Written Assessment Report

Advertisement number: 44081

Job title: Applied Scientist

Department: Manitoba Public Service Commission

**Applicant: Chandan Saha**

Part 1: Plain Language Summary (≈200 words, bullet-style)

**Business Problem:**

* The city of Toronto wants to predict ferry ticket redemptions and sales to support operational planning (staffing, inventory, capacity).
* The goal is to anticipate demand more accurately using available timestamped transaction data.

**What I Did:**

* Built two models: one to forecast Redemption Count (actual boardings) and another for Sales Count (tickets sold).
* Used a machine learning approach (XGBoost) and compared with a simple baseline model.
* Created useful features from the timestamp: weekday, month, holiday, weekend indicator.
* Added more detailed features using past values (lag features) and trends (rolling averages, differences) to detect trends and spikes.
* Evaluated model performance using industry-standard metrics: MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error).

**Why It’s Better:**

* Original baseline model used only seasonality (day of year). It missed trends and recent behaviors.
* Both redemption count and sales count models reduced MAPE to under 10%, showing much better accuracy.
* These models would assist the city better to prepare for the busy days and optimize the ferry operations

**Part 2: Detailed Technical Summary (≈500 words)**

This project aimed to improve forecasting of ferry demand, using historical ticketing data from the Toronto Island Ferry service. Accurate forecasting is crucial for several operations, including scheduling ferries, planning for peak days, and staffing ticket booths. Two outcomes were modeled: Redemption Count (individuals who boarded the ferry) and Sales Count (tickets sold). The forecasting task has some challenges, including seasonal data, holiday fluctuations, day-of-week effects, and unexpected spikes because of special days.

The original baseline model was also simple, using only seasonal averages based on the day of year; however, it does not consider recent trends, holidays, and sudden demand changes.

To address this, two forecasting models for redemption and sales counts using XGBoost, a tree-based machine learning algorithm that performs well with structured time series data, have been developed that incorporate the following features:

* **Calendar-based features** included weekday, month, weekend, and official Ontario holidays (using the holidays package).
* **Lag features** included the sales or redemptions 1, 7, and 14 days prior.
* **Rolling features** included3-day and 7-day moving averages and 7-day standard deviation, giving the model a short-term pattern.
* **Difference features** measured the change in demand from the previous day and the previous week.
* **Log-transformation** of the target variable helped reduce the influence of extreme outliers (spikes), stabilizing model learning. T

Both model pipelines use TimeSeriesSplit to simulate real-world training and validation: the model is trained on earlier periods and tested on future periods, avoiding data leakage. For evaluation metrics, we used Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), widely used metrics for forecasting tasks.

The improvements were substantial. After implementing the XGBoost models with the above features, we found MAPE **8.7% for Redemption** and **9.9% for Sales**, with much lower MAE. This makes the models far more actionable for daily planning.

The following business benefits are expected to be achieved:

* Staffs can prepare for peak days without over-committing resources.
* Operations teams can detect early signs of surges or slumps in demand.
* Trends (e.g., weekend spikes or holiday dips) are detected automatically.

In conclusion, the models are designed to be interpretable and extensible. The forecasts help staff anticipate high-demand days and optimize resources. They also allow the City of Toronto to detect unusual usage trends, plan ferry capacities more accurately, and improve the rider experience. Furthermore, future improvements of these models could include integrating weather, special events, or economic variables. The potentiality of other models, such as Prophet and LSTM, could also be investigated.

**Disclaimer:**

The forecasting models and documentations (part 1 and part 2) were partially developed using ChatGPT (Models GPT-4o/o3) by OpenAI for coding support and explanation. The final implementation and review have been done by the applicant.