WARWICK BUSINESS SCHOOL
THE UNIVERSITY OF WARWICK

# Masters Programmes: Group Assignment Cover Sheet

| | |
|---|---|
| **Student Numbers:** Please list numbers of all group members | **5582680, 5504970, 5550485, 5544857, 5586034** |
| **Module Code:** | **IB9190** |
| **Module Title:** | **Advanced Analytics: Models and Applications** |
| **Submission Deadline:** | **12:00 (UK time) Tuesday, 12 March 2024** |
| **Date Submitted:** | **10 March 2024** |
| **Word Count:** | **1522** |
| **Number of Pages:** | **4** |
| **Question Attempted:** *(question number/title, or description of assignment)* | **3** |
| **Have you used Artificial Intelligence (AI) in any part of this assignment?** | **Grammar checking** |


**Academic Integrity Declaration**
We're part of an academic community at Warwick. Whether studying, teaching, or researching, we're all taking part in an expert conversation which must meet standards of academic integrity. When we all meet these standards, we can take pride in our own academic achievements, as individuals and as an academic community.

Academic integrity means committing to honesty in academic work, giving credit where we've used others' ideas and being proud of our own achievements.

In submitting my work, I confirm that:

- I have read the guidance on academic integrity provided in the Student Handbook and understand the University regulations in relation to Academic Integrity. I am aware of the potential consequences of Academic Misconduct.
- I declare that this work is being submitted on behalf of my group and is all our own, , except where I have stated otherwise.
- No substantial part(s) of the work submitted here has also been submitted by me in other credit bearing assessments courses of study (other than in certain cases of a resubmission of a piece of work), and I acknowledge that if this has been done this may lead to an appropriate sanction.
- Where a generative Artificial Intelligence such as ChatGPT has been used I confirm I have abided by both the University guidance and specific requirements as set out in the Student Handbook and the Assessment brief. I have clearly acknowledged the use of any generative Artificial Intelligence in my submission, my reasoning for using it and which generative AI (or AIs) I have used. Except where indicated the work is otherwise entirely my own.
- I understand that should this piece of work raise concerns requiring investigation in relation to any of points above, it is possible that other work I have submitted for assessment will be checked, even if marks (provisional or confirmed) have been published.
- Where a proof-reader, paid or unpaid was used, I confirm that the proof-reader was made aware of and has complied with the University's proofreading policy.

**Upon electronic submission of your assessment you will be required to agree to the statements above**

**Table of Contents**

## Introduction

The COVID-19 pandemic has reshaped education, prompting the need for adaptable scheduling systems. Moallemi and Patange's hybrid scheduling model for Columbia Business School (2023), outlined in the INFORMS Journal on Applied Analytics, inspires our project to explore its feasibility for Warwick Business School's MSBA course. Our task involves designing a hybrid timetable for primarily in-person classes, blending in-person and online attendance. Through extensive data analysis and optimization, we aim to tailor a scheduling solution for MSBA students. This project not only addresses immediate scheduling needs but also contributes to the broader discussion on post-pandemic education, showcasing the practical application of advanced analytics in hybrid learning models.

## Description of Student-Level Scheduling

The scheduling optimization problem allocates students to in-person classes, balancing social distancing and equitable attendance distribution. Through mathematical formulation, it utilizes parameters and decision variables to achieve an efficient solution. The parameters are the value to be held constant in the model which can be seen in Table 1.

| | | |
|---|---|---|
| Student: i | Total number of students: S | Number of students in a class: $A_k$ |
| Class: k | Total number of classes: C | Social distancing capacity in a class: $c_k$ |
| Group: j | The number of group division (integer): M | Number of classes ongoing at time t: $C_t$ |
| Time: t | Total number of hours in a week: T | Weight of the total deviation metric: λ |
| Weight of the surplus simultaneous excess: μ | | |

**Table 1**. Parameters

The decision variable is the decision of allocation of a student in a group, assigned as $\pi_{ij}$, while the objective function is formulated as follows.

$$\min \sum_{k\in C}\sum_{j=1}^{M} e_{jk} + \lambda \sum_{k\in C}\sum_{j=1}^{M} \sigma_{jk} + \mu s$$

To assess the model's performance, three key metrics have been established, utilizing parameters and decision variables. These metrics measure both the excess capacity within classes and the discrepancies in attendance across student groups. By minimizing these metrics, the model aims to optimize allocation and ensure equitable attendance distribution. Achieving smaller values indicates improved performance in minimizing excess capacity and ensuring fair attendance opportunities for all students. These metrics, as auxiliary variables, provide detailed insights into the allocation process and attendance distribution.

- Surplus Simultaneous Excess (SSE) is the maximum number of students assigned to the excess room at any given time: $s = (max_{tj}\ s_j^t - E)^+$.
- Total Excess (TE) is the sum of excess from each class when each of the groups is scheduled: $e_{jk} = (\sum_{i \in A_k} \pi_{ij} - c_k)^+$.
- Total Deviation (TD) is the sum of the absolute difference between an assignment and the uniform fractional student assignment: $\sigma_{j,k} = |\ \sum_{i \in A_k} \pi_{ij} - \frac{|A_k|}{M}|$.

Meanwhile, the constraints for this problem are listed below.

- Allocation constraint dictates that each student can only be assigned to one group: $\sum_{j=1}^{M} \pi_{ij} = 1$.
- Notation constraints mean that the auxiliary variables are converted into equalities to ensure the positive part notation and the absolute notation of them are compatible with the optimisation algorithm.
  - For TE: $\sum_{i \in A_k} \pi_{ij} - c_k \leq e_{jk}$
  - For TD: $-\sigma_{jk} \leq \sum_{i \in A_k} \pi_{ij} - \frac{|A_k|}{M} \leq \sigma_{jk}$
  - For SSE: $s \geq s_j^t - E$
- The binary constraint for $\pi_{i,j}$ is where 1 indicates allocation and 0 indicates non-allocation to the group: $\pi_{ij} \in \{0,1\}$.
- Non-negative constraint ensures that these variables can never be negative, as it would be illogical to have a negative number of students: $\sigma_{jk},\ e_{jk},\ s \geq 0$.

There are several assumptions used for this optimisation model.

1. Students have the flexibility to enrol in various classes and are thus assigned to classes individually.
2. Each day only one group of students is allocated to in-person attendance for all classes scheduled on that day (the same group attends all classes on the given day).
3. Any additional students beyond the class capacity are randomly assigned to an excess room for the duration of that class.
4. Surplus students are unable to participate in classes online while physically present on campus.
5. There exists a single excess room designated for accommodating students from all classes.

### Hybrid Timetable Scheduling In-person for MSBA Students

In our scenario, there are 229 MSBA students and 6 main modules. S is the set of these 229 students. Each module is divided into different types of classes for students to take, for example, AAMA module has 1 lecture and 4 workshops, which means 5 classes in total for

AAMA. In a similar way, ADA, DM, SCA, F, and FA have 5, 4, 3, 3, and 3 classes separately resulting in 23 classes after all. Thus, C is the set of these 23 classes.

With the allocation file provided by the Programme Team, we transformed the information into a new data frame. The columns are the 23 different classes mentioned in previous paragraph and the rows are all 229 MSBA students. If student i is allocated for class k, then the value of the cell in row i and column k is 1, otherwise 0. This table is named A. Each column Ak is the set of students enrolled in class k. For the social distancing capacity, we assumed that COVID-19 pandemic will decrease the capacity of each classroom in our university by 30% and this proportion will be further explored to see how it will influence our scheduling. Therefore, we built the table ***ck*** to store the new capacity for each class which is the original capacity times 30%.

Based on the timetable given, we created a data frame with the column being 23 classes and the row being a time series ***T*** from 0 to 167 representing every hour in a week. If class k is ongoing at time t, then the value for the cell in row t and column k is 1, otherwise 0. This table is named ***Ct***. In terms of excess room, we chose 2.007 that has capacity of 40 students in Warwick Business School (WBS). So, ***E*** is 40 times 30% in this question. E is dynamic due to the change of proportion of social distancing limitation. As for the other parameters, we set $\boldsymbol{\lambda}$ to 0.25 and $\boldsymbol{\mu}$ as 0 according to the paper. M is set to be 2 and will be tried different values as comparison.

In Python implementation, we followed the mathematical formulation step by step. Initially, we defined decision variables: $\pi_{ij}$ as binary, $e_{jk}$ as nonnegative integers, $\delta_{jk}$ as nonnegative Reals, $s$ as nonnegative Reals, and $s_{tj}$ as nonnegative integers. Subsequently, we constructed the objective function and constraints based on the paper's formulations and inequalities. Due to slow solving with the "glpk" solver for large variable sizes, we switched to the "gurobi" solver, which yielded an objective value of 337.

**Comparing Different Solutions**

After solving the scheduling problem in Python and adjusting parameters, results show a notable correlation between parameters and outcomes. Key parameters analysed were social distancing capacity, excess room capacity, and parameter M, representing workshop in-person attendance frequency.

As M increases from 2 to 10, varying social distancing capacities (p) at 0.3, 0.25, and 0.2, alongside an excess room capacity set at 120, objective values fluctuate notably. Initially dropping sharply, the objective value decreases from 337 at M=2 to 46.33 at M=3. It continues to decrease, reaching a low of 9.20 at M=5, before a slight increase to 11.67 at M=6. Various solutions were explored by adjusting M and social distancing capacity to achieve SSE=0 and minimize the objective value. The best objective values for p=0.3, p=0.25, and p=0.2 are as follows.

| Social Distancing Capacity (p) | M | TE | TD | SSE | Objective Value |
|---|---|---|---|---|---|
| 0.3 | 5 | 0 | 36.8 | 0 | 9.2 |
| 0.25 | 4 | 16 | 32 | 0 | 24 |
| 0.2 | 3 | 40 | 25.33 | 0 | 46.33 |

**Table 2**. The best results of different parameters

The table showing the full results with different values of social distancing capacity and M can be seen in the Appendix 1. Comparing the previous result when p=0.3 and M=2, with the new solution p=0.3 and M=5, changing one parameter (M) impacted the overall result greatly with 337 and 9.2 objective values respectively with SSE remained 0. Assuming that the p is decreased to 0.25, we should set M=4, so that the SSE will be 0, TE=16, TD=32 and the objective value will be equal to 24. If the social distancing capacity continuously decreases to 0.2, M should be equal to 3, so that SSE=0, TE=40, TD=25.33 and objective value will be equal to 46.33.

## Conclusions

The analysis highlights the significant impact of adjusting parameters, particularly social distancing capacity (p) and the number of workshop attendances (M), on scheduling efficiency. Results demonstrate the necessity for meticulous parameter selection to achieve optimal outcomes. However, it is time-consuming to find the best result when changing the value of parameters manually as we did. Therefore, to improve the proposed formulation, Sensitivity Analysis (SA) can be used for identifying the best parameters. SA can identify parameters of which a reduction in uncertainty specification will have the most significant impact on improving model performance measures. As a result, the computational cost will be decreased without reducing the model performance (Gan *et al.,* 2014).

Additionally, there are limitations included, such as fixed values for certain parameters (λ at 0.25 and μ at 0) and a singular room capacity consideration (40). Despite this, the study underscores the critical role of parameter selection in optimizing scheduling models, paving the way for future enhancements in computational efficiency and performance.

**References**

Gan, Y. *et al.* (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling and Software*, 51, pp. 269–285. Available at https://doi.org/10.1016/j.envsoft.2013.09.031 (Accessed 01 March 2024)

## Appendix 1. The Result Comparison of Different Parameters

| Social Distancing Capacity (p) | M | TE | TD | SSE | Objective Value |
|---|---|---|---|---|---|
| 0.3 | 2 | 335 | 8 | 0 | 337 |
| 0.3 | 3 | 40 | 25.33 | 0 | 46.33 |
| 0.3 | 4 | 16 | 32 | 0 | 24 |
| 0.3 | 5 | 0 | 36.8 | 0 | 9.2 |
| 0.3 | 6 | 0 | 46.67 | 1 | 11.67 |
| 0.3 | 7 | 0 | 57.71 | 2 | 14.43 |
| 0.3 | 8 | 0 | 58 | 3 | 14.5 |
| 0.3 | 9 | 0 | 71.11 | 4 | 17.78 |
| 0.3 | 10 | 0 | 74.4 | 5 | 18.6 |
| 0.25 | 2 | 335 | 8 | 0 | 337 |
| 0.25 | 3 | 40 | 25.33 | 0 | 46.33 |
| 0.25 | 4 | 16 | 32 | 0 | 24 |
| 0.25 | 5 | 0 | 36.8 | 1 | 9.2 |
| 0.25 | 6 | 0 | 46.67 | 2 | 11.67 |
| 0.25 | 7 | 0 | 57.71 | 3 | 14.43 |
| 0.25 | 8 | 0 | 58 | 4 | 14.5 |
| 0.25 | 9 | 0 | 71.11 | 5 | 17.78 |
| 0.25 | 10 | 0 | 74.4 | 6 | 18.6 |
| 0.2 | 2 | 335 | 8 | 0 | 337 |
| 0.2 | 3 | 40 | 25.33 | 0 | 46.33 |
| 0.2 | 4 | 16 | 32 | 1 | 24 |
| 0.2 | 5 | 0 | 36.8 | 2 | 9.2 |
| 0.2 | 6 | 0 | 46.67 | 3 | 11.67 |
| 0.2 | 7 | 0 | 57.71 | 4 | 14.43 |
| 0.2 | 8 | 0 | 58 | 5 | 14.5 |
| 0.2 | 9 | 0 | 71.11 | 6 | 17.78 |
| 0.2 | 10 | 0 | 74.4 | 7 | 18.6 |

Cells highlighted as yellow are the best solutions (objective value should be minimum when SSE is 0).