

# STORE SALES PREDICTION

**SANJUSHA SURESH**

**(HALL TICKET NO.: 120919467001)**

**ROHAN AGARWAL**

**(HALL TICKET NO.: 120919467022)**

**DEVI VARAPRASAD GUNDRALA**

**(HALL TICKET NO.: 120919467029)**

**UNDER THE GUIDANCE OF**

**MS. PANAM SRAVANI**

**SUPERVISOR, LECTURER IN STATISTICS, DEPT. OF SCIENCES, ST. MARY'S COLLEGE.**



# CONTENT

ABSTRACT

INTRODUCTION

LITERATURE REVIEW

DATA COLLECTION

DATA ANALYSIS

METHODOLOGY

RESULTS

CONCLUSION

REFERENCES

# ABSTRACT



- ▶ Sales forecasting is a fundamental assignment in retailing which helps businesses to allocate capital effectively, to forecast realistic sales revenues as well as to prepare a better plan for potentially increasing the business.
- ▶ In this work we proposed a data analytics-based statistical method that is time series analysis, moving average method, regression analysis and correlation for predicting the sales based on the last three years data and it will generate reports.
- ▶ This approach is carried out on quantitative data from five Walmart stores which is taken from Kaggle.
- ▶ This prediction performs various tasks like, finding best months for sales, in which month the sales will increase and decrease and how the holidays affect the sales of the stores.



# INTRODUCTION

## ► Objective of the study

- Sales forecasting facilitates business in right decision making by providing relevant market and production information leading to purchase of materials and accomplishment of pre-established targets.
- Therefore in this study our aim is to understand if the sales will increase or decrease in the future using the past data.
- And also in which month or year the sales will be high so that the result can be used in future for increasing the sales and production of the store.



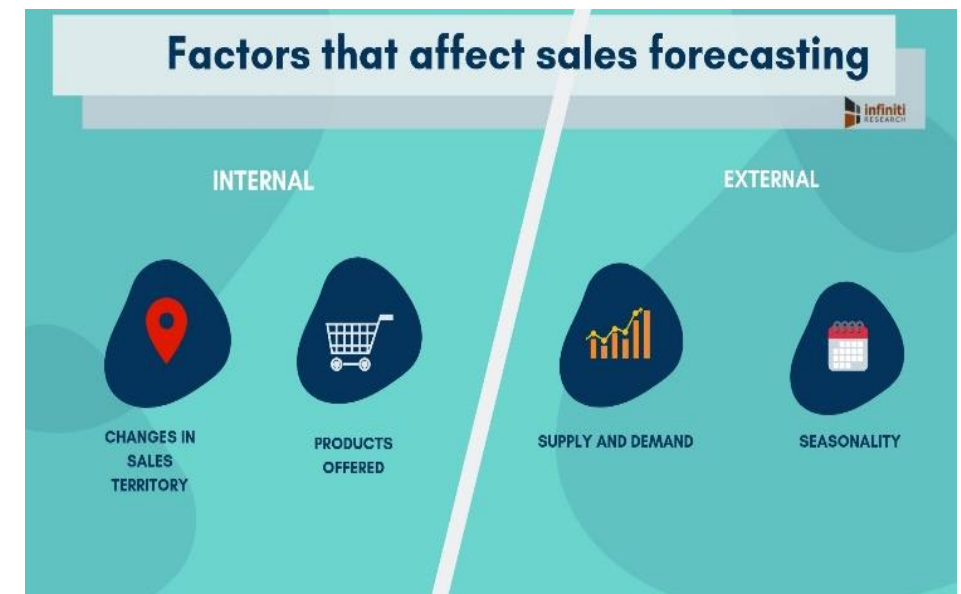


### ► Scope of the study

- This study will cover the data of the sales of past three years which is analyzed using the time series model, moving average method, regression analysis and correlation. Then we examine the factors that affect the future sales.
- Subsequently with the acquired result, it can be estimated as to how to increase the sales, in which months the products should be in extra stock and what changes should be made so that the store gets profitable without much loss and eventually advance the business.
- This study can also be referred for time ahead purposes and changes can be made according to the future necessity.

## ► Limitations of the study

- One of the limitation is that the standard deviation of error in regression analysis made in actual forecasting, is likely to be larger than the computed error of estimates, due to a set of causes persisting in future years.
- And the data which is analyzed using the time series is not suitable for short-term forecasting and this method is difficult to use in cases where irrelevant forces disrupt the regularity of sales.



# LITERATURE REVIEW

- ▶ "Walmart's Sales Data Analysis - A Big Data Analytics Perspective"
- ▶ "Forecast of Sales of Walmart Store Using Big Data Applications"

# DATA COLLECTION

- ▶ We used wide market sales data as a dataset in our work where the dataset consists of 9 attributes, where holidays, weekly sales, monthly sales, months, dates are the main factors on which our dataset focuses on.
- ▶ Attributes such as temperature, fuel price, CPI, unemployment have been included in store level and valuable data information is drawn-out from the dataset. Outlet year of establishment ranges from 2019 to 2021.
- ▶ There are 6436 different values of the sales present in the dataset out of which the data is divided into 715 values of five Walmart stores for easy and better calculations and the data is collected from Kaggle.





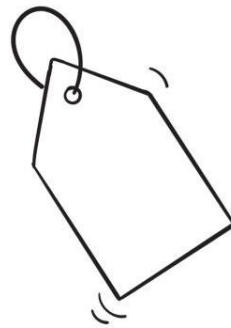
# DATA ANALYSIS

- ▶ In this work we proposed a data analytics-based statistical model that is time series analysis, moving average method, regression analysis and correlation for predicting the sales based on the last three years data and it will generate reports.
- ▶ Time series analysis is a series of techniques that make forecasts based on past patterns of data. These data are collected, observed, and recorded at regular intervals of time.
- ▶ The moving average method suggests drawing an average of the sales of number of years to predict the sales of a coming period.
- ▶ Regression analysis is used in determining and measuring the association between company sales and other variables and it involves fitting an equation to explain sales fluctuations.
- ▶ Correlation looks at the strength of a relationship between two variables. It is used to know if there is a predictable relationship between sales and factors like temperature, holidays, advertising etc.



# METHODOLOGY

- ▶ Explanation using Excel sheet
- ▶ Analysis - Excel.xlsx

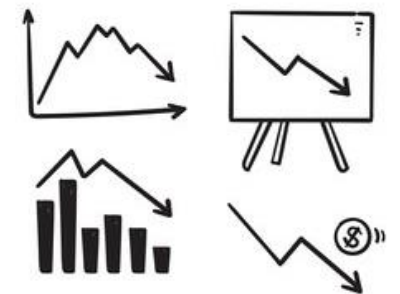


# RESULT



- ▶ Using the **time series analysis**, we plotted the monthly total sales plot and monthly average sales plot of five stores of Walmart. From the monthly total sales plot January 2019 had the highest sales. And the lowest sales was in the month of January 2020.
- ▶ From the monthly average sales plot January 2019 to February 2019 had the highest sales. And the lowest sales was in the month of January 2020.
- ▶ Similarly, from the weekly total sales plot December 2019 and between December 2020 and January 2021 highest sales can be observed. And the lowest sales were in the months of February 2020 and February 2021.
- ▶ From the weekly average plot, the highest sales again in December 2019 and between December 2020 and January 2021. And the lowest sales are in the months of February 2020 and February 2021.

- ▶ Using the **moving average method**, the highest and lowest sales of store 1 is determined. The highest sales are in the month of December 2019 whereas the lowest sales is in the month of January 2020.
- ▶ For store 2, The highest sales is in the month of December 2019 whereas the lowest sales is in the month of January 2020.
- ▶ For store 3, the highest sales are again in the month of December and the lowest sales are constant from May to June 2019 and August to September 2019. And also is lowest in January 2020 and January 2021.
- ▶ In store 4, there is a peak increase in the sales from December 2019 to February 2020. And the lowest sales were observed in January 2021.
- ▶ In store 5, there is a continuous increase and decrease pattern of sales. The highest sales is in December 2020 and the lowest is in January 2020.





- ▶ Using **regression analysis** we found out in which year the sales were less and the year the sales was more. In Store 1, there is a uniform increase in the sales from 2019 till 2021. Hence the sales in 2022 of store 1 will also increase uniformly.
- ▶ In store 2, there is a sudden decrease in the sales in the year 2019 till 2020. From 2020 to 2021 it is observed that the sales remained constant. Hence the sales might remain constant in the year 2022.
- ▶ In store 3, there is an increase in the sales from 2019 till 2021. Hence there are chances that the sales in 2022 of store 3 will also increase.
- ▶ In store 4, there was an increase in the sales from 2019 till 2020 and a decrease from 2020 to 2021. Hence there are chances that the sales might increase or decrease for the store 4 in 2022.
- ▶ In store 5, there was an increase in the sales from 2019 till 2020 and a decrease from 2020 to 2021. Hence there are chances that the sales might increase or decrease for the store 5 in 2022.
- ▶ In the next part **correlation** between sales and holidays were obtained. Since the correlation is positive it can be said that holidays are directly proportional to the sales. That is, holidays increase sales.

# CONCLUSION

- ▶ The research carried out in this project has concluded that product sales prediction systems are needed to manage massive volumes of data by businesses and that company's decision-making is focused on data processing technologies' speed and precision.
- ▶ For businesses to be competent, it is necessary that they equip themselves with specialized techniques in order to take account of diverse forms of consumer behavior by making predictions for their product sales and demand to increase profit.
- ▶ Accuracy, which plays a key role in prediction-based systems, can be significantly increased as the number of parameters used are increased.



# REFERENCES

1. A.S. Harsoor And A. Patil, “Forecast Of Sales Of Walmart Store Using Big Data Application”, International Journal Of Research In Engineering And Technology, Vol. 4, P. 6, June 2015.
2. Manpreet Singh, Bhawick Ghutla, Reuben Lilo Jnr, Aeesan Fs Mohammed, Mahmood A Rashid, “Walmart Sales Data Analysis- A Big Data Analytics Perspective”, 4th Asia-pacific World Congress On Computer Science And Engineering, December 2017.
3. Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. Time Series Analysis: Forecasting And Control; John Wiley & Sons: Hoboken, NJ, USA, 2015.
4. Doganis, P.; Alexandridis, A.; Patrinos, P.; Sarimveis, H. Time Series Sales Forecasting For Short Shelf-life Food Products Based On Artificial Neural Networks And Evolutionary Computing. J. Food Eng. 2006, 75, 196–204
5. Hyndman, R.J.; Khandakar, Y. Automatic Time Series For Forecasting: The Forecast Package For R; Number 6/07; Monash University, Department Of Econometrics And Business Statistics: Melbourne, Australia, 2007.



THANK YOU