

Daily Public Transport Journey Analysis Report

Introduction

Efficient management of public transport services requires **accurate forecasting of passenger demand**. Daily ridership fluctuates based on **weekdays, weekends, school terms, and special events**. This report analyzes **multi-year ridership data** for various service types and provides a **7-day forecast** using the **SARIMA (Seasonal ARIMA) model**.

Objectives:

- Analyze historical ridership trends and seasonality
- Identify key insights for operational planning
- Forecast the next 7 days of passenger demand
- Evaluate model performance using standard metrics

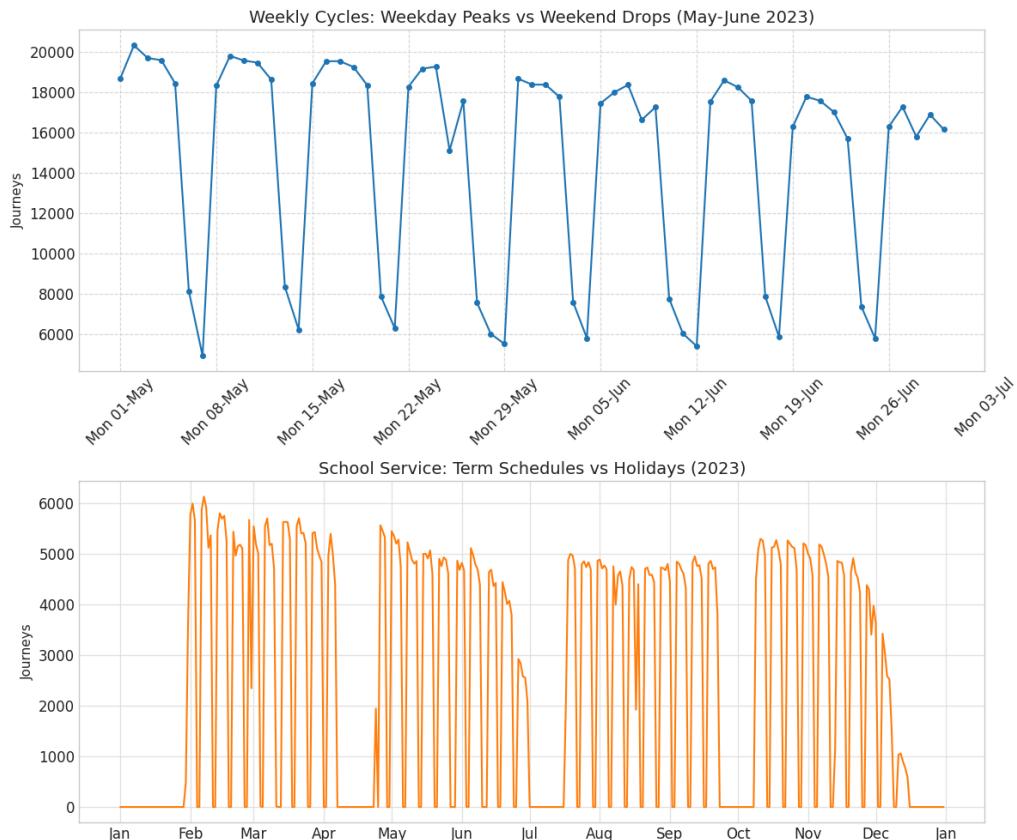
Dataset Description

The dataset contains daily passenger counts for the following services:

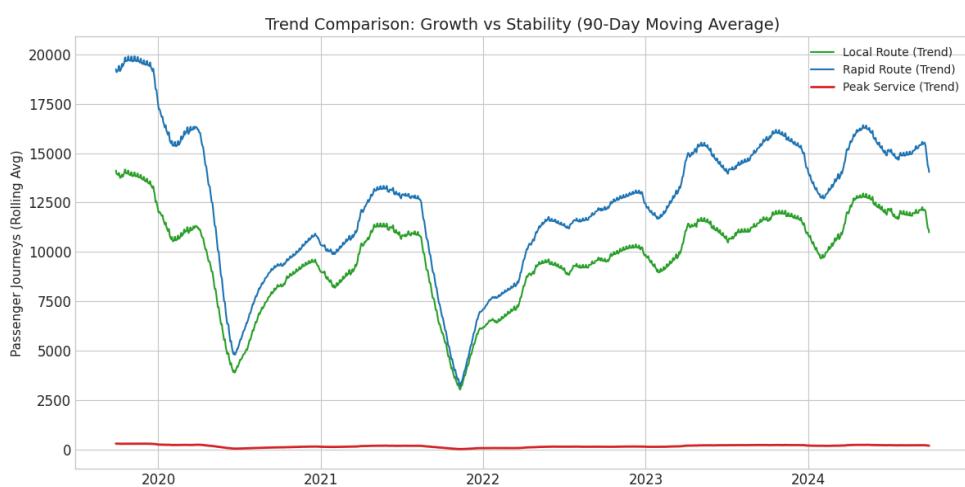
Column	Description
Date	Daily calendar dates
Local Route	Ridership on local bus routes
Light Rail	Ridership for light rail services
Peak Service	Passenger counts during peak hours
Rapid Route	Ridership on fast bus routes
School	Passenger counts on school-focused services

Key Insights

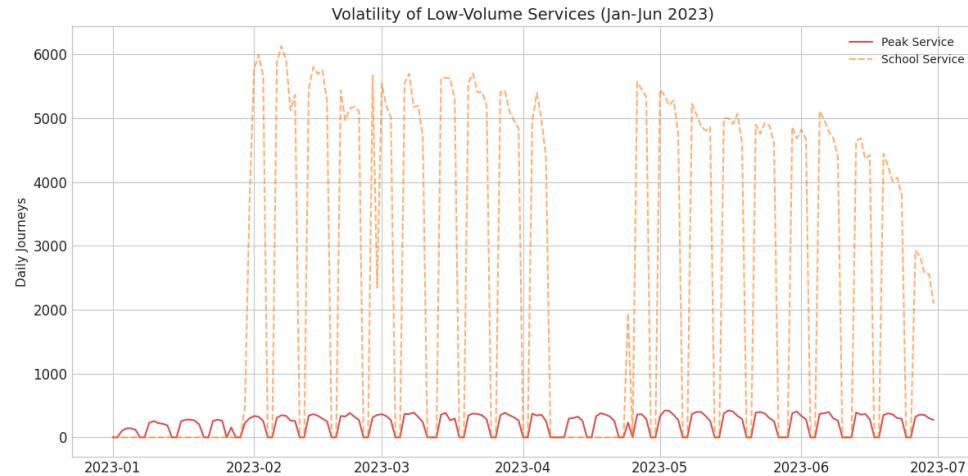
1. Weekly Cycles: Ridership peaks on weekdays, drops on weekends; school services follow term schedules.



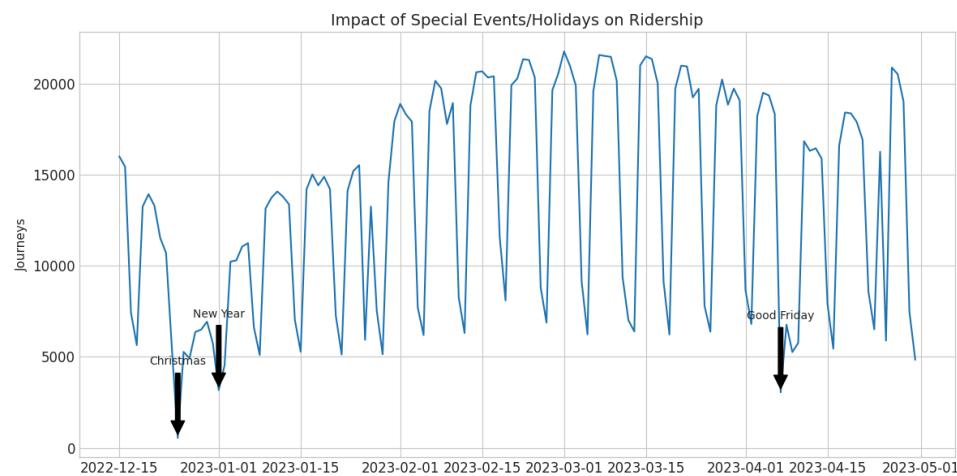
2. Trend Differences: Local and Rapid Routes show gradual increases over time, while Peak Service remains fairly stable.



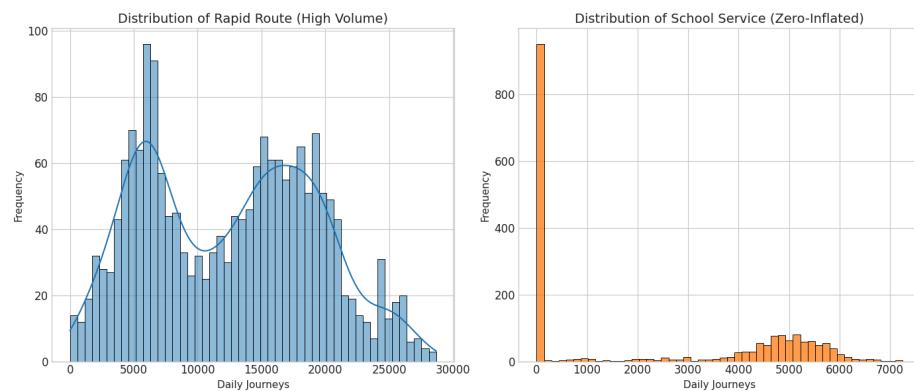
3. High Volatility in Small Services: Peak Service and School show erratic day-to-day variation.



4. Special Event Effects: Holidays and events create outliers in passenger numbers.



5. Right-Skewed Distribution: High-volume routes show a long-tail distribution; smaller services have frequent zero or near-zero counts.



Data Cleaning & Preprocessing

- Removed missing values and ensured **no negative passenger counts**
- Converted Date to datetime format and set as **index**
- Resampled to **daily frequency** to handle missing dates
- Checked for **outliers** using statistical summaries and plots

Exploratory Data Analysis (EDA)

Trend Analysis

- Local Route and Rapid Route show **steady growth over years**
- Peak Service remains **relatively stable**

Seasonality

- **Weekday peaks** and **weekend dips** are consistent across all services
- School service spikes during **school term periods**

Variability & Outliers

- Peak Service and School services show **high day-to-day volatility**
- Holidays and events cause **abnormal spikes**

Distribution

- High-volume services are **right-skewed**
- Low-volume services frequently have **zeros or near-zero values**

Forecasting Methodology

Model Used:

SARIMA (Seasonal ARIMA)

Rationale:

- Captures **trend + seasonality**
- Handles **non-stationary series** via differencing
- Provides **confidence intervals** and route-specific forecasts

Steps:

1. Data preprocessing and cleaning
2. Stationarity check (ADF test)
3. SARIMA fitting per route
4. 7-day forecast generation
5. Post-processing (negative values replaced with 0)
6. Evaluation using **MAE, RMSE, SMAPE**

Operational Insights & Recommendations

- Deploy **more buses on weekdays** and during **school terms**
- Reduce fleet on **weekends and holidays**
- Reallocate idle buses to **high-demand routes**
- Monitor anomalies in the “Other” category for **data quality improvements**

Technical Stack

- **Programming:** Python
- **Data Handling:** Pandas, NumPy
- **Forecasting:** Statsmodels SARIMAX
- **Visualization:** Matplotlib, Seaborn
- **Evaluation Metrics:** MAE, RMSE, SMAPE

Future Improvements

- Tune SARIMA hyperparameters for better accuracy
- Compare SARIMA with **Prophet or LSTM models**
- Include **holiday/event effects** for more precise forecasting
- Develop **real-time dashboard** using Flask or Streamlit
- Cluster routes for **similar demand patterns**