Zewen Li / **Final Project** Published at Jun 29, 2021

```
import pandas as pd
df = pd.read_csv('University_Rank.csv')
df
```

	Rank	University	Country	Overall_Score	International_Student_Ratio
0	1	Massachusetts Institute of Technology (MIT)	United States	100.0	91.4
1	2	University of Oxford	United Kingdom	99.5	98.5
2	3	Stanford University	United States	98.7	67.0
3	3	University of Cambridge	United Kingdom	98.7	97.7
4	5	Harvard University	United States	98.0	70.1
5	6	California Institute of Technology (Caltech)	United States	97.4	87.7
6	7	Imperial College London	United Kingdom	97.3	100.0
7	8	ETH Zurich (Swiss Federal Institute of Technol...	Switzerland	95.4	98.2
8	8	UCL (University College London)	United Kingdom	95.4	100.0
9	10	University of Chicago	United States	94.5	84.9
10	11	National University of Singapore (NUS)	Singapore	93.9	70.3
11	12	Nanyang Technological University, Singapore (NTU)	Singapore	90.8	69.6
12	13	University of Pennsylvania	United States	90.7	58.1
13	14	Ecole Polytechnique F��d��rale de Lausanne (EPFL)	Switzerland	90.2	100.0
14	14	Yale University	United States	90.2	69.8
15	16	University of Edinburgh	United Kingdom	89.9	99.6
16	17	Tsinghua University	Mainland China	89.0	26.3
17	18	Peking University	Mainland China	88.8	38.5
18	19	Columbia University	United States	88.7	98.4

```
df['Citations_per_Faculty'].mean()
```

```
84.67333333333335
```

```
import numpy as np
from sklearn.linear_model import LinearRegression

x = [df['International_Student_Ratio'], df['International_Faculty_Ratio'], df['Faculty_Student_Ratio'],
 df['Citations_per_Faculty'], df['Academic_Reputation'], df['Employer_Reputation']]
x = np.array(x).reshape((-1,6))
y = np.array(df['Overall_Score'])

model = LinearRegression().fit(x,y)

r_sq = model.score(x, y)
r_sq
```

```
0.4351086541063228
```

```
model.intercept_
```

```
94.52425072719008
```

```
model.coef_
```

```
array([ 0.09806118,  0.07019878, -0.04552736,  0.00383022, -0.09100006,
       -0.09265162])
```

```
y_pred = model.predict(x)
y_pred
```

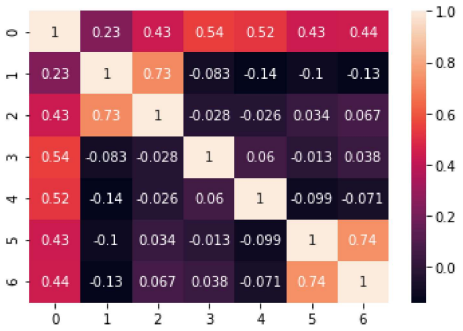
```
array([ 93.22085122,  94.15048223,  98.48477517, 100.15244116,
        92.52698222,  90.2827635 ,  88.73049989, 101.53295744,
        92.43140187,  91.7008673 ,  88.89726556,  89.62219124,
        90.77128537,  89.92371745,  90.23726057,  88.50886434,
        89.87232838,  89.29740265,  89.0105878 ,  91.493941  ,
        89.12111584,  89.45246085,  87.3871178 ,  88.96363372,
        89.76724501,  89.75114683,  90.14467705,  87.15718237,
        89.4619538 ,  86.64460034])
```

```
df
```

	Rank	U
0	1	Massachusetts Institute of Technology (M
1	2	University of Oxford
2	3	Stanford University
3	3	University of Cambridge
4	5	Harvard University
5	6	California Institute of Technology (Calt
6	7	Imperial College London
7	8	ETH Zurich (Swiss Federal Institute of T
8	8	UCL (University College London)
9	10	University of Chicago
10	11	National University of Singapore (NUS)
11	12	Nanyang Technological University, Singap
12	13	University of Pennsylvania
13	14	Ecole Polytechnique F��d��rale de Lausan
14	14	Yale University
15	16	University of Edinburgh
16	17	Tsinghua University
17	18	Peking University
18	19	Columbia University

```
del df['Rank']
del df['University']
del df['Country']
```

```
import seaborn as sns
import matplotlib.pyplot as plt
correlation_coefficients = np.corrcoef(df, rowvar=False )
sns.heatmap( correlation_coefficients, annot=True )
plt.show()
```


0 1 2 3 4 5 6
1 0.23 0.43 0.54 0.52 0.43 0.44
0.23 1 0.73 -0.083 -0.14 -0.1 -0.13
0.43 0.73 1 -0.028 -0.026 0.034 0.067
0.54 -0.083 -0.028 1 0.06 -0.013 0.038
0.52 -0.14 -0.026 0.06 1 -0.099 -0.071
0.43 -0.1 0.034 -0.013 -0.099 1 0.74
0.44 -0.13 0.067 0.038 -0.071 0.74 1
1.0 0.8 0.6 0.4 0.2 0.0