</talentlabs>

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/1w34wPaf72IEulNsAcMFjkex5e7z2qVY52t39bVZ4y60/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/1lmWatpZXMkHYJ-YnC81o-ViJnvka5XEcY1VJL2N2xDE/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:

1. What are the top-performing universities globally, and how do they rank across different countries?
2. How does the quality of education correlate with research performance and alumni employment?
3. What trends can be observed in the national ranks of universities compared to their global ranks?

Dataset Description:

- The dataset contains information on the top 2000 universities in the world, including their world rank, institution name, country, national rank, and various quality metrics such as Quality of Education Rank, Alumni Employment Rank, Quality of Faculty Rank, Research Performance Rank, and the overall score. The dataset allows for a comprehensive analysis of the higher education landscape across different countries and institutions, providing insights into global and national rankings and the factors that contribute to the success of these universities.

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:

The dataset contains 2000 rows and 9 columns.

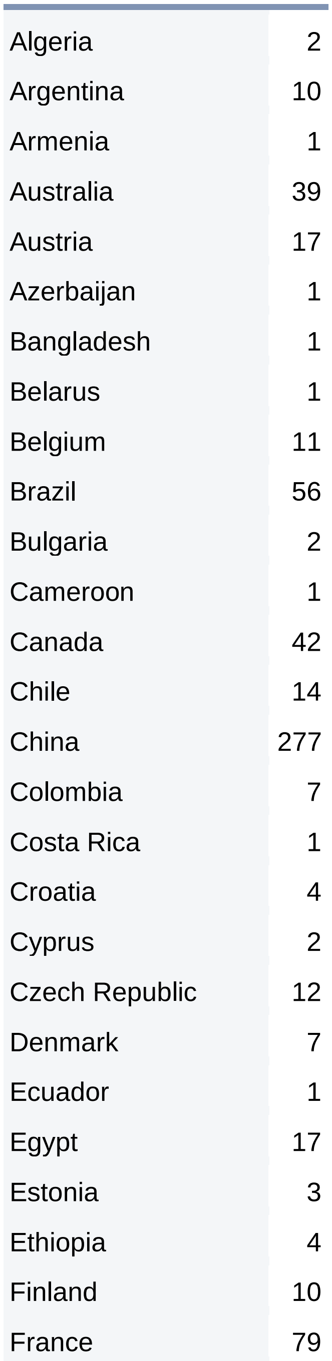
Shape of Dataset:

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)



Algeria	2
Argentina	10
Armenia	1
Australia	39
Austria	17
Azerbaijan	1
Bangladesh	1
Belarus	1
Belgium	11
Brazil	56
Bulgaria	2
Cameroon	1
Canada	42
Chile	14
China	277
Colombia	7
Costa Rica	1
Croatia	4
Cyprus	2
Czech Republic	12
Denmark	7
Ecuador	1
Egypt	17
Estonia	3
Ethiopia	4
Finland	10
France	79

Georgia	3
Germany	70
Ghana	2
Greece	12
Hungary	8
Iceland	2
India	68
Indonesia	2
Iran	39
Ireland	9
Israel	9
Italy	66
Japan	124
Jordan	2
Kazakhstan	1
Kenya	1
Kuwait	1
Latvia	2
Lebanon	2
Lithuania	4
Luxembourg	1
Malawi	1
Malaysia	8
Malta	1
Mexico	18
Morocco	4
Netherlands	15
New Zealand	8
Nigeria	2
North Macedonia	1
Northern Cyprus	1
Norway	13
Oman	1

Pakistan	10
Peru	2
Philippines	3
Poland	40
Portugal	14
Qatar	2
Romania	10
Russia	46
Saudi Arabia	9
Senegal	1
Serbia	4
Singapore	4
Slovak Republic	3
Slovenia	3
South Africa	13
South Korea	60
Spain	53
Sri Lanka	2
Sweden	19
Switzerland	12
Taiwan	35
Tanzania	1
Thailand	10
Tunisia	5
Turkey	55
Uganda	1
Ukraine	6
United Arab Emirates	3
United Kingdom	95
Uruguay	1
USA	347
Venezuela	1
Vietnam	5

Zambia	1
Zimbabwe	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.**

VAR of Score	25.80431355
STDEV of Score	5.079794637
MIN of Score	65.7
MAX of Score	100
SUM of Score	143172.9
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.

On Google Spreadsheet, Column T & U

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):

- How do university rankings vary across different countries, and what are the trends in the rankings of universities within specific countries?

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

- **Rows:** Countries, University
- **Columns:** Score
- **Values:** Average Score

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

- The average of world rank for each country.

- **Reason**: Averaging the World Rank for universities within a country provides insight into the general ranking trend within that country. This helps to identify whether a country typically has high-ranking universities or a mix of different ranks.

4. Data Filters if needed (1 point):

- Top 100 Universities

5. Plot the data in pivot table (3 points)
Link of the Pivot table:

- https://docs.google.com/spreadsheets/d/1lmWatpZXMkHYJ-YnC81o-ViJnvka5XEcY1VJL2N2xDE/edit?usp=sharing

6: Conclusions (1 point):

- The pivot table reveals that certain countries consistently have universities that rank higher globally, indicating strong educational systems and possibly better funding, research facilities, or international collaboration. Conversely, countries with lower average ranks may indicate areas for improvement in their higher education systems. This analysis helps identify trends and outliers in global university rankings.