

Q1. Download the Titanic dataset and perform the Exploratory data analysis using pandas. Read the dataset (df= pd.read_csv(r'.....\Titanic.csv')

(i) Display the first and last 10 instances from the dataset

(ii) Acquire the necessary information using the df.info() and df. Describe().

(iii) Retrieve the number of columns and rows. (using shape)

```
In [ ]: import pandas as pd
df= pd.read_csv('Titanic.csv')
df
```

Out[]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599 7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803 5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536 1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053 3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607 2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369 3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376

891 rows × 12 columns



In []:

```
# (i)Display the first and last 10 instances from the datasetdf.head(10)
print("First 10 Rows")
df.head(10)
```

First 10 Rows

Out[]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	71.0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.0
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	71.0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.0
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.0
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.0
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.0
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0

In []:

```
print("Last 10 Rows")
df.tail(10)
```

Last 10 Rows

Out[]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376

In []:

```
print("Information about Dataset")
df.info()
```

```

Information about Dataset
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column        Non-Null Count  Dtype
---  -
 0   PassengerId   891 non-null    int64
 1   Survived      891 non-null    int64
 2   Pclass        891 non-null    int64
 3   Name          891 non-null    object
 4   Sex           891 non-null    object
 5   Age           714 non-null    float64
 6   SibSp         891 non-null    int64
 7   Parch         891 non-null    int64
 8   Ticket        891 non-null    object
 9   Fare          891 non-null    float64
10   Cabin         204 non-null    object
11   Embarked      889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB

```

```

In [ ]: print("Description About Dataset")
        df.describe()

```

Description About Dataset

```

Out[ ]:

```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.200000
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693298
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.912500
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.450000
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329167

```

In [ ]: print("Shape of the Dataset")
        print(df.shape)

```

Shape of the Dataset
(891, 12)

Q2. Create the data visualization using the matplotlib.

(i) Visualize the Gender of Passengers using the Bar graph.

(ii) Visualize the Survival Count of Passengers using the Bar graph.

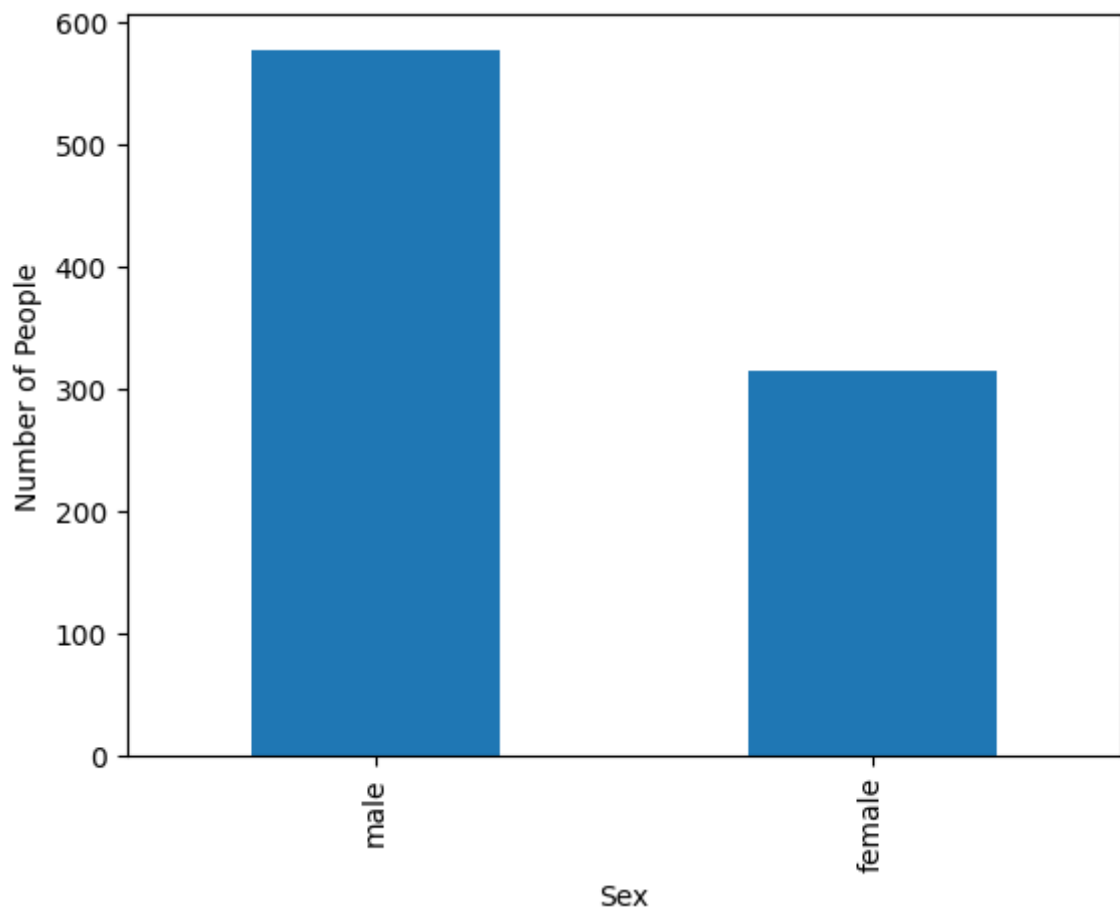
(iii) Visualize the Age of Passengers using the Bar/Histogram graph.

(iv) Visualize the comparison of Age and Fare of Passengers using the Scatterplot.

```
In [ ]: print("Visualization on Gender")
import matplotlib.pyplot as plt
df['Sex'].value_counts().plot(kind='bar',ylabel="Number of People")
```

Visualization on Gender

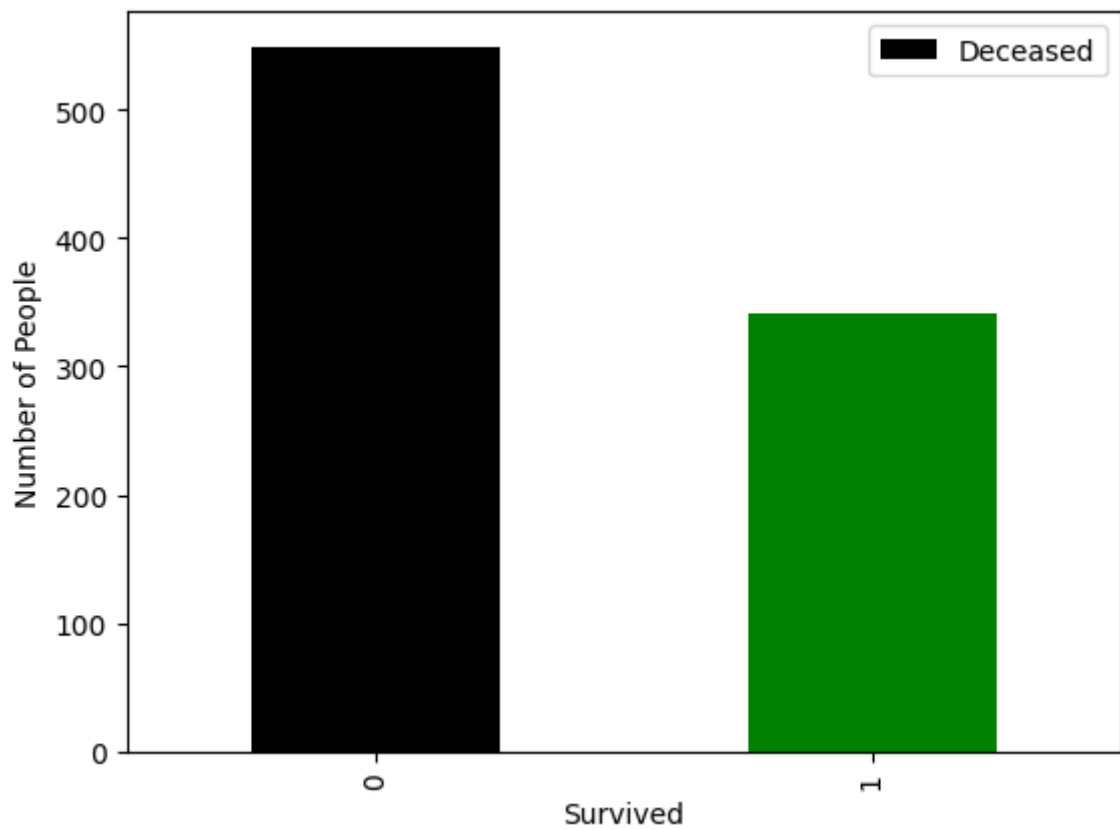
```
Out[ ]: <Axes: xlabel='Sex', ylabel='Number of People'>
```



```
In [ ]: print("Visualization on Survival")
import matplotlib.pyplot as plt
df['Survived'].value_counts().plot(kind='bar',ylabel="Number of People",color=["
```

Visualization on Survival

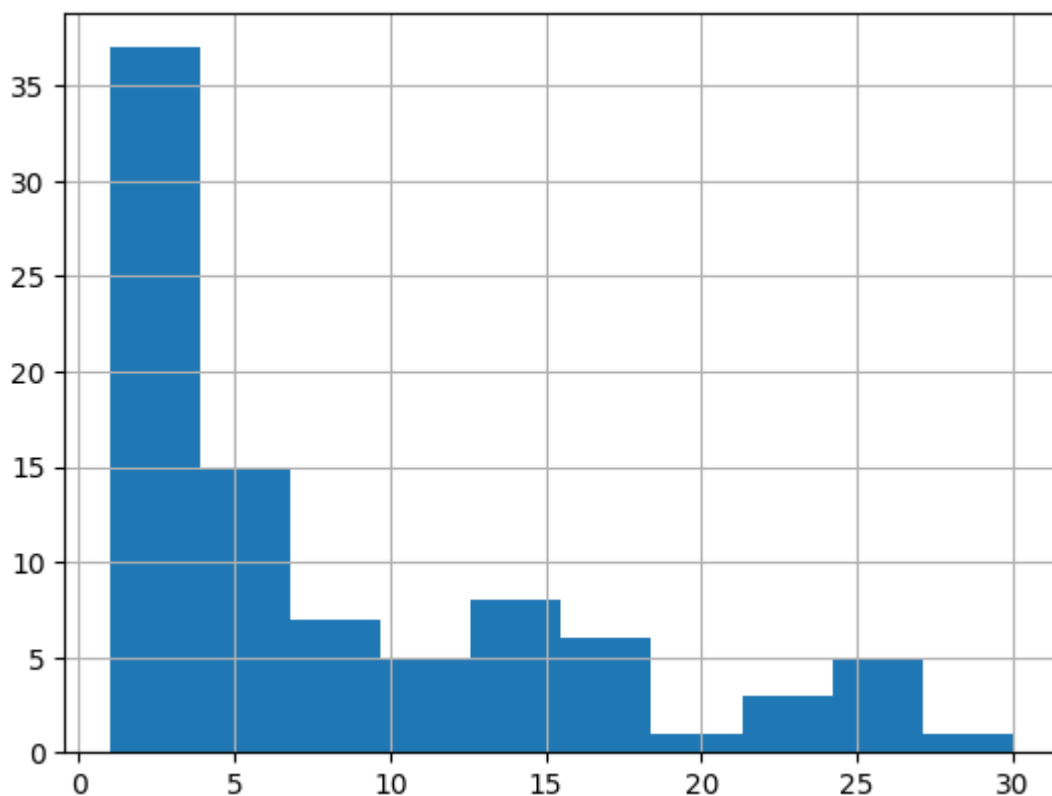
```
Out[ ]: <matplotlib.legend.Legend at 0x26253919970>
```



```
In [ ]: print("Visualization on Age")
import matplotlib.pyplot as plt
df['Age'].value_counts().hist()
```

Visualization on Age

Out[]: <Axes: >



```
In [ ]: print("Comparison Between Age and Fare")
```

```
df.plot.scatter(x='Age', y='Fare')
```

Comparison Between Age and Fare

Out[]: <Axes: xlabel='Age', ylabel='Fare'>

