# Daily Public Transport Forecasting Report

## About Dataset:

The dataset tracks how many people use different transport types each day:

- The dataset records daily ridership counts for several public transport service types, including **Local Route, Light Rail, Peak Service, Rapid Route, School services, and other category.** Each entry corresponds to a single calendar date and contains the number of passengers (or trips) observed on that day for each route type.
- The data forms a multivariate time series, where multiple route-specific ridership variables are tracked simultaneously across time. Clear weekly patterns are present, with differences between weekdays and weekends, and additional school-related effects visible in the School and Peak Service routes during term time. Some dates contain missing values or irregular entries, which are resolved during preprocessing before model training.

## Problem Statement

Efficient public transport operations depend on knowing how passenger demand changes across different routes and days. Without reliable forecasting:
• Buses may remain under-utilized on low-demand routes while other routes become overcrowded.
• Peak periods such as school days and weekday office hours may suffer from insufficient service.
• Fuel and staffing costs may rise due to unnecessary vehicle deployment.

To Develop accurate 7-day route-level ridership forecasts to support smarter scheduling, vehicle allocation, and cost-efficient operational planning.

## Time Series Forecasting

**Time series forecasting** predicts future values based on historical data patterns. Key elements include:

- **Trend:** Long-term increase or decrease in ridership.
- **Seasonality:** Regular repeat patterns (e.g., weekday/weekend cycles).
- **Noise:** Random variations in the data.

For public transport, **weekly seasonality** is significant, making simple regression insufficient.

## Chosen Algorithm: SARIMA

SARIMA (Seasonal ARIMA) is a time-series forecasting model designed to capture both **long-term trends** and **repeating seasonal patterns**. It expands ARIMA by adding seasonal terms, making it suitable for datasets like public transport ridership where weekday and weekend cycles repeat every week.

**ARIMA Components**

- **AR (p):** Influence of past values
- **I (d):** Differencing used to remove trend and make the data stationary
- **MA (q):** Influence of past forecast errors

**Seasonal Components**

- **P, D, Q:** Seasonal versions of AR, I, and MA
- **m:** Length of seasonal cycle (7 here, representing weekly seasonality)

**Why SARIMA was chosen**

- Effectively models weekday–weekend seasonal behaviour in ridership
- Handles autocorrelation, trend, and irregular fluctuations
- Produces realistic route-level forecasts that align with real commuting patterns

## Model Parameters

| Parameter | Value | Meaning |
|---|---|---|
| **order (p, d, q)** | **(1, 1, 1)** | Includes 1 lag, first differencing, and 1 MA term |
| **seasonal_order (P, D, Q, m)** | **(1, 1, 0, 7)** | Weekly seasonality: 1 seasonal lag, seasonal differencing, and cycle length of 7 days |
| **enforce_stationarity** | False | Allows model flexibility even if data isn't strictly stationary |
| **enforce_invertibility** | False | Helps achieve stable fitting without strict invertibility constraints |

## Implementation

1. **Data Preparation**
    - Converted Date column to datetime and set it as the index
    - Cleaned missing values and ensured numeric consistency across ridership columns
2. **Model Execution**
    - Fitted a SARIMAX model **for each route individually**
    - Forecasted ridership for the **next 7 days** per route
3. **Fallback Strategy**
    - If SARIMA fails to converge for any route, the model automatically uses the **last week's ridership pattern** as a backup forecast
4. **Post-processing**
    - Converted predictions to **non-negative integer passenger counts**
    - Combined all forecasts into a **single table and visualization**

**Benefits of Forecasting**

- **Operational Efficiency:** Reassign buses from low-demand routes to high-demand ones
- **Capacity Planning:** Prepare for school days and peak commuter periods
- **Cost Optimization:** Reduce unnecessary weekend over-provisioning
- **Decision Support:** Enable data-driven scheduling and resource allocation

**Advantages of SARIMA for This Project**

- **Captures Seasonality:** Understands repeating weekly travel behaviour, unlike plain ARIMA
- **Handles Trends & Autocorrelation:** Learns long-term ridership growth/decline and weekday-weekend dependency
- **Flexible & Tunable:** Parameters allow fine control of model behaviour to match real-world patterns
- **Noise-Resilient:** Smooths out random fluctuations and produces stable, actionable forecasts