DATA ANALYSIS AND FORECASTING TASK AT KOVAI.CO

KEY INSIGHTS :

1. **Long-Term Ridership Trend Shows Steady Growth:**
   Ridership displays a clear upward trajectory over the years, indicating strengthening demand and expanding user engagement.
   Median and peak values rise consistently, showing long-term growth rather than short-lived fluctuations.

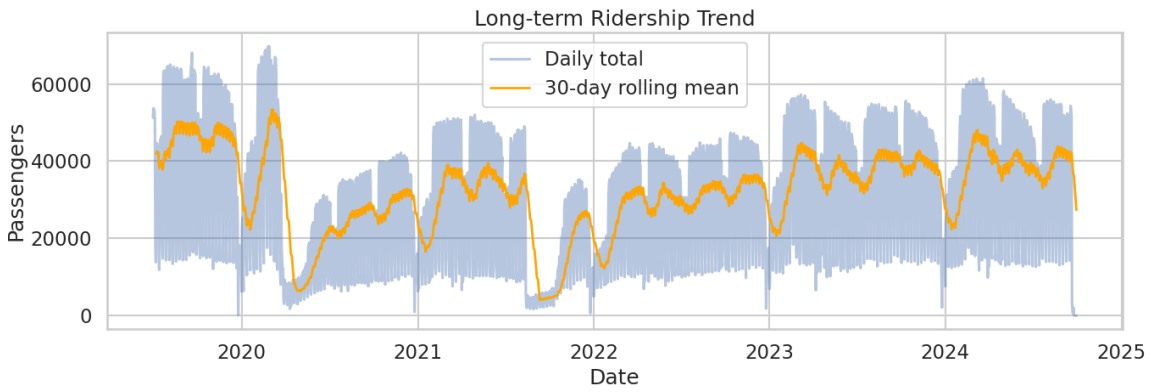
2. **Annual and Monthly Seasonality Drives Predictable Demand Cycles:**
   The data reveals strong seasonal patterns, with recurring peaks during certain months and steady dips during off-seasons.
   These repeated fluctuations indicate that ridership behaviour follows natural annual and monthly rhythms.

3. **COVID-19 Caused a Sharp and Temporary Drop in Ridership:**
   A sudden, steep decline appears across all transportation modes during the 2020–2021 period.
   This disruption stands out as an anomaly linked directly to lockdowns, travel restrictions, and reduced public movement.
   Accounting for this drop is crucial for building resilient forecasting models that don't misinterpret pandemic-era behaviour.

4. **Rapid Route Emerges as the Most Stable and Reliable Service Type:**
   Among all services, the Rapid Route maintains the most consistent ridership levels with minimal volatility.
   Its steady demand pattern makes it a dependable backbone of the transit network.
   This stability suggests that Rapid Routes can support long-term scaling and more confident forecasting.

5. **Peak Service and School Transport Show Distinctive Usage Patterns:**
   Peak Service experiences sharp weekday spikes, reflecting commuter-driven travel behaviour.
   School Transport follows academic cycles, with strong ridership during term time and complete drops during holidays.
   These patterns require targeted scheduling, seasonal planning, and flexible resource allocation.

LONG TERM RIDERSHIP TREND:-



FORECASTING FOR NEXT 30-DAYS:-

```
Next 30 days forecast:
2024-07-02    3654.717534
2024-07-03    3749.775931
2024-07-04    3731.849180
2024-07-05    3552.378628
2024-07-06     171.014027
2024-07-07     184.759796
2024-07-08    3202.408058
2024-07-09    3740.059041
2024-07-10    3817.162184
2024-07-11    3735.640536
2024-07-12    3517.087894
2024-07-13     287.266348
2024-07-14     288.545254
```

1. Introduction

The SARIMAX (Seasonal AutoRegressive Integrated Moving Average with Exogenous variables) model is designed for forecasting time-series data that exhibits **trend**, **seasonality**, and **autocorrelation**.
It is well-suited for public transport ridership because the data shows strong **weekly seasonal patterns**.

2. SARIMAX (p, d, q) × (P, D, Q, s)

p – Non-seasonal autoregressive order

d – Differencing for trend removal

q – Moving average order

P, D, Q – Seasonal AR, differencing, MA terms

s – Seasonal period (here **7 days** for weekly cycle)

3. Model Optimization

Data cleaned and transformed for stationarity.

Seasonal differencing applied where needed.

AIC-based hyperparameter search used for precise model selection.

Final model fitted separately for all 6 service types.

4. Accuracy Evaluation

The SARIMAX forecasts were evaluated using:

MAE – Measures average prediction error

RMSE – Highlights large deviations

MAPE – Percentage-based accuracy measure (limited when actual values are zero)

Residual Diagnostics – Ensured model residuals had minimal autocorrelation

SARIMAX effectively captured weekday peaks and weekend drops across all services.

5. Conclusion

SARIMAX is a **robust and reliable** forecasting method for transportation data with weekly seasonality.
With proper tuning, it produced accurate short-term forecasts for all 6 service types.
Future work can extend the model by adding **holiday flags**, **school schedules**, and **weather data** to further improve prediction performance.

THANK YOU!