Course title: software project V

Teacher Information:

Fariha Jahan

Lecturer, Department of Computer Science & Engineering

Daffodil International University (DIU)

Dhaka, Bangladesh

Course code: CSE-317

Group Name

Little Giant's

Group Members

Nusrat Jahan Marjana

ID: 191-15-12999

Azmain Iqutedar Anik

ID: 191-15-12988

Applying Data Visualization on Supermarket Sales Prediction

Abstraction: This project is based on Data visualization. It makes data more natural and cleaner. To visualize the data, creator find a good dataset. This project done in Google colab. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. By input a xlsx or csv file in Colab can preprocessed by some library function call pandas and matplotlib, to make some charts such as Line Chart, Histogram, Bar Chart, Pie Chart, Horizontal Bar chart, scatter plot, Count plot, Horizontal Sns chart and Heapmap. *Supermarket Sales Prediction*"- dataset is used in this project. This Supermarket Sales is a Prediction on 2019.

Introduction: Data visualization is the graphical representation of processed data or information. It is the easiest way to understand Big data or information in short time. By using visual elements like charts, graphs and cards, data visualization platform provides an approachable way to see and understand results, trends, outliers, and patterns in data. This is a report which based on Data visualization. It gives us a clear idea of what the information means by giving it visual context through graphs. This makes the data more natural and easier to understand. Regular visualization helps to focus human mind on what they want. When anyone visualize on a regular basis, it helps them to reset goal and it focuses their mind.

In this project. Datasets are collected from Kaggle, name of dataset "*Supermarket Sales Prediction*". Kaggle allows user to find and publish dataset, it can still be a great learning tool for beginners.

Here, Google Colab platform is use to data visualization by coding in Python. Now, the Goal is to connect dataset into Google Colab and Visualize the data using Python libraries like Pandas, NumPy, Matplotlib. After insert selected dataset, user Pre-preprocessed the data. Then create different types of charts and graphs. The result of this dataset quit satisfying.

Methodology: In this project, user visualize the dataset "*Supermarket Sales Prediction*". This sample data module contains representative Supermarket Sales Prediction. (Source: IBM). Here are Csv file of this dataset:

- Supermarket Sales Prediction (Columns:1002, Rows-14)

To enter new dataset in Google Colab, we call a function "read_csv" in python code to import the data set
"*Supermarket Sales Prediction*"
From the function call users can define any csv file to import. Then select Dataset file path from Google Drive which can find from any data site. In this project user input the selected dataset. After input any dataset, user must have Pre-preprocessed the data. Data Prep reprocessing includes data cleaning, data integration, data transformation and data reduction. Before data visualization user try to prepare a Clean data set which has no Error or null Value. This all process is called Data Pre-preprocessing.

Results & Discussion: Results are the presentation of data and hence findings or investigations. In here user create 5 Dashboard/Report for visualization.

Report-1: This report is based on Supermarket Sales Prediction. There are 3 type of visuals: Column Chart, Clustered Bar Chart and line chart.

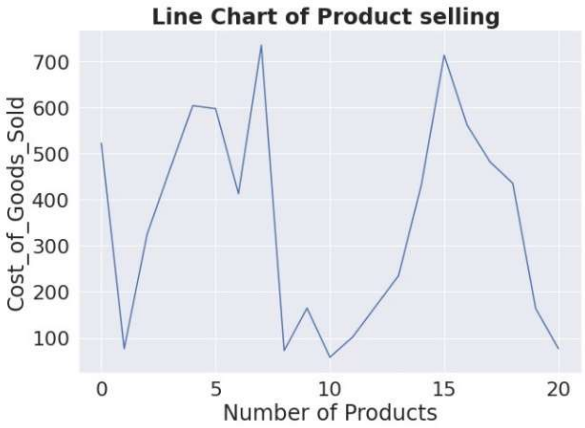

Fig 1: Line chart of Product Selling

In this line chart (Fig-1), 2 fields are used, Number of product (X axis) and Cost of Goods sold (Y axis).

Here goods sold 100 to 700 and number of products 0 to 20. Hare easily understand the selling rate of the products.

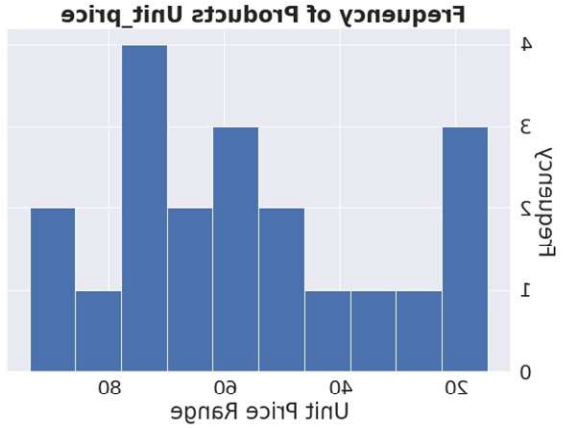

Fig 2: Histogram chart Frequency of products unit price

In this Histogram chart (Fig-2), 2 fields are used, Unit price range (X axis) and Frequency (Y axis).

Here Frequency 0 to 4 and Unit price 20 to 80. Hare easily understand the Frequency of products unit price.

Fig 3: Horizontal simple bar charts of Store inventory

In this Horizontal simple bar charts (Fig-3), 2 fields are used, profit (X axis) and Product (Y axis).

Here 4 product in product y axis and profit range is in x axis. Hare easily understand the store inventory product by profit range.

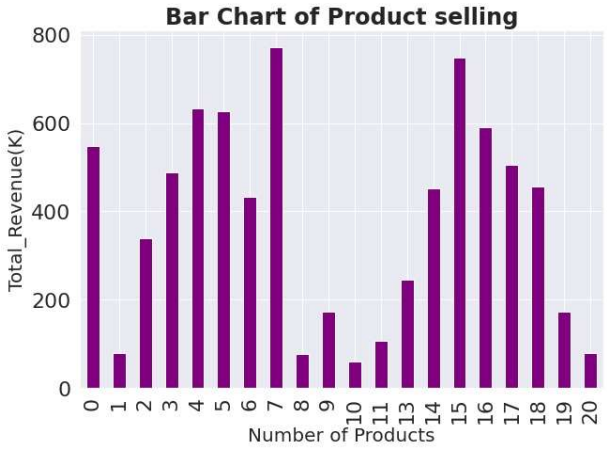

Fig 4: bar charts of product selling

In this bar charts (Fig-4), 2 fields are used, Number of products (X axis) and Total revenue in (Y axis).

Here revenue range is 0 to 800 and the x axis 20 products. Hare easily understand the product selling range.

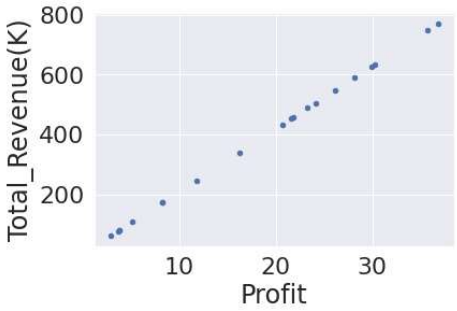

Fig 5: scatter plot of total revenue and profit

In this bar charts (Fig-5), 2 fields are used, Profit (X axis) and Total revenue in (Y axis).

Here revenue range is 200 to 800 and the x axis profit range is 10 to 30. Hare easily understand the revenue and profit growth.

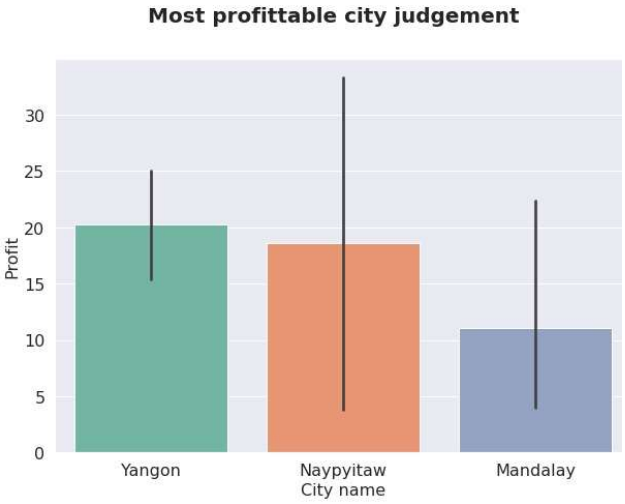

Fig 5: bar plot of Most profitable city

In this bar charts (Fig-5), 2 fields are used, City name (X axis) and Profit in (Y axis).

Here profit range is 0 to 30 and the x axis profit range has 3 cities. Hare easily understand the profit growth in city.

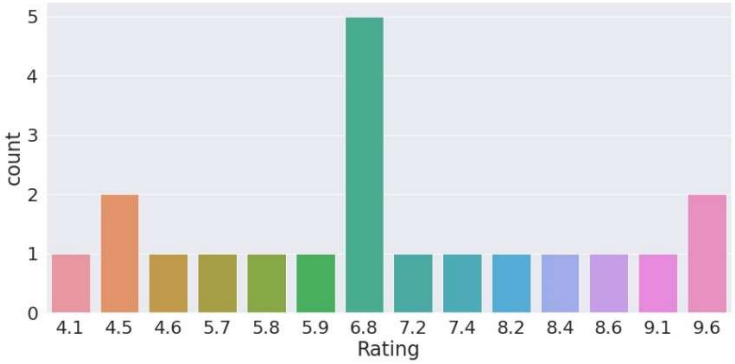

Fig 6: count plot of Rating

In this count charts (Fig-6), 1 field are used, Rating (X axis) and count in (Y axis).

Here rating range is 0 to 5 and the x axis rating range. Hare easily understand the rating.

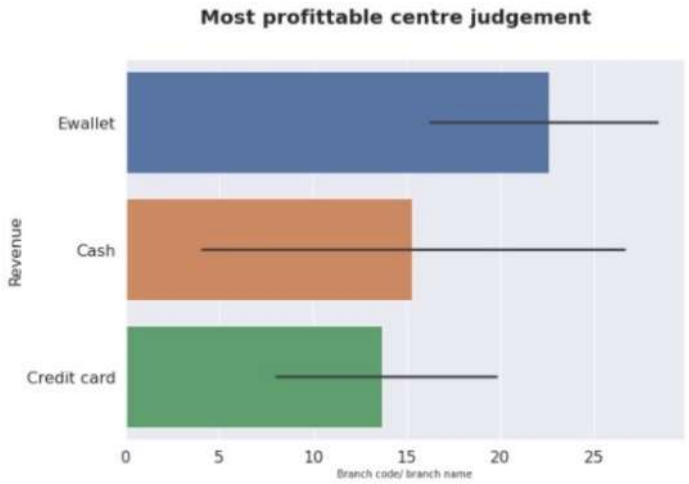

Fig 7: Horizontal bar plot of Most profitable center

In this bar charts (Fig-7), 3 fields are used, branch code and branch name in (X axis) and revenue in (Y axis).

Hare 3 Payment methods in y axis and the x axis branch code and branch name. Hare easily understand the profitable center.

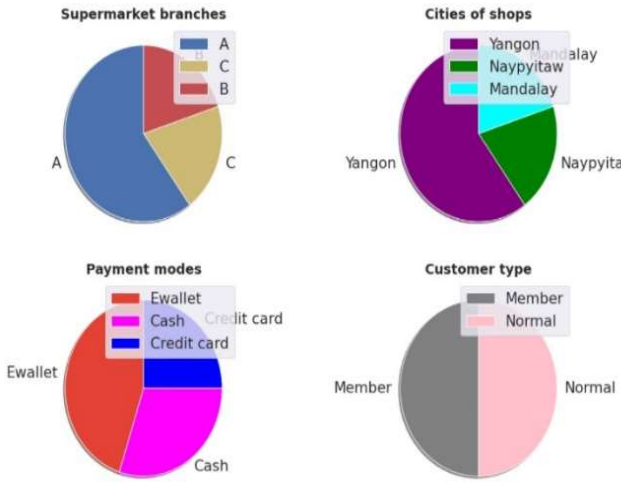

Fig 8: Pie plot of Most profitable center

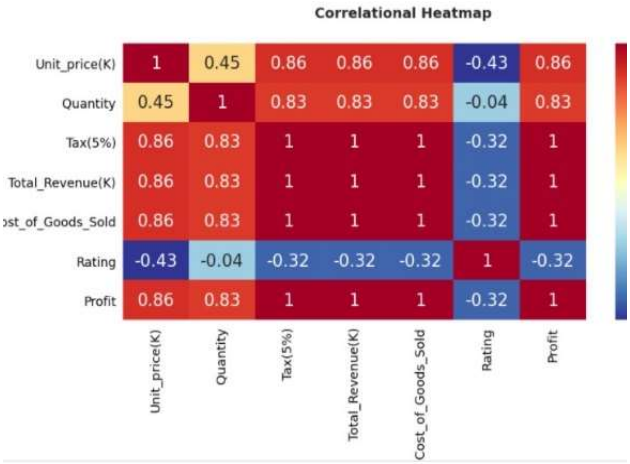

Fig 9: Correlational Heatmap plot of Most profitable center

Conclusion: Data visualization helps people organize, understand, and use data to its fullest potential. This might include pie charts, bar graphs or any other visual aid that helps its audience find the answers they need. In this project, they use a medium dataset of Supermarket Sales Prediction on (2019). They try their best to get maximum profit. If they can follow their stapes properly it can say they will make some good profits from the supermarket business.