Exploratory Data Analysis

Stroke Prediction Based on Risk Factors

# Project Description:

The objective of this project is to utilize stroke data to create a predictive model for identifying individuals who have a higher likelihood of experiencing a stroke. Our team will be analyzing the correlation and impact between risk factors and the occurrence of strokes in individuals.

# Team Members:

* Aisha Henderson
* Kelly Blake
* Rajeshwari Radharkrishna
* Elona Barjami

# Research Questions:

1. Is there a correlation between obesity (high BMI) and the likelihood of experiencing a stroke?
2. Does a person’s residency type increase the risk of having a stroke? Are individuals that live in Urban settings more likely to have a stroke vs. those that live in a Rural setting?
3. Are males over 60 at a higher risk of having a stroke compared to males overall? Does smoking status (current smoker) have an impact on the likelihood of stroke for both populations?
4. Are patients with hypertension vs patients with heart disease more likely to have a stroke?

# Dataset to Be Used

# **Description:** The dataset below contains data for over 5000 unique stroke patients, with columns included for individual clinical diagnoses, and demographic information.

# [Stroke Prediction Dataset](https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset)

Source: kaggle.com

# Rough Breakdown of Tasks

1. **GitHub & ReadMe File** 
   1. Owner: Raji
2. **Visualizations** 
   1. Owner: Elona
3. **Data Cleaning** 
   1. Owner: Kelly
4. **Analysis & Conclusion –** *Each Group Member will be responsible for 1 research question to complete the statistical analysis and write up.* 
   1. Summary/Write Up
   2. Statistical Analysis
   3. Research Questions Answered
5. **Presentation Slides**
   1. Owner: Aisha