Walmart Sales Forecasting Data Science Project

Description

Every Departmental store chain like Walmart wants to predict the store sales in the future so that inventory planning can be done. Along with that, sales prediction helps to increase/decrease store staff based on the rush (More sales can mean more customers are coming to the stores). Also, it is always a good idea to do sales and revenue forecasting to better understand the company's cash-flows and overall growth.

For inventory planning, you also need to know what products (or category of products aka department) will be utilized more. Under-stock some products and your sales are hit. Over-stock items like perishables and you run into losses if the product expires. That's why the sales prediction is done at a combination of store and department level (and sometimes even at product level for high-selling products).

Objective of Retail Analysis with Walmart Data Project

In this sales forecasting project, we will use the historical sales data of 45 Walmart stores based on store location, department, and week. Each store's land area and type have been provided, and holiday weeks have been marked. Along with these, price markdown data (almost like discount data) has been given. A few macro-indicators like CPI, Unemployment rate, Fuel price, etc., are also provided. Our goal is to analyze the **Walmart sales data** and predict their stores' sales for all the weeks in a year.

Data Science Techniques used for Walmart Store Sales Forecasting

For this **Walmart project**, we will use exciting data science tools and methods that data scientists use almost every day in their job. Here is an overview of the techniques used in this project.

Exploratory Data Analysis (EDA)

Handling large amounts of data is not an easy task. Before deciding upon which mathematical methods to use, one must draw statistical inferences from the data. Parameters like mean, median, mode, minimum and maximum value, etc., assist in understanding which models should be used for preprocessing. This project will use the **Walmart dataset** and plot various graphs to understand the overall distribution of different variables using Python libraries like NumPy, Pandas, and Seaborn.

Data Preprocessing

Once the data has been analyzed, the next task is to prepare the data for applying machine learning (ML) models. Many negative values do not make sense in the real world. Data scientists use methods like averaging the neighborhood values, making them all zero, adding a specific number to all the variables, etc., to pre-process data. In this **Walmart sales prediction project**, you will learn how to handle negative values, missing values, and duplicate values

of different variables. Additionally, you will learn how to convert variables of varying data types into a specific data type for easy application of ML algorithms.

Data Merging

Information is often stored in separate condensed data files to save on hardware. A data scientist then merges the different data files to draw inferences. In this **retail analysis with Walmarts dataset project**, you will merge three separate data files for creating testing and validation datasets. The dataset will then be used for sales prediction.

Statistical Modelling

If we look at the dataset in this project at a more granular level, one can quickly build a Walmart sales forecasting model using elementary statistics. This project will guide you on choosing specific variables for forecasting sales with the help of more straightforward statistical methods.

Time Series Forecasting

After implementing essential statistical tools for this Walmart project, you will learn about various time series forecasting models as this project aims to predict sales with respect to time. The method discussed in detail in this project solution is the AutoRegressive Integrated Moving Average (ARIMA) model. You will get to explore how to prepare the dataset for implementing the ARIMA model and draw valuable inferences.

Machine Learning Algorithms

The most crucial part of any data sci project is the implementation of a machine learning/deep learning model. You will learn how to implement regression models like the random forest for sales forecasting using the Walmart dataset in this project. You will learn how to perform hyperparameter tuning for training the model efficiently.

Comparing Different Machine Learning Models

Finally, the project will help you prepare the **Walmart sales prediction project report** by comparing the performance of different models discussed above. You will use a few statistical formulae to analyze their performance and some metrics to choose the best model.

Frequently Asked Questions on Walmart Data Science Project

What type of forecasting does Walmart use?
Walmart uses its data for time series analysis and prediction methods for forecasting its sales across its different stores across the United States. As per the company's distinguished data scientist and director of data science, John Bowman, Walmart estimates the predictions for a full 52-week horizon in weekly increments, and produces a new set of forecasts every week, over the weekend. Additionally, the company has a 12-hour window to complete all of the forecasting tasks, and about three days to evaluate all of the training tasks. The methods include Deep Learning models as well and Walmart also assesses the forecasts which it has received from its vendors for comparison and evaluation.

What are the methods of sales forecasting?
Here are a few methods used for sales forecasting:

- Time-series forecasting models like ARIMA.
- Machine learning models like Random forests.
- Deep Learning models like Neural Prophet and SilverKite

What we will learn in this project

1. Understand the Problem Statement
2. Perform basic EDA to familiarize with the data
3. Take care of missing values and datatype issues in the data
4. Understand the unique key in different data and merging the data
5. Perform Univariate analysis for both numeric and categorical variables
6. Perform Bi-variate analysis to identify redundant variables
7. Plot Trend of each predictor with the target variable
8. Do in-depth analysis on the impact of Date/Week on Sales
9. Create new features that might add value to the model
10. Define a function for each set of code that might need to be repeated again
11. Prepare the data for modelling
12. Make prediction using statistical techniques
13. Make model using machine learning techniques
14. Create time series ARIMA models and learn to give their parameters
15. Perform Hyper-parameter tuning to get the best parameters
16. Learn how to make predictions where data is sparse
17. Compare the performance of different models using multiple metrics