**Optimizing Course Scheduling at Marshall**

**DSO 570 Interim Deliverable 1 (Spring 2018)**

**Tarun Gorowara**

**Likhith Gunjal**

**Mahalakshmi Raghavan**

**Libby He You**

**Xingjian Zheng**

**Nikki Zhu**

**Objective**

Develop a metric to optimize and improve the current USC Marshall scheduling system. This metric will efficiently utilize the available space to output a feasible schedule for all courses.

**Possible Metrics**

**Satisfaction based on professor time preference:**

Each professor may have his/her own preferred timing. There is a satisfaction factor or the “satisfactor” that comes with the timing assigned to them. There are two scales that are required to compute this metric. One is the Satisfaction factor and the other is their time preference ranking. The professors are asked to enlist their three most preferred timings in descending order and each timing is ranked on a scale of 1 to 3. The “satisfactor” is on a scale of 1 to 4. If a professor is assigned his/her most preferred class, then the “satisfactor” is 4. Similarly, if assigned a timing not in the preferred timings, the “satisfactor” is 1. The goal is to achieve a high expected “satisfactor” and to include more professors closer to the higher “satisfactor” side.

**Waitlisted students:**

The number of students on the waitlist helps in determining the popularity of a class.

**Biggest Opportunities for Improvement**

**Minimize class waitlist and minimize waste in classroom capacity**

The current process of assigning courses and classrooms at USC Marshall is primarily through historical allocation from each department. This doesn’t account for new courses being added in that have never been taught before. And potentially, with temporary closures to classrooms in Bridge and Accounting buildings; efficiency will be very important to scheduling courses in the near future. In addition, these courses and classrooms have different sizes and capacities, respectively. During the scheduling process, these courses are assigned manually and while they can satisfy a certain time/day; the space could be better utilized by using an optimization model. The biggest opportunity for improvement is to minimize these conflicts by developing a model to satisfy a number of factors. This would include efficiently utilizing available space to schedule courses based on popularity and minimizing the class waitlist.

**Definition of Metric**

USC Marshall provides hundreds of courses for both undergraduate and graduate students every semester. There can be conflicts when assigning classes to certain times and days because of the sheer amount of courses. Using the datasets provided, we can assign classes efficiently through several factors. These factors include the number of students enrolled in a particular class versus the maximum number of seats. We can use a scoring model by taking the percentage enrollment over capacity. For those courses that end up having the same score, we can use a tiebreaker by using an absolute number to find out which class is larger. Once we find which classes are the most popular, we can assign the class to a room that best utilizes an assigned room. This metric would help minimize the number of undesirable conflict in sections for a program.

**Justification of Metric**

This metric is appropriate because it is computable, actionable, simple and enlightening (CASE).

**A. Computable:** We can compute this data by using excel files: Marshall\_Course\_Enrollment\_1516\_1617.xlsx and Marshall\_Room\_Capacity\_Chart.xlsx. From the Course Enrollment data, we can determine the max number of students allowed and number of students enrolled to find out the the percentage of how filled the class is. We can determine which classes are full and then find out of those classes, which class is bigger. Using the Capacity Chart data, we can use the rooms within USC Marshall and determine the max capacity for each in order to assign classes.

**B. Actionable:** The metric can be affected by the course scheduling decision as it will prioritize the most popular classes to be registered first. Instead of courses being assigned based on historical data, we can use current data to better optimize how courses are put in which classrooms. The variables involved would be the number of enrolled students with the category of the size of a classroom.

**C. Simple:** This metric is simple because we have defined a rank of which courses are popular based on our optimization model. We define a popular course by the percentage of students in the course over the total seats available. The courses are then ranked by their popularity; higher the percentage, the higher the class is ranked. Based on their rank, the top popular classes will be given priority in assigning classes first. These classes will then be assigned to a classroom that best fits the closest capacity.

**D. Enlightening:** There is a close relationship between the output of the metric and how we can solve the classroom utilization. By looking at the number of popular classes and giving them the higher preference, we are in turn satisfying the maximum amount of students possible.

**Impact of the Metric**

Assuming that on an average, each class is of 3 units at a tuition cost of $1900, USC Marshall is losing $5700 per waitlisted student. Assuming about 20 waitlisted students per class (which is a commonly observed number for many popular classes), Marshall is losing $114000 per class.with just 10 such classes, the school is potentially losing **$1 Million**. With better class utilization, this can be significantly improved.

**Analysis of Metric**

1. **Data Cleaning:**

* Removed records where the Reg Count > Seats. This was done to get a better understanding of the data set. If a greater registration count indicates a waitlist, we would like to deal with it as a separate column in our future steps.
* Values representing the classes that are not held on the USC Campus (Los Angeles) were removed. They are- OFFICE, DEN@VITERBI, ONLINE, SAN DIEGO, SHANGHAI, OCC, OFF CAMPUS.

1. **Data Manipulation:**

**Categorical Group Creation**

* Selected 18 unique values from classroom capacity dataset.
* Regarded 18 unique values as categorical variables. Assigned every record to the smallest capacity, which is higher than the number enrolled. We can then ensure the difference between enrollment and room capacity is minimized so all students have a seat.
* Joined the classroom capacity dataset, the joined dataset will show all classrooms available for every class with minimized empty seats.

1. **Current Inefficiency**

* Students on the waitlist
  + We don’t have the data now, but there are 596 course sections that have 100% enrollment/seats capacity. So there is a large space of improvement.
* Waste of Classroom Capacity
  + For each course section, we are calculating the waste by finding the difference of the classroom capacity and actual enrollment. We then sum the difference up to get the current waste of classroom capacity of 36,212 seats. The python code of calculation will be shown in the appendix.

1. **Benchmark**

* Students on the waitlist
  + This is to be decided once we have the dataset.
* Waste of Classroom Capacity
  + Best possible value

We matched up the classroom capacity category with the actual enrollment data. This would be the best possible value because it doesn’t include constraints like the limit slots of each classroom. The number is 9551, which is a save of 26,661 vacant seats. Code will be included in the appendix.

* + Worst possible value

When we don’t have any improvements, the current inefficiency is 36,212 vacant seats.

* + Reasonable aim

Based on the constraints, we are aiming at 70% of the best possible value, which is a total of 18,663 seats saved from going waste.

**Potential Questions**

1. Capacities of several classrooms are not offered. Will the assumption make sense that fixed courses are assigned to buildings outside Marshall, so that we only focus on those currently assigned to Marshall classrooms?

Further analysis: find out what kind of classes are assigned to classrooms outside of Marshall and their capacities.

2. To optimize the model further, will we be given the Waitlist data or should we assign random values to it?

Further analysis: obtain waitlist dataset or identify what values should be assigned to each course.

3. If we enroll more students from the waitlist, what are the limits of seats in the course? Since for some courses, professors and students may not expect too many students in the course.

Further analysis: find related preference data to determine limits for extra seats offered for really popular courses.

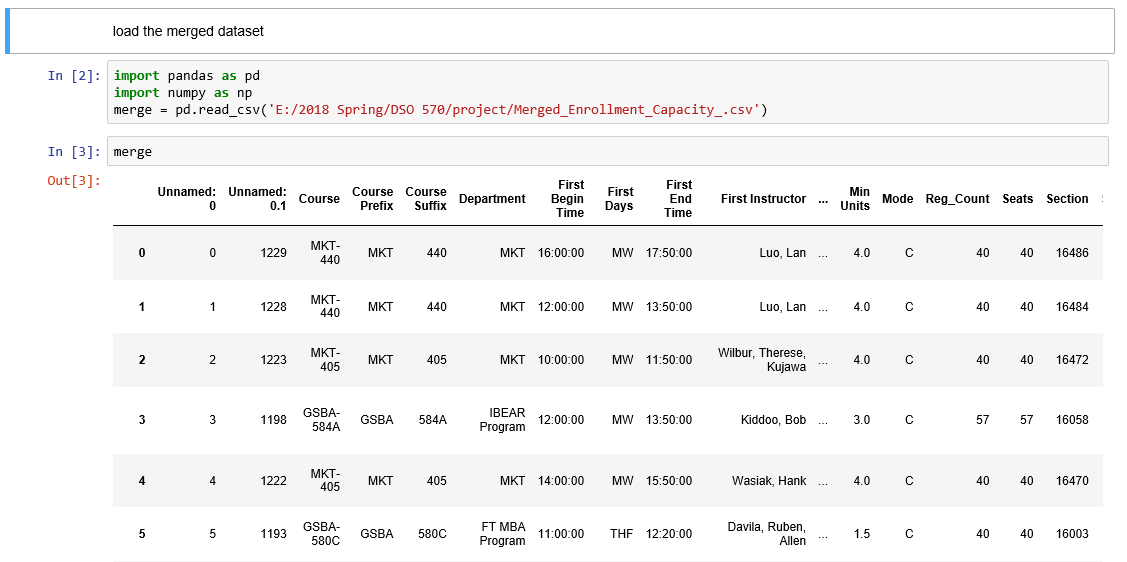
**Next Steps & Conclusion**

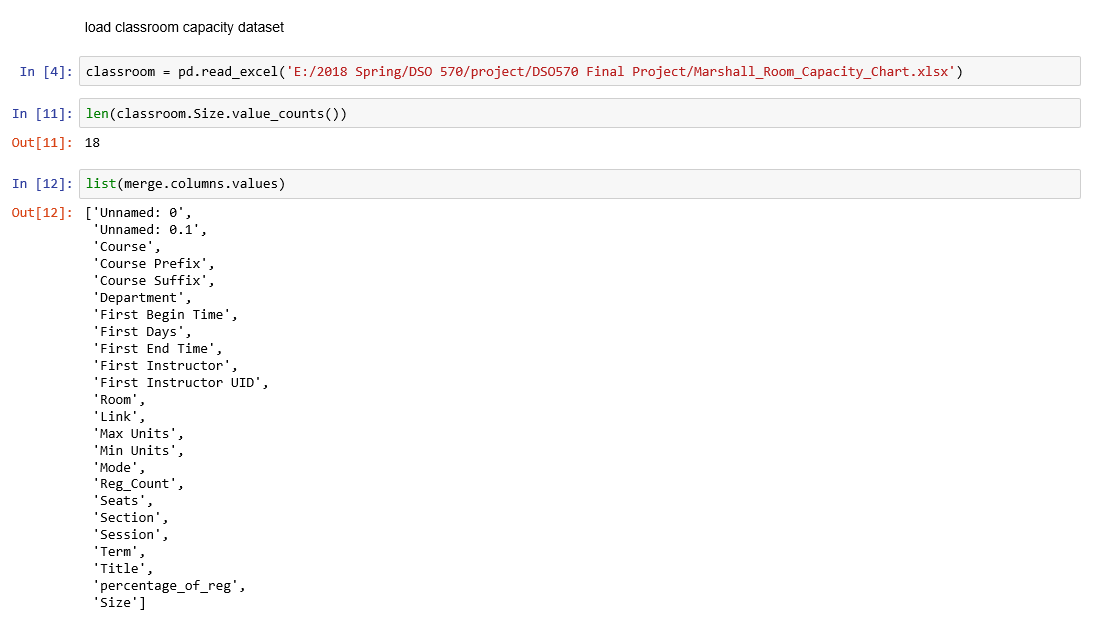
Our next step will involve deciding the number of constraints to be included in the model. We will explore several constraints such as popularity of classes, timing slots available per day, professor preferences on timings, undergraduate and graduate class timings. Due to time constraints, the scope of the project might only allow us to work on maximum of two constraints that have been stated above.

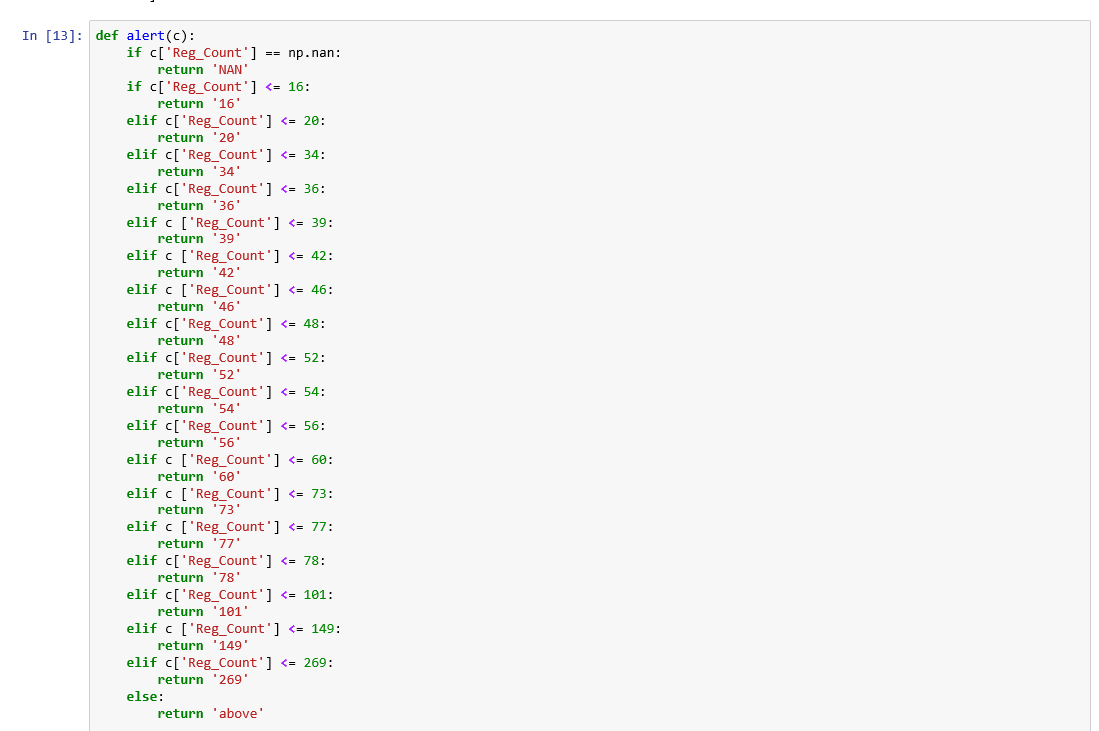
By identifying the metric of class size utilization, a model can be created and with the right interface, we can help make an easy and effective way of assigning classes without any wastage in space. By minimizing this wastage, popular courses can accommodate more students which will lead to higher levels of satisfaction among students.

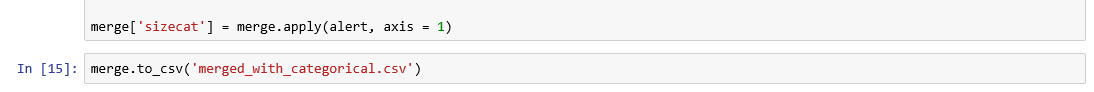
**Appendix**

Appendix 1

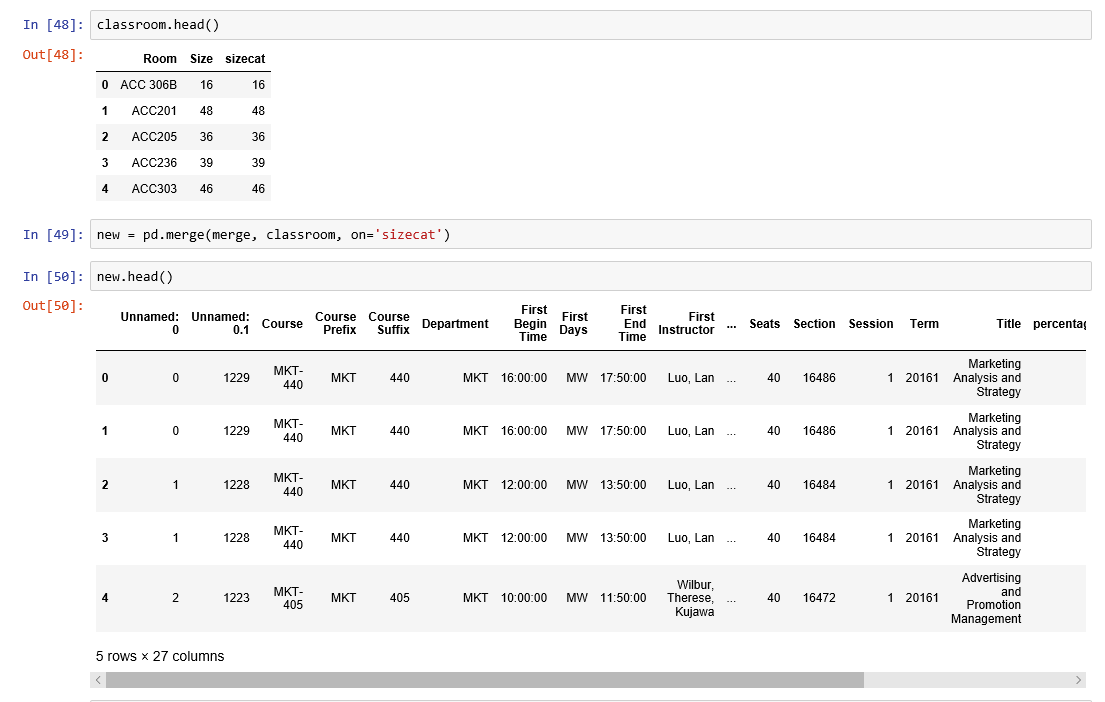




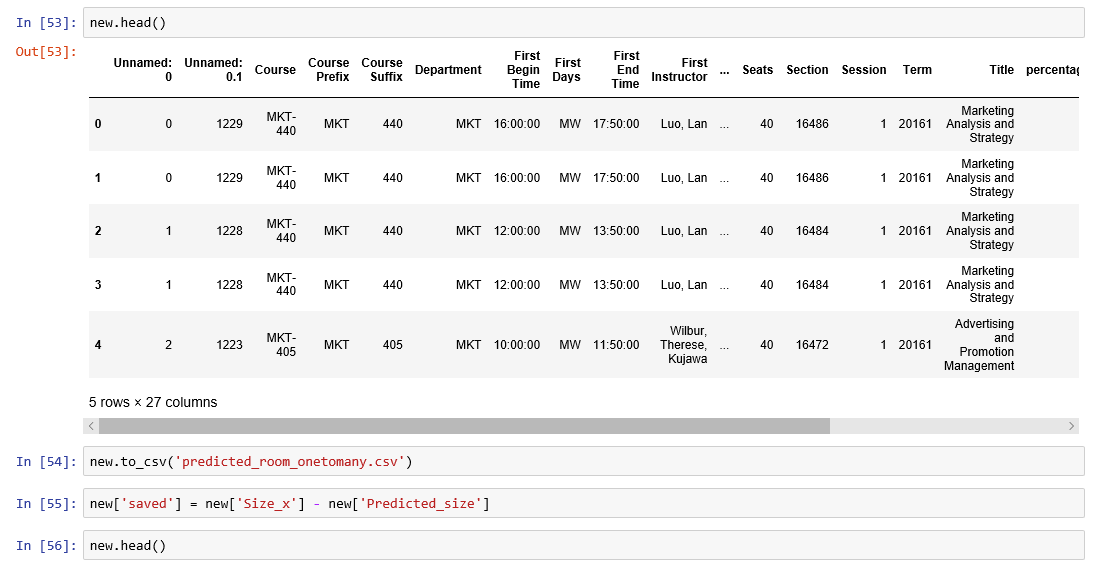


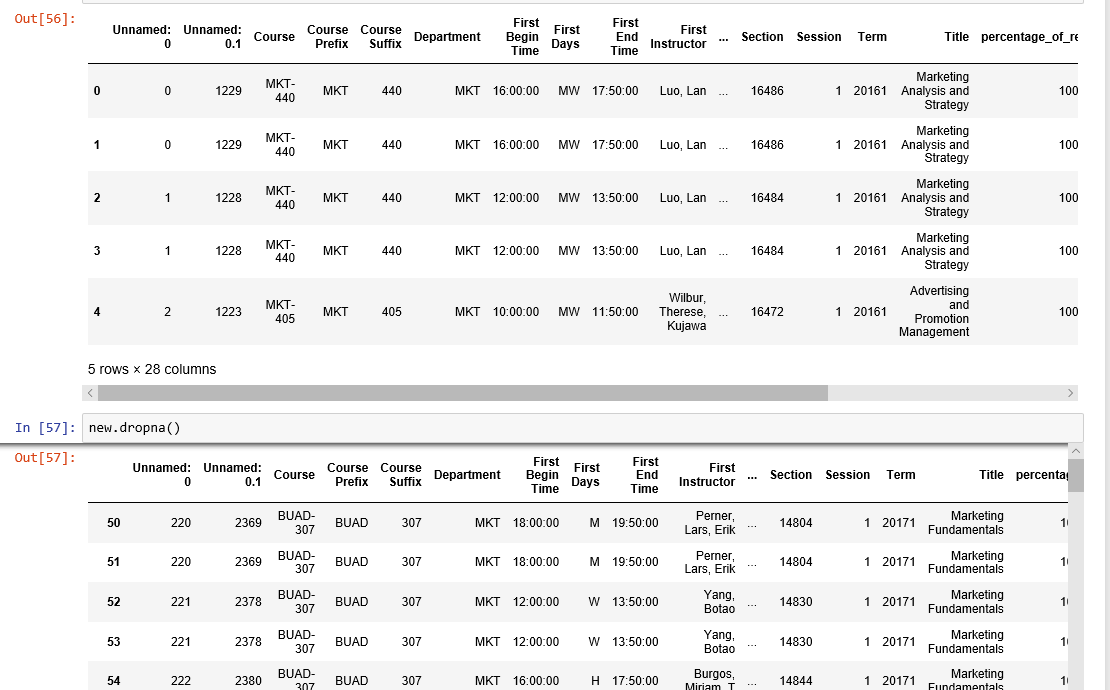


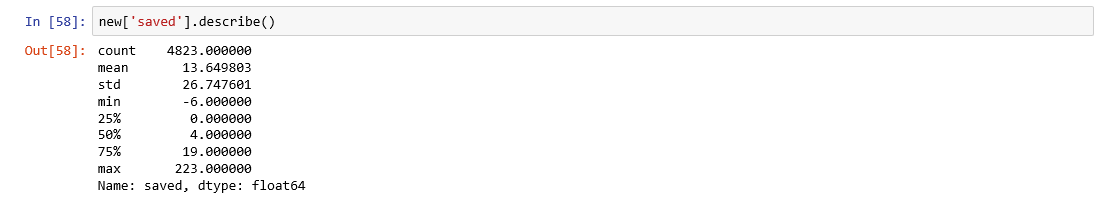












Appendix 2 - Calculation for Metric

