

Experiment - 04

Aim: To perform Data Visualization techniques

Theory: Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, 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.

Our eyes are drawn to colors and patterns. We can quickly identify red from blue, and squares from circles. 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.

Data visualization techniques are ways to represent data visually to make it easier to understand. Some techniques include:

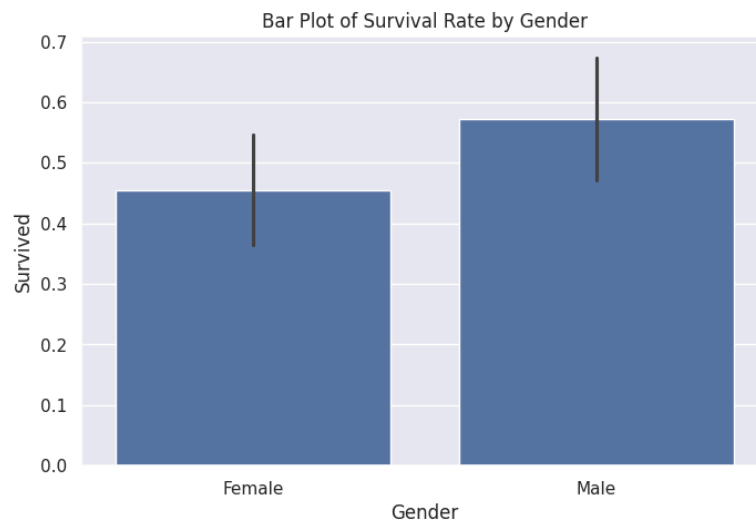
- **Bar plots:** A bar plot uses rectangular bars to represent data categories, with bar length or height proportional to their values. It compares discrete categories, with one axis for categories and the other for values.
- **Histograms:** Histogram is a type of graphical representation used in statistics to show the distribution of numerical data. It looks somewhat like a bar chart, but unlike bar graphs, which are used for categorical data, histograms are designed for continuous data, grouping it into logical ranges which are also known as "bins."
- **Box plots:** Box Plot is a graphical method to visualize data distribution for gaining insights and making informed decisions. Box plot is a type of chart that depicts a group of numerical data through their quartiles.
- **Count plots:** The countplot is used to represent the occurrence(counts) of the observation present in the categorical variable. It uses the concept of a bar chart for the visual depiction.
- **Scatter plots:** A scatter plot (aka scatter chart, scatter graph) uses dots to represent values for two different numeric variables. The position of each dot on the horizontal and vertical axis indicates values for an individual data point. Scatter plots are used to observe relationships between variables.
- **Pie Charts:** A pie chart is a type of graph representing data in a circular form, with each slice of the circle representing a fraction or proportionate part of the whole. All slices of the pie add up to make the whole equaling 100 percent.
- **Line plots:** A line plot is a type of graph that displays data points along a number line. It is basically useful to provide a clear and concise representation of trends, patterns, and changes that occur over time.

Code & Output:

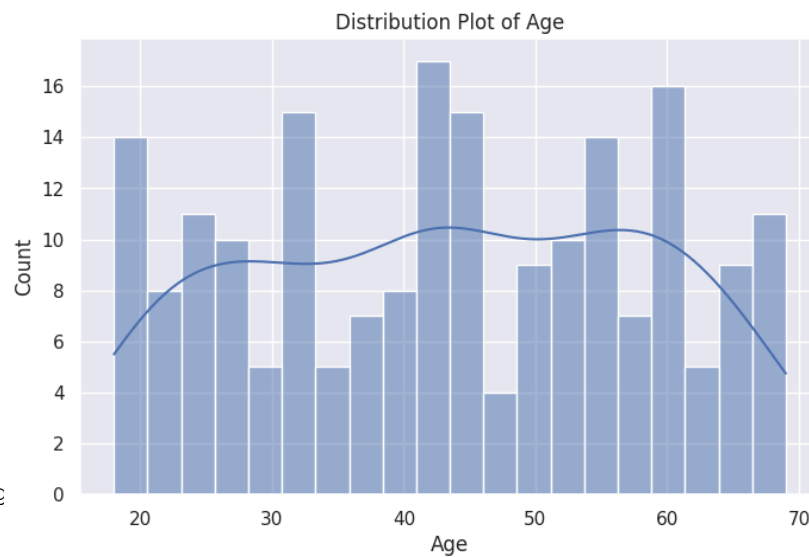
```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
import pandas as pd
import kagglehub

path = kagglehub.dataset_download("himelsarder/road-accident-survival-dataset")
df = pd.read_csv(path + "/accident.csv")
df.shape
df.head()

plt.figure(figsize=(8,5))
sns.barplot(x="Gender", y="Survived", data=df)
plt.title("Bar Plot of Survival Rate by Gender")
plt.show()
```

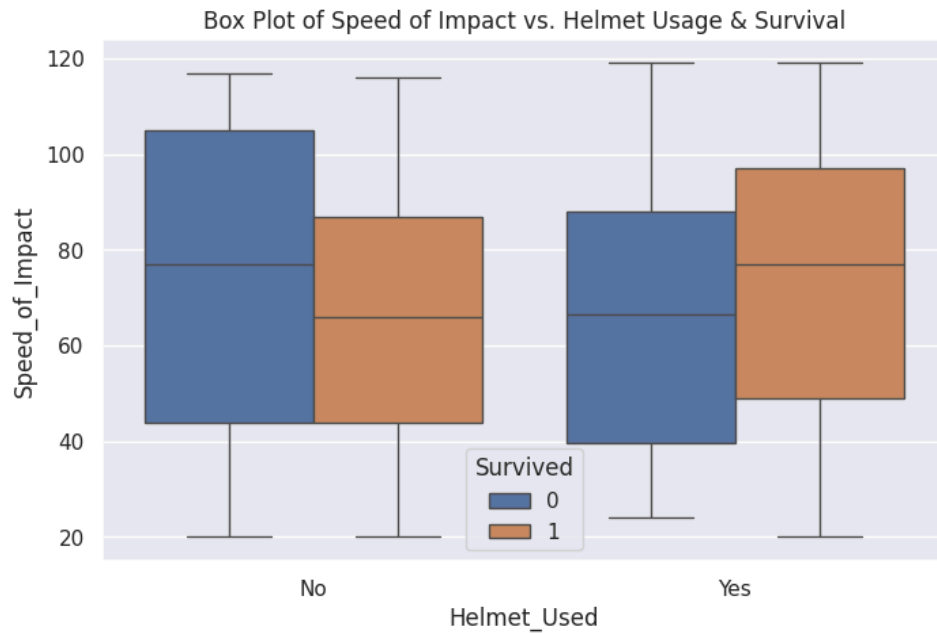


```
plt.figure(figsize=(8,5))
sns.histplot(df["Age"], bins=20, kde=True)
plt.title("Distribution Plot of Age")
plt.show()
```

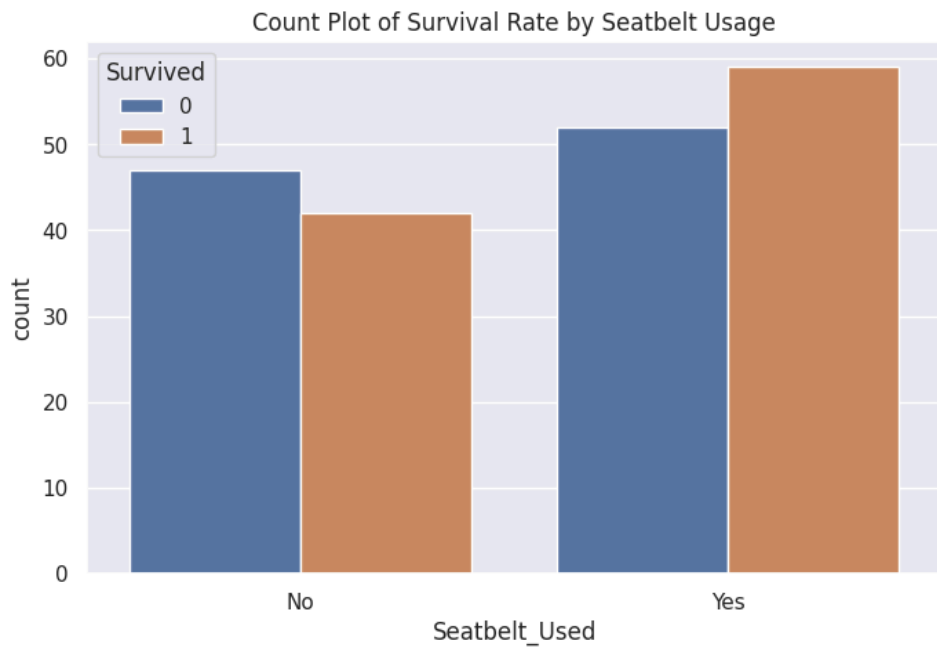


```
plt.figure(fig
```

```
sns.boxplot(x="Helmet_Used", y="Speed_of_Impact", hue="Survived", data=df)
plt.title("Box Plot of Speed of Impact vs. Helmet Usage & Survival")
plt.show()
```

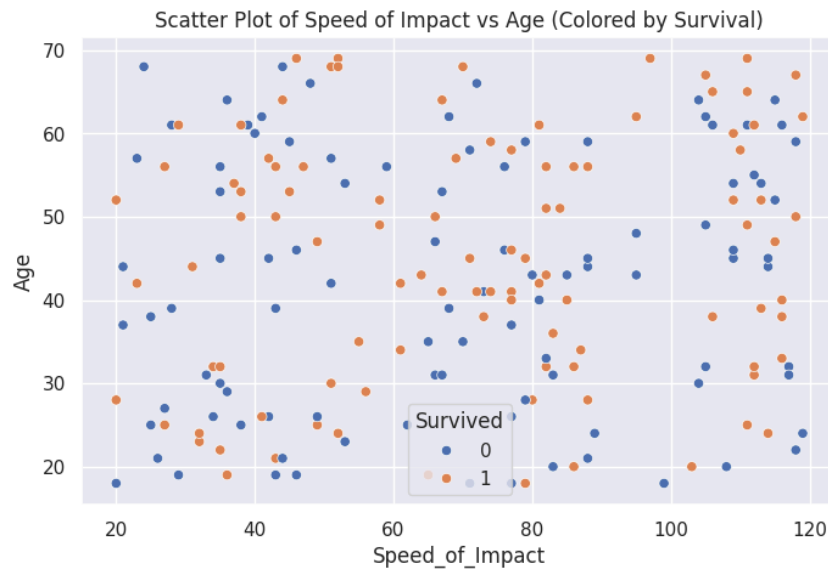


```
plt.figure(figsize=(8, 5))
sns.countplot(x="Seatbelt_Used", hue="Survived", data=df)
plt.title("Count Plot of Survival Rate by Seatbelt Usage")
plt.show()
```

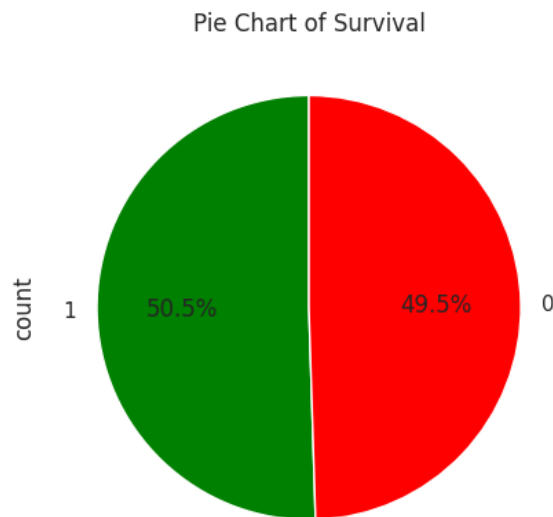


```
plt.figure(figsize=(8, 5))
```

```
sns.scatterplot(x="Speed_of_Impact", y="Age", hue="Survived", data=df)
plt.title("Scatter Plot of Speed of Impact vs Age (Colored by Survival)")
plt.show()
```



```
plt.figure(figsize=(8, 5))
counts = df['Survived'].value_counts()
counts.plot(kind='pie', autopct='%1.1f%%', startangle=90,
            colors=['green', 'red'])
plt.title('Pie Chart of Survival')
plt.show()
```



Conclusion: Hence, we performed data visualization techniques on a Road Accidents dataset