MONASH University

FIT5137 Assignment 4- S2 2023

PTV

STUDENT NAME & STUDENT ID:
Wanru Xiang 33729220;
Linhao Wang 31273327;
Ziqi Pei 33429472 ;

FIT5137

Assignment 4 Group Contract & Contribtution form

Lab No and time: Tuesday 6:00-8:00pm
Tutor: David Daniel Cheng Zarate

Name:Wanru Xiang
Email address:wxia0021@student.monash.edu

Name:Linhao Wang
Email address:lwan0191@student.monash.edu

Name:Ziqi Pei
Email address: zpei0003@student.monash.edu

As a member of the team, I understand that:

- I will contribute to ***each*** of the tasks within Assignment 3;
- I will attend and contribute at all agreed team meetings;
- I will respond in a timely manner (within 24 hours) to my fellow team members when they make contact;
- I will make every effort to resolve any issues that arise within the group, and raise, if necessary, any problems with my tutor ***before*** the due date;
- If I do not participate adequately, the tutor will be informed and will take appropriate action;
- I understand that part of the grade for this assignment will involve peer review where my partner will grade me on my participation and quality of contribution;
- **My mark for this assignment will reflect the quality of my work and my participation within the team.**

Signed: Wanru Xiang **Name:** Wanru Xiang **Date:** 27/10/2023

Signed: Linhao Wang **Name:** Linhao Wang **Date:** 27/10/2023

Signed: Ziqi Pei **Name:** Ziqi Pei **Date:** 27/10/2023

Contribution Declaration Form
(to be completed by all team members)

Please fill in the form with the contribution from each student towards the assignment.

Note: A sample contribution declaration form is available on the Ed Forum site.

1 NAME AND CONTRIBUTION DETAILS

Student ID	Student Name	Contribution Percentage
33729220	Wanru Xiang	40%
31273327	Linhao Wang	30%
33429472	Ziqi Pei	30%

Please list the tasks you have done in this table		
Student ID:33729220	Student ID:31273327	Student ID:33429472

Task 1 Data Restoration Creating ptv schema.20% Restore GTFS dataset in ptv schema.20% Restore LGA2021 in ptv schema.25% Restore Suburb2021 in ptv shcema.20% Restore Suburb 2021 in ptv shcema .20% Task2 Data Preprocessing Mesh Blocks filiteing. 30% Melbourene Boundary creation. 20% Denormalise GTFS. 20% Task3 Data Analytics Suburbs Accessibility 70% LGA Blankspot. 100% Task4 Data Visualisation Creating iga_blankspot table.70% Heatmap.50%	Task 1 Data Restoration Creating ptv schema.20% Restore GTFS dataset in ptv schema.20% Restore LGA2021 in ptv schema.25% Restore Suburb2021 in ptv shcema.20% Restore Suburb 2021 in ptv shcema .20% Task2 Data Preprocessing Mesh Blocks filiteing. 50% Melbourene Boundary creation. 60% Denormalise GTFS. 60% Task4 Data Visualisation Creating iga_blankspot table.30% Heatmap.50%	Task1 Data Restoration Creating ptv schema.60% Restore GTFS dataset in ptv schema.60% Restore LGA2021 in ptv schema.50% Restore Suburb2021 in ptv shcema.60% Restore Suburb 2021 in ptv shcema .60% Task 2 Data Preprocessing Mesh Blocks filiteing. 20% Melbourene Boundary creation. 20% Denormalise GTFS. 20% Task3 Data Analytics Suburbs Accessibility 30%

2 DECLARATION

We declare that:

- The information we have supplied in or with this form is complete and correct.
- We understand that the information we have provided in this form will be used for individual assessment of the assignment.
- The contribution percentage cannot be changed once you submit.

3 SIGNATURE

Linhao Wang	Wanru Xiang

Signatures

Ziqi Pei

Date 27/10/2023

GROUP ASSIGNMENT COVER SHEET

Student ID Number	Surname	Given Names
33729220	Xiang	Wanru
31273327	Wang	Linhao
33429472	Pei	Ziqi

***** Please include the names of all other group members.

Unit name and code	FIT5137 Advanced Database Technology	
Title of assignment	FIT5137 Assignment 4 - S2 2023	
Lecturer/tutor	David Daniel Cheng Zarate	
Tutorial day and time	Tuesday 6:00-8:00pm	**Campus** Clayton
Is this an authorised group assignment? Yes No ✓		
Has any part of this assignment been previously submitted as part of another unit/course? Yes No ✓		
Due Date 27 October 2023, 11:55pm		**Date submitted** 27 October 2023

All work must be submitted by the due date. If an extension of work is granted this must be specified with the signature of the lecturer/tutor.

Extension granted until (date) ..... Signature of lecturer/tutor .....

Please note that it is your responsibility to retain copies of your assessments.

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means taking and using another person's ideas or manner of expressing them and passing them off as one's own. For example, by failing to give appropriate acknowledgement. The material used can be from any source (staff, students or the internet, published and unpublished works).

Collusion: Collusion means unauthorised collaboration with another person on assessable written, oral or practical work and includes paying another person to complete all or part of the work.

Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or delegate, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity Policy and Procedures.
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations http://adm.monash.edu/legal/legislation/statutes
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.

- I acknowledge and agree that the assessor of this assignment may for the purposes of assessment, reproduce the assignment and:
  - i. provide to another member of faculty and any external marker; and/or
  - ii. submit it to a text matching software; and/or
  - iii. submit it to a text matching software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.
- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration when preparing this assignment.

Signature .Wanru Xiang Linhao Wang Ziqi Pei..... Date…27 Oct, 2023……

* delete (iii) if not applicable

Signature __ Wanru Xiang _____Date: ___ 27 Oct, 2023_____

Signature ___ Linhao Wang _____Date: _ 27 Oct, 2023____

Signature ____ Ziqi Pei_____Date: ____27 Oct, 2023_____

Privacy Statement

The information on this form is collected for the primary purpose of assessing your assignment and ensuring the academic integrity requirements of the University are met. Other purposes of collection include recording your plagiarism and collusion declaration, attending to course and administrative matters and statistical analyses. If you choose not to complete all the questions on this form it may not be possible for Monash University to assess your assignment. You have a right to access personal information that Monash University holds about you, subject to any exceptions in relevant legislation. If you wish to seek access to your personal information or inquire about the handling of your personal information, please contact the University Privacy Officer: privacyofficer@adm.monash.edu.au

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

PLEASE SUBMIT ANSWER SHEET IN PDF FORMAT

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

Version: 2.0 – 25/09/2023

Assignment Task list

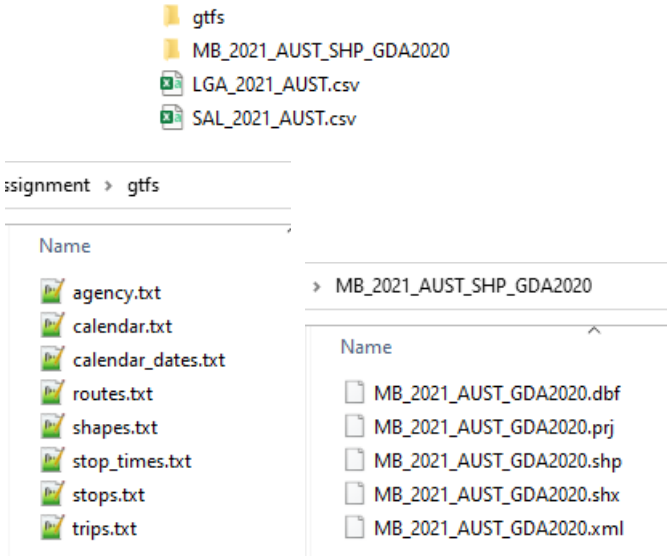
Your assignment consists of several parts. Always read the instruction one by one. Do not move to the step without completing the previous step:

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

```
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.

Do not edit or remove any content on the answer sheet, including the questions. Please write your answers in the answer boxes provided. If necessary, you can adjust the size of the answer box to fit your answer.

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. Do not utilize foreign tables, as the data must be stored directly within the PostgreSQL database – updated 27/09/2023
- Make sure all 8 GTFS tables are restored successfully
- Index or constraints can be added to the table after the data has been restored completely (Note: This index or constraint requirement is **NOT** mandatory in this Task 1 – updated 25/09/2023)
- **No data cleaning required for this assignment**
- For more information, see the FAQ for Assignment 4.

1.1 PTV schema

Write the SQL script to create the destination schema named "**ptv**".

```
create schema ptv;
```

1.2 GTFS

Write the SQL script to restore **ALL** tables in GTFS files.

```
create table ptv.agencies(
agency_id int primary key,
agency_name varchar(255),
agency_url varchar(255),
agency_timezone varchar(255),
agency_lang char(2)
);

copy ptv.agencies from '/data/adata/gtfs/agency.txt' delimiter ',' csv header;

create table ptv.routes(
route_id varchar(15) primary key,
```

```
agency_id int not null,
route_short_name varchar(30),
route_long_name varchar(255) not null,
route_type int not null,
route_color varchar(6) not null,
route_text_color varchar(6) not null
);

copy ptv.routes from '/data/adata/gtfs/routes.txt' delimiter ',' csv header;

create table ptv.calendar(
service_id varchar(20) primary key,
monday boolean not null,
tuesday boolean not null,
wednesday boolean not null,
thursday boolean not null,
friday boolean not null,
saturday boolean not null,
sunday boolean not null,
start_date DATE not null,
end_date DATE not null
);

copy ptv.calendar from '/data/adata/gtfs/calendar.txt' delimiter ',' csv header;

create table ptv.calendar_dates(
service_id varchar(20) not null,
date DATE not null,
exception_type int not null,
primary key (service_id,
```

```
date),
foreign key (service_id) references ptv.calendar(service_id)
);

copy ptv.calendar_dates from '/data/adata/gtfs/calendar_dates.txt' delimiter ',' csv header;

create table ptv.shapes(
shape_id varchar(20) not null,
shape_pt_lat decimal(16,
13) not null,
shape_pt_lon decimal(16,
13) not null,
shape_pt_sequence int not null,
shape_dist_traveled decimal(10,
2) not null,
primary key (shape_id,
shape_pt_sequence)
);

copy ptv.shapes from '/data/adata/gtfs/shapes.txt' delimiter ',' csv header;

create table ptv.trips(
route_id varchar(15) not null,
service_id varchar(20) not null,
trip_id varchar(30) not null primary key,
shape_id varchar(20),
trip_headsign varchar(255),
direction_id int not null,
foreign key (route_id) references ptv.routes(route_id),
```

```
foreign key (service_id) references ptv.calendar(service_id)
);

copy ptv.trips from '/data/adata/gtfs/trips.txt' delimiter ',' csv header;


create table ptv.stops(
stop_id int,
stop_name varchar(255),
stop_lat decimal(16, 13),
stop_lon decimal(16, 13)
);

copy ptv.stops from '/data/adata/gtfs/stops.txt' delimiter ',' csv header;


create table ptv.stop_times(
trip_id varchar(30) not null,
arrival_time char(8) not null,
departure_time char(8) not null,
stop_id int not null,
stop_sequence int not null,
stop_headsign varchar(255),
pickup_type int not null,
drop_off_type int not null,
shape_dist_traveled varchar(10),
primary key (trip_id,
stop_sequence)
);

copy ptv.stop_times from '/data/adata/gtfs/stop_times.txt' delimiter ',' null as '' csv header
quote '"';
```

1.3 ABS Mesh Blocks

Scripts to restore the Mesh Blocks files by using correct dataset file. Restore the file using ogr2ogr into table "**mb2021**"

```
ogr2ogr PG:"dbname=gisdb user=postgres"
"/data/adata/MB_2021_AUST_SHP_GDA2020/MB_2021_AUST_GDA2020.shp" -nln
ptv.mb2021 -overwrite -nlt MULTIPOLYGON
```

1.4 ABS Allocation Files

Write the SQL script to restore the LGA2021 Allocation file.

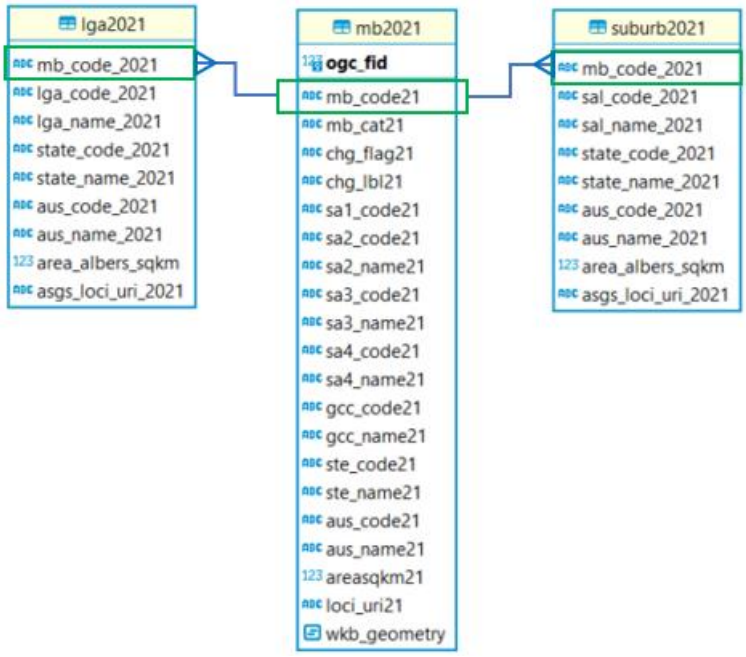
```
create table ptv.lga2021 (
mb_code_2021 char(11) primary key,
lga_code_2021 char(5) not null,
lga_name_2021 varchar(50) not null,
state_code_2021 char(1) not null,
state_name_2021 varchar(50) not null,
aus_code_2021 varchar(10) not null,
aus_name_2021 varchar(50) not null,
area_albers_sqkm decimal(10,
4),
asgs_loci_uri_2021 varchar(255) not null
);

copy ptv.lga2021 from '/data/adata/LGA_2021_AUST.csv' delimiter ',' csv header;
```

Write the SQL script to restore the SAL 2021 Allocation file for suburb2021.

```
create table ptv.suburb2021 (
mb_code_2021 char(11) primary key,
sal_code_2021 char(5) not null,
sal_name_2021 varchar(50) not null,
```

```
state_code_2021 char(1) not null,
state_name_2021 varchar(50) not null,
aus_code_2021 varchar(10) not null,
aus_name_2021 varchar(50) not null,
area_albers_sqkm decimal(10,
4),
asgs_loci_uri_2021 varchar(255) not null
);

copy ptv.suburb2021 from '/data/adata/SAL_2021_AUST.csv' delimiter ',' csv header;
```

The allocation tables have 1-N relationship with the mb2021 in mb_code21 – mb_code_2021. Although there are no PK – FK defined in the table, the relationship rule still apply.



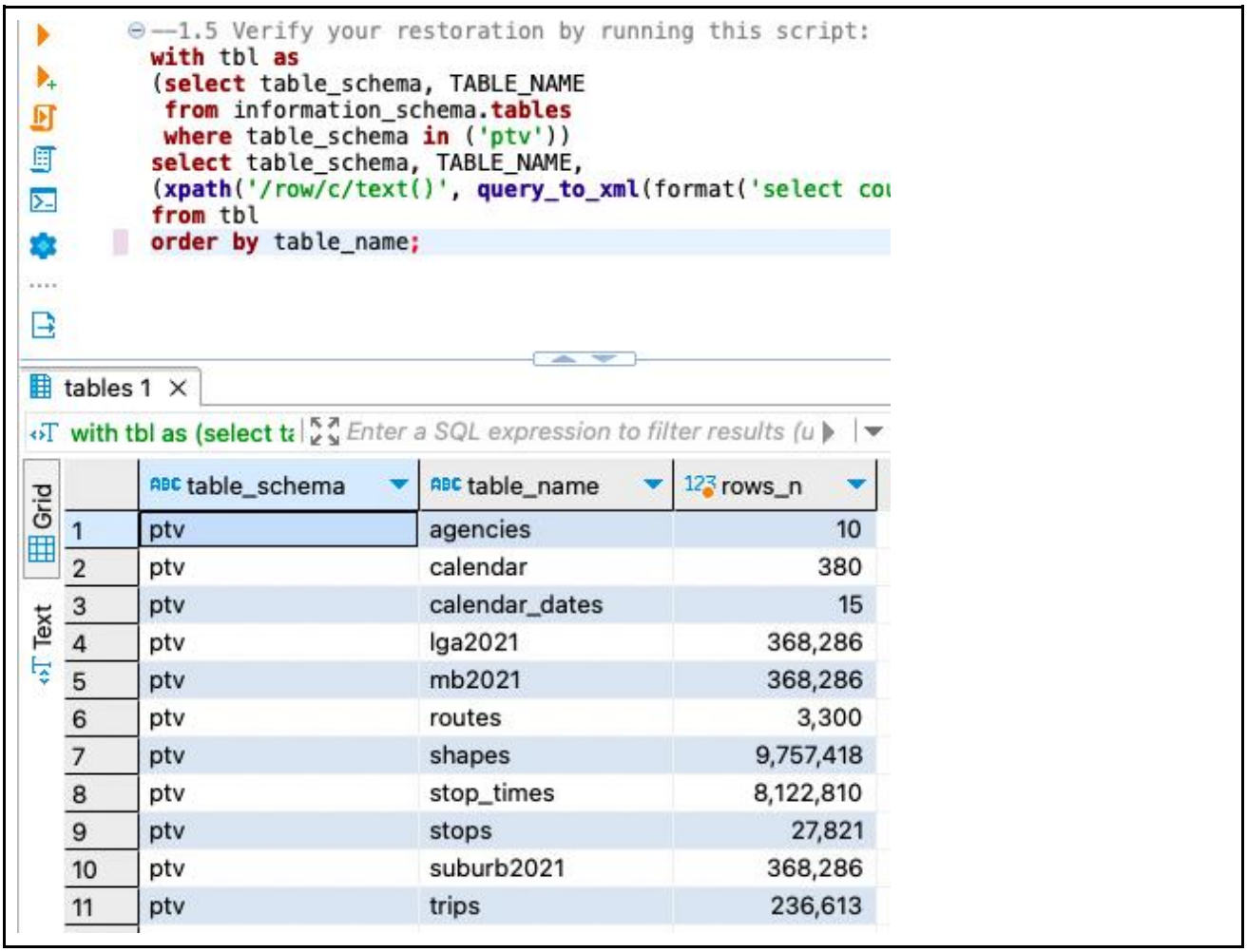
1.5 Data Verification

Verify your restoration by running this script. 1.Do not modify the verification script. 2.Make sure that the table name is consistent with the table we provided.

Output: Attach a screenshot of the results to include all tables you have restored in Task 1.

```
with tbl as
(select table_schema, TABLE_NAME
 from information_schema.tables
 where table_schema in ('ptv'))
select table_schema, TABLE_NAME,
(xpath('/row/c/text()', query_to_xml(format('select count(*) as c from %I.%I', table_schema, TABLE_NAME),
FALSE, TRUE, ''))[1]::text::int AS rows_n
from tbl
order by table_name;
```

Screenshot:



```
--1.5 Verify your restoration by running this script:
with tbl as
(select table_schema, TABLE_NAME
 from information_schema.tables
 where table_schema in ('ptv'))
select table_schema, TABLE_NAME,
(xpath('/row/c/text()', query_to_xml(format('select cou
from tbl
order by table_name;
```

tables 1

with tbl as (select ta | Enter a SQL expression to filter results (u

Grid | Text

	table_schema	table_name	rows_n
1	ptv	agencies	10
2	ptv	calendar	380
3	ptv	calendar_dates	15
4	ptv	lga2021	368,286
5	ptv	mb2021	368,286
6	ptv	routes	3,300
7	ptv	shapes	9,757,418
8	ptv	stop_times	8,122,810
9	ptv	stops	27,821
10	ptv	suburb2021	368,286
11	ptv	trips	236,613

Task 2: Data Preprocessing

2.1 Filter Melbourne Metropolitan area

The mb2021 table contains whole mesh blocks in Australia. To minimise the query cost, we want to ensure that you only use the mesh blocks in Melbourne Metropolitan. 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.

Create a table named "**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

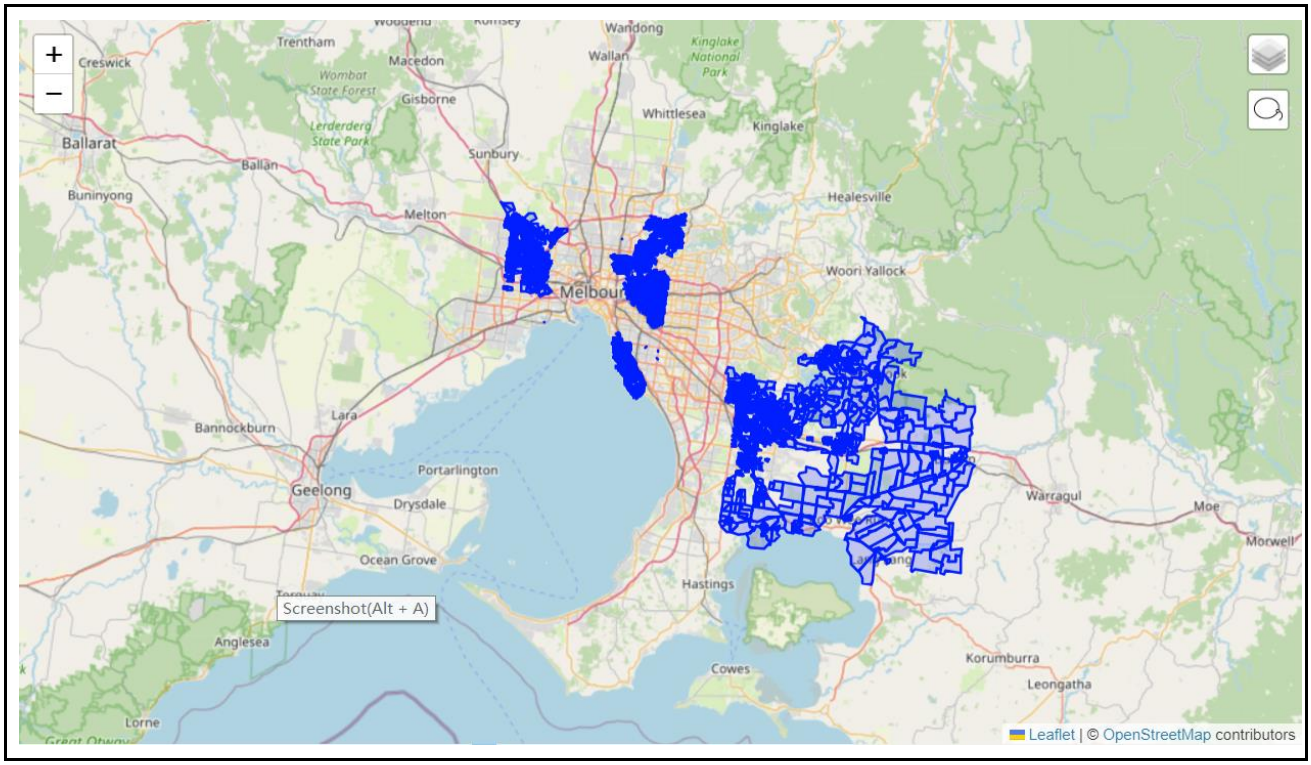
Write the SQL script to do this.

```
CREATE TABLE ptv.mb2021_mel AS SELECT * FROM ptv.mb2021 WHERE
gcc_name21 ILIKE '%Greater Melbourne%';


SELECT * FROM ptv.mb2021_mel;


select count(*) from ptv.mb2021_mel;
```

Attach a screenshot of the Spatial Map results.

Screenshot:

	123 count
1	59,483



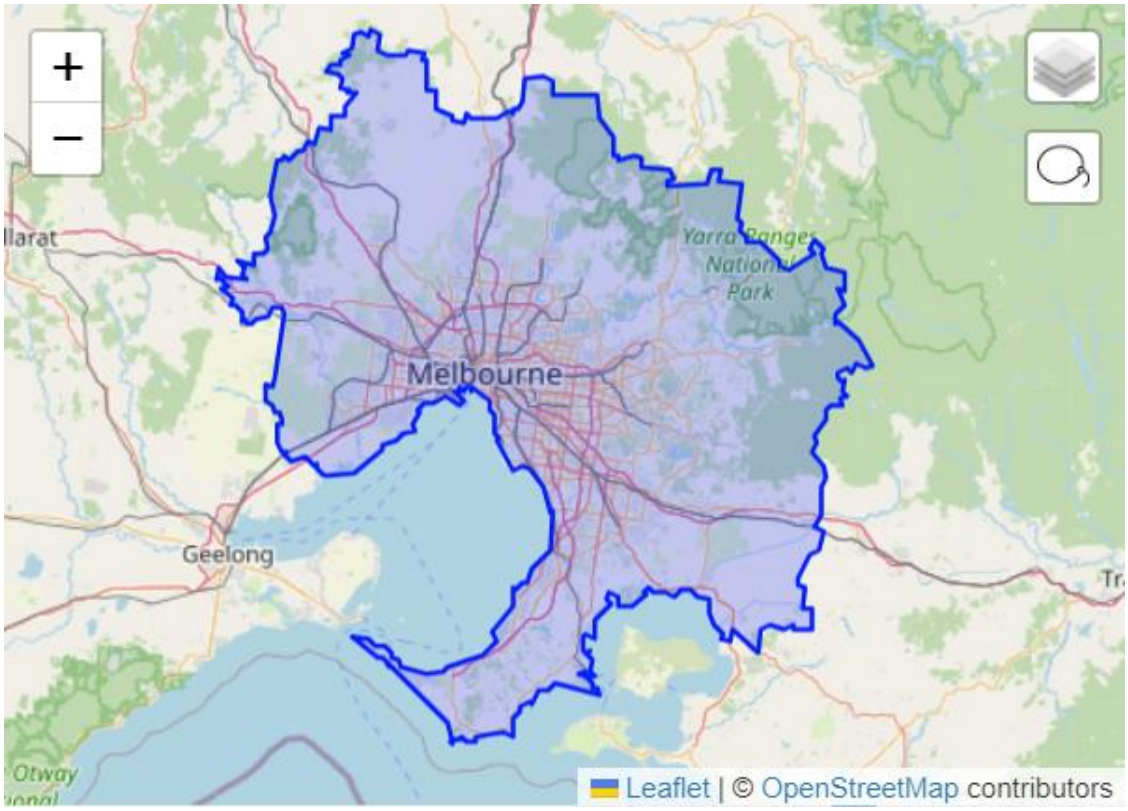
2.2 Melbourne Metropolitan Boundary

Since the working area will be Melbourne Metropolitan, it is important to have a polygon for the boundary of our working area. Create a table, named "**melbourne**" for Melbourne Metropolitan boundary. Hint: aggregate all mesh blocks polygon to create one large polygon for Melbourne Metropolitan boundary.

Write the SQL script to do this.

```
CREATE TABLE ptv.melbourne AS
SELECT ST_UNION(wkb_geometry) AS geom
FROM ptv.mb2021_mel;


select * from ptv.melbourne;
```

Attach a screenshot of the Spatial Map results.



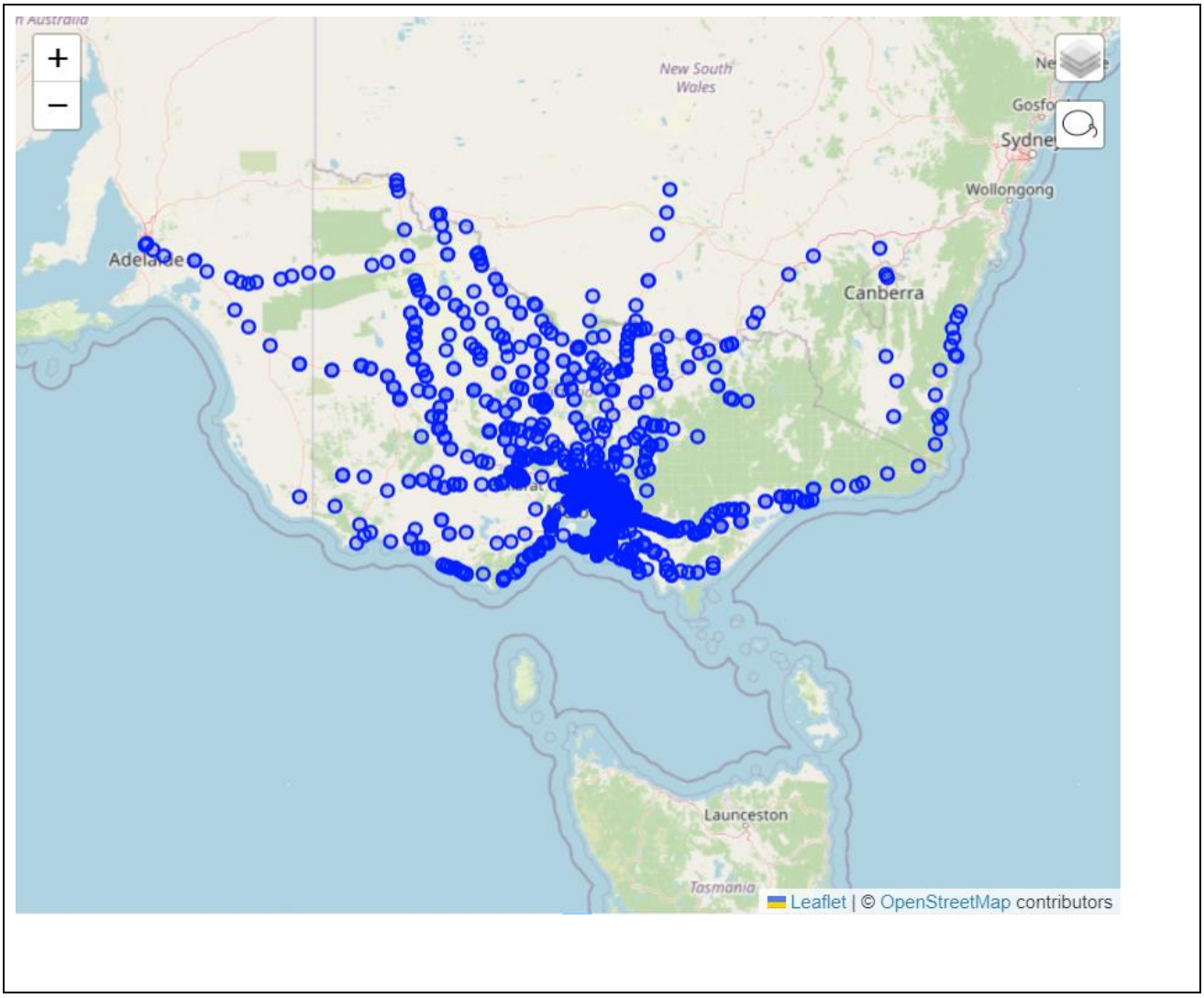
2.3 Add Geometry column to Stops table

Stops table does not have any geometry column. Add a geometry column by using the latitude and longitude value that are available in the table. Make sure you use GDA2020 (SRID:7844) for this column.

Write the SQL script to do this.

```
ALTER TABLE ptv.stops
ADD COLUMN geom geometry(Point, 7844);



UPDATE ptv.stops
SET geom = ST_SetSRID(ST_MakePoint(stop_lon, stop_lat), 7844);
```

Attach a screenshot of the Spatial Map results.



2.4 Denormalise GTFS structure

The ptv.stops table does not show direct information regarding the vehicle types, routes_short_name and routes_long_name. These information are stored in the routes table.

Create a table called "stops_routes_mel" to encompass the following attributes: stop_id, stop_name, coordinates, route number (derived from routes_short_name), route name (derived from routes_long_name), and vehicle type. This data set should encompass all stops within the Melbourne Metropolitan area.

The vehicle type is determined by the corresponding route type, where:

- 0 corresponds to tram
- 2 corresponds to train
- 3 corresponds to bus
- Any other route type is labeled as 'Unknown'.

Use this figure as an example of expected result. (Note:Data value is for demonstration purposes only.)

	stop_id	stop_name	geom	route_number	route_name	vehicle
1	1000	Dole Ave/Cheddar Rd (Reservoir)	POINT (145.018951051008 -37.7007748061772)	556	Northland SC - Epping Plaza SC	Bus
2	10001	Rex St/Taylors Rd (Kings Park)	POINT (144.776152425766 -37.7269752097248)	418	St Albans Station - Caroline Springs	Bus
3	10002	Yuille St/Centenary Ave (Melton)	POINT (144.595789405033 -37.6761595024019)	458	Melton Station - Kurunjang	Bus
4	10002	Yuille St/Centenary Ave (Melton)	POINT (144.595789405033 -37.6761595024019)	943	Watergardens Station - Melton	Bus
5	10009	Gum Rd/Main Rd West (Albanvale)	POINT (144.775899388911 -37.7414971143014)	424	St Albans Station - Brimbank Central SC	Bus
6	1001	Lloyd Ave/Cheddar Rd (Reservoir)	POINT (145.019685286526 -37.6991830099504)	556	Northland SC - Epping Plaza SC	Bus
7	10010	Kings Rd/Main Rd West (St Albans)	POINT (144.78008474429 -37.7419455261211)	424	St Albans Station - Brimbank Central SC	Bus
8	10011	Moffat St/Main Rd West (St Albans)	POINT (144.783466504334 -37.7423246041254)	424	St Albans Station - Brimbank Central SC	Bus
9	10012	Washington St/Main Rd West (St Albans)	POINT (144.787912291551 -37.7427956612577)	424	St Albans Station - Brimbank Central SC	Bus
10	10013	Kate St/Main Rd West (St Albans)	POINT (144.79457341719 -37.7435693788456)	424	St Albans Station - Brimbank Central SC	Bus
11	10013	Kate St/Main Rd West (St Albans)	POINT (144.79457341719 -37.7435693788456)	425	St Albans Station - Watergardens Station	Bus
12	10014	Raleighs Rd/Centenary Ave (Melton)	POINT (144.588776428553 -37.6753043554238)	458	Melton Station - Kurunjang	Bus

Make sure you remove any duplications in your result.

Write your SQL query here

```
CREATE TABLE ptv.stops_routes_mel AS
SELECT DISTINCT
s.stop_id,
s.stop_name,
s.geom,
r.route_short_name AS route_number,
r.route_long_name AS route_name,
CASE
WHEN r.route_type = 0 THEN 'Tram'
WHEN r.route_type = 2 THEN 'Train'
WHEN r.route_type = 3 THEN 'Bus'
ELSE 'Unknown'
END AS vehicle
FROM
ptv.stops s
JOIN
ptv.stop_times st ON st.stop_id = s.stop_id
JOIN
ptv.trips t ON t.trip_id = st.trip_id
JOIN
ptv.routes r ON r.route_id = t.route_id
JOIN
```

```
ptv.melbourne m ON ST_Within(s.geom, m.geom);
```

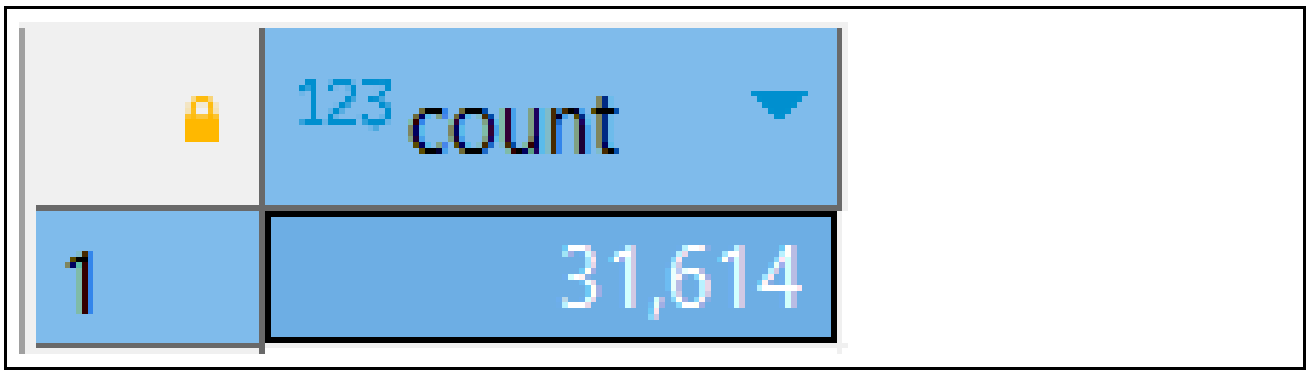
Please complete the following statistics from **stops_routes_mel** table:

Question 2.4.1:How many rows do you have in the stops_routes_mel table?

Write the SQL script to do this and attach a screenshot of the query

```
select count(*) from ptv.stops_routes_mel;
```

Screenshot

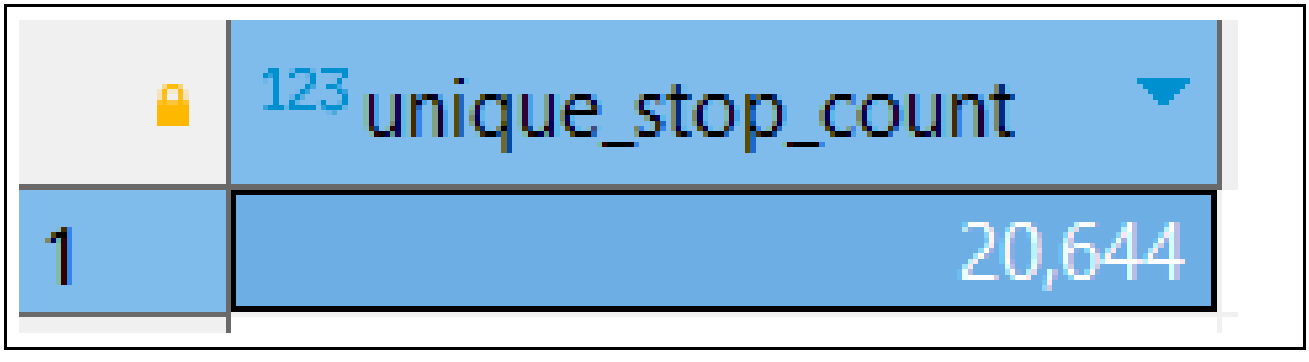


Question 2.4.2:How many unique stop_ids do you have in the stops_routes_mel table?

Write the SQL script to do this and attach a screenshot of the query

```
SELECT COUNT(DISTINCT stop_id) AS unique_stop_count
FROM ptv.stops_routes_mel_test;
```

Screenshot



Task 3: Data Analytics

3.1 Suburbs Accessibility

Create an SQL query to identify the **number of bus stops** in each Suburb. Your result should have the suburb name and the total bus stops in it.

Hint :

- identify the mesh block location of a bus stop. Then, aggregate the number in Suburb level.
- One suburb consists of multiple mesh blocks

Write your SQL query here

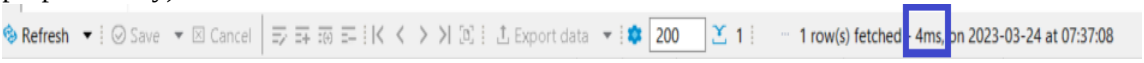
```
select
sal_name_2021,
COUNT(distinct s.stop_id) as total_bus_stops
from
ptv.stops s
join
ptv.stop_times st on
s.stop_id = st.stop_id
join
ptv.trips t on
t.trip_id = st.trip_id
join
ptv.routes r on
r.route_id = t.route_id
join ptv.melbourne m on
ST_Within(s.geom,
m.geom)
join
ptv.mb2021 mb
on
ST_Within(s.geom,
```

```
mb.wkb_geometry)
join ptv.suburb2021 s2 on
s2.mb_code_2021 = mb.mb_code21
where
r.route_type = 3
group by
sal_name_2021;
```

Provide the screenshot of your **execution plan** and the real **execution time** for this query. You can get the real execution time under your result set.

Note:

- Execution plan can be found in SQL Editor -> Explain Execution Plan
- The execution time is shown in the screenshot below.(Note: the screenshot is for demonstration purposes only)



Execution plan :

Execution time :

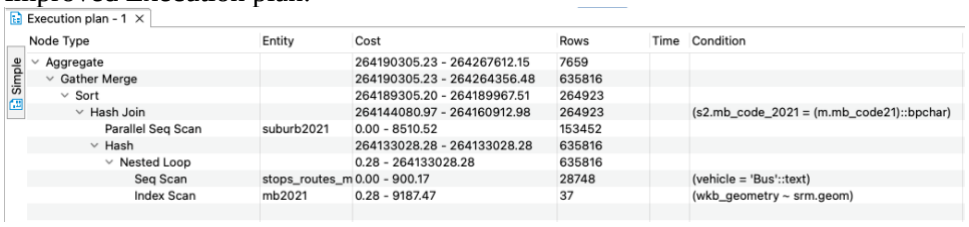
You are now tasked with devising an approach to enhance your query execution and minimize execution time. Provide a comprehensive explanation of your strategy, accompanied by the SQL script outlining the measures you've taken to optimize query performance. Additionally, include a screenshot showcasing the execution plan and execution time, effectively visualizing the enhancements achieved following the optimizatio.

Strategy and SQL script:
Strategy:
Table Consolidation: Instead of joining multiple tables (stops, stop_times, trips, routes), the optimized query uses a consolidated table (stops_routes_mel) from the previous task. This reduces the number of JOIN operations.

SQL script:

```
select
sal_name_2021,
count(distinct srm.stop_id) as total_bus_stops
from
ptv.stops_routes_mel srm
join ptv.mb2021 m on
st_within(srm.geom,
m.wkb_geometry)
join ptv.suburb2021 s2 on
s2.mb_code_2021 = m.mb_code21
where
srm.vehicle = 'Bus'
group by
sal_name_2021;
```

Improved Execution plan:



Execution plan - 1

Node Type	Entity	Cost	Rows	Time	Condition
Aggregate		264190305.23 - 264267612.15	7659		
Gather Merge		264190305.23 - 264264356.48	635816		
Sort		264189305.20 - 264189967.51	264923		
Hash Join		264144080.97 - 264160912.98	264923		(s2.mb_code_2021 = (m.mb_code21)::bpchar)
Parallel Seq Scan	suburb2021	0.00 - 8510.52	153452		
Hash		264133028.28 - 264133028.28	635816		
Nested Loop		0.28 - 264133028.28	635816		
Seq Scan	stops_routes_m	0.00 - 900.17	28748		(vehicle = 'Bus'::text)
Index Scan	mb2021	0.28 - 9187.47	37		(wkb_geometry ~ srm.geom)

Improved Execution Time :

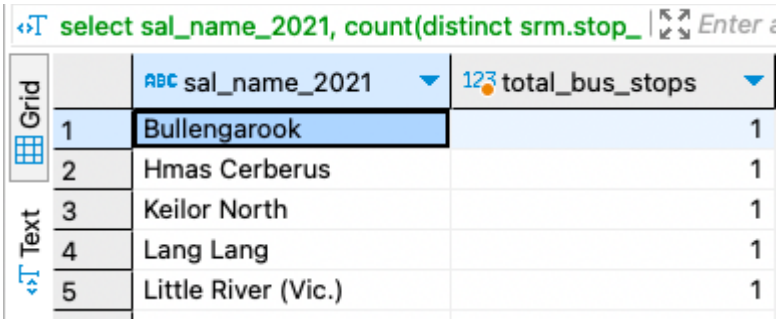
Question 3.1.1: Provide a list of the five suburbs with the lowest count of stops. In case multiple suburbs share the same minimum number of stops in your findings, arrange them in ascending order based on their suburb names.

Write the SQL script to do this and attach a screenshot of the query

```
select
```

```
sal_name_2021,
count(distinct srm.stop_id) as total_bus_stops
from
ptv.stops_routes_mel srm
join ptv.mb2021 m on
st_within(srm.geom,
m.wkb_geometry)
join ptv.suburb2021 s2 on
s2.mb_code_2021 = m.mb_code21
where
srm.vehicle = 'Bus'
group by
sal_name_2021
order by
total_bus_stops asc,
sal_name_2021 asc
limit 5;
```

Screenshot



	sal_name_2021	total_bus_stops
1	Bullengarook	1
2	Hmas Cerberus	1
3	Keilor North	1
4	Lang Lang	1
5	Little River (Vic.)	1

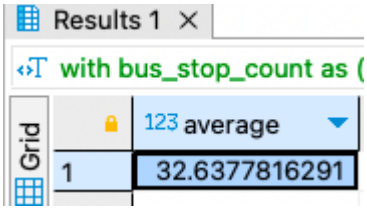
Question 3.1.2: Average number of distinct stops in suburb

Write the SQL script to do this and attach a screenshot of the query

```
with bus_stop_count as (
select
```

```
count(distinct srm.stop_id) as bus_stop_ct
from
ptv.stops_routes_mel srm
where
srm.vehicle = 'Bus'),
sub_count as (
select
count(distinct sal_name_2021) as sub_ct
from
ptv.mb2021_mel m
join ptv.suburb2021 s2 on
s2.mb_code_2021 = m.mb_code21)
select
bus_stop_count.bus_stop_ct::float / sub_count.sub_ct::float as average
from
bus_stop_count,
sub_count;
```

Screenshot



3.2 LGA Blankspot

The next step is to evaluate the **residential area without Bus Stops**. The residential area without any Bus Stops in it is considered as the **Blankspot**. Each mesh block has a distinct category. The category is defined in "**mb_cat21**" column, mb2021_mel table. In this task, your duty is to identified the percentage of blankspot in every city council or LGA. Below is the blankspot example in Kingsbury.

Let B as the number of residential blankspot, R as the total number of Residential Mesh Blocks, where B ⊆ R. The percentage of blankspot X in every LGA can be calculated using the following formula

$$X = \frac{\sum B}{\sum R} * 100\%$$

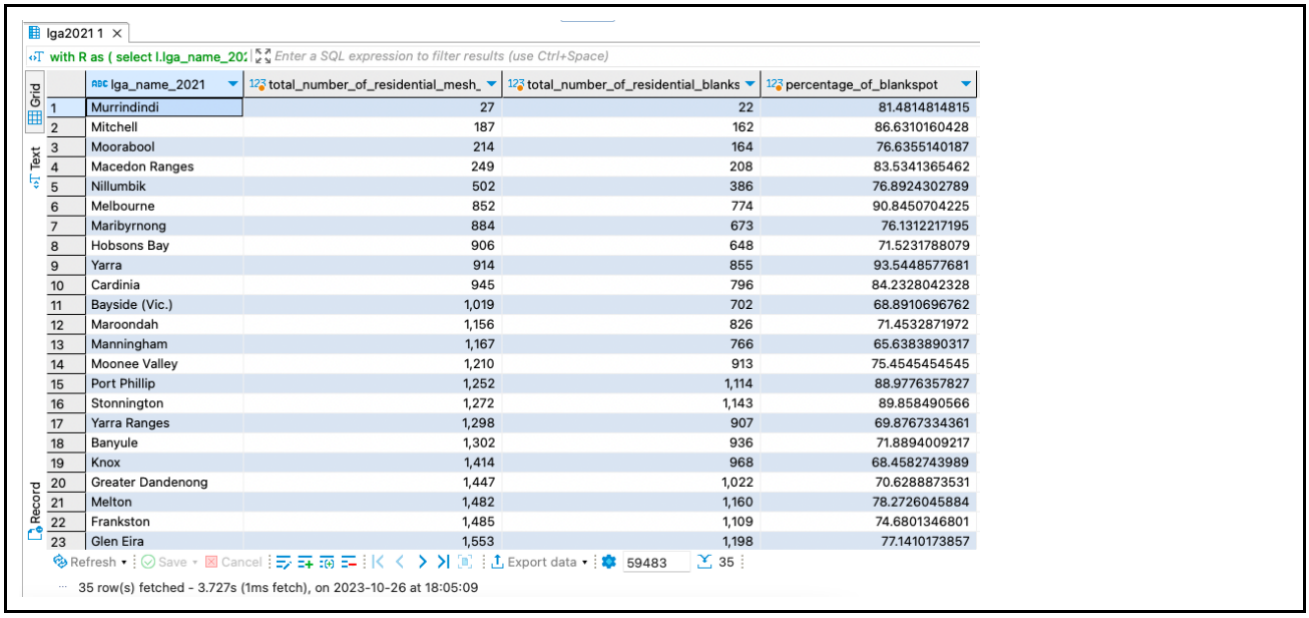
Display the LGA name, total number of Residential Mesh Blocks, total number of residential blankspot, percentage of blankspot in Melbourne Metropolitan. Sort results in ascending order by total number of Residential Mesh Blocks.

Write the SQL script to do this and attach a screenshot of the query

```
create index mb2021_mel_wkb_geometry_geom_idx on
ptv.mb2021_mel
using gist (wkb_geometry);

create index stops_routes_mel_geom_idx on
ptv.stops_routes_mel
using gist (geom);


with R as (
select
l.lga_name_2021,
count(distinct mm.mb_code21) as ct
from
ptv.mb2021_mel mm
join ptv.lga2021 l on
l.mb_code_2021 = mm.mb_code21
where
mm.mb_cat21 = 'Residential'
group by
l.lga_name_2021),
NB as (
select
l.lga_name_2021,
count(distinct mm.mb_code21) as ct
from
ptv.mb2021_mel mm
```

```
join ptv.stops_routes_mel srm on
st_within(srm.geom,
mm.wkb_geometry)
join ptv.lga2021 l on
l.mb_code_2021 = mm.mb_code21
where
srm.vehicle = 'Bus'
and mm.mb_cat21 = 'Residential'
group by
l.lga_name_2021)
select
R.lga_name_2021,
R.ct as total_number_of_Residential_Mesh_Blocks,
R.ct - NB.ct as total_number_of_residential_blankspot,
((R.ct - NB.ct)/ R.ct::float)* 100 as percentage_of_blankspot
from
R
join NB on
R.lga_name_2021 = NB.lga_name_2021
order by
total_number_of_Residential_Mesh_Blocks asc;
```

Screenshot


lga2021 1

with R as (select l.lga_name_20: Enter a SQL expression to filter results (use Ctrl+Space)

	lga_name_2021	total_number_of_residential_mesh_	total_number_of_residential_blanks	percentage_of_blankspot
1	Murrindindi	27	22	81.4814814815
2	Mitchell	187	162	86.6310160428
3	Moorabool	214	164	76.6355140187
4	Macedon Ranges	249	208	83.5341365462
5	Nillumbik	502	386	76.8924302789
6	Melbourne	852	774	90.8450704225
7	Maribyrnong	884	673	76.1312217195
8	Hobsons Bay	906	648	71.5231788079
9	Yarra	914	855	93.5448577681
10	Cardinia	945	796	84.2328042328
11	Bayside (Vic.)	1,019	702	68.8910696762
12	Maroondah	1,156	826	71.4532871972
13	Manningham	1,167	766	65.6383890317
14	Moonee Valley	1,210	913	75.4545454545
15	Port Phillip	1,252	1,114	88.9776357827
16	Stonnington	1,272	1,143	89.858490566
17	Yarra Ranges	1,298	907	69.8767334361
18	Banyule	1,302	936	71.8894009217
19	Knox	1,414	968	68.4582743989
20	Greater Dandenong	1,447	1,022	70.6288873531
21	Melton	1,482	1,160	78.2726045884
22	Frankston	1,485	1,109	74.6801346801
23	Glen Eira	1,553	1,198	77.1410173857

Refresh Save Cancel Export data 59483 35

35 row(s) fetched - 3.727s (1ms fetch), on 2023-10-26 at 18:05:09

Question 3.2.1: Complete the following statistical data based on the result.

Note:

- The query and screenshot are not required for this section. You can write down your results directly
- If more than one suburb has the same percentage of blankspots, sort them by suburb name in ascending order.

Criteria	Answer
Top 5 LGAs with the highest % of blankspots	Yarra Melbourne Stonnington Port Phillip Mitchell
Top 5 LGAs with the lowest % of blankspots	Manningham Whitehorse Knox Bayside (Vic.) Yarra Ranges
Average % of blankspots	76.81374928268006

Task 4: Data Visualisation

In this task, you are required to incorporate a heatmap visualization.

4.1 LGA Blankspot Analysis

In this task, you will put the blankspot percentage in the heatmap. Provide the segmentation as follow:

- $X \leq 20\ \%$
- $20\% < X \leq 40\%$
- $40\% < X \leq 60\%$
- $60\% < X \leq 80\%$
- $X > 80\%$

Write an SQL query to create a table named 'lga_blankspot' containing the blankspot percentages categorised by the LGA from the previous task 3.2 and sufficient spatial data. And attach a screenshot of the table contents.

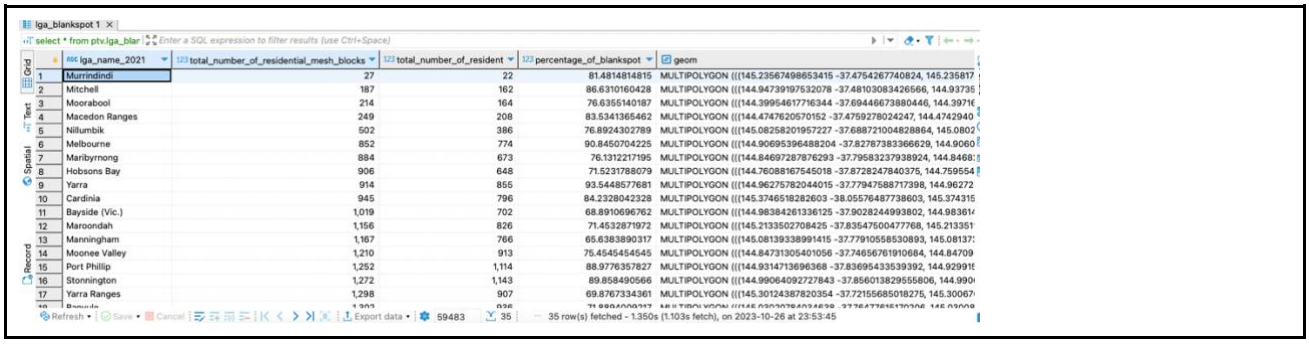
Note: Ensure that this table is structured to facilitate the creation of a visual heat map specifically for the Melbourne region.

```
create table ptv.lga_blankspot as (
with filtered as (
select * from ptv.mb2021_mel mm where mm.mb_cat21 = 'Residential'
),
R as (
select
l.lga_name_2021,
count(distinct f.mb_code21) as ct
from
filtered f
join ptv.lga2021 l on
l.mb_code_2021 = f.mb_code21
group by
l.lga_name_2021),
NB as (
select
l.lga_name_2021,
count(distinct f.mb_code21) as ct
from
filtered f
join ptv.stops_routes_mel srm on
```

```
st_within(srm.geom,
f.wkb_geometry)
join ptv.lga2021 l on
l.mb_code_2021 = f.mb_code21
where
srm.vehicle = 'Bus'
group by
l.lga_name_2021),
layer as (
select
l.lga_name_2021,
st_union(f.wkb_geometry) as geom
from
filtered f
join ptv.lga2021 l on
l.mb_code_2021 = f.mb_code21
group by
l.lga_name_2021
)
select
R.lga_name_2021,
R.ct as total_number_of_Residential_Mesh_Blocks,
R.ct - NB.ct as total_number_of_residential_blankspot,
((R.ct - NB.ct)/ R.ct::float)* 100 as percentage_of_blankspot,
layer.geom
from
R
join NB on
R.lga_name_2021 = NB.lga_name_2021
```

```
join layer on layer.lga_name_2021 = R.lga_name_2021
order by
total_number_of_Residential_Mesh_Blocks asc
);


select * from ptv.lga_blankspot;
```

Screenshot



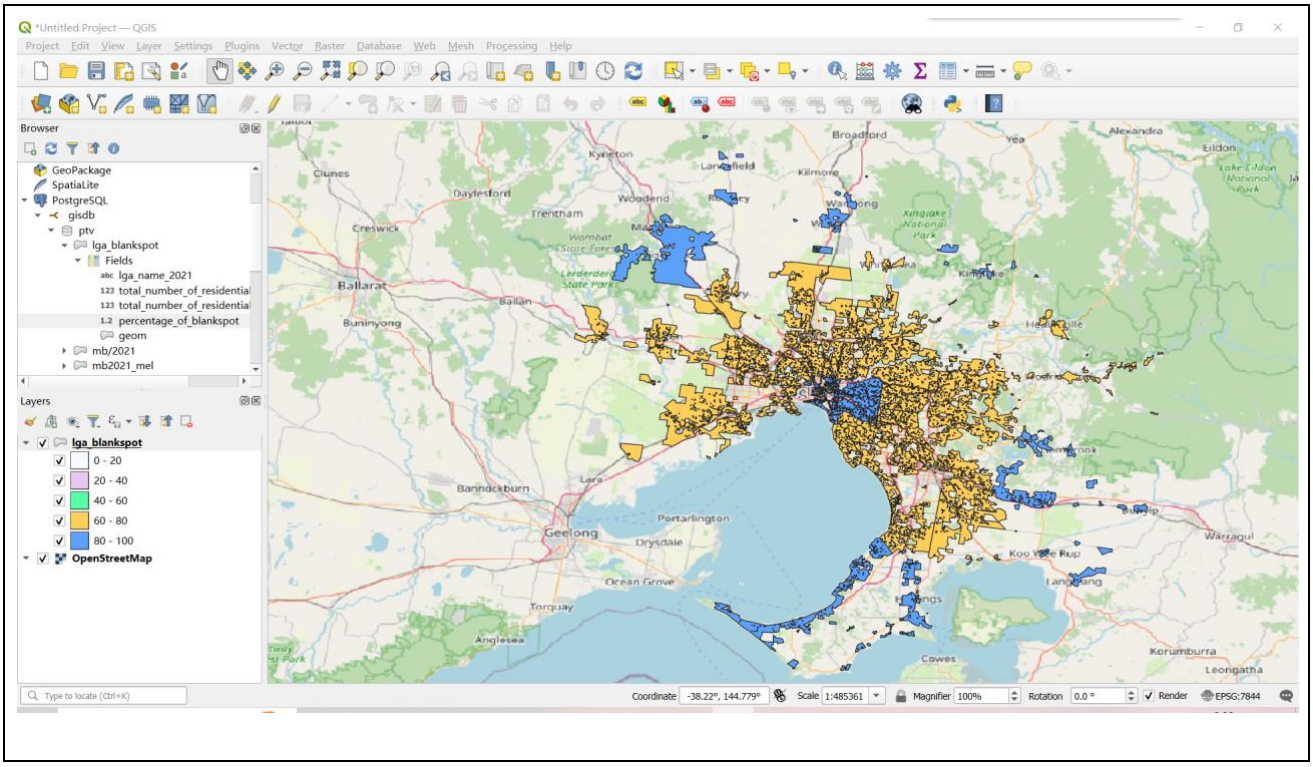
lga_blankspot 1

	lga_name_2021	total_number_of_residential_mesh_blocks	total_number_of_resident	percentage_of_blankspot	geom
1	Murrindindi	27	22	81.4814814815	MULTIPOLYGON (((145.23567498653415 -37.4754267740824, 145.235817
2	Mitchell	187	162	86.6310160428	MULTIPOLYGON (((144.94739197532078 -37.48103083426566, 144.93735
3	Moorabool	214	164	76.6355140187	MULTIPOLYGON (((144.39954617716344 -37.69446673880446, 144.39716
4	Macedon Ranges	249	208	83.5341365462	MULTIPOLYGON (((144.474762057015 2 -37.4759278024247, 144.4742940
5	Nillumbik	502	386	76.8924302789	MULTIPOLYGON (((145.08258201957227 -37.688721004828864, 145.0802
6	Melbourne	852	774	90.8450704225	MULTIPOLYGON (((144.90695396488204 -37.82787383366629, 144.9060
7	Maribyrnong	884	673	76.1312217195	MULTIPOLYGON (((144.84697287876293 -37.79583237938924, 144.8468:
8	Hobsons Bay	906	648	71.5231788079	MULTIPOLYGON (((144.76088167545018 -37.8728247840375, 144.759554
9	Yarra	914	855	93.5448577681	MULTIPOLYGON (((144.96275782044015 -37.77947588717398, 144.96272
10	Cardinia	945	796	84.2328042328	MULTIPOLYGON (((145.3746518282603 -38.05576487738603, 145.374315
11	Bayside (Vic.)	1,019	702	68.8910696762	MULTIPOLYGON (((144.98384261336125 -37.90282449938 02, 144.98361
12	Maroondah	1,156	826	71.4532871972	MULTIPOLYGON (((145.2133502708425 -37.83547500477768, 145.213351
13	Manningham	1,167	766	65.6383890317	MULTIPOLYGON (((145.08139338991415 -37.77910558530893, 145.08137:
14	Moonee Valley	1,210	913	75.4545454545	MULTIPOLYGON (((144.84731305401056 -37.74656761910684, 144.84709
15	Port Phillip	1,252	1,114	88.9776357827	MULTIPOLYGON (((144.9314713696368 -37.83695433539392, 144.92991
16	Stonnington	1,272	1,143	89.858490566	MULTIPOLYGON (((144.99064092727843 -37.85601382955 806, 144.990
17	Yarra Ranges	1,298	907	69.8767334361	MULTIPOLYGON (((145.30124387820354 -37.72155685018275, 145.30067

Refresh • Save • Cancel Export data • 59483 35 35 row(s) fetched - 1.350s (1.103s fetch), on 2023-10-26 at 23:53:45

Provide the screenshot for the heatmap by using QGIS. Please note that the heatmap is map-based and visually represents the distribution of blankspot percentages across different LGAs in the Melbourne region.

Remember to include appropriate labels, titles, and legends in your visualizations to make them easy to understand (Updated 05/10/2023)


*Untitled Project — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Browser
GeoPackage
SpatiaLite
PostgreSQL
gisdb
ptv
lga_blankspot
Fields
lga_name_2021
total_number_of_residential
total_number_of_residential
percentage_of_blankspot
geom
mb/2021
mb2021_mel
Layers
lga_blankspot
0 - 20
20 - 40
40 - 60
60 - 80
80 - 100
OpenStreetMap
Broadford
Kyneton
Clunes
Daylesford
Trentham
Woodend
Creswick
Ballarat
Bunninyong
Bannockburn
Geelong
Lara
Portarlington
Drysdale
Ocean Grove
Torquay
Anglesea
Cowes
Korumburra
Leongatha
Alexandra
Eildon
Yea
Kilmore
Kinglake National Park
Warragul
Langwarrin
Coordinate -38.22°, 144.779°
Scale 1:485361
Magnifier 100%
Rotation 0.0 °
Render
EPSG:7844

End