Dr. B R Ambedkar National Institute of Technology
Jalandhar

IEA-502 Data Analytics

Project: Inventory Management of a Product

Submitted by:
Raj Sekhar Chakraborty
(24205307)

Submitted To: Dr. Ajay Gupta

MTech Industrial Engineering and Data Analytics

Industrial and Production Engineering Department

Introduction:

This report summarizes an inventory management project focused on optimizing the stock of food items within a specific retail store, CA_1. Effective inventory management is crucial for businesses to meet customer demand, minimize carrying costs, and prevent stockouts. This project leverages data analysis and forecasting techniques to gain insights into sales patterns, identify high-value and high-variability items, and ultimately improve inventory planning. The primary objectives include understanding sales trends, categorizing inventory based on revenue contribution and demand variability, and applying time series forecasting models to predict future demand.

1. Data Import and Initial Exploration:
    a. The project imports essential libraries like numpy and pandas.
    b. It then loads three datasets: calendar.csv, sales_train_validation.csv, and sell_prices.csv.
    c. Initial head() and info() calls are used to inspect the structure and data types of each DataFrame.

2. Data Merging and Preprocessing:
    a. The sales_train_validation and sell_prices DataFrames are merged.
    b. A crucial step involves calculating the actual sales for each item, likely by multiplying sell_price with demand.
    c. The calendar DataFrame is merged to incorporate date-related features and event information.
    d. Missing values (NaN) in event-related columns are handled, likely by filling them with 'No Event'.
    e. Date features such as day, month, year, weekday, and dayofyear are extracted from the date column to enrich the dataset for time series analysis.

3. Feature Engineering:
    a. **Lag Features:** The notebook generates lag features for sales (e.g., sales_lag_7, sales_lag_28) to capture past sales patterns, which are highly predictive in time series forecasting.
    b. **Rolling Mean Features:** Rolling mean features (e.g., rolling_mean_7, rolling_mean_28) are computed to smooth out short-term fluctuations and highlight trends.
    c. **Event Indicators:** Binary indicators are created for different event

types (event_type_1, event_type_2) to account for their impact on sales.

d. **Snap Day Indicators:** Features related to SNAP (Supplemental Nutrition Assistance Program) days in California, Texas, and Wisconsin are included, as these often influence consumer purchasing behavior.

4. Exploratory Data Analysis (EDA):
    a. **Sales and Revenue Analysis:** The notebook performs extensive EDA on sales and revenue data.
        i. It calculates total sales and revenue for different categories (e.g., FOODS, HOUSEHOLD, HOBBIES) and sub-categories.
        ii. Visualizations (e.g., bar plots) are used to show the distribution of sales/revenue across these categories.
        iii. Time series plots are generated to visualize daily sales trends for specific items or categories, often revealing seasonality and trends.
    b. **Event Impact Analysis:** The impact of various events on sales is analyzed, likely by comparing sales on event days versus non-event days.
    c. **Snapshot Day Impact:** The influence of SNAP days on sales is also investigated.

5. ABC Analysis:
    a. The project implements ABC analysis to categorize inventory items based on their revenue contribution.
    b. Items are ranked by cumulative revenue, and then classified into A, B, and C categories (e.g., A: top 80% revenue, B: next 15%, C: bottom 5%).
    c. This analysis helps in prioritizing inventory management efforts, focusing more attention on high-value 'A' items.

6. Demand Variability Analysis (Coefficient of Variation - CoV):
    a. The Coefficient of Variation (μσ) is calculated for sales of each item to measure demand variability.
    b. Items with higher CoV indicate more volatile demand, requiring different forecasting and inventory strategies (e.g., safety stock).
    c. The notebook likely categorizes items based on their CoV, identifying highly variable 'A' items that need special attention.

7. Time Series Forecasting:
    a. The notebook demonstrates the application of several forecasting

models, particularly for a high-volume, high-variability item (FOODS_3_578).
   b. **SARIMA (Seasonal Autoregressive Integrated Moving Average):** This model is used to capture seasonality and trends in the time series data.
   c. **Prophet:** Developed by Facebook, Prophet is another popular model for forecasting time series data, especially those with strong seasonal components and holidays.
   d. **XGBoost (Extreme Gradient Boosting):** A machine learning model is also employed, leveraging the engineered features (lag sales, rolling means, event indicators) to predict future sales.

8. Model Evaluation:
   a. The performance of the forecasting models is evaluated using appropriate metrics (e.g., Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE)).
   b. Visualizations (e.g., actual vs. predicted plots) are used to assess how well the models capture the actual sales patterns.

Conclusion:

The analysis effectively revealed significant variability in sales and revenue across different food items within the CA_1 store's FOODS category. The ABC analysis, based on revenue contribution, effectively categorized items, highlighting that a small percentage of "A" items contribute to a large portion of the revenue, while a significant number of "C" items contribute minimally. The demand variability analysis, using the Coefficient of Variation, further identified "A" category items with high volatility, indicating they might require closer inventory management and more sophisticated forecasting techniques. Time series forecasting models like SARIMA and Prophet showed promising results for forecasting the sales of a high-volume, high-variability item (FOODS_3_578), demonstrating their potential for capturing seasonality and trend. XGBoost, utilizing lag features and event indicators, also provided reasonable forecasts, emphasizing the importance of recent historical sales and special events in predicting demand. These insights can inform targeted inventory strategies, marketing efforts, and improved forecasting models for specific item categories.