

Analytics Report

1. Introduction

This report presents an end-to-end analysis of trip data, from data extraction and transformation to deriving actionable insights visualized in Power BI. The focus of this project was to analyze trip patterns, evaluate key metrics, and derive insights for decision-making. The data pipeline included loading raw trip records into SQL Server, transforming the data using dbt (Data Build Tool), and visualizing the insights on a Power BI dashboard.

2. Methodology

2.1 Data Loading

- The source data, provided in **Parquet format**, was ingested into SQL Server using Python scripts.
- Key tasks included:
 - Establishing a connection to SQL Server using **pyodbc**.
 - Filtering invalid datetime values and ensuring data consistency before insertion.
 - Batch processing to load data efficiently into the **FHV_Trip_Records** table in SQL Server.

2.2 Data Transformation

- dbt was used for data cleaning and transformation to create a robust data model.
- Key dbt models:
 - **base_trip_data.sql**: Created a cleaned and structured table, ensuring invalid or incomplete records were excluded.
 - **avg_duration_distance.sql**: Calculated average trip duration and distance segmented by the time of day.
 - **daily_trip_trend.sql**: Analyzed daily trip counts to identify trends over time.
 - **top_5_pickup_locations.sql**: Identified the most frequent pickup locations based on trip volumes.

2.3 Data Visualization

- The transformed data was imported into Power BI for visualization.
 - A dynamic dashboard was designed to showcase trip trends, average metrics, and key insights.
-

3. Data Cleaning and Transformation

3.1 Cleaning the Raw Data

- Removed rows with missing or invalid trip metrics such as `trip_miles`, `trip_time`, and `base_passenger_fare`.
- Excluded records with zero values for critical metrics to ensure meaningful analysis.

3.2 Key Transformations

1. **Datetime Conversions:**
 - Converted datetime columns into SQL-compatible formats.
 2. **Segmentation by Time of Day:**
 - Categorized trips into time slots (Morning, Afternoon, Evening, Late Night) based on pickup times.
 3. **Aggregations:**
 - Calculated averages and trends for trip durations, distances, and trip counts.
-

4. Insights and Findings

4.1 Daily Trip Trends

- **Key Observation:** The number of trips fluctuates significantly over time. Peak days were identified, potentially aligning with external factors like events or holidays.
- **Insight:** Understanding these trends can help optimize resource allocation for dispatching services.

4.2 Average Trip Duration and Distance by Time of Day

- **Key Observation:**
 - Morning trips had the shortest average duration and distance.
 - Late Night trips showed the longest durations, likely due to less traffic or longer travel routes.
- **Insight:** This segmentation highlights periods of high activity versus longer trips, aiding in operational planning.

4.3 Top Pickup Locations

- **Key Observation:**
 - Specific locations dominated the trip origins, indicating high-demand areas.
 - The top 5 pickup locations accounted for a significant percentage of all trips.
- **Insight:** These areas can be targeted for marketing or improving service efficiency.

4.4 Revenue-Driving Metrics

- Aggregated metrics like **base passenger fare**, **tolls**, and **driver pay** revealed critical revenue and cost components for the service.

5. Power BI Dashboard Visualizations

The dashboard provides interactive insights into the dataset, enabling stakeholders to explore key metrics. Below are the visualizations and their corresponding interpretations:

5.1 Daily Trip Trends

- A line chart illustrating the trip count over time, with noticeable spikes and dips.
- **Actionable Insight:** Monitor trends to predict peak days and plan for increased demand.

5.2 Average Trip Duration and Distance

- A bar chart segmented by time of day, highlighting variations in trip metrics.
- **Actionable Insight:** Optimize fleet distribution based on trip characteristics at different times.

5.3 Top Pickup Locations

- A bar chart ranking pickup locations by trip volume.
- **Actionable Insight:** Prioritize resources and advertisements in high-demand locations.

5.4 Revenue Metrics

- A summary table showing aggregated revenue components such as fares, tips, and driver pay.
- **Actionable Insight:** Assess financial performance and identify areas for cost optimization.

6. Conclusion and Recommendations

The analysis provided actionable insights into trip patterns and key metrics. Here are the recommendations based on the findings:

1. **Optimize Fleet Allocation:**
 - Focus on high-demand pickup locations and peak times for efficient resource utilization.
2. **Enhance Marketing Strategies:**
 - Target campaigns in areas with high trip volumes to attract more customers.
3. **Monitor Trends Regularly:**
 - Use the daily trip trend visualization to anticipate demand fluctuations.
4. **Improve Operational Efficiency:**
 - Adjust operations based on trip duration and distance metrics to reduce costs and improve service.

The combination of data engineering, transformation, and visualization has unlocked valuable insights, demonstrating the potential for data-driven decision-making.