6-Usecase 2 (Lab)

By: eng. Esraa Madhi

So we defined data science as: It's the process of asking interesting questions, and then answering those questions using data.

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

1. Defining the Problem Statement
2. Collecting Data
3. Data Quality Checking and Remediation
4. Exploratory Data Analysis
5. Building Machine Learning Models
6. Model Evaluation
7. Communicating Results
8. Model Deployment
9. Model Performance Maintenance in Production

Usecase 2

Step 1: Defining the Problem Statement

We are eager to learn about the **World Happiness**. We would like to know:

1. What countries or regions rank the highest in overall happiness and each of the six factors contributing to happiness?
2. How did country ranks or scores change between the 2015 and 2016 as well as the 2016 and 2017 reports?
3. Did any country experience a significant increase or decrease in happiness?
4. Bounce: Please begin your analysis, and don't hesitate to consider additional relevant questions.

Step 2: Collecting Data

World Happiness Reports was collected for 5 years starting from 2015 until 2019

Step 3: Data Quality Checking and Remediation

To do in the notebook

Step 4: Exploratory Data Analysis

To do in the notebook

Step 5: Building Machine Learning Models

Not applicable

Step 6: Model Evaluation

Not applicable

Step 7: Communicating Results

To do in the notebook

Step 8: Model Deployment

Not applicable

Step 9 : Model Performance Maintenance in Production

Not applicable

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, 4 Aug, 09:00 AM.**

Final Deliverables:

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