3/5/2018

Group 8: ByungJae Kang, Cindy Wu, Kayla Xie, Murali Avvari, Ying Liu

Interim Deliverable 1

DSO570 Final Project - Spring 2018

# Opportunities for Improvement

The course scheduling system at USC Marshall is a complex process that involves many components, stakeholders, goals, and constraints. It is difficult to solve all the issues pertaining to system inefficiencies, so our group strove to identify one component that has a significant impact on the entire system.

After exploratory data analysis of the provided data, we noticed that many courses and sections were cancelled from various terms. As students ourselves, we knew the impact of a course section being cancelled after registering for it. We then thought about how cancellations of course sections affect the system and other stakeholders, and saw it as a significant contributor to the scheduling problems.

In Phase 1, a department is given slots that indicate a classroom and a time for that slot. The department then allocates these slots to their courses, so if a course section gets cancelled, then there is an unused slot that will need to be manually filled at some future point. These cancellations affect many other parts of the scheduling process, as well as the stakeholders involved. If a student registers for a course that ends up being cancelled, they may be disappointed, and their schedule may change in unfavorable ways. It affects faculty that might be teaching a course that gets cancelled, as they could lose the opportunity to teach a course at a specific time and day. It also affects the department since they would have to find another way to efficiently use that slot. These are just a few examples of how a cancelled course section affects the system. In summary, a cancelled course section has a domino effect on other components and stakeholders in USC Marshall.

A course cancellation can happen at various points throughout the scheduling process. Causes of cancellations include but are not limited to: not being able to find a suitable/available professor for the course section, not having enough registered students in the section, or unavailability of slots due to other course sections receiving higher priority in the slot allocations. Since cancellations are so influential on other parts of the process, minimizing their occurrences will also minimize uncertainties and inefficiencies across the system.

# Definition of Metric

The metric that we created is cancellation rate, which is the number of cancelled sections divided by the total number of sections originally offered. For this analysis, it was calculated by course. For example, GSBA 528 originally had two sections available, and one section was cancelled. Thus, the cancellation rate for this course is 50%.

To calculate the metric, the data from Cancelled\_Courses and Marshall\_Course\_Enrollment was grouped by term and course. Both datasets were then merged together by term and course. Columns were then created to count the number of cancellations, active sections, and total sections originally available (cancellations + active sections). Our final metric, the cancellation rate, was calculated by diving the number of cancelled sections by the number of total sections. Any section with a cancellation rate of 0% (no sections cancelled) was filtered out of the dataset since they are not the focus of our analysis. As previously mentioned, a course section cancellation can happen at any point in the scheduling process and has a negative domino effect on other parts of the system. Minimizing this metric significantly improves “goodness” for all stakeholders involved.

# Justification of Appropriateness

This metric is computable since all the required information is present from the original datasets that were provided to us. After merging the datasets and mutating certain columns, the metric was calculated. This metric can be used in future course scheduling analysis as well since the data used to compute it will continuously be collected for future terms and course sections.

There are various steps that can be taken to minimize the cancellation rate. Departments can look at historical data to see which courses and sections have been cancelled often. They can then determine what actions to take for each course, such as finding the appropriate professor for that course much earlier in the process, marketing the course better to students to ensure that enough register, or choosing not to offer the course/section at all.

The metric is also simple to understand and interpret because it requires no technical background or outside information to use it. It is versatile since it can be applied to any course section within any department or term. Simply put, the cancellation rate is the ratio of cancelled sections and total sections originally available. Using a percentage across all courses, departments, and terms allows us to standardize the metric.

Minimizing the cancellation rate allows students to confidently register for classes with the best schedule possible, faculty to have a clear understanding of their teaching schedule, and departments to avoid troublesome slot reallocation. These are just a few examples of the positive benefits. It is not a complete fix for the entire scheduling system but can significantly improve “goodness” for the various stakeholders involved.

# Analysis of Available Data Using Metric

The best possible value for cancellation rate is obviously 0%, which would indicate that a course had no sections cancelled. This may be unrealistic for every single course offered in a semester, but there is significant room for improvement to minimize the metric. Each department can come up with a reasonable goal for what the cancellation rate should be based on historical data and their own resources/constraints. Though the cancellation rate being a percentage value allows us to standardize the metric across all data points, the absolute counts matter as well. For example, if a course has ten sections available and only four were cancelled, the cancellation rate would be 40%. If there was a second course that had one section cancelled out of two available sections, its cancellation rate of 50% is higher than the first course. The marginal difference in cancellation rate is not meaningful since there were more total cancellations for the first course. The cancellations for each course should be evaluated within proper context.

# Next Steps and Future Analysis

Our next steps include aggregating the cancellation rate by department and term to see what the total cancellation rate was for each department within a given term. We will also dive deeper into what exactly causes a course section to be cancelled. Hopefully this insight gives us the ability to craft better solutions to minimize the cancellation rate.

Other metrics we are currently exploring in addition to cancellation rate include the daily utilization rate for each classroom, and weekly campus arrivals for professors. For example, some professors prefer to come to campus only twice a week and teach all their classes on those days. This metric will try to optimize the scheduling preferences of faculty.

# Appendix

The following files contain Python code, outputs for the code, and some visualizations for our metrics and other aspects of data analysis. These files were uploaded to the Blackboard submission link along with this report.

* Cancellation Rate Metric: Cancelled + Rate.pdf
* Weekly Campus Arrivals Metric: EDA\_Student\_Instructors.pdf
* Room Utilization Rate for Enrollment Dataset: Room\_Utilization\_Rate.pdf