Task 2: To visualize and perform Univariate analysis using continuous and categorical data

Categorical Data - Bar chart, Pie Chart

Continuous data – Scatterplot, Line Plot, Strip Plot, Swarm Plot, Histogram, Density Plot, Rug Plot.

Tools: Tableau, Language :Python

2.a) Choose any one of the dataset from website. Identify and utilize the categorical data(attribute) and continuous data(attribute). Construct a Bar chart, Pie Chart using Univariate **analysis of Categorical** data using identified data. Construct a Scatter plot, Line Plot, Strip Plot, Swarm Plot using Univariate **analysis of continuous data** using identified data.

Aim:

To visualize categorical and continuous data from a chosen dataset using Python visualization libraries such as Matplotlib and Seaborn, applying univariate analysis.

Algorithm:

Choose Dataset: Select a dataset with categorical and continuous attributes.

Identify Data Types: Distinguish between categorical (non-numeric) and continuous (numeric) attributes.

Construct Bar and Pie Charts: Create frequency tables for categorical data, then plot bar and pie charts to visualize distribution.

Construct Scatter, Line, Strip, and Swarm Plots: Utilize appropriate plots for continuous data analysis, considering relationships, trends, and distributions.

Interpret Visualizations: Analyze insights gained from visualizations, identifying patterns, dominant categories, and correlations.

Consider Further Analysis: Reflect on implications for further exploration or decision-making based on the analysis.

Document Findings: Summarize findings and observations for reporting or presentation purposes.

Program:

Bar chart:

import numpy as np

import pandas as pd

```
import matplotlib.pyplot as plt

import seaborn as sns

df=pd.read_csv("/content/sentimentdataset.csv")

platform_likes_top5 =
df.groupby('Platform')['Likes'].sum().sort_values(ascending=False).head(5)


# Create a pie chart

plt.figure(figsize=(4, 4))

platform_likes_top5.plot(kind='pie', autopct='%1.1f%%', startangle=140, colors=['skyblue',
'lightcoral', 'lightgreen', 'lightsalmon', 'lightsteelblue'])

plt.title('Top 5 Platforms by Total Likes')

plt.ylabel('')  # Remove the 'Platform' label on the y-axis

plt.show()
```
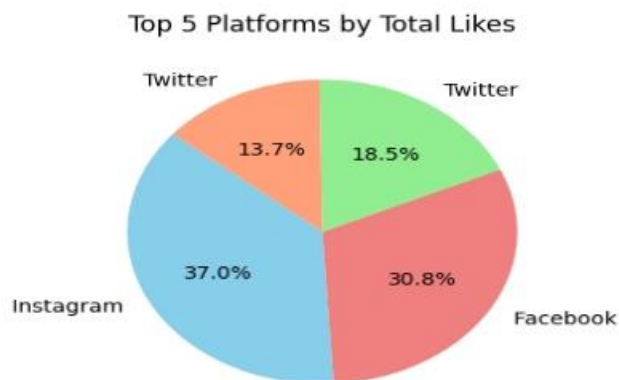
Output:

```
top5_countries = df.nlargest(5, 'Likes')
```


Pie chart:

```
# Create a bar chart

plt.figure(figsize=(10, 6))

plt.bar(top5_countries['Country'], top5_countries['Likes'], color=['skyblue', 'lightcoral',
'lightgreen', 'lightsalmon', 'lightsteelblue'])

plt.xlabel('Country')

plt.ylabel('Total Likes')
```
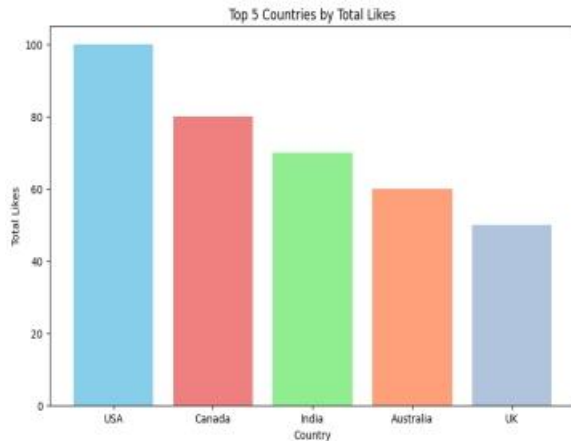
```
plt.title('Top 5 Countries by Total Likes')

plt.show()
```

Output:


```
data = {

    'Text': ['Enjoying a beautiful day at the park!', 'Traffic was terrible this morning.', 'Just
finished an amazing workout!', 'Excited about the upcoming weekend getaway!', 'Trying out a
new recipe for dinner tonight.'],

    'Retweets': [15, 5, 20, 8, 12],

    'Likes': [30, 10, 40, 15, 25]

}


df = pd.DataFrame(data)
```

Scatter Plot:

```
# Set the Seaborn style

sns.set(style="whitegrid")

# Create a scatter plot using Seaborn

plt.figure(figsize=(10, 6))

sns.scatterplot(x='Retweets', y='Likes', data=df, color='blue', alpha=0.7)

plt.title('Scatter Plot of Retweets vs Likes')

plt.show()
```
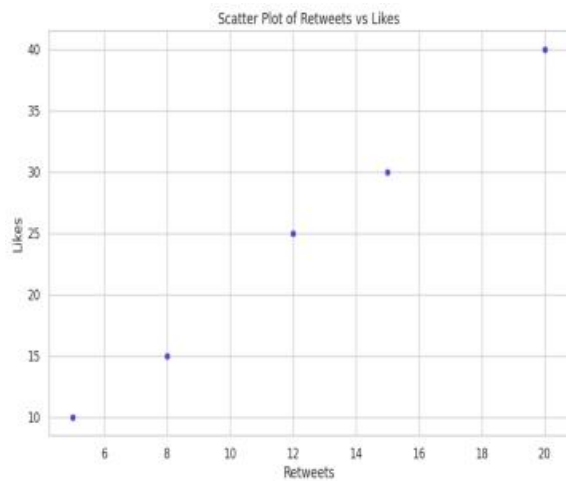
Output:

Scatter Plot of Retweets vs Likes

Result:

The dataset was successfully visualized using bar, pie, scatter, line, strip, and swarm plots. The univariate analysis clearly showed category distributions and relationships between continuous variables.

2. b) Utilize a given source of dataset that contains employee details.

Identify the **categorical data (attribute) and continuous data (attribute).**

Construct a **Bar chart, Pie Chart using Univariate analysis of Categorical Data**

Construct a **Histogram, Density Plot, and Rug Plot using Univariate analysis of Categorical Data**. Dataset Source: Employee dataset (kaggle.com)

Aim:

To analyze categorical and continuous employee data using Python visualization tools through univariate plots such as bar, pie, histogram, density, and rug plots.

Algorithm:

1. **Start** the program and import necessary libraries (pandas, matplotlib, seaborn).

2. **Load** the Employee dataset from a CSV file.

3. **Identify** categorical attributes (e.g., Gender, City) and continuous attributes (e.g., Age, Joining Year).

4. **Plot Bar Chart** and **Pie Chart** to visualize the distribution of categorical data.

5. **Plot Histogram**, **Density Plot**, and **Rug Plot** to analyze the spread and distribution of continuous data.

6. **Observe** and interpret patterns or trends from the visualizations.

7. **End** the program.

Program:

Pie chart:

import pandas as pd

import matplotlib.pyplot as plt

Read the dataset from a CSV file

file_path = 'Employee.csv' # Replace 'your_dataset.csv' with the actual file path

df = pd.read_csv(file_path

Extract the column with categorical data for plotting (e.g., 'Gender')

categorical_column = 'Gender' # Replace with the actual categorical column name

Count the occurrences of each category

category_counts = df[categorical_column].value_counts(

Create a pie chart

```python
plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', startangle=90,
colors=['skyblue', 'lightcoral'])

plt.title(f'Pie Chart of {categorical_column}')

plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle

# Show the plot

plt.show()
```
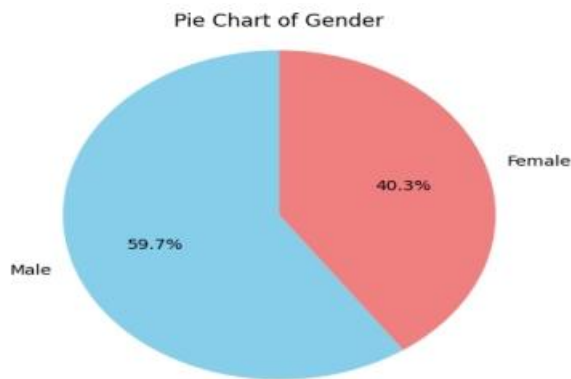
Output:

Pie Chart of Gender

 Bar chart:

```python
import pandas as pd

import matplotlib.pyplot as plt


# Read the dataset from a CSV file

file_path = 'Employee.csv'  # Replace 'your_dataset.csv' with the actual file path

df = pd.read_csv(file_path)


# Extract the column with categorical data for plotting (e.g., 'City')

categorical_column = 'City'  # Replace with the actual categorical column name


# Count the occurrences of each category

category_counts = df[categorical_column].value_counts()
```

```
# Create a bar chart
plt.bar(category_counts.index, category_counts, color=['skyblue', 'lightcoral', 'lightgreen',
'orange'])
plt.title(f'Bar Chart of {categorical_column}')
plt.xlabel(categorical_column)
plt.ylabel('Count')


# Show the plot
plt.show()
```
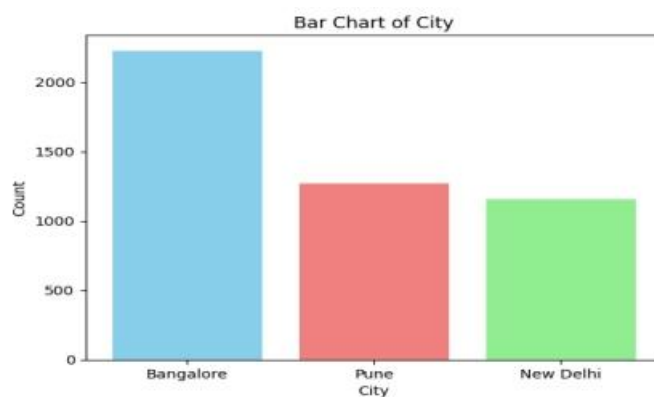
Output:

 Histogram:

```
import pandas as pd
import matplotlib.pyplot as plt


# Read the dataset from a CSV file
file_path = '/content/Employee.csv'  # Replace 'your_dataset.csv' with the actual file path
df = pd.read_csv(file_path)


# Extract the column with continuous data for plotting (e.g., 'Age' column)
continuous_column = 'Age'  # Replace 'Age' with the actual column name
# Create a histogram
plt.hist(df[continuous_column], bins=10, edgecolor='black')  # You can adjust the number of
bins as needed, bins=blue bars
```
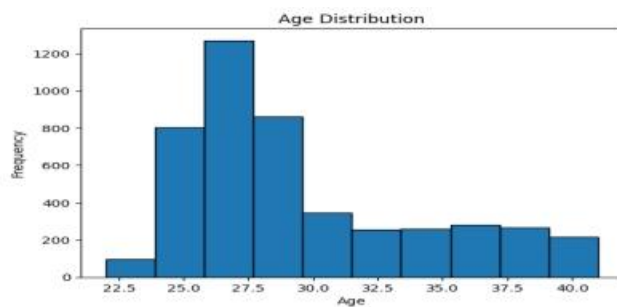
```python
plt.title(f'{continuous_column} Distribution')

plt.xlabel(continuous_column)

plt.ylabel('Frequency')

# Show the plot

plt.show()
```
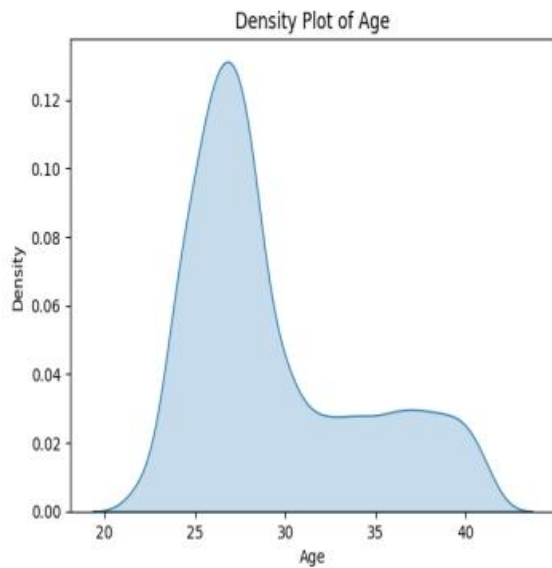
Output:

Density Plot:

```python
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Read the dataset from a CSV file

file_path = 'Employee.csv'  # Replace 'your_dataset.csv' with the actual file path

df = pd.read_csv(file_path)

# Extract the column with continuous data for plotting (e.g., 'Age' column)

continuous_column = 'Age'  # Replace 'Age' with the actual column name

# Create a density plot

sns.kdeplot(df[continuous_column], fill=True)

plt.title(f'Density Plot of {continuous_column}')

plt.xlabel(continuous_column)

plt.ylabel('Density')

# Show the plot

plt.show()
```

Output:

Density Plot of Age

Rug Plot:

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Read the dataset from a CSV file

file_path = 'Employee.csv'  # Replace 'your_dataset.csv' with the actual file path

df = pd.read_csv(file_path)

# Extract the column with continuous data for plotting (e.g., 'ExperienceInCurrentDomain')

continuous_column = 'JoiningYear'  # Replace with the desired column name

# Create a rug plot

sns.rugplot(df[continuous_column], height=0.5)

plt.title(f'Rug Plot of {continuous_column}')

plt.xlabel(continuous_column)

plt.yticks([])  # Remove y-axis ticks for better clarity (optional)

# Show the plot

plt.show()
```
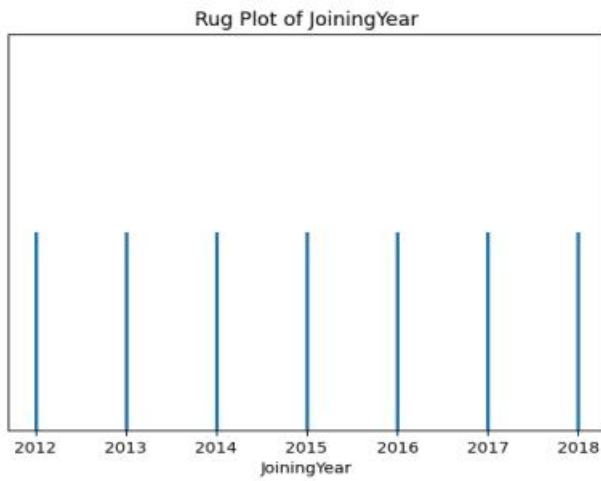
Output:

Rug Plot of JoiningYear

Result:

Employee data was analyzed using bar, pie, histogram, density, and rug plots, revealing clear patterns in categorical attributes and the spread of continuous variables like age and joining year.