In [1]:

```
import numpy as np
import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
%matplotlib inline
```

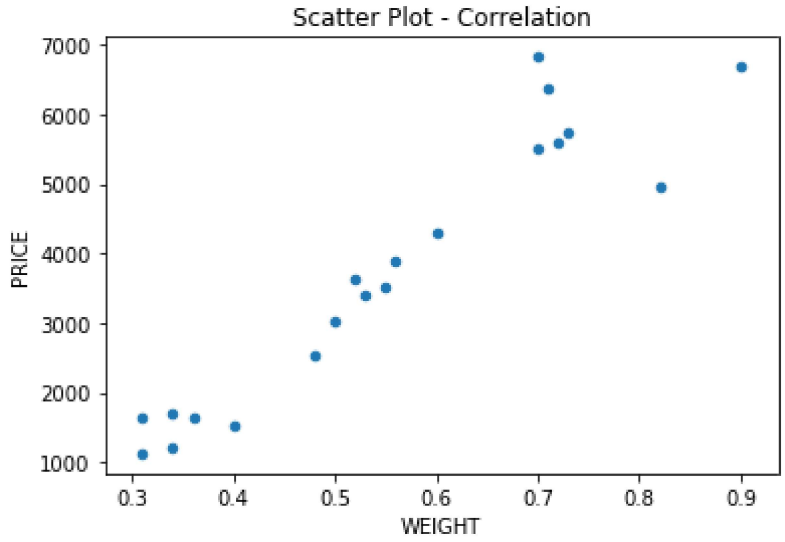
In [2]:

```
diamond = pd.read_excel('datasets/Inferential_Statistics.xlsx', sheet_name = 'Sheet1')
```

Correlation Coefficients

In [3]:

```
ScatterPlot = diamond.plot(kind='scatter',x='WEIGHT',y='PRICE',title='Scatter Plot - Correl
```



In [4]:

```
diamond['WEIGHT'].corr(diamond['PRICE']) #positive correlation
```

Out[4]:

```
0.9457713913032358
```

Correlation Coefficient Matrix

Here all the diagonal values will be 1 while correlation coefficient will be there for all the combination of the numerical variables. There are two methods of calculating Correlation Coefficient and its matrix – Pearson and Spearman.
diamond.corr(method= 'pearson')

In [5]:

```
diamond.corr(method= 'spearman')
```

Out[5]:

	IDNO.	WEIGHT	PRICE
IDNO.	1.000000	0.928115	0.921805
WEIGHT	0.928115	1.000000	0.925104
PRICE	0.921805	0.925104	1.000000

In [6]:

```
diamond.corr(method= 'pearson')
```

Out[6]:

	IDNO.	WEIGHT	PRICE
IDNO.	1.000000	0.893409	0.925019
WEIGHT	0.893409	1.000000	0.945771
PRICE	0.925019	0.945771	1.000000