Task 3: To visualize and perform Bivariate analysis using continuous and categorical data

Categorical vs. Categorical: Stacked Bar Chart, Grouped Bar chart, Segmented Bar Chart, Mosaic Plots .

Continuous vs. Continuous: Scatterplot Fit Lines

Categorical vs. Continuous : Bar Chart (Summary statistics), Grouped Kernel Density Plots, Box Plots, Violin Plots, Ridgeline Plots, Beeswarm Plots. CO2,S3

Tools: Tableau, Language :Python

Link for dataset: Student Alcohol Consumption (kaggle.com)

Identify Columns holding categorical data
and Columns holding continuous data.

Aim:

Construct a Stacked Bar Chart, Grouped Bar chart, Segmented Bar Chart using Bivariate analysis of categorical vs categorical data for the attributes of approved and gender in above data set.

1. Construct a Scatterplot Fit Lines using bivariate analysis of Continuous vs. Continuous data.

2. Identify the categorical vs **Continuous data to plot the** Bar Chart,Grouped Kernel Density Plots, Box Plots, Violin Plots, Ridgeline Plots, Beeswarm Plots.

Algorithm:

1. Select Dataset: Choose a dataset containing both categorical and continuous variables.

2. Differentiate Variables: Identify categorical and continuous variables.

3.Categorical vs. Categorical: Create stacked, grouped, or segmented bar charts, along with mosaic plots, to visualize relationships between categorical variables.

4. Continuous vs. Continuous: Generate scatterplots with fit lines to explore relationships between two continuous variables.

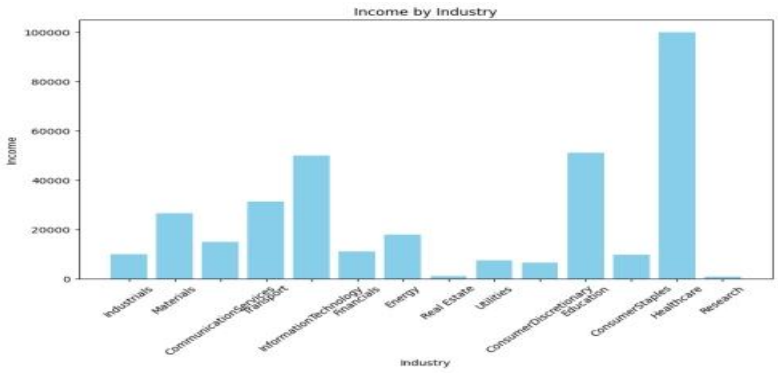
5. Categorical vs. Continuous: Construct bar charts for summary statistics and use grouped kernel density plots, box plots, violin plots, ridgeline plots, or beeswarm plots to visualize distributions across categories.

6. Interpretation: Analyze visualizations for insights into relationships, distributions, and patterns, drawing conclusions for further analysis or decision-making.

Program:

```
import pandas as pd
import matplotlib.pyplot as plt
# Read data from CSV file
df = pd.read_csv('/content/clean_dataset.csv')
# Extract the two columns
industry = df['Industry']
income = df['Income']
# Create a bar chart
plt.figure(figsize=(10, 6))
plt.bar(industry, income, color='skyblue')
plt.xlabel('Industry')
plt.ylabel('Income')
plt.title('Income by Industry')
plt.xticks(rotation=45)
plt.show()
```

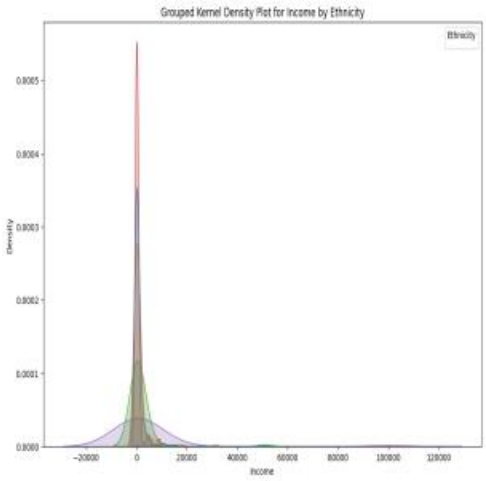
Output:




```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Read data from CSV file
df = pd.read_csv('/content/clean_dataset.csv')
```

```
# Extract the two columns
categorical_column = 'Ethnicity'
continuous_column = 'Income'
# Create grouped kernel density plots
plt.figure(figsize=(12, 8))
sns.kdeplot(data=df, x=continuous_column, hue=categorical_column, fill=True, common_norm=False)
plt.xlabel(continuous_column)
plt.title(f'Grouped Kernel Density Plot for {continuous_column} by {categorical_column}')
plt.legend(title=categorical_column)
plt.show()
```

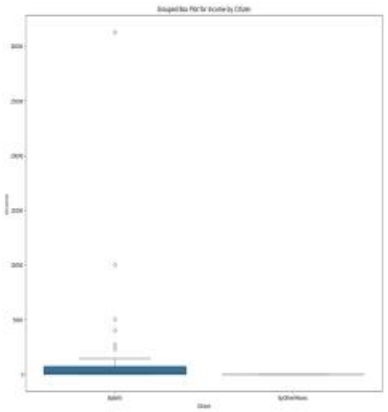
Output:




```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Read data from CSV file
df = pd.read_csv('/content/clean_dataset.csv', nrows=50)
# Extract the two columns
```

```
categorical_column = 'Citizen'
continuous_column = 'Income'
# Create grouped box plots
plt.figure(figsize=(16, 12))
sns.boxplot(data=df, x=categorical_column, y=continuous_column)
plt.xlabel(categorical_column)
plt.ylabel(continuous_column)
plt.title(f'Grouped Box Plot for {continuous_column} by {categorical_column}')
plt.show()
```

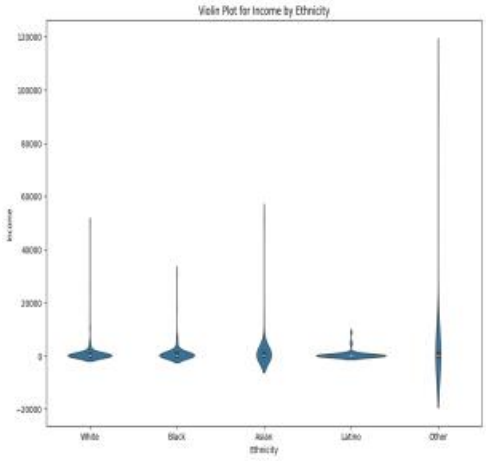
Output:



```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Read data from CSV file
df = pd.read_csv('/content/clean_dataset.csv')
# Extract the two columns
categorical_column = 'Ethnicity'
continuous_column = 'Income'
# Create violin plots
plt.figure(figsize=(12, 8))
```

```
sns.violinplot(data=df, x=categorical_column, y=continuous_column)
plt.xlabel(categorical_column)
plt.ylabel(continuous_column)
plt.title(f'Violin Plot for {continuous_column} by {categorical_column}')
plt.show()
```

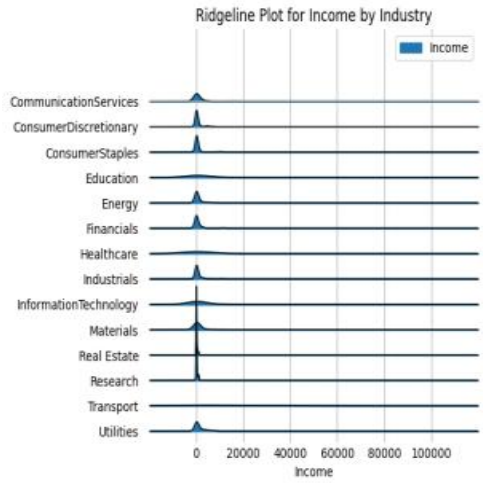
Output:




```
import pandas as pd
from joypy import joyplot
import matplotlib.pyplot as plt

# Read data from CSV file
df = pd.read_csv('/content/clean_dataset.csv')

# Extract the two columns
categorical_column = 'Industry'
continuous_column = 'Income'

# Create ridgeline plots
plt.figure(figsize=(12, 8))
```

```
joyplot(
    data=df,
    by=categorical_column,
    column=continuous_column,
    kind='kde',  # Use KDE (Kernel Density Estimation) for the ridgeline plot
    fill=True,
    linecolor="black",
    grid=True,
    linewidth=1,
    legend=True,
)

plt.xlabel(continuous_column)
plt.title(f'Ridgeline Plot for {continuous_column} by {categorical_column}')
plt.show()
```

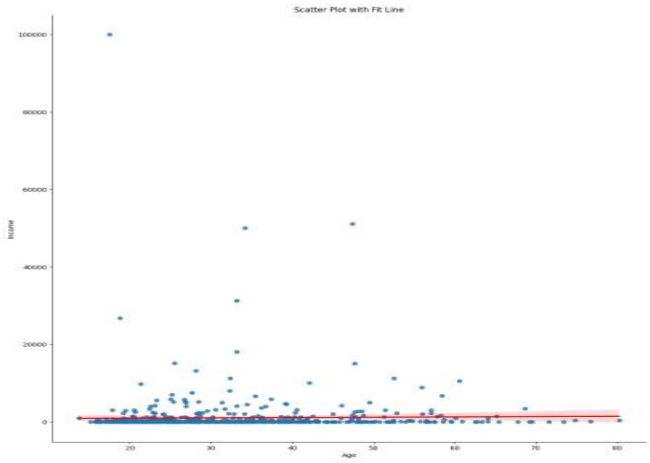
Output:




```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv('/content/DV task-4.csv')
```

```
# Assuming 'Age' is on the x-axis and 'Income' is on the y-axis
sns.lmplot(data=df, x='Age', y='Income', height=6, line_kws={'color': 'red'})
plt.title('Scatter Plot with Fit Line')
plt.xlabel('Age')
plt.ylabel('Income')
plt.show()
```

Output:



Result:

Bivariate analysis was successfully performed using various plots. Relationships between categorical and continuous variables were clearly visualized, showing patterns and correlations among attributes like gender, approval, age, and income.