

#### **Minor Project Report**

on

### [University Ranking and Accreditation Analysis]

Submitted to

#### LOVELY PROFESSIONAL UNIVERSITY

in partial fulfilment of the requirements for the award of degree of

### **Master of Computer Applications**

Submitted By Supervised By

Mohd Shah Wasif 12223538 Rafiq Sir

**Anmol Gupta 12223484** 

**Laxman Rao 12223537** 

### LOVELY FACULTY OF TECHNOLOGY & SCIENCES

LOVELY PROFESSIONAL UNIVERSITY

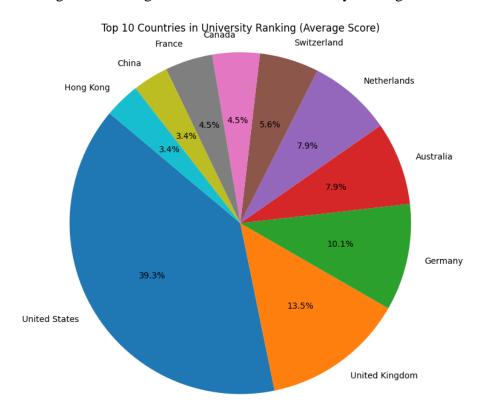
**PUNJAB** 

[November 2023]

Table of Content		
Sr. No.	Content	Page Index
1	Introduction about project	1-3
2	Project Modules and its description	3-6
3	Coding	6-27
4	Screenshot output	27-28

### Introduction

In this project we analyse the World University Rankings 2023. The Times Higher Education World University Rankings 2023 includes 1,799 universities across 104 countries and regions, making them the largest and most diverse university rankings to date.



This dataset includes the following 13 columns:

- 1. **University Rank**: The global ranking of the university.
- 2. Name of University: The name of the university.
- 3. **Location**: The country where the university is located.
- 4. **No of student**: The total number of students at the university.
- 5. **No of student per staff**: The ratio of students to staff at the university.
- 6. **International Student**: The percentage of students who are international.
- 7. **Female: Male Ratio**: The ratio of female to male students at the university.
- 8. **Overall Score**: The overall score of the university, presumably out of 100.
- 9. **Teaching Score**: The university's score for teaching, presumably out of 100.
- 10. **Research Score**: The university's score for research, presumably out of 100.
- 11. **Citations Score**: The university's score for citations, presumably out of 100.
- 12. **Industry Income Score**: The university's score for industry income, presumably out of 100.
- 13. **International Outlook Score**: The university's score for international outlook, presumably out of 100.

•

The goal of this project is to take a deep dive into the dataset and answer the following four questions:

Which country has the highest number of universities in the top 100 of the World University Rankings 2023

What is the university ranking distribution of the top 10 scoring Countries?

What is the correlation between each score and overall ranking? Which factor is the most important?

## **Project Module**

- 1. **University Rank**: There are 162 unique university ranks in the dataset. The most frequent rank is 'Reporter', which appears 512 times.
- 2. Name of University: There are 2233 unique universities in the dataset. The most frequent university is the University of Oxford, which appears once.
- 3. **Location**: The universities are located in 116 unique locations. The most frequent location is the United States, which is the location of 173 universities in the dataset.

- 4. **No of student**: The number of students ranges from 17,080 to an unknown maximum value. The most frequent number of students is 17,080, which appears twice.
- 5. **No of student per staff**: The ratio of students to staff ranges from 0.4 to 232.2. The average ratio is approximately 19. The standard deviation is approximately 12.13, indicating a wide spread of the data.
- 6. **International Student**: There are 79 unique values for the percentage of international students. The most frequent value is 0%, which appears 336 times.
- 7. **Female:Male Ratio**: There are 87 unique values for the female-to-male ratio. The most frequent ratio is 58:42, which appears 95 times.
- 8. **OverAll Score**: There are 160 unique overall scores. The most frequent score is 18.4–24.3, which appears 305 times.
- 9. **Teaching Score**: The teaching score ranges from 11.6 to 94.8, with an average of approximately 27. The standard deviation is approximately 13.28, indicating a wide spread of the data.
- 10. **Research Score**: The research score ranges from 7.4 to 99.7, with an average of approximately 23. The standard deviation is approximately 16.76, indicating a wide spread of the data.
- 11. **Citations Score**: The citations score ranges from 0.8 to 100, with an average of approximately 48.5. The standard deviation is approximately 27.97, indicating a wide spread of the data.
- 12. **Industry Income Score**: The industry income score ranges from 36.9 to 100, with an average of approximately 47.1. The standard deviation is approximately 15.09, indicating a wide spread of the data.
- 13. **International Outlook Score**: The international outlook score ranges from 14.1 to 99.7, with an average of approximately 46.9. The standard deviation is approximately 22.58, indicating a wide spread of the data.

### Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
university= pd.read_csv('World University Rankings 2023.csv')
university.head()
university.info()
missing_values = university.isnull().sum()
print("Missing Values:\n", missing_values)
missing_values = university.isnull().sum()
print("Missing Values:\n", missing_values)
university.drop_duplicates(inplace=True)
university['Location'].value_counts()
plt.figure(figsize=(12, 8))
sns.countplot(data=university, x='Location', order=university['Location'].value_counts().index)
plt.title('Regional Distribution of Universities')
plt.xlabel('Region')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()
# Get the top ten countries with the most universities in the top 100
top_100= university.head(100)
countries = top_100['Location']
# Find rows with null values in the 'Location' column
null_locations = top_100[top_100['Location'].isnull()]
# Display the rows with null values
print(null_locations)
#fill in the missing country data
universities_to_countries = {
  'Tsinghua University': 'China',
  'Peking University': 'China',
  'National University of Singapore': 'Singapore',
```

```
'Technical University of Munich': 'Germany',
  'University of Hong Kong': 'Hong Kong',
  'LMU Munich': 'Germany',
  'KU Leuven': 'Belgium',
  'Universität Heidelberg': 'Germany',
  'Chinese University of Hong Kong': 'Hong Kong',
  'McGill University': 'Canada',
  'The University of Queensland': 'Australia',
  'University of Manchester': 'United Kingdom',
  'The Hong Kong University of Science and Technology': 'Hong Kong',
  'Zhejiang University': 'China',
  'UNSW Sydney': 'Australia',
  'University of Science and Technology of China': 'China',
  'University of Groningen': 'Netherlands',
  'University of Bristol': 'United Kingdom',
  'Leiden University': 'Netherlands',
  'Yonsei University (Seoul campus)': 'South Korea',
  'Hong Kong Polytechnic University': 'Hong Kong',
  'Erasmus University Rotterdam': 'Netherlands',
  'University of Glasgow': 'United Kingdom',
  'McMaster University': 'Canada',
  'University of Adelaide': 'Australia',
  'City University of Hong Kong': 'Hong Kong'
# Iterate through the universities and assign the corresponding country
for university_name, country in universities_to_countries.items():
  top_100.loc[top_100['Name of University'] == university_name, 'Location'] = country
# number of universities in each country
country_counts = countries.value_counts()
# top ten countries
top_ten_countries = country_counts.head(10)
```

}

```
print(top_ten_countries)
# Convert 'OverAll Score' column to float using .loc for assignment
top_100.loc[:, 'OverAll Score'] = top_100['OverAll Score'].astype(float)
# Create a pie chart
plt.figure(figsize=(8, 8))
plt.pie(top_ten_countries, labels=top_ten_countries.index, autopct='%1.1f%%', startangle=140)
plt.title('Top 10 Countries in University Ranking (Average Score)')
plt.axis('equal')
# Show the pie chart
plt.show()
# top ten countries and their count of universities in the top 100
top_ten_countries_df = top_ten_countries.reset_index()
# Rename the columns
top_ten_countries_df.columns = ['Country', 'University Count']
#create the barchart
plt.figure(figsize=(10, 6))
bar_chart = sns.barplot(x='Country', y='University Count', data=top_ten_countries_df,
palette='viridis')
bar_chart.set_title('Top 10 Countries with Universities in Top 100')
bar_chart.set_xlabel('Country')
bar_chart.set_ylabel('Number of Universities')
plt.xticks(rotation=45, ha='right')
plt.show()
# Filter the data to include only the top 10 countries
filtered_data = top_100[top_100['Location'].isin(top_ten_countries.index)]
# Create a distribution box plot
plt.figure(figsize=(10, 6))
```

```
sns.boxplot(x='Location', y='OverAll Score', data=filtered_data)
plt.title('Distribution of University Rankings in Top 10 Countries')
plt.xticks(rotation=45)
plt.xlabel('Country')
plt.ylabel('OverAll Score')
plt.show()
# Scatterplot for Teaching Score vs. OverAll Score
plt.scatter(top_100['Teaching Score'], top_100['OverAll Score'])
plt.title('Teaching Score vs. OverAll Score')
plt.xlabel('Teaching Score')
plt.ylabel('OverAll Score')
plt.show()
# Scatterplot for Research Score vs. OverAll Score
plt.scatter(top_100['Research Score'], top_100['OverAll Score'])
plt.title('Research Score vs. OverAll Score')
plt.xlabel('Research Score')
plt.ylabel('OverAll Score')
plt.show()
# Scatterplot for Citations Score vs. OverAll Score
plt.scatter(top_100['Citations Score'], top_100['OverAll Score'])
plt.title('Citations Score vs. OverAll Score')
plt.xlabel('Citations Score')
plt.ylabel('OverAll Score')
plt.show()
# Scatterplot for Industry Income Score vs. OverAll Score
plt.scatter(top_100['Industry Income Score'], top_100['OverAll Score'])
plt.title('Industry Income Score vs. OverAll Score')
plt.xlabel('Industry Income Score')
plt.ylabel('OverAll Score')
plt.show()
```

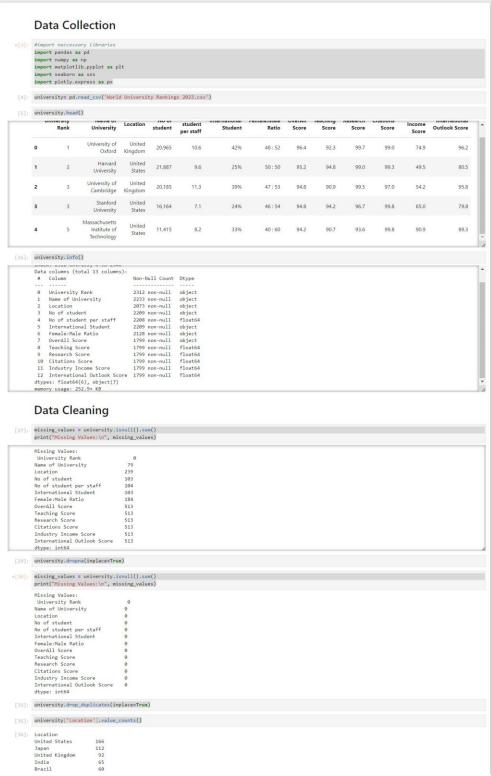
```
# Scatterplot for International Outlook Score vs. OverAll Score
plt.scatter(top_100['International Outlook Score'], top_100['OverAll Score'])
plt.title('International Outlook Score vs. OverAll Score')
plt.xlabel('International Outlook Score')
plt.ylabel('OverAll Score')
plt.show()
top_n = 50
top_scores_teaching_rank = university.nlargest(top_n, 'Teaching Score')
plt.figure(figsize=(10, 6))
sns.barplot(x='Name of University', y='Teaching Score', data=top_scores_teaching_rank)
plt.title(f'Top {top_n} Universities by Teaching Scores')
plt.xlabel('University Name')
plt.xticks(rotation=90)
plt.ylabel('Teaching Score Rank')
plt.show()
top_n = 50
top_scores_research_rank = university.nlargest(top_n, 'Research Score')
plt.figure(figsize=(10, 6))
sns.barplot(x='Name of University', y='Research Score', data=top_scores_research_rank,
color='green')
plt.title(f'Top {top_n} Universities by Research Scores')
plt.xlabel('University Name')
plt.xticks(rotation=90)
plt.ylabel('Research Score')
plt.show()
top_n = 50
top_scores_citation = university.nlargest(top_n, 'Citations Score')
```

```
plt.figure(figsize=(10, 6))
sns.barplot(x='Name of University', y='Citations Score', data=top_scores_research_rank, color='skyblue')
plt.title(f'Top {top_n} Universities by Citations Scores')
plt.xlabel('Name of Universitys')
plt.xticks(rotation=90)
plt.ylabel('Citations Score')
plt.show()
```

## **Screenshot Output**



JupyterLab 🗗 🐞 Python 3 (ipykernel) 🔘



```
Costa Rica
                                        Jamaica
Brunei Darussalam
                                        Montenegro 1
Name: count, Length: 99, dtype: int64
                                 sns.countplot(data=university, x='location', order=university['location'].value_counts().index)
plt.title('Regional Distribution of Universities')
plt.xlabel('Region')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.xticks(rotation=90)
 [37]: plt.figure(figsize=(12, 8))
                                   C:\Users\mewho\AppData\local\Programs\Python\Python311\tib\site-packages\seaborn\oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
if pd.api.types.is_categorical_dtype(vector):
C:\Users\mewho\AppData\local\Programs\Python\Python311\tib\site-packages\seaborn\oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
if pd.api.types.is_categorical_dtype(vector):
                                                                                                                                                                                                                                                                                                                                Regional Distribution of Universities
                                                         160
                                                         120
                                                         100
                                          Count
                                                             80
                                                             60
                                                              40
                                                                                                                                                                                                                                                                    United strategy of the property of the propert
                                                                                                                                                                                                                                                                                                                                                                                                                              Region
 [38]: # Get the top ten countries with the most universities in the top 100 top_100= university.head(100) countries = top_100['location']
[39]: # Find rows with null values in the 'Location' column null_locations = top_100[top_100['Location'].isnull()]
                                     # Display the rows with null values
print(null_locations)
                                      Empty DataFrame
Columns: [University Rank, Name of University, Location, No of student, No of student per staff, International Student, Female:Male Ratio, OverAll Score, T acaching Score, Research Score, Citations Score, Industry Income Score, International Outlook Score]
Index: []
                                      Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

Index: []

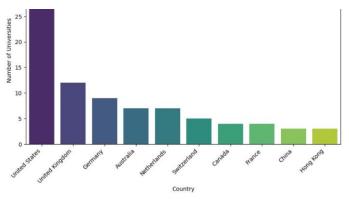
Index: []
```

```
# Iterate through the universities and assign the corresponding country
for university_name, country in universities_to_countries.items():
top_100.lc(top_100'\text{Memo of University'} |= university_name, 'tocation') = country
                Data Analysis
[41]: # number of universities in each country
country_counts = countries.value_counts()
                  top_ten_countries = country_counts.head(10)
                print(top_ten_countries)
                United States
United Kingdom
Germany
Australia
Netherlands
Switzerland
Canada
France
China
Hone Kone
                  Hong Kong 3
Name: count, dtype: int64
[42]: # Convert 'OverAll Score' column to float using .loc for assignment
top_100.loc[:, 'OverAll Score'] = top_100['OverAll Score'].astype(float)
[43]: # Create a pie chart
plt.figure(figsize(8, 8))
plt.pie(foot ten.countries, labels=top_ten.countries.index, autopct='%1.1f%%', startangle=140)
plt.title('Top 10 Countries in University Ranking (Average Score)')
plt.saig('equal')
# Show the pie chart
plt.show()
                                                            Top 10 Countries in University Ranking (Average Score)
                                                                                                                                                                   Switzerland
                                                                                        France
                                                                    China
                                                                                                                                                                                                        Netherlands
                                        Hong Kong
                                                                                                                4.5% 4.5%
                                                                                                                                                                                                                                    Australia
                                                                                                                                                               13.5%
                                                                                                                                                                                                 United Kingdom
                # Top 10 Countries and their Count of Universities in the top 100
[44]: # top ten countries and their count of universities in the top 100
top_ten_countries_df = top_ten_countries.reset_index()
                  # Rename the columns
top_ten_countries_df.columns = ['Country', 'University Count']
                @create the barchart
plt.figure(figstzer(10, 6))
bar_chart = sns.barplot(xs'(country', y='University Count', datastop_ten_countries_df, palette='viridis')
bar_chart.set_title('Top 10 Countries with Universities in Top 100')
bar_chart.set_xlabel('Country')
bar_chart.set_xlabel('Ulumber of Universities')
plt.xticks(rotation=45, ha='right')
plt.shou()
                C:\Users\men\mo\appData\local\Programs\Python\Python311\Lib\site-packages\seaborn\oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is_categorical_dtype(vector):

C:\Users\men\modeln\appData\local\Programs\Python\Python311\Lib\site-packages\seaborn\oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is_categorical_dtype(vector):

C:\Users\men\modeln\appData\loca\Loca\Lip\mograms\Python\Python\Python311\Lib\site-packages\seaborn\oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is_categorical_dtype(vector):
                                                                                                 Top 10 Countries with Universities in Top 100
```

35



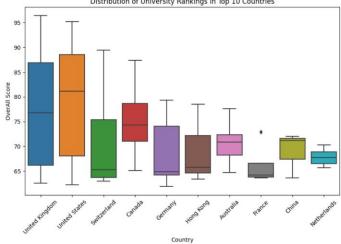
[45]: # Filter the data to include only the top 10 countries filtered\_data = top\_100[top\_100['Location'].isin(top\_ten\_countries.index)] # Create a distribution box plot plt.figure(figsize(10, 6)) sns.boxplot(xc'location', yc'overAll Score', data=filtered\_data) plt.title('Distribution of University Rankings in Top 10 Countries') plt.xticks(rotation=45) plt.xlabel('Country') plt.ylabel('OverAll Score') plt.ylabel('OverAll Score')

C:\Users\menho\AppData\\cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, Categorical) instead if pd. api.types.is\_categorical\_dtype(vector):

C:\Users\menho\AppData\cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is\_categorical\_dtype(vector):

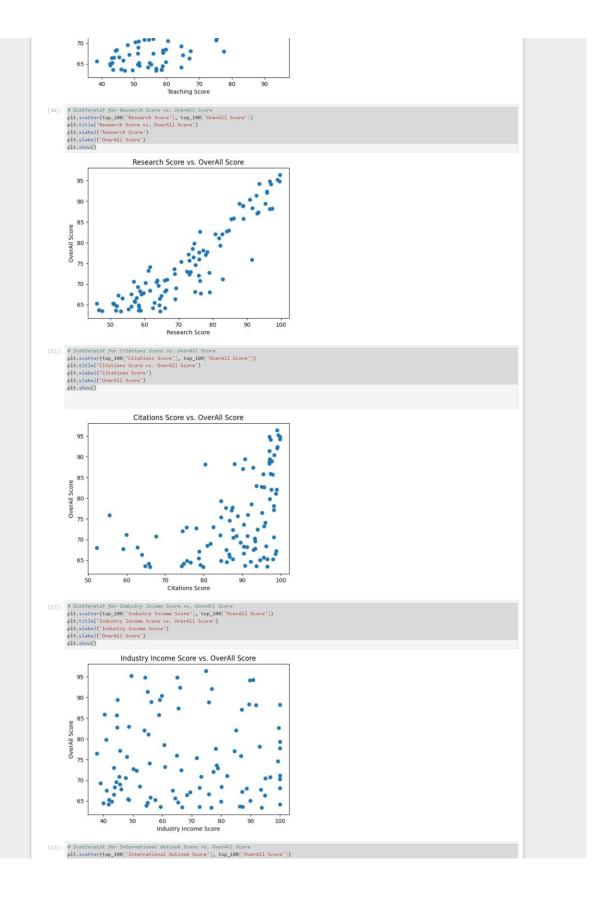
C:\Users\menho\AppData\\cal\Cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is\_categorical\_dtype(vector):

#### Distribution of University Rankings in Top 10 Countries

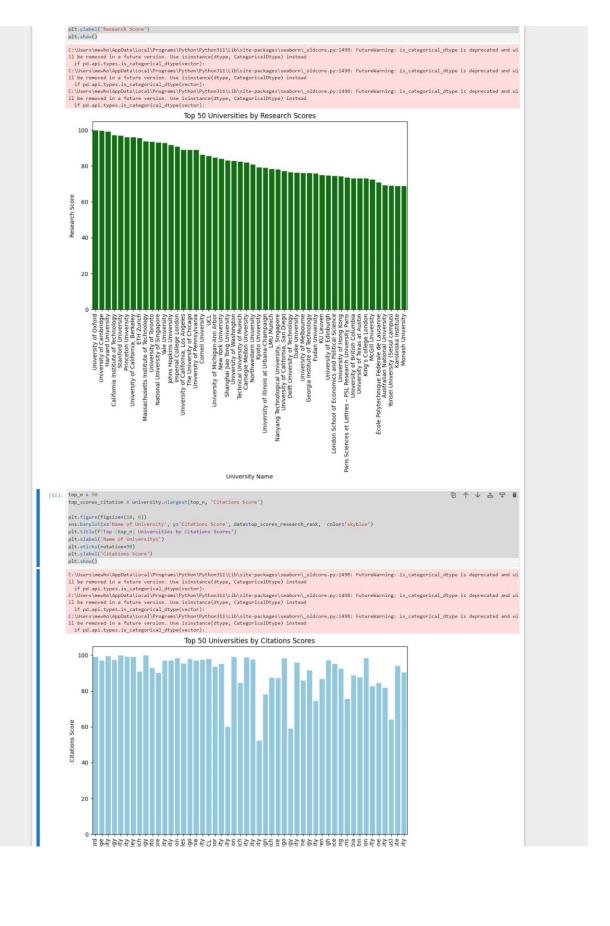


[19]: # Scatterplot for Teaching Score vs. OverAll Score
plt.scatter(top\_100! 'Teaching Score'), top\_100! 'OverAll Score'))
plt.title('Teaching Score vs. OverAll Score')
plt.xlabel('Teaching Score')
plt.ylabel('OverAll Score')
plt.show()





# plt.title('International Outlook Score vs. OverAll Score') plt.xlabel('International Outlook Score') plt.ylabel('OverAll Score') plt.ylabel('OverAll Score') International Outlook Score vs. OverAll Score 95 90 85 OverAll Score 80 75 70 65 International Outlook Score [60]: top\_n = 50 top\_scores\_teaching\_rank = university.nlargest(top\_n, 'Teaching Score') plt.figure(figsize=(10, 6)) sns.barplot(xc"Name of University', y='Teaching Score', data=top\_scores\_teaching\_rank) plt.title(f'Top\_(top\_n) Universities by Teaching Scores') plt.xlabel('University Name') plt.xlabel('University Name') plt.xlabel('Teaching Score Rank') plt.show() C:\Users\menho\AppOata\\cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is\_categorical\_dtype(vector): C:\Users\menho\AppOata\\cal\Cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is\_categorical\_dtype(vector): C:\Users\menho\AppOata\\cal\Cal\Programs\Python\Python311\\ib\site-packages\seaborn\\_oldcore.py:1498: FutureWarning: is\_categorical\_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead if pd.api.types.is\_categorical\_dtype(vector): Top 50 Universities by Teaching Scores 60 40 20 Havard University Etanford University of Cambridge Californal Institute of Technology Massachusetts Institute of Technology Princeton University of Cathology University of California, Berkeley University of California, Les Angeles University of California, Les Angeles Comel University of Michigan-Ann Arbor Dinnersity of Michigan-Ann Arbor University of Michigan-Ann Arbor Bond University of Michigan-Ann Arbor North Comel University of Michigan-Ann Arbor University of Washington University of Michigan University of California, San Dego Ecole Polytechnique Féderale de Lausanne Camage Mellon University of Hong Kong Ecole Polytechnique Féderale de Lausanne Camage Mellon University of Hong Kong Ecole Polytechnique Féderale de Lausanne Camage Mellon University of Hong Kong Ecole Polytechnique Féderale de Lausanne Camage Mellon University of California, San Dego Georgia Institute of Science of Receiver of Technology (Activity University of California, San Dego Georgia Institute of Science of Michigan Institute of Science of Michigan Institute of Science of Georgia Institute of Technology College Michigan University of California, San Dego Georgia Institute of Technology California, Davis Advanced University Name [59]: top\_n = 50 top\_scores\_research\_rank = university.nlargest(top\_n, 'Research Score') plt.figure(figsizes(10, 6)) sns.barplot(x='Name of University', y='Research Score', data=top\_scores\_research\_rank, color='green') plt.fila(f'Top (top.n) Universities by Research Scores') plt.xabel('university Name') plt.xabel('university Name')



University of Cambrid

Brand Interest of Technols

California institute of Technols

Princeton University of Caulfornia, Berfee

Massaehusetts institute of Technols

Massaehusetts institute of Technols

University of Caulfornia, Des Ange

The University of Singaph

University of California, Los Ange

The University of Mensylva

University of Mensylva

Camegle design University of Mensylva

University of Mensylva

Camegle design University of Mensylva

University of Illinois at University of Mensylva

Camegle design University of Mensylva

University of Illinois at University of Menbour

Georgia Institute of Technolo

Delte University of Illinois at University of Menbour

Georgia Institute of Technolo

Delte University of Illinois at University of Pension

University of Texas at Aus

Miniversity of Texas at Aus

Miniversity of Texas at Aus

University of Texas at Aus

University of Texas at Aus

University of Pension

University of Texas at Aus

Mercial University Soul camp

Westill University Soul camp

Westill University Soul camp

Mercial University Soul camp

Mercial University Soul camp

Westilland Matonal University

Mercial University Soul camp

Mercial Camp