Data Analysis Project - Using Python and Pandas

Name – Sarita Roy

Sap id – 590018406

Batch - 5

Introduction

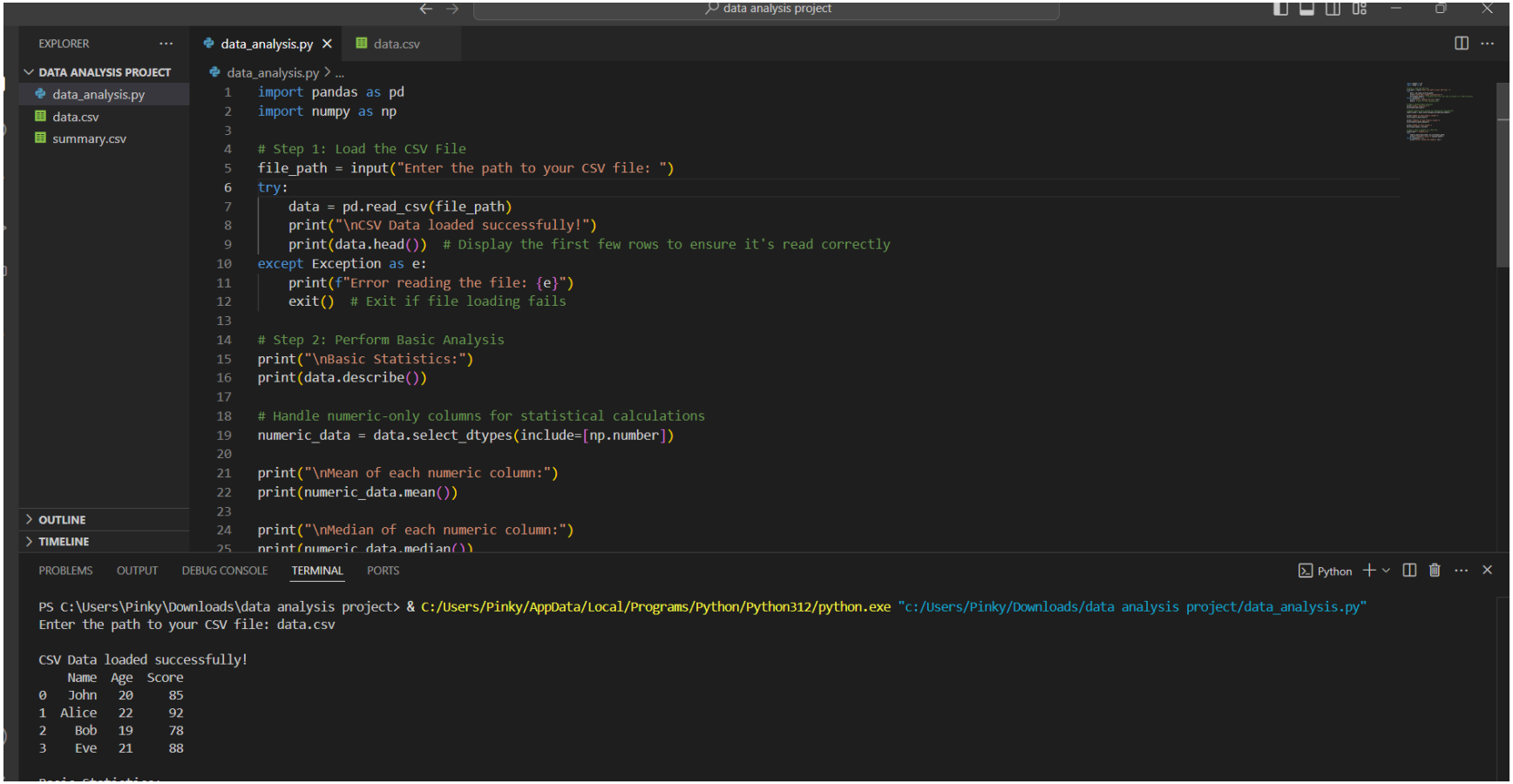
- This project demonstrates data analysis on a simple CSV dataset.
- Dataset: Name, Age, Score
- Tools: Python, pandas, numpy
- Goal: Calculate basic statistics (mean, median, mode) and save a summary.

Dataset

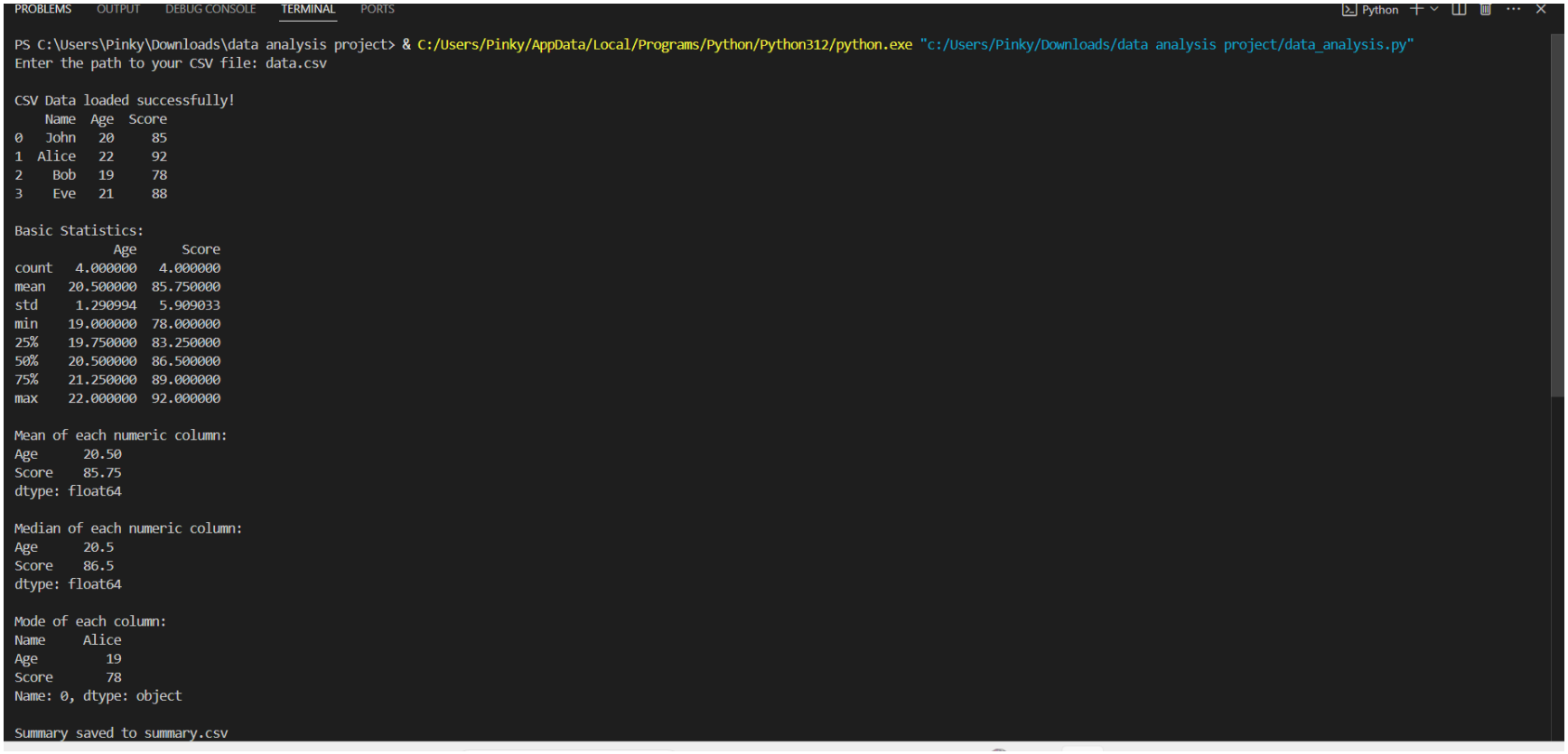
- Sample Dataset (example.csv):

- Name, Age, Score
- John, 20, 85
- Alice, 22, 92
- Bob, 19, 78
- Eve, 21, 88

Python Code


data analysis project
EXPLORER
data_analysis.py
data.csv
DATA ANALYSIS PROJECT
data_analysis.py
data.csv
summary.csv
OUTLINE
TIMELINE
import pandas as pd
import numpy as np
# Step 1: Load the CSV File
file_path = input("Enter the path to your CSV file: ")
try:
data = pd.read_csv(file_path)
print("\nCSV Data loaded successfully!")
print(data.head()) # Display the first few rows to ensure it's read correctly
except Exception as e:
print(f"Error reading the file: {e}")
exit() # Exit if file loading fails
# Step 2: Perform Basic Analysis
print("\nBasic Statistics:")
print(data.describe())
# Handle numeric-only columns for statistical calculations
numeric_data = data.select_dtypes(include=[np.number])
print("\nMean of each numeric column:")
print(numeric_data.mean())
print("\nMedian of each numeric column:")
PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
Python
PS C:\Users\Pinky\Downloads\data analysis project> & C:/Users/Pinky/AppData/Local/Programs/Python/Python312/python.exe "c:/Users/Pinky/Downloads/data analysis project/data_analysis.py"
Enter the path to your CSV file: data.csv
CSV Data loaded successfully!
Name Age Score
0 John 20 85
1 Alice 22 92
2 Bob 19 78
3 Eve 21 88

OUTPUT


PROBLEMS
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
Python
PS C:\Users\Pinky\Downloads\data analysis project> & C:/Users/Pinky/AppData/Local/Programs/Python/Python312/python.exe "c:/Users/Pinky/Downloads/data analysis project/data_analysis.py"
Enter the path to your CSV file: data.csv
CSV Data loaded successfully!
Name Age Score
0 John 20 85
1 Alice 22 92
2 Bob 19 78
3 Eve 21 88
Basic Statistics:
Age Score
count 4.000000 4.000000
mean 20.500000 85.750000
std 1.290994 5.909033
min 19.000000 78.000000
25% 19.750000 83.250000
50% 20.500000 86.500000
75% 21.250000 89.000000
max 22.000000 92.000000
Mean of each numeric column:
Age 20.50
Score 85.75
dtype: float64
Median of each numeric column:
Age 20.5
Score 86.5
dtype: float64
Mode of each column:
Name Alice
Age 19
Score 78
Name: 0, dtype: object
Summary saved to summary.csv

Results

- Descriptive Statistics:
- - Mean: Age=20.5, Score=85.75
- - Median: Age=20.5, Score=86.5
- - Mode: Name=Alice, Age=19, Score=78

- Summary saved to 'summary.csv'.

Visualization

- Using matplotlib:

- import matplotlib.pyplot as plt

- plt.bar(data['Name'], data['Score'], color='skyblue')
- plt.title('Scores by Name')
- plt.xlabel('Name')
- plt.ylabel('Score')
- plt.show()

Conclusion

- This project demonstrates:
- - How to load and analyze a CSV dataset using Python.
- - Calculating statistics: mean, median, mode.
- - Saving analysis results to a CSV file.

- Further steps could include advanced visualizations or more complex analysis.