# Milestone 2 Report: Data Analysis and Visualization

## 1. Objective

The primary goal of this milestone was to conduct exploratory data analysis (EDA) and visualization on the stroke prediction dataset to better understand the underlying patterns and relationships between variables, especially in relation to the 'stroke' outcome.

## 2. Data Preparation

The preprocessed dataset (preprocessed\_stroke\_data.csv) was loaded for analysis. Necessary libraries such as pandas, matplotlib, seaborn, and sklearn.preprocessing were imported.

## 3. Data Cleaning

Null value analysis was performed to ensure the dataset was clean. It was confirmed that no missing values were present, ensuring reliable analysis.

## 4. Exploratory Data Analysis (EDA)

A visual analysis using Seaborn’s pairplot was conducted to explore the relationships between features with respect to the stroke target. The dataset was divided into stroke and non-stroke groups to compare key features like age.

## 5. Statistical Analysis

A T-test was performed to compare the mean age between stroke and non-stroke individuals. The results indicated whether there was a statistically significant difference in age distribution between the two groups.

## 6. Key Insights

Visualizations revealed that certain variables such as age may have a noticeable impact on stroke probability. The T-test provided a statistical foundation for confirming this hypothesis.

## 7. Conclusion

This milestone successfully established a solid understanding of the dataset through visualization and basic statistical analysis. These insights will help in guiding feature selection and model development in the next phases.