

Visualizing the Titanic Disaster

Introduction:

This exercise is based on the titanic Disaster dataset avaiable at [Kaggle](#).
To know more about the variables check [here](#)

Step 1. Import the necessary libraries

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

%matplotlib inline
```

Step 2. Import the dataset from this [address](#)

Step 3. Assign it to a variable titanic

```
In [2]: titanic = pd.read_csv(r'C:\Users\Yasin\Desktop\Machine Excercise\pandas_exercises-master\07_Visualization\Titanic_Desaster\train.csv')
titanic.head()
```

```
Out[2]:
```

	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

Step 4. Set PassengerId as the index

```
In [3]: titanic.set_index('PassengerId')
```

```
Out[3]:
```

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

891 rows × 11 columns

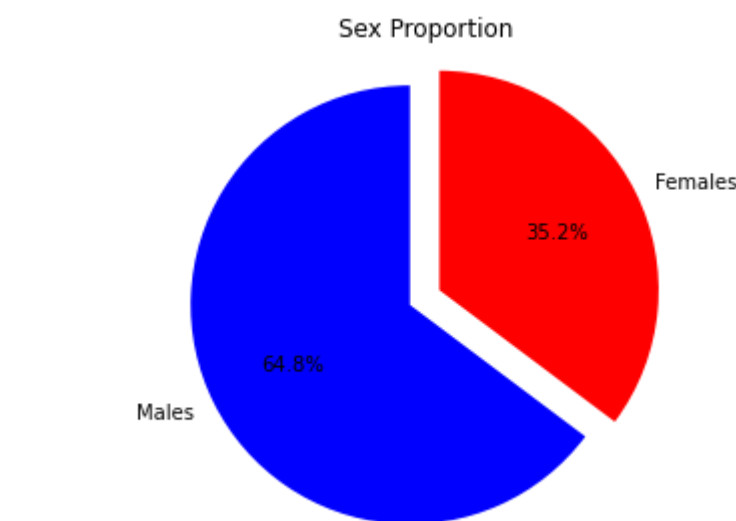
Step 5. Create a pie chart presenting the male/female proportion

```
In [4]: # sum the instances of males and females
males = (titanic['Sex'] == 'male').sum()
females = (titanic['Sex'] == 'female').sum()

# put them into a list called proportions
proportions = [males, females]

# Create a pie chart
plt.pie(x=proportions, labels = ['Males', 'Females'], shadow = False,
       colors = ['blue','red'],
       # with one slide exploded out
       explode = (0.15 , 0),
       # with the start angle at 90%
       startangle = 90,
       # with the percent listed as a fraction
       autopct = '%1.1f%%'
       )

# View the plot drop above
plt.axis('equal')
# Set labels
plt.title("Sex Proportion")
# View the plot
plt.tight_layout()
plt.show()
```



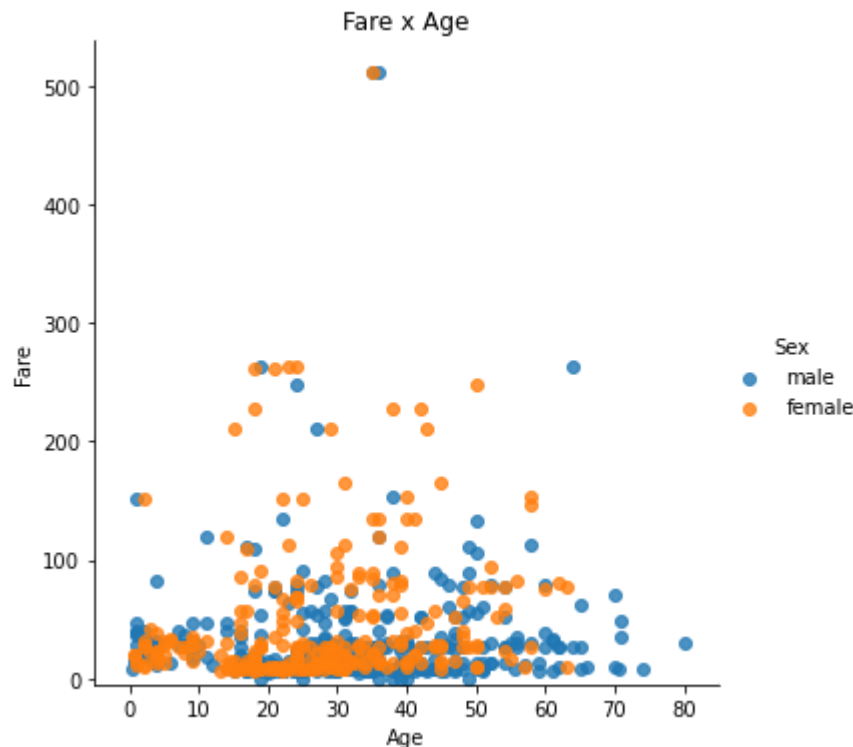
Step 6. Create a scatterplot with the Fare paid and the Age, differ the plot color by gender

```
In [5]: # creates the plot using
lm = sns.lmplot(x = 'Age', y = 'Fare', data = titanic, hue = 'Sex', fit_reg=False)

# set title
lm.set(title = 'Fare x Age')

# get the axes object and tweak it
axes = lm.axes
axes[0,0].set_ylim(-5,)
axes[0,0].set_xlim(-5,85)
```

```
Out[5]: (-5.0, 85.0)
```



Step 7. How many people survived?

```
In [6]: titanic.Survived.sum()
```

```
Out[6]: 342
```

Step 8. Create a histogram with the Fare paid

```
In [7]: # sort the values from the top to the least value and slice the first 5 items
df = titanic.Fare.sort_values(ascending = False)
df
```

```
Out[7]:
```

258	512.3292
737	512.3292
679	512.3292
88	263.0000
27	263.0000
...	...
633	0.0000
413	0.0000
822	0.0000
732	0.0000
674	0.0000

Name: Fare, Length: 891, dtype: float64

```
In [8]: # create bins interval using numpy
binsVal = np.arange(0,600,10)
binsVal
```

```
Out[8]: array([  0,  10,  20,  30,  40,  50,  60,  70,  80,  90, 100, 110, 120,
 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250,
 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380,
 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510,
 520, 530, 540, 550, 560, 570, 580, 590])
```

```
In [9]: # create the plot
plt.figure(figsize=(12,5))
plt.hist(df, bins = binsVal)
plt.show()
```

