

# Project Proposal Template

## Question/need:

- The purpose of this model is to predict whether or not someone will be diagnosed with a stroke given some characteristics
- Patients at high risk from stroke may benefit from this model because it will allow them to understand the contributing factors of stroke (so they may mitigate those factors to reduce their chances of having a stroke). Doctors may also benefit from this model because it can be applied in preventative medicine.

## Data Description:

- I will be downloading a publicly available stroke prediction dataset off of kaggle.

Link: <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset>

- The target for this model is whether or not the person (each row) was diagnosed with stroke, while features include: gender, age, hypertension, heart disease, marital status, work type, residence type, glucose level, bmi, and smoking status. There are a total of 5110 rows and 10 columns.

## Tools:

- In order to complete this project I plan on using the following tools: pandas, numpy, seaborn, matplotlib, scikit learn, and tableau

## MVP Goal:

- An MVP for this project would be a baseline classification model such as logistic regression with a few features and evaluation metrics such as confusion matrix and f1 score.