General Sir John Kotelawala Defence University

Faculty of Management, Social Sciences, and Humanities

Department of Languages

BSc in Applied Data Science Communication

1st Year: 2nd Semester

SQL for Data Science – LB 1224

Assignment 03

Lecturer: Dr. Charith Silva

Task 01

Australian Government Visualization Projects

National Passenger Travel,

by transport mode dataset

TABLE OF CONTENT

1. Introduction

The transportation sector is very crucial for Australia's economy as it helps in linking the people and the products throughout the country. It would also be easier to develop better structures and improve services by integrating long-term sustainable solutions if an understanding of how people move can be achieved. This report analysis data from the National Passenger Movement Dataset which is a very useful data set in monitoring the movements of people in Australia using different means of transport. And our goal in this exercise is to analyze the data and be able to understand the patterns of travel, the normal modes of transport and areas that need more attention or revamping.

This study has used SQL to prepare and manage its data for analysis and other processes. Among the key factors that can be analyzed using SQL include different geographical distributions, seasonal variations in volume of travel and the number of times each means of transport is used during a period. Once the data is cleaned and prepared for the analysis, Power BI is employed to create simple and easy to understand graphics which can portray complex concepts. When these methods are combined, they help to convert large volumes of data into usable information for management.

This report aims to:

1. Understand the preferred modes chosen by people.

2. Analyze travel behavior in different periods and different locations.

3. Analyze transportation systems and recommend how they can be improved.

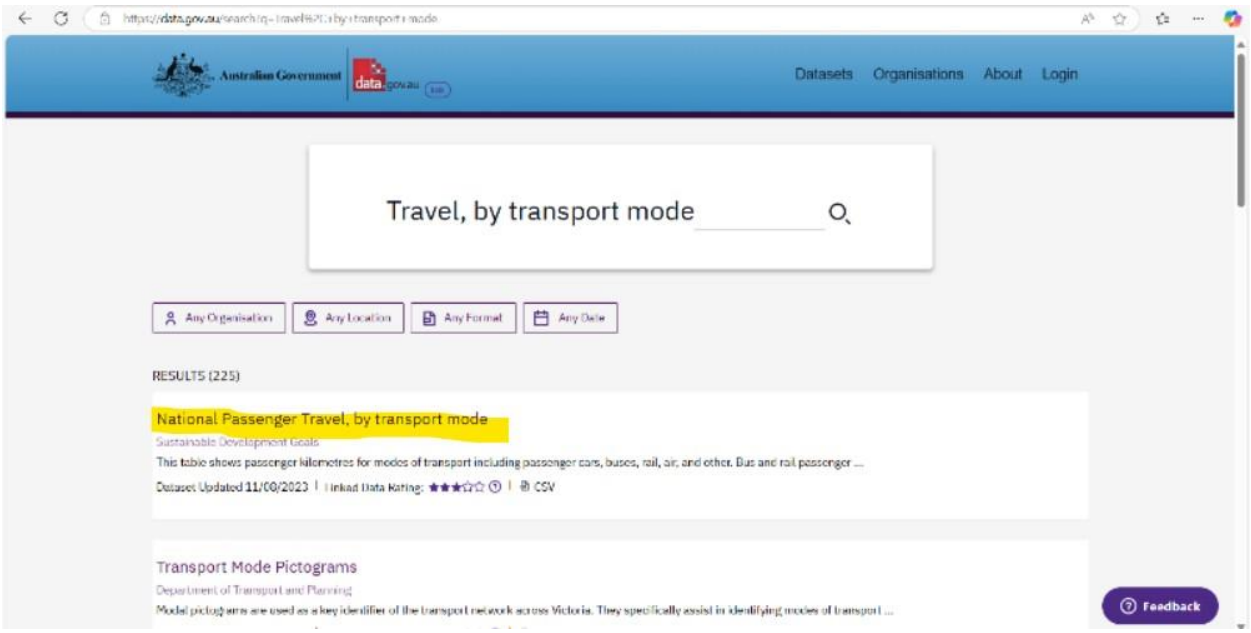
2.Exploration of Data Set

- **Review of the data set**

The dataset depicts how Australians travelled domestically, and the volume of traffic for different modes of transport is measured in billion passenger kilometers.

You can find the database / resources

Data.gov.au



The following is a brief overview of the structure and content of the dataset:

Structure:

• 17 entries mainly represent data for the financial years from 1999 - 2000 to 2015 - 2016 for rows.

• 7 columns where each one represents either a certain means of transportation or an aggregated estimation.

Columns:

1. Financial year - Appears as the year of the data in the form of “YYYY-MM”.

2. Passenger cars - Cars traveled distance by billion passenger kilometers.

3. Buses - Bus traveled distance in billions of passenger kilometers.

4. Rail - Completed rail traveling distance in billion passenger kilometers.

5. Air - The distance traveled via air in billions of passenger kilometers.

6. Other - Distance traveled via any other means in billion passenger kilometers.

7. Total - Cumulatively all passengers traveled across all modes of transportation in billions of passenger kilometers.

Key Observations:

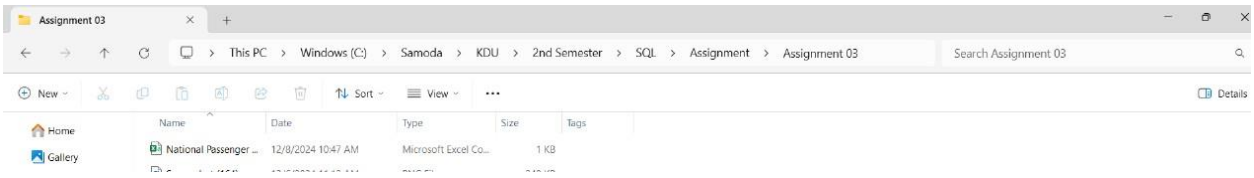
• From the data, passenger vehicles have the highest number of usages followed by air transport. Values of passenger kilometers from this region range from as low as237.16 up to 279.07billion passenger kilometers.

• Total distance traveled by planes has increased significantly over the years from 32.84 to 71.44 billion passenger kilometers with gaps of a few years.

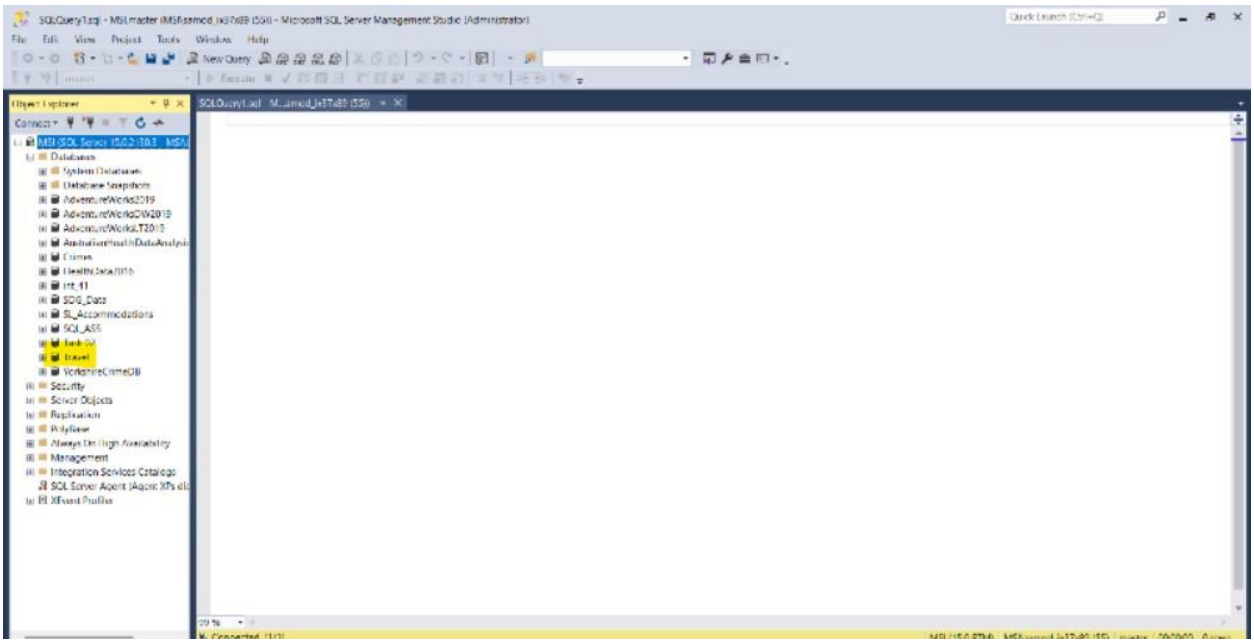
• The percentage of usage of buses and rail averaged some value over the years whereas the percentage of cars and air is said to have averaged lower in comparison.

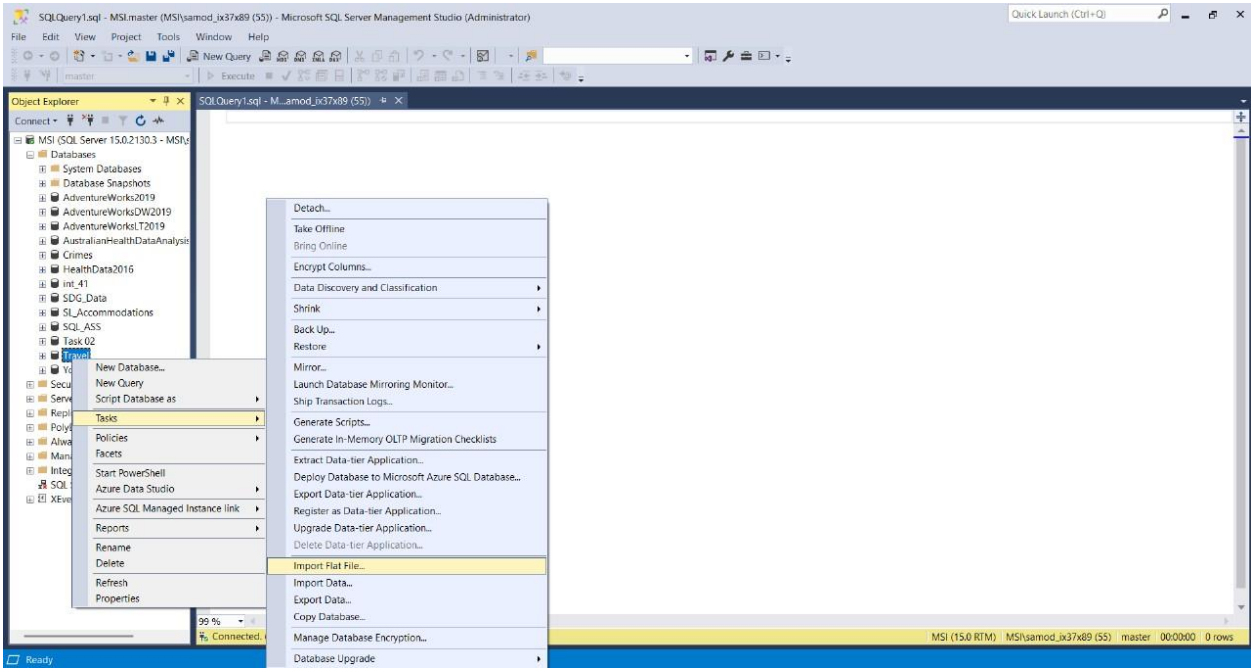
•All the other modes seem to be not significantly different.

- **Importing data set to SQL Server**

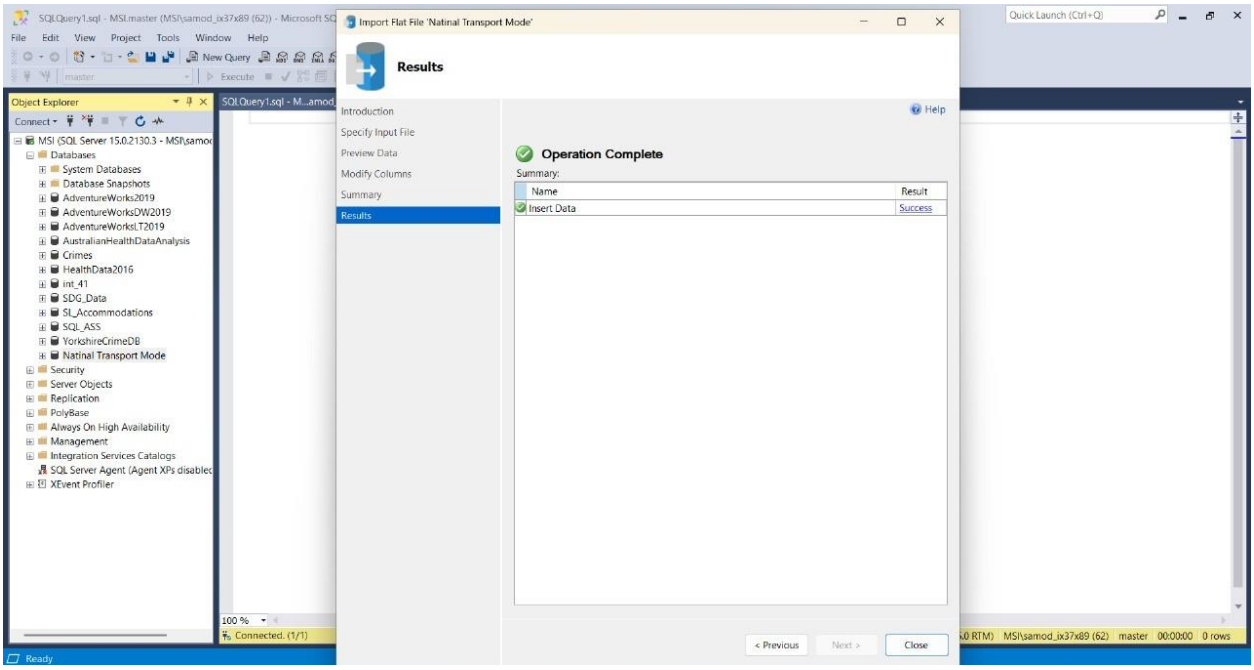


• Create the database



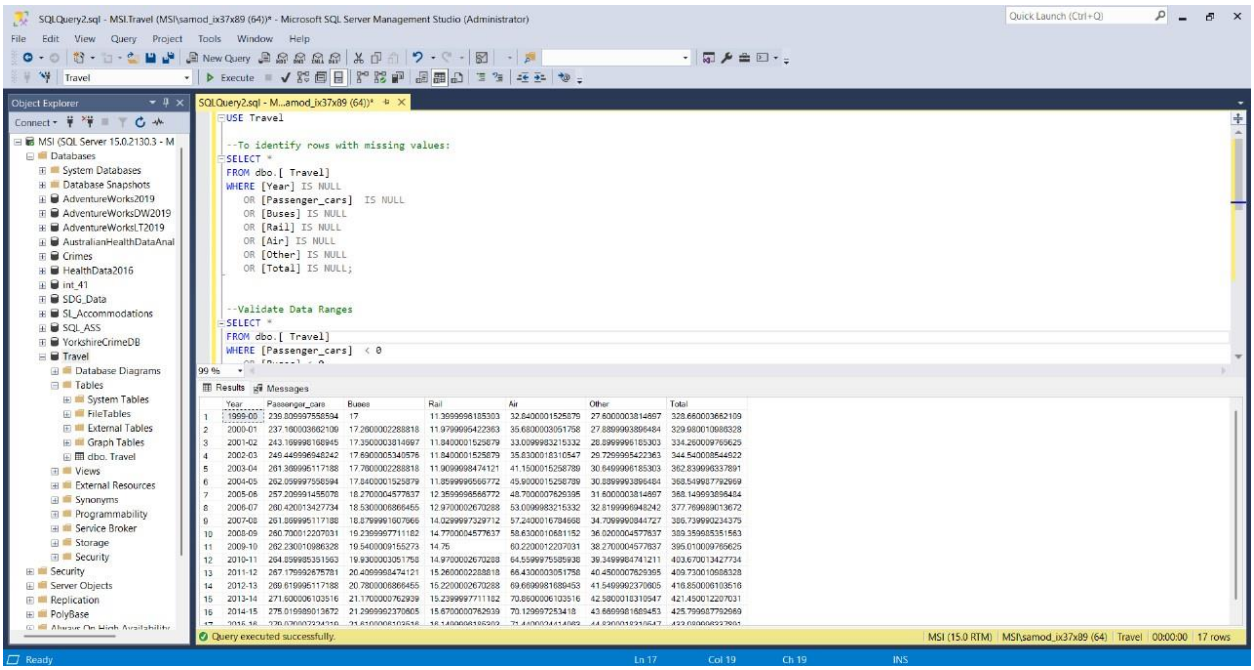


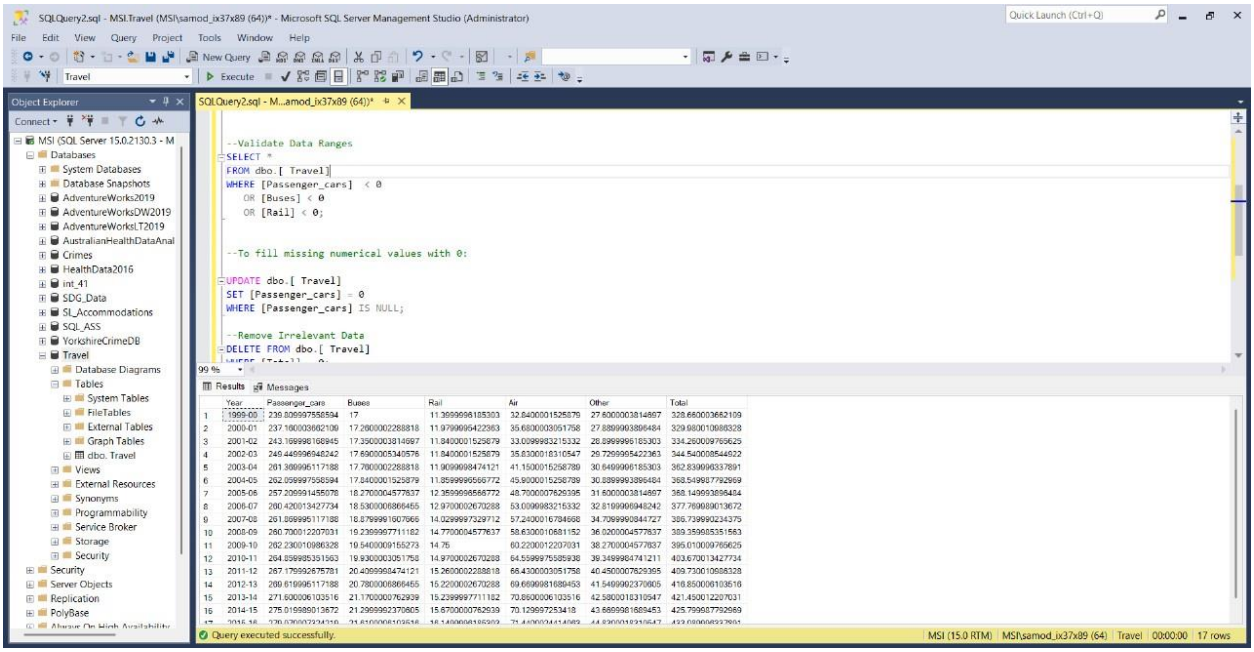
• After inserting data, click close.

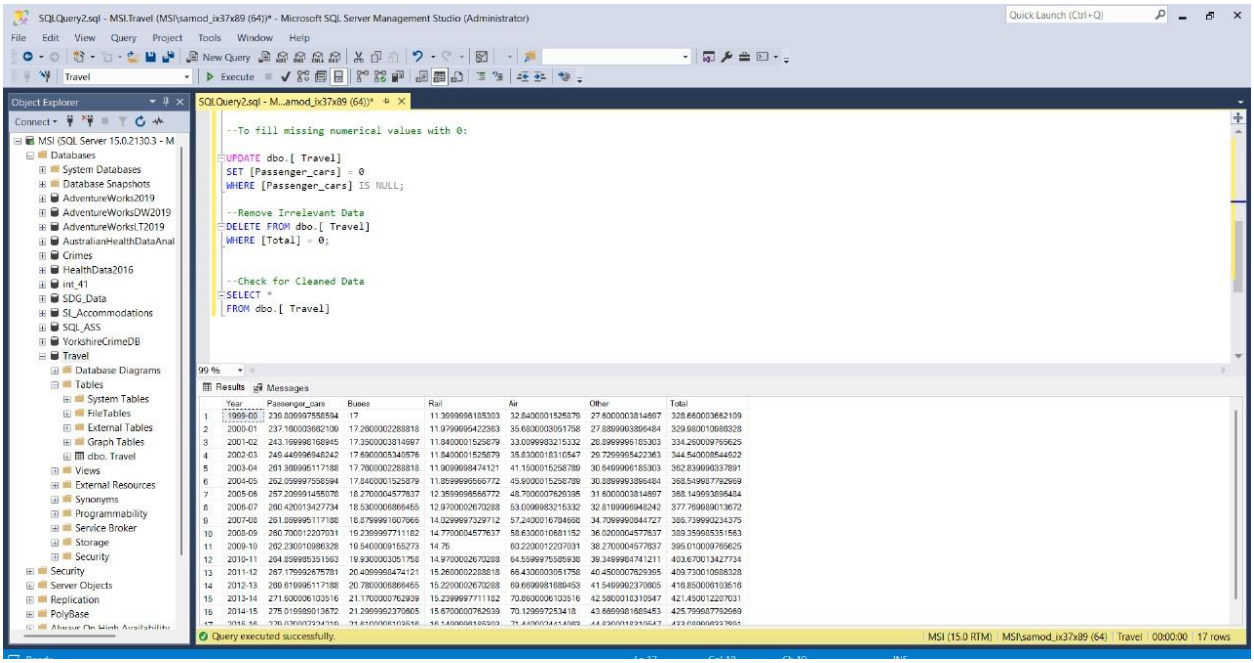


- **Create View table.**

These are the SQL queries wrote to create tables and pivot tables for creating databases.

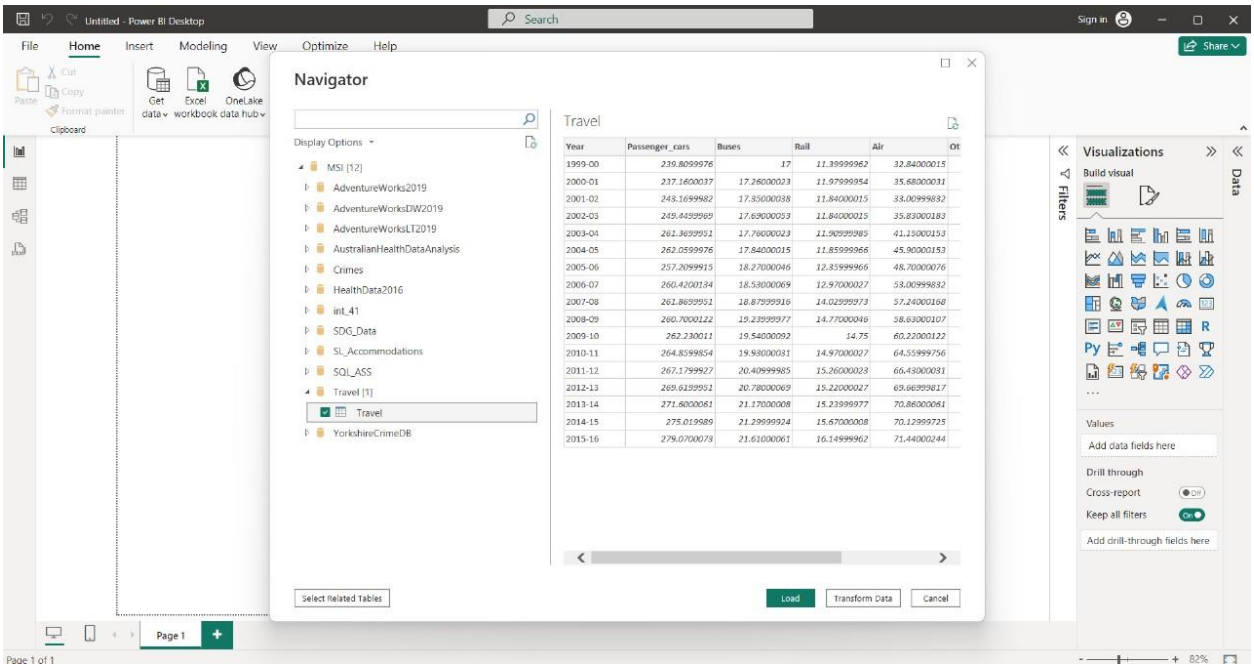




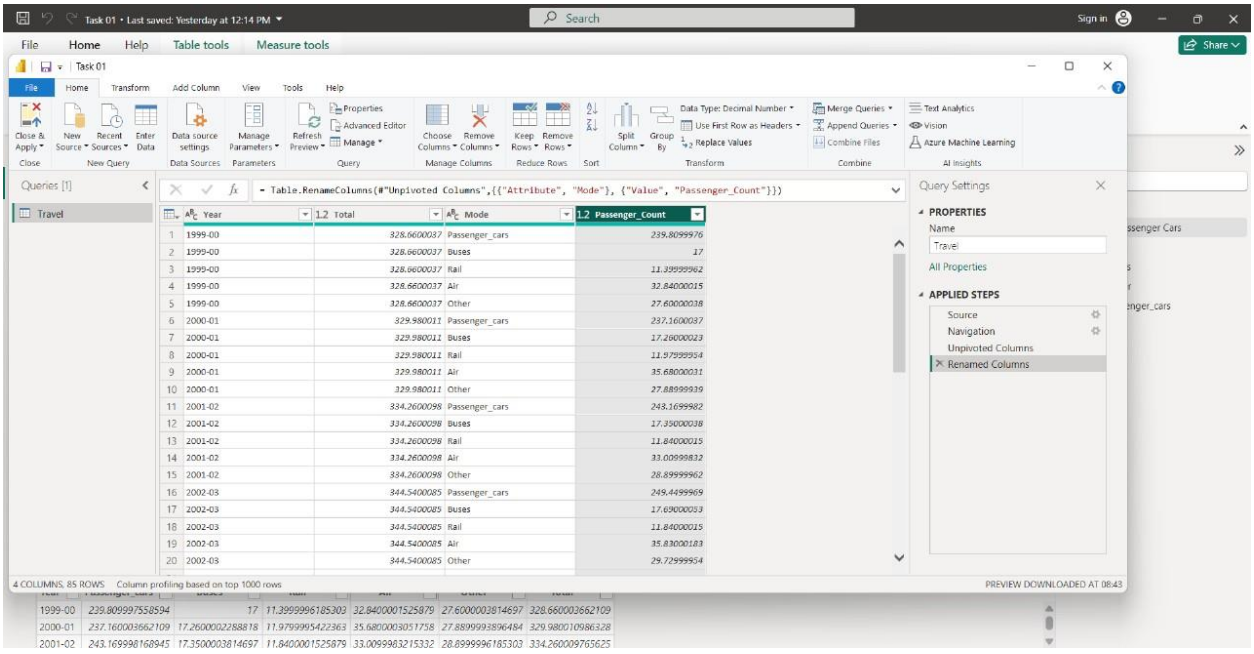

SQLQuery2.sql - MSI.Travel (MSI\samod_ix37x89 (64))* - Microsoft SQL Server Management Studio (Administrator)
Quick Launch (Ctrl+Q)
File Edit View Query Project Tools Window Help
New Query
Travel
Execute
Object Explorer
Connect
MSI (SQL Server 15.0.2130.3 - M
Databases
System Databases
Database Snapshots
AdventureWorks2019
AdventureWorksDW2019
AdventureWorksLT2019
AustralianHealthDataAnal
Crimes
HealthData2016
int_41
SDG_Data
SL_Accommodations
SQL_ASS
YorkshireCrimeDB
Travel
Database Diagrams
Tables
System Tables
FileTables
External Tables
Graph Tables
dbo.Travel
Views
External Resources
Synonyms
Programmability
Service Broker
Storage
Security
Security
Server Objects
Replication
PolyBase
SQLQuery2.sql - M...amod_ix37x89 (64))*
--To fill missing numerical values with 0:
UPDATE dbo.[Travel]
SET [Passenger_cars] = 0
WHERE [Passenger_cars] IS NULL;
--Remove Irrelevant Data
DELETE FROM dbo.[Travel]
WHERE [Total] = 0;
--Check for Cleaned Data
SELECT *
FROM dbo.[Travel]
99 %
Results
Messages
Year Passenger_cars Buses Rail Air Other Total
1 1999-00 239.899997558594 17 11.3999996185303 32.6400001525879 27.8000003814697 328.660003662109
2 2000-01 237.160003662109 17.2600002288818 11.9799995422363 35.6800003051758 27.8899993896484 329.680010986328
3 2001-02 243.199996185945 17.3500003814697 11.8400001525879 33.0299983215332 28.8999996185303 334.260009765625
4 2002-03 249.449996948242 17.6900005340576 11.8400001525879 35.6300018310547 29.7299995422363 344.540008544922
5 2003-04 261.399996117188 17.7800002288818 11.9099998474121 41.1500015258789 30.6499996185303 362.839999337891
6 2004-05 262.059997558594 17.8400001525879 11.8599996566772 45.9000015258789 30.8999993896484 368.549987792969
7 2005-06 257.209991455078 18.2700004577637 12.3599996566772 48.7000007629395 31.6000003814697 368.149993896484
8 2006-07 260.420013427734 18.5300006866455 12.9700002670288 53.0099983215332 32.8199996948242 377.769989013672
9 2007-08 261.899995117188 18.8799991607666 14.0299997329712 57.2400016784668 34.7099990844727 386.739990234375
10 2008-09 260.700012207031 19.2399997711182 14.7700004577637 58.6300010681152 36.0200004577637 389.359985351563
11 2009-10 262.230010986328 19.5400009155273 14.75 60.2200012207031 38.2700004577637 395.010009765625
12 2010-11 264.899993515563 19.9300003051758 14.9700002670288 64.5999975585938 39.3499984741211 403.670013427734
13 2011-12 267.179992675781 20.4099998474121 15.2600002288818 66.4300003051758 40.4500007629395 409.730010986328
14 2012-13 269.619995117188 20.7800006866455 15.2200002670288 69.6699981689453 41.5499992370605 416.850006103516
15 2013-14 271.600006103516 21.1700000762939 15.2399997711182 70.8600006103516 42.5800018310547 421.450012207031
16 2014-15 275.019988013672 21.2999992370605 15.6700000762939 70.1299972534180 43.6699981689453 425.799987792969
Query executed successfully.
MSI (15.0 RTM) MSI\samod_ix37x89 (64) Travel 00:00:00 17 rows

3.Dashboard Design & Implementation

- Transform the data to Power BI

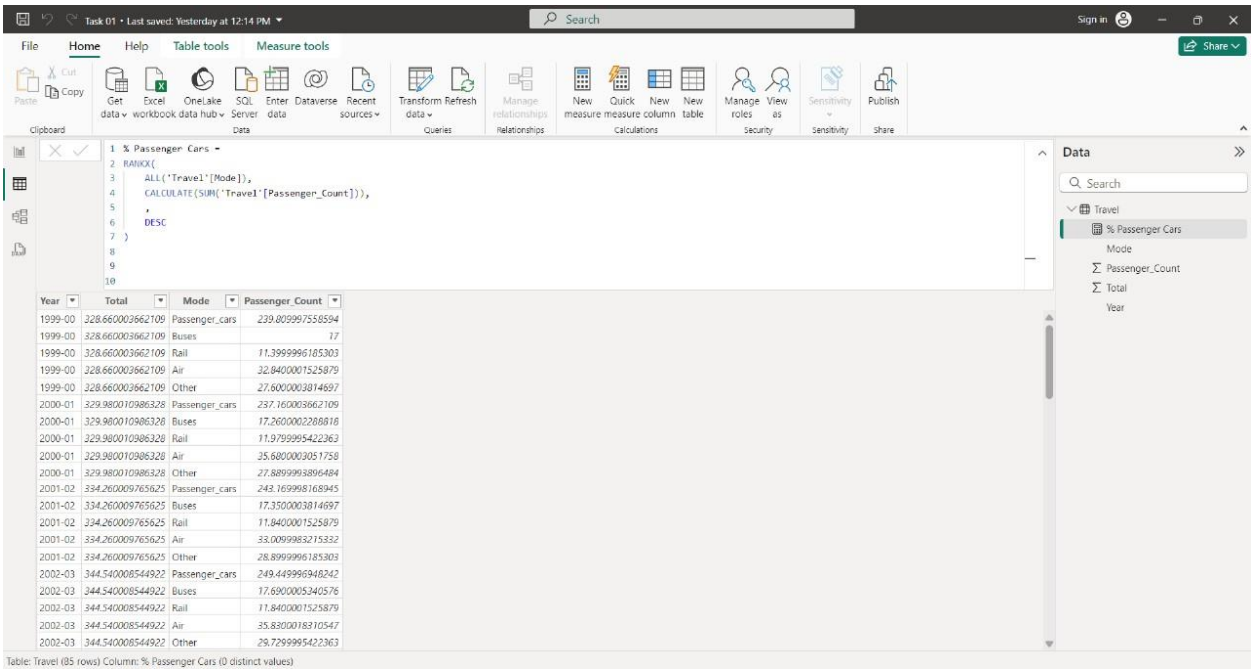


- Unpivoting Data

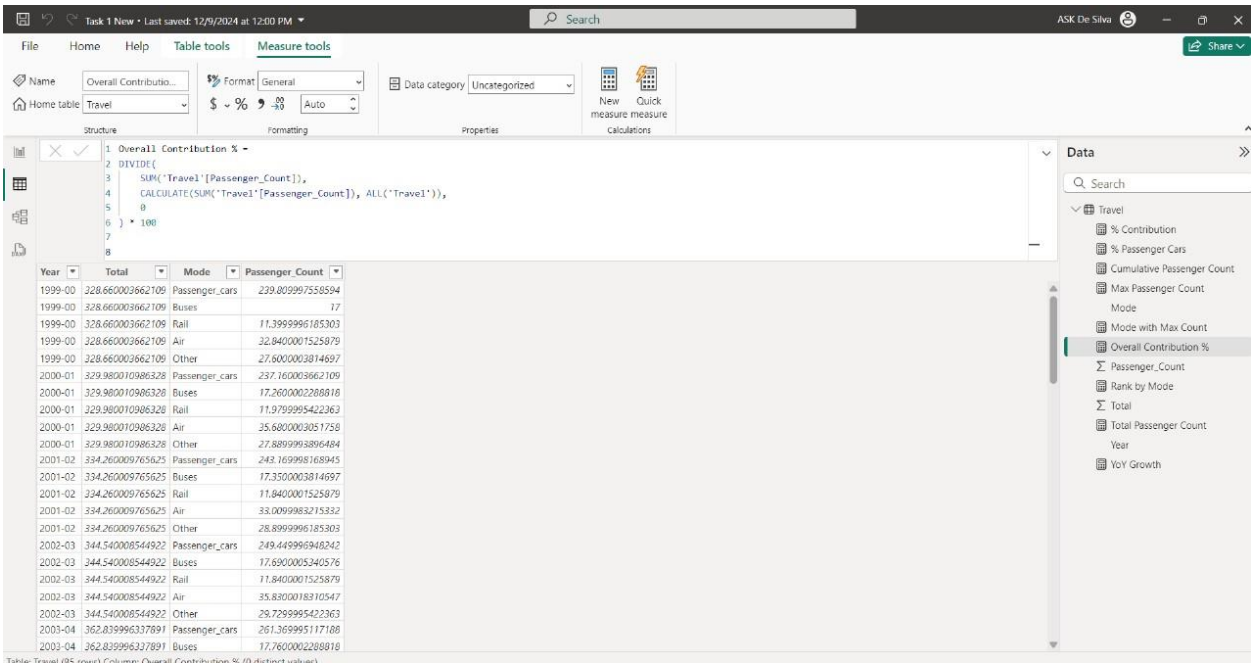


Advanced DAX measures for Power BI

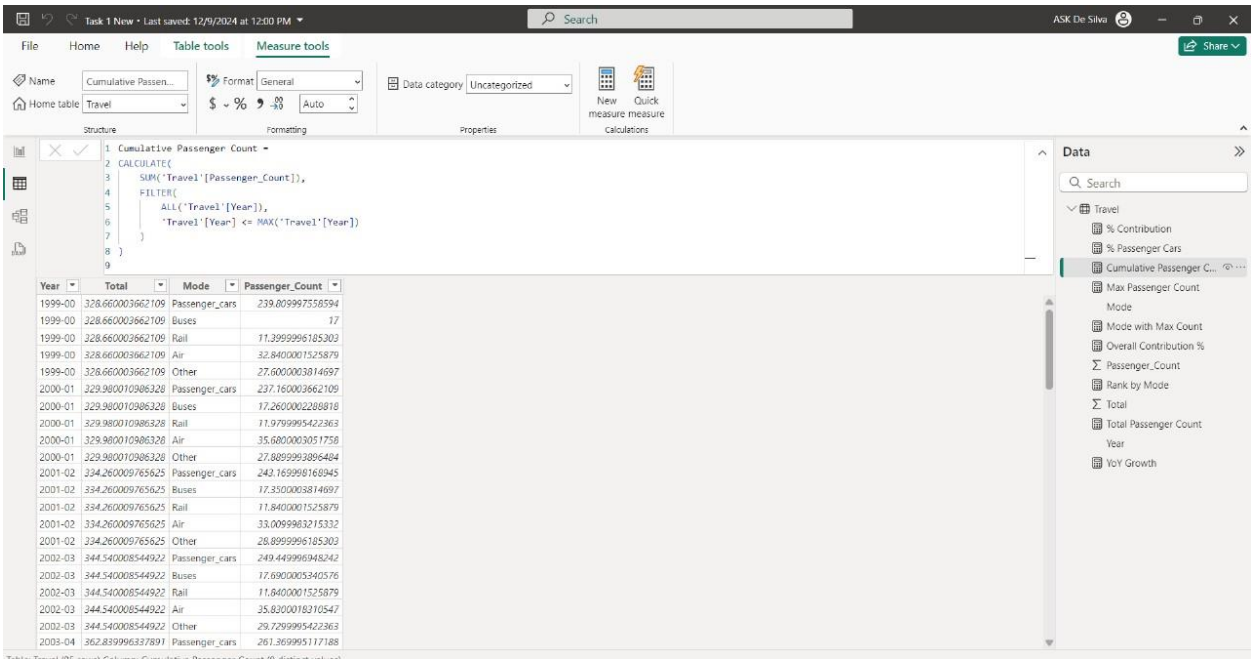
- Passenger cars



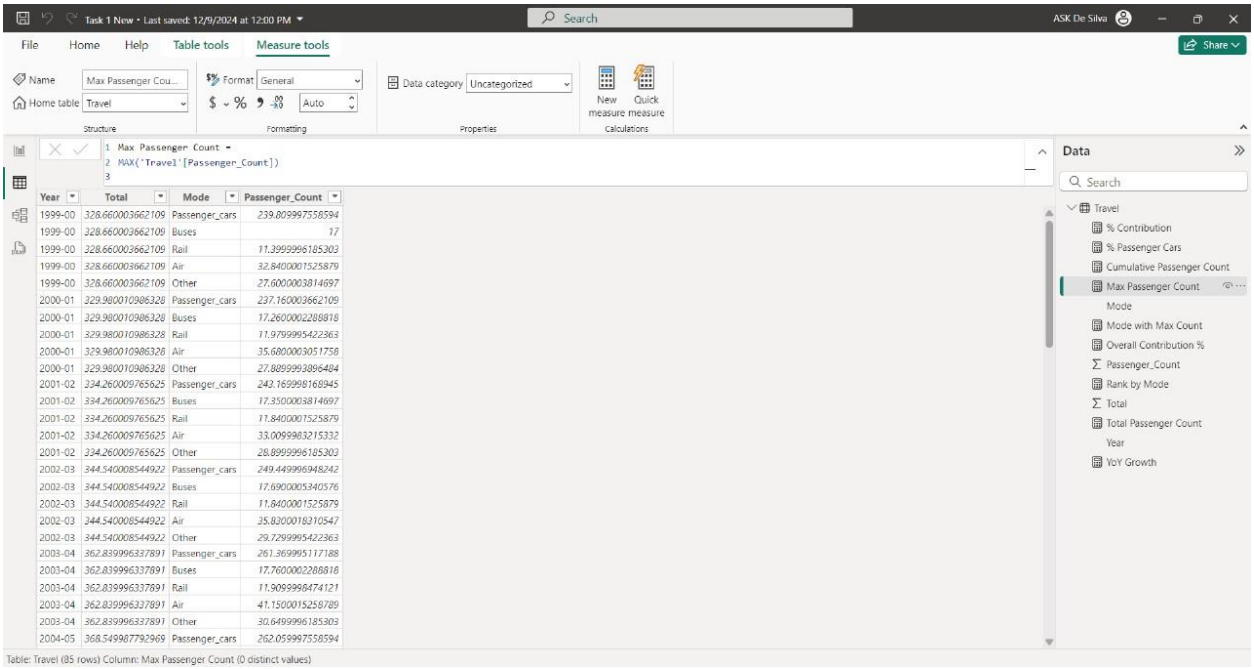
- Overall contribution.



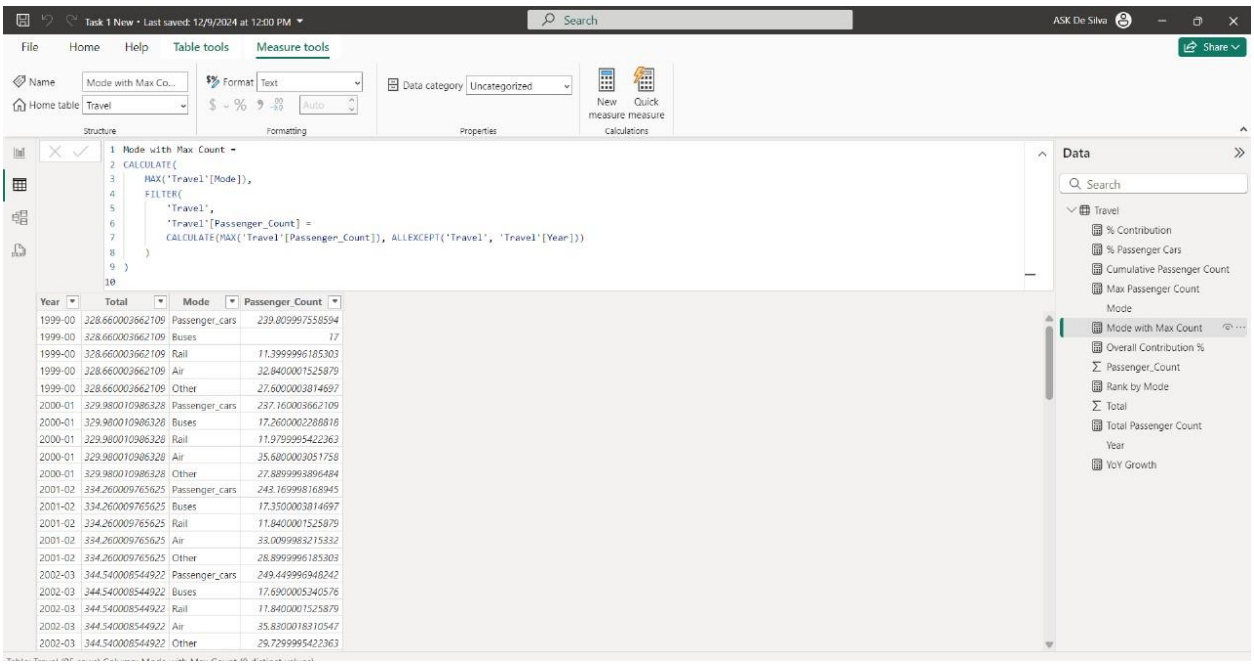
- Cumulative passenger count.



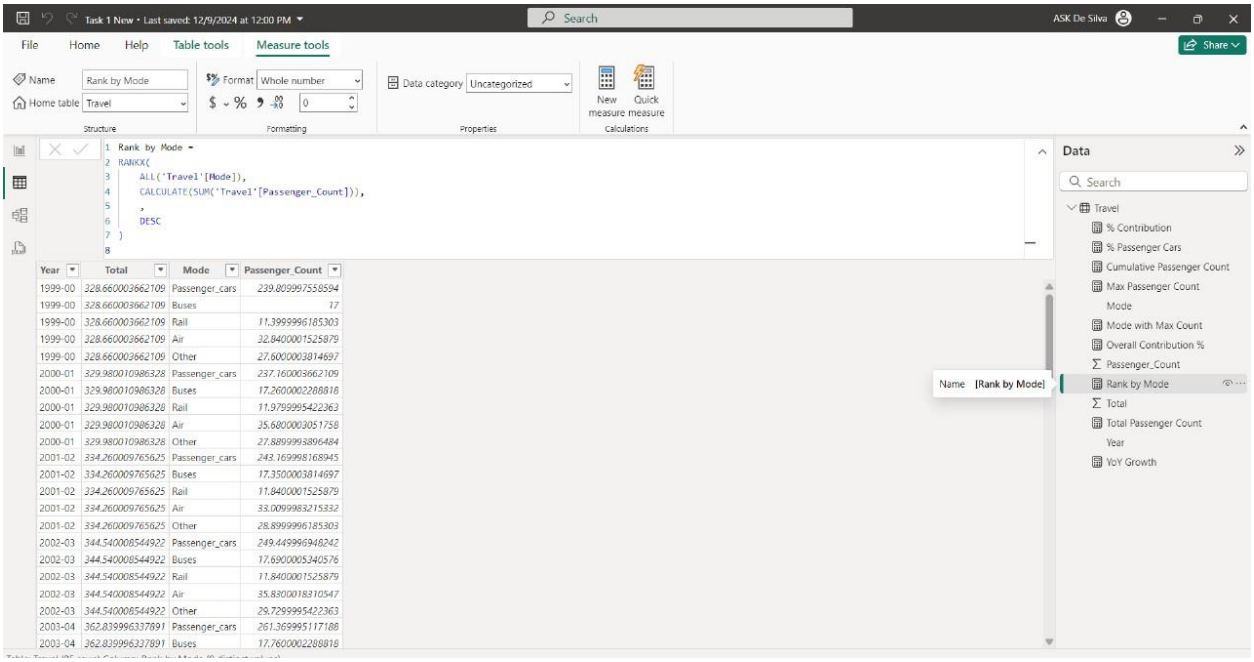
- Max passenger count.



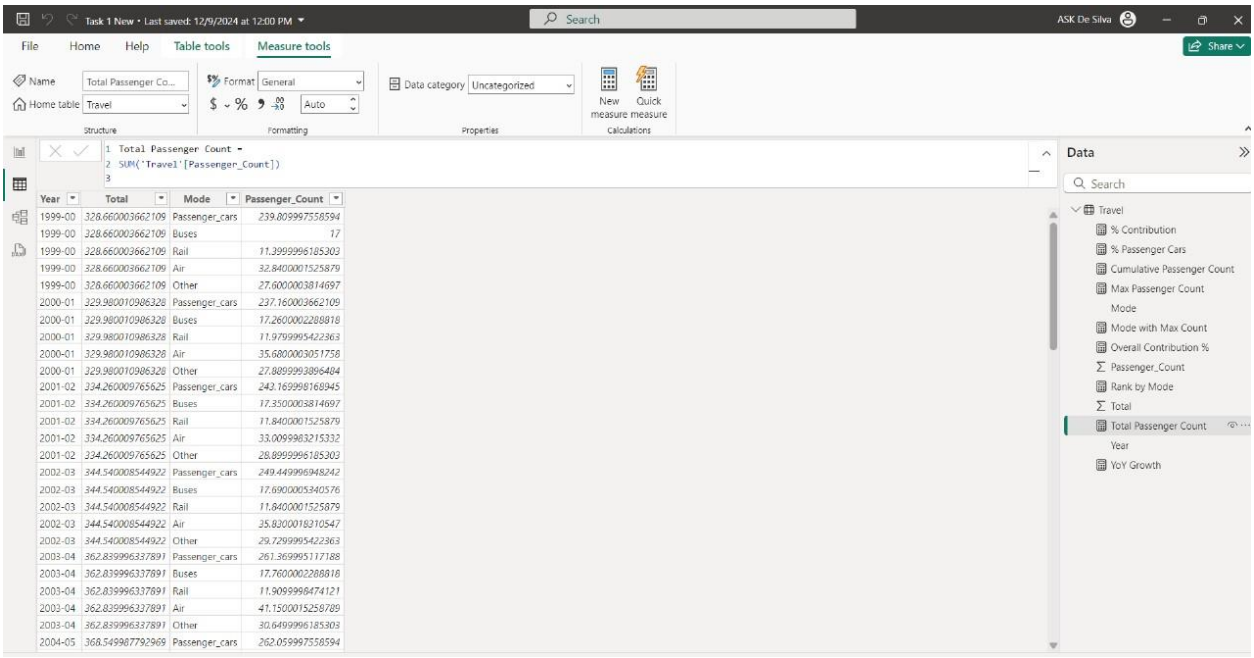
- Mode with max count.



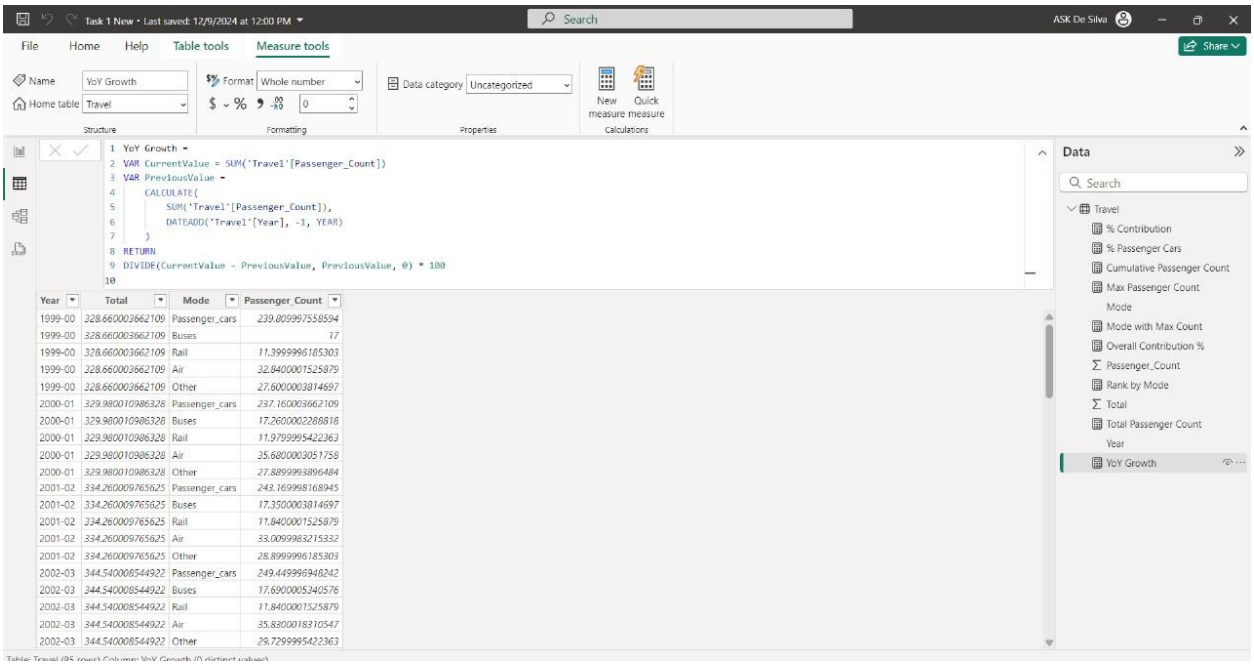
- Rank by mode.



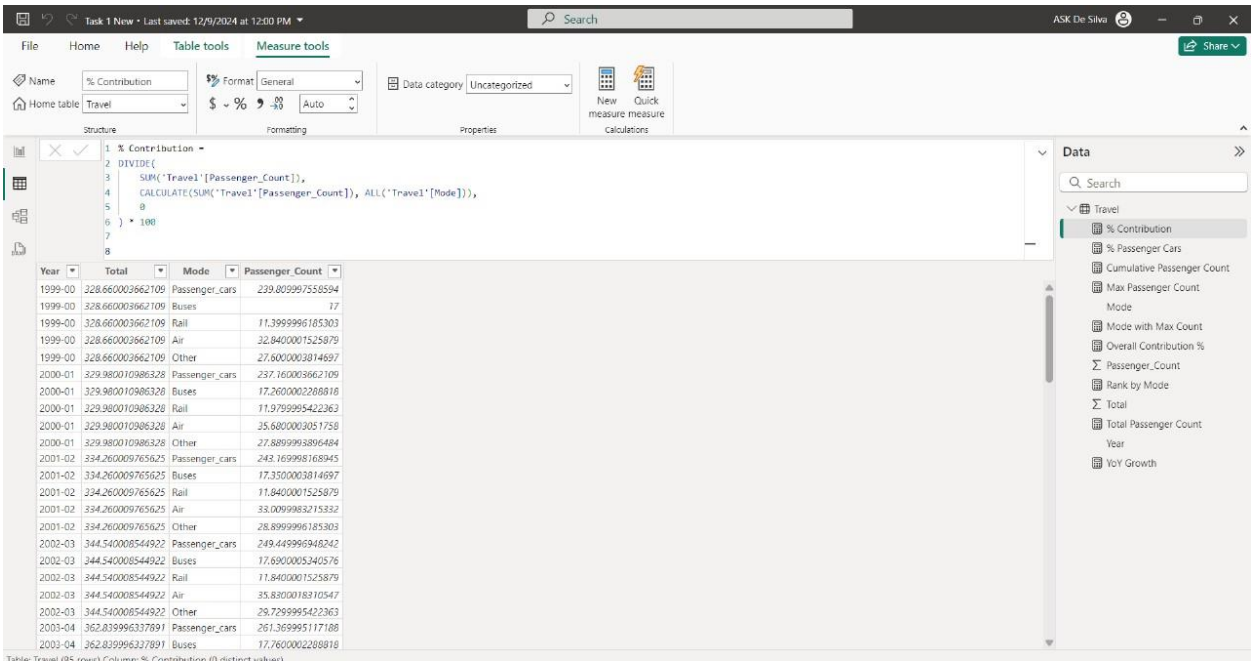
- Total passenger count.



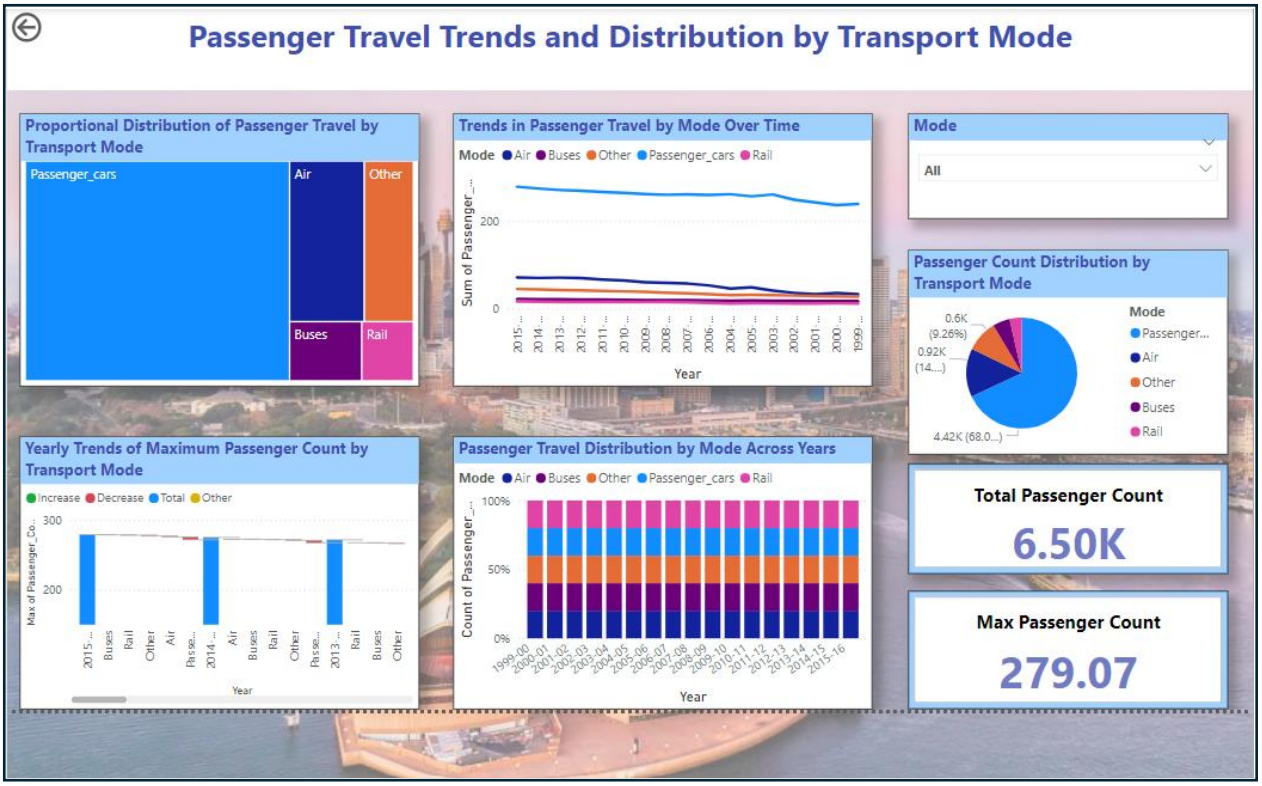
- Year-over-year Growth for passenger count.



- Contribution.



4.Final Dashboard



With a presentation of travel behaviors in different modes of transport, this dashboard will give an interactive, data-rich perspective travel trend covering the distribution, patterns with time, and insights using visual representations such as treemaps, line charts, bar charts, and pie charts. It completely permits stakeholders, analysts, and decision-makers to recognize emerging patterns in travel behavior, understand mode performance, and observe changes over years.

The transport modes analyzed are:

Passenger Cars, Buses, Rail, Air Travel, Other

The summary metrics and accompanying visual trends would allow us to make decisions based on data regarding transport infrastructure, allocation of resources, and performance evaluation.

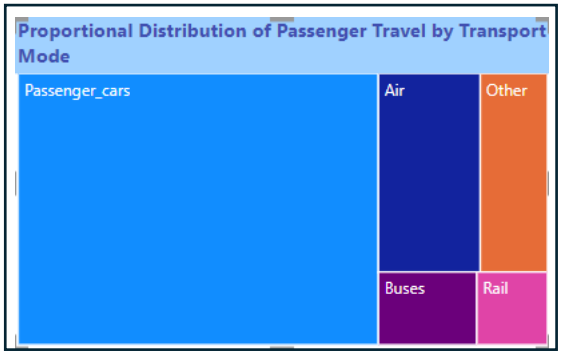
The dashboard gives good passenger travel trends and their distribution by transport modes. In this, the most dominant factor is passenger cars, while air, rail, and bus contribute much less but still significant portions. From these visualizations, interested stakeholders can:

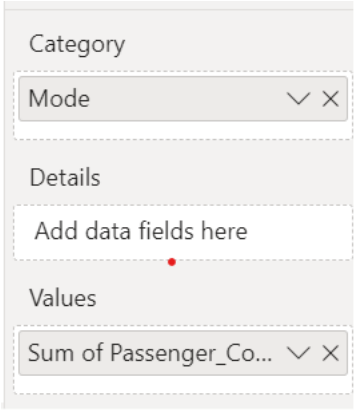
Analyze trends in years and along proportions.

Identify dominant modes, even detecting shifts over time. Make informed decisions with respect to transport policy, infrastructure development, or resource allocation.

This indicates a trend towards informing stakeholders using evidence on the understanding of the mobility behavior and improvement of performance in the transport sector.

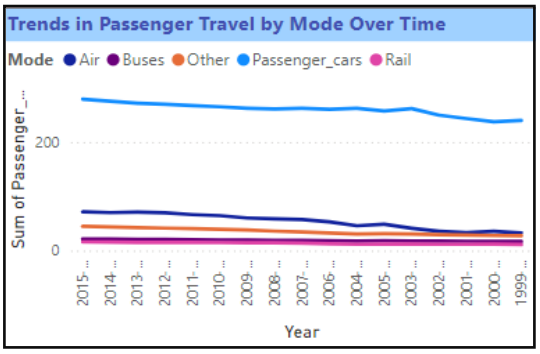
Treemap

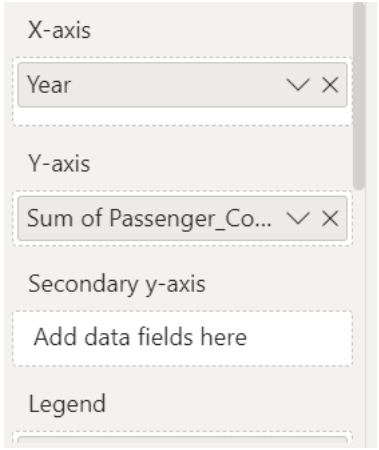




Treemap is a representation used to visualize hierarchical data in a very compact way and it's intuitive to view. One can easily compare the different proportions of hierarchy. The values are represented as rectangles in their sizes corresponding to relative importance. It would be a pointer in indicating a pattern or an outlier of categorical data.

Line chart





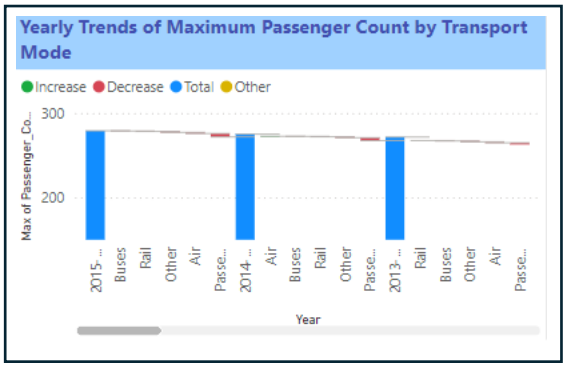
We use line charts to demonstrate trends or patterns over a specified time (time-series data). It is best for comparing multiple series of data. It highlights the progression of the data or seasonal variations. It would represent continuous data ideally.

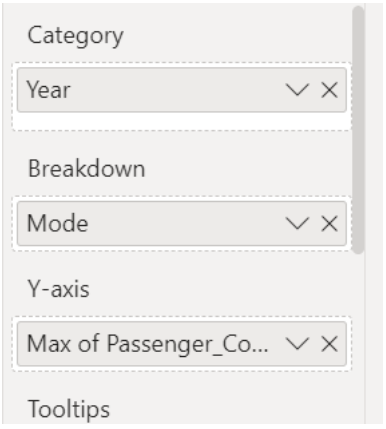
This line chart illustrates trends in passenger counts over the years (1999 to 2015) for each transport mode.

Observations include Passenger cars consistently lead with a high passenger count but show a slight declining trend over time. Other modes such as buses, rail, and air travel remain relatively stable, with marginal variations.

This trend analysis enables the identification of long-term shifts and usage patterns

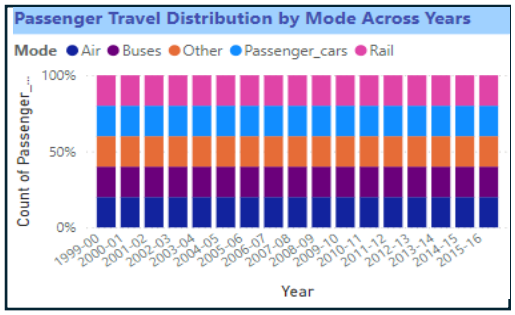
Waterfall chart

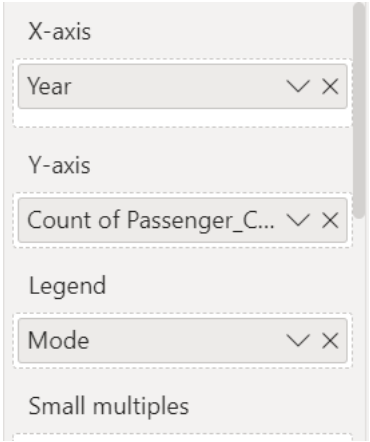




Waterfall charts describe the effect accumulated sequentially with positive and negative values. It's a great analysis tool for financial metric changes (like revenue and profit). It would shed light on the components of the total, for example, cost structure. Most suitable for direct contributions or losses within a series.

Stacked column chart

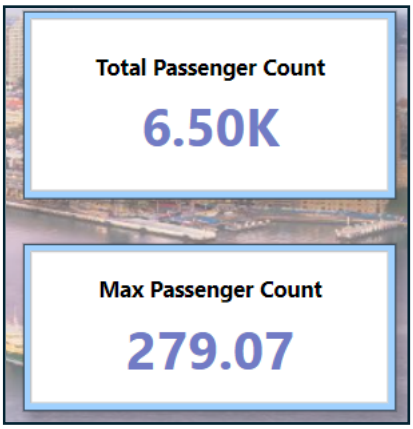


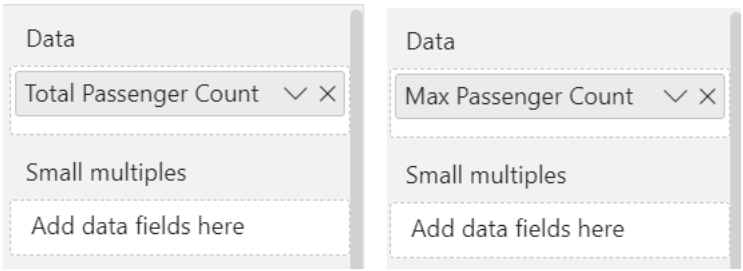


The complete data contents at the total are easy to be shown with stacked column charts, which facilitate a smooth comparison of its parts. The ideal tool to monitor time related trends, patterns that can easily relate, or trends can be better visualized. Combining multiple data sets in a single chart consumes space-efficiently, thus maximizing data analysis efficiency.

Insights include Passenger cars maintain a dominant share. Modes like Air and Rail hold a smaller but consistent portion.

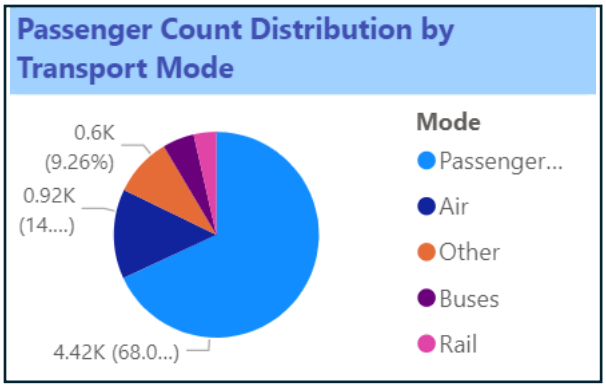
Cards

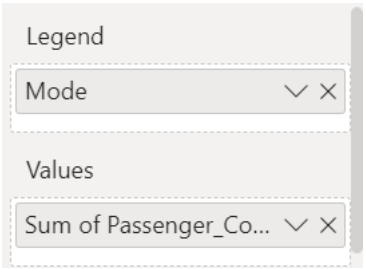




Cards are thus used to present key metrics such as totals or percentages at speed with clarity. cards strive to highlight such data points as vital. Delivery of short insights into dashboards is part of the function, the purpose for which the current and previous performance can be compared is improved decision-making, visual appeal emphasizing performance indicators that are critical.

Pie chart





Pie charts are used to show proportions of data, indicating how much of that data should be allocated to each category. They are for making effortless comparisons, emphasizing dominating segments, and simplifying usually more complicated data.

The pie chart on the right summarizes passenger count distribution as a percentage: Passenger Cars account for the majority (68%). Air and Other modes make up smaller portions at 14% and 9%, respectively.

This visualization effectively conveys the relative contribution of each transport mode to the total passenger count.

5. Key findings

Key Insights from the Dashboard

1. The Domination of Passenger's Cars

The dominance of other Motor vehicles underscored by transportation landscape trends, is one of the prominent observations. The Proportional Distribution of Passenger Travel visualization corroborates this, with passengers' cars remaining the most popular mode compared to air, buses, rail and other modes. This trend suggests high dependence on personal vehicles, at least for these users, for convenience, availability or even cost for the users.

2. Other Modes Have Very Decreasing Trends

In Trends in Passenger Travel by Mode Over Time diagram, other alternative modes which include bus and rail have been declining in the number of passengers using them through the time being analyzed. While other forms of air travel have remained in a steady state level, surging more than other forms are passenger cars. This decrease in the other modes could result from the change in the users, economic changes or even changes in investments in the relevant infrastructure systems.

3. Air Travel Stability

Air travel has been minimum flinches in the usage pattern through the years. This could be due to its role in long-distance travel since it offers almost no alternatives. However, it contributes far less than passenger cars.

4. Annual Trends in Maximum Passenger Counts

This chart of Yearly Trends of Maximum Passenger Count by Transport Mode does indeed reinforce what we know about passenger cars. Every year, passenger cars report the highest maximum passenger counts and show a consistent increase in their usage as

compared to other transport modes like buses and rail-, which display sporadic or declining patterns with no major spike in comparative rise.

5.Over the Years Passenger Distribution

The distribution of passengers by travel mode across the years is represented in a bar chart. It gives a time-series picture of the usage modes. It has been visibly reconfirmed that the greater part of travel activity over the years is passengers traveling by private cars. There are alternative modes, yet, as the division emphasizes, they have been so limited over the years that most of travel activity still featured that of the private vehicle.

The dashboards present realities on transport trends that are largely dominated by passenger cars as well as the stable or declining share of other modes. All would indicate specific areas where policies and investments could be made to foster the use of more sustainable and diverse transport alternatives. This data highlights the need for greater attention to be paid to issues of over-reliance on cars and the benefits of improved public transport systems to better balance modal distribution.

6.Challenges

Difficulties in Importing, Modeling, and Visualizing Data

Formatting Issues: Through cleaning and transformation were necessary for the SQL database data to be transfer-friendly and adapt correctly with Power BI.

Connection Setup: In connecting oneself to SQL Server regarding Safe and Reliable Connections, the right credentials and database permissions are to join.

Data Modeling Issues: To build connections, must be taken into consideration on main and foreign key dependencies that can be defined in SQL database schema.

Avoiding Circles: No loops and duplicate joints should be found in the way of keeping data integrity during analysis.

Challenges in Visualization: The Appropriate Visuals: This means having and presenting complex metrics (for example: years passenger distribution, trends in max passenger counts) in an intelligible manner.

Interactivity Issues: Uninterrupted functioning of slicers, drill through, cross filtering within dashboards.

Applying the advanced features of Power BI and making personalized visualizations.

Performance indicators for the passenger distribution and actual results have been added to give an image of performance parameters.

DAX (Data Analysis Expressions) Developed a DAX-based measure for measuring cumulated increase in transport mode as follows:

Drill through:

To provide extremely explicit research, create drill through pages enabling a user to click at summary figures and do intense investigations into specific transport mode.

Knowledge Acquired and Its Worth in Reality

Connecting Power BI to SQL Server would offer benefits like

Streamlining Data Analysis Workflow:

The Live SQL Server Integration was a clear advantage as it made the dashboards updated instantly without any need for export activities.

SQL Server also made it easy to cleanse and transform data before loading that into Power BI, producing datasets that were of a consistent and high-quality kind.

The analysis workflow became much more efficient after resolving challenges, harnessing advanced features, and integrating Power BI with SQL Server delivering ready-to-use actionable insights in real-world business scenarios on which decisions would be made.

5.Conclusion

This report is about the whole process starting from importing data and molding it into a model and then visualizing it using Power BI end to end. The sole source for this is SQL Server. Advanced features, techniques of Power BI such as Using the custom visuals, DAX, Drill through, etc., were brought into play to identify significant insights and trouble-free, streamlined workflow real-time analysis.

One major challenge of data import is data consistency and compatibility between SQL-based databases, which provide raw data entry through Power BI. While most SQL databases mostly involve issues such as formatting and redundant data, these were taken care of, either through SQL queries or Power BI transformation tools by simply cleaning and structuring the data before analysis. Another critical step involved was creating a secure connection to the SQL Server, which needed the right database credentials and permissions to reach the database but without tampering with the security.

Data modelling is somewhat of a challenge itself in terms of defining relationship between tables and breaking up all types of going around in circles. Constructing an ironclad data model was essential for any efficient, post-analytical approach. The complete understanding of the relationships between the primary and foreign keys of the tables allowed creating a data structure efficient and intuitive enough to encourage performance in Power BI.

The visualization stage has something to do about visuals to be selected carefully that will really communicate trends, comparisons and other important measures in measurement. Moreover, the clean "working" around slicers, drillthroughs, and just about any other interactive feature posed a little bit of a challenge but further improved the user experience.

Advanced Power BI functionality, primarily revolving around DAX calculations and drill through, has had great value in coming up with actionable insights. DAX or dynamic measures such as cumulative transport modes. Drill-throughs let users into specific details like individual transport profiles.

Power BI has proved effective in getting integrated with SQL Server to bring forth seamless data analysis transport modes over the years. Now, with dynamically connected dashboards updated with data, real-time decisions could very well be made. Data preparation, especially cleansing and filtering were done in SQL Server, and this prepared the data to be sent to Power BI to minimize redundancy and improve data quality before exporting it to Power BI. Such a live and integrated approach is more relevant in retail areas; the dynamic store dashboard will show passenger counts, yearly trends in transport modes.

Towards achieving that, the project also demonstrates how modern tools and techniques coalesce with the existing strength within SQL Server to bring the raw data into actionable insights via Power BI. Such insights may prove to be indispensable in driving strategic decisions and optimizing business operations in addition to securing a competitive edge in the marketplace. It builds the case for a streamlined, integration approach using data analytics in the overcoming of challenges with data import, modeling, and visualization while leveraging the advantages of advanced analytics.

Team Members:

- D/ADC/24/0003: Thesanya Lamahewa
- D/ADC/24/0013: Samoda De Silva
- D/ADC/24/0028: Thisari Perera
- D/ADC/24/0048: Dilakna Godagamage