Descriptive Analytics for Stroke

Abstract:

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. The goal of this project is determined 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.

Design:

The project was originated to decide which diseases that associated with stroke from dataset that have multi columns named by diseases, so it will help to choose best clinic to establish an awareness campaign for stroke.

Data:

The stroke dataset is containing many features such as ID, Gender, ever_married, work_type, hypertension, avg_glucose_level and BMI (Body mass index).

Rows: 5111 rows.

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).

Algorithms:

Generate new columns from BMI-to-BMI type using the domain knowledge based on some condition. (In ppt more details).

Tools:

Programs: Python programming language.

Libraries: Pandas, Matplotlib, seaborn, and SQLAchemy.

Communication: