOBILE DATA MANAGEMENT (MDM 2021), pp 236-239, 2021
28. Binthaisong A., Srichan J., Phithakkitnukoon S. Wi-Crowd: Sensing and visualizing crowd on campus using Wi-Fi access point data. UbiComp/ISWC 2017 - Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp 441-447,
29. Grgurević I, Juršić K, Rajič I (2020) Review of automatic passenger counting Systems in Public Urban Transport. In: Proceedings of 5th EAI international conference on Management of Manufacturing Systems, EAI MMS 2020, cyberspace

30. Wenhua Shao, Haiyong Luo, Fang Zhao, Hui Tian: Floor Identification in Large Scale Environments with Wi-Fi Autonomous Block Models. In: IEEE Transactions on Industrial Informatica, Vol 18, No. 2, February 2022
31. Junqi Wang, Norman Chung Fai Tse, John Yau Chung Chan: Wi-Fi-based occupancy detection in a complex indoor space under discontinuous wireless communication: A robust filtering based on event-triggered updating. In: Building and Environment 151, pp 228-239, 2019
32. Gloria Ciavarrini, Francesco Marcelloni, Alessio Vecchio: Improving Wi-Fi-based localization using external constraints. In 2015 9th international Conference on next generation Mobile Applications, Services and Technologies, 2015.
33. Yu Ichifuji, Noriaki Koide, Yoshitaka Terai, Yoshihide Matsuo, Nobuhiro Akashi, Toru Kobayashi: A study for understanding of tourist trip patter based on log data of Wi-Fi access points. In 2016 IEEE international Conference on Big Data, 2016
34. Hamed Nabizadeh Rafsanjani, Ali Ghahramani: Extracting occupants' energy-use patterns from Wi-Fi networks in office buildings. In Journal of Building Engineering 26 (2019).
35. Calabrese, F., J. Reades, C. Ratti. "Eigne places: Segmenting Space through Digital Signatures." Pervasive Computing, IEEE 9.1 (2010): pp 78-84

Appendix A: List of Acronyms

Acronyms	Description
AP	Access Point
COU	Council of Ontario Universities
SPO	Space Planning Office
IST	Information Systems and Technology
SOC	Security Operations Centre
ICT	Information & Communication Technology

Appendix B:
AP Counts for Buildings with Known (Y) and Unknown (N) AP Locations; Log Counts for Buildings with Known (Y) and Unknown (N) AP Locations

Building	N	Y
AL		21
ARC		31
B1		26
B2		21
BMH	1	89
C2		33
CIF	2	21
COM		7
CPH	1	49
CSB	1	
DC	4	110
DMS	1	28
DMS2	9	
DWE		54
E10	3	
E2	1	69
E3		45
E5		81
E6		36
E7	1	199
ECH		26
EIT	1	59
ERC	1	11
ESC	1	31
EV1		41
EV2		22
EV3	5	19
FED		16
FRL	1	
GH	3	
GSC		20
HH		97
HS	4	11
IHB		21
LIB	7	60
M3	9	32
MC	2	158
MED		4
ML		32
NH		43
OPT		59
PAC		10
PAS	3	102
PHR		43
PHY		55
QNC	2	75
RCH		36
SCH		25
SLC	1	28
STC		51
TC		18
UC		3
Totals	**64**	**2,128**

Buildi	N	Y
AL		33,454
ARC		52,572
B1		44,612
B2		35,988
BMH	1,728	152,346
C2		57,442
CIF	3,504	35,582
COM		12,302
CPH	1,746	82,840
CSB	1,778	
DC	6,998	180,350
DMS	1,692	47,010
DMS2	9,462	
DWE		90,186
E10	5,000	
E2	1,738	118,602
E3		77,978
E5		138,694
E6		63,336
E7	1,704	181,156
ECH		44,728
EIT	1,694	100,844
ERC	1,734	19,334
ESC	1,738	52,962
EV1		70,044
EV2		37,738
EV3	8,528	32,296
FED		26,424
FRL	1,650	
GH	5,156	
GSC		34,840
HH		160,520
HS	6,852	18,758
IHB		35,432
LIB	11,942	102,184
M3	15,718	55,568
MC	3,450	269,972
MED		6,876
ML		53,974
NH		73,942
OPT		100,992
PAC		17,238
PAS	5,168	173,918
PHR		72,548
PHY		93,456
QNC	3,238	130,718
RCH		60,686
SCH		43,628
SLC	1,766	49,066
STC		85,936
TC		30,982
UC		816
	103,984	**3,460,870**

Appendix C: Council of Ontario Universities Space Categories

Category	Category Description	Subcategory	Subcategory Description
'1.0	Classroom Facilities	'1.1	Tiered classroom
		'1.2	Non-tiered classroom
		'1.3	Active Learning classroom space
		'1.3A	Tiered Space
		'1.3B	Non-Tiered Space
		'1.4	Classroom service space
'2.0	Undergraduate Laboratories	'2.1	Scheduled class laboratory
		'2.2	Unscheduled class laboratory
		'2.3	Laboratory (undergraduate) support
'3.0	Research Laboratories	'3.1	Research laboratory space
		'3.2	Research (graduate & faculty) support space
'4.0	Academic Departmental Offices	'4.1	Academic Offices
		'4.2	Research office / project space
		'4.3	Graduate student offices
		'4.4	Departmental administrative & Support staff offices
		'4.5	Office support space
'5.0	Library Facilties & Study Space	'5.1	Library collection space
		'5.2	Library/office space
		'5.3	Library support space
		'5.4	Study Space under the Jurisdiction of the Univ Lib System
		'5.5	Study space Not under the jurisdiction of the univ lib system
'6.0	Athletic / Recreation	'6.1	Athletic activity areas
		'6.2	Athletic seating areas
		'6.3	Athletic service space
'7.0	Food Services	'7.1	Food facilities
		'7.2	Food facilities services
'8.0	Bookstore & Merchandising Facilities	'8.1	Bookstore/Merchandising
'9.0	Plant Maintenance	'9.1	Plant maintenance
'10.0	Central Administrative Offices	'10.1	Administrative office areas
		'10.2	Administrative office support space
'11.0	Non-Library Study Space	'11.1	Formal Study Space
		'11.2	Informal Study Space
'12.0	Central Services	'12.1	Computing facilities
		'12.2	Other central services
		'12.3	Central student and student support services
'13.0	Health Services	'13.1	Health service facilities
'14.0	Common Use / Student Activity	'14.1	Student offices and support space
		'14.2	Recreational facilities and service
		'14.3	Lounge and Service Space
'15.0	Asembly & Exhibition Facilities	'15.1	Assembly facilities
		'15.2	Exhibition facilities
'16.0	Non-Assignable	'16.1	Central utility plant
		'16.2	Other non-assignable areas
		'16.3	Inactive Unassignable
		'16.4	Parking Structures
'17.0	Residences	'17.1	Residence living space
		'17.2	Residence service space
'18.0	Animal Space	'18.1	Specialized central animal areas
		'18.2	Farm-type animal areas
'19.0	Other University Facilities	'19.1	Day care facilities
		'19.2	Rifle ranges and military training
		'19.3	Extra-university merchandising facilities
		'19.4	Demonstration schools
		'19.5	Inactive assignable
		'19.6	Non-institutional agencies occupying univerity space
		'19.7	Instructional service activities to external community
'20.0	Health Science Clinical Facilities	'20.1	Health Science clinical facilities
		'20.2	All space in support of clinical instruction, research and service

Appendix D:

Correlation Coefficients for User, User30 and User60 vs. Expected Users for Building Floors

Building	Building_Floor	GrossNASMRatio	ConstructionYear	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
AL	AL_00	2	12/1/2002	80	1.031	0.825	0.572	
AL	AL_01	2	9/1/1965	660	1.667	1.366	0.870	
AL	AL_02	2	9/1/1965	377	0.727	0.579	0.353	
ARC	ARC_01	1.5	9/1/1965	248	0.668	0.576	0.561	
ARC	ARC_02	1.5	9/1/2004	1212	1.389	1.210	1.068	
ARC	ARC_03	1.5	9/1/2004	504	1.194	0.884	0.568	
B1	B1_02	2.5	9/1/2004	984	1.373	0.970	0.391	
B1	B1_03	2.5	9/1/1964	168	1.147	0.953	0.826	
B2	B2_01	1.6	9/1/1964	96	0.833	0.614	0.346	
B2	B2_02	1.6	9/1/1967	380	8.324	5.855	4.145	
B2	B2_03	1.6	9/1/1967	404	1.361	1.089	0.679	
BMH	BMH_01	1.9	9/1/1967	1376	-0.940	-0.579	0.053	
BMH	BMH_02	1.9	9/1/1972	1500	1.933	1.543	0.808	
BMH	BMH_03	1.9	9/1/1972	508	0.234	0.354	0.234	
BMH	BMH_04	1.9	9/1/1972	380	14.012	11.382	7.772	
C2	C2_00	1.7	9/1/1972	380	7.053	4.584	2.995	
C2	C2_01	1.7	1/1/1971	380	6.776	4.926	2.455	
C2	C2_02	1.7	1/1/1971	760	1.068	0.981	0.839	
C2	C2_03	1.7	1/1/1971	377	10.729	8.024	4.926	
CIF	CIF_01	1.3	1/1/1971	380	9.389	7.963	4.934	
COM	COM_01	1.6	5/1/1987	380	1.443	1.296	0.659	
CPH	CPH_01	1.7	9/1/1966	878	1.863	1.466	0.902	
CPH	CPH_03	1.7	9/1/1972	440	1.105	0.943	0.693	
CPH	CPH_04	1.7	9/1/1972	398	1.740	1.540	1.315	
DC	DC_01	2	9/1/1972	1471	2.165	1.679	1.047	
DMS	DMS_01	1.8	3/1/1987	380	6.371	5.497	4.718	
DMS	DMS_02	1.8	12/1/2012	380	5.805	2.855	0.729	
DMS	DMS_03	1.8	12/1/2012	380	6.184	3.853	2.621	
DWE	DWE_01	1.9	12/1/2012	340	1.521	1.118	0.504	
DWE	DWE_02	1.9	9/1/1958	688	1.255	1.049	0.731	
DWE	DWE_03	1.9	9/1/1958	992	1.192	0.974	0.611	
E2	E2_01	1.6	9/1/1958	760	4.421	2.686	0.855	
E2	E2_02	1.6	9/1/1961	162	1.325	1.114	0.842	
E2	E2_03	1.6	9/1/1961	384	1.216	1.023	0.788	
E3	E3_02	1.5	9/1/1961	904	0.145	0.077	0.036	
E3	E3_03	1.5	9/1/1961	856	0.630	0.592	0.399	
E5	E5_03	2	9/1/1961	564	1.310	1.183	0.992	
E5	E5_04	2	7/1/2010	224	0.186	0.184	0.145	
E5	E5_06	2	7/1/2010	332	1.619	1.453	1.189	

E6	E6_02	2	7/1/2010	256	0.384	0.349	0.277	
E6	E6_04	2	7/1/2011	184	1.039	0.952	0.830	
E7	E7_01	2	7/1/2011	380	44.211	21.958	8.808	
E7	E7_02	2	8/16/2018	328	20.882	18.402	16.058	
E7	E7_03	2	8/16/2018	1116	52.608	40.052	27.392	
E7	E7_04	2	8/16/2018	1014	-17.085	-7.696	1.414	
E7	E7_05	2	8/16/2018	1127	-29.917	-24.658	-18.147	
E7	E7_06	2	8/16/2018	805	10.389	8.770	6.334	
E7	E7_07	2	8/16/2018	1432	-3.764	0.082	1.567	
ECH	ECH_01	1.4	8/16/2018	1214	1.122	0.854	0.556	
EIT	EIT_01	2.1	1/1/1959	807	1.679	1.255	0.592	
EIT	EIT_02	2.1	9/1/2003	192	1.251	1.105	1.225	
EIT	EIT_03	2.1	9/1/2003	240	1.061	0.880	0.671	
EIT	EIT_05	2.1	9/1/2003	758	2.636	2.876	3.050	
ESC	ESC_01	1.8	9/1/2003	84	-0.084	-0.135	-0.244	
ESC	ESC_02	1.8	9/1/1964	27	11.796	6.185	4.519	
ESC	ESC_03	1.8	9/1/1964	1163	0.730	0.625	0.480	
EV1	EV1_01	1.8	9/1/1964	296	1.185	1.084	0.864	
EV1	EV1_02	1.8	9/1/1965	252	0.781	0.681	0.453	
EV1	EV1_03	1.8	9/1/1965	208	1.385	1.151	0.740	
EV2	EV2_01	1.6	9/1/1965	611	0.883	0.802	0.518	
EV2	EV2_02	1.6	11/1/1981	276	0.634	0.501	0.258	
EV3	EV3_03	2.2	11/1/1981	224	0.930	0.781	0.582	
FED	FED_01	2.1	9/1/2011	756	0.746	1.069	0.817	
GSC	GSC_01	1.3	12/1/1984	2280	3.040	-1.039	-1.417	
GSC	GSC_02	1.3	9/1/1966	760	-0.999	-0.277	-0.131	
HH	HH_01	2.1	9/1/1966	1874	2.091	1.660	1.036	
HH	HH_02	2.1	9/1/1968	666	2.810	2.540	2.114	
HH	HH_03	2.1	9/1/1968	472	1.302	1.052	0.769	
HS	HS_01	1.8	9/1/1968	760	20.958	15.932	7.600	
IHB	IHB_03	1.9	9/1/1968	380	7.725	6.804	5.539	
LIB	LIB_01	1.6	9/1/2009	380	7.984	6.166	3.713	
MC	MC_01	1.8	9/1/1965	2412	1.418	1.307	1.058	
MC	MC_02	1.8	9/1/1967	2217	1.649	1.407	0.989	
MC	MC_03	1.8	9/1/1967	960	1.468	1.252	0.892	
MC	MC_04	1.8	9/1/1967	2276	1.382	1.205	0.927	
MC	MC_05	1.8	9/1/1967	7200	8.326	4.520	2.619	
MC	MC_06	1.8	9/1/1967	4180	12.242	8.595	5.812	

ML	ML_01	2	9/1/1967	328	1.534	1.311	1.041	
ML	ML_02	2	9/1/1962	344	1.464	1.257	1.024	
ML	ML_03	2	9/1/1962	360	1.096	0.858	0.580	
NH	NH_01	1.9	9/1/1962	380	4.586	2.362	0.637	
NH	NH_02	1.9	8/1/1972	1060	1.216	1.371	1.398	
NH	NH_03	1.9	8/1/1972	1140	0.875	2.853	2.866	
OPT	OPT_01	1.7	8/1/1972	176	2.213	1.585	1.246	
OPT	OPT_02	1.7	9/1/1973	380	3.739	1.166	0.239	
OPT	OPT_03	1.7	9/1/1973	183	2.287	1.785	1.113	
OPT	OPT_04	1.7	9/1/1973	708	1.697	1.447	1.094	
PAS	PAS_01	1.8	9/1/1973	526	1.537	1.289	0.922	
PAS	PAS_02	1.8	9/1/1972	701	1.422	1.232	0.907	
PAS	PAS_03	1.8	9/1/1972	116	0.976	0.934	0.951	
PAS	PAS_04	1.8	9/1/1972	114	0.970	1.002	1.067	
PHR	PHR_01	2.4	9/1/1972	428	2.874	2.499	1.266	
PHR	PHR_02	2.4	1/1/2009	156	-0.163	-0.148	-0.147	
PHR	PHR_06	2.4	1/1/2009	380	4.958	3.332	2.379	
PHY	PHY_01	1.8	1/1/2009	472	1.175	0.956	0.452	
PHY	PHY_02	1.8	9/1/1959	1020	1.487	1.150	0.677	
PHY	PHY_03	1.8	9/1/1959	801	2.053	1.481	0.714	
PHY	PHY_04	1.8	9/1/1959	380	2.336	1.329	0.863	
QNC	QNC_01	2.3	9/1/1959	344	2.262	1.824	1.365	
QNC	QNC_02	2.3	8/1/2011	316	2.101	1.643	1.212	
RCH	RCH_01	2.1	8/1/2011	622	1.540	1.169	0.655	
RCH	RCH_02	2.1	9/1/1967	659	1.924	1.449	0.872	
RCH	RCH_03	2.1	9/1/1967	360	1.805	1.373	0.998	
SCH	SCH_01	1.6	9/1/1967	760	14.434	4.745	0.716	
SCH	SCH_02	1.6	9/1/1967	380	5.559	2.410	1.040	
SLC	SLC_00	2	9/1/1967	760	8.094	4.878	2.886	
SLC	SLC_01	2	2/1/1967	760	6.328	0.681	1.952	
SLC	SLC_02	2	2/1/1967	380	23.439	15.532	7.213	
SLC	SLC_03	2	2/1/1967	348	21.871	18.011	10.434	
STC	STC_01	2.1	2/1/1967	380	112.650	68.613	37.155	
STC	STC_02	2.1	8/1/2015	380	17.425	13.257	8.378	
TC	TC_01	2.2	8/1/2015	760	0.045	0.140	0.019	
TC	TC_02	2.2	12/1/2002	380	0.954	0.788	0.595	
TC	TC_03	2.2	12/1/2002	380	8.307	6.999	5.391	

Appendix E: Correlation Coefficient Values for User, User30 and User60 for Building Floor Zones

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
AL	AL_00	AL_00_00	MixedUse	0.63	80	1.031	0.825	0.572	
AL	AL_01	AL_01_10	MixedUse	0.63	160	1.233	0.911	0.381	
AL	AL_01	AL_01_11	Classroom	0.04	276	1.566	1.263	0.808	
AL	AL_01	AL_01_12	Classroom	0.1	224	0.964	0.748	0.412	
AL	AL_02	AL_02_20	Classroom	0.26	102	-0.001	0	0	
AL	AL_02	AL_02_21	Classroom	0.1	275	1.176	0.936	0.572	
ARC	ARC_01	ARC_01_100	UndergradLab	0	96	0.318	-0.037	-0.528	
ARC	ARC_01	ARC_01_110	Classroom	0	152	0.59	0.394	0.677	
ARC	ARC_02	ARC_02_200	AcademicOffice	0	380	2.322	2.053	1.543	
ARC	ARC_02	ARC_02_201	AcademicOffice	0	380	4.324	3.368	2.355	
ARC	ARC_02	ARC_02_202	AcademicOffice	0	72	1.38	1.284	1.15	
ARC	ARC_02	ARC_02_210	Library	0	380	5.284	3.126	1.282	
ARC	ARC_03	ARC_03_300	UndergradLab	0	392	0.281	0.199	0.087	
ARC	ARC_03	ARC_03_310	UndergradLab	0.06	112	1.292	0.95	0.605	
B1	B1_02	B1_02_27	Classroom	0.34	224	1.326	0.942	0.395	
B1	B1_02	B1_02_28	ResearchLab	0	380	15.884	14.405	10.874	
B1	B1_02	B1_02_29	ResearchLab	0.19	380	8.808	7.261	4.682	
B1	B1_03	B1_03_37	UndergradLab	0	168	1.147	0.953	0.826	
B2	B2_01	B2_01_15	ResearchLab	0	96	0.833	0.614	0.346	
B2	B2_02	B2_02_25	ResearchLab	0	380	8.324	5.855	4.145	
B2	B2_03	B2_03_35	ResearchLab	0	404	1.361	1.089	0.679	
BMH	BMH_01	BMH_01_100	Classroom	0	104	2.073	1.591	0.946	
BMH	BMH_01	BMH_01_101	Classroom	0	232	1.553	1.425	1.322	
BMH	BMH_01	BMH_01_161	ResearchLab	0	380	8.016	3.739	1.266	
BMH	BMH_01	BMH_01_162	Classroom	0.1	124	1.824	1.527	1.085	
BMH	BMH_01	BMH_01_168	Classroom	0	380	112.403	88.097	47.384	
BMH	BMH_01	BMH_01_170	AcademicOffice	0	156	0.742	0.626	0.428	
BMH	BMH_02	BMH_02_230	AcademicOffice	0	380	10.068	6.282	4.1	
BMH	BMH_02	BMH_02_240	ResearchLab	0.03	116	1.14	1.09	0.756	
BMH	BMH_02	BMH_02_268	AcademicOffice	0	380	9.55	6.434	3.021	
BMH	BMH_02	BMH_02_269	ResearchLab	0	380	4.705	2.979	2.192	
BMH	BMH_02	BMH_02_270	Classroom	0	244	1.127	0.804	0.204	
BMH	BMH_03	BMH_03_302	AcademicOffice	0	128	0.847	0.815	0.701	
BMH	BMH_03	BMH_03_368	UndergradLab	0	380	14.937	11.826	8.142	
BMH	BMH_04	BMH_04_468	ResearchLab	0	380	14.012	11.382	7.772	
C2	C2_00	C2_00_06	ResearchLab	0	380	7.053	4.584	2.995	
C2	C2_01	C2_01_16	UndergradLab	0	380	6.776	4.926	2.455	
C2	C2_02	C2_02_27	UndergradLab	0	380	0.994	0.939	0.832	
C2	C2_02	C2_02_28	AcademicOffice	0	380	4.14	2.453	1.343	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
C2	C2_03	C2_03_38	AcademicOffice	0	377	10.729	8.024	4.926	
CIF	CIF_01	CIF_01_121	Athletics	0	380	9.389	7.963	4.934	
COM	COM_01	COM_01_11	AdministrativeOf	0.09	380	1.443	1.296	0.659	
CPH	CPH_01	CPH_01_132	MixedUse	0.42	378	11.37	7.619	3.534	
CPH	CPH_01	CPH_01_133	ResearchLab	0.17	380	1.408	1.187	0.816	
CPH	CPH_01	CPH_01_134	UndergradLab	0.06	120	1.96	1.634	1.053	
CPH	CPH_03	CPH_03_360	CommonUse	0	192	-5.16	-4.424	-5.611	
CPH	CPH_03	CPH_03_368	UndergradLab	0	248	0.67	0.553	0.344	
CPH	CPH_04	CPH_04_430	AcademicOffice	0	380	9.176	7.595	4.753	
CPH	CPH_04	CPH_04_433	UndergradLab	0.07	18	-7.867	-5.7	-4.367	
DC	DC_01	DC_01_135	Classroom	0	360	2.239	1.735	1.106	
DC	DC_01	DC_01_155	Library	0	380	7.655	5.855	4.629	
DC	DC_01	DC_01_156	Library	0.01	380	2.179	1.711	0.643	
DC	DC_01	DC_01_170	ResearchLab	0.05	351	2.399	1.946	0.996	
DMS	DMS_01	DMS_01_100	Classroom	0	380	6.371	5.497	4.718	
DMS	DMS_02	DMS_02_201	AcademicOffice	0	380	5.805	2.855	0.729	
DMS	DMS_03	DMS_03_312	Classroom	0	380	6.184	3.853	2.621	
DWE	DWE_01	DWE_01_150	ResearchLab	0.05	200	1.451	1.075	0.48	
DWE	DWE_01	DWE_01_151	UndergradLab	0.02	140	-0.261	-0.353	-0.767	
DWE	DWE_02	DWE_02_251	AcademicOffice	0.21	380	19.511	11.753	6.039	
DWE	DWE_02	DWE_02_252	Classroom	0.02	308	1.673	1.359	0.923	
DWE	DWE_03	DWE_03_350	UndergradLab	0	380	10.197	6.476	3.334	
DWE	DWE_03	DWE_03_351	Classroom	0.02	348	1.221	0.976	0.555	
DWE	DWE_03	DWE_03_352	Classroom	0.04	264	1.204	0.958	0.61	
E2	E2_01	E2_01_177	AcademicOffice	0.16	380	17.518	9.9	4.671	
E2	E2_01	E2_01_178	CommonUse	0.16	380	9.197	7.318	5.766	
E2	E2_02	E2_02_235	AcademicOffice	0.05	72	-6.12	-5.388	-3.867	
E2	E2_02	E2_02_236	UndergradLab	0	90	1.895	1.597	1.205	
E2	E2_03	E2_03_333	AcademicOffice	0	108	0.29	0.287	0.333	
E2	E2_03	E2_03_334	UndergradLab	0	108	2.081	1.864	1.602	
E2	E2_03	E2_03_335	UndergradLab	0.01	168	0.72	0.651	0.692	
E3	E3_02	E3_02_210	ResearchLab	0.03	380	0.23	0.125	0.067	
E3	E3_02	E3_02_211	ResearchLab	0.01	144	0.302	0.218	0.131	
E3	E3_02	E3_02_212	ResearchLab	0.25	380	1.854	1.378	0.862	
E3	E3_03	E3_03_310	AcademicOffice	0	380	2.766	1.005	0.282	
E3	E3_03	E3_03_311	AcademicOffice	0.14	380	11.568	9.068	6.016	
E3	E3_03	E3_03_316	UndergradLab	0	96	-0.232	-0.06	0.14	
E5	E5_03	E5_03_303	AcademicOffice	0	380	2.384	1.192	0.761	
E5	E5_03	E5_03_310	Classroom	0.16	184	1.142	1.051	0.92	
E5	E5_04	E5_04_412	Classroom	0	224	0.186	0.184	0.145	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
E5	E5_06	E5_06_600	Classroom	0	332	1.619	1.453	1.189	
E6	E6_02	E6_02_202	AcademicOffice	0	256	0.384	0.349	0.277	
E6	E6_04	E6_04_402	Classroom	0	184	1.039	0.952	0.83	
E7	E7_01	E7_01_142	Classroom	0	380	44.211	21.958	8.808	
E7	E7_02	E7_02_230	AcademicOffice	0	157	6.408	5.293	4.045	
E7	E7_02	E7_02_240	Classroom	0	171	25.847	22.344	18.854	
E7	E7_03	E7_03_331	AcademicOffice	0	152	9.138	6.967	3.539	
E7	E7_03	E7_03_333	AcademicOffice	0	171	8.906	7.292	5.269	
E7	E7_03	E7_03_334	Classroom	0	153	120.582	103.895	67.542	
E7	E7_03	E7_03_335	Classroom	0	155	184.09	141.335	104.239	
E7	E7_03	E7_03_340	AcademicOffice	0	153	8.922	6.039	3.353	
E7	E7_03	E7_03_341	ResearchLab	0.03	171	11.17	7.07	4.953	
E7	E7_03	E7_03_345	AcademicOffice	0	161	34.416	15.255	7.944	
E7	E7_04	E7_04_404	Classroom	0	161	177.286	155.919	113.472	
E7	E7_04	E7_04_405	Classroom	0	380	81.711	65.926	47.721	
E7	E7_04	E7_04_431	AcademicOffice	0	153	9.072	7.098	5.183	
E7	E7_04	E7_04_441	Classroom	0	159	65.327	45.698	31.925	
E7	E7_04	E7_04_443	Classroom	0	161	62.21	56.592	46.758	
E7	E7_05	E7_05_534	Classroom	0	155	167.09	147.129	111.787	
E7	E7_05	E7_05_535	Classroom	0	161	78.578	65.621	50.832	
E7	E7_05	E7_05_541	ResearchLab	0	155	7.535	6.394	5.484	
E7	E7_05	E7_05_542	ResearchLab	0	167	13.316	11.91	9.49	
E7	E7_05	E7_05_543	ResearchLab	0	153	7.183	5.902	4.359	
E7	E7_05	E7_05_544	ResearchLab	0	171	3.614	2.784	2.006	
E7	E7_05	E7_05_545	AcademicOffice	0	165	9.382	4.63	2.739	
E7	E7_06	E7_06_631	ResearchLab	0	169	8.68	7.497	5.663	
E7	E7_06	E7_06_640	AcademicOffice	0	153	17.235	13.02	6.778	
E7	E7_06	E7_06_641	ResearchLab	0	171	10.351	9.292	7.409	
E7	E7_06	E7_06_643	AcademicOffice	0	159	4.799	4.208	3.409	
E7	E7_06	E7_06_644	AcademicOffice	0	153	11.281	10.085	8.471	
E7	E7_07	E7_07_731	AcademicOffice	0.06	171	3.026	2.599	2.032	
E7	E7_07	E7_07_732	AcademicOffice	0	152	4.388	2.546	1.783	
E7	E7_07	E7_07_733	AcademicOffice	0	167	9.579	7.757	6.711	
E7	E7_07	E7_07_734	AcademicOffice	0	153	39.255	17.379	7.046	
E7	E7_07	E7_07_740	AcademicOffice	0	150	4.627	3.62	2.867	
E7	E7_07	E7_07_741	AcademicOffice	0	171	4.515	3.772	2.953	
E7	E7_07	E7_07_742	AcademicOffice	0	169	6.787	6.142	5.036	
E7	E7_07	E7_07_744	AcademicOffice	0	146	3.014	2.411	1.712	
E7	E7_07	E7_07_745	AcademicOffice	0	153	22.928	12.366	8.314	
ECH	ECH_01	ECH_01_10	CentralServices	0.32	380	2.647	1.87	1.17	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
ECH	ECH_01	ECH_01_120	UndergradLab	0	112	0.844	0.726	0.511	
ECH	ECH_01	ECH_01_121	UndergradLab	0	158	1.626	1.274	1.028	
ECH	ECH_01	ECH_01_122	UndergradLab	0	144	1.264	0.759	0.378	
ECH	ECH_01	ECH_01_123	UndergradLab	0	40	1.438	1.077	0.597	
ECH	ECH_01	ECH_01_13	CentralServices	0.01	380	2.179	1.511	0.691	
EIT	EIT_01	EIT_01_100	UndergradLab	0	120	1.84	1.691	1.485	
EIT	EIT_01	EIT_01_101	UndergradLab	0	340	1.584	1.14	0.497	
EIT	EIT_01	EIT_01_102	ResearchLab	0.01	347	2.072	1.024	0.455	
EIT	EIT_02	EIT_02_201	ResearchLab	0	96	0.354	0.09	0.007	
EIT	EIT_02	EIT_02_205	AcademicOffice	0	96	0.975	0.953	0.986	
EIT	EIT_03	EIT_03_314	AcademicOffice	0	84	3.848	3.354	2.823	
EIT	EIT_03	EIT_03_315	Classroom	0	156	1.109	0.961	0.75	
EIT	EIT_05	EIT_05_500	ResearchLab	0	380	3.372	2.946	2.48	
EIT	EIT_05	EIT_05_501	ResearchLab	0	378	4.108	3.016	1.91	
ESC	ESC_01	ESC_01_14	UndergradLab	0.02	84	-0.084	-0.135	-0.244	
ESC	ESC_02	ESC_02_25	AcademicOffice	0	27	11.796	6.185	4.519	
ESC	ESC_03	ESC_03_32	ResearchLab	0.22	380	9.437	6.168	4.121	
ESC	ESC_03	ESC_03_33	AcademicOffice	0.06	368	11.076	7.563	4.353	
ESC	ESC_03	ESC_03_34	UndergradLab	0.17	36	0.793	0.62	0.412	
ESC	ESC_03	ESC_03_35	AcademicOffice	0.32	379	18.863	13.409	8.259	
EV1	EV1_01	EV1_01_13	UndergradLab	0.04	296	1.185	1.084	0.864	
EV1	EV1_02	EV1_02_22	AcademicOffice	0.01	76	-0.406	-0.264	-0.08	
EV1	EV1_02	EV1_02_24	MixedUse	0.37	176	1.094	0.913	0.563	
EV1	EV1_03	EV1_03_35	AcademicOffice	0.03	208	1.385	1.151	0.74	
EV2	EV2_01	EV2_01_100	UndergradLab	0	208	0.828	0.675	0.412	
EV2	EV2_01	EV2_01_101	UndergradLab	0	23	1.292	1.028	0.671	
EV2	EV2_01	EV2_01_102	ResearchLab	0	380	5.486	3.497	1.954	
EV2	EV2_02	EV2_02_200	Classroom	0	276	0.634	0.501	0.258	
EV3	EV3_03	EV3_03_341	Classroom	0	224	0.93	0.781	0.582	
FED	FED_01	FED_01_103	FoodServices	0	376	1.227	0.891	0.587	
FED	FED_01	FED_01_110	FoodServices	0	380	2.189	0.534	0.126	
GSC	GSC_01	GSC_01_10	PlantMaintenanc	0	380	19.605	14.029	6.658	
GSC	GSC_01	GSC_01_11	PlantMaintenanc	0	380	3.94	1.955	0.486	
GSC	GSC_01	GSC_01_110	AdministrativeOf	0	380	14.671	10.663	7.261	
GSC	GSC_01	GSC_01_112	AdministrativeOf	0	380	3.375	1.816	0.692	
GSC	GSC_01	GSC_01_114	OtherSpace	0	380	6.106	1.641	0.255	
GSC	GSC_01	GSC_01_116	CentralServices	0	380	11.216	7.283	4.279	
GSC	GSC_02	GSC_02_22	AdministrativeOf	0	380	0.831	0.677	0.388	
GSC	GSC_02	GSC_02_26	AdministrativeOf	0.08	380	3.271	1.949	1.08	
HH	HH_01	HH_01_10	AcademicOffice	0.08	380	20.937	11.003	2.368	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
HH	HH_01	HH_01_11	AcademicOffice	0	173	0.901	0.691	0.236	
HH	HH_01	HH_01_110	Classroom	0	383	1.457	1.215	0.836	
HH	HH_01	HH_01_13	Classroom	0	296	1.416	1.197	0.851	
HH	HH_01	HH_01_15	AssemblyExhibiti	0.06	240	2.369	1.784	0.98	
HH	HH_01	HH_01_16	AssemblyExhibiti	0.15	378	2.915	0.466	0.209	
HH	HH_01	HH_01_18	UndergradLab	0	24	1.413	1.179	0.788	
HH	HH_02	HH_02_210	Classroom	0	308	2.525	2.284	1.911	
HH	HH_02	HH_02_22	AcademicOffice	0	224	1.006	0.892	0.684	
HH	HH_02	HH_02_25	AcademicOffice	0	134	1.579	1.378	0.474	
HH	HH_03	HH_03_33	Classroom	0	312	1.056	0.814	0.538	
HH	HH_03	HH_03_34	AcademicOffice	0.06	160	0.42	0.251	0.025	
HS	HS_01	HS_01_130	HealthServices	0	380	13.325	10.259	4.716	
HS	HS_01	HS_01_141	HealthServices	0.22	380	5.692	4.587	1.832	
IHB	IHB_03	IHB_03_302	OtherSpace	0	380	7.725	6.804	5.539	
LIB	LIB_01	LIB_01_13	Library	0.18	380	7.984	6.166	3.713	
MC	MC_01	MC_01_100	AdministrativeOf	0.13	373	5.765	3.882	2.548	
MC	MC_01	MC_01_101	OtherSpace	0	380	4.211	3.845	3.408	
MC	MC_01	MC_01_102	AdministrativeOf	0.17	305	3	2.148	1.925	
MC	MC_01	MC_01_105	AdministrativeOf	0	320	1.985	1.805	1.517	
MC	MC_01	MC_01_106	AdministrativeOf	0.1	380	3.789	3.055	2.343	
MC	MC_01	MC_01_107	AdministrativeOf	0	380	3.685	2.832	1.492	
MC	MC_01	MC_01_108	Classroom	0.07	274	1.065	0.959	0.689	
MC	MC_02	MC_02_201	Classroom	0	392	1.726	1.537	1.179	
MC	MC_02	MC_02_202	CommonUse	0	380	5.524	2.261	1.45	
MC	MC_02	MC_02_203	Classroom	0.07	344	2.329	2.04	1.69	
MC	MC_02	MC_02_204	AdministrativeOf	0.19	377	9.812	3.645	1.69	
MC	MC_02	MC_02_205	Classroom	0	380	1.559	1.39	1.137	
MC	MC_02	MC_02_206	Classroom	0.08	344	1.611	1.34	0.861	
MC	MC_03	MC_03_300	UndergradLab	0	176	1.046	0.846	0.545	
MC	MC_03	MC_03_302	UndergradLab	0	24	9.083	8.1	6.533	
MC	MC_03	MC_03_304	AcademicOffice	0	380	11.613	8.616	4.711	
MC	MC_03	MC_03_305	AcademicOffice	0	380	7.182	5.811	4.642	
MC	MC_04	MC_04_400	AcademicOffice	0	380	15.911	8.524	5.421	
MC	MC_04	MC_04_401	AcademicOffice	0.1	380	10.55	6.184	2.376	
MC	MC_04	MC_04_402	Classroom	0	404	1.607	1.4	1.087	
MC	MC_04	MC_04_404	Classroom	0.05	356	1.581	1.355	1.08	
MC	MC_04	MC_04_405	Classroom	0	400	1.015	0.864	0.635	
MC	MC_04	MC_04_406	Classroom	0	356	1.365	1.168	0.837	
MC	MC_05	MC_05_500	AcademicOffice	0.19	373	8.702	2.043	0.976	
MC	MC_05	MC_05_501	AcademicOffice	0	380	6.572	2.855	1.654	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
MC	MC_05	MC_05_503	AcademicOffice	0	380	10.961	8.4	5.492	
MC	MC_05	MC_05_504	AcademicOffice	0	380	8.192	4.25	3.032	
MC	MC_05	MC_05_510	AcademicOffice	0	380	12.942	8.232	4.1	
MC	MC_05	MC_05_512	AcademicOffice	0	380	8.634	6.063	4.737	
MC	MC_05	MC_05_520	AcademicOffice	0	380	9.122	4.791	2.908	
MC	MC_05	MC_05_522	AcademicOffice	0	380	7.066	5.316	3.384	
MC	MC_05	MC_05_523	AcademicOffice	0	367	7.717	6.24	4.428	
MC	MC_05	MC_05_524	AcademicOffice	0	380	11.299	8.738	5.842	
MC	MC_05	MC_05_530	AcademicOffice	0	380	5.895	4.082	2.608	
MC	MC_05	MC_05_532	AcademicOffice	0	380	8.858	2.824	1.332	
MC	MC_05	MC_05_533	AcademicOffice	0	380	6.289	3.229	2.079	
MC	MC_05	MC_05_542	AcademicOffice	0	380	7.25	5.861	3.542	
MC	MC_05	MC_05_543	AcademicOffice	0	380	9.547	6.297	4.247	
MC	MC_05	MC_05_544	AcademicOffice	0	380	15.097	12.026	9.711	
MC	MC_05	MC_05_545	AcademicOffice	0	380	13.913	10.234	7.947	
MC	MC_05	MC_05_546	AcademicOffice	0	380	16.911	12.003	8.4	
MC	MC_05	MC_05_548	AcademicOffice	0	380	6.655	5.255	3.005	
MC	MC_06	MC_06_600	AcademicOffice	0	380	6.061	1.126	0.379	
MC	MC_06	MC_06_602	AcademicOffice	0	380	9.905	3.955	2.053	
MC	MC_06	MC_06_604	AcademicOffice	0	380	2.482	2.132	0.829	
MC	MC_06	MC_06_610	AcademicOffice	0	380	30.566	24.458	14.853	
MC	MC_06	MC_06_611	AcademicOffice	0	380	9.2	5.721	3.9	
MC	MC_06	MC_06_612	AcademicOffice	0	380	7.297	5.974	3.589	
MC	MC_06	MC_06_620	AcademicOffice	0	380	9.605	7.853	6.218	
MC	MC_06	MC_06_623	AcademicOffice	0	380	14.442	12.426	9.682	
MC	MC_06	MC_06_624	AcademicOffice	0	380	14.674	12.861	10.279	
MC	MC_06	MC_06_631	AcademicOffice	0	380	19.597	10.189	6.318	
MC	MC_06	MC_06_644	AcademicOffice	0.13	380	10.829	7.85	5.832	
ML	ML_01	ML_01_11	UndergradLab	0.16	328	1.534	1.311	1.041	
ML	ML_02	ML_02_24	Classroom	0	248	1.768	1.528	1.284	
ML	ML_02	ML_02_25	AssemblyExhibiti	0.3	96	1.976	1.336	0.553	
ML	ML_03	ML_03_34	Classroom	0.45	144	1.353	1.124	0.857	
ML	ML_03	ML_03_35	Classroom	0.16	216	0.98	0.723	0.414	
NH	NH_01	NH_01_102	AdministrativeOf	0	380	4.586	2.362	0.637	
NH	NH_02	NH_02_241	HealthServices	0	300	2.753	1.893	1.077	
NH	NH_02	NH_02_242	HealthServices	0	380	2.667	1.895	1.384	
NH	NH_02	NH_02_244	AdministrativeOf	0	380	1.766	1.56	1.299	
NH	NH_03	NH_03_300	AdministrativeOf	0.12	380	9.181	7.147	5.602	
NH	NH_03	NH_03_304	AdministrativeOf	0.09	380	2.504	1.632	1.228	
NH	NH_03	NH_03_307	AdministrativeOf	0	380	25.711	10.537	5.608	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
OPT	OPT_01	OPT_01_112	Classroom	0	176	2.213	1.585	1.246	
OPT	OPT_02	OPT_02_23	ResearchLab	0	380	3.739	1.166	0.239	
OPT	OPT_03	OPT_03_30	Classroom	0.11	39	-1.368	-1.252	-1.049	
OPT	OPT_03	OPT_03_34	Classroom	0.2	144	0.665	0.391	0.357	
OPT	OPT_04	OPT_04_40	UndergradLab	0	296	2.046	1.734	1.386	
OPT	OPT_04	OPT_04_43	UndergradLab	0	32	2.083	1.854	1.49	
OPT	OPT_04	OPT_04_45	AcademicOffice	0	380	14.474	6.874	4.763	
PAS	PAS_01	PAS_01_110	ResearchLab	0.17	16	1.625	1.188	0.76	
PAS	PAS_01	PAS_01_122	Classroom	0.15	224	1.383	1.173	0.822	
PAS	PAS_01	PAS_01_123	UndergradLab	0	142	1.018	0.908	0.667	
PAS	PAS_01	PAS_01_124	Classroom	0	144	2.809	2.435	1.822	
PAS	PAS_02	PAS_02_208	Classroom	0	321	1.586	1.327	0.956	
PAS	PAS_02	PAS_02_243	AcademicOffice	0	380	22.868	16.242	11.303	
PAS	PAS_03	PAS_03_302	AcademicOffice	0	116	0.976	0.934	0.951	
PAS	PAS_04	PAS_04_403	AcademicOffice	0	114	0.97	1.002	1.067	
PHR	PHR_01	PHR_01_100	Classroom	0	308	3.386	2.957	1.638	
PHR	PHR_01	PHR_01_101	Classroom	0	120	-0.32	-0.275	-0.221	
PHR	PHR_02	PHR_02_201	UndergradLab	0.02	120	21.181	17.785	12.84	
PHR	PHR_02	PHR_02_202	Classroom	0.12	36	2.44	2.03	1.609	
PHR	PHR_06	PHR_06_601	AcademicOffice	0	380	4.958	3.332	2.379	
PHY	PHY_01	PHY_01_14	Classroom	0.39	228	2.281	1.775	0.999	
PHY	PHY_01	PHY_01_15	Classroom	0	244	0.08	0.013	-0.044	
PHY	PHY_02	PHY_02_20	AcademicOffice	0.44	380	9.521	7.851	5.983	
PHY	PHY_02	PHY_02_22	ResearchLab	0	380	10.479	5.6	2.289	
PHY	PHY_02	PHY_02_23	Classroom	0.07	260	1.358	1.099	0.732	
PHY	PHY_03	PHY_03_30	MixedUse	0.47	32	1.583	1.02	0.449	
PHY	PHY_03	PHY_03_31	Classroom	0.03	381	1.983	1.422	0.683	
PHY	PHY_03	PHY_03_32	ResearchLab	0	380	1.241	0.982	0.645	
PHY	PHY_03	PHY_03_35	AcademicOffice	0.43	8	3.396	1.917	1.125	
PHY	PHY_04	PHY_04_40	AcademicOffice	0.11	380	2.336	1.329	0.863	
QNC	QNC_01	QNC_01_150	Classroom	0	344	2.262	1.824	1.365	
QNC	QNC_02	QNC_02_250	Classroom	0	232	0.98	0.866	0.752	
QNC	QNC_02	QNC_02_261	UndergradLab	0	84	1.061	0.978	0.798	
RCH	RCH_01	RCH_01_10	Classroom	0.03	346	1.796	1.358	0.745	
RCH	RCH_01	RCH_01_11	Classroom	0.28	276	1.019	0.799	0.517	
RCH	RCH_02	RCH_02_20	Classroom	0	344	1.92	1.485	0.916	
RCH	RCH_02	RCH_02_21	Classroom	0.34	315	1.524	1.184	0.686	
RCH	RCH_03	RCH_03_30	Classroom	0.03	360	1.805	1.373	0.998	
SCH	SCH_01	SCH_01_11	Bookstore	0.29	380	11.499	4.871	1.711	
SCH	SCH_01	SCH_01_12	MixedUse	0.48	380	10.032	4.934	2.208	

Building	Building_Floor	Building_Zone	ClusterNames	COU16Pct	nbr_obs_Users	coeff_Users	coeff_Users30	coeff_Users60	Sparklines
SCH	SCH_02	SCH_02_20	FoodServices	0.18	380	5.559	2.41	1.04	
SLC	SLC_00	SLC_00_010	Bookstore	0.01	380	5.076	4.036	2.242	
SLC	SLC_00	SLC_00_013	Bookstore	0	380	7.088	4.597	2.671	
SLC	SLC_01	SLC_01_111	FoodServices	0.01	380	21.618	11.241	3.408	
SLC	SLC_01	SLC_01_112	FoodServices	0	380	23.496	15.882	2.189	
SLC	SLC_02	SLC_02_210	CommonUse	0	380	23.439	15.532	7.213	
SLC	SLC_03	SLC_03_310	StudySpace	0.03	348	21.871	18.011	10.434	
STC	STC_01	STC_01_101	Classroom	0.02	380	112.65	68.613	37.155	
STC	STC_02	STC_02_203	AcademicOffice	0	380	17.425	13.257	8.378	
TC	TC_01	TC_01_120	AdministrativeOf	0	380	0.369	0.241	0.076	
TC	TC_01	TC_01_121	AdministrativeOf	0	380	20.829	6.616	3.655	
TC	TC_02	TC_02_210	AdministrativeOf	0	380	0.954	0.788	0.595	
TC	TC_03	TC_03_311	AdministrativeOf	0	380	8.307	6.999	5.391	

Appendix F: Comparison of Coefficient of Correlation Values for User vs Expected Occupants, by Building Floors and Building Floor Zones.

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
AL	AL_00	2002	AL_00_00	MixedUse	80	80	1.030	1.030	-
AL	AL_01	1965	AL_01_10	MixedUse	660	160	1.660	1.230	N
AL	AL_01	1965	AL_01_11	MixedUse	660	276	1.660	1.560	N
AL	AL_01	1965	AL_01_12	MixedUse	660	224	1.660	0.960	N
AL	AL_02	1965	AL_02_20	Classroom	377	102	0.720	-0.001	N
AL	AL_02	1965	AL_02_21	Classroom	377	275	0.720	1.170	Y
ARC	ARC_01	1965	ARC_01_100	UndergradLab	248	96	0.660	0.310	N
ARC	ARC_01	1965	ARC_01_110	UndergradLab	248	152	0.660	0.590	N
ARC	ARC_02	2004	ARC_02_200	Library	1212	380	1.380	2.320	Y
ARC	ARC_02	2004	ARC_02_201	Library	1212	380	1.380	4.320	Y
ARC	ARC_02	2004	ARC_02_202	Library	1212	72	1.380	1.380	N
ARC	ARC_02	2004	ARC_02_210	Library	1212	380	1.380	5.280	Y
ARC	ARC_03	2004	ARC_03_300	UndergradLab	504	392	1.190	0.280	N
ARC	ARC_03	2004	ARC_03_310	UndergradLab	504	112	1.190	1.290	Y
B1	B1_02	2004	B1_02_27	ResearchLab	984	224	1.370	1.320	N
B1	B1_02	2004	B1_02_28	ResearchLab	984	380	1.370	15.880	Y
B1	B1_02	2004	B1_02_29	ResearchLab	984	380	1.370	8.800	Y
B1	B1_03	1964	B1_03_37	UndergradLab	168	168	1.140	1.140	-
B2	B2_01	1964	B2_01_15	ResearchLab	96	96	0.830	0.830	-
B2	B2_02	1967	B2_02_25	ResearchLab	380	380	8.320	8.320	-
B2	B2_03	1967	B2_03_35	ResearchLab	404	404	1.360	1.360	-
BMH	BMH_01	1967	BMH_01_100	ResearchLab	1376	104	-0.940	2.070	Y
BMH	BMH_01	1967	BMH_01_101	ResearchLab	1376	232	-0.940	1.550	Y
BMH	BMH_01	1967	BMH_01_161	ResearchLab	1376	380	-0.940	8.010	Y
BMH	BMH_01	1967	BMH_01_162	ResearchLab	1376	124	-0.940	1.820	Y
BMH	BMH_01	1967	BMH_01_168	ResearchLab	1376	380	-0.940	112.400	Y
BMH	BMH_01	1967	BMH_01_170	ResearchLab	1376	156	-0.940	0.740	Y
BMH	BMH_02	1972	BMH_02_230	ResearchLab	1500	380	1.930	10.060	Y
BMH	BMH_02	1972	BMH_02_240	ResearchLab	1500	116	1.930	1.140	N
BMH	BMH_02	1972	BMH_02_268	ResearchLab	1500	380	1.930	9.550	Y
BMH	BMH_02	1972	BMH_02_269	ResearchLab	1500	380	1.930	4.700	Y
BMH	BMH_02	1972	BMH_02_270	ResearchLab	1500	244	1.930	1.120	N
BMH	BMH_03	1972	BMH_03_302	UndergradLab	508	128	0.230	0.840	Y
BMH	BMH_03	1972	BMH_03_368	UndergradLab	508	380	0.230	14.930	Y
BMH	BMH_04	1972	BMH_04_468	ResearchLab	380	380	14.010	14.010	-
C2	C2_00	1972	C2_00_06	ResearchLab	380	380	7.050	7.050	-
C2	C2_01	1971	C2_01_16	UndergradLab	380	380	6.770	6.770	-
C2	C2_02	1971	C2_02_27	UndergradLab	760	380	1.060	0.990	N
C2	C2_02	1971	C2_02_28	UndergradLab	760	380	1.060	4.140	Y

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
C2	C2_03	1971	C2_03_38	AcademicOffice	377	377	10.720	10.720	-
CIF	CIF_01	1971	CIF_01_121	Athletics	380	380	9.380	9.380	-
COM	COM_01	1987	COM_01_11	AdministrativeOf	380	380	1.440	1.440	-
CPH	CPH_01	1966	CPH_01_132	UndergradLab	878	378	1.860	11.370	Y
CPH	CPH_01	1966	CPH_01_133	UndergradLab	878	380	1.860	1.400	N
CPH	CPH_01	1966	CPH_01_134	UndergradLab	878	120	1.860	1.960	Y
CPH	CPH_03	1972	CPH_03_360	UndergradLab	440	192	1.100	-5.160	N
CPH	CPH_03	1972	CPH_03_368	UndergradLab	440	248	1.100	0.670	N
CPH	CPH_04	1972	CPH_04_430	UndergradLab	398	380	1.740	9.170	Y
CPH	CPH_04	1972	CPH_04_433	UndergradLab	398	18	1.740	-7.860	N
DC	DC_01	1972	DC_01_135	ResearchLab	1471	360	2.160	2.230	Y
DC	DC_01	1972	DC_01_155	ResearchLab	1471	380	2.160	7.650	Y
DC	DC_01	1972	DC_01_156	ResearchLab	1471	380	2.160	2.170	Y
DC	DC_01	1972	DC_01_170	ResearchLab	1471	351	2.160	2.390	Y
DMS	DMS_01	1987	DMS_01_100	Classroom	380	380	6.370	6.370	-
DMS	DMS_02	2012	DMS_02_201	AcademicOffice	380	380	5.800	5.800	-
DMS	DMS_03	2012	DMS_03_312	Classroom	380	380	6.180	6.180	-
DWE	DWE_01	2012	DWE_01_150	UndergradLab	340	200	1.520	1.450	N
DWE	DWE_01	2012	DWE_01_151	UndergradLab	340	140	1.520	-0.260	N
DWE	DWE_02	1958	DWE_02_251	Classroom	688	380	1.250	19.510	Y
DWE	DWE_02	1958	DWE_02_252	Classroom	688	308	1.250	1.670	Y
DWE	DWE_03	1958	DWE_03_350	UndergradLab	992	380	1.190	10.190	Y
DWE	DWE_03	1958	DWE_03_351	UndergradLab	992	348	1.190	1.220	Y
DWE	DWE_03	1958	DWE_03_352	UndergradLab	992	264	1.190	1.200	Y
E2	E2_01	1958	E2_01_177	CommonUse	760	380	4.420	17.510	Y
E2	E2_01	1958	E2_01_178	CommonUse	760	380	4.420	9.190	Y
E2	E2_02	1961	E2_02_235	UndergradLab	162	72	1.320	-6.120	N
E2	E2_02	1961	E2_02_236	UndergradLab	162	90	1.320	1.890	Y
E2	E2_03	1961	E2_03_333	UndergradLab	384	108	1.210	0.290	N
E2	E2_03	1961	E2_03_334	UndergradLab	384	108	1.210	2.080	Y
E2	E2_03	1961	E2_03_335	UndergradLab	384	168	1.210	0.720	N
E3	E3_02	1961	E3_02_210	ResearchLab	904	380	0.140	0.230	Y
E3	E3_02	1961	E3_02_211	ResearchLab	904	144	0.140	0.300	Y
E3	E3_02	1961	E3_02_212	ResearchLab	904	380	0.140	1.850	Y
E3	E3_03	1961	E3_03_310	UndergradLab	856	380	0.630	2.760	Y
E3	E3_03	1961	E3_03_311	UndergradLab	856	380	0.630	11.560	Y
E3	E3_03	1961	E3_03_316	UndergradLab	856	96	0.630	-0.230	N
E5	E5_03	1961	E5_03_303	Classroom	564	380	1.310	2.380	Y
E5	E5_03	1961	E5_03_310	Classroom	564	184	1.310	1.140	N
E5	E5_04	2010	E5_04_412	Classroom	224	224	0.180	0.180	-

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
E5	E5_06	2010	E5_06_600	Classroom	332	332	1.610	1.610	-
E6	E6_02	2010	E6_02_202	AcademicOffice	256	256	0.380	0.380	-
E6	E6_04	2011	E6_04_402	Classroom	184	184	1.030	1.030	-
E7	E7_01	2011	E7_01_142	Classroom	380	380	44.210	44.210	-
E7	E7_02	2018	E7_02_230	Classroom	328	157	20.880	6.400	N
E7	E7_02	2018	E7_02_240	Classroom	328	171	20.880	25.840	Y
E7	E7_03	2018	E7_03_331	ResearchLab	1116	152	52.600	9.130	N
E7	E7_03	2018	E7_03_333	ResearchLab	1116	171	52.600	8.900	N
E7	E7_03	2018	E7_03_334	ResearchLab	1116	153	52.600	120.580	Y
E7	E7_03	2018	E7_03_335	ResearchLab	1116	155	52.600	184.090	Y
E7	E7_03	2018	E7_03_340	ResearchLab	1116	153	52.600	8.920	N
E7	E7_03	2018	E7_03_341	ResearchLab	1116	171	52.600	11.170	N
E7	E7_03	2018	E7_03_345	ResearchLab	1116	161	52.600	34.410	N
E7	E7_04	2018	E7_04_404	Classroom	1014	161	-17.080	177.280	Y
E7	E7_04	2018	E7_04_405	Classroom	1014	380	-17.080	81.710	Y
E7	E7_04	2018	E7_04_431	Classroom	1014	153	-17.080	9.070	Y
E7	E7_04	2018	E7_04_441	Classroom	1014	159	-17.080	65.320	Y
E7	E7_04	2018	E7_04_443	Classroom	1014	161	-17.080	62.210	Y
E7	E7_05	2018	E7_05_534	ResearchLab	1127	155	-29.910	167.090	Y
E7	E7_05	2018	E7_05_535	ResearchLab	1127	161	-29.910	78.570	Y
E7	E7_05	2018	E7_05_541	ResearchLab	1127	155	-29.910	7.530	Y
E7	E7_05	2018	E7_05_542	ResearchLab	1127	167	-29.910	13.310	Y
E7	E7_05	2018	E7_05_543	ResearchLab	1127	153	-29.910	7.180	Y
E7	E7_05	2018	E7_05_544	ResearchLab	1127	171	-29.910	3.610	Y
E7	E7_05	2018	E7_05_545	ResearchLab	1127	165	-29.910	9.380	Y
E7	E7_06	2018	E7_06_631	ResearchLab	805	169	10.380	8.680	N
E7	E7_06	2018	E7_06_640	ResearchLab	805	153	10.380	17.230	Y
E7	E7_06	2018	E7_06_641	ResearchLab	805	171	10.380	10.350	N
E7	E7_06	2018	E7_06_643	ResearchLab	805	159	10.380	4.790	N
E7	E7_06	2018	E7_06_644	ResearchLab	805	153	10.380	11.280	Y
E7	E7_07	2018	E7_07_731	AcademicOffice	1432	171	-3.760	3.020	Y
E7	E7_07	2018	E7_07_732	AcademicOffice	1432	152	-3.760	4.380	Y
E7	E7_07	2018	E7_07_733	AcademicOffice	1432	167	-3.760	9.570	Y
E7	E7_07	2018	E7_07_734	AcademicOffice	1432	153	-3.760	39.250	Y
E7	E7_07	2018	E7_07_740	AcademicOffice	1432	150	-3.760	4.620	Y
E7	E7_07	2018	E7_07_741	AcademicOffice	1432	171	-3.760	4.510	Y
E7	E7_07	2018	E7_07_742	AcademicOffice	1432	169	-3.760	6.780	Y
E7	E7_07	2018	E7_07_744	AcademicOffice	1432	146	-3.760	3.010	Y
E7	E7_07	2018	E7_07_745	AcademicOffice	1432	153	-3.760	22.920	Y
ECH	ECH_01	2018	ECH_01_10	UndergradLab	1214	380	1.120	2.640	Y

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
ECH	ECH_01	2018	ECH_01_120	UndergradLab	1214	112	1.120	0.840	N
ECH	ECH_01	2018	ECH_01_121	UndergradLab	1214	158	1.120	1.620	Y
ECH	ECH_01	2018	ECH_01_122	UndergradLab	1214	144	1.120	1.260	Y
ECH	ECH_01	2018	ECH_01_123	UndergradLab	1214	40	1.120	1.430	Y
ECH	ECH_01	2018	ECH_01_13	UndergradLab	1214	380	1.120	2.170	Y
EIT	EIT_01	1959	EIT_01_100	UndergradLab	807	120	1.670	1.840	Y
EIT	EIT_01	1959	EIT_01_101	UndergradLab	807	340	1.670	1.580	N
EIT	EIT_01	1959	EIT_01_102	UndergradLab	807	347	1.670	2.070	Y
EIT	EIT_02	2003	EIT_02_201	ResearchLab	192	96	1.250	0.350	N
EIT	EIT_02	2003	EIT_02_205	ResearchLab	192	96	1.250	0.970	N
EIT	EIT_03	2003	EIT_03_314	Classroom	240	84	1.060	3.840	Y
EIT	EIT_03	2003	EIT_03_315	Classroom	240	156	1.060	1.100	Y
EIT	EIT_05	2003	EIT_05_500	ResearchLab	758	380	2.630	3.370	Y
EIT	EIT_05	2003	EIT_05_501	ResearchLab	758	378	2.630	4.100	Y
ESC	ESC_01	2003	ESC_01_14	UndergradLab	84	84	-0.080	-0.080	-
ESC	ESC_02	1964	ESC_02_25	AcademicOffice	27	27	11.790	11.790	-
ESC	ESC_03	1964	ESC_03_32	UndergradLab	1163	380	0.730	9.430	Y
ESC	ESC_03	1964	ESC_03_33	UndergradLab	1163	368	0.730	11.070	Y
ESC	ESC_03	1964	ESC_03_34	UndergradLab	1163	36	0.730	0.790	Y
ESC	ESC_03	1964	ESC_03_35	UndergradLab	1163	379	0.730	18.860	Y
EV1	EV1_01	1964	EV1_01_13	UndergradLab	296	296	1.180	1.180	-
EV1	EV1_02	1965	EV1_02_22	MixedUse	252	76	0.780	-0.400	N
EV1	EV1_02	1965	EV1_02_24	MixedUse	252	176	0.780	1.090	Y
EV1	EV1_03	1965	EV1_03_35	AcademicOffice	208	208	1.380	1.380	-
EV2	EV2_01	1965	EV2_01_100	UndergradLab	611	208	0.880	0.820	N
EV2	EV2_01	1965	EV2_01_101	UndergradLab	611	23	0.880	1.290	Y
EV2	EV2_01	1965	EV2_01_102	UndergradLab	611	380	0.880	5.480	Y
EV2	EV2_02	1981	EV2_02_200	Classroom	276	276	0.630	0.630	-
EV3	EV3_03	1981	EV3_03_341	Classroom	224	224	0.930	0.930	-
FED	FED_01	2011	FED_01_103	FoodServices	756	376	0.740	1.220	Y
FED	FED_01	2011	FED_01_110	FoodServices	756	380	0.740	2.180	Y
GSC	GSC_01	1984	GSC_01_10	PlantMaintenanc	2280	380	3.040	19.600	Y
GSC	GSC_01	1984	GSC_01_11	PlantMaintenanc	2280	380	3.040	3.940	Y
GSC	GSC_01	1984	GSC_01_110	PlantMaintenanc	2280	380	3.040	14.670	Y
GSC	GSC_01	1984	GSC_01_112	PlantMaintenanc	2280	380	3.040	3.370	Y
GSC	GSC_01	1984	GSC_01_114	PlantMaintenanc	2280	380	3.040	6.100	Y
GSC	GSC_01	1984	GSC_01_116	PlantMaintenanc	2280	380	3.040	11.210	Y
GSC	GSC_02	1966	GSC_02_22	AdministrativeOf	760	380	-0.990	0.830	Y
GSC	GSC_02	1966	GSC_02_26	AdministrativeOf	760	380	-0.990	3.270	Y
HH	HH_01	1966	HH_01_10	UndergradLab	1874	380	2.090	20.930	Y

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
HH	HH_01	1966	HH_01_11	UndergradLab	1874	173	2.090	0.900	N
HH	HH_01	1966	HH_01_110	UndergradLab	1874	383	2.090	1.450	N
HH	HH_01	1966	HH_01_13	UndergradLab	1874	296	2.090	1.410	N
HH	HH_01	1966	HH_01_15	UndergradLab	1874	240	2.090	2.360	Y
HH	HH_01	1966	HH_01_16	UndergradLab	1874	378	2.090	2.910	Y
HH	HH_01	1966	HH_01_18	UndergradLab	1874	24	2.090	1.410	N
HH	HH_02	1968	HH_02_210	Classroom	666	308	2.810	2.520	N
HH	HH_02	1968	HH_02_22	Classroom	666	224	2.810	1.000	N
HH	HH_02	1968	HH_02_25	Classroom	666	134	2.810	1.570	N
HH	HH_03	1968	HH_03_33	Classroom	472	312	1.300	1.050	N
HH	HH_03	1968	HH_03_34	Classroom	472	160	1.300	0.420	N
HS	HS_01	1968	HS_01_130	HealthServices	760	380	20.950	13.320	N
HS	HS_01	1968	HS_01_141	HealthServices	760	380	20.950	5.690	N
IHB	IHB_03	1968	IHB_03_302	OtherSpace	380	380	7.720	7.720	-
LIB	LIB_01	2009	LIB_01_13	Library	380	380	7.980	7.980	-
MC	MC_01	1965	MC_01_100	OtherSpace	2412	373	1.410	5.760	Y
MC	MC_01	1965	MC_01_101	OtherSpace	2412	380	1.410	4.210	Y
MC	MC_01	1965	MC_01_102	OtherSpace	2412	305	1.410	3.000	Y
MC	MC_01	1965	MC_01_105	OtherSpace	2412	320	1.410	1.980	Y
MC	MC_01	1965	MC_01_106	OtherSpace	2412	380	1.410	3.780	Y
MC	MC_01	1965	MC_01_107	OtherSpace	2412	380	1.410	3.680	Y
MC	MC_01	1965	MC_01_108	OtherSpace	2412	274	1.410	1.060	N
MC	MC_02	1967	MC_02_201	CommonUse	2217	392	1.640	1.720	Y
MC	MC_02	1967	MC_02_202	CommonUse	2217	380	1.640	5.520	Y
MC	MC_02	1967	MC_02_203	CommonUse	2217	344	1.640	2.320	Y
MC	MC_02	1967	MC_02_204	CommonUse	2217	377	1.640	9.810	Y
MC	MC_02	1967	MC_02_205	CommonUse	2217	380	1.640	1.550	N
MC	MC_02	1967	MC_02_206	CommonUse	2217	344	1.640	1.610	N
MC	MC_03	1967	MC_03_300	UndergradLab	960	176	1.460	1.040	N
MC	MC_03	1967	MC_03_302	UndergradLab	960	24	1.460	9.080	Y
MC	MC_03	1967	MC_03_304	UndergradLab	960	380	1.460	11.610	Y
MC	MC_03	1967	MC_03_305	UndergradLab	960	380	1.460	7.180	Y
MC	MC_04	1967	MC_04_400	Classroom	2276	380	1.380	15.910	Y
MC	MC_04	1967	MC_04_401	Classroom	2276	380	1.380	10.550	Y
MC	MC_04	1967	MC_04_402	Classroom	2276	404	1.380	1.600	Y
MC	MC_04	1967	MC_04_404	Classroom	2276	356	1.380	1.580	Y
MC	MC_04	1967	MC_04_405	Classroom	2276	400	1.380	1.010	N
MC	MC_04	1967	MC_04_406	Classroom	2276	356	1.380	1.360	N
MC	MC_05	1967	MC_05_500	AcademicOffice	7200	373	8.320	8.700	Y
MC	MC_05	1967	MC_05_501	AcademicOffice	7200	380	8.320	6.570	N

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
MC	MC_05	1967	MC_05_503	AcademicOffice	7200	380	8.320	10.960	Y
MC	MC_05	1967	MC_05_504	AcademicOffice	7200	380	8.320	8.190	N
MC	MC_05	1967	MC_05_510	AcademicOffice	7200	380	8.320	12.940	Y
MC	MC_05	1967	MC_05_512	AcademicOffice	7200	380	8.320	8.630	Y
MC	MC_05	1967	MC_05_520	AcademicOffice	7200	380	8.320	9.120	Y
MC	MC_05	1967	MC_05_522	AcademicOffice	7200	380	8.320	7.060	N
MC	MC_05	1967	MC_05_523	AcademicOffice	7200	367	8.320	7.710	N
MC	MC_05	1967	MC_05_524	AcademicOffice	7200	380	8.320	11.290	Y
MC	MC_05	1967	MC_05_530	AcademicOffice	7200	380	8.320	5.890	N
MC	MC_05	1967	MC_05_532	AcademicOffice	7200	380	8.320	8.850	Y
MC	MC_05	1967	MC_05_533	AcademicOffice	7200	380	8.320	6.280	N
MC	MC_05	1967	MC_05_542	AcademicOffice	7200	380	8.320	7.250	N
MC	MC_05	1967	MC_05_543	AcademicOffice	7200	380	8.320	9.540	Y
MC	MC_05	1967	MC_05_544	AcademicOffice	7200	380	8.320	15.090	Y
MC	MC_05	1967	MC_05_545	AcademicOffice	7200	380	8.320	13.910	Y
MC	MC_05	1967	MC_05_546	AcademicOffice	7200	380	8.320	16.910	Y
MC	MC_05	1967	MC_05_548	AcademicOffice	7200	380	8.320	6.650	N
MC	MC_06	1967	MC_06_600	AcademicOffice	4180	380	12.240	6.060	N
MC	MC_06	1967	MC_06_602	AcademicOffice	4180	380	12.240	9.900	N
MC	MC_06	1967	MC_06_604	AcademicOffice	4180	380	12.240	2.480	N
MC	MC_06	1967	MC_06_610	AcademicOffice	4180	380	12.240	30.560	Y
MC	MC_06	1967	MC_06_611	AcademicOffice	4180	380	12.240	9.200	N
MC	MC_06	1967	MC_06_612	AcademicOffice	4180	380	12.240	7.290	N
MC	MC_06	1967	MC_06_620	AcademicOffice	4180	380	12.240	9.600	N
MC	MC_06	1967	MC_06_623	AcademicOffice	4180	380	12.240	14.440	Y
MC	MC_06	1967	MC_06_624	AcademicOffice	4180	380	12.240	14.670	Y
MC	MC_06	1967	MC_06_631	AcademicOffice	4180	380	12.240	19.590	Y
MC	MC_06	1967	MC_06_644	AcademicOffice	4180	380	12.240	10.820	N
ML	ML_01	1967	ML_01_11	UndergradLab	328	328	1.530	1.530	-
ML	ML_02	1962	ML_02_24	Classroom	344	248	1.460	1.760	Y
ML	ML_02	1962	ML_02_25	Classroom	344	96	1.460	1.970	Y
ML	ML_03	1962	ML_03_34	Classroom	360	144	1.090	1.350	Y
ML	ML_03	1962	ML_03_35	Classroom	360	216	1.090	0.980	N
NH	NH_01	1962	NH_01_102	AdministrativeOf	380	380	4.580	4.580	-
NH	NH_02	1972	NH_02_241	HealthServices	1060	300	1.210	2.750	Y
NH	NH_02	1972	NH_02_242	HealthServices	1060	380	1.210	2.660	Y
NH	NH_02	1972	NH_02_244	HealthServices	1060	380	1.210	1.760	Y
NH	NH_03	1972	NH_03_300	AdministrativeOf	1140	380	0.870	9.180	Y
NH	NH_03	1972	NH_03_304	AdministrativeOf	1140	380	0.870	2.500	Y
NH	NH_03	1972	NH_03_307	AdministrativeOf	1140	380	0.870	25.710	Y

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
OPT	OPT_01	1972	OPT_01_112	Classroom	176	176	2.210	2.210	-
OPT	OPT_02	1973	OPT_02_23	ResearchLab	380	380	3.730	3.730	-
OPT	OPT_03	1973	OPT_03_30	Classroom	183	39	2.280	-1.360	N
OPT	OPT_03	1973	OPT_03_34	Classroom	183	144	2.280	0.660	N
OPT	OPT_04	1973	OPT_04_40	UndergradLab	708	296	1.690	2.040	Y
OPT	OPT_04	1973	OPT_04_43	UndergradLab	708	32	1.690	2.080	Y
OPT	OPT_04	1973	OPT_04_45	UndergradLab	708	380	1.690	14.470	Y
PAS	PAS_01	1973	PAS_01_110	UndergradLab	526	16	1.530	1.620	Y
PAS	PAS_01	1973	PAS_01_122	UndergradLab	526	224	1.530	1.380	N
PAS	PAS_01	1973	PAS_01_123	UndergradLab	526	142	1.530	1.010	N
PAS	PAS_01	1973	PAS_01_124	UndergradLab	526	144	1.530	2.800	Y
PAS	PAS_02	1972	PAS_02_208	Classroom	701	321	1.420	1.580	Y
PAS	PAS_02	1972	PAS_02_243	Classroom	701	380	1.420	22.860	Y
PAS	PAS_03	1972	PAS_03_302	AcademicOffice	116	116	0.970	0.970	-
PAS	PAS_04	1972	PAS_04_403	AcademicOffice	114	114	0.970	0.970	-
PHR	PHR_01	1972	PHR_01_100	Classroom	428	308	2.870	3.380	Y
PHR	PHR_01	1972	PHR_01_101	Classroom	428	120	2.870	-0.320	N
PHR	PHR_02	2009	PHR_02_201	UndergradLab	156	120	-0.160	21.180	Y
PHR	PHR_02	2009	PHR_02_202	UndergradLab	156	36	-0.160	2.440	Y
PHR	PHR_06	2009	PHR_06_601	AcademicOffice	380	380	4.950	4.950	-
PHY	PHY_01	2009	PHY_01_14	Classroom	472	228	1.170	2.280	Y
PHY	PHY_01	2009	PHY_01_15	Classroom	472	244	1.170	0.080	N
PHY	PHY_02	1959	PHY_02_20	ResearchLab	1020	380	1.480	9.520	Y
PHY	PHY_02	1959	PHY_02_22	ResearchLab	1020	380	1.480	10.470	Y
PHY	PHY_02	1959	PHY_02_23	ResearchLab	1020	260	1.480	1.350	N
PHY	PHY_03	1959	PHY_03_30	ResearchLab	801	32	2.050	1.580	N
PHY	PHY_03	1959	PHY_03_31	ResearchLab	801	381	2.050	1.980	N
PHY	PHY_03	1959	PHY_03_32	ResearchLab	801	380	2.050	1.240	N
PHY	PHY_03	1959	PHY_03_35	ResearchLab	801	8	2.050	3.390	Y
PHY	PHY_04	1959	PHY_04_40	AcademicOffice	380	380	2.330	2.330	-
QNC	QNC_01	1959	QNC_01_150	Classroom	344	344	2.260	2.260	-
QNC	QNC_02	2011	QNC_02_250	UndergradLab	316	232	2.100	0.980	N
QNC	QNC_02	2011	QNC_02_261	UndergradLab	316	84	2.100	1.060	N
RCH	RCH_01	2011	RCH_01_10	Classroom	622	346	1.540	1.790	Y
RCH	RCH_01	2011	RCH_01_11	Classroom	622	276	1.540	1.010	N
RCH	RCH_02	1967	RCH_02_20	Classroom	659	344	1.920	1.920	N
RCH	RCH_02	1967	RCH_02_21	Classroom	659	315	1.920	1.520	N
RCH	RCH_03	1967	RCH_03_30	Classroom	360	360	1.800	1.800	-
SCH	SCH_01	1967	SCH_01_11	MixedUse	760	380	14.430	11.490	N
SCH	SCH_01	1967	SCH_01_12	MixedUse	760	380	14.430	10.030	N

Building	Bldg_Flr	Const Year	Building_Zone	ClusterNames	nbr_Obs_Floor	nbr_Obs_Zone	c_Users_Flr	c_Users_Zon	Zones_Corr_Higher
SCH	SCH_02	1967	SCH_02_20	FoodServices	380	380	5.550	5.550	-
SLC	SLC_00	1967	SLC_00_010	Bookstore	760	380	8.090	5.070	N
SLC	SLC_00	1967	SLC_00_013	Bookstore	760	380	8.090	7.080	N
SLC	SLC_01	1967	SLC_01_111	FoodServices	760	380	6.320	21.610	Y
SLC	SLC_01	1967	SLC_01_112	FoodServices	760	380	6.320	23.490	Y
SLC	SLC_02	1967	SLC_02_210	CommonUse	380	380	23.430	23.430	-
SLC	SLC_03	1967	SLC_03_310	StudySpace	348	348	21.870	21.870	-
STC	STC_01	1967	STC_01_101	Classroom	380	380	112.650	112.650	-
STC	STC_02	2015	STC_02_203	AcademicOffice	380	380	17.420	17.420	-
TC	TC_01	2015	TC_01_120	AdministrativeOf	760	380	0.040	0.360	Y
TC	TC_01	2015	TC_01_121	AdministrativeOf	760	380	0.040	20.820	Y
TC	TC_02	2002	TC_02_210	AdministrativeOf	380	380	0.950	0.950	-
TC	TC_03	2002	TC_03_311	AdministrativeOf	380	380	8.300	8.300	-