Task Report

Dataset Overview

The dataset contains daily passenger counts for various public transport services, including:

- Local Route
- Light Rail
- Peak Service
- Rapid Route
- School Service

The data spans multiple months or years, capturing both short-term and long-term trends.

Key Insights:

Passenger Volume Trends

- Local Route and Peak Service consistently record the highest average passenger counts, reflecting heavy daily usage.
- School Service exhibits distinct spikes on school days with sharp drops during weekends and holidays.

Seasonality & Weekly Patterns

- Clear weekly seasonality is evident, with significantly higher ridership on weekdays across all services.
- Light Rail shows relatively steady ridership with less variability compared to bus routes.

Growth & Decline Periods

- Rapid Route demonstrates a gradual upward trend over the past year, likely linked to route expansions or population growth.
- Peak Service experiences fluctuations, potentially influenced by holidays or weather conditions.

Data Quality & Anomalies

- Missing or zero passenger counts were present and cleaned from the dataset.

- Occasional outliers with unusually high counts correlate with special events.

Correlation Between Services

- A strong positive correlation exists between Local Route and Peak Service passenger volumes, indicating overlapping demand during rush hours.

Technical Report — Forecasting with Prophet

Chosen Algorithm: Facebook Prophet

Facebook Prophet is an open-source forecasting tool designed to handle time series data characterized by strong seasonal effects and multiple seasonal patterns. It is widely used for business forecasting due to its flexibility and robustness.

Why Prophet?

- Handles **multiple seasonality** including daily, weekly, and yearly cycles.

- Robust to **missing data and outliers**, reducing the need for extensive data cleaning.

- Provides **uncertainty intervals** with predictions, allowing confidence assessment.

- Easy to tune and extend with **custom seasonalities** and holidays.

Model Parameters and Configuration

Parameter	Description	Value/Setting
daily_seasonality	Captures daily recurring patterns	Enabled (True)
weekly_seasonality	Captures weekly cycles (weekday vs weekend effects)	Enabled (True)
yearly_seasonality	Captures annual seasonal patterns	Enabled if data > 365 days
seasonality_mode	Type of seasonal effect (additive or multiplicative)	multiplicative
interval_width	Confidence interval for prediction uncertainty	0.95 (95%)
custom seasonality	Added monthly seasonality (period = 30.5 days)	Added if data > 30 days

Data Preparation

- Converted the date column to a datetime format required by Prophet.

- Reshaped the dataset into a long format with columns: ds (date), y (passenger count), and service_type.

- Removed missing and negative passenger counts to ensure data quality.

Forecasting Process

- Aggregated data by date and service type for model input.

- Trained a separate Prophet model for each public transport service category.

- Generated 7-day ahead forecasts to predict passenger counts.

- Clipped negative predictions to zero to maintain logical consistency.

- Saved visualizations including forecast plots and seasonal component plots for analysis.

Model Evaluation and Limitations

- Forecasts matched historical seasonal patterns, validating model accuracy.

- Services with sparse data or low variability yielded less reliable predictions.

- The multiplicative seasonality mode effectively captured proportional seasonal effects.

- Future enhancements could include adding external regressors such as weather conditions or holiday effects to improve forecast accuracy.