

Estimate the Unit Sales of Walmart Retail Foods

Problem statement

This project aims to estimate the point forecasts of unit sales for ten food products sold by Walmart in three US States (California, Texas, and Wisconsin) and provides detailed insights at the item level and specific store attributes. The data contains 1913 days from 2011 to 2016, and the goal is to forecast daily sales for the next 28 days.

Context

Walmart, the world's largest company by revenue, operates numerous stores across the United States. The company aims to predict the daily sales of selected food items in three key states: California, Texas, and Wisconsin. Precise sales forecasts are essential for Walmart to maintain optimal inventory levels, preventing stockouts or overstocking. By leveraging these forecasts, Walmart can adopt data-driven approaches for inventory management, make informed pricing choices, and strategically plan promotions based on projected sales trends. This optimization enhances inventory turnover rates and boosts overall operational efficiency, enabling Walmart to maximize revenue while minimizing inventory holding expenses.

Data Source

The data are sourced from <https://www.kaggle.com/competitions/m5-forecasting-accuracy> and consists of the following four files: calendar.csv, sell_prices, sales_train.csv, and sales_train_evaluation.csv.

- calendar.csv - Contains information about the dates on which the products are sold (2011-2016).
- sell_prices.csv - Contains information about the price of the products sold per store and date.
- sales_train.csv - Contains the historical daily unit sales data per product and store for day-1 to day-1913.
- sales_train_evaluation.csv - Includes sales data for day-1 to day-1941 (contains the next 28 days).

This data contains three categories (Hobbies, Food, and Household) in various departments. For this project, we only use 10 items from the Food category.

Criteria for success

- Determine the key features within the sales data that significantly influence the unit sales of food products.
- Build robust forecasting models based on the identified key features to estimate daily sales for the next 28 days with minimal error.
- Provide detailed insights at both the item level and specific store attributes.
- Demonstrate the effective use of historical sales data spanning 1913 days from 2011 to 2016 to improve forecasting accuracy and enhance decision-making processes.

Scope of solution space

- Conduct an in-depth exploration of the sales data to understand trends, patterns, and correlations.
- Implement advanced time series forecasting models, such as ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), Prophet, or others to forecast daily sales for the next 28 days.
- Fine-tune the models to capture seasonality, trends, and any irregular patterns in the sales data.
- Evaluate the performance of the time series models using appropriate metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE).

Deliverables

The project deliverables will include Jupyter Notebooks containing each step, methods used, and codes to support the analysis of the project. In addition, the final report will be presented as a slide deck and project report.