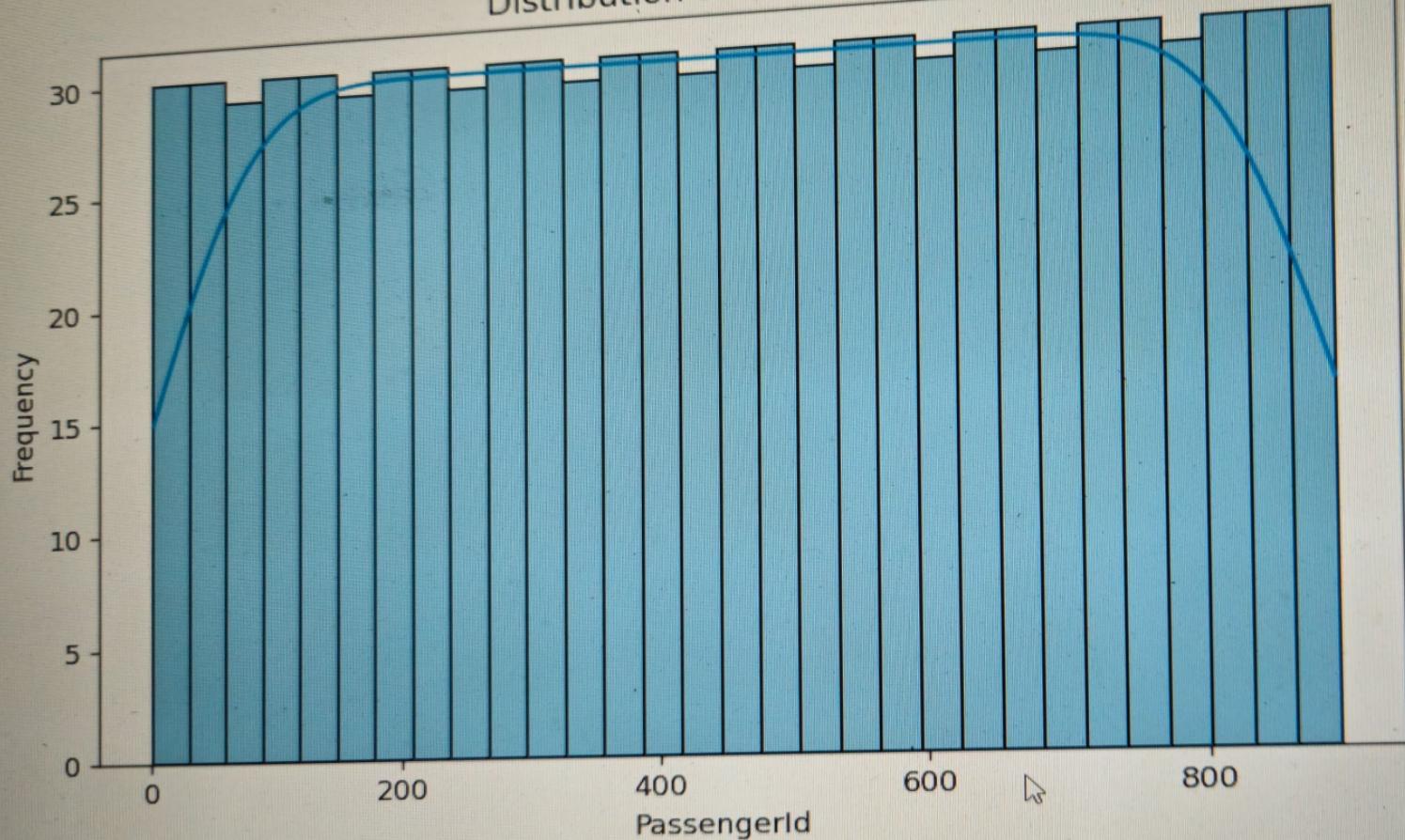
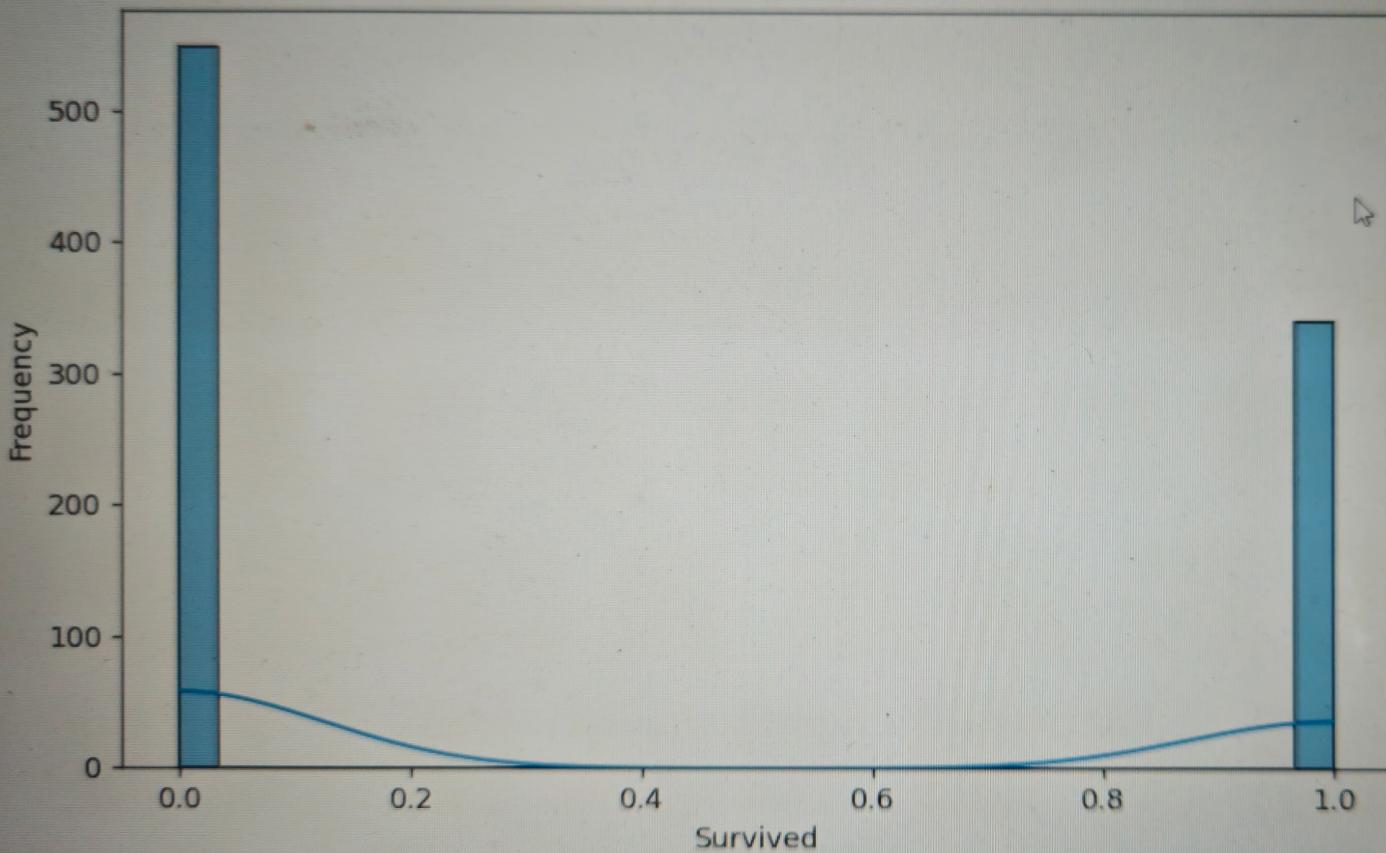


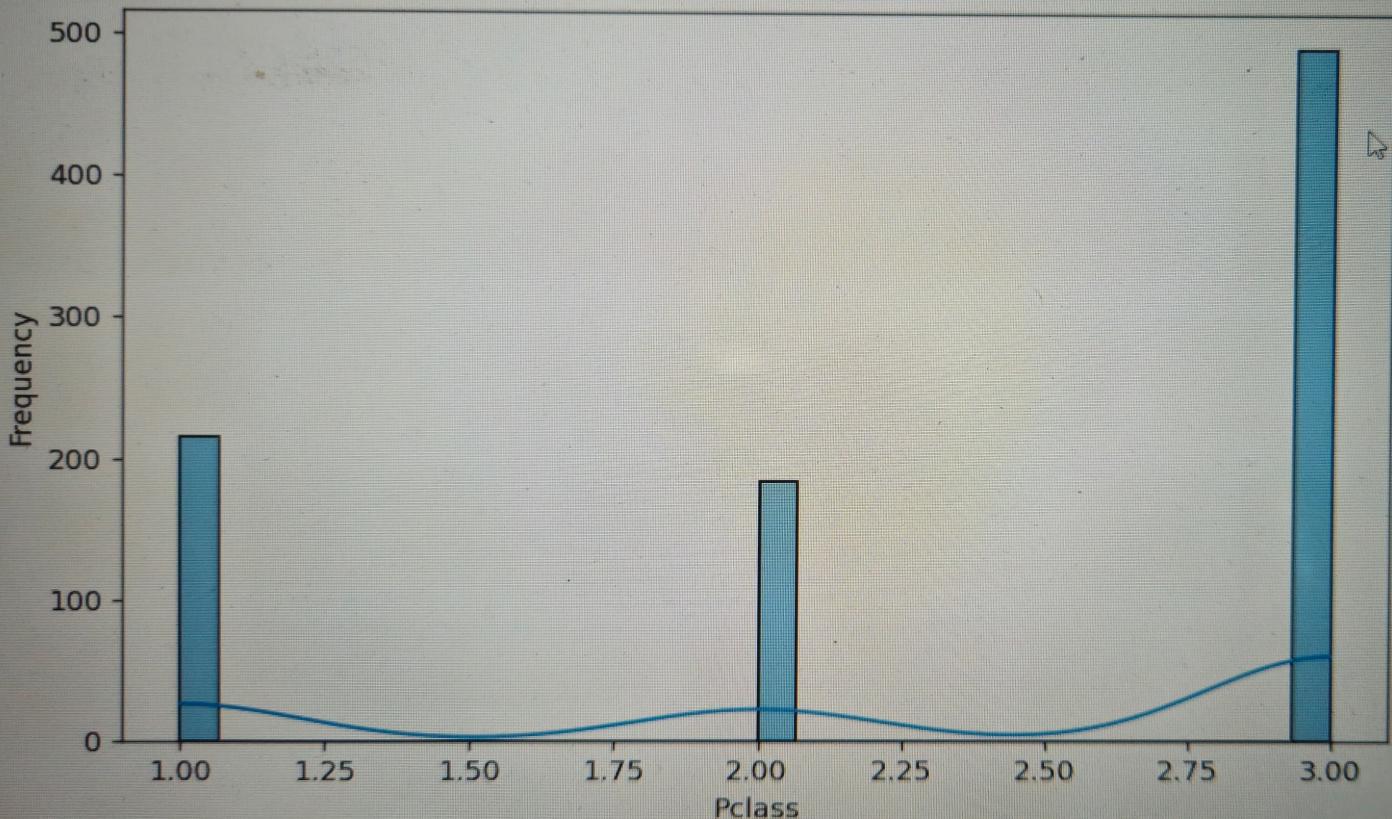
Distribution of PassengerId

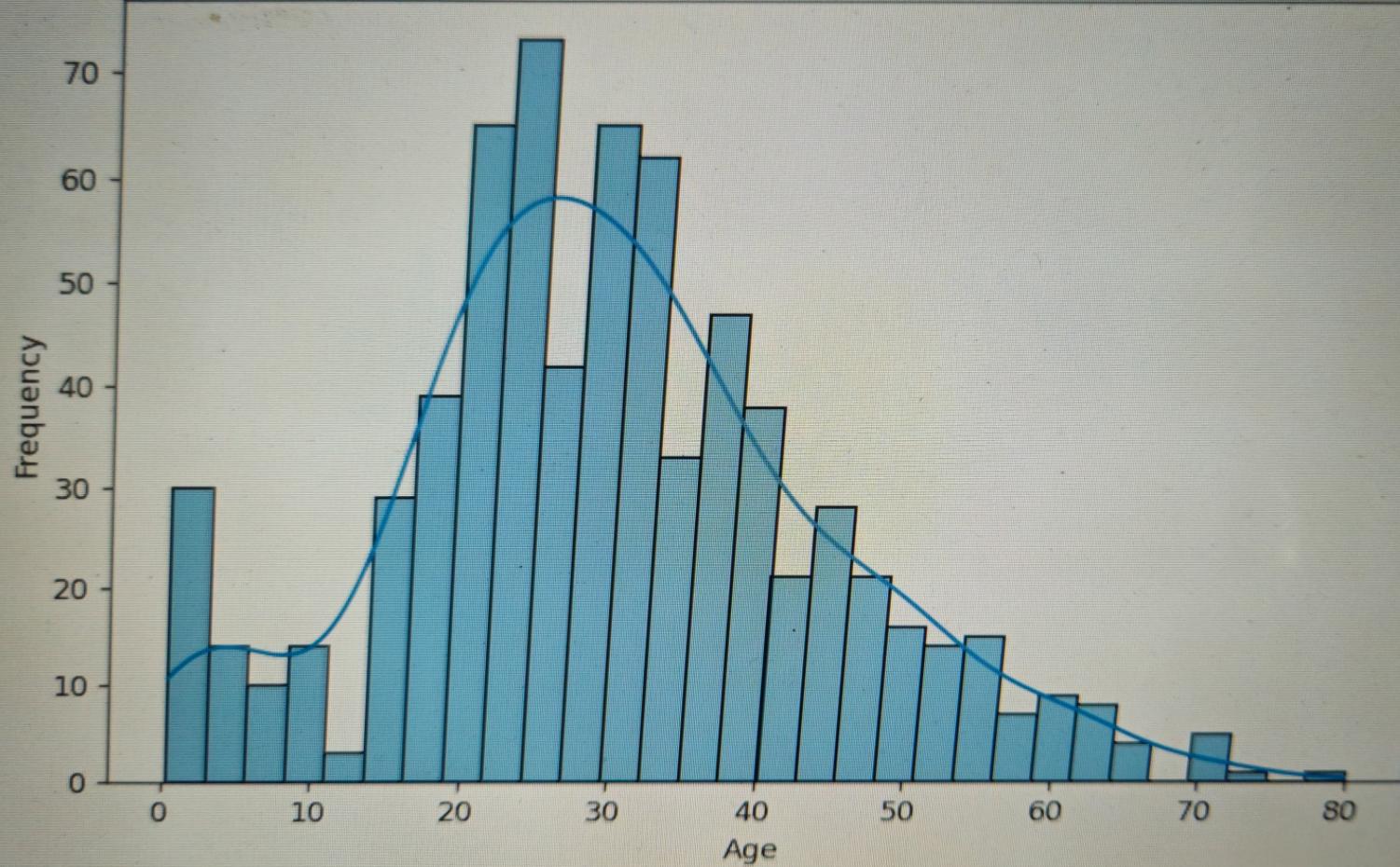


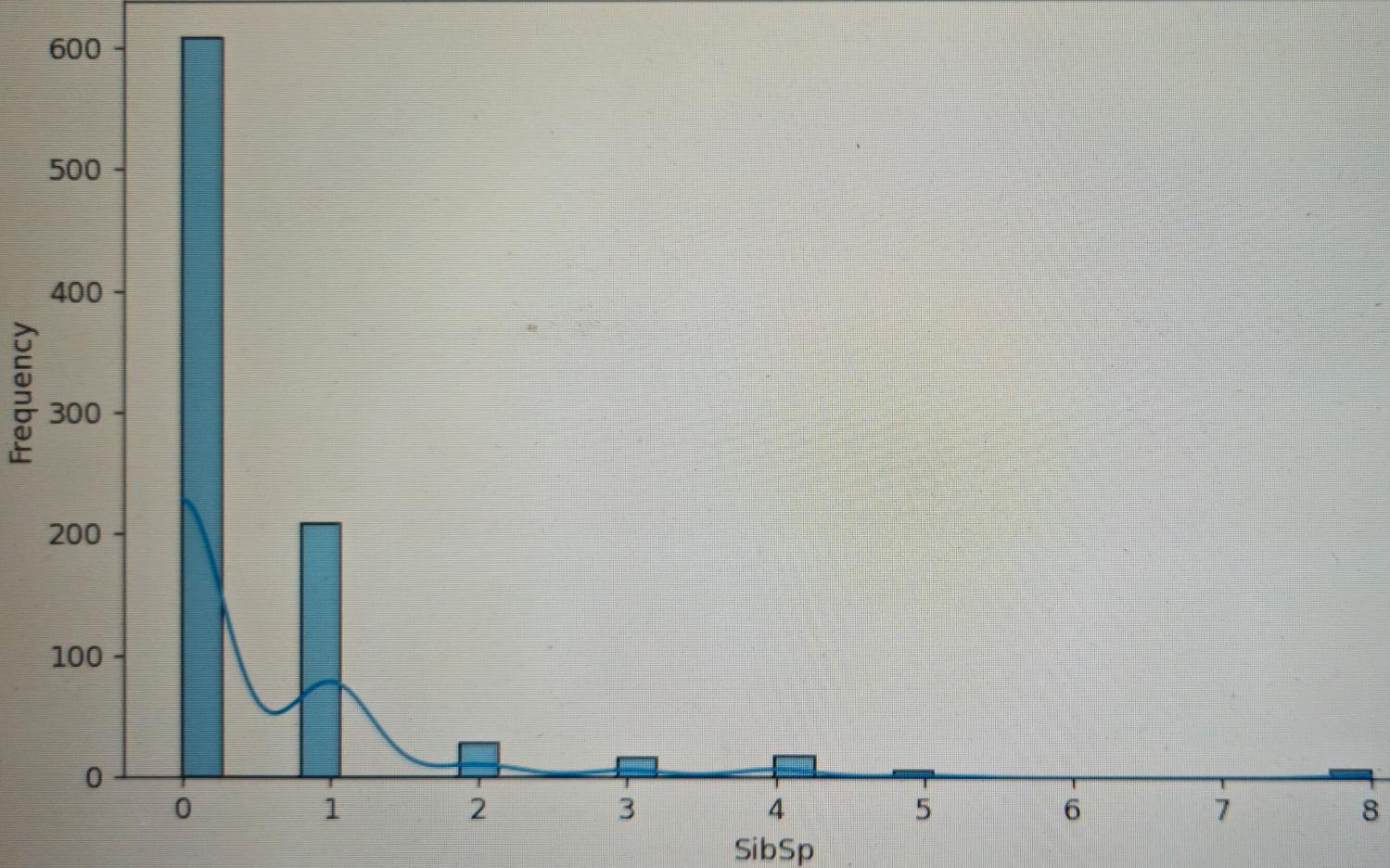
Distribution of Survived

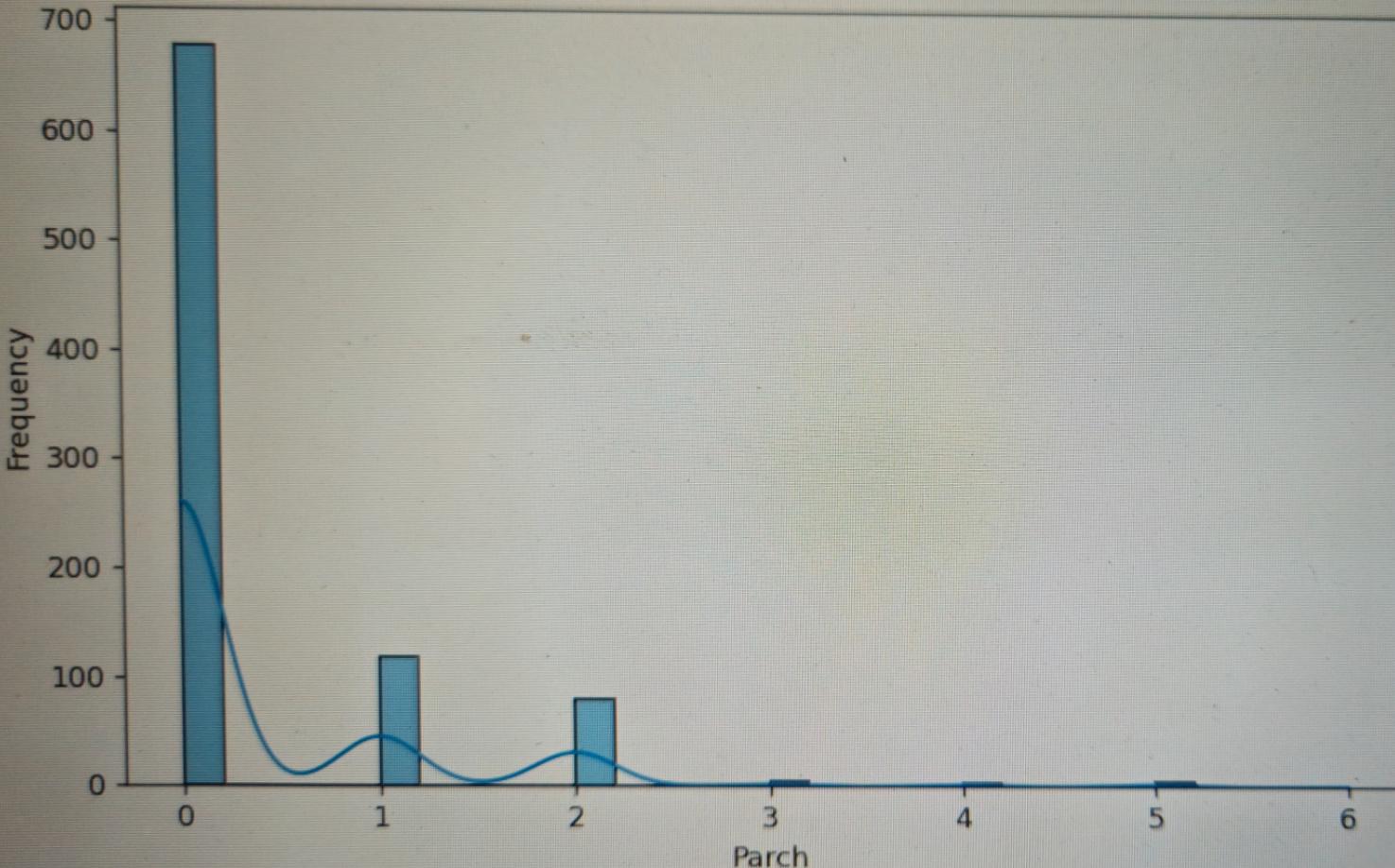


Distribution of Pclass









Distribution of Fare

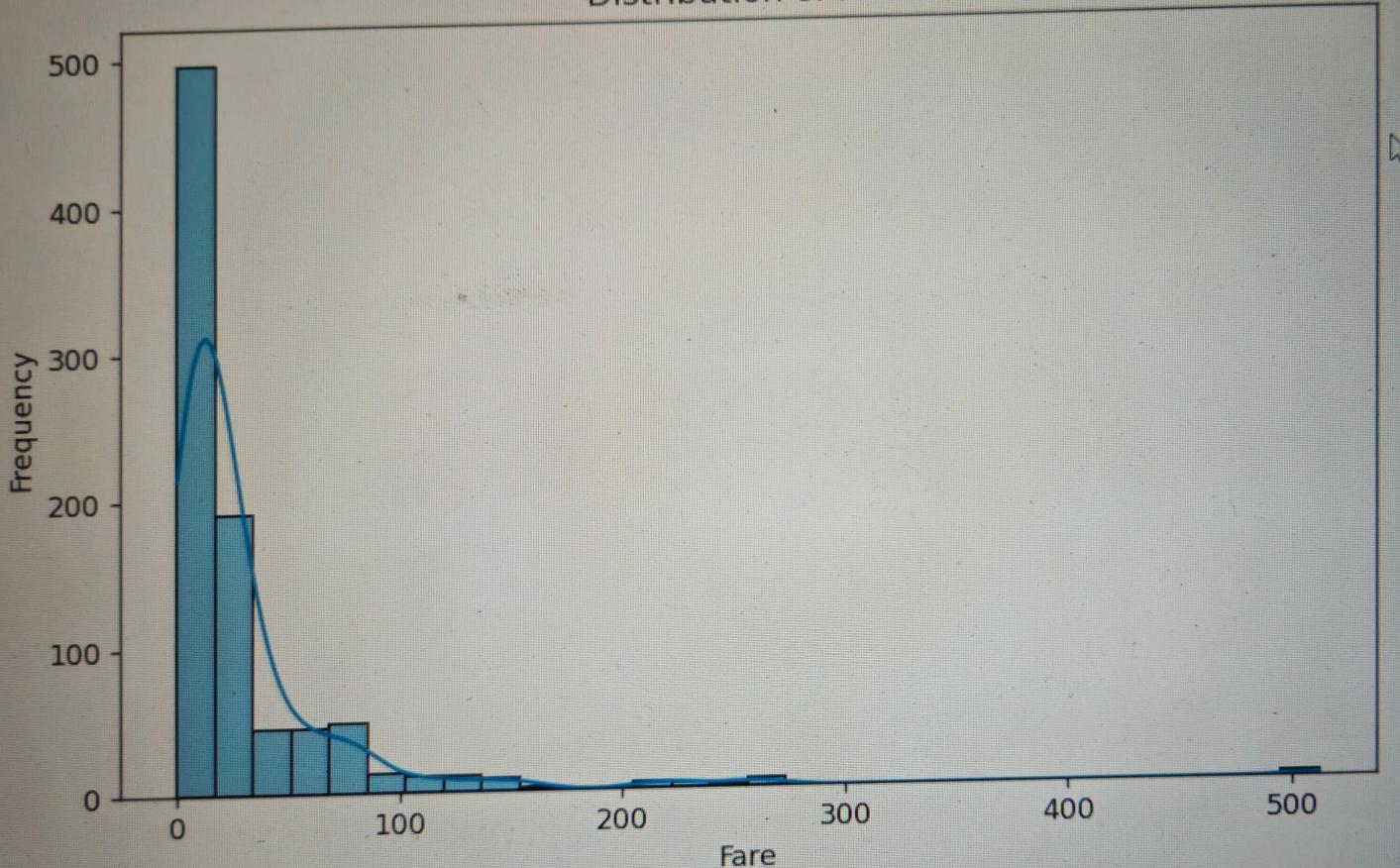
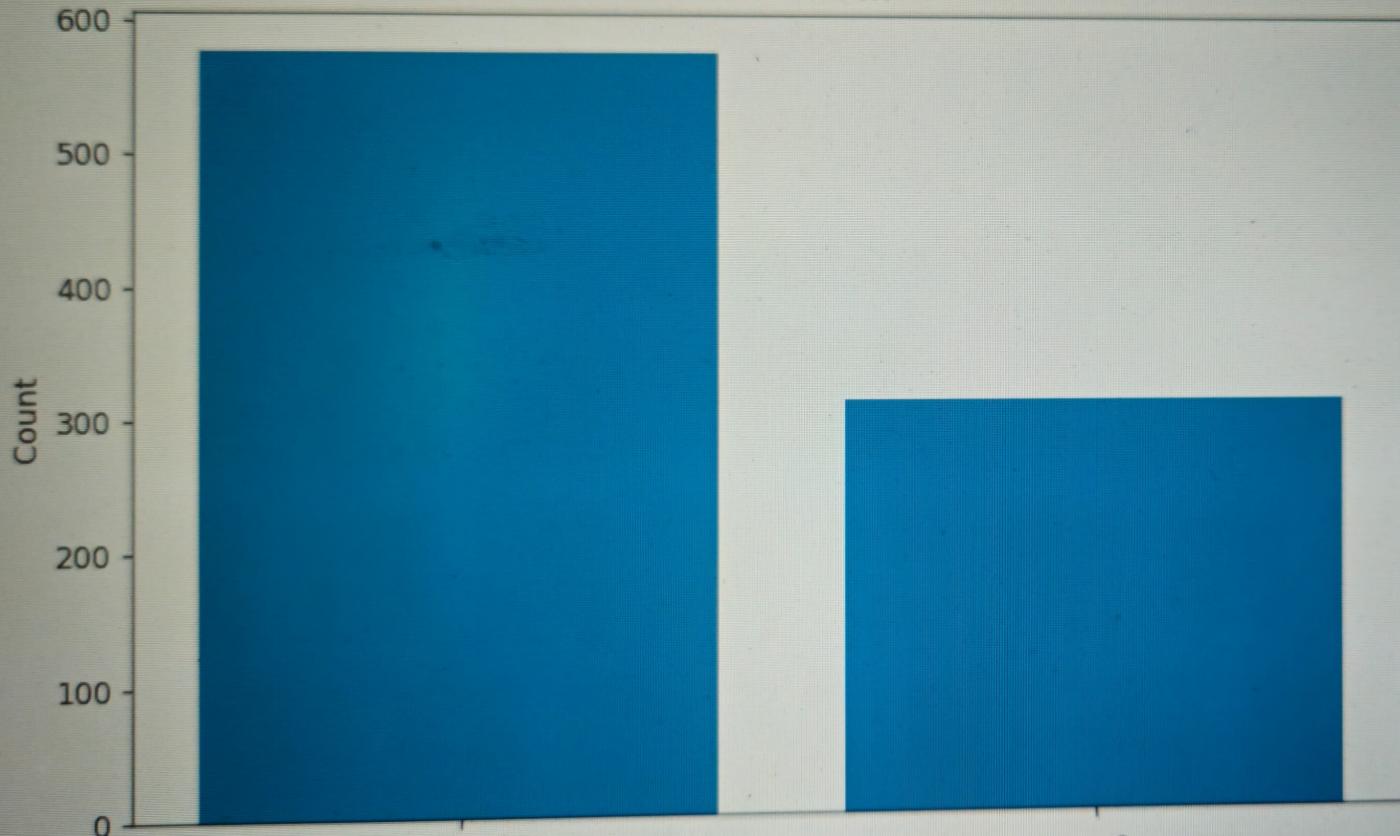


Figure 1

Count of Sex



Count of Embarked

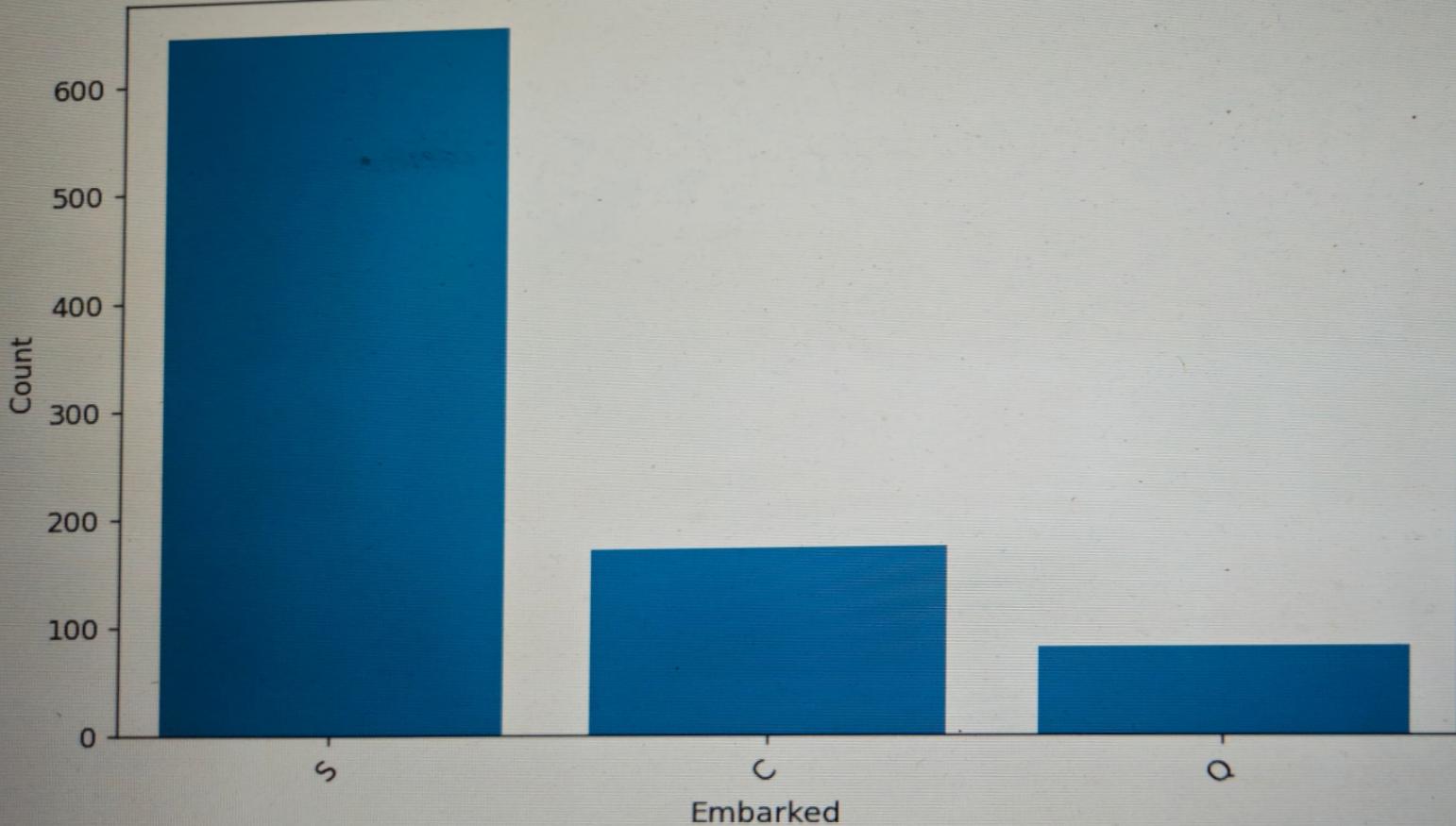
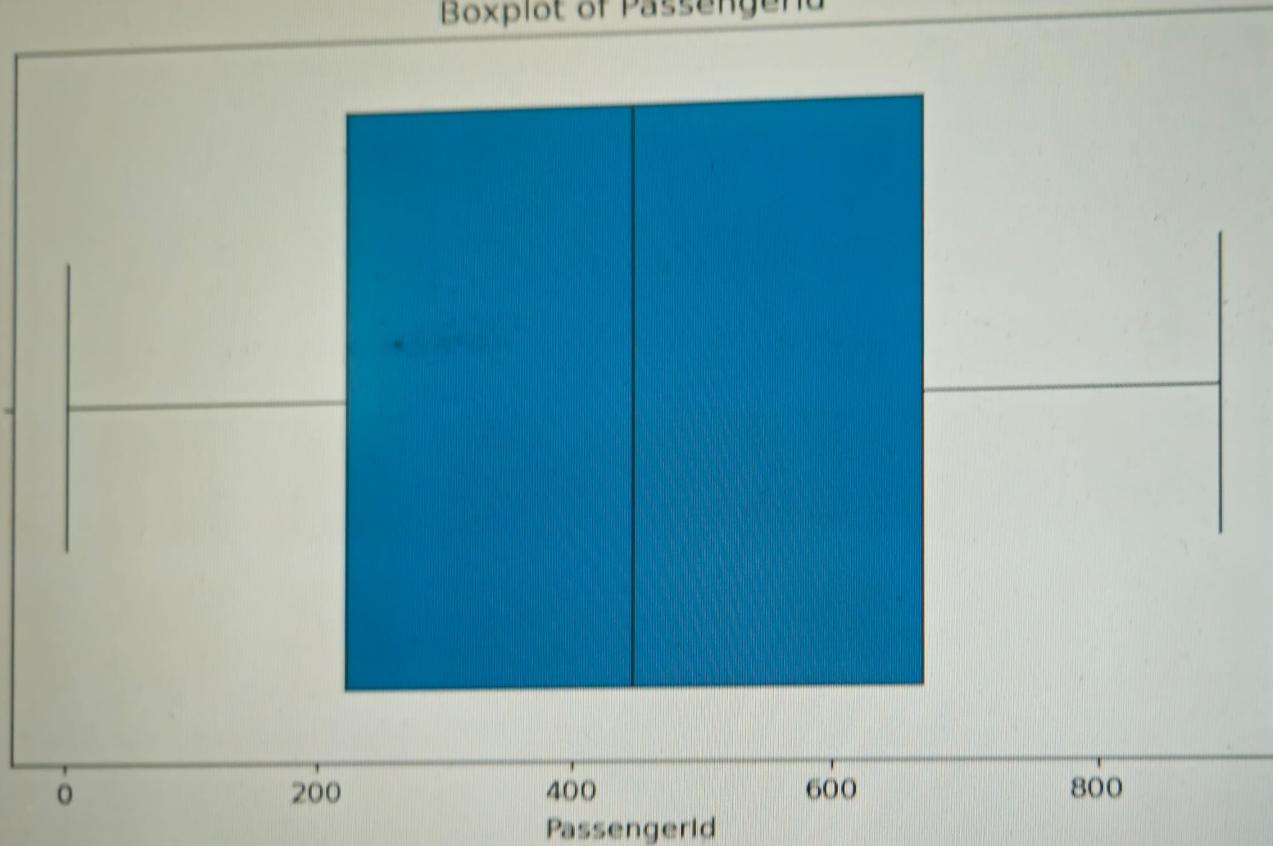




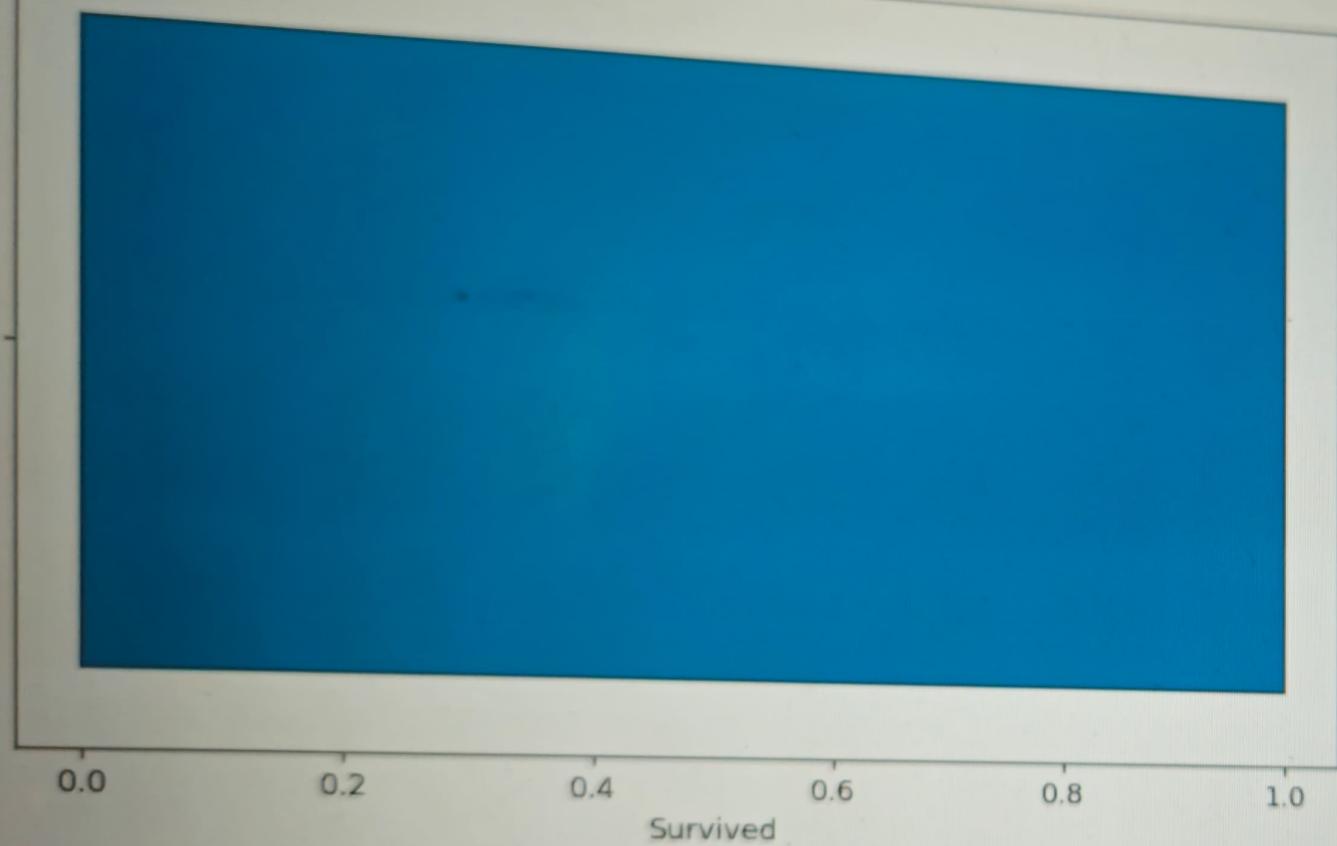
Figure 1

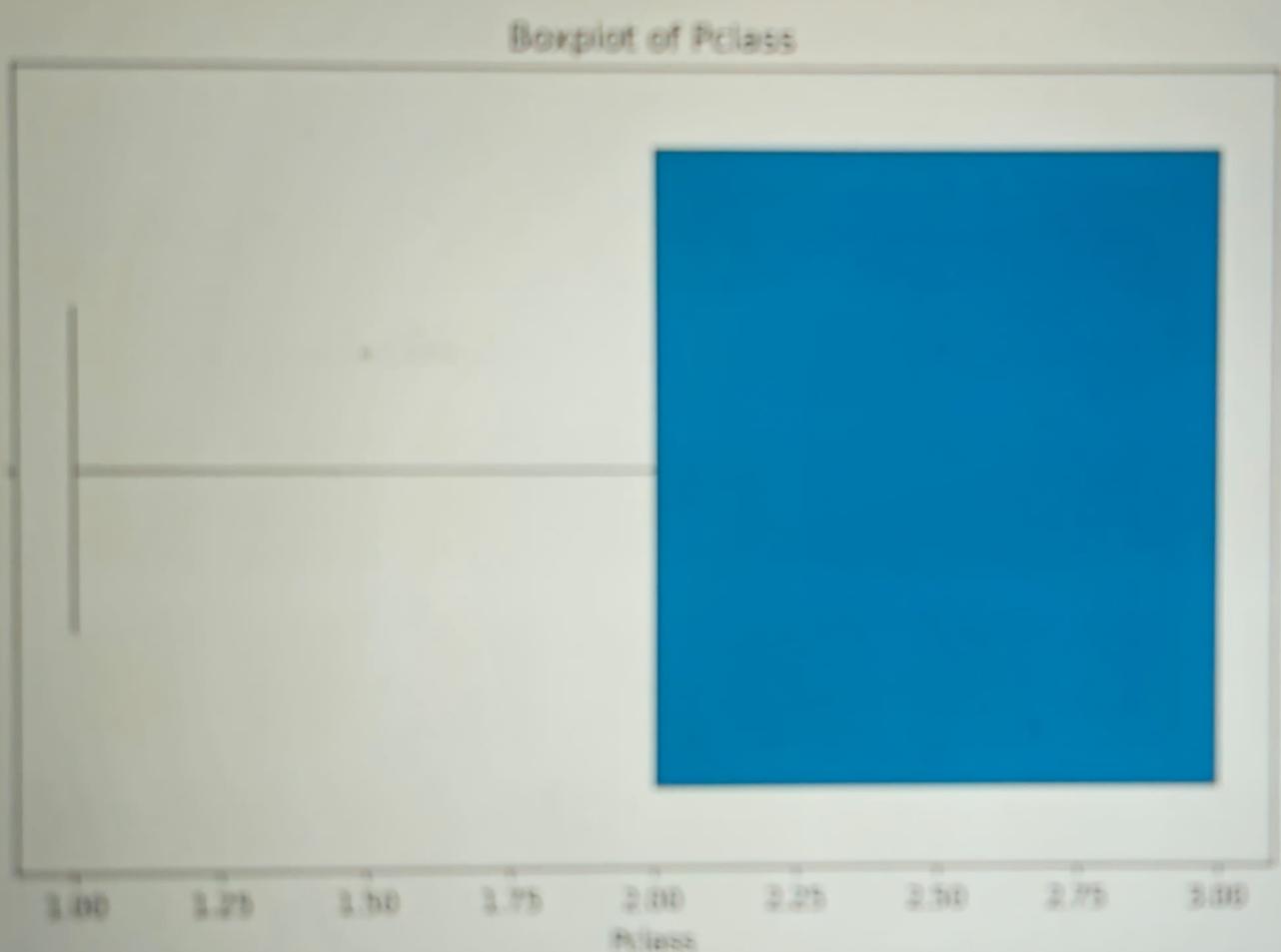


Boxplot of PassengerId



Boxplot of Survived





Python 3.10.2

(tags/v3.10.2-a58ebc)

Figure 1

Boxplot of Age

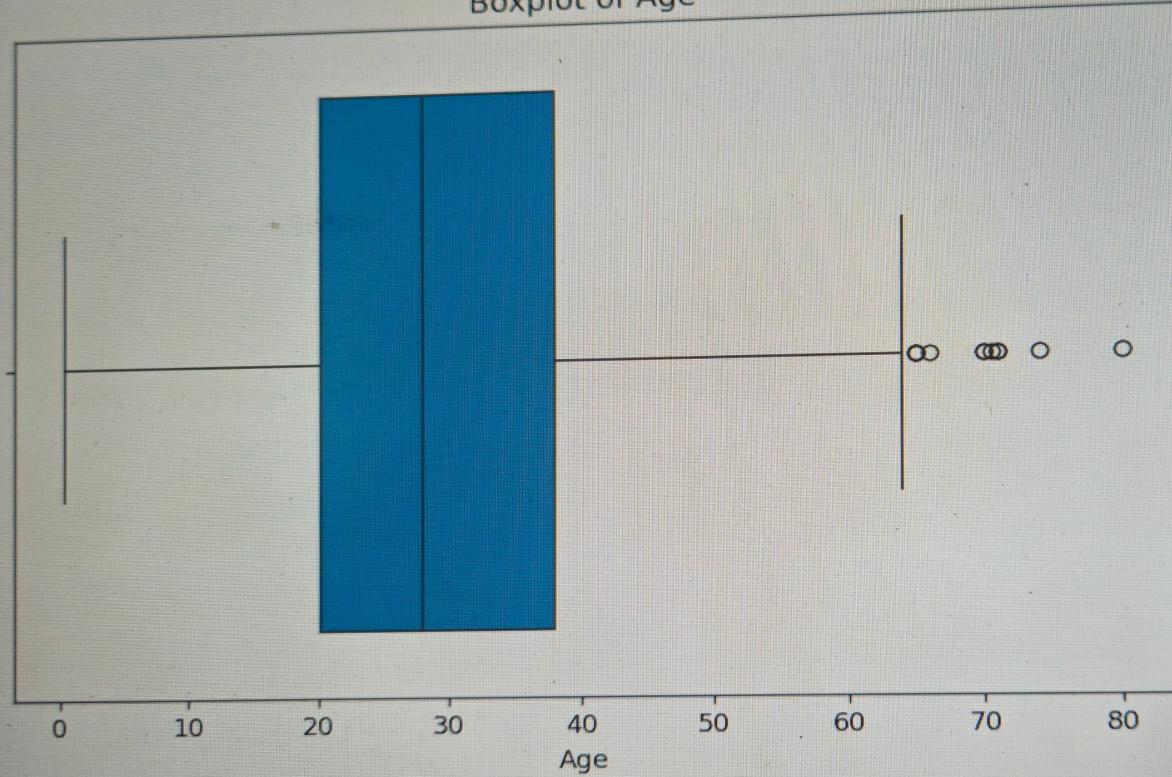
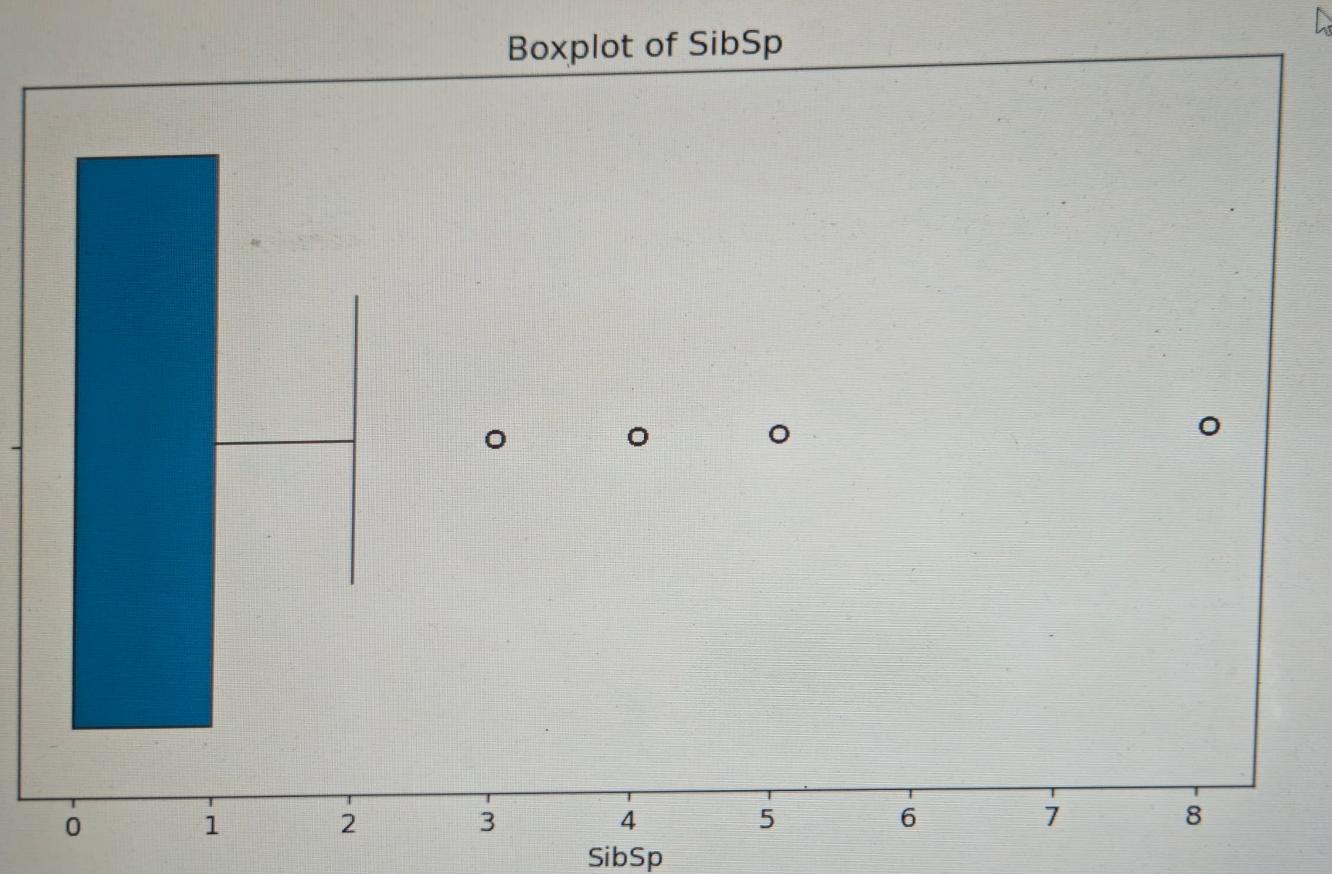
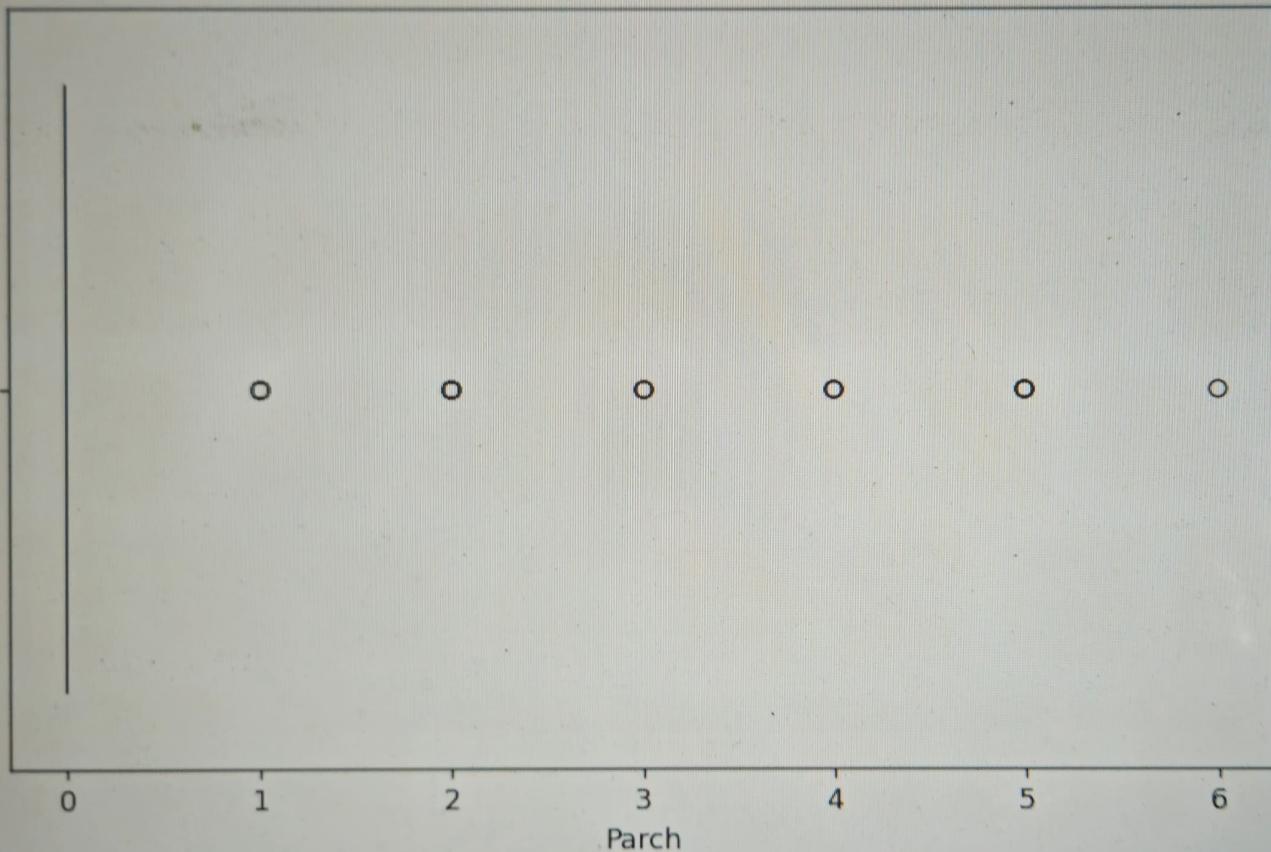


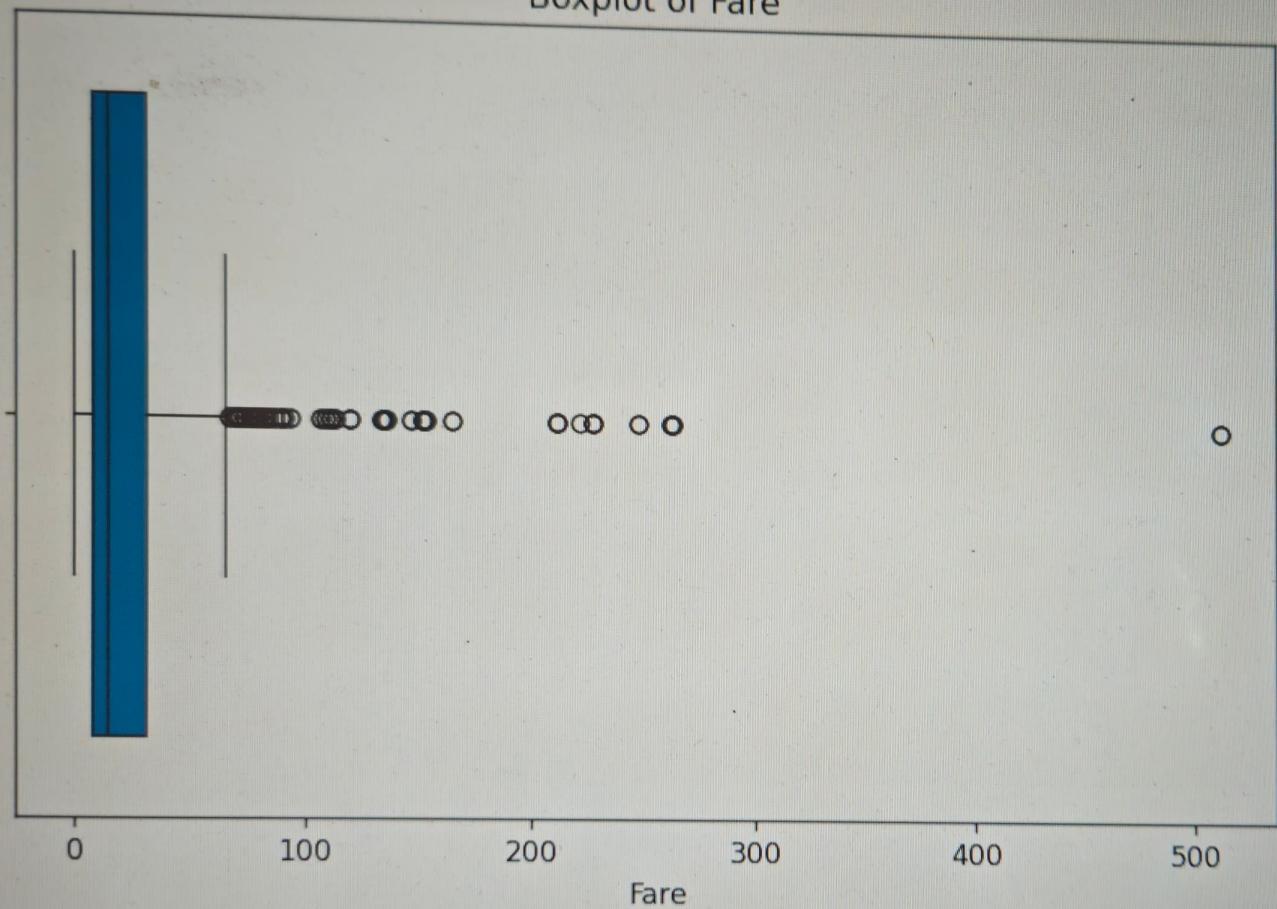
Figure 1



Boxplot of Parch



Boxplot of Fare



===== RESTART: C:/Users/vk171/Downloads/ds_T02.py =====

First 5 rows of the dataset:

	PassengerId	Survived	Pclass	...	Fare	Cabin	Embarked
0	1	0	3	...	7.2500	NaN	
1	2	1	1	...	71.2833	C85	
2	3	1	3	...	7.9250	NaN	
3	4	1	1	...	53.1000	C123	
4	5	0	3	...	8.0500	NaN	

[5 rows x 12 columns]

Dataset Info:

[5 rows x 12 columns]

Dataset Info:

<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	201 non-null	object

```
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

None

Missing Values:

PassengerId	0
Survived	1
Pclass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0

```
Age          177
SibSp         0
Parch         0
Ticket        0
Fare          0
Cabin        687
Embarked      2
dtype: int64
```

Summary Statistics:

	PassengerId	Survived	Pclass	...	SibSp
	Parch	Fare			
count	891.000000	891.000000	891.000000	...	891.000000
mean	446.000000	0.383838	2.308642	...	0.523008
std	257.353842	0.486592	0.836071	...	1.102743
min	1.000000	0.000000	1.000000	0.000000	0.000000

File Edit Shell Debug Options Window Help

25%	223.500000	0.000000	2.000000	...	0.000000
	0.000000	7.910400			
50%	446.000000	0.000000	3.000000	...	0.000000
	0.000000	14.454200			
75%	668.500000	1.000000	3.000000	...	1.000000
	0.000000	31.000000			
max	891.000000	1.000000	3.000000	...	8.000000
	6.000000	512.329200			

[8 rows x 7 columns]

Number of duplicate rows: 0

Numerical columns: ['PassengerId', 'Survived', 'Pclass', 'Age', 'SibSp', 'Parch', 'Fare']

Categorical columns: ['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked']

Handling Missing Values

For example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value) instead, to perform the operation inplace on the original object.

Filled missing values in Age with mean.

Dropped rows with missing values in Cabin.

Dropped rows with missing values in Embarked.

Cleaned Dataset Info:

<class 'pandas.core.frame.DataFrame'>

Index: 202 entries, 1 to 889

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	PassengerId	202 non-null	int64
1	Survived	202 non-null	int64
2	Pclass	202 non-null	int64

Data columns (total 12 columns):

#	Column	Non-Null Count	Dtype
0	PassengerId	202 non-null	int64
1	Survived	202 non-null	int64
2	Pclass	202 non-null	int64
3	Name	202 non-null	object
4	Sex	202 non-null	object
5	Age	202 non-null	float64
6	SibSp	202 non-null	int64
7	Parch	202 non-null	int64
8	Ticket	202 non-null	object
9	Fare	202 non-null	float64
10	Cabin	202 non-null	object
11	Embarked	202 non-null	object

dtypes: float64(2), int64(5), object(5)

memory usage: 20.5+ KB

None