

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

Matplotlib is building the font cache; this may take a moment.

```
import pandas as pd

# Load the dataset
df = pd.read_csv(r'C:\Users\M.MATHAN\Downloads\archive (5)\Titanic-Dataset.csv')

# Display the first few rows to check
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs T. B.)	female	38.0	1	0	PC 17599	71.2

```
df.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        204 non-null    object
11  Embarked     889 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
df.describe()
```



	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
<b>count</b>	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
<b>mean</b>	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204200
<b>std</b>	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693422
<b>min</b>	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
<b>25%</b>	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910452
<b>50%</b>	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
<b>75%</b>	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
<b>max</b>	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329000

```
df.head()
```



	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
<b>0</b>	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
<b>1</b>	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs T. B.)	female	38.0	1	0	PC 17599	71.2833

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



```

PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age           177
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin         687
Embarked       2
dtype: int64

```

```
df['Sex'].value_counts()
```



```

Sex
male      577
female    314
Name: count, dtype: int64

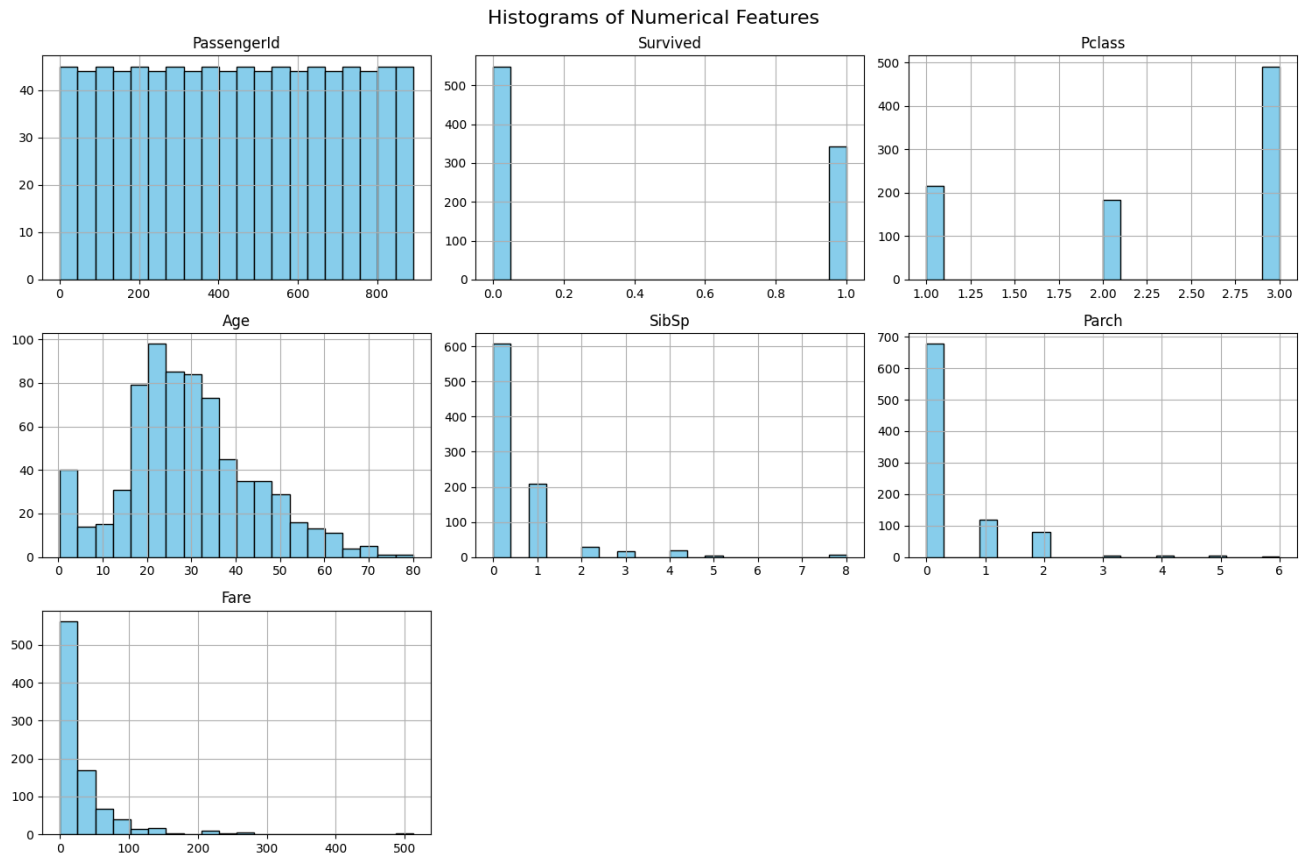
```

```
df['Pclass'].value_counts()
```

```
↵ Pclass  
3    491  
1    216  
2    184  
Name: count, dtype: int64
```

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

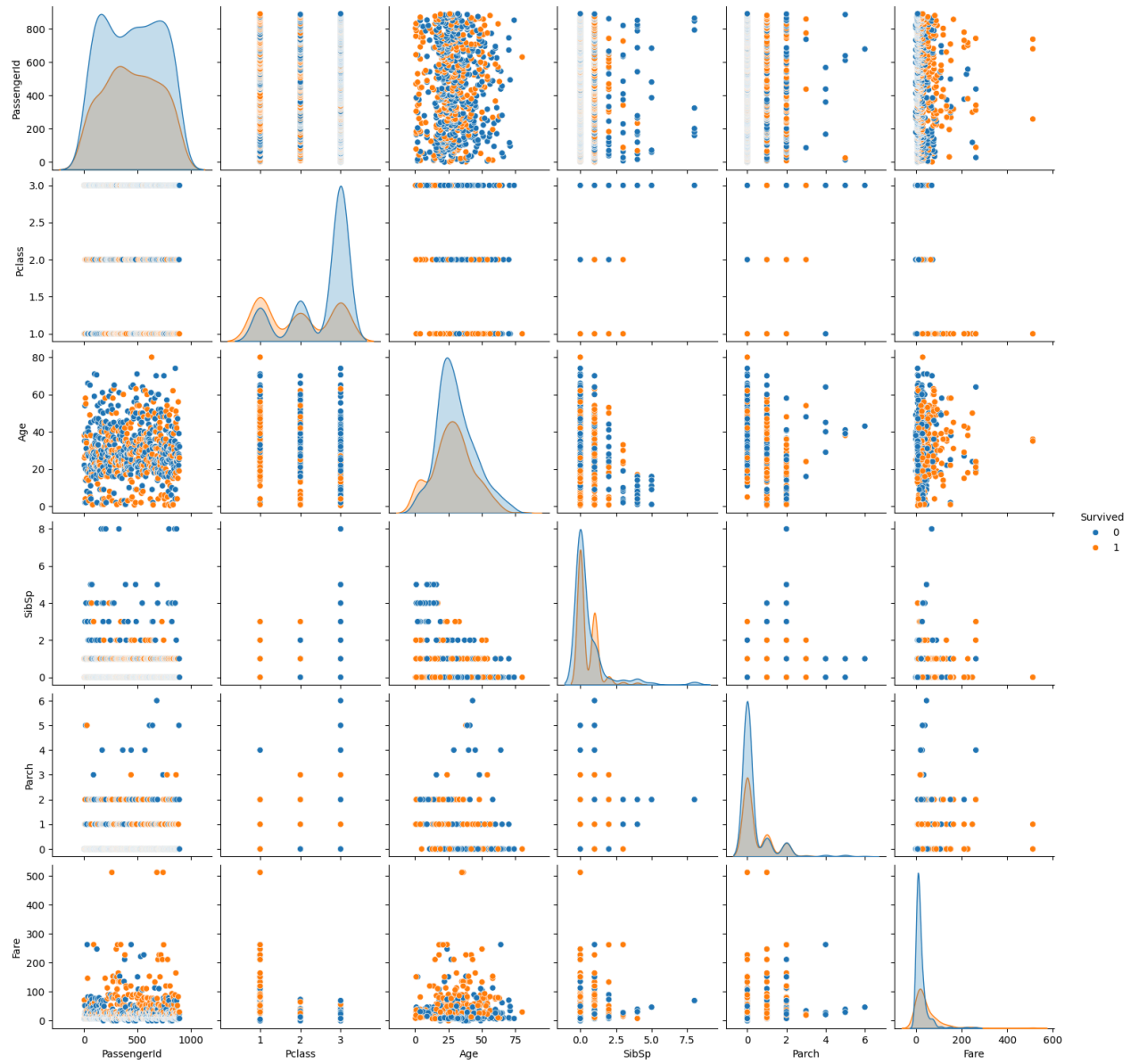
```
df.hist(bins=20, figsize=(15, 10), color='skyblue', edgecolor='black')  
plt.suptitle('Histograms of Numerical Features', fontsize=16)  
plt.tight_layout()  
plt.show()
```



```
import seaborn as sns
```

```
sns.pairplot(df, hue='Survived')
```

↗ <seaborn.axisgrid.PairGrid at 0x1de90bfd190>



```
plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap='coolwarm', linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()
```