**Assumptions:**

**Stationarity**: We assume that the series is stationary, so it has consistent statistical properties over time.

**Linearity**: Based on the plots, we assume linear relationships between variables (both for redemption count and sales count)

**Data Quality**: Based on the website’s data quality report we assume that the data is accurate, consistent data without significant missing values.

**Stable External Environment**: We also excluded external factors so that redemption / sales are not influenced by other factors.

**Independent Residuals**: We assumed that the forecast errors are uncorrelated.

**Approach to the Business Problem**

To address the business objective of accurately forecasting daily redemption counts, I began with thorough data inspection and exploratory analysis to identify trends, seasonal patterns, and key correlations. I performed feature engineering to enrich the dataset with relevant calendar variables such as year, month, week, day of week, quarter, and season, alongside lagged and rolling features for both redemption and sales counts. Special attention was given to handling data uncertainties, irregular patterns and outliers were detected and removed, timestamps were aggregated to daily frequency, and appropriate metrics (refined MAPE and MSE) were selected given the presence of zero values in sales/ redemption count.

Patterns of pronounced seasonality and trend were confirmed through visualizations and autocorrelation analysis (ACF, PACF). I computed autocorrelations to identify relevant lag structures. The PACF plot showed direct correlation at each lag interval, aiding in identifying autoregressive terms. To prevent data leakage and ensure robust model assessment, I employed rolling and expanding window cross-validation throughout. Sales count can affect redemption count, and forecasting redemption helps authorities prepare for peaks. Similarly, previous redemption data aids in optimizing inventory and resources based on sales trends.

**Summary of Applied methods**

The modeling process started with a simple seasonal average model as a baseline. Statistical models – Auto ARIMA can capture shorter trend and ETS can deal with seasonality explicitly. So, statistical models such as ARIMA and Exponential Smoothing were utilized, but their performance was limited by the complex, seasonality present in the data and due to ignoring long term trends. The Facebook Prophet model was introduced to better handle dynamic trends, seasonal amplitude, holiday effects, and uncertainty intervals, and external regressors such as Canadian holidays and special events were incorporated for greater accuracy.

Advanced tree-based machine learning models including Decision Tree, Random Forest, XGBoost, and LightGBM were developed to learn dynamic and varying seasonal effects, capture nonlinear relationships and interactions among engineered features. Model hyperparameters were systematically tuned to maximize accuracy and minimize overfitting, and ensemble approaches were explored for greater robustness. I systematically tuned model parameters (max depth, no of trees, no of leaves, feature fraction, learning rate and regularization) using grid search to capture indirect uncertainty and to come up with custom models that are well tailored and generalized for the redemption / sales data.

**Why our approach is better: model Comparison and Results**

The baseline static averaging noisy model failed to adapt to dynamic seasonal patterns and long-term trends. In contrast, machine learning models demonstrated a significantly higher capacity for generalization and adaptability. Base model underperformed as this static baseline approach ignored uncommon dynamic seasonal patten (leap year / covid / shifting holiday) and long-term historical trend. Machine learning was able to generalize and adapt to recent behaviour and complex seasonal trend whereas the base model lacked learning power or intelligence to come up with flat seasonal predictions ignoring long term trend.

**Empirical results:**

Sales Model gained 44–50% improvement and Redemption Model gained 57–57.84% improvement in forecast accuracy using tree-based methods.

The methodology combining comprehensive feature engineering, ensemble machine learning, and cross-validation substantially improved time series forecasting accuracy for both sales and redemption counts, outperforming traditional baseline approaches and addressing the complexities inherent in real-world.

**Notes**

I used Google Collab for notebook-based coding and leveraged generative AI tool- Gemini to assist with code auto-completion and syntax suggestions, ensuring all implementations were written and customized by me.