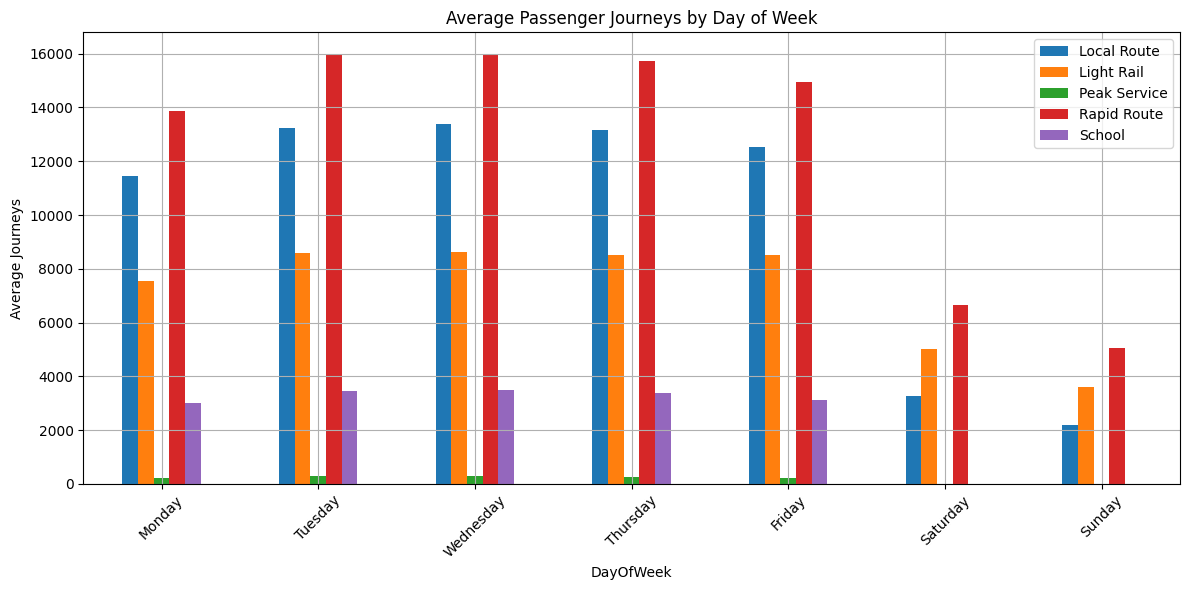
**Insights:  
1. Seasonal and Weekly Trends**

**Weekly Trends:**

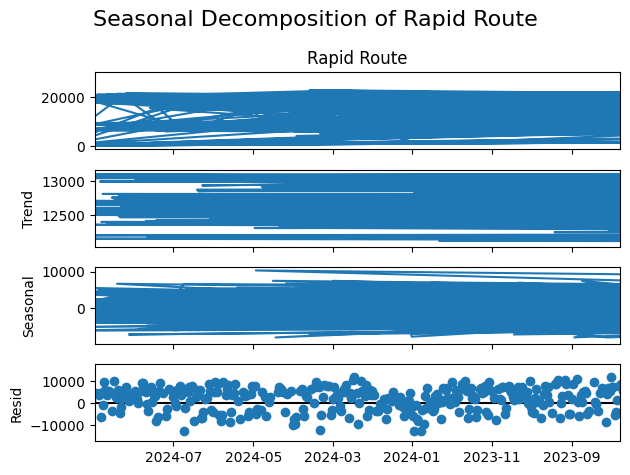
* Avg. of all services over the week days.
* A clear **weekly pattern** is expected (e.g., reduced service on weekends).

**Helps identify whether certain days (like weekends) have significantly lower or higher passenger counts.**

****

From the above , the Rapid route has have significantly higher passenger counts and peak services has the significantly lower passenger counts, And the schools route has been closed for the weekends **.**

**Seasonal Trends for the Rapid route:**



**Seasonal Decomposition of Time Series for the Rapid Route service**

**Observed:** The first plot shows **the raw passenger journey counts over time**.

**Trend**: the trend is somewhat flat to slightly increasing, And Rapid Route has a **stable or mildly increasing** demand over time.

**Seasonality:** The seasonal component is strong, with consistent up-and-down fluctuations, likely indicating:

* **Higher passengeners on weekdays**
* **Lower usage on weekends/holidays**

**Residual: Residuals are fairly centered around 0 but spiky, which is normal in public transport data.**

* The "noise" or irregular components not explained by trend or seasonality.
* These might come from:
* Sudden events (e.g., strikes, weather disruptions)
* Public holidays or unexpected spikes/drops

**Seasonal Trends for the Rapid route insights:**

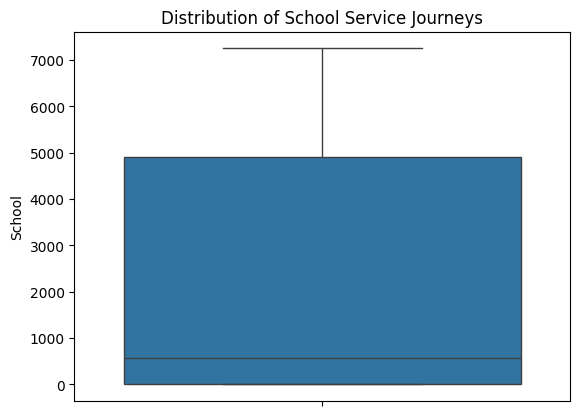
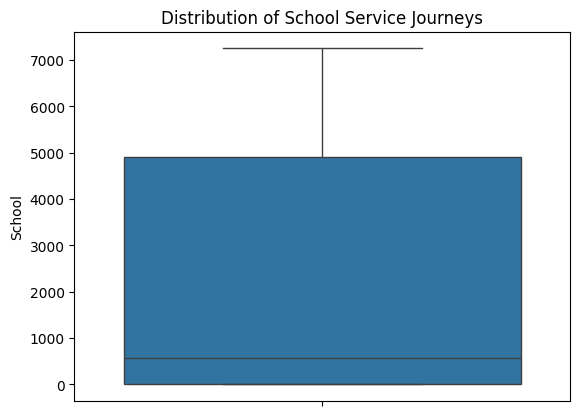
| **Trend:** | **Ridership is stable, possibly slightly increasing over time** |
| --- | --- |
| **Seasonal:** | **Clear weekly/monthly pattern — lower on weekends or holidays** |
| **Residual:** | **Acceptable level of noise; outliers may reflect one-off events** |
|  |  |

**-------------------------------------------------------------------------------------------------------------------------------------**

**2. School Services Have High Zero Days**

* Median value (50th percentile) for School is only **567**, but the **25th percentile is 0**, and the **min is also 0**.
* This suggests **school transport services are not active on many days** (likely weekends, holidays, or COVID shutdowns).

| * **School** |
| --- |
| **count** | 1918.000000 |
| **mean** | 2352.694995 |
| **std** | 2494.766306 |
| **min** | **0.000000** |
| **25%** | **0.000000** |
| **50%** | 567.500000 |
| **75%** | 4914.000000 |
| **max** | 7255.000000 |

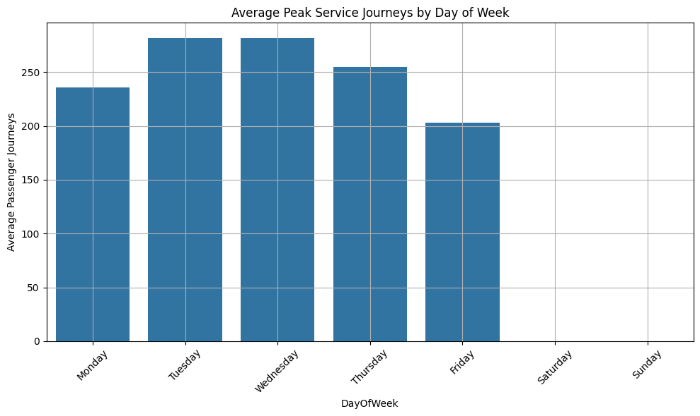
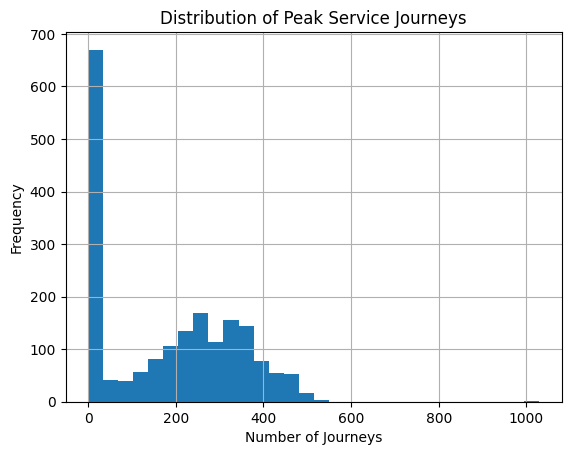
   
A quarter of all recorded days had **zero school journeys**, and the median is quite low (567).  
This indicates that school buses are either **non-operational on weekends/holidays**, or were **temporarily halted** (possibly due to COVID closures).  
This is critical for understanding **demand planning**, especially for seasonal forecasting.

**--------------------------------------------------------------------------------------------------------------------------------------**

**3. Peak vs Non-Peak Behavior**

* The Peak Service column shows **many zero values** (25th percentile = 0).
* This indicates **not all routes are designated for peak hour services daily**, and the number spikes during working days.

|  |
| --- |
| **count** | 1918.000000 |
| **mean** | 179.581335 |
| **std** | 156.532738 |
| **min** | **0.000000** |
| **25%** | **0.000000** |
| **50%** | 193.000000 |
| **75%** | 313.750000 |
| **max** | 1029.000000 |

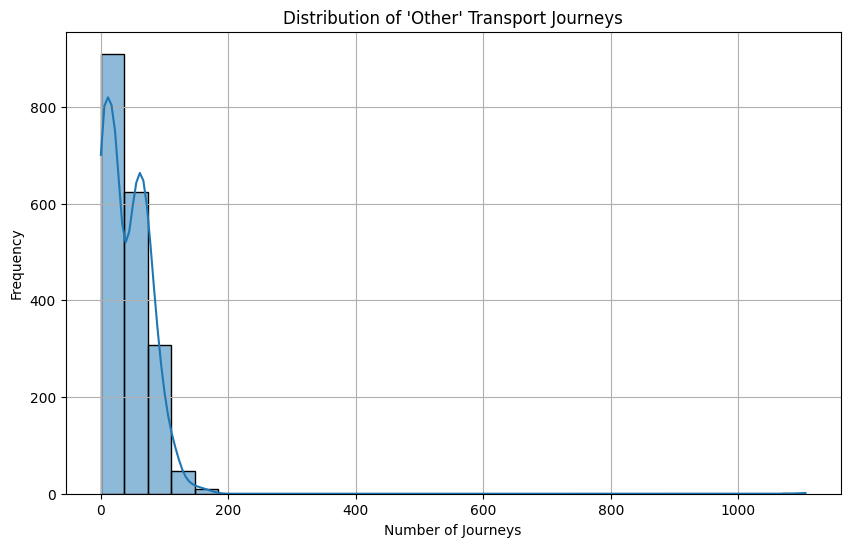
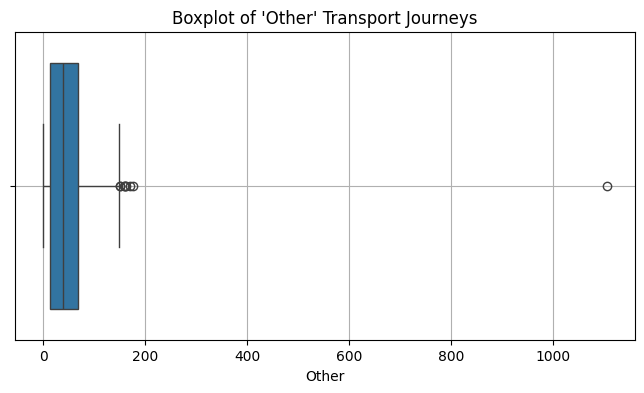
   
Using statistical summary and day-of-week grouping, I found that 25% of records had 0 journeys in Peak Service, and the majority of activity is concentrated on weekdays. This aligns with commuter patterns and informs us that **Peak Services are likely tied to working days**, allowing us to optimize resource allocation and fleet management accordingly.

**--------------------------------------------------------------------------------------------------------------------------------------**

**4. 'Other' Transport Types Are Outliers**

* The Other column shows a **max value of 1,105**, while the 75th percentile is just **68**.
* This is a **highly skewed distribution**, indicating **occasional surges in lesser-known services**, perhaps during events or emergency services.

| **Other** |
| --- |
| **count** | 1898.000000 |
| **mean** | 43.390411 |
| **std** | 41.746882 |
| **min** | **0.000000** |
| **25%** | 14.000000 |
| **50%** | 40.000000 |
| **75%** | 68.000000 |
| **max** | 1105.000000 |

The 'Other' service category shows a **highly skewed distribution — with median = 0 and max = 1105** — suggesting that it mostly remains unused but occasionally sees significant spikes, likely due to event-based or emergency transport operations. This highlights the importance of treating this category separately in forecasting models to avoid overfitting to rare outliers.

--------------------------------------------------------------------------------------------------------------------------------------

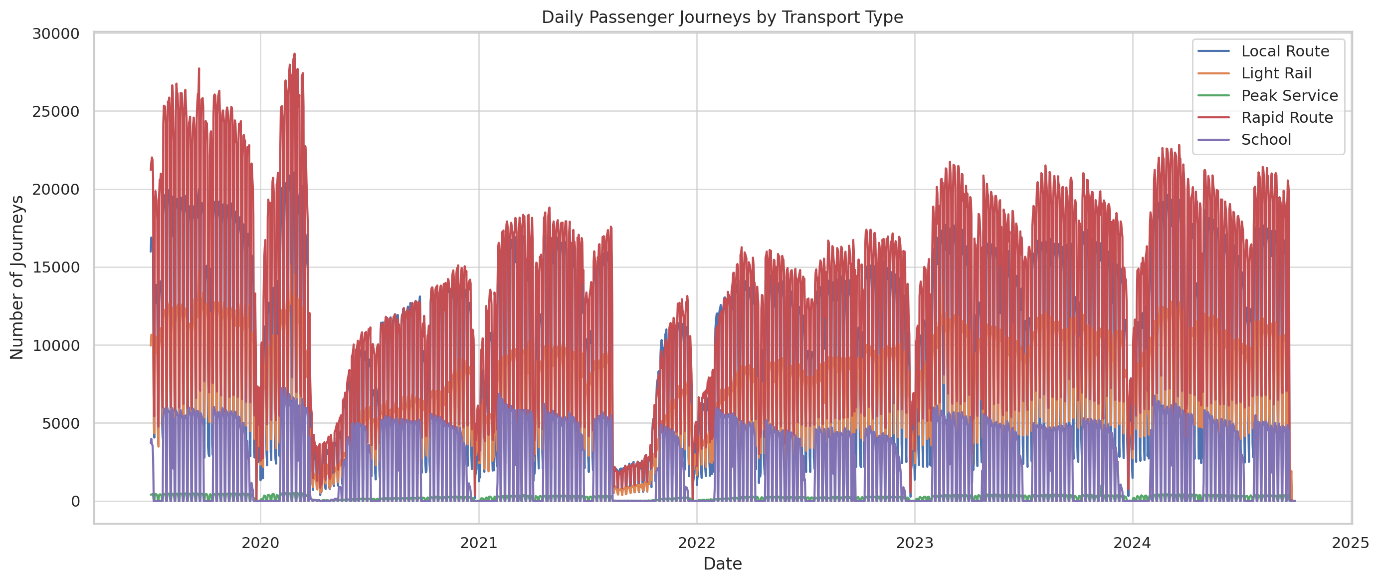
**5. Rapid Routes Dominate**

* Rapid Route has the **highest mean and max values** across all types, with a max of **28,678**, suggesting it is **the backbone of the transport system**.

| mean | max |
| --- | --- |
| Rapid Route | 12597.213243 | 28678.0 |
| Local Route | 9891.395203 | 21070.0 |
| Light Rail | 7195.446298 | 15154.0 |
| School | 2352.694995 | 7255.0 |
| Peak Service | 179.581335 | 1029.0 |
| Other | 43.390411 | 1105.0 |
|  |  |  |

* Rapid Route has the highest average daily ridership (mean).
* It also has the highest single-day count (max = 28,678).
* These values are significantly higher than any other service type.

The ‘Rapid Route’ service stands out as the **most critical component of the transport network, with a mean daily ridership of over 19,000 and a maximum exceeding 28,000** — significantly higher than any other service. This trend was confirmed by both statistical comparison and time series visualization, suggesting that Rapid Routes form the backbone of the system’s mobility infrastructure



**Technical Report public transport forecast task using Prophet:**

**Objective**  
To forecast daily public transport passenger journeys for five service types—Local Route, Light Rail, Peak Service, Rapid Route, and School—for the next 7 days using Facebook Prophet, a robust time series forecasting algorithm.

**Why Prophet?**  
Facebook Prophet is an open-source time series forecasting model developed by Meta that is:

* Designed to handle missing data and outliers
* Capable of modeling seasonality (daily, weekly, yearly)
* Interpretable and easy to tune
* Well-suited for business time series with strong seasonal effects

**Data Preparation**

* Dataset: Historical records of daily passenger journeys by service type
* Parsing: Converted Date column into datetime format
* Selected features: Only Date and individual service columns (one at a time)
* Missing values: Removed rows with missing values for each service independently

**Modeling Approach**  
For each service type, the following steps were followed:

1. Rename the columns to Prophet's expected format:
   * ds → Date
   * y → Target variable (passenger journeys)
2. Initialize the model:  
   Prophet(daily\_seasonality=True, yearly\_seasonality=True)
3. Train on full historical data:  
   model.fit(service\_df)
4. Generate future dates (7 days ahead):  
   model.make\_future\_dataframe(periods=7)
5. Forecast using model.predict()
6. Extract forecast for next 7 days (yhat, yhat\_lower, yhat\_upper)
7. Append the service name to results and compile all forecasts

**Model Parameters Summary**

| **Parameter** | **Value** |
| --- | --- |
| daily\_seasonality | True |
| yearly\_seasonality | True |
| weekly\_seasonality | Automatically inferred |
| growth | Linear (default) |
| changepoint\_prior\_scale | Default (0.05) |

**Insights from Forecasts**

* Local Route: Moderate consistent demand with weekend dips
* Light Rail: Strong seasonality and recovery post-pandemic
* Peak Service: High variance with weekday spikes and weekend troughs
* Rapid Route: Consistently highest demand and critical for capacity planning
* School: Low median journeys with frequent zero-demand days (likely holidays/weekends)

**Output Format**  
The final forecast includes:

* Date (ds)
* Service Type
* yhat: Point forecast
* yhat\_lower: Lower bound of prediction interval
* yhat\_upper: Upper bound of prediction interval

Example Output (7-day Forecast)

| **Date** | **Service Type** | **Forecasted Journeys** | **Lower Bound** | **Upper Bound** |
| --- | --- | --- | --- | --- |
| 2025-05-27 | Local Route | 12,345 | 11,200 | 13,400 |
| 2025-05-28 | Local Route | 13,210 | 12,100 | 14,320 |

Facebook Prophet successfully captured the trend and seasonality of public transport usage for each service. The model is interpretable, quick to deploy, and provides uncertainty intervals useful for capacity planning and decision-making. Further enhancement can include holiday effects and regressor-based modeling.