FIT5137 S2 2023 Assignment 4: PTV Assignment Scenario (Weight = 30%)

DO NOT SUBMIT THIS DOCUMENT

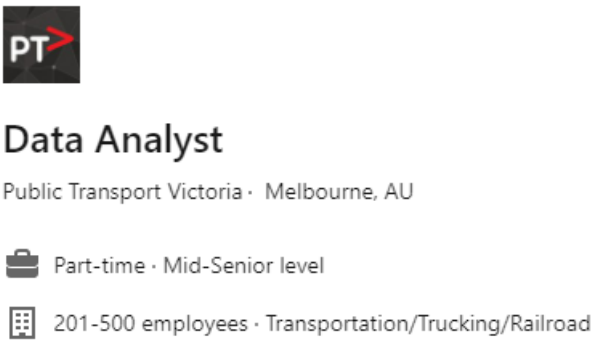
Due date: Wednesday, 25 October 2023, 11:55pm

Version: 1.0 – 16/07/2023

A. General Information and Submission

- This is a group assignment. One group consists of **TWO or THREE** students from the same lab you have enrolled in. You need to register your group composition through the **Assignment 4:PTV Group Self Selection Form[Assessment page, Moodle site]** as soon as possible.
- *Submission method*: Submission is online through Moodle.
- *Penalty for late submission*: 10% deduction for each day.
- *Assignment coversheet*: You will need to sign the assignment coversheet.
- *Contribution form*: **The contribution form needs to be completed by all members and signed (e-signature is acceptable) as an agreement between members.**
- *Assignment FAQ*: There is a Assignment 4 FAQ page set up on the EdStem forum.

B.Assignment Background



You have been hired as a data analyst at Public Transport Victoria (PTV), the Victorian Government authority responsible for public transport in the state. Some of your duties are data extraction, integration and analysis to provide good understanding regarding the public transportation condition in Victoria to the stakeholders.

After the COVID-19 restrictions were lifted, most companies are switching the workstyle from work-from-home to face-to-face. Therefore, transportation infrastructure and network is one of the most important aspects. While some of the people prefer to drive to work, some of other people prefer to use public transportation network as their main transportation mode. PTV as the sole provider for the public transportation network reduced their services during the lockdown period. Now, PTV has restored the services to cover as many areas as possible in the whole regions. However, some questions remained mysteries. How good is the current PTV coverage? Is there any uncovered spots? Which area has the best public transportation options?

Therefore, as a data analyst, your task is to evaluate the data and provide the spatial data analysis to the stakeholders of PTV. The data should be presented in an area level, such as municipality, suburbs or postcode. For example, you may present “The number of bus services in Bundoora” or "The number of Trains or Trams network in Bundoora”.

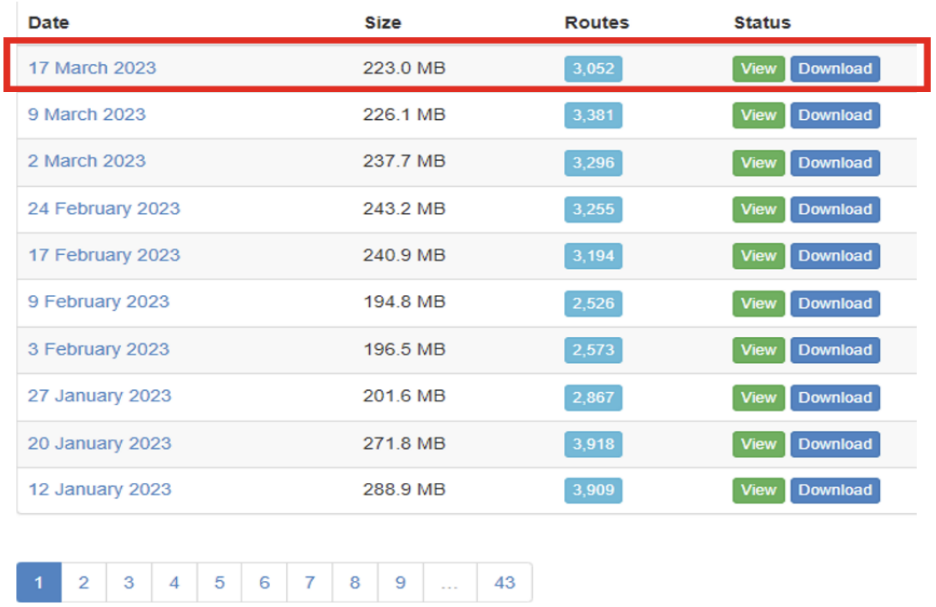
Data

There are two datasets that you have to obtain in this assignment, which are the PTV/GTFS dataset and Australian Boundary data.

The **General Transit Feed Specification (GTFS)** is a data specification that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications. Today, the GTFS data format is used by thousands of public transport providers.

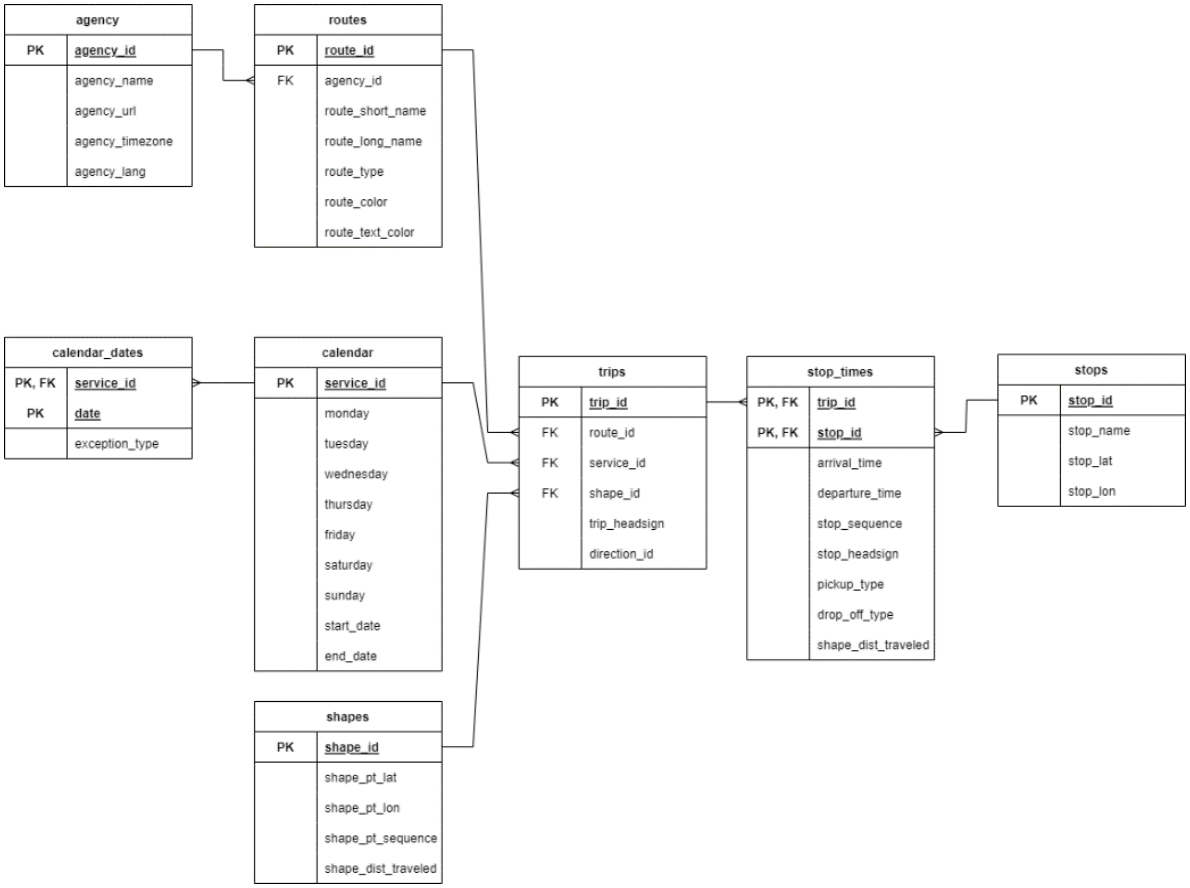
GTFS is split into a schedule component that contains schedule, fare, and geographic transit information and a real-time component that contains arrival predictions, vehicle positions and service advisories. A GTFS feed is composed of a series of text files collected in a ZIP file. Each file models a particular aspect of transit information: stops, routes, trips, and other schedule data.

For more detailed information about GTFS, you can refer to the official documentation provided by Google at https://developers.google.com/transit/gtfs. Additionally, You can read further explanation about the PTV-GTFS data from https://transitfeeds.com/p/ptv/497. For this assignment, we will be using the 17th March 2023 version of the dataset.



Date	Size	Routes	Status
17 March 2023	223.0 MB	3,052	View Download
9 March 2023	226.1 MB	3,381	View Download
2 March 2023	237.7 MB	3,296	View Download
24 February 2023	243.2 MB	3,255	View Download
17 February 2023	240.9 MB	3,194	View Download
9 February 2023	194.8 MB	2,526	View Download
3 February 2023	196.5 MB	2,573	View Download
27 January 2023	201.6 MB	2,867	View Download
20 January 2023	271.8 MB	3,918	View Download
12 January 2023	288.9 MB	3,909	View Download

The GTFS data structure is shown below:

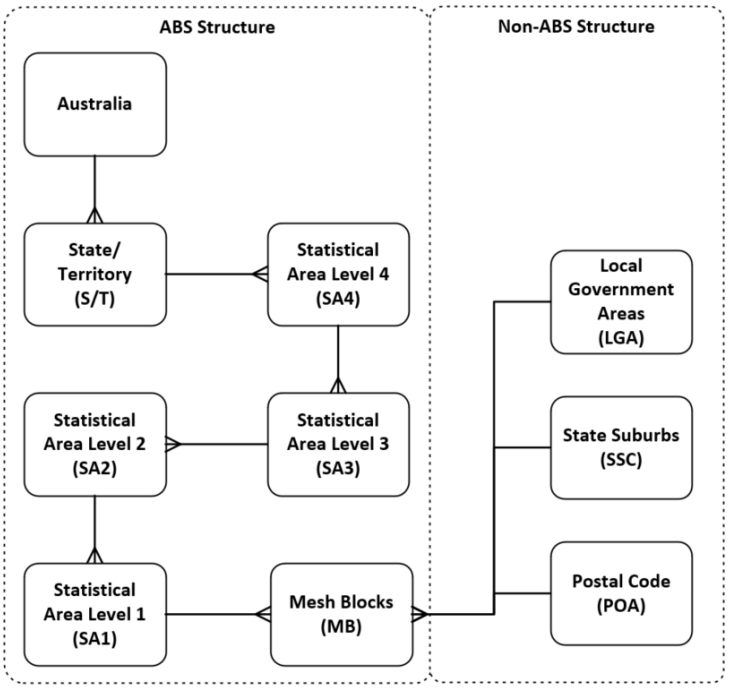


The **Australian digital boundary** is defined by the Australian Bureau of Statistics using the **Australian Statistical Geography Standard (ASGS)**. The ASGS is a classification of Australia into a hierarchy of statistical areas. It is a social geography, developed to reflect the location of people and communities. It is used for the publication and analysis of official statistics and other data. The ASGS is updated every 5 years to account for growth and change in Australia's population, economy and infrastructure. For the 2021 release, the ASGS will be re-named to the Australian Statistical Geography Standard (ASGS) Edition 3.

The ASGS is split into two parts, the ABS and Non ABS Structures. The **ABS Structures** are geographies that the ABS designs specifically for the release and analysis of statistics. This means that the statistical areas are designed to meet the requirements of statistical collections as well as geographic concepts relevant to those statistics. This helps to ensure the confidentiality, accuracy and relevance of ABS data.The **Non ABS Structures** generally represent administrative regions which are not defined or maintained by the ABS, but for which the ABS is committed to directly providing a range of statistics.

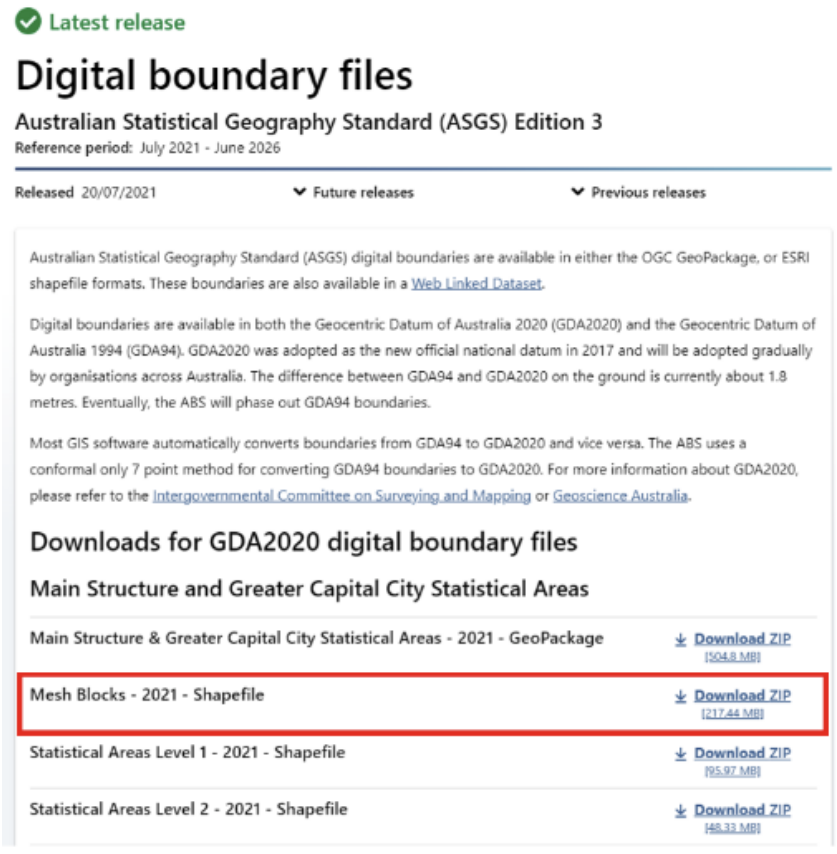
The Main Structure is developed by the ABS and is used to release and analyse a broad range of social, demographic and economic statistics. It is a nested hierarchy of geographies, and each level directly aggregates to the next level. **Mesh Blocks (MBs) are the smallest geographic areas defined by the ABS and form the building blocks for the larger regions of the ASGS. Most Mesh Blocks contain 30 to 60 dwellings.**

Below is the simplified ABS and Non ABS Structure. You can read further explanation about the structure here https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026#overview



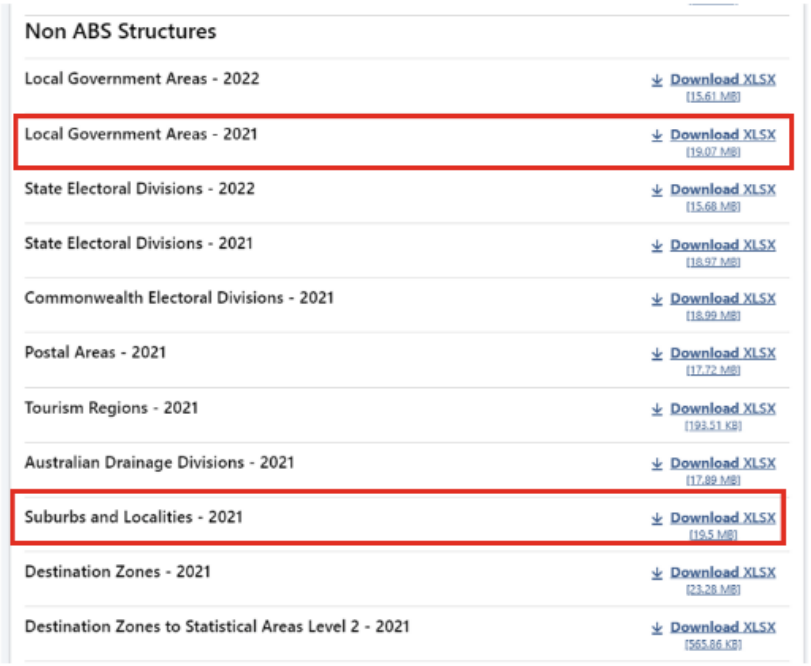
The Digital boundary files that you have to get is the Mesh Blocks dataset. The Mesh Blocks dataset is available as Shape file. You can read further explanation about the Mesh Blocks dataset here

https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/digital-boundary-files



Allocation files are non-spatial representations of how each geography is aggregated from their building block geography. You can also read further explanation about the Allocation files dataset here

https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/access-and-downloads/allocation-files



Assignment Task list

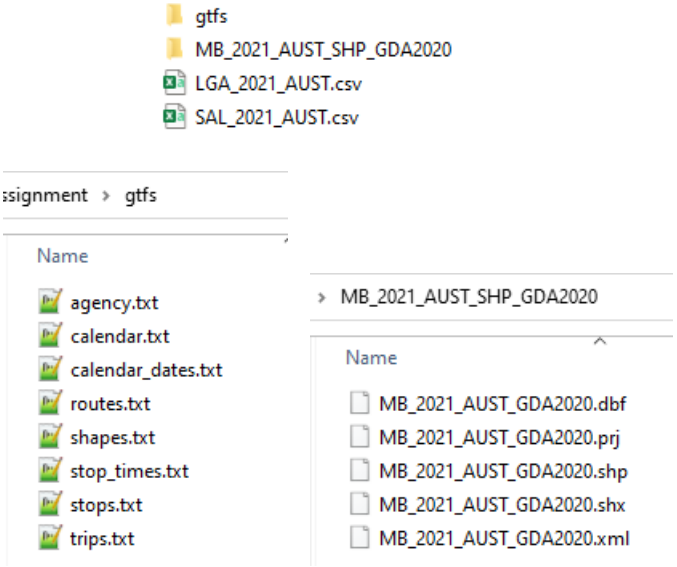
Your assignment consists of several parts. Always read the instruction one by one. Do not move to the step without completing the previous step:

- Task 1: Data Restoration - Restore the data to the database. Monitor the success indicator to ensure successful restoration of the data.
- Task 2: Data Preprocessing - Perform necessary structure maintenance and create result tables for further processing.
- Task 3: Data Analytics - Develop SQL queries to analyze the data and evaluate performance.
- Task 4: Data Visualization - Create visualizations to present the results of the data analytics.

For simplicity, all the data required for this assignment is readily available in the PostGIS Docker container. You can access these datasets within the container by navigating to the /data/adata folder.If you don't know how to do it, refer to the labs 10 activities.

```
root@db94d38b7162:/home/student# ls /data/adata
gtfs  LGA_2021_AUST.csv  MB_2021_AUST_SHP_GDA2020  SAL_2021_AUST.csv
```

Verify your data before the restoration process.



As a data analyst, it is your responsibility to understand and explore these publicly available data.

Important note:
The question and answer box is available in the FIT5137_A4_PTV_Answer sheet[Assessment page, Moodle site]. Please carefully follow the provided instructions to answer each question.

- **When submitting your assignment, ensure that the FIT5137_A4_PTV_Answer sheet is in PDF format.**
- **There is no need to submit this assignment scenario file.**

C. Submission Checklist

1. One **combined** .**pdf file** containing all tasks mentioned above:
    - □ Cover page
    - □ A signed coversheet
    - □ A Group Contract & contribution declaration form:

Note:
Each student must state the parts of the assignment that they completed. An example is as follows:

Note:
- If you group shared workload equally, the contribution percentage will be 50% or 33%. Otherise, the contribution percentage must end in 0, e.g, 80%, 20% [85% is not acceptable]
- The example is based on a Group of 2 scenario, and the Contribution Declaration template can also be found on the Assignment 4 FAQ page, Ed forum.

Example:

Percentage of contribution:

1. Name: Adam, ID: 210008, Contribution: 60%
2. Name: Ben, ID: 230933, Contribution: 40%

List of parts that each student completed:
1. Adam: list the parts that Adam did
2. Ben: list the parts that Ben did

- □ Assingment 4 Answer Sheet

Note:
Ensure that each answer box on the Assingment 4 Answer Sheet contains a well-organised and clear answer.
- The SQL script or commend should be properly structured and,
- The screenshot must be coherent, concise and aligned with the query.

2. Consolidate all SQL scripts from tasks 1 to 4 into a single .sql file for submission. Verify the provided file's executability to guarantee its functionality.

Use this table as a guidance to make sure that you don’t miss anything in this assignment

Task #	Activity	FIT5137_A4_PTV_Answer sheet [PDF]		SQL script
		Commend/SQL script	Screenshot	
Task 1:Data Restoration	Creating ptv schema	Yes	No	Yes
	Restore GTFS dataset in ptv schema	Yes	No	Yes
	Restore Mesh Blocks in ptv schema(commend)	Yes	No	No
	Restore LGA2021 in ptv schema	Yes	No	Yes
	Restore Suburb2021 in ptv schema	Yes	No	Yes
	Data Verification	No	Yes	Yes
Task 2:Data Preprocessing	Mesh Blocks filtering	Yes	Yes	Yes
	Melbourne Boundary creation	Yes	Yes	Yes
	Adding geometry column	Yes	Yes	Yes
	Denormalise GTFS table	Yes	Yes	Yes
Task 3:Data Analytics (Note: Question 3.2.1 required results only)	Suburbs Accessibility	Yes	Yes	Yes
	LGA Blankspot	Yes	Yes	Yes
Task 4:Data Visualisation	Creating lga_blankspot table	Yes	Yes	Yes
	Heatmap	No	Yes	No

D. Submission Method

1. **Upload the PDF file from Checklist #1 and the SQL file from Checklist #2 to Moodle by the due date:Wednesday, 25 October 2023, 11:55pm.**
    - The submission of this assignment must be in the form of **a single PDF file AND a single .sql file**. No other forms will be accepted.
    - One member of your group can upload the submission. However, **please note that all group members must click the submit button and accept the submission statement** (failure to do so will mean your assignment will not be submitted and will incur late penalties).
    - **It's important to note that our support hours are limited, and we don't have the capacity to address submission issues outside of working hours. You must ensure that you have all the files listed in this checklist before submitting your assignment to Moodle. Failure to submit a complete list of files will result in a mark penalty.**s.

2. Penalty for late submission: 10% deduction for each day, including weekends

3. Submission cut-off time: **Wednesday, 1 Novemeber 2023, 11:55 pm**. The submission link will be unavailable after this time.

E. Late Penalty

Late assignments submitted without an approved extension may be accepted up to a maximum of **seven days** with the approval of the Chief Examiner and/or Lecturer but will be **penalised at the rate of 10% per day (including weekends and public holidays)**. Assignments submitted more than seven days after the due date will receive a zero mark for that assignment and may **not receive any feedback**.

***Please note(late penalty and extension)*:**

1. An inability to manage your time or computing resources will not be accepted as a valid excuse. (Several assignments being due at the same time are a fact of university life.)

2. Group issues, hardware failures, whether of personal or university equipment, are not normally recognised as valid excuses. Failure to back up assignment files is also not recognised as a valid excuse.

F. Authorship

This assignment is an **group assignment** and the final submission must be identifiably your own group work. Breaches of this requirement will result in an assignment not being accepted for assessment and may result in disciplinary action.

As per the University's policy on the guidelines and practice pertaining to the usage of Generative AI, this assignment restricts all use of generative AI. In this assessment, **you must not use generative artificial intelligence (AI) to generate any materials or content in relation to the assessment task.**

G. Special Consideration

From this semester onwards, students will no longer seek extensions from the Chief Examiner/Unit Teaching Team. All extensions / special considerations will now be handled by the central Spec Con team. **Please do not email teaching staff to request an extension or special consideration.**

Extensions and other individual alterations to the assessment regime will only be considered using the University Special Consideration Policy. Students should carefully read the Special Consideration website, especially the details about what formal documentation is required.

All special consideration requests should be made using the Special Consideration Application.

Please do not assume that submission of a Special Consideration application guarantees that it will be granted – you must receive an official confirmation that it has been granted.

H. Getting help and support

What can you get help for?

- ***Consultations with the Teaching Team***

Talk to the Teaching Team: https://lms.monash.edu/course/view.php?id=162086&section=2

- ***English language skills***

Talk to English Connect: https://www.monash.edu/english-connect

- ***Study skills***

Talk to a learning skills advisor: https://www.monash.edu/library/skills/contacts

- ***Counselling***

Talk to a counsellor: https://www.monash.edu/health/counselling/appointments

I. Plagiarism and Collusion:

Monash University is committed to upholding standards and academic integrity and honesty. Please take the time to view these links.

Academic Integrity Module

Student Academic Integrity Policy

Test your knowledge, collusion (FIT No Collusion Module)

All the best for your Assignment!