CALGARY TRANSIT

Siva Sakthi Balan Sudhakar Sundarambal
sivasakthibalanss@gmail.com

Mentor: Dr. Junaid Qazi

TABLE OF CONTENTS

Introduction

1.1 Overview of the Open-Source Data

1.2 Purpose and Goals

1.3 Source and Dataset Selection

1.4 Expectations

Research Questions

Metrics and KPIs

3.1 Key Performance Indicators

3.2 Additional Metrics

Data Model

4.1 Overview of the Data

4.2 Entity-Relationship Diagram (ERD)

Results

5.1 Analysis Results

5.2 Key Findings

5.3 Business Implications

Conclusions

Appendices

7.1 Power BI Dashboard

7.2 Insights

LinkedIn: Siva Sakthi Balan

1. INTRODUCTION

1.1 Overview of the Open-Source Data

The Calgary Transit Ridership dataset provides insights into how people use public transportation across different stops and routes. It includes detailed information on ridership trends, fare types, trip durations, and service efficiency. By analysing this data, transit authorities can make data-driven decisions to enhance service quality, improve efficiency, and optimize the commuter experience. The dataset serves as a crucial tool for city planners and policymakers to ensure the transit system remains reliable and sustainable.

1.2 Purpose and Goals

The primary goal of this study is to leverage data analytics to improve transit operations and commuter satisfaction. The specific objectives include:

- Enhancing commuter satisfaction by optimizing transit services.
- Identifying high and low-demand routes to ensure efficient resource allocation.
- Analysing fare strategies to balance affordability and revenue generation.
- Using real-time data to enhance transit efficiency and reduce delays.

1.3 Source and Dataset Selection

The dataset for this analysis was sourced from the Calgary Open Data Portal, specifically:

1. Transit Ridership Data – This dataset contains information on passenger counts, routes, and transit usage patterns. (LINK)
2. Transit Stops Data – Provides location-based data on boarding and alighting counts at different stops. (LINK)

These datasets were selected for their ability to provide insights into ridership trends, helping transit authorities improve operations and commuter experiences.

1.4 Expectations

By analysing the Calgary Transit dataset, the study aims to:

- Identify peak ridership hours to optimize service frequency.
- Understand fare-type preferences (monthly pass vs. single ride) to enhance pricing models.
- Assess the frequency and impact of bus delays on overall ridership.
- Analyse boarding rates at different stops to improve route planning.

2. RESEARCH QUESTIONS

1. Which transit stops experience the highest and lowest boarding rates?
2. What are the peak hours for transit ridership?
3. How often are buses delayed, and what impact do these delays have on ridership?
4. How does fare type affect ridership numbers and overall transit revenue?

3. METRICS AND KPIS

To evaluate transit performance and ridership trends, the following key performance indicators (KPIs) were analysed:

1. Total Ridership – The overall number of passengers using the transit system over a given period.
2. On-Time Performance – The percentage of buses that arrive on time relative to the scheduled arrival time.
3. Total Rides – The total number of trips made by passengers.
4. Average Trip Duration – The average time taken per transit ride, helping to assess efficiency.

Additional Metrics:

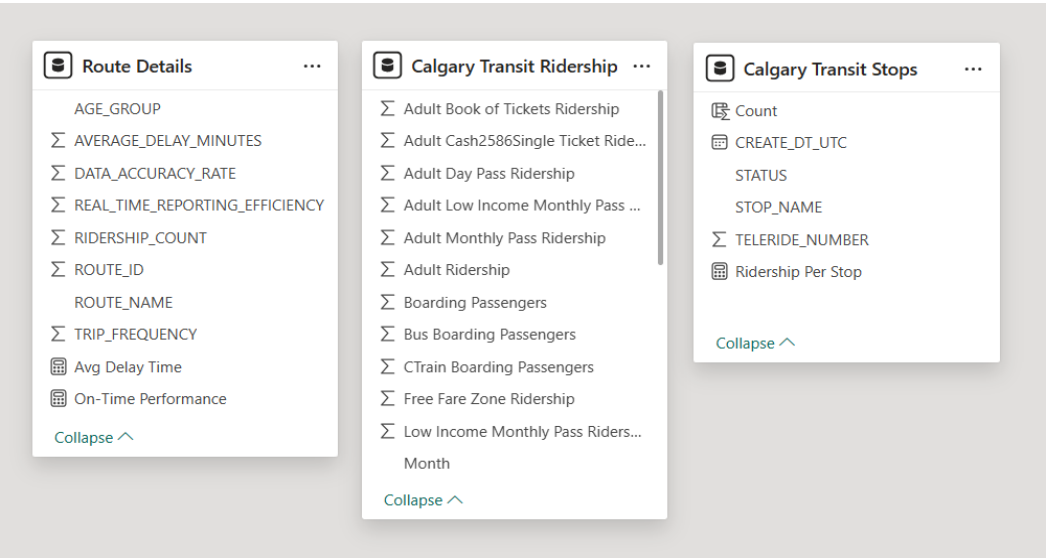
- Monthly Pass Usage – The percentage of riders using monthly passes compared to single tickets.
- Ridership by Route – Passenger volume breakdown across different routes.
- Average Delay Time – The average deviation of bus arrival times from the schedule.

4. DATA MODEL

4.1 Overview of the Data

The dataset consists of multiple attributes, including ridership counts, fare types, stop locations, and trip details. This data helps in understanding commuter behaviour, identifying service bottlenecks, and optimizing transit routes.

4.2 Entity-Relationship Diagram (ERD)



The data model includes:

- Transit Routes Table – Contains information about different bus and train routes.
- Ridership Data Table – Tracks the number of passengers using each stop and trip and differentiates between single-ride, monthly pass, and other fare types.
- Stop Information Table – Provides details on transit stop locations, usage statistics, and boarding counts.

5. RESULTS

5.1 Analysis Results

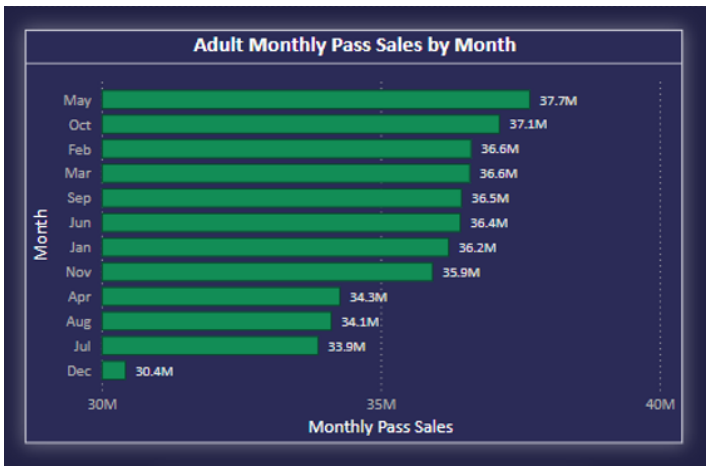
The analysis of the dataset provided several key insights:

1. Total Ridership Card

- A card visualization displays the Total Ridership for the selected date range.
- This metric helps evaluate overall transit usage trends, providing insights into commuter demand and service utilization.

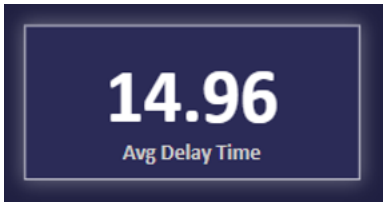
2. Monthly Pass Sales Bar Chart

- A bar graph represents monthly pass sales trends, with May and October showing the highest sales (37.7M and 37.1M, respectively).
- This data helps in understanding fare preferences and adjusting pricing strategies.



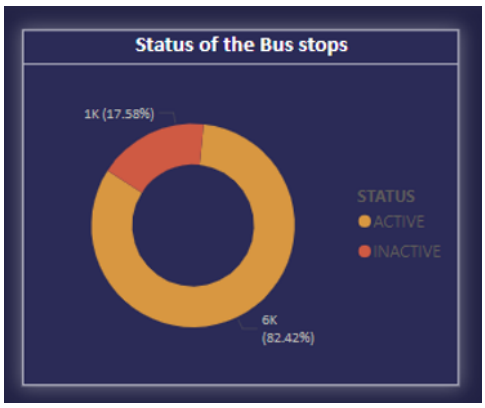
3. Average Delay Time Card

- A card visualization dynamically displays the Average Delay Time, recorded at 14.96 minutes.
- This metric provides insights into service reliability and operational efficiency.



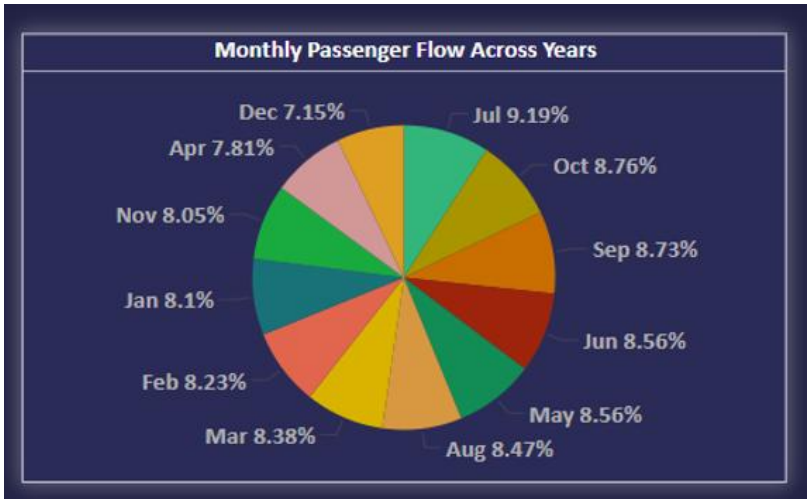
4. Active vs. Inactive Bus Stops Pie Chart

- A pie chart displays the proportion of active (82.42%) and inactive (17.58%) bus stops.
- This visualization highlights areas where transit accessibility could be improved.



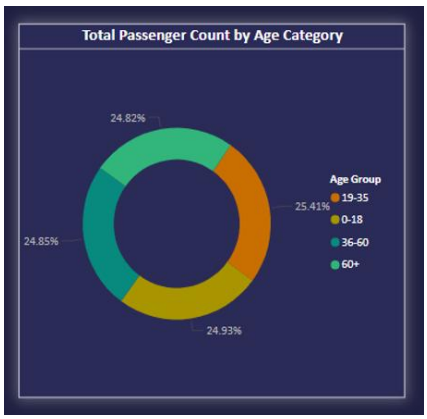
5. Monthly Ridership Line Chart

- A line graph represents monthly ridership trends, revealing peak months (July, September, and October) and low ridership months (December and February).
- This information helps in seasonal transit planning and operational adjustments.



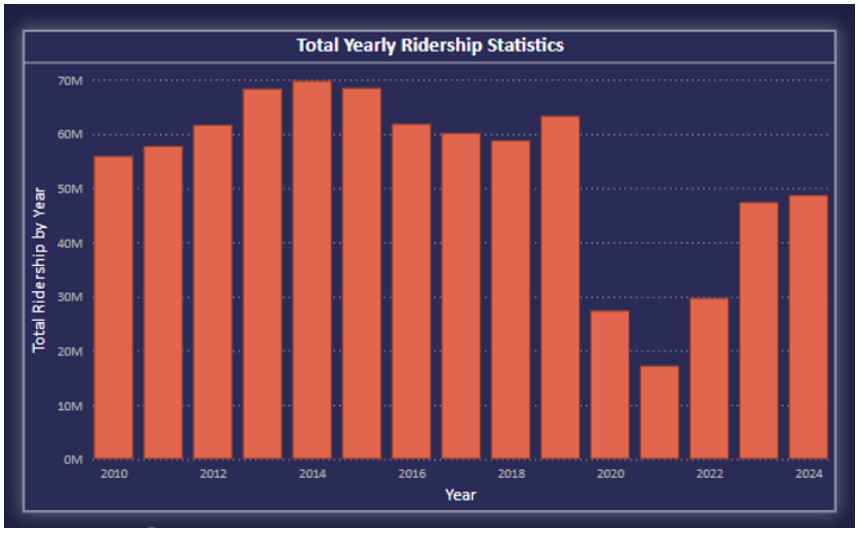
6. Ridership by Age Group Bar Chart

- A bar chart categorizes ridership based on age groups, helping identify key commuter demographics.
- This data is useful for targeted fare policies and accessibility improvements.



7. Yearly Ridership Bar Chart

- A bar chart visualizes total ridership by year, showing fluctuations over time.
- The data highlights pre-pandemic, pandemic, and post-pandemic trends, offering insights into transit recovery and long-term growth patterns.



5.2 Key Findings

- Ridership peaked in July, September, and October, while December showed lower transit usage.
- Yearly trends indicate a gradual recovery in ridership following the pandemic.
- Average delay time is 14.96 minutes, impacting service reliability.
- Monthly pass usage remains strong, with peak sales occurring in May and October.
- A majority (82.42%) of bus stops are active, but 17.58% remain inactive, indicating potential service gaps.

5.3 Business Implications

- Optimize service scheduling based on peak ridership months and high-traffic routes.
- Implement strategies to reduce transit delays and improve on-time performance.
- Evaluate inactive bus stops to determine if they should be reactivated or removed.
- Adjust fare policies to encourage more commuters to adopt long-term transit passes.

6. CONCLUSIONS

By leveraging data analytics and visualization, Calgary Transit can implement strategic improvements to enhance service quality, commuter satisfaction, and operational efficiency. Identifying ridership trends and external influences enables better resource allocation and informed decision-making for long-term transit planning. A data-driven approach ensures a smoother, more reliable, and sustainable public transportation system, ultimately benefiting the entire community.

7. APPENDICES

7.1 Power BI Dashboard

Visualizations:

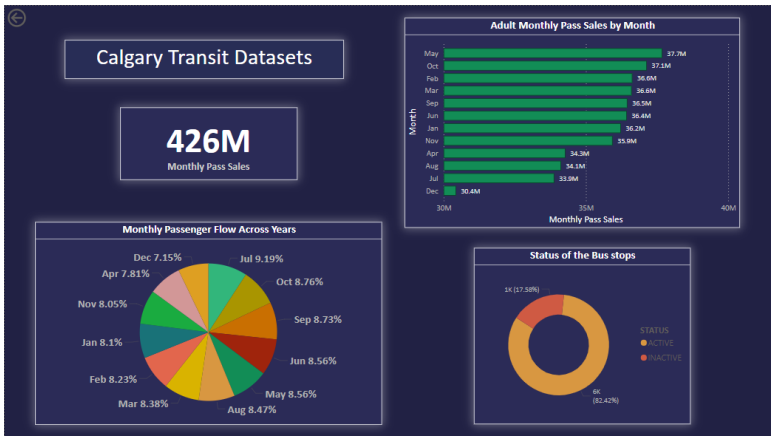
- **Total Ridership Card**: Displays the total ridership for a selected date range, providing insights into commuter demand and overall transit utilization.
- **Monthly Pass Sales Bar Chart**: Visualizes monthly pass sales trends, with notable peaks in May and October (37.7M and 37.1M respectively), helping to understand fare preferences.
- **Average Delay Time Card**: Shows the average delay time (14.96 minutes), reflecting service reliability and operational efficiency.
- **Active vs. Inactive Bus Stops Pie Chart**: Displays the proportion of active (82.42%) and inactive (17.58%) bus stops, pointing out areas where service accessibility could be improved.
- **Monthly Ridership Line Chart**: Illustrates monthly ridership trends, highlighting peak months (July, September, October) and lower ridership

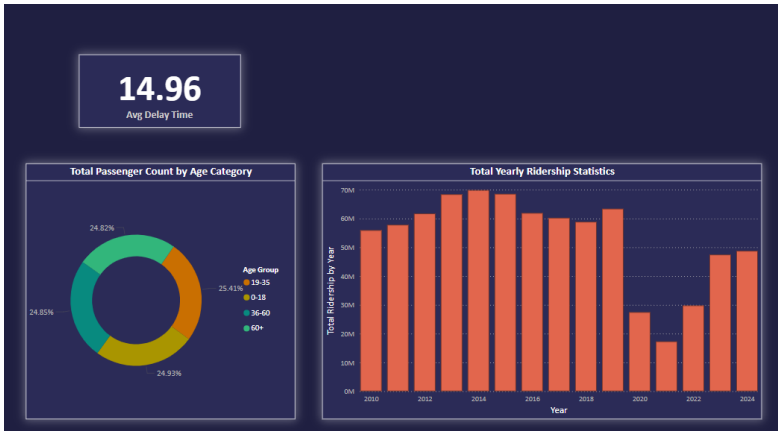
months (December and February), offering insights into seasonal variations.

- **Ridership by Age Group Bar Chart**: Categorizes ridership by age groups, helping identify key commuter demographics for targeted fare policies and service adjustments.
- **Yearly Ridership Bar Chart**: Highlights yearly ridership trends, focusing on pre-pandemic, pandemic, and post-pandemic fluctuations.

Insights:

- The **Total Ridership Card** helps identify commuter trends and can assist in optimizing service based on demand.
- The **Monthly Pass Sales Bar Chart** highlights the months with the highest monthly pass sales, aiding in fare strategy adjustments to optimize revenue generation.
- The **Average Delay Time Card** offers a clear view of service delays, signalling areas for improving on-time performance.
- The **Active vs. Inactive Bus Stops Pie Chart** emphasizes the need to evaluate underutilized bus stops and optimize service deployment.
- The **Monthly Ridership Line Chart** can guide scheduling adjustments for higher ridership months, ensuring resource allocation matches demand.
- The **Ridership by Age Group Bar Chart** helps identify the most frequent riders, enabling better targeting of fare policies and improved services for specific demographics.
- The **Yearly Ridership Bar Chart** offers insights into long-term ridership trends, aiding in recovery planning and forecasting future transit demand.





LinkedIn: Siva Sakthi Balan