datascience_assignment_11.1

July 14, 2018

0.1 Visualize titanic data using matplotlib: data to use:

https://raw.githubusercontent.com/Geoyi/Cleaning-Titanic-Data/master/titanic_original.csv '

0.2 Solution:

- Import all the libaries numpy, pandas, matplotlib
- Read the data from URL provided using pandas read_csv method
- Display few records of the data

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

```
In [3]: # Read the titanic data using pandas read_csv method
        url='https://raw.githubusercontent.com/Geoyi/Cleaning-Titanic-Data/master/titanic_orig
        titanic = pd.read_csv(url)

        # Display first few records of data
        titanic.head()
```

Out[3]:

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1.0	1.0	Allen, Miss. Elisabeth Walton	female	29.0000	0.0	0.0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1.0	1.0	Allison, Master. Hudson Trevor	male	0.9167	1.0	2.0	113781	151.5500	C22 C26	S	11	NaN	
2	1.0	0.0	Allison, Miss. Helen Loraine	female	2.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	
3	1.0	0.0	Allison, Mr. Hudson Joshua Creighton	male	30.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	135.0	
4	1.0	0.0	Allison, Mrs. Hudson J C (Bessie Waldo Daniels)	female	25.0000	1.0	2.0	113781	151.5500	C22 C26	S	NaN	NaN	

```
1  Montreal, PQ / Chesterville, ON
2  Montreal, PQ / Chesterville, ON
3  Montreal, PQ / Chesterville, ON
4  Montreal, PQ / Chesterville, ON
```

0.3 1. Create a pie chart presenting the male/female proportion

0.4 Steps:

- Count records by sex using pandas value_counts method and store the values in counts_by_sex
- Use labels for Male and Female
- Use colors, male with color blue and female with color red
- Use matplotlib pie method to create a pie chart using counts_by_sex, labels and colors and plot it

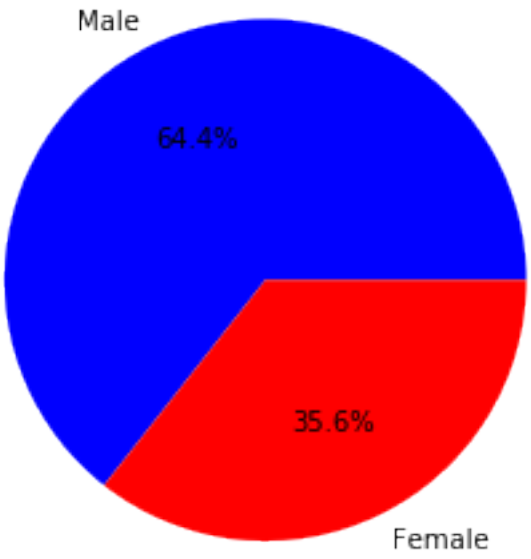
In [4]:
```
# Count records by sex using pandas value_counts method and store the values in count_
val_counts=pd.value_counts(titanic["sex"])
count_by_sex = val_counts.values
print(count_by_sex)
```

```
[843 466]
```

In [5]:
```
# Use labels for Male and Female
labels = ['Male', 'Female']

# Use colors, male with color blue and female with color red
colors = ['blue',  'red']

# Use matplotlib pie method to create a pie chart using count_by_sex, labels and color
plt.pie(count_by_sex, labels=labels, colors=colors,
        autopct='%1.1f%%',  startangle=0)
plt.axis('equal')
plt.show()
```



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

0.6 Steps:

- Get the records for whcih sex is Male and store it as male_df
- Get the records for whcih sex is Female and store it as female_df
- Create scatter plot for fare vs age for Male using male_df and color blue
- Create scatter plot for fare vs age for Male using male_df and color red

```
In [6]: # Get the records for whcih sex is Male and store it as male_df
        male_df = titanic[titanic['sex'] == 'male']

        # Get the records for whcih sex is Female and store it as female_df
        female_df = titanic[titanic['sex'] == 'female']


        plt.figure(figsize=(18,10))

        # Create scatter plot for fare vs age for Male using male_df and color blue
        plt.scatter(male_df['fare'], male_df['age'], c='blue', label='Male')

        #Create scatter plot for fare vs age for Male using female_df and color red
        plt.scatter(female_df['fare'], female_df['age'], c='red', label='Female')
        plt.title('Matplot scatter plot with the Fare paid and the Age')
        plt.xlabel('Fare')
        plt.ylabel('Age')
```

```
plt.legend(loc=2)
plt.show()
```