DSO 570 : Final Project
Optimizing Course Scheduling at Marshall

Authored by :
Tianqi He
Naiyuan Xiao
Pratyush Shankar
Wei Tang
Yu Luo

Section 1: Summary

In this, phase-1 of the final project we were able to successfully collate all datasets and establish meaningful insights and valuable data visualization in Python Jupyter notebook. We followed a meticulous scientific approach to understand the problem statement, given the various technical and structural constraints.

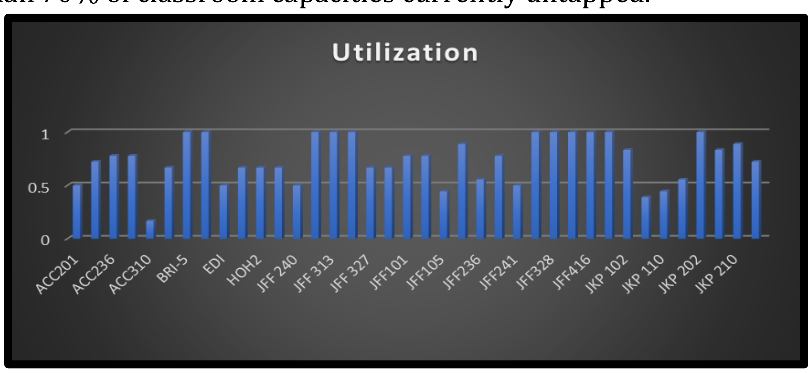
Subsequently, we defined metrics for establishing the "goodness" of our optimization models as well as decision variables that we will need to optimize in order to create an optimal model for the course allocation process in Marshall Business School in the near future.

Section 2.1 : Improvement Opportunities

After analyzing all the data provided to us, we were able to identify the following revenues that could be better optimized so as to improve efficiency of the scheduling process. Going forward, we will be working on metrics that can help translate these opportunities into quantifiable targets to make the scheduling process more efficient in the coming semesters.

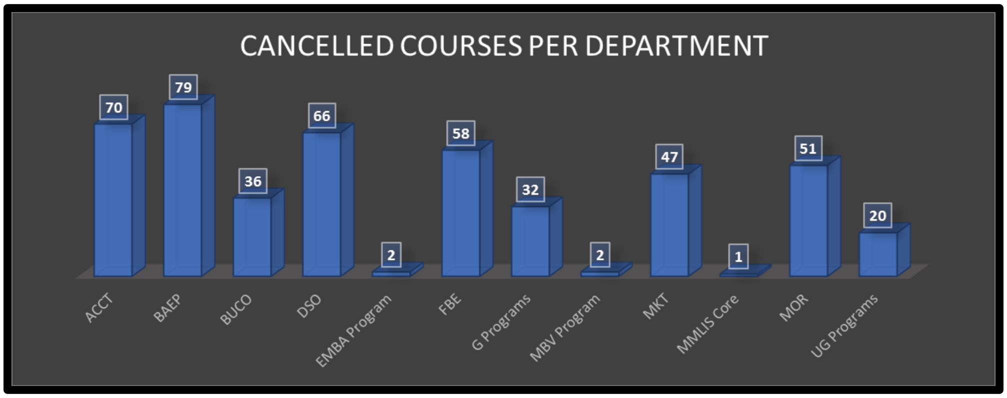
2.1.1 Classroom Utilization

Given the constraint that the Accounting Building and Bridge Hall will be undergoing renovation soon, we tried to identify the classroom utilization based on distance near these buildings, so as to identify possible alternatives for conducting classes in the future semesters. One of the first observations we could derive from the given data was that most of the classrooms were being under-utilized. Either their entire space was not being utilized or there were certain days that had a large majority of classrooms in a given building being left unused. We calculated that currently Marshall Business School had buildings with a utilization of only 28% overall. Which means that with optimization modeling and scheduling algorithm, there was a scope of utilizing more than 70% of classroom capacities currently untapped.



2.1.2 Inefficiency in Course Allocation

From the given data we could analyze that every semester a large number of courses end up being cancelled. This may be due to a number of reasons such as insufficient number of students opting for the course, unavailability of faculty or infrastructure required for the course or simply inefficiencies in demand forecasting. As stated in the introduction of the project, the final allocation of courses usually takes place in phase-2, by which time a lot of valuable resources (including time, money and man power) might have already been invested into planning, organizing the courses. In the event that course does indeed end up getting cancelled, further resources need to be invested in scrapping the course schedule and providing alternative solutions for the same. With better demand sensing and shaping algorithms better course forecasting can be made in the future.



2.1.3 Need for Automation

As discussed in previous sections, there are a number of ways in which course optimization may be done. While universities like MIT employ a central allocation algorithm, other universities still have a manual or first-come-first-serve allocation system.

Taking an example from the current allocation system at USC Marshall, we noticed that often times a faculty's seniority may play a large factor in determining the course allocation algorithm. To create a more efficient optimization algorithm we approached the problem a little differently. Instead of going with the previous status quo, we decided to treat student and faculty preference like metrics with weighted averages. With their first preference being of value 3, their second preference being of value 2 and their third preference being of value 1. Furthermore, we took a faculty's seniority into consideration and allocated a similar weight to the same. Our aim in doing this is to create and maximize a metric called Satisfaction Index which would be defined as a sum-product of all these weights and give the optimal solution to meeting the preference of most, if not all, stakeholders.

Section 3.1 : Data Cleaning and Transformation

We have adopted the various datasets into Python. And we used Jupyter Notebook to clean the data, create meaningful summary tables as well as insightful visualizations to help us better understand the data.

➢ We cleaned and transformed the data in python and attached the codes in the appendix.

Section 4.1 : Variable Creation and Metric Definition

We identified the five major stakeholders involved in the course allocation process. Our final aim in this project will be to create and maximize a metric called Satisfaction Index which would be defined as a sum-product of all these weights and give the optimal solution to meeting the preference of most, if not all, stakeholders.

We conclude some important matric in terms of each stakeholder: Students, Faculty, Department, Management, Program.

Ø For student:

1. **Enrollment rate for core courses**: each student should be able to take all core courses. Student preference will influence the course scheduling decision, we can calculate the percentage of the enrollment number and the registration number of each core courses. A high percentage (100%) means the course scheduling decision is good.

2. **Enrollment rate for elective courses**: each student should be able to take electives in a high enrollment rate. Student preference will influence the course scheduling decision, we can calculate the percentage of the enrollment number and the registration number of each core courses. We assume that more than 70% is a high enrollment rate for electives. A high percentage means the result is good.

3. **Student time preference**: Student preference will influence the course scheduling decision, each one will have different time preference. In order to measure "goodness", assuming we know students time preference, we compare the course scheduling decision with students' time preference, if the time of a certain course is same, we mark it as 1; otherwise, 0. And then we calculate the average of time satisfactory of each student. And then we calculate average satisfactory rate for all student.

Ø For faculty:

1. **Time Span between classes:** In order to improve faculty time efficiency, a good course scheduling decision should consider the time span between classes in one day. To compute this index, we assume that if the break is less than half hour, the index is 2; if the break is less than two hours but more than half hour, the index is 1; if the break is more than two hours, the index is 0. We add the index of each professor of each day, and a lower total index is better.

2. **Physical distance between classes.** A measurement of faculty time efficiency, similar with the one above. a good course scheduling decision should consider the location of two continuous classes. We calculate the physical distance,

3. **Faculty time preference**: Faculty preference will influence the course scheduling decision, each one will have different time preference. In order to measure "goodness", assuming we know faculty time preference, we compare the course scheduling decision with faculty time preference, if the time of a certain course is same, we mark it as 1; otherwise, 0. And then we calculate the average of time satisfactory of each faculty. And then we calculate average satisfactory rate for all faculty. A higher rate is better.

Ø For department:

1. **Total number of classes per professor:** this metric shows total number of sessions each professor teaches. We can measure if the allocation of classes are reasonable for professors and whether their workload is heavy or not.

2. **Student enrollment rate per course:** number of enrolled students for each course divided by capacity of the course. It is an important index showing the popularity of the course and can be taken into account when avoiding conflicts between core courses and popular electives. We assume

that course has an enrollment rate lower than 30% is less popular and mark it 0; 30%-70% is normal and mark it 1; higher than 70% is popular and mark it 2.

3. **Student enrollment rate per session:** number of enrolled students for each session of a course divided by capacity of the session. We set a threshold (30%) and if the minimum student enrollment rate is not met, the session should be cancelled. Besides, we set a threshold (100%) and if maximum student enrollment rate is exceeded, new section needs to be created.

4. **Prerequisite for taking course:** take value as TRUE (1) or FALSE (0). Some courses are open only to students from specific programs. If a student meets the prerequisite, value 1 is assigned, otherwise, value 0 is assigned.

5. **Course cancellation rate per department:** number of cancelled courses per department divided by total number of courses each department. The cancellation rate is the lower the better. The best possible value is 0% and worst possible value is 100%. Cancelled course occupies time slots and classrooms during allocation process, and leads to a waste of time and effort. We should try to minimize the number of cancelled courses and focus on departments with a high cancellation rate. We assume that cancellation rate higher than 10% is bad and mark 0; cancellation rate lower than 10% is good and mark 1.

Ø For Program:

1. **No conflict between core courses and popular electives:** a good course scheduling decision should not make confliction between core courses and popular electives. Students are free to choose what they want to learn. It is assumed that we know the core courses and popular electives of each program. We measure it by calculate the conflict rate of each program.

2. **Special requirement of classroom:** the course scheduling decision will be affected by the requirements of classroom, such as equipment and layout. We compute this matrix by satisfaction rate.

Ø For Management:

1. **Minimize cost to allocate:** cost includes all cost incurred from course scheduling process. For instance, the salary paid to office staff and data analysts.

2. **Minimize time to allocate:** time includes all time spent in course scheduling process. For instance, time spent before all courses are arranged and time spent on running the course scheduling model.

3. **Utilization of classroom:** capacity of the classroom divided by number of registered students. This metric shows the utilization rate of each classroom. 0% is the worst value possible and 100% is the best value possible. We assume that rate higher than 70% is a suitable case. We should change the classroom when the rate is lower than 50%.

4. **Percentage time a building each classroom is occupied:** total amount of time a building or a classroom is occupied divided by total prime time (10 a.m. – 4 p.m.) in a week. This metric shows

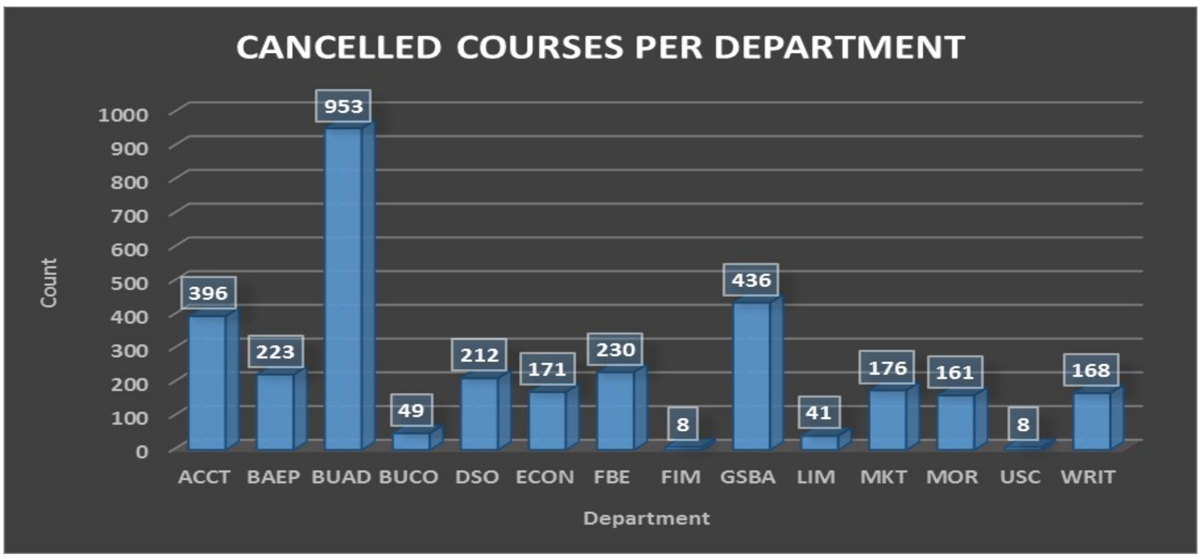
the utilization of building and classroom. High percentage indicates a high utilization and low percentage indicates a waste of resources. The worst value possible is 0% and the best value possible is 100%. Besides, since Marshall plans to undergo a large-scale renovation of Bridge and Accounting buildings in next few years, we should try to use other buildings first and minimize the use of BRI and ACC.

Section 5.1 : Exploratory Analysis and Data Visualization

After analyzing all the data provided to us, using Python, we came up with the following insights.

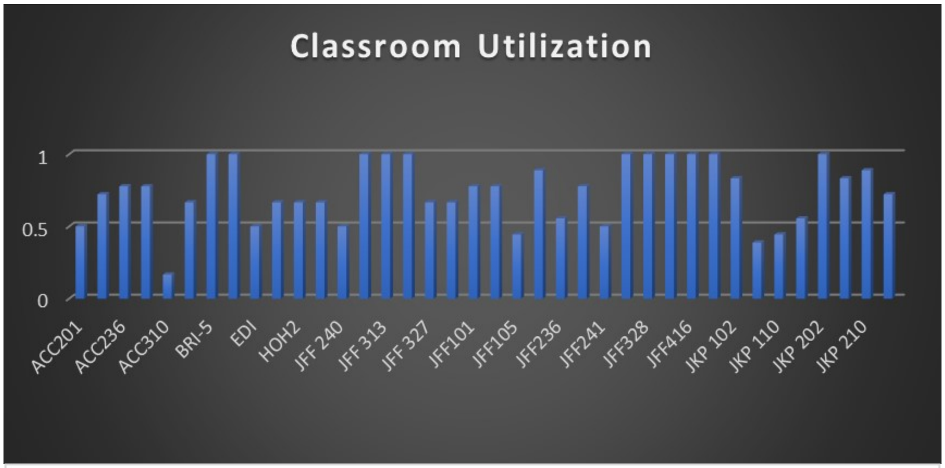
1. Cancelled Courses:

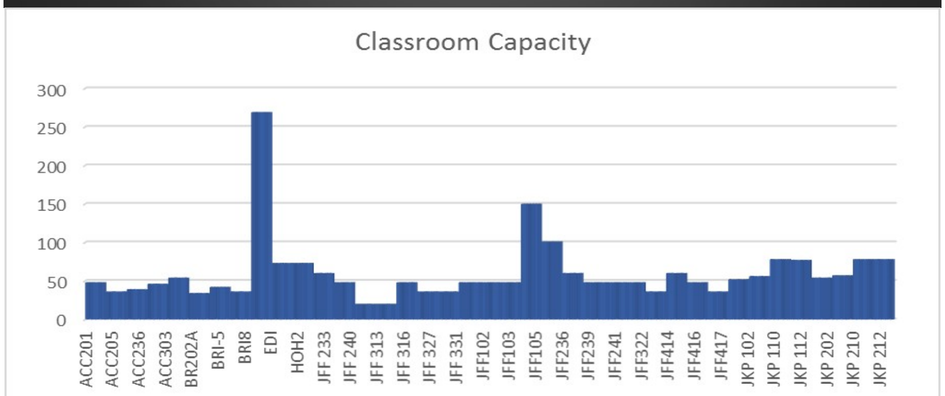
This graph shows the cancelled courses for each department. A total of 3064 courses were cancelled across all programs in the past two years, which indicated high inefficiency. A maximum number of 953 courses were cancelled in the BUCO department, followed by MKT, ACCT, BAEP and EMBA departments. In contrast, MBV, MMIS Core and DSO program nearly had no cancelled courses. Thus, our goal is to decrease the number of cancelled courses for all departments and programs so as to reduce unnecessary wastes of labor and time.



2. Room Utilization:

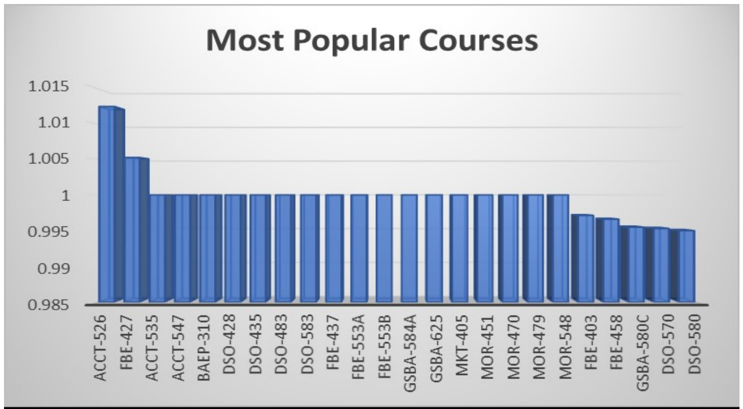
These two graphs show the utilization rate and capacity of each room. For this metric, we think over 50% of utilization is efficient and utilization rate of less than 30% is inefficient. The total utilization rate was only 28.2% which suggested large space for improvement. For all types of buildings, JFF and JKP tended to have higher utilization rate compared to ACC, BRI, EDI and HOH. Also, we notice that classrooms with larger capacity usually had small utilization rate. Thus, our goal is to increase the utilization rate of each room and make full use of the resources. We could increase the utilization rate of the room by matching appropriate sizes of classes to corresponding rooms.





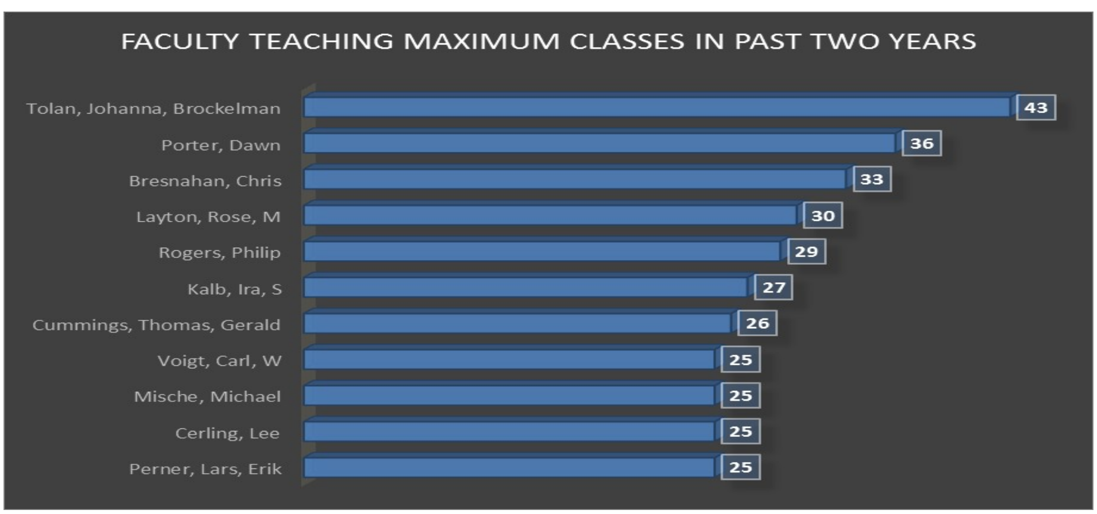
3. Course Popularity:

These two graphs show the most and least popular courses. For this metric, over 70% is classified as popular, 30%-70% is classified as normal while less than 30% is unpopular. We mainly focused on the most and least popular courses because we should increase the capacity of those popular courses while trying to reduce those unpopular courses. For instance, ACCT-526 is over-registered and the capacity of this course could be increased in the future. For those courses that had fewer than 15% registration rate, they may be cancelled later and we should also try to reduce the openings of these courses.Currently, there are still some courses that are either over-popular or too unpopular, which implicates large room for improvement. Thus, our goal is to make better allocations of the courses based on the popularity.



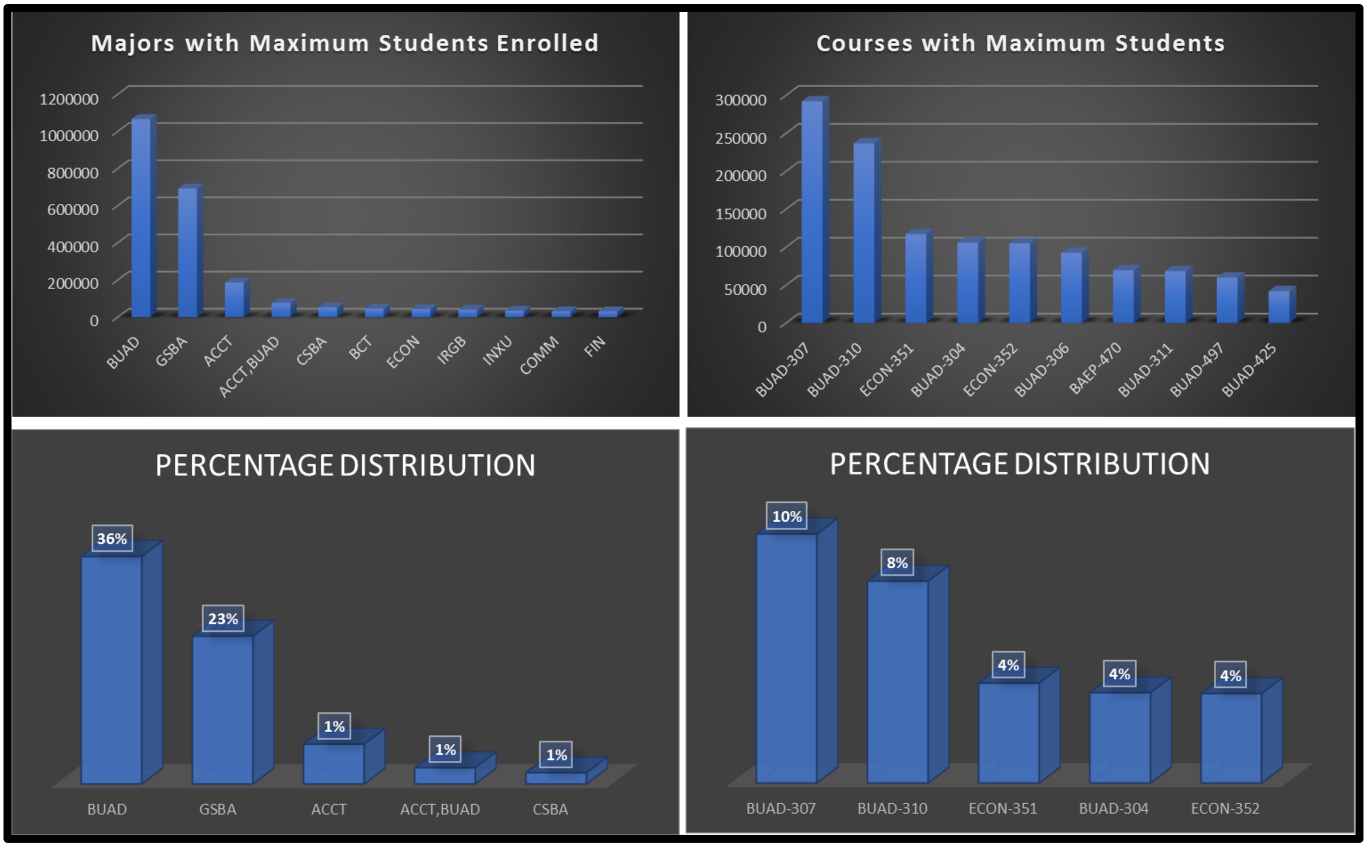
4. Faculty Utilization:

This graph shows the total number of courses taught by each faculty in the past two years and we focused on the faculty teaching most courses per semester. We noticed that some faculty had much greater workload than other faculty with a maximum of 43 courses while some faculty only taught one course for the semester. Thus, our goal is to even the number of courses taught by every faculty to ensure the standard and quality of the classes.



5. Student Distribution:

The following graphs show the distribution of students enrollment regarding specific majors and courses. We could see that majors with maximum students enrolled were BUAD, GSBA and ACCT. Also, courses with maximum students were mainly BUAD and ECON. Since BUAD and GSBA have most students involved, we could prioritize these two majors when optimizing. Also, we could better allocate courses to specific time slots after knowing their popularity.



Section 5.2: Python Codes

```
In [119]: print("The number of popular courses is:", len(popular_course))
     ...: print("The number of normal courses is:", len(normal_course))
     ...: print("The number of less popular courses is:", len(less_popular_course))
The number of popular courses is: 279
The number of normal courses is: 118
The number of less popular courses is: 38
```

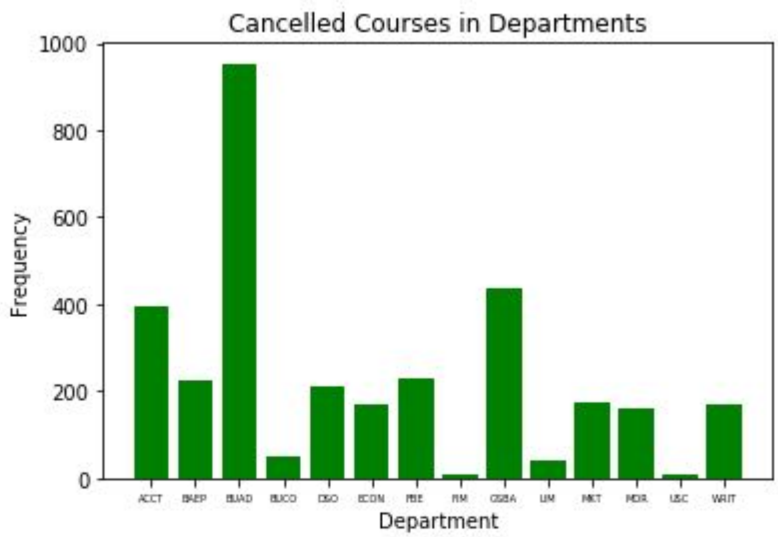
Here we defined the enrollment rate which is more than 70% as popular class, between 30% and 70% as normal class and less than 30% as unpopular class.

```
In [123]: print("The average utilization is:", average(output['Utilization']))
The average utilization is: 0.35695473251
```

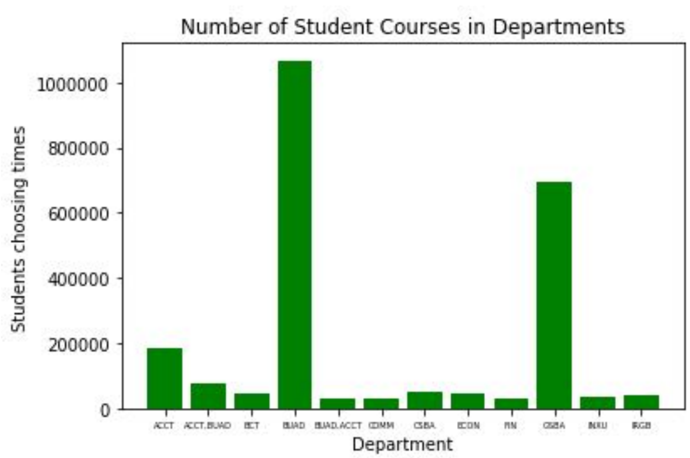
During the prime hours from 10 am to 4 pm, the average utilization rate of classrooms is 0.36.

```
'Rogers, Philip': 1686,
'Selby, Richard': 1005,
'Suh, Julie': 1320,
'Mische, Michael': 1077,
'Tolan, Johanna, Brockelman': 1478,
'Perner, Lars, Erik': 1568,
'Porter, Dawn': 2141,
'Voigt, Carl, W': 1367,
'Kalb, Ira, S': 1589,
'Hopkins, Merle, W': 1054,
'Layton, Rose, M': 1355,
'Bemis, Nimfa, Abarquez': 1058,
'Bresnahan, Chris': 1437,
'Fields, Kevin, Clark': 1007,
'Bristow, Duke': 1047,
'Burgos, Miriam, T': 1141}
```

Here, we get the instructors with largest number of students enrolled.



Here we get the cancelled courses for each department.



Here we get the number of student courses for each department.

Appendix:

See the Python codes in the Zip File.