# Ride-Sharing Analytics & Insights

Data Engineering and Analysis Project

Satvik Chandra



## Summary

This report presents a comprehensive analysis of ride-sharing data using a modern data stack consisting of **PostgreSQL**, **dbt**, **Apache Airflow**, and **Apache Superset**. The primary objective is to uncover key patterns in rider behavior, service performance, and operational efficiency to enable data-driven decision-making.

Data was ingested from raw CSV files and loaded into a PostgreSQL database hosted in a Docker environment. A robust transformation layer was implemented using **dbt**, enabling clean, tested, and modular data models. **Airflow** was used to orchestrate the pipeline for automated and reliable execution. The transformed data was analyzed and visualized through **Superset dashboards**.

## **XProject Architecture**

This project is built on a modular, scalable data pipeline, integrating several modern tools:

#### **Architecture Overview:**

- Data Source: Raw CSV file with ride-level transactional data
- Data Warehouse: PostgreSQL running in a Docker container
- Transformation Layer: dbt (staging → marts with tests)
- Orchestration: Apache Airflow to schedule and automate ETL
- Visualization: Apache Superset for building interactive dashboards

## Data Modeling & Pipeline

#### Source Schema

Raw data contains the following columns:

[services, date, time, ride\_status, source, destination, duration, ride\_id, distance, ride\_charge, misc\_charge, total\_fare, payment\_method]

#### Staging Transformations (stg\_rides)

- Parsed date and time to form ride\_datetime
- Casted numeric fields: distance, duration, total\_fare, etc.
- Standardized payment\_method, especially for cancelled rides
- Removed or handled nulls using dbt not\_null tests

## Exploratory Data Analysis (EDA)

### **Key Metrics:**

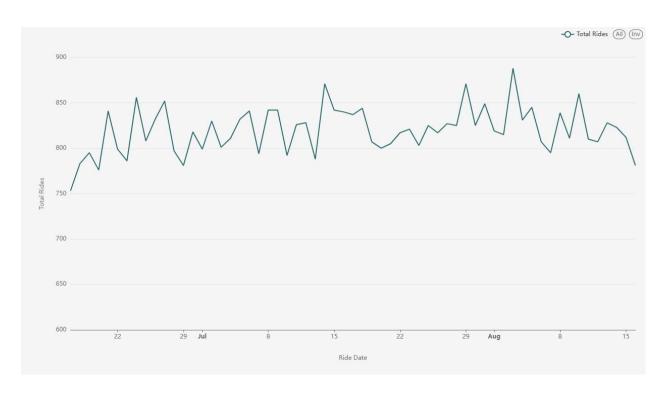
• Total Rides: 50,000

• Average Fare: \$12.34

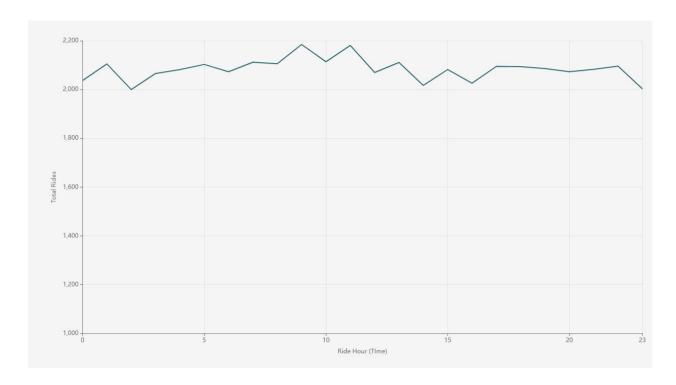
• Avg Distance: 6.5 km

• Cancellation Rate: 6.2%

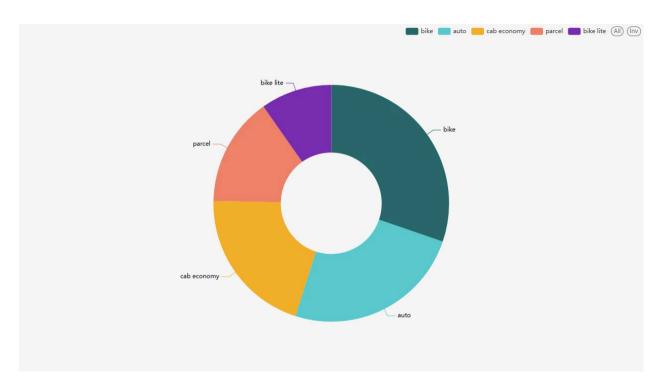
#### Daily Ride Trends



## Ride Volume by Hour of Day



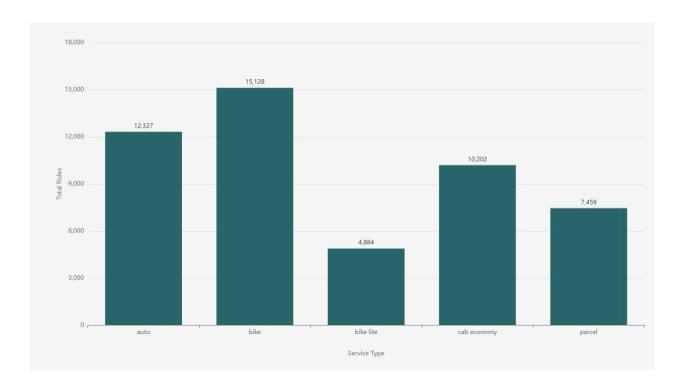
## Service Type Distribution



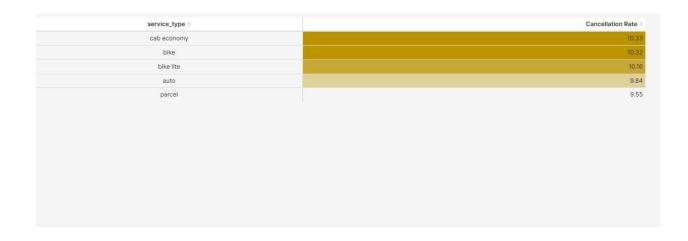
## Key Analyses and Visualizations

## **Rides by Service**

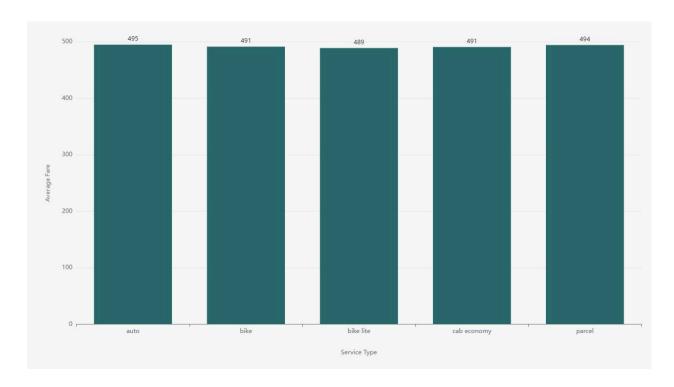
### Ride Volume per Service



### Cancellation Rate by Service

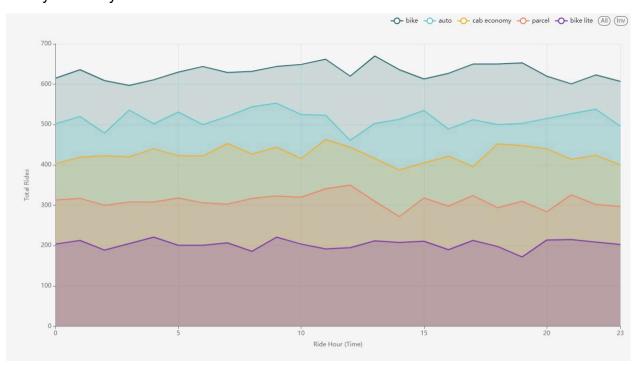


### Average Fare per Service

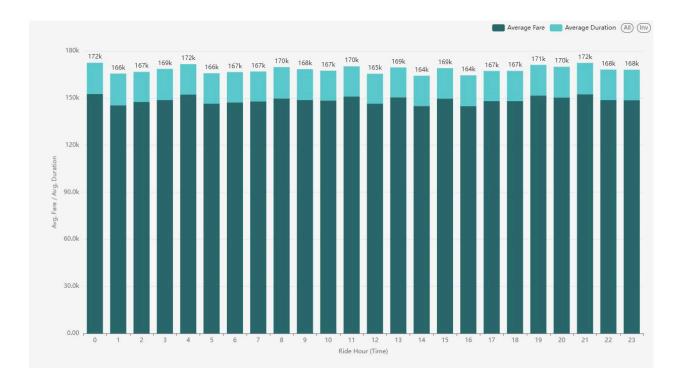


## **Hourly Trends by Service**

## Hourly Rides by Service

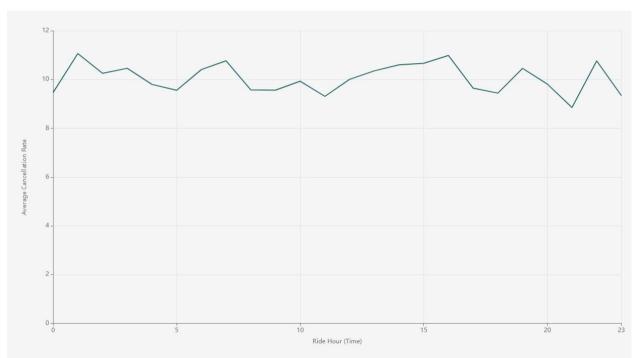


### Average Fare and Duration by Hour

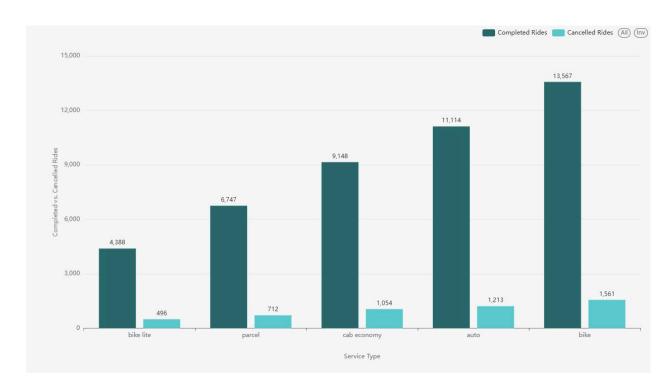


### **Cancellation Factors**

### Cancellation Rate by Hour of Day



### Completed vs. Cancelled Rides



# ✓ KPIs Summary

Total Rides	50000
Average Fare	₹492.25
Average Distance	25.52 km
Cancellation Rate	10.04 %
Top Service	Bike