Lab-5

Correlation Analysis

- Name:Bhavesh Waghela
- Student ID:N01639685

Import Libraries

In [2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [6]:
```
import seaborn as sns
%matplotlib inline
```

Import Titanic dataset

In [3]:
```
data = pd.read_csv('Titanic.csv')
```

Read head of the dataset

In [4]:
```
data
```

Out[4]:

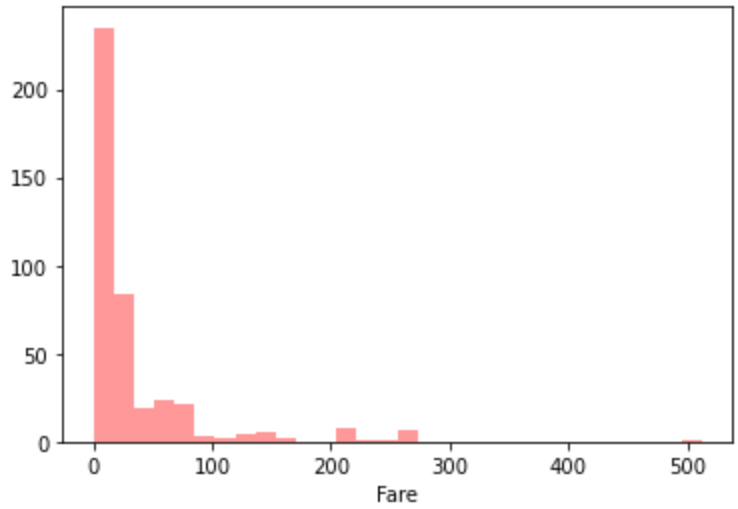
	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN
...	...	...	...	...	...	...	...	...	...	...
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN

418 rows × 11 columns

Exercise 1

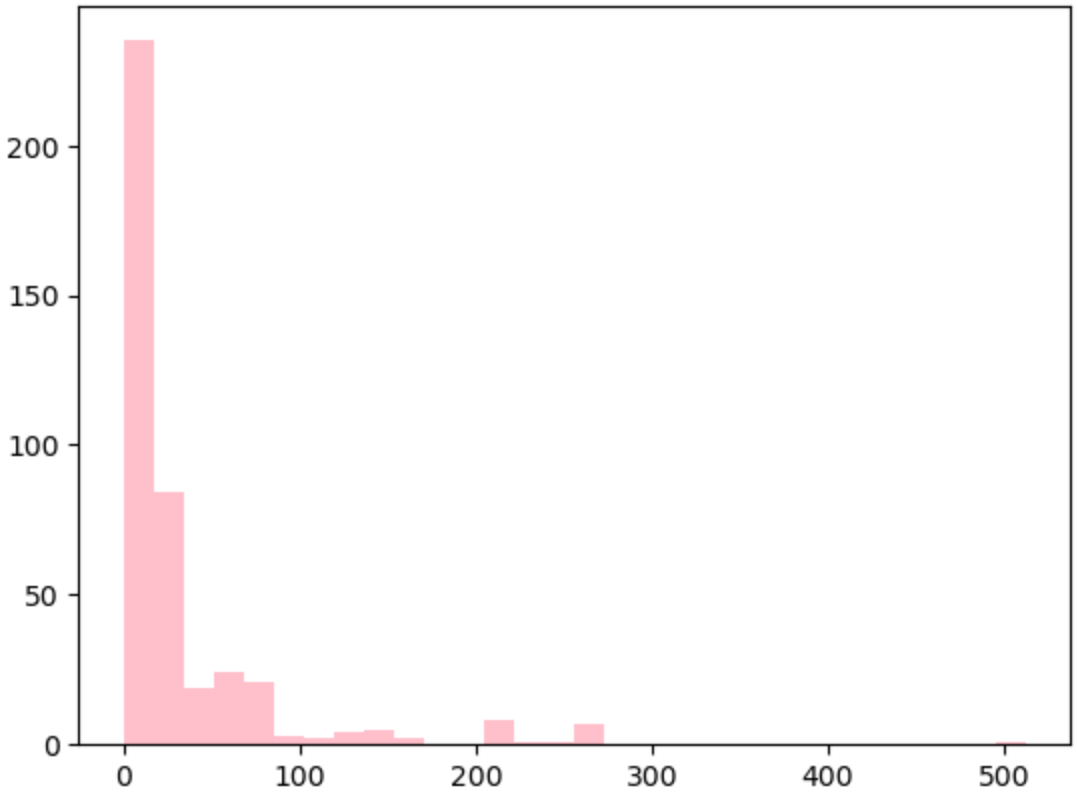
In [10]:
```
# CODE HERE
```

Out[10]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x21196b17208>
```



In [5]:
```
plt.hist(data['Fare'].values, color='pink', bins=30)
```

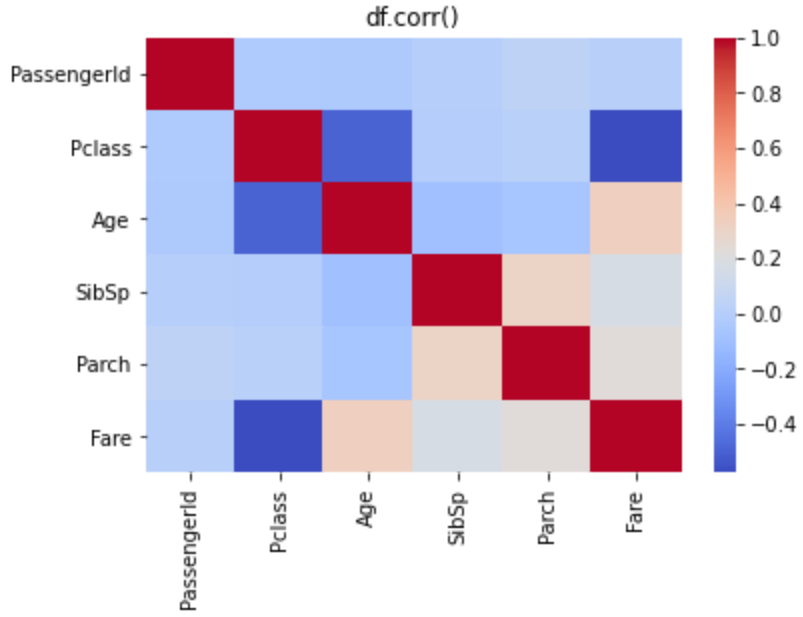
Out[5]:
```
(array([235.,  84.,  19.,  24.,  21.,   3.,   2.,   4.,   5.,   2.,   0.,
          0.,   8.,   1.,   1.,   7.,   0.,   0.,   0.,   0.,   0.,   0.,
          0.,   0.,   0.,   0.,   0.,   0.,   0.,   1.]),
 array([  0.     ,  17.07764,  34.15528,  51.23292,  68.31056,  85.3882 ,
        102.46584, 119.54348, 136.62112, 153.69876, 170.7764 , 187.85404,
        204.93168, 222.00932, 239.08696, 256.1646 , 273.24224, 290.31988,
        307.39752, 324.47516, 341.5528 , 358.63044, 375.70808, 392.78572,
        409.86336, 426.941  , 444.01864, 461.09628, 478.17392, 495.25156,
        512.3292 ]),
 <BarContainer object of 30 artists>)
```



Exercise 2

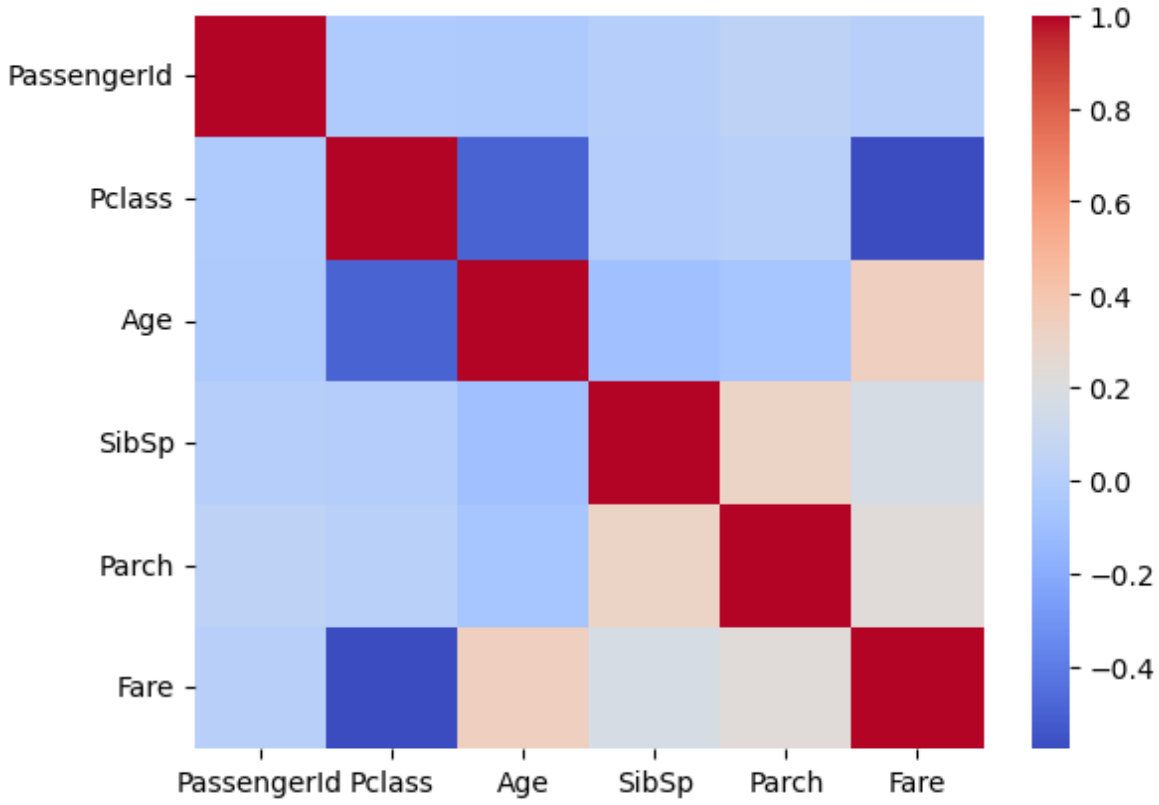
In [9]:
```
# CODE HERE
```

Out[9]: Text(0.5, 1.0, 'df.corr()')



In [8]:
```
sns.heatmap(data.corr(),cmap='coolwarm')
```

Out[8]: <AxesSubplot:>



Exercise 3

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

In [17]:
```
# CODE HERE
import scipy
from scipy.stats.stats import pearsonr
from scipy.stats.stats import spearmanr


data.dropna(subset=['Age'], inplace=True)

age = data['Age']
parch = data['Parch']

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


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

```
PeasonR Correlation Coefficient -0.061
Spearman Rank Correlation Coefficient -0.130

C:\Users\bhave\AppData\Local\Temp\ipykernel_13400\2819273084.py:3: Deprecatio
nWarning: Please use `pearsonr` from the `scipy.stats` namespace, the `scipy.
stats.stats` namespace is deprecated.
  from scipy.stats.stats import pearsonr
C:\Users\bhave\AppData\Local\Temp\ipykernel_13400\2819273084.py:4: Deprecatio
nWarning: Please use `spearmanr` from the `scipy.stats` namespace, the `scip
y.stats.stats` namespace is deprecated.
  from scipy.stats.stats import spearmanr
```

Exercise 4

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

In [23]:
```
# CODE HERE

mean_fare = data['Fare'].mean()
mean_age = data['Age'].mean()

std_dev_fare = data['Fare'].std()
std_dev_age = data['Age'].std()

variance_fare = data['Fare'].var()
variance_age = data['Age'].var()

print(f"Mean Fare: {mean_fare}")
print(f"Mean Age: {mean_age}")
print(f"\nStandard Deviation of Fare: {std_dev_fare}")
print(f"Standard Deviation of Age: {std_dev_age}")
print(f"\nVariance of Fare: {variance_fare}")
print(f"Variance of Age: {variance_age}")
```

```
Mean Fare: 40.98208731117823
Mean Age: 30.272590361445783

Standard Deviation of Fare: 61.22855822554924
Standard Deviation of Age: 14.18120923562442

Variance of Fare: 3748.9363423794734
Variance of Age: 201.10669538455937
```

In [ ]:
```

```