

```
!pip install pypyodbc
```

```
!pip install pyodbc
```

```
Requirement already satisfied: pypyodbc in c:\users\asus\anaconda3\lib\site-packages (1.3.6)
```

```
Requirement already satisfied: setuptools in c:\users\asus\anaconda3\lib\site-packages (from pypyodbc) (63.4.1)
```

```
Requirement already satisfied: pyodbc in c:\users\asus\anaconda3\lib\site-packages (4.0.34)
```

```
import pandas as pd
```

```
import numpy as np
```

```
import pyodbc as odbc
```

```
import sqlite3
```

```
import matplotlib.pyplot as plt
```

```
# Task 1
```

```
data = pd.read_excel('WorldUniversity.xlsx')
```

```
# Task 2
```

```
data.isnull().sum()
```

world_rank	0
institution	0
country	0
national_rank	0
quality_of_education	0
alumni_employment	0
quality_of_faculty	0
publications	0
influence	0
citations	0
broad_impact	1800
patents	0
score	0
year	0
dtype: int64	

```
df = data.drop_duplicates()
```

```
Connection_string = (
```

```
    r'DRIVER={ODBC Driver 17 for SQL Server};'
```

```
    r'SERVER= ASUS-ROG\SQLEXPRESS;'
```

```
    r'DATABASE=Mega-Project;'
```

```
    r'Trusted_Connection=yes;'
```

```
)
```

```

con = odbc.connect(Connection_string)
print(con)

<pyodbc.Connection object at 0x000001E64D1F4850>

from sqlalchemy.engine import URL
connection_url = URL.create("mssql+pyodbc",
query = {"odbc_connect": Connection_string})

from sqlalchemy import create_engine
engine = create_engine(connection_url)

```

## Task 4 - SQL

```

import sqlalchemy as sa
with engine.begin() as conn:
    df= pd.read_sql_query(sa.text("SELECT TOP 10 * FROM world_rank"),
conn)
    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						

```
# 1
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" select * from world_rank where
country = 'USA'"), conn)
    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	5.0	California Institute of Technology	USA
4	6.0	Princeton University	USA
...	...		...
5117	655.0	University of Idaho	USA
5118	656.0	Old Dominion University	USA
5119	661.0	Portland State University	USA
5120	665.0	Loma Linda University	USA
5121	668.0	University of Maine, Orono	USA

	national_rank	quality_of_education	alumni_employment	\
0	1.0	7.0	9.0	
1	2.0	9.0	17.0	
2	3.0	17.0	11.0	
3	4.0	2.0	29.0	
4	5.0	8.0	14.0	
...	...	...	...	
5117	190.0	367.0	567.0	
5118	191.0	367.0	567.0	
5119	192.0	367.0	245.0	
5120	193.0	367.0	567.0	
5121	194.0	345.0	567.0	

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	1.0	1.0	1.0	1.0
None				
1	3.0	12.0	4.0	4.0
None				
2	5.0	4.0	2.0	2.0
None				
3	7.0	37.0	22.0	22.0
None				
4	2.0	53.0	33.0	26.0
None				
...	...	...	...	...
...				
5117	218.0	664.0	417.0	511.0
579				
5118	218.0	719.0	590.0	511.0
590				
5119	218.0	800.0	548.0	812.0
741				
5120	218.0	848.0	755.0	645.0
565				
5121	218.0	788.0	427.0	368.0
637				

	patents	score	year
0	5.0	100.00	2012.0
1	1.0	91.67	2012.0

```

2      15.0    89.50  2012.0
3      18.0    85.21  2012.0
4     101.0    82.50  2012.0
...
5117   469.0    44.42  2015.0
5118   379.0    44.42  2015.0
5119   484.0    44.41  2015.0
5120   555.0    44.41  2015.0
5121   346.0    44.40  2015.0

```

[5122 rows x 14 columns]

```

# 2
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" select top 10 * from world_rank
where year = 2012 "), conn)
    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						

```
# 3
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" select top 10 * from world_rank
where country = 'United Kingdom' and score>=80 and year = 2013 "),
conn)
    print(df)
```

	world_rank	institution	country	national_rank
\				
0	3.0	University of Oxford	United Kingdom	1.0
1	5.0	University of Cambridge	United Kingdom	2.0
2	3.0	University of Oxford	United Kingdom	1.0

3	5.0	University of Cambridge	United Kingdom	2.0
4	3.0	University of Oxford	United Kingdom	1.0
5	5.0	University of Cambridge	United Kingdom	2.0
6	3.0	University of Oxford	United Kingdom	1.0
7	5.0	University of Cambridge	United Kingdom	2.0
8	3.0	University of Oxford	United Kingdom	1.0
9	5.0	University of Cambridge	United Kingdom	2.0

	quality_of_education publications \	alumni_employment	quality_of_faculty
0	7.0	12.0	10.0
11.0			
1	3.0	15.0	5.0
9.0			
2	7.0	12.0	10.0
11.0			
3	3.0	15.0	5.0
9.0			
4	7.0	12.0	10.0
11.0			
5	3.0	15.0	5.0
9.0			
6	7.0	12.0	10.0
11.0			
7	3.0	15.0	5.0
9.0			
8	7.0	12.0	10.0
11.0			
9	3.0	15.0	5.0
9.0			

	influence	citations	broad_impact	patents	score	year
0	7.0	13.0	None	15.0	92.54	2013.0
1	11.0	10.0	None	39.0	90.24	2013.0
2	7.0	13.0	None	15.0	92.54	2013.0
3	11.0	10.0	None	39.0	90.24	2013.0
4	7.0	13.0	None	15.0	92.54	2013.0
5	11.0	10.0	None	39.0	90.24	2013.0
6	7.0	13.0	None	15.0	92.54	2013.0
7	11.0	10.0	None	39.0	90.24	2013.0
8	7.0	13.0	None	15.0	92.54	2013.0
9	11.0	10.0	None	39.0	90.24	2013.0

```
# 4
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select country, count(*) as
count_of_university from world_rank group by country order by
count_of_university desc"), conn)
    print(df)
```

	country	count_of_university
0	USA	5122
1	China	1457
2	Japan	1400
3	United Kingdom	1279
4	Germany	1026
5	France	961
6	Italy	852
7	Spain	709
8	Canada	640
9	South Korea	635
10	Australia	514
11	Taiwan	401
12	Brazil	313
13	India	272
14	Netherlands	260
15	Switzerland	233
16	Sweden	216
17	Austria	212
18	Israel	197
19	Finland	179
20	Belgium	177
21	Turkey	173
22	Poland	156
23	Ireland	140
24	Iran	136
25	Portugal	125
26	Greece	124
27	Hong Kong	108
28	Denmark	108
29	Norway	107
30	New Zealand	107
31	Hungary	105
32	South Africa	89
33	Czech Republic	86
34	Russia	79
35	Chile	70
36	Saudi Arabia	69
37	Egypt	68
38	Argentina	61
39	Malaysia	54
40	Thailand	53
41	Singapore	45



42	Colombia	35
43	Slovenia	35
44	Mexico	35
45	Romania	25
46	Iceland	18
47	Croatia	18
48	Estonia	18
49	Slovak Republic	17
50	Lebanon	17
51	Lithuania	17
52	Uganda	17
53	Puerto Rico	17
54	Serbia	17
55	United Arab Emirates	17
56	Uruguay	17
57	Cyprus	17
58	Bulgaria	17

```
# 5
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select country,
round(avg(score),2) as AVG_Score from world_rank where year = 2014
group by country order by AVG_Score desc"), conn)
    print(df)
```

	country	AVG_Score
0	Israel	52.14
1	Switzerland	51.66
2	Singapore	51.44
3	USA	50.64
4	Russia	49.10
5	United Kingdom	48.45
6	Netherlands	48.41
7	Denmark	48.35
8	Sweden	48.25
9	Canada	47.29
10	Belgium	47.26
11	Japan	46.87
12	South Africa	46.70
13	Germany	46.65
14	Norway	46.60
15	Hong Kong	46.56
16	South Korea	46.37
17	France	46.33
18	Australia	46.05
19	Finland	45.86
20	Italy	45.58
21	Ireland	45.35
22	Thailand	45.31
23	Malaysia	45.31

24	New Zealand	45.30
25	Mexico	45.29
26	Austria	45.29
27	Portugal	45.24
28	Spain	45.13
29	China	45.11
30	Taiwan	45.06
31	Iceland	45.05
32	Greece	44.98
33	Estonia	44.94
34	Brazil	44.93
35	Lebanon	44.93
36	Croatia	44.92
37	India	44.89
38	Chile	44.86
39	Czech Republic	44.84
40	Hungary	44.80
41	Slovenia	44.74
42	Saudi Arabia	44.73
43	Argentina	44.73
44	Poland	44.65
45	Turkey	44.63
46	Colombia	44.62
47	Slovak Republic	44.60
48	Serbia	44.51
49	Bulgaria	44.48
50	Lithuania	44.46
51	Uganda	44.40
52	Iran	44.38
53	Egypt	44.36
54	United Arab Emirates	44.36
55	Uruguay	44.35
56	Cyprus	44.32
57	Romania	44.32
58	Puerto Rico	44.29

```
# 6
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
quality_of_education >= 20"), conn)
    print(df)
```

	world_rank	institution	country
national_rank \			
0	9.0	Columbia University	USA
7.0			
1	12.0	Cornell University	USA
10.0			
2	13.0	University of Pennsylvania	USA
11.0			

3	14.0	University of Tokyo	Japan			
1.0						
4	15.0	Johns Hopkins University	USA			
12.0						
...	...	...	...			
..						
18819	668.0	University of Maine, Orono	USA			
194.0						
18820	669.0	Graz University of Technology	Austria			
8.0						
18821	670.0	Gifu University	Japan			
43.0						
18822	671.0	University of Jyväskylä	Finland			
8.0						
18823	672.0	University of Paris 13	France			
29.0						
	quality_of_education	alumni_employment	quality_of_faculty \			
0	23.0	21.0	10.0			
1	21.0	42.0	14.0			
2	31.0	16.0	24.0			
3	32.0	19.0	31.0			
4	34.0	77.0	20.0			
...	...	...	...			
18819	345.0	567.0	218.0			
18820	367.0	567.0	218.0			
18821	367.0	567.0	218.0			
18822	367.0	525.0	218.0			
18823	367.0	567.0	218.0			
	publications	influence	citations	broad_impact	patents	score
year						
0	13.0	12.0	14.0	None	5.0	78.86
2012.0						
1	22.0	21.0	16.0	None	10.0	73.69
2012.0						
2	9.0	10.0	8.0	None	9.0	73.64
2012.0						
3	8.0	19.0	23.0	None	3.0	69.49
2012.0						
4	11.0	9.0	9.0	None	7.0	66.94
2012.0						
...	...	...	...	...	...	...
...						
18819	788.0	427.0	368.0	637	346.0	44.40
2015.0						
18820	677.0	715.0	368.0	622	404.0	44.40
2015.0						
18821	701.0	658.0	645.0	606	403.0	44.40

2015.0						
18822	456.0	443.0	812.0	622	805.0	44.40
2015.0						
18823	669.0	744.0	645.0	579	700.0	44.40
2015.0						

[18824 rows x 14 columns]

```
# 7
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
score between 70 and 80"), conn)
    print(df)
```

	world_rank	institution	country
national_rank \			
0	8.0	Yale University	USA
6.0			
1	9.0	Columbia University	USA
7.0			
2	10.0	University of California, Berkeley	USA
8.0			
3	11.0	University of Chicago	USA
9.0			
4	12.0	Cornell University	USA
10.0			
..	...	...	...
...			
193	19.0	Johns Hopkins University	USA
14.0			
194	13.0	University of Tokyo	Japan
1.0			
195	14.0	University of Pennsylvania	USA
11.0			
196	15.0	University of California, Los Angeles	USA
12.0			
197	16.0	Johns Hopkins University	USA
13.0			

	quality_of_education	alumni_employment	quality_of_faculty	\
0	14.0	31.0	12.0	
1	23.0	21.0	10.0	
2	16.0	52.0	6.0	
3	15.0	26.0	8.0	
4	21.0	42.0	14.0	
..	...	...	...	
193	24.0	75.0	18.0	
194	16.0	3.0	38.0	
195	20.0	4.0	28.0	
196	28.0	27.0	13.0	

197		18.0		84.0		16.0
	publications	influence	citations	broad_impact	patents	score
year						
0	14.0	6.0	15.0	None	66.0	79.14
2012.0						
1	13.0	12.0	14.0	None	5.0	78.86
2012.0						
2	6.0	5.0	3.0	None	16.0	78.55
2012.0						
3	34.0	20.0	28.0	None	101.0	73.82
2012.0						
4	22.0	21.0	16.0	None	10.0	73.69
2012.0						
..	...	...	...	...	...	...
...						
193	4.0	10.0	6.0	3	3.0	71.17
2014.0						
194	14.0	19.0	31.0	29	7.0	78.23
2015.0						
195	8.0	18.0	14.0	9	14.0	77.60
2015.0						
196	6.0	14.0	8.0	6	9.0	76.91
2015.0						
197	4.0	11.0	5.0	3	2.0	71.60
2015.0						

[198 rows x 14 columns]

# 8

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select top 5 * from world_rank
order by alumni_employment desc"), conn)
    print(df)
```

	world_rank		institution	country
national_rank \				
0	21.0		University of California, San Diego	USA
16.0				
1	26.0		University of California, San Francisco	USA
19.0				
2	29.0		Rockefeller University	USA
21.0				
3	39.0		Weizmann Institute of Science	Israel
2.0				
4	71.0		Karolinska Institute	Sweden
1.0				

	quality_of_education	alumni_employment	quality_of_faculty
publications \			

0	36.0	567.0	19.0
16.0			
1	367.0	567.0	20.0
22.0			
2	49.0	567.0	12.0
372.0			
3	19.0	567.0	22.0
246.0			
4	149.0	567.0	36.0
51.0			

	influence	citations	broad_impact	patents	score	year
0	5.0	15.0	15	17.0	66.59	2015.0
1	7.0	15.0	11	56.0	63.69	2015.0
2	28.0	115.0	70	289.0	61.28	2015.0
3	63.0	234.0	143	45.0	57.17	2015.0
4	54.0	52.0	34	743.0	52.79	2015.0

```
# 9
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select top 1 * from world_rank
where year = 2013 order by quality_of_faculty desc"), conn)
    print(df)
```

	world_rank	institution	country	national_rank	\
0	28.0	Northwestern University	USA	21.0	

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	101.0	21.0	101.0
27.0			

	influence	citations	broad_impact	patents	score	year
0	26.0	24.0	None	30.0	56.12	2013.0

```
# 10
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select count(*) as
count_of_universities from world_rank where national_rank <=5"), conn)
    print(df)
```

	count_of_universities
0	4679

```
# 11
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
quality_of_education = 1"), conn)
    print(df)
```

	world_rank	institution	country	national_rank	\
0	21.0	Rockefeller University	USA	15.0	
1	1.0	Harvard University	USA	1.0	
2	1.0	Harvard University	USA	1.0	
3	1.0	Harvard University	USA	1.0	
4	21.0	Rockefeller University	USA	15.0	
5	1.0	Harvard University	USA	1.0	
6	1.0	Harvard University	USA	1.0	
7	1.0	Harvard University	USA	1.0	
8	21.0	Rockefeller University	USA	15.0	
9	1.0	Harvard University	USA	1.0	
10	1.0	Harvard University	USA	1.0	
11	1.0	Harvard University	USA	1.0	
12	21.0	Rockefeller University	USA	15.0	
13	1.0	Harvard University	USA	1.0	
14	1.0	Harvard University	USA	1.0	
15	1.0	Harvard University	USA	1.0	
16	21.0	Rockefeller University	USA	15.0	
17	1.0	Harvard University	USA	1.0	
18	1.0	Harvard University	USA	1.0	
19	1.0	Harvard University	USA	1.0	
20	21.0	Rockefeller University	USA	15.0	
21	1.0	Harvard University	USA	1.0	
22	1.0	Harvard University	USA	1.0	
23	1.0	Harvard University	USA	1.0	
24	21.0	Rockefeller University	USA	15.0	
25	1.0	Harvard University	USA	1.0	
26	1.0	Harvard University	USA	1.0	
27	1.0	Harvard University	USA	1.0	
28	21.0	Rockefeller University	USA	15.0	
29	1.0	Harvard University	USA	1.0	
30	1.0	Harvard University	USA	1.0	
31	1.0	Harvard University	USA	1.0	
32	21.0	Rockefeller University	USA	15.0	
33	1.0	Harvard University	USA	1.0	
34	1.0	Harvard University	USA	1.0	
35	1.0	Harvard University	USA	1.0	
	quality_of_education	alumni_employment	quality_of_faculty		
0	1.0	101.0	16.0		
101.0					
1	1.0	1.0	1.0		
1.0					
2	1.0	1.0	1.0		
1.0					
3	1.0	1.0	1.0		
1.0					
4	1.0	101.0	16.0		
101.0					

5	1.0	1.0	1.0
1.0			
6	1.0	1.0	1.0
1.0			
7	1.0	1.0	1.0
1.0			
8	1.0	101.0	16.0
101.0			
9	1.0	1.0	1.0
1.0			
10	1.0	1.0	1.0
1.0			
11	1.0	1.0	1.0
1.0			
12	1.0	101.0	16.0
101.0			
13	1.0	1.0	1.0
1.0			
14	1.0	1.0	1.0
1.0			
15	1.0	1.0	1.0
1.0			
16	1.0	101.0	16.0
101.0			
17	1.0	1.0	1.0
1.0			
18	1.0	1.0	1.0
1.0			
19	1.0	1.0	1.0
1.0			
20	1.0	101.0	16.0
101.0			
21	1.0	1.0	1.0
1.0			
22	1.0	1.0	1.0
1.0			
23	1.0	1.0	1.0
1.0			
24	1.0	101.0	16.0
101.0			
25	1.0	1.0	1.0
1.0			
26	1.0	1.0	1.0
1.0			
27	1.0	1.0	1.0
1.0			
28	1.0	101.0	16.0
101.0			
29	1.0	1.0	1.0
1.0			



30	1.0	1.0	1.0
1.0			
31	1.0	1.0	1.0
1.0			
32	1.0	101.0	16.0
101.0			
33	1.0	1.0	1.0
1.0			
34	1.0	1.0	1.0
1.0			
35	1.0	1.0	1.0
1.0			

	influence	citations	broad_impact	patents	score	year
0	28.0	96.0	None	101.0	61.74	2012.0
1	1.0	1.0	None	7.0	100.00	2013.0
2	1.0	1.0	1	2.0	100.00	2014.0
3	1.0	1.0	1	3.0	100.00	2015.0
4	28.0	96.0	None	101.0	61.74	2012.0
5	1.0	1.0	None	7.0	100.00	2013.0
6	1.0	1.0	1	2.0	100.00	2014.0
7	1.0	1.0	1	3.0	100.00	2015.0
8	28.0	96.0	None	101.0	61.74	2012.0
9	1.0	1.0	None	7.0	100.00	2013.0
10	1.0	1.0	1	2.0	100.00	2014.0
11	1.0	1.0	1	3.0	100.00	2015.0
12	28.0	96.0	None	101.0	61.74	2012.0
13	1.0	1.0	None	7.0	100.00	2013.0
14	1.0	1.0	1	2.0	100.00	2014.0
15	1.0	1.0	1	3.0	100.00	2015.0
16	28.0	96.0	None	101.0	61.74	2012.0
17	1.0	1.0	None	7.0	100.00	2013.0
18	1.0	1.0	1	2.0	100.00	2014.0
19	1.0	1.0	1	3.0	100.00	2015.0
20	28.0	96.0	None	101.0	61.74	2012.0
21	1.0	1.0	None	7.0	100.00	2013.0
22	1.0	1.0	1	2.0	100.00	2014.0
23	1.0	1.0	1	3.0	100.00	2015.0
24	28.0	96.0	None	101.0	61.74	2012.0
25	1.0	1.0	None	7.0	100.00	2013.0
26	1.0	1.0	1	2.0	100.00	2014.0
27	1.0	1.0	1	3.0	100.00	2015.0
28	28.0	96.0	None	101.0	61.74	2012.0
29	1.0	1.0	None	7.0	100.00	2013.0
30	1.0	1.0	1	2.0	100.00	2014.0
31	1.0	1.0	1	3.0	100.00	2015.0
32	28.0	96.0	None	101.0	61.74	2012.0
33	1.0	1.0	None	7.0	100.00	2013.0

34	1.0	1.0	1	2.0	100.00	2014.0
35	1.0	1.0	1	3.0	100.00	2015.0

```
# 12
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select top 10 * from world_rank
order by citations desc"), conn)
    print(df)
```

	world_rank	institution	country
\			
0	104.0	Mines ParisTech	France
1	226.0	National Cheng Kung University	Taiwan
2	250.0	Moscow Institute of Physics and Technology	Russia
3	341.0	Indian Institute of Technology Delhi	India
4	351.0	Wesleyan University	USA
5	379.0	University of Delhi	India
6	383.0	University of Turku	Finland
7	389.0	Xi'an Jiaotong University	China
8	395.0	Innsbruck Medical University	Austria
9	422.0	Chonbuk National University	South Korea

	national_rank	quality_of_education	alumni_employment
quality_of_faculty \			
0	5.0	278.0	23.0
169.0			
1	2.0	367.0	52.0
218.0			
2	2.0	23.0	339.0
218.0			
3	1.0	367.0	59.0
218.0			
4	124.0	154.0	61.0
218.0			
5	2.0	240.0	72.0
218.0			
6	3.0	367.0	531.0
218.0			
7	13.0	367.0	154.0
218.0			
8	4.0	367.0	567.0

218.0						
9	13.0		367.0		567.0	
218.0						

	publications	influence	citations	broad_impact	patents	score
year						
0	977.0	834.0	812.0	906	871.0	50.34
2015.0						
1	164.0	485.0	812.0	388	436.0	46.97
2015.0						
2	949.0	605.0	812.0	1000	839.0	46.55
2015.0						
3	635.0	943.0	812.0	781	625.0	45.54
2015.0						
4	983.0	506.0	812.0	850	871.0	45.48
2015.0						
5	703.0	763.0	812.0	796	797.0	45.30
2015.0						
6	320.0	319.0	812.0	330	568.0	45.27
2015.0						
7	271.0	583.0	812.0	606	174.0	45.21
2015.0						
8	469.0	413.0	812.0	324	678.0	45.18
2015.0						
9	507.0	774.0	812.0	646	83.0	45.09
2015.0						

```
# 13
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select avg(influence) as
AVG_influence_score from world_rank where country = 'USA'"), conn)
    print(df)
```

	AVG_influence_score
0	277.301835

```
# 14
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
broad_impact <= 50"), conn)
    print(df)
```

	world_rank	institution \
0	1.0	Harvard University
1	2.0	Stanford University
2	3.0	Massachusetts Institute of Technology
3	4.0	University of Cambridge
4	5.0	University of Oxford
...	...	...
895	85.0	Erasmus University Rotterdam

896	87.0	Utrecht University
897	90.0	University of Texas MD Anderson Cancer Center
898	121.0	Baylor College of Medicine
899	139.0	University of Texas at Dallas

	country	national_rank	quality_of_education
alumni_employment \			
0	USA	1.0	1.0
1.0			
1	USA	2.0	11.0
2.0			
2	USA	3.0	3.0
11.0			
3	United Kingdom	1.0	2.0
10.0			
4	United Kingdom	2.0	7.0
12.0			
..	...	...	...
...			
895	Netherlands	1.0	367.0
46.0			
896	Netherlands	2.0	84.0
199.0			
897	USA	53.0	367.0
567.0			
898	USA	62.0	367.0
567.0			
899	USA	66.0	367.0
159.0			

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	1.0	1.0	1.0	1.0
1				
1	4.0	5.0	3.0	3.0
4				
2	2.0	15.0	2.0	2.0
2				
3	5.0	10.0	9.0	12.0
13				
4	10.0	11.0	12.0	11.0
12				
..	...	...	...	...
..				
895	189.0	87.0	147.0	53.0
44				
896	115.0	39.0	44.0	59.0
46				
897	80.0	93.0	72.0	42.0

```

29
898          218.0          113.0          41.0          101.0
44
899          218.0          250.0          151.0          59.0
49

```

```

      patents    score    year
0         2.0  100.00  2014.0
1         6.0   99.09  2014.0
2         1.0   98.69  2014.0
3        48.0   97.64  2014.0
4        16.0   97.51  2014.0
..         ...      ...      ...
895      227.0   51.78  2015.0
896      429.0   51.74  2015.0
897       71.0   51.51  2015.0
898      199.0   49.70  2015.0
899      313.0   49.05  2015.0

```

```
[900 rows x 14 columns]
```

```

# 15
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
country = 'Japan' and score > 60 and year = 2012"), conn)
    print(df)

```

```

      world_rank    institution country  national_rank \
0          14.0  University of Tokyo    Japan          1.0
1          17.0    Kyoto University    Japan          2.0
2          14.0  University of Tokyo    Japan          1.0
3          17.0    Kyoto University    Japan          2.0
4          14.0  University of Tokyo    Japan          1.0
5          17.0    Kyoto University    Japan          2.0
6          14.0  University of Tokyo    Japan          1.0
7          17.0    Kyoto University    Japan          2.0
8          14.0  University of Tokyo    Japan          1.0
9          17.0    Kyoto University    Japan          2.0
10         14.0  University of Tokyo    Japan          1.0
11         17.0    Kyoto University    Japan          2.0
12         14.0  University of Tokyo    Japan          1.0
13         17.0    Kyoto University    Japan          2.0
14         14.0  University of Tokyo    Japan          1.0
15         17.0    Kyoto University    Japan          2.0
16         14.0  University of Tokyo    Japan          1.0
17         17.0    Kyoto University    Japan          2.0

```

```

      quality_of_education  alumni_employment  quality_of_faculty
publications \
0              32.0              19.0              31.0

```

8.0			
1	42.0	38.0	19.0
25.0			
2	32.0	19.0	31.0
8.0			
3	42.0	38.0	19.0
25.0			
4	32.0	19.0	31.0
8.0			
5	42.0	38.0	19.0
25.0			
6	32.0	19.0	31.0
8.0			
7	42.0	38.0	19.0
25.0			
8	32.0	19.0	31.0
8.0			
9	42.0	38.0	19.0
25.0			
10	32.0	19.0	31.0
8.0			
11	42.0	38.0	19.0
25.0			
12	32.0	19.0	31.0
8.0			
13	42.0	38.0	19.0
25.0			
14	32.0	19.0	31.0
8.0			
15	42.0	38.0	19.0
25.0			
16	32.0	19.0	31.0
8.0			
17	42.0	38.0	19.0
25.0			

	influence	citations	broad_impact	patents	score	year
0	19.0	23.0	None	3.0	69.49	2012.0
1	36.0	43.0	None	23.0	65.76	2012.0
2	19.0	23.0	None	3.0	69.49	2012.0
3	36.0	43.0	None	23.0	65.76	2012.0
4	19.0	23.0	None	3.0	69.49	2012.0
5	36.0	43.0	None	23.0	65.76	2012.0
6	19.0	23.0	None	3.0	69.49	2012.0
7	36.0	43.0	None	23.0	65.76	2012.0
8	19.0	23.0	None	3.0	69.49	2012.0
9	36.0	43.0	None	23.0	65.76	2012.0
10	19.0	23.0	None	3.0	69.49	2012.0
11	36.0	43.0	None	23.0	65.76	2012.0

12	19.0	23.0	None	3.0	69.49	2012.0
13	36.0	43.0	None	23.0	65.76	2012.0
14	19.0	23.0	None	3.0	69.49	2012.0
15	36.0	43.0	None	23.0	65.76	2012.0
16	19.0	23.0	None	3.0	69.49	2012.0
17	36.0	43.0	None	23.0	65.76	2012.0

```
# 16
```

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 5 * FROM world_rank
WHERE year = 2013 ORDER BY patents DESC "), conn)
    print(df)
```

	world_rank	institution	country
national_rank \			
0	8.0	Princeton University	USA
6.0			
1	9.0	University of Chicago	USA
7.0			
2	29.0	University of Toronto	Canada
1.0			
3	36.0	University of Paris-Sud	France
1.0			
4	37.0	University of California, Santa Barbara	USA
25.0			

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	4.0	14.0	3.0
58.0			
1	9.0	19.0	8.0
37.0			
2	91.0	81.0	34.0
2.0			
3	26.0	101.0	26.0
73.0			
4	101.0	101.0	28.0
89.0			

	influence	citations	broad_impact	patents	score	year
0	35.0	27.0	None	101.0	82.17	2013.0
1	21.0	30.0	None	101.0	79.16	2013.0
2	13.0	12.0	None	101.0	56.11	2013.0
3	101.0	101.0	None	101.0	51.72	2013.0
4	91.0	38.0	None	101.0	51.67	2013.0

```
# 17
```

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(*) AS
quality_of_faculty FROM world_rank WHERE quality_of_faculty BETWEEN 5
```

```
AND 10"), conn)
```

```
print(df)
```

```
quality_of_faculty
0      216
```

```
# 18
```

```
with engine.begin() as conn:
```

```
df = pd.read_sql_query(sa.text(" SELECT country, AVG(score) AS  
Avg_score FROM world_rank WHERE country = 'United Kingdom' AND year =  
2014 GROUP BY country ")), conn)
```

```
print(df)
```

```
country Avg_score  
0  United Kingdom  48.445937
```

```
# 19
```

```
with engine.begin() as conn:
```

```
df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE  
year = 2012 AND national_rank BETWEEN 1 AND 3 ")), conn)
```

```
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	7.0	University of Oxford	United Kingdom
..	...	...	...
310	91.0	University of Oslo	Norway
311	92.0	University of Queensland	Australia
312	93.0	University of Copenhagen	Denmark
313	94.0	University of Sydney	Australia
314	99.0	Utrecht University	Netherlands

	national_rank	quality_of_education	alumni_employment	\
0	1.0	7.0	9.0	
1	2.0	9.0	17.0	
2	3.0	17.0	11.0	



3	1.0	10.0	24.0
4	2.0	13.0	28.0
..	...	...	...
310	1.0	101.0	101.0
311	1.0	101.0	101.0
312	1.0	101.0	101.0
313	2.0	101.0	101.0
314	2.0	100.0	101.0

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	1.0	1.0	1.0	1.0
None				
1	3.0	12.0	4.0	4.0
None				
2	5.0	4.0	2.0	2.0
None				
3	4.0	16.0	16.0	11.0
None				
4	9.0	15.0	13.0	19.0
None				
..	...	...	...	...
..				
310	39.0	101.0	101.0	101.0
None				
311	101.0	67.0	90.0	89.0
None				
312	60.0	85.0	66.0	101.0
None				
313	101.0	56.0	92.0	85.0
None				
314	101.0	65.0	101.0	60.0
None				

	patents	score	year
0	5.0	100.00	2012.0
1	1.0	91.67	2012.0
2	15.0	89.50	2012.0
3	50.0	86.17	2012.0
4	26.0	82.34	2012.0
..	...	...	...
310	101.0	44.26	2012.0
311	46.0	44.18	2012.0
312	101.0	44.15	2012.0
313	69.0	44.13	2012.0
314	101.0	43.47	2012.0

[315 rows x 14 columns]

```
# 20
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_Rank WHERE
citations < 10 "), conn)
    print(df)
```

	world_rank	institution	country
national_rank \			
0	1.0	Harvard University	USA
1.0			
1	2.0	Massachusetts Institute of Technology	USA
2.0			
2	3.0	Stanford University	USA
3.0			
3	10.0	University of California, Berkeley	USA
8.0			
4	13.0	University of Pennsylvania	USA
11.0			
..	...	...	...
...			
319	7.0	University of California, Berkeley	USA
5.0			
320	15.0	University of California, Los Angeles	USA
12.0			
321	16.0	Johns Hopkins University	USA
13.0			
322	19.0	University of Michigan, Ann Arbor	USA
15.0			
323	32.0	University of Toronto	Canada
1.0			

	quality_of_education	alumni_employment	quality_of_faculty	\
0	7.0	9.0	1.0	
1	9.0	17.0	3.0	
2	17.0	11.0	5.0	
3	16.0	52.0	6.0	
4	31.0	16.0	24.0	
..	...	...	...	
319	5.0	21.0	6.0	
320	28.0	27.0	13.0	
321	18.0	84.0	16.0	
322	24.0	17.0	140.0	
323	70.0	51.0	32.0	

	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	6.0	5.0	3.0	None	16.0	78.55
2012.0						
4	9.0	10.0	8.0	None	9.0	73.64
2012.0						
...	...	...	...	...	...	...
...						
319	10.0	4.0	4.0	7	29.0	92.25
2015.0						
320	6.0	14.0	8.0	6	9.0	76.91
2015.0						
321	4.0	11.0	5.0	3	2.0	71.60
2015.0						
322	3.0	20.0	6.0	14	12.0	68.36
2015.0						
323	2.0	17.0	9.0	8	145.0	60.04
2015.0						

[324 rows x 14 columns]

```
# 21
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 1 * FROM world_rank
WHERE year = 2013 ORDER BY alumni_employment DESC "), conn)
    print(df)
```

world_rank	institution	country
national_rank \		
0	11.0 California Institute of Technology	USA
9.0		

quality_of_education	alumni_employment	quality_of_faculty
publications \		
0	5.0	101.0
38.0		7.0

influence	citations	broad_impact	patents	score	year
0	24.0	25.0	None	17.0	77.59
					2013.0

```
# 22
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 5 * FROM world_rank
WHERE year = 2012 ORDER BY publications DESC "), conn)
    print(df)
```

world_rank	institution	country
\		
0	18.0 Weizmann Institute of Science	Israel
1	21.0 Rockefeller University	USA

2	22.0	Hebrew University of Jerusalem	Israel
3	29.0	University of Texas Southwestern Medical Center	USA
4	43.0	Carnegie Mellon University	USA

	national_rank	quality_of_education	alumni_employment
quality_of_faculty \			
0	1.0	4.0	101.0
22.0			
1	15.0	1.0	101.0
16.0			
2	2.0	24.0	93.0
13.0			
3	21.0	19.0	101.0
32.0			
4	31.0	30.0	81.0
26.0			

	publications	influence	citations	broad_impact	patents	score
year						
0	101.0	67.0	101.0	None	29.0	65.09
2012.0						
1	101.0	28.0	96.0	None	101.0	61.74
2012.0						
2	101.0	91.0	101.0	None	28.0	60.76
2012.0						
3	101.0	43.0	84.0	None	101.0	56.43
2012.0						
4	101.0	101.0	61.0	None	101.0	51.60
2012.0						

# 23

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT year, COUNT(*) AS
count_of_university FROM world_rank GROUP BY year "), conn)
    print(df)
```

	year	count_of_university
0	2015.0	8672
1	2014.0	9000
2	2013.0	900
3	2012.0	900

# 24

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("SELECT
ROUND(AVG(alumni_employment), 2) AS alumni_employment_score FROM
```

```
world_rank WHERE country = 'USA' "), conn)
print(df)
```

```
    alumni_employment_score
0                256.01
```

```
# 25
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
broad_impact > 70"), conn)
print(df)
```

	world_rank	institution \
0	18.0	Swiss Federal Institute of Technology in Zurich
1	22.0	Hebrew University of Jerusalem
2	24.0	Seoul National University
3	29.0	University of Texas at Austin
4	34.0	Keio University
...	...	...
16389	668.0	University of Maine, Orono
16390	669.0	Graz University of Technology
16391	670.0	Gifu University
16392	671.0	University of Jyväskylä
16393	672.0	University of Paris 13

	country	national_rank	quality_of_education
alumni_employment \			
0	Switzerland	1.0	16.0
105.0			
1	Israel	1.0	15.0
255.0			
2	South Korea	1.0	355.0
9.0			
3	USA	22.0	92.0
39.0			
4	Japan	3.0	271.0
5.0			
...	...	...	...
...			
16389	USA	194.0	345.0
567.0			
16390	Austria	8.0	367.0
567.0			
16391	Japan	43.0	367.0
567.0			
16392	Finland	8.0	367.0
525.0			
16393	France	29.0	367.0
567.0			

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	13.0	42.0	28.0	45.0
86				
1	16.0	114.0	94.0	493.0
151				
2	210.0	38.0	165.0	87.0
107				
3	24.0	51.0	54.0	26.0
74				
4	210.0	299.0	243.0	310.0
266				
...	...	...	...	...
...				
16389	218.0	788.0	427.0	368.0
637				
16390	218.0	677.0	715.0	368.0
622				
16391	218.0	701.0	658.0	645.0
606				
16392	218.0	456.0	443.0	812.0
622				
16393	218.0	669.0	744.0	645.0
579				

	patents	score	year
0	84.0	72.18	2014.0
1	40.0	66.76	2014.0
2	5.0	66.06	2014.0
3	45.0	62.57	2014.0
4	159.0	59.84	2014.0
...	...	...	...
16389	346.0	44.40	2015.0
16390	404.0	44.40	2015.0
16391	403.0	44.40	2015.0
16392	805.0	44.40	2015.0
16393	700.0	44.40	2015.0

[16394 rows x 14 columns]

```
# 26
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 10 * FROM world_rank
WHERE year = 2014 ORDER BY influence DESC "), conn)
    print(df)
```

world_rank	institution	country
national_rank \		
0	855.0	National Chung Cheng University Taiwan
14.0		

1	865.0	Nanjing University of Technology	China			
50.0						
2	904.0	Jadavpur University	India			
14.0						
3	928.0	Feng Chia University	Taiwan			
18.0						
4	947.0	Northwestern Polytechnical University	China			
67.0						
5	950.0	Chung Shan Medical University	Taiwan			
19.0						
6	960.0	Beijing Jiaotong University	China			
69.0						
7	975.0	China Pharmaceutical University	China			
72.0						
8	976.0	Tarbiat Modares University	Iran			
7.0						
9	979.0	University of Pau and Pays de l'Adour	France			
49.0						
quality_of_education alumni_employment quality_of_faculty publications \						
0	355.0	478.0	210.0			
760.0						
1	355.0	478.0	210.0			
810.0						
2	355.0	478.0	210.0			
871.0						
3	355.0	376.0	210.0			
869.0						
4	355.0	417.0	210.0			
628.0						
5	355.0	478.0	210.0			
750.0						
6	355.0	478.0	210.0			
971.0						
7	355.0	478.0	210.0			
919.0						
8	355.0	478.0	210.0			
811.0						
9	355.0	478.0	210.0			
988.0						
influence citations broad_impact patents score year						
0	987.0	800.0	783	737.0	44.44	2014.0
1	987.0	800.0	948	227.0	44.44	2014.0
2	987.0	493.0	849	737.0	44.39	2014.0
3	987.0	800.0	956	737.0	44.36	2014.0
4	987.0	800.0	985	737.0	44.33	2014.0
5	987.0	609.0	923	737.0	44.32	2014.0

6	987.0	800.0	997	280.0	44.31	2014.0
7	987.0	800.0	956	552.0	44.30	2014.0
8	987.0	800.0	956	737.0	44.29	2014.0
9	987.0	800.0	956	737.0	44.29	2014.0

```
# 27
```

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
patents = 1"), conn)
    print(df)
```

	world_rank national_rank \	institution	country
0	2.0	Massachusetts Institute of Technology	USA
2.0			
1	4.0	Massachusetts Institute of Technology	USA
3.0			
2	3.0	Massachusetts Institute of Technology	USA
3.0			
3	3.0	Massachusetts Institute of Technology	USA
3.0			
4	2.0	Massachusetts Institute of Technology	USA
2.0			
5	4.0	Massachusetts Institute of Technology	USA
3.0			
6	3.0	Massachusetts Institute of Technology	USA
3.0			
7	3.0	Massachusetts Institute of Technology	USA
3.0			
8	2.0	Massachusetts Institute of Technology	USA
2.0			
9	4.0	Massachusetts Institute of Technology	USA
3.0			
10	3.0	Massachusetts Institute of Technology	USA
3.0			
11	3.0	Massachusetts Institute of Technology	USA
3.0			
12	2.0	Massachusetts Institute of Technology	USA
2.0			
13	4.0	Massachusetts Institute of Technology	USA
3.0			
14	3.0	Massachusetts Institute of Technology	USA
3.0			
15	3.0	Massachusetts Institute of Technology	USA
3.0			
16	2.0	Massachusetts Institute of Technology	USA
2.0			
17	4.0	Massachusetts Institute of Technology	USA
3.0			
18	3.0	Massachusetts Institute of Technology	USA



3.0			
19	3.0	Massachusetts Institute of Technology	USA
3.0			
20	2.0	Massachusetts Institute of Technology	USA
2.0			
21	4.0	Massachusetts Institute of Technology	USA
3.0			
22	3.0	Massachusetts Institute of Technology	USA
3.0			
23	3.0	Massachusetts Institute of Technology	USA
3.0			
24	2.0	Massachusetts Institute of Technology	USA
2.0			
25	4.0	Massachusetts Institute of Technology	USA
3.0			
26	3.0	Massachusetts Institute of Technology	USA
3.0			
27	3.0	Massachusetts Institute of Technology	USA
3.0			
28	2.0	Massachusetts Institute of Technology	USA
2.0			
29	4.0	Massachusetts Institute of Technology	USA
3.0			
30	3.0	Massachusetts Institute of Technology	USA
3.0			
31	3.0	Massachusetts Institute of Technology	USA
3.0			
32	2.0	Massachusetts Institute of Technology	USA
2.0			
33	4.0	Massachusetts Institute of Technology	USA
3.0			
34	3.0	Massachusetts Institute of Technology	USA
3.0			
35	3.0	Massachusetts Institute of Technology	USA
3.0			

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	9.0	17.0	3.0
12.0			
1	2.0	16.0	2.0
16.0			
2	3.0	11.0	2.0
15.0			
3	3.0	11.0	2.0
15.0			
4	9.0	17.0	3.0
12.0			
5	2.0	16.0	2.0

16.0			
6	3.0	11.0	2.0
15.0			
7	3.0	11.0	2.0
15.0			
8	9.0	17.0	3.0
12.0			
9	2.0	16.0	2.0
16.0			
10	3.0	11.0	2.0
15.0			
11	3.0	11.0	2.0
15.0			
12	9.0	17.0	3.0
12.0			
13	2.0	16.0	2.0
16.0			
14	3.0	11.0	2.0
15.0			
15	3.0	11.0	2.0
15.0			
16	9.0	17.0	3.0
12.0			
17	2.0	16.0	2.0
16.0			
18	3.0	11.0	2.0
15.0			
19	3.0	11.0	2.0
15.0			
20	9.0	17.0	3.0
12.0			
21	2.0	16.0	2.0
16.0			
22	3.0	11.0	2.0
15.0			
23	3.0	11.0	2.0
15.0			
24	9.0	17.0	3.0
12.0			
25	2.0	16.0	2.0
16.0			
26	3.0	11.0	2.0
15.0			
27	3.0	11.0	2.0
15.0			
28	9.0	17.0	3.0
12.0			
29	2.0	16.0	2.0
16.0			

30	3.0	11.0	2.0
15.0			
31	3.0	11.0	2.0
15.0			
32	9.0	17.0	3.0
12.0			
33	2.0	16.0	2.0
16.0			
34	3.0	11.0	2.0
15.0			
35	3.0	11.0	2.0
15.0			

	influence	citations	broad_impact	patents	score	year
0	4.0	4.0	None	1.0	91.67	2012.0
1	3.0	3.0	None	1.0	91.45	2013.0
2	2.0	2.0	2	1.0	98.69	2014.0
3	2.0	2.0	2	1.0	97.54	2015.0
4	4.0	4.0	None	1.0	91.67	2012.0
5	3.0	3.0	None	1.0	91.45	2013.0
6	2.0	2.0	2	1.0	98.69	2014.0
7	2.0	2.0	2	1.0	97.54	2015.0
8	4.0	4.0	None	1.0	91.67	2012.0
9	3.0	3.0	None	1.0	91.45	2013.0
10	2.0	2.0	2	1.0	98.69	2014.0
11	2.0	2.0	2	1.0	97.54	2015.0
12	4.0	4.0	None	1.0	91.67	2012.0
13	3.0	3.0	None	1.0	91.45	2013.0
14	2.0	2.0	2	1.0	98.69	2014.0
15	2.0	2.0	2	1.0	97.54	2015.0
16	4.0	4.0	None	1.0	91.67	2012.0
17	3.0	3.0	None	1.0	91.45	2013.0
18	2.0	2.0	2	1.0	98.69	2014.0
19	2.0	2.0	2	1.0	97.54	2015.0
20	4.0	4.0	None	1.0	91.67	2012.0
21	3.0	3.0	None	1.0	91.45	2013.0
22	2.0	2.0	2	1.0	98.69	2014.0
23	2.0	2.0	2	1.0	97.54	2015.0
24	4.0	4.0	None	1.0	91.67	2012.0
25	3.0	3.0	None	1.0	91.45	2013.0
26	2.0	2.0	2	1.0	98.69	2014.0
27	2.0	2.0	2	1.0	97.54	2015.0
28	4.0	4.0	None	1.0	91.67	2012.0
29	3.0	3.0	None	1.0	91.45	2013.0
30	2.0	2.0	2	1.0	98.69	2014.0
31	2.0	2.0	2	1.0	97.54	2015.0
32	4.0	4.0	None	1.0	91.67	2012.0
33	3.0	3.0	None	1.0	91.45	2013.0

34	2.0	2.0	2	1.0	98.69	2014.0
35	2.0	2.0	2	1.0	97.54	2015.0

```
# 28
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(*) AS
broad_impact_count FROM world_rank WHERE broad_impact BETWEEN 20 AND
30 "), conn)
    print(df)
```

	broad_impact_count
0	189

```
# 29
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(quality_of_faculty) AS
avg_quality_of_faculty FROM world_rank WHERE country = 'United
Kingdom' "), conn)
    print(df)
```

	avg_quality_of_faculty
0	166.068804

```
# 30
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
influence < 5 "), conn)
    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	24.0	University of California, San Francisco	USA	
4	1.0	Harvard University	USA	
...	...			
139	7.0	University of California, Berkeley	USA	
140	1.0	Harvard University	USA	
141	2.0	Stanford University	USA	
142	3.0	Massachusetts Institute of Technology	USA	
143	7.0	University of California, Berkeley	USA	

	national_rank	quality_of_education	alumni_employment	\
0	1.0	7.0	9.0	
1	2.0	9.0	17.0	
2	3.0	17.0	11.0	
3	17.0	101.0	101.0	
4	1.0	1.0	1.0	
...	...			
139	5.0	4.0	22.0	
140	1.0	1.0	1.0	

141	2.0	9.0	2.0
142	3.0	3.0	11.0
143	5.0	5.0	21.0

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	1.0	1.0	1.0	1.0
None				
1	3.0	12.0	4.0	4.0
None				
2	5.0	4.0	2.0	2.0
None				
3	21.0	19.0	3.0	13.0
None				
4	1.0	1.0	1.0	1.0
None				
..	...	...	...	...
..				
139	6.0	7.0	4.0	3.0
7				
140	1.0	1.0	1.0	1.0
1				
141	4.0	5.0	3.0	3.0
4				
142	2.0	15.0	2.0	2.0
2				
143	6.0	10.0	4.0	4.0
7				

	patents	score	year
0	5.0	100.00	2012.0
1	1.0	91.67	2012.0
2	15.0	89.50	2012.0
3	33.0	59.70	2012.0
4	7.0	100.00	2013.0
..	...	...	...
139	28.0	92.84	2014.0
140	3.0	100.00	2015.0
141	10.0	98.66	2015.0
142	1.0	97.54	2015.0
143	29.0	92.25	2015.0

[144 rows x 14 columns]

```
# 31
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
score > 75 AND national_rank <= 10 AND year = 2013 "), conn)
    print(df)
```

	world_rank	institution	country
\			
0	1.0	Harvard University	USA
1	2.0	Stanford University	USA
2	3.0	University of Oxford	United Kingdom
3	4.0	Massachusetts Institute of Technology	USA
4	5.0	University of Cambridge	United Kingdom
..	...	...	...
112	9.0	University of Chicago	USA
113	10.0	Yale University	USA
114	11.0	California Institute of Technology	USA
115	12.0	University of Pennsylvania	USA
116	14.0	University of Tokyo	Japan

	national_rank	quality_of_education	alumni_employment	\
0	1.0	1.0	1.0	
1	2.0	11.0	2.0	
2	1.0	7.0	12.0	
3	3.0	2.0	16.0	
4	2.0	3.0	15.0	
..	...	...	...	
112	7.0	9.0	19.0	
113	8.0	10.0	25.0	
114	9.0	5.0	101.0	
115	10.0	22.0	5.0	
116	1.0	17.0	3.0	

	quality_of_faculty	publications	influence	citations	broad_impact	\
0	1.0	1.0	1.0	1.0		
None						
1	4.0	6.0	2.0	2.0		
None						
2	10.0	11.0	7.0	13.0		
None						
3	2.0	16.0	3.0	3.0		
None						
4	5.0	9.0	11.0	10.0		
None						
..	...	...	...	...		.

```

..
112          8.0          37.0          21.0          30.0
None
113          13.0         20.0          12.0          22.0
None
114          7.0          38.0          24.0          25.0
None
115          25.0          10.0          10.0           9.0
None
116          32.0          14.0          23.0          29.0
None

```

```

      patents  score  year
0         7.0  100.00  2013.0
1        11.0   93.94  2013.0
2        15.0   92.54  2013.0
3         1.0   91.45  2013.0
4        39.0   90.24  2013.0
..
112       101.0   79.16  2013.0
113       42.0   78.83  2013.0
114       17.0   77.59  2013.0
115       14.0   77.24  2013.0
116        5.0   76.23  2013.0

```

[117 rows x 14 columns]

```

# 32
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 5 * FROM world_rank
WHERE year = 2014 ORDER BY patents DESC "), conn)
    print(df)

```

```

      world_rank      institution country
national_rank \
0         35.0  Åcole normale supÅrieure - Paris  France
1.0
1         48.0  Lomonosov Moscow State University  Russia
1.0
2         73.0      Karolinska Institute  Sweden
1.0
3        116.0      Lund University  Sweden
3.0
4        124.0      Uppsala University  Sweden
4.0

```

```

      quality_of_education  alumni_employment  quality_of_faculty
publications \
0          8.0          478.0          59.0
362.0

```

1	14.0	210.0	50.0
264.0			
2	126.0	478.0	37.0
55.0			
3	132.0	478.0	165.0
81.0			
4	62.0	336.0	111.0
94.0			

	influence	citations	broad_impact	patents	score	year
0	156.0	493.0	311	737.0	59.72	2014.0
1	247.0	310.0	361	737.0	56.42	2014.0
2	61.0	38.0	36	737.0	53.64	2014.0
3	133.0	48.0	70	737.0	50.34	2014.0
4	77.0	99.0	102	737.0	49.96	2014.0

```
# 33
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(*) AS count_of_univ
FROM world_rank WHERE alumni_employment > 50 "), conn)
    print(df)
```

	count_of_univ
0	18113

```
# 34
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(quality_of_education)
AS avg_education_japan FROM world_rank WHERE country = 'Japan' "),
conn)
    print(df)
```

	avg_education_japan
0	290.470714

```
# 35
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
quality_of_education = 1 AND year = 2014 "), conn)
    print(df)
```

	world_rank	institution	country	national_rank	\
0	1.0	Harvard University	USA	1.0	
1	1.0	Harvard University	USA	1.0	
2	1.0	Harvard University	USA	1.0	
3	1.0	Harvard University	USA	1.0	
4	1.0	Harvard University	USA	1.0	
5	1.0	Harvard University	USA	1.0	
6	1.0	Harvard University	USA	1.0	
7	1.0	Harvard University	USA	1.0	
8	1.0	Harvard University	USA	1.0	



	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	1.0	1.0	1.0
1.0			
1	1.0	1.0	1.0
1.0			
2	1.0	1.0	1.0
1.0			
3	1.0	1.0	1.0
1.0			
4	1.0	1.0	1.0
1.0			
5	1.0	1.0	1.0
1.0			
6	1.0	1.0	1.0
1.0			
7	1.0	1.0	1.0
1.0			
8	1.0	1.0	1.0
1.0			

	influence	citations	broad_impact	patents	score	year
0	1.0	1.0	1	2.0	100.0	2014.0
1	1.0	1.0	1	2.0	100.0	2014.0
2	1.0	1.0	1	2.0	100.0	2014.0
3	1.0	1.0	1	2.0	100.0	2014.0
4	1.0	1.0	1	2.0	100.0	2014.0
5	1.0	1.0	1	2.0	100.0	2014.0
6	1.0	1.0	1	2.0	100.0	2014.0
7	1.0	1.0	1	2.0	100.0	2014.0
8	1.0	1.0	1	2.0	100.0	2014.0

# 36

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
score > 80 AND alumni_employment < 5 "), conn)
    print(df)
```

	world_rank	institution	country	national_rank
0	1.0	Harvard University	USA	1.0
1	2.0	Stanford University	USA	2.0
2	1.0	Harvard University	USA	1.0
3	2.0	Stanford University	USA	2.0
4	13.0	University of Tokyo	Japan	1.0
...	...	...	...	...
58	1.0	Harvard University	USA	1.0
59	2.0	Stanford University	USA	2.0
60	13.0	University of Tokyo	Japan	1.0
61	1.0	Harvard University	USA	1.0

62	2.0	Stanford University	USA	2.0
	quality_of_education	alumni_employment	quality_of_faculty	
publications \				
0	1.0	1.0	1.0	
1.0				
1	11.0	2.0	4.0	
6.0				
2	1.0	1.0	1.0	
1.0				
3	11.0	2.0	4.0	
5.0				
4	17.0	3.0	33.0	
12.0				
..	...	...	...	
...				
58	1.0	1.0	1.0	
1.0				
59	11.0	2.0	4.0	
5.0				
60	17.0	3.0	33.0	
12.0				
61	1.0	1.0	1.0	
1.0				
62	9.0	2.0	4.0	
5.0				

	influence	citations	broad_impact	patents	score	year
0	1.0	1.0	None	7.0	100.00	2013.0
1	2.0	2.0	None	11.0	93.94	2013.0
2	1.0	1.0	1	2.0	100.00	2014.0
3	3.0	3.0	4	6.0	99.09	2014.0
4	16.0	28.0	26	14.0	80.64	2014.0
..	...	...	...	...	...	...
58	1.0	1.0	1	2.0	100.00	2014.0
59	3.0	3.0	4	6.0	99.09	2014.0
60	16.0	28.0	26	14.0	80.64	2014.0
61	1.0	1.0	1	3.0	100.00	2015.0
62	3.0	3.0	4	10.0	98.66	2015.0

[63 rows x 14 columns]

```
# 37
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT TOP 10 * FROM world_rank
WHERE year = 2013 ORDER BY publications DESC "), conn)
    print(df)
```

world_rank	institution
country \	

0	21.0	Hebrew University of Jerusalem
Israel		
1	33.0	Weizmann Institute of Science
Israel		
2	46.0	University of Texas Southwestern Medical Center
USA		
3	49.0	Rockefeller University
USA		
4	51.0	Carnegie Mellon University
USA		
5	56.0	University of Geneva
Switzerland		
6	64.0	École normale supérieure - Paris
France		
7	66.0	Technion " Israel Institute of Technology
Israel		
8	70.0	Keio University
Japan		
9	73.0	Arizona State University
USA		

	national_rank	quality_of_education	alumni_employment
quality_of_faculty \			
0	1.0	15.0	101.0
15.0			
1	2.0	18.0	101.0
23.0			
2	32.0	46.0	101.0
29.0			
3	33.0	55.0	101.0
17.0			
4	35.0	30.0	101.0
27.0			
5	2.0	52.0	101.0
30.0			
6	3.0	8.0	101.0
57.0			
7	4.0	50.0	101.0
36.0			
8	4.0	101.0	6.0
101.0			
9	46.0	101.0	101.0
73.0			

	publications	influence	citations	broad_impact	patents	score
year						
0	101.0	96.0	101.0	None	24.0	59.98
2013.0						
1	101.0	85.0	101.0	None	48.0	54.34

2013.0						
2	101.0	28.0	82.0	None	101.0	50.79
2013.0						
3	101.0	34.0	101.0	None	101.0	50.43
2013.0						
4	101.0	101.0	74.0	None	101.0	49.56
2013.0						
5	101.0	78.0	101.0	None	101.0	48.83
2013.0						
6	101.0	101.0	101.0	None	101.0	47.70
2013.0						
7	101.0	101.0	101.0	None	73.0	47.45
2013.0						
8	101.0	101.0	101.0	None	73.0	47.05
2013.0						
9	101.0	101.0	89.0	None	20.0	46.72
2013.0						

```
# 38
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT COUNT(*) AS count_of_univ
FROM world_rank WHERE broad_impact BETWEEN 40 AND 50 "), conn)
    print(df)
```

	count_of_univ
0	198

```
# 39
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT AVG(score) AS
avg_score_of_univ FROM world_rank WHERE country = 'Australia' "),
conn)
    print(df)
```

	avg_score_of_univ
0	45.850584

```
# 40
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" SELECT * FROM world_rank WHERE
influence = 1 AND year = 2012 "), conn)
    print(df)
```

	world_rank	institution	country	national_rank	\
0	1.0	Harvard University	USA	1.0	
1	1.0	Harvard University	USA	1.0	
2	1.0	Harvard University	USA	1.0	
3	1.0	Harvard University	USA	1.0	
4	1.0	Harvard University	USA	1.0	
5	1.0	Harvard University	USA	1.0	
6	1.0	Harvard University	USA	1.0	

7	1.0	Harvard University	USA	1.0
8	1.0	Harvard University	USA	1.0

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	7.0	9.0	1.0
1.0			
1	7.0	9.0	1.0
1.0			
2	7.0	9.0	1.0
1.0			
3	7.0	9.0	1.0
1.0			
4	7.0	9.0	1.0
1.0			
5	7.0	9.0	1.0
1.0			
6	7.0	9.0	1.0
1.0			
7	7.0	9.0	1.0
1.0			
8	7.0	9.0	1.0
1.0			

	influence	citations	broad_impact	patents	score	year
0	1.0	1.0	None	5.0	100.0	2012.0
1	1.0	1.0	None	5.0	100.0	2012.0
2	1.0	1.0	None	5.0	100.0	2012.0
3	1.0	1.0	None	5.0	100.0	2012.0
4	1.0	1.0	None	5.0	100.0	2012.0
5	1.0	1.0	None	5.0	100.0	2012.0
6	1.0	1.0	None	5.0	100.0	2012.0
7	1.0	1.0	None	5.0	100.0	2012.0
8	1.0	1.0	None	5.0	100.0	2012.0

```
# 41
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text(" select * from world_rank where
quality_of_faculty > 15 "),conn)
    print(df)
```

	world_rank	institution	country
national_rank \			
0	13.0	University of Pennsylvania	USA
11.0			
1	14.0	University of Tokyo	Japan
1.0			
2	15.0	Johns Hopkins University	USA
12.0			
3	17.0	Kyoto University	Japan

2.0						
4	18.0	Weizmann Institute of Science	Israel			
1.0						
...	...			...	...	.
..						
18927	668.0	University of Maine, Orono	USA			
194.0						
18928	669.0	Graz University of Technology	Austria			
8.0						
18929	670.0	Gifu University	Japan			
43.0						
18930	671.0	University of Jyväskylä	Finland			
8.0						
18931	672.0	University of Paris 13	France			
29.0						

	quality_of_education	alumni_employment	quality_of_faculty	\
0	31.0	16.0	24.0	
1	32.0	19.0	31.0	
2	34.0	77.0	20.0	
3	42.0	38.0	19.0	
4	4.0	101.0	22.0	
...	...	...	...	
18927	345.0	567.0	218.0	
18928	367.0	567.0	218.0	
18929	367.0	567.0	218.0	
18930	367.0	525.0	218.0	
18931	367.0	567.0	218.0	

	publications	influence	citations	broad_impact	patents	score
year						
0	9.0	10.0	8.0	None	9.0	73.64
2012.0						
1	8.0	19.0	23.0	None	3.0	69.49
2012.0						
2	11.0	9.0	9.0	None	7.0	66.94
2012.0						
3	25.0	36.0	43.0	None	23.0	65.76
2012.0						
4	101.0	67.0	101.0	None	29.0	65.09
2012.0						
...	...	...	...	...	...	...
...						
18927	788.0	427.0	368.0	637	346.0	44.40
2015.0						
18928	677.0	715.0	368.0	622	404.0	44.40
2015.0						
18929	701.0	658.0	645.0	606	403.0	44.40
2015.0						

18930	456.0	443.0	812.0	622	805.0	44.40
2015.0						
18931	669.0	744.0	645.0	579	700.0	44.40
2015.0						

[18932 rows x 14 columns]

# 42

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select top 5 * from world_rank
where year = 2014 order by alumni_employment desc"), conn)
    print(df)
```

	world_rank		institution	country
national_rank \				
0	20.0		University of California, San Diego	USA
15.0				
1	27.0		University of California, San Francisco	USA
20.0				
2	35.0		École normale supérieure - Paris	France
1.0				
3	37.0		Rockefeller University	USA
25.0				
4	38.0		Weizmann Institute of Science	Israel
2.0				

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	32.0	478.0	20.0
16.0			
1	355.0	478.0	23.0
24.0			
2	8.0	478.0	59.0
362.0			
3	72.0	478.0	14.0
347.0			
4	18.0	478.0	22.0
233.0			

	influence	citations	broad_impact	patents	score	year
0	5.0	19.0	16	13.0	68.36	2014.0
1	8.0	17.0	9	57.0	63.36	2014.0
2	156.0	493.0	311	737.0	59.72	2014.0
3	26.0	107.0	62	227.0	59.49	2014.0
4	63.0	250.0	137	51.0	59.17	2014.0

# 43

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select count(*) as
nos_of_University from world_rank where national_rank < 20 and year =
```

```
2013"), conn)
    print(df)
```

```
   nos_of_University
0                558
```

```
# 44
```

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select avg(citations) as
avg_citations from world_rank where country = 'Germany'"), conn)
    print(df)
```

```
   avg_citations
0      362.476608
```

```
# 45
```

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
publications = 1"), conn)
    print(df)
```

	world_rank	institution	country	national_rank	\
0	1.0	Harvard University	USA	1.0	
1	1.0	Harvard University	USA	1.0	
2	1.0	Harvard University	USA	1.0	
3	1.0	Harvard University	USA	1.0	
4	1.0	Harvard University	USA	1.0	
5	1.0	Harvard University	USA	1.0	
6	1.0	Harvard University	USA	1.0	
7	1.0	Harvard University	USA	1.0	
8	1.0	Harvard University	USA	1.0	
9	1.0	Harvard University	USA	1.0	
10	1.0	Harvard University	USA	1.0	
11	1.0	Harvard University	USA	1.0	
12	1.0	Harvard University	USA	1.0	
13	1.0	Harvard University	USA	1.0	
14	1.0	Harvard University	USA	1.0	
15	1.0	Harvard University	USA	1.0	
16	1.0	Harvard University	USA	1.0	
17	1.0	Harvard University	USA	1.0	
18	1.0	Harvard University	USA	1.0	
19	1.0	Harvard University	USA	1.0	
20	1.0	Harvard University	USA	1.0	
21	1.0	Harvard University	USA	1.0	
22	1.0	Harvard University	USA	1.0	
23	1.0	Harvard University	USA	1.0	
24	1.0	Harvard University	USA	1.0	
25	1.0	Harvard University	USA	1.0	
26	1.0	Harvard University	USA	1.0	
27	1.0	Harvard University	USA	1.0	



28	1.0	Harvard University	USA	1.0
29	1.0	Harvard University	USA	1.0
30	1.0	Harvard University	USA	1.0
31	1.0	Harvard University	USA	1.0
32	1.0	Harvard University	USA	1.0
33	1.0	Harvard University	USA	1.0
34	1.0	Harvard University	USA	1.0
35	1.0	Harvard University	USA	1.0

	quality_of_education	alumni_employment	quality_of_faculty
publications \			
0	7.0	9.0	1.0
1.0			
1	1.0	1.0	1.0
1.0			
2	1.0	1.0	1.0
1.0			
3	1.0	1.0	1.0
1.0			
4	7.0	9.0	1.0
1.0			
5	1.0	1.0	1.0
1.0			
6	1.0	1.0	1.0
1.0			
7	1.0	1.0	1.0
1.0			
8	7.0	9.0	1.0
1.0			
9	1.0	1.0	1.0
1.0			
10	1.0	1.0	1.0
1.0			
11	1.0	1.0	1.0
1.0			
12	7.0	9.0	1.0
1.0			
13	1.0	1.0	1.0
1.0			
14	1.0	1.0	1.0
1.0			
15	1.0	1.0	1.0
1.0			
16	7.0	9.0	1.0
1.0			
17	1.0	1.0	1.0
1.0			
18	1.0	1.0	1.0
1.0			

19	1.0	1.0	1.0
1.0			
20	7.0	9.0	1.0
1.0			
21	1.0	1.0	1.0
1.0			
22	1.0	1.0	1.0
1.0			
23	1.0	1.0	1.0
1.0			
24	7.0	9.0	1.0
1.0			
25	1.0	1.0	1.0
1.0			
26	1.0	1.0	1.0
1.0			
27	1.0	1.0	1.0
1.0			
28	7.0	9.0	1.0
1.0			
29	1.0	1.0	1.0
1.0			
30	1.0	1.0	1.0
1.0			
31	1.0	1.0	1.0
1.0			
32	7.0	9.0	1.0
1.0			
33	1.0	1.0	1.0
1.0			
34	1.0	1.0	1.0
1.0			
35	1.0	1.0	1.0
1.0			

	influence	citations	broad_impact	patents	score	year
0	1.0	1.0	None	5.0	100.0	2012.0
1	1.0	1.0	None	7.0	100.0	2013.0
2	1.0	1.0	1	2.0	100.0	2014.0
3	1.0	1.0	1	3.0	100.0	2015.0
4	1.0	1.0	None	5.0	100.0	2012.0
5	1.0	1.0	None	7.0	100.0	2013.0
6	1.0	1.0	1	2.0	100.0	2014.0
7	1.0	1.0	1	3.0	100.0	2015.0
8	1.0	1.0	None	5.0	100.0	2012.0
9	1.0	1.0	None	7.0	100.0	2013.0
10	1.0	1.0	1	2.0	100.0	2014.0
11	1.0	1.0	1	3.0	100.0	2015.0
12	1.0	1.0	None	5.0	100.0	2012.0

13	1.0	1.0	None	7.0	100.0	2013.0
14	1.0	1.0	1	2.0	100.0	2014.0
15	1.0	1.0	1	3.0	100.0	2015.0
16	1.0	1.0	None	5.0	100.0	2012.0
17	1.0	1.0	None	7.0	100.0	2013.0
18	1.0	1.0	1	2.0	100.0	2014.0
19	1.0	1.0	1	3.0	100.0	2015.0
20	1.0	1.0	None	5.0	100.0	2012.0
21	1.0	1.0	None	7.0	100.0	2013.0
22	1.0	1.0	1	2.0	100.0	2014.0
23	1.0	1.0	1	3.0	100.0	2015.0
24	1.0	1.0	None	5.0	100.0	2012.0
25	1.0	1.0	None	7.0	100.0	2013.0
26	1.0	1.0	1	2.0	100.0	2014.0
27	1.0	1.0	1	3.0	100.0	2015.0
28	1.0	1.0	None	5.0	100.0	2012.0
29	1.0	1.0	None	7.0	100.0	2013.0
30	1.0	1.0	1	2.0	100.0	2014.0
31	1.0	1.0	1	3.0	100.0	2015.0
32	1.0	1.0	None	5.0	100.0	2012.0
33	1.0	1.0	None	7.0	100.0	2013.0
34	1.0	1.0	1	2.0	100.0	2014.0
35	1.0	1.0	1	3.0	100.0	2015.0

# 46

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
broad_impact > 60 and national_rank <= 5"), conn)
    print(df)
```

	world_rank	institution \
0	18.0	Swiss Federal Institute of Technology in Zurich
1	22.0	Hebrew University of Jerusalem
2	24.0	Seoul National University
3	34.0	Keio University
4	35.0	École normale supérieure - Paris
...	...	...
3414	641.0	University of Patras
3415	646.0	Warsaw University of Technology
3416	652.0	University of Los Andes (Colombia)
3417	658.0	Massey University
3418	667.0	University of Debrecen

	country	national_rank	quality_of_education
alumni_employment	\		
0	Switzerland	1.0	16.0
105.0			
1	Israel	1.0	15.0
255.0			
2	South Korea	1.0	355.0

9.0			
3	Japan	3.0	271.0
5.0			
4	France	1.0	8.0
478.0			
...	...	...	...
...			
3414	Greece	5.0	367.0
567.0			
3415	Poland	3.0	344.0
545.0			
3416	Colombia	1.0	367.0
189.0			
3417	New Zealand	5.0	367.0
496.0			
3418	Hungary	3.0	367.0
567.0			

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	13.0	42.0	28.0	45.0
86				
1	16.0	114.0	94.0	493.0
151				
2	210.0	38.0	165.0	87.0
107				
3	210.0	299.0	243.0	310.0
266				
4	59.0	362.0	156.0	493.0
311				
...	...	...	...	...
...				
3414	218.0	510.0	704.0	511.0
565				
3415	218.0	696.0	757.0	287.0
579				
3416	218.0	898.0	561.0	428.0
837				
3417	218.0	570.0	554.0	812.0
579				
3418	218.0	687.0	621.0	428.0
590				

	patents	score	year
0	84.0	72.18	2014.0
1	40.0	66.76	2014.0
2	5.0	66.06	2014.0
3	159.0	59.84	2014.0
4	737.0	59.72	2014.0

```

...
3414      731.0    44.44    2015.0
3415      824.0    44.43    2015.0
3416      749.0    44.43    2015.0
3417      723.0    44.42    2015.0
3418      499.0    44.40    2015.0

```

```
[3419 rows x 14 columns]
```

```
# 47
```

```

with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select top 10 * from world_rank
where year = 2012 order by quality_of_education desc"), conn)
    print(df)

```

	world_rank	institution	country	\
0	24.0	University of California, San Francisco	USA	
1	30.0	University of Texas at Austin	USA	
2	33.0	Northwestern University	USA	
3	35.0	University of Toronto	Canada	
4	36.0	University of North Carolina at Chapel Hill	USA	
5	39.0	University of Washington - Seattle	USA	
6	40.0	University of California, Santa Barbara	USA	
7	44.0	University of Southern California	USA	
8	47.0	University of California, Irvine	USA	
9	49.0	University of Minnesota, Twin Cities	USA	

	national_rank	quality_of_education	alumni_employment
0	17.0	101.0	101.0
1	22.0	101.0	78.0
2	23.0	101.0	32.0
3	1.0	101.0	101.0
4	25.0	101.0	86.0
5	28.0	101.0	101.0
6	29.0	101.0	101.0
7	32.0	101.0	101.0
8	35.0	101.0	101.0
9	36.0	101.0	101.0

	publications	influence	citations	broad_impact	patents	score
year						
0	19.0	3.0	13.0	None	33.0	59.70
2012.0						
1	41.0	47.0	40.0	None	57.0	56.18
2012.0						
2	24.0	25.0	20.0	None	35.0	54.40
2012.0						
3	7.0	14.0	18.0	None	101.0	53.43
2012.0						
4	31.0	29.0	31.0	None	29.0	53.09
2012.0						
5	5.0	7.0	5.0	None	101.0	52.25
2012.0						
6	68.0	72.0	36.0	None	101.0	52.15
2012.0						
7	46.0	48.0	32.0	None	23.0	51.38
2012.0						
8	59.0	57.0	52.0	None	65.0	50.64
2012.0						
9	18.0	31.0	17.0	None	84.0	50.30
2012.0						

# 48

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select count(*) as
score_greaterThan_90 from world_rank where score > 90"), conn)
    print(df)
```

	score_greaterThan_90
0	207

# 49

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select avg(influence) as
avg_influence from world_rank where country = 'United Kingdom'"),
conn)
    print(df)
```

	avg_influence
0	343.802971

# 50

```
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select * from world_rank where
quality_of_education <= 5 and year = 2014"), conn)
    print(df)
```

	world_rank	institution	country
0	1.0	Harvard University	USA

1	3.0	Massachusetts Institute of Technology	USA
2	4.0	University of Cambridge	United Kingdom
3	7.0	University of California, Berkeley	USA
4	9.0	Princeton University	USA
5	1.0	Harvard University	USA
6	3.0	Massachusetts Institute of Technology	USA
7	4.0	University of Cambridge	United Kingdom
8	7.0	University of California, Berkeley	USA
9	9.0	Princeton University	USA
10	1.0	Harvard University	USA
11	3.0	Massachusetts Institute of Technology	USA
12	4.0	University of Cambridge	United Kingdom
13	7.0	University of California, Berkeley	USA
14	9.0	Princeton University	USA
15	1.0	Harvard University	USA
16	3.0	Massachusetts Institute of Technology	USA
17	4.0	University of Cambridge	United Kingdom
18	7.0	University of California, Berkeley	USA
19	9.0	Princeton University	USA
20	1.0	Harvard University	USA
21	3.0	Massachusetts Institute of Technology	USA
22	4.0	University of Cambridge	United Kingdom
23	7.0	University of California, Berkeley	USA
24	9.0	Princeton University	USA
25	1.0	Harvard University	USA
26	3.0	Massachusetts Institute of Technology	USA

27	4.0	University of Cambridge	United Kingdom
28	7.0	University of California, Berkeley	USA
29	9.0	Princeton University	USA
30	1.0	Harvard University	USA
31	3.0	Massachusetts Institute of Technology	USA
32	4.0	University of Cambridge	United Kingdom
33	7.0	University of California, Berkeley	USA
34	9.0	Princeton University	USA
35	1.0	Harvard University	USA
36	3.0	Massachusetts Institute of Technology	USA
37	4.0	University of Cambridge	United Kingdom
38	7.0	University of California, Berkeley	USA
39	9.0	Princeton University	USA
40	1.0	Harvard University	USA
41	3.0	Massachusetts Institute of Technology	USA
42	4.0	University of Cambridge	United Kingdom
43	7.0	University of California, Berkeley	USA
44	9.0	Princeton University	USA
	national_rank	quality_of_education	alumni_employment \
0	1.0	1.0	1.0
1	3.0	3.0	11.0
2	1.0	2.0	10.0
3	5.0	4.0	22.0
4	7.0	5.0	16.0
5	1.0	1.0	1.0
6	3.0	3.0	11.0
7	1.0	2.0	10.0
8	5.0	4.0	22.0
9	7.0	5.0	16.0
10	1.0	1.0	1.0
11	3.0	3.0	11.0



12	1.0	2.0	10.0
13	5.0	4.0	22.0
14	7.0	5.0	16.0
15	1.0	1.0	1.0
16	3.0	3.0	11.0
17	1.0	2.0	10.0
18	5.0	4.0	22.0
19	7.0	5.0	16.0
20	1.0	1.0	1.0
21	3.0	3.0	11.0
22	1.0	2.0	10.0
23	5.0	4.0	22.0
24	7.0	5.0	16.0
25	1.0	1.0	1.0
26	3.0	3.0	11.0
27	1.0	2.0	10.0
28	5.0	4.0	22.0
29	7.0	5.0	16.0
30	1.0	1.0	1.0
31	3.0	3.0	11.0
32	1.0	2.0	10.0
33	5.0	4.0	22.0
34	7.0	5.0	16.0
35	1.0	1.0	1.0
36	3.0	3.0	11.0
37	1.0	2.0	10.0
38	5.0	4.0	22.0
39	7.0	5.0	16.0
40	1.0	1.0	1.0
41	3.0	3.0	11.0
42	1.0	2.0	10.0
43	5.0	4.0	22.0
44	7.0	5.0	16.0

	quality_of_faculty	publications	influence	citations
broad_impact \				
0	1.0	1.0	1.0	1.0
1				
1	2.0	15.0	2.0	2.0
2				
2	5.0	10.0	9.0	12.0
13				
3	6.0	7.0	4.0	3.0
7				
4	3.0	70.0	25.0	19.0
41				
5	1.0	1.0	1.0	1.0
1				
6	2.0	15.0	2.0	2.0

2				
7	5.0	10.0	9.0	12.0
13				
8	6.0	7.0	4.0	3.0
7				
9	3.0	70.0	25.0	19.0
41				
10	1.0	1.0	1.0	1.0
1				
11	2.0	15.0	2.0	2.0
2				
12	5.0	10.0	9.0	12.0
13				
13	6.0	7.0	4.0	3.0
7				
14	3.0	70.0	25.0	19.0
41				
15	1.0	1.0	1.0	1.0
1				
16	2.0	15.0	2.0	2.0
2				
17	5.0	10.0	9.0	12.0
13				
18	6.0	7.0	4.0	3.0
7				
19	3.0	70.0	25.0	19.0
41				
20	1.0	1.0	1.0	1.0
1				
21	2.0	15.0	2.0	2.0
2				
22	5.0	10.0	9.0	12.0
13				
23	6.0	7.0	4.0	3.0
7				
24	3.0	70.0	25.0	19.0
41				
25	1.0	1.0	1.0	1.0
1				
26	2.0	15.0	2.0	2.0
2				
27	5.0	10.0	9.0	12.0
13				
28	6.0	7.0	4.0	3.0
7				
29	3.0	70.0	25.0	19.0
41				
30	1.0	1.0	1.0	1.0
1				

31	2.0	15.0	2.0	2.0
2				
32	5.0	10.0	9.0	12.0
13				
33	6.0	7.0	4.0	3.0
7				
34	3.0	70.0	25.0	19.0
41				
35	1.0	1.0	1.0	1.0
1				
36	2.0	15.0	2.0	2.0
2				
37	5.0	10.0	9.0	12.0
13				
38	6.0	7.0	4.0	3.0
7				
39	3.0	70.0	25.0	19.0
41				
40	1.0	1.0	1.0	1.0
1				
41	2.0	15.0	2.0	2.0
2				
42	5.0	10.0	9.0	12.0
13				
43	6.0	7.0	4.0	3.0
7				
44	3.0	70.0	25.0	19.0
41				

	patents	score	year
0	2.0	100.00	2014.0
1	1.0	98.69	2014.0
2	48.0	97.64	2014.0
3	28.0	92.84	2014.0
4	204.0	88.56	2014.0
5	2.0	100.00	2014.0
6	1.0	98.69	2014.0
7	48.0	97.64	2014.0
8	28.0	92.84	2014.0
9	204.0	88.56	2014.0
10	2.0	100.00	2014.0
11	1.0	98.69	2014.0
12	48.0	97.64	2014.0
13	28.0	92.84	2014.0
14	204.0	88.56	2014.0
15	2.0	100.00	2014.0
16	1.0	98.69	2014.0
17	48.0	97.64	2014.0
18	28.0	92.84	2014.0

19	204.0	88.56	2014.0
20	2.0	100.00	2014.0
21	1.0	98.69	2014.0
22	48.0	97.64	2014.0
23	28.0	92.84	2014.0
24	204.0	88.56	2014.0
25	2.0	100.00	2014.0
26	1.0	98.69	2014.0
27	48.0	97.64	2014.0
28	28.0	92.84	2014.0
29	204.0	88.56	2014.0
30	2.0	100.00	2014.0
31	1.0	98.69	2014.0
32	48.0	97.64	2014.0
33	28.0	92.84	2014.0
34	204.0	88.56	2014.0
35	2.0	100.00	2014.0
36	1.0	98.69	2014.0
37	48.0	97.64	2014.0
38	28.0	92.84	2014.0
39	204.0	88.56	2014.0
40	2.0	100.00	2014.0
41	1.0	98.69	2014.0
42	48.0	97.64	2014.0
43	28.0	92.84	2014.0
44	204.0	88.56	2014.0

```
# test_query
with engine.begin() as conn:
    df = pd.read_sql_query(sa.text("select count(*) from world_rank
where country = 'Japan' and year = 2012"), conn)
    print(df)
```

0 45

## Task 5 - Python

```
# 1
total_count = len(data)
print("Total records in dataset :", total_count)
```

Total records in dataset : 19472

```
# 2
cols = data.shape[1]
print("Total Columns :", cols)
```

Total Columns : 14

```
# 3
typ = data['world_rank'].dtype
print("Datatype of world_rank :", typ)

Datatype of world_rank : int64

# 4
ctr = data['country'].mode().iloc[0]
print("Most featured country :", ctr)

Most featured country : USA

# 5
avg_edu = data['quality_of_education'].mean()
print("Average of quality_of_education : ", avg_edu)

Average of quality_of_education : 273.8289338537387
```

```
# 6
inst = data.loc[data['alumni_employment'].idxmax()]
inst

world_rank                21
institution    University of California, San Diego
country                USA
national_rank            16
quality_of_education      36
alumni_employment        567
quality_of_faculty        19
publications              16
influence                  5
citations                 15
broad_impact              15.0
patents                   17
score                     66.59
year                      2015
Name: 1220, dtype: object
```

```
# 7
endd = max(data['year'])
print("Dataset ends in year :", endd)

Dataset ends in year : 2015
```

```
# 8
grp = data.groupby('country')
countt = len(grp)
print("Total unique countries in dataset :", countt)

Total unique countries in dataset : 59
```

# 9

```
filter_year = 2012
print("Institution with highest score in year", filter_year)
temp_data = data[data['year'] == filter_year]
temp_data.loc[data['score'].idxmax()]
```

Institution with highest score in year 2012

world_rank	1
institution	Harvard University
country	USA
national_rank	1
quality_of_education	7
alumni_employment	9
quality_of_faculty	1
publications	1
influence	1
citations	1
broad_impact	NaN
patents	5
score	100.0
year	2012

Name: 0, dtype: object

# 10

```
year2012 = 2012
temp_data = data.loc[data['year'] == filter_year]
temp2 = temp_data.loc[data['institution'] == 'Harvard University']
print("National rank for 'Harvard University' in year", year2012, '=',
temp2['world_rank'].iloc[0])
```

National rank for 'Harvard University' in year 2012 = 1

# 11

```
data_uk = data.loc[data['country'] == 'United Kingdom']
avg_value = data_uk['publications'].mean()
print("Avegrae publications value for institutions in United
Kingdom :", avg_value)
```

Avegrae publications value for institutions in United Kingdom :  
391.01016419077405

# 12

```
year2013 = 2013
print("University with highest influence in :", year2013)
dt = data.loc[data['year'] == year2013]
dt.loc[dt['influence'].idxmax()]
```

University with highest influence in : 2013

world_rank	36
institution	University of Paris-Sud
country	France
national_rank	1
quality_of_education	26
alumni_employment	101
quality_of_faculty	26
publications	73
influence	101
citations	101
broad_impact	NaN
patents	101
score	51.72
year	2013

Name: 135, dtype: object

# 13

```
bi = data['broad_impact']
mn = bi.dropna().min()
print("Minimum broad impact value in dataset : " ,mn)
```

Minimum broad impact value in dataset : 1.0

# 14

```
yr12 = 2012
jp_data = data.loc[(data['country'] == 'Japan') & (data['year'] == yr12)]
print(f"Total number of university in Japan in year {yr12} is : {len(jp_data)}")
```

Total number of university in Japan in year 2012 is : 45

# 15

```
dt_usa = data.loc[(data['country'] == 'USA') & (data['year'] == yr12)]
print(f"Avegrage patents value for institutions in USA in year {yr12} {dt_usa['patents'].mean()}")
```

Avegrage patents value for institutions in USA in year 2012  
56.293103448275865

# 16

```
yr15 = 2015
cit = data.loc[data['year'] == yr15]
cit_lrg = cit.nlargest(5, 'citations')
cit_lrg
```

country \	world_rank	institution	
1303	104	Mines ParisTech	France
1425	226	National Cheng Kung University	Taiwan
1449	250	Moscow Institute of Physics and Technology	Russia
1540	341	Indian Institute of Technology Delhi	India
1550	351	Wesleyan University	USA

	national_rank	quality_of_education	alumni_employment \
1303	5	278	23
1425	2	367	52
1449	2	23	339
1540	1	367	59
1550	124	154	61

broad_impact \	quality_of_faculty	publications	influence	citations
1303	169	977	834	812
906.0				
1425	218	164	485	812
388.0				
1449	218	949	605	812
1000.0				
1540	218	635	943	812
781.0				
1550	218	983	506	812
850.0				

	patents	score	year
1303	871	50.34	2015
1425	436	46.97	2015
1449	839	46.55	2015
1540	625	45.54	2015
1550	871	45.48	2015

# 17

# there is no such column : "international\_students"

# 18

# there is no income column but we can consider : low score = low income

yr14 = 2014

inc = data.loc[data['year'] == yr14]

income = inc.nsmallest(1, 'score')



```
print(f"Institution with lowest income in {yr14}")
income
```

Institution with lowest income in 2014

	world_rank	institution	country	national_rank	\
1199	1000	Yanbian University	China	84	

	quality_of_education	alumni_employment	quality_of_faculty	\
1199	355	478	210	

	publications	influence	citations	broad_impact	patents	score
year						
1199	890	790	800	1000.0	737	44.18
2014						

# 19

```
miss_val = data['score'].isnull().sum()
print(f"Total missing values in score column : {miss_val}")
```

Total missing values in score column : 0

# 20

```
c_count = data['country'].value_counts()
c_count.head(3)
```

USA	5122
China	1457
Japan	1400

Name: country, dtype: int64

# 21

```
res_scr = data.loc[data['score'] > 70]
x = len(res_scr)
y = len(data)
```

```
percentage = (x/y)*100
```

```
print(f"percentage of institutions with score greater than 70 : {percentage:.2f}%")
```

percentage of institutions with score greater than 70 : 2.87%

# 22

*# there is no year 2016 in data so i am calculating for year 2015*  
*# aslo we dont have teaching score column so : teaching score = quality\_of\_education*

```
yr15 = 2015
```

```
MIT = 'Massachusetts Institute of Technology'
```

```

SU = 'Stanford University'

data15 = data.loc[data['year'] == yr15]

mit = data15.loc[data15['institution'] == MIT].head(1)
su = data15.loc[data15['institution'] == SU].head(1)

mit_score = mit['quality_of_education'].values[0]
su_score = su['quality_of_education'].values[0]

score = (su_score - mit_score)

print(f"Difference between teaching score between MIT and SU :
{score}")

Difference between teaching score between MIT and SU : 6

# 22 OR

# there is no year 2016 in data so i am calculating for year 2015
# aslo we dont have teaching score column so : teaching score = score

yr15 = 2015
MIT = 'Massachusetts Institute of Technology'
SU = 'Stanford University'

data15 = data.loc[data['year'] == yr15]

mit = data15.loc[data15['institution'] == MIT].head(1)
su = data15.loc[data15['institution'] == SU].head(1)

mit_score = mit['score'].values[0]
su_score = su['score'].values[0]

score = (su_score - mit_score)

print(f"Difference between teaching score between MIT and SU :
{score:.2f}")

Difference between teaching score between MIT and SU : 1.12

# 23
# there is no data for year 2011, but just to get ouput i am trying
with year 2012

rank = len(data.loc[(data['world_rank'] >= 50) & (data['world_rank']
<=100 ) & (data['year'] == 2012)])
print(f"Total institutions with world rank between 50 and 100 in year
2012 : {rank}")

```

Total institutions with world rank between 50 and 100 in year 2012 :  
459

# 24

```
avg_scores = data.groupby('country')['score'].mean()
print("according to the highest average score, country with the  
highest average industry_income across all years : ")
print(f"{avg_scores.idxmax()} : {avg_scores.max():.2f} ")
```

according to the highest average score, country with the highest  
average industry\_income across all years :  
Israel : 52.70

# 25

```
usa_data = data.loc[data['country'] == 'USA']
std = usa_data['influence'].std()
print(f" standard deviation of research scores for institutions in the  
United States : {std:.2f} ")
```

standard deviation of research scores for institutions in the United  
States : 254.65

# 26

```
data15 = data.loc[data['year'] == yr15]
print("Institution with highest alumni_employment in 2105 : ")
data15.nlargest(1, 'alumni_employment')
```

Institution with highest alumni\_employment in 2105 :

	world_rank		institution	country
national_rank \				
1220	21	University of California, San Diego	USA	
16				

	quality_of_education	alumni_employment	quality_of_faculty \
1220	36	567	19

	publications	influence	citations	broad_impact	patents	score
year						
1220	16	5	15	15.0	17	66.59
2015						

# 27

# lets assume research = patents

```
corr = data['score'].corr(data['patents'])
print(f"Correlation between 'Score' and 'Research' : {corr} ")
```

Correlation between 'Score' and 'Research' : -0.4737216892717171

# 28

```
data14 = data.loc[data['year'] == yr14]
data14.loc[data['broad_impact'] > 800]
```

	world_rank	institution
country \		
366	167	Mines ParisTech
France		
506	307	Wesleyan University
USA		
514	315	University of Science, Malaysia
Malaysia		
619	420	King Saud University Saudi
Arabia		
627	428	Sogang University South
Korea		
...	...	...
...		
18795	996	National Dong Hwa University
Taiwan		
18796	997	National Taipei University of Technology
Taiwan		
18797	998	Shaanxi Normal University
China		
18798	999	National University of Defense Technology
China		
18799	1000	Yanbian University
China		

	national_rank	quality_of_education	alumni_employment \
366	6	249	31
506	117	130	54
514	1	355	478
619	1	355	133
627	14	355	478
...	...	...	...
18795	24	355	478
18796	25	355	478
18797	82	355	478
18798	83	355	478
18799	84	355	478

	quality_of_faculty	publications	influence	citations
broad_impact \				
366	165	980	825	609
917.0				
506	210	986	538	609
832.0				
514	210	700	877	609

819.0				
619	210	611	537	800
897.0				
627	210	781	792	800
889.0				
...	...	...	...	...
...				
18795	210	901	934	800
989.0				
18796	210	867	987	800
994.0				
18797	210	956	965	800
994.0				
18798	210	860	973	800
999.0				
18799	210	890	790	800
1000.0				

	patents	score	year
366	737	48.81	2014
506	737	46.30	2014
514	42	46.23	2014
619	115	45.45	2014
627	68	45.43	2014
...	...	...	...
18795	737	44.24	2014
18796	737	44.24	2014
18797	737	44.23	2014
18798	637	44.21	2014
18799	737	44.18	2014

[1638 rows x 14 columns]

# 29

# lets assume 1 patent for every student

```
aus_data = data.loc[data['country'] == 'USA']
avg_stud = aus_data['patents'].mean()
print(f"Avegrage international_student percentage for Australia is
{avg_stud:.2f}%")
```

Avegrage international\_student percentage for Australia is 293.75%

# 30

# there is no year 2016 in dataset so answer will be (null)

```
yr16 = 2016
data16 = data.loc[data['year'] == yr16]
high_scr_16 = data16.nlargest(1, 'score')
```

```
print(f"Institution with highest score in year 2016 is : ")
high_scr_16
```

Institution with highest score in year 2016 is :

Empty DataFrame

Columns: [world\_rank, institution, country, national\_rank, quality\_of\_education, alumni\_employment, quality\_of\_faculty, publications, influence, citations, broad\_impact, patents, score, year]

Index: []

*# 31*

*# there is female\_male\_ratio column in dataset*

*# 32*

yr13 = 2013

data13 = data.loc[data['year'] == yr13]

scr\_90 = data13.loc[data13['score'] > 90]

```
print(f"Institutions having score greater than 90 in the year 2013 : {len(scr_90)}")
```

Institutions having score greater than 90 in the year 2013 : 45

*# 33*

*# there is no research\_score column in dataset*

top\_10 = data15.nlargest(10, 'score')

avg\_top10 = top\_10['influence'].mean()

```
print(f"average research score for institutions in the top 10 of score in 2015 : {avg_top10}")
```

average research score for institutions in the top 10 of score in 2015 : 1.2

*# 34*

*#there is no international\_students column in dataset*

*# 35*

*# there is income coumn in dataset*

*# 36*

res = len(data14.loc[data14['national\_rank'] == 1])

```
print(f" Number of institutions having national_rank of 1 in the year 2014 : {res}")
```

Number of institutions having national\_rank of 1 in the year 2014 : 531

```
# 37
# there is no information about international students in dataset
# 38
# there is no data of 2016 and there is no column teaching score
# 39
```

```
high_scr = int(data['score'].nlargest(1))
res14 = data14.loc[data14['score'] == high_scr]
print("institution with the highest score in the year 2014")
res14
```

institution with the highest score in the year 2014

	world_rank	institution	country	national_rank	\
200	1	Harvard University	USA	1	
2400	1	Harvard University	USA	1	
4600	1	Harvard University	USA	1	
6800	1	Harvard University	USA	1	
9000	1	Harvard University	USA	1	
11200	1	Harvard University	USA	1	
13400	1	Harvard University	USA	1	
15600	1	Harvard University	USA	1	
17800	1	Harvard University	USA	1	

	quality_of_education	alumni_employment	quality_of_faculty	\
200	1	1	1	
2400	1	1	1	
4600	1	1	1	
6800	1	1	1	
9000	1	1	1	
11200	1	1	1	
13400	1	1	1	
15600	1	1	1	
17800	1	1	1	

	publications	influence	citations	broad_impact	patents
score	year				
200		1	1	1	1.0 2
100.0	2014				
2400		1	1	1	1.0 2
100.0	2014				
4600		1	1	1	1.0 2
100.0	2014				
6800		1	1	1	1.0 2
100.0	2014				
9000		1	1	1	1.0 2
100.0	2014				

11200		1	1	1	1.0	2
100.0	2014					
13400		1	1	1	1.0	2
100.0	2014					
15600		1	1	1	1.0	2
100.0	2014					
17800		1	1	1	1.0	2
100.0	2014					

# 40

```
res_15 = len(data15.loc[data15['world_rank'] > 200])
print(f"Number of institutions with world_rank greater than 200 in
year 2015 : {res_15}")
```

Number of institutions with world\_rank greater than 200 in year 2015 :  
6872

# 41

```
mode_country = data['country'].mode().tolist()
print(f"Mode of country column : {mode_country}")
```

Mode of country column : ['USA']

# 42

*# there is industry\_income column in dataset*

# 43

```
bott_15 = data13.nlargest(10, 'world_rank')
avg_cit = bott_15['citations'].mean()
print(f"Average citation score for bottom 10 institutions :
{avg_cit}")
```

Average citation score for bottom 10 institutions : 48.8

# 44

*# there is no data available for 2016*

```
data16 = data.loc[data['year'] == 2016]
res80 = data16.loc[data16['score'] > 80]
print(f"Number of institutions with score greater than 80 in 2016")
res80
```

Number of institutions with score greater than 80 in 2016

Empty DataFrame

Columns: [world\_rank, institution, country, national\_rank, quality\_of\_education, alumni\_employment, quality\_of\_faculty, publications, influence, citations, broad\_impact, patents, score,



```

year]
Index: []

# 45

grp_data = data.groupby('country')['score'].mean()
res = grp_data.idxmin()
res2 = grp_data.min()
print(f"country with lowest average score of all time = {res} : {res2}")

```

country with lowest average score of all time = Romania : 44.1408

```

# 46
# there is international_students column in table

```

```

# 47

res47 = data15.nlargest(1,'quality_of_faculty')
print(f"institution with the lowest teaching score in the year 2015 :")
res47

```

institution with the lowest teaching score in the year 2015 :

	world_rank	institution	country
national_rank \			
1223	24	Seoul National University	South Korea

1

	quality_of_education	alumni_employment	quality_of_faculty \
1223	367	9	218

	publications	influence	citations	broad_impact	patents	score
year						
1223	36	163	146	112.0	6	64.82
2015						

```

# 48
# there is no research column

ms_val = data['patents'].isnull().sum()
print(f"Total missing values in patents column : {ms_val} ")

Total missing values in patents column : 0

```

```

# 49
# there is no female_male_ratio column in dataset

```

```

# 50

```

```
res50 = data['influence'].max()
print(f"Highest influence score in dataset : {res50}")
```

Highest influence score in dataset : 991

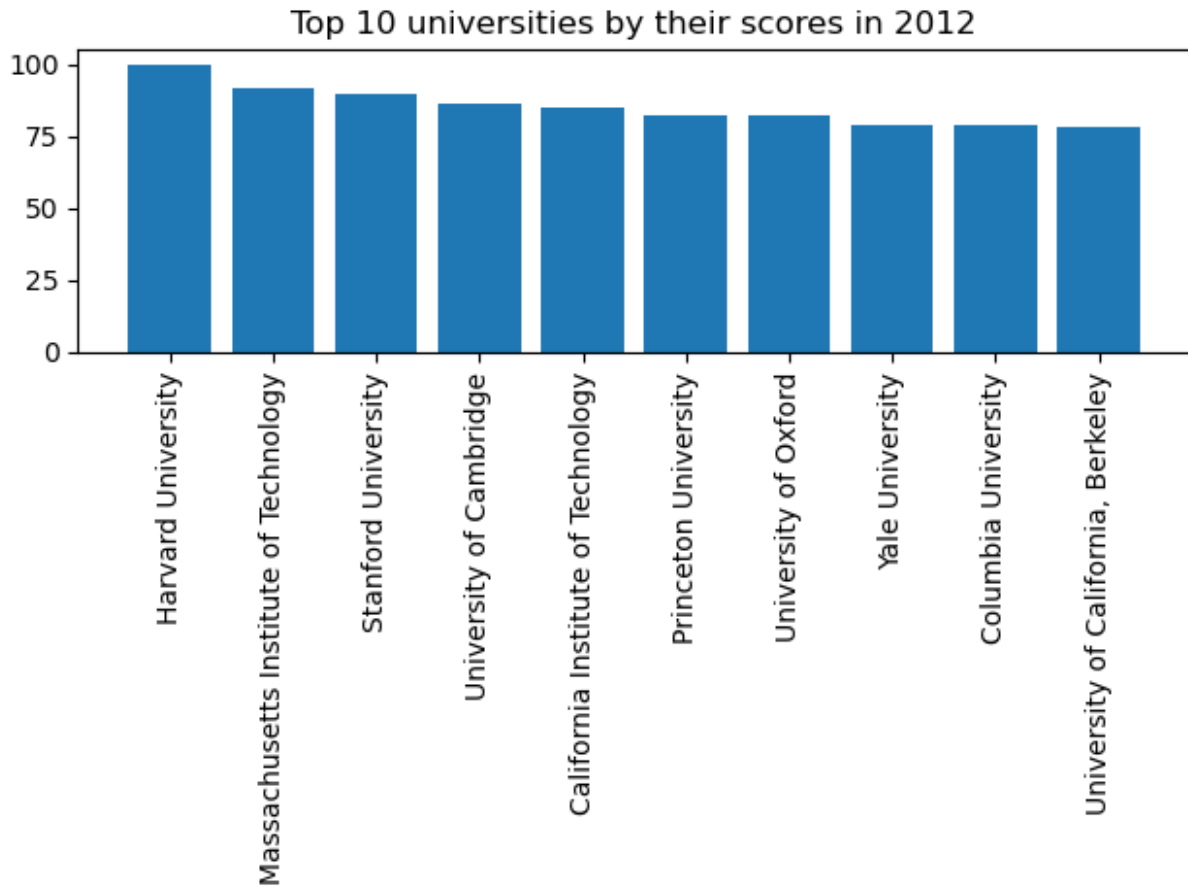
## Task 6 - Visualization

```
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib.ticker import MaxNLocator
import altair as alt

# 1
data12 = data.loc[data['year'] == 2012]
data121 = data12.drop_duplicates()
data_top10 = data121.nlargest(10, 'score')

x_ax = data_top10['institution']
y_ax = data_top10['score']

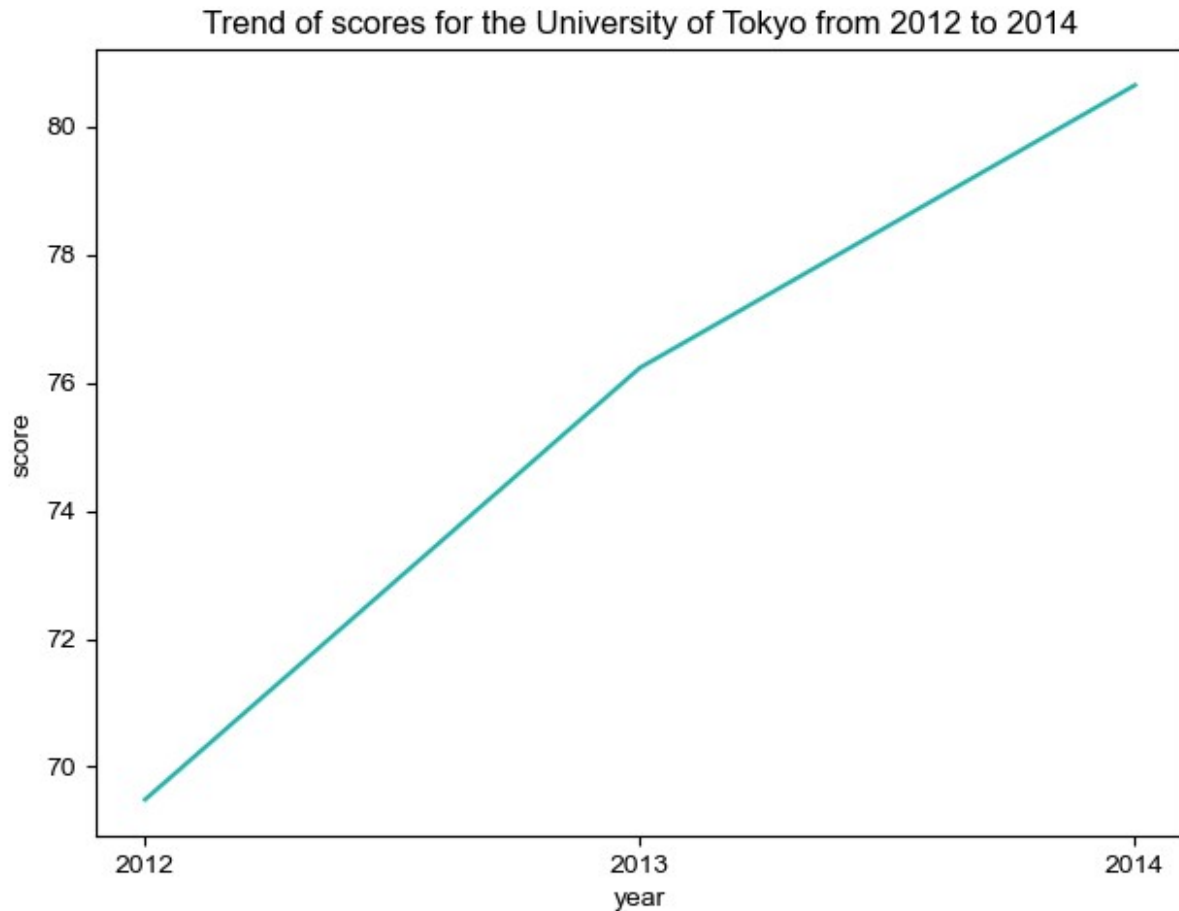
plt.bar(x_ax , y_ax)
plt.title('Top 10 universities by their scores in 2012')
plt.xticks(rotation='vertical')
plt.tight_layout()
```



```
# 2
```

```
yr_data = data.loc[data['year'].isin([2012,2013,2014])]
tokyo_data = yr_data.loc[yr_data['institution'] == 'University of Tokyo']

sline = sns.lineplot(data = tokyo_data, x = 'year', y = 'score', color = 'lightseagreen')
sline.xaxis.set_major_locator(MaxNLocator(integer=True))
sns.set(rc={'figure.figsize':(7,3)})
plt.tight_layout()
plt.title('Trend of scores for the University of Tokyo from 2012 to 2014')
plt.show()
```

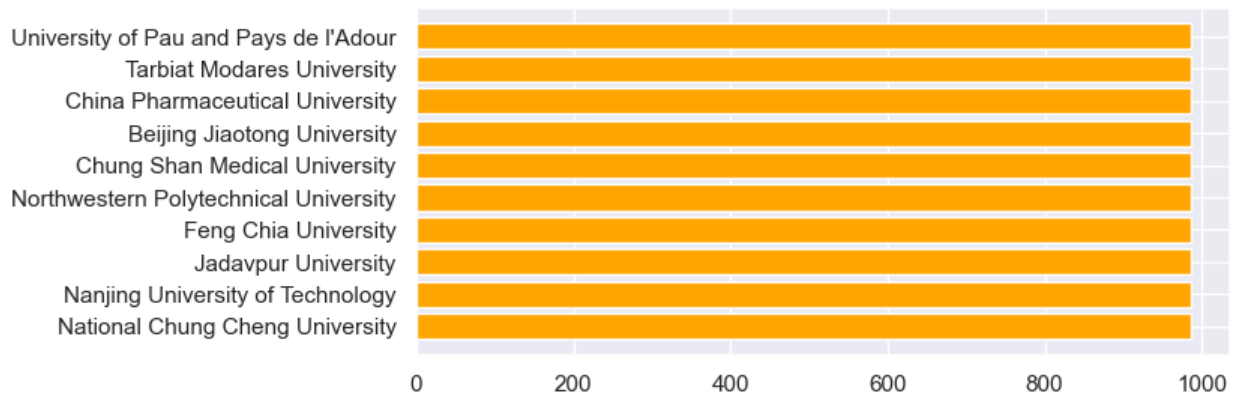


```
# 3
data13 = data.loc[data['year'] == 2013]
alt.Chart(data13).mark_circle(size = 50).encode(
    x = 'quality_of_faculty',
    y = 'alumni_employment', color = 'country').interactive()
alt.Chart(...)
```

```
# 4
data14 = data.loc[data['year'] == 2014]
dt_inf = data14.nlargest(10, 'influence')

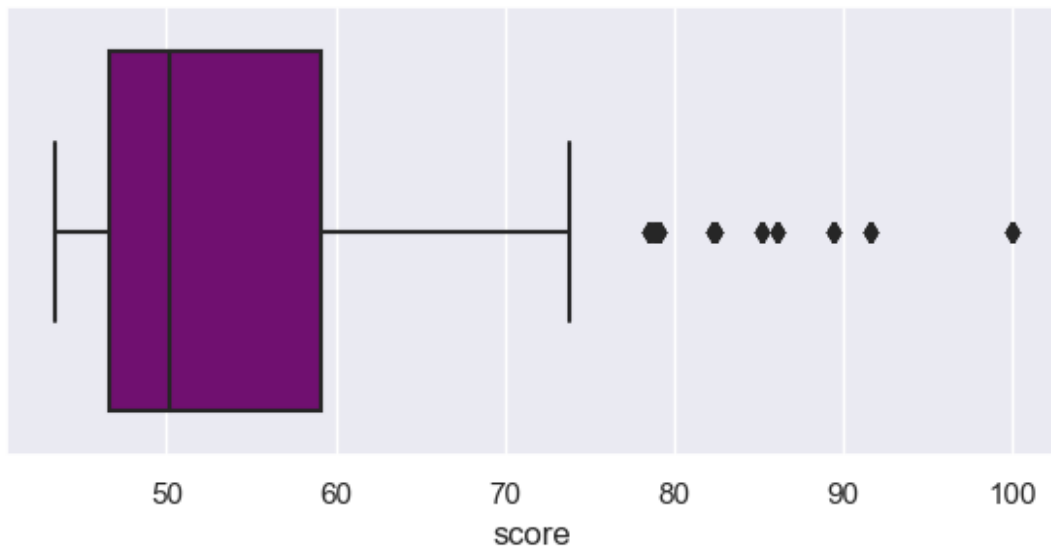
inst = dt_inf['institution']
inf = dt_inf['influence']

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



# 5

```
data12 = data.loc[data['year'] == 2012]
sns.boxplot(x='score', data=data12, color='purple')
plt.show()
```



# 6

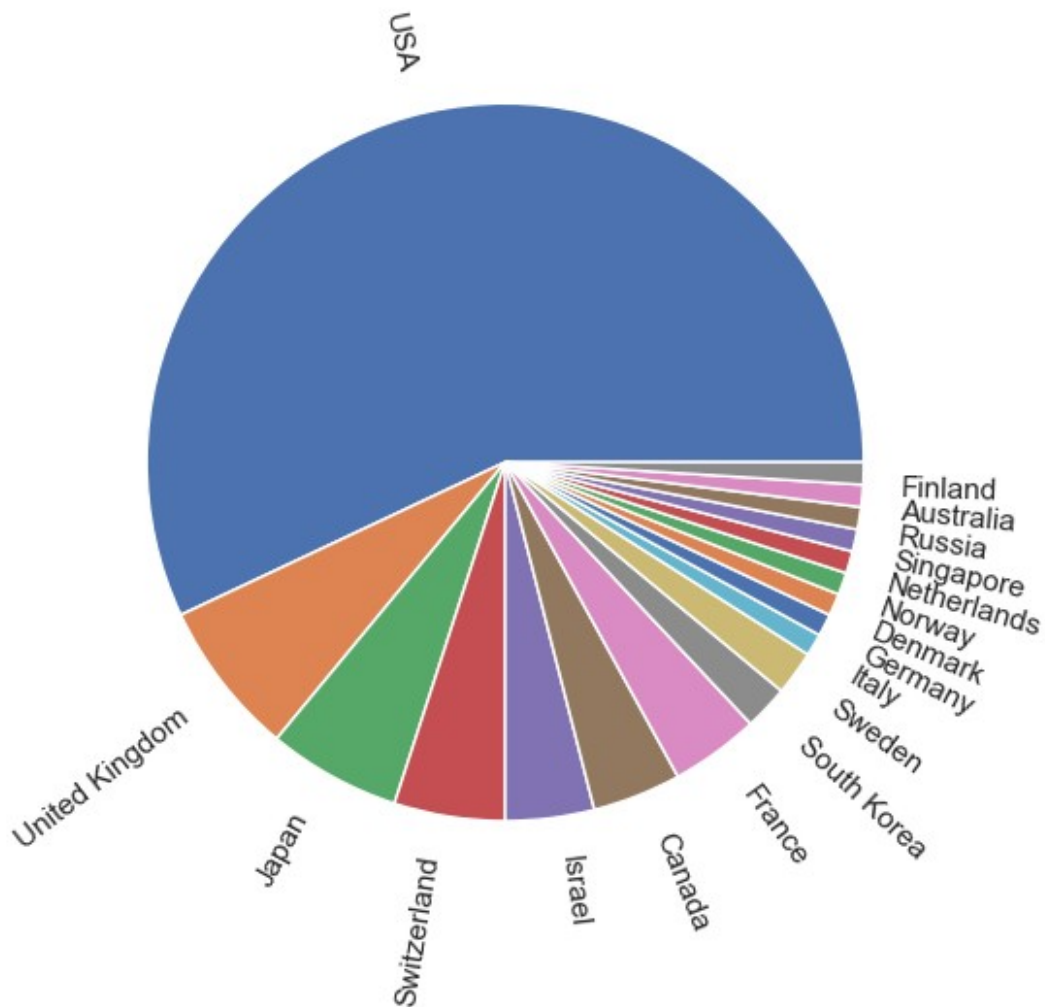
```
temp6 = yr_data.drop_duplicates()
top5uni = temp6.nsmallest(5, 'world_rank')
alt.Chart(top5uni).mark_area().encode(
    x = 'year',
    y = 'score',
    color='institution:N'
).properties(width = 800, title = 'Change in scores for the top 5
universities from 2012 to 2014.')
```

```
alt.Chart(...)
```

```
# 7
```

```
x = data13['country'].value_counts()  
lbl = data13['country'].drop_duplicates()
```

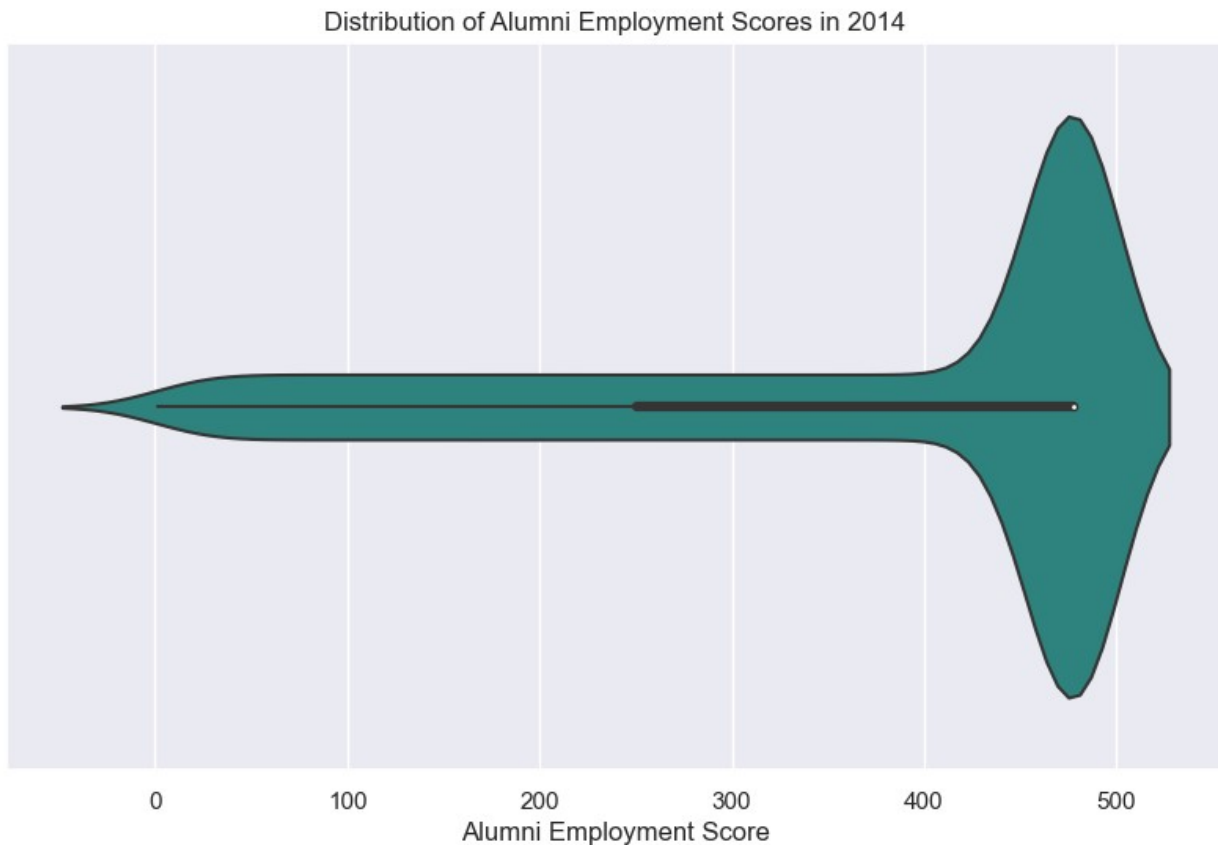
```
fig , piep = plt.subplots()  
piep.pie(x,labeldistance = 1.1, radius = 2, labels = lbl,  
rotatelabels=True)  
plt.show()
```



```
# 8
```

```
plt.figure(figsize=(10, 6))  
sns.violinplot(data = data14, x = 'alumni_employment',  
palette='viridis')
```

```
plt.title('Distribution of Alumni Employment Scores in 2014')
plt.xlabel('Alumni Employment Score')
plt.show()
```



# 9

```
srt = data14.drop_duplicates()
topdata14 = srt.nlargest(10, 'score').drop_duplicates()

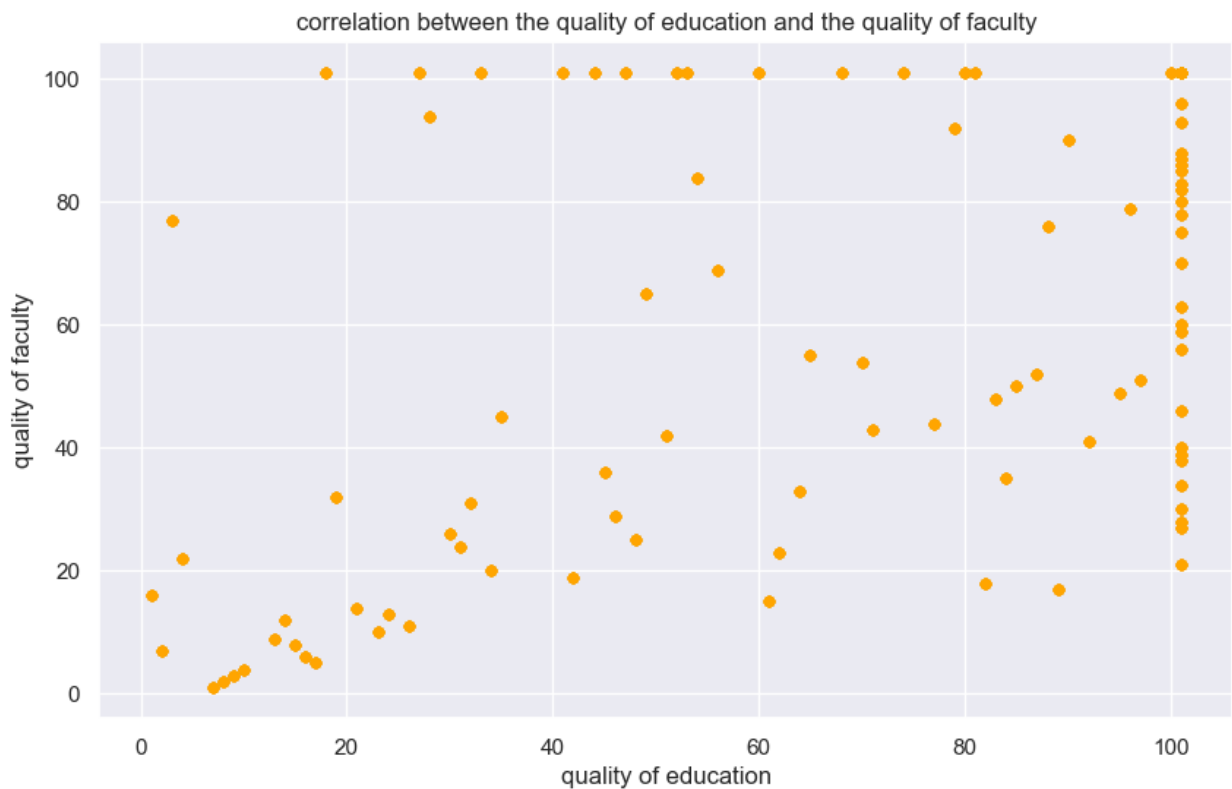
alt.Chart(topdata14, title = 'top 10 universities with the highest
scores in 2014').mark_bar(color = 'lightseagreen').encode(
    x = 'institution',
    y = 'score'
).properties(width = 600, height = 300, padding={"left": 50, "right":
50, "top": 40, "bottom": 20})

alt.Chart(...)
```

# 10

```
x = data12['quality_of_education']
y = data12['quality_of_faculty']
```

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



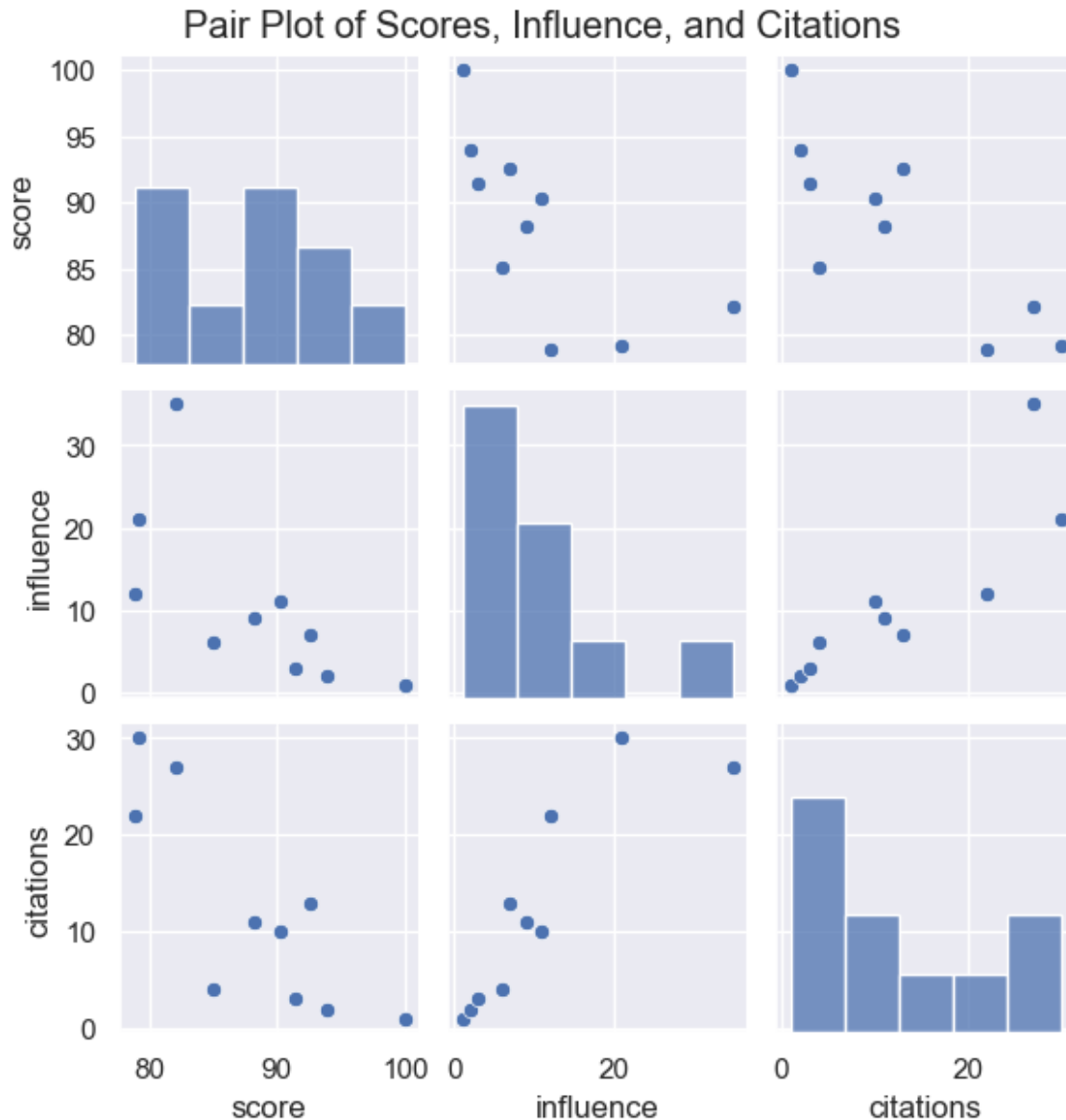
```
# 11
```

```
tmp = data13.drop_duplicates()
topdt13 = tmp.nsmallest(10, 'world_rank')

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

sns.pairplot(topdt13[cols], height=2)
plt.suptitle('Pair Plot of Scores, Influence, and Citations', y=1.02)
plt.show()
```





```
# 12
```

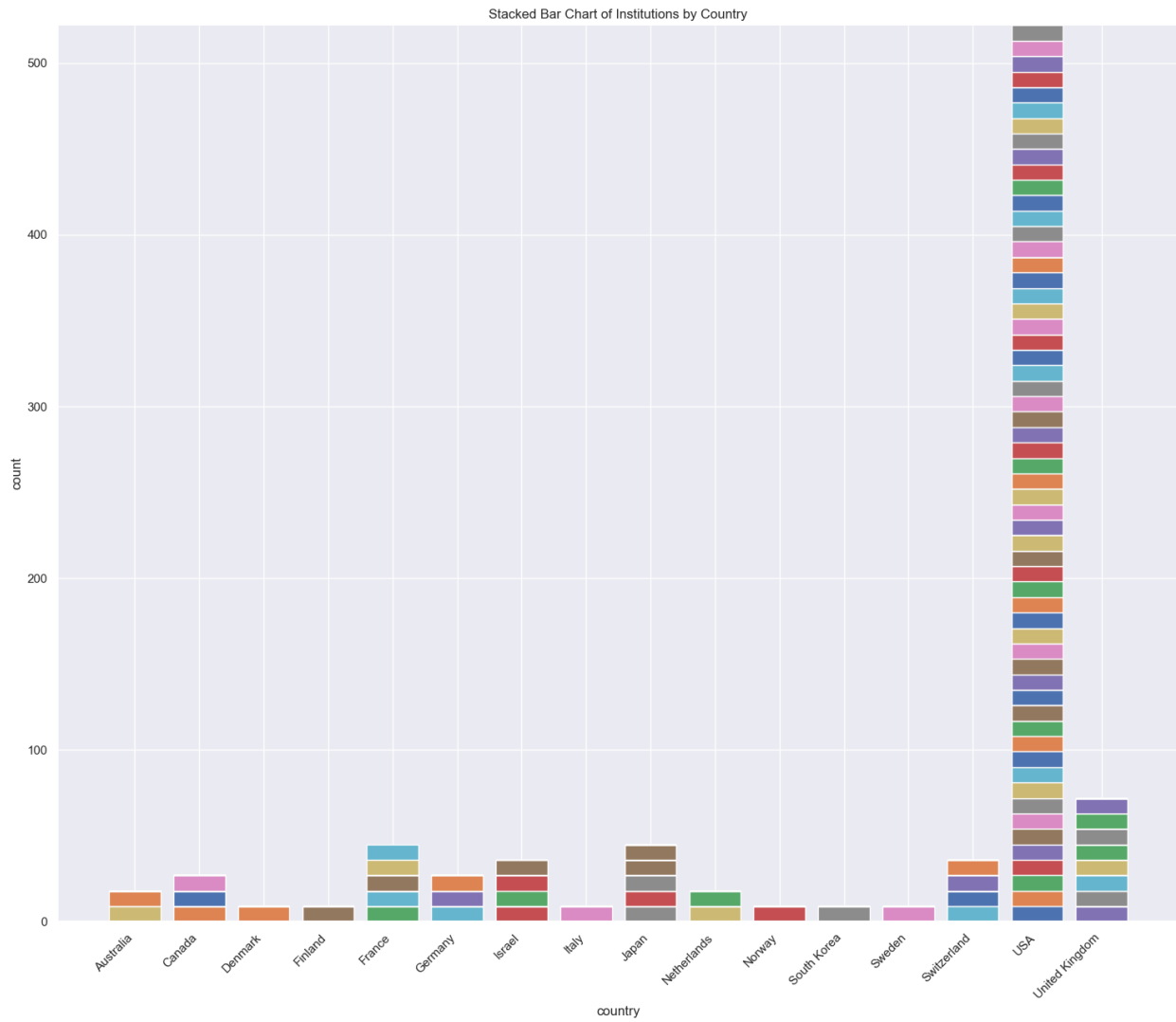
```
country_filter = data14.loc[data14['country'].isin(['USA', 'United
Kingdom'])]
clean_data = country_filter.drop_duplicates()

alt.Chart(clean_data, title = 'universities in the United States and
the United Kingdom').mark_bar().encode(
    x = 'institution',
    y = 'score',
    color = 'country'
)

alt.Chart(...)
```

```
# 13
```

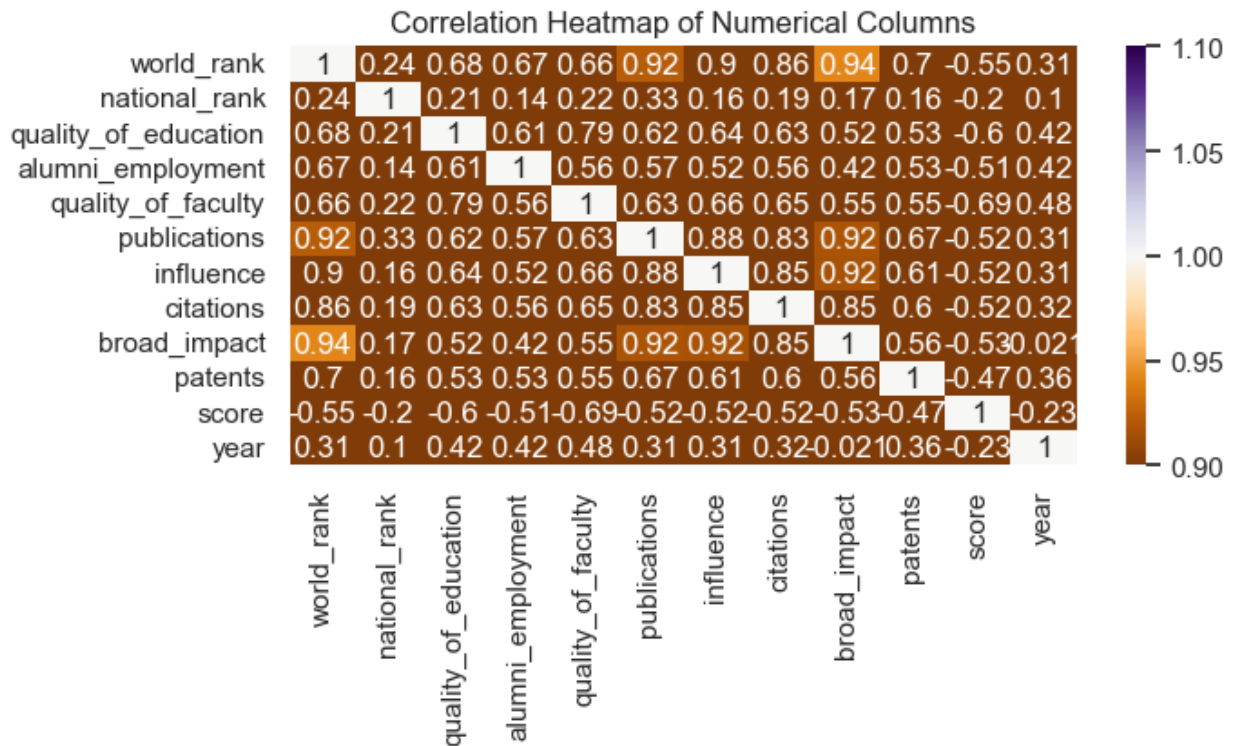
```
grp_data = data12.groupby(['country',  
    'institution']).size().unstack(fill_value=0)  
  
plt.figure(figsize=(15,13))  
bottom_counts = 0  
for inst in grp_data.columns:  
    plt.bar(grp_data.index, grp_data[inst], bottom=bottom_counts,  
        label=inst)  
    bottom_counts += grp_data[inst]  
  
plt.ylabel('count')  
plt.xlabel('country')  
plt.title('Stacked Bar Chart of Institutions by Country')  
plt.xticks(rotation=45, ha='right')  
  
plt.tight_layout()  
plt.show()
```



# 14

```
num_cols = data.select_dtypes(include=[np.number]).columns
corr_mat = data[num_cols].corr()

sns.heatmap(corr_mat, annot=True, cmap='PuOr', vmin=1, vmax=1)
plt.title('Correlation Heatmap of Numerical Columns')
plt.show()
```



```
# 15
import altair_viewer as av

data14 = data14.drop_duplicates()
top_5 = data14.nsmallest(5, 'world_rank')

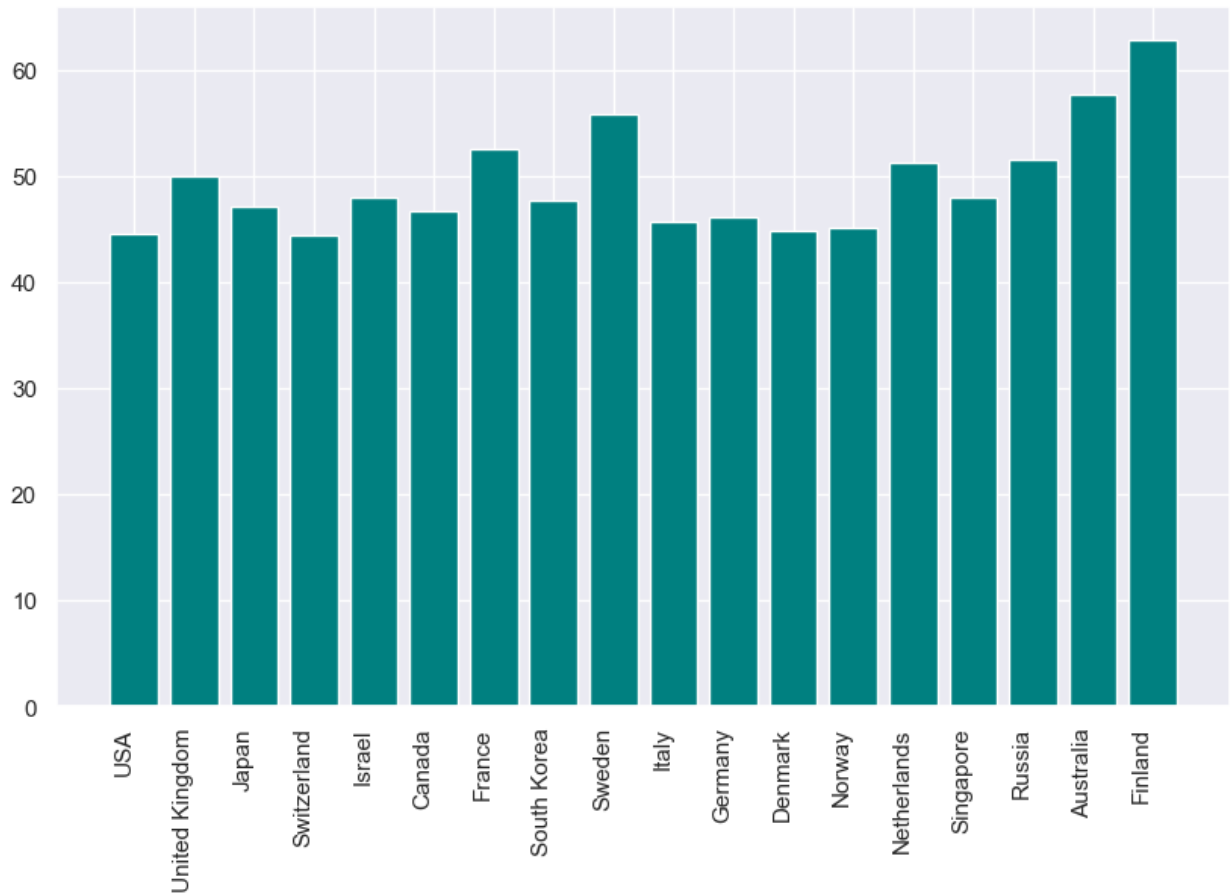
alt.Chart(top_5).mark_line(color='lightseagreen').encode(
    x = 'institution:N',
    y = alt.Y("score:Q", scale=alt.Scale(domain=[40,130]))
).properties(title = "University Scores", width=500)

alt.Chart(...)
```

```
# 16

yax = data13.groupby('country')['score'].mean().tolist()
xax = data13['country'].drop_duplicates()

plt.figure(figsize=(10,6))
plt.bar(xax ,yax, color='teal')
plt.xticks(rotation=90, ha='right')
plt.show()
```

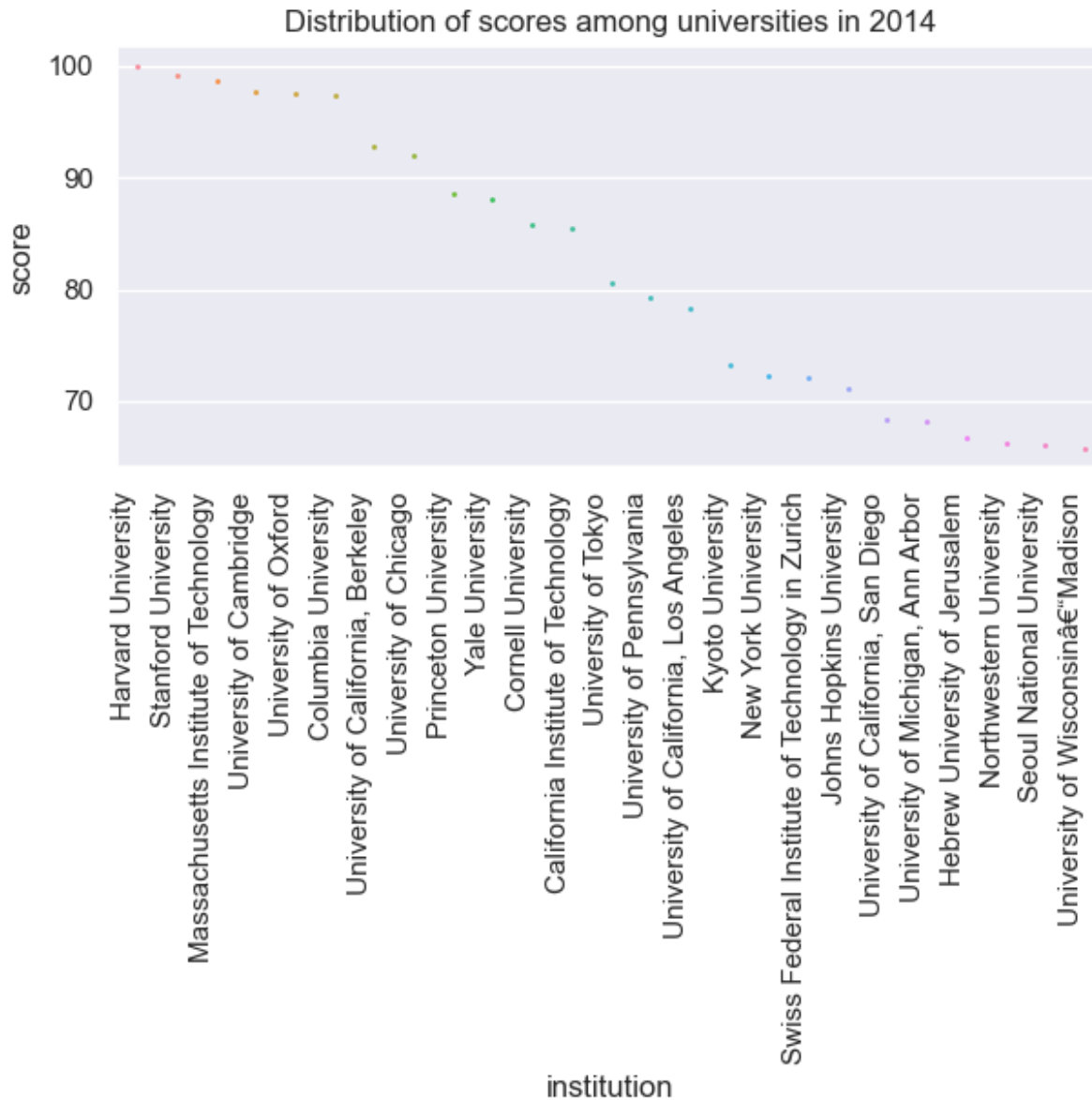


# 17

```
d = data[data['year'] == 2014]
dt14 = d.drop_duplicates()
dt14 = dt14.nsmallest(25, 'world_rank')

sns.swarmplot(size = 2, x = 'institution', y = 'score', data=dt14)

plt.xlabel('institution')
plt.ylabel('score')
plt.title('Distribution of scores among universities in 2014')
plt.xticks
plt.xticks(rotation=90, ha='right')
plt.show()
```



```
# 18
import plotly.express as px

cz = data12.groupby('country')
vct = cz['institution'].value_counts()
df = vct.to_frame(name="value").reset_index()

fig = px.treemap(
    df,
    values="value",
    names="institution",
    path=["country"],
    color="country",
    color_continuous_scale="Viridis",
```

```

    title=" proportion of universities in each country in 2012",
)

fig.show()

{"config":{"plotlyServerURL":"https://plot.ly"},"data":
[{"branchvalues":"total","customdata":["Australia"],["Canada"],
["Denmark"],["Finland"],["France"],["Germany"],["Israel"],["Italy"],
["Japan"],["Netherlands"],["Norway"],["South Korea"],["Sweden"],
["Switzerland"],["USA"],["United Kingdom"]],"domain":{"x":[0,1],"y":
[0,1]},"hvertemplate":"labels=%{label}<br>value=%{value}<br>parent=%
{parent}<br>id=%{id}<br>country=%{customdata[0]}<extra></
extra>","ids":
["Australia","Canada","Denmark","Finland","France","Germany","Israel",
"Italy","Japan","Netherlands","Norway","South
Korea","Sweden","Switzerland","USA","United Kingdom"],"labels":
["Australia","Canada","Denmark","Finland","France","Germany","Israel",
"Italy","Japan","Netherlands","Norway","South
Korea","Sweden","Switzerland","USA","United Kingdom"],"marker":
{"colors":
["#636efa","#EF553B","#00cc96","#ab63fa","#FFA15A","#19d3f3","#FF6692",
"#B6E880","#FF97FF","#FECB52","#636efa","#EF553B","#00cc96","#ab63fa",
"#FFA15A","#19d3f3"]},"name":"","parents":
["","","","","","","","","","","","","","","","","",""],"type":"treemap","va
lues":[18,27,9,9,45,27,36,9,45,18,9,9,9,36,522,72]}],"layout":
{"legend":{"tracegroupgap":0},"template":{"data":{"bar":{"error_x":
{"color":"#2a3f5f"},"error_y":{"color":"#2a3f5f"},"marker":{"line":
{"color":"#E5ECF6","width":0.5},"pattern":
{"fillmode":"overlay","size":10,"solidity":0.2},"type":"bar"}},{"barpo
lar":{"marker":{"line":{"color":"#E5ECF6","width":0.5},"pattern":
{"fillmode":"overlay","size":10,"solidity":0.2},"type":"barpolar"}},
"carpet":{"aaxis":
{"endlinecolor":"#2a3f5f","gridcolor":"white","linecolor":"white","min
orgridcolor":"white","startlinecolor":"#2a3f5f"},"baxis":
{"endlinecolor":"#2a3f5f","gridcolor":"white","linecolor":"white","min
orgridcolor":"white","startlinecolor":"#2a3f5f"},"type":"carpet"}},{"ch
oropleth":{"colorbar":
{"outlinewidth":0,"ticks":"","type":"choropleth"}},{"contour":
{"colorbar":{"outlinewidth":0,"ticks":"","colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],
[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]],"type":"contour"}},{"contourcarpet":{"colorbar":
{"outlinewidth":0,"ticks":"","type":"contourcarpet"}},{"heatmap":
{"colorbar":{"outlinewidth":0,"ticks":"","colorscale":
[[0,"#0d0887"],[0.1111111111111111,"#46039f"],
[0.2222222222222222,"#7201a8"],[0.3333333333333333,"#9c179e"],
[0.4444444444444444,"#bd3786"],[0.5555555555555556,"#d8576b"],

```

```

[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]], "type": "heatmap"}], "heatmapgl": [{"colorbar":
{"linewidth":0, "ticks":""}, "colorscale": [[0, "#0d0887"],
[0.1111111111111111, "#46039f"], [0.2222222222222222, "#7201a8"],
[0.3333333333333333, "#9c179e"], [0.4444444444444444, "#bd3786"],
[0.5555555555555556, "#d8576b"], [0.6666666666666666, "#ed7953"],
[0.7777777777777778, "#fb9f3a"], [0.8888888888888888, "#fdca26"],
[1, "#f0f921"]], "type": "heatmapgl"}], "histogram": [{"marker": {"pattern":
{"fillmode": "overlay", "size": 10, "solidity": 0.2}}, "type": "histogram"}],
"histogram2d": [{"colorbar": {"linewidth":0, "ticks":""}, "colorscale":
[[0, "#0d0887"], [0.1111111111111111, "#46039f"],
[0.2222222222222222, "#7201a8"], [0.3333333333333333, "#9c179e"],
[0.4444444444444444, "#bd3786"], [0.5555555555555556, "#d8576b"],
[0.6666666666666666, "#ed7953"], [0.7777777777777778, "#fb9f3a"],
[0.8888888888888888, "#fdca26"],
[1, "#f0f921"]], "type": "histogram2d"}], "histogram2dcontour":
[{"colorbar": {"linewidth":0, "ticks":""}, "colorscale":
[[0, "#0d0887"], [0.1111111111111111, "#46039f"],
[0.2222222222222222, "#7201a8"], [0.3333333333333333, "#9c179e"],
[0.4444444444444444, "#bd3786"], [0.5555555555555556, "#d8576b"],
[0.6666666666666666, "#ed7953"], [0.7777777777777778, "#fb9f3a"],
[0.8888888888888888, "#fdca26"],
[1, "#f0f921"]], "type": "histogram2dcontour"}], "mesh3d": [{"colorbar":
{"linewidth":0, "ticks":""}, "type": "mesh3d"}], "parcoords": [{"line":
{"colorbar": {"linewidth":0, "ticks":""}}, "type": "parcoords"}], "pie":
[{"automargin": true, "type": "pie"}], "scatter": [{"fillpattern":
{"fillmode": "overlay", "size": 10, "solidity": 0.2}, "type": "scatter"}], "scatter3d": [{"line": {"colorbar": {"linewidth":0, "ticks":""}}, "marker":
{"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scatter3d"}], "scattercarpet":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scattercarpet"}], "scattergeo":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scattergeo"}], "scattergl":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scattergl"}], "scattermapbox":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scattermapbox"}], "scatterpolar":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scatterpolar"}], "scatterpolargl":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scatterpolargl"}], "scatterternary":
[{"marker": {"colorbar":
{"linewidth":0, "ticks":""}}, "type": "scatterternary"}], "surface":
[{"colorbar": {"linewidth":0, "ticks":""}, "colorscale":
[[0, "#0d0887"], [0.1111111111111111, "#46039f"],
[0.2222222222222222, "#7201a8"], [0.3333333333333333, "#9c179e"],
[0.4444444444444444, "#bd3786"], [0.5555555555555556, "#d8576b"],

```



```

[0.6666666666666666,"#ed7953"],[0.7777777777777778,"#fb9f3a"],
[0.8888888888888888,"#fdca26"],
[1,"#f0f921"]], "type": "surface"}], "table": [{"cells": {"fill":
{"color": "#EBF0F8"}, "line": {"color": "white"}}, "header": {"fill":
{"color": "#C8D4E3"}, "line":
{"color": "white"}}, "type": "table"}]], "layout": {"annotationdefaults":
{"arrowcolor": "#2a3f5f", "arrowhead": 0, "arrowwidth": 1}, "autotypenumbers
": "strict", "coloraxis": {"colorbar":
{"linewidth": 0, "ticks": ""}, "colorscale": {"diverging":
[[0, "#8e0152"], [0.1, "#c51b7d"], [0.2, "#de77ae"], [0.3, "#f1b6da"],
[0.4, "#fde0ef"], [0.5, "#f7f7f7"], [0.6, "#e6f5d0"], [0.7, "#b8e186"],
[0.8, "#7fb341"], [0.9, "#4d9221"], [1, "#276419"]], "sequential":
[[0, "#0d0887"], [0.1111111111111111, "#46039f"],
[0.2222222222222222, "#7201a8"], [0.3333333333333333, "#9c179e"],
[0.4444444444444444, "#bd3786"], [0.5555555555555556, "#d8576b"],
[0.6666666666666666, "#ed7953"], [0.7777777777777778, "#fb9f3a"],
[0.8888888888888888, "#fdca26"], [1, "#f0f921"]], "sequentialminus":
[[0, "#0d0887"], [0.1111111111111111, "#46039f"],
[0.2222222222222222, "#7201a8"], [0.3333333333333333, "#9c179e"],
[0.4444444444444444, "#bd3786"], [0.5555555555555556, "#d8576b"],
[0.6666666666666666, "#ed7953"], [0.7777777777777778, "#fb9f3a"],
[0.8888888888888888, "#fdca26"], [1, "#f0f921"]]]}, "colorway":
["#636efa", "#EF553B", "#00cc96", "#ab63fa", "#FFA15A", "#19d3f3", "#FF6692",
"#B6E880", "#FF97FF", "#FECB52"]}, "font": {"color": "#2a3f5f"}, "geo":
{"bgcolor": "white", "lakecolor": "white", "landcolor": "#E5ECF6", "showlake
s": true, "showland": true, "subunitcolor": "white"}, "hoverlabel":
{"align": "left"}, "hovermode": "closest", "mapbox":
{"style": "light"}, "paper_bgcolor": "white", "plot_bgcolor": "#E5ECF6", "po
lar": {"angularaxis":
{"gridcolor": "white", "linecolor": "white", "ticks": ""}, "bgcolor": "#E5ECF
6", "radialaxis":
{"gridcolor": "white", "linecolor": "white", "ticks": ""}}, "scene":
{"xaxis":
{"backgroundcolor": "#E5ECF6", "gridcolor": "white", "gridwidth": 2, "lineco
lor": "white", "showbackground": true, "ticks": "", "zerolinecolor": "white"},
"yaxis":
{"backgroundcolor": "#E5ECF6", "gridcolor": "white", "gridwidth": 2, "lineco
lor": "white", "showbackground": true, "ticks": "", "zerolinecolor": "white"},
"zaxis":
{"backgroundcolor": "#E5ECF6", "gridcolor": "white", "gridwidth": 2, "lineco
lor": "white", "showbackground": true, "ticks": "", "zerolinecolor": "white"}
}, "shapedefaults": {"line": {"color": "#2a3f5f"}}, "ternary": {"aaxis":
{"gridcolor": "white", "linecolor": "white", "ticks": ""}, "baxis":
{"gridcolor": "white", "linecolor": "white", "ticks": ""}, "bgcolor": "#E5ECF
6", "caxis":
{"gridcolor": "white", "linecolor": "white", "ticks": ""}}, "title":
{"x": 5.0e-2}, "xaxis":
{"automargin": true, "gridcolor": "white", "linecolor": "white", "ticks": "",
"title":

```

```

{"standoff":15}, "zerolinecolor": "white", "zerolinewidth": 2}, {"yaxis":
{"automargin": true, "gridcolor": "white", "linecolor": "white", "ticks": "",
"title":
{"standoff": 15}, {"zerolinecolor": "white", "zerolinewidth": 2}}}, {"title":
{"text": " proportion of universities in each country in 2012"}]}

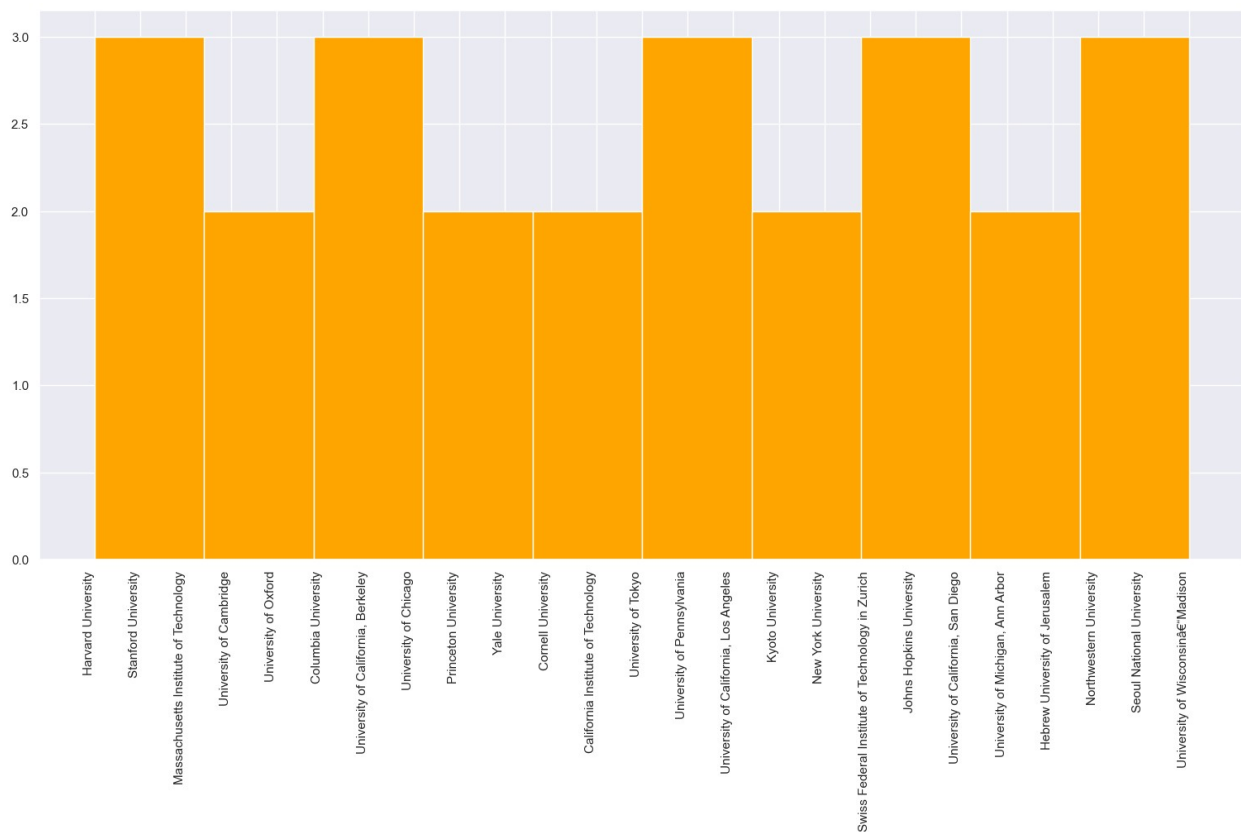
# 19

his_data = data14.drop_duplicates()
top25his = his_data.nsmallest(25, 'world_rank')

x = top25his['score'].tolist()
y = top25his['institution'].tolist()

plt.figure(figsize=(15,10), constrained_layout=True)
plt.hist(data = top25his, x='institution', color='orange')
plt.xticks(rotation=90, ha='right')
plt.show()

```



```

# 20

aus_cad = data13.loc[data13['country'].isin(['Australia', 'Canada'])]

plt.figure(figsize=(15,6))

```

```
alt.Chart(aus_cad).mark_bar().encode(  
    x = 'institution:N',  
    y = 'influence:Q',  
    color='country'  
).properties(title='Influence scores of universities in Canada and  
Australia in 2013', width=800, height=300)
```

```
alt.Chart(...)
```

```
<Figure size 1500x600 with 0 Axes>
```