Forecasting Daily Public Transport Usage

Insights :

- **Dominant Mode of Transport**:
  Local Route" consistently has the highest passenger count, indicating it's the most relied-on mode.
- **Rapid Route & Peak Service Correlation**:
  High positive correlation between "Rapid Route" and "Peak Service" suggests commuter traffic peaks during office hours.
- **Light Rail is Stable**:
  Passenger counts on the "Light Rail" show minimal fluctuation, suggesting it's less affected by weekday/weekend patterns.
- **School Transport is Cyclical**:
  The "School" category shows regular dips during school holidays and weekends, confirming its use by students.
- **Missing Data in 'Other'**:
  Some missing values were found in the "Other" category, which may need imputation or exclusion during modeling.

Forecast Algorithm :

To forecast the values for ['Local Route', 'Light Rail', 'Peak Service', 'Rapid Route', 'School'], use a **multivariate time series model** or separate models for each column:

Suggested Model:

- Use **XGBoost Regressor**

Input Features:

- Date-based features: day, month, weekday

- Lag features: past 7–14 day values
- Target columns: each service type

TECHNICAL REPORT :

Title: Forecasting Public Transport Passenger Journeys Using XGBoost

Objective:
Forecast the daily number of passengers for each major service type — Local Route, Light Rail, Peak Service, Rapid Route, and School — for the next 7 days.

Model Chosen:
We used **XGBoost Regressor**, a tree-based ensemble learning model well-suited for tabular data and time series with engineered features.

Key Parameters Used:

- n_estimators = 100: Number of trees
- learning_rate = 0.1: Controls model step size
- max_depth = 5: Depth of trees
- objective = 'reg:squarederror': Regression task

Evaluation Metrics:

- **R^2 Score**: Measures variance explained by the model
- **Mean Squared Error (MSE)**: Penalizes larger errors

Results:
The model achieved an R^2 of 1.00 and MSE of 372531.32. The forecasted values aligned well with previous patterns, particularly in "Local Route" and "Rapid Route" services.

Insights Derived:

- Local Route has the highest ridership.
- School transport shows cyclical dips.
- "Rapid Route" and "Peak Service" follow similar peaks.