Data Visualization- Bank Details Dataset

Prabesh Pokhrel   
*Department of Computer Science and Engineering  
Kathmandu University*Dhulikhel, Nepal  
pp02410118@student.ku.edu.np

Pradeep Banjade  
 *Department of Computer Science and Engineering*   
*Kathmandu Univerisity* Dhulikhel, Nepal  
pb02406918@student.ku.edu.np

*Abstract*— With the increase in usage of internet, the amount of data is also increasing day by day. If this big amount of data can be utilized properly many organization can be benifited. Data visualization helps in utilizing the big data to find out patterns and relations between different data which can further help the organization in their business and different work. This report gives a quick summary on data visualization on bank account details dataset and uses it to show it visually using streamlit framework. The report will show experimental analysis on different visualization chart and how the user can interact with it.

Keywords - pandas, plotly, streamlit, dataset, visualization

# Introduction

In our increasingly data-driven world, it’s more important than ever to have accessible ways to view and understand data. After all, the [demand for data skills in employees is steadily increasing](https://www.tableau.com/sites/default/files/2021-06/Tableau_Data_Literacy_Report.pdf) each year. Employees and business owners at every level need to have an understanding of data and of its impact. That’s where data visualization comes in handy. With the goal of making data more accessible and understandable, data visualization in the form of dashboards is the go-to tool for many businesses to analyze and share information. [1]

Data visualization is the graphical representation of information and data. By using v[isual elements like charts, graphs, and maps](https://www.tableau.com/data-insights/reference-library/visual-analytics), data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion. [1]

Our eyes are [drawn to colors and patterns](https://www.tableau.com/learn/whitepapers/tableau-visual-guidebook). We can quickly identify red from blue, and squares from circles. Our culture is visual, including everything from art and advertisements to TV and movies. Data visualization is another form of visual art that grabs our interest and keeps our eyes on the message. When we see a chart, we [quickly see trends and outliers](https://www.tableau.com/reports/business-intelligence-trends). If we can see something, we internalize it quickly. It’s storytelling with a purpose. If you’ve ever stared at a massive spreadsheet of data and couldn’t see a trend, you know how much more effective a visualization can be. [1]

Data visualization tools can be used in a variety of ways. The most common use today is as a business intelligence ([BI](https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI)) reporting tool. Users can set up visualization tools to generate automatic dashboards that track company performance across key performance indicators ([KPIs](https://www.techtarget.com/searchbusinessanalytics/definition/key-performance-indicators-KPIs)) and visually interpret the results. In the early days of visualization, the most common visualization technique was using a [Microsoft Excel](https://www.techtarget.com/searchenterprisedesktop/definition/Excel) spreadsheet to transform the information into a [table](https://www.techtarget.com/whatis/definition/table), bar graph or [pie chart](https://www.techtarget.com/whatis/definition/pie-graph-or-pie-chart). While these visualization methods are still commonly used, more intricate techniques are now available like line chart, area chart, scatter plot, tree maps, population pyramids, etc. [2]

This report summarizes about different types of data visualization techniques using bank detail dataset and also shows the importance of the graphical representation of the big data.

# technical background

Python along with its libraries like streamlit, plotly express, numpy and pandas are used to show the graphical representation of the bank details dataset.

* 1. *Python*

Python is a high-level, general-purpose and a very popular programming language. Python programming language (latest Python 3) is being used in web development, Machine Learning applications, along with all cutting edge technology in Software Industry. Python Programming Language is very well suited for Beginners, also for experienced programmers with other programming languages like C++ and Java.[3]

* 1. *Streamlit*

[Streamlit](https://www.streamlit.io/) is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science. We can build and deploy powerful data apps. [4]

* 1. *Plotly*

The Plotly Python library is an interactive open-source library. This can be a very helpful tool for data visualization and understanding the data simply and easily. plotly graph objects are a high-level interface to plotly which are easy to use. It can plot various types of graphs and charts like scatter plots, line charts, bar charts, box plots, histograms, pie charts, etc.[5]

* 1. *Pandas*

Pandas is an open-source library that is built on top of NumPy library. It is a Python package that offers various data structures and operations for manipulating numerical data and time series. It is mainly popular for importing and analyzing data much easier. Pandas is fast and it has high-performance & productivity for users. [6]

# related works

## Tablaeu

One of the most widely used data visualization tools tablaeu offers interactive visualization solutions to more than 57,000 companies. Providing integration for advanced databases, including Teradata, SAP, My SQL, [Amazon AWS](https://www.simplilearn.com/tutorials/aws-tutorial/what-is-aws" \t "_blank" \o "Amazon AWS), and [Hadoop](https://www.simplilearn.com/tutorials/hadoop-tutorial/what-is-hadoop" \t "_blank" \o "Hadoop), Tableau efficiently creates visualizations and graphics from large, constantly-evolving datasets used for [artificial intelligence](https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-artificial-intelligence" \t "_blank" \o "artificial intelligence), [machine learning](https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-machine-learning" \t "_blank" \o "machine learning), and [Big Data applications](https://www.simplilearn.com/tutorials/big-data-tutorial/big-data-applications" \t "_blank" \o "Big Data applications). [7]

## Dundas BI

Dundas BI offers highly-customizable data visualizations with interactive scorecards, maps, gauges, and charts, optimizing the creation of ad-hoc, multi-page reports. By providing users full control over visual elements, Dundas BI simplifies the complex operation of cleansing, inspecting, transforming, and modeling big datasets.[7]

## Google Charts

One of the major players in the data visualization market space, Google Charts, coded with SVG and [HTML5](https://www.simplilearn.com/tutorials/html-tutorial/html-vs-html5" \t "_blank" \o "HTML5), is famed for its capability to produce graphical and pictorial data visualizations. Google Charts offers zoom functionality, and it provides users with unmatched cross-platform compatibility with iOS, Android, and even the earlier versions of the Internet Explorer browser.[7]

# proposed system/ methodology

There are thousands of datasets in kaggle on many topics. We used bank details dataset in this project.

After thoroughly studying the dataset to determine which framework should be used to build a dashboard. There are many framework which can be used for this project but we used streamlit for this project. Streamlit is an open source python framework which helps to visualize the data easily.

##### V. experimental analysis

Here we can see different graphical visualization chart and their analysis and importance.

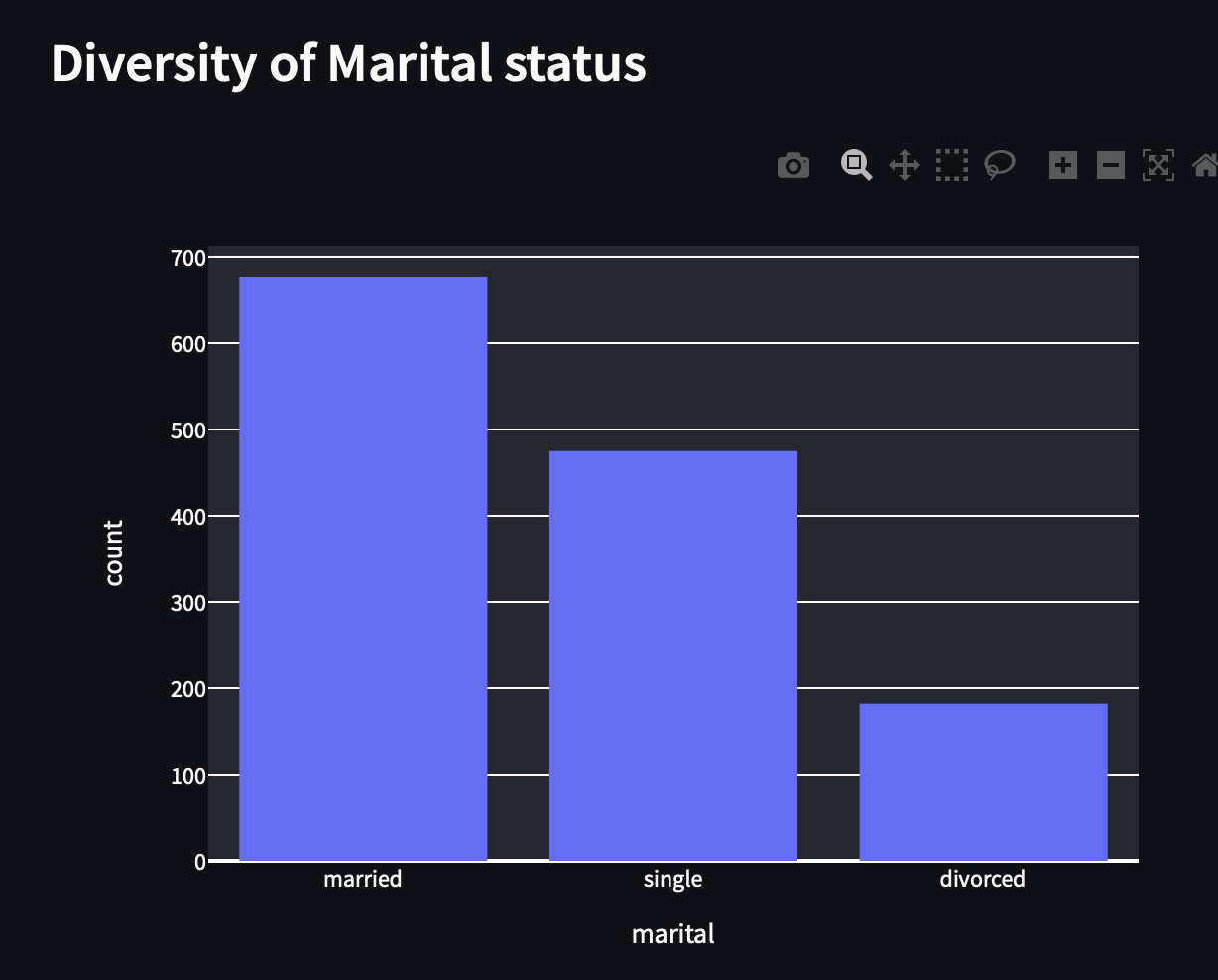


Figure 1: Bar Chart for marital status number

Figure 1 shows the Bar chart for the number of married, single and divorced people in the bank dataset. We can change the job as shown in fig 1 and can see their chart change in Bar chart.

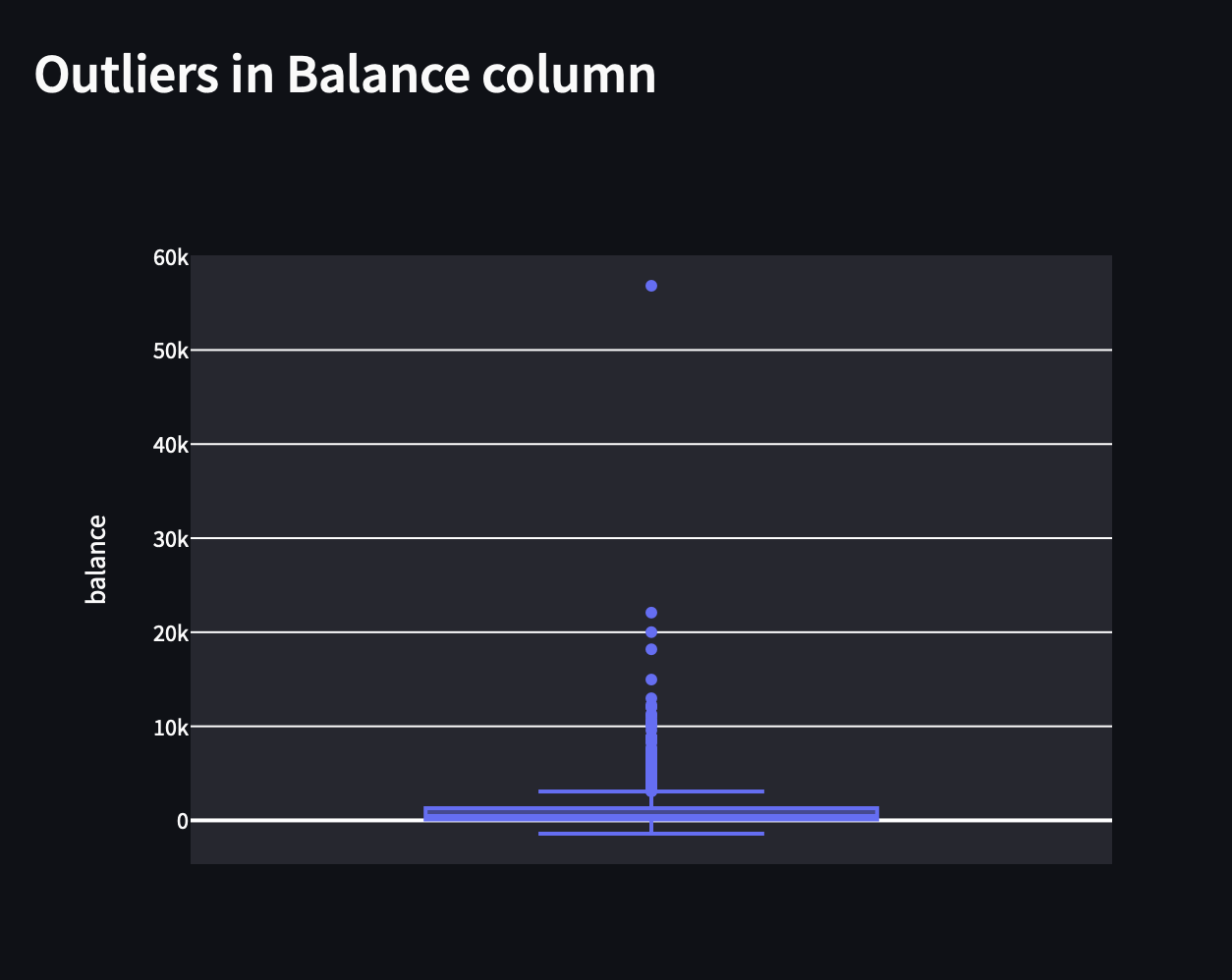


Figure 2: Outliers in Balance Column

Figure 2 shows the balance of user of bank. With this we can find out the outliers on the balance.

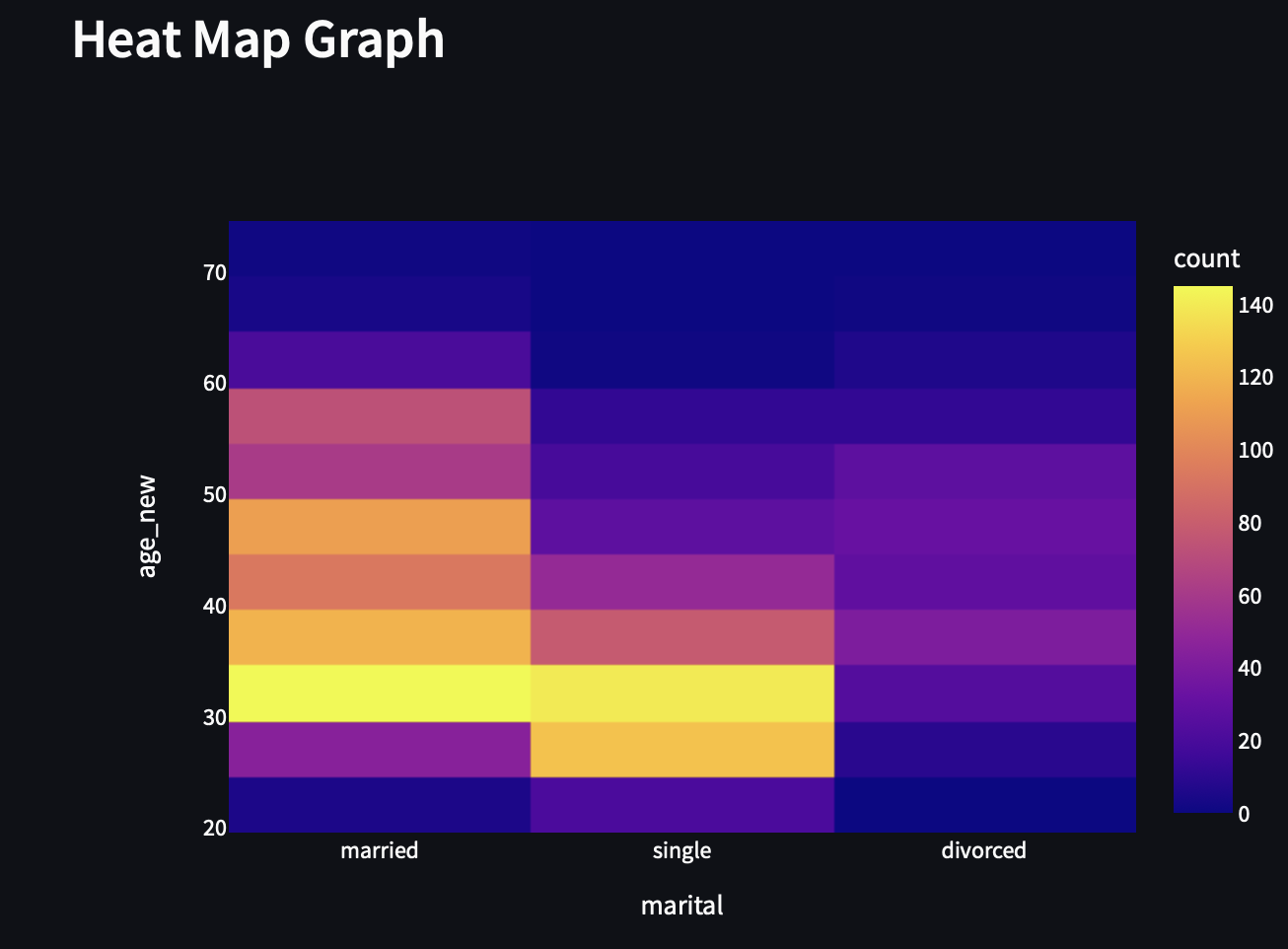


Figure 3: Heat Map Graph

Figure 3 shows the marital status like married, single and divorced on x-axis and their age on y-axis on the heat map graph. A heat map graph is is a [data visualization](https://en.wikipedia.org/wiki/Data_visualization" \o "Data visualization) technique that shows magnitude of a phenomenon as color in two dimensions. This heat map can be further changed by the people’s job as shown in figure 1.

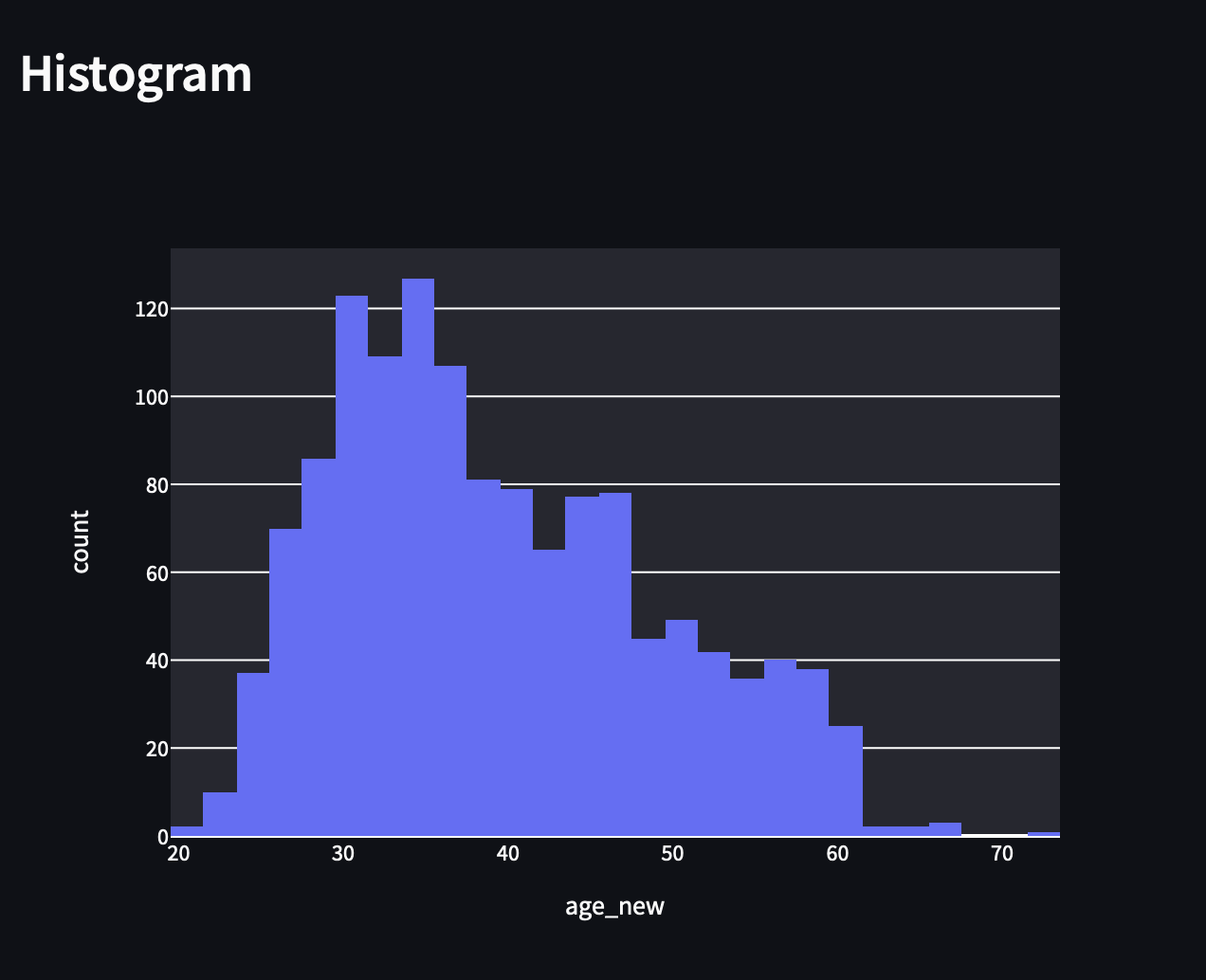


Figure 4: Histogram

Figure 4 shows the age count in a histogram. This shows the number of people with same and different age group from the bank dataset.

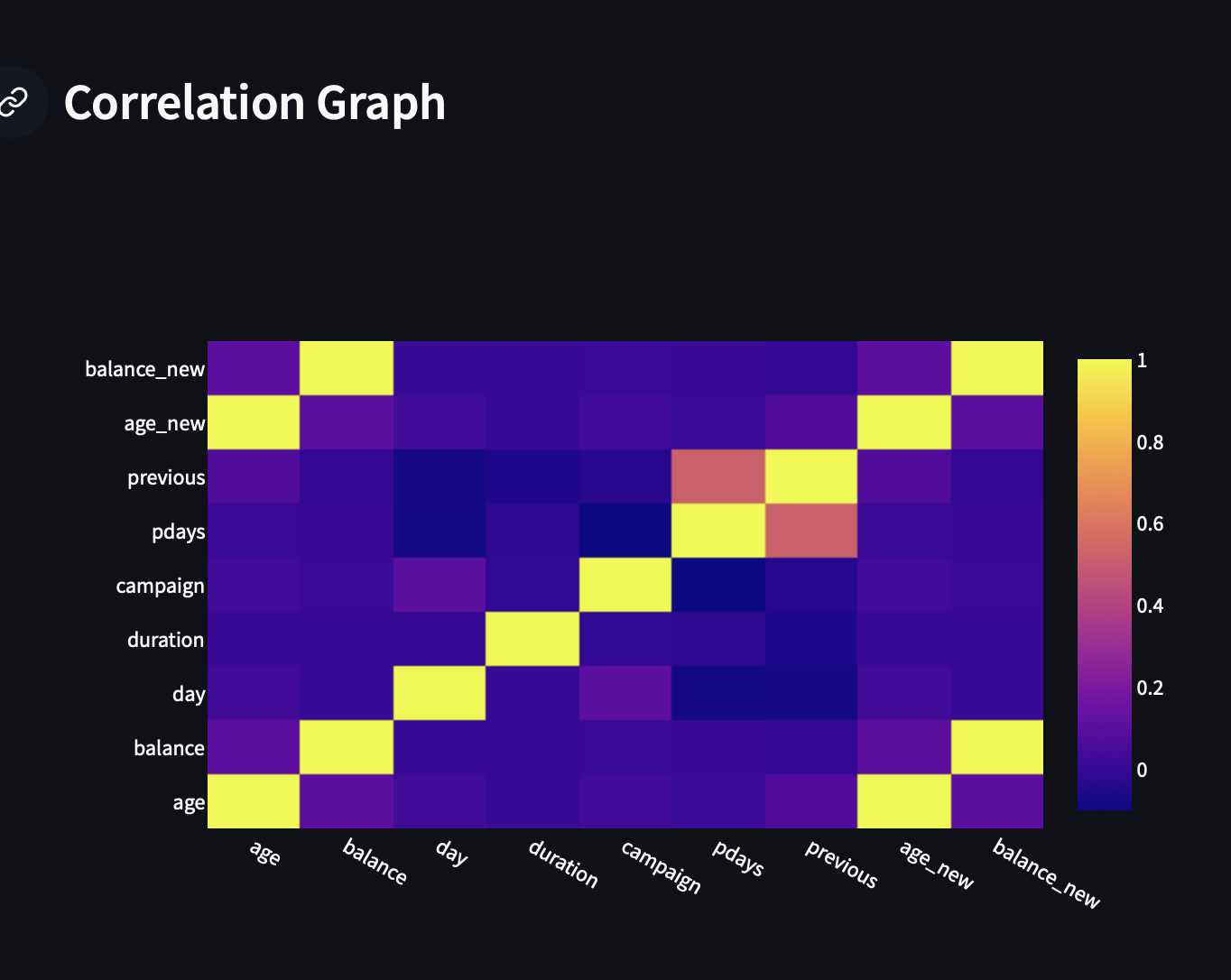
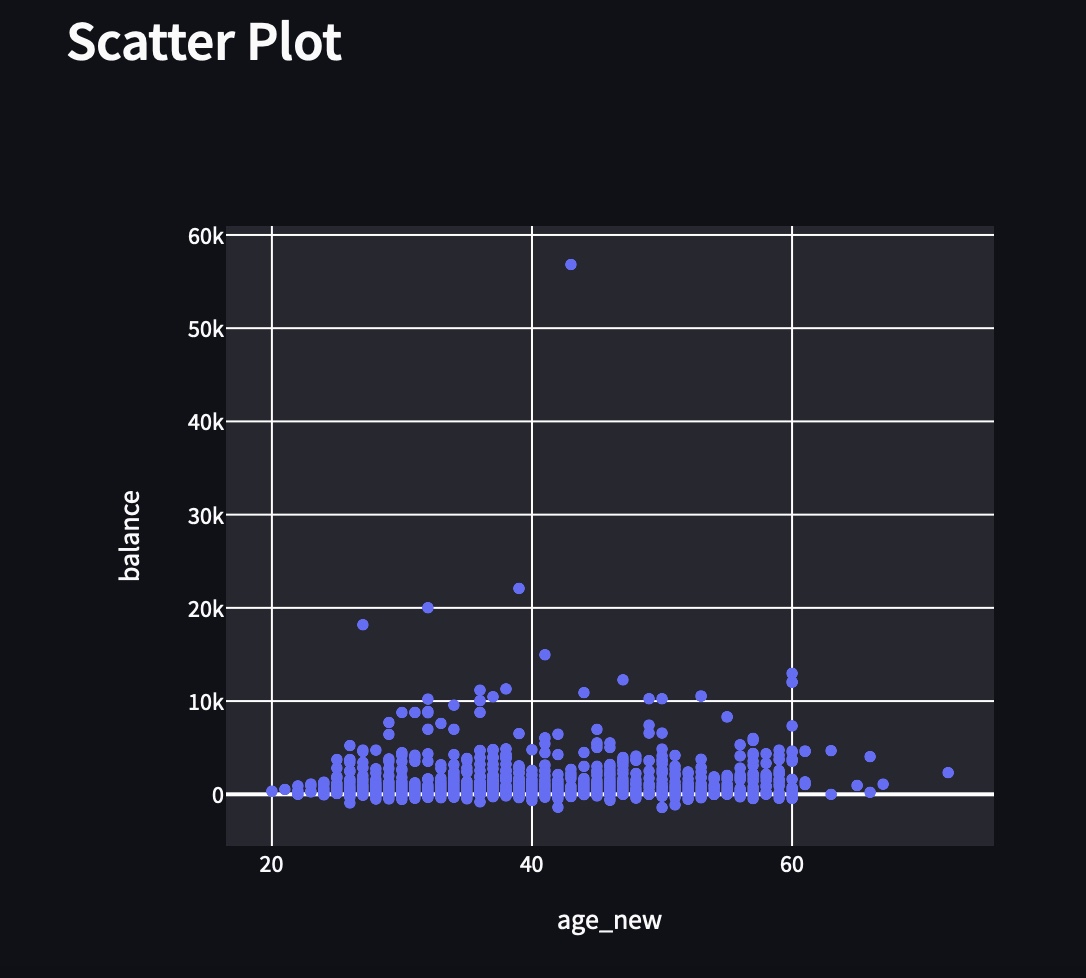


Figure 5: Correlation Graph

Figure 5 shows the correlation between different sectors like age, balance, day, duration, campaign, pdays and previous. This chart helps in knowing which sector is more related to another.



*Figure 6: Scatter Plot*

Figure 6 shows the difference in balance in different age groups using scatter plot. Here we can find how much a certain age group earns and also find out the outlier.

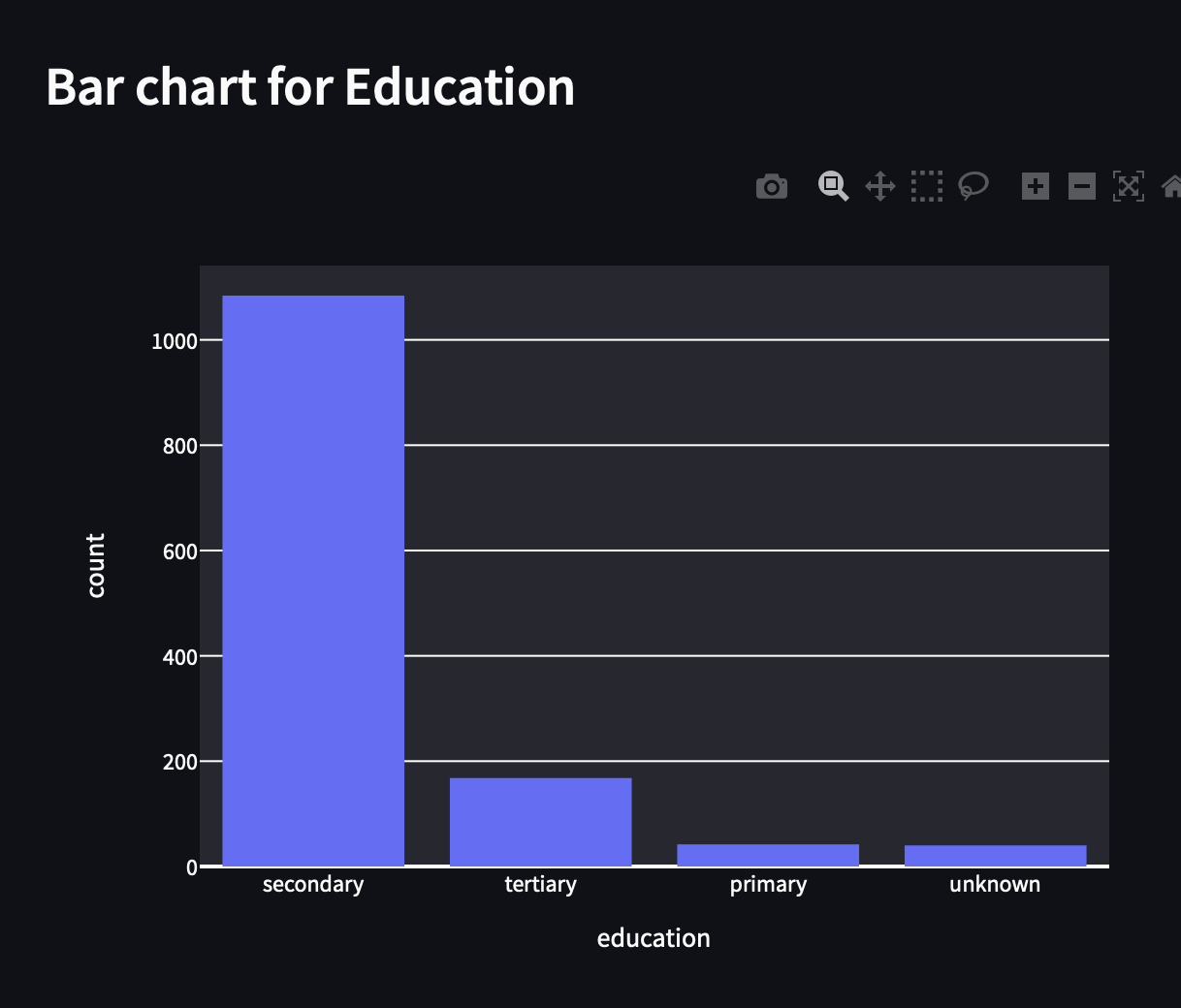


Figure 7: Histogram for Education Count

Figure 7 shows the bar chart representation of the secondary, primary,tertiary and unknown education count of the people using bank services.

##### Screenshot 2022-11-23 at 2.39.41 PM

Figure 10: Distribution for weekly demand data

Figure 10 shows the bar chart representation for housing with their count.

##### vi. discussion

We applied Exploratory Data Analysis technique which clarified the records and attributes more clearly and visually present the dataset by plotting graphs and charts. There are different types of charts: Bar, Scatter Plot, outliers plot etc. which can visually present the data in a more understandable way.

 Exploratory Data Analysis (EDA) can be used to understand data prior to doing formal statistical analyses such as creating tests of significance, creating confidence intervals, developing statistical models, We started with simple descriptive statistics that aid in characterizing the Age, Marital status and their balance according to their job .then outliers graph is plotted to identify noise in the data.accordingly we made are relationship between maritial and age in heat map. EDA is inherently graphical so well known and widely useful graphical methods are discussed. Next the use of scatter plots in the visualization of relationships between age and balance is discussed. Special attention is paid to the use of housing histogram in studying the variation in the data over time. This often leads to the identification of important aspect of data. Assessment of the stability of the underlying process is also a critical outcome of the analysis.

##### vii. conclusion

Data mining can be utilized to find covered up, obscure, however valuable learning from huge, fuzzy, uproarious, fragmented, and irregular data. In this paper, we present our task of visualizing the data set of bank details .By building a dashboard using streamlit we could understand, analyze and visualize in an efficient manner. When you are working with Streamlit, you don’t have to worry about your front-end knowledge. Streamlit framework will easily convert data scripts into a shareable web application with just a few lines of coding. From this project using the respective framework, we could understand more about the quantitative analysis  and the exploratory data analysis. By implementing this project, we could successfully conclude that data visualization provides a good, organized pictorial representation of the data which makes it easier to understand, observe, analyze.

##### **References**

1. “Data visualization,” Tableau. [Online]. Available: https://www.tableau.com/learn/articles/data-visualization#:~:text=Data%20visualization%20is%20the%20graphical,outliers%2C%20and%20patterns%20in%20data. [Accessed: 21-Nov-2022].
2. “An K. Brush and E. Burns, “Data visualization and it's important,” SearchBusinessAnalytics, 20-Feb-2020. [Online]. Available: https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization. [Accessed: 23-Nov-2022].
3. “Python programming language,” GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/python-programming-language/. [Accessed: 23-Nov-2022].
4. “Streamlit docs,” Streamlit documentation. [Online]. Available: https://docs.streamlit.io/. [Accessed: 23-Nov-2022].
5. “Getting started with plotly-python,” GeeksforGeeks, 15-Jul-2021. [Online]. Available: https://www.geeksforgeeks.org/getting-started-with-plotly-python/. [Accessed: 23-Nov-2022].
6. “Pandas tutorial,” GeeksforGeeks, 29-Feb-2020. [Online]. Available: https://www.geeksforgeeks.org/pandas-tutorial/. [Accessed: 23-Nov-2022].
7. Simplilearn, “23 best data visualization tools for 2023,” Simplilearn.com, 28-Oct-2022. [Online]. Available: https://www.simplilearn.com/data-visualization-tools-article. [Accessed: 23-Nov-2022].