**7PAM2000 APPLIED DATA SCIENCE 1**

**ASSIGNMENT 1: VISUALIZATION**

**STUDENT NAME: SUBASH REDDY PADIGELA**

**STUDENT ID:22088785**

**GIT HUB LINK: https://github.com/subashreddy222/ADS--1-Assignment.git**

# Dataset

This dataset contains information regarding the sales in an organization which contains information regarding the sales, order info, customer, shipping, date etc. (Kaggle, 2023) It is used for performing different visualizations, customer segmentation., clustering etc.

# Line Plot

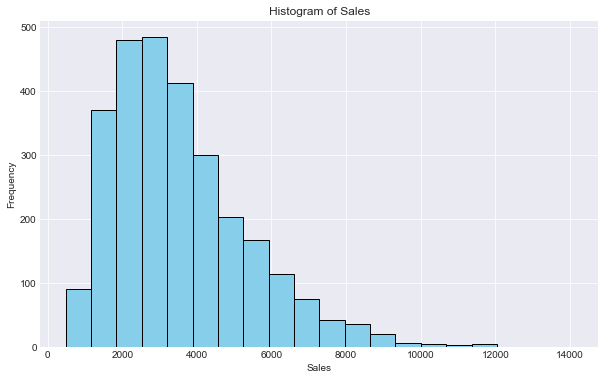


A line graph which is also known as a line chart is most popularly known for plotting time series data or showing data which are in a series form. The data points are known as markers which are connected using a straight line. If we are having a time series data or some continuous data then we can use this line chart in showing the trends in the data and changes in the data can be shown. It is useful in showing the trend analysis, comparisons, change over time and also in displaying the forecasting of data. A flexible and popular technique for data visualisation, line graphs give changing values over time or across categories which is an understandable and straightforward representation of the data points.

Since a line chart is a useful tool for illustrating trends over time, it is the preferred method in this data. In this instance, the graph illustrates the evolution of monthly sales over time. The months are represented by the x-axis, and each line denotes a distinct year. Each data point has markers added for easier viewing and understanding. The inclusion of a legend that links each sentence to the appropriate year facilitates audience comprehension of the story. The addition of grid lines makes it easier to understand the chart's data.

This line chart shows the sales on y-axis and the month on x-axis which shows three different years which is blue line shows the year 2003, orange line shows the year 2004 and green line shows the year 2005. The trend in the sales for these three years are represented using the line chart. From the line chart it could be observed that the year 2003 and 2004 had good sale and in these two years the sale was maximum in the month of November. But while observing the blue line which shows 2005 had made sale only till the month of July.

# Histogram

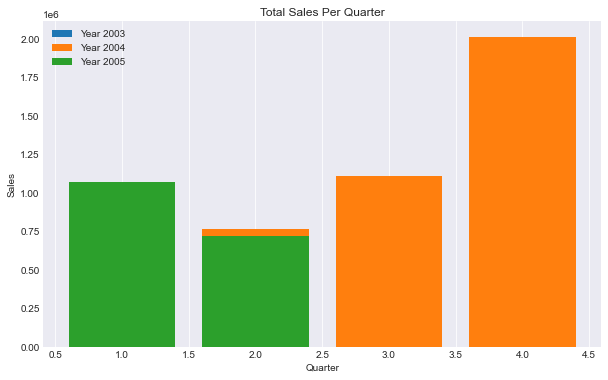


Histogram is a graph which is used in showing the distribution of data points. It basically shows the frequencies of different values or ranges of values in a dataset. Which helps in understanding the different patterns in the data, shows the central tendency of the data points and also helps in understanding the spread of the data. There are different bins in the this plot where the data points are basically divided into different intervals or bins and the different frequencies of data points are then represented using different heights of the bins. Every bar in the histogram shows how frequently data points fall into a certain category. Usually, there are no spaces between the bars; they are closely spaced.

When attempting to visualise the distribution of a single variable, histograms are especially helpful. Here the ‘Sales’ attribute is used for showing the data distribution. There are 20 bins created. This histogram plot would help us in knowing the range of sales and the maximum sales value. This makes it possible for us to see how sales values are distributed and how frequently sales fall into each category. In exploratory data analysis, the histogram is a useful tool for understanding the general distribution of a numerical variable.

The histogram shows the range of sales from 500 to 12000. The maximum sales is between 2000 to 4000.

# Bar Plot



A bar chart or a bar graph is used for representing the categorical data points using rectangular bars. Each bar’s length shows the value it is representing. It is mostly used in comparing the values of different categories or groups. Bars are main element in any bar chart. The quantity or value associated with each bar would be represented by the height of the bar chart.

In this bar graph the total sales per quarter is displayed where the quarter is displayed on the x-axis and the y-axis shows the sales. When comparing the amounts or values linked to several groups or categories, bar charts work well. The objective of the code supplied is to compare the overall sales for every quarter across several years. When displaying categorical data with unique labels or names for each category, bar plots work well. The code's specific use case is to display the total revenues for every quarter across several years. The sales performance for each quarter may be easily compared thanks to the bar plot.

The box plot represents the sales in different quarter. It is showing that quarter four for the year 2004 was the best and had made good sales while it had a minimum sales for the 2nd quarter. While 2005 made good sale in the first and second quarter.

# References

Kaggle, 2023. *Sample Sales Data.* [Online]   
Available at: https://www.kaggle.com/datasets/kyanyoga/sample-sales-data/data  
[Accessed 08 November 2023].