MAST90106

Project Proposal

Optimisation of university space based on a supply and demand analysis

A case study on staff meeting rooms and student toilet facilities on campus

THE UNIVERSITY OF MELBOURNE

Group 3

Abhinav Sharma, 1009225

sharma.a1@student.unimelb.edu.au

Advait Deshpande, 1005024

advaitd@student.unimelb.edu.au

Xinyi Xu, 900966

xinyix4@student.unimelb.edu.au

Yanming Wang, 881330

yanmwang@student.unimelb.edu.au

Contents

1 Introduction **1**

2 Related Work **2**

3 Data Analysis **3**

3.1 Exploratory Data Analysis 3

3.1.1 What are the different properties, size and scale of data? 3

3.1.2 How is the data distributed across problem statement? 4

3.1.3 How is the data connected with the problem? 4

3.2 Data Preprocessing 5

3.3 Data Correlations 6

3.3.1 Correlations between supply of meeting rooms and demand by employees . . . 6

3.3.2 Correlations between supply of toilet facilities and demand by students 7

3.4 Spatial Data Analysis 8

4 Proposed Methods **10**

4.1 Method 1: Finding the most optimal nearest building 10

4.2 Method 2: Finding the most optimal nearest floor for a building 11

5 Timeline **13**

1 Introduction

The university has spent its second-largest expense to space allocation and the arrangement of meeting rooms and toilets has long been recognised as a major concern in campus planning. It is important to ensure optimal space utilisation as under-utilisation of these facilities entails extra cost penalties for maintenance. In this project, the space arrangement of staff meeting rooms and student toilets will be optimised by proposing solutions that can be efficiently use the current supply of resources. Generally, the number of existing meeting rooms and toilets is considered as “supply” while the number of staffs and enrolled students are considered as “demand”. By analysing space, employee and timetabling data, we will firstly explore if the supply meets current demand, then propose different predictive models that will help our client to suggest the usage of current resources more efficiently.

Our client for this project is the Spatial Analytics and Space Management department of the University of Melbourne. This department works in future space design, better space allocation and optimising usage of resources for the university. As stated by our client, the expected outcome from this project can be summarized as:

- We need to suggest how the **space arrangement** of meeting rooms and toilets can be optimised, and advise how the overall space on campus can be better planned.
- We are expected to deliver a **detailed analysis report**. The analysis is expected to be conducted by campus, by building, and by different meeting room and toilet types, etc. The report should include interactive maps, charts with interpretation of findings.
- We need to use different **analytical methods** such as spatial analysis, correlation analysis, etc.
- We need to provide **reasonable recommendations** of space optimisation opportunities based on the analysis.

From the data science perspective, this project involves extensive exploratory data analysis of supply and demand, complicated data mutations, joins and preprocessing. We also need to perform correlation analysis of different factors and spatial analysis using QGIS. In order to suggest current usage of resources more effectively, we need to do predictive statistical modelling using well-defined constraints. This kind of modelling poses an integration challenge of python models with QGIS spatial layers. We also need to make our models extremely generic so that they can provide support for analyzing any building in any campus. Moreover, identifying appropriate factors for correlation analysis from the provided data is a difficult and daunting task that we'll be exploring in this project.

This report is organised as follows: Section 2 describes relevant work. We present the results of our basic analysis using the provided datasets, including data preprocessing, correlations and spatial analysis in section 3. Section 4 discusses our solutions for suggesting how to use the current supply of resources more efficiently. Finally, a summarized timeline of part-1 and a clear plan for next semester is discussed in section 5.

2 Related Work

In this section, we have researched multiple literature resources which tackle similar problems. The optimization of meeting rooms and toilets allocation is a classical scheduling problem and considered to be NP-complete[1]. Since students' activities are limited by their selected subjects, staff are usually based in a single location, and other than the physical distance, the actual utilization of meeting rooms and toilets should also be considered. Our goal is taking all of those variables into consideration and see how the arrangement and maintenance services can be optimised given current supply and demand.

Ahmed Wasfy and Fadi A. Aloul proposed a complete approach using integer linear programming (ILP) to solve similar scheduling problem[2]. The goal of this algorithm is satisfying all of the university's constraints to assigning courses to classrooms and optimizing the utilization of existing facilities effectively and efficiently.

Burke et al.[3] also proposed an integer linear programming method to solve this problem. This method decomposes the problem into multiple sub-problems. In each sub-problem, only one part of the optimization constraint or criteria is considered. These sub-problems will be used to work outbounds in their respective optimization criterion. Finally, the ultimate solution to the problem is incorporated by each solution of the sub-problems.

Vermuyten et al.[4] proposed a two-stage approach to optimize demand allocation using integer linear programming. Firstly, the approach focuses on event assignments to rooms and time slots, and the second focuses on the reassignments of meeting rooms to events with previous time slots from the previous stage.

Beyrouthy et al.[5] researched when planning to build a campus, how teaching space can be utilized to improve usage efficiency. The study shows that most rooms have overcapacity, which means the supply (seat number) is greater than the demand (students' number).

Song et al.[6] proposed three stages iterative algorithm, the three stages are initialization, intensification, and diversification. This algorithm is used to solve the course scheduling problem. The initialization stage is to use a greedy algorithm to find a roughly feasible solution to schedule the maximum possible events given specific constraints and variables. The second stage is to use a simulated annealing method to find a local optimal solution. The diversification stage implements random perturbations (exchange some rooms) to improve the result. The final result developed by this algorithm is used as the new initial scenario and implement these stages iteratively.

Greedy heuristics have been implemented to course timetabling and meeting room scheduling problems successfully. For example, "Greedy Randomized Adaptive Search Procedure (GRASP)" [7]. The greedy algorithm is designed through a cost function that takes all constraints violated into accounts. However, this algorithm cannot guarantee to reach an optimal solution. Furthermore, this algorithm is not considering room optimization as its constrains.

3 Data Analysis

The goal of this project is to analyse and optimise the current supply and demand of meeting rooms and toilets on campus. We have been provided with the following data to perform this analysis:

#	Dataset Name	Dataset Description
1	`uom-space`	Space metadata of all rooms across campuses and buildings
2	`rm-category-type`	Definition of all UoM standard room categories and types
3	`fl-name`	Dataset to provide more information about building floors
4	`av-equipment`	Audio Visual equipment data including its location information
5	`em-location`	De-identified employee/staff location data
6	`2020-timetable-v2`	Latest class scheduling data including a class's time and its location
7	`meeting-room-usage`	Collected data of meeting room usage

Table 1: Provided datasets

3.1 Exploratory Data Analysis

In this section, we have explored all the provided datasets to understand their properties, size and scale. We have also performed an analysis of how these datasets correlate with the provided problem.

3.1.1 What are the different properties, size and scale of data?

After importing all the provided datasets using `pandas` python package, we deduced following summary of the data as shown in the below tables.

#	Dataset Name	Dataset Description
1	`uom-space`	7 campuses, 331 buildings, 28 floor codes, 5703 rooms, 185 room types
2	`rm-category-type`	209 different room types
3	`fl-name`	floor information of all possible floor codes
4	`av-equipment`	1964 equipments, 32 manufacturers in 11 campuses across 142 buildings
5	`em-location`	7709 employees across 130 buildings and 1565 room codes
6	`2020-timetable-v2`	52 departments, 1577 modules across 248 activity dates
7	`meeting-room-usage`	890 meeting rooms across 8 campuses, 125 buildings

Table 2: Important categorical variables data summary

#	Dataset Name	Dataset Description
1	`uom-space`	Room Capacity: 0-599 with an μ of 4.0627 and σ of 17.2592
2	`uom-space`	Room Area m^2: 0.22-5696.90 with an μ of 30.70 and σ of 118.3070
3	`2020-timetable-v2`	Planned Size: 0-684 with average of 50 students
4	`2020-timetable-v2`	Class Duration(min): 30-675 with an average of 94.336
5	`meeting-room-usage`	Meetings: 0-1000 with an average of 241 meetings

Table 3: Important numerical variables data summary

3.1.2 How is the data distributed across problem statement?

In our analysis, we explored different categorical and numerical features of the datasets to get more depth about the data. Initially, we saw that most of the data is provided for the `Parkville` campus as shown in Figure 1. Due to this huge skewness, we have performed our data and correlation analysis on the Parkville campus in this report, which can be easily extended to other campuses in the future phase of this project.

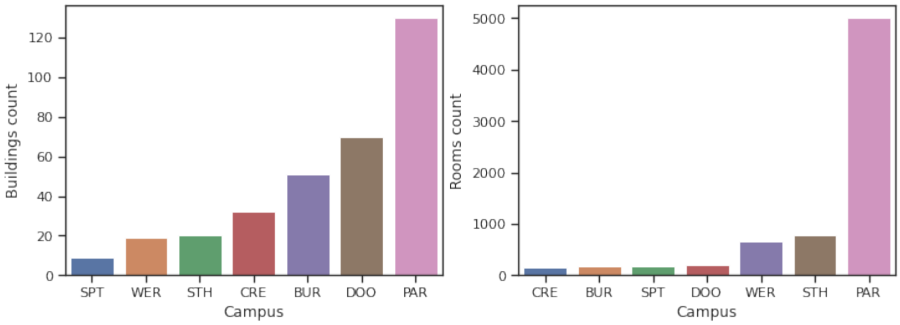

Figure 1: Distribution of buildings and rooms across campus

Using room category data and merging it with space metadata, we were able to figure out the distribution of meeting rooms and toilet facilities across buildings in the Parkville campus as shown in Figure 2. We saw that `333 Exhibition st` buildings have the highest number of meeting rooms and `The spot` building has the highest number of toilet facilities.

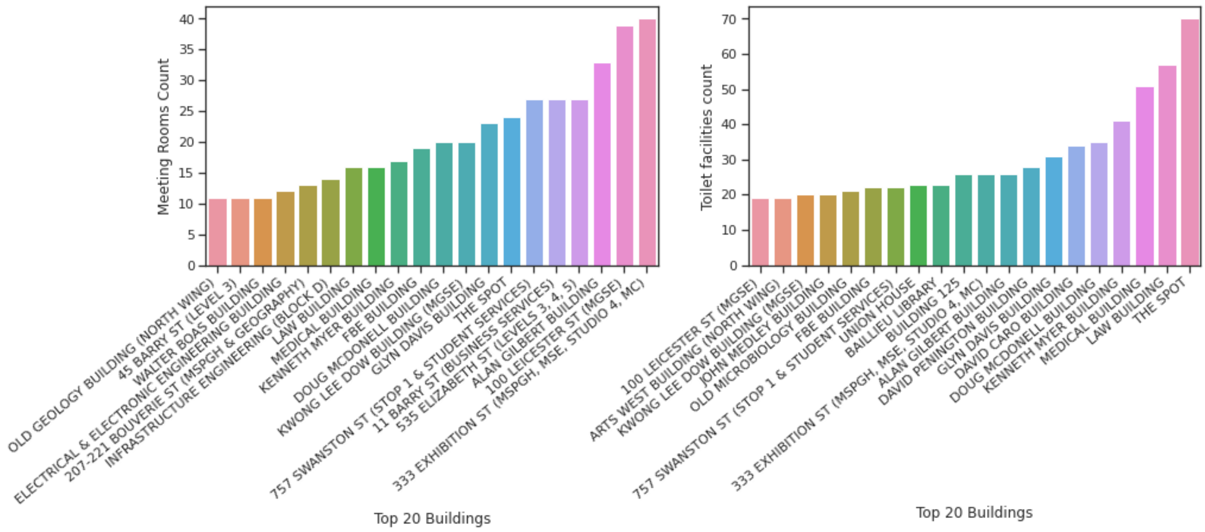

Figure 2: Distribution of meeting rooms and toilet facilities across buildings

3.1.3 How is the data connected with the problem?

After exploring different properties and aspects of the data, we tried to see and understand how data can be used to get the basic intuition of the problem.

To do that, we created several supply-demand plots across various covariates which helped to see the need for space optimization. Initially, we created supply-demand plots based on the very trivial preference of staff members trying to book a meeting room in the same building where they are located. This plot is shown in Figure 3. As per the plot, we can see space optimization problem in terms of supply and demand proportion especially in the `Law Building`, `Doug Mcdonell building` and `Kenneth Myer building`.

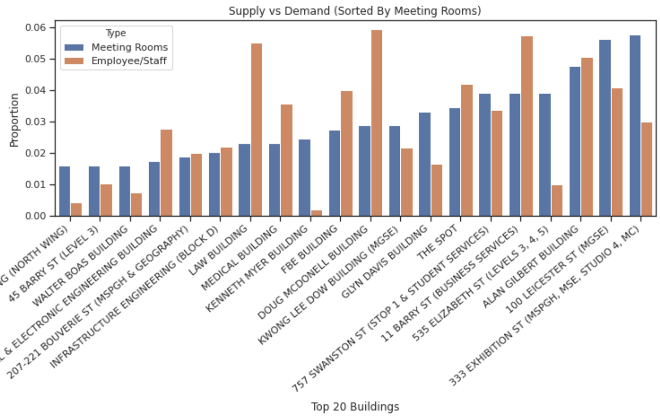

Figure 3: Supply vs Demand of meeting rooms across buildings

Similarly, we created a supply-demand plot on the same trivial preference of students trying to access a toilet facility in the same building which is shown in Figure 4. Again, we can see that the space optimization problem in terms of supply and demand proportion, especially for `Redmond barry building`.

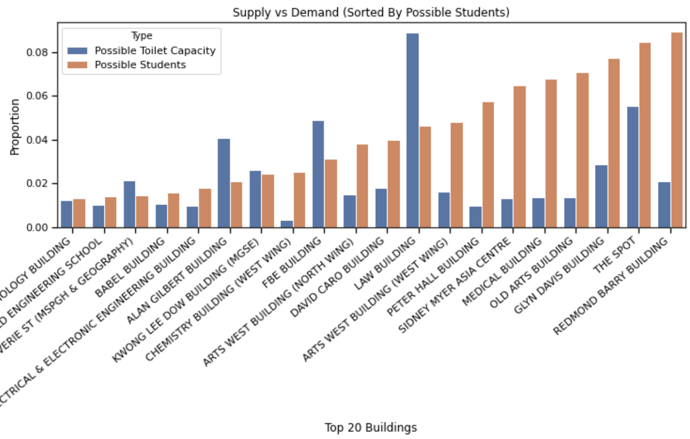

Figure 4: Supply vs Demand of toilet facilities across buildings

We have explored several other covariates concerning supply-demand in Section 3.3 which will eventually help us to perform space optimization on the underlying problem.

3.2 Data Preprocessing

Data Preprocessing is one of the main components of Data Analysis as it enhances the data quality of data. If the data is not cleaned or normalised, it may lead to wrong model selection and prediction outputs. This includes the cleaning of data, data transformation and feature selections. The primary aim of preprocessing is to minimise or, eventually, eliminate those small data contributions associated with the experimental error if they are systematic, like for example, baseline drifts, or random, for example, the noise contributions related to the instrumental measures[8]. It helps to declutter the data which would help us to define the features of the model.

We approached data preprocessing by converting all the data to lower case for the text values which makes it easier in the later part of the model selection. We also removed null values and added default values to the null values which helped us to analyze and remove unnecessary columns. Moreover, we eliminated duplicate rows and unwanted records like *online* and *off-site* classes option in the timetable data.

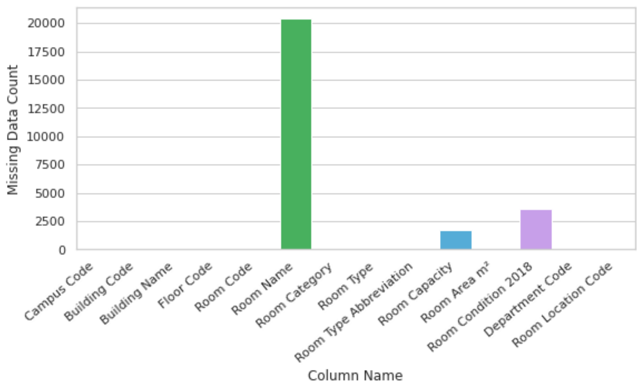

Figure 5: Missing data summary of the timetable dataset

3.3 Data Correlations

Correlation is an important tool used in Machine Learning for feature extraction on which all the models are based. It involves measuring the relationship between two or more variables in the data set. Quality of the data plays a crucial role in determining the correlations. We will explain different correlations we have identified in the below section.

3.3.1 Correlations between supply of meeting rooms and demand by employees

In this section, the supply and demand analysis is carried on meeting rooms and employees.

1. If the staff wishes to book a meeting room in the same building where they work, it is observed that building on *333 Exhibition St* has the highest number of meeting rooms(**29**) and it accommodates only 3% of the total employee from the data set.

2. It can be inferred that number of employees located on a particular floor and the corresponding meeting rooms on the floor have a supply-demand problem. For example, **Doug Mcdonell** building has 20% meeting rooms on level 8 out of total meeting rooms in the building for which there is 10% staff located on that level.

3. Meeting room usage also depends upon the condition of the room: Out of total meeting rooms across the parkville campus, 81% of them are in *excellent* condition while 2% are in very-good condition, 1% poor and 17% good.

4. Meeting room usage and booking can be related to the usage of the meeting room and if they are located at a longer distance from the employee location. **Stop 1** has the most number of excellent rooms based on the usage data and out of all the meeting rooms, only 3% can be booked. Buildings with low demand have low usage of the meeting rooms as shown in Figure 6.

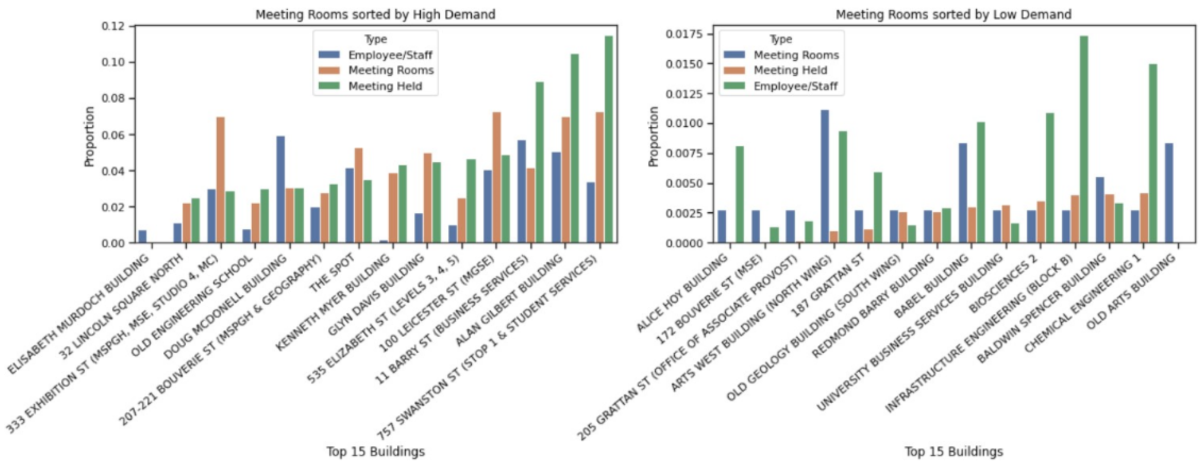

Figure 6: Supply vs demand analysis with meeting room usage correlation

Few correlations are not that strong and can be mentioned as part of the findings of the correlation between data sets. Average attendance, capacity of the meeting room, time of the meeting and the attendance of the meeting are some of the other factors for correlation analysis. An employee would like to book a room with better equipment to enhance the experience.

3.3.2 Correlations between supply of toilet facilities and demand by students

In this section, we explore different preferences and factors that correlates with the problem of supply-demand analysis for the toilet facilities.

1. Students usually prefer accessing the toilet facilities which are in good condition and in the same building where their respective classes are conducted. We can observe this correlation with respect to supply demand as shown in Figure 7. It can be seen that **Redmond Barry** has the most number of classes with 8% of total student and it has mere 2% of the total excellent toilet capacity to hold possible students. In contrast, Law building has approximately 12% of the total toilet capacity and around 3% of total students are attending classes in that building.

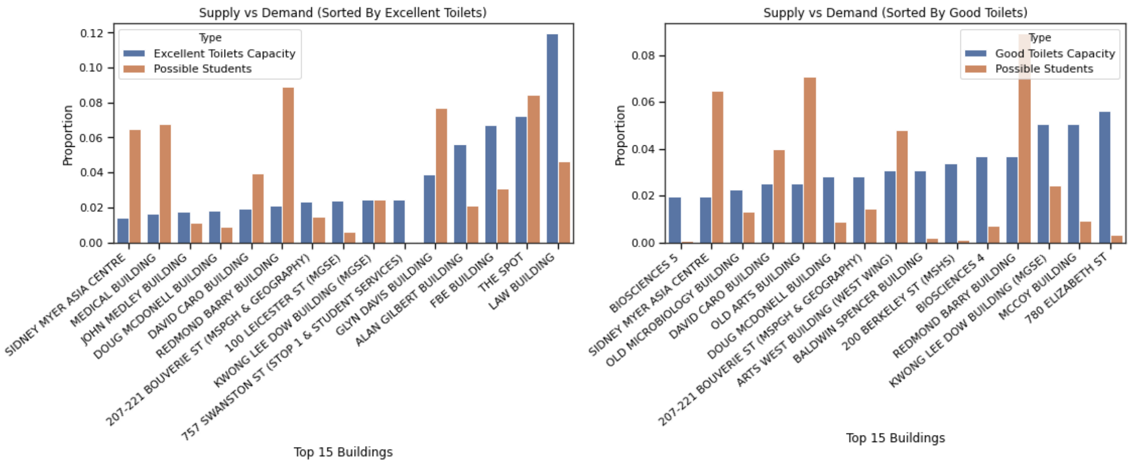

Figure 7: Supply vs demand analysis with toilet conditions correlation

2. Students prefer to use the washrooms on the same floor where the lecture hall is situated. Considering Medical Building, it is observed that 30% of the toilet capacity is available on level 1 while only 5% of students have lectures on the same floor.

3. Students can decide to avail the facilities based on the size of the washroom. It is evident that Law building has 5% of total students and the size of the facilities are not enough with respect to the demand. In contrast, Arts West building has larger toilet facility sizes as compared to the student demand proportion(4.5%).

4. Duration of classes is also an important correlation to consider in terms of supply and demand of toilet facilities. If classes have longer duration in a building, then that building should have facilities with better capacity. As per our analysis, it can be seen that Old Physics building has very long duration of lectures but it has 1% of the total capacity of toilets.

3.4 Spatial Data Analysis

We observed in the previous section that the number of staffs and meeting rooms varies from building to building. This presents an opportunity for us to consider if the current allocation for meeting rooms can offer flexibility for staffs (i.e. if meeting rooms for that building are fully booked, can the staff find another meeting room nearby?). In this section, we map the staffs and meeting rooms data with building outlines and building GPS using QGIS.

QGIS provides visualisation of the "heat-map" of the average meeting room capacity per staff, which is calculated using the total capacity of meeting rooms divided by the amount of staff as shown in Figure 10. The dark (blue) cell represents high availability and the green colour indicates that either the meeting room or staff number is zero. Using this analysis, we can deduce following results:

- `Doug McDonell building(168)`, `11 Barry Street(266)` and `Law building(106)` have a significant number of staffs and therefore stronger demand for meeting rooms. The following analysis assumes that 3-minute walk is the distance people are willing to walk. And based on the average walking speed of an adult, a three-minute walk is about 240 meters.

- For staffs in `Doug McDonell(168)`, `Old Engineering School(173)`and `Walter Boas(163)` are merely 150 meters away and both buildings have a comparatively small number of staffs and high meeting room capacities.
- `11 Barry Street(266)` is near `Law building(106)`, and other surrounding buildings such as `Kwong Lee Dow building(263)`, `The Spot(110)` and `FBE building(105)` also have lots of staffs.
- Also, all of these buildings have low ratios, indicating that those staffs may need to use meeting rooms in other buildings as well. Choices are including `MDHS(207)`, `Statistical Consulting Center(394)` and `University Business Service building(384)`, however, their meeting rooms capacity is around 20, and it might be difficult for them to support those high-demand buildings nearby.

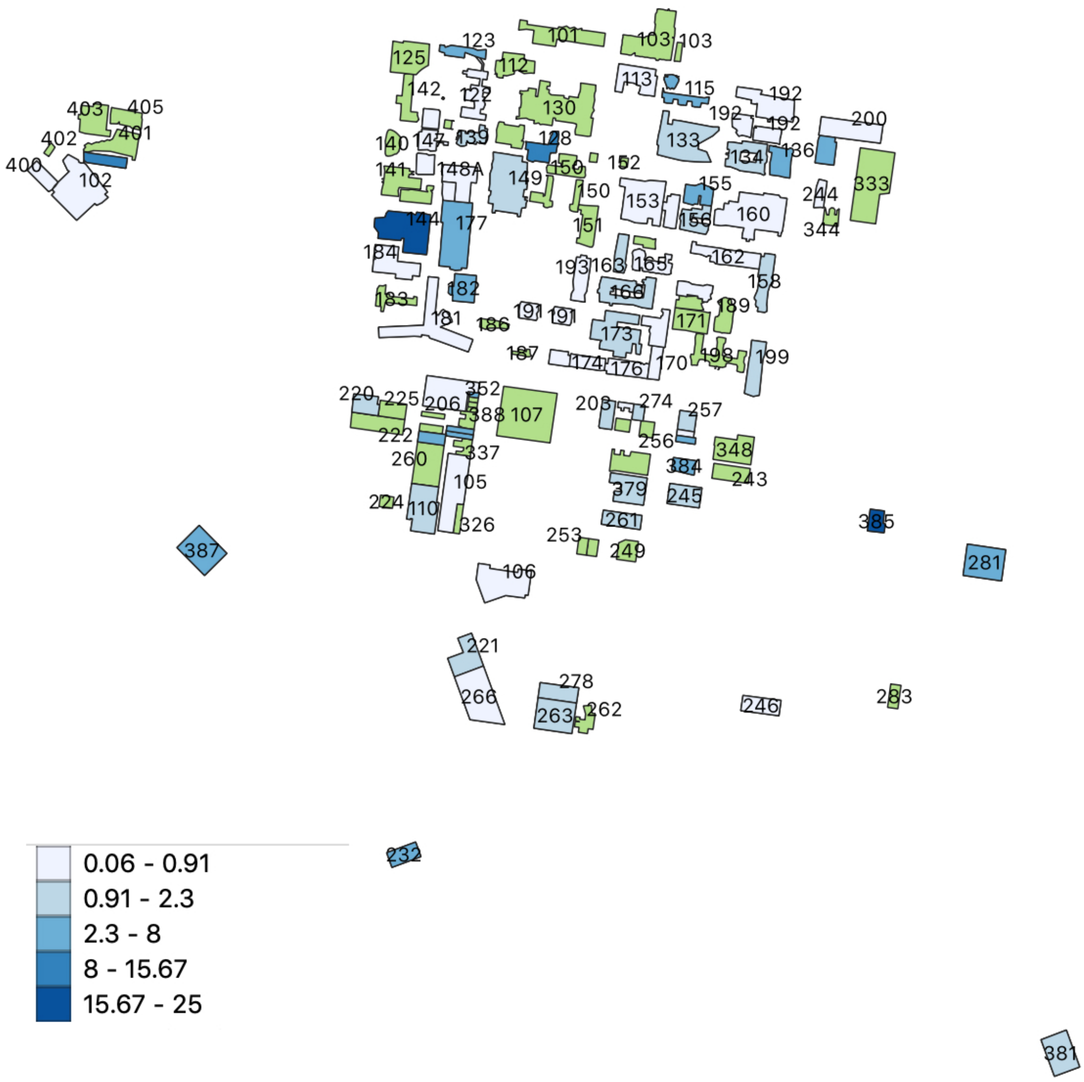

Figure 8: Spatial Analysis: Capacity of meeting rooms with respect to Staff

4 Proposed Methods

In this section, we propose 2 methods to solve the space optimization problem for effective meeting rooms allocation and toilet facilities usage. Our objective is to devise a solution that can efficiently use the current supply of resources. So, in our methods, we aim to find the `most optimal nearest building` and the `most optimal nearest floor` to use for booking a meeting room or using a toilet facility as per the demand and supply.

4.1 Method 1: Finding the most optimal nearest building

In this method, we use spatial analysis using QGIS to predict the best nearest building for booking a meeting room or using a toilet facility. We propose an algorithm which is inspired by Song et al.[6] three-stage approach and greedy heuristics as explained in Section 2.

Algorithm 1 Find the most optimal nearest building

Inputs: Current Building **B**, Radius **R**, Objective **O**, Penalty **P**

1: Find all the nearest buildings within the radius R
2: Calculate the *weights* of these buildings: $weights = \frac{\text{Supply of the objective}}{\text{Demand of the objective}}$
3: Update *weights* by including correlations of the objective.
4: Calculate $probabilities = softmax(weights)$
5: Penalize *probabilities* to get *scores* as per their physical distance from **B** and using penalty **P**.
6: Optimal nearest building: $argmax(scores)$

The initial results of this algorithm on Current Building **B** as Alan gilbert building (Building No: 104), Radius **R** as 400m and 600m, and Objective **O** as meeting rooms are shown below.

Figure 9: Preliminary results for Building 104 with 400m and 600m radius

The concept of this algorithm is to find the most effective weights for the buildings within the provided radius. These weights are initially calculated based on the supply and demand of the objectives such as meeting rooms or toilet facility. Then, these weights are penalized on how far they are from the current building and updated based on the different correlations for the objective as discussed in Section 3.3.

The initial results as shown in Figure 9 suggests that there is a high chance that user is not able to book a meeting room in building 104 as the weight is quite less (shown with little shade of red colour). ***Currently, these weights and probabilities are not getting updated as per step 3 and 5 of the algorithm.*** So, based entirely on the supply-demand weighting scheme, algorithm running with 400m radius will suggest `building 207 (MDHS)` and 600m radius will suggest `building 144 (Kenneth Myer)` as the most optimal nearest buildings.

4.2 Method 2: Finding the most optimal nearest floor for a building

In this method, we have used `python` to create a model based on the greedy heuristics which can predict the most optimal nearest floor for booking a meeting room or using a toilet facility.

Algorithm 2 Find the most optimal nearest floor

Inputs: Current Building **B**, Current Floor **F**, Objective **O**, Penalty **P**

1: Get all the floors in the provided current building
2: Calculate *props* for supply on each floor: $props = \frac{\text{Supply of the objective at each level}}{\text{Total Supply of the objective in building}}$
3: Calculate *props* for demand on each floor: $props = \frac{\text{Demand of the objective at each level}}{\text{Total Demand of the objective in building}}$
4: Calculate *weights* for each floor: $weights = \frac{\text{Supply of the objective} \times props \text{ for supply}}{\text{Demand of the objective} \times props \text{ for demand}}$
5: Update *weights* by including correlations of the objective.
6: Calculate $probabilities = softmax(weights)$
7: Penalize *probabilities* to get *scores* as per their physical distance from **F** and using penalty **P**.
8: Optimal nearest floor: $argmax(scores)$

The model predictions using above algorithm for Current Building **B** as Kwong lee dow building (Building 263), Current Floor **F** where the user is located as Level 3, Objective **O** as meeting rooms and Penalty **P** as 0.005 is shown below.

	Floor Code	Floor Name	Room Code	weights	probs	penalize	scores
0	0	Ground	50	0.228571	0.046933	−0.015	0.031933
1	1	Level 1	27	0.500000	0.061569	−0.010	0.051569
2	2	Level 2	38	2.400000	0.411641	−0.005	0.406641
3	3	Level 3	83	0.602540	0.068217	0.000	0.068217
5	5	Level 5	31	2.400000	0.411641	−0.010	0.401641

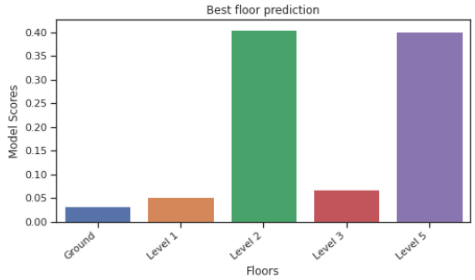

Figure 10: Model scores prediction for Building 263

Using the above predictions, we can say that `Level 2` will be the most optimal nearest floor for booking a meeting room after considering supply-demand weighting scheme. These results are still not considering **Step 5** of the algorithm which will further enhance the accuracy of the scores. Similarly, we can do model predictions for Current Building **B** as Redmond barry building (Building 115), Current Floor **F** where user is located as Level 1, Objective **O** as toilet facilities and Penalty **P** as 0.005. The results are shown below.

	Floor Code	Floor Name	weights	probs	penalize	scores
1	0.2	Basement 2	0.823529	0.157057	-0.015	0.142057
7	2.0	Level 2	0.000006	0.068930	-0.075	-0.006070
8	3.0	Level 3	0.000096	0.068936	-0.090	-0.021064
9	4.0	Level 4	0.001872	0.069058	-0.105	-0.035942
10	5.0	Level 5	0.000395	0.068956	-0.120	-0.051044
11	6.0	Level 6	0.000419	0.068958	-0.135	-0.066042
12	7.0	Level 7	0.000611	0.068971	-0.150	-0.081029
13	8.0	Level 8	0.000970	0.068996	-0.165	-0.096004
14	9.0	Level 9	0.588235	0.124128	-0.180	-0.055872
3	10.0	Level 10	0.000273	0.068948	-0.015	0.053948
4	11.0	Level 11	0.000416	0.068958	-0.030	0.038958
5	12.0	Level 12	0.352941	0.098103	-0.045	0.053103

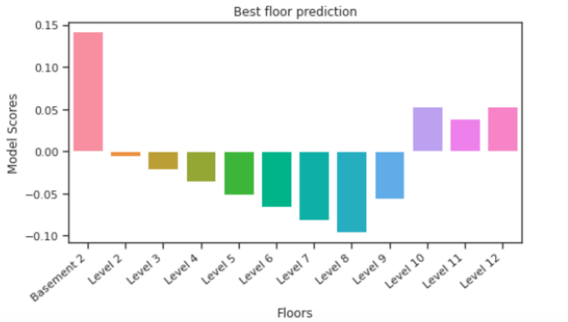

Figure 11: Model scores prediction for Building 115

Using the above predictions, we can say that `Basement 2` will be the most optimal nearest floor for using a toilet facility, considering all the supply-demand constraints. Again, these results are still not considering **Step 5** of the algorithm which will be implemented in the next phase of the project.

5 Timeline

In this section, we will provide the summary of the project part-1 timeline with task and their respective assignees as shown in Figure 12. This is followed by rough timeline for part-2 of the project. The next phase of the project is divided into 2 phases, where first phase will focus on method proposal enhancements and better prediction modelling as shown in Figure 13. The second phase will focus on detailed analysis report which will present all important findings as shown in Figure 14.

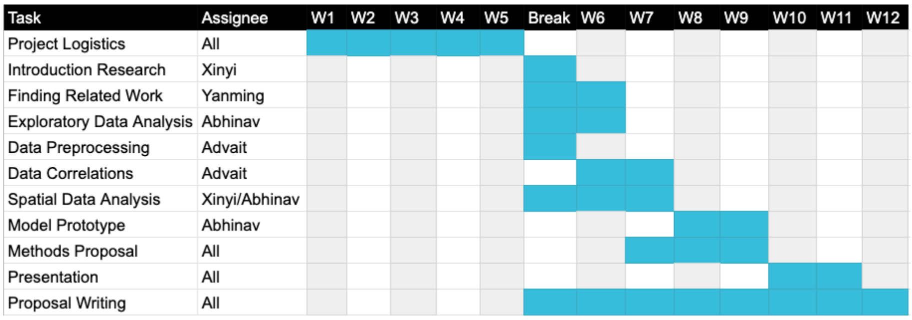

Figure 12: Project Part-1 Timeline Summary

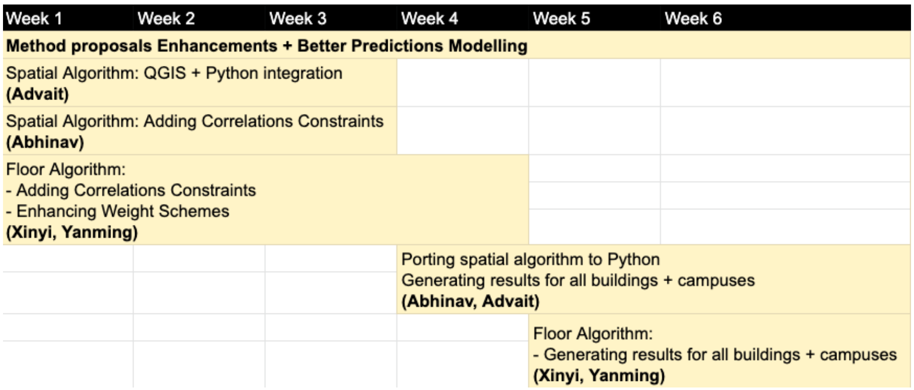

Figure 13: Project Part-2 Timeline for Week 1-6 (Phase 1)

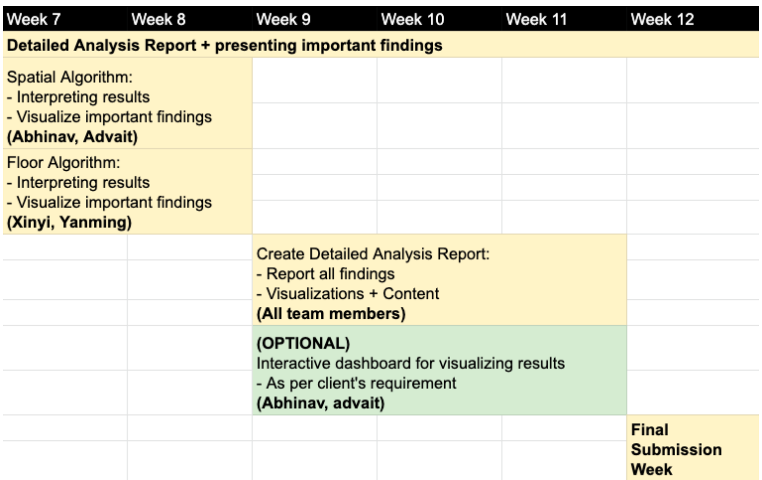

Figure 14: Project Part-2 Timeline for Week 7-12 (Phase 2)

References

[1] S. Even, A. Itai, and A. Shamir. "On the Complexity of Timetable and Multicommodity Flow Problems". In: *SIAM Journal on Computing* 5.4 (1976), pp. 691–703. DOI: `10.1137/0205048`. eprint: `https://doi.org/10.1137/0205048`. URL: `https://doi.org/10.1137/0205048`.

[2] Ahmed Wasfy and Fadi Aloul. "Solving the University Class Scheduling Problem Using Advanced ILP Techniques". In: (Jan. 2007).

[3] Edmund K Burke et al. "Decomposition, reformulation, and diving in university course timetabling". In: *Computers & Operations Research* 37.3 (2010), pp. 582–597.

[4] Hendrik Vermuyten et al. "Developing compact course timetables with optimized student flows". In: *European Journal of Operational Research* 251.2 (2016), pp. 651–661.

[5] C Beyrouthy et al. "Towards improving the utilization of university teaching space". In: *Journal of the Operational Research Society* 60.1 (2009), pp. 130–143. DOI: `10.1057/palgrave.jors.2602523`.

[6] Ting Song et al. "An iterated local search algorithm for the University Course Timetabling Problem". In: *Applied Soft Computing* 68 (2018), pp. 597–608.

[7] Arnaldo Vieira Moura and Rafael Augusto Scaraficci. "A GRASP strategy for a more constrained School Timetabling Problem". In: *International Journal of Operational Research* 7.2 (2010), pp. 152–170.

[8] Tahir Cagin Arnab Chakrabarty Sam Mannan. *Multiscale Modeling for Process Safety Applications*. 2016, pp. 339–396. DOI: `https://doi.org/10.1016/B978-0-12-396975-0.00008-5`.