



Green University Of Bangladesh
Department Of Computer Science and Engineering (CSE)
Faculty of Sciences and Engineering
Semester: (Fall, Year: 2023), B.Sc. in CSE (DAY)

LAB REPORT NO - 04
Course Title: Data Mining Lab
Course Code: CSE-436 **Section:** D6

Lab Experiment Name: Data Visualization

Student Details

| Name | | ID |
|------|----------------|-----------|
| 1 | Md Zahid Hasan | 201902060 |

Lab Date : 20/10/2023
Submission Date : 27/10/2023
Course Teacher Name : Abdullah Al Farhad

Lab Report Status

| | |
|-----------------------|------------------------|
| Mark:..... | Signature:..... |
| Comments:..... | Date:..... |

1 INTRODUCTION

Data visualization is defined as a graphical representation that contains the information and the data. By using visual elements like charts, graphs, and maps, data visualization techniques provide an accessible way to see and understand trends, outliers, and patterns in data.

2 OBJECTIVE

Here we have two tasks to do.

- Visualize datasets columns in different ways.
- To show two data-frames correlation using Pearson's correlation.

3 IMPLEMENTATION

3.1 Load Dataset & Distribution Plot

```
1 #Load the library and dataset
2 import pandas as pd
3 import numpy as np
4 df = pd.read_csv("/kaggle/input/titanic/train_and_test2.csv")
5 # visualize displot
6 updated_df = df
7 plot = sns.displot(updated_df['Age'], bins=15, kde=True, rug=False)
8 plt.savefig('/kaggle/working/displot.png')
9 plt.show()
```

3.2 Box & Violin Plot

```
1 plot=sns.catplot(x="Sex", y="Age", kind="box", data=updated_df) #Box &
   Whisker plot
2 plt.savefig('/kaggle/working/boxplot.png')
3 plt.show()
4 #Violin plot
5 plot=sns.violinplot(x="Sex",y="Age",data=updated_df,size=6)
6 plt.savefig('/kaggle/working/violinplot.png')
7 plt.show()
```

3.3 Bar Plot & Scatter Plot

```

1 #bar plot
2 plot=sns.catplot(x="Sex", y="Age", kind="bar", data = updated_df)
3 plt.savefig('/kaggle/working/barplot.png')
4 plt.show()
5 #Scatter plot
6 plot=sns.jointplot(x="Sex", y="Age", data=updated_df)
7 plt.savefig('/kaggle/working/scatterplot.png')
8 plt.show()

```

3.4 Multivariate Analysis

```

1 plot=sns.catplot(x="Sex", y="Age", hue="Pclass", kind="bar", data =
    updated_df)
2 plt.savefig('/kaggle/working/multiplot.png')
3 plt.show()

```

3.5 Pearson's Correlation with visualize

```

1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 # Creating DataFrames
5 dt = {'A': np.random.rand(100), 'B': np.random.rand(100)}
6 dt2 = {'C': np.random.rand(100), 'D': np.random.rand(100)}
7 data_frame1 = pd.DataFrame(dt)
8 data_frame2 = pd.DataFrame(dt2)
9 # Calculating Pearson's correlation
10 correlation = data_frame1.corrwith(data_frame2)
11 # Visualize the correlation using a scatter plot
12 plt.figure(figsize=(6, 6))
13 plt.scatter(data_frame1['A'], data_frame2['C'], label='A vs C', color='b',
    marker='o')
14 plt.scatter(data_frame1['B'], data_frame2['D'], label='B vs D', color='r',
    marker='x')
15 plt.title("Scatter Plot of Correlation")
16 plt.xlabel("DataFrame 1")
17 plt.ylabel("DataFrame 2")
18 plt.legend()
19 plt.grid(True)
20 plt.show()

```

3.6 Pearson's Correlation

```

1 print(updated_df["Age"].corr(updated_df["Sex"]))
2 print(updated_df["Age"].corr(updated_df["Sex"], method='spearman'))
3 # pearson', 'spearman', 'kendall', or a callable, 'zero'

```

4 OUTPUT

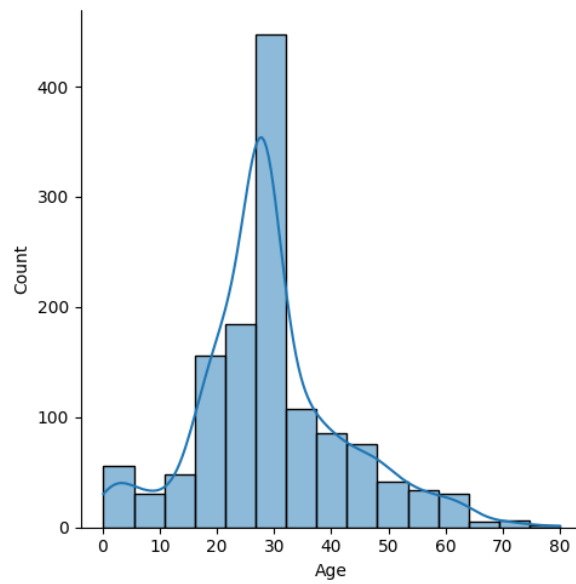


Figure 1: Distplot visualization

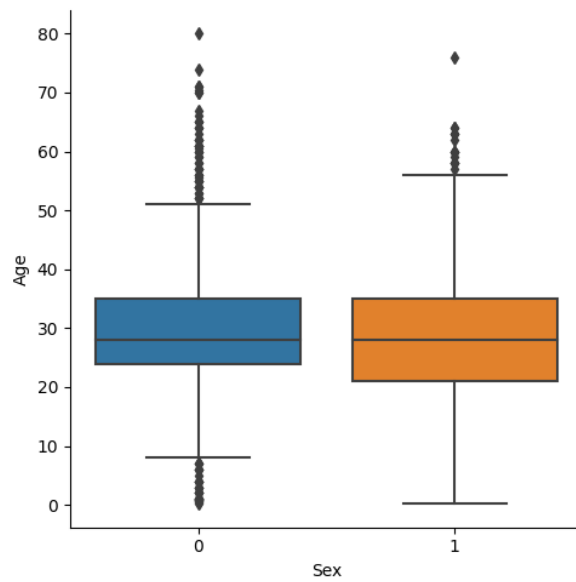


Figure 2: Box Plot

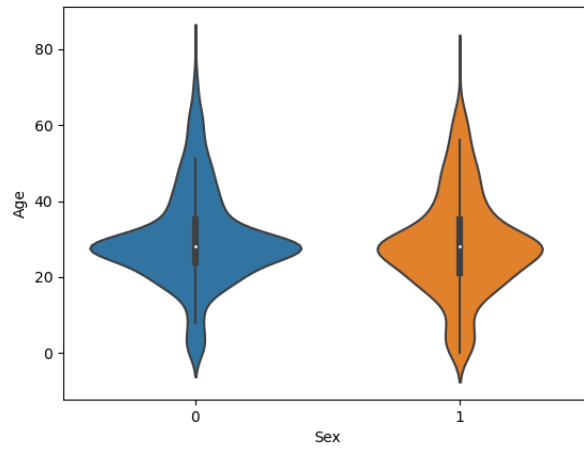


Figure 3: Violin Plot

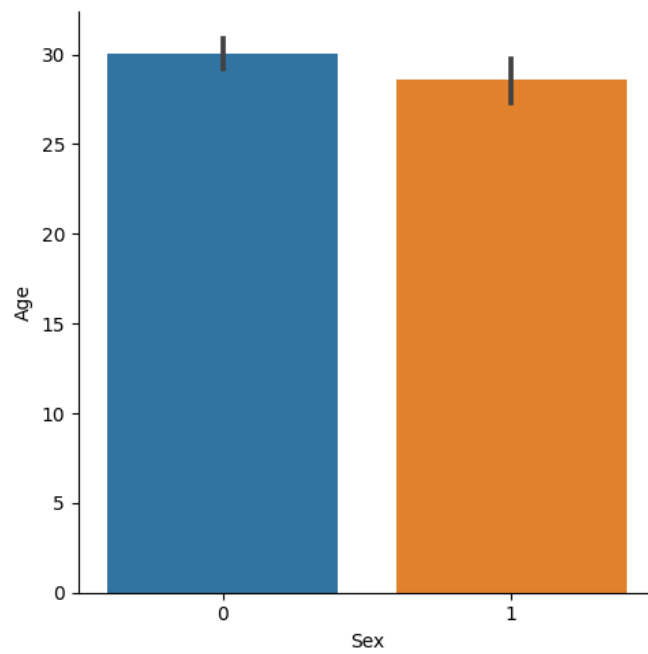


Figure 4: Bar Plot

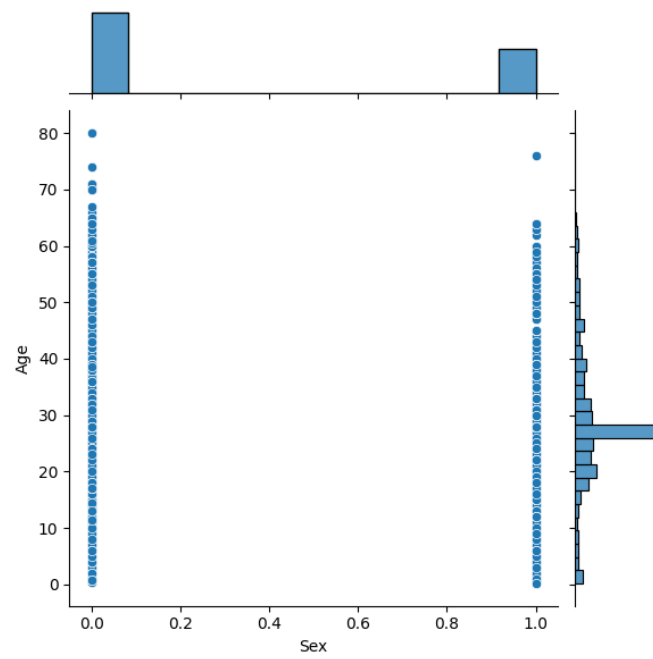


Figure 5: Scatter Plot

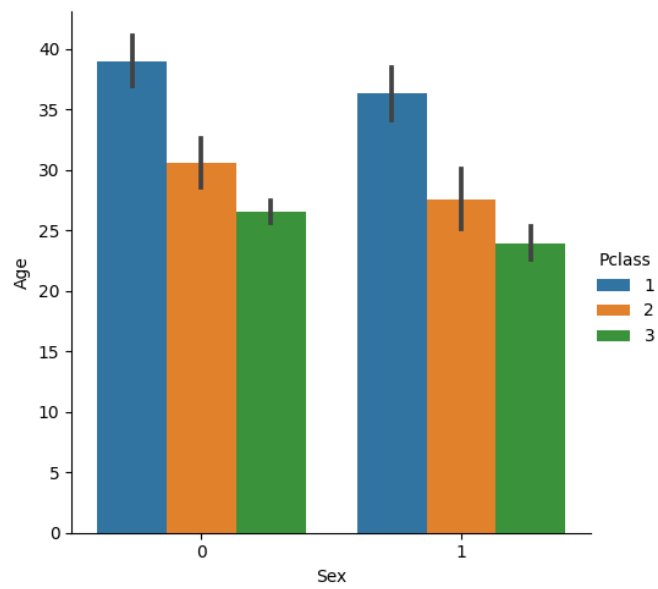


Figure 6: Multivariate

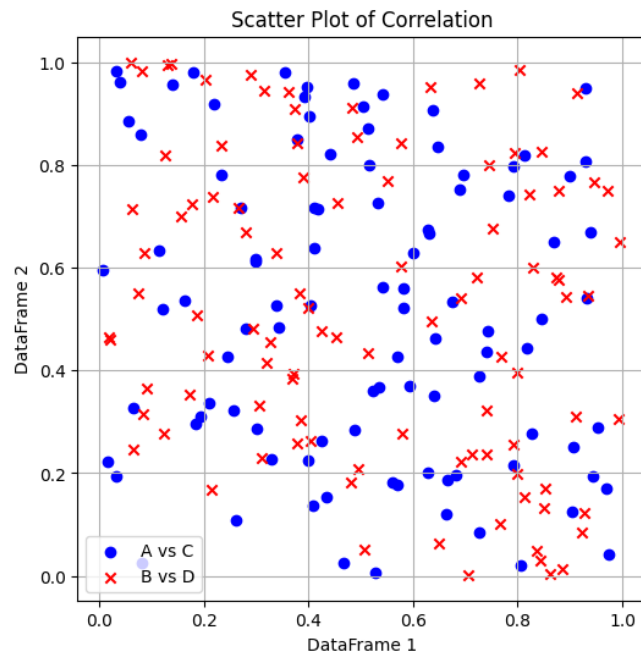


Figure 7: Pearson's Correlation with visualize

```
-0.05366328477824959
-0.050203064400914835
```

Figure 8: Pearson's Correlation

5 DISCUSSION & ANALYSIS

I conducted data visualization and analysis using Kaggle's platform, focusing on the Titanic dataset. Several plots were created, including distplots, boxplots, scatter plots, bar plots, and a violin plot. The distplot revealed that the Age count. Boxplots showed variations in Sex(gender), particularly among Age. Scatter plots helped identify relationships between variables, such as Sex(gender) and Age. Bar plots demonstrated the Sex(gender) distribution across Age. Then, the violin plot indicated the age distribution among different Sex(gender). Finally the Pearson's Correlation.

In summary, these visualizations provided valuable insights into the Titanic dataset, showcasing the age and fare distribution, class demographics, and gender-based age variations. Kaggle's platform facilitated this analysis, making it accessible and efficient.