

Data Analysis

Assignment Chapter 3

Instructions

1. You can take help from the lecture notes to revise the concepts that we have covered
2. Choose the best suitable answer and submit the word document
3. You have been provided a google sheet named “Top 2000 Universities of the World”, this is your dataset for this assignment.
4. For these questions, you need to work on the Google Sheet which are named per your question number, and you have to write the short summary of answer in this word document is needed.
5. To get started with the assignment, you need to make a copy of the assignment Google Sheet here:
<https://docs.google.com/spreadsheets/u/1/d/1w34wPaf72IEuINsAcMFjkex5e7z2qVY52t39bVZ4y60/copy>
6. Each question’s answer sheet should have the correct formulas applied and headings mentioned alongside the calculations. (Refer to lecture slides if you need help with sheet formatting)
7. Please submit the assignment through TalentLabs Learning System. You will need to submit this word document (with answers). **Make sure you include the link to the Google Sheet and set the permission of the Spreadsheet to “viewable by everyone”.**

Completed Spreadsheet Link:

<https://docs.google.com/spreadsheets/d/1YhHWH-egD2vT8t67oJNFPPU9nw6wofHx-HanUi72KpU/edit?usp=sharing>

Question 1 (2 points):

Data Analysis starts with the formulation of research questions and description of the dataset.

You have been provided the dataset of “Top 2000 Universities of the World”, after having a look at the data set, formulate your research questions and provide a description of the data set. You can also take help from the data plots to have an idea of the data set description.

Research Questions:

What countries have the highest average world ranking of their universities, based on their education quality, alumni employment performance and research performance?

Dataset Description:

The dataset includes the information on the world ranking of 2000 universities. The columns provide the information as below:

- **World rank:** The overall world ranking of the university.
- **Institution:** The name of the university.
- **Country:** The country where the university is located.
- **Quality of Education Rank:** The ranking based on the quality of education provided by the university.
- **Alumni Employment Rank:** The ranking based on the graduate employability.
- **Quality of Faculty Rank:** The ranking based on the quality of the faculty.
- **Research Performance Rank:** The ranking based on the research performance of the university.
- **Score:** The overall score of the university, contributing to its world ranking.

Question 2 (4 points):

When we jump on working with the data sets in Data Analysis, after verifying the data for missing values and formulating our research questions. We always observe our dataset, describe it and check the dimensions of the dataset.

For the dataset provided to you, identify the size and shape of your dataset by taking help from different data aggregation techniques?

Size of Dataset:

- 9 columns
- 2000 rows

Shape of Dataset:

Max of Score: 100

Min of Score: 65.7

Question 3 (3 points):

Whenever we have large datasets; it's always a good practice to observe the unique values in our dataset so that we can understand the records in our dataset, and we can choose our data summarizing strategies accordingly.

For the dataset provided to you, write the count of Institutions per unique country. (Hint: You can take help from data aggregation)

Number of Institutions Per Country (In Descending Order)

Country	Count
USA	347
China	277
Japan	124
United Kingdom	95
France	79
Germany	70
India	68
Italy	66
South Korea	60
Brazil	56
Turkey	55
Spain	53
Russia	46
Canada	42
Poland	40
Australia	39
Iran	39
Taiwan	35
Sweden	19
Mexico	18
Austria	17
Egypt	17
Netherlands	15
Portugal	14
Chile	14
Norway	13
South Africa	13
Switzerland	12
Czech Republic	12
Greece	12
Belgium	11
Finland	10

Argentina	10
Thailand	10
Pakistan	10
Romania	10
Israel	9
Ireland	9
Saudi Arabia	9
New Zealand	8
Malaysia	8
Hungary	8
Denmark	7
Colombia	7
Ukraine	6
Tunisia	5
Vietnam	5
Singapore	4
Serbia	4
Croatia	4
Lithuania	4
Morocco	4
Ethiopia	4
Slovenia	3
Estonia	3
Slovak Republic	3
Georgia	3
United Arab Emirates	3
Philippines	3
Iceland	2
Lebanon	2
Cyprus	2
Bulgaria	2
Qatar	2
Nigeria	2
Ghana	2
Jordan	2
Peru	2
Latvia	2
Sri Lanka	2
Indonesia	2
Algeria	2
Armenia	1
Belarus	1
Luxembourg	1
Uganda	1
Uruguay	1

Azerbaijan	1
Cameroon	1
Kuwait	1
Malawi	1
Costa Rica	1
Oman	1
Kenya	1
Zambia	1
Malta	1
Ecuador	1
Zimbabwe	1
Venezuela	1
Bangladesh	1
Northern Cyprus	1
Tanzania	1
Senegal	1
Kazakhstan	1
North Macedonia	1

Question 4 (5 points):

Data Summarizing strategies always help us understand the trends in our dataset. We have different strategies of summarizing dataset which provide insights on different aspects of dataset such as symmetry, location of data points, etc. For the dataset provided to you, calculate the measures of location. **Your answer should cover variance, standard deviation, minimum, maximum and range of score.**

Variance of Score	25.8043
Standard Deviation of Score	5.0798
Minimum score	65.7
Maximum score	100
Range of Score	34.3

Question 5 (3 points):

Inter quartile Range provides us a good estimation for the range of the dataset and helps us identify the outlier using the Minimum and the Maximum of IQR. For the dataset provided to you; identify if we have any outlier institution based on Score column.

Here are the outlier institution based on Score column:

Harvard University
Massachusetts Institute of Technology
Stanford University
University of Cambridge
University of Oxford
Princeton University
Columbia University
University of Chicago
University of Pennsylvania
Yale University
California Institute of Technology
University of California, Berkeley
University of Tokyo
Cornell University
University of Michigan, Ann Arbor
Johns Hopkins University
Northwestern University
University of California, Los Angeles
University College London
Duke University
PSL University
University of Illinois at Urbanaâ€”Champaign
University of Washington - Seattle
University of Toronto
University of Wisconsinâ€”Madison
New York University
McGill University
Kyoto University
ETH Zurich
Imperial College London

Seoul National University
Paris-Saclay University
University of Texas at Austin
University of California, San Diego
Karolinska Institute
Sorbonne University
University of Copenhagen
University of California, San Francisco
University of North Carolina at Chapel Hill
Dartmouth College
Institut Polytechnique de Paris
University of Paris
King's College London
Ludwig Maximilian University of Munich
University of Edinburgh
University of Minnesota - Twin Cities
Washington University in St. Louis
University of British Columbia
Rockefeller University
Rutgers University—New Brunswick
University of Southern California
University of California, Davis
Vanderbilt University
Pennsylvania State University
Free University of Berlin
Ohio State University
University of Manchester
Tsinghua University
Peking University
Humboldt University of Berlin
Purdue University
University of Texas Southwestern Medical Center

University of Melbourne
Hebrew University of Jerusalem
University of Zurich
Brown University
Weizmann Institute of Science
University of Colorado Boulder
University of Virginia
Heidelberg University
Utrecht University
University of Pittsburgh
University of Chinese Academy of Sciences
Texas A&M University, College Station
University of Maryland, College Park
Georgia Institute of Technology
University of Oslo
Osaka University
University of Birmingham
Technical University of Munich
University of Alberta
University of California, Irvine
University of Rochester
University of Amsterdam
Carnegie Mellon University
University of Florida
Leiden University
Uppsala University
National University of Singapore
National Taiwan University
University of Bristol
University of Göttingen
University of Arizona
Boston University

University of California, Santa Barbara
Erasmus University Rotterdam
Keio University
University of Sydney
University of Geneva
University of Basel
University of Leeds
Shanghai Jiao Tong University
University of Groningen
École Polytechnique Fédérale de Lausanne
University of São Paulo
University of Texas MD Anderson Cancer Center
Aarhus University
University of Tübingen
University of Queensland
Zhejiang University
Indiana University Bloomington
University of Utah
Sapienza University of Rome
University of New South Wales
Fudan University
KU Leuven
University of Southampton
University of Helsinki
INSEAD
University of Montreal
Ghent University
University of Freiburg
University of Bonn
Emory University
Australian National University
University of Glasgow

Baylor College of Medicine
Monash University
Case Western Reserve University
Nagoya University
University of Barcelona

Question 6 (9 points):

Pivot tables are a good source of analyzing the general trends and variable correlation in the dataset. Create a pivot table to analyze the trends of university ranks based on the universities and countries level information.

1. Research Question (1 point):

What countries have the highest average world ranking of their universities, based on their education quality, alumni employment performance and research performance?

2. Identification of Rows and Columns of interest (1 point):

Our data of interest include columns such as "Country", "Quality Education Rank", "Quality Education Rank", "Alumni Employment Rank", "Research Performance Rank", "Score" and "World Ranking".

3. Data Summary chosen and reason (2 points):

We will summarize the data with an average, grouped by "Country".

4. Data Filters if needed (1 point):

- Filter out any columns with "-" symbol inside the dataset, as it symbolizes null values.

5. Plot the data in pivot table (3 points)

Link of the Pivot table: <https://docs.google.com/spreadsheets/d/1YhHWH-egD2vT8t67oJNFPPU9nw6wofHx-HanUi72KpU/edit#gid=990890014>

6: Conclusions (1 point):

From the chart, the average research performance is most correlated with the average world ranking of university, followed by average education quality and average alumni employment rank.

We could also see the average ranking of university is likely impacted by which country they located.