Public Transport Forecasting: Technical Report

1 Project Overview

This project implements time series forecasting for public transport ridership using three models: Prophet, SARIMA, and LightGBM. Analyzing 61.87M passenger journeys (2020-2026) across 5 service types, SARIMA emerged as champion with 3/5 wins. An interactive HTML dashboard visualizes insights and forecasts for operational planning.

2 Dataset & Key Insights

Services: Rapid Route (24.16M, 39%), Local Route (18.97M, 31%), Light Rail (13.80M, 22%), School (4.51M, 7.3%), Peak Service (0.34M)

Insights:

- Strong weekly cycle: mid-week peaks 41.5K journeys/day, weekends drop 68% to 13K
- 2023 peak ridership (13.55M), declined 23% in 2024
- School services highly volatile with term-time dependency
- Top 3 services account for 92% of demand

3 Model Implementation

3.1 SARIMA (Winner: 3/5)

Config: Order (1,1,1), Seasonal (0,1,1,7)

Why it won: Perfectly captures 7-day weekly cycle through seasonal differencing. Achieved 33% MAE improvement over Prophet on Rapid Route (9,555 vs 14,196) and 36% on Light Rail (5,307 vs 8,272). Statistical rigor provides confidence intervals for risk assessment.

Best for: Light Rail, Peak Service, Rapid Route - stable services with strong weekly patterns.

3.2 LightGBM (Winner: 2/5)

Config: 100 trees, learning rate 0.1, features include day/week/year, 7-day and 14-day lags, rolling mean

Why it won: Excels at non-linear patterns and high variance. Achieved 41% better MAE than SARIMA on School services (1,276 vs 2,977) by capturing term-time vs holiday fluctuations. Won highest-volume Local Route (MAE 9,171).

Best for: School and Local Route - volatile services with irregular patterns.

3.3 Prophet (Winner: 0/5)

Config: Linear growth, multiplicative seasonality, automatic changepoint detection

Why it failed: Overestimated demand across all services by focusing on long-term trends while missing critical short-term weekly dynamics needed for operational planning.

4 7-Day Forecast Results

SARIMA forecasts (Sept 30 - Oct 6, 2024) show:

- Mon-Thu: High demand 13-14.5K daily journeys
- Friday: 58% drop to 4.5K (early weekend behavior)
- Weekend: Zero demand predicted (requires verification)
- Local Route most stable: 5,080-5,926 weekday journeys
- Rapid Route high weekday: 4,573-5,336 with sharp Friday drop

5 Interactive Dashboard

Developed a comprehensive HTML dashboard to visualize the complete analysis and enable operational decision-making.

5.1 Technologies Used

- **Frontend:** HTML5, CSS3, JavaScript
- **Visualization:** Plotly.js and Chart.js for interactive charts
- **Data Processing:** Papa Parse for CSV handling
- **Design:** Responsive layout with CSS Grid/Flexbox

5.2 Dashboard Features

Overview Section: Displays key statistics including total journeys (61.87M), service distribution, and model performance scoreboard (SARIMA: 3, LightGBM: 2, Prophet: 0).

Insights Panel:

- Service type ridership bar chart
- Weekly seasonality line graph
- School vs public ridership pie chart
- Yearly trends time series
- Service correlation heatmap

Model Comparison: Interactive performance matrix showing MAE, RMSE, MAPE, SMAPE for all models across services. Color-coded cells (green=best, red=worst) with tooltips explaining metrics.

Forecasting Section:

- Multi-line chart plotting 7-day forecasts for all services
- Toggle buttons to show/hide individual service predictions
- Side-by-side comparison of Prophet, SARIMA, LightGBM forecasts
- Weekly totals and weekday averages for planning

Interactivity: Hover tooltips, zoom/pan on charts, dynamic filtering, smooth animations, and real-time data updates from CSV files.

Deployment: Fully self-contained single-page application with no server requirements. CDN-hosted libraries enable easy sharing and deployment.

6 Conclusions

Best Model Choice: SARIMA for 60% of services (Light Rail, Peak, Rapid), LightGBM for 40% (School, Local Route). This hybrid approach leverages each model's strengths.

Key Findings:

- SARIMA's 7-day seasonality perfectly matches ridership cycles
- LightGBM handles volatility 41% better than statistical models
- Prophet unsuitable for tactical forecasting
- Friday capacity should be reduced 50-60%
- Weekend zero-demand needs operational verification

Recommendations: Deploy winning models to production, retrain weekly with 180-day rolling window, monitor forecast accuracy (target: MAE ¡10K for high-volume routes), automate pipeline, and incorporate external features (weather, events). The interactive dashboard provides real-time operational insights for transport planners.