Report on Data Science

1. Key Insights from the Dataset

1. General Trends: Local and Rapid Routes Dominate Usage

Observation: The Local Route and Rapid Route service types consistently record the highest number of passenger journeys across the dataset. Their mean daily usage is approximately 9,891 and 12,597 journeys, respectively.

Inference: These services are likely the backbone of the transport system, covering essential and frequently used routes. The high numbers suggest a stable demand for daily commuting, possibly due to their coverage of urban and suburban areas.

---

2. School Services Show High Variability

Observation: Passenger journeys for School services range from 0 to a maximum of 7,255. The median value is only 568, indicating a skewed distribution with frequent low usage.

Inference: The large variability is indicative of a strong dependence on school term calendars and holidays. On days when schools are open, usage spikes, suggesting efficient targeting of this demographic. However, low median values imply that these services might be underutilized on non-school days, raising questions about cost-effectiveness during those times.

---

3. Peak Services: Small Contribution with Notable Spikes

Observation: The Peak Service category averages 179 passenger journeys per day, with occasional spikes to 1,029 journeys.

Inference: These spikes align with rush hours, suggesting that Peak Services effectively target high-demand periods. Despite this, their overall contribution to daily totals is small, indicating potential room for optimization to address under-utilization outside peak times.

---

4. Light Rail: Steady Usage with Seasonal Increases

Observation: Light Rail services maintain moderate and steady usage, with an average of 7,195 journeys per day. Usage shows periodic increases, possibly tied to seasonal or event-based demand.

Inference: The steadiness reflects its reliability and possibly its role as a feeder service. However, periodic increases hint at opportunities to align scheduling or capacity adjustments with peak seasons or special events.

---

5. Seasonal and Event-Driven Fluctuations

Observation: Passenger journeys for most service types show clear fluctuations corresponding to seasons, holidays, or specific dates.

Inference: These patterns emphasize the impact of external factors, such as festivals, school vacations, or weather conditions. By analyzing such trends further, the transport authority can better allocate resources and manage capacity.

---

2. Forecasting

Model Overview: Long Short-Term Memory (LSTM)

LSTM is a type of Recurrent Neural Network (RNN) capable of learning long-term dependencies in sequential data. Traditional machine learning models, such as linear regression, often struggle with sequential data as they lack memory to retain information from earlier data points. LSTM, however, is designed to address this issue by using gates to control the flow of information over time, allowing the model to remember important past data and forget irrelevant data.

Steps Taken for Forecasting using LSTM

- **Data Preprocessing**
- **Outlier Removal**
- **Data Normalization**
- **Creating Time Series Sequences**
- **Splitting the Data**
- **Building the LSTM Model**
- **Training the Model**
- **Making Predictions**
- **Inverse Transformation**
- **Visualization of Predictions**

Output Explanation

The model's predictions are compared with historical data using graphs. The blue line represents the historical data, while the red dashed line shows the forecasted values.

The table below displays the predicted values for the next 7 days. Note: Negative values might indicate parameter issues, which can be resolved through hyperparameter tuning.