Data Science Bootcamps

SDAIA Academy

❖ The Design

The project will be around education area where the goal of budling the solution is to discover data about students' level and disruption and predict students' level of study and graduation over the next years and how this will impact

- Education in how to distribute student around the schools based on predicted result.
- economy in how to compare number of graduates on next years with the amount of unemployment.
- how to utilize disruption into different collages and majors that help actual need.

Questions to be discovered:

- How many numbers of students currently we have?
- How the students disrupted around the kingdom?
- How students we expect to have on next years?

❖ The dataset

Will have on this project data of students on all levels that contains almost 100k rows and below features:

- Student gender
- Student school location
- Study Level
- Student class
- Student school city
- Schools' fullness

❖ The algorithm

Given the requirement Linear regression algorithm seems to be suitable to be used on this project.

❖ The tools to be used in EDA, Modeling and Visualization

EDA will be used as follow:

- Pandas to importing dataset and to be used on manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features.
- Seaborn to be used for data visualization and exploratory data analysis.
- Folium is a used for visualizing geospatial data.