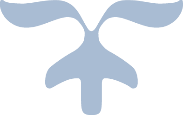
FINAL REPORT

DSO 570





**EXECUTIVE SUMMARY**

Shannon and Hal are in charge of scheduling classes for the USC Marshall School of Business, which contains 7 departments and 22 academic programs. In the scheduling process, there are many aspects that need to be taken into consideration, like the utilization rate of the classroom, the student’s preference, the professor’s preference, cancellation rate, etc. All these aspects make the scheduling process a very complex problem. A perfect schedule that satisfies everyone’s preferences and maximizes classrooms’ utilization rate is almost impossible to achieve. On the other hand, an improperly planned schedule could result in a series of complications for both the school and the university.

The current process of scheduling class for Marshall School is not efficient enough. Shannon and Hal are seeking the opportunity to optimize the scheduling process as well as decrease their workload by partnering with the DSO-570 students from the M.S. in Business Analytics program.

This analysis will focus on optimizing the classroom utilization rate which is a main factor to consider when formulating a final schedule. We used Marshall’s classrooms capacity data and class enrollment data as inputs to our statistical analysis and created a Mixed Integer Programing (MIP) formulation in python.

Our MIP builds a robust model that allows changes in course sections or classrooms. This feature gives the register office (Shannon and Hal) more flexibility when making decisions in the future. The output is a sheet containing binary values indicating whether we designate a specific classroom for a specific course section in a specific day pattern. Based on the result sheet we created a table which contains the course sections and the corresponding optimal combination of classroom and day pattern.

With our formulation, we were able to increase the total utilization rate of Marshall classrooms by 19.2%. This means that more classrooms will be available. Therefore, the classroom system could both intake more course sections and increase its flexibility to satisfy temporary adjustment.

**BUSINESS PROBLEM**

We identified that the most time-consuming part is assigning sections to classrooms. In the current course scheduling system, sections are manually assigned to each classroom. The inefficiency includes both the heavy workload for Shannon and her team and the relatively inefficient classroom utilization rate.

Shannon, Hal, and their team need to put lots of effort when scheduling the classroom due to both the large number of sections and classrooms and the complexity of course conflicts. There are 484 sections (809 sessions) which are supposed to be assigned to 45 Marshall classrooms. Since the number of classrooms is limited, administrative office need to consider increasing classrooms utilization rate as far as possible to allow assignment of more courses, which makes the process even more complicated.

In order to ease the burden for Shannon and Hal, our team design a MIP (Mixed Integer Program) that can automatically assign course sections to classrooms on a specific day of week/days of week and also maximize the classroom utilization rate. This can substantially solve the inefficiency mentioned above and moreover, reduce human labor.

**VISION**

The purpose of this project is to improve the efficiency of Marshall’s phase one scheduling process. For our analysis, efficiency is measured in terms of classroom utilization rate. In other words, from our analysis we hope to increase the overall classroom utilization rate by the end of phase one.

To do so, our team will use Marshall’s classrooms capacity data and previous class enrollment data as inputs to our statistical analysis and formulate an optimization program in python. We will only focus on optimizing the utilization rate of Marshall classes that are usually offer from Monday to Thursday. We decided to leave out Friday and the weekend from our formulation as the number of classes that usually meets on these days are much smaller compared to the rest. Hence, the scheduling of classes on these days could be done manually by Shannon and Hal.

The final output from our code will be a sheet containing all the section numbers with their designated classroom and day of the week.

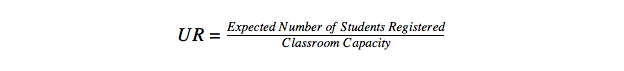
We believe our analysis would be of great help to Shannon and Hal because with our result, their workload during phase one and phase two should decrease substantially. Moreover, our analysis could help Marshall make a better use of their resources at hand, making a significant profit by reducing the amount of empty seats per class.

The details of our analysis are explained on the following sections.

**METRIC**

As previously mentioned, our main performance metric will be the *Overall Classroom Utilization Rate* of all Marshall classrooms. Given that our purpose, from a financial and administrative point of view, is to minimize the number of empty seats in each classroom, our metricfits perfectly our business model. By maximizing the overall utilization rate, Marshall can make a better use of its resources which translates into a more profitable and flexible business.

In order to calculate the overall utilization rate we first compute the utilization rate for each course section given its assigned Marshall classroom. Each section will have an utilization rate equal to the expected number of students to be enrolled in that section divided by the capacity of the classroom to which it was assigned. The number of students enrolled in each section can be obtained from the Marshall\_Course\_Enrollment dataset and the capacity of each Marshall classroom is available from the Marshall\_Room\_Capacity\_Chart. The formula for calculating utilization rate for a given section and a given classroom is shown below:



When calculating the utilization rate for each section given a classroom for a given term, the expected number of students to be enrolled in a given section will be equal to the number of students previously enrolled in such section in the previous term of its kind. In other word, for calculating the expected number of students per section for the term spring 2018, we will use the actual number of students registered for those sections in the term spring 2017. Same logic applies when calculating the expected number of students for terms summer 2018 and fall 2018.

Finally, the *Overall Classroom Utilization Rate (OCUR)* will be the sum of all utilization rates across all sections given their designated classrooms.

Screen%20Shot%202018-05-01%20at%2011.59.04.png

**ASSUMPTIONS**

In our formulation, we made a few important assumptions to conduct a thorough analysis:

* All Marshall classrooms are available from 8 AM - 10 PM (14 hours)
* Reserve certain hours for one-session sections and the rest hours for two-session sections for each classroom every day.
* Use spring 17 data to predict the expected number of student registered for each course on spring 18
* Ignore non-USC campus course (online, office class, San Diego, and Shanghai)
* There must be a 10 mins break between different classes
* Exclude the records with more than 2 sessions per week
* Exclude the Friday and weekend course sections
* Instead of considering the day of week, we consider six day patterns, M, T, W, H, MW, and TH

**FORMULATION**

Based on the information above, we have formulated a mixed integer program (MIP) that applies optimization to solve this business case.

**Decision Variable**

Our decision variable is *xij*. The notation i represents each course section and the notation j is the combination of classroom and day pattern. There are 6 day patterns for j: M, T, W, H, MW, and TH. The letters M, T, W, and H stand for Monday, Tuesday, Wednesday, and Thursday respectively. This subcategory of day patterns are reserved for sections that meet once a week since they only need a classrooms once every week. MW stands for Monday & Wednesday and TH stands for Tuesday & Thursday. This subcategory of day patterns are reserved for sections that meet twice a week. We assume that this kind of sections meet either on Monday and Wednesday or on Tuesday and Thursday.

The decision variable *xij* is a binary variable that indicates whether the course section i should be assigned to a specific classroom in a specific day pattern j or not. There are two values for our decision variable: 1 means section i will be assigned to a specific classroom in a specific day of week j; 0 means not. For example, if the decision variable *x14028,ACC306B\_M* = 1, it indicates that course section 14028 will be assigned to classroom ACC 306B on Monday; if *x14028,ACC306B\_MW* = 1, it indicates that course section 14028 meets two sessions per week and it will be assigned to classroom ACC 306B on Monday and Wednesdays.

**Data Preparation:**

* **I**: set of all sections that need to be assigned a classroom (I = I1 + I2)
* **I1**: set of sections that have one session per week
* **I2**: set of sections that have two sessions per week
* **J**: set of combinations of classroom and day pattern (J = J1 + J2)
* **J1**: set of combinations of classroom and day pattern: M(Monday), T(Tuesday), W(Wednesday), H(Thursday)
* **J2**: set of combinations of classroom and day pattern: MW(Monday & Wednesday), TH(Tuesday & Thursday)
* **C*j***: classroom capacity of j
* **ES*ij***: expected number of student registered in section *i* for each classroom and day pattern *j*
* **T*i***: duration in hours of section *i*
* ***uij***: classroom utilization rate of section *i* assigned to classroom and day pattern *j*
* **a**: amount of hours reserved for one-session sections for each classroom everyday
* **b**: amount of hours reserved for two-session sections for each classroom everyday

**Objective**

In this analysis, our objective is to maximize the sum of the classroom’s utilization rate for all sections. To do so, we need the utilization rate, which is computed using the expected number of student registered in section i divided by the capacity of its assigned classroom j. We named this utilization rate *uij*. We use our decision variable *xij* to multiply it by *uij* and then calculate sum across the products. Our objective is to determine *xij* that maximizes the sum of the product.

**Constraints**

Given the decision variable and objective, we have built certain constraints based on the real-life situation to limit certain levels of optimization.

*Time constraint*

For each classroom, we assume that total time period that can be used every day is 14 hours, from 8am to 10pm. Among these 14 hours, we reserve “**a**” hours for sections that meet once a week and “**b**” hours for sections that meet twice a week. Therefore, we have two time constraints that limit the sum of the duration of sections that are assigned to the same classroom every day. For each classroom with assigned section, the sum of the duration for sections that meet once a week must be less than or equal to “**a**” hours and the sum of the duration for sections that meet twice a week must be less than or equal to “**b**” hours.

*Capacity constraint*

Since different classrooms have different capacities. We want to make sure that each section will be assigned to a classroom that have enough seats. Thus, the capacity constraint limits that the capacity of the assigned classroom should be larger than or equal to the expected number of students registered in the section.

*One classroom constraint*

Each section only needs one classroom. This means that the sum of decision variable *xij* for the same section *i* across all possible classrooms *j* needs to be equal to 1.

*Mutually Exclusive constraint*

For section that meet once a week, they cannot be assigned to classrooms during the timeframes that are reserved for section that meet twice a week. Similarly, for section that meet twice a week, they cannot be assigned to classrooms during the timeframes that are reserved for section that meet once a week. For example, if section 14028 only meets once a week, the decision variable *x14028,ACC306B\_MW*  cannot be equal to one.

**INPUT DATA**

**Original Data Source**

The input data was abstracted from the following two datasets:

1. “Marshall\_Room\_Capacity\_Chart.xlsx” (hereinafter referred to as “**capacity data**”)
2. “Marshall\_Course\_Enrollment\_1516\_1617.xlsx” (hereinafter referred to as “**course data**”).

**Data Reorganization**

The following are the datasets we generated from the two original datasets mentioned above and input to our python code:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Description** | **Source** | **Related Index** |
| **data3** | Marshall course enrollment - 2017 Spring | course data | I, I1, I2, T, S |
| **ES\_final** | Expected Students for each section: a dataset duplicates the Registered Students for each section 270 times for future coding | course data | related to decision variables and constraints |
| **classroom** | Classroom capacity | capacity data | J1, J2, J, C |

**RESULTS**

**1. Description of Output**

The analysis gives two output tables.

The first one is a table called “BIG\_OUTPUT.xlsx”. It shows the results for all the binary decision variable in the model. The table has 484 rows indicating all course sections, and 270 columns indicating all combinations of classrooms and day patterns. The cell value is binary with “1” indicating to assign the section to a certain combination of classroom and day pattern and “0” indicating not to do so.

The second output table is called “FINAL\_SCHEDULE.xlsx”. It only has two columns. The first one lists all the 484 course sections, and the second one lists the optimal combination of classroom and day pattern that course section will be assigned to.

For visual and simplicity reasons, the second table is easier for decision makers to have a quick overview of what the optimal decisions should be.

**2. Discussion about Results and Model Performance**

After applying the optimization, many course sections achieve higher classroom utilization rate. For instance, in the first term in 2017, course section 14042 had 32 students registered and was assigned to room “JFF 236” with capacity 60 on both Tuesday and Thursday. The room utilization rate was only 53.3% for both days. After applying the MIP, the course section 14042 is assigned to “BRI 202A” room with capacity 34 on both Monday and Wednesday. The room utilization rate increases to 94.1% if the section is registered by the same number of students.

Although not every course section achieves increase in classroom utilization rate, in fact, there are some course sections remain the same classroom utilization rate or even suffer a slight decrease in classroom utilization rate after applying the optimization. However, the overall sum of classroom utilization rate for all course sections increases significantly. In the first term of 2017, the sum of classroom utilization rate is 357.48. After applying the MIP formulation, the optimal value reaches 426.19, which represents a 19.2% increase.

The increase in utilization rate is gained from assigning course sections to the classrooms with capacity larger than but as close to the number of expected registered students as possible.

Please notice that for course sections held on two days per week, we only calculate the unitization rate once for both original schedule and optimized schedule. Meanwhile, course sections previously held on Fridays and Weekends are regarded as special sections (usually MBA courses) and are thus excluded from both original and optimal schedule while calculating the overall classroom utilization rate. The analysis ensures that the overall calculation rates for both historical schedule and optimal schedule are calculated for the same course sections and based on same methodology. Therefore, the results for the two schedules are completely comparable.

Because there is a large number of course sections and classrooms, the optimization is pretty complex. Even a slight increase in utilization rate is a great progress. Therefore, the 19.2% increase is significant enough. In other words, increasing the utilization rate by 19.2% also means releasing more classrooms or more classrooms with relatively large capacity. Therefore, the classroom system could both intake more course sections and increase its flexibility to satisfy temporary adjustment.

Meanwhile, we separated the 14 daily work hours for all classrooms into two parts, “a” and “b”. The value of “a” means how many hours are reserved for one-session course sections from Monday to Thursday. The value of “b” means how many hours are reserved for two-session course sections on Monday/Wednesday or Tuesday/Thursday. For instance, section 14226 is optimally assigned to classroom “HOH 706”only on Tuesday, and thus it occupies some time in “a”. Section 14028 is optimally assigned to classroom “JFF LL101” on both Tuesday and Thursday, and thus it occupies some time in “b”. By doing so, we achieved the significant increase in overall sum of classroom utilization rate (19.2%) both while ensuring that course sections held twice per week will be scheduled to the same classroom either on Monday/Wednesday or Tuesday/Thursday, which could significantly increase the convenience for both students and professors.

During the analysis process, we tried different reasonable combination of “a” and “b”, for example, (a,b) = (5,9), (a,b) = (4,10), etc. Given the current information about course sections and expected number of registered students, (4,10) combination gives us the maximum optimal value. However, the optimal combination of “a” and “b” might change in different situation. Decision makers could always adjust “a” and “b” under different situations.

We ignore the time order for sections assigned to a certain combination of classroom and day-pattern in this current stage. However, this will not affect the decision-making efficiency because we did the “a” and “b” separation thus the time slot is detailed enough. It is simple for administrative office to sort the sections within a given time slot.

Furthermore, our MIP builds a robust model that allows changes in course sections or in classrooms. This feature gives register office (Shannon and Hal) more flexibility when they make decisions in the future.

**RECOMMENDATIONS**

In light of the analysis above, Marshall is able to maximize the overall classroom utilization rate by applying the optimization. As next steps, we suggest Marshall to go further on the following aspects:

**1. Pursue even more precise estimation on the number of incoming registered students.**

After applying the optimization, the scheduled classroom always has a capacity larger than but very close to the number of expected registered students. For instance, section 14042 with 32 estimated students will be assigned to “BRI 202A” with capacity 34. If the estimated number of registered students is far smaller than true value, the scheduled classrooms cannot accommodate all the students registered and thus cause great trouble.

To avoid this troublesome situation, we recommend administration office to do more detailed analysis to estimate the number of registered students for each course section. Administrative office could also put stricter limit during the pre-register process to achieve more precise estimation on the number of registered students.

**2. Keep MBA courses in a fix day pattern.**

The analysis does not take Friday/Weekend and MBA courses held on those days into consideration, because MBA students tend to only have time to attend the class on those certain days. Therefore, we suggest administrative office to lock related course sections on Friday or weekends as far as possible to avoid affecting the optimization process for other courses.

**Limitations & Next Steps**

This project is a simplification of a real problem and only focuses on the utilization rate of the classroom, but as we mentioned before, there are many other factors need to be considered when making the schedule. We listed some factors that could affect the scheduling process and that should be taken into consideration for the next steps

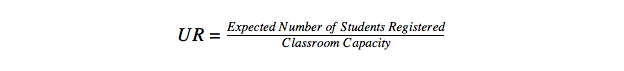
* **Relied on historical data**: We use the historical data of registered student to calculate the expected class size. However, if we change the location or the time of the course, the popularity of the course may also change.
* **Faculty Schedule**: Many professors are not full-time professors. Thus, for some courses, we need to consider the professor’s schedule when scheduling the courses.
* **Manual work**: Our model could assign the class to a specific classroom and specific day pattern. However, the staff needs to assign a specific start time for each class.

**TECHNICAL APPENDIX**

**Metric**

**1. Utilization Rate (UR) for Each Course Section:**

Utilization rate equals to the expected number of students to be enrolled in that section divided by the capacity of the classroom to which it was assigned.



**2. Overall Classroom Utilization Rate (OCUR) for all Course Sections:**

Screen%20Shot%202018-05-01%20at%2011.59.04.png

**Assumption**

* All Marshall classrooms are available from 8 AM - 10 PM (14 hours)
* Reserve certain hours for one-session sections and the rest hours for two-session sections for each classroom every day.
* Use spring 17 data to predict the expected number of student registered for each course on spring 18
* Ignore non-USC campus course (online, office class, San Diego, and Shanghai)
* There must be a 10 mins break between different classes
* Exclude the records with more than 2 sessions per week
* Exclude the Friday and weekend course sections
* Instead of considering the day of week, we consider six day patterns, M, T, W, H, MW, and TH

**Mathematical Formulation**

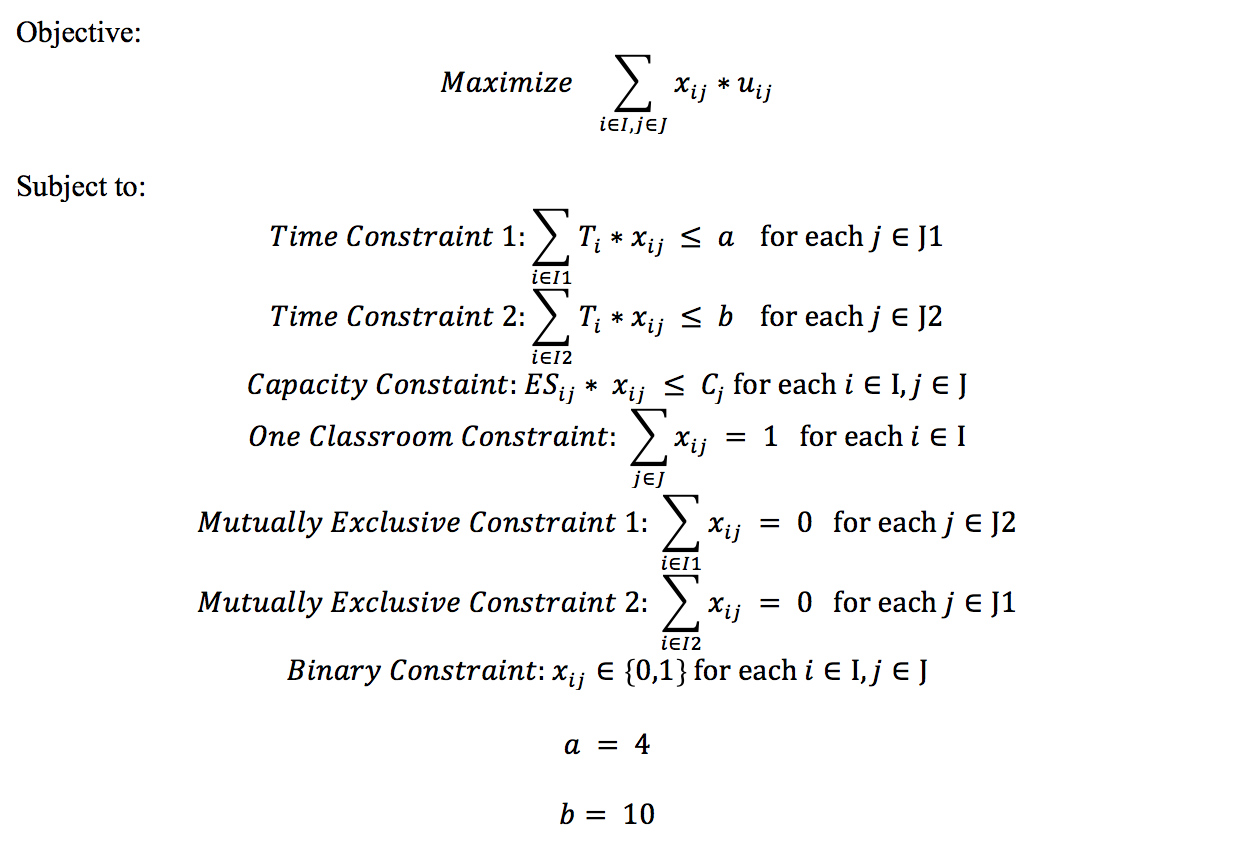
**Data Preparation**:

* **I**: set of all sections that need to be assigned a classroom (I = I1 + I2)
* **I1**: set of sections that have one session per week
* **I2**: set of sections that have two sessions per week
* **J**: set of combinations of classroom and day pattern (J = J1 + J2)
* **J1**: set of combinations of classroom and day pattern: M(Monday), T(Tuesday), W(Wednesday), H(Thursday)
* **J2**: set of combinations of classroom and day pattern: MW(Monday & Wednesday), TH(Tuesday & Thursday)
* **C*j***: classroom capacity of j
* **ES*ij***: expected number of student registered in section *i* for each classroom and day pattern *j*
* **T*i***: duration in hours of section *i*
* ***uij***: classroom utilization rate of section *i* assigned to classroom and day pattern *j*
* **a**: amount of hours reserved for one-session sections for each classroom everyday
* **b**: amount of hours reserved for two-session sections for each classroom everyday

**Decision Variable**: *xij* (binary)

Indicates whether the section *i* should be assigned to a specific classroom in a specific day of week/specific days of week *j* or not. 1 means course *i* will be assigned to a specific classroom in a specific day of week/specific days of week I; 0 means otherwise.

**MIP:**

****

**Input Data and Format:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Format** |
| **data3** | Marshall course enrollment - 2017 Spring | xlsx |
| **ES\_final** | Expected Students for each section: a dataset duplicates the Registered Students for each section 270 times for future coding | xlsx |
| **classroom** | Classroom capacity | xlsx |

**Output Data and Format:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Description** | **Format** |
| **BIG\_OUTPUT.xlsx** | Contains all optimal decision variable values | xlsx |
| **FINAL\_SCHEDULE.xlsx** | Provides information of final schedule for each course section | xlsx |