



**UNSW Business School**

**School of Information Systems and Technology  
Management**

## **INFS2822 Programming for Data Analytics (T3 2023)**

### **Group Assignment**



*Image source: <https://www.itnews.com.au/news/transport-for-nsw-ditches-scats-traffic-system-commercialisation-566722>*

## **Table of Contents**

<b>1. Overview .....</b>	<b>3</b>
<b>2. Key Dates .....</b>	<b>4</b>
<b>3. Assessment Instructions.....</b>	<b>5</b>
3.1. Tasks .....	5
3.2. Requirements for the Report.....	10
3.3. A 12-minute Presentation .....	12
<b>4. Diaries and Reflection (Optional).....</b>	<b>13</b>
<b>5. Within-Group Peer Review (Optional) .....</b>	<b>14</b>
<b>6. General Rules .....</b>	<b>15</b>
6.1. Proper Academic Conduct.....	15
6.2. Assignment Submission .....	15
6.3. Professional Group Work.....	16
6.4. Use of AI tools.....	17

## 1. Overview

In this assessment task, you will take the role as a data analyst and prepare a technical report for an external client, e.g., Transport for NSW, to detail your analytics of a dataset of interest. The task is designed to test your Python skills and understanding of data analytics, basic machine learning knowledge, and overall problem-solving skills. You must present your findings, supported by data visualisations, and modelling **in the form of a written report and a presentation/pitch, together with your Jupyter notebook(s).**

- The assignment is to be undertaken in groups of four (three or five fits) students, all of which must be from **the same** lab/tutorial class.
- In this assignment, you would use the data analytics knowledge you gain from the lectures, labs/tutorials, and class exercises to develop your solution(s).
- We will conduct a within-group **peer review** to review each teammate in the group work.
- The group will be awarded a **group mark as a baseline mark**. This baseline mark may or may not be adjusted for the individual group members (based on the within-group peer review).
- Tips based on previous experience:
  - Start **immediately** after group formation.
  - Keep **backup copies** of all your work.
  - Hold **group meetings** at least once, better twice, per week.
  - Define roles and responsibilities within the team; especially you will need a **project leader/group co-ordinator** for handling submissions, scheduling meetings, etc.

## 2. Key Dates

What?	When?
Project group formation – A group of 4 (3 or 5 fits, but no more than 5 people)	<b>Week 3 Lab</b> (Email the signed Group Formation Form to your tutor by Thursday, 28 <sup>th</sup> September 2023, 10:00 am (Sydney Time) – One form per group)
Assignment Due	<b>Week 10 Thursday, 16<sup>th</sup> November 2023, 3:00 pm (Sydney Time)</b> – Submit the <b>written report, Jupyter Notebook(s), and presentation slide deck</b> via Moodle (Submission details are shown in Section 3)

### 3. Assessment Instructions

*“Australia’s roads have been more deadly in every state and territory bar NSW and the Northern Territory in the past year and only NSW has met its national road death toll reduction target.”* ([Sydney Morning Herald, 2023](#)). Even in NSW, the recent increasing number of road fatalities has gain significant attention from the public. To achieve the 2026 [Towards Zero target](#), Transport for NSW has approached your company because you are the best in the industry. They want your team to conduct research on the provided data (**NSW\_Road\_Crash\_Data\_2017-2021\_CRASH**, see Moodle for details) so that they can better allocate their campaign budget in the next 12 months. Transport for NSW has given you two tasks:

**Part A: Use Pandas, and Matplotlib or Seaborn for data visualisation.** In this part, you will need to complete **SEVEN (7)** data visualisation questions.

**Part B: Building models on target variables to predict the most important variables that contribute to recent crash incidents, especially fatal incidents.**

Along with the target variables, the dataset contains a variety of variables detailing information, such as the weather, terrain condition, date, time, etc. However, though there are a lot of variables, not all these variables might be useful in building a promising predictive model. On top of that, some variables might be useful predictors but have data quality issues, e.g., missing values, and require cleaning and pre-processing before they can be used as input into the model.

As the business analytics team, it is up to you to figure out and understand which variables make sense to be used as accurate predictors of the target, and which ones would be best used to build the best predictive model to address the task.

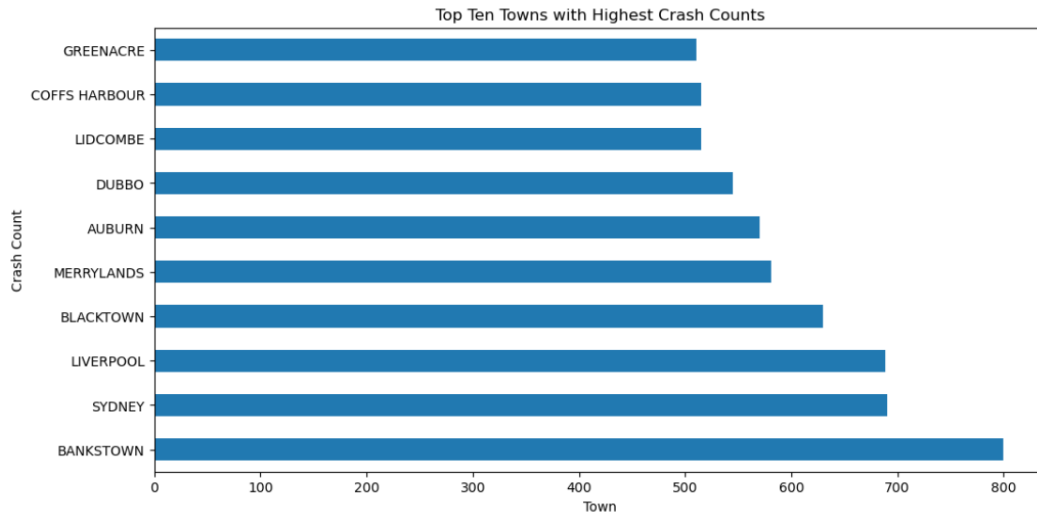
#### 3.1. Tasks

##### 3.1.1. Data Visualisation

*Note: You do NOT need to match the exact colours or styles, but you need to match all the labels, legends, title, xticks and yticks. If you prepare your data properly, you plots will be similar to the graphs in this document. Do NOT include any incidents happened in 2016.*

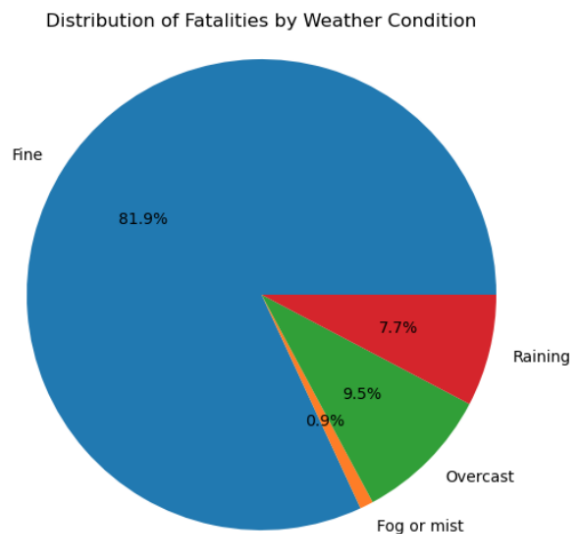
*Use `plt.figure(figsize=(12, 6))` to size your figure.*

1. Which towns/suburbs have the highest number of crashes, and can we compare the crash counts for the top ten towns/suburbs using a horizontal bar chart?



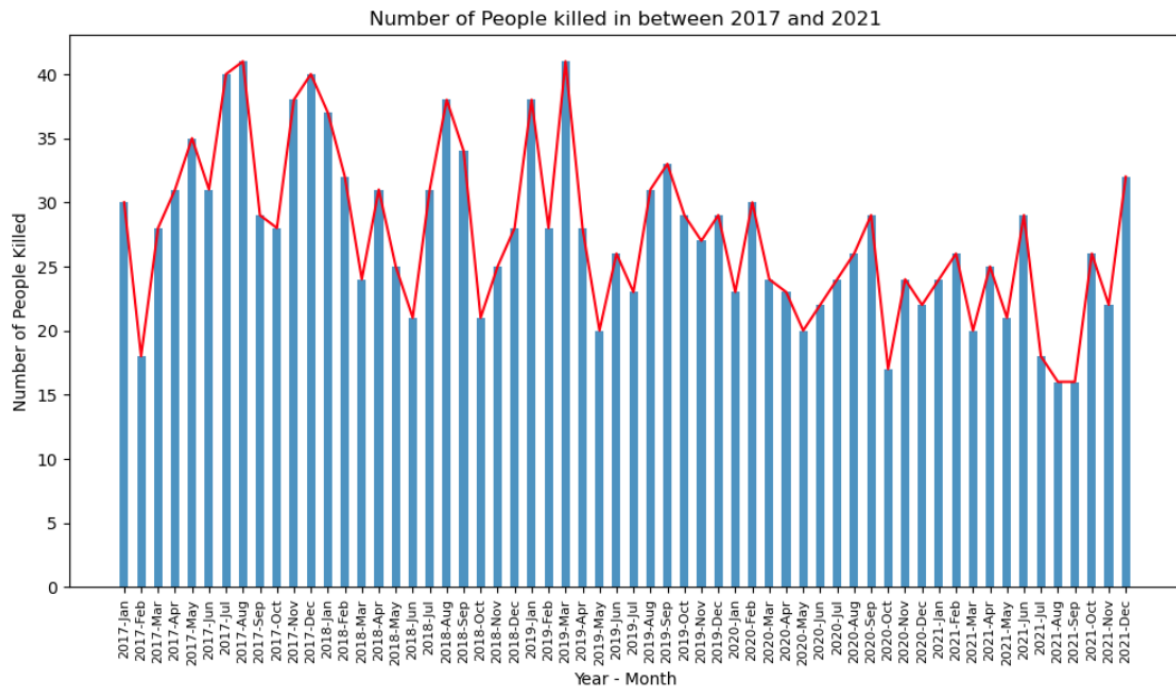
2. Develop a pie chart to display the proportion of fatalities occurring in different weather conditions (e.g., fine, rainy, foggy, etc.), emphasising the most prevalent weather conditions. A possible example representation is provided below.

*Hint: You need to display the percentage on the pie chart.*



3. How does the distribution of the number of people killed on the road vary by month and year? Use a line plot and a bar chart to achieve the output shown below.

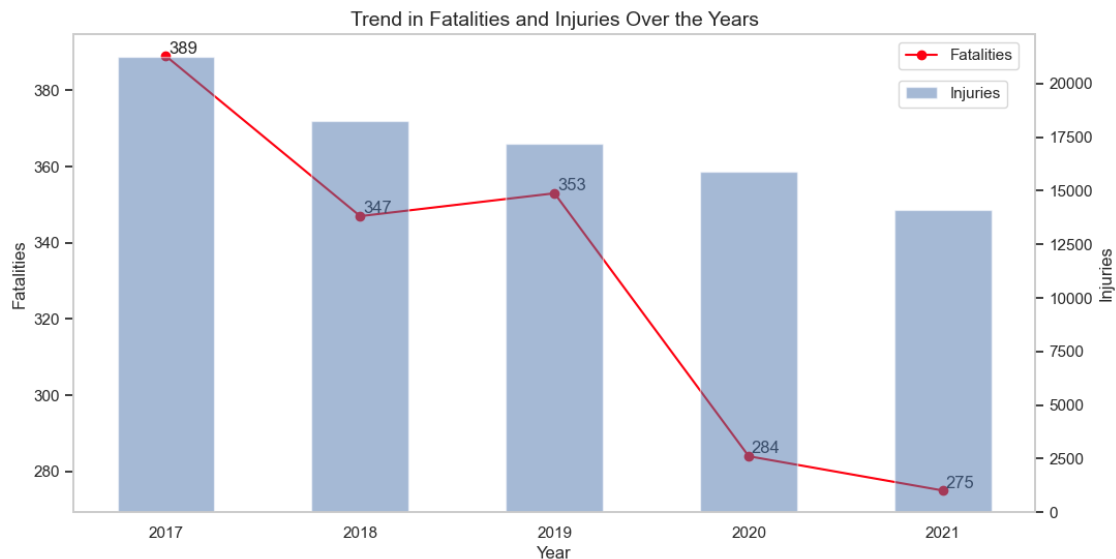
*Hint: You need to sort the year-month in chronological order and only show the abbreviation of months.*



4. Create a line plot and a bar chart showing the trend in the number of fatalities and injuries over the years. Are there any noticeable fluctuations or patterns?

You should aim to match the output provided below.

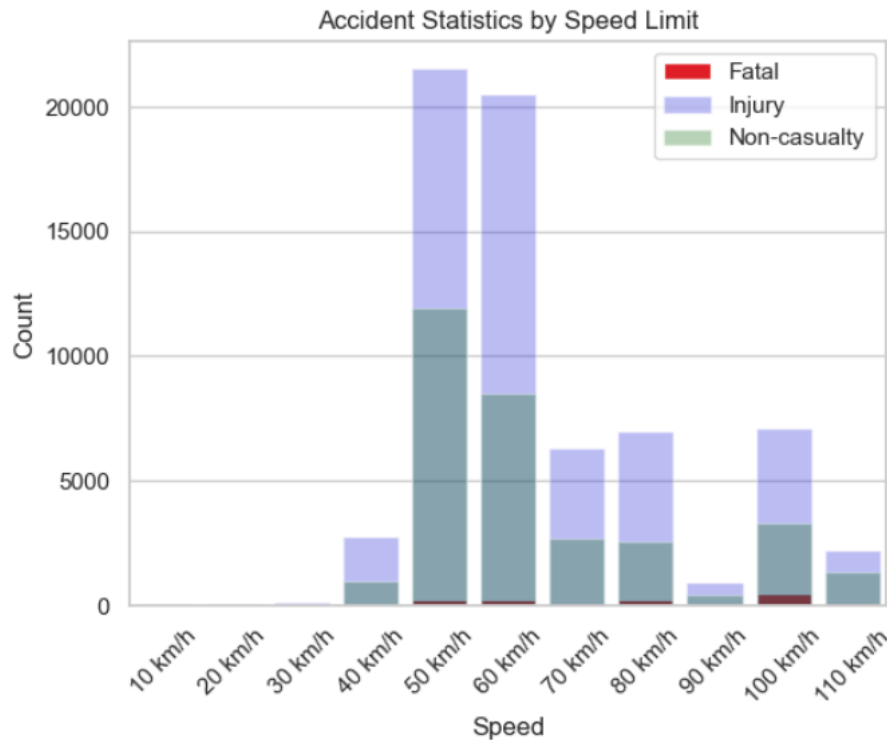
*Hint: You also need to display the yearly fatality number on the line plot.*



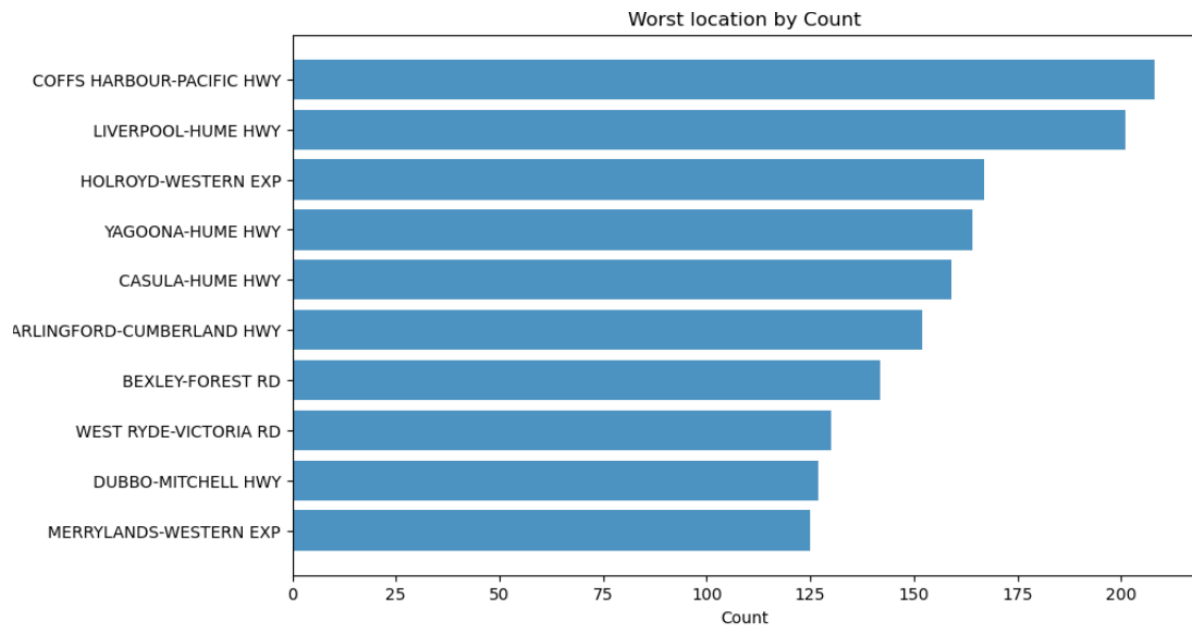
5. What is the relationship between the speed limit and accident statistics?

Utilise a stacked bar chart to depict the figure shown below, using the 'degree\_of\_crash' and 'speed\_limit' columns from the provided Excel spreadsheet.

*Hint: you need to sort the speed in ascending order. Use  $\alpha=0.5$  for readability.*



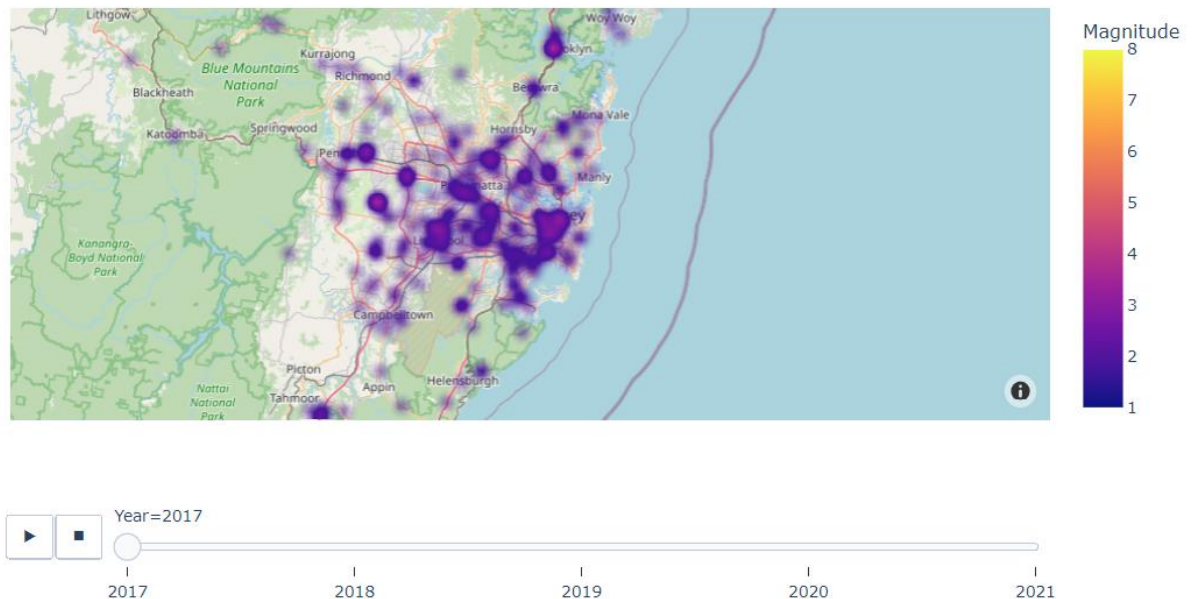
6. What are the most notorious locations in terms of number of crashes recorded over the past five years, use a horizontal bar chart to match the output below.



7. Your client also heard about a popular library called [plotly](#). Plotly can offer many advanced features such as animation and density heatmap. In this question, you have to conduct your own research from these two websites [density heatmap](#) and [animation](#) to complete this task. The magnitude is the sum of 'no\_killed' and 'no\_seriously\_injured' for each crash. Do NOT include



rows where magnitude is 0. You do NOT have to match the exact colour, centre of map, zoom level, or map type.



### 3.1.2. Data Exploration and Predictive Modelling

You need to answer **ALL** the questions below. Each question can have different interpretations, and some of them may have multiple solutions. **We will evaluate not only the correctness of your solutions but also the optimisation of your model.** You also want to make sure you understand the dataset. For instance, when we talk about severity of crashes, any incidents with injuries will be treated as severe.

1. Compare the crash severity distribution for different speed limits (e.g., 30 km/h, 60 km/h, 100 km/h). Does crash severity vary significantly depending on the speed limit? In other words, is it possible to predict the severity of crashes (e.g., whether there is any injury and fatality or just towaway) using a **linear regression** model with speed feature?
2. Explore the relationship between crash severity and the day of the week (2-hour interval) using **linear regression**. Are certain days and time associated with more severe crashes?
3. Can you use **logistic regression** to predict the likelihood of a crash being severe based on features like weather, road alignment, and road classification?
4. Utilise **Decision Tree Classification** to improve the performance of a model predicting severe crash using features like weather and speed limit in Q1.

5. Observe your results from the previous 4 questions, summarise your findings. If you believe other features are relevant to whether a crash is severe or not, include them in a new logistic analysis. If not, explain why.

Hints: Attempt to answer each question by analysing the data. Explore visualisation methods that might assist in addressing each question, and perform any necessary calculations as needed. These visualisations may include, but are not limited to, confusion matrices, ROC curves, and formatted print statements. You may include an appendix, in which you describe how you conducted your analysis so that it can be replicated. If you have made any assumptions in your analysis, state all those assumptions in an appendix to your report.

### 3.2. Requirements for the Report

The written report is worth **25 out of 30 marks** in this group assignment. You need to complete the listed tasks in Section 3.1.1 Data Visualisation and Section 3.1.2 Data Exploration and Predictive Modelling, breakdown of marking criteria is provided in the next section.

- **UNSW Coversheet.** Submit your assignment with a signed coversheet of all group members. (Note: Typed signatures are allowed.) Failure to include the UNSW coversheet with signatures will lead to a penalty of 5% of the awarded marks, and no marks will be released until the coversheet is received.
- **Length.** The total length of the report must not exceed **2,500 words** (excluding the cover page, table of contents, abstract/executive summary, Python code, and appendix). You can stay well below this limit.
- **Table of Contents.** Should not exceed one page, restricted to two levels of headline.
- **Format.** The style/format of the report can be as you find it appropriate and useful. You should use headings, sub-headings, bullet points, diagrams, and tables as appropriate. The file format of the report is **only PDF**.
- **ZIP File.** The ZIP file should contain all Python code of the project. The Python code must be stored in one or more Jupyter Notebooks that must be able to run without any bug/error.
- **References.** References and citations should follow the [Harvard system](#).

- **File Naming.** The files should be “GroupName\_Report”.PDF (e.g., T09A02\_Report.pdf) and “GroupName\_Code”.ZIP (e.g., T09A02\_Code.zip). The group name includes your lab session code. Please refer to Moodle and confirm with your tutor to get your complete correct GroupName.
- **Only one submission per group.** The Group Leader (or a delegated person) will submit the report and the code through Turnitin on Moodle. Failure to comply will lead to a penalty of 5% of the awarded marks.
- Your report should be written in clear and simple language suitable for the board of management.
- Words within figures are not included in the word count.
- Text inserted as pictures will NOT be marked. Only figures can be inserted as pictures.
- Figures and tables should be appropriately labelled.
- Your language should be free of bias, including but not limited to race, gender, sexual orientation or disability.

### **3.2.1. Breakdown of Criteria**

You will be assessed according to the following criteria:

#### **Criteria 1 - Problem definition and Analysis (70%):**

##### ***Problem definition***

How well you have understood the tasks, what problems you are trying to solve, or what questions you are trying to answer.

##### ***Analysis***

You need to transform your problem into analytics. How well you have used analysis and modelling to address the listed tasks. Is your analysis, including data visualisations, appropriate, rigorous, accurate, and comprehensive?

You also need to pay attention to the overall quality of your code. You need to consider things such as whether your code is readable for other people, and whether people can understand your code by reading the comments. A small percentage of the marks will be awarded towards coding style.

### **Criteria 2 - Recommendations (15%):**

How well you have drawn conclusions from your analysis. Are your conclusions appropriate? Are they supported by your analysis?

Your client essentially is interested in how to allocate their campaign budget to different LGA (local government areas) for things such as education program, signs on the road, and random breath test. You need to consider how to use limited funds to achieve effective campaign.

### **Criteria 3 - Written communication (15%):**

How well you have written your report. Is it clear and concise? Is quantitative information presented effectively and professionally?

### **3.3. A 12-minute Presentation**

The presentation/pitch is worth **5 out of 30 marks** in this group assignment. This presentation is your opportunity to engage directly with the Board and present them with the most relevant information. To complete a successful pitch, your team will need to:

- Prepare a pitch deck (max 12 slides, excluding cover page and reference list).
- Present your pitch to the audience, i.e., the board of management.
- Answer questions that raised from the audience.

You will need to submit your pitch deck as a PowerPoint presentation through Turnitin on Moodle. **The file name should be *GroupName\_Pitch.PPT*, e.g., T09A02\_Pitch.ppt, or *GroupName\_Pitch.PPTX*, e.g., T09A02\_Pitch.pptx.**

Requirements for your pitch deck and presentation:

- You must address your proposed research questions and your predictive models.
- You need to transform your problem into analytics.
- Your pitch deck must contain all critical information required to support your recommendations.

- You must assume that some of your audience may NOT have the time to read through your written report and will therefore rely solely on the pitch deck and your pitch to make a decision.
- You can reuse as many or as little content from your report in your pitch deck.
- You have the flexibility to decide what would be the best in this scenario.
- Overall, your pitch deck and the presentation should be concise, logical, and professional.

Any content **beyond 12 slides (for the pitch deck) or beyond the 12-minute mark (for the pitch presentation)** will NOT be assessed. **At least two team members have to present in this pitch, but team members who are not presenting should still help in the preparation.** Note: All team members must be present during the Week 10 tutorial to participate in Q&A.

#### 4. Diaries and Reflection (Optional)

An important part of a project is to record and to evaluate the teamwork. The purpose of the diary is to reflect and learn from the project, in addition to achieving the project's aims in a narrow sense. The project requires that each team member should **keep a personal diary**, and the group should **keep a group diary (diaries might be asked for in cases of disputes)** and that the group includes a critical reflection in their report.

**Individual Diary:** You, as an individual member of a group, are required to keep a diary of your work activities. The individual diary should record your work on the assignment and should note details such as what you did and when you did it. You may also record meeting notes in the diary. The individual diary is NOT to be submitted with the assignment and will NOT be marked. The individual diary may be consulted by the LiC/Lecturer in case of group disputes. Failure to produce an individual diary on request by the LiC/Lecturer will be taken as evidence for a lack of a structured effort in the project.

**Group Diary:** Your group's diary must record all the activities that take place in completing the assignment. The group diary is NOT to be submitted with the assignment and will NOT be marked. The group diary may be consulted by the LiC/Lecturer in case of group disputes. The format of your group's diary is up to you, but **it must record the following details for each group activity:**

- The **Group Leader is responsible for keeping the group diary** unless other arrangements are made within the group.
- Record what the activity (meetings, work) entailed.
- Record location, time, date, and duration of a group activity.
- Record who was present at the activity.
- For "next actions": Specify **who is doing what by when**.
- Signatures of all members, or other suitable forms of demonstrating that all members have seen and approved meeting minutes and other records in the group diary.

## **5. Within-Group Peer Review (Optional)**

**In general, equal contribution of group members is expected.** Significantly unequal contributions should be discussed in due time within the group. In the case that a dispute emerges, the group needs to discuss with the LiC. At the end of the assignment period, you need to perform a within-group peer review to evaluate the contribution of all group members (including yourself) to the group project. **The purpose of the within-group peer review is to critically reflect on the group work, to prevent "free-riding" and to redistribute marks between group members in cases where free-riding or unequal contributions occur.** Claims of unequal contributions, especially if contradicting the evaluation of others, will need to be substantiated with evidence, e.g., group diary, personal diary, meeting notes, emails, Facebook messages, Line messages, WeChat messages, WhatsApp messages, etc. The LiC will make a final judgement in the case that a dispute emerges by calling in the group and/or considering earlier discussions and submitted evidence.

## 6. General Rules

### 6.1. Proper Academic Conduct

All assignments need to follow UNSW's guidelines regarding proper academic conduct. The submission of materials that are non-original or have been submitted elsewhere will be considered plagiarism. **Plagiarism is unacceptable.** All instances of plagiarism or other academic misconduct will be pursued. **Plagiarism may lead to you failing this course and may have negative consequences for your studies at UNSW.** The general UNSW guideline on academic conduct is available online. **For group assignments: If your group suspects that a group member's work contains plagiarism then you should raise this with the group member concerned and have the problem rectified. If the problem is not rectified, notify the LiC who will call in a group meeting.**

### 6.2. Assignment Submission

Assignments are to be submitted via Moodle on, or better before, the due date. Late submission of assignments is not desirable. Late submission of an assignment disrupts the course timelines and is a sign of poor time management and will lead to reduced marks. The late submission of assignments carries **a penalty of 5% of the awarded marks for that assignment per day of lateness** (including weekends and holidays). For example, a 70 marking would be reduced by 3.5 marks per day of lateness.

An extension of time to complete an assignment may be granted by submitting a Special Consideration in the case of illness or misadventure. **For group assignments: Groups are expected to plan ahead and to be able to balance out a missing member WITHOUT an extension. An extension is very unlikely to be granted for group assignments.**

Even if an extension is granted, parts of the marks that are dependent on a timely submission and timely progression of the course (especially marks for participation in the peer review process) cannot be achieved at all. The general UNSW guidelines for special considerations are available online.

### 6.3. Professional Group Work

The membership of groups is at your discretion. It is your responsibility to join a group, otherwise we will assign you to a group. **If you do NOT join a group, you are still expected to complete the assignment in full and no allowance in marking standards is made for the fact that the assignment becomes a solo effort in this case.**

Groups must plan, schedule, and conduct activities in due time. Groups must meet on a regular basis (at least twice per week) while the assignment is being undertaken and keep records (diaries, meeting minutes) of such meetings. The groups must ensure that all members are involved in the completion of the assignment. The work is to be divided equally among the group members.

All group members are expected to behave professionally and work diligently. Group members should contribute in a useful and constructive way to the teamwork.

Deadlines should be kept, and work should be delivered at a professional standard.

**If problems emerge in your group, then these problems should in the first instance openly be discussed in the group (different members might have different views) and resolutions should be agreed on. If internal arrangements repeatedly fail to remedy the situation, then you should bring the issues to the attention of the LiC.** The LiC may call a meeting of the group in which each group member will be asked to describe in details his/her input into the assignment and provide supporting documentation of this effort, e.g., individual diary, group diary, meeting notes, emails, Facebook messages, Line messages, WeChat messages, WhatsApp messages, texts labelled with author names, etc. If group members are found to be making inadequate effort or delivering poor quality, then they will be counselled to improve their effort. If sufficient improvement is not made despite group efforts and LiC interventions, then the mark of under-performing group member(s) may be moderated to reflect the relative lower input into the assignment. Note that the inability to resolve internal group conflicts without involving the LiC does NOT reflect well on the group's project management and teamwork skills.



#### **6.4. Use of AI tools**

As AI applications continue to develop, and technology rapidly progresses around us, we remain committed to our values around academic integrity at UNSW. Your work must be your own and where the use of AI tools, such as ChatGPT, have been permitted by your course convener, they must be properly credited and your submissions must be substantially your own work.

A simple principle to remember for all assessments, regardless of what level of AI-use is permitted, is:

**If you use AI in the writing of your assessment, you should always clearly acknowledge this (e.g., via a footnote).**

For more details about the impact of generative AI tools, please refer to [Generative AI in Learning, Teaching and Assessment at UNSW](#).