

KOVAI.CO DATA SCIENCE INTERN ASSESSMENT

NITHYA SREE RS 22AM030

Use Case: Forecasting Public Transport Passenger Journeys

CHAPTER 1

TASK INSIGHTS

1.1 Key Insights from Public Transport Data

1. Local and Rapid Routes Move Together

The number of passengers using Local Routes and Rapid Routes tend to rise and fall almost in sync. This means people likely use these services in similar ways or for similar trips.

2. School Service Has Clear Busy and Quiet Periods

The School service shows big spikes when schools are open and drops to almost zero during holidays. This matches what we'd expect since kids don't travel on school transport during breaks.

3. Light Rail and Peak Services Slow Down During Holidays

Both Light Rail and Peak services have steady use but show noticeable drops during holidays or vacations. This tells us travel decreases when people are off work or on holiday.

4. All Services Show Regular Patterns Over Time

Each service shows clear seasonal trends and overall growth or decline patterns. This makes them great candidates for forecasting models that can pick up on these repeating cycles.

5. The Data Is Reliable and Mostly Complete

There were very few missing values, and where needed, we filled gaps by carrying forward the last known data. This keeps the dataset solid and trustworthy for making predictions.

CHAPTER 2

TECHNICAL REPORT

2.1 Prophet Algorithm for Time Series Forecasting

Prophet is an open-source forecasting tool developed by Facebook (now Meta) designed for time series data with strong seasonal effects and historical trends. It is especially useful when working with daily data and provides reliable forecasts even with missing values or outliers. Prophet automates much of the complexity behind time series modeling while allowing users to fine-tune various components as needed.

2.1.1 Why Prophet ?

- Handles **missing data and outliers** without special preprocessing
- Easily models **seasonality, holidays, and custom events**
- Provides **forecast intervals**, improving decision-making confidence
- **Simple syntax**, making it suitable for quick experimentation and production use

2.1.3 Key Model Parameters

Parameter	Description
interval_width	Sets the confidence interval for forecast uncertainty.
growth	Determines the trend type; defaults to linear if not specified.
seasonality_mode	Defines how seasonal effects are added to the model; additive by default.
changepoints	Automatically detects points where the time series trend changes.
yearly_seasonality	Captures recurring yearly patterns in the data.
weekly_seasonality	Captures weekly patterns and cycles in the time series.
daily_seasonality	Skipped for daily data; used only for sub-daily data like hourly.

2.1.5 Application in This Project

Prophet was applied separately for each transport service (e.g., Local Route, Peak Service) using 365-day historical data. A 7-day future forecast was generated with 95% confidence intervals. The model captured both trend and seasonality effectively, and cross-validation confirmed its reliability with performance metrics like MAPE and RMSE.

LINK:

https://colab.research.google.com/drive/1fd3N6Zpf_rEAvUXhTN3KxDwPG5ptWcyp?usp=sharing