

Interim Deliverable 1

Group 4



Zimeng Cao

Ying Chen

Ruoxian Jia

Anqi Pan

Cristian Quintero

**Part A: Opportunity for Improvement**

Based on the primary analysis of the course schedule at Marshall, we have identified one opportunity for improvement on initial department allocations by reducing the course cancellation rate. Standing from the Finance and Administration office’s perspective, our goal is to decrease human resource and other cost by reducing the workload of Shannon's team and increasing the utilization rate of the classroom.

Every year, each department is given a certain amount of time slots to make their initial allocations. The distribution of these time slots is mostly done based on historical data. During this process, departments sometimes request more slots than needed, and return the surplus after their initial allocations. As consequence, Shannon's team have to deal with the re-distribution of these remaining slots. Due to the complexity of the task, phase one is usually prolonged. Historical data shows that during the past two years, at least hundreds of scheduled courses were eventually canceled, which was an enormous waste of administrative effort. Many classrooms might not be well utilized after course cancellation, which resulted in a waste of Marshall resources. At the beginning stage of this project, we calculate and compare the cancellation rate, investigate the impact, find the hidden reasons, and present recommended changes on the initial department allocation.

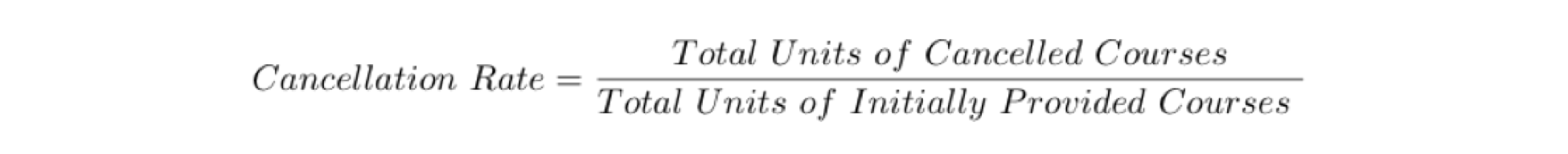
Upon our analysis, we determined that the Lloyd Greif Center for Entrepreneurial Studies Department (BAEP) has the highest cancellation rate among Marshall Departments. This department is followed by the Marketing (MKT) and Accounting (ACCT) departments. Among the programs that control their own courses, the Graduate Program has the highest cancellation rate, followed by the Master of Business for Veterans program (MBV) and the Undergraduate Program (UG). We suggest that for future semesters, the administration office may lower the initial number of time slots assigned to these departments.

**Part B: Definition of Metric**

Our group came up with a metric called Cancellation Rate (CR). This metric measures the percentage of courses that have been cancelled among all initially provided courses. Since the register office spends a significant amount of time scheduling and distributing classrooms for these courses, their cancellation results in a waste of human efforts and classroom spaces.Those cancelled courses create a burden on the register office and reduce the proper utilization of Marshall resources. Therefore, the register office would benefit if all departments had a low CR.

In order to reduce the average cancellation ratio, the register office can prevent departments from initially requesting courses that are likely to be cancelled. Our metric is a good measurement since it can be easily Computable for each department, it is Actionable, Simple to understand, and Enlightening. For departments that have high cancellation rate compared to the average, the register office should limit the number of courses that can be requested based on historical data. The departments will then have the incentive to use their quota efficiently and avoid requesting courses that are likely to be cancelled later on. In this way, the average Cancellation Rate for all department will be lower.

We calculated the Cancellation Rate for each department though the following formula:



To be more specific,



To compute the metric, we need to use the **Cancelled\_Courses\_1516\_1617** dataset and **Marshall\_Course\_Enrollment\_1516\_1617** dataset. We noticed that some courses appear in both datasets. Thus, we assumed that only the courses that are included in the **Cancelled\_Courses\_1516\_1617** but not in the **Marshall\_Course\_Enrollment\_1516\_1617** are the real cancelled courses.

Steps to Compute the Cancellation Rate:

1. Group the course by department in the **Cancelled\_Courses\_1516\_1617** dataset and **Marshall\_Course\_Enrollment\_1516\_1617** dataset.
2. Sum the minimum units of courses that have been cancelled for each department using the **Cancelled\_Courses\_1516\_1617** dataset.
3. Calculate the minimum units of courses initially proposed by adding the actual enrolled courses and the courses that have been cancelled for each department.
4. Compute the Cancellation Rate metric for each department by dividing the results from step 2 by the result from step 3.
5. Calculate the overall average cancellation rate from step 4.

**Part C: Metric Justification**

**Computable**

The defined metric is computable because all data needed is available. By grouping the data by department, we can simply calculate the following parts:

Total Units of Cancelled Courses for each department is the sum of the field "Min Units" from the **Cancelled\_Courses\_ 1516\_1617** dataset.

Total Units of Enrolled Courses for each department is the sum of courses’ "Min Units" from the **Marshall\_Course\_Enrollment\_1516\_1617** dataset.

A fundamental assumption lying in our logic is that high cancellation rate negatively affects the effectiveness and efficiency of course scheduling.

**Actionable**

The cancellation rate is directly associated with the course scheduling decision.

Specifically, department with a relatively high cancellation rate requests classrooms superfluously, increasing the complexity of scheduling process. After assigning classrooms or time slots to that department, the administration office will face the significant uncertainty that many courses might be canceled and that resources might need to be reallocated.

However, if the administration office has the knowledge of cancellation rate for all departments, it can prioritize departments with low cancellation rate, and thus can significantly increase the efficiency of the first-round resource allocation to reduce the trouble following.

**Simple**

The cancellation rate metric is quite intuitive and straightforward to understand and interpret. A high cancellation rate indicates a high uncertainty lying in the scheduling process and vice versa.

In the algorithm, we use the total units of courses rather than the total number of courses to take the overall demand or importance of the course into account. Meanwhile, using course units rather than simple count is standing from the administration office's perspective since USC holds a fixed tuition rate per credit and administration office cares about financial issues.

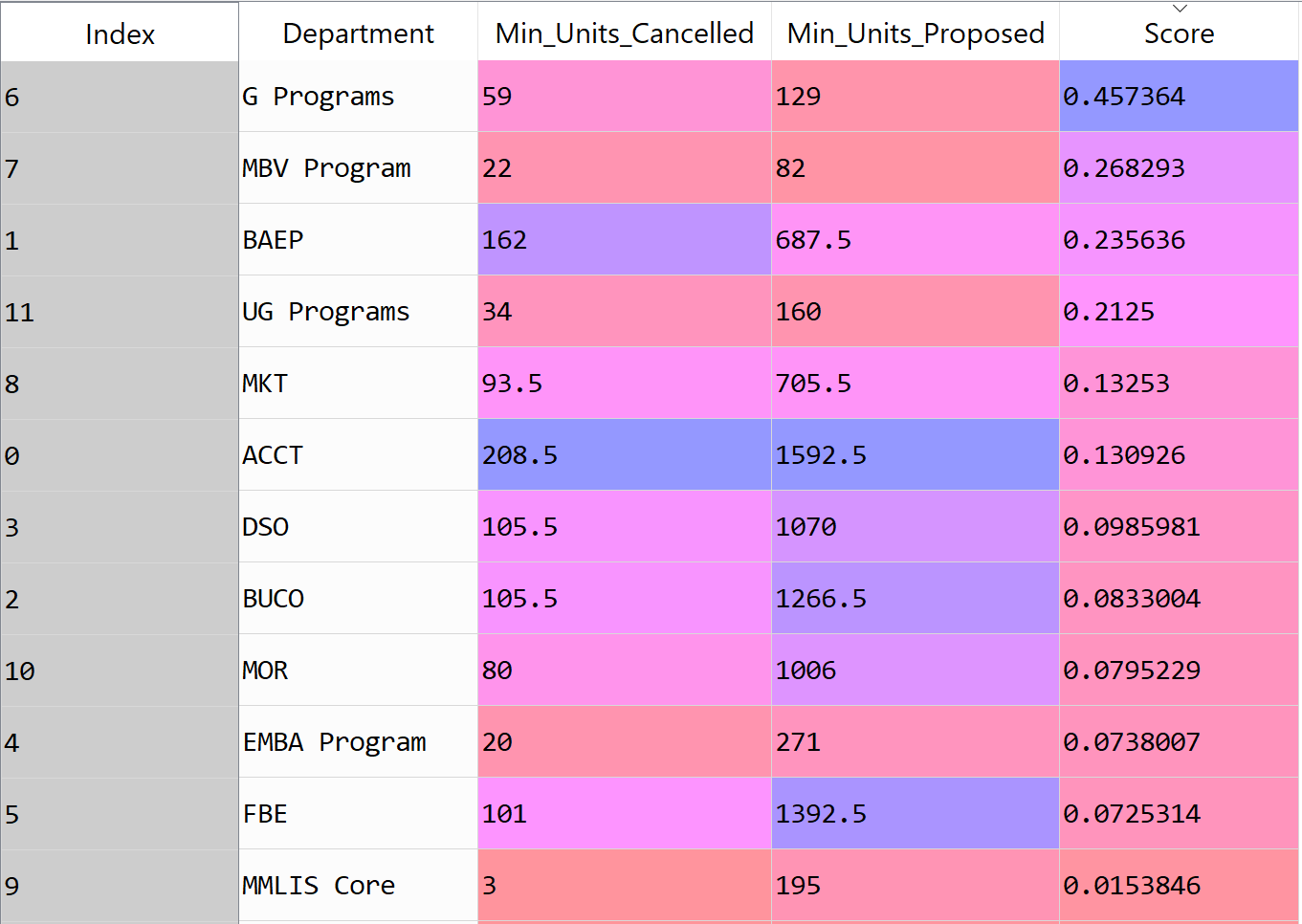
**Enlightening for "goodness"**

The output metric is a set of numbers indicating the cancellation rate for each department. By aggregating the numbers, one can conclude an average cancellation rate for all Marshall departments.

The average cancellation rate serves as a benchmark. Simply by comparing the cancellation rate for each department with the average cancellation rate, one can get an intuitive judgment on whether that department's scheduling strategy is efficient or not efficient. If the cancellation rate is lower than average we will categorize the strategy as "good" and "not good" otherwise.

Our goal is to lower down the overall cancellation rate after applying this optimization. Therefore, one can scale the data, identify the department with the highest distance, and thus to give it higher “limitation” in the next scheduling process. By doing so, the administration office is able to significantly improve the “goodness” this metric enlightens.

**Part D: Analysis On data**



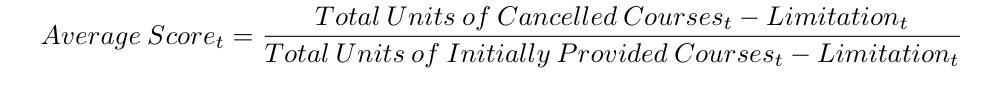


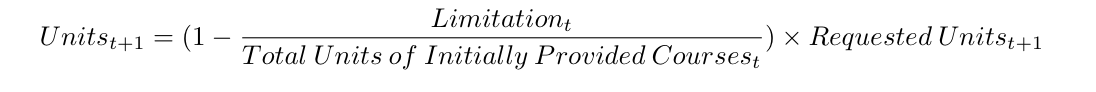
According to the python analysis, the average score for all the departments is 0.930. There are six departments have the higher score than average which could be identified as “not good.” For these departments, the administration office should give them a limited number of courses that can be requested.

More specifically, our goal is to decrease the cancellation rate of these 'not good' departments to our benchmark (average score).

Since the cancellation rate is less than one, we could reduce it by deducting a constant (Limitation) both in numerator and denominator. By the using Limitation divided by Total requested units, we get a percentage. In our opinion, in the next year, for these 'not good' departments, the AD office should decrease their requested units by 'percentage' rather than give these departments all units they request.

The mathematical formula shows below:





**Appendix**

# loading Data

import pandas as pd

import numpy as np

cancelledCourses = pd.read\_excel('Cancelled\_Courses\_1516\_1617.xlsx')

courseEnrollment = pd.read\_excel('Marshall\_Course\_Enrollment\_1516\_1617.xlsx')

# Group by Department and aggregate the field 'Min Units' --> this will give a series

cancelledUnitsPerDep = cancelledCourses.groupby('Department')['Min Units'].sum()

enrolledUnitsPerDep = courseEnrollment.groupby('Department')['Min Units'].sum()

# Transform back to a panda Dataframe

cancelledUnitsPerDepDF = cancelledUnitsPerDep.to\_frame(name=None)

enrolledUnitsPerDepDF = enrolledUnitsPerDep.to\_frame(name=None)

# create new indexes for dataframes

newCancelledIndex = np.arange(0,len(cancelledUnitsPerDepDF))

newEnrolledIndex = np.arange(0,len(enrolledUnitsPerDepDF))

# change current 'Department' index into a column and update new index

cancelledUnitsPerDepDF['Department'] = cancelledUnitsPerDepDF.index

cancelledUnitsPerDepDF = cancelledUnitsPerDepDF.set\_index(newCancelledIndex)

enrolledUnitsPerDepDF['Department'] = enrolledUnitsPerDepDF.index

enrolledUnitsPerDepDF = enrolledUnitsPerDepDF.set\_index(newEnrolledIndex)

# merge (right join) the two dataframes

s1 = pd.merge(cancelledUnitsPerDepDF, enrolledUnitsPerDepDF, how='right',

on=['Department'])

# change nan values to zeros

s1 = s1.fillna(0)

# calculate the number of units proposed by each department

s1['Min Units Proposed'] = s1['Min Units\_x'] + s1['Min Units\_y']

# change names of colums

s1.columns = ['Min\_Units\_Cancelled', 'Department', 'Min\_Units\_Enrolled', 'Min\_Units\_Proposed']

# save names of colums into a list in order to change order

cols = s1.columns.tolist()

cols = [cols[1],cols[0],cols[2],cols[3]]

s1 = s1[cols]

# create the score metric

s1['Score'] = s1['Min\_Units\_Cancelled']/s1['Min\_Units\_Proposed']

# sort by descending score

depCancelRate = s1.sort\_values('Score', ascending=[False])

# final dataframe

depScore = depCancelRate[['Department','Min\_Units\_Cancelled','Min\_Units\_Proposed','Score']]

# calculate average Cancellation Rate

print('The average Cancellation Rate for all departments is', depScore['Score'].mean()\*100,'%')