



Done By : Rawan Al-Qahtani

Instructor : Mejdal Al-Qahtani

T5 Bootcamp

Data scientists

# Project Proposal

## Heart Disease Dataset

- **Question/Problem statement:**

Predicted the Patients are have heart disease or not?

- **Data Description:**

Main data source is Heart Disease from kaggle site .

([https://www.kaggle.com/captainozlem/framingham-chd-preprocessed-data?select=CHD\\_preprocessed.csv](https://www.kaggle.com/captainozlem/framingham-chd-preprocessed-data?select=CHD_preprocessed.csv))

- **Columns:** contains 16 features.  
Male, Age, Education, Current, Smoker, CigsPerDay, BPMeds, PrevalentStroke, PrevalentHyp, Diabetes, TotChol, SysBP, DiaBP, BMI, HeartRate, Glucose, TenYearCHD.
- **Data Type:** Integer, Float.
- **Rows:** contains 4133 rows
- **Data size:** contains 4133 rows x 16 columns.

- **Tools:**

- **Programs:**  
we will use Python .
- **Libraries:**  
Pandas, matplotlib, seaborn, numpy, KNeighborsClassifier, metrics.
- **Functions:**  
We will use several functions that we define in the project.
- **Plots:**  
We will use one of them (lineplot, Scatter, bar, hits, pie, circle).

- **MVP Goal:**

The goal of this project is to Predict the Patients are have heart disease or not.