



T5 Boot Camp

Data Science

Project Proposal

No.1

Surayyi Alqahtani

Instructor:

Mejdal Al-Qahtani

- **Question/Problem statement:**

Stroke is a life-threatening condition that is caused by a lack of blood and oxygen flow to the brain cells. It is associated with long-term disabilities.

Scenario: The Ministry of Health intends to establish an awareness campaign to raise awareness about the risk of stroke and wants to choose the right clinic that contributes to reaching the largest possible number of patients likely to have a stroke.

Q\ What are the most common diseases that associated with stroke?

- **Data Description:**

The Dataset is Called healthcare dataset stroke from Kaggle.

Columns: ID (Int), Gender (Object), Age (Float), Hypertension (Int), heart_disease (Int), ever_married (Object), work_type (Object), Residence_type (Object), avg_glucose_level (Float), BMI: body mass index (Float), smoking_status (Object), stroke (Int).

Data size: (5110 rows × 12 columns).

- **Tools:**

Programs: Python programming language, SQL.

Libraries: Pandas, NumPy, Matplotlib, seaborn, SQLAlchemy.

- **MVP Goal:**

The goal of this project is to determine what is most common diseases that associated with stroke, so the Ministry of Health can decide which is the proper clinic to establish an awareness campaign. Such as, Cardiac Center, Anti-Smoking Clinics and Diabetes Clinic.