



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

Task 01

1. Introduction

2. Exploration of Data Set

- **Review of the data set**
- **Importing data set to SQL server**
- **Create View table**

3. Dashboard Design & Implementation

- **Import the data to Power BI**
- **Advanced DAX measures for Power BI**

4. Final Dashboard

- **Dashboard Explanation**
- **Treemap -Proportional Distribution of Passenger Travel by transport mode**
- **Line chart-Trends in passenger travel by mode over time**
- **Waterfall chart- Yearly trends of Maximum Passenger Count by Transport mode**
- **Stacked column chart - Passenger Travel Distribution by Mode Across Years**
- **Cards -Total Passenger Count and Max Passenger Count**
- **Pie Chart- Passenger Count Distribution by Transport Mode**

5. Key findings

6. Challenges

7. Conclusion

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 screenshot shows the Data.gov.au website interface. At the top, there's a blue header bar with the Australian Government logo, the data.gov.au logo, and navigation links for Datasets, Organisations, About, and Login. Below the header, a search bar contains the query "Travel, by transport mode". Underneath the search bar are four filter buttons: "Any Organisation", "Any Location", "Any Format", and "Any Date". A "RESULTS (225)" section follows, featuring a highlighted result titled "National Passenger Travel, by transport mode". This result includes a link to "Sustainable Development Goals" and a brief description: "This table shows passenger kilometres for modes of transport including passenger cars, buses, rail, air, and other. Bus and rail passenger ...". It also shows the last update date as "Datasets Updated 11/09/2023" and a "Linked Data Rating" of ★★★☆☆. There's a "CSV" download link. Below this result, there's a section titled "Transport Mode Pictograms" with a link to "Department of Transport and Planning". A note states: "Model pictograms are used as a key identifier of the transport network across Victoria. They specifically assist in identifying modes of transport ...". On the right side of this section is a purple "Feedback" button.

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.

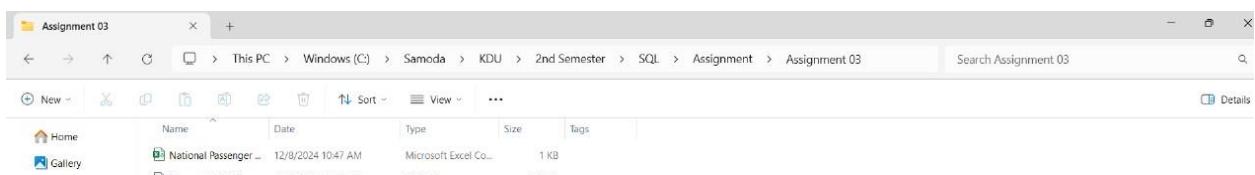
Columns:

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.

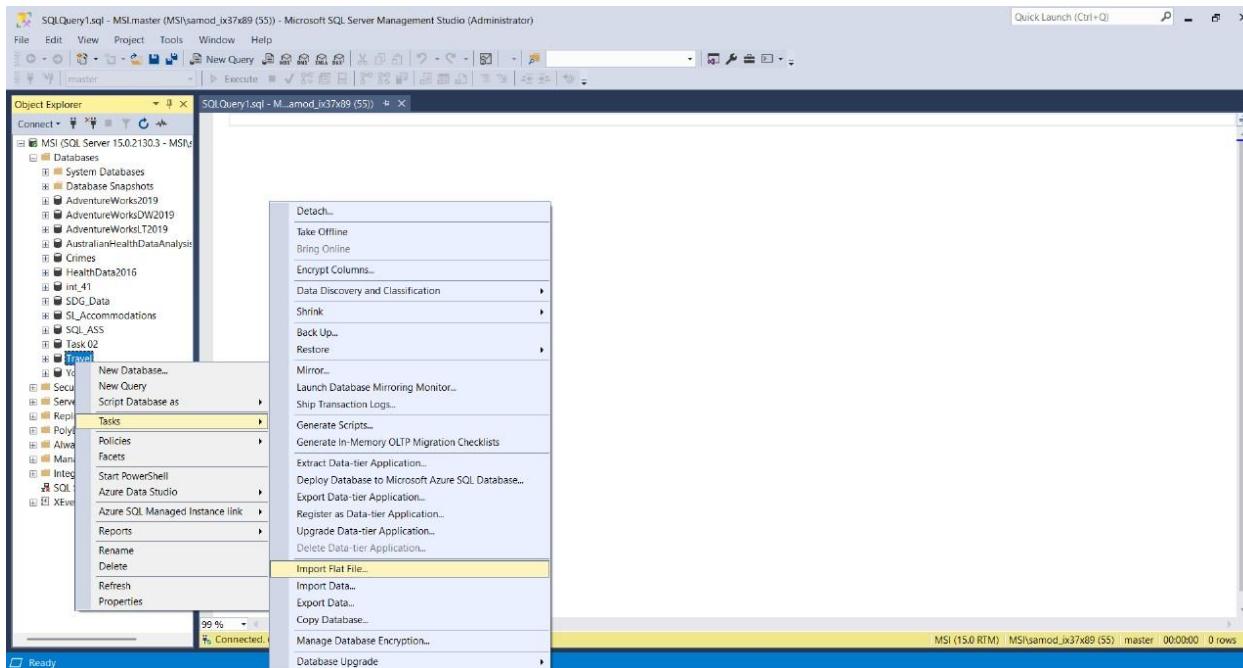
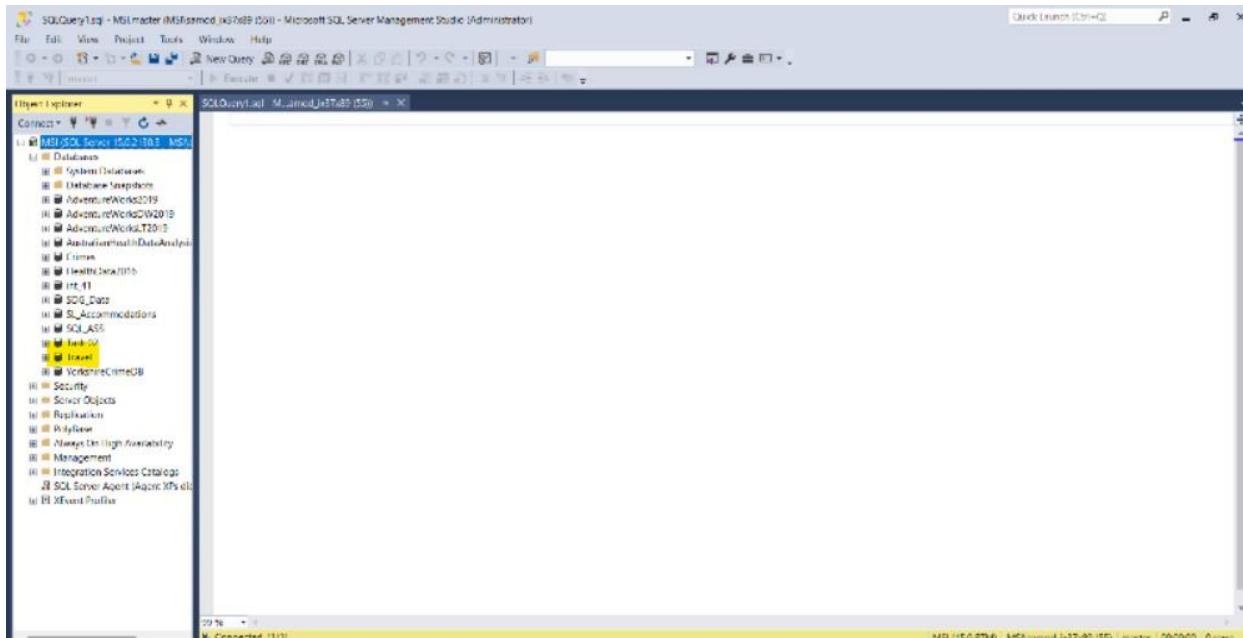
Key Observations:

- 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 as 237.16 up to 279.07 billion 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.

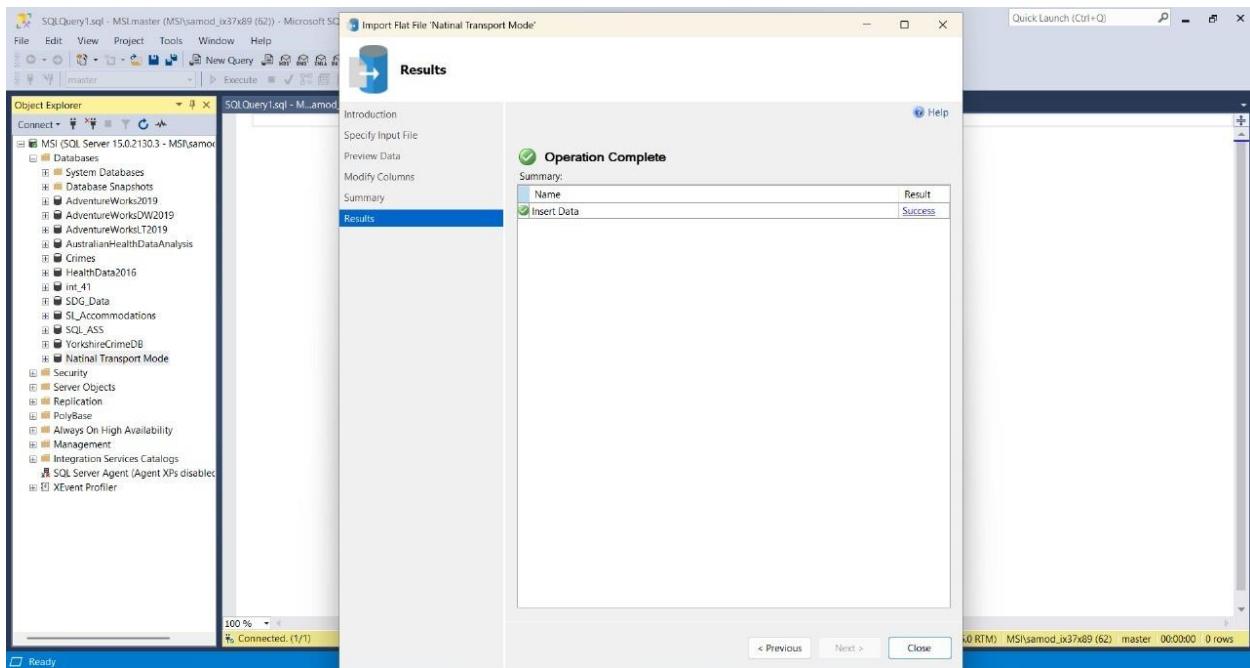
- **Importing data set to SQL Server**



- 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.

```

--To identify rows with missing values:
SELECT *
FROM dbo.[Travel]
WHERE [Year] IS NULL
OR [Passenger_cars] IS NULL
OR [Buses] IS NULL
OR [Rail] IS NULL
OR [Air] IS NULL
OR [Other] IS NULL
OR [Total] IS NULL;

--Validate Data Ranges
SELECT *
FROM dbo.[Travel]
WHERE [Passenger_cars] < 0

```

Year	Passenger_cars	Buses	Rail	Air	Other	Total	
1995-00	236.009997585894	17	11.3999998158303	32.640000152579	27.8000003814867	328.6600003652109	
2000-01	237.190003962109	17.280000288818	19.799999542283	35.680000305178	27.8899993886484	329.9800103698328	
2001-02	243.19998168945	17.3500003814867	11.840000152579	33.0099993821532	26.899999616303	334.260007795625	
4	2002-03	249.44099648242	17.6900005340578	11.840000152579	35.83000181310547	27.9299995422363	344.54000954492
5	2003-04	261.39999517189	17.780000288818	11.9099989474121	41.1500015258789	30.6499996016303	362.830009337891
6	2004-05	262.099979585894	17.8400001925879	11.859999566772	45.950001525879	30.8899993886484	368.54997792969
7	2005-06	257.200001455078	18.2700004577837	12.359999566772	48.7000007262939	31.8000003814867	388.149995389484
8	2006-07	260.190001455078	18.3500004577837	12.470000207288	50.8000007262939	32.7000003814867	399.319995389484
9	2007-08	261.0999517189	10.8799991007965	12.970000207288	50.8000007262939	32.7000003814867	399.659995389484
10	2008-09	262.2000010986328	19.540000165273	14.75	60.2200012207031	38.2700004577837	395.010009796625
11	2009-10	264.599995301563	19.830000305759	14.75	61.3200010881152	36.020004577837	398.350002031543
12	2010-11	267.759924575731	20.4099993474121	15.260002288818	64.430000205178	40.450000723096	409.730005058328
13	2011-12	269.819999751738	20.780000869465	14.560000205178	64.699998189465	41.549999237065	416.850000103616
14	2012-13	271.600006103516	21.170000762939	15.23999711112	70.8800006103516	42.5800013310547	421.4500012207031
15	2013-14	275.01998013672	21.2999952370605	15.870000762939	70.129997253418	43.6899981689453	425.799957792969
16	2014-15	275.51998013672	21.6100008105268	18.14000081852919	71.4400007414092	44.8200010319547	429.8000009257904
17	2015-16	276.01998013672	21.6100008105268	18.14000081852919	71.4400007414092	44.8200010319547	429.8000009257904

Query executed successfully.

```

--Validate Data Ranges
SELECT *
FROM dbo.[Travel]
WHERE [Passenger_cars] < 0
OR [Buses] < 0
OR [Rail] < 0;

--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]

```

Year	Passenger_cars	Buses	Rail	Air	Other	Total	
1995-00	236.009997585894	17	11.3999998158303	32.640000152579	27.8000003814867	328.6600003652109	
2000-01	237.190003962109	17.280000288818	19.799999542283	35.680000305178	27.8899993886484	329.9800103698328	
2001-02	243.19998168945	17.3500003814867	11.840000152579	33.0099993821532	26.899999616303	334.260007795625	
4	2002-03	249.44099648242	17.6900005340578	11.840000152579	35.83000181310547	27.9299995422363	344.54000954492
5	2003-04	261.39999517189	17.780000288818	11.9099989474121	41.1500015258789	30.6499996016303	362.830009337891
6	2004-05	262.099979585894	17.8400001925879	11.859999566772	45.950001525879	30.8899993886484	368.54997792969
7	2005-06	257.200001455078	18.2700004577837	12.359999566772	48.7000007262939	31.8000003814867	388.149995389484
8	2006-07	260.2000013427734	18.530000869465	12.970000207288	50.8000007262939	32.7000003814867	399.319995389484
9	2007-08	261.0999517189	18.610000207288	14.249997329712	51.9200004577837	34.020004577837	400.850000389484
10	2008-09	262.2000010986328	19.540000165273	14.75	60.2200012207031	38.2700004577837	395.010009796625
11	2009-10	264.599995301563	19.830000305759	14.75	61.3200010881152	36.020004577837	398.350002031543
12	2010-11	267.759924575731	20.4099993474121	15.260002288818	64.430000205178	40.450000723096	409.730005058328
13	2011-12	269.819999751738	20.780000869465	14.560000205178	64.699998189465	41.549999237065	416.850000103616
14	2012-13	271.600006103516	21.170000762939	15.23999711112	70.8800006103516	42.5800013310547	421.4500012207031
15	2013-14	275.01998013672	21.2999952370605	15.870000762939	70.129997253418	43.6899981689453	425.799957792969
16	2014-15	275.51998013672	21.6100008105268	18.14000081852919	71.4400007414092	44.8200010319547	429.8000009257904
17	2015-16	276.01998013672	21.6100008105268	18.14000081852919	71.4400007414092	44.8200010319547	429.8000009257904

Query executed successfully.

3. Dashboard Design & Implementation

- Transform the data to Power BI

The screenshot shows the Power BI Desktop interface. The left side features the 'Navigator' pane, which lists various data sources and tables. A table named 'Travel' is selected. The right side shows the 'Data' pane with a preview of the 'Travel' table data. The table has columns: Year, Passenger_cars, Buses, Rail, Air, and OT. The preview shows data from 1999-00 to 2015-16. Below the preview are buttons for 'Load', 'Transform Data', and 'Cancel'. The bottom of the screen shows a ribbon with tabs like File, Home, Insert, Modeling, View, Optimize, Help, and a search bar.

Year	Passenger_cars	Buses	Rail	Air	OT
1999-00	239.8099976	17	11.39999962	32.8400015	
2000-01	237.1600037	17.2600023	11.9799954	35.6800031	
2001-02	243.169982	17.3500038	11.8400015	33.0099832	
2002-03	249.4499905	17.6900035	12.8400015	35.8300183	
2003-04	261.869993	17.7600033	12.9099985	41.1500153	
2004-05	262.059976	17.8400015	11.8599966	45.8000153	
2005-06	257.8099915	18.2700046	12.8599966	48.7000076	
2006-07	260.4200134	18.5300069	12.9700027	53.0099832	
2007-08	261.869993	18.8799916	14.0299973	57.1400156	
2008-09	260.7000122	19.3399977	14.7700049	58.6300107	
2009-10	262.2300113	19.5400092	14.75	60.22000122	
2010-11	264.8599854	19.5800031	14.9700027	64.5599956	
2011-12	267.1799927	20.4099985	15.2600023	66.4300031	
2012-13	269.8129932	20.7800069	15.2200027	69.8699917	
2013-14	271.8000061	21.1700008	15.2399977	70.8600081	
2014-15	275.019989	21.2999924	15.8700008	70.1299975	
2015-16	279.0700073	21.6100061	16.1499962	71.44000244	

- Unpivoting Data

The screenshot shows the Power Query Editor interface. The left side displays a query named 'Travel' with a preview of the data. The right side shows the 'Applied Steps' pane, which includes a step named 'Travel' under 'Renamed Columns'. The main area shows the 'Table.RenameColumns' step, which is mapping columns to new names: 'Attribute' to 'Mode', 'Value' to 'Passenger_Count', and 'Passenger_Count' to 'Passenger_Cars'. The preview shows the transformed data with three columns: 'Year', 'Mode', and 'Passenger_Cars'.

Year	Mode	Passenger_Cars
1999-00	Passenger_cars	239.8099976
1999-00	Buses	17
1999-00	Rail	11.39999962
1999-00	Air	32.8400015
1999-00	Other	27.6000038
2000-01	Passenger_cars	237.1600037
2000-01	Buses	17.2600023
2000-01	Rail	11.9799954
2000-01	Air	35.6800031
2000-01	Other	28.8599962
2001-02	Passenger_cars	243.169982
2001-02	Buses	17.3500038
2001-02	Rail	11.8400015
2001-02	Air	33.0099832
2001-02	Other	28.8599962
2002-03	Passenger_cars	249.4499905
2002-03	Buses	17.6900035
2002-03	Rail	11.8400015
2002-03	Air	35.8300183
2002-03	Other	28.8599962

Advanced DAX measures for Power BI

● Passenger cars

The screenshot shows the Power BI desktop interface with a table named "% Passenger Cars". The table contains 85 rows of data with columns: Year, Total, Mode, and Passenger_Count. The formula bar displays the DAX code for the measure:

```
1 % Passenger Cars =  
2 RANK(  
3 ALL('Travel'[Mode]),  
4 CALCULATE(SUM('Travel'[Passenger_Count])),  
5  
6 DESC  
7 )  
8  
9  
10
```

The Data pane on the right shows the structure of the table, including the measure "% Passenger Cars" and its dependencies.

● Overall contribution.

The screenshot shows the Power BI desktop interface with a table named "Overall Contribution %". The table contains 85 rows of data with columns: Year, Total, Mode, and Passenger_Count. The formula bar displays the DAX code for the measure:

```
1 Overall Contribution % =  
2 DIVIDE(  
3 SUM('Travel'[Passenger_Count]),  
4 CALCULATE(SUM('Travel'[Passenger_Count]), ALL('Travel')),  
5  
6 ) * 100  
7  
8
```

The Data pane on the right shows the structure of the table, including the measure "Overall Contribution %" and its dependencies.

- Cumulative passenger count.

The screenshot shows the Power BI Editor interface with the 'Measure tools' tab selected. A new measure named 'Cumulative Passenger Count' is being defined:

```

1 Cumulative Passenger Count =
2 CALCULATE(
3     SUM('Travel'[Passenger_Count]),
4     FILTER(
5         ALL('Travel'[Year]),
6         'Travel'[Year] <= MAX('Travel'[Year])
7     )
8 )
9

```

The Data pane on the right lists various measures, including 'Cumulative Passenger Count' and 'Max Passenger Count'.

Year	Total	Mode	Passenger_Count
1999-00	328.660003662109	Passenger_cars	239.05997558594
1999-00	328.660003662109	Buses	17
1999-00	328.660003662109	Rail	11.3999996185303
1999-00	328.660003662109	Air	32.8400001525879
1999-00	328.660003662109	Other	27.600003814697
2000-01	329.980010986320	Passenger_cars	237.160003662109
2000-01	329.980010986320	Buses	17.260002268818
2000-01	329.980010986320	Rail	11.7999995422363
2000-01	329.980010986320	Air	35.680003051758
2000-01	329.980010986320	Other	27.8899993895404
2001-02	334.260009765625	Passenger_cars	243.16599916945
2001-02	334.260009765625	Buses	17.350003814697
2001-02	334.260009765625	Rail	11.8400001525879
2001-02	334.260009765625	Air	33.099983215332
2001-02	334.260009765625	Other	28.899995105303
2002-03	344.540008544922	Passenger_cars	249.449996948242
2002-03	344.540008544922	Buses	17.69000340576
2002-03	344.540008544922	Rail	11.8400001525879
2002-03	344.540008544922	Air	35.8200016310547
2002-03	344.540008544922	Other	29.299995422363
2003-04	362.839996337891	Passenger_cars	261.369995117188

- Max passenger count.

The screenshot shows the Power BI Editor interface with the 'Measure tools' tab selected. A new measure named 'Max Passenger Cou...' is being defined:

```

1 Max Passenger Count =
2 MAX('Travel'[Passenger_Count])
3

```

The Data pane on the right lists various measures, including 'Cumulative Passenger Count' and 'Max Passenger Count'.

Year	Total	Mode	Passenger_Count
1999-00	328.660003662109	Passenger_cars	239.05997558594
1999-00	328.660003662109	Buses	17
1999-00	328.660003662109	Rail	11.3999996185303
1999-00	328.660003662109	Air	32.8400001525879
1999-00	328.660003662109	Other	27.600003814697
2000-01	329.980010986320	Passenger_cars	237.160003662109
2000-01	329.980010986320	Buses	17.260002268818
2000-01	329.980010986320	Rail	11.7999995422363
2000-01	329.980010986320	Air	35.680003051758
2000-01	329.980010986320	Other	27.8899993895404
2001-02	334.260009765625	Passenger_cars	243.16599916945
2001-02	334.260009765625	Buses	17.350003814697
2001-02	334.260009765625	Rail	11.8400001525879
2001-02	334.260009765625	Air	33.099983215332
2001-02	334.260009765625	Other	28.899995105303
2002-03	344.540008544922	Passenger_cars	249.449996948242
2002-03	344.540008544922	Buses	17.69000340576
2002-03	344.540008544922	Rail	11.8400001525879
2002-03	344.540008544922	Air	35.8200016310547
2002-03	344.540008544922	Other	29.299995422363
2003-04	362.839996337891	Passenger_cars	261.369995117188
2003-04	362.839996337891	Buses	17.760002268818
2003-04	362.839996337891	Rail	11.8099996474121
2003-04	362.839996337891	Air	41.1500015259789
2003-04	362.839996337891	Other	30.6499996185303
2004-05	368.549987792909	Passenger_cars	262.059997558594

Table: Travel (85 rows) Column: Max Passenger Count (0 distinct values)

- Mode with max count.

The screenshot shows the Power BI desktop interface with a calculated column named 'Mode with Max Count'. The formula is:

```

1 Node with Max Count
2 CALCULATE(
3     MAX('Travel'[Mode]),
4     FILTER(
5         'Travel',
6         'Travel'[Passenger_Count] =
7             CALCULATE(MAX('Travel'[Passenger_Count]), ALLEXCEPT('Travel', 'Travel'[Year])))
8     )
9 }
10

```

The Data pane on the right lists various measures and calculations, including 'Mode with Max Count' under the 'Mode' category.

Year	Total	Mode	Passenger_Count
1999-00	328.660003562109	Passenger_cars	239.009997550594
1999-00	328.660003562109	Buses	17
1999-00	328.660003562109	Rail	11.399996185303
1999-00	328.660003562109	Air	32.400001525879
1999-00	328.660003562109	Other	27.600003814697
2000-01	329.980010986328	Passenger_cars	237.160003662109
2000-01	329.980010986328	Buses	17.260002288818
2000-01	329.980010986328	Rail	11.799995422363
2000-01	329.980010986328	Air	35.600003051758
2000-01	329.980010986328	Other	27.8899998996484
2001-02	334.260009765625	Passenger_cars	243.169995168945
2001-02	334.260009765625	Buses	17.3500003814697
2001-02	334.260009765625	Rail	11.8400001525879
2001-02	334.260009765625	Air	33.0099983215332
2001-02	334.260009765625	Other	28.899996185303
2002-03	344.540008544922	Passenger_cars	249.449996948242
2002-03	344.540008544922	Buses	17.6900005340576
2002-03	344.540008544922	Rail	11.8400001525879
2002-03	344.540008544922	Air	35.8200018310547
2002-03	344.540008544922	Other	29.7299995422363

- Rank by mode.

The screenshot shows the Power BI desktop interface with a calculated column named 'Rank by Mode'. The formula is:

```

1 Rank by Mode
2 RANKX(
3     ALL('Travel'[Mode]),
4     CALCULATE(SUM('Travel'[Passenger_Count])),
5     ,
6     DESC
7 )
8

```

The Data pane on the right lists various measures and calculations, including 'Rank by Mode' under the 'Mode' category.

Year	Total	Mode	Passenger_Count
1999-00	328.660003562109	Passenger_cars	239.009997550594
1999-00	328.660003562109	Buses	17
1999-00	328.660003562109	Rail	11.399996185303
1999-00	328.660003562109	Air	32.400001525879
1999-00	328.660003562109	Other	27.600003814697
2000-01	329.980010986328	Passenger_cars	237.160003662109
2000-01	329.980010986328	Buses	17.260002288818
2000-01	329.980010986328	Rail	11.799995422363
2000-01	329.980010986328	Air	35.600003051758
2000-01	329.980010986328	Other	27.8899998996484
2001-02	334.260009765625	Passenger_cars	243.169995168945
2001-02	334.260009765625	Buses	17.3500003814697
2001-02	334.260009765625	Rail	11.8400001525879
2001-02	334.260009765625	Air	33.0099983215332
2001-02	334.260009765625	Other	28.899996185303
2002-03	344.540008544922	Passenger_cars	249.449996948242
2002-03	344.540008544922	Buses	17.6900005340576
2002-03	344.540008544922	Rail	11.8400001525879
2002-03	344.540008544922	Air	35.8200018310547
2002-03	344.540008544922	Other	29.7299995422363
2003-04	362.839996337891	Passenger_cars	261.36999517108
2003-04	362.839996337891	Buses	17.600002288818

- Total passenger count.

The screenshot shows the Power BI Measure tools interface. A new measure named 'Total Passenger Count' has been created. The formula is:

```

1 Total Passenger Count =
2 SUM('Travel'[Passenger_Count])
3

```

The Data pane on the right shows the 'Travel' table with columns: Year, Total, Mode, and Passenger_Count. The table includes data from 1999-00 to 2004-05 across various modes of transport: Passenger_cars, Buses, Rail, Air, Other, and a summary row for 'Passenger_cars'.

- Year-over-year Growth for passenger count.

The screenshot shows the Power BI Measure tools interface. A new measure named 'YoY Growth' has been created. The formula is:

```

1 YoY Growth =
2 VAR CurrentValue = SUM('Travel'[Passenger_Count])
3 VAR PreviousValue =
4 CALCULATE(
5     SUM('Travel'[Passenger_Count]),
6     DATEADD('Travel'[Year], -1, YEAR)
7 )
8 RETURN
9 DIVIDE(CurrentValue - PreviousValue, PreviousValue, 0) * 100
10

```

The Data pane on the right shows the 'Travel' table with columns: Year, Total, Mode, and Passenger_Count. The table includes data from 1999-00 to 2004-05 across various modes of transport: Passenger_cars, Buses, Rail, Air, Other, and a summary row for 'Passenger_cars'.

- Contribution.

The screenshot shows the Microsoft Power BI Data Editor interface. The top navigation bar includes File, Home, Help, Table tools, Measure tools, and a search bar. The 'Measure tools' tab is selected. On the left, there's a code editor pane showing the DAX formula for the measure:

```

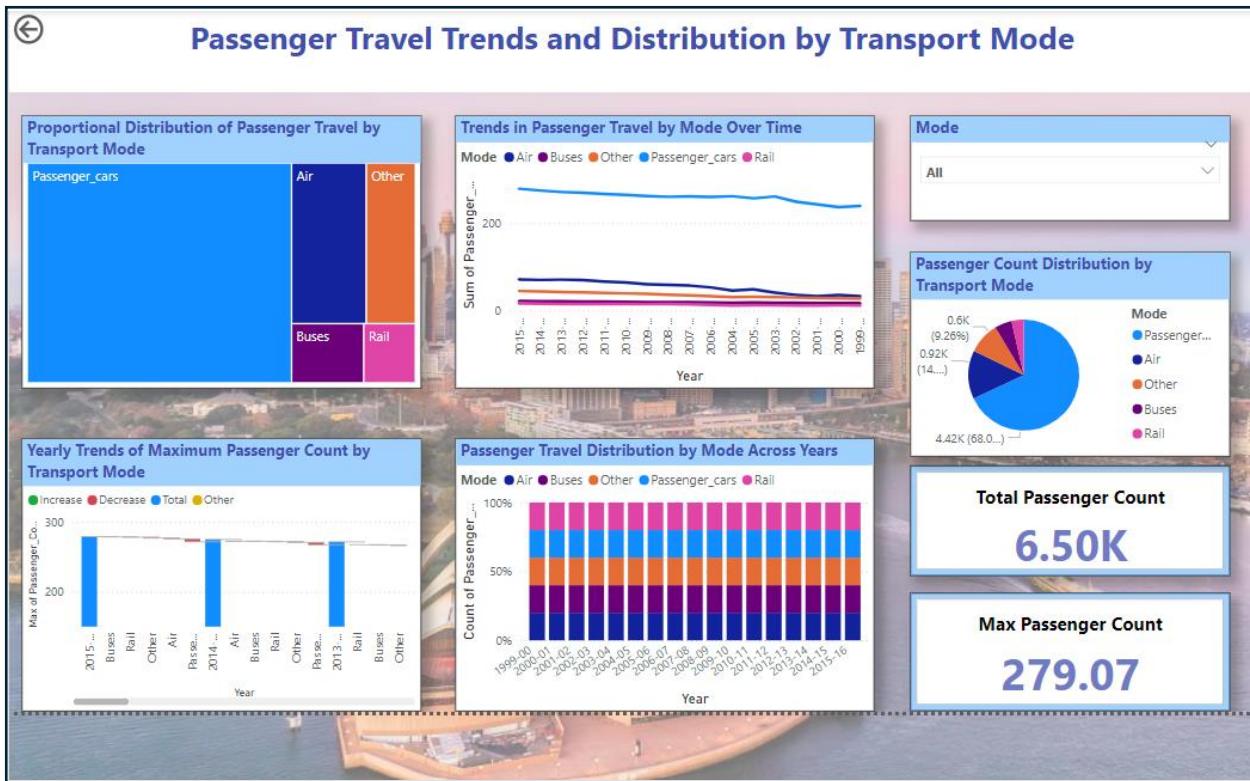
1 % Contribution =
2 DIVIDE(
3     SUM('Travel'[Passenger_Count]),
4     CALCULATE(SUM('Travel'[Passenger_Count]), ALL('Travel'[Mode])),
5     8
6 ) * 100
7
8

```

Below the code editor is a table view with columns: Year, Total, Mode, and Passenger_Count. The table displays data from 1999-00 to 2003-04, showing values for different modes of transport (Buses, Rail, Air, Other) and their corresponding passenger counts. To the right of the table is a 'Data' pane containing a list of measures and calculations, with '% Contribution' highlighted.

Table: Travel DS (Rows: 105, Columns: 4, Contribution of distinct values)

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.

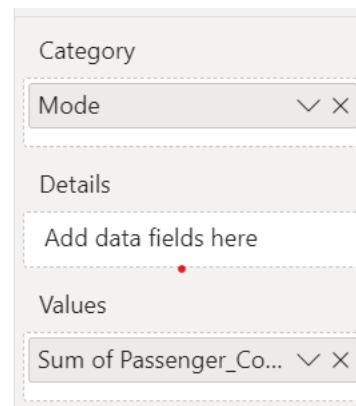
The dashboard gives good passenger travel trends and their distribution by transport modes. In this, the most dominant factor is passenger cars, while air, rail, and bus contribute much less but still significant portions. From these visualizations, interested stakeholders can:

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.

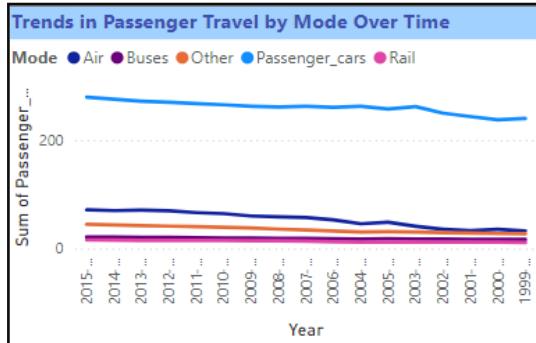
This indicates a trend towards informing stakeholders using evidence on the understanding of the mobility behavior and improvement of performance in the transport sector.

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



X-axis
Year ✓ X

Y-axis
Sum of Passenger_Co... ✓ X

Secondary y-axis
Add data fields here

Legend

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



Category
Year ✓ X

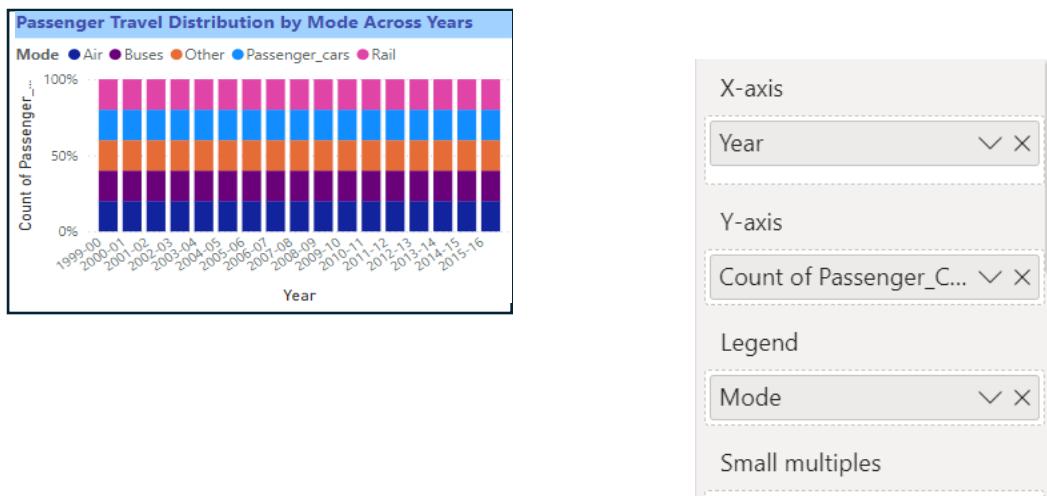
Breakdown
Mode ✓ X

Y-axis
Max of Passenger_Co... ✓ X

Tooltips

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.

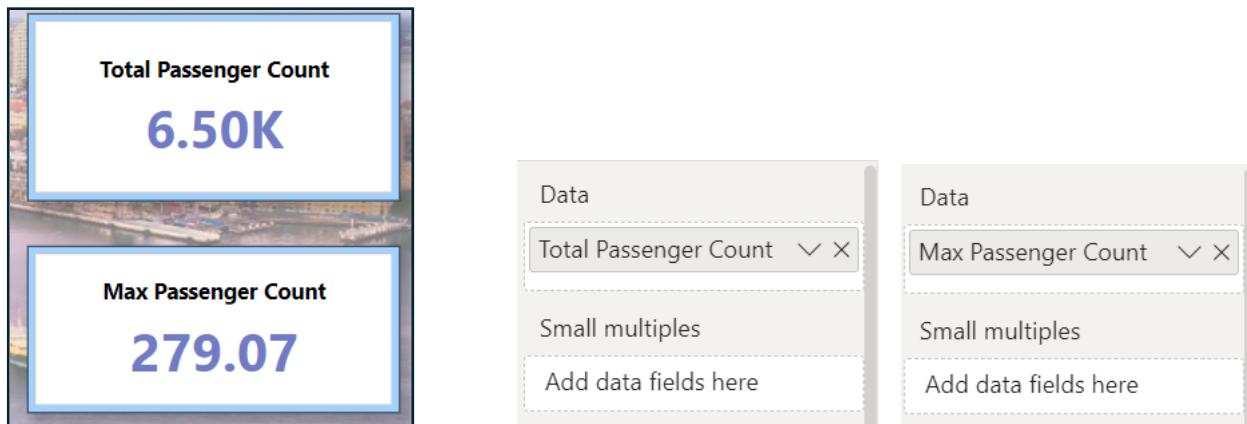
Stacked column chart



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.

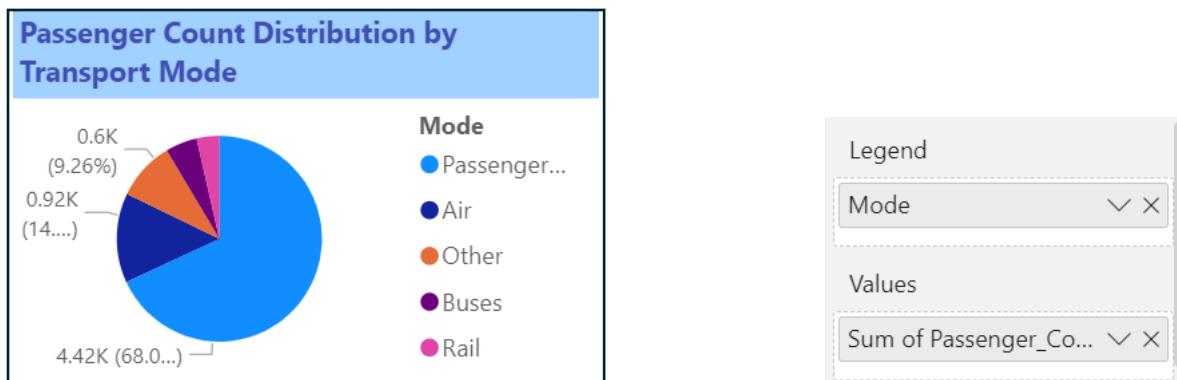
Insights include Passenger cars maintain a dominant share. Modes like Air and Rail hold a smaller but consistent portion.

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