Assignment 1

Context

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths.

This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relevant information about the patient.

Dataset: **Stroke Prediction Dataset**

Attribute Information

1) id: unique identifier
2) gender: "Male", "Female" or "Other"
3) age: age of the patient
4) hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
5) heart_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
6) ever_married: "No" or "Yes"
7) work_type: "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
8) Residence_type: "Rural" or "Urban"
9) avg_glucose_level: average glucose level in blood
10) bmi: body mass index
11) smoking_status: "formerly smoked", "never smoked", "smokes" or "Unknown"*
12) stroke: 1 if the patient had a stroke or 0 if not
*Note: "Unknown" in smoking_status means that the information is unavailable for this patient

Task 1: Exploratory Data Analysis

Step 1: Download the github codes
Step 2: Perform the Exploratory Data Analysis by running this code (00_data_exploration.py)

Task 2: Create XAI explainer
Task 3: Visualize local Explainer using Force plot
Task 4: Visualize global Explainer using Summary plot
Task 5: Prepare in three slides, explaining the model using SHAP

- **As data scientist**

- **As Doctor**

- **As Patient**

Bonus:
Task 6 : Create Deep SHAP or Kernel SHAP as an explainer and explain a chosen black box model different from Task 2.