**Maged Eid - Toronto Island Ferry Ticket Counts**

**Applied Scientist (Data Scientist) - (SX2)**

**Non-Technical Summary - (for Business Stakeholders)**

**What We Did:**

* Enhanced the existing ferry ticket forecasting system with five advanced prediction models
* Built a completely new forecasting system for ticket sales (previously unavailable)
* Added uncertainty estimates to help with capacity planning and risk management
* Tested all models using rigorous time-series validation to ensure real-world performance

**Key Improvements:**

* Significantly improved prediction accuracy compared to the basic seasonal model
* Added multiple forecasting approaches including machine learning and statistical methods
* Created an ensemble model that combines the best features of all approaches
* Implemented confidence intervals showing prediction uncertainty ranges

**Business Impact:**

* Better daily capacity planning with more accurate passenger volume forecasts
* Separate sales predictions help optimize ticket inventory and staffing
* Uncertainty estimates support risk-based decision making
* Automated model comparison identifies the best approach for different scenarios

**Practical Applications:**

* Daily operational staffing decisions based on predicted passenger volumes
* Inventory management for ticket sales across different channels
* Capacity planning for peak periods with confidence interval guidance
* Performance monitoring through automated model evaluation metrics

**Technical Implementation Report – (Detailed)**

**Problem Approach and Methodology**

The original forecasting system used only basic seasonal decomposition, extracting day-of-year patterns without considering trends, multiple seasonalities, or external factors. Our enhanced approach addresses these limitations through a comprehensive multi-model framework. This work included multiple trail and error and I included all of them till the best outcome were achieved after testing different approaches and techniques.

**Data Engineering Enhancements:** We expanded the feature engineering pipeline beyond basic monthly/quarterly indicators to include day-of-week patterns, weekend indicators, seasonal trigonometric features, and time-of-day patterns. For sales modeling, we incorporated redemption counts as a predictor variable, recognizing the business relationship between advance sales and actual ferry usage.

**Model Architecture:** The solution implements five distinct forecasting approaches: (1) Enhanced Seasonal Decomposition with linear trend extrapolation, (2) Triple Exponential Smoothing (Holt-Winters) with additive seasonality and damped trends, (3) ARIMA modeling for capturing autocorrelation patterns, (4) Random Forest regression leveraging engineered temporal features, and (5) Weighted ensemble combining all approaches.

**Uncertainty Quantification:** Statistical models (Holt-Winters, ARIMA) provide native prediction intervals through parameter uncertainty propagation. For machine learning models, we implement bootstrap-based confidence intervals. The ensemble approach uses weighted combinations optimized on validation performance.

**Validation Framework:** We employ time-series cross-validation with expanding windows to respect temporal dependencies, using multiple splits to assess model stability across different time periods. MAPE serves as the primary metric given its interpretability for business stakeholders, supplemented by MAE for absolute error assessment.

**Sales Model Innovation:** The sales forecasting component introduces a novel application of the same framework to ticket sales, incorporating redemption patterns as predictive features. This captures the business dynamic where advance sales patterns may differ from day-of-use redemption patterns.

**Technical Implementation:** The solution maintains the existing class structure while extending functionality through inheritance and composition patterns. All models include robust error handling with graceful fallbacks to simpler approaches when complex models fail to converge. The code follows reproducible research principles with fixed random seeds and comprehensive logging.

**Performance Optimization:** Model selection includes automatic hyperparameter optimization where computationally feasible, with ensemble weights determined through cross-validation performance. The framework supports easy addition of new models through the established interface pattern.

This comprehensive approach transforms a single-model system into a robust forecasting platform capable of handling multiple business scenarios with quantified uncertainty, providing the Toronto Island Ferry service with significantly enhanced predictive capabilities for operational planning.

Kind regards,  
**Maged Eid**