INSY 442 - 001
Data Analysis and Visualization

Submitted to: Professor Geneviève Bassellier
Addressed to Teaching and Learning Services McGill

April 10, 2024

Group 5

Blue Namba - 260958599
Gia Luong - 261119805
Khushi Toprani - 261018346
Timmy-Tin Tran - 260388719
Yanfei Cao – 261014438

1. Client Description

For this project, our clients are undergraduate students at McGill, they are from a diverse range of academic disciplines and levels of study. These students, who are at different stages of their undergraduate journey, share a common interest in the quality and equity of their educational environments.

The project utilized classroom data from McGill University's Teaching and Learning Services (TLS), which is dedicated to enhancing the educational experience by designing and maintaining learning spaces that foster active engagement. TLS's initiatives range from renovating classrooms with cutting-edge tools to reconfiguring spaces for interactive learning, all aimed at enhancing the teaching and learning environment.

2. Business Problem and Context

Recognizing the vital role of TLS in student success, we focused our analysis on whether the distribution and accessibility of renovated learning spaces across all faculties for undergraduate students was equitable. In this context, we defined equitable as fair and equal allocation of renovated classrooms to undergraduate students at McGill University, irrespective of their faculty, field of study, or year in their academic journey. This meant consistent access to renovated spaces, that are assumed to be modernized and improved learning spaces.

Guiding Questions

- Are certain faculties more likely to have renovated classrooms?
- Is there a pattern in when classrooms are renovated (e.g., more incidents, specific years with trends, locations)

Dashboard specific questions

Our dashboard creation was guided by a question-first methodology to ensure that each visualization directly addressed our guiding questions:

1. How are undergraduate students distributed across faculties at McGill in 2023?
2. What is the probability a student will have a renovated classroom, based on their faculty?
3. What type of classroom is more likely to get renovated?
4. Does the number of incidents cause the likelihood for renovations?

3. Data Description

The TLS dataset provides a holistic view of McGill's academic infrastructure, with information on property inventory of rooms and buildings, faculty, students, renovation records, incidents reported, and service requests. We focused on four key tables: MEETTIME, ALL_INSTRUCTORS, LSRS Data - V2, and ITSM Query. While the former have complete datasets, LSRS Data - V2 and ITSM Query tables presented challenges with many missing or 'N/A' values, notably within the 'renovation year' column, leading to an incomplete dataset. Due to this, we had to exclude values which reduced the data dimensionality and impacted the scope of our findings.

In our analysis, we encountered inconsistencies across tables; similar data were formatted differently, complicating the integration process. For example, the ITSM Query and LSRS Data – V2 tables, was an example where building and room information could only be aligned by synthesizing a new column that

concatenated building names with room numbers, facilitating a connection to the ConcatBuilding column from LSRS Data – V2. This lack of standardized formatting across tables necessitated a manual approach to data integration. To work around the locked Power Query on existing data tables, we took the initiative to incorporate supplementary data files processed through Excel, to enhance the overall data model.

4. Findings

We translated our analysis into fifteen visualizations, using nine distinct types of visualizations, with each crafted to answer a facet of our guiding questions. Our aim was to create a suite of four interconnected dashboards, each offering clear insights into specific, measurable aspects of the dashboard-specific questions, laid out in the Business Problem and Context section of our report. This approach allowed us to paint a complete picture of the data landscape from 2008 to 2019 and the analysis was anchored in the renovation data provided by the LSRS table, particularly utilizing the Renovation Year column to quantify and track renovations.

Our dashboards were designed from the standpoint of a McGill student, prioritizing ease of use and relevance. We employed interactive elements such as slicers and dropdown menus, ensuring users can tailor the information to their unique queries. A navigation system was also integrated to facilitate smooth transitions between different perspectives within our analysis.

Dashboard 1: Summary

How are Undergraduate students distributed across faculties at McGill in 2023?

To get a better understanding of a student's experience with renovated classrooms, we first wanted to have a general analysis of the McGill student body and how they are distributed across faculties, subjects, courses, and buildings.

We found that out of all the McGill undergraduate students, they have classes in 46 buildings in 413 classrooms. To then visualize the number of students in each faculty, we created a bar chart with the number of students registered for courses in the different faculties. Because we did not have data on the specific number of students belonging to each faculty, we had to estimate it with the number of course registrations. We found that the three largest faculties are the Faculty of Science, Arts, and Engineering. The faculty of science has the highest membership with 83,954 students registered in classes in 2023. The faculty with the lowest membership is the faculty of Arts and Science with 102 students registered in classes.

The next step to gain an overview of students' experience with classrooms was to look at the number of students in each building and classroom. Leacock is by far the building with the highest number of students, with 24,000 seats, followed by McConnell Engineering at 13,000 and Stewart Biology at 11,000. More specifically the most used classrooms are Leacock 132 with 15,000 students, Adams auditorium with 10,000 students and Stewart Bio with 8,000 students.

The initial analysis helped us understand if there are disproportionate experiences among different Faculties. Moving forward, we can ask ourselves questions like, are buildings with higher student

capacities likely to get renovated, or are students in smaller faculties less likely to see their classrooms renovated?

Dashboard 2: Renovation Probability

What is the probability a student will have a renovated classroom, based on their faculty?

In our analysis to determine the probability of students attending classes in renovated classrooms across different faculties, we created two measures: Renovation Classroom Count, by conditionally counting not "N/A" or blank values in the RENOVATION YEAR column in the LSRS Data-V2 table, and Total Classrooms Count by counting distinct BLDGC+ROOMC entries in the MEETTIME table.

To assess the equity of the student experience across faculties, we compared the ratio of Renovation Classroom Count over Total Classrooms Count, indicating the likelihood that a student will be in a renovated classroom during their undergraduate studies.

Over the span of 11 years (from 2008 to 2019), the overall probability that a classroom is renovated across all faculties is 17%, which reflects a small share of classrooms. A student in the Faculty of Education has the highest chance of being in a renovated classroom, with a 34% likelihood. For the Faculty of Agricultural and Environmental Sciences, Faculty of Arts, Faculty of Science, and Faculty of Engineering, the probability is around 23%. However, The Faculty of Arts and Faculty of Science have the highest numbers of total classrooms, hence it indicates that the number of renovated classrooms does not correlate directly with the total number of classrooms within each faculty. On the other hand, students in the Desautels Faculty of Management have the lowest likelihood, with a 10% chance of being in a renovated classroom.

These figures suggest there is an imbalance in the distribution of renovated classrooms, with the Faculty of Education being the most advantaged and the Desautels Faculty Management being the least. This disparity indicates that not all students equally benefit from renovated and improved learning environments. Ultimately, this discrepancy could affect the student's overall educational experience and lead to differences in engagement and focus. Thus, our findings highlight areas of improvement in facility management and resource allocation to achieve a more equitable student experience.

Dashboard 3: Renovation

What type of classroom is more likely to get renovated?

The renovation dashboard encapsulates an insightful overview of the renovation trends across various classrooms between 2008 and 2019, revealing that 94 classrooms underwent renovations during this period in total.

From the number of renovations for each type of classroom, we found that classrooms (traditional lecture spaces) are more frequently renovated with 57 renovated rooms. Auditoriums, which accommodate larger student groups, were the next most renovated space, with 10 refurbishments. Active Learning Classrooms (ALC), which are designed to foster a more engaging learning environment, also saw significant updates, with 9 renovations. Other specialized spaces, such as teaching labs and computer labs, witnessed a lower number of renovations, 6 and 4 respectively, indicating that laboratories specific to a few faculties are renovated less frequently compared to general-use classrooms.

The time series analysis of the renovations delineates fluctuating patterns with peaks in 2012 and 2018. The peak in 2012, with 18 renovations, followed by a gradual decrease and subsequent rise to 14 in 2018, could indicate cycles of funding, strategic planning cycles, or responses to emerging pedagogical trends.

Moreover, a capacity analysis for renovated classrooms shows that most renovations have been carried out in smaller classrooms, accommodating fewer than 100 students, which suggests a potential focus on small to medium-sized classrooms. The scatter plot clusters notably thin out beyond the 200-student capacity mark, indicating such large-capacity spaces are less frequently chosen for renovations.

Dashboard 4: Incidents
Does the number of incidents cause the likelihood for renovations?

We aim to explore the relationship between the number of incidents and the likelihood of renovated classrooms. In other words, we would like to see if the number of incidents is the reason for renovating classrooms across McGill University.

Based on the new dataset being derived from joining ITSM Query tables (Number of incidents, building, room number) with LSRS Data – V2 (Year of Innovation), we found that there are 62 out of 94 renovated buildings that have incidents. In other words, there are 32 buildings that were renovated despite having incidents reported. Additionally, there are only incidents codes that are recorded instead of details about the incidents, we assume the reported incidents can be solved without the need of renovation.

The azure map pictures the location of buildings at McGill downtown campus and denotes the different number of incidents across faculties. We find that Sciences & Arts faculty has the highest number of incidents and Agriculture & Environmental Sciences has the least incidents. By identifying hotspots where more incidents occur, we can look for patterns/common issues emerging and prioritize allocating resources in these buildings. Additionally, clustered column chart breaks down into top 10 rooms with the highest number of incidents despite their faculties. This chart shows one classroom can have up to 130 incidents.

According to the clustered bar chart, the number of incidents outweighs the number of renovated classrooms per faculty. Sherbrooke 680 has the highest reported incidents with the highest number of renovated classrooms of 9, whereas Parasitology and CINE buildings report 0 incidents and 0 renovated classrooms. It seems like the more incidents there are, the more renovated classrooms there are in a building. However, when we look at the top 5 buildings with most incidents as reference, there's a big difference in the total reported incidents while the difference in the number of renovated classrooms is very small. The data given is not significant for us to support our argument.

The ratio suggests that there should be more renovations across faculties such as Sciences & Arts (1079 incidents with only 10 renovated classrooms), Engineering (1050 incidents with only 14 renovated classrooms, Art (967 incidents with 11 renovated classrooms), etc. as they have the highest number of incidents being reported. We can see that faculties with more incidents should have more renovated classrooms (e.g. Sciences & Arts, Engineering). However, it is difficult to assume incidents are the main reason to renovate classrooms, as we see in our probability dashboard that the faculties with the most

incidents are not necessarily the most likely to be renovated. For instance, students in Education are more likely to encounter renovated classrooms (34%). However, the incidents dashboard suggests that Sciences & Arts buildings should have more renovations as they have the highest number of incidents. There are two possible reasons explaining this contradiction. Firstly, we lack data on the number of renovations per year in LSRS Data-V2. Secondly, reported incidents do not have a causal relationship with the likelihood of renovation.

5. Recommendations

Drawing on the analysis through the dashboards, we can conclude that McGill University's classroom renovation efforts, while substantial, indicate an uneven distribution of modernized learning spaces. While faculties, like the Faculty of Education, benefit from a higher likelihood of renovated spaces, others, particularly the Desautels Faculty of Management, are less favored. Thus, we have three recommendations for McGill to recalibrate its approach to ensure fair allocation of renovated classrooms for all faculties:

1. Optimize Renovation Resources: The TLS should target renovation efforts towards spaces with high traffic and incident rates, such as the Leacock and McConnell Engineering buildings, to benefit a larger segment of the student population.
2. Enhance Data Collection: To empower more precise and comprehensive analysis, it is crucial to improve data integrity and standardization, particularly for the 'renovation year' data. Currently limited by the use of only 94 renovation entries, future data gathering efforts should be more exhaustive, including details on the types of renovations and incidents to foster deeper insights and stronger recommendations.
3. Integrate Detailed Student Membership Data: Incorporate a data table reflecting the number of students actively associated with each faculty, rather than merely the number of course registrations. This will provide a more accurate estimate of students consistently impacted by renovations within their respective faculties, thus allowing for a more nuanced assessment of renovation needs and priorities.

By implementing these recommendations, McGill University can create a fair distribution of renovated learning environments. A more equitable approach to updating learning environments is not only essential for leveling the academic playing field but also for affirming the university's commitment to providing an exceptional educational experience for all students.

Appendix

Section 1: Dashboards and Visualizations

Dashboard	Visualizations	Business Question	Difficulties
Summary	1. Stacked bar chart 2. Tree map 3. Cards (2)	How are Undergraduate students distributed across faculties at McGill in 2023?	Unremovable duplicated rows as Power Query is locked, see Section 2A for detailed challenges
Probability	4. Clustered column chart 5. Funnel 6. Slicer 7. Card	What is the probability students will have a renovated classroom?	Lack of direct relationships, see Section 2B for detailed challenges
Renovation	8. Tree map 9. Area chart 10. Scatter chart 11. Card	What type of classroom is more likely to get renovated?	Number of missing values outweighs usable data, see Section 2C
Incident	12. Clustered bar chart 13. Azure map 14. Clustered column chart	Does the number of incidents cause the likelihood for renovations?	Missing values, see Section 2D for details Longitudes and Latitudes were not provided, see Section 2D Missing and confusing data see Section 2D for details

Section 2: List of Challenges

2A. Summary Visualizations Dashboard

- There were many duplicate rows with the same primary key (CRN and Term) in the MEETIME dataset which inflated the number of students registered in each class. Because we did not have access to Power Query, we had to delete the duplicates in a separate Excel file, reimport it and recreate the connections to other tables.

2B. Probability Calculation Dashboard

- The relationship between the MEETTIME table, which held faculty and classroom information, and the LSRS Data-V2 table, containing renovation details, was indirect. This necessitated the creation of calculated measures or the use of virtual relationships to bridge the data.
- Explanation for two measures we create:
    - Renovation Classroom Count: the count of renovated classrooms within each faculty. To compute this, we used a DAX formula that employed a conditional count to only include classrooms that were not marked as "N/A" or blank in the renovation year column, ensuring accurate representation of available renovated spaces.
    - Total Classrooms Count: formulated to tally the total number of classrooms available for each faculty by counting distinct BLDGC+ROOMC entries in the MEETTIME table.

With these foundational measures established, we proceeded to calculate the ratio of renovated to total classrooms for each faculty using a simple division in DAX.

2C. Renovations Dashboard

- The renovation dataset does not provide details on the types of renovations performed. This lack of granularity limited our ability to analyze the impact of different renovation types on the educational environment and prevented us from making targeted recommendations based on the type of incident or improvement.

2D. Incidents Dashboard

- Power Query is locked, so we must export the necessary data and manually modify/join datasets on Excel to reimport into PowerBI.
- The number of students registered does not represent the students taking the course, as you can register for one and then drop it. As a result, the number of students has inflated.
- We identified instances within the dataset where identical Course Reference Numbers (CRNs) paired with the same term exhibit varying enrollment figures. This anomaly could indicate several scenarios such as multiple sections of a course offered within the term that are not distinctly separated in the data, changes in enrollment over time not captured discretely, or potential inaccuracies within the data recording process. The inconsistencies in data reduced the potential of targeted Fall and Winter specific recommendations.
- Longitudes and Latitudes are missing in the provided data. We geocoded the specific buildings' locations ourselves and then imported them with an excel file, creating a relationship with the TLS dataset's buildings column.
- The granularity of the dataset with regard to the types of reported incidents was limited, with no specifics provided that would allow us to identify patterns or predict reactive measures such as renovations. Detailed information on incident types could have been valuable; for instance, categorizing an incident like the asbestos discovery in February 2023 as a safety issue could have enabled us to anticipate and analyze subsequent renovations. Such predictive analysis would be instrumental in understanding how certain types of incidents potentially trigger prioritization in the renovation process.

2E. Overall Challenges

- Hard to collaborate effectively with most team members using Power BI for the first time.
- Finding a color scheme that is visually appealing while portraying every element rightfully.
- Balancing quantity and quality of information while keeping the dashboards not too busy.
- Staying in the scope of our research questions was hard, we often found interesting information that wasn't useful to us.
- The number of missing and N/A values outweighs that of useful data which makes our conclusion less significant.

Section 3: References

"Geocoder.ca: Geocoding for North America - USA and Canada." *Geocoder.ca*, geocoder.ca/?locate=&geoit=GeoCode. Accessed 6 Apr. 2024.

"McGill Interactive Accessible Network Map." *Maps.mcgill.ca*, maps.mcgill.ca/.

"McGill University Closes 3 Buildings, Moves Classes Online after Asbestos Found." *Montreal*, 1 Feb. 2023, montreal.ctvnews.ca/mcgill-university-closes-3-buildings-moves-classes-online-after-asbestos-found-1.6255683. Accessed 6 Apr. 2024.