Sales Forecasting

1. Introduction:
   * Sales forecasting is an essential component of any business that deals with sales and inventory. Accurate sales forecasting can help companies optimize their inventory levels, plan promotions, and ultimately increase profits. In this case study, we will be using the Walmart Sales Dataset of 45 stores to build a sales forecasting model.
   * Background:

Walmart is one of the world's largest retail chains, serving millions of customers every week across the globe. With over 11,000 stores worldwide, Walmart's revenue in 2020 exceeded $500 billion. As a retail giant, Walmart faces many challenges, such as managing inventory, optimizing supply chain management, and accurately forecasting sales.

* + Purpose:

The purpose of this case study is to build a sales forecasting model that can accurately predict future sales for Walmart stores. This model will help Walmart optimize its inventory levels, plan promotions, and increase profits by ensuring that they always have the right amount of stock to meet customer demand.

* + Research Questions:

The following research questions will guide our analysis:

* + - What are the key factors that influence weekly sales at Walmart stores?
    - How accurate can we forecast sales for Walmart stores using historical sales and other relevant data?
    - How can Walmart use the sales forecasting model to optimize its inventory levels and plan promotions?
  + Significance:

The significance of this case study lies in its potential to improve Walmart's sales forecasting capabilities, which in turn can lead to increased profits. Accurate sales forecasting can help Walmart optimize its inventory levels, which means they can reduce the amount of overstock and understock situations, and ultimately reduce waste. Additionally, a better understanding of the key factors that influence weekly sales can help Walmart make informed decisions about inventory management and promotional planning.

1. Business Problem:
   * As one of the world's largest retailers, Walmart needs to manage a vast amount of inventory across its numerous stores. Accurate sales forecasting is essential to ensure that the right products are available in the right stores at the right time. Walmart's inventory management and supply chain operations are complex, involving many factors that can impact sales, including weather conditions, store location, promotions, and pricing. To address these challenges, Walmart needs to build a robust sales forecasting model that can accurately predict future sales.
   * The business problem is to develop an accurate sales forecasting model that can help Walmart improve its inventory management, optimize its supply chain, and enhance its pricing strategies. An accurate sales forecasting model can help Walmart ensure that it has enough inventory to meet customer demand while avoiding overstocking, which can lead to waste and losses. Additionally, an accurate sales forecasting model can help Walmart optimize its supply chain by ensuring that products are delivered to stores in a timely and cost-effective manner. Finally, a reliable sales forecasting model can also help Walmart improve its pricing strategies by providing insights into the impact of different pricing strategies on sales.
   * To address these challenges, Walmart has provided a dataset containing historical sales data for 45 of its stores, along with information on store location, the size of the store, the type of store, and various promotional events that occurred during the sales period. The goal of this case study is to use this data to build an accurate sales forecasting model that can help Walmart achieve its business objectives.
2. Data Description:
   * Data Source:

‘sales-forecasting.csv’ - Walmart Recruiting - Store Sales Forecasting is a publicly available dataset on Kaggle (<https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data?select=stores.csv>). The dataset was released as part of a recruiting competition organized by Walmart Labs to identify talented data scientists who could help Walmart develop better sales forecasting models.

* + Data Characteristics:

The Walmart Recruiting - Store Sales Forecasting dataset contains historical sales data for 45 Walmart stores located in different regions of the United States. The data covers a period of two years, from 2010 to 2012, and includes information on store location, the size of the store, the type of store, and various promotional events that occurred during the sales period. The dataset contains a total of 421,570 records, with each record representing the sales of a particular item in a particular store on a particular date.

* + Data Dictionary:

For this case study there are 3 datasets:

* “sales-forecasting.csv”
* “stores.csv”
* “features.csv”

The datasets include the following variables:

“sales-forecasting.csv”

* “Store”: The unique store number (numeric)
* “Dept”: The unique department number (numeric)
* “Date”: Whether the customer is a senior citizen or not (date)
* “IsHoliday”: Whether the customer has a partner or not (categorical: ‘FALSE’, ‘TRUE’)
* “Weekly\_Sales”: target (numeric)

“stores.csv”

* “Store”: The unique store number (numeric)
* “Type”: The type of store (categorical: ‘A’, ‘B’,’C’)
* “Size”: The size of the store in square feet (numeric)

“features.csv”

* “Store”: The unique store number (numeric)
* “Date”: Whether the customer is a senior citizen or not (date)
* “Temperature”: Average temperature in the region (numeric)
* “Fuel\_Price”: Cost of fuel in the region (numeric)
* “CPI”: Consumer Price Index (numeric)
* “Unemployment”: Unemployment rate (numeric)
* “IsHoliday”: Whether the customer has a partner or not (categorical: ‘FALSE’, ‘TRUE’)

Some important notes on the variables:

* The “Weekly\_Sales” variable is the response variable in this case study.
* The Holiday\_Flag variable indicates whether the week of sales includes a holiday or not.
* The Type variable indicates the type of store, with Type A stores being the largest and Type C stores being the smallest.
* The Size variable indicates the size of the store in square feet, with larger stores generally having higher sales.

1. Data Science Process:
   * Data Preparation:
     + The first step is to gather and clean the data. This involves loading the data into a data frame and checking for missing values and data inconsistencies. In this case, we need to merge the given datasets and also create some additional features from the existing ones. Once we have clean data, we can proceed with exploratory data analysis.
   * Exploratory Data Analysis (EDA):
     + In this stage, we will perform exploratory data analysis to gain insights into the data and identify patterns or relationships between variables. We will use various statistical and visualization techniques to examine the distribution of the variables, the correlation between the variables, and any potential outliers or anomalies.
   * Feature Engineering:
     + In this stage, we will create new features or transform existing features to improve the performance of the model. This includes creating interaction terms, transforming variables, and scaling or standardizing the features.
     + We will also conduct feature selection to identify the most important variables that contribute to the prediction of term deposit subscription. We will use techniques and feature importance from the selected model to select the most relevant features.
   * Model Selection:
     + Once we have created the necessary features, the next step is to select an appropriate model to predict future sales. There are several models that can be used for time series forecasting, including ARIMA, XGBoost, Prophet, LSTM, and AutoML. We can evaluate the performance of each model on the training data and select the one that provides the best results.
   * Model Evaluation:
     + The selected model is evaluated on the test set to estimate its performance on unseen data. This involves computing evaluation metrics like RMSE (root mean squared error), MAE (mean absolute error), and MAPE (mean absolute percentage error) to compare the predicted values with the actual values. If the model does not perform well, we may need to revisit the previous steps and make changes to the data or model.
   * Model Deployment:
     + The final step is to deploy the model into production. This involves setting up a pipeline that can take new data and feed it into the trained model to make predictions. The model needs to be monitored over time to ensure that it continues to perform well and any necessary updates can be made.

In conclusion, the process of sales forecasting for Walmart Recruiting - Store Sales dataset involved multiple steps such as data preparation, exploratory data analysis, feature engineering, model selection, evaluation, and deployment. The dataset provided valuable insights into the sales trends, customer behavior, and external factors affecting the sales of the Walmart stores. The machine learning models developed for sales forecasting showed promising results. Overall, this case study demonstrates the importance of data-driven decision making and the effectiveness of machine learning techniques in solving real-world business problems.