Lab-6

Correlation Analysis ¶

* Name: Meet Hiteshkumar Trivedi
* Student ID: N01520331

Import Libraries

In [2]:

```
import numpy as np
import pandas as pd


import matplotlib.pyplot as plt
```

In [3]:

```
import seaborn as sns
%matplotlib inline
```

Import Titanic dataset

In [4]:

```
t_df = pd.read_csv('E:\Programming\Humber college\Humber Sem 2\Data Analytics\Week-6/Titani
```

Read head of the dataset

In [5]:

```
t_df.head()
```

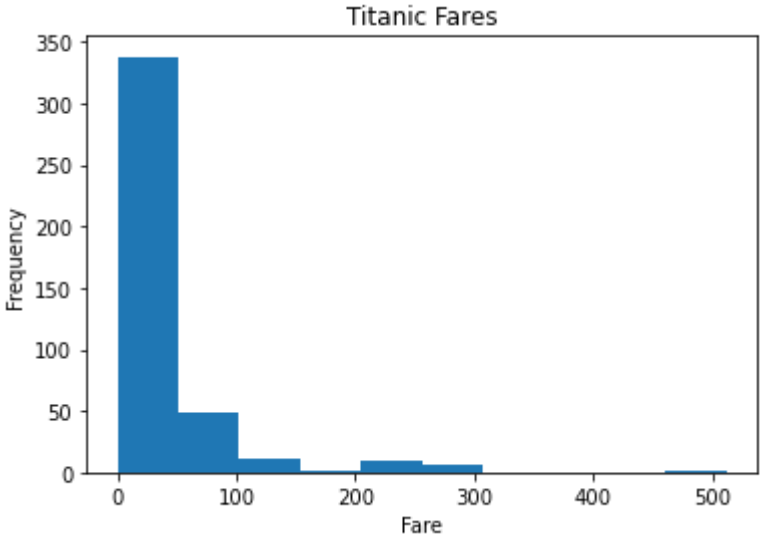
Out[5]:

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Emba
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	

Exercise 1

In [38]:

```
# CODE HERE
fare = t_df['Fare']
fare.plot(kind='hist')
plt.xlabel('Fare')
plt.title('Titanic Fares')
plt.show()
```



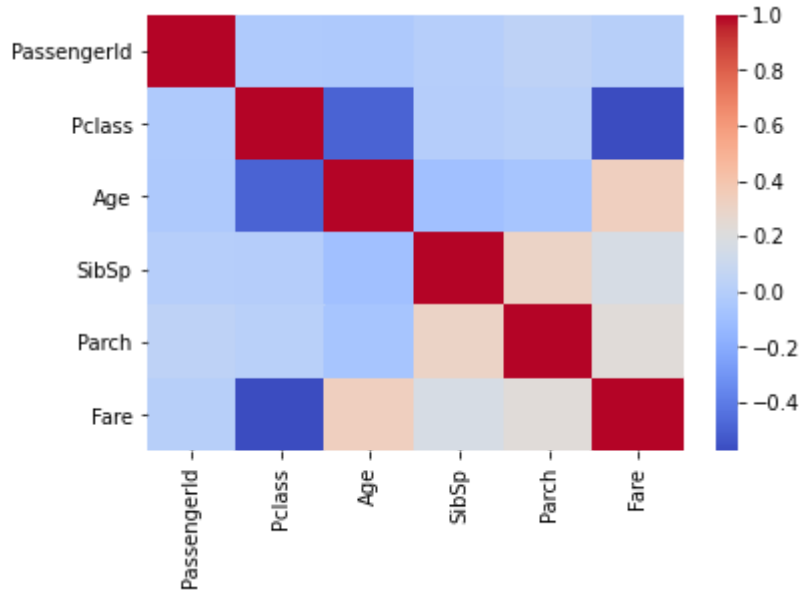
Exercise 2

In [11]:

```
# CODE HERE
sns.heatmap(t_df.corr(),cmap='coolwarm')
```

Out[11]:

```
<AxesSubplot:>
```



Exercise 3

Find "Pearson correlation" and "Spearman correlation" between "Age" and "Parch" column?

In [14]:

```
# CODE HERE
import scipy
from scipy.stats.stats import pearsonr
from scipy.stats.stats import spearmanr
```

In [26]:

```
age = t_df['Age']
parch = t_df['Parch']

age = age.fillna(t_df['Age'].mean())

parch = parch.fillna(t_df['Parch'].mean())
```

In [29]:

```
pearsonr_coefficient, p_value = pearsonr(age, parch)
print('PeasonR Correlation Coefficient %0.3f'% (pearsonr_coefficient))

# age.isnull()
```

```
PeasonR Correlation Coefficient -0.045
```

In [30]:

```
spearmanr_coefficient, p_value = spearmanr(age, parch)
print('Spearman Rank Correlation Coefficient %0.3f' % (spearmanr_coefficient))
```

```
Spearman Rank Correlation Coefficient -0.110
```

Exercise 4

Calculate the standard deviation, variance and mean of column "Fare" and "Age"

In [32]:

```
# CODE HERE
t1 = t_df[['Age', 'Fare']]

t1.std()
```

Out[32]:

```
Age     14.181209
Fare    55.907576
dtype: float64
```

In [33]:

```
t1.var()
```

Out[33]:

```
Age      201.106695
Fare    3125.657074
dtype: float64
```

In [35]:

```
t1.mean()
```

Out[35]:

```
Age     30.272590
Fare    35.627188
dtype: float64
```

In [ ]:

```

```