ASSIGNMENT - 2

Finding the independent and dependent parameters using Correlation - 2

In [1]:
```
# importing packages pandas and numpy
import pandas as pd
import numpy as np
```

In [2]:
```
# loading the dataset
income = pd.read_csv("Income2.csv")
```

In [3]:
```
income.head()
```

Out[3]:

	Unnamed: 0	Education	Seniority	Income
0	1	21.586207	113.103448	99.917173
1	2	18.275862	119.310345	92.579135
2	3	12.068966	100.689655	34.678727
3	4	17.034483	187.586207	78.702806
4	5	19.931034	20.000000	68.009922

In [4]:
```
# removing column "unnamed: 0" since it is of no use
income = income.drop(['Unnamed: 0'], axis=1)
```

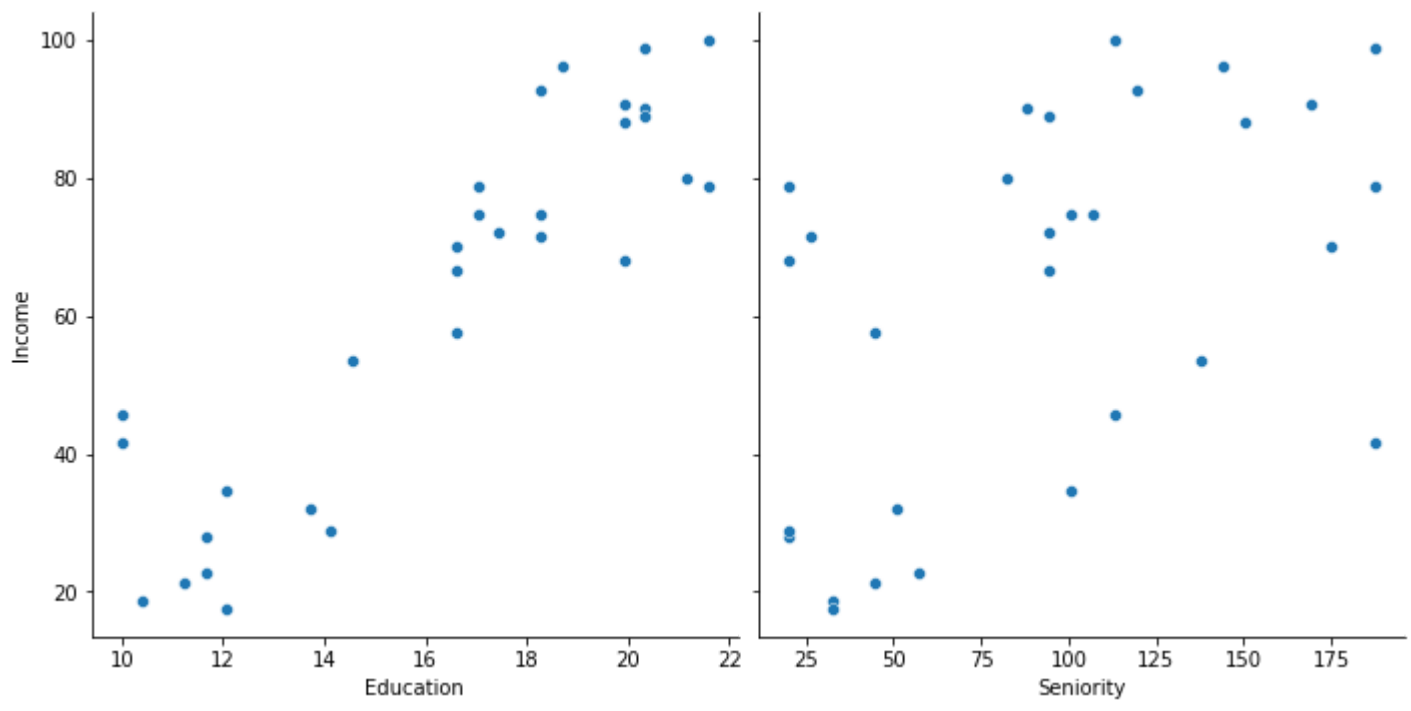
In [5]:
```
income.info()
income.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   Education  30 non-null     float64
 1   Seniority  30 non-null     float64
 2   Income     30 non-null     float64
dtypes: float64(3)
memory usage: 848.0 bytes
```

Out[5]:

	Education	Seniority	Income
0	21.586207	113.103448	99.917173
1	18.275862	119.310345	92.579135
2	12.068966	100.689655	34.678727
3	17.034483	187.586207	78.702806
4	19.931034	20.000000	68.009922

In [6]:
```
# determining the relation between Education & Seniority, Seniority & Income and Education & Income
# importing packages seaborn and matplotlib
import seaborn as sns
import matplotlib.pyplot as plt
sns.pairplot(income, x_vars=['Education', 'Seniority'], y_vars='Income', height = 5)
plt.show()
```



In [7]:
```
# visualizing the data using heatmap
sns.heatmap(income.corr(), cmap="YlGnBu", annot = True)
plt.show()
```

From the above correlation matrix, we could see that the Education is highly correlated to Income with 0.9 as its correlation value. That is, higher the education, the income would be higher. Hence Education can be considered as independent parameter in order to predict income which would be the dependent parameter.

In [8]:
```
X = income[['Education']] # taking independent parameter as X
y = income['Income'] # taking dependent parameter as y
```