SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

# AUTOMATED SALES ANALYSIS AND FORECASTING

**INT 625 – INTERNSHIP 3 PROJECT REPORT**

***Submitted by***

**MARIA ANTHONY RAJA – E7321009**

***In partial fulfilment for the award of the degree of***

**MASTER OF SCIENCE**

(Data Analytics)

**Sri Ramachandra Faculty of Engineering and Technology**

**Sri Ramachandra Institute of Higher Education and Research, Porur,**

**Chennai -600116**

**MAY, 2023**

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

# BONAFIDE CERTIFICATE

Certified that this project report **"AUTOMATED SALES ANALYSIS AND FORECASTING"** is the bonafide record of work done by **"MARIA ANTHONY RAJA – E7321009"** who carried out the internship work under my supervision.

**Signature of the Supervisor**

**Dr. P.A.Abhinand**

**Lecturer,**

Department of Bioinformatics

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

**Signature of Programme Coordinator**

**Dr. Pitchumani Angayarkani**

**Associate Professor,**

Department of Computer Science and Engineering

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

**Evaluation Date: 4-May-23**

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

# ACKNOWLEDGEMENT

I express my sincere gratitude to our Chancellor, Vice-Chancellor, our Dean **Dr. T. Ragunathan** and Programme Coordinator **Dr. Pitchumani Angayarkani** for their support and for providing the required facilities for carrying out this study.

I wish to thank my faculty supervisor, **Dr. P.A.Abhinand**, Department of Bioinformatics, Sri Ramachandra faculty of Engineering and Technology for extending help and encouragement throughout the project. Without his continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to all the members of Sri Ramachandra Faculty of Engineering and Technology, my beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

# TABLE OF CONTENTS

# ABSTRACT

The ability to analyse the company's sales is essential for boosting revenue. The exploratory data analysis approach will be used to get insights into a company's performance and enhance it. The company's future success can be forecasted using historical data. Important business decisions can be made by studying the company's prior performance. This project aims to tackle business-related challenges by evaluating the company's historical data with advanced statistical tools in R and data visualization tools in Power BI. Furthermore, the dashboard generated by Power BI will be automatically updated as new data is received.

# CHAPTER 1
# INTRODUCTION

## 1.1. INTRODUCTION TO PROJECT

The multinational retail chain Walmart operates several supermarkets, discount department stores, and hypermarkets. Walmart is the one of largest retailers in the world, with more than 11,000 outlets across 27 countries. The enterprise is based in the United States. The company employs around 2.3 million people worldwide, and it serves millions of customers every day. The project's goal is to examine historical sales information for 45 Walmart locations spread across several areas, from February 2010 to November 2012. Predict sales for the future and try to derive insight from the data. The methodologies throughout this project comprise business intelligence tools, data analytics, and statistics.

## 1.2. TECHNIQUES INVOLVED

The data preparation stage is where it goes after the data collection stage. Data preparation—often referred to as "pre-processing"—is the stage where raw data is cleaned up and organised for the next stage of data processing.

### 1.2.1. Tool: Power BI

PowerBI used to load and transform data using various function in it. Especially Data Analysis Expression (DAX) function has been used to create a calculated column. The data analysis results are consolidated as dashboard.

### 1.2.2. Tool: R-Script

Power integrates with R-Script which in turn let the access to create more advanced statistical visuals and analysis.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1. OUTLINE OF EXISTING WORK

Table2.1.

| | |
|---|---|
| Project in Analytics 2022 | Walmart’s Sales Analysis through Data Visualization, Aman Preet Gulati, et al. |
| Case study in Analytics 2023 | Superstore Sales & Profit Report Using Power BI, Chaitanya Shah, et al. |
| Kaggle competition 2014 | Walmart Sales Forecast projects |
| NovyPro platform 2023 | Walmart's sales data analysis, K.D.Manoj Kumar, et al. |

## 2.2. OVERVIEW

There are many projects in the internet especially in Kaggle website for the Walmart Sales Analysis. However, the proposed project deals with more automation with the help of Power BI.

## 2.3. COMPARATIVE REVIEW

The available project featured in the Kaggle mostly built with the help of Python. The proposed project uses R programming for statistical analysis and forecasting.

Same way, all available projects are using Python for data visualization whereas proposed project uses Power BI for advanced and customized data visualization into a dashboard.

## 2.4. GAPS IN AVAILABLE RESEARCH

There are few some below gaps in the existing research.

### 2.4.1. LIMITATION IN DATA VISUALIZATION

Data analysis results can be more easily understood by users when they are visualized. Most of the projects used Python, where it is impossible to showcase the charts; instead, they can only be generated as images.

### 2.4.2. MISSING STATISTICAL ANALYSIS

Businesses can use statistical analysis to help them make better decisions that can drive growth, profitability, and provide them a competitive edge in their market. Available research fails to implement more on statistical analysis.

### 2.4.3. START FROM ZERO

The entire process of data preparation, forecasting, and reporting must be restarted from scratch whenever fresh data is received.

### 2.4.4. NOT AN END USER PRODUCT

With the Python data visualization, the charts only can be prepared and reported. These charts are static and end users should edit the code if they want to interact with the visuals.

# CHAPTER 3

# PROPOSED METHODOLOGY

## 3.1. DASHBOARD

Using Power BI, all the charts can be consolidated to form a dashboard. This gives a great presentation experience for end users.

## 3.2. INTEGRATING R SCRIPT

Power BI can support in-built R script access which can be used to perform advanced statistical analysis. Businesses can use the statistical methods below as useful tools to understand their historical performance and make future decisions.

### 3.2.1. Descriptive Statistics

The historical data on the performance of the company is summarized using fundamental statistical measures including mean, median, mode, standard deviation, and range. Descriptive statistics, such as the average revenue or profit margins for a specific time, can be utilized to gain a general understanding of the company's previous performance.

### 3.2.2. Time Series Analysis

Based on historical trends, time series analysis can be used to forecast and predict future performance. Moving averages, exponential smoothing, and ARIMA models are typical time series analysis methods.

### 3.2.3. Regression Models

The proposed project uses regression models to examine the relationship between independent parameters and how they affect the performance of the business. Regression analysis can be used to identify important factors that affect a company's success, including changes in consumer preferences, the state of the economy, and market rivalry.

# CHAPTER 4

# IMPLEMENTATION

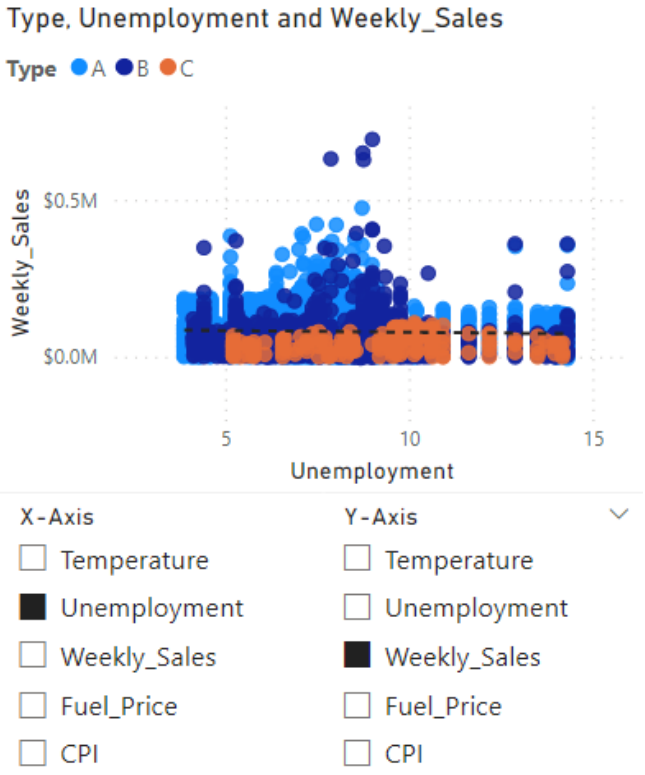
## 4.1. DYNAMIC SCATTER PLOT

Data points are shown visually as a series of dots or markers on a two-dimensional coordinate system in a dynamic scatter plot. The plot is referred described as "dynamic" since it changes in real-time or as a result of user interactions like clicking, hovering, or choosing various options.

Each data point is represented by a dot that is placed in accordance with the values of the respective variables on the scatter plot, which commonly displays two variables on the X and Y axes. Additional variables or categories can also be represented by changing the dots' size, colour, or shape.

Users can explore and analyse the relationships between variables in a dynamic scatter plot in a variety of ways because the data can be filtered or sorted based on numerous criteria. Data analysis, exploratory data visualisation, and interactive dashboards frequently employ this style of visualisation.

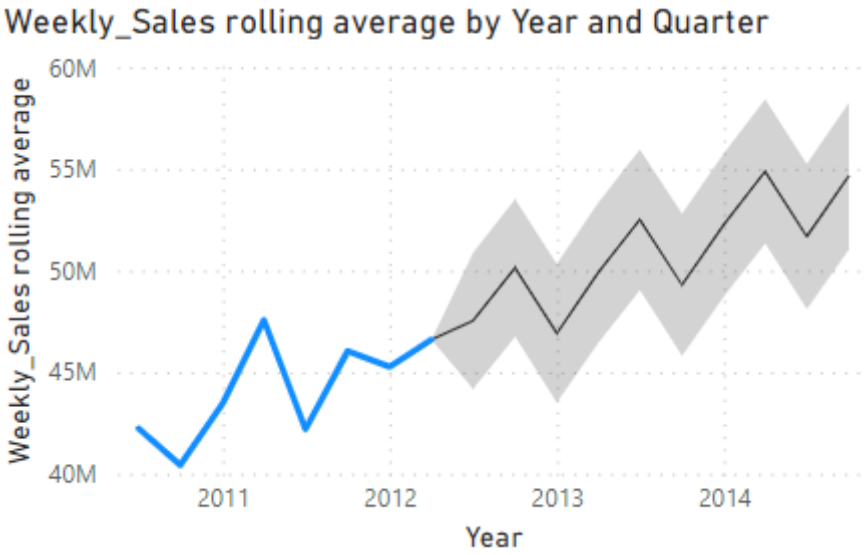
Fig4.1. Dynamic scatter plot



## 4.2. FORECASTING

A forecast chart in Power BI may be used to produce predictions and projections for next periods as well as to spot trends, patterns, and anomalies in data. It is a helpful tool for planning, budgeting, and making decisions in business.

Fig4.2. Forecast plot




## 4.3. CORRELATION MATRIX USING R-SCRIPT

The correlation coefficients between several variables in a dataset are shown in a table called a correlation matrix. Each variable is compared to every other variable, and a correlation coefficient is used to quantify the strength and direction of their linear relationship.

The correlation coefficient, abbreviated "r," has a range of values from -1 to 1, with -1 denoting a perfect negative correlation, 0 denoting no correlation, and 1 denoting a perfect positive correlation. A positive correlation indicates a tendency for the variables to rise or fall together, whereas a negative correlation indicates a tendency for them to move in the opposite ways.

A correlation matrix normally consists of a square grid with the variables being analysed represented by the rows and columns, and the correlation coefficients between each pair of variables contained in the cells. Due to the complete correlation between each variable and itself, the diagonal of the matrix is always 1.

In order to discover links between variables and investigate patterns and trends in the data, correlation matrices are frequently employed in data analysis and research. They can also be used to identify multicollinearity, which is when two or more variables have a high degree of correlation with one another and it is challenging to separate out their individual impacts on the end variable.

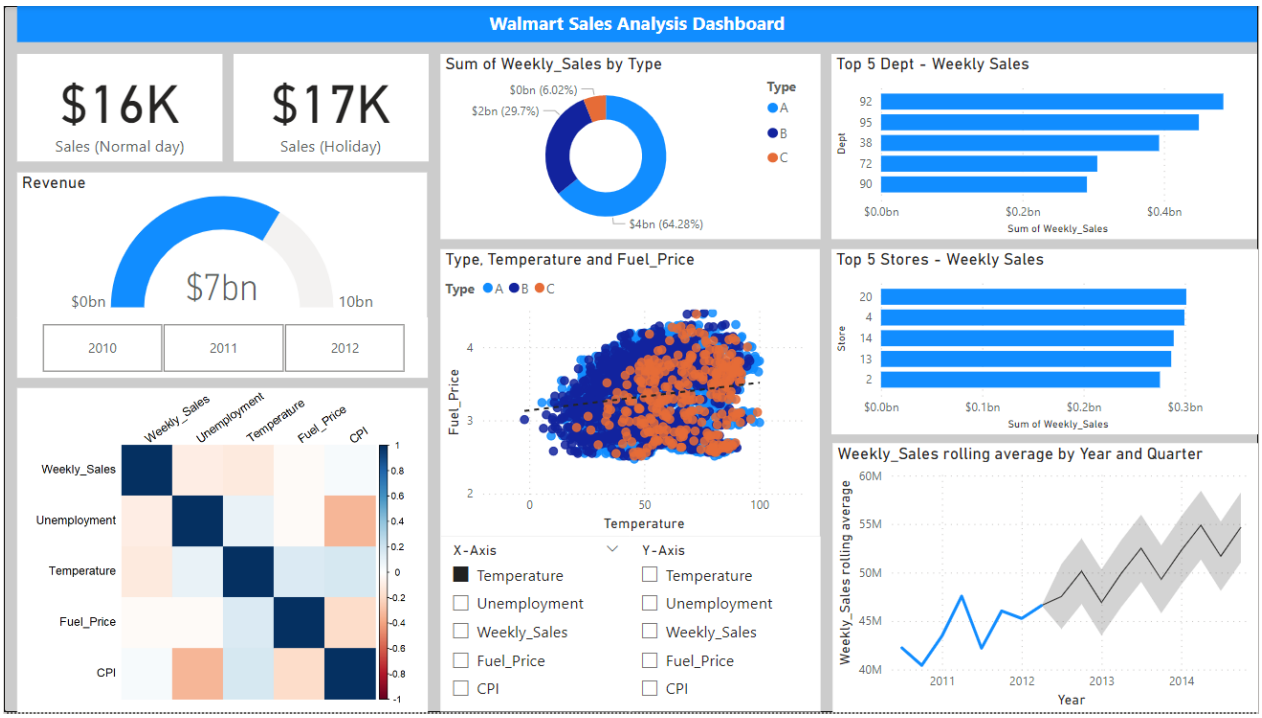
Fig4.3. Correlation matrix

# CHAPTER 5

# RESULTS AND DISCUSSIONS

## 5.1. FINAL DASHBOARD

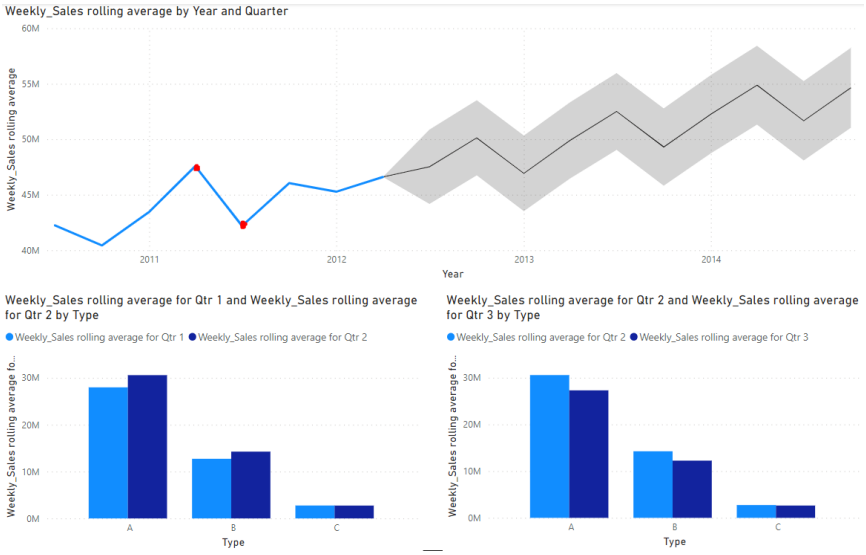
Fig5.1. Dashboard




## 5.2. RESULTS

- The store type A produced almost 64% revenue all-time.
- Sales on holiday is higher than usual days has been proved evidently.
- The future sales have been forecasted.
- CPI and Unemployment variables have 0.25 (approx.) negative correlation.
- Top performing Department – 92
- Top performing Store – 20

Fig5.2. Result chart




- As the store type A and B received more sales in 2011 Quarter 1. Hence, there is a 9.48% increase in Overall Sales between 2011 Quarter 1 and 2011 Quarter 2.
- As the store type A and B have lower sales in 2011 Quarter 3. Hence, there is a 11.31% decrease in Overall Sales between 2011 Quarter 2 and 2011 Quarter 3.

# APPENDICES

## APPENDIX-1: Rolling Average of Sales – DAX

```
Weekly_Sales rolling average =
IF(
    ISFILTERED('train'[Date]),
    ERROR("Time intelligence quick measures can only be grouped or filtered by the Power BI-provided date hierarchy or primary date column."),
    VAR __LAST_DATE = LASTDATE('train'[Date].[Date])
    RETURN
        AVERAGEX(
            DATESBETWEEN(
                'train'[Date].[Date],
                DATEADD(__LAST_DATE, -1, DAY),
                DATEADD(__LAST_DATE, 1, DAY)
            ),
            CALCULATE(SUM('train'[Weekly_Sales]))
        )
)
```

## APPENDIX-2: Correlation Matrix – R-Script

# The following code to create a dataframe and remove duplicated rows is always executed and acts as a preamble for your script:

# dataset <- data.frame(Weekly_Sales, features.Unemployment, features.Temperature, features.Fuel_Price, features.CPI)

# dataset <- unique(dataset)

# Paste or type your script code here:

require("corrplot")

library(corrplot)

M<- cor(dataset)

corrplot(M, method = "color",  tl.cex=1, tl.srt = 35, tl.col = "black")

#,col = colorRampPalette(c("blue", "green"))(100)

# REFERENCES

**Web References:**


1. Walmart’s Sales Analysis through Data Visualization: https://www.analyticsvidhya.com/blog/2022/01/walmarts-sales-analysis-through-data visualization/
2. Walmart Sales Dataset: https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/data?select=features.csv.zip
3. Walmart Sales Forecast projects in Kaggle: https://www.kaggle.com/competitions/walmart-recruiting-store-sales-forecasting/code
4. Superstore Sales & Profit Report Using Power BI: https://www.analyticsvidhya.com/blog/2023/01/a-case-study-superstore-sales-profit-report-using-power-bi/
5. NovyPro - Walmart's sales data analysis:
6. https://www.novypro.com/project/performed-kpiskey-performance-indicators-on-various-given-data-from-walmart-sales-report-for-the-time-period-between-012011-012015

# WORKLOG

| Date | Task Done |
|---|---|
| 25-Mar-23 | Project idea discussion |
| 01-Apr-23 | Abstract and Literature survey submission |
| 02-Apr-23 | Guide's Suggestion in Literature survey |
| 04-Apr-23 | Amendment in Literature survey |
| 17-Apr-23 | First review presentation and submission |
| 17-Apr-23 | Amendment in review presentation |
| 02-May-23 | Final review submitted |
| 02-May-23 | Final report submission |