

PRACTICAL NO 6

Aim: Implementation to perform data mining using WEKA and Python separately

Part A: Data mining using Python

Step 1: Get the dataset about the titanic passengers in the excel file and name it as titanic.csv

	A	B	C	D	E	F	G	H	I
1	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
2	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171
3	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599
4	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2.
5	4	1	1	Futelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803
6	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450
7	6	0	3	Moran, Mr. James	male		0	0	330877
8	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463

Step 2: Perform the data mining

1. Find the number of people survived above the age 20

Code:

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

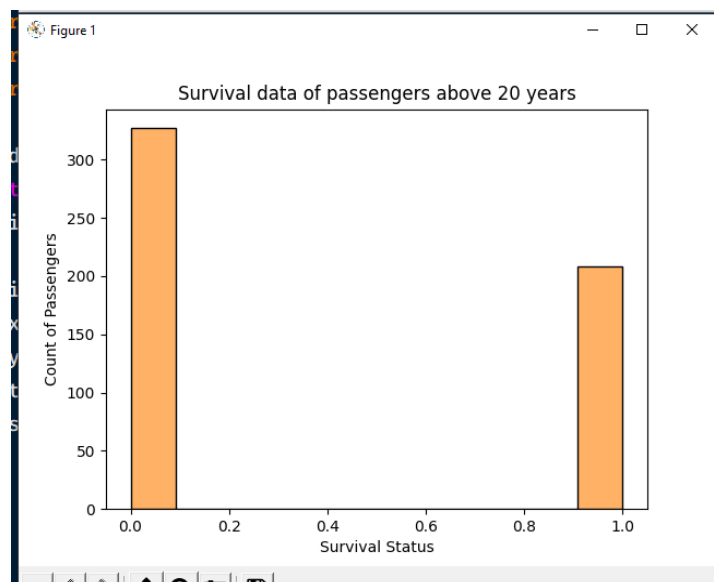
```
import seaborn as sns
```

```
df = pd.read_csv("D:\\TYCS\\DWDM\\titanic  
(1).csv")
```

```
survived_above_20 = df[df['Age'] >  
20]['Survived']
```

```
sns.histplot(data=survived_above_20,  
color='#ff9933')  
plt.xlabel('Survival Status')  
plt.ylabel('Count of Passenger')  
plt.title('Survival data of passengers above  
20 years')  
plt.show()
```

Output:



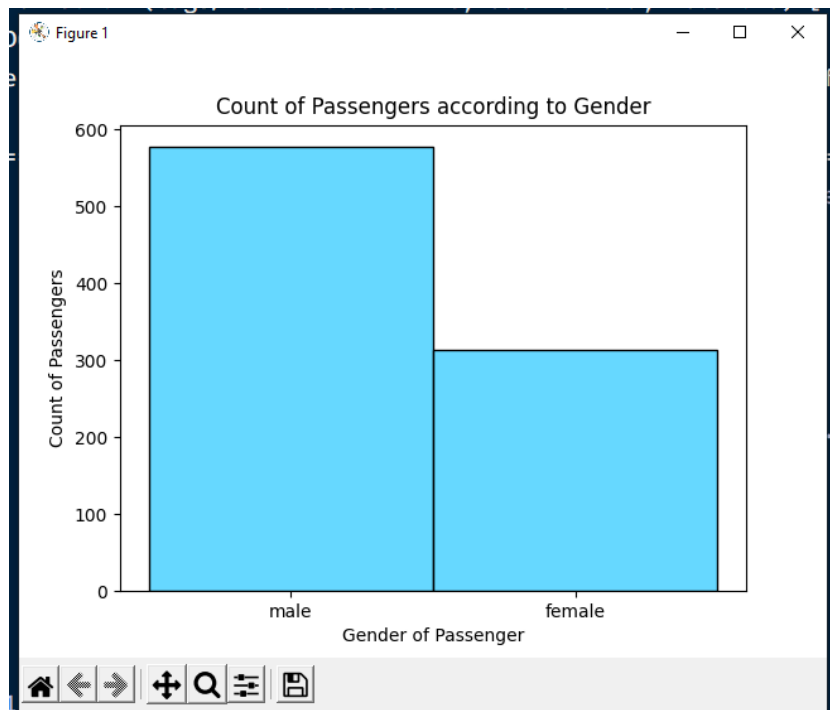
2. Visualize the distribution of gender using a histogram

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data =
pd.read_csv("D:\\TYCS\\DWDM\\titanic
(1).csv")
```

```
sns.histplot(data=data, x='Sex',
color='#33ccff')
plt.xlabel('Gender of Passenger')
plt.ylabel('Count of Passenger')
plt.title('Count of Passenger according to
gender')
plt.show()
```

Output:



3. Show the count of passengers per class (1,2,3 etc)

Code:

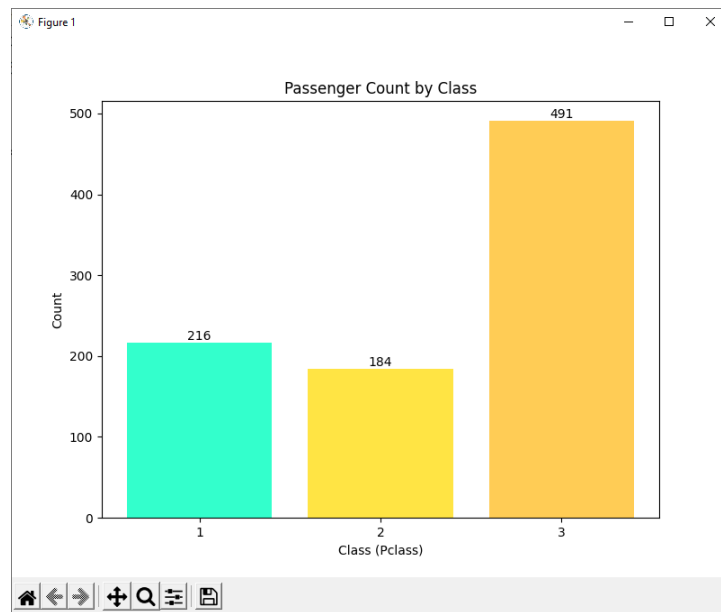
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data =
pd.read_csv("D:\\TYCS\\DWDM\\titanic
(1).csv")

class_counts =
data['Pclass'].value_counts().sort_index()
plt.figure(figsize=(8, 6))
```

```
plt.bar(class_counts.index,
class_counts.values, color=['skyblue',
'orange', 'teal'])
plt.title('Passenger Count by Class')
plt.xlabel('Class (Pclass)')
plt.ylabel('Count')

for i, count in enumerate(class_counts):
    plt.text(i + 1, count, str(count),
    ha='center', va='bottom')
plt.xticks(class_counts.index)
plt.show()
```

Output:



4. Show the Scatter plot of age Vs survived

Code:

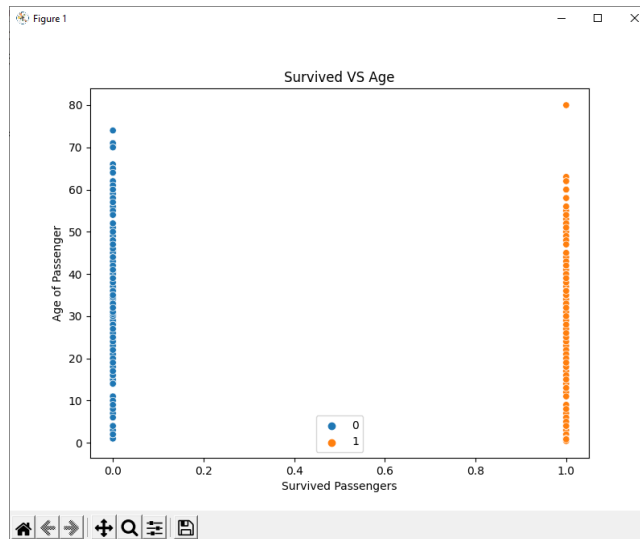
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
data =
pd.read_csv("D:\\TYCS\\DWDM\\titanic
(1).csv")
plt.figure(figsize=(8, 6))
```

```
sns.scatterplot(x='Survived', y='Age',  
hue='Survived', data=data)  
plt.title("Survived VS Age")  
plt.xlabel("Survived Passengers")
```

```
plt.ylabel("Age of Passenger")  
plt.legend()  
plt.show()
```

Output:



Part B: Data mining using WEKA