

```
In [1]: import plotly as py
import plotly.express as px
import plotly.graph_objects as go
import pandas as pd
import numpy as np
import plotly.figure_factory as ff

In [2]: df = pd.read_csv('Titanic Data.csv')

In [3]: df.head()

Out[3]:
```

| PassengerId | Survived | Pclass | Name  | Sex    | Age  | SibSp | Parch | Ticket           | Fare    | Cabin | Embarked |
|-------------|----------|--------|---|--------|------|-------|-------|------------------|---------|-------|----------|
| 0           | 1        | 0      | Braund, Mr. Owen Harris                           | male   | 22.0 | 1     | 0     | A/5 21171        | 7.250   | NaN   | S        |
| 1           | 2        | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1     | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2           | 3        | 1      | Heikkinen, Miss. Laina                            | female | 26.0 | 0     | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3           | 4        | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35.0 | 1     | 0     | 113803           | 53.1000 | C123  | S        |
| 4           | 5        | 0      | Allen, Mr. William Henry                          | male   | 35.0 | 0     | 0     | 373450           | 8.0500  | NaN   | S        |

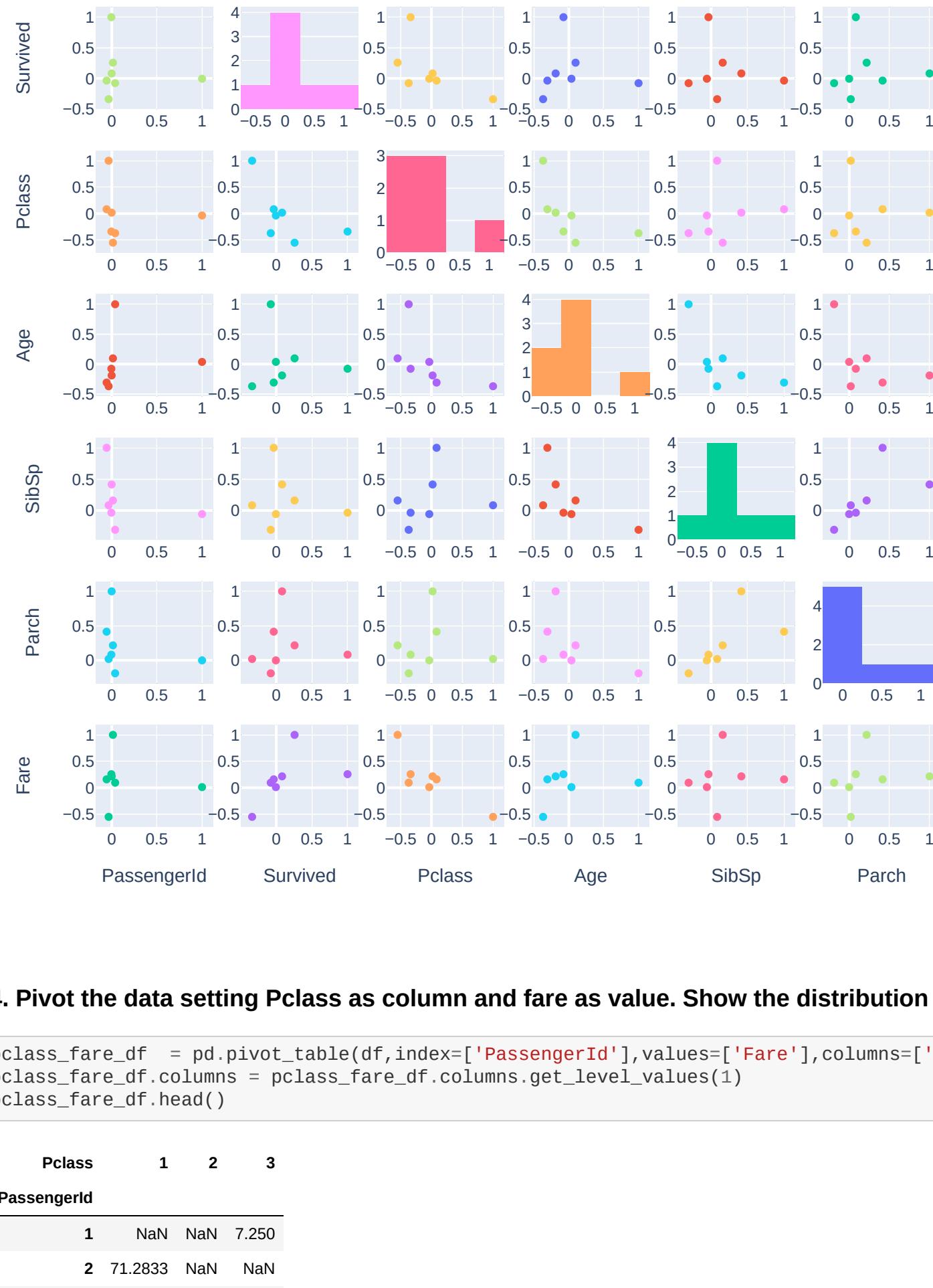
### 1. Show the age distribution using a histogram

```
In [4]: df['Age'] = np.floor(pd.to_numeric(df['Age'], errors='coerce')).astype('Int64')

Out[4]:
```

| PassengerId | Survived | Pclass | Name  | Sex    | Age  | SibSp | Parch | Ticket           | Fare    | Cabin | Embarked |
|-------------|----------|--------|---|--------|------|-------|-------|------------------|---------|-------|----------|
| 0           | 1        | 0      | Braund, Mr. Owen Harris                           | male   | 22.0 | 1     | 0     | A/5 21171        | 7.250   | NaN   | S        |
| 1           | 2        | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1     | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2           | 3        | 1      | Heikkinen, Miss. Laina                            | female | 26.0 | 0     | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3           | 4        | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35.0 | 1     | 0     | 113803           | 53.1000 | C123  | S        |
| 4           | 5        | 0      | Allen, Mr. William Henry                          | male   | 35.0 | 0     | 0     | 373450           | 8.0500  | NaN   | S        |

Passenger ages



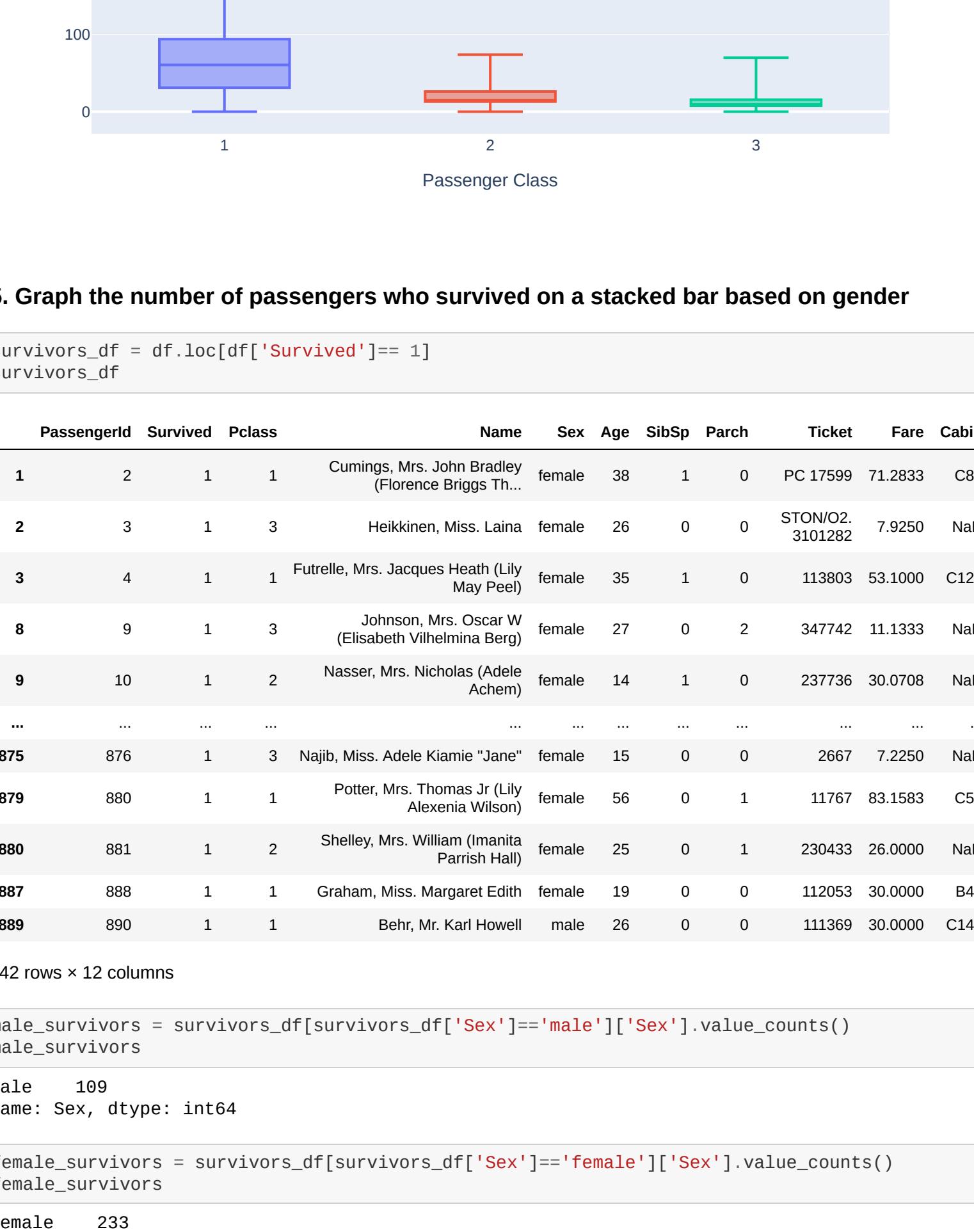
### 2. Show the age distribution by gender

```
In [6]: px.histogram(data_frame=age_df,x='Age',color='Sex',title='Passenger ages',color_discrete_sequence = ['LightSkyBlue','Pink'],nbins=100)

Out[6]:
```

| PassengerId | Survived | Pclass | Name  | Sex    | Age  | SibSp | Parch | Ticket           | Fare    | Cabin | Embarked |
|-------------|----------|--------|---|--------|------|-------|-------|------------------|---------|-------|----------|
| 0           | 1        | 0      | Braund, Mr. Owen Harris                           | male   | 22.0 | 1     | 0     | A/5 21171        | 7.250   | NaN   | S        |
| 1           | 2        | 1      | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38.0 | 1     | 0     | PC 17599         | 71.2833 | C85   | C        |
| 2           | 3        | 1      | Heikkinen, Miss. Laina                            | female | 26.0 | 0     | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 3           | 4        | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35.0 | 1     | 0     | 113803           | 53.1000 | C123  | S        |
| 4           | 5        | 0      | Allen, Mr. William Henry                          | male   | 35.0 | 0     | 0     | 373450           | 8.0500  | NaN   | S        |

Passenger ages



### 3. Create a matrix plot of the correlation within the data

```
In [7]: correlation_df = df.corr()
correlation_df

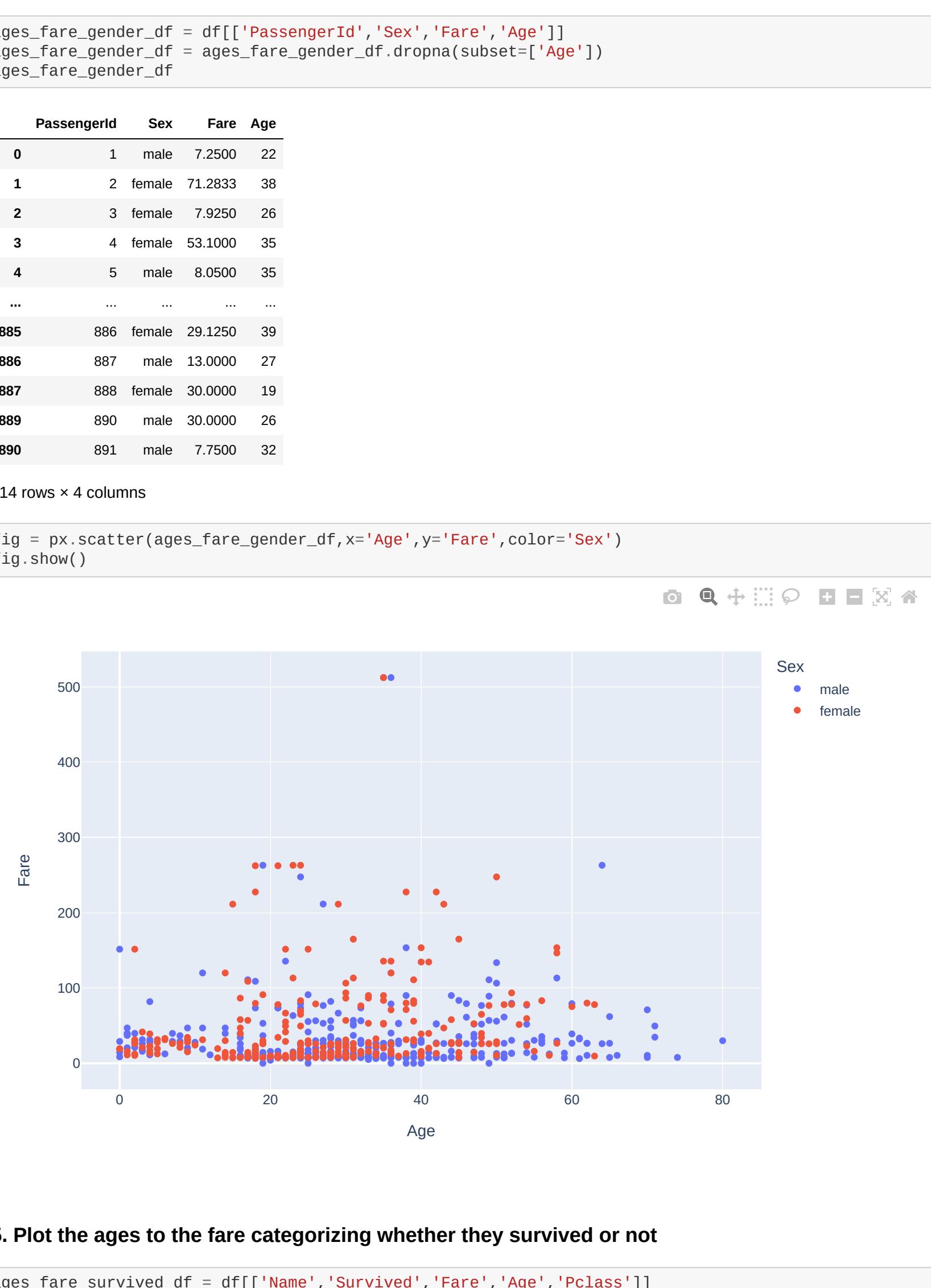
Out[7]:
```

| PassengerId | Survived  | Pclass    | Age       | SibSp     | Parch     | Fare      |           |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| PassengerId | 1.000000  | -0.050007 | -0.035144 | 0.036982  | -0.057527 | -0.001652 | 0.012658  |
| Survived    | -0.005007 | 1.000000  | -0.338481 | -0.077151 | -0.035322 | 0.081629  | 0.257307  |
| Pclass      | -0.035144 | -0.338481 | 1.000000  | -0.369450 | 0.083081  | 0.018443  | -0.549500 |
| Age         | 0.036982  | -0.077151 | -0.369450 | 1.000000  | -0.307947 | -0.189307 | 0.096288  |
| SibSp       | -0.057527 | -0.035322 | 0.083081  | -0.307947 | 1.000000  | 0.414838  | 0.159651  |
| Parch       | -0.001652 | 0.081629  | 0.018443  | -0.189307 | 0.414838  | 1.000000  | 0.216225  |
| Fare        | 0.012658  | 0.257307  | -0.549500 | 0.096288  | 0.159651  | 0.216225  | 1.000000  |

```
In [8]: fig = ff.create_scatterplotmatrix(correlation_df, diag='histogram', height=900, width=900)
py.offline.iplot(fig, filename='Titanic correlation matrix')

Out[8]:
```

Scatterplot Matrix



### 4. Pivot the data setting Pclass as column and fare as value. Show the distribution on a box plot

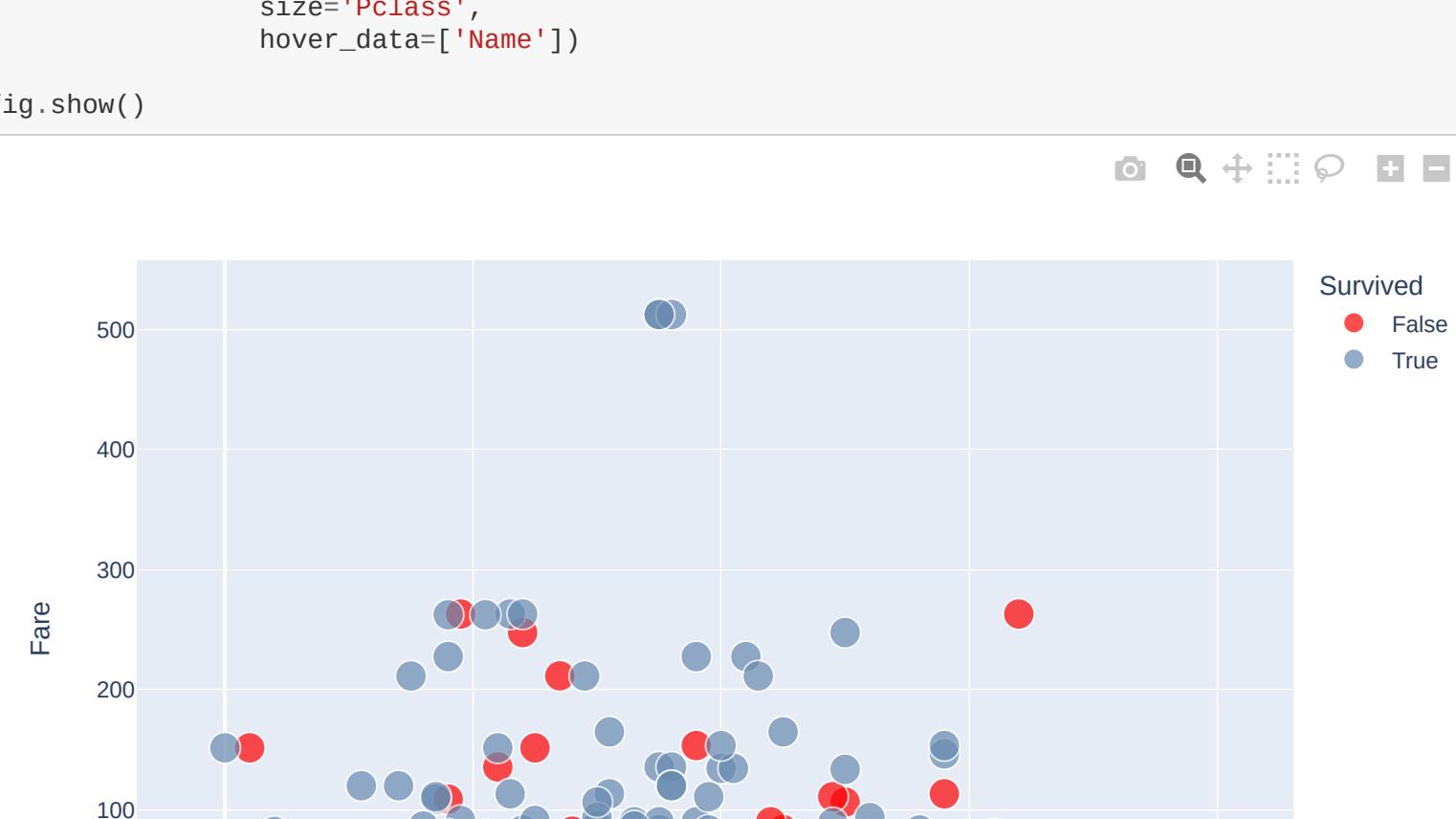
```
In [9]: pclass_fare_df = pd.pivot_table(df,index=['PassengerId'],values=['Fare'],columns=['Pclass'])
pclass_fare_df.columns = pclass_fare_df.columns.get_level_values(1)
pclass_fare_df.head()

Out[9]:
```

| Pclass      | 1       | 2   | 3     |
|-------------|---------|-----|-------|
| PassengerId |         |     |       |
| 1           | NaN     | NaN | 7.250 |
| 2           | 71.2833 | NaN | NaN   |
| 3           | NaN     | NaN | 7.925 |
| 4           | 53.1000 | NaN | NaN   |
| 5           | NaN     | NaN | 8.050 |

```
In [10]: fig = go.Figure()
for column in pclass_fare_df.columns:
    fig.add_trace(go.Box(y = pclass_fare_df[column].values, name = column, boxpoints=False))
fig.layout['title'].update(
    title='Fare paid by each passenger class',
    xaxis=dict(title='Passenger Class'),
    yaxis=dict(title='Fare'),
)
fig.show()
```

Fare paid by each passenger class



### 5. Graph the number of passengers who survived on a stacked bar based on gender

```
In [11]: survivors_df = df.loc[df['Survived'] == 1]
survivors_df

Out[11]:
```

| PassengerId | Survived | Pclass | Name  | Sex    | Age | SibSp | Parch | Ticket           | Fare    | Cabin | Embarked |
|-------------|----------|--------|---|--------|-----|-------|-------|------------------|---------|-------|----------|
| 0           | 1        | 0      | Cumings, Mrs. John Bradley (Florence Briggs Th... | female | 38  | 1     | 0     | PC 17599         | 71.2833 | C85   | C        |
| 1           | 2        | 1      | Heikkinen, Miss. Laina                            | female | 26  | 0     | 0     | STON/O2. 3101282 | 7.9250  | NaN   | S        |
| 2           | 3        | 1      | Futrelle, Mrs. Jacques Heath (Lily May Peel)      | female | 35  | 1     | 0     | 113803           | 53.1000 | C123  | S        |
| 3           | 4        | 1      | Graham, Mrs. Thomas Jr (Lily Alexenia Wilson)     | female | 27  | 0     | 2     | 347742           | 11.1333 | NaN   | S        |
| 4           | 5        | 1      | Nasser, Mrs. Nicholas (Adele Achen)               | female | 14  | 1     | 0     | 237736           | 30.0708 | NaN   | C        |
| ...         | ...      | ...    | ...   | ...    | ... | ...   | ...   | ...              | ...     | ...   | ...      |
| 875         | 1        | 3      | Najib, Miss. Adele Kiamie "Jane"                  | female | 15  | 0     | 0     | 2667             | 7.2250  | NaN   | C        |
| 876         | 1        | 1      | Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)     | female | 56  | 0     | 1     | 11767            | 83.1583 | C50   | C        |
| 877         | 1        | 2      | Shelley, Mrs. William (Manoria Parish Hall)       | female | 25  | 0     | 1     | 230433           | 26.0000 | NaN   | S        |
| 878         | 1        | 1      | Graham, Miss. Margaret Edith                      | female | 19  | 0     | 0     | 112053           | 30.0000 | B42   | S        |
| 879         | 1        | 1      | Behr, Mr. Karl Howell                             | male   | 26  | 0     | 0     | 111369           | 30.0000 | C148  | C        |

742 rows × 12 columns

```
In [12]: male_survivors = survivors_df[survivors_df['Sex']=='male']['Sex'].value_counts()
male_survivors

Out[12]:
```

| Sex  | count |
|------|-------|
| male | 109   |

```
In [13]: female_survivors = survivors_df[survivors_df['Sex']=='female']['Sex'].value_counts()
female_survivors

Out[13]:
```

| Sex    | count |
|--------|-------|
| female | 233   |

```
In [14]: plot = go.Figure()
plot.add_trace(go.Bar(
    name = 'Men',
    x = ['Survived'],
    y = male_survivors.values,
))
plot.add_trace(go.Bar(
    name = 'Women',
    x = ['Survived'],
    y = female_survivors.values,
))
plot.update_layout(barmode='stack',title=dict(text='Survivors'))
plot.show()
```

Survivors



### 6. Create a scatter plot of the ages and fare paid based on gender

```
In [15]: ages_fare_gender_df = df[['PassengerId','Sex','Fare','Age']]
ages_fare_gender_df = ages_fare_gender_df.dropna(subset=['Age'])
ages_fare_gender_df
```

```
Out[15]:
```

| PassengerId | Sex | Fare    | Age |
|-------------|-----|---------|-----|
| 0           | 1   | 7.2500  | 22  |
| 1           | 2   | 71.2833 | 38  |
| 2           | 3   | 7.9250  | 26  |
| 3           | 4   | 53.1000 | 35  |
| 4           | 5   | 8.0500  | 35  |
| ...         | ... | ...     | ... |
| 885         | 6   | 29.1250 | 39  |
| 886         | 7   | 13.0000 | 27  |
| 887         | 8   | 30.0000 | 19  |
| 888         | 9   | 30.0000 | 26  |
| 889         | 10  | 7.7500  | 32  |

714 rows × 4 columns

```
In [16]: fig = px.scatter(ages_fare_gender_df,x='Age',y='Fare',color='Sex')
fig.show()
```



```
In [17]: ages_fare_survived_df = df[['Name','Survived','Fare','Age','Pclass']]
ages_fare_survived_df = ages_fare_survived_df.dropna(subset=['Age'])
ages_fare_survived_df
```

```
Out[17]:
```

| Name | Survived | Fare    | Age | Pclass |
|------|----------|---------|-----|--------|
| 0    | 1        | 7.2500  | 22  | 3      |
| 1    | 2        | 71.2833 | 38  | 1      |
| 2    | 3        | 7.9250  | 26  | 3      |
| 3    | 4        | 53.1000 | 35  | 1      |
| 4    | 5        | 8.0500  | 35  | 3      |
| ...  | ...      | ...     | ... | ...    |
| 885  | 6        | 29.1250 | 39  | 3      |
| 886  | 7        | 13.0000 | 27  | 2      |
| 887  | 8        | 30.0000 | 19  | 1      |
| 888  | 9        | 30.0000 | 26  | 1      |
| 889  | 10       | 7.7500  | 32  | 3      |

714 rows × 5 columns