University_Rankings

October 12, 2025

[1]:
```
import pandas as pd
```

[2]:
```
df = pd.read_csv('qs-world-rankings-2025.csv')
```

[3]:
```
df.info
```

[3]:
```
<bound method DataFrame.info of      2025 Rank  2024 Rank
Institution Name  \
0             1          1  Massachusetts Institute of Technology (MIT)
1             2          6                        Imperial College London
2             3          3                           University of Oxford
3             4          4                             Harvard University
4             5          2                        University of Cambridge
…           …          …                                              …
1498      1401+  1201-1400                 University of Montana Missoula
1499      1401+      1401+                           University of Oradea
1500      1401+  1201-1400                      University of San Carlos
1501      1401+      1401+        University Politehnica of Timisoara, UPT
1502      1401+      1401+                   Western Washington University

     Location   Location Full Size  Academic Reputation  Employer Reputation  \
0          US   United States    M                100.0                100.0
1          UK  United Kingdom    L                 98.5                 99.5
2          UK  United Kingdom    L                100.0                100.0
3          US   United States    L                100.0                100.0
4          UK  United Kingdom    L                100.0                100.0
…       …              …  …                    …                    …
1498       US   United States    M                  3.0                  2.2
1499       RO         Romania    L                  5.6                  2.2
1500       PH     Philippines    M                  7.2                  9.4
1501       RO         Romania    L                  4.1                  4.2
1502       US   United States    L                  2.6                  2.6

      Faculty Student  Citations per Faculty  International Faculty  \
0               100.0                  100.0                   99.3
1                98.2                   93.9                  100.0
2               100.0                   84.8                   98.1
```

```
3                96.3                  100.0                  74.1
4               100.0                   84.6                 100.0
…                 …                      …                     …
1498             10.6                    6.1                   1.3
1499              4.0                    1.9                   1.5
1500              3.3                    1.8                   2.1
1501              7.2                    3.9                   1.4
1502              7.3                    3.5                   9.7

      International Students  International Research Network  \
0                       86.8                            96.0
1                       99.6                            97.4
2                       97.7                           100.0
3                       69.0                            99.6
4                       94.8                            99.3
…                        …                               …
1498                     1.9                             6.5
1499                     5.2                            34.5
1500                     2.1                             6.4
1501                     2.5                            18.6
1502                     1.6                            12.4

      Employment Outcomes  Sustainability QS Overall Score
0                   100.0            99.0              100
1                    93.4            99.7             98.5
2                   100.0            85.0             96.9
3                   100.0            84.4             96.8
4                   100.0            84.8             96.7
…                    …                …                …
1498                  3.1             1.0                -
1499                  6.2             2.3                -
1500                  9.6             1.0                -
1501                  3.9             1.1                -
1502                  1.5             1.1                -

[1503 rows x 16 columns]>
```

[4]:
```
df.head()
```

[4]:
```
   2025 Rank 2024 Rank                             Institution Name Location  \
0          1         1  Massachusetts Institute of Technology (MIT)       US
1          2         6                        Imperial College London       UK
2          3         3                           University of Oxford       UK
3          4         4                             Harvard University       US
4          5         2                        University of Cambridge       UK

    Location Full Size  Academic Reputation  Employer Reputation  \
```

```
0   United States    M                100.0                100.0
1  United Kingdom    L                 98.5                 99.5
2  United Kingdom    L                100.0                100.0
3   United States    L                100.0                100.0
4  United Kingdom    L                100.0                100.0

   Faculty Student  Citations per Faculty  International Faculty  \
0            100.0                  100.0                   99.3
1             98.2                   93.9                  100.0
2            100.0                   84.8                   98.1
3             96.3                  100.0                   74.1
4            100.0                   84.6                  100.0

   International Students  International Research Network  \
0                    86.8                             96.0
1                    99.6                             97.4
2                    97.7                            100.0
3                    69.0                             99.6
4                    94.8                             99.3

   Employment Outcomes  Sustainability QS Overall Score
0                100.0            99.0              100
1                 93.4            99.7             98.5
2                100.0            85.0             96.9
3                100.0            84.4             96.8
4                100.0            84.8             96.7
```

[5]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1503 entries, 0 to 1502
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   2025 Rank                       1503 non-null   object
 1   2024 Rank                       1482 non-null   object
 2   Institution Name                1503 non-null   object
 3   Location                        1503 non-null   object
 4   Location Full                   1503 non-null   object
 5   Size                            1503 non-null   object
 6   Academic Reputation             1503 non-null   float64
 7   Employer Reputation             1503 non-null   float64
 8   Faculty Student                 1503 non-null   float64
 9   Citations per Faculty           1503 non-null   float64
 10  International Faculty           1403 non-null   float64
 11  International Students          1445 non-null   float64
 12  International Research Network  1502 non-null   float64
```

```
 13  Employment Outcomes             1503 non-null   float64
 14  Sustainability                  1484 non-null   float64
 15  QS Overall Score                1503 non-null   object
dtypes: float64(9), object(7)
memory usage: 188.0+ KB
```

[6]:
```
df.isnull()
```

[6]:

	2025 Rank	2024 Rank	Institution Name	Location	Location Full	Size	\
0	False	False	False	False	False	False	
1	False	False	False	False	False	False	
2	False	False	False	False	False	False	
3	False	False	False	False	False	False	
4	False	False	False	False	False	False	
…	…	…	…	…	…	…	
1498	False	False	False	False	False	False	
1499	False	False	False	False	False	False	
1500	False	False	False	False	False	False	
1501	False	False	False	False	False	False	
1502	False	False	False	False	False	False	

	Academic Reputation	Employer Reputation	Faculty Student	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
…	…	…	…	
1498	False	False	False	
1499	False	False	False	
1500	False	False	False	
1501	False	False	False	
1502	False	False	False	

	Citations per Faculty	International Faculty	International Students	\
0	False	False	False	
1	False	False	False	
2	False	False	False	
3	False	False	False	
4	False	False	False	
…	…	…	…	
1498	False	False	False	
1499	False	False	False	
1500	False	False	False	
1501	False	False	False	
1502	False	False	False	

```
      International Research Network  Employment Outcomes  Sustainability  \
0                                 False                False           False
1                                 False                False           False
2                                 False                False           False
3                                 False                False           False
4                                 False                False           False
…                                   …                    …               …
1498                              False                False           False
1499                              False                False           False
1500                              False                False           False
1501                              False                False           False
1502                              False                False           False

      QS Overall Score
0                False
1                False
2                False
3                False
4                False
…                    …
1498             False
1499             False
1500             False
1501             False
1502             False

[1503 rows x 16 columns]
```

[7]:
```
print("---Missing Value Counts---")
print(df.isnull().sum())
```

```
---Missing Value Counts---
2025 Rank                          0
2024 Rank                         21
Institution Name                   0
Location                           0
Location Full                      0
Size                               0
Academic Reputation                0
Employer Reputation                0
Faculty Student                    0
Citations per Faculty              0
International Faculty            100
International Students            58
International Research Network     1
Employment Outcomes                0
Sustainability                    19
```

```
QS Overall Score                    0
dtype: int64
```

[8]:
```
missing_value = df.isnull().sum() / len(df) * 100
print("---Missing Value Percentages---")
print(missing_value.sort_values(ascending = False))
```

```
---Missing Value Percentages---
International Faculty              6.653360
International Students             3.858949
2024 Rank                          1.397206
Sustainability                     1.264138
International Research Network     0.066534
2025 Rank                          0.000000
Institution Name                   0.000000
Location                           0.000000
Location Full                      0.000000
Size                               0.000000
Academic Reputation                0.000000
Employer Reputation                0.000000
Faculty Student                    0.000000
Citations per Faculty              0.000000
Employment Outcomes                0.000000
QS Overall Score                   0.000000
dtype: float64
```

[9]:
```
median_Faculty = df['International Faculty'].median()
df['International Faculty'].fillna(median_Faculty, inplace = True)
```

[10]:
```
median_research = df['International Research Network'].median()
df['International Research Network'].fillna(median_research, inplace = True)
```

[11]:
```
print("---Column Data Types---")
print(df.dtypes)
```

```
---Column Data Types---
2025 Rank                          object
2024 Rank                          object
Institution Name                   object
Location                           object
Location Full                      object
Size                               object
Academic Reputation               float64
Employer Reputation               float64
Faculty Student                   float64
Citations per Faculty             float64
International Faculty             float64
International Students            float64
```

```
International Research Network    float64
Employment Outcomes               float64
Sustainability                    float64
QS Overall Score                   object
dtype: object
```

[12]:
```
median_Students = df['International Students'].median()
df['International Students'].fillna(median_Students, inplace = True)
```

[13]:
```
median_Sustain = df['Sustainability'].median()
df['Sustainability'].fillna(median_Sustain, inplace = True)
```

[14]:
```
df.isnull().sum()
```

[14]:
```
2025 Rank                          0
2024 Rank                         21
Institution Name                   0
Location                           0
Location Full                      0
Size                               0
Academic Reputation                0
Employer Reputation                0
Faculty Student                    0
Citations per Faculty              0
International Faculty              0
International Students             0
International Research Network     0
Employment Outcomes                0
Sustainability                     0
QS Overall Score                   0
dtype: int64
```

[15]:
```
column_to_impute = '2024 Rank'
df[column_to_impute].fillna('Unknown', inplace=True)

print(f"Missing values in '{column_to_impute}' are now filled with 'Unknown'.")
print(df[column_to_impute].value_counts(dropna=False).head())
```

```
Missing values in '2024 Rank' are now filled with 'Unknown'.
2024 Rank
1201-1400    206
1001-1200    199
1401+         78
951-1000      54
801-850       52
Name: count, dtype: int64
```

[16]:
```
df.head()
```

[16]:
```
   2025 Rank 2024 Rank                              Institution Name Location  \
0          1         1  Massachusetts Institute of Technology (MIT)        US
1          2         6                        Imperial College London       UK
2          3         3                           University of Oxford       UK
3          4         4                             Harvard University       US
4          5         2                        University of Cambridge       UK

     Location Full Size  Academic Reputation  Employer Reputation  \
0   United States    M                100.0                100.0
1  United Kingdom    L                 98.5                 99.5
2  United Kingdom    L                100.0                100.0
3   United States    L                100.0                100.0
4  United Kingdom    L                100.0                100.0

   Faculty Student  Citations per Faculty  International Faculty  \
0            100.0                  100.0                   99.3
1             98.2                   93.9                  100.0
2            100.0                   84.8                   98.1
3             96.3                  100.0                   74.1
4            100.0                   84.6                  100.0

   International Students  International Research Network  \
0                    86.8                            96.0
1                    99.6                            97.4
2                    97.7                           100.0
3                    69.0                            99.6
4                    94.8                            99.3

   Employment Outcomes  Sustainability QS Overall Score
0                100.0           99.0              100
1                 93.4           99.7             98.5
2                100.0           85.0             96.9
3                100.0           84.4             96.8
4                100.0           84.8             96.7
```

[17]:
```
df.isnull().sum()
```

[17]:
```
2025 Rank                  0
2024 Rank                  0
Institution Name           0
Location                   0
Location Full              0
Size                       0
Academic Reputation        0
Employer Reputation        0
```

```
Faculty Student                  0
Citations per Faculty            0
International Faculty            0
International Students           0
International Research Network   0
Employment Outcomes              0
Sustainability                   0
QS Overall Score                 0
dtype: int64
```

[18]:
```
import numpy as np
```

[19]:
```
column_name = 'QS Overall Score'
print(f"Original d-type: {df[column_name].dtype}")
print(f"Example Values: {df[column_name].unique()[:5]}")
```

```
Original d-type: object
Example Values: ['100' '98.5' '96.9' '96.8' '96.7']
```

[20]:
```
df[column_name] = df[column_name].str.replace('*', '', regex=False)
```

[21]:
```
df[column_name] = df[column_name].str.replace(r'[^0-9\.]', '', regex=True)
```

[22]:
```
df[column_name] = df[column_name].str.strip()
```

[23]:
```
df[column_name] = pd.to_numeric(df[column_name], errors='coerce')
```

[24]:
```
print(f"\nNew d-type: {df[column_name].dtype}")
print(f"NaN count after conversion: {df[column_name].isnull().sum()}")
```

```
New d-type: float64
NaN count after conversion: 903
```

[25]:
```
median_score = df[column_name].median()
```

[26]:
```
df[column_name].fillna(median_score, inplace=True)
```

[27]:
```
print(f"\nNaN count after imputation: {df[column_name].isnull().sum()}")
```

```
NaN count after imputation: 0
```

[34]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1503 entries, 0 to 1502
Data columns (total 17 columns):
```

```
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   2025 Rank                       1503 non-null   object
 1   2024 Rank                       1503 non-null   object
 2   Institution Name                1503 non-null   object
 3   Location                        1503 non-null   object
 4   Location Full                   1503 non-null   object
 5   Size                            1503 non-null   object
 6   Academic Reputation             1503 non-null   float64
 7   Employer Reputation             1503 non-null   float64
 8   Faculty Student                 1503 non-null   float64
 9   Citations per Faculty           1503 non-null   float64
 10  International Faculty           1503 non-null   float64
 11  International Students          1503 non-null   float64
 12  International Research Network  1503 non-null   float64
 13  Employment Outcomes             1503 non-null   float64
 14  Sustainability                  1503 non-null   float64
 15  QS Overall Score                1503 non-null   float64
 16  QS Overall Score Z-Score        1503 non-null   float64
dtypes: float64(11), object(6)
memory usage: 199.7+ KB
```

[29]:
```
mean_score = np.mean(df[column_name])
print(f"\nMean {column_name}: {mean_score:.2f}")
```

```
Mean QS Overall Score: 38.54
```

[30]:
```
std_dev_score = np.std(df[column_name])
print(f"Standard Deviation {column_name}: {std_dev_score:.2f}")
```

```
Standard Deviation QS Overall Score: 12.18
```

[31]:
```
df[f'{column_name} Z-Score'] = (df[column_name] - mean_score) / std_dev_score
```

[32]:
```
print(f"\nFirst 5 Z-Scores:")
print(df[f'{column_name} Z-Score'].head())
```

```
First 5 Z-Scores:
0    5.043966
1    4.920861
2    4.789549
3    4.781342
4    4.773135
Name: QS Overall Score Z-Score, dtype: float64
```

[40]:
```
grouped = df.groupby(['Location', 'Size']).agg({
    'Academic Reputation': 'mean',
    'Employer Reputation': 'mean',
    'QS Overall Score': 'mean'})
```

[41]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1503 entries, 0 to 1502
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   2025 Rank                       1503 non-null   object
 1   2024 Rank                       1503 non-null   object
 2   Institution Name                1503 non-null   object
 3   Location                        1503 non-null   object
 4   Location Full                   1503 non-null   object
 5   Size                            1503 non-null   object
 6   Academic Reputation             1503 non-null   float64
 7   Employer Reputation             1503 non-null   float64
 8   Faculty Student                 1503 non-null   float64
 9   Citations per Faculty           1503 non-null   float64
 10  International Faculty           1503 non-null   float64
 11  International Students          1503 non-null   float64
 12  International Research Network  1503 non-null   float64
 13  Employment Outcomes             1503 non-null   float64
 14  Sustainability                  1503 non-null   float64
 15  QS Overall Score                1503 non-null   float64
 16  QS Overall Score Z-Score        1503 non-null   float64
dtypes: float64(11), object(6)
memory usage: 199.7+ KB
```

[43]:
```
df.tail()
```

[43]:

	2025 Rank	2024 Rank	Institution Name	Location	Location Full	Size	Academic Reputation	Employer Reputation
1498	1401+	1201-1400	University of Montana Missoula	US	United States	M	3.0	2.2
1499	1401+	1401+	University of Oradea	RO	Romania	L	5.6	2.2
1500	1401+	1201-1400	University of San Carlos	PH	Philippines	M	7.2	9.4
1501	1401+	1401+	University Politehnica of Timisoara, UPT	RO	Romania	L	4.1	4.2
1502	1401+	1401+	Western Washington University	US	United States	L	2.6	2.6

```
      Faculty Student  Citations per Faculty  International Faculty  \
1498             10.6                    6.1                    1.3
1499              4.0                    1.9                    1.5
1500              3.3                    1.8                    2.1
1501              7.2                    3.9                    1.4
1502              7.3                    3.5                    9.7

      International Students  International Research Network  \
1498                     1.9                             6.5
1499                     5.2                            34.5
1500                     2.1                             6.4
1501                     2.5                            18.6
1502                     1.6                            12.4

      Employment Outcomes  Sustainability  QS Overall Score  \
1498                  3.1             1.0             36.35
1499                  6.2             2.3             36.35
1500                  9.6             1.0             36.35
1501                  3.9             1.1             36.35
1502                  1.5             1.1             36.35

      QS Overall Score Z-Score
1498                  -0.17979
1499                  -0.17979
1500                  -0.17979
1501                  -0.17979
1502                  -0.17979
```

[44]:
```python
print("\n--- 3. Top 10 Countries with Ranked Universities (Distribution) ---")
country_counts = df['Location'].value_counts()
print(country_counts.head(10))

print("\n--- 4. University Region Distribution ---")
region_counts = df['Location Full'].value_counts(normalize=True) * 100
print(region_counts)
```

```
--- 3. Top 10 Countries with Ranked Universities (Distribution) ---
Location
US    197
UK     90
CN     71
JP     49
DE     48
RU     47
IN     46
```

```
KR     43
IT     42
AU     38
Name: count, dtype: int64

--- 4. University Region Distribution ---
Location Full
United States       13.107119
United Kingdom       5.988024
China (Mainland)     4.723886
Japan                3.260146
Germany              3.193613
                      …
Northern Cyprus      0.066534
Iceland              0.066534
Cyprus               0.066534
Luxembourg           0.066534
Honduras             0.066534
Name: proportion, Length: 106, dtype: float64
```

[45]:
```
score_cols = ['QS Overall Score', 'Academic Reputation', 'International␣
 ↪Research Network']
score_df = df[score_cols]

print("\n--- 5. Correlation Matrix for University Scores ---")
correlation_matrix = score_df.corr()
print(correlation_matrix)
```

```
--- 5. Correlation Matrix for University Scores ---
                              QS Overall Score  Academic Reputation  \
QS Overall Score                      1.000000             0.801470
Academic Reputation                   0.801470             1.000000
International Research Network        0.359102             0.588874

                              International Research Network
QS Overall Score                                     0.359102
Academic Reputation                                  0.588874
International Research Network                       1.000000
```

[48]:
```
z_score_col = 'QS Overall Score Z-Score'

outliers = df[np.abs(df[z_score_col]) > 3]

print(f"\n--- 6. Identifying Top Outliers (|Z-Score| > 3) ---")
print(outliers[['Institution Name', 'Location', 'QS Overall Score',␣
 ↪z_score_col]])
```

```
--- 6. Identifying Top Outliers (|Z-Score| > 3) ---
                                    Institution Name Location  \
0        Massachusetts Institute of Technology (MIT)        US
1                             Imperial College London       UK
2                                University of Oxford       UK
3                                  Harvard University       US
4                             University of Cambridge       UK
5                                 Stanford University       US
6   ETH Zurich - Swiss Federal Institute of Techno…       CH
7              National University of Singapore (NUS)       SG
8                                                 UCL       UK
9        California Institute of Technology (Caltech)       US
10                         University of Pennsylvania       US
11           University of California, Berkeley (UCB)       US
12                        The University of Melbourne       AU
13                                Peking University       CN
14  Nanyang Technological University, Singapore (NTU)       SG
15                                Cornell University       US
16                        The University of Hong Kong       HK
17                           The University of Sydney       AU
18    The University of New South Wales (UNSW Sydney)       AU
19                                Tsinghua University       CN
20                              University of Chicago       US
21                               Princeton University       US
22                                    Yale University       US
23                                     Université PSL       FR
24                              University of Toronto       CA
25                                              EPFL       CH
26                           The University of Edinburgh       UK
27                       Technical University of Munich       DE
28                                    McGill University       CA
29                  The Australian National University       AU
30                             Seoul National University       KR
31                              Johns Hopkins University       US
32                               The University of Tokyo       JP
33                                  Columbia University       US
34                         The University of Manchester       UK
35          The Chinese University of Hong Kong (CUHK)       HK
36                                  Monash University       AU
37                       University of British Columbia       CA
38                                   Fudan University       CN
39                               King's College London       UK
40                          The University of Queensland       AU
41       University of California, Los Angeles (UCLA)       US
42                             New York University (NYU)       US
43                    University of Michigan-Ann Arbor       US
44                       Shanghai Jiao Tong University       CN
```

45	Institut Polytechnique de Paris	FR
46	The Hong Kong University of Science and Techno…	HK
47	Zhejiang University	CN
48	Delft University of Technology	NL
49	Kyoto University	JP
50	Northwestern University	US
51	The London School of Economics and Political S…	UK
52	KAIST - Korea Advanced Institute of Science & …	KR
53	University of Bristol	UK

	QS Overall Score	QS Overall Score Z-Score
0	100.0	5.043966
1	98.5	4.920861
2	96.9	4.789549
3	96.8	4.781342
4	96.7	4.773135
5	96.1	4.723893
6	93.9	4.543339
7	93.7	4.526925
8	91.6	4.354578
9	90.9	4.297129
10	90.3	4.247887
11	90.1	4.231473
12	88.9	4.132989
13	88.5	4.100161
14	88.4	4.091954
15	87.9	4.050919
16	87.6	4.026298
17	87.3	4.001677
18	87.1	3.985263
19	86.5	3.936021
20	86.2	3.911400
21	85.5	3.853951
22	85.2	3.829330
23	84.7	3.788295
24	84.1	3.739053
25	83.5	3.689811
26	83.3	3.673397
27	83.2	3.665190
28	83.0	3.648776
29	82.4	3.599534
30	82.3	3.591327
31	82.1	3.574913
32	82.1	3.574913
33	82.0	3.566706
34	82.0	3.566706
35	81.3	3.509257
36	81.2	3.501050

37	81.0	3.484636
38	80.3	3.427187
39	80.2	3.418980
40	80.2	3.418980
41	79.8	3.386152
42	79.6	3.369738
43	79.0	3.320496
44	77.8	3.222012
45	77.5	3.197391
46	77.1	3.164563
47	77.1	3.164563
48	77.0	3.156356
49	76.0	3.074286
50	76.0	3.074286
51	76.0	3.074286
52	75.7	3.049665
53	75.4	3.025044

[50]:
```
top_10_universities = df.sort_values(by='QS Overall Score', ascending=False).
  ↪head(10)

mean_faculty_top_10 = top_10_universities['Faculty Student'].mean()

mean_faculty_all = df['Faculty Student'].mean()

print(f"\n--- 7. Conditional Hypothesis Test (Faculty student) ---")
print(f"Average Faculty student (Top 10): {mean_faculty_top_10:.0f}")
print(f"Average Faculty student (All Universities): {mean_faculty_all:.0f}")
```

```
--- 7. Conditional Hypothesis Test (Faculty student) ---
Average Faculty student (Top 10): 93
Average Faculty student (All Universities): 28
```

[53]:
```
top_10_universities = df.sort_values(by='QS Overall Score', ascending=False).
  ↪head(10)

# Calculate the mean Faculty Count for the Top 10
mean_faculty_top_10 = top_10_universities['International Faculty'].mean()

# Calculate the mean Faculty Count for ALL universities
mean_faculty_all = df['International Faculty'].mean()

print(f"\n--- 7. Conditional Hypothesis Test (International Faculty Count) ---")
print(f"Average International Faculty Count (Top 10): {mean_faculty_top_10:.
  ↪0f}")
```

```
print(f"Average International Faculty Count (All Universities):␣
  ↪{mean_faculty_all:.0f}")
```

```
--- 7. Conditional Hypothesis Test (International Faculty Count) ---
Average International Faculty Count (Top 10): 94
Average International Faculty Count (All Universities): 30
```

[ ]: