FIT5137 S2 2024 Assignment 3: PTV Assignment Scenario (Weight = 35%)
Due date: Friday, 25 October 2024, 4:30 PM

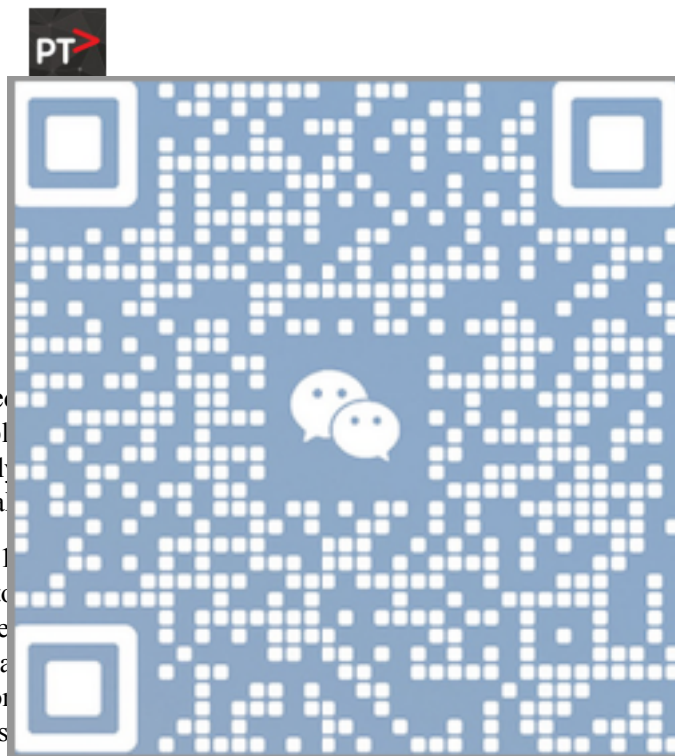# General Information and Submission

- This is an *individual assignment*.
- *Submission method*: Submission is online through Moodle.
- *Penalty for late submission*: 5% deduction for each day.
- *Assignment FAQ*: There is an Assignment 3 FAQ page set up on the EdStem forum.

Assignment Background

You have been hired [...] Victoria Government authority responsibl[...] are data extraction, integration and anal[...] nsportation condition in Victoria to the sta[...]

After the COVID-1[...] the workstyle from work-from-home to[...] network is one of the most important aspe[...] people prefer to use the public transporta[...] e sole provider for the public transportation[...] riod. Now, PTV has restored the services[...] ever, some questions remained mysteries. How good is the current PTV coverage? Are 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".
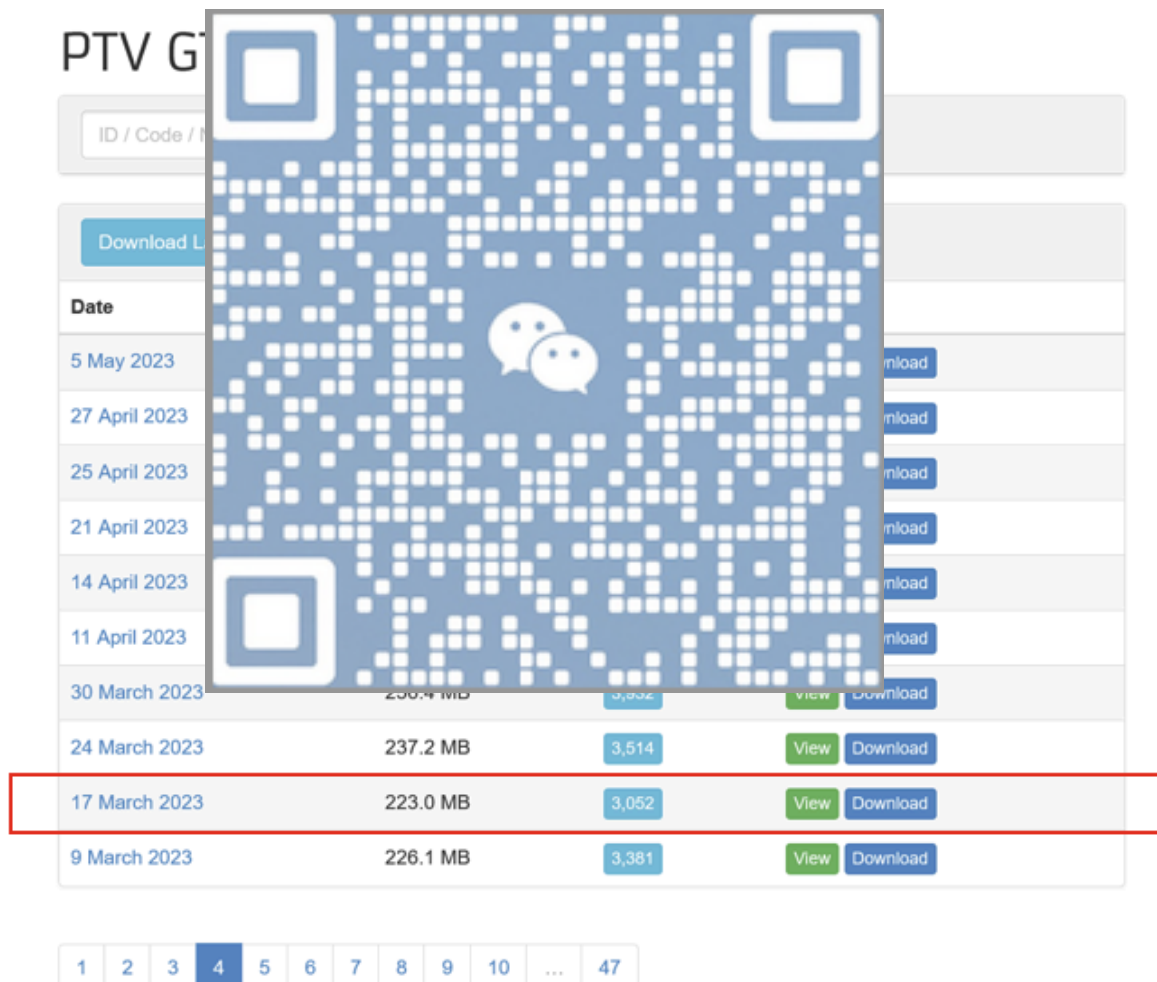
Data

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.**

The GTFS data structure is shown below:

agency

PK	agency_id
	agency_name
	agency_url
	agency_timezone
	agency_lang

routes

PK	route_id
FK	agency_id
	route_short_name
	route_long_name
	route_type
	route_color
	route_text_color

calendar_dates

PK, FK	service_id
PK	date
	exception_type

calendar

PK	service_id
	monday
	tuesday
	wednesday
	thursday
	friday
	saturday
	sunday

trips

PK	trip_id
FK	route_id
FK	service_id
FK	shape_id
	trip_headsign
	direction_id

stop_times

PK, FK	trip_id
PK, FK	stop_id
	arrival_time
	departure_time
	stop_sequence
	stop_headsign
	pickup_type

stops

PK	stop_id
	stop_name
	stop_lat
	stop_lon

The **Australian di**   Statistics using the
Australian Statisti   fication of Australia
into a hierarchy of   eflect the location of
people and commu   al statistics and other
data. The ASGS i   ange in Australia's
population, econom   l be re-named to the
Australian Statistica

The ASGS is split   **ABS Structures** are
geographies that th   is 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
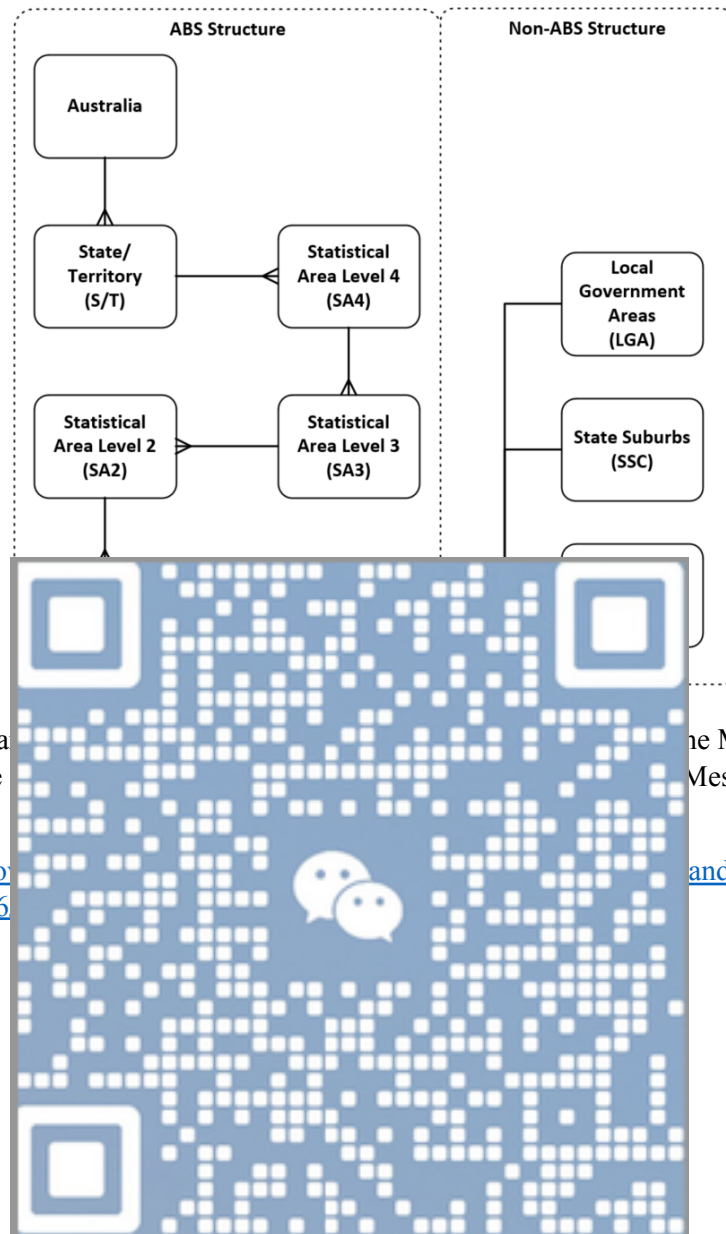
https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026#overview

The Digital bounda... ...he Mesh Blocks dataset is available ... ...Mesh Blocks dataset here

https://www.abs.gov... ...andard-asgs-edition-3/jul2021-jun2026...

✅ Latest release

Digital boundary files

Australian Statistical Geography Standard (ASGS) Edition 3
Reference period: July 2021 - June 2026

Released 20/07/2021   ⌄ Future releases   ⌄ Previous releases

Australian Statistical Geography Standard (ASGS) digital boundaries are available in either the OGC GeoPackage, or ESRI shapefile formats. These boundaries are also available in a Web Linked Dataset.

Digital boundaries are available in both the Geocentric Datum of Australia 2020 (GDA2020) and the Geocentric Datum of Australia 1994 (GDA94). GDA2020 was adopted as the new official national datum in 2017 and will be adopted gradually by organisations across Australia. The difference between GDA94 and GDA2020 on the ground is currently about 1.8 metres. Eventually, the ABS will phase out GDA94 boundaries.

Most GIS software automatically converts boundaries from GDA94 to GDA2020 and vice versa. The ABS uses a conformal only 7 point method for converting GDA94 boundaries to GDA2020. For more information about GDA2020, please refer to the Intergovernmental Committee on Surveying and Mapping or Geoscience Australia.

Downloads for GDA2020 digital boundary files

Main Structure and Greater Capital City Statistical Areas

Main Structure & Greater Capital City Statistical Areas - 2021 - GeoPackage   ↓ Download ZIP

Mesh   d ZIP

Statist   d ZIP

Statist   d ZIP

Allocation files are ... regated from their building block geog... location files dataset here

https://www.abs.gov... andard-asgs-edition-3/jul2021-jun2026...

Nor...

Loca... XLSX

Loca... XLSX

State... XLSX

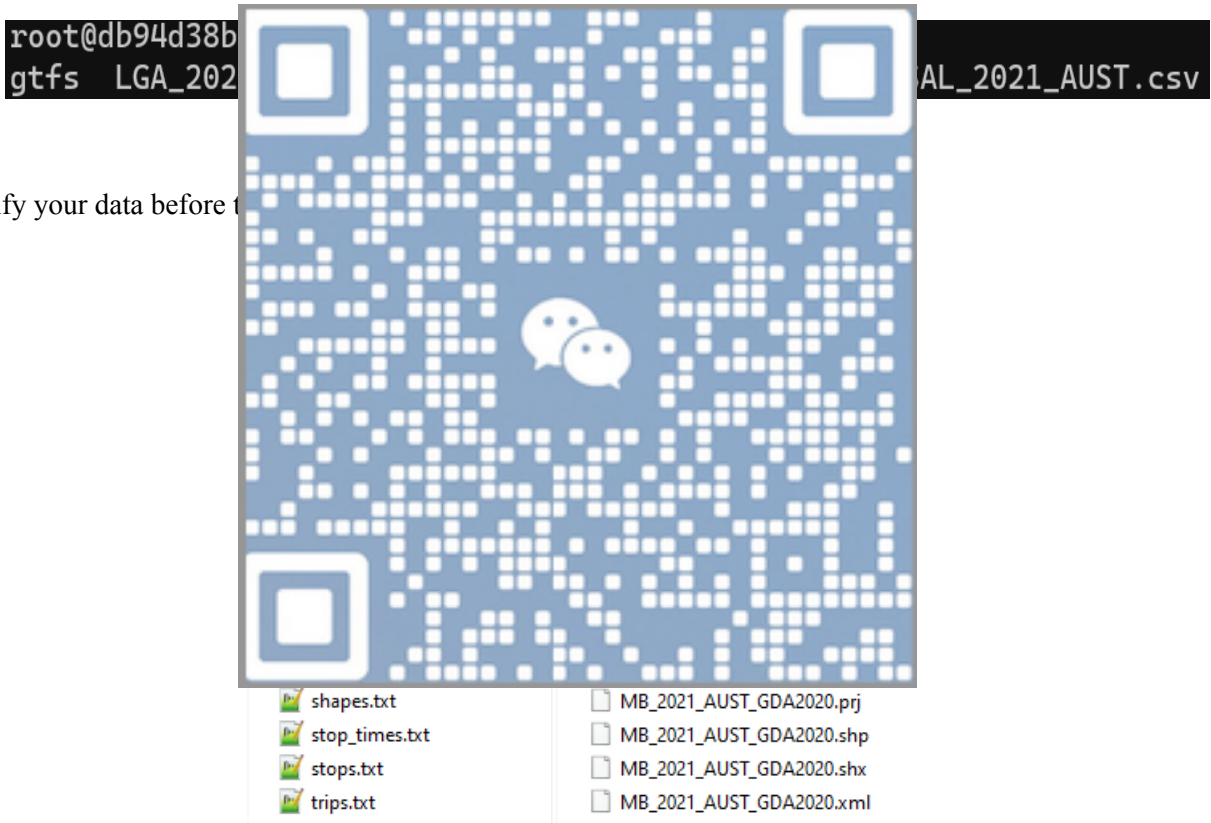|	
State Electoral Divisions - 2021	↓ Download XLSX [18.97 MB]
Commonwealth Electoral Divisions - 2021	↓ Download XLSX [18.99 MB]
Postal Areas - 2021	↓ Download XLSX [17.72 MB]
Tourism Regions - 2021	↓ Download XLSX [193.51 KB]
Australian Drainage Divisions - 2021	↓ Download XLSX [17.89 MB]
Suburbs and Localities - 2021	↓ Download XLSX [19.5 MB]
Destination Zones - 2021	↓ Download XLSX [23.28 MB]
Destination Zones to Statistical Areas Level 2 - 2021	↓ Download XLSX [565.86 KB]

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 and Visualization - Develop SQL queries to analyze the data and evaluate performance & Create visualizations to present the results of the data analytics.
- **No data cleaning required for this assignment.**
- **For more information, see the FAQ for Assignment 3.**

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@db94d38b   ...   AL_2021_AUST.csv
gtfs LGA_202   ...

Verify your data before t

shapes.txt	MB_2021_AUST_GDA2020.prj
stop_times.txt	MB_2021_AUST_GDA2020.shp
stops.txt	MB_2021_AUST_GDA2020.shx
trips.txt	MB_2021_AUST_GDA2020.xml

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

Assignment Task

Task 1: Data Restoration

Before you can start the data analytic processes, the first thing you have to do is to restore the external data to your database. Make sure you prepare a destination schema to restore your data. The destination schema for your assignment is "**ptv**".

Note:

- Before initiating the data restoration process, **it is essential to thoroughly explore the dataset**. This exploration involves identifying appropriate data types, determining field lengths, and making other relevant considerations that will inform the creation of the table structure.

- **Ensure that you restore the data into the PTV schema using regular (local) tables.** ~~[illegible]~~ **directly within the Postgre** ~~[illegible]~~

- **Ensure** ~~[illegible]~~ **bles from GTFS and 3 ta** ~~[illegible]~~ **vely.**

The outputs of ~~[illegible]~~

a) Attach a scr ~~[illegible]~~ in Task 1, including th ~~[illegible]~~ owing code:

```
with tbl as
(select table_schema, T
 from information_sch
 where table_schema i
select table_schema, T
(xpath('/row/c/text()', e                                        chema,
TABLE_NAME), FAL
from tbl
order by table_name;
```

Task 2: Data Preprocessing for Melbourne Metropolitan area

The purpose of this section is to manipulate the data into a suitable format for the following task analysis. This task has two parts: Mandatory requirement and Optional requirements.

Mandatory requirement
[You must meet the mandatory requirements described in this section.]

In this assignment, we aim to explore the **transportation accessibility [Topic of report]** of the **Melbourne Metropolitan area exclusively [Scope of report]**. The *mb_2021* table

contains mesh blocks for the entire country of Australia. To minimise query costs, ensure that you <u>only use the mesh blocks within the Melbourne Metropolitan area for this assignment</u>. The Melbourne Metropolitan's mesh blocks can be identified from the gcc_name21. If the column contains "Greater Melbourne", this mesh block is located in Melbourne Metropolitan. As a result, you need to create a table called "mb2021_mel" that contains ONLY the mesh blocks in Melbourne Metropolitan.

	sa1_code21	sa2_code21	sa2_name21	sa3_code21	sa3_name21	sa4_code21	sa4_name21	gcc_code21	gcc_name21	ste_code21	ste_name21	
22	10901117322	109011173	Albury - North	10901	Albury	109	Murray	1RNSW	Rest of NSW	1	New South Wales	AU
23	10901117322	109011173	Albury - North	10901	Albury	109	Murray	1RNSW	Rest of NSW	1	New South Wales	AU
24	21401137143	214011371	Frankston	21401	Frankston	214	Mornington Peninsula	2GMEL	Greater Melbourne	2	Victoria	AU
25	10901117325	109011173	Albury - North	10901	Albury	109	Murray	1RNSW	Rest of NSW	1	New South Wales	AU
26	10901117301	109011173	Albury - North	10901	Albury	109	Murray	1RNSW	Rest of NSW	1	New South Wales	AU
27	10901117323	109011173	Albury - North	10901	Albury	109	Murray	1RNSW	Rest of NSW	1	New South Wales	AU

Optional requirements:

[You are free to explore and manipulate the data creatively within the mandatory requirements, which are limited to Melbourne Metropolitan for the topic of transport accessibility.]

For ***optional re***   nalysis needs. Make sure to include a   optional requirements you choose.

> **Question:**
> Do I have to ar
>
> **Answer:**
> No, you are f   ong as the data is analysed in Me   ty.

The following   lysis transportation accessibility of l

1.Since the wor   nt to have a polygon for the boundary   olygon to create one large polygon fo

2.*Stops* table does not have any geometry column. It might be useful to add a geometry column, using the latitude and longitude values available in the table. Make sure you use GDA2020 (SRID:7844) for this column.

3.The *Stops* table does not show direct information regarding the vehicle types, routes_short_name and routes_long_name. These informations are stored in the *routes* table.

4.If you want to explore the transportation situation for different vehicle types, such as tram, train, or bus, the vehicle type is determined by the corresponding route type in the routes table, where:

- 0 corresponds to tram
- 2 corresponds to train
- 3 corresponds to bus

- Any other route type is labelled as 'Unknown'.

The outputs of this task for **Report** are:

b) Attach a screenshot of SQL script for creating a table named "**mb2021_mel**" that contains ONLY the mesh blocks in Melbourne Metropolitan.

c) Provide a detailed explanation of the **remaining** data processing steps you have conducted, including screenshots of the SQL scripts and the rationale behind your choices in the report.

Task 3: Data Analytics and Visualisation

In this section you will need to perform data analysis on the tables you have restored, focusing on transport accessibility in metropolitan Melbourne. Use the techniques you have learned in the spatial database part to carry out your analysis. You are free to choose any specific perspectives or aspects of data analysis relevant to your dataset, but ensure that your analysis relates to the **main topic: transport accessibility in metropolitan Melbourne.**

This could include exploring different statistical measures or carrying out other relevant analyses. Presen understanding of the dataset and high

As part of this ast one map-based headmap using nain topics. These visualisations w summary report. To support your ana rectly in the report.

Be sure to incl visualisation in the appendix of you how the analysis was performed and the data. This will ensure that your include appropriate labels, titles, ar to understand. The visualisations sh onvey your analysis findings.

Note:

- Use SQ
- Conduct a thorough descriptive analysis to uncover insights within the data.
- Summarise and Visualise your findings clearly and concisely.
- Highlight key observations and patterns discovered during the analysis.
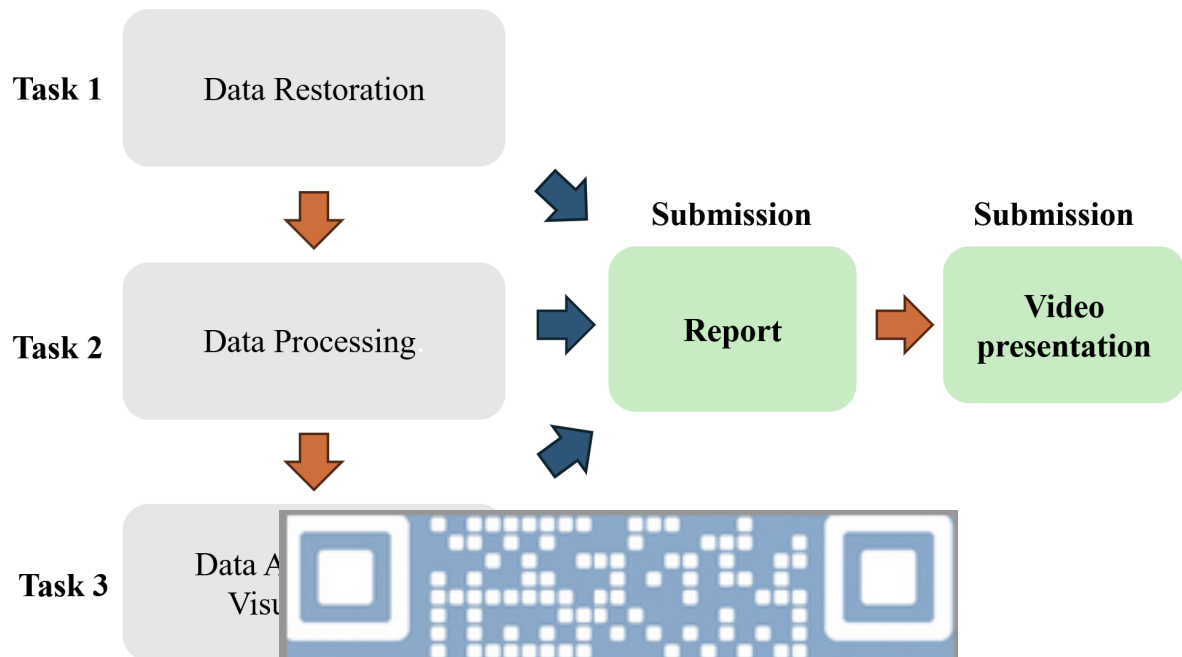- Ensure your findings reflect a deep understanding of the data.

The outputs of this task for **Report** are:

d) Data analysis and visualisation, including the screenshot of SQL script and visualisation. For the visualisation, it must contain at least one map base figure.

Submission Checklist

Assignment Flowchart

Task 1 — Data Restoration

Task 2 — Data Processing

Task 3 — Data A... Visu...

Submission — **Report**

Submission — **Video presentation**

Summary Repo...

As a professiona... ...tasks, including data restoration, proc... ...e written report. The report should ad... ...ed format, consisting of an introduct... **Please note that a question-and-a... ...and marks will be deducted for us...**

Please ensure th... ...l-organised, concise, and coherent.Th...

Title: Write your title here on a separate page, (Note: Abstract is not required)

1. Introduction, such as

- Briefly explain the purpose of the report and what you aim to achieve with your analysis.
- Highlight the key questions you want to investigate through your analysis.

2. Methodology

This section should provide a clear explanation of the different stages of your work.

- **Dataset Overview, such as**

Provide an overview of the data and its source.

- **Data Restoration and Preprocessing, such as**

  Explain how you imported and initially explored the data. Include the software and libraries used.

  Provide a detailed explanation of the **remaining** data processing steps you have conducted, and the rationale behind your choices in the report.

- **Data Analysis and Visualization**

  Describe which area of transport accessibility in metropolitan Melbourne you are primarily exploring.

  Describe the analysis you conducted and the types of visualisations you chose to use, and why you felt they would effectively represent your data and findings. What software or libraries were used to create these visualisations?

3. Results:

Presen[...]these results mean and ho[...]

For the[...].

4. Discussion

Discuss your fi[...]

- Restate[...]yses.
- Discus[...]otheses.
- Reflec[...]ed during your analysi[...]

5. References

If you have u[...]y according to the chosen style g[...]

6. Appendix [Excluded from the 2000-word limit]

The screenshots of the following tasks:
- Attach a screenshot of the results to include all the tables you restored in Task 1, including the number of rows for each table you restored.

- Attach screenshots of the SQL scripts used in Task 2, including the SQL scripts for creating a table called "mb2021_mel" and screenshots of the SQL scripts you used for the remaining data processing steps.
- The SQL script for Data analysis and visualisation
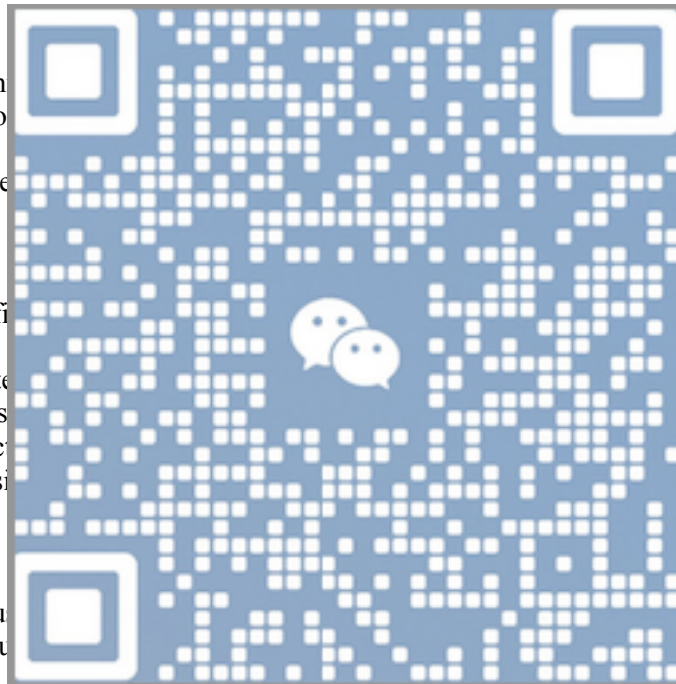
Video presentation

A five minute video presentation in mp4 format save as: YourstudentID_A3_video.mp4

Based on the report you have created, present your design and findings in a _five-minute video presentation_. Ensure you thoroughly understand both the dataset and the report to effectively extract and communicate the key points.

Assignment Submission

1. A combined pdf file save as: YourstudentID_A3_report.pdf, containing all of the above tasks 1 to 3.
2. A five minute video presentation in mp4 format save as: YourstudentID

Zip all above files fro   tudentID.zip.

- The su   a single ZIP file. Only P   . No other formats will be
- You m   s checklist before submitt   a complete list of files wi
- It's imp   d we don't have the capacit   hours. You must ensure   re submitting your assignn   iles will result in a mark pe
- Penalty   uding weekends
- Submis   . Submissions will not be accepted after this time unless there are special considerations.

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

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 5% 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. 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.
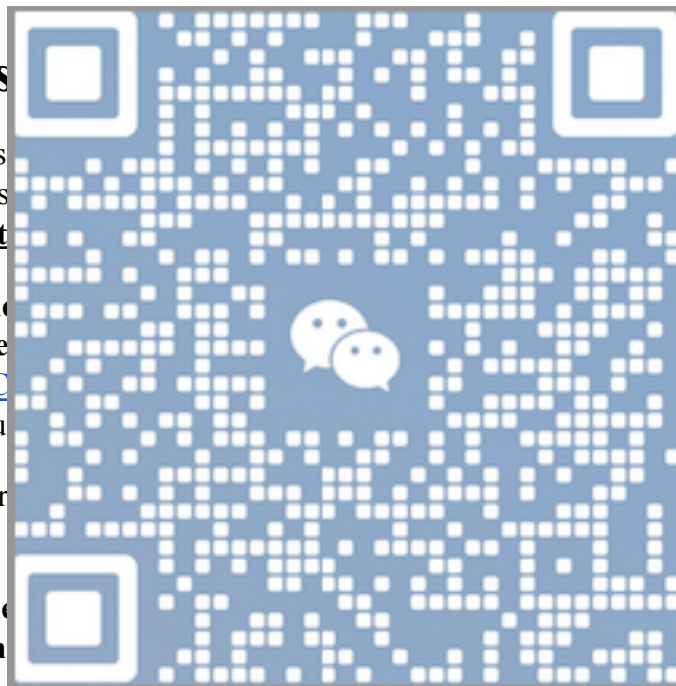
Special Cons

Students no longer s . All extensions / special considerations m. **Please do not email teaching staff t**

Extensions and othe **ime will only be considered using the** ts should carefully read the Special C out what formal documentation is requ

All special consider cial Consideration Application.

Please do not assume **cation guarantees that it will be gran** **that it has been granted.**

Getting help and support

What can you get help for?

- ***Consultations with the Teaching Team***
Talk to the Teaching Team:
https://learning.monash.edu/course/view.php?id=19675&section=5
- ***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

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!