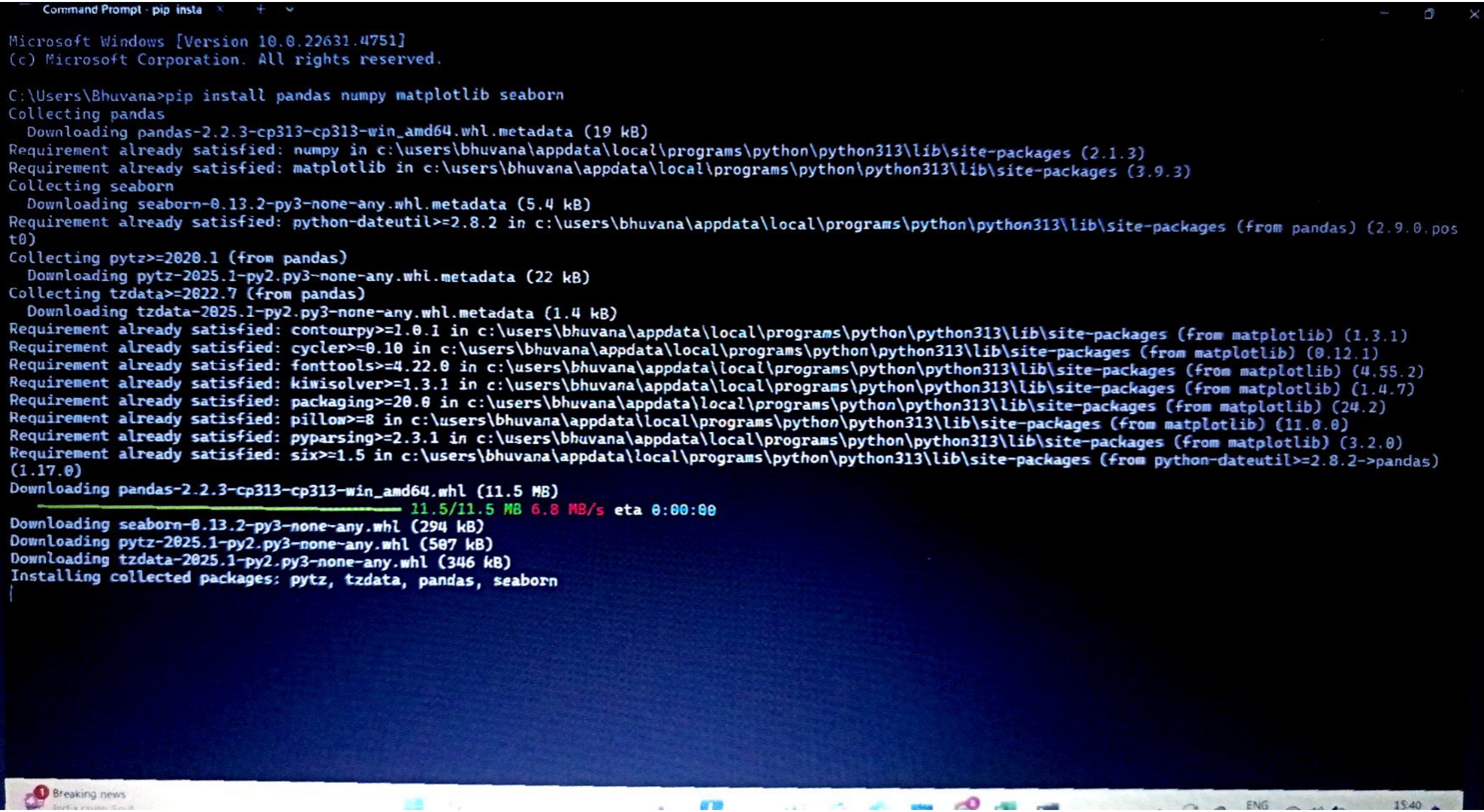
TASK 1

Working With Data

Manipulate and Analyze data using Python libraries


Command Prompt - pip insta
Microsoft Windows [Version 10.0.22631.4751]
(c) Microsoft Corporation. All rights reserved.
C:\Users\Bhuvana>pip install pandas numpy matplotlib seaborn
Collecting pandas
Downloading pandas-2.2.3-cp313-cp313-win_amd64.whl.metadata (19 kB)
Requirement already satisfied: numpy in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (2.1.3)
Requirement already satisfied: matplotlib in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (3.9.3)
Collecting seaborn
Downloading seaborn-0.13.2-py3-none-any.whl.metadata (5.4 kB)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from pandas) (2.9.0.post0)
Collecting pytz>=2020.1 (from pandas)
Downloading pytz-2025.1-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)
Downloading tzdata-2025.1-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (1.3.1)
Requirement already satisfied: cycler>=0.10 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (4.55.2)
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (1.4.7)
Requirement already satisfied: packaging>=20.0 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (24.2)
Requirement already satisfied: pillow>=8 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (11.0.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from matplotlib) (3.2.0)
Requirement already satisfied: six>=1.5 in c:\users\bhuvana\appdata\local\programs\python\python313\lib\site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
Downloading pandas-2.2.3-cp313-cp313-win_amd64.whl (11.5 MB)
11.5/11.5 MB 6.8 MB/s eta 0:00:00
Downloading seaborn-0.13.2-py3-none-any.whl (294 kB)
Downloading pytz-2025.1-py2.py3-none-any.whl (507 kB)
Downloading tzdata-2025.1-py2.py3-none-any.whl (346 kB)
Installing collected packages: pytz, tzdata, pandas, seaborn
Breaking news
ENG
15:40

IDLE Shell 3.13.0

File Edit Shell Debug Options Window Help

```
Python 3.13.0 (tags/v3.13.0:60403a5, Oct  7 2024, 09:38:07) [MSC v.1941 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> import pandas as pd
>>> df=pd.read_csv('C:/Users/Bhuvana/OneDrive/Desktop/Demo1.csv')
>>> print(df.head())
    Id     Names   Age
0  111    Baskar  26.0
1  109  Sandhiya   NaN
2  105   Sumathi  46.0
3  201    Harshi   1.0
4  110     Krish  11.0
>>> print("\nDatasets Info:")

Datasets Info:
>>> print(df.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 3 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Id      6 non-null      int64
 1   Names   6 non-null      object
 2   Age     5 non-null      float64
dtypes: float64(1), int64(1), object(1)
memory usage: 276.0+ bytes
None
>>> print("\nSummary Statistics:")

Summary Statistics:
>>> print(df.describe())
              Id        Age
count    6.000000   5.000000
mean   124.333333  20.600000
std     37.617372  16.979399
min    105.000000   1.000000
25%    109.250000  11.000000
50%    110.000000  19.000000
75%    110.750000  26.000000
max    201.000000  46.000000
>>> print("\n Missing Values:")

 Missing Values:
>>> print(df.isnull().sum())
Id       0
Names    0
Age      1
dtype: int64
```

Ln: 87 Col: 0

IDLE Shell 3.13.0

File Edit Shell Debug Options Window Help

```
 Missing Values:
>>> print(df.isnull().sum())
Id       0
Names    0
Age      1
dtype: int64
>>> print("\nFilter:")

Filter:
>>> data={'Names':['Baskar','Bhuvana','Krish','Harshi'],'Age':['26','19','11','1']}
>>> df=pd.DataFrame(data)
>>> df['Age']=pd.to_numeric(df['Age'],errors='coerce')
>>> filtered_df=df[df['Age']>15]
>>> print(filtered_df)
     Names  Age
0   Baskar   26
1  Bhuvana   19
>>> print("\nSort:")

Sort:
>>> df_sorted=df.sort_values(by='Age',ascending=True)
>>> print(df_sorted)
     Names  Age
3   Harshi    1
2    Krish   11
1  Bhuvana   19
0   Baskar   26
>>> df_sorted=df.sort_values(by='Age',ascending=False)
>>> print(df_sorted)
     Names  Age
0   Baskar   26
1  Bhuvana   19
2    Krish   11
3   Harshi    1
>>> print("\nGroup:")

Group:
>>> grouped_data=df.groupby('Names')['Age'].mean()
>>> print(grouped_data)
Names
Baskar     26.0
Bhuvana    19.0
Harshi      1.0
Krish      11.0
Name: Age, dtype: float64
>>>
```

Ln: 40 Col: 27

79°F Mostly cloudy

Search

ENG IN

20:32 02-02-2025

IDLE Shell 3.13.0

File Edit Shell Debug Options Window Help

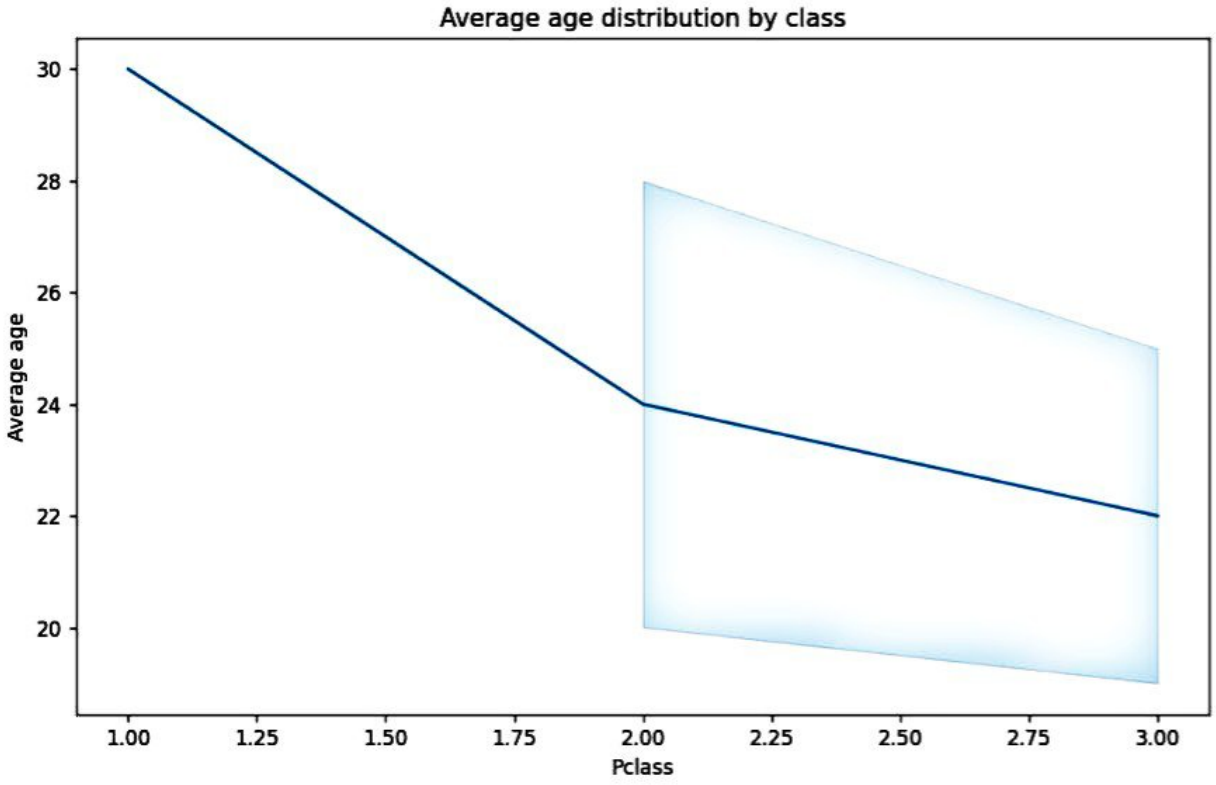
```
Python 3.13.0 (tags/v3.13.0:60403a5, Oct  7 2024, 09:38:07) [MSC v.1941 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> import  pandas as pd
>>> import numpy as np
>>> import seaborn as sns
>>> import matplotlib.pyplot as plt
>>> df=pd.read_csv(r"C:/Users/Bhuvana/OneDrive/Dokumen/titanic dataset.csv")
>>> print(df.head())
   Passenger Id  Survived  Pclass  Name  Gender  Age   Fare
0             1         0       3  Alex    male   19   7.25
1             2         1       1  John    male   30  71.28
2             3         1       3  Mary  female   25   7.92
3             4         1       2  Jack  female   28   7.88
4             5         2       2   Bhu  female   20   8.77
>>> print("\n 1.Line Chart:")

 1.Line Chart:
>>> plt.figure(figsize=(10,6))
<Figure size 1000x600 with 0 Axes>
>>> sns.lineplot(x='Pclass',y='Age',data=df,estimator='mean')
<Axes: xlabel='Pclass', ylabel='Age'>
>>> plt.title('Average age distribution by class')
Text(0.5, 1.0, 'Average age distribution by class')
>>> plt.xlabel('Pclass')
Text(0.5, 0, 'Pclass')
>>> plt.ylabel('Average age')
Text(0, 0.5, 'Average age')
>>> plt.show()
>>>
```

Ln: 29 Col: 0

IDLE Shell 3.13.0

File

Figure 1



>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>
>>>

Ln: 30 Col: 0

79°F
Mostly cloudy

Search

ENG
IN

21:30
02-02-2025

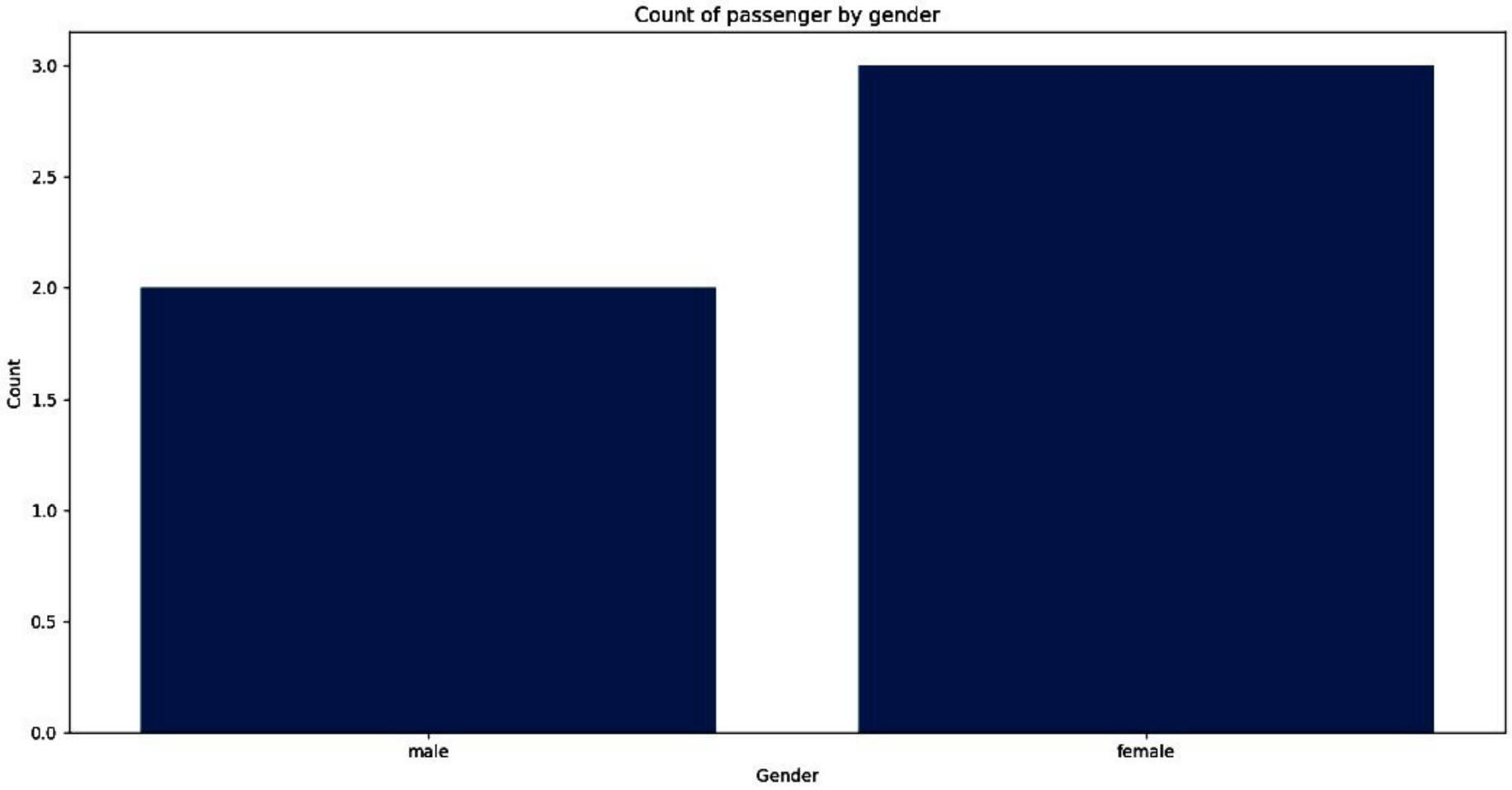
barchart.py - C:/Users/Bhuvana/AppData/Local/Programs/Python/Python313/barchart.py (3.13.0)

File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:/Users/Bhuvana/OneDrive/Dokumen/titanic dataset.csv")
print(df.head())
plt.figure(figsize=(10,6))
sns.countplot(x='Gender',data=df)
plt.title('Count of passenger by gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```

Ln: 1 Col: 0

Figure 1



Search

ENG
IN

21:48
02-02-2025

barchart.py - C:/Users/Bhuvana/AppData/Local/Programs/Python/Python313/barchart.py (3.13.0)

File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:/Users/Bhuvana/OneDrive/Dokumen/titanic dataset.csv")
print(df.head())

#Histogram
plt.figure(figsize=(10,6))
sns.histplot(df['Age'],kde=True,bins=30)
plt.title('Age of Distribution of passenger')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```

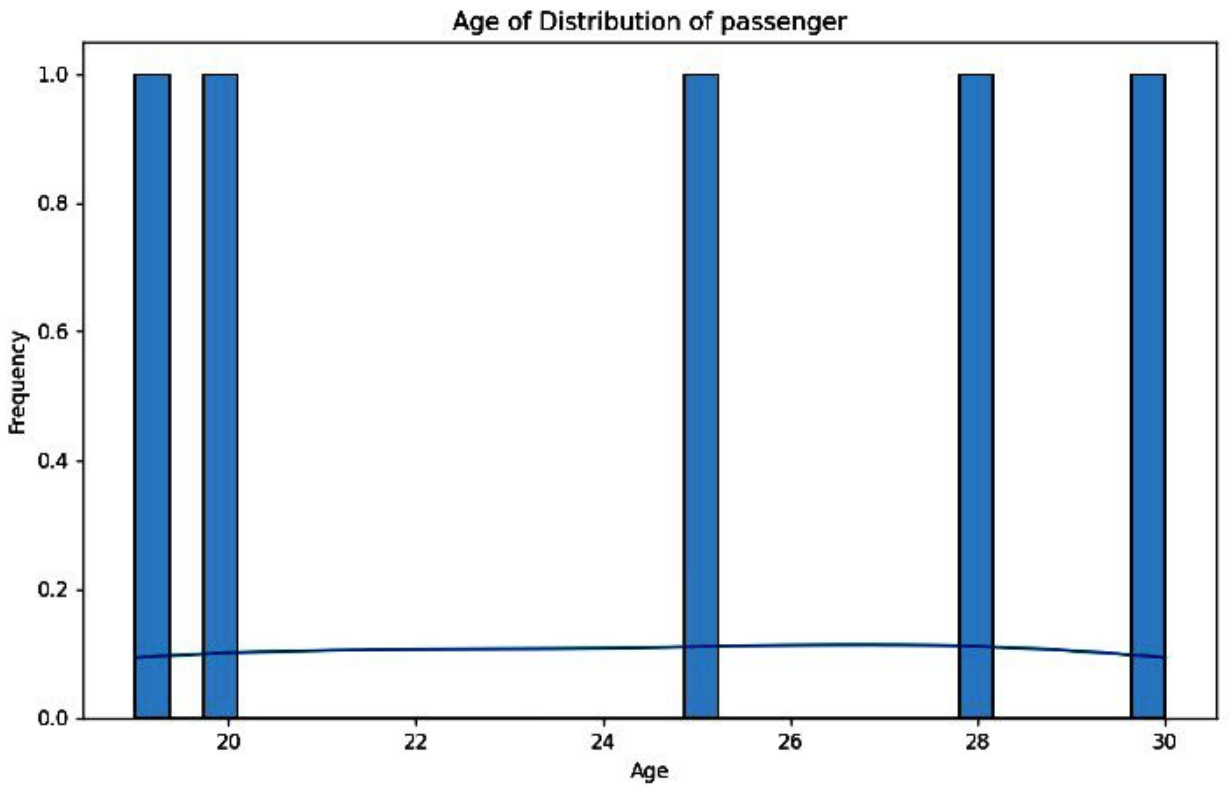
Ln: 10 Col: 15

barchart.py - C:/Users/Bhuvana/AppData/Local/Programs/Python/Python313/barchart.py (3.13.0)

File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_cs
print(df.head

#Histogram
plt.figure(fi
sns.histplot(
plt.title('Ag
plt.xlabel('A
plt.ylabel('F
plt.show()
```

Figure 1



Ln: 15 Col: 0

Search

ENG IN

21:53 02-02-2025

barchart.py - C:/Users/Bhuvana/AppData/Local/Programs/Python/Python313/barchart.py (3.13.0)

File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:/Users/Bhuvana/OneDrive/Dokumen/titanic dataset.csv")
print(df.head())

#Scatter plot
plt.figure(figsize=(10,6))
sns.scatterplot(x='Age',y='Fare',data=df,hue='Pclass',palette='Set1')
plt.title('Scatter plot of age vs fare')
plt.xlabel('Age')
plt.ylabel('Fare')
plt.show()
```

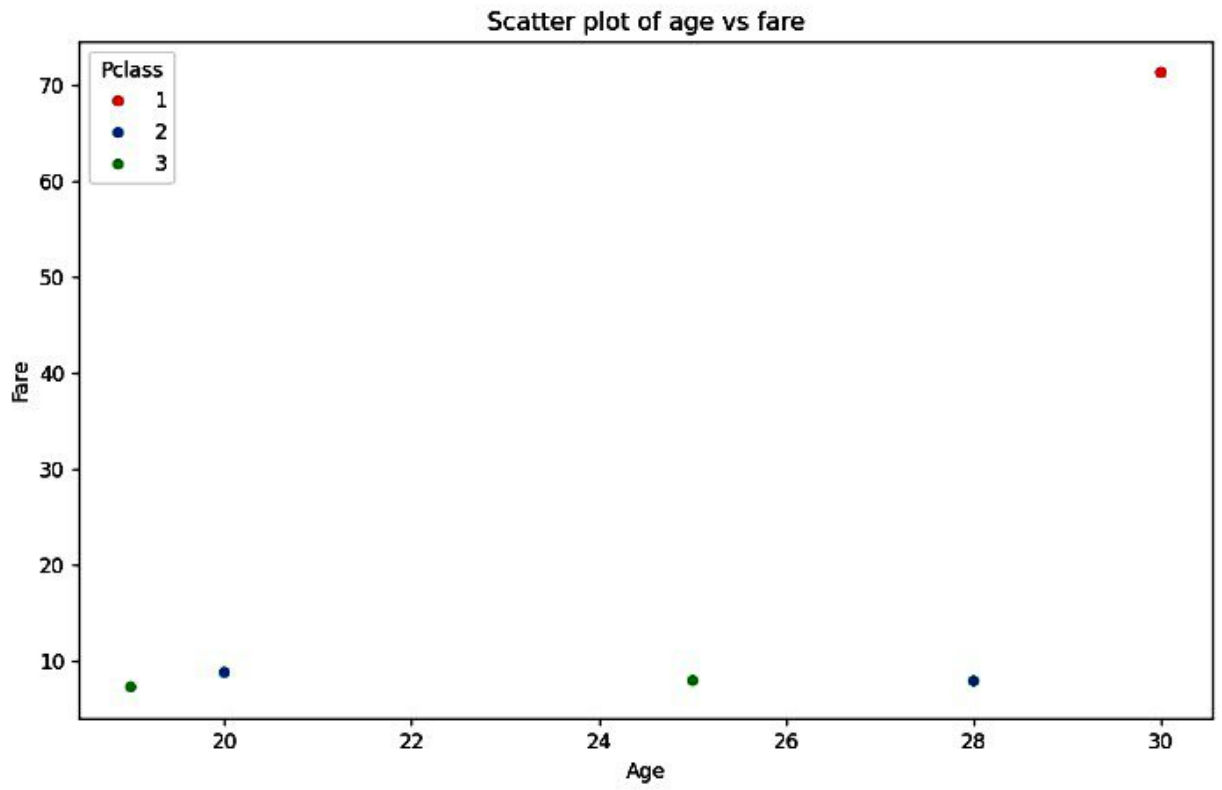
Ln: 10 Col: 47

barchart.py - C:/Users/Bhuvana/AppData/Local/Programs/Python/Python313/barchart.py (3.13.0)

File Edit Format Run Options Window Help

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv(r"C:/Users/Bhuvana/OneDrive/Dokumen/titanic dataset.csv")
print(df.head())

#Scatter plot
plt.figure(figsize=
sns.scatterplot(x='
plt.title('Scatter
plt.xlabel('Age')
plt.ylabel('Fare')
plt.show()
```

Figure 1



Ln: 15 Col: 0

Search

ENG IN

22:01 02-02-2025