

```
In [77]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly.express as px
import plotly.graph_objs as go
```

```
In [78]: df = pd.read_csv('C:\\Users\\EagleCORS\\Assignment 5\\ADS-Assignment-5\\Titanic Data.csv')
```

```
In [79]: df
```

Out[79]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

In [80]: df.describe

```
Out[80]: <bound method NDFrame.describe of      PassengerId  Survived  Pclass  \
0              1         0        3
1              2         1        1
2              3         1        3
3              4         1        1
4              5         0        3
..          ...     ...     ...
886           887         0        2
887           888         1        1
888           889         0        3
889           890         1        1
890           891         0        3

      Name      Sex  Age  SibSp  \
0      Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2      Heikkinen, Miss. Laina    female  26.0      0
3  Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4      Allen, Mr. William Henry    male  35.0      0
..          ...     ...     ...     ...
886      Montvila, Rev. Juozas    male  27.0      0
887      Graham, Miss. Margaret Edith    female  19.0      0
888  Johnston, Miss. Catherine Helen "Carrie"    female   NaN      1
889      Behr, Mr. Karl Howell    male  26.0      0
890      Dooley, Mr. Patrick    male  32.0      0

      Parch      Ticket    Fare Cabin Embarked
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
..      ...     ...     ...     ...
886         0      211536  13.0000   NaN      S
887         0      112053  30.0000   B42      S
888         2      W./C. 6607  23.4500   NaN      S
889         0      111369  30.0000  C148      C
890         0      370376   7.7500   NaN      Q

[891 rows x 12 columns]>
```

In [81]: df.head()

```
Out[81]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [82]: df.tail()

```
Out[82]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

```
In [83]: df.corr()
```

Out[83]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.012658
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.257307
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.549500
Age	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.096067
SibSp	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.159651
Parch	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.216225
Fare	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.000000

```
In [84]: 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
```

```
In [85]: df_encoded = pd.get_dummies(df, columns=['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked'])
```

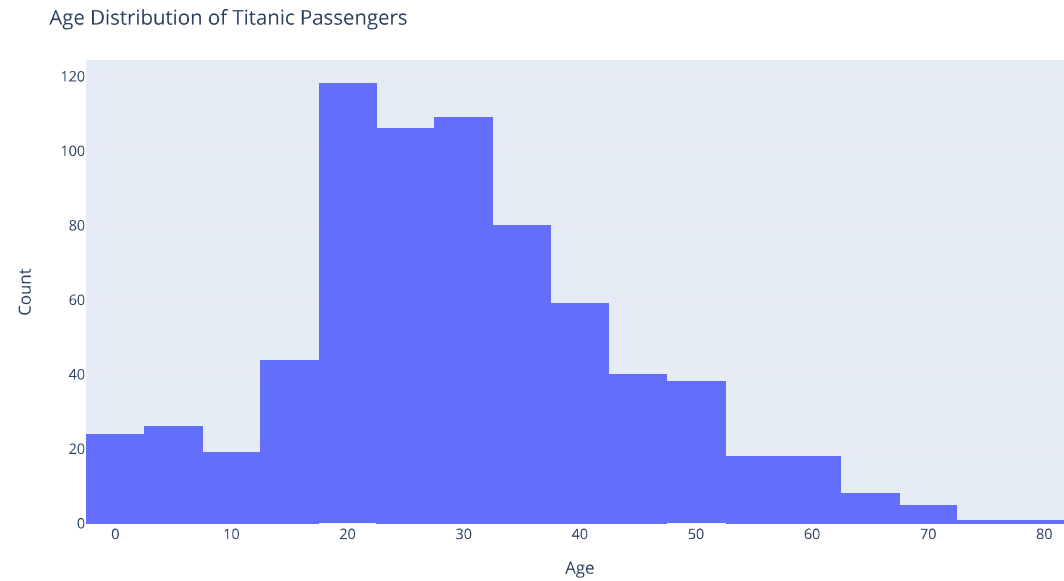
```
In [86]: df_encoded
```

Out[86]:

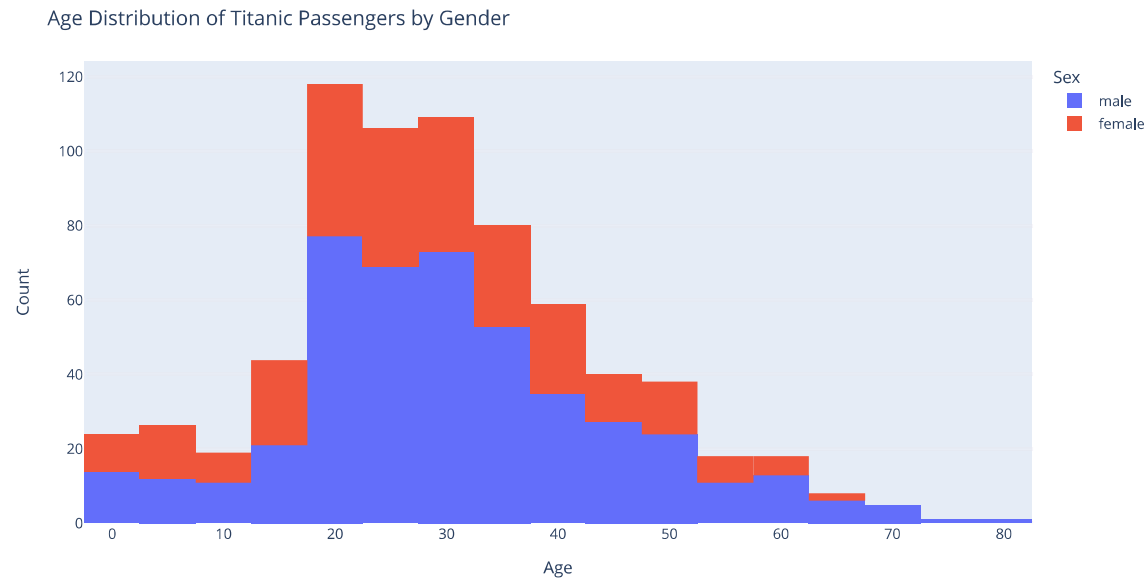
	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	Name_Abbing, Mr. Anthony	Name_Abbott, Mr. Rossmore Edward	Name_Abbott, Mrs. Stanton (Rosa Hunt)	...	Cabin_F G73	Cabin_F2	Cabin_F33	Cabin_F38	Cabin_F4	Cabin_G6	Cabin_T	Embarked_C	Embarked_Q	Embarked_S
0	1	0	3	22.0	1	0	7.2500	0	0	0	...	0	0	0	0	0	0	0	0	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0	...	0	0	0	0	0	0	0	1	0	0
2	3	1	3	26.0	0	0	7.9250	0	0	0	...	0	0	0	0	0	0	0	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	0	...	0	0	0	0	0	0	0	0	0	1
4	5	0	3	35.0	0	0	8.0500	0	0	0	...	0	0	0	0	0	0	0	0	0	1
...
886	887	0	2	27.0	0	0	13.0000	0	0	0	...	0	0	0	0	0	0	0	0	0	1
887	888	1	1	19.0	0	0	30.0000	0	0	0	...	0	0	0	0	0	0	0	0	0	1
888	889	0	3	NaN	1	2	23.4500	0	0	0	...	0	0	0	0	0	0	0	0	0	1
889	890	1	1	26.0	0	0	30.0000	0	0	0	...	0	0	0	0	0	0	0	1	0	0
890	891	0	3	32.0	0	0	7.7500	0	0	0	...	0	0	0	0	0	0	0	0	1	0

891 rows × 1731 columns

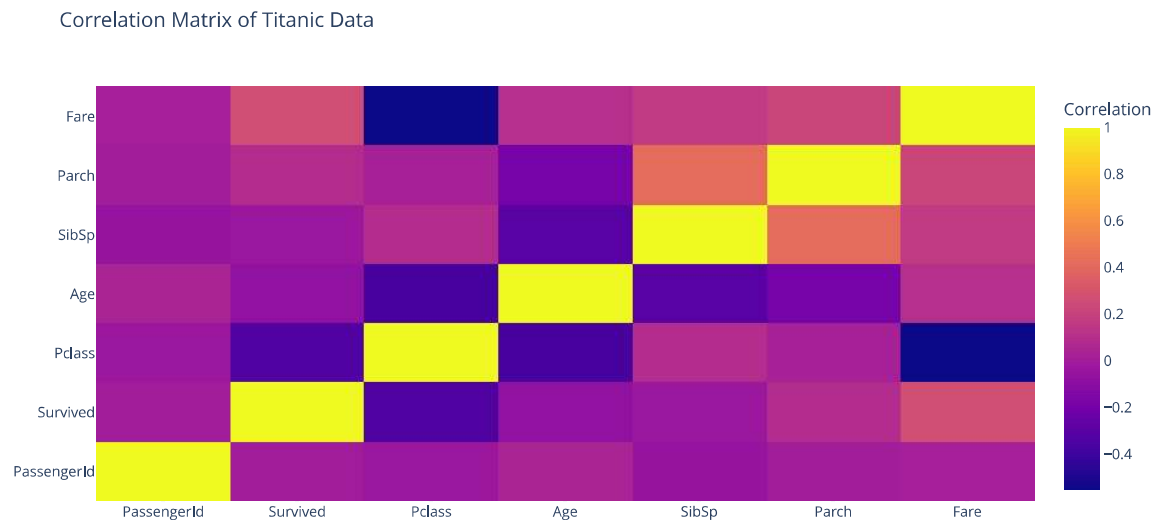
```
In [87]: # 1. Show the age distribution in the data using a histogram
fig1 = px.histogram(df, x='Age', nbins=30)
fig1.update_layout(title='Age Distribution of Titanic Passengers', xaxis_title='Age', yaxis_title='Count')
```



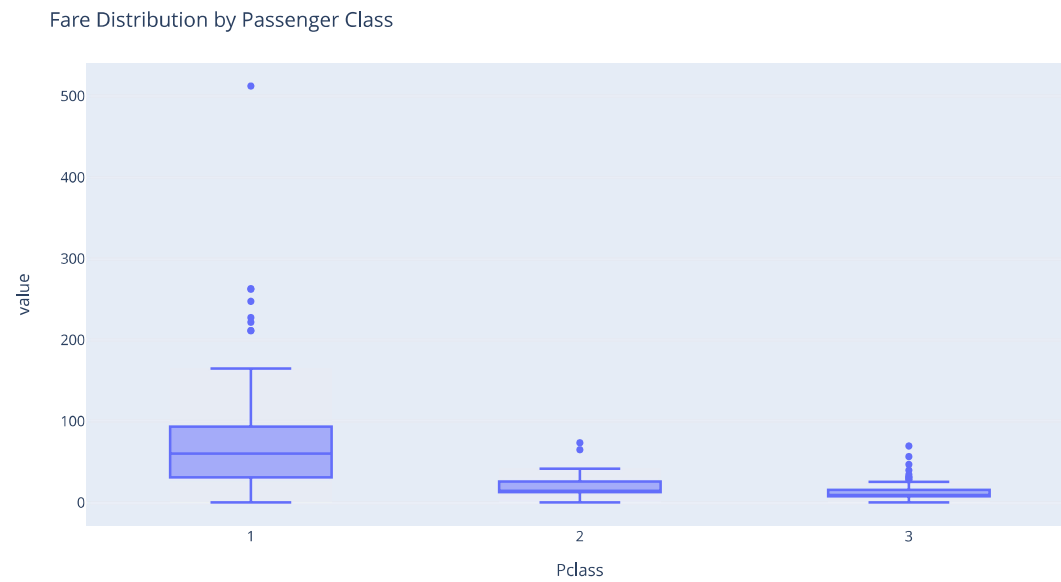
```
In [88]: # 2. Show the age distribution based on gender using a histogram
fig2 = px.histogram(df, x='Age', color='Sex', nbins=30)
fig2.update_layout(title='Age Distribution of Titanic Passengers by Gender', xaxis_title='Age', yaxis_title='Count')
```



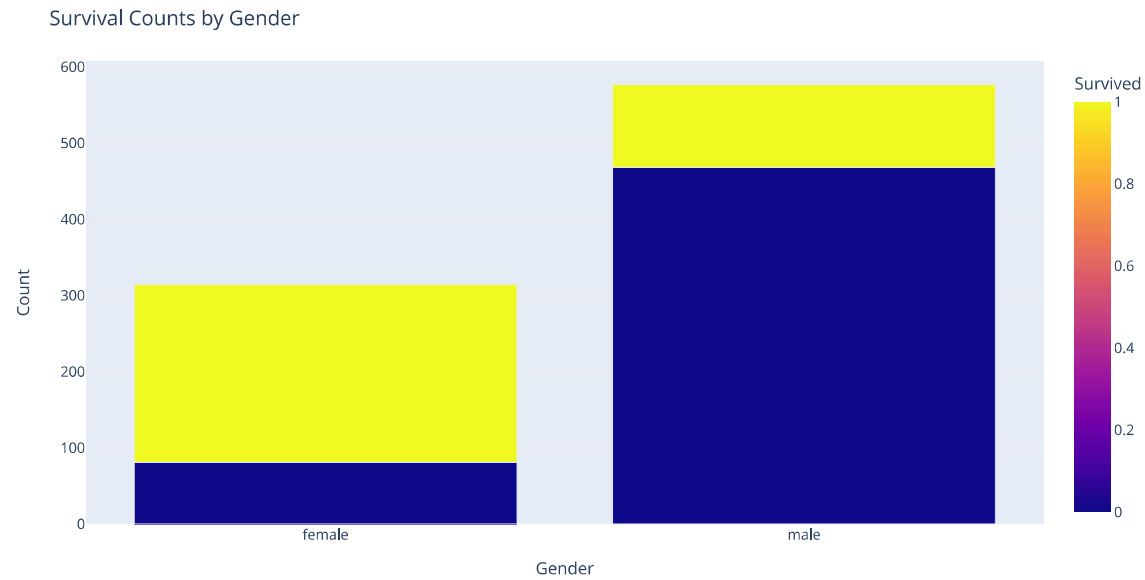
```
In [89]: # 3. Represent the correlation results using a matrix plot
corr = df.corr()
fig3 = go.Figure(data=go.Heatmap(z=corr, x=corr.columns, y=corr.columns, colorbar=dict(title='Correlation')))
fig3.update_layout(title='Correlation Matrix of Titanic Data')
```



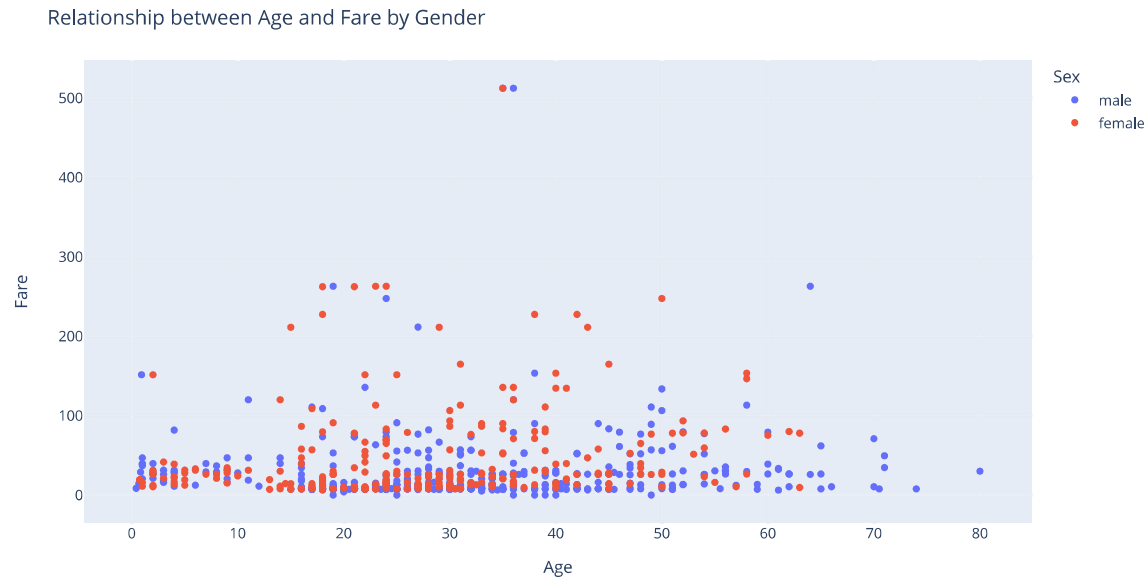
```
In [90]: # 4. Pivot the data and show the distribution using a boxplot
pivot = df.pivot(index='PassengerId', columns='Pclass', values='Fare')
fig4 = px.box(pivot, boxmode='group')
fig4.update_layout(title='Fare Distribution by Passenger Class')
```



```
In [91]: # 5. Graph the value counts of survivors and non-survivors by gender using a stacked bar graph
group = df.groupby(['Survived', 'Sex'])['PassengerId'].count().reset_index()
fig5 = px.bar(group, x='Sex', y='PassengerId', color='Survived', barmode='stack')
fig5.update_layout(title='Survival Counts by Gender', xaxis_title='Gender', yaxis_title='Count')
```



```
In [92]: # 6. Plot the ages to the fare paid by each passenger based on their gender using a scatter plot
fig6 = px.scatter(df, x='Age', y='Fare', color='Sex')
fig6.update_layout(title='Relationship between Age and Fare by Gender', xaxis_title='Age', yaxis_title='Fare')
```



```
In [93]: # 7. Plot a bubble plot of the ages to the fare paid by each passenger categorizing whether they survived or not
fig7 = px.scatter(df, x='Age', y='Fare', color='Survived', size='Pclass', hover_name='Name')
fig7.update_layout(title='Relationship between Age, Fare, Survival and Passenger Class', xaxis_title='Age', yaxis_title='Fare')
```

