KEY INSIGHTS :

1. Rapid and Local Routes Drive Overall Ridership: Rapid Route posts the highest average (12,597) and peak (28,678) ridership, making it the backbone of the transit system. Local Routes also contribute significantly but show high variability (standard deviation: 6,120), requiring flexible scheduling and targeted forecasting.

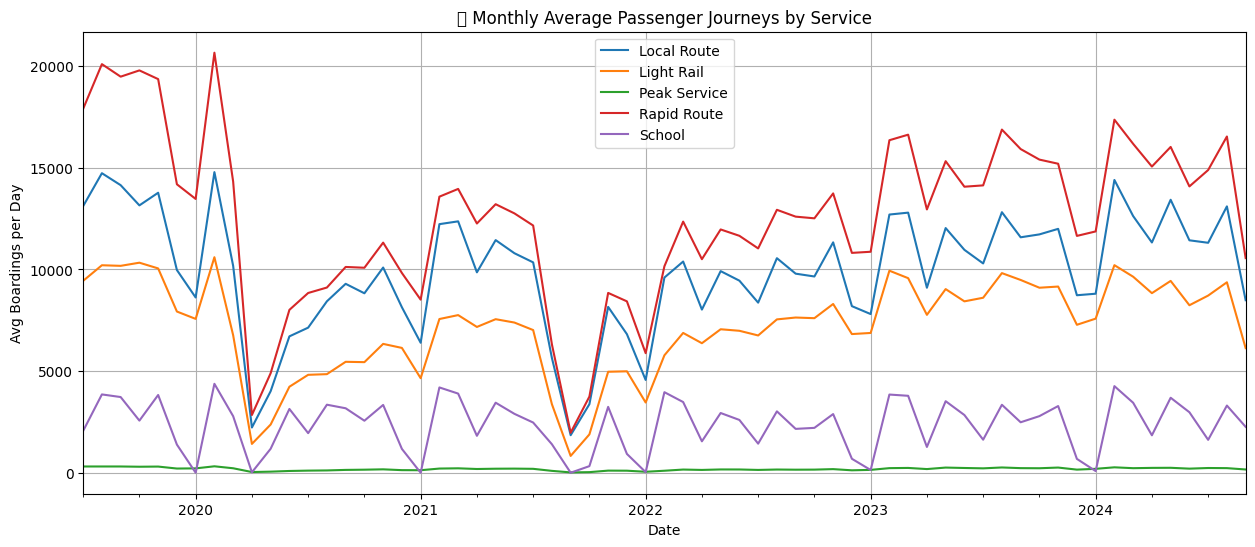
2. Peak Service and School Transport Are Highly Variable : Peak Service has low average numbers but experiences sharp mid-week spikes (standard deviation: 156.5). Similarly, School Transport shows strong weekday and academic term patterns, with usage dropping to zero during holidays—demanding seasonal and weekday-based planning.

3. “Other” Transit Modes and Light Rail Reflect Unique Patterns : The “Other” category has a low average (43) but can spike up to 1,105, often driven by events or seasonal factors—highlighting the need for contingency strategies. Light Rail, on the other hand, shows more stable and gradually increasing usage, suggesting long-term growth potential.

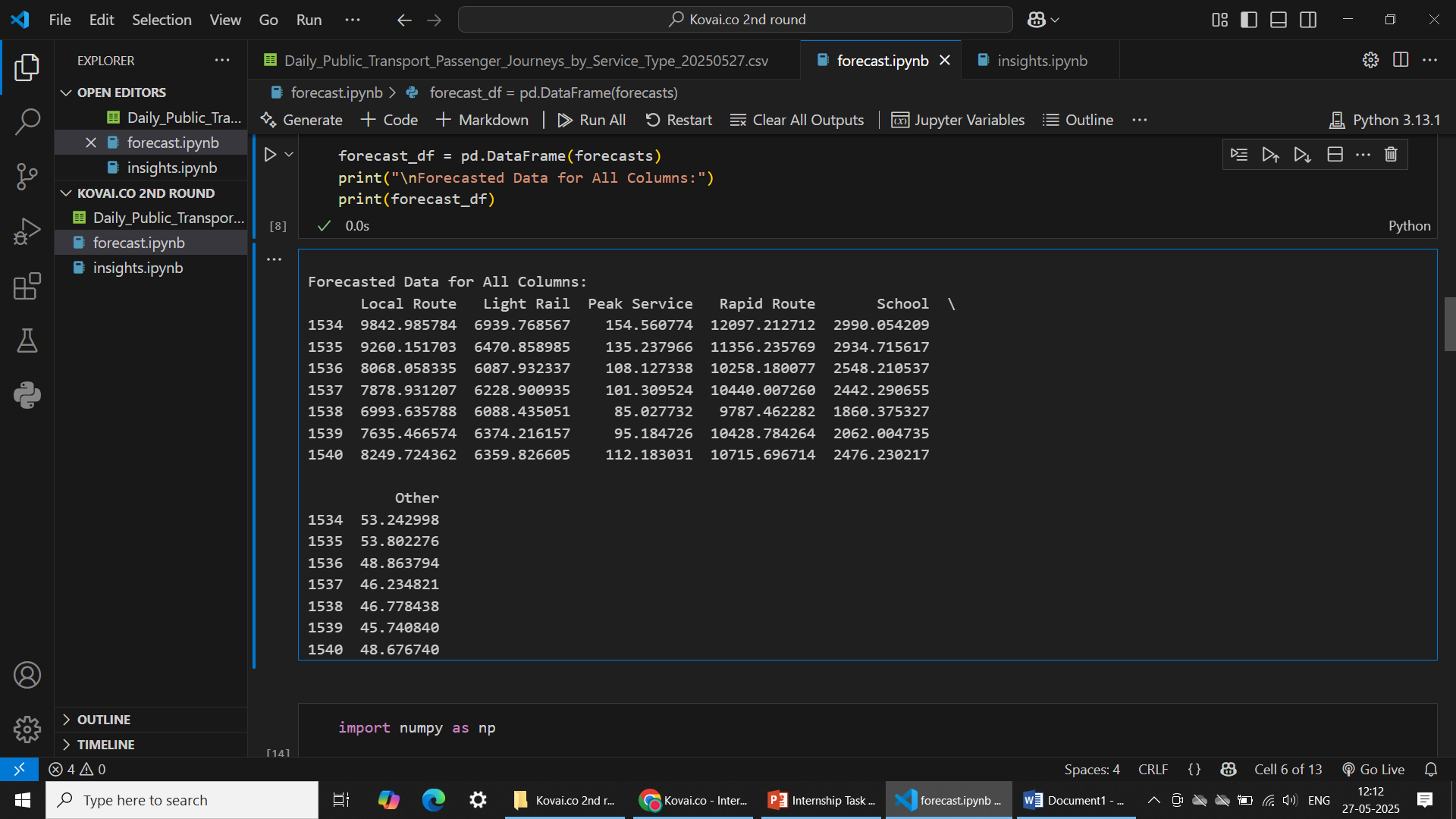
4. COVID-19 Caused Sharp Temporary Disruptions : A significant dip in ridership across all services between 2020–2021 reveals the clear impact of COVID-related disruptions, offering valuable context for anomaly detection in forecasting models.

5. Total Ridership Is Rising Over Time : The consistent increase in median and peak values across transit types indicates a clear upward trend, underscoring the need for scalable infrastructure and forward-looking transit investments.

**MONTHLY AVERAGE PASSENGER JOURNEYS BY SERVICE:**



2.FORECASTING FOR NEXT 7-DAYS:



ARIMA Model and Its Parameters

1. Introduction

The ARIMA (Autoregressive Integrated Moving Average) model is widely used for forecasting non-seasonal time-series data by capturing trends and past patterns.

2. ARIMA(p, d, q)

p: Lag order (past values)

d: Differencing order (for stationarity)

q: Moving average order (past errors)

3. Parameter Selection

p: Based on PACF plot

d: Chosen via stationarity tests (e.g., ADF test)

q: From ACF plot

Auto-ARIMA or manual tuning can be used. In this case, GridSearchCV was applied to find the optimal (p, d, q) combination by minimizing error metrics (e.g., AIC or RMSE).

4. Model Optimization

Checked stationarity and applied differencing

Used GridSearchCV for systematic hyperparameter tuning

#### **5. Accuracy Evaluation & Improvements**

The final **optimized ARIMA(3,0,2)** model was evaluated using:

* **MAE:** Measures absolute errors.
* **RMSE:** Quantifies overall error magnitude.
* **MAPE:** Assessed the forecasting accuracy across different dataset columns

6.Challenges include:

Poor handling of non-linear or highly volatile patterns

Needs sufficient historical data for accurate modeling

7. Conclusion

ARIMA is a reliable forecasting model when properly tuned. With techniques like GridSearchCV, prediction accuracy can be significantly improved. Future work can explore hybrid models for better performance on complex datasets.

THANK YOU!