**Overview of Methodology**

To analyze ferry ticket sales over the past ten years, I adopted a structured, data science-driven approach:

1. **Data Cleaning & Integration**:
   * Loaded 15 minutes’ interval data of Ferry ticket sales. Aggregated to daily interval for better trend and seasonality analysis.
   * Web scraped daily weather data of ‘Toronto Ontario City’ from <https://climate.weather.gc.ca/>
   * Merged daily ferry ticket datasets (sales and redemptions) with daily weather records using a left join.
   * Handled non-numeric anomalies (e.g., "LegendMM", "LegendEE") and imputed missing values where necessary using surrounding averages. Checked for duplicates, outliers, inconsistencies and data-types.
2. **Exploratory Data Analysis (EDA)**:
   * Conducted trend analysis by day, month, and season to uncover usage spikes.
   * Visualized ticket activity against weather factors (mean temperature, precipitation).
   * Used heatmaps and correlation plots to assess relationships among key variables.
3. **Feature Engineering**:
   * Introduced flags for weekends, holidays, and season changes.
   * Calculated rolling averages to smooth short-term volatility.
4. **Visualization for Insights**:
   * Created seasonal line charts, calendar heatmaps, and Prophet forecast plot with labeled axes, legends, and confidence bands to support storytelling.
5. **Time Series Modeling**:
   * Performed **decomposition** to isolate trend, seasonality, and residuals.
   * Built forecasting model:
     + **Prophet** (because it takes note of seasonality): included **holiday effects** like Canada Day, Civic Holidays, and Christmas. Achieved **R² = 0.73**, **RMSE = 2178**, and **MAE = 1539**, indicating solid predictive performance.

**Assumptions Made**

* Historical sales patterns are consistent and predictive of future behavior.
* Holidays have a significant effect on demand — particularly Canada Day and long weekends.
* Weather impacts ferry demand linearly (e.g., warmer days’ correlate with higher sales).
* Operational changes (e.g., fare increases or schedule changes) are not explicitly available and thus not modeled.
* Covid-19 lockdown created a significant anomaly in sales and redemption count in year 2020-2021
* There was also decline in ticket sales in 2017 possibly due to internal restructuring.
* Assumptions were made for seasons. There could be some overlaps in due to climate change

**Next Steps (with More Data)**

I would recommend:

1. **Customer Segmentation**
   * Incorporate ticket types (adult, senior, group, etc.) or payment methods to tailor insights for different rider groups.
2. **Fleet Capacity and size data**
   * Include ferry capacity, delays, local events, or service changes to distinguish between demand-side and supply-side impacts for better fleet allocation and cost reduction while meeting demands.
3. **Advanced Forecasting Enhancements**
   * Fine tune models or try SARIMA or other advanced models for better forecast.
   * demands.
4. **Dashboard Monitoring**
   * Incorporate BI dashboards/reports to monitor Fleet Demands, Fleet in use/repair/break down/maintenance. Also incorporate total sales, demand, weather, precipitation and so on for optimal fleet service delivery.