In [192]:
```
!pip install pypyodbc
```

```
Requirement already satisfied: pypyodbc in c:\users\prathamesh ghorpade\anaconda3\lib\site-packages (1.3.6)
Requirement already satisfied: setuptools in c:\users\prathamesh ghorpade\anaconda3\lib\site-packages (from pypyodbc) (68.0.0)
```

In [193]:
```
import pyodbc as odbc
import numpy as np
import pandas as pd
import sqlite3
import matplotlib.pyplot as plt
```

In [194]:
```
Connection_string = (

        r'DRIVER={ODBC Driver 17 for SQL Server};'
        r'SERVER=DESKTOP-HB0ANNM\SQLEXPRESS;'
        r'DATABASE=WorldUniversity;'
        r'Trusted_Connection=yes;'


)
Conn = odbc.connect(Connection_string)
print(Conn)
```

```
<pyodbc.Connection object at 0x0000015C28A1F430>
```

In [ ]:

In [195]:
```
from sqlalchemy.engine import URL
connection_url = URL.create("mssql+pyodbc", query={"odbc_connect": Connection_

from sqlalchemy import create_engine
engine = create_engine(connection_url)
```

In [196]:
```
import sqlalchemy as sa

with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT TOP 10  * FROM Worldu"), conn)
    print(df)
```

```
   world_rank                       institution         country  national_rank
\
0       673.0               University of Girona           Spain           21.0
1       674.0                   Keele University  United Kingdom           49.0
2       675.0             University of Poitiers          France           30.0
3       676.0             University of Insubria           Italy           36.0
4       677.0  Wuhan University of Technology           China           38.0
5       678.0             University of Pretoria    South Africa            5.0
6       679.0            Banaras Hindu University           India           10.0
7       680.0             University of L'Aquila           Italy           37.0
8       681.0             University of Alicante           Spain           22.0
9       682.0   American University of Beirut         Lebanon            1.0

   quality_of_education  alumni_employment  quality_of_faculty  publications
\
0                 367.0              567.0               218.0         662.0
1                 367.0              567.0               218.0         659.0
2                 255.0              567.0               148.0         636.0
3                 367.0              567.0               218.0         705.0
4                 367.0              499.0               218.0         738.0
5                 367.0              262.0               218.0         620.0
6                 243.0              260.0               218.0         754.0
7                 367.0              567.0               218.0         655.0
8                 367.0              567.0               218.0         666.0
9                 114.0              247.0               218.0         915.0

   influence  citations broad_impact  patents  score    year
0      843.0      645.0         None    756.0  44.40  2015.0
1      533.0      321.0         None    853.0  44.40  2015.0
2      470.0      645.0         None    754.0  44.39  2015.0
3      564.0      645.0         None    871.0  44.39  2015.0
4      931.0      428.0         None    763.0  44.39  2015.0
5      513.0      511.0         None    604.0  44.39  2015.0
6      904.0      511.0         None    871.0  44.39  2015.0
7      678.0      645.0         None    871.0  44.38  2015.0
8      779.0      812.0         None    262.0  44.38  2015.0
9      892.0      812.0         None    871.0  44.38  2015.0
```

```
In [197]: #1)Retrieve all columns for universities in the USA.
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text("Select * from Worldu"), conn)
              print(df)
```

```
      world_rank                     institution         country  \
0          673.0            University of Girona           Spain   
1          674.0                Keele University  United Kingdom   
2          675.0          University of Poitiers          France   
3          676.0          University of Insubria           Italy   
4          677.0  Wuhan University of Technology           China   
...          ...                             ...             ...   
2195       668.0      University of Maine, Orono             USA   
2196       669.0    Graz University of Technology         Austria   
2197       670.0                 Gifu University           Japan   
2198       671.0         University of JyvÃ¤skylÃ¤         Finland   
2199       672.0          University of Paris 13          France   

      national_rank  quality_of_education  alumni_employment  \
0              21.0                 367.0              567.0   
1              49.0                 367.0              567.0   
2              30.0                 255.0              567.0   
3              36.0                 367.0              567.0   
4              38.0                 367.0              499.0   
...             ...                   ...                ...   
2195          194.0                 345.0              567.0   
2196            8.0                 367.0              567.0   
2197           43.0                 367.0              567.0   
2198            8.0                 367.0              525.0   
2199           29.0                 367.0              567.0   

      quality_of_faculty  publications  influence  citations broad_impact  \
0                  218.0         662.0      843.0      645.0         None   
1                  218.0         659.0      533.0      321.0         None   
2                  148.0         636.0      470.0      645.0         None   
3                  218.0         705.0      564.0      645.0         None   
4                  218.0         738.0      931.0      428.0         None   
...                  ...           ...        ...        ...          ...   
2195               218.0         788.0      427.0      368.0         None   
2196               218.0         677.0      715.0      368.0         None   
2197               218.0         701.0      658.0      645.0         None   
2198               218.0         456.0      443.0      812.0         None   
2199               218.0         669.0      744.0      645.0         None   

      patents  score    year  
0       756.0  44.40  2015.0  
1       853.0  44.40  2015.0  
2       754.0  44.39  2015.0  
3       871.0  44.39  2015.0  
4       763.0  44.39  2015.0  
...       ...    ...     ...  
2195    346.0  44.40  2015.0  
2196    404.0  44.40  2015.0  
2197    403.0  44.40  2015.0  
2198    805.0  44.40  2015.0  
2199    700.0  44.40  2015.0  

[2200 rows x 14 columns]
```

In [198]:
```
#2)Find the top 10 universities with the highest scores in 2012.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 10 * FROM Worldu where year=20
    print(df)
```

```
   world_rank                           institution         country  \
0         1.0                    Harvard University             USA   
1         2.0  Massachusetts Institute of Technology             USA   
2         3.0                   Stanford University             USA   
3         4.0               University of Cambridge  United Kingdom   
4         5.0     California Institute of Technology             USA   
5         6.0                  Princeton University             USA   
6         7.0                  University of Oxford  United Kingdom   
7         8.0                       Yale University             USA   
8         9.0                   Columbia University             USA   
9        10.0     University of California, Berkeley             USA   

   national_rank  quality_of_education  alumni_employment  quality_of_faculty  
\
0            1.0                   7.0                9.0                 1.0   
1            2.0                   9.0               17.0                 3.0   
2            3.0                  17.0               11.0                 5.0   
3            1.0                  10.0               24.0                 4.0   
4            4.0                   2.0               29.0                 7.0   
5            5.0                   8.0               14.0                 2.0   
6            2.0                  13.0               28.0                 9.0   
7            6.0                  14.0               31.0                12.0   
8            7.0                  23.0               21.0                10.0   
9            8.0                  16.0               52.0                 6.0   

   publications  influence  citations broad_impact  patents   score    year  
0           1.0        1.0        1.0         None      5.0  100.00  2012.0  
1          12.0        4.0        4.0         None      1.0   91.67  2012.0  
2           4.0        2.0        2.0         None     15.0   89.50  2012.0  
3          16.0       16.0       11.0         None     50.0   86.17  2012.0  
4          37.0       22.0       22.0         None     18.0   85.21  2012.0  
5          53.0       33.0       26.0         None    101.0   82.50  2012.0  
6          15.0       13.0       19.0         None     26.0   82.34  2012.0  
7          14.0        6.0       15.0         None     66.0   79.14  2012.0  
8          13.0       12.0       14.0         None      5.0   78.86  2012.0  
9           6.0        5.0        3.0         None     16.0   78.55  2012.0  
```

In [199]:
```
#3)List universities in the United Kingdom with a score above 80 in 2013.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from worldu where coun
    print(df)
```

```
               institution
0     University of Oxford
1  University of Cambridge
```

In [200]:

```
#4)Count the number of universities in each country.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT country, count (institution) from
    print(df)
```

```
                 country
0              Argentina    7
1              Australia   58
2                Austria   24
3                Belgium   20
4                 Brazil   36
5               Bulgaria    2
6                 Canada   72
7                  Chile    8
8                  China  167
9               Colombia    4
10               Croatia    2
11                Cyprus    2
12        Czech Republic   10
13               Denmark   12
14                 Egypt    8
15               Estonia    2
16               Finland   20
17                France  109
18               Germany  115
19                Greece   14
20             Hong Kong   12
21               Hungary   12
22               Iceland    2
23                 India   31
24                  Iran   16
25               Ireland   16
26                Israel   22
27                 Italy   96
28                 Japan  159
29               Lebanon    2
30             Lithuania    2
31              Malaysia    6
32                Mexico    4
33           Netherlands   29
34           New Zealand   12
35                Norway   12
36                Poland   18
37              Portugal   14
38           Puerto Rico    2
39               Romania    3
40                Russia    9
41          Saudi Arabia    8
42                Serbia    2
43             Singapore    5
44       Slovak Republic    2
45              Slovenia    4
46          South Africa   10
47           South Korea   72
48                 Spain   81
49                Sweden   24
50           Switzerland   26
51                Taiwan   46
52              Thailand    6
53                Turkey   20
54                Uganda    2
55  United Arab Emirates    2
56        United Kingdom  144
57               Uruguay    2
58                   USA  573
```

In [201]:

```
#5) Calculate the average score for universities in each country in 2014.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT country, AVG (score) from worldu w
    print(df)
```

```
                 country
0              Argentina  44.732500
1              Australia  46.050741
2                Austria  45.291667
3                Belgium  47.256000
4                 Brazil  44.929444
5               Bulgaria  44.480000
6                 Canada  47.287500
7                  Chile  44.862500
8                  China  45.109286
9               Colombia  44.620000
10               Croatia  44.920000
11                Cyprus  44.320000
12        Czech Republic  44.842000
13               Denmark  48.350000
14                 Egypt  44.365000
15               Estonia  44.940000
16               Finland  45.862222
17                France  46.332600
18               Germany  46.654909
19                Greece  44.984286
20             Hong Kong  46.556667
21               Hungary  44.795000
22               Iceland  45.050000
23                 India  44.888667
24                  Iran  44.376250
25               Ireland  45.353750
26                Israel  52.138571
27                 Italy  45.580426
28                 Japan  46.873649
29               Lebanon  44.930000
30             Lithuania  44.460000
31              Malaysia  45.310000
32                Mexico  45.285000
33           Netherlands  48.412308
34           New Zealand  45.300000
35                Norway  46.596000
36                Poland  44.653333
37              Portugal  45.235714
38           Puerto Rico  44.290000
39               Romania  44.320000
40                Russia  49.096667
41          Saudi Arabia  44.730000
42                Serbia  44.510000
43             Singapore  51.435000
44       Slovak Republic  44.600000
45              Slovenia  44.740000
46          South Africa  46.700000
47           South Korea  46.370000
48                 Spain  45.133171
49                Sweden  48.251818
50           Switzerland  51.661111
51                Taiwan  45.061600
52              Thailand  45.313333
53                Turkey  44.628000
54                Uganda  44.400000
55  United Arab Emirates  44.360000
56        United Kingdom  48.445937
57               Uruguay  44.350000
58                   USA  50.642489
```

In [202]:

```
#6) Find universities with a quality_of_education score greater than 20.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where qual
    print(df)
```

```
                         institution
0               University of Girona
1                   Keele University
2             University of Poitiers
3             University of Insubria
4     Wuhan University of Technology
...                              ...
2121      University of Maine, Orono
2122   Graz University of Technology
2123                 Gifu University
2124       University of JyvÃ¤skylÃ¤
2125          University of Paris 13

[2126 rows x 1 columns]
```

In [203]:

```
#7) Retrieve universities with a score between 70 and 80.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where scor
    print(df)
```

```
                                        institution
0                                   Yale University
1                               Columbia University
2               University of California, Berkeley
3                             University of Chicago
4                                Cornell University
5                        University of Pennsylvania
6                             University of Chicago
7                                   Yale University
8                California Institute of Technology
9                        University of Pennsylvania
10                               Cornell University
11                              University of Tokyo
12                       University of Pennsylvania
13            University of California, Los Angeles
14                                 Kyoto University
15                              New York University
16  Swiss Federal Institute of Technology in Zurich
17                         Johns Hopkins University
18                              University of Tokyo
19                       University of Pennsylvania
20            University of California, Los Angeles
21                         Johns Hopkins University
```

In [204]:
```
#8) List the top 5 universities with the highest alumni employment scores in 2
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT  top 5 institution from Worldu whe
    print(df)
```

```
            institution
0  Ã‰cole Polytechnique
1       Mines ParisTech
2    Harvard University
3   Stanford University
4     Dartmouth College
```

In [205]:
```
#9) Find the university with the highest quality_of_faculty in 2013.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 1 institution from Worldu wher
    print(df)
```

```
               institution
0  Northwestern University
```

In [206]:
```
#10)    Count the number of universities that have a national_rank less than 5
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT count (*) as institution from Worl
    print(df)
```

```
   institution
0          448
```

In [207]:
```
#11)    Retrieve universities with a quality_of_education rank equal to 1.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 1 institution from Worldu  ord
    print(df)
```

```
        institution
0  Keele University
```

In [208]:
```
#12)    List the top 10 universities with the highest citations in 2014.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 10 institution from Worldu whe
    print(df)
```

```
                            institution
0        National Cheng Kung University
1                    Leipzig University
2  Indian Institute of Technology Delhi
3                 University of Navarra
4                    Tianjin University
5                  Cranfield University
6             Xi'an Jiaotong University
7           Chonbuk National University
8                  King Saud University
9                     Sogang University
```

In [209]:
```
#13)    Calculate the average influence score for universities in the USA.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(influence) as avg_influence FR
    print(df)
```

```
   avg_influence
0      280.13438
```

In [210]:
```
#14)    Find universities with a broad_impact rank less than or equal to 50.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution, country, broad_impact
    print(df)
```

```
Empty DataFrame
Columns: [institution, country, broad_impact]
Index: []
```

In [211]:
```
#15)    Retrieve universities in Japan with a score greater than 60 in 2012.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where coun
    print(df)
```

```
           institution
0  University of Tokyo
1     Kyoto University
2  University of Tokyo
3     Kyoto University
4  University of Tokyo
5     Kyoto University
6  University of Tokyo
7     Kyoto University
```

In [212]:
```
#16)    List the top 5 universities with the highest patents in 2013.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 5 institution from Worldu wher
    print(df)
```

```
                               institution
0                     Princeton University
1                    University of Chicago
2                    University of Toronto
3                  University of Paris-Sud
4  University of California, Santa Barbara
```

In [213]:
```
#17)    Count the number of universities with a quality_of_faculty score betwe
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT count(institution) as COUN from Wo
    print(df)
```

```
   COUN
0    24
```

In [214]:

```
#18)    Calculate the average score for universities in the United Kingdom in
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT  AVG(score) as AVGSCR  from Worldu
    print(df)
```

```
      AVGSCR
0  48.445937
```

In [215]:

```
#19)    Find universities with a national_rank between 1 and 3 in 2012.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT  institution from Worldu where yea
    print(df)
```

```
                                          institution
0                                  Harvard University
1               Massachusetts Institute of Technology
2                                 Stanford University
3                             University of Cambridge
4                                University of Oxford
5                                 University of Tokyo
6     Swiss Federal Institute of Technology in Zurich
7                                    Kyoto University
8                       Weizmann Institute of Science
9                      Hebrew University of Jerusalem
10                            Imperial College London
11                                   Osaka University
12                              University of Toronto
13                                  McGill University
14                            University of Paris-Sud
15        Technion â€“ Israel Institute of Technology
16                 Ã‰cole normale supÃ©rieure - Paris
17                            University of Edinburgh
18                              Ã‰cole Polytechnique
19                               University of Geneva
20                     University of British Columbia
21  Swiss Federal Institute of Technology in Lausanne
22                               Karolinska Institute
23                          Seoul National University
24                        Sapienza University of Rome
25             Ruprecht Karl University of Heidelberg
26             Ludwig Maximilian University of Munich
27                                  Leiden University
28                             University of Helsinki
29                     Technical University of Munich
30                                 University of Oslo
31                           University of Queensland
32                           University of Copenhagen
33                               University of Sydney
34                                 Utrecht University
```

In [216]:

```
#20)    List universities with a citations rank less than 10.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("   SELECT Institution, country, citations
    print(df)
```

```
                              Institution         country  citations
0                      Harvard University             USA        1.0
1   Massachusetts Institute of Technology             USA        4.0
2                     Stanford University             USA        2.0
3      University of California, Berkeley             USA        3.0
4              University of Pennsylvania             USA        8.0
5                Johns Hopkins University             USA        9.0
6   University of California, Los Angeles             USA        6.0
7       University of Michigan, Ann Arbor             USA        7.0
8      University of Washington - Seattle             USA        5.0
9                      Harvard University             USA        1.0
10                    Stanford University             USA        2.0
11  Massachusetts Institute of Technology             USA        3.0
12     University of California, Berkeley             USA        4.0
13             University of Pennsylvania             USA        9.0
14  University of California, Los Angeles             USA        6.0
15               Johns Hopkins University             USA        7.0
16     University of Washington - Seattle             USA        5.0
17      University of Michigan, Ann Arbor             USA        8.0
18                     Harvard University             USA        1.0
19                    Stanford University             USA        3.0
20  Massachusetts Institute of Technology             USA        2.0
21                    Columbia University             USA        9.0
22     University of California, Berkeley             USA        3.0
23  University of California, Los Angeles             USA        7.0
24               Johns Hopkins University             USA        6.0
25      University of Michigan, Ann Arbor             USA        5.0
26     University of Washington - Seattle             USA        8.0
27                     Harvard University             USA        1.0
28                    Stanford University             USA        3.0
29  Massachusetts Institute of Technology             USA        2.0
30                   University of Oxford  United Kingdom        7.0
31     University of California, Berkeley             USA        4.0
32  University of California, Los Angeles             USA        8.0
33               Johns Hopkins University             USA        5.0
34      University of Michigan, Ann Arbor             USA        6.0
35                  University of Toronto          Canada        9.0
```

In [217]:

```
#21)    Retrieve the university with the highest alumni_employment in 2013.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("   SELECT Institution, country FROM Worldu
    print(df)
```

```
                                     Institution  country
0             California Institute of Technology      USA
1            University of California, San Diego      USA
2                 Hebrew University of Jerusalem   Israel
3       University of California, San Francisco      USA
4   University of Illinois at Urbanaâ€“Champaign      USA
..                                           ...      ...
95                    University of Pennsylvania      USA
96                          Ã‰cole Polytechnique   France
97                           University of Tokyo    Japan
98                           Stanford University      USA
99                            Harvard University      USA

[100 rows x 2 columns]
```

In [218]:

```
#22)    Find the top 5 universities with the highest publications in 2012.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("   SELECT Top 5 Institution, country FROM
    print(df)
```

```
                                       Institution  country
0                    Weizmann Institute of Science   Israel
1                           Rockefeller University      USA
2                   Hebrew University of Jerusalem   Israel
3  University of Texas Southwestern Medical Center      USA
4                       Carnegie Mellon University      USA
```

In [219]:

```
#23)    Count the number of universities in each year.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("   SELECT count(institution), year from Wo
    print(df)
```

```
           year
0  1000  2015.0
1  1000  2014.0
2   100  2013.0
3   100  2012.0
```

In [220]:

```
#24)    Calculate the average alumni_employment score for universities in the
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT AVG(alumni_employment) as avgemploy
    print(df)
```

```
   avgemployment
0     257.514834
```

In [221]:
```
#25)    Retrieve universities with a broad_impact score greater than 70.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT institution, country from Worldu wh
    print(df)
```

```
Empty DataFrame
Columns: [institution, country]
Index: []
```

In [222]:
```
#26)    List the top 10 universities with the highest influence in 2014.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT Top 10  institution, country from W
    print(df)
```

```
                        institution country
0      National Chung Cheng University  Taiwan
1     Nanjing University of Technology   China
2                 Jadavpur University   India
3                Feng Chia University  Taiwan
4  Northwestern Polytechnical University   China
5        Chung Shan Medical University  Taiwan
6          Beijing Jiaotong University   China
7      China Pharmaceutical University   China
8           Tarbiat Modares University    Iran
9  University of Pau and Pays de l'Adour  France
```

In [223]:
```
#27)    Find universities with a patents rank equal to 1.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT   institution, country from Worldu
    print(df)
```

```
                        institution country
0  Massachusetts Institute of Technology     USA
1  Massachusetts Institute of Technology     USA
2  Massachusetts Institute of Technology     USA
3  Massachusetts Institute of Technology     USA
```

In [224]:
```
#28)    Count the number of universities with a broad_impact between 20 and 30
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT COUNT(*) as num FROM Worldu WHERE b
    print(df)
```

```
   num
0    0
```

In [225]:
```
#29)    Calculate the average quality_of_faculty score for universities in the
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("  SELECT AVG(quality_of_faculty) FROM Worl
    print(df)
```

```
         
0  166.75
```

In [226]:

```
#30)    Find universities with an influence rank less than 5.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("  SELECT institution from Worldu where inf
    print(df)
```

```
                                institution
0                        Harvard University
1     Massachusetts Institute of Technology
2                       Stanford University
3   University of California, San Francisco
4                        Harvard University
5                       Stanford University
6     Massachusetts Institute of Technology
7   University of California, San Francisco
8                        Harvard University
9                       Stanford University
10    Massachusetts Institute of Technology
11       University of California, Berkeley
12                       Harvard University
13                      Stanford University
14    Massachusetts Institute of Technology
15       University of California, Berkeley
```

In [227]:

```
#31)    Retrieve universities with a score greater than 75 and a national_rank
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT   institution, country from Worldu
    print(df)
```

```
                              institution         country
0                      Harvard University             USA
1                     Stanford University             USA
2                    University of Oxford  United Kingdom
3   Massachusetts Institute of Technology             USA
4                 University of Cambridge  United Kingdom
5                     Columbia University             USA
6      University of California, Berkeley             USA
7                    Princeton University             USA
8                   University of Chicago             USA
9                         Yale University             USA
10     California Institute of Technology             USA
11             University of Pennsylvania             USA
12                    University of Tokyo           Japan
```

In [228]:

```
#32)    List the top 5 universities with the highest patents in 2014.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT Top 5 institution from Worldu where
    print(df)
```

```
                          institution
0  Ã‰cole normale supÃ©rieure - Paris
1   Lomonosov Moscow State University
2                Karolinska Institute
3                     Lund University
4                  Uppsala University
```

```
In [229]: #33)    Count the number of universities with an alumni_employment score great
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text("SELECT COUNT(institution) from Worldu wher
              print(df)
```

```
0  2049
```

```
In [230]: #34)    Calculate the average quality_of_education score for universities in J
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text("SELECT AVG(quality_of_education) as QUALIT
              print(df)
```

```
   QUALITYEDU
0  291.930818
```

```
In [231]: #35)    Find universities with a quality_of_education rank equal to 1 in 2014.
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text("SELECT institution from Worldu where quali
              print(df)
```

```
          institution
0  Harvard University
```

```
In [232]: #36)    Retrieve universities with a score greater than 80 and an alumni_emplo
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text("SELECT institution, country from Worldu wh
              print(df)
```

```
           institution country
0   Harvard University     USA
1  Stanford University     USA
2   Harvard University     USA
3  Stanford University     USA
4  University of Tokyo   Japan
5   Harvard University     USA
6  Stanford University     USA
```

```
In [233]: #37)    List the top 10 universities with the highest publications in 2013.
          with engine.begin() as conn:
              df = pd.read_sql_query(sa.text(" SELECT top 10 Institution,publications FR
              print(df)
```

```
                                 Institution  publications
0                 Hebrew University of Jerusalem         101.0
1                  Weizmann Institute of Science         101.0
2  University of Texas Southwestern Medical Center         101.0
3                          Rockefeller University         101.0
4                      Carnegie Mellon University         101.0
5                            University of Geneva         101.0
6                Ã‰cole normale supÃ©rieure - Paris         101.0
7      Technion â€“ Israel Institute of Technology         101.0
8                                 Keio University         101.0
9                         Arizona State University         101.0
```

In [234]:
```
#38)    Count the number of universities with a broad_impact score between 40
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(institution) from Worldu whe
    print(df)
```

```
   
0  0
```

In [235]:
```
#39)    Calculate the average score for universities in Australia.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(score) as AVGSCR from Worldu w
    print(df)
```

```
      AVGSCR
0  45.825517
```

In [236]:
```
#40)    Find universities with an influence rank equal to 1 in 2012.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where infl
    print(df)
```

```
          institution
0  Harvard University
```

In [237]:
```
#41)    Retrieve universities with a quality_of_faculty score greater than 15.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where Qual
    print(df)
```

```
                          institution
0                University of Girona
1                    Keele University
2              University of Poitiers
3              University of Insubria
4     Wuhan University of Technology
...                               ...
2135      University of Maine, Orono
2136   Graz University of Technology
2137                 Gifu University
2138        University of JyvÃ¤skylÃ¤
2139          University of Paris 13

[2140 rows x 1 columns]
```

In [238]:
```
#42)    List the top 5 universities with the highest alumni_employment in 2014
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT top 5 institution from Worldu wher
    print(df)
```

```
                             institution
0      University of California, San Diego
1  University of California, San Francisco
2       Ã‰cole normale supÃ©rieure - Paris
3                   Rockefeller University
4             Weizmann Institute of Science
```

In [239]:

```
#43)    Count the number of universities with a national_rank less than 20 in
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(institution) as cntinst from
    print(df)
```

```
   cntinst
0       62
```

In [240]:

```
#44)    Calculate the average citations score for universities in Germany.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(citations) as AVGCIT from Worl
    print(df)
```

```
      AVGCIT
0  364.73913
```

In [241]:

```
#45)    Find universities with a publications rank equal to 1.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution, country from Worldu w
    print(df)
```

```
          institution country
0  Harvard University     USA
1  Harvard University     USA
2  Harvard University     USA
3  Harvard University     USA
```

In [242]:

```
#46)    Retrieve universities with a broad_impact score greater than 60 and a
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution, country from Worldu w
    print(df)
```

```
Empty DataFrame
Columns: [institution, country]
Index: []
```

In [243]:

```
#47)    List the top 10 universities with the highest quality_of_education in
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 10 institution from Worldu whe
    print(df)
```

```
                                   institution
0      University of California, San Francisco
1                University of Texas at Austin
2                      Northwestern University
3                      University of Toronto
4  University of North Carolina at Chapel Hill
5            University of Washington - Seattle
6      University of California, Santa Barbara
7             University of Southern California
8              University of California, Irvine
9          University of Minnesota, Twin Cities
```

In [244]:

```
#48)    Count the number of universities with a score greater than 90.
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT Count(institution)as countt from W
    print(df)
```

```
   countt
0      23
```

In [245]:

```
#49)    Calculate the average influence score for universities in the United K
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(influence) as avginf from Worl
    print(df)
```

```
       avginf
0  347.791667
```

In [246]:

```
#50)    Find universities with a quality_of_education rank less than or equal
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT institution from Worldu where qual
    print(df)
```

```
                               institution
0                       Harvard University
1  Massachusetts Institute of Technology
2                  University of Cambridge
3     University of California, Berkeley
4                     Princeton University
```

In [247]:

```
#TASK 5(using python)
#1) What is the total number of records in the dataset?
data=pd.read_excel("D:\python projects\WorldUniversity.xlsx")
total_records = len(data)
print("Total number of records:", total_records)
```

```
Total number of records: 2200
```

In [248]:

```
#2) How many columns are there in the dataset?
df = pd.DataFrame(data)
column_names = list(df.columns)
num_columns = len(column_names)
print("Number of columns in the dataset:", num_columns)
```

```
Number of columns in the dataset: 14
```

In [249]:

```
#3) What is the datatype of the "world_rank" column?
Data_type = df["world_rank"].dtype
print("Datatype of the 'world_rank' column:", Data_type)
```

```
Datatype of the 'world_rank' column: int64
```

In [250]:

```
#4) Which country is represented the most in the dataset?
m_country = df["country"].value_counts().idxmax()
print("The country represented the most in the dataset is:", m_country)
```

```
The country represented the most in the dataset is: USA
```

In [251]:

```
#5) What is the average "quality_of_education" across all institutions?
AVG_education=df["quality_of_education"].mean()
print(AVG_education)
```

```
275.10045454545457
```

In [252]:

```
#6) Find the institution with the highest "alumni_employment" value.
MAX_AE = df["alumni_employment" ].idxmax
print(MAX_AE)
```

```
<bound method Series.idxmax of 0         9
1        17
2        11
3        24
4        29
       ...
2195    567
2196    566
2197    549
2198    567
2199    567
Name: alumni_employment, Length: 2200, dtype: int64>
```

In [253]:

```
#7) In which year does the dataset end?
End_year = df["year"].max
("Ending year",End_year)
```

Out[253]:

```
('Ending year',
 <bound method NDFrame._add_numeric_operations.<locals>.max of 0       2012
 1       2012
 2       2012
 3       2012
 4       2012
         ...
 2195    2015
 2196    2015
 2197    2015
 2198    2015
 2199    2015
 Name: year, Length: 2200, dtype: int64>)
```

In [254]:

```
#8) How many unique countries are present in the dataset?
unique_countries = df["country"].nunique()
print("Number of unique countries in the dataset:", unique_countries)
```

```
Number of unique countries in the dataset: 59
```

In [255]:

```
#9) Which institution has the highest "score" in the year 2012?
df_2012 = df[df["year"] == 2012]

max_score_institution_2012 = df_2012.loc[df_2012["score"].idxmax()]["instituti

print("The institution with the highest score in the year 2012 is:", max_score
```

```
The institution with the highest score in the year 2012 is: Harvard Universit
y
```

In [256]:

```
#10)    What is the national rank of Harvard University in the year 2012?
harvard_2012 = df[(df["institution"] == "Harvard") & (df["year"] == 2012)]
harvard_national_rank_2012 = harvard_2012["national_rank"].values[0] if not ha

print("The national rank of Harvard University in the year 2012 is:", harvard_
```

```
The national rank of Harvard University in the year 2012 is: None
```

In [257]:

```
#11)    Find the average "publications" value for institutions in the United K
uk_institutions = df[df["country"] == "UK"]
average_publications_uk = uk_institutions["publications"].mean()

print( average_publications_uk)
```

```
nan
```

In [258]:

```
#12)    Identify the institution with the highest "influence" in the year 2013
df_2013 = df[df['year'] == 2013]

highest_influence_2013 = df_2013[df_2013['influence'] == df_2013['influence'].
print(highest_influence_2013[['institution', 'influence']])
```

```
                                           institution  influence
135                            University of Paris-Sud        101
137                  Purdue University, West Lafayette        101
139                          Seoul National University        101
150                         Carnegie Mellon University        101
156                                Tel Aviv University        101
161                        Sapienza University of Rome        101
163                 Ã‰cole normale supÃ©rieure - Paris        101
165        Technion â€“ Israel Institute of Technology        101
169                                    Keio University        101
172                           Arizona State University        101
179              Texas A&M University, College Station        101
180                                 University of Oslo        101
181                                  Dartmouth College        101
182  Swiss Federal Institute of Technology in Lausanne        101
185                               Ã‰cole Polytechnique        101
186                   National University of Singapore        101
188                             Stony Brook University        101
189                  Lomonosov Moscow State University        101
190                                  Nagoya University        101
191                               University of Sydney        101
193                                    Rice University        101
194                                  Tohoku University        101
195                     Australian National University        101
196                              University of Alberta        101
199                    Georgia Institute of Technology        101
```

In [259]:

```
#13)    What is the minimum "broad_impact" value in the dataset?
min_impact = df["broad_impact"].min()
print("minimum broad impact",min_impact)
```

```
minimum broad impact 1.0
```

In [260]:
```
#14)    How many institutions are from Japan in the year 2012?
insti_count= df[(df['year'] == 2012) & (df['country'] == 'Japan')]
num_insti_count = insti_count.shape[0]
print(num_insti_count)
```

```
5
```

In [261]:
```
#15)    What is the average "patents" value for institutions in the USA in the
usa_2013 = df[(df['year'] == 2013) & (df['country'] == 'USA')]
average_patents = usa_2013['patents'].mean()
print("Average patents value for institutions in the USA in 2013:", average_pa
```

```
Average patents value for institutions in the USA in 2013: 56.75438596491228
```

In [262]:
```
#16)    Find the top 5 institutions with the highest "citations" in the year 2
df_2015 = df[df['year'] == 2015]
sorted_df = df_2015.sort_values(by='citations', ascending=False)

top_5_citations = sorted_df.head(5)

(top_5_citations[['institution', 'citations']])
```

Out[262]:

	institution	citations
2199	China Pharmaceutical University	812
2084	Manchester Metropolitan University	812
2030	University of Rouen	812
2027	Jiangnan University	812
1793	Cranfield University	812

In [263]:
```
#17)    What is the median "international_students" percentage across all inst

#there is no column is there named as international_students
```

In [264]:
```
#18)    Identify the institution with the lowest "income" in the year 2014.
#there is no column is there named as income
```

In [265]:
```
#19)    How many missing values are there in the "total_score" column?
missing_values=df['score'].isnull().sum()
print(missing_values)
```

```
0
```

In [266]:
```
#20)    Find the top 3 countries with the most institutions in the data
country_cnt=df['country'].value_counts()
top_3=country_cnt.head(3)
print(top_3)
```

```
country
USA      573
China    167
Japan    159
Name: count, dtype: int64
```

In [267]:
```
#21)    Calculate the percentage of institutions with a "research" score great
high_score = df[df['score']>70]
percentage_h = (len(high_score) / len(df)) * 100
print(percentage_h)
```

```
2.8181818181818183
```

In [268]:
```
#22)    What is the difference in "teaching" score between MIT and Stanford Un
#data is not available
```

In [269]:
```
#23)    How many institutions have a "world_rank" between 50 and 100 in the ye
df_rank = df[(df['year'] == 2011) & (df['world_rank'] >= 50) & (df['world_rank
count_institutions = len(df_rank)
print(count_institutions)
```

```
0
```

In [270]:
```
#24)    Find the country with the highest average "industry_income" across all
#no column name as industry income
```

In [271]:
```
#25)    What is the standard deviation of "research" scores for institutions i
df_usa = df[df['country'] == 'USA']
standerd_dev = df_usa['score'].std()
print(standerd_dev)
```

```
11.91542515497582
```

In [272]:
```
#26)    Identify the institution with the highest "alumni_employment" in the y
df_2015 = df[df['year'] == 2015]

highest_alumni = df_2015.loc[df_2015['alumni_employment'].idxmax()]['instituti

print(highest_alumni)
```

```
University of California, San Diego
```

In [273]:
```
#27)    Calculate the correlation between "score" and "research" across all ye
#research column not available
```

In [274]:

```
#28)    How many institutions have a "broad_impact" greater than 800 in the ye
filtered_df = df[(df['year'] == 2014) & (df['broad_impact'] > 800)]
num_institutions = len(filtered_df)
print("institution count is",num_institutions)
```

```
institution count is 182
```

In [275]:

```
#29)    Find the average "quality_of_eduacation" percentage for institutions i
country_aus = df[df['country']=='Australia']
avg_qual_edu=country_aus['quality_of_education'].mean()
print(avg_qual_edu)
```

```
287.12068965517244
```

In [276]:

```
#30)    Identify the institution with the highest "total_score" in the year 20
df_2016 = df[df['year'] == 2016]

if not df_2016.empty:

    highest_total_score_institution = df_2016.loc[df_2016['total_score'].idxma
    print(f"The institution with the highest 'total_score' in the year 2016 is
else:
    print("No data available for the year 2016.")
```

```
No data available for the year 2016.
```

In [277]:

```
#31)    What is the percentage of missing values in the "female_male_ratio" co
#data insufficiant
```

In [278]:

```
#32)    How many institutions have a "score" greater than 90 in the year 2013?
high_score_2013 = df[(df['year'] == 2013) & (df['score'] > 90)].shape[0]

print(high_score_2013)
```

```
5
```

In [279]:

```
#33)    Calculate the average "research" score for institutions in the top 10
top_10_2015 = df[df['year'] == 2015].nlargest(10, 'score')
average_top_10_2015 = top_10_2015['publications'].mean()

print(average_top_10_2015)
```

```
17.5
```

In [ ]:

```
```

In [280]:

```
#34)    Identify the institution with the lowest "international_students" perc
#no column is there in international_students
```

In [281]:

```
#35)    What is the range of "influence" values in the dataset?
inf_range=df['influence'].max()-df['influence'].min()
print(inf_range)
```

```
990
```

In [282]:

```
#36)    How many institutions have a "national_rank" of 1 in the year 2014?
count_rank_1_2014 = df[(df['year'] == 2014) & (df['national_rank'] == 1)].shap

print(f"The number of institutions with a national rank of 1 in the year 2014
```

```
The number of institutions with a national rank of 1 in the year 2014 is: 59
```

In [283]:

```
#37)    Find the country with the highest average "international_students" per
#there is no column name as international_students
```

In [284]:

```
#38)    Calculate the average "teaching" score for institutions in the top 5 o
top_5_2016 = df[(df['year'] == 2016)].nlargest(5, 'world_rank')
average_teaching = top_5_2016['score'].mean()
print("The average teaching score is",average_teaching)
```

```
The average teaching score is nan
```

In [285]:

```
#39)    Identify the institution with the highest "research" score in the year
#research not available
```

In [286]:
```
#40)    How many institutions have a "world_rank" greater than 200 in the year
World_r=df[(df['year']==2015) & (df['world_rank']>200)]
(World_r)
```

Out[286]:

	world_rank	institution	country	national_rank	quality_of_education	alumni_employm
1400	201	Cardiff University	United Kingdom	16	321	2
1401	202	George Washington University	USA	87	192	
1402	203	University of Illinois at Chicago	USA	88	367	5
1403	204	University of Erlangen-Nuremberg	Germany	13	231	5
1404	205	University of MÃ¼nster	Germany	14	113	4
...	...	...	...	...	...	
2195	996	University of the Algarve	Portugal	7	367	5
2196	997	Alexandria University	Egypt	4	236	5
2197	998	Federal University of CearÃ¡	Brazil	18	367	5
2198	999	University of A CoruÃ±a	Spain	40	367	5
2199	1000	China Pharmaceutical University	China	83	367	5

800 rows × 14 columns

In [287]:
```
#41)    What is the mode of the "country" column in the dataset?
mode_ctry=df['country'].mode()
print(mode_ctry)
```

```
0    USA
Name: country, dtype: object
```

In [288]:
```
#42)    Identify the institution with the highest "industry_income" in the yea
#industry income not available
```

In [289]:
```
#43)    Calculate the average "citations" score for institutions in the bottom
bottom_10 = df[(df['year'] == 2013)].nsmallest(10, 'world_rank')
average_bottom_10 = bottom_10['citations'].mean()

print(f"The average citations score for bottom 10",average_bottom_10)
```

```
The average citations score for bottom 10 12.3
```

In [290]:
```
#44)    How many institutions have a "total_score" greater than 80 in the year
count_high_total_score_2016 = df[(df['year'] == 2016) & (df['score'] > 80)].sh

print(count_high_total_score_2016)
```

```
0
```

In [291]:
```
#45)    Find the country with the lowest average "score" across all years.
average_score = df.groupby('country')['score'].mean()
lowest_avg_score = average_score.idxmin()

print(f"The country with the lowest average 'score' across all years is: {lowe
```

```
The country with the lowest average 'score' across all years is: Romania
```

In [292]:
```
#46)    Calculate the average "publications" percentage for institutions in th

top_5_alumni_employment_2014 = df[(df['year'] == 2014)].nlargest(5, 'alumni_em
average_international_students_top_5_2014 = top_5_alumni_employment_2014['publ

print("The average publications percentage for institutions in the top 5 of al
```

```
The average publications percentage for institutions in the top 5 of alumni e
mployment in 2014 is  196.4
```

In [293]:
```
#47)    Identify the institution with the lowest "quality_of_education" score
lowest_quality_edu = df[df['year'] == 2015].nsmallest(1, 'quality_of_education

# Print the institution with the lowest quality of education score in 2015
(lowest_quality_edu)
```

Out[293]:

	world_rank	institution	country	national_rank	quality_of_education	alumni_employment
1200	1	Harvard University	USA	1	1	1

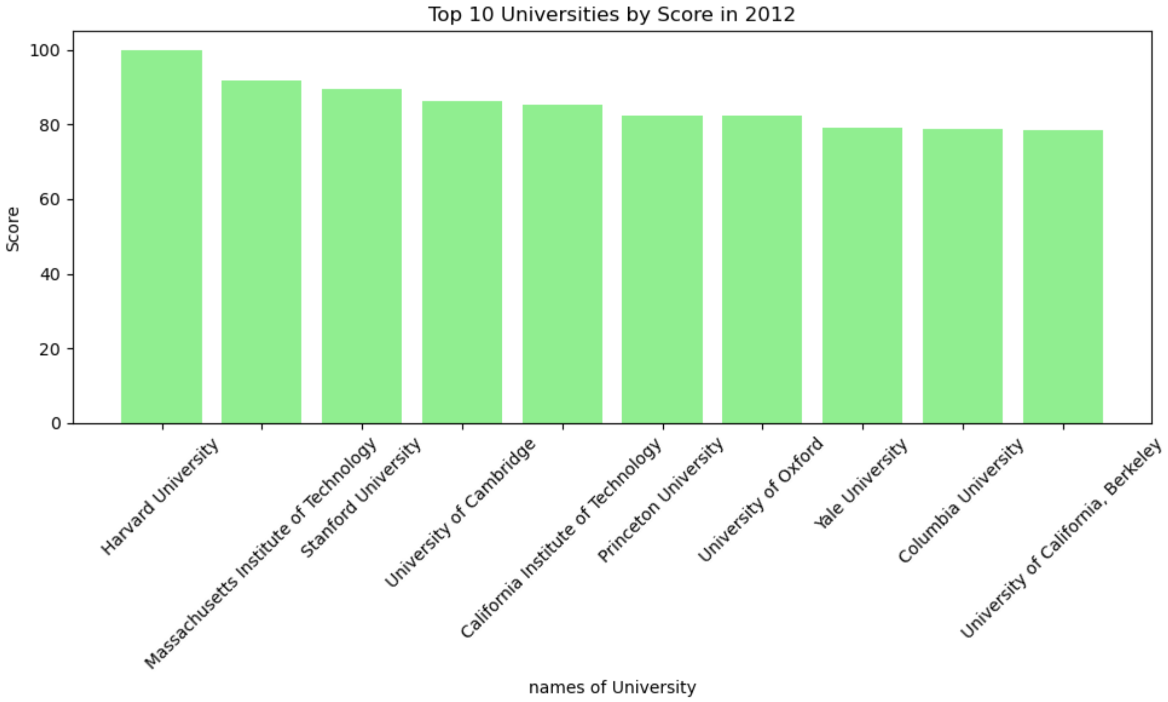
In [294]:
```
#48)    How many missing values are there in the "research" column?
#no data available
```

In [295]:
```
#49)    Find the average "female_male_ratio" for institutions in the United St
#no column is there named as female_male_ratio
```

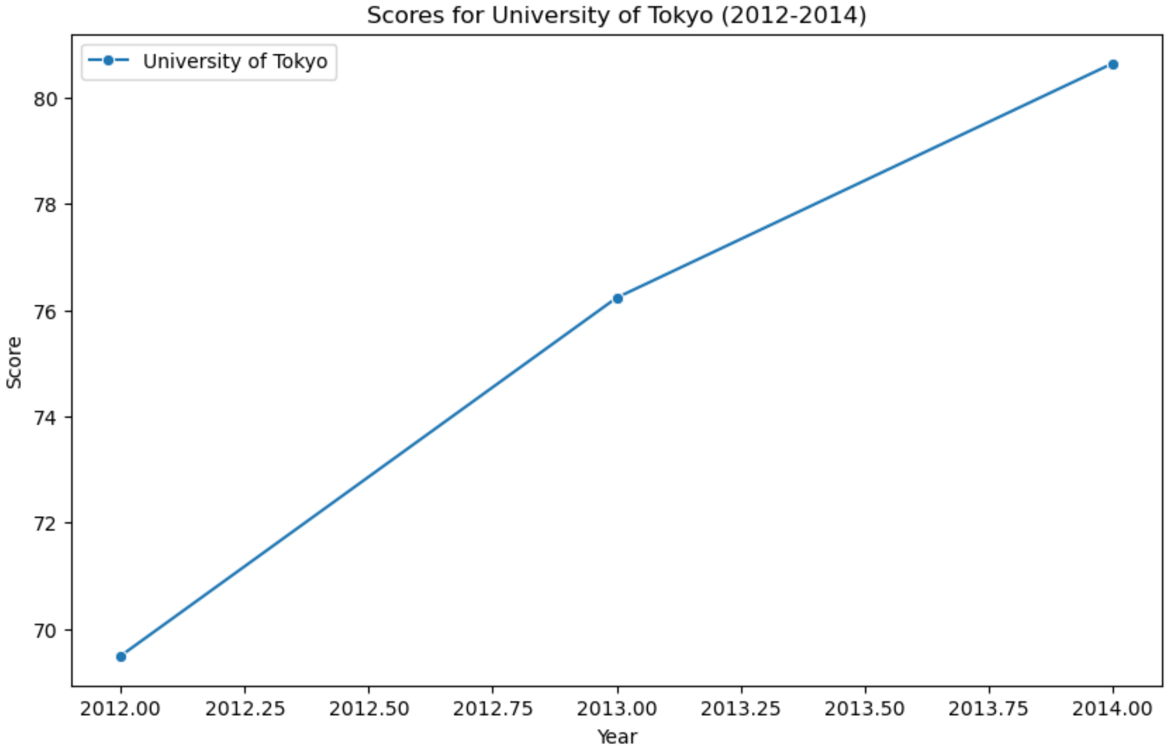
In [296]:
```
#50)    What is the highest "influence" score in the dataset?
high_influence=df['influence'].max()
(high_influence)
```

Out[296]:
```
991
```

In [297]:
```
#task 6:
#visualization:
#1) How can you use matplotlib to create a bar chart showing the top 10 univer
top_10_2012 = df[df['year'] == 2012].nlargest(10, 'score')

plt.figure(figsize=(10, 6))
plt.bar(top_10_2012['institution'], top_10_2012['score'], color='lightgreen')
plt.xlabel('names of University')
plt.ylabel('Score')
plt.title('Top 10 Universities by Score in 2012')
plt.xticks(rotation=45)
plt.tight_layout()

# Show the plot
plt.show()
```



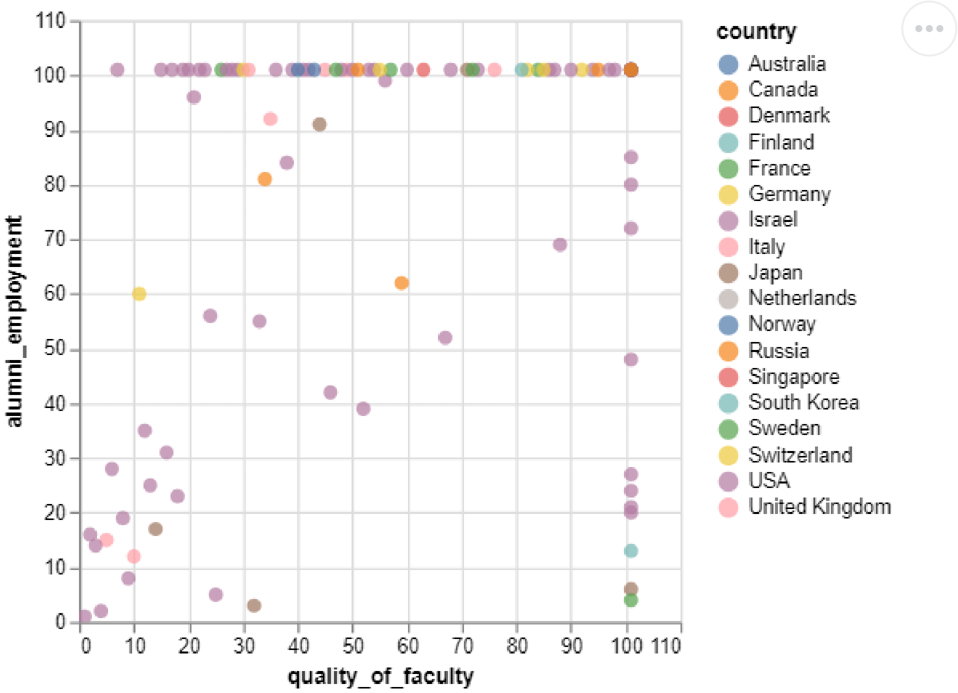
In [298]:
```
#2) Create a line plot using seaborn to visualize the trend of scores for the
import seaborn as sns
university_of_tokyo= df[(df['institution'] == 'University of Tokyo')]
subset_years = [2012, 2013, 2014]
subset_data = university_of_tokyo[university_of_tokyo['year'].isin(subset_year

# Create a line plot using Seaborn
plt.figure(figsize=(10, 6))
sns.lineplot(x='year', y='score', data=subset_data, marker='o', label='Univers
plt.xlabel('Year')
plt.ylabel('Score')
plt.title('Scores for University of Tokyo (2012-2014)')
plt.show()
```



In [299]:

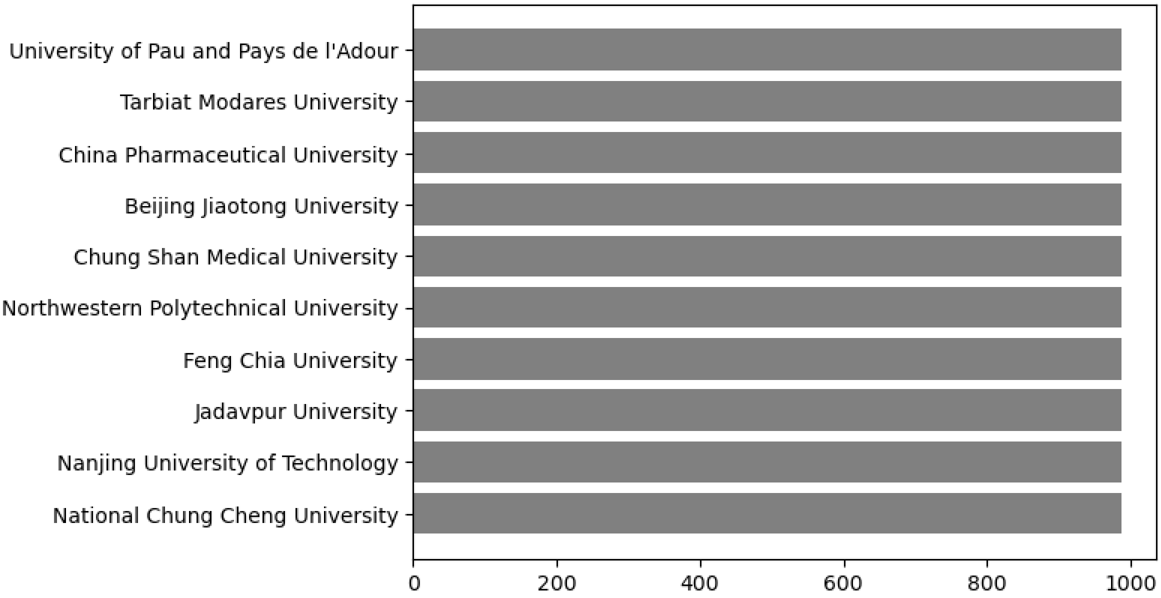
```
#3) How can you use Altair to create a scatter plot comparing the quality of f
import altair as alt
data_13= df.loc[df['year']==2013]
alt.Chart(data_13).mark_circle(size = 50).encode(
    x='quality_of_faculty',
    y='alumni_employment',color='country').interactive()
```

Out[299]:

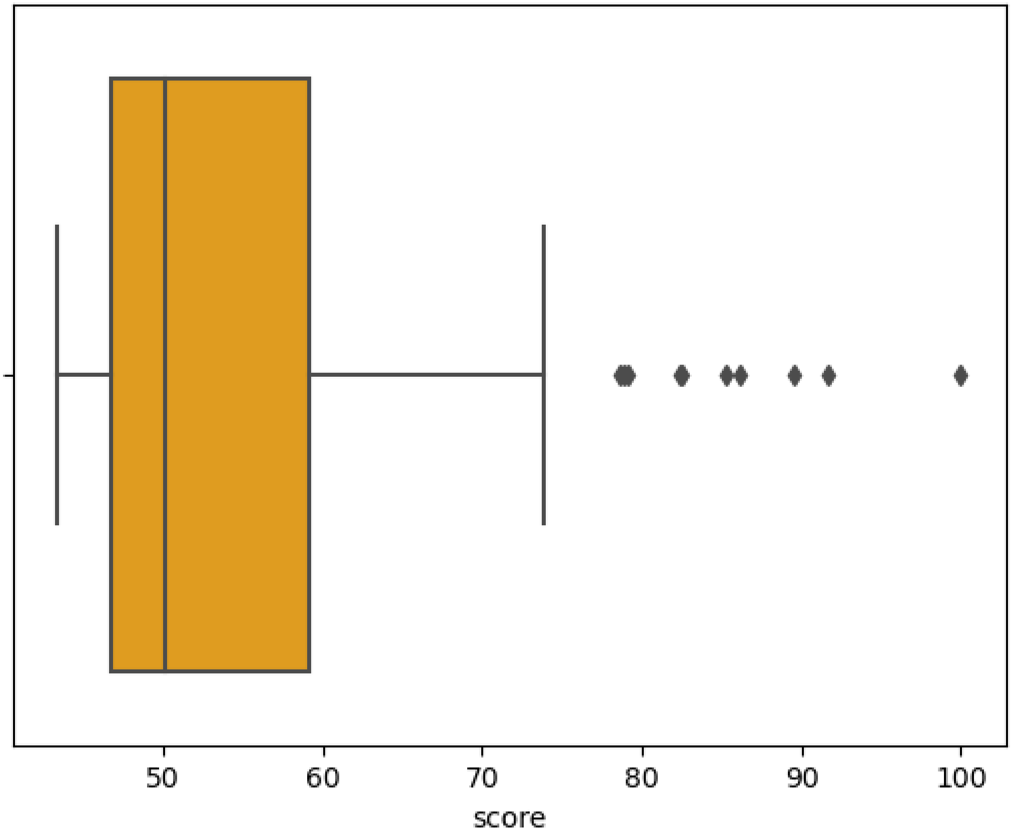


In [300]:

```
#4) Using matplotlib, create a horizontal bar chart to display the top 10 univ
data_14= df.loc[df['year']==2014]
influence_dt= data_14.nlargest(10,'influence')

inst= influence_dt['institution']
inf= influence_dt['influence']

plt.barh(inst,inf,color='grey')
plt.show()
```



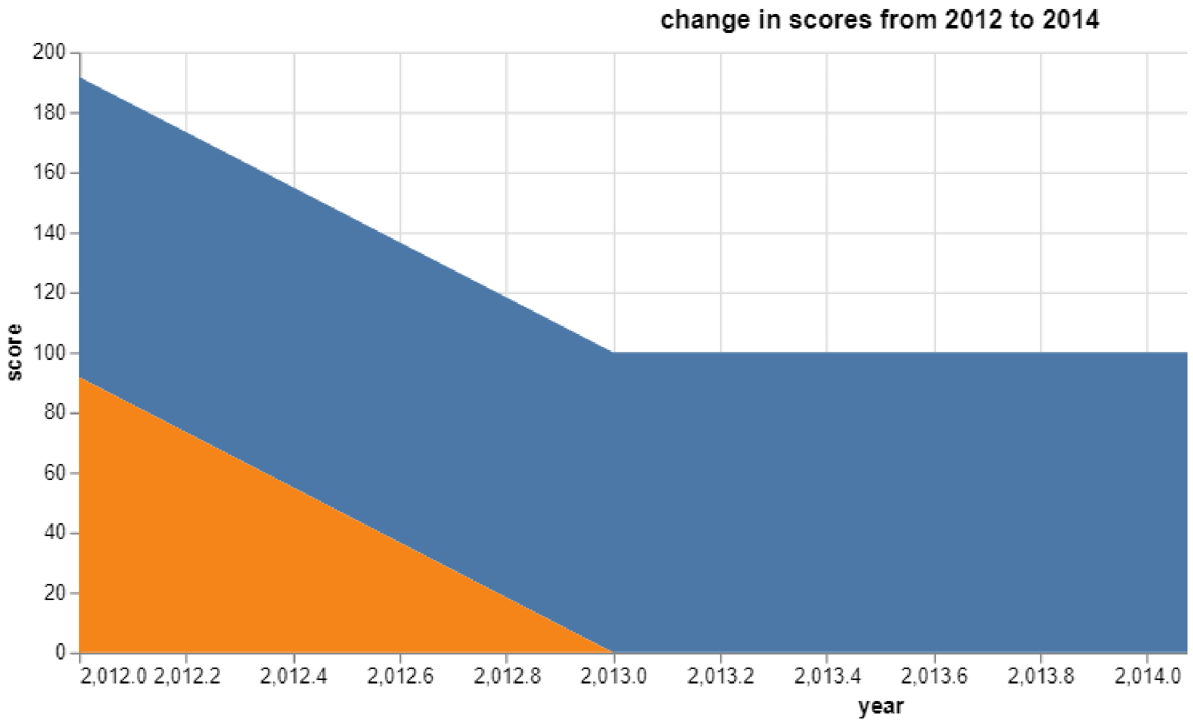
In [301]:
```
#5) How can you use seaborn to create a boxplot for the distribution of scores
data_2012= df.loc[df['year'] == 2012]
sns.boxplot(x='score',data = data_2012,color='orange')
plt.show()
```



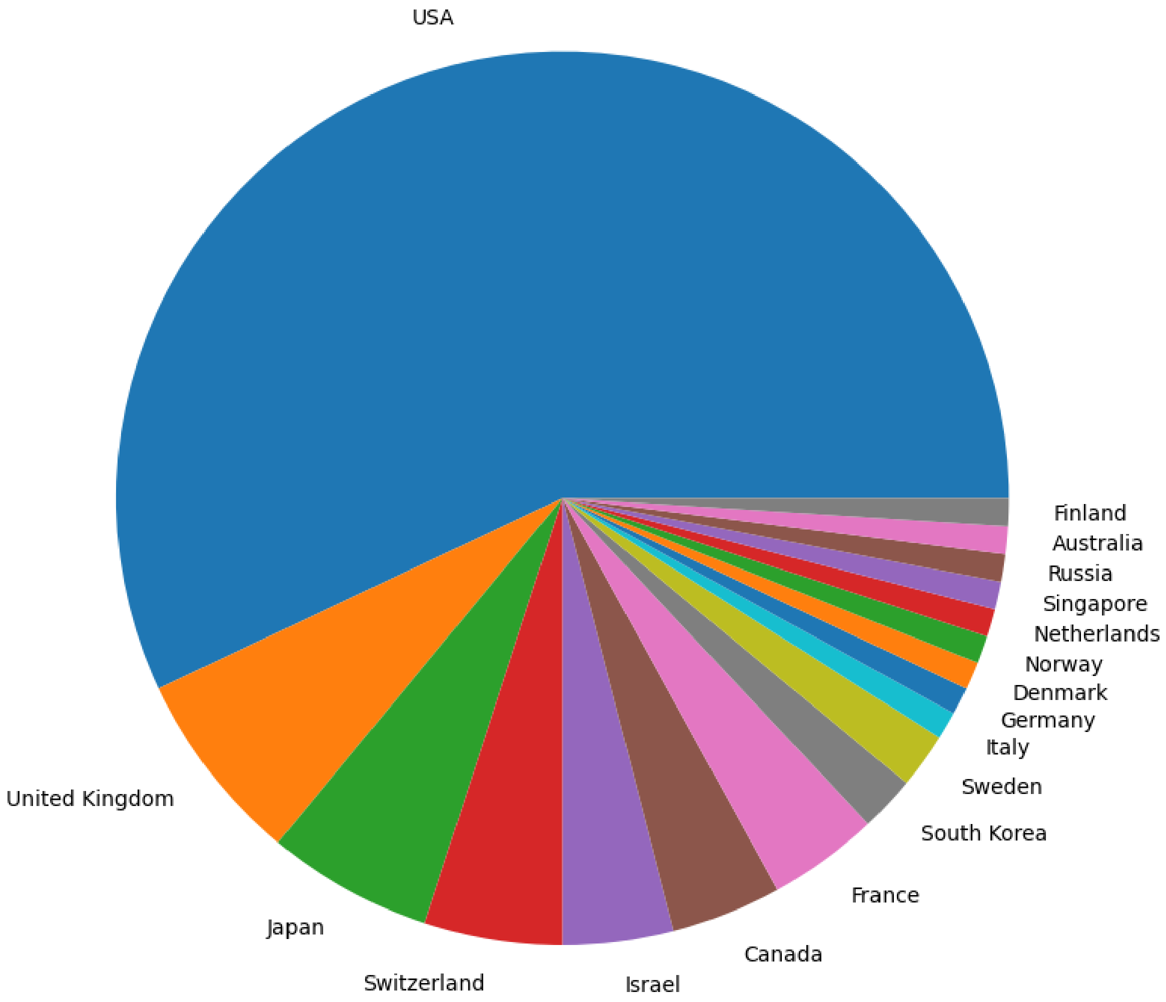
In [302]:

```
#6) Create a stacked area plot using Altair to represent the change in scores
change_scr = df.drop_duplicates()
top_5 = change_scr.nsmallest(5,'world_rank')

alt.Chart(top_5).mark_area().encode(
    x = 'year',
    y = 'score',
    color='institution:N'
).properties(width = 800,title = 'change in scores from 2012 to 2014')
```

Out[302]:

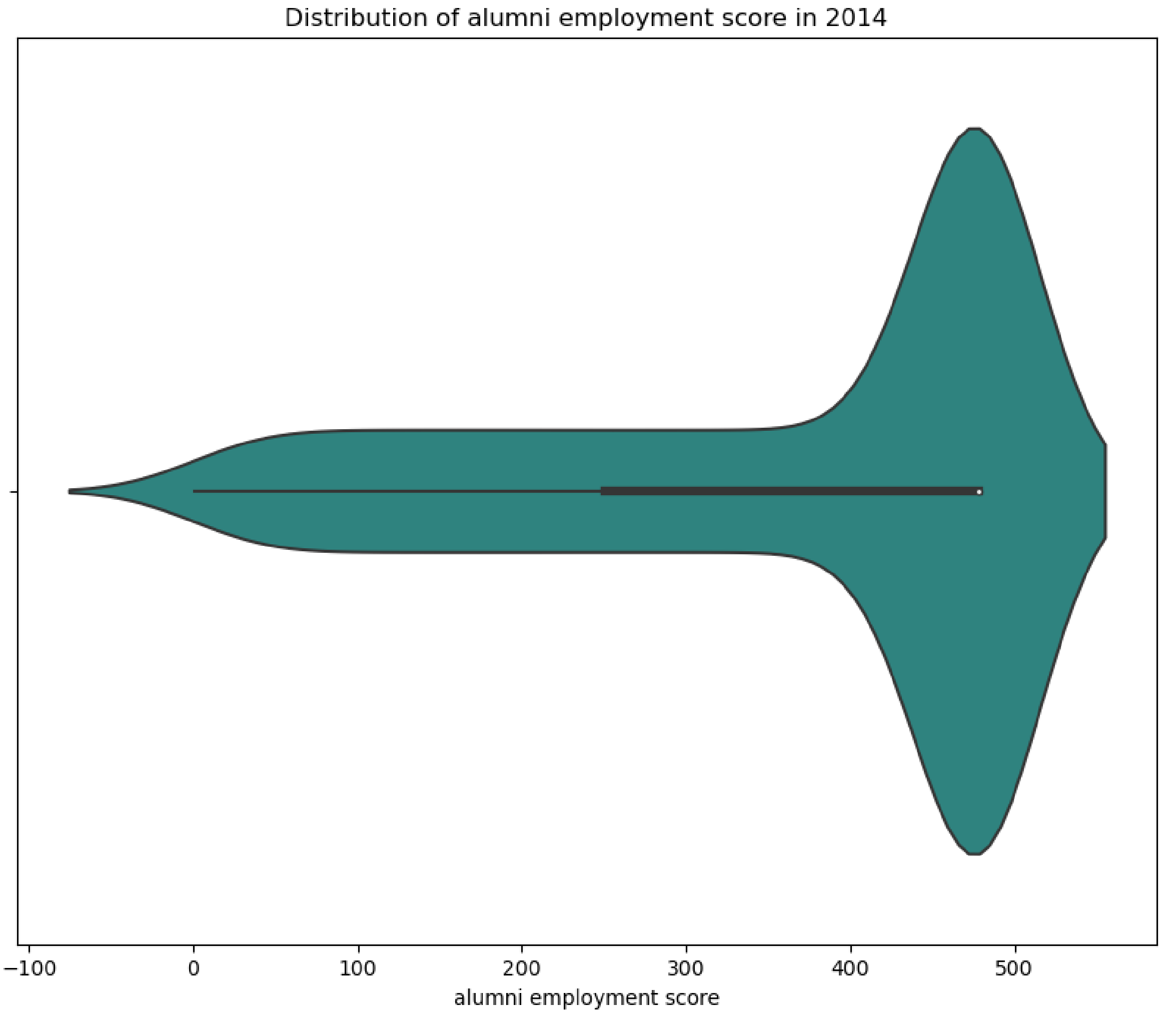


In [303]:
```
#7) How can you use matplotlib to create a pie chart illustrating the distribu
pie_crt = data_13['country'].value_counts()
lbl = data_13['country'].drop_duplicates()

fig , piep = plt.subplots()
piep.pie(pie_crt,labeldistance = 1.1, radius = 2, labels = lbl)
plt.show()
```



In [304]:

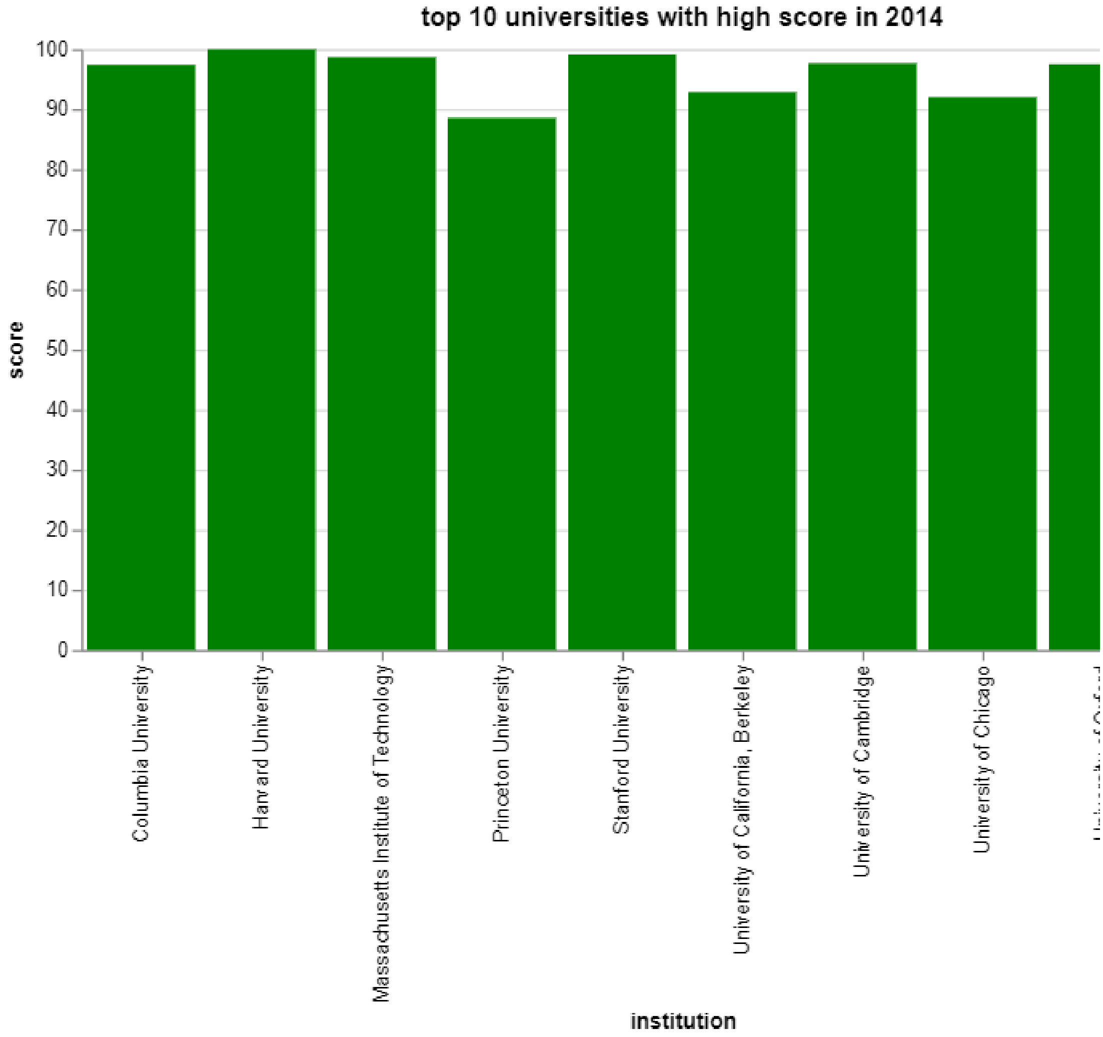
```
#8) Using seaborn, create a violin plot to show the distribution of alumni emp
plt.figure(figsize=(10,8))
sns.violinplot(data = data_14 , x = 'alumni_employment',palette='viridis')
plt.title('Distribution of alumni employment score in 2014')
plt.xlabel('alumni employment score')
plt.show()
```



In [305]:

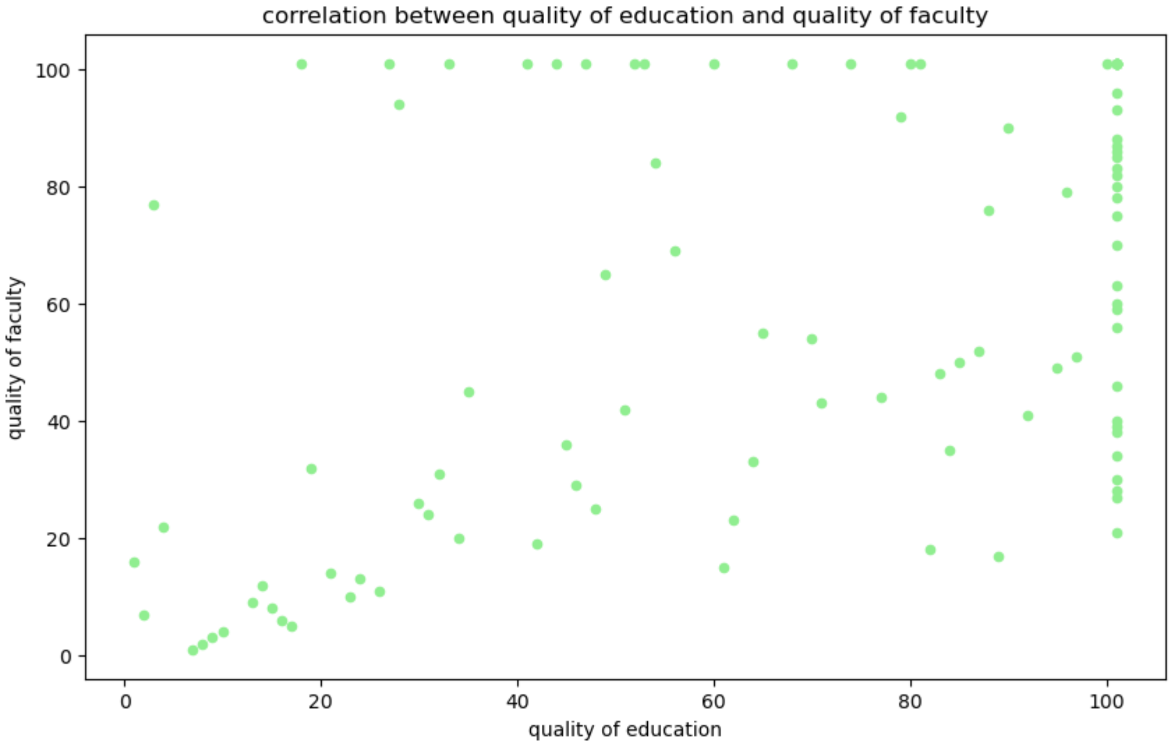
```
#9) How can you use Altair to create a bar chart for the top 10 universities w
srt= data_14.drop_duplicates()
top_data = srt.nlargest(10,'score')
alt.Chart(top_data,title= 'top 10 universities with high score in 2014').mark_
    x = 'institution',
    y = 'score'
).properties(width = 600 , height = 300 , padding ={"left":50 , "right":50,"to
```

Out[305]:



In [306]:

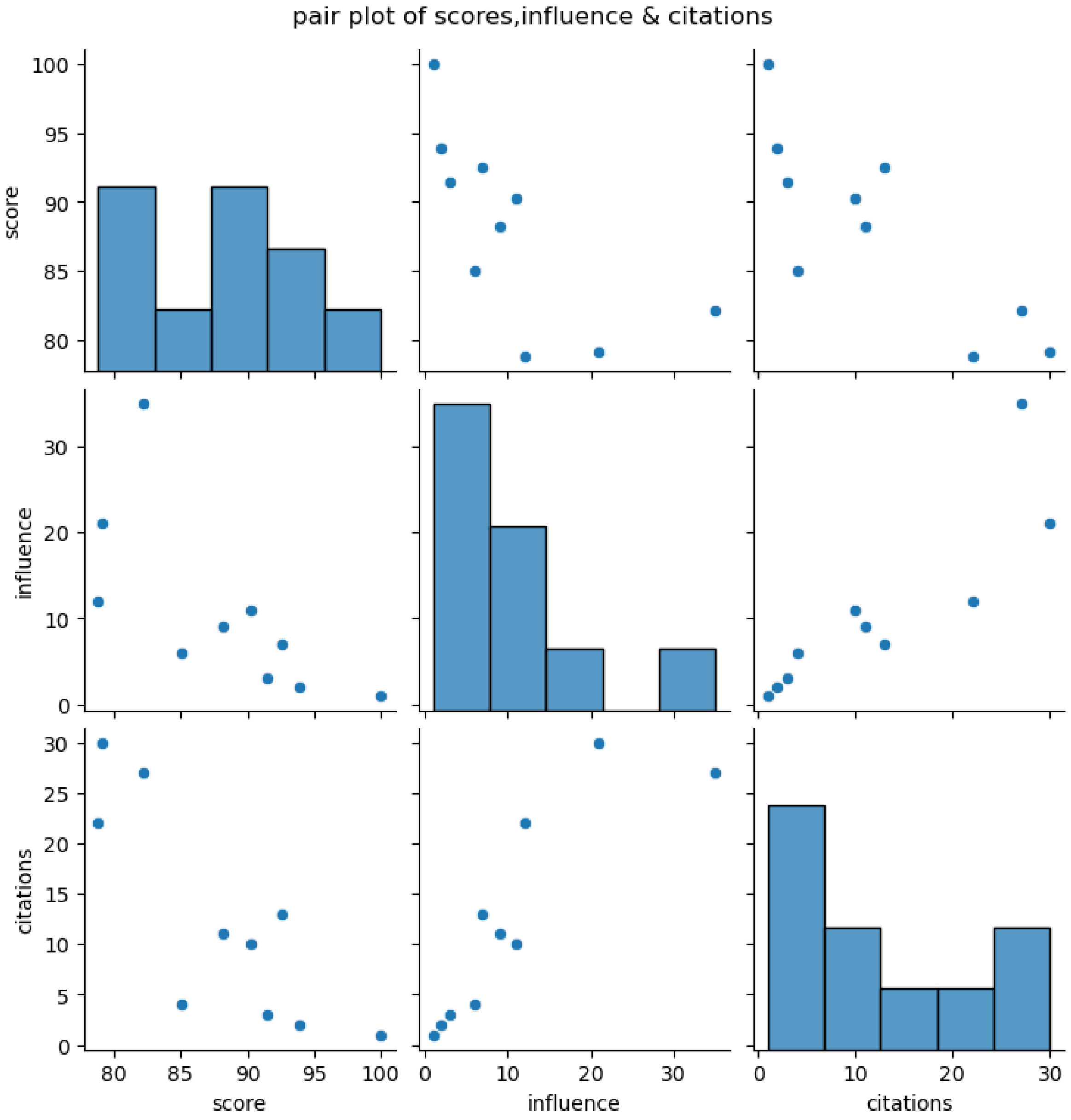
```
#10)    Create a scatter plot using matplotlib to visualize the correlation be
x= data_2012['quality_of_education']
y= data_2012['quality_of_faculty']

plt.figure(figsize=(10,6))
plt.scatter(x,y,s=20,color='lightgreen')
plt.xlabel('quality of education')
plt.ylabel('quality of faculty')
plt.title('correlation between quality of education and quality of faculty')
plt.show()
```



In [307]:

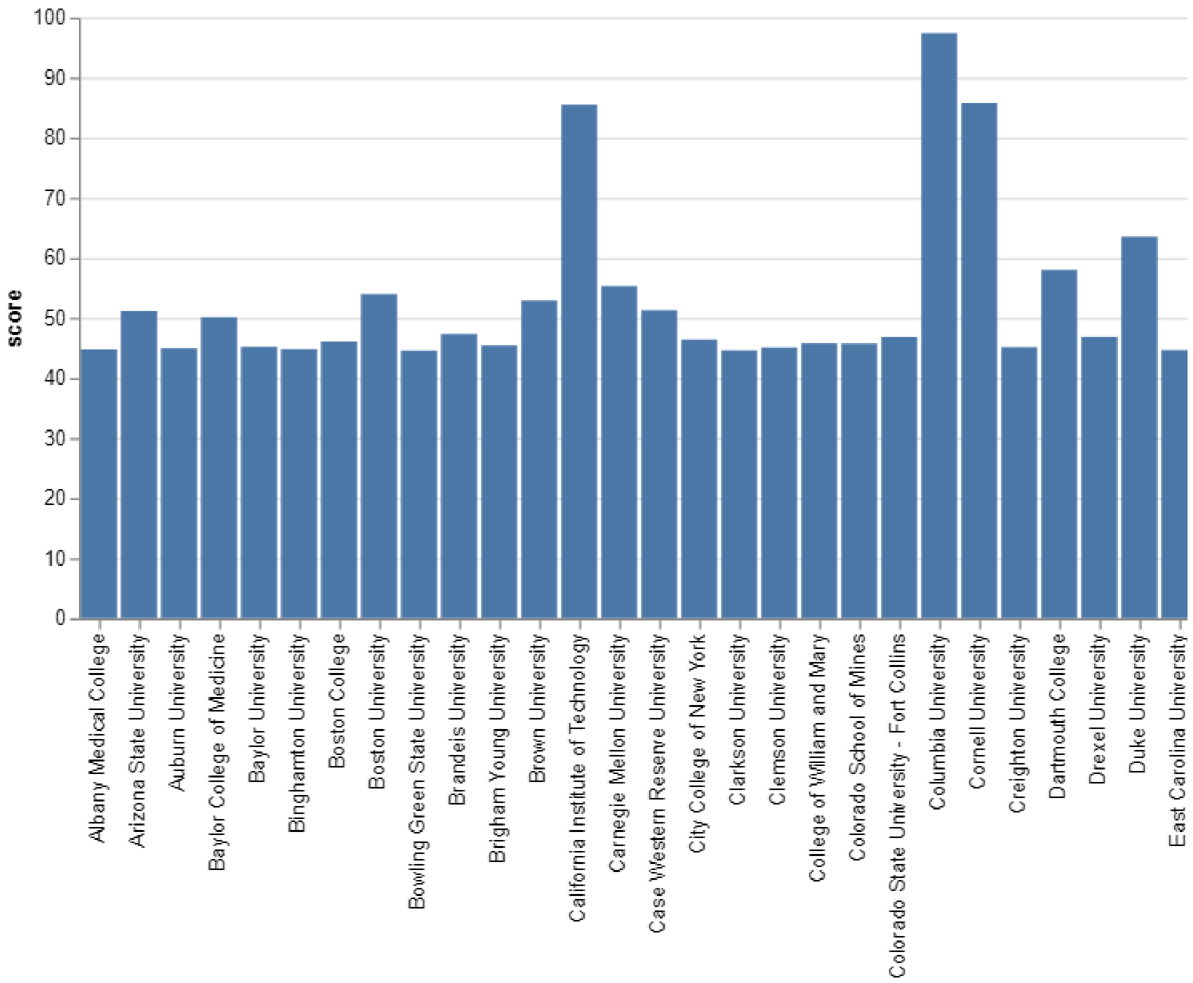
```
#11)    How can you use seaborn to create a pair plot to visualize the relatio
top13= data_13.nsmallest(10,'world_rank')

cols = ['score' , 'influence' , 'citations']

sns.pairplot(top13[cols], height=2.5)
plt.suptitle('pair plot of scores,influence & citations', y = 1.02)
plt.show()
```

```
C:\Users\Prathamesh Ghorpade\anaconda3\Lib\site-packages\seaborn\axisgrid.py:
118: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



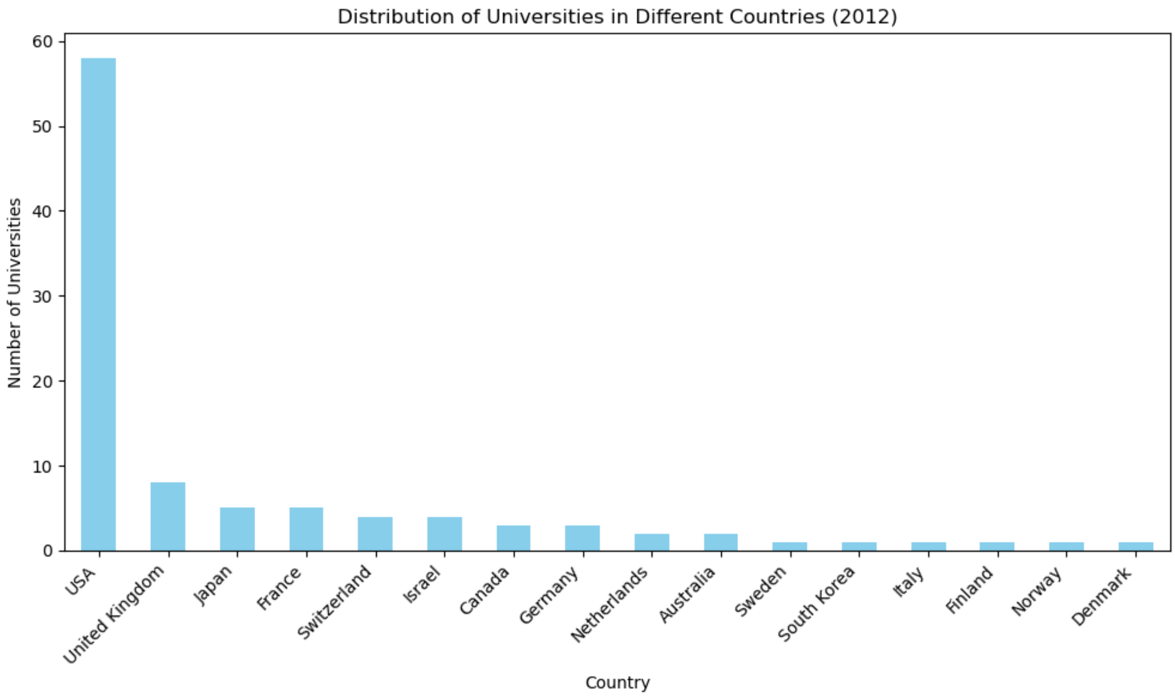
In [308]:
```
#12)    Using Altair, create a bar chart to compare the scores of universities
cntry_filt = data_14.loc[data_14['country'].isin(['USA','United kingdom'])]
clean_dt = cntry_filt.drop_duplicates()

alt.Chart(clean_dt, title = 'universities in USA and united kingdom ').mark_ba
    x= 'institution',
    y= 'score',
    color = 'country'
)
```

Out[308]:

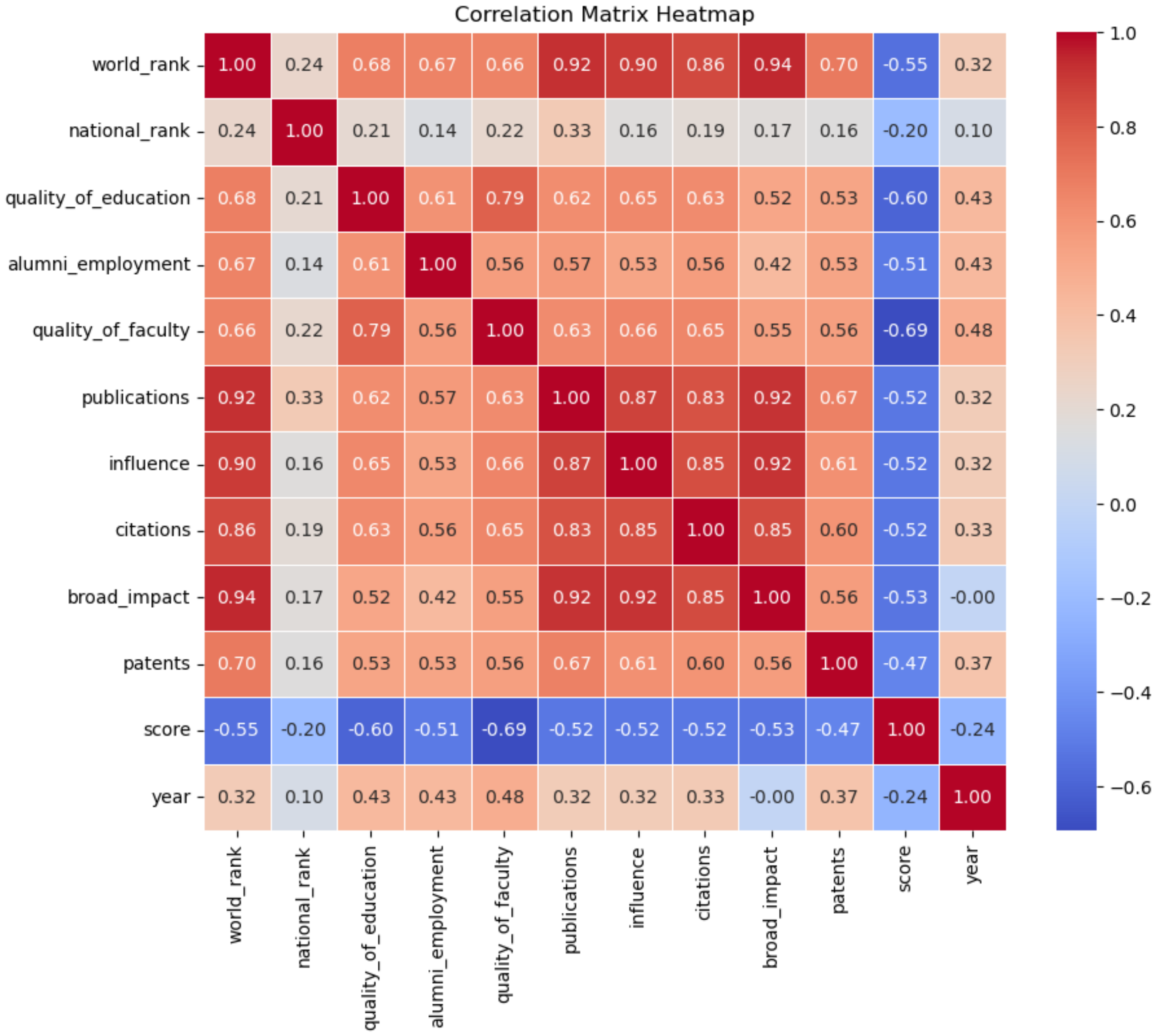


In [309]:

```
#13)    How can you use matplotlib to create a stacked bar chart illustrating
country_counts = data_2012['country'].value_counts()

plt.figure(figsize=(10, 6))
country_counts.plot(kind='bar', stacked=True, color='skyblue')
plt.xlabel('Country')
plt.ylabel('Number of Universities')
plt.title('Distribution of Universities in Different Countries (2012)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```



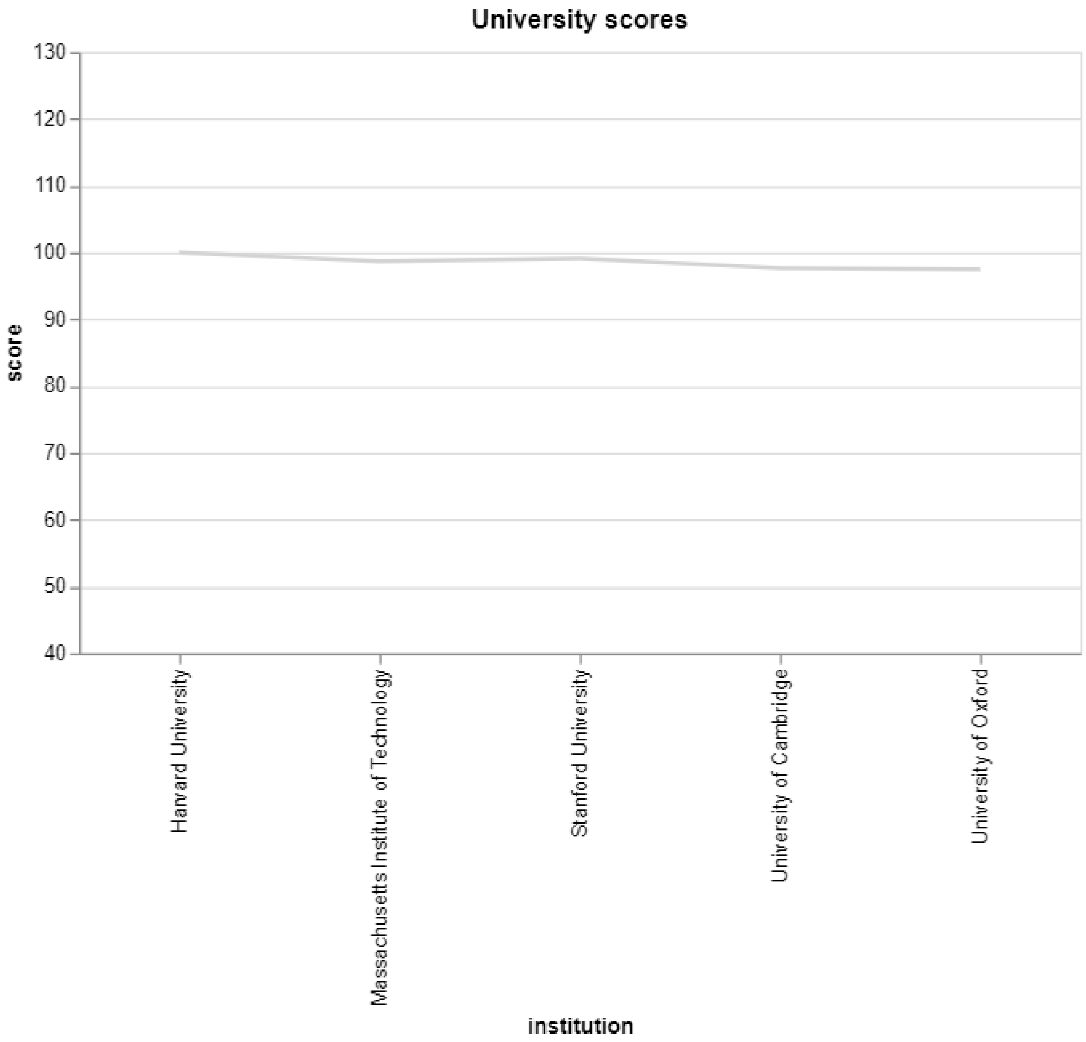
```
In [310]: #14)    Create a heatmap using seaborn to visualize the correlation matrix of
          numerical_columns = df.select_dtypes(include=['float64', 'int64'])

          correlation_matrix = numerical_columns.corr()

          plt.figure(figsize=(10, 8))
          sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f", linewi
          plt.title('Correlation Matrix Heatmap')
          plt.show()
```



	world_rank	national_rank	quality_of_education	alumni_employment	quality_of_faculty	publications	influence	citations	broad_impact	patents	score	year
world_rank	1.00	0.24	0.68	0.67	0.66	0.92	0.90	0.86	0.94	0.70	-0.55	0.32
national_rank	0.24	1.00	0.21	0.14	0.22	0.33	0.16	0.19	0.17	0.16	-0.20	0.10
quality_of_education	0.68	0.21	1.00	0.61	0.79	0.62	0.65	0.63	0.52	0.53	-0.60	0.43
alumni_employment	0.67	0.14	0.61	1.00	0.56	0.57	0.53	0.56	0.42	0.53	-0.51	0.43
quality_of_faculty	0.66	0.22	0.79	0.56	1.00	0.63	0.66	0.65	0.55	0.56	-0.69	0.48
publications	0.92	0.33	0.62	0.57	0.63	1.00	0.87	0.83	0.92	0.67	-0.52	0.32
influence	0.90	0.16	0.65	0.53	0.66	0.87	1.00	0.85	0.92	0.61	-0.52	0.32
citations	0.86	0.19	0.63	0.56	0.65	0.83	0.85	1.00	0.85	0.60	-0.52	0.33
broad_impact	0.94	0.17	0.52	0.42	0.55	0.92	0.92	0.85	1.00	0.56	-0.53	-0.00
patents	0.70	0.16	0.53	0.53	0.56	0.67	0.61	0.60	0.56	1.00	-0.47	0.37
score	-0.55	-0.20	-0.60	-0.51	-0.69	-0.52	-0.52	-0.52	-0.53	-0.47	1.00	-0.24
year	0.32	0.10	0.43	0.43	0.48	0.32	0.32	0.33	-0.00	0.37	-0.24	1.00

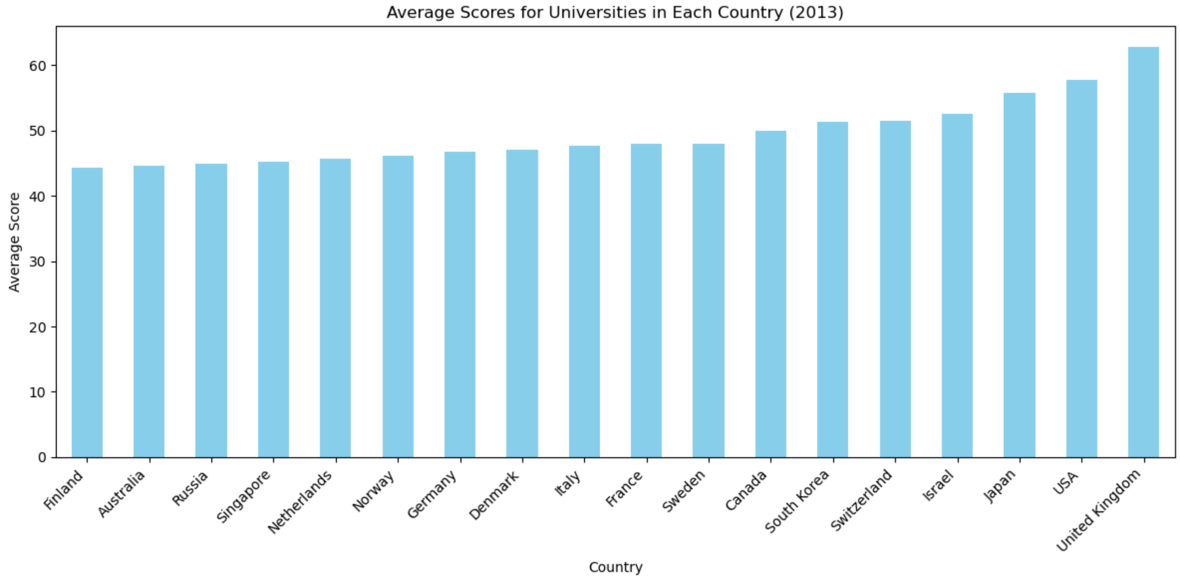
In [339]:
```
#15)    How can you use Altair to create a line chart showing the trend of sco
top_5 = data_14.nsmallest(5, 'world_rank')

alt.Chart(top_5).mark_line(color='lightgrey').encode(
    x='institution:N',
    y=alt.Y("score:Q",scale=alt.Scale(domain=[40,130]))
).properties(title='University scores',width=500)
```

Out[339]:



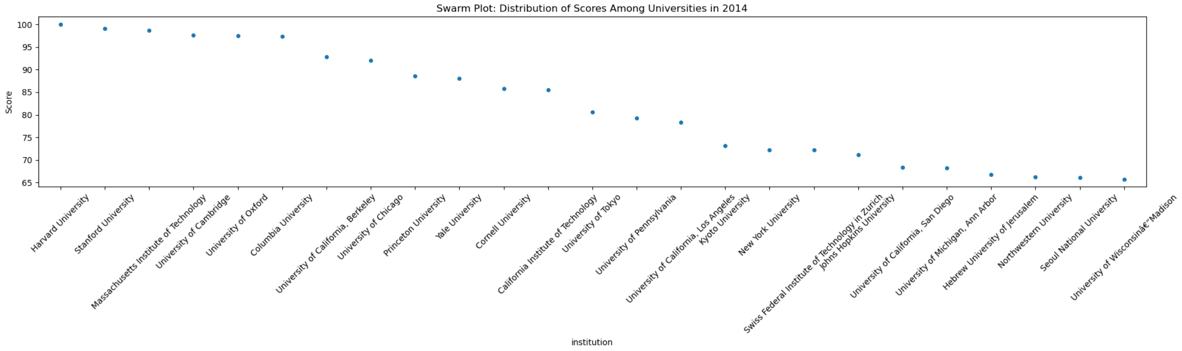
In [329]:

```
#16)    Using matplotlib, create a bar chart to show the average scores for un
average_scores = data_13.groupby('country')['score'].mean()

plt.figure(figsize=(12, 6))
average_scores.sort_values().plot(kind='bar', color='skyblue')
plt.xlabel('Country')
plt.ylabel('Average Score')
plt.title('Average Scores for Universities in Each Country (2013)')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()

plt.show()
```

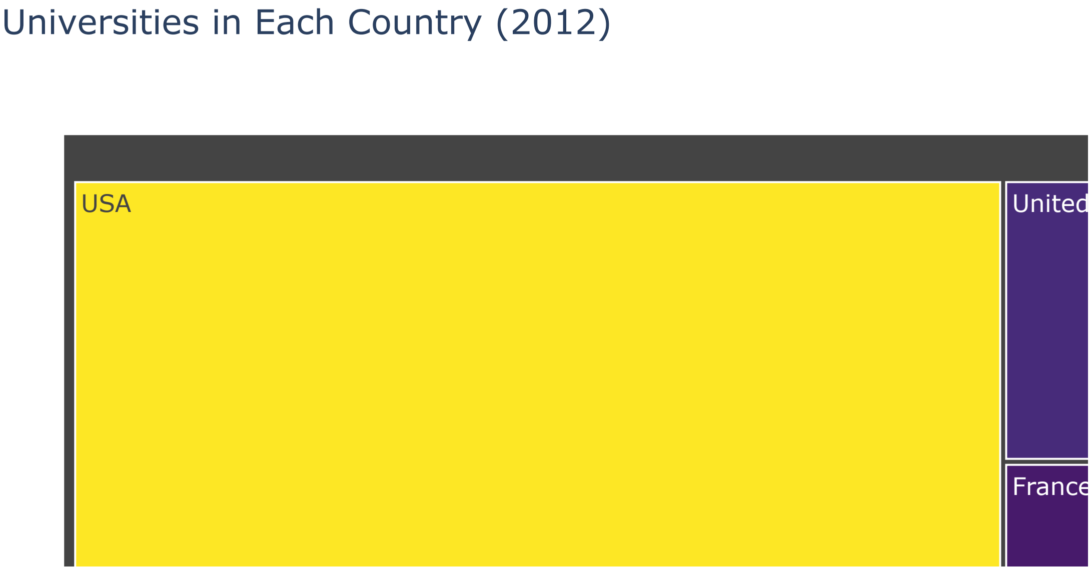


In [336]:

```
#17)    How can you use seaborn to create a swarm plot to visualize the- distr
data_2014=data_14.nsmallest(25,'world_rank')
plt.figure(figsize=(20,6))
sns.swarmplot(size=5,x='institution', y='score', data=data_2014)
plt.xlabel('institution')
plt.ylabel('Score')
plt.title('Swarm Plot: Distribution of Scores Among Universities in 2014')
plt.xticks(rotation=45)
plt.tight_layout()

plt.show()
```

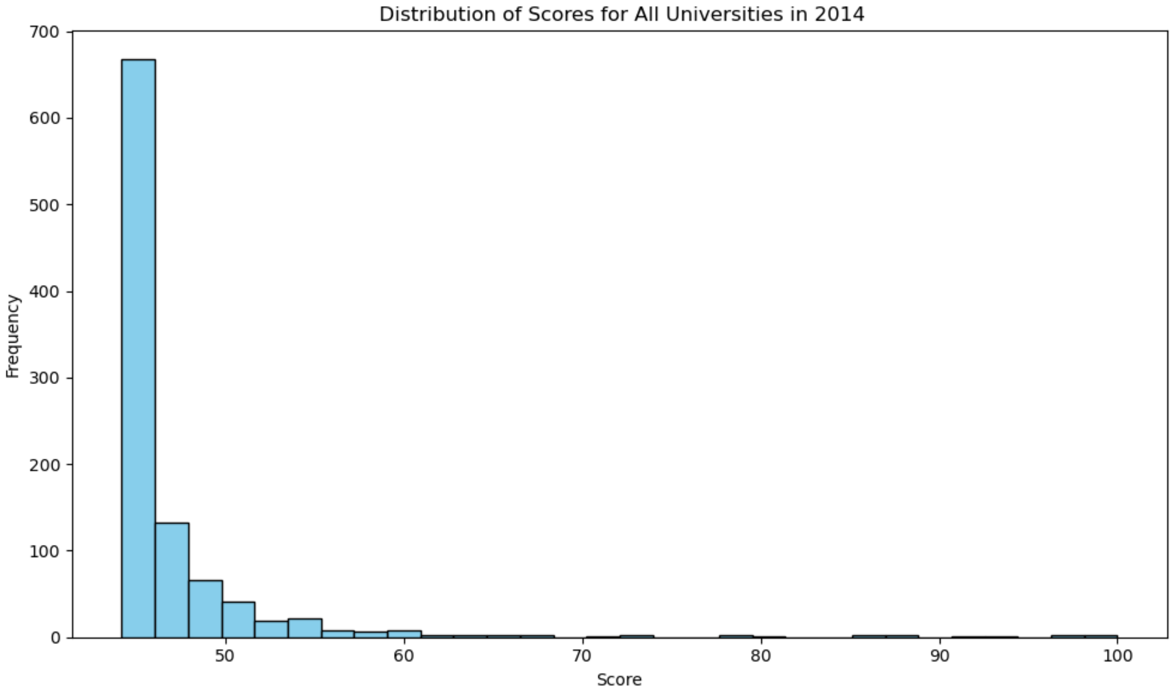


In [342]:
```
#18)    Create a treemap using Altair to represent the proportion of universit

import plotly.express as px
uni_country = data_2012['country'].value_counts().reset_index()
uni_country.columns = ['country', 'count']

fig = px.treemap(uni_country, path=['country'], values='count',
                 title=' Universities in Each Country (2012)',
                 color='count', color_continuous_scale='viridis')

fig.show()
```



In [344]:

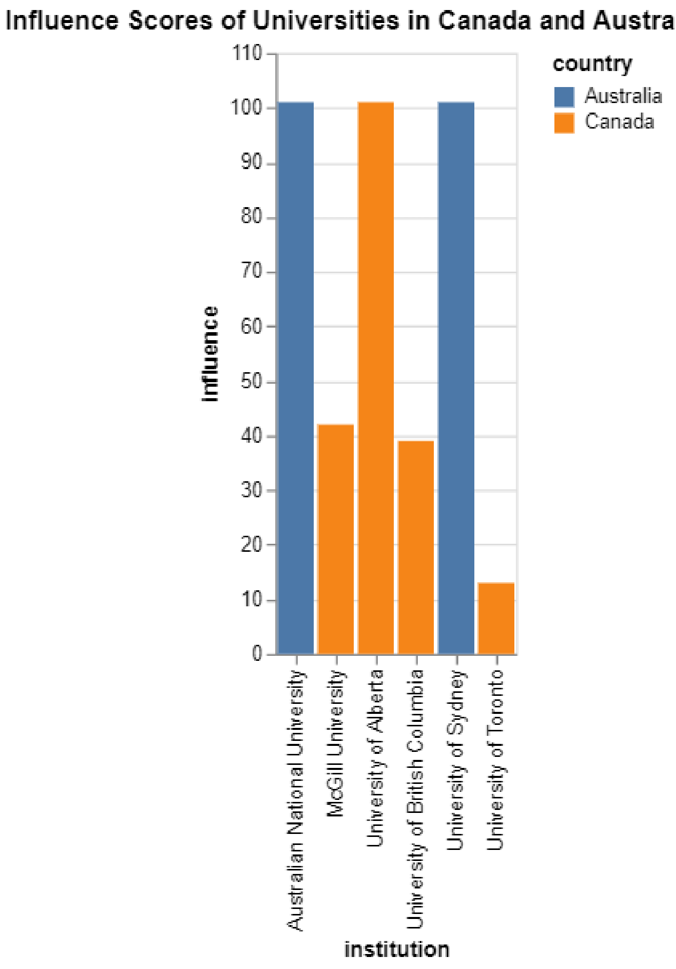
```
#19)    How can you use matplotlib to create a histogram showing the distribut
plt.figure(figsize=(10, 6))
plt.hist(data_14['score'], bins=30, color='skyblue', edgecolor='black')
plt.xlabel('Score')
plt.ylabel('Frequency')
plt.title(' Distribution of Scores for All Universities in 2014')
plt.tight_layout()

plt.show()
```



In [351]:

```
#20)    Using Altair, create a bar chart to compare the influence scores of un

df_ca_au = df_2013[df_2013['country'].isin(['Canada', 'Australia'])]

bar_chart = alt.Chart(df_ca_au).mark_bar().encode(
    x='institution',
    y='influence',
    color='country'
).properties(
    title='Influence Scores of Universities in Canada and Australia (2013)'
)

bar_chart
```

Out[351]:



Influence Scores of Universities in Canada and Australia (2013)

In [ ]: