Interim Deliverable 1

Team 8: Shiyu Ding, Stan Hsu, William Staudenmeier, Xi Yuan, Haoshen Yu

Executive Summary

Yearly course and classroom scheduling at the USC Marshall School of Business is a painstaking assignment that takes the administration many months to complete. In fact, it is an ad-hoc process, requiring lots of back-and-forth communication among the program offices, departments, professors, and administrators. Despite the massive amount of effort in coordination among departments, the final schedule does not always fulfills the needs of students and faculty alike. During our initial examination of this process, we discovered that stakeholders who have the most push and are the most vocal have the majority impact on the final schedule. As a result, we believe that there are inefficiencies in the system that we can improve to deliver a schedule that best satisfies both student and faculty needs. The purpose of this paper is to define the inefficiencies in the course/classroom scheduling and provide metrics to measure them as a first step in creating our optimization model.

A. Description of Inefficiencies Found

Inefficiency One - Registration and Seat Allocation Mismatch:

For some courses, we discovered a mismatch between the number of registered students and the number of seats available in the alloted classroom. For example, the course ACCT-470 has 20 students registered yet was assigned a room with 40 seats; meanwhile, the course ACCT-430 has 36 students registered yet was assigned a room with 36 seats.

Clearly, the above scenarios are suboptimal because they result in unnecessary waste and unbalance of classroom utilization, and decrease the possibility of scheduling more courses in one semester. This inefficiency has a negative impact on all stakeholders. In the case of ACCT-470, faculty may find it more difficult to organize the classroom layout with half of the classroom empty, while students may find it more difficult to participate in group discussions and become distracted. Conversely, in the case of ACCT-430, some students may not be able to take the course due to over-capacity.

Inefficiency Two - Course Popularity and Seat Allocation:

From our analysis of the dataset, we found that course popularity is not deterministic of seat allocation since many courses are full while having limited seating. We are assuming that not all of these courses have limited seating due to faculty assignment. That is, in some cases, the faculty member teaching the course may not be able to accommodate more students, but we assume that in other cases the overflow is due to suboptimal seat allocation. From the students' perspective, making course popularity a stronger driver of seat allocation will improve efficiency.

Inefficiency Three - Cancelled Course Inefficiency:

Sometimes over-capacity issues can result in course cancellations. After studying the cancelled course dataset, we realized that many cancelled classes were not assigned classrooms. We believe that the limitation of classroom utilization may cause those classes are cancelled. Unfortunately, many of the cancelled classes had high initial registration rates, indicating that their cancellation negatively impacts students. We will attempt to minimize this deficiency over the course of this project.

B. Description of Metrics

The three inefficiencies described above all have ineffective space-utilization in common. Though we began our analysis with data on the students and the classrooms, we did not have a quick means of identifying which classrooms had effective seat-utilization or not. For that reason, we created the following metrics to better quantify the inefficiencies we observed.

Metric One: Our first metric is called "class_util," which is the ratio of the number of registered students, "num_reg," and the capacity of the classroom, "class_cap." Using ACCT-470 again as an example, we observed that the utilization of the class was at 50%, indicating that half the space is being wasted and not generating any value for students or the university. We can also compare this to ACCT-436, which is at 100%. Class_util gives us an easy way of quantifying how well the school's classrooms are being utilized by percentage of seats filled.

Metric Two: Since more efficient use of space may yield additional room to host more classes, we do not want to limit the possible combination of courses only to those that are already in place. We also needed a metric to identify which classes are in highest demand. To accomplish this, we create a new variable called "expand_clas," which takes the current number of registrations for those popular classes and increases them by a certain rate. We calculated this rate off of the assumption that class demand is normally distributed (the number of students in the waitlist of populated classes has a mean of $0.1 \times$ current enrollment, and standard deviation of $.03 \times$ current enrollment). Next, we divided the expand_clas variable by the capacity of a newly assigned classroom to create the metric, "expand_class_util."

Metric Three: Previously, we also described an instance where student registration was high, but the class was still canceled. From the canceled class lists we can consider those with the highest number of registrations, since these represent canceled classes that were in high demand. Again we divide this number by the capacity of some newly assigned classroom to derive our third metric, "canceled_class_util."

C. Appropriateness of the Metrics

In addressing the inefficiencies of the current scheduling process, we leverage what we already know about the number of available and canceled courses, the corresponding number of registrations, and the capacities of all available classrooms. By dividing the number of registrations by the classroom capacity, we created metrics that not only show how well the current schedule utilizes space but also provide benchmarks for how well future schedules utilize space.

Computable: The following code in Python proves the computability of above metrics.

$$class\_util = num\_reg \div class\_cap$$
$$expand\_class\_util = expand\_clas \div class\_cap$$
$$canceled\_class\_util = canceled\_num\_reg \div class\_cap$$

Note that we cleaned and joined our datasets as follows:
1. Removed every course that was taken online in Marshall_Course_Enrollment_1516_1617.xlsx
2. Merged Marshall_Course_Enrollment_1516_1617.xlsx with Marshall_Room_Capacity_Chart.xlsx
3. Simulate the scenario of waitlists based on the data in Marshall_Course_Enrollment_1516_1617.xlsx
4. Merged Marshall_Room_Capacity_Chart.xlsx with Cancelled_Courses_1516_1617.xlsx

Actionable: We can affect the seats available in each classroom, and increase class utilization by better allocating more seats to more popular classes.

Simple: Keeping in mind all stakeholders, we consider students as our first priority. As the administration determines the future schedule, the above metrics will provide concise feedback on a scale that is simple and easy to understand. They will have a better grasp of utilization without having to scan back and forth between registrations and capacity. Moreover they can utilize this metric to facilitate the decision making process.

Enlightening: When we try to optimize scheduling, we should not only focus on the time matrix, but also utilization of space. If we want to use classrooms more efficiently and even accommodate more courses in a constant number of classrooms, improving space utilization is a very effective method to measure and improve goodness.

Though the administration can continue to pair courses and classrooms manually, the addition of the above metrics provides the means to further simplify scheduling through optimization. Once we set a utilization target, a computer can iterate through many possible combinations of courses and classrooms, yielding the optimal outcome.

D. Analysis of the Data Using the Metrics

We used Python to analyze the data, paying particular attention to the three inefficiencies described in Part A through the use of our defined metrics (see Appendix). In this manner, we calculated *class_util* to be 68.22%, meaning that the current classroom utilization rate is 68.22%. Next, we computed *expand_class_util* to be 1.0887, meaning that 8.9% of students in the dataset could not register for the courses they wanted due to lack of seating. Finally, we computed *canceled_class_util* at 82.39%, which indicates that these classes would have had high classroom utilization had they not been cancelled in the first place. This last metric also indicates that classroom utilization rate would most likely increase simply as a result of reducing cancelled classes through linear optimization.

Based off these results, our goal is to increase *class_util* to 80%, decrease class cancellations, and reduce *expand_class_util* to be as close to one as possible. When we consider the needs of all the stakeholders at USC Marshall, we observed that students want more classes or more seats in high demand classes to meet their program requirements, programs/departments want higher enrollment to help cover overhead, and faculty want convenience. By computing the metrics in Python, we were able to better measure these inefficiencies, in order to create opportunities that more adequately address all stakeholder needs.

Appendix. Computation of the Metrics Using Python

To compute the metrics, we used the pandas, numpy, and scipy.stats packages.

Step 1: Fill out the missing value in the course column in the cancelled class table:

```
import pandas as pd
df = pd.read_csv("Cancelled_Courses_1516_1617.csv")
df.fillna(value = 0)
for i in range(0,3232):
    if df["Course"][i] == 0:
        Prefix = df["Course Prefix"][i].strip()
        Suffix = df["Course Suffix"][i].strip()
        df["Course"][i] = '-'.join([Prefix,Suffix])
    else: continue
```

Step Two: Calculate the utilization rate of classrooms for the classes that are not cancelled based on the registered students in each class and the classroom capacity.

```
from scipy.stats import norm
df1 = pd.read_csv("Enrollment.csv")
df2 = pd.read_csv("Room.csv")
for i in range(0,45):
    df2['Room'][i] = df2['Room'][i].strip()
for i in range(0,2899):
    df1['First Room'][i] = str(df1['First Room'][i]).strip()
utilization = pd.merge(df1, df2, how = ‘inner’, left_on= 'First Room', right_on='Room', copy = False)
sum(utilization['Reg Count'])/sum(utilization['Size'])
```

The above scripts yields an output of 0.6822, or 68.22 percent utilization.

Step 3: We assume that if the number of registered student is equal to or higher than 95% of the seats provided in the course, there will be a waitlist. The number of student in the waitlist is about 10% of the registered students. We want to calculate the percentage of students, both registered and on the waitlist, who actually obtained seats in the class.

```
sum_count = 0
```

```
sum_seats = 0
for i in range(0,2899):
    if df["Reg Count"][i] >= 0.95* df["Seats"][i]:
        sum_count +=  max(df["Reg Count"][i],norm(loc = 0.1*df["Reg Count"][i],
                                          scale = 0.03*df["Reg Count"][i]).rvs()+df["Reg Count"][i])
        sum_seats +=  df["Seats"][i]
    else: continue
sum_count/sum_seats
```

The above script yields an output of 1.0887. The output indicates that around 8.9% of students, out of all the students that wanted to register for those courses, are not registered in the classes they want because of the limited seats.

Step 4: Last, we calculate the utilization rate of classrooms for the cancelled classes based off the registered students in each class and the classroom capacity

```
import numpy as np
for i in range(0,3222):
    df['First Room'][i] = str(df['First Room'][i]).strip()
utilization = pd.merge(df, df2, how = 'inner', left_on= 'First Room', right_on='Room', copy = False)
np.nansum(utilization['Reg Count'])/np.nansum(utilization['Size'])
```

The output is 0.8239, which indicates that classroom utilization for the cancelled classes is 82.39%.