DSO 570 Final Project
Phase 1

Group 7

Executive Summary

Historically, scheduling courses and arranging classrooms for the entire Marshall business school has proven to be a challenging task for the administrative team. During phase 1, the group examined all existing data and leveraged external sources. The group identified a possible measure to improve the current course scheduling system by reducing conflicts of popular courses. The three defined metric scores: popularity of courses, conflict and popular course conflict are utilized to measure 'goodness'. At the culmination of this report, an initial analysis is presented using the metrics defined above.

Opportunities

There are 5000 undergraduate and 1000 graduate students in Marshall. From the students perspective, the chief complaint is lacking the option to enroll in majority of the popular courses. The current course scheduling places most of the 'popular courses' in the same time slot, leaving students with their secondary choices. Additionally, a large section of the student population, barring the part time students, want to have all courses during 'prime time', which is during 9am to 5pm.

Therefore, the group recommends reducing conflicts of popular courses as a possible remedy. Reducing conflicts of popular courses can serve to improve student satisfaction rate, increase registration rate, and potentially reduce dropout rate. Thereby, enhancing overall student satisfaction.

Metrics

The goal of the project is to ensure that students have the ability to select all the popular courses they desire. Therefore, the solution would be to first define the popularity level of all courses, and then analyze whether the top 10 percent are clashing with each other. Finally, securing a solution to minimize the conflicts between popular courses.

Data Source

In addition to existing data files, the group used two more files provided by MBA students - 'Peer Course Evaluation Spreadsheet Spring 2018' and 'Spring 2017 MBA Elective Registration Information - as of 2016.10.25'.
In the file 'Peer Course Evaluation Spreadsheet Spring 2018', the two columns used: 'Would you recommend this course (regardless of professor?)' and 'Would you recommend this professor

(regardless of course?)'. Conversely, in the file 'Spring 2017 MBA Elective Registration Information - as of 2016.10.25.csv', the three columns used: 'Max Cap' (The maximum number of students that a certain course could hold), 'Max Bid' (The maximum number of points that a student put on this course. The higher the 'Max Bid', the more popular the course.) and 'Number of bids' (The number of students who want to choose a certain course.)

Metrics

1. **Popularity Score= Professor Rating + Class Rating + Utilization (Number of bids/Max Cap) + Max Bid**
    - For the 'Professor Rating' and 'Class Rating', the group used columns 'Would you recommend this course (regardless of professor?)' and 'Would you recommend this professor (regardless of course?)' in file 'Peer Course Evaluation Spreadsheet Spring 2017'. If the record is 'strongly recommend', it was changed to 2, if it is 'recommend', it was changed to 1, if is it 'strongly not recommend', it was changed to -2, and if it is 'not recommend', it was changed to -1. The last step is to z-scale all these records.
    - The 'Utilization' is set equal to the 'Number of bids / Max Cap'
    - The 'Max Bid', we z- scale all the records and use the new ones after z- scale
2. **Conflict score = # of classes in the same department in the same time slot**
   Here, the file 'Merged_Enrollment' is used to illustrate number of classes in the same department during the same time slot
3. **Popular conflict score = # of popular classes in the same time slot**
   After getting the 'Popularity Score' for all records, they are rearranged from high to low, to get the top 10 percent of popular courses. The next task is to calculate the conflict score for these top 10 percent of popular courses

Justification

The objective is to maximize the popularity score throughout the week, while ensuring that that students have a option to select of all/most popular courses at once. For this objective function, the metrics considered: Popularity Score, Conflict Score and Popular Conflict Score.

1. **Popularity of the class**
    a. To quantify the data, the group considered the professor ratings from the student feedback dataset. This dataset contains the feedback given by students after taking the class. As such a dataset is readily available, computation of this metric becomes easy. Here, the ratings such as "Strongly recommend, recommend, do not recommend and strongly do not recommend" were encoded into numeric codes. The new data set allows for a direct judgement of whether the class is popular amongst the students or not, and if optimization can be done accordingly. Furthermore, the interpretation of the data set directly impacts the decision-making of the stakeholders; thereby, making it an actionable metric.

2. **Conflict Score**
    a. This measure indicates the number of classes offered in the same time slot by the same department. As we have the datafile “Merged_Enrollment” that tells us the number of classes offered by a department in the same time slot, this metric becomes computable. If the conflict score for a class is high, it means that other classes are being offered at the same time by the department. This is problematic for the students who want to take both classes as then they have to pick between them. Hence, this metric directly affects the scheduling decisions.
3. **Popular Conflict Score**
    a. This metric is similar to the previous one except for the fact that an extra layer of “popularity” is added to the conflict score. This is intended to optimize the scheduling of popular classes such that no two popular classes are conflicted. The metric is solely aimed at enriching the college experience of students, by ensuring that all the popular classes are offered at separate times.

Analysis

Based on the defined metric above, the aggregated popularity scores for each of the classes is calculated. In this report, the analysis remains focused on term 2017. Below are top 10 most popular classes and top 10 least popular classes, based on the popularity score.

This methodology of scoring allows comparison of different classes quantitatively using the preference of students. However, as the data source is limited to MBA peer evaluations, the data is bereft of student ratings for undergraduate courses. Under the given scenario, the extreme values may appear to be more extreme than they really are, compared to the whole population.

	class_scores	max_bid_scores	utilization_scores	popularity_score	very_popular
DSO-570	1.244623	2.093239	4.975856	8.313718	True
DSO-580	0.945341	3.013637	3.167168	7.126146	True
DSO-583	0.817355	2.421545	3.365251	6.604150	True
BAEP-564	0.463368	1.730634	2.508811	4.702813	True
DSO-547	0.874437	1.995705	1.714980	4.585121	True
MOR-588	-0.486806	0.513237	4.339821	4.366252	True
DSO-545	0.287410	2.109573	1.897796	4.294779	True
ACCT-559	0.991243	1.650055	1.425155	4.066453	True
BAEP-559	0.590058	1.592615	1.751173	3.933846	True
DSO-586	0.463368	0.484381	2.234750	3.182500	True

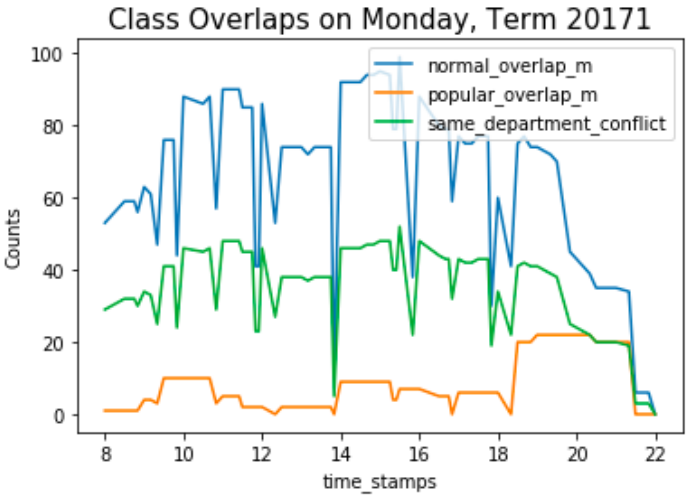
Top 10 most popular classes

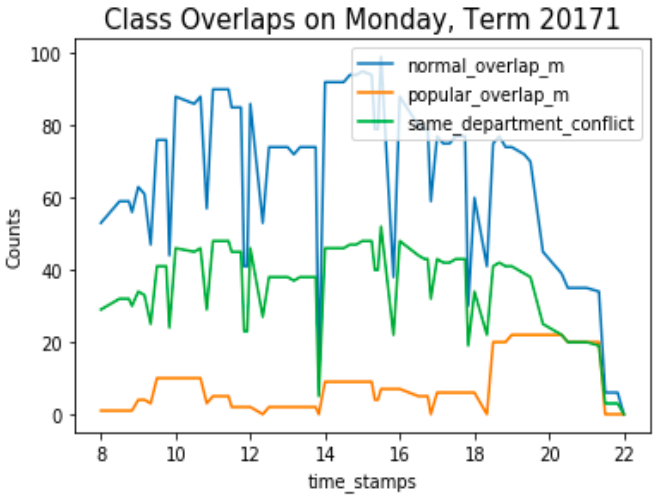
	class_scores	max_bid_scores	utilization_scores	popularity_score	very_popular
FBE-559	-3.654054	-0.667136	-0.910099	-5.231290	False
ACCT-528	-2.387155	-1.444615	-1.121282	-4.953051	False
MOR-554	-3.654054	0.133863	-0.251009	-3.771201	False
ACCT-558	0.000000	-2.199227	-1.401522	-3.600749	False
DSO-534	-1.120256	-1.447882	-1.012578	-3.580715	False
DSO-581	-1.120256	-1.528243	-0.847571	-3.496070	False
ACCT-574	0.000000	-1.157144	-1.247712	-2.404856	False
BAEP-555	1.413543	-2.300495	-1.507598	-2.394550	False
DSO-516	-1.492873	-0.832106	-0.063791	-2.388769	False
MKT-526	-1.880395	0.007589	-0.454898	-2.327704	False

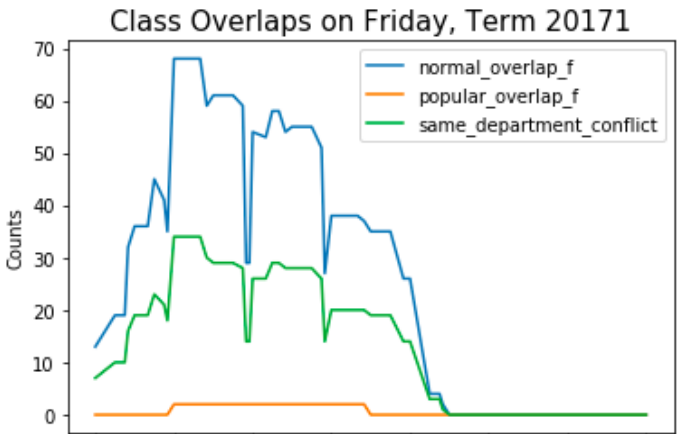
Top 10 least popular classes

Despite working with limited data and a biased sample set, the group tried to explore this scoring system as a benchmark of measuring the goodness of scheduling (for students). If the result turns out to be informative, making this course bidding / peer evaluation process universal to all Marshall may be a feasible solution to acquiring sufficient data for enacting the scheduling algorithm.

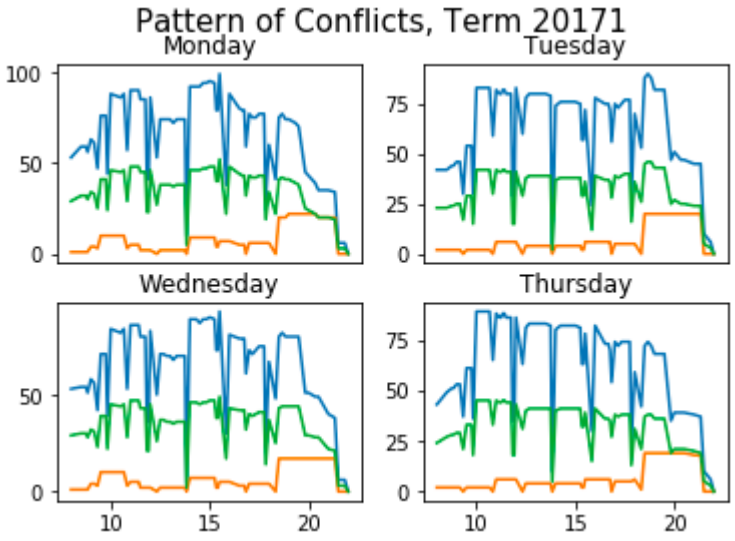
Below are two typical graphs of class overlaps. Classrooms in Marshall are generally busy whole day from Monday to Thursday, and come to lower utilization on Friday.

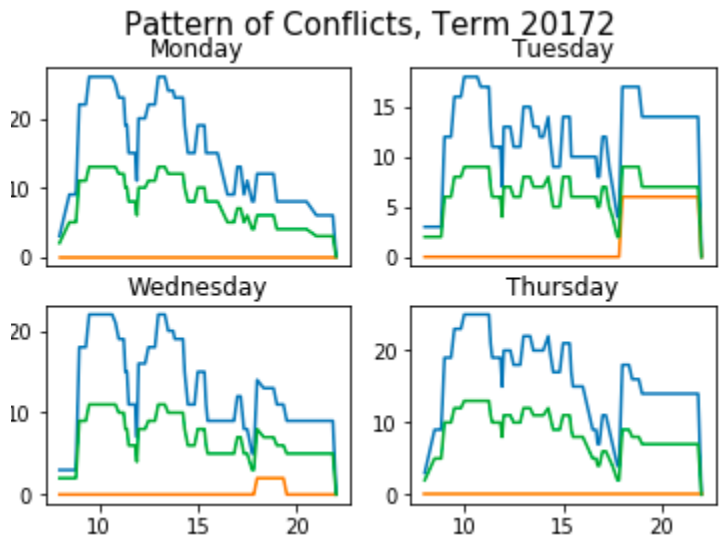






Note that since the calculated 'popular_overlap' is based on MBA peer evaluation, it is evident that in term 20171, a bump of counts of popular classes overlapping each others appears consistently on the time slot 18:30 to 21:30. In term 20172, there are fewer overlaps because total class count is lower. However, on Tuesday we can still discover a stressful bump of overlaps in the time slot 18:30 to 21:30.





The above graphs demonstrate how popularity score and the derived popular conflict counts are computable and simple to interpret. A fair actionable objective is to lower popular conflict counts throughout the week. Based on the popularity score system and derived conflict counts, without further constraints, it is reasonable to move popular classes to Friday, especially in term 20171.

There are several possible improvements that can made the metric to make this popularity score system a reliable representation of the student preferences at Marshall. In addition to acquiring representative ratings from students, the score derived conflict counts should be subject to departments, or at least be grouped by undergraduate or graduate, for the conflicts not considering undergraduate or graduate difference are possibly just false positives. After the necessary adjustments, a holistic system rooted in student preferences, professor preferences and school limitations can be formulated.

References

Spring 2017 MBA Elective Registration Information:
https://docs.google.com/spreadsheets/d/1je_aKSwAiFx7NAP1LDxV5YzkUEwMBTsm-RbOhh7J4Qg/edit#gid=1646438476

Peer Course Evaluation Spreadsheet Spring 2018:
https://docs.google.com/spreadsheets/d/1_ttx75T1FxP8ZfhfaMOgQ_gjJ4VwNc6edZpsDtCX8So/edit#gid=97807236

Appendix

Codes and outputs

Codes and outputs are stored here.