|  |
| --- |
|  |
| Sale Forecasting for Walmart  Forecasting for 12 weeks |
| |  |  |  | | --- | --- | --- | | SOUVIK SIKDER | 12/17/23 | Intellipaat | |

# TABLE OF CONTENTS: -

**1. Problem Statement**

**2. Project Objective**

**3. Data Description**

**4. Data Pre-processing Steps and Inspiration**

**5. Choosing the Algorithm for the Project**

**6. Motivation for Choosing the Algorithm**

**7.Sales Forecasting Of Different Models for Definite Period**

**8. Model Evaluation and Techniques**

**9. Inferences from the Same**

**10. Future Possibilities of the Project**

**11. Conclusion**

**12. Reference**

### *PROBLEM STATEMENT: -*

A retail store that has multiple outlets across the country are facing issues in managing the inventory - to match the demand with respect to supply. You are a data scientist, who has to come up with useful insights using the data and make prediction models to forecast the sales for X number of months/years.

*PROJECT OBJECTIVE: -*

Using Walmart dataset, come up with useful insights to improve sales in each and every store of Walmart. Also forecast sales for last 12 weeks.

*DATA DESCRIPTION: -*

Walmart dataset has 6435 rows and 8 columns. Columns description is as follows-

* Store-Store Number
* Date-Week of Sales
* Weekly Sales-Sales for a given store in that week.
* Holiday Flag- If it is a Holiday week.
* Temperature-Temperature of the day.
* Fuel Price-Cost of fuel in that place.
* CPI-Consumer Price Index
* Unemployment-Unemployment Rate.

*DATA PREPROCESSING: -*

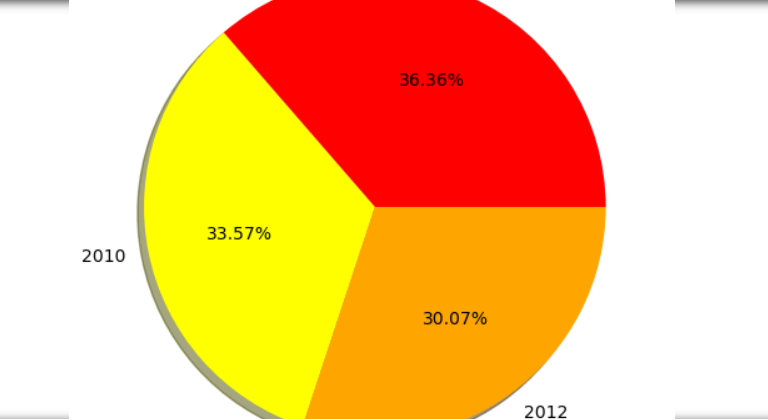
* Check for null values and found none.
* Statistical measures were seen for the dataset like count,min,max, average.
* Convert Date column to datetime format.
* Extract month, week, year, day from the date column and finally drop date column.
* For each column we find out the unique values for each column.

*VISUALIZATION: -*

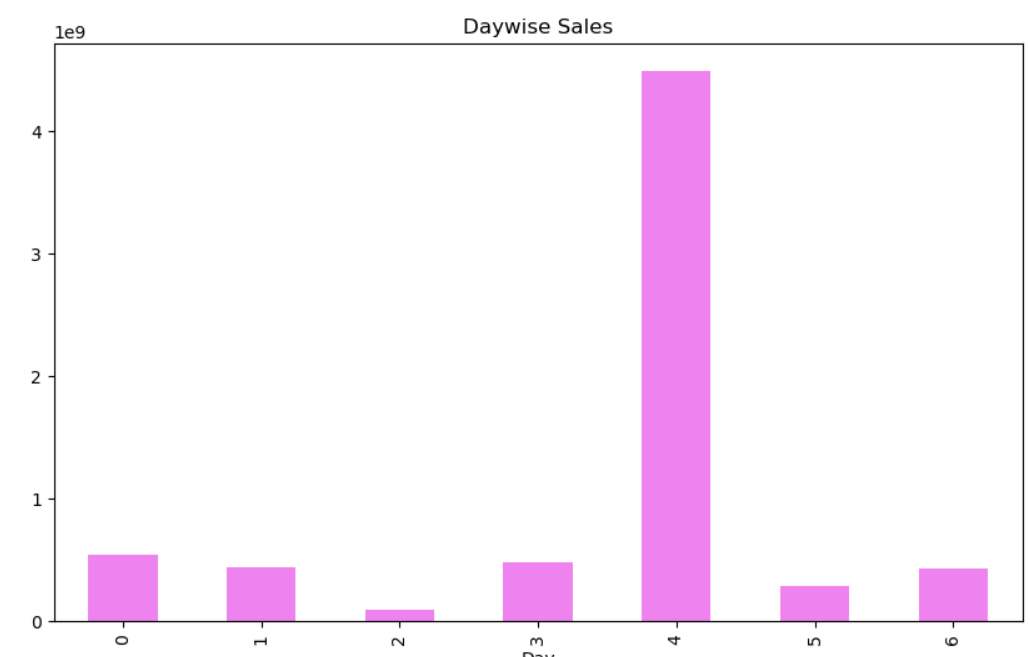
* Find annual sales for each year-2010,2011,2012. Annual sales in the year 2011 is very good compared to 2010 and 2011
* Store wise weekly sales, Monthly sales, yearly sales, day wise sales were found out and insights found out that store number 4,13,14,20,27 is doing exceptionally well.
* Month wise sales has definite improve in the month of April.
* Day wise sales has sales maximum in the 4-th day of the week.

All graph and charts are depicted in next page.

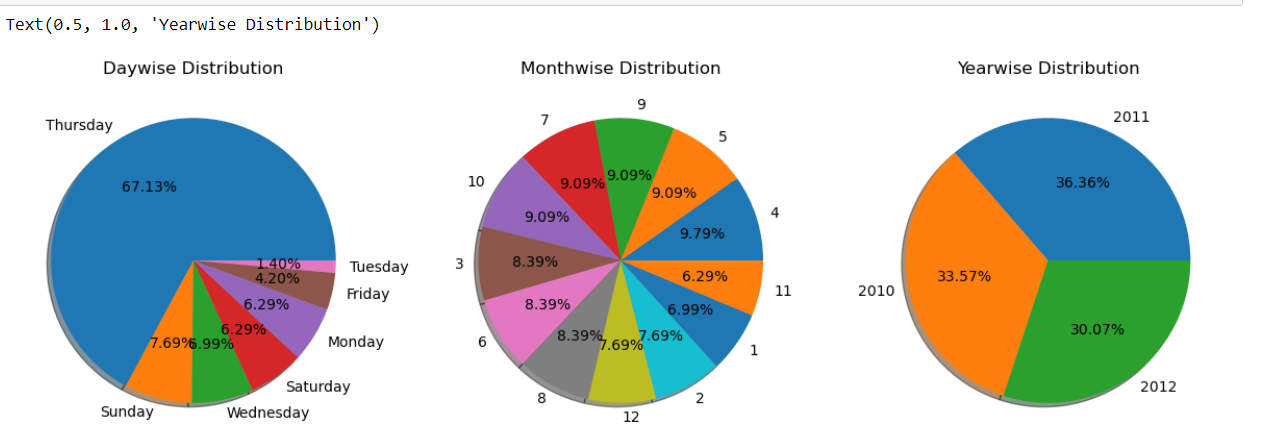
# Important Graphs and Charts: -



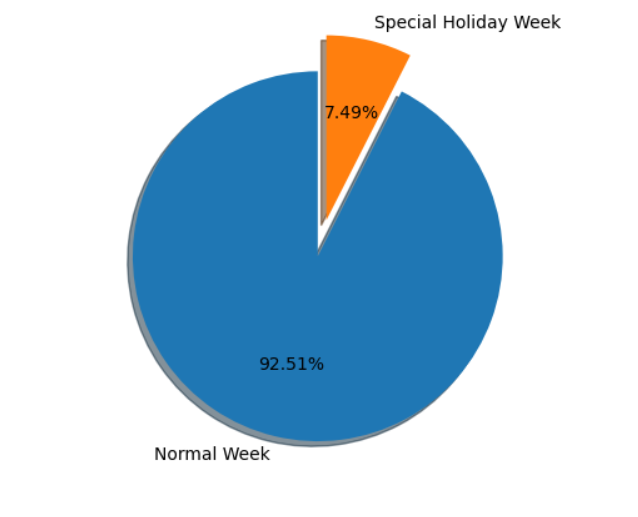
* Pie chart showing maximum sales in the year 2011.



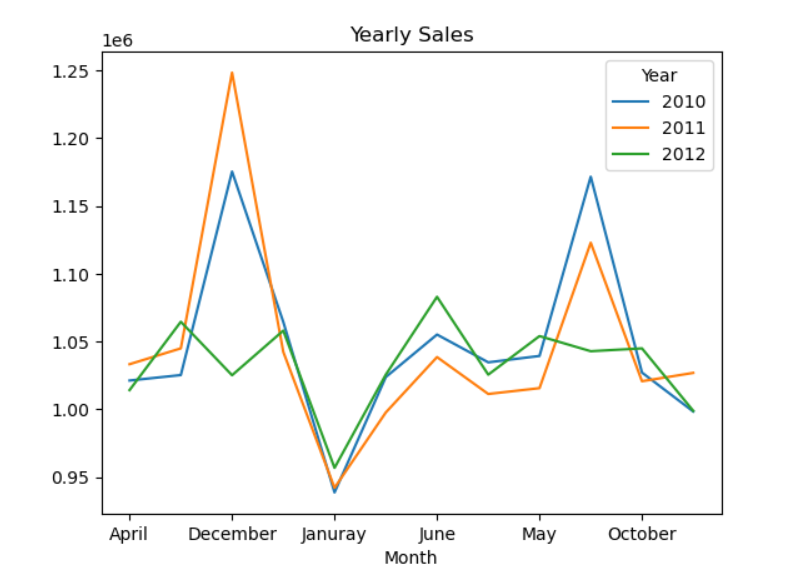
* Day wise sales showing maximum sales in 4th day of the week.



* Day wise, Month wise, Year wise sales for Walmart.



* Normal Week And Special Holiday Week Sales



* Sales across year showing how it changes across month.
* Minimum Sales in the Walmart: 37160221.96
* Maximum Sales in the Walmart: 299066335.2575

# Choosing Algorithm for The Project: -

**Linear Regression: -**

* **Algorithm Type:** Linear regression is a type of linear model, making predictions based on a linear relationship between the independent variables (features) and the dependent variable (target).
* **Working Principle:** It assumes that the relationship between the variables is linear and tries to find the best-fitting line through the data points.
* **Pros and Cons:**
  + *Pros:* Simple, interpretable, computationally efficient for large datasets with a linear relationship.
  + *Cons:* Assumes a linear relationship, sensitive to outliers.

**Random Forest Regression:**

* **Algorithm Type:** Random Forest is an ensemble learning method, specifically a bagging technique.
* **Working Principle:** It builds multiple decision trees and combines their predictions to improve accuracy and generalization. Each tree is constructed using a random subset of the data and features.
* **Pros and Cons:**
  + *Pros:* Robust to overfitting, handles non-linear relationships well, less sensitive to outliers.

**Gradient Boosting Regressor:**

* **Algorithm Type:** Gradient Boosting is an ensemble learning method, specifically a boosting technique.
* **Working Principle:** It builds a series of weak learners (usually decision trees) sequentially, with each tree correcting the errors of the previous ones. It combines their predictions to create a strong learner.
* **Pros and Cons:**
  + *Pros:* High predictive accuracy, handles complex relationships, less prone to overfitting.
  + *Cons:* Can be computationally expensive, sensitive to hyperparameters.

# **SALES FORECAST FOR DIFFFERENT MODELS: -**

We are predicting sales for next 12 weeks using different models and showcasing the result of sales for each model below: -

Linear Regression-

[629590.45370633 633329.55009985 636429.06606315 634124.34344998

705278.83793917 641649.04500406 644861.32315798 644470.24775963

644335.87339788 654422.97369447 652084.829575 648776.70397814]

**Random Forest Regressor-**

[729838.5284 736645.7081 752438.0586 743977.9908 799472.0967 735649.88

734454.1493 737796.337 736966.5064 740072.3351 734679.915 735416.0455]

Gradient Boosting Regressor-

[762979.17814089 753172.85170758 753172.85170758 753172.85170758

806224.66100662 753172.85170758 753172.85170758 753172.85170758

753172.85170758 751987.42221738 753172.85170758 753172.85170758]

# Model Evaluation: -

Each model is evaluated and model having mean absolute error is chosen as best algorithm to predict sales for Walmart retail shop in different stores. Certainly Random Forest Regressor is chosen as the best algorithm.

* **Mean Absolute Error of Random Forest Regressor: 17844.50442499996**
* **Mean Absolute Error of Gradient Boosting Regressor:** **29546.8527764958**
* **Mean Absolute Error of Gradient Boosting Regressor:** **82504.0501811969**

# Future Possibilities of the Project:-

The future possibilities for a Walmart sales forecasting project can be diverse, incorporating advancements in technology, data analytics, and artificial intelligence. Here are some potential avenues for future development:

1. **Advanced Predictive Analytics:**
   * **Machine Learning Improvements:** Explore more advanced machine learning models and algorithms for sales forecasting. This may involve using deep learning techniques or more sophisticated ensemble methods to capture complex patterns in the data.
2. **Incorporating External Data:**
   * **External Factors:** Integrate additional external data sources, such as economic indicators, weather data, and local events, to enhance the accuracy of sales forecasts. This can help the model adapt to various influencing factors.
3. **Real-Time Forecasting:**
   * **Real-Time Updates:** Develop real-time forecasting capabilities to provide up-to-the-minute insights into sales trends. This could involve streaming analytics and continuous model updating to adapt to changing market conditions promptly.
4. **Enhanced User Interface and Visualization:**
   * **Interactive Dashboards:** Build user-friendly dashboards that allow stakeholders to interact with the data, explore different scenarios, and gain deeper insights. Visualization tools can aid in understanding patterns and making informed decisions.
5. **Supply Chain Integration:**
   * **End-to-End Integration:** Integrate the sales forecasting system with the supply chain management system to optimize inventory levels, reduce stockouts, and improve overall supply chain efficiency.
6. **Customer Segmentation:**
   * **Personalized Forecasting:** Implement customer segmentation techniques to tailor sales forecasts based on the behavior and preferences of different customer segments. This can lead to more personalized marketing and inventory strategies.
7. **AI-driven Automation:**
   * **Automated Decision-Making:** Explore the use of artificial intelligence for automated decision-making. This could involve automated reordering systems, dynamic pricing strategies, and other AI-driven processes to optimize sales and inventory management.
8. **Feedback Loops and Continuous Improvement:**
   * **Model Refinement:** Establish feedback loops to continuously improve the forecasting model. Regularly update the model based on new data, monitor its performance, and refine it to ensure it remains effective over time.
9. **Blockchain for Transparency:**
   * **Blockchain Technology:** Consider leveraging blockchain technology to enhance transparency in the supply chain. This can improve traceability and accountability, reducing the risk of fraud and errors.
10. **Collaboration and Partnerships:**
    * **Collaboration with Tech Partners:** Explore collaborations with technology partners, startups, or research institutions to stay at the forefront of innovation in sales forecasting and related technologies.
11. **Integration with Emerging Technologies:**
    * **IoT and RFID Integration:** Utilize Internet of Things (IoT) and Radio-Frequency Identification (RFID) technologies to gather real-time data on product movement and consumer behavior, contributing to more accurate sales forecasts.

# Conclusion: -

In summary, the future possibilities for a Walmart sales forecasting project involve embracing advanced technologies, incorporating external factors, enhancing user interfaces, integrating with supply chain processes, and maintaining a focus on continuous improvement and innovation. The dynamic nature of the retail industry and technological advancements provide numerous opportunities for growth and optimization in sales forecasting.

# References: -

* [User guide and tutorial — seaborn 0.12.2 documentation (pydata.org)](https://seaborn.pydata.org/tutorial.html)
* [Data Science Courses Online | Learn Fundamentals to Advanced | Intellipaat](https://intellipaat.com/course-cat/data-science/)