Task 4: To visualize and perform Multivariate analysis using Multiple variables involving Multiple measures

Scatterplot Matrix, Parallel Coordinates, Line Graph, Stacked Bar Chart

Tools: Tableau, Language :Python

Link for dataset:https://www.kaggle.com/datasets/samuelcortinhas/credit-card-approval-clean-data. This dataset contains a cleaned version of [this dataset from UCI machine learning repository](#) on **credit card approvals**. Missing values have been filled and feature names and categorical names have been inferred, resulting in more context and it being easier to use.

Aim:

Construct a Scatterplot Matrix, Parallel Coordinates, Line Graph, Stacked Bar Chart, using Multiple variables involving Multiple measures

Algorithm:

1. Import necessary libraries including NumPy, pandas, seaborn, matplotlib.pyplot, and plotly.express.

2. Load the dataset from a zip file using pandas `read_csv` function.

3. Print the loaded dataset to inspect its structure and contents.

4. Select numeric columns (Age, Debt, YearsEmployed), create a pairplot using seaborn to visualize pairwise relationships, and display the plot.

5. Use plotly.express to create a parallel coordinates plot, color-coded by the 'Approved' column, with dimensions as Age, Debt, and CreditScore, and show the plot.

6. Extract the first 20 entries of the dataset and plot the Age against Debt and CreditScore using a line plot with different colors for Debt and CreditScore, then display the plot.

7. Group the first 20 entries by Age, summing up Debt and CreditScore for each Age group, create a stacked bar chart representing Debt and CreditScore for each Age group, and display the chart.

Program:

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

import plotly.express as px

```python
data = {
    'Age': [22, 25, 28, 30, 32, 35, 38, 40, 42, 45, 48, 50, 52, 55, 58, 60, 62, 65, 67, 70],
    'Debt': [500, 800, 1200, 1500, 1800, 2000, 2200, 2500, 2600, 3000, 3200, 3500, 3600,
3700, 4000, 4200, 4300, 4500, 4600, 4800],
    'YearsEmployed': [1, 2, 3, 4, 6, 7, 8, 10, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 25],
    'CreditScore': [600, 620, 650, 670, 680, 700, 710, 730, 740, 750, 760, 770, 780, 790, 800,
810, 820, 830, 840, 850],
    'Approved': [1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1]
}
df = pd.DataFrame(data)
print("Sample Dataset:\n", df.head())
numeric_subset = df[['Age', 'Debt', 'YearsEmployed']]
sns.pairplot(numeric_subset)
plt.show()
attributes = ["Age", "Debt", "CreditScore"]
fig = px.parallel_coordinates(
    df,
    color="Approved",
    dimensions=attributes,
    color_continuous_scale=px.colors.diverging.Tealrose,
    color_continuous_midpoint=0.5
)
fig.show()
first_20_entries = df.head(20)
age = first_20_entries['Age']
debt = first_20_entries['Debt']
credit_score = first_20_entries['CreditScore']
plt.figure(figsize=(10, 6))
plt.plot(age, debt, label='Debt', color='blue')
plt.plot(age, credit_score, label='Credit Score', color='green')
```
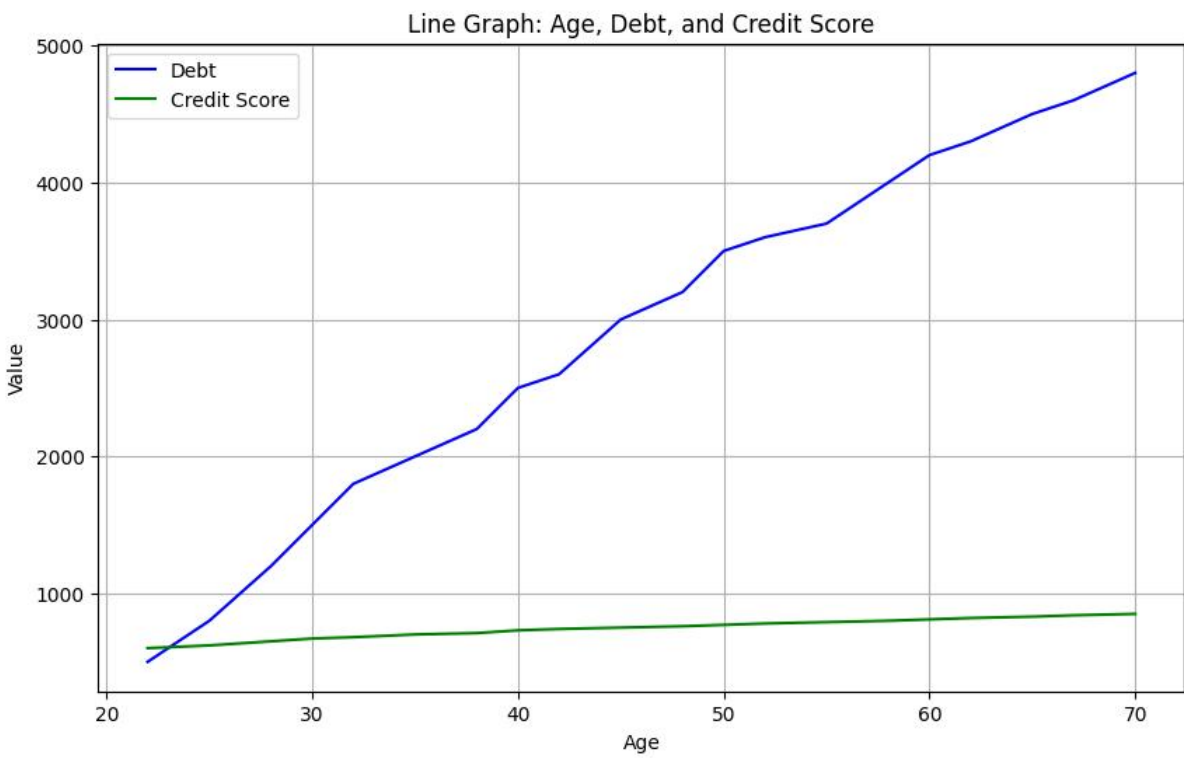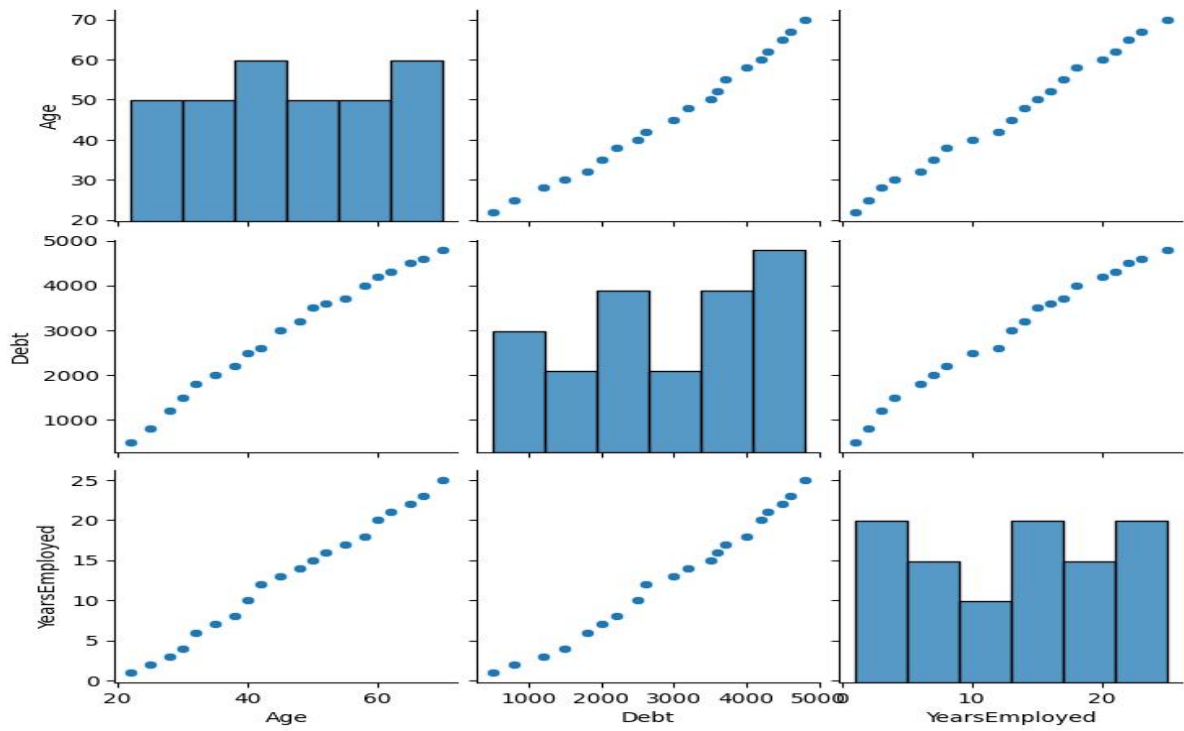
plt.xlabel('Age')

plt.ylabel('Value')

plt.title('Line Graph: Age, Debt, and Credit Score')

plt.legend()

plt.grid(True)

plt.show()grouped_data = first_20_entries.groupby('Age').sum()[['Debt', 'CreditScore']]

grouped_data.plot(kind='bar', stacked=True, figsize=(12, 8))

plt.xlabel('Age')

plt.ylabel('Value')

plt.title('Stacked Bar Chart: Debt and Credit Score by Age')

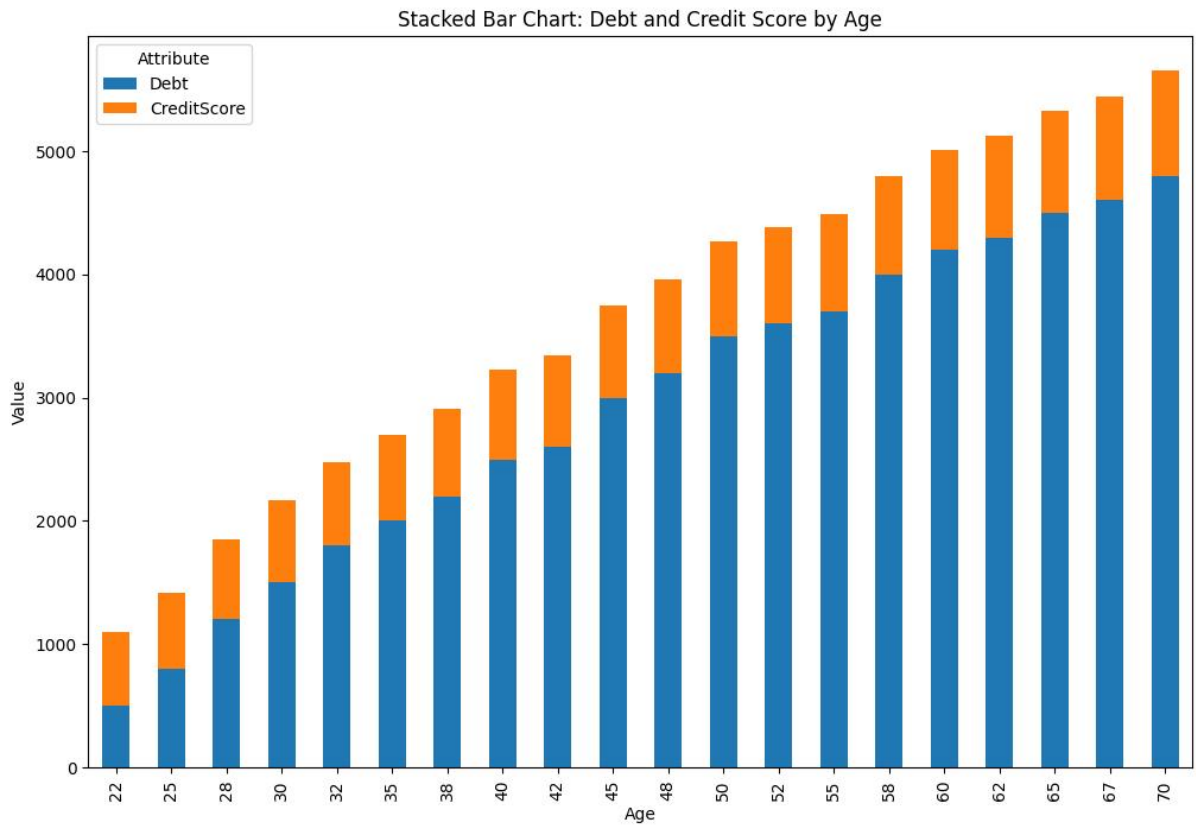plt.legend(title='Attribute')

plt.show()

Output:

Sample Dataset:

Age	Debt	YearsEmployed	CreditScore	Approved
22	500	1	600	1
25	800	2	620	0
28	1200	3	650	1
30	1500	4	670	0
32	1800	6	680	1

Line Graph: Age, Debt, and Credit Score

Stacked Bar Chart: Debt and Credit Score by Age

Result:

The visualizations display clear patterns — as Age increases, both Debt and Credit Score tend to rise. The stacked bar chart and line plots show a positive relationship among variables, while the parallel coordinates highlight how approval varies with financial attributes.