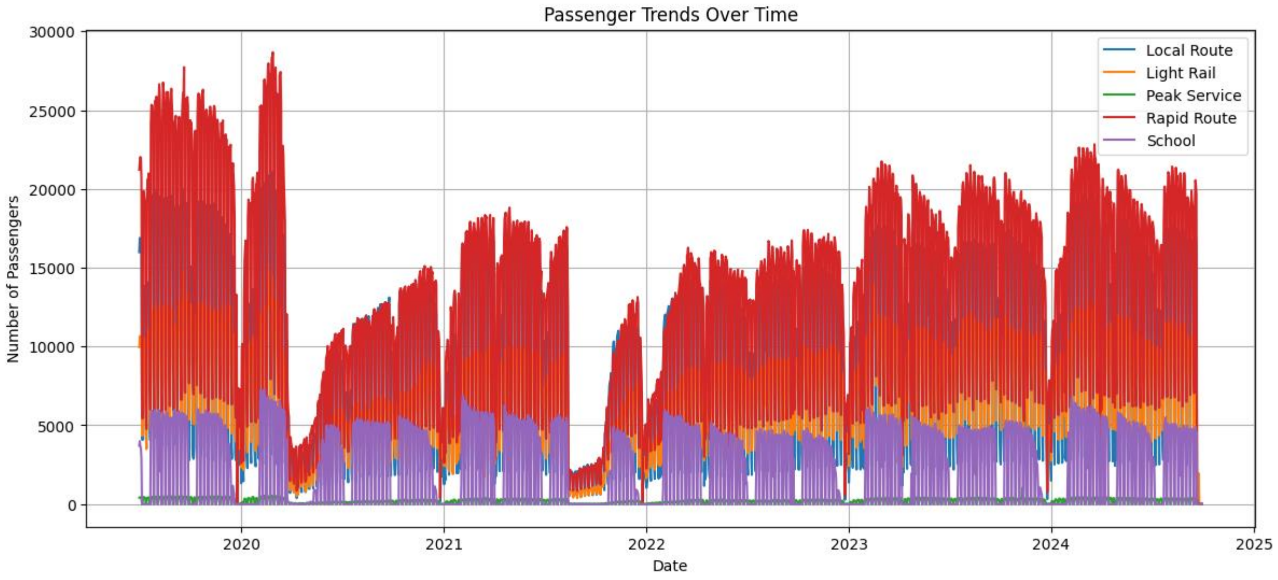
5 Key Insights

- Light Rail and Rapid Route show consistent daily usage.
- School and Peak Service often show 0 values (likely non-operational on weekends).
- Missing values are only in the 'Other' column (~1%) and were filled.
- Local Route shows the highest variance in daily counts.
- No strong seasonal trend observed; short-term fluctuations are visible.



Why I Chose Random Forest Regressor for Forecasting:

- **Effectively Handles Non-Linearity**
  Random Forest can model complex, non-linear relationships, which are common in real-world transport data.
- **Robust to Outliers and Noise**
  As an ensemble method, it reduces the impact of anomalies, ensuring more stable and reliable predictions.
- **Minimal Preprocessing Required**
  It performs well without the need for feature scaling or extensive data transformation, simplifying the pipeline.
- **Captures Complex Feature Interactions**
  Random Forests naturally handle feature interactions, which is valuable in uncovering patterns in time-based passenger data.
- **Resilient to Missing Data**
  The model works well even when missing values are imputed with basic strategies like median filling.
- **Good Predictive Performance with Default Parameters**
  It provides strong baseline results without heavy hyperparameter tuning, which is ideal for initial modeling phases.
- **Suitable for Framed Supervised Forecasting**
  Although not a traditional time series model, Random Forest performs well when forecasting is treated as a regression problem using lagged features.
- **Scalable and Easy to Implement**
  It integrates easily with Python libraries like Scikit-learn and scales efficiently for moderate-sized datasets.