**Title: Basic data analysis project: slab 1**

**Question:** Using the Pandas library, load a CSV file and perform basic data analysis tasks, such as calculating the average of a selected column. Additionally, use Matplotlib to create visualizations, including bar charts, scatter plots, and heatmaps, to analyze the data. Provide insights and observations based on the analysis and visualizations.

Csv file:



**Code:**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from datetime import datetime

# 1. Load CSV file

df = pd.read\_csv('data.csv.csv')

print("First 5 rows of the dataframe:")

print(df.head())

# 2. Calculate 'Age' Column based on 'Date of birth'

def calculate\_age(birthdate\_str):

    # birthdate\_str format should be 'dd-mm-yyyy'

    birthdate = datetime.strptime(birthdate\_str, "%d-%m-%Y")

    today = datetime(2025, 8, 18)  # Use current date for age calculation

    age = today.year - birthdate.year - ((today.month, today.day) < (birthdate.month, birthdate.day))

    return age

df['Age'] = df['Date of birth'].apply(calculate\_age)

average\_age = df['Age'].mean()

print("\nAverage age in the data:", average\_age)

# 3. Show summary statistics for Age

print("\nSummary statistics for Age:")

print(df['Age'].describe())

# 4. Bar chart: Count by Sex

plt.figure(figsize=(6,4))

df['Sex'].value\_counts().plot(kind='bar')

plt.title('Count by Sex')

plt.xlabel('Sex')

plt.ylabel('Count')

plt.tight\_layout()

plt.savefig('bar\_chart\_sex.png')

plt.show()

# 5. Scatter plot: Age vs Index

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

plt.scatter(df['Index'], df['Age'], alpha=0.7)

plt.title('Age vs Index')

plt.xlabel('Index')

plt.ylabel('Age')

plt.tight\_layout()

plt.savefig('scatter\_age\_index.png')

plt.show()

# 6. Heatmap: Correlation matrix (for numerical columns)

plt.figure(figsize=(6,5))

corr = df[['Index', 'Age']].corr()

sns.heatmap(corr, annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap')

plt.tight\_layout()

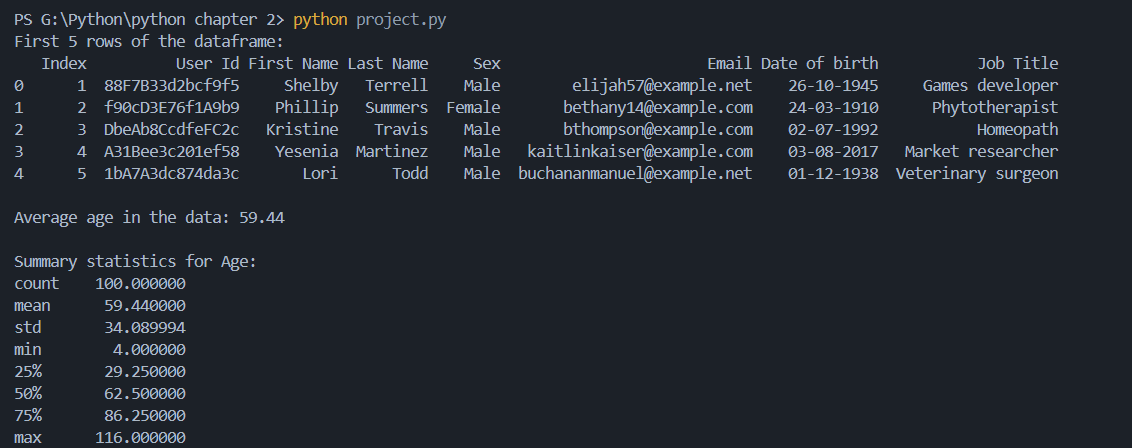
plt.savefig('heatmap\_corr.png')

plt.show()

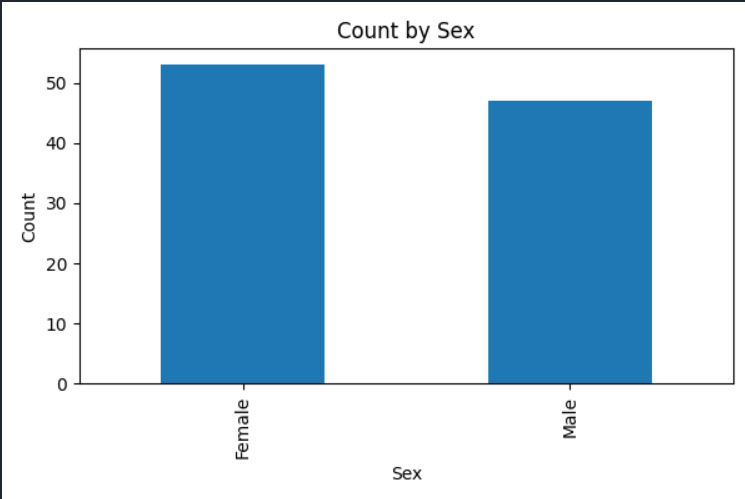
# 7. Save new CSV with the Age column

df.to\_csv('data\_with\_age.csv', index=False)

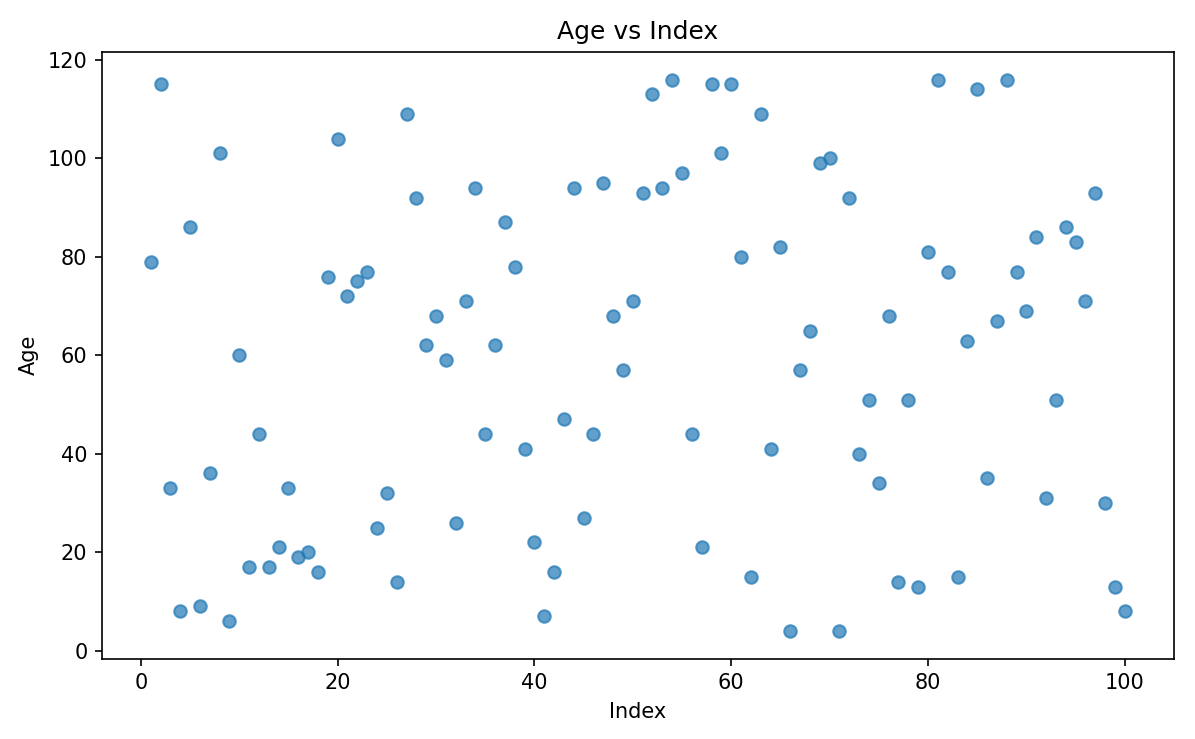
print("\nNew CSV with 'Age' column saved as 'data\_with\_age.csv'")

**output:**

**(*Figure no:1)***

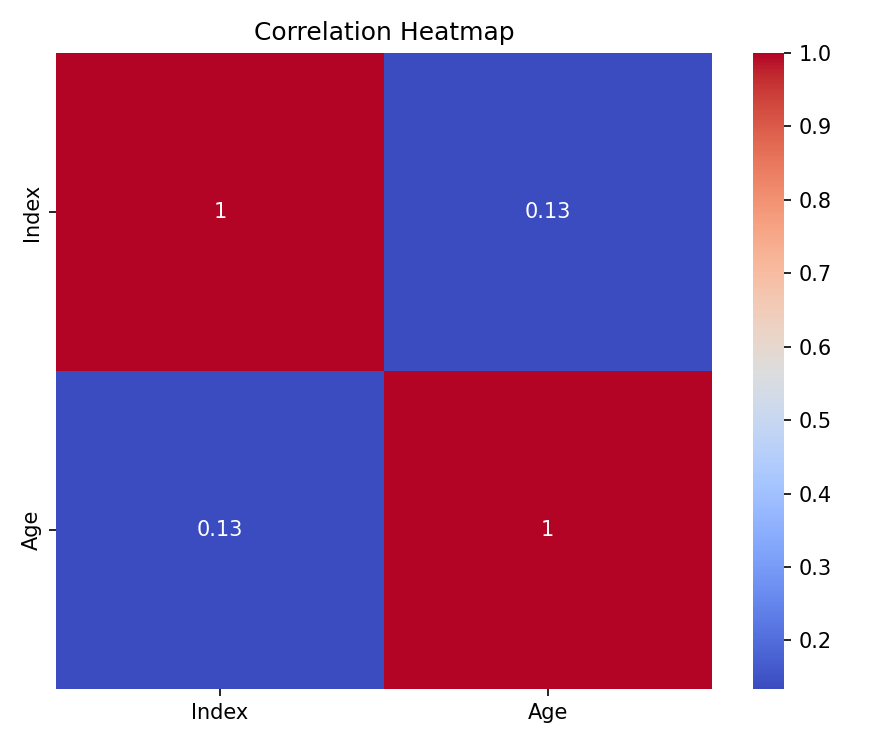
**bar chart:**

***(figure no:2)***

**Scatter plot:**

***(figure no:2)***

**Heatmap:**

****

***(figure no:3)***

**conclusion:**   
During this project, I successfully applied Python libraries such as Pandas and Matplotlib to analyze and visualize real-world data. By loading a CSV file, I explored the dataset and performed basic operations like calculating averages and generating descriptive statistics. I also created bar charts, scatter plots, and heatmaps to identify patterns and correlations.

This project not only enhanced my technical skills in data handling and visualization but also gave me practical exposure to how data analysis works in real scenarios.