Masters Programmes: Assignment Cover Sheet

Student Number:	5519743, 5554086, 5568440, 5587165, 5588409
Module Code:	IB9190
Module Title:	Advanced Analytics: Models and Applications
Submission Deadline:	Tuesday, 12th of March 2024
Date Submitted:	Saturday, 9th of March 2024
Word Count:	1566
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?	No

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 the work is all my 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

Task 1: Student-Level Scheduling Optimisation Problem Simulation

1.1. Introduction of the Problem

Due to the pandemic, Moallemi and Patange (2023) proposed a hybrid scheduling system with a Student-Level Scheduling approach at Columbia Business School. This system schedules students individually and places them into groups considering social distancing guidelines. On each day, only one group is allowed to attend classes in-person, which rotates daily.

1.2. Parameters

The problem parameters include sets of students, classes, student class enrolment, class capacity after COVID, hours on each day of the term; term timetable; the specified number of groups and the excess room capacity (see Appendix 1.A).

1.3. Objective Function

The methods for evaluating group allocation can be ascertained using three relevant metrics:

1. **Surplus Simultaneous Excess (SSE)** is the maximum number of students assigned to the excess room at any given time, which exceeds its capacity.
2. **Total Excess (TE)** represents the total excess from each class, assuming each group is already scheduled for in-person attendance. `
3. **Total Deviation (TD)** measures the absolute difference between actual allocation and the ideal scenario, where every student is equally placed in all the groups.

The objective is to minimise the weighted sum of the three metrics. However, since SSE is minimal, they set its weight to 0, examining the optimal results afterward. TE's weight is set to be 1 and TD's weight is set between 0 and 1, practically, 0.25 (see Appendix 1.A. and 1.C.).

1.4. Decision Variables

The key variable is binary representing whether a student is allocated to a specific group. The second variable is auxiliary for simultaneous excess (SE), the maximum students exceeding the classroom capacity for all classes at the same time. The last three variables are auxiliary for TE, TD and SSE respectively. All auxiliary variables are non-negative while SE and TE are integers as the number of students cannot be fractional (see Appendix 1.B.).

1.5. Constraints

The problem consists of five constraints (see Appendix 1.D. for further details):

1. **Unique Group Assignment** ensures that each student is assigned to only one group.
2. **Excess Class Capacity** links the key variable to TE's auxiliary variable by ensuring that the total excess for each class does not exceed TE.
3. **Allocation Equality** transforms the value of TD into a linear constraint by linking the key variable to TD's auxiliary variable.

4. **Simultaneous Excess** links TE's auxiliary variable to SE's auxiliary variable by ensuring SE is the maximum of the total TE of all classes at any given time.
5. **Surplus Simultaneous Excess** links SE's auxiliary to SSE's auxiliary variable by ensuring SSE is the maximum of SE over excess room capacity at any given time.

1.6. Assumptions

The authors stated three key assumptions. The first assumption is that all classes, currently in-person, will be changed to hybrid mode. Additionally, they assume that any one of the groups can be assigned in-person attendance on any given day. The last assumption is that each class follows a schedule that repeats weekly.

Task 2: Hybrid Timetable Scheduling using WBS Case

We apply the hybrid scheduling model of Task 1 for MSBA student at University of Warwick for Term 2 and consult the MSBA for its application, assuming a similar situation as COVID.

2.1. Key Assumptions

- The timetable provided is fixed and student must attend their allocated classes for modules selected by them including electives.
- All classes will be moved to hybrid mode and no student opted out of in person classes.
- The timetable remains the same for all weeks with wbs.live changed to 0.004.
- The distance between two seats is 1 metre and total seats are equivalent to total capacity provided for the room in timetable.

  There are more than 1 excess rooms available throughout the week (Room 0.006 and 1.009 with the total capacity before COVID of 221).
- Number of groups formed and number of days classes are scheduled are co-primed. Allocated group attend all classes in person for that day and 1 group come each day.

2.2. Data Structuring

We have created 3 data frames to define our parameters (detailed in Appendix 2).

1. **Student Class Enrolment** represents individual students in rows and all lectures and workshops in columns. The value is 1 if they are enrolled in the respective class by the programme team, otherwise 0. This table defines the set of students (S), set of classes (C), and set of students enrolled for each class A_k.
2. **Room Capacity** represents the designated rooms for each class, along with their original capacities. Following guidelines by the UK Government (COVID-19 response: Summer 2021, 2021) and the University of Warwick (Potter, 2020) being 1 or 2 meter distancing, the seating capacities is reduced to 50% and 33%, defining C_k.

3. **Class Timetable** outlines the weekly schedule from Sunday 12AM to Saturday 12AM, with 30-minute time slots (T). The value of each timeslot (t) is '1' if a class is scheduled during that time and '0' otherwise. This table also defines term timetable C_t.

We followed the paper in assigning the weights of TE, SSE and TD in the objective function. Considering the assumption of multiple excess rooms, SSE can be minimal, thus, its weight will be set at 0. Besides, the number of groups (M) is set to be co-primed with weekly working days (e.g. 5 days) so that the rotation sequence can be maintained across the term.

2.3. Implementation Issue

- The allocation of weights for objective function was difficult to set as they can be changed based on requirements of hybrid scheduling.
- Final value of number of groups is ascertained after trial and error. M should be small so that many students attend classes but not too small that objective value is large.
- The benchmark that every student attend classes in person a fraction of times which indicates every student allocated a class will attend equal number of times as other student using Uniform fractional student allocation which is not possible in real life implementation of allocation.

Task 3: Problem Solving and Analysis

3.1. Analysis

Our primary goal is to keep students safe while maximising their in-person participation. Therefore, it turns out that this is an optimal student assignment for the integer linear program. We used gurobi solver version 11.0 in Pyomo in Python to implement this optimisation project at WBS since gurobi outperformed glpk in computational capacity. We have run glpk for 10 hours and did not get the result. It is worth noting that we set the MIP equal to 0, and all cases achieved optimal results in less than 10 seconds of runtime (see Table 1).

Furthermore, as M increases, almost all SEs and TEs are 0 and the value of TD gradually increases, which also proportionately increases objective value. However, when the social distance protocol is 2 metres (33% original capacity), we observed TE becomes 0 when M changes from 2 to 3, suggesting that too small M can potentially lead to high TE, increasing the objective value. Additionally, the changing of social distancing protocol from 50% to 33% does not have a significant impact on the objective value.

Table 1 – Result summary with different social-distancing protocols and the number of groups

Social-distancing protocol	M	TE	SSE	TD	Objective value
50%	2	0	0	8.00	2.00
50%	3	0	0	25.33	6.33

50%	4	0	0	32.00	8.00
50%	6	0	0	46.67	11.67
33%	2	211	0	8.00	213.00
33%	3	0	0	25.33	6.33
33%	4	0	0	32.00	8.00
33%	6	0	0	46.67	11.67

3.2. Implementation Improvements

- Optimisation objective can be updated to consider: (1) maximising student engagement in class (e.g. timetable scheduling is optimised so that students can join as many classes in person as possible on the day they are allowed to), (2) maximising students' satisfaction, measured by individual attendance preference (e.g. online versus in-person) and effective timetables which can minimise students' disutility such as long gaps between classes, (3) optimising classroom utilisation (e.g. allocating classes to their best-fit room size).
- Integrate machine learning algorithms (e.g. genetic algorithms) to predict scheduling preferences, resource allocation and identify patterns from historical scheduling and feedback data from previous cohorts to improve schedule effectiveness for the upcoming cohorts.
- The current formulation neglects the interaction between students from different groups as only one group is scheduled to attend classes in person on a specific day. Therefore, there might be a lack of connection among the cohort, which is essential for students of a business school. Since the current TE and SSE result is 0, the model can be updated to allow more than 1 group to attend classes in one day or students can be rotated to different groups during the term to promote a sense of community across the cohort.
- Currently, TD is the factor driving the objective value of the model. While trying different weights for TD, we noted that the objective value changed proportionally, meaning that the current model has already optimised the value of TD given the current constraints. We recommend that for the classes with high TD (classes with empty seats available), the empty seats for each class on each day can be offered to students who want to attend class in person on a first-come-first-serve basis at the beginning of the term. Therefore, not only TD will be reduced while health safety is maintained but also students have more opportunities to attend classes and mingle with other classmates.

Reference

Moallemi, C. and Patange, U. (2023). 'Hybrid Scheduling with Mixed-Integer Programming at Columbia Business School', INFORMS Journal on Applied Analytics, Articles in Advance, pp. 1-19.

Potter, A. (2020). New measures to make WBS campus 'Covid-secure'. Available at: https://www.wbs.ac.uk/news/new-measures-to-make-wbs-campus-covid-secure/ (Accessed: 22 February 2024).

Tinsley, B. (2020). 'Coronavirus and the impact on students in higher education in England: September to December 2020', Office for National Statistics. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/articles/coronavirusandtheimpactonstudentsinhighereducationinenglandseptembertodecember2020/2020-12-21 (Accessed: 22 February 2024).

UK Government (2021). COVID-19 response: Summer 2021. Available at: https://www.gov.uk/government/publications/covid-19-response-summer-2021-roadmap/covid-19-response-summer-2021 (Accessed: 21 February 2024).

University of Warwick (2020). Latest Government guidance on reopening spaces. Available at: https://warwick.ac.uk/insite/news/guidance_on_reopening_spaces/ (Accessed: 21 February 2024).

Warwick Business School (2020). How face masks can affect social distancing. Available at: https://www.wbs.ac.uk/news/how-face-masks-can-affect-social-distancing1/ (Accessed: 21 February 2024).

Appendix 1: Formulation

A. Required Parameters

No	Parameters	Description
1	S	a set of students
2	C	a set of classes
3	A_k	set of students enrolled in each class $k \in C$
4	C_k	class capacity after COVID, for each class $k \in C$
5	T	the set of hours in a week
6	t	an hour in a week; $t \in T$, t = 0, 1, …, 167 (t=0: Sunday 12 AM; t=167: Saturday 11 PM)
7	C_t	set of classes scheduled to be ongoing at time t, $t \in [start\ time\ \pm \frac{1}{6}, end\ time\ \pm \frac{1}{6}]$
8	M	number of groups (M > 0)
9	d	each day of the term
10	$T_d \subseteq T$	set of hours on day d
11	E	capacity of excess room
12	$w_{TE} = 1$	weight of TE is set to be 1
13	$w_{TD} = \lambda = 0.25$	weight of TD is set between 0 and 1, but in practice, it is set to 0.25
14	$w_{SSE} = \mu = 0$	weight of SSE is initially set to 0, since SE is typically smaller than the excess-room capacity

B. Decision Variables

No	Decision Variables	Description
1	$\pi_{i,j} \in \{0,1\}$, for i = 1, …, S; j = 1, …, M	$\pi_{i,j} = 1$ if the student i is assigned to group j, otherwise is 0.
2	$s_j^t \geq 0$, for j = 1, …, M; t = 0, 1, …, 167	The number of students in the excess room at time t for group j if group j is assigned to in-person attendance (related to SE).

	Auxiliary variable representing: $\sum_{k \in C_t}(\sum_{i \in A_k} \pi_{ij} - c_k)$ if $\sum_{i \in A_k} \pi_{ij} \geq c_k$	
3	$e_{jk} \geq 0$ for j = 1, …, M; k = 1, …, C Auxiliary variable representing $$\sum_{i \in A_k} \pi_{ij} - c_k$$ when $\sum_{i \in A_k} \pi_{ij} \geq c_k$	The number of students exceeding the social distancing capacity of class k, group j (related to TE).
4	$\delta_{jk} \geq 0$ for j = 1, …, M; k = 1, …,C Auxiliary variable representing $\left\|\sum_{i \in A_k} \pi_{ij} - \frac{\|A_k\|}{M}\right\|$	The absolute value of the difference in the number of students assigned and the uniform fractional student assignment (related to TD).
5	$s \geq 0$ Auxiliary variable representing $\max s_j^t - E$ if $\max s_j^t > E$	The number of students in the excess room that exceeding the excess room capacity (related to SSE).

C. Objective Function

$$\text{minimise } \sum_{k \in C} \sum_{j=1}^{M} e_{jk} + \lambda \sum_{k \in C} \sum_{j=1}^{M} \delta_{jk} + \mu s$$

Minimise the weighted sum of SSE, TE and TD

D. Constraints

No	Constraint	Description
1	**Unique Group Assignment** $$\sum_{j=1}^{M} \pi_{ij} = 1$$ for $\pi_{i,j} \in \{0,1\}$, $i \in S$, $1 \leq j \leq M$	Each student must be assigned to exactly one group.
2	**Excess Class Capacity** $$\sum_{i \in A_k} \pi_{ij} - \mathrm{c}_k \leq e_{jk}$$ for $1 \leq j \leq M, \forall k \in C,\ e_{jk} \geq 0$	Due to social distancing capacities, they try to minimise the number of students assigned within group j that exceed the capacity of class k, so the number of excess constraints linking e_{jk} to π_{ij} (TE). e_{jk} is the maximum excess students for class k from group j. e_{jk} must be greater than every $\sum_{i \in A_k} \pi_{ij} - \mathrm{c}_k$. It ensures the total excess for each class does not exceed TE.
3	**Allocation Equality** $$\left| \sum_{i \in A_k} \pi_{ij} - \frac{\lvert A_k \rvert}{M} \right| \leq \delta_{jk}$$ Equivalent to $\sum_{i \in A_k} \pi_{ij} - \frac{\lvert A_k \rvert}{M} \leq \delta_{jk}$ and -$\sum_{i \in A_k} \pi_{ij} + \frac{\lvert A_k \rvert}{M} \leq \delta_{jk}$ for $1 \leq j \leq M, \forall k \in C, \delta_{jk} \geq 0$	This transforms the value of TD into a linear constraint by linking the key variable to TD's auxiliary variable. They try to make sure that each group holds an equal number of students (total deviation constraint).
4	**Simultaneous Excess** $$s_j^t \geq \sum_{k \in C_t} e_{jk}$$ for $0 \leq t \leq T, 1 \leq j \leq M,$ $e_{jk} \geq 0, s \geq 0$	The authors try to minimise the number of students assigned to the excess room at any given time, so the number of excess students in the excess room at time t for group j must be larger than or equal to the total number of excess students from group j for attending all classes k at time t.

		It links TE's auxiliary variable to SE's auxiliary variable by ensuring SE is the maximum of the total TE of all classes at any given time.
5	**Surplus Simultaneous Excess** $s \geq s_j^t - E$ for $1 \leq j \leq M, 0 \leq t \leq T, s \geq 0$	Due to the limited excess-room capacity, they try to minimise the number of students assigned to the excess room that exceeds its capacity. Surplus simultaneous excess constraint: linking s and $\max s_j^t - E$ if $\max s_j^t > E$. It links SE's auxiliary variable to SSE's auxiliary variable by ensuring SSE is the maximum of SE over excess room capacity at any given time.

Appendix 2: Data Structuring

Data Frame	Index*Columns	Explanation
Student - Class	229 * 23	1. 229 students in MSBA provided with student ID (1, 2, 3, 4, …, 229) which presents 'S' decision variable i.e., set of students 2. There are 6 lecture (1 for each module) Module codes: Core Module - IB9190, IB98D0, IB9HP0 Elective Module - IB98E0, IB9MJ0, IB9BS0 3. 2 modules (IB9190, IB98D0) have 4 workshop each week, 1 module (IB9HP0) has 3 workshop and 3 elective modules (IB98E0, IB9MJ0, IB9BS0) have 2 workshop each week. In Total there are 23 classes which presents 'C' i.e., set of classes and 1 class is denoted as 'k' (IB9190-L, IB9190-W1, IB9190-W2, IB9190-W3, IB9190-W4, IB98D0-L, IB98D0-W1, IB98D0-W2, IB98D0-W3, IB98D0-W4, IB9HP0-L, IB9HP0-W1, IB9HP0-W2, IB9HP0-W3, IB98E0-L, IB98E0-W1, IB98E0-W2, IB9MJ0-L, IB9MJ0-W1, IB9MJ0-W2, IB9BS0-L, IB9BS0-W1, IB9BS0-W2) 4. Each student takes up 4 Modules – 3 core and 1 elective.
Class - Capacity	23*4	1. There are 23 classes in total for which rooms are to be assigned in timetable (as notes in point 3 above) 2. The room no. and capacity are traced from timetable shared by programme team and covid capacity is calculated as 50% or 33% of total capacity represented as C_k capacity of a class with restriction due to covid under hybrid scheduling model.

Time - Class	336*23	1. There are total 168 hours (24hr for 7 days) in which classes can be held which we have divided into 336 time slots (1/2 hr is equal to 1 slot), 0 is 12 am to 12:30 am on Monday, 1 is 12:30 am to 1 am on Monday,…….,335 is 11:30 pm to 12 am on Sunday. 'T' presents the set of half an hour slot for the week. 2. There are 23 classes in total which are held at certain time.

Table A: Based on the allocation shared for each student ID, this dataframe represents A_k which is set of students enrolled in a class.

Student ID	IB9BS0 - L	IB9BS0 - W1	IB9BS0 - W2	IB98D0 - L	Until 23 classes
1	0	0	**0**	**1**	
2	1	0	**1**	**1**	
3	0	0	**0**	**1**	
4	0	0	**0**	**1**	
5	0	0	**0**	**1**	
Until 229 students					

Table B: Based on the actual capacity of room, we have derived covid capacity. C_k represents covid capacity of the class k

Class	Allocated Room	Capacity	COVID capacity (50%)
IB9BS0 - L	2.004	120	60
IB9BS0 - W1	2.007	40	20
IB9BS0 - W2	2.007	40	20
IB98D0 - L	0.004	292	146
Until 23 classes			

Table C: Based on the timing of the class k we created the below dataframe. If class held in the time (t) then assign 1 otherwise we assign 0.

half an hour in a week; $t \in T$, t = 0, 1, …, 335 (t=0: Sunday 12 AM to 12:30 AM; t=335: Saturday 11:30 PM to 12:00 AM)

C_t represents the set of classes schedules at particular time t.

Time	IB9BS0 - L	IB9BS0 - W1	IB9BS0 - W2	IB98D0 - L	Until 23 classes
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
Until 335					