

INTERIM DELIVERABLE

Analytics Edge Project

By:

Zihao Geng

Shivani Arun

Feifei Shao

Ting Gu

Ruoxu Wang

Problem Statement

This report reveals a challenge of scheduling courses and classrooms at USC Marshall School of Business, which is how to utilize the available classroom space efficiently as the demand for Marshall courses is high and is expected to rise in the future. Furthermore, Marshall plans to undergo a large-scale renovation of Bridge and Accounting buildings, which would remove several medium-sized classrooms for one academic year. Thus, classroom utilization must become more efficient during this time to make up for the supply shortage.

Based on a preliminary data analysis, we find that 48.9% of all courses have a gap between the classroom seating capacity and the number of students enrolled in the session. Particularly, 103 courses have a gap exceeding 50% of the classroom capacity.

Improving classroom space utilization will drive operational efficiency and reduce the operation cost at Marshall. It will also allow the availability of more classes within the same time frame, resulting in a more compact schedule that may benefit both students and professors.

Therefore, we explored the course scheduling problem in this report, built a metric that measures classroom utilization, and defined a suitable benchmark to influence future decision making.

Description of the Metric

Our metric measures the benchmark capacity utilization rate, and our goal is to optimize the utilization of the limited classroom space. It aims to reduce the gap between classroom seating capacity and the number of seats offered in a class session¹.

The metric is defined as:

$$\text{benchmark utilization rate} = \text{median}\left(\frac{\text{registration count}}{\text{total number of seats}}\right)$$

The median utilization rate can be improved if the utilization rate of individual classes is improved.

¹Every class can have multiple sessions and each has its own time slot. For instance, DSO 545 has three sessions which run on different days and time slots.

Appropriateness of the Metric

The metric we picked is computable from data and has plausible assumptions to back it up. The main goal of the metric is to measure the seats utilization in a classroom and can be computed by dividing the 'Reg Count' column by the 'Seats' column² and finding the median of this distribution over all class sessions to arrive at the final value.

This metric can be used to:

- a. Develop a benchmark/goal that we want to achieve. From our data, we observe that the median is at a 94% utilization rate. We consider improving the utilization rate for all class sessions that are below the benchmark without changing those that are above the median.
- b. Using this metric, we can check whether and to what extent classroom utilization has been improved after implementing new course schedules.

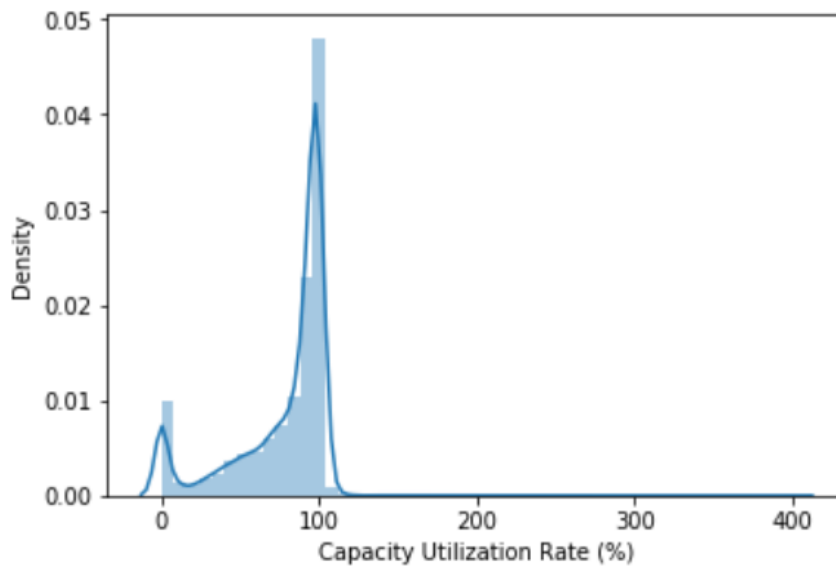
The metric is actionable. Course schedulers can monitor the enrollment count and move a class with a low utilization rate to a classroom with a smaller capacity, and in other cases, move a class with utilization around 100% to a bigger classroom. The closer the metric output is to 100%, the more effectively the classrooms are being utilized. The administrators can therefore monitor both individual and overall classroom utilizations to support their decision making.

This metric is also simple since its computation involves only the use of two variables which are easily accessible in the Marshall Enrollment dataset. Also, it doesn't involve any complex formula or computation which makes it easy to understand.

Data Analysis & Recommendations

By using the Marshall Enrollment dataset, we calculated the utilization rate using the enrollment count divided by classroom capacity for each session. The following graph shows the distribution plot of individual capacity utilization rates (i.e. classroom utilization rate/capacity utilization rate).

Graph 1: Capacity Utilization Rate Distribution



From the above graph, we notice that there are some abnormal values. First, zero utilization rate should not exist, because it implies the classroom has no student at all. For our analysis, we have excluded all records with 0 utilization rate. Second, the utilization rate should not exceed 100%. We investigated the data and found that, in absolute numbers, rates above 100% represent a maximum of 6 more students than the number of seats in a class. We believe this is not a big problem as most of these classes have multiple sessions and we expect a small number of students to be absent on a particular day. Also, among the records that show more than 100 % utilization, some of them have only three or four seats in total and represent borderline cases. Hence, any value above 100% can be ignored.

Based on the previous section, we see that since our median value is 94%, this can serve as a benchmark to rearrange classrooms with capacity utilization rates below the benchmark³.

Based on our analysis, some actionable insights/recommendations are given below:

1. An example of classroom with low utilization rate is ACCT-581. It has only 2 registered students but takes a classroom with 20 seats. To improve scheduling efficiency, we recommend shifting it to a classroom with smaller size (10-15 seats). The other option is that, for classes with one or two enrolled students, USC should consider canceling this session.

Course	First Begin Time	First End Time	First Room	Reg Count	Seats
ACCT-581	15:30:00	16:50:00	JKP202	2	20

³ Some may argue that this value is very high. Depending on the amount of leeway available in our model, we could consider working with only the bottom 25% of the data instead of the bottom 50% and in that case, our benchmark would be the 1st quartile value of this distribution i.e. 75% capacity utilization

- Another example is a session for class ACCT-470. This session has only 73 enrolled students, but it was assigned to a classroom with 269 seats. We recommend removing this class to a smaller classroom. For example, JKP has classrooms with approximately 75 seats that we believe are more suitable for this class's size.

Course	First Begin Time	First End Time	First Room	Reg Count	Seats
ACCT-470	14:00:00	15:50:00	HOH EDI	73	269

Anticipated Questions

- Why not use the average utilization rate?

We considered using the average utilization rate, but we realized that the average utilization rate is affected by extreme observations and may not be a suitable metric to optimize. Also, the distribution of the utilization rates (Graph 1, Page 4) shows that it is left skewed. This means our average value would not have been representative of the sample.

- Why not have separate benchmarks for different class sizes?

This is a good point and given additional time, we would like to examine the distribution of seats and speak to USC course scheduling team to understand how they work with large classrooms (more than 100 seats). It is because there are fewer number of classrooms with large sizes, so it is less flexible to move them around. The majority of the classrooms have a seating capacity of 100 or less. Hence, we should have separate benchmark for these classrooms as the infrastructural constraints are different.

Conclusion

The goal of our report was to build a metric to improve Marshall course scheduling efficiency. By analyzing the available data, we saw that the capacity utilization rate varies greatly among different class sessions. Our metric provides a benchmark for administrators to target classes with utilization rate lower than the benchmark, while not reducing the utilization rates of the ones that are higher than 94%. We also considered that achieving this goal is dependent on how much leeway we have to move classes around, given the constraints and interests of other stakeholders like students and faculty. Finally, we considered certain anticipated questions and answered them to make our analysis more robust.