

```
import pandas as pd
```

```
df = pd.read_csv('/content/Titanic-Dataset.csv')
```

```
print(df.head())
print(df.info())
print(df.describe())
```

```
0      1      0      3
1      2      1      1
2      3      1      3
3      4      1      1
4      5      0      3
```

```
0      Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley    female  38.0      1
2      Heikkinen, Miss. Laina    female  26.0      0
3  Futrelle, Mrs. Jacques Heath (L...     NaN      0
4      Allen, Mr.                 male  35.0      1
```

```
0      0      A/5 21171    7.2500    NaN      S
1      0      PC 17599   71.2833    C85      C
2      0  STON/O2. 3101282    7.9250    NaN      S
3      0      113803   53.1000   C123      S
4      0      373450    8.0500    NaN      S
```

```
<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
```

```
None
```


	PassengerId	Survived	Pclass	Age	SibSp
count	891.000000	891.000000	891.000000	714.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008
std	257.353842	0.486592	0.836071	14.526497	1.102743
min	1.000000	0.000000	1.000000	0.420000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000
50%	446.000000	0.000000	3.000000	28.000000	0.000000
75%	668.500000	1.000000	3.000000	38.000000	1.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000

	Parch	Fare
count	891.000000	891.000000
mean	0.381594	32.204208
std	0.806057	49.693429
min	0.000000	0.000000
25%	0.000000	7.910400
50%	0.000000	14.454200
75%	0.000000	31.000000
max	6.000000	512.329200

```
summary_stats = df.describe(include='all')
print(summary_stats)
```

```
print(df.isnull().sum())
```

```
print(df.dtypes)
```



	PassengerId	Survived	Pclass	Name	Sex	\
count	891.000000	891.000000	891.000000	891	891	
unique	NaN	NaN	NaN	891	2	
top	NaN	NaN	NaN	Dooley, Mr. Patrick	male	
freq	NaN	NaN	NaN	1	577	
mean	446.000000	0.383838	2.308642	NaN	NaN	
std	257.353842	0.486592	0.836071	NaN	NaN	
min	1.000000	0.000000	1.000000	NaN	NaN	
25%	223.500000	0.000000	2.000000	NaN	NaN	
50%	446.000000	0.000000	3.000000	NaN	NaN	
75%	668.500000	1.000000	3.000000	NaN	NaN	
max	891.000000	1.000000	3.000000	NaN	NaN	

	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
count	714.000000	891.000000	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	681	NaN	147	3
top	NaN	NaN	NaN	347082	NaN	G6	S
freq	NaN	NaN	NaN	7	NaN	4	644
mean	29.699118	0.523008	0.381594	NaN	32.204208	NaN	NaN
std	14.526497	1.102743	0.806057	NaN	49.693429	NaN	NaN
min	0.420000	0.000000	0.000000	NaN	0.000000	NaN	NaN
25%	20.125000	0.000000	0.000000	NaN	7.910400	NaN	NaN
50%	28.000000	0.000000	0.000000	NaN	14.454200	NaN	NaN
75%	38.000000	1.000000	0.000000	NaN	31.000000	NaN	NaN
max	80.000000	8.000000	6.000000	NaN	512.329200	NaN	NaN

PassengerId0

Survived0

Pclass0

Name0

Sex0

Age177

SibSp0

Parch0

Ticket0

Fare0

Cabin687

Embarked2

dtype: int64

PassengerIdint64

Survivedint64

Pclassint64

Nameobject

Sexobject

Agefloat64

SibSpint64

Parchint64


Ticketobject

Farefloat64

Cabinobject

Embarkedobject

dtype: object

 Generate

randomly select 5 items from a list



Close

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

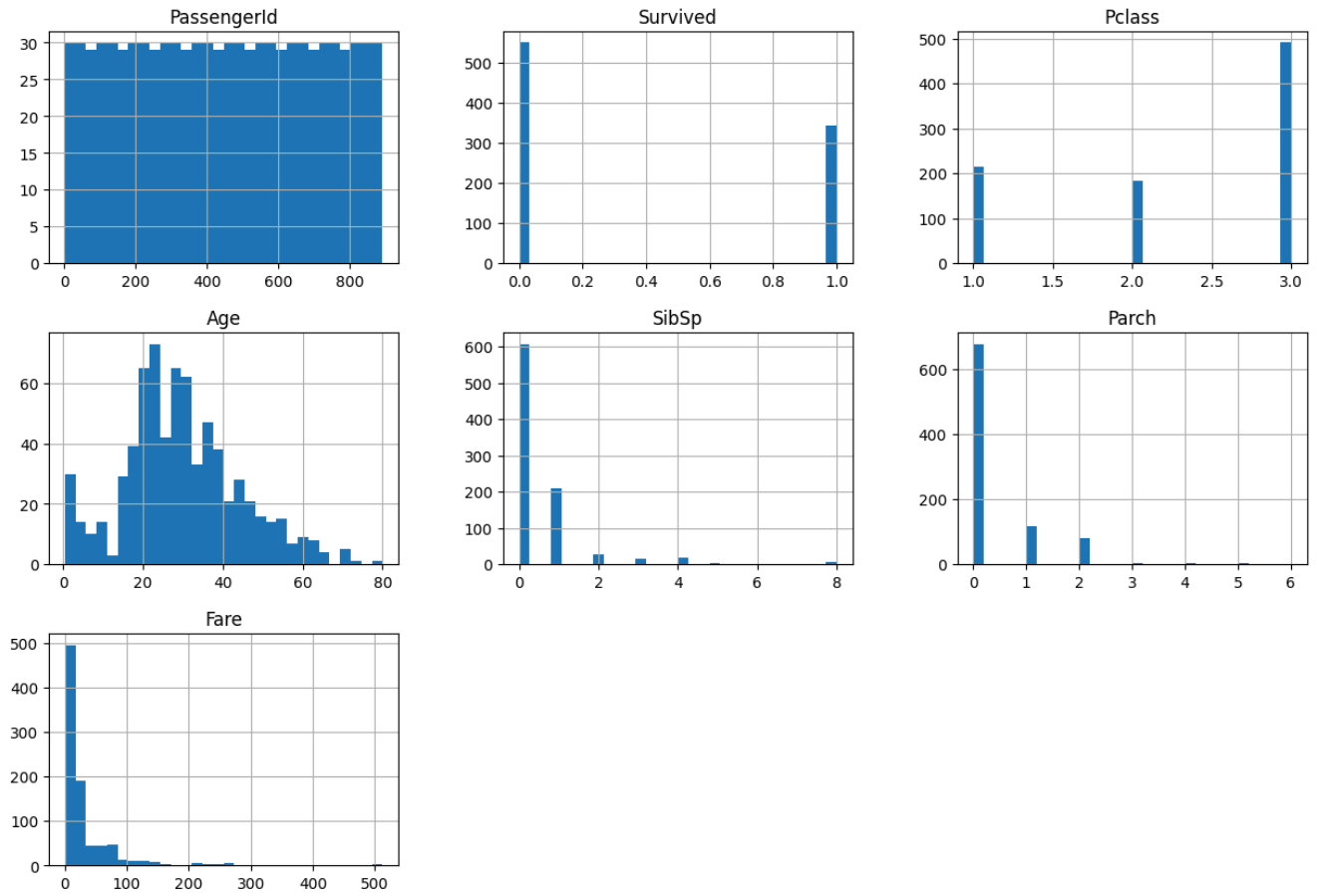
df.hist(bins=30, figsize=(15, 10))
plt.suptitle('Histograms of Numeric Features')
plt.show()

numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns

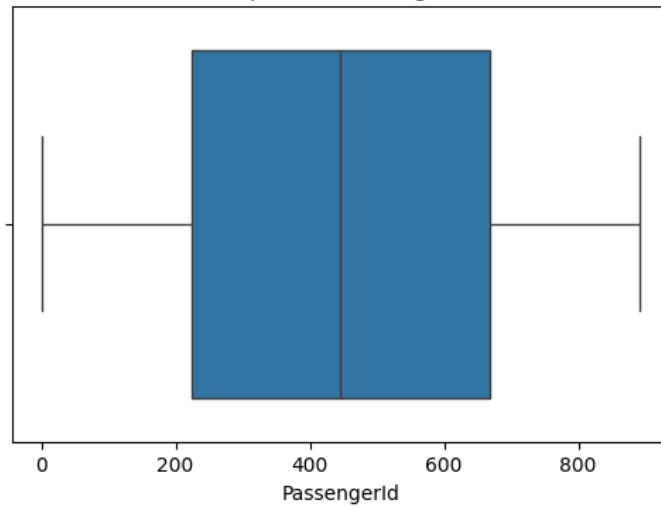
for col in numeric_cols:
    plt.figure(figsize=(6, 4))
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot of {col}')
    plt.show()
```



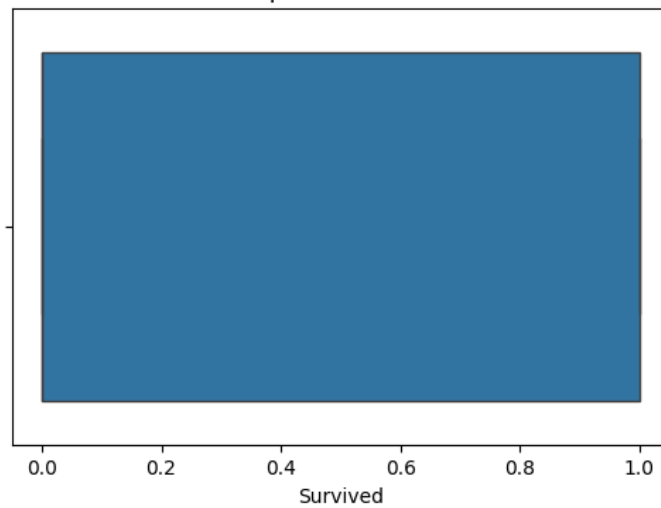
Histograms of Numeric Features



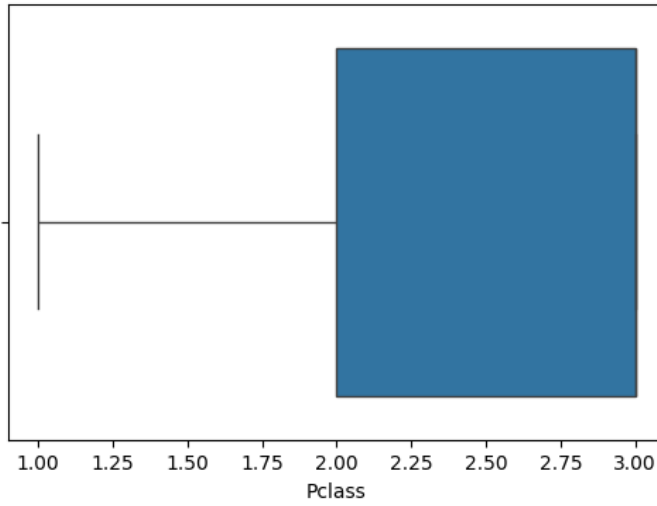
Boxplot of PassengerId



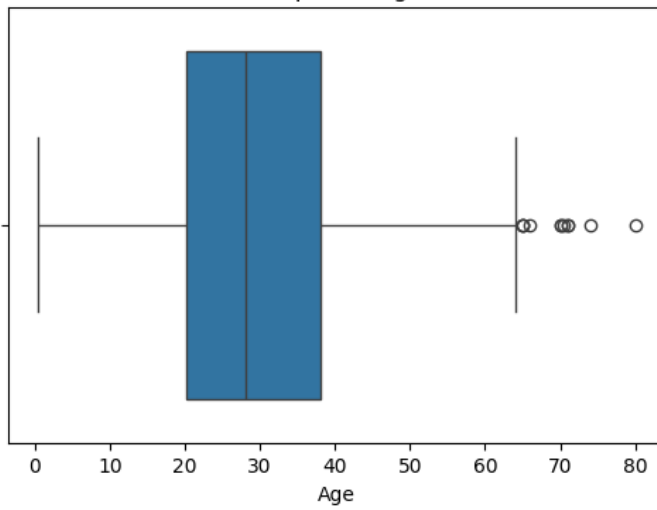
Boxplot of Survived



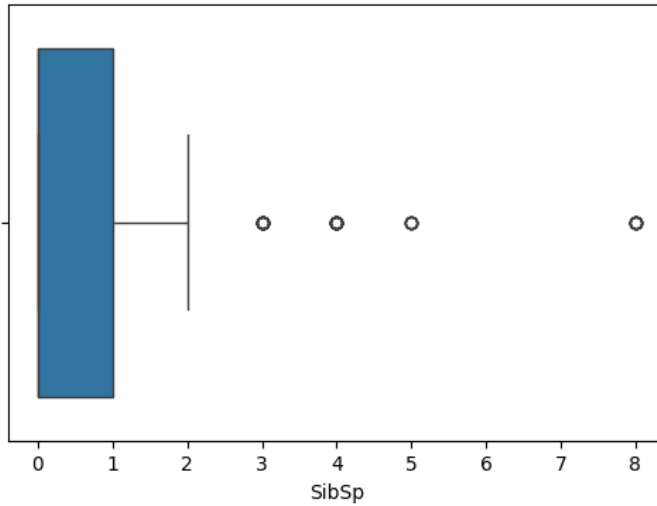
Boxplot of Pclass



Boxplot of Age



Boxplot of SibSp

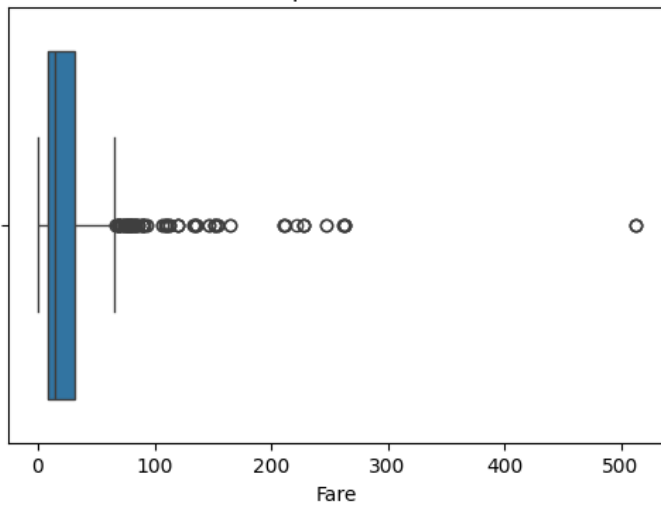


Boxplot of Parch





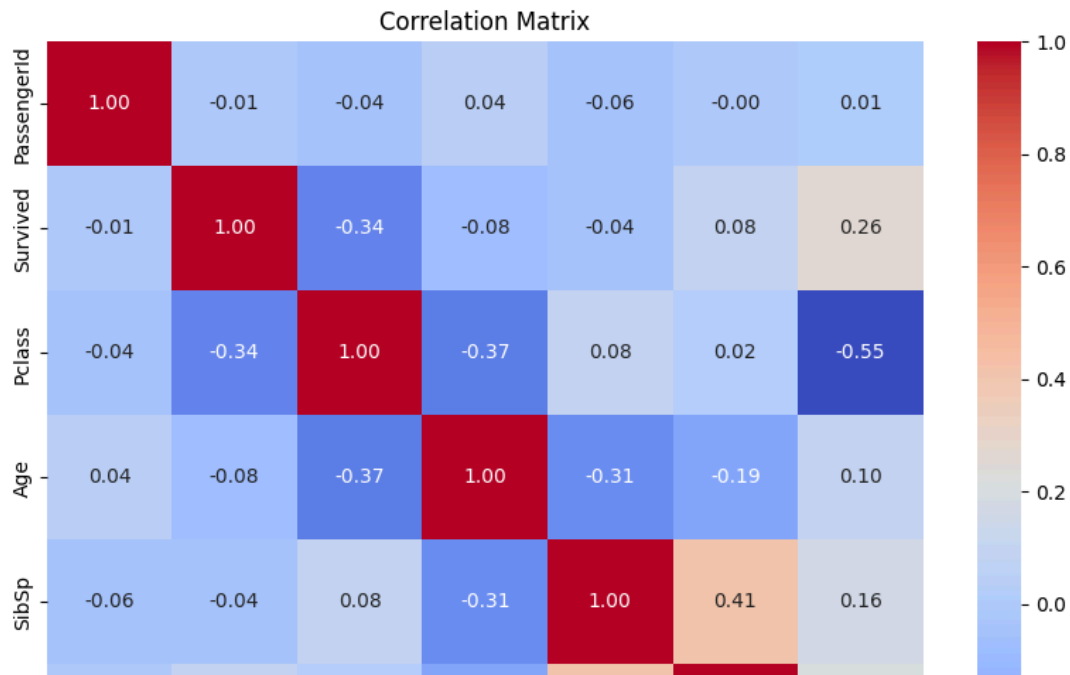
Boxplot of Fare



```
plt.figure(figsize=(10, 8))

numeric_df = df.select_dtypes(include=['float64', 'int64'])
corr = numeric_df.corr()
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

sns.pairplot(numeric_df)
plt.show()
```



```
import plotly.express as px
```

```
fig = px.histogram(df, x='Age', nbins=30, title='Interactive Histogram of Age')  
fig.show()
```

```
fig = px.scatter_matrix(df[numeric_cols])  
fig.show()
```



Interactive Histogram of Age

