1. Seasonal Fluctuations in Route Frequency

One of the key insights observed from the data is the seasonal variation in route frequencies. Routes like Light Rail and Peak Service show clear increases during certain times of the year, such as school holidays or weekends. This indicates that the demand for these services may be influenced by external factors like public holidays, school schedules, and special events.

2. Increased Demand During Weekdays for Peak Services

For routes such as Peak Service, the data reveals a consistent spike in frequency during weekday morning and evening hours. This could be indicative of commuter patterns, where the service is primarily used for work-related travel. The ARIMA model successfully captured these trends, which can help optimize service during peak hours.

3. Correlation Between Different Routes

There is a significant correlation between different types of routes, especially between the Rapid Route and Light Rail. Both show similar patterns, with higher frequencies during certain parts of the day. This suggests that these services may cater to similar groups of passengers, possibly for similar destinations or routes.

4. Stability in School Route Frequencies

The School route exhibits relatively stable frequencies, with only slight fluctuations. This indicates that the demand for school-related transportation is relatively consistent and less influenced by external factors like holidays or events, unlike the Peak or Rapid routes.

5. Opportunities for Service Optimization

The insights from the time series data indicate that there may be opportunities to optimize services based on forecasted demand. For example, routes like Local Route and Light Rail show consistent fluctuations during specific hours. This pattern could help service planners adjust timetables, allocate resources more efficiently, and improve overall passenger satisfaction.

6. Model Performance and Accuracy

The ARIMA model showed good predictive accuracy, with minimal residual error. However, some routes (like Peak Service) showed small but consistent deviations between historical and forecasted values. This suggests that there could be underlying seasonal factors or external events influencing demand that were not fully captured by the ARIMA model.

Data Preprocessing:

To begin the analysis, we first preprocessed the dataset. The Date column was converted into a datetime format to ensure accurate time-based indexing and analysis. We then set the Date column as the index, which is crucial for time series forecasting.

Handling missing values was another important step. The dataset had missing values that were filled using forward filling, where the most recent known value was used to fill in the gaps.

```python
Python
Copy
Edit
df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
df.set_index('Date', inplace=True)
df.fillna(method='ffill', inplace=True)
```

Forecasting with ARIMA:

Once the data was cleaned and prepared, we applied the ARIMA model for forecasting the next 7 days. ARIMA was chosen due to its ability to model and forecast time series data that shows trends and patterns over time. The model was trained with parameters (p=5, d=1, q=0). These parameters were selected based on the nature of the data, where p refers to the lag order (5), d refers to differencing (1 for stationarity), and q is the moving average (set to 0).

The ARIMA model was fitted to each route's historical data, and the forecast was generated for the next 7 days. The following code was used for forecasting:

```python
python
Copy
Edit
model = ARIMA(df[column], order=(5, 1, 0))
model_fit = model.fit()
forecast = model_fit.forecast(steps=7)
```

Visualizing Forecasts:

The historical data and forecasted values were plotted to visually assess the model's performance. The forecast_all DataFrame, which contains the forecasts for all routes, was plotted against the historical data to compare the actual vs forecasted values.

Route Forecasts for Next 7 Days

Local Route Forecast for Next 7 Days