

A PROJECT REPORT ON
“Walmart Time Series Forecasting”

SUBMITTED TO

Department of Computer Science & Engineering (Data Science) in fulfillment of Project Based
Learning for the semester-V of academic year 2023-2024

SUBMITTED BY

Sr.No	RollNo	Name of the Student
1	01	Bhakti Ayarekar
2	10	Ankita Yadav
3	21	Rohan Chopade
4	29	Soham Mangore

UNDER THE GUIDANCE OF

Mr. Vinay Prabhavalkar

Asst.Professor



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

SPECIALIZATION IN DATA SCIENCE

KIT'S COLLEGE OF ENGINEERING,

KOLHAPUR.

YEAR : 2023-2024

CERTIFICATE



KIT's COLLEGE OF ENGINEERING

This is to certify that, the project entitled “Walmart Time Series Forecasting”, has been satisfactorily completed by,

Bhakti Ayarekar (1)
Ankita Yadav (10)
Rohan Chopade (21)
Soham Manogre (29)

The students of SY B.Tech, Department of Computer Science & Engineering, Specialization in Data Science in fulfillment of PBL Project for the semester-V of academic year 2023-2024.

This project report is a record of student's own work carried by him/her under my supervision and guidance in satisfactory manner.

Date: 28/11/2023

Mr. Vinay Prabhavalkar

Asst Professor

Dr. Uma Gurav

HOD

Dr. Mohan Vanarotti

Director

Kolhapur Institute of Technology's
College of Engineering, Kolhapur.

Year 2023-2024

ACKNOWLEDGEMENT

We are highly grateful to the Dr. Uma Gurav, HOD CSE (AIML & DS), KIT's College of Engineering, Kolhapur, for providing this opportunity to carry out the Project Based Learning at CSE(AIML & DS) department. We would like to express our gratitude to other faculty members of department for providing academic inputs, guidance and encouragement throughout this period. We would like to express a deep sense of gratitude and thank to Dr. Uma Gurav and Mr. Vinay Prabhavalkar without whose permission, wise counsel and able guidance, it would have not been possible to carryout our project in this manner.

We express our indebtedness to all who have directly or indirectly contributed to the successful completion of Project Based Learning.

Date:28/11/2023

Place: KIT, Kolhapur

Sincerely by,

Name	Roll No
Bhakti Ayarekar	DS01
Ankita Yadav	DS10
Rohan Chopade	DS21
Soham Manogre	DS29

Project Overview

This project encompasses a systematic approach to time series analysis using Walmart data sourced from Kaggle. The methodology begins with meticulous data collection and preparation, involving the acquisition of time series data from Kaggle, followed by rigorous cleaning and processing to ensure high data quality. Utilizing **Python** and libraries such as Pandas, NumPy, Matplotlib, and Seaborn, an extensive exploratory data analysis (EDA) is conducted to unveil patterns and insights, visually represented through various graphs and charts. Feature engineering, facilitated by the Sci-kit Learn library, enhances model performance and captures crucial information. The subsequent step involves the application of **ARIMA**, **SARIMA**, and **FB Prophet** models, with the data split into training and testing sets to comprehensively capture different aspects of the time series. Visualizations are then dynamically presented using the **Plotly** library for forecasting. To facilitate efficient storage and retrieval, a **Postgre SQL** system is implemented as the backend database. The project culminates in the creation of an insightful and comprehensive dashboard using **PowerBI**, seamlessly integrating data rendered from Postgre SQL for a holistic analysis of the time series trends and predictions.

INDEX

SR.NO	CONTENTS	PAGE NO.
1	Introduction	6
2	Problem Statement	6
3	Objectives	6
4	Methodology	6
5	Forecasting Algorithms	8
6	Flowchart	8
7	Snapshot	9-10
8	Observations & Suggestions	11
10	Conclusion	12

INTRODUCTION

As many of you are aware, Walmart is a global retail giant with a vast network of stores and a wide range of products. Effective inventory management and sales prediction are critical for Walmart's operations. Our project focuses on leveraging time series forecasting techniques to optimize these processes.

PROBLEM STATEMENT

The objective involves analyzing historical sales data to unveil intricate temporal patterns and construct resilient forecasting models. The goal is to empower Walmart to achieve ideal inventory levels, make strategic pricing decisions, and maximize resource efficiency.

OBJECTIVES

- Accurate Sales Forecasting: Develop a reliable time series forecasting model that can accurately predict future sales for Walmart products based on historical data.
- Optimize Inventory Levels: Utilize the forecasting model to optimize inventory levels across Walmart's network of stores, ensuring that stock levels are aligned with anticipated demand.
- Enhance Decision-Making: Provide Walmart's management with valuable insights and forecasts to make informed decisions regarding marketing strategies, and store operations.
- Improve Customer Service: Ensure that products are consistently available to meet customer demand, enhancing the overall shopping experience and customer satisfaction.

METHODOLOGY

1) Data Collection & Preparation:

- Source: Kaggle
- Collection: Acquired Time Series data from Walmart on Kaggle
- Preparation: Data Cleaning and formatting, handling Null values, and ensuring data quality.

2) Exploratory Data Analysis (EDA)

- Tools: Python (Pandas, Numpy, Matplotlib & Seaborn)
- Analysis: Conducted throughout EDA, exploring patterns, trends, and insights using various graphs and charts.

3) Feature Engineering:

- Library: Sci-kit Learn
- Enhancement: Engineered relevant features to improve model performance and capture additional information.

4) Model Training:

- Models: ARIMA, SARIMAX, FB Prophet.
- Procedure: Split data into training and testing sets, and trained time series models to capture different aspects of the data.

5) Forecasting and Visualization:

- Library: Plotly
- Visualizationions: Utilized Plotly for dynamic and interactive visualizations of time series forecasts.

6) Database:

- System: PostgreSQL
- Integration: Implemented Postgre SQL on the backend for efficient storage, retrieval, and management of time series data.

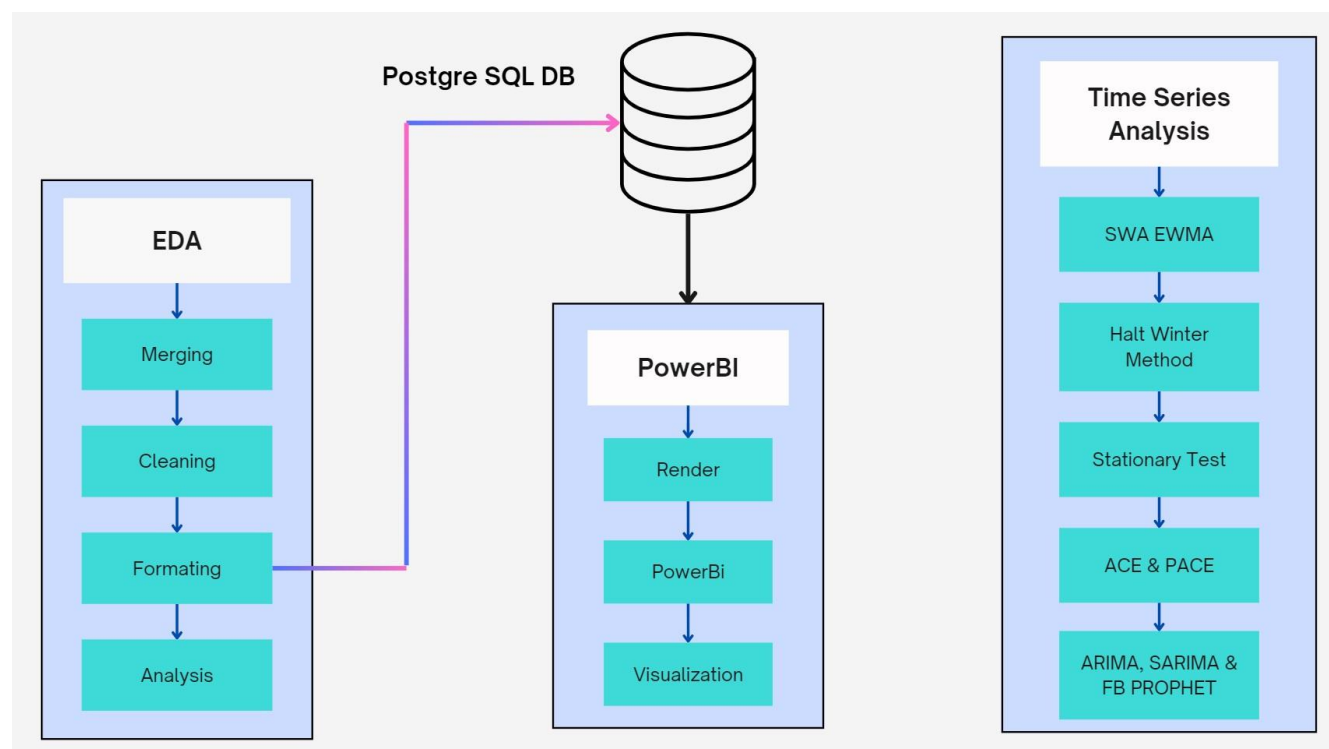
7) Dashboard:

- Tool: PowerBI
- Integration: Rendered data from Postgre SQL in PowerBi to create an insightful dashboard for comprehensive analysis.

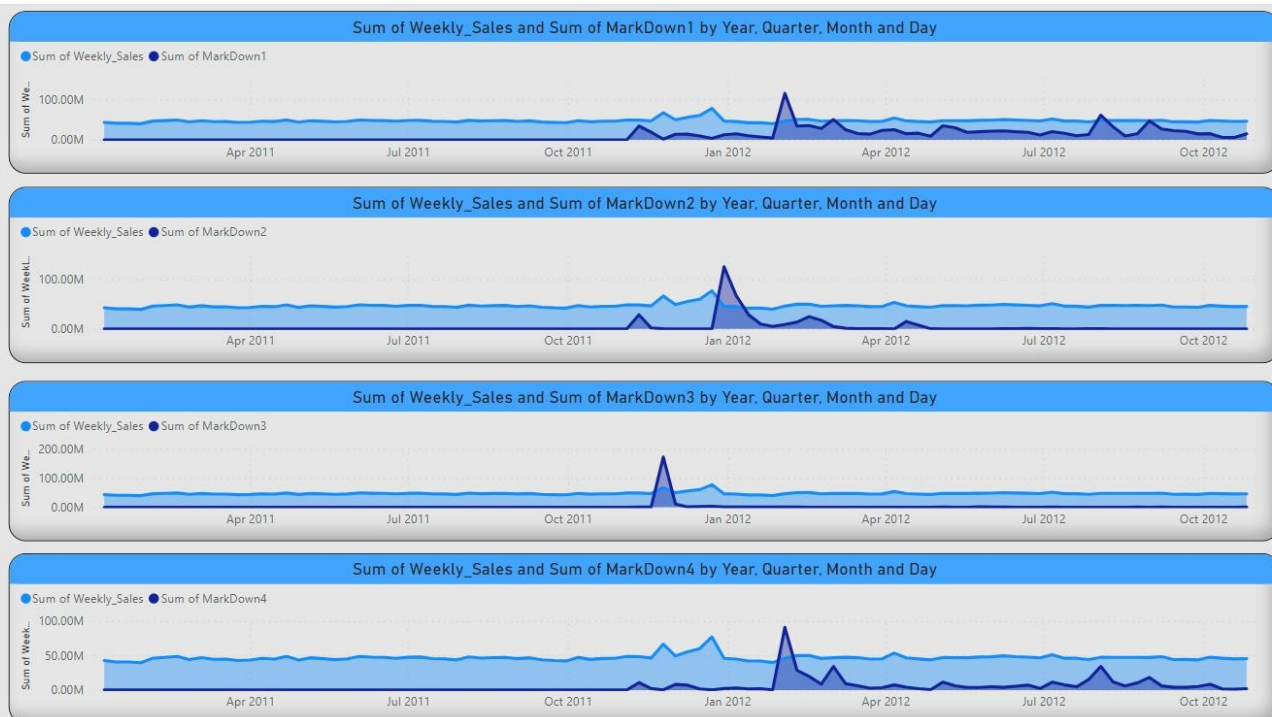
FORECASTING ALGORITHMS

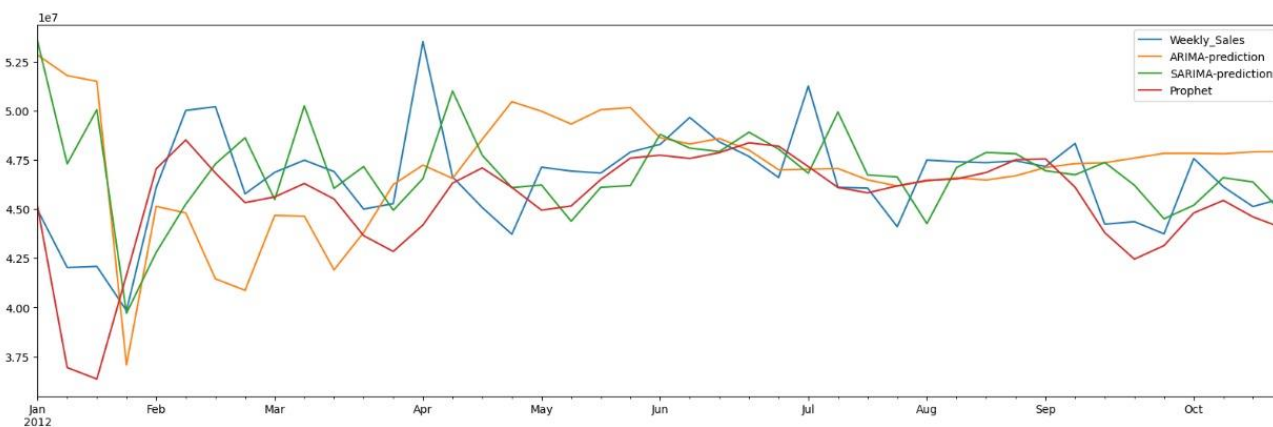
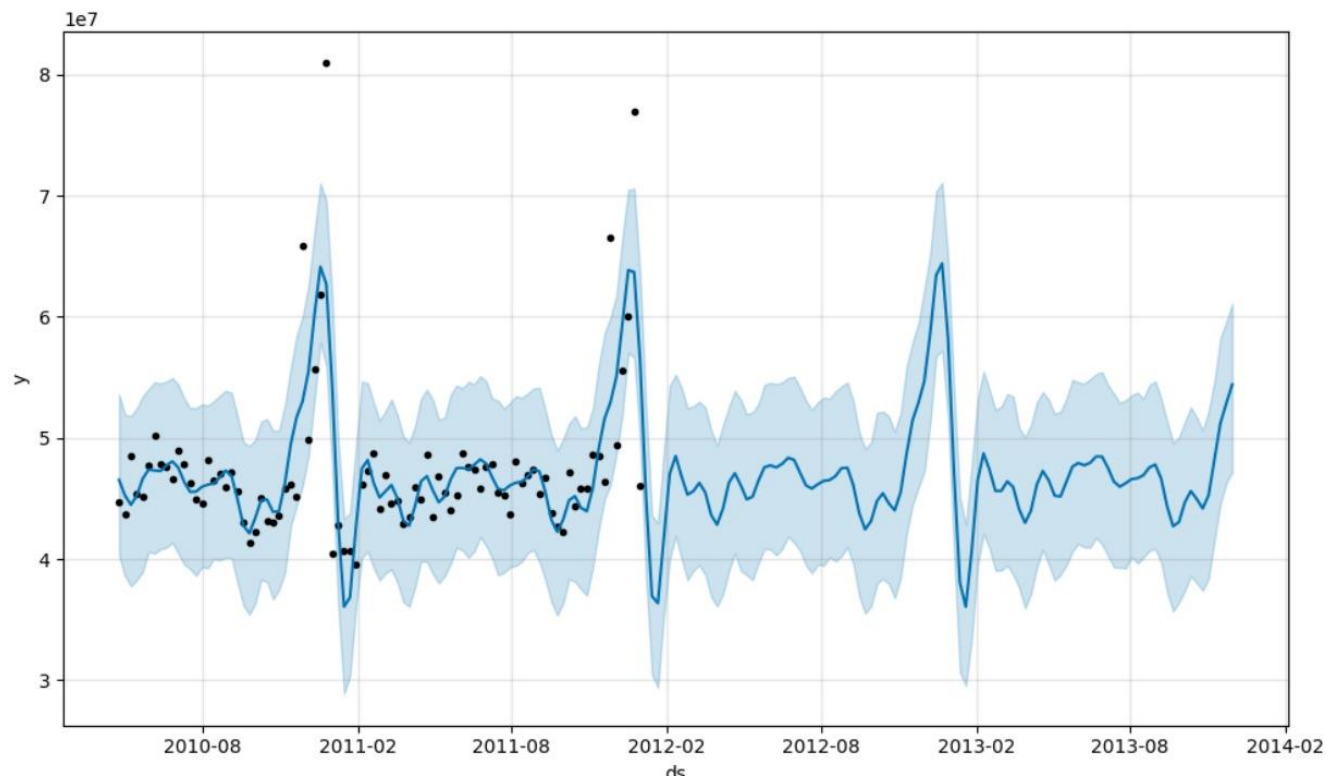
- 1) ARIMA: An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.
- 2) SARIMA: The SARIMA model is a statistical model that captures seasonal trends in the past values of the data to predict future discounts.
- 3) FB PROPHET: Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects.

FLOWCHART



SNAPSHOTS





OBSERVATIONS & SUGGESTIONS

- Observation 1: Out of all the store types, Type A stores gathered the most sales for Walmart.
- Suggestion 1: Type A stores deserve additional resources to support their growth and expansion.
 - Increase staffing levels, invest in store renovations, and provide targeted marketing support to maximize the potential of these high-performing stores.
 - Consider expanding the number of Type A stores in areas with strong market potential.
- Observation 2: There is a positive correlation between store size and sales.
- Suggestion 2: Build larger stores in areas with a high population density.
 - Walmart should consider building larger stores in areas with a high population density. This will allow them to offer a wider variety of products and services, which will attract more customers.
- Observation 3: There is a seasonal trend in both fuel prices and sales.
- Suggestion 3: Track fuel prices and adjust prices accordingly.
 - Walmart should track fuel prices and adjust their prices accordingly. This will help them to stay competitive and maximize profits.
- Observation 4: There is an evident hike in sales in weeks 47 and 51 that correspond to *Thanksgiving* and *Christmas* respectively.
- Suggestion 4: Walmart can consider:
 - Early holiday promotions.
 - Extended store hours.
 - In-store events: product demonstrations, gift-wrapping workshops, and holiday celebrations.
 - Enhanced online shopping experience.

- bservation 5: The highest sales occur for most store types between the range of 40 to 80 degrees Fahrenheit.
- Suggestion 5: Walmart can consider implementing strategies such as:
 - a) Seasonal merchandise.
 - b) Outdoor promotions.
 - c) Temperature-based marketing: Utilize weather data to tailor marketing messages and promotions based on the current and forecasted temperature.
 - d) In-store comfort: Maintain a comfortable temperature inside stores to enhance the shopping experience and encourage customers to linger and browse.

CONCLUSION

Factors significantly affecting the sales are:

- a) Holidays.
- b) Store size.
- c) Temperature.
- d) Fuel price.
- e) Department.

Factors having less impact on sales are:

- a) CPI Index.
- b) Unemployment