

Usecase 3 - (Project 2)

By: eng. Esraa Madhi

Utilizing your knowledge of NumPy, pandas, matplotlib, seaborn, and Plotly, proceed with the analysis of the provided data.

This project must at least satisfy the following minimum requirements:

Usecase 3

For any **Data project** we should go through these steps:

Step 1: Defining the Problem Statement

- a. Which universities are ranked in the top 10 globally?
- b. Which universities are ranked in the top 10 for employment outcomes?
- c. What positions do universities in Saudi Arabia hold within the global rankings?
- d. Considering various factors such as employment rankings, research rankings, and others, which has the most significant impact on a university's overall ranking?
- e. Is there a correlation between national and global university rankings, and based on this information, can you recommend a country that appears to have a high concentration of top-ranked universities?
- f. Bonus: Develop two additional questions that could be explored using the data set at your disposal.

Step 2: Collecting Data

- Use the following dataset.
 - <https://www.kaggle.com/datasets/ourfuture/world-university-rankings>

Step 3: Data Quality Checking and Remediation

Step 4: Exploratory Data Analysis

- For these two steps, make sure to do:
 - a. Data Profiling: apply the 7 types of data profiling
 - b. Data Cleaning: handle missing values, correcting errors, and dealing with outliers.
 - c. Univariate Analysis & Bivariate/Multivariate Analysis: to understand their distribution and look at the relationships between variables. For your visualizations make sure to:
 - Drive meaningful insights (at least 10 different charts).
 - Choose a specific style for your charts.
 - Apply one color palette from your choice on all charts.
 - Use the title, x-y labels, font size, figure size, and legends.
 - Bonus: Create your charts using Plotly.

Step 5: Building Machine Learning Models

Not applicable

Step 6: Model Evaluation

Not applicable

Step 7: Communicating Results

- Report your final conclusion and findings in one page (readme markdown file).
 - Team members.
 - Introduction (Problem, Objectives)
 - Dataset Overview and Source.
 - List of EDA steps that applied on data with description
 - Describe the final ten insights with their charts

Step 8: Model Deployment

Not applicable

Step 9 : Model Performance Maintenance in Production

Not applicable

Note: the red steps means they are Not applicable in the project

- The Final presentation will be on Sunday.
- Due Date: Sun, 12 Jan, 08:00 AM.

Final Deliverables:

- Notebook file(.ipynb).
- Presentation of the result in 3 slides
- README.md file.