**Project**

**2006-commercial-vehicle-survey-traffic-volumes-at-survey-stations**

Icon

Description automatically generated

Rayavarapu Satya Sai Manish, 8869731

Business Analytics, Conestoga College

INFO8076-23F-Sec1-SQL and Data Analysis

**Jey Kumaresan**

October 22, 2023

Table of Contents

1. **Introduction**
2. **Data Import and Data Cleaning**
3. **ERD Diagram**
4. **SQL Query**
5. **Views**
6. **Conclusions**

**Introduction:**

This dataset is essential for transportation and logistics analysis, as it contains information related to the flow of commercial vehicles at specific locations within the province.

The 2006 Commercial Vehicle Survey Traffic Volumes at Survey Stations dataset offers a comprehensive record of traffic volumes for commercial vehicles at specific survey stations throughout Ontario. It includes a wide range of information about the movement of commercial vehicles, which can be particularly valuable for transportation planners, policymakers, researchers, and businesses involved in logistics.

The 2006 Commercial Vehicle Survey Traffic Volumes at Survey Stations dataset provides a valuable resource for understanding the movement of commercial vehicles within Ontario, supporting informed decision-making, research, and policy development in the field of transportation and logistics. Researchers and professionals interested in this domain may find this dataset particularly useful for their work. Below are the terms used in table

**Station ID**: Unique CVS station number

**Station name:** Station name

**Direction:** Direction of traffic

**MTO region:** Five regions of MTO (Central, Eastern, etc)

**Highway or road:** Highway number or road name

**Location:** Description of location

**Day of week number:** A number between 1 and 7 representing the day of week. 1=Sunday

**Hour:** Hour of the day, 0 to 23 represents starting hour of the day (e.g. 12 represents 12 P.M)

**Single:** Number of single-unit trucks

**Multi:** Number of multi-unit trucks

**Auto:** Number of cars and other passenger vehicles

**Total trucks:** Sum of single and multi-unit vehicles

**Total vehicles:** Number of total vehicles

**Url**: <https://data.ontario.ca/dataset/2006-commercial-vehicle-survey-traffic-volumes-at-survey-stations/resource/489a8758-a183-4a96-8087-4a1c183b56d9>

**Data Import & Data Cleaning**

Data import refers to the process of bringing data from external sources into a PostgreSQL database. This is commonly done when you have data in various formats that you want to store and manipulate within your database. We have taken csv as a data source.

Data cleaning is the process of identifying and rectifying inconsistencies, errors, and anomalies in your data to ensure its accuracy and reliability. In the context of PostgreSQL, you can perform data cleaning using SQL queries and various functions.

We have divided the given dataset into four tables namely as below.

**Survey table**

**City table**

**Locations table**

**Region table**

Here Survey table is the fact table and three tables are known as dimension tables.

The attributes in the survey tables are

**Survey\_ID, Location\_ID, Station\_ID, City\_ID, Region\_ID, Direction, City\_Name, Day of Week Number, Hour, Single, Multi, Auto, total\_trucks, Total vehicles.**

The attributes in the Location tables are

**Location\_ID, Station\_ID, Station\_Name, Direction**

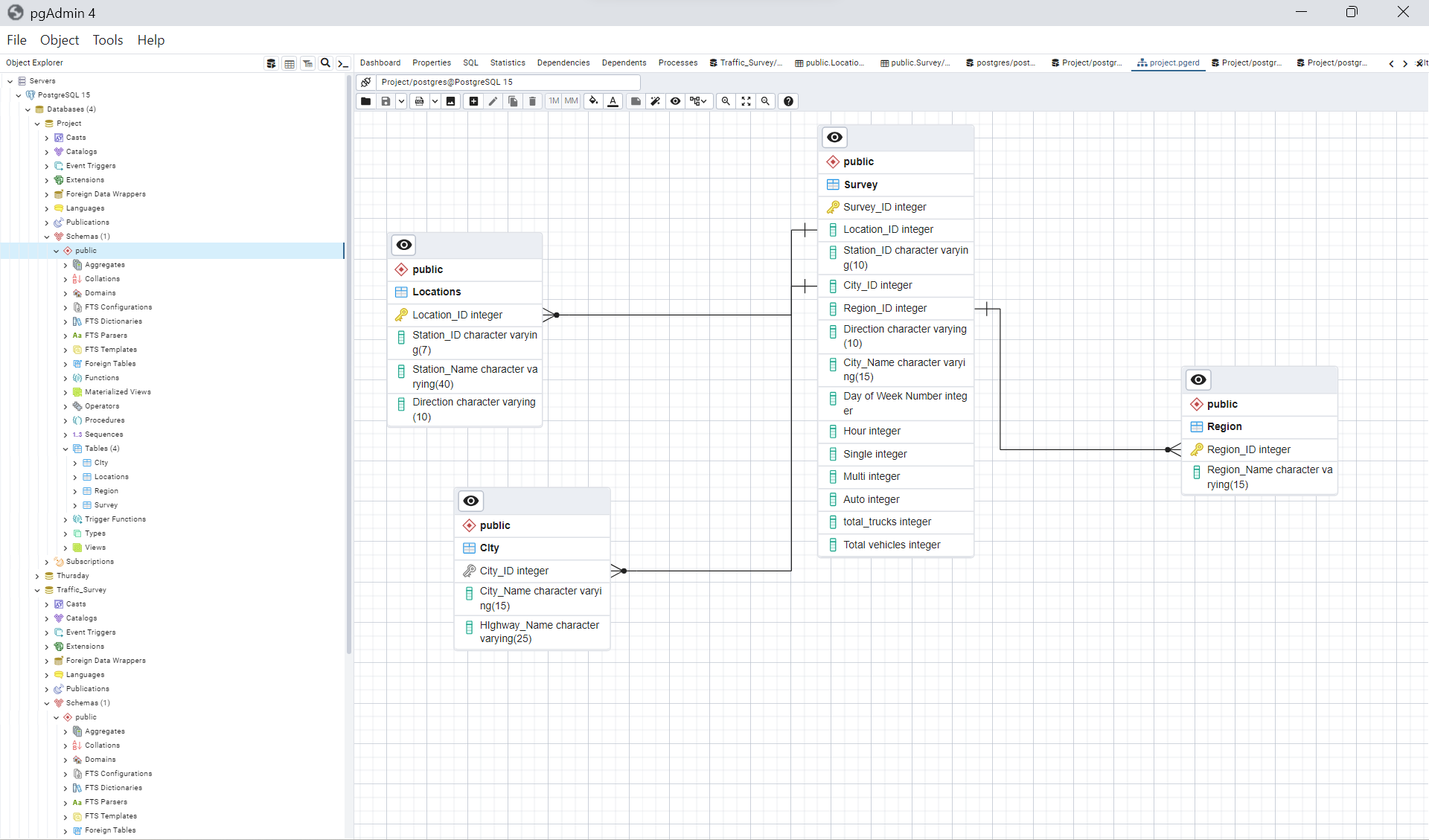
The attributes in the City tables are

**City\_ID, City\_Name, Highway\_Name**

The attributes in the Location tables are

**Region\_ID, Region\_Name**

**ERD Diagram**



Here I have created four tables for which I have created the relationships for each table to fetch the data without any delay. I have created the one-to-many relationships for the primary keys like city ID, location\_id, and region\_id.

**SQL Query**

**Query-1**

SELECT m."Survey\_ID",

m."City\_Name",

n."Highway\_Name",

r."Station\_Name",

r."Direction",

o."Region\_Name",

case when

m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday'

when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday'

when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday'

when m."Day of Week Number" = 7 then 'Saturday'

end "Day of Week Number",

case when

m."Hour" = 0 then '12 AM-1 AM'

when m."Hour" = 1 then '1 AM-2 AM' when m."Hour" = 2 then '2 AM-3 AM'

when m."Hour" = 3 then '3 AM-4 AM' when m."Hour" = 4 then '4 AM-5 AM'

when m."Hour" = 5 then '5 AM-6 AM' when m."Hour" = 6 then '6 AM-7 AM'

when m."Hour" = 7 then '7 AM-8 AM' when m."Hour" = 8 then '8 AM-1 AM'

when m."Hour" = 9 then '9 AM-10 AM' when m."Hour" = 10 then '10 AM-11 AM'

when m."Hour" = 11 then '11 AM-12 PM' when m."Hour" = 12 then '12 PM-1 PM'

when m."Hour" = 13 then '12 PM-1 PM' when m."Hour" = 14 then '12 PM-1 PM'

when m."Hour" = 15 then '12 PM-1 PM' when m."Hour" = 16 then '12 PM-1 PM'

when m."Hour" = 17 then '12 PM-1 PM' when m."Hour" = 18 then '12 PM-1 PM'

when m."Hour" = 19 then '12 PM-1 PM' when m."Hour" = 20 then '12 PM-1 PM'

when m."Hour" = 21 then '12 PM-1 PM' when m."Hour" = 22 then '12 PM-1 PM'

when m."Hour" = 23 then '12 PM-1 PM' when m."Hour" = 24 then '12 PM-1 PM'

end Timezone,

m."Single",

m."Multi",

m."Auto",

"Total vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

A screenshot of a computer

Description automatically generated

**Query-2**

SELECT

m."City\_Name", n."Highway\_Name", o."Region\_Name",

case when m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday' when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday' when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday' when m."Day of Week Number" = 7 then 'Saturday'

end as "Day of Week Number",

case when

m."Hour" in ('6','7','8','9','10','11') then 'Morning' when m."Hour" in ('12','13','14','15','16','17') then 'Afternoon'

when m."Hour"in ('18','19','20','21') then 'Evening' when m."Hour" in ('22','23','24','0','1','2','3','4','5') then 'Night'

end "Time of Day",

sum(m."Single") as "Total\_Single\_Trucks", sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles", sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1,2,3,4,5

A screenshot of a computer

Description automatically generated

**Query-3 – Region Data**

SELECT

o."Region\_Name",

sum(m."Single") as "Total\_Single\_Trucks",

sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

**A screenshot of a computer

Description automatically generated**

**Query-4 - City Data**

SELECT

m."City\_Name",

sum(m."Single") as "Total\_Single\_Trucks",

sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

A screenshot of a computer

Description automatically generated

**Query-5 - Highway Data**

SELECT

n."Highway\_Name",

sum(m."Single") as "Total\_Single\_Trucks",

sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

**A screenshot of a computer

Description automatically generated**

**Query-6 - Data recorded for a number of vehicles as per Day**

SELECT

case when

m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday' when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday' when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday' when m."Day of Week Number" = 7 then 'Saturday'

end as "Day of Week Number",

sum(m."Single") as "Total\_Single\_Trucks", sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles", sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

order by "Total\_Vehicles" desc

A screenshot of a computer

Description automatically generated

**Query-7 – Data recorded for a number of vehicles as per time**

SELECT

case when

m."Hour" in ('6','7','8','9','10','11') then 'Morning'

when m."Hour" in ('12','13','14','15','16','17') then 'Afternoon'

when m."Hour"in ('18','19','20','21') then 'Evening'

when m."Hour" in ('22','23','24','0','1','2','3','4','5') then 'Night'

end "Time of Day",

sum(m."Single") as "Total\_Single\_Trucks",

sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

order by "Total\_Vehicles" desc

A screenshot of a computer

Description automatically generated

**Query-8 - Data for vehicles as per direction**

SELECT

m."Direction",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi") as "Total\_Trucks"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1

order by "Total\_Auto\_Vehicles" desc

A screenshot of a computer

Description automatically generated

**Query-9 - Maximum auto vehicles traveled for a day**

SELECT

m."Direction",

m."City\_Name",

n."Highway\_Name",

o."Region\_Name",

case when

m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday'

when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday'

when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday'

when m."Day of Week Number" = 7 then 'Saturday'

end as "Day of Week Number",

case when

m."Hour" in ('6','7','8','9','10','11') then 'Morning'

when m."Hour" in ('12','13','14','15','16','17') then 'Afternoon'

when m."Hour"in ('18','19','20','21') then 'Evening'

when m."Hour" in ('22','23','24','0','1','2','3','4','5') then 'Night'

end "Time of Day",

m."Hour",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi") as "Total\_Trucks"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

where m."Auto" = (select max("Auto") from "Survey")

group by 1,2,3,4,5,6,7

order by "Total\_Auto\_Vehicles" desc

A screenshot of a computer

Description automatically generated

**Query-10 - Maximum trucks traveled for a day**

SELECT

m."Direction",

m."City\_Name",

n."Highway\_Name",

o."Region\_Name",

case when

m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday' when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday' when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday' when m."Day of Week Number" = 7 then 'Saturday'

end as "Day of Week Number",

case when

m."Hour" in ('6','7','8','9','10','11') then 'Morning'

when m."Hour" in ('12','13','14','15','16','17') then 'Afternoon'

when m."Hour"in ('18','19','20','21') then 'Evening'

when m."Hour" in ('22','23','24','0','1','2','3','4','5') then 'Night'

end "Time of Day",

m."Hour",

sum(m."Auto") as "Total\_Auto\_Vehicles",

sum(m."Single"+m."Multi") as "Total\_Trucks"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

where m."Total vehicles" = (select max("Total vehicles") from "Survey")

group by 1,2,3,4,5,6,7

order by "Total\_Auto\_Vehicles" desc

**A screenshot of a computer

Description automatically generated**

**Views**

**View 1**

create view Survey\_Data as SELECT m."Survey\_ID",

m."City\_Name",

n."Highway\_Name",

r."Station\_Name",

r."Direction",

o."Region\_Name",

case when

m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday'

when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday'

when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday'

when m."Day of Week Number" = 7 then 'Saturday'

end "Day of Week Number",

case when

m."Hour" = 0 then '12 AM-1 AM'

when m."Hour" = 1 then '1 AM-2 AM' when m."Hour" = 2 then '2 AM-3 AM'

when m."Hour" = 3 then '3 AM-4 AM' when m."Hour" = 4 then '4 AM-5 AM'

when m."Hour" = 5 then '5 AM-6 AM' when m."Hour" = 6 then '6 AM-7 AM'

when m."Hour" = 7 then '7 AM-8 AM' when m."Hour" = 8 then '8 AM-1 AM'

when m."Hour" = 9 then '9 AM-10 AM' when m."Hour" = 10 then '10 AM-11 AM'

when m."Hour" = 11 then '11 AM-12 PM' when m."Hour" = 12 then '12 PM-1 PM'

when m."Hour" = 13 then '12 PM-1 PM' when m."Hour" = 14 then '12 PM-1 PM'

when m."Hour" = 15 then '12 PM-1 PM' when m."Hour" = 16 then '12 PM-1 PM'

when m."Hour" = 17 then '12 PM-1 PM' when m."Hour" = 18 then '12 PM-1 PM'

when m."Hour" = 19 then '12 PM-1 PM' when m."Hour" = 20 then '12 PM-1 PM'

when m."Hour" = 21 then '12 PM-1 PM' when m."Hour" = 22 then '12 PM-1 PM'

when m."Hour" = 23 then '12 PM-1 PM' when m."Hour" = 24 then '12 PM-1 PM'

end Timezone,

m."Single",

m."Multi",

m."Auto",

"Total vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

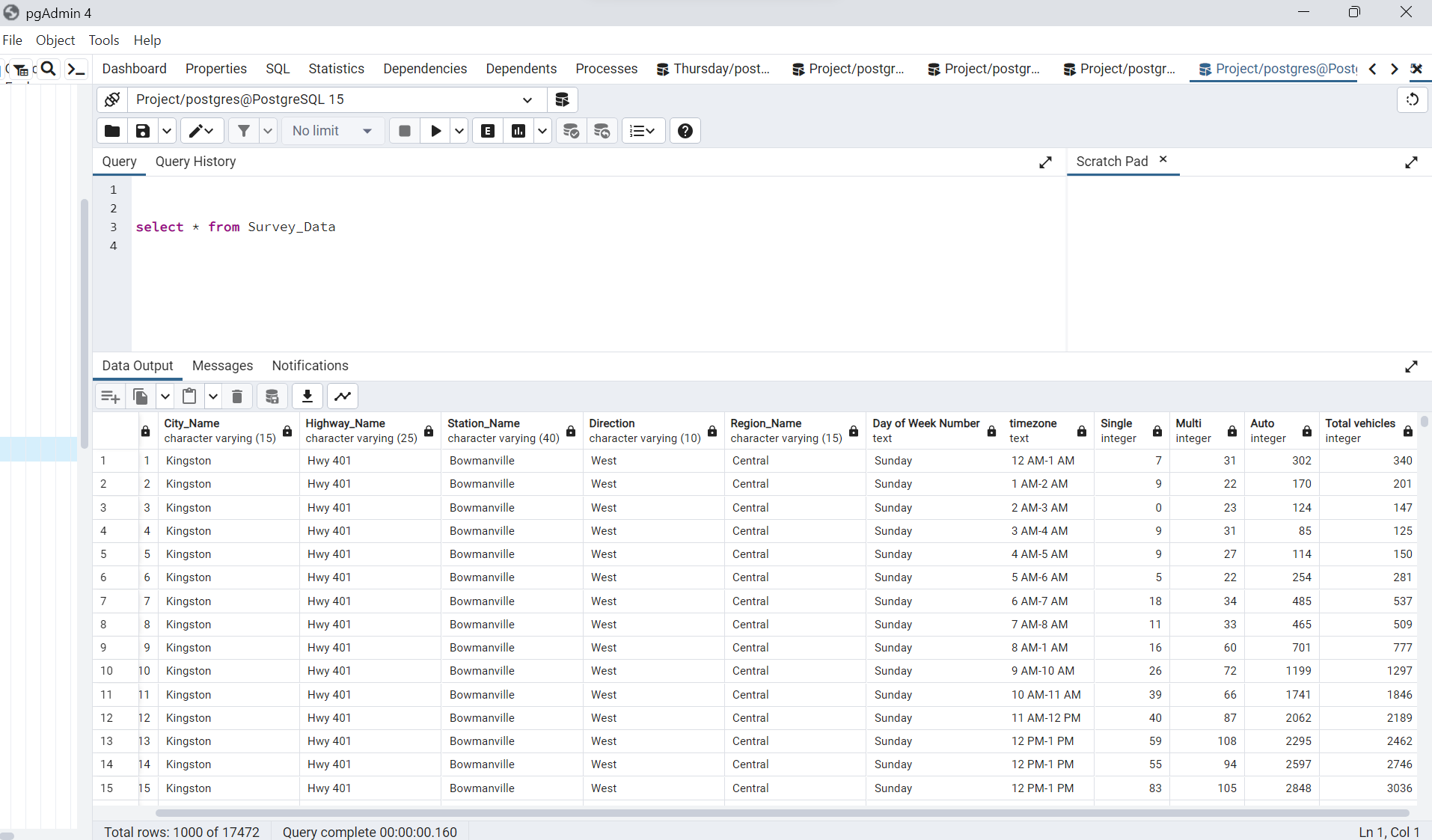
on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID" ;



**View 2**

Create view Day\_wise\_Data as SELECT

m."City\_Name", n."Highway\_Name", o."Region\_Name",

case when m."Day of Week Number" = 1 then 'Sunday'

when m."Day of Week Number" = 2 then 'Monday' when m."Day of Week Number" = 3 then 'Tuesday'

when m."Day of Week Number" = 4 then 'Wednesday' when m."Day of Week Number" = 5 then 'Thursday'

when m."Day of Week Number" = 6 then 'Friday' when m."Day of Week Number" = 7 then 'Saturday'

end as "Day of Week Number",

case when

m."Hour" in ('6','7','8','9','10','11') then 'Morning' when m."Hour" in ('12','13','14','15','16','17') then 'Afternoon'

when m."Hour"in ('18','19','20','21') then 'Evening' when m."Hour" in ('22','23','24','0','1','2','3','4','5') then 'Night'

end "Time of Day",

sum(m."Single") as "Total\_Single\_Trucks", sum(m."Multi") as "Total\_Multi\_Trucks",

sum(m."Auto") as "Total\_Auto\_Vehicles", sum(m."Single"+m."Multi"+m."Auto") as "Total\_Vehicles"

from "Survey" m

inner join "City" n

on

m."City\_ID" = n."City\_ID"

inner join "Region" o

on

m."Region\_ID" = o."Region\_ID"

inner join "Locations" r

on

m."Location\_ID" = r."Location\_ID"

group by 1,2,3,4,5;

A screenshot of a computer

Description automatically generated

**Conclusion**

This dataset is essential for understanding the composition of traffic on the surveyed roads. It

allows for a detailed examination of how different types of vehicles interact with the road

infrastructure and provides insights into any variations or trends that may exist. To perform a

meaningful analysis, it is important to preprocess the data, ensuring data quality and consistency.

I have created two queries(Query 1 &amp; Query 2) which help to provide insights regarding the

count of vehicles(trucks and auto) for different days of the week. These queries also give the data

for different cities, regions and directions where the traffic is observed.

Analyzing the dataset can reveal the composition of traffic during the surveyed week. It is likely

to show variations in traffic volumes between trucks and autos throughout the day. I have used

case statements to provide the name of the day of the week and utilized the hour column to

differentiate the data for different times like morning, afternoon, evening, and night. By identifying

peak hours for each vehicle type, transportation authorities and planners can optimize traffic

management strategies.

When trying to segregate the data for various regions I have found that the maximum count of auto

vehicles was **8168552** and the maximum total count of single and multi-unit trucks was **9025305**.

This was observed for the **central** region. When I tried to look for the highway traffic I observed

the maximum count of auto vehicles was **6291687** and the maximum total count of single and

multi-unit trucks were **7098170**. It was for **Highway 401**.

Understanding the hourly volumes of trucks and autos is crucial for capacity planning and ensuring

the efficient use of road infrastructure. On analyzing the data I found that the maximum count of

auto vehicles was **1760527** and the maximum total count of single and multi-unit trucks was

**1996384**. It was recorded on **Friday**.

Analyzing the data can help assess the road capacity during different times of the day and week.

When analyzing data for different times in the day I found that the maximum count of auto vehicles

was **4343572** and the maximum total count of single and multi-unit trucks was **4822439**. It was

recorded during the **afternoon (6 am -12 pm).** It can also contribute to the development of

transportation models that forecast future traffic demands.

The direction of the traffic is important to know in order to figure out where the traffic occurs

more. When analyzing the data I found that the maximum count of auto vehicles was **4594138** and

the maximum total count of single and multi-unit trucks was **635855**. It was recorded in the **east**

direction.

Overall, when analyzing the data it is found that the maximum count of total trucks vehicles was

**13129** and the maximum total count of single and multi-unit trucks **964** was found on **Thursday**

**morning** time in **Mississauga** driving towards the **central** region in the **east** direction on **Highway**

**401**. It was also noticed that the maximum count of total truck vehicles was **13168** and the

maximum total count of single and multi-unit trucks **400** was found on **Wednesday afternoon**

time in **Mississauga** driving towards the **central** region in the **east** direction on **Highway 401**.