1. What research question are you trying to answer and why is it important?

Can we accurately predict if a person is going to have a stroke? Strokes can have a very large impact in lifestyles.

1. What is the description of the data you will use, what are examples of input variables and output variable (variables) in the data?

Dataset from Kaggle

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

Input Variables

* Gender
* Age
* Hypertension
* Heart disease
* Married
* Work
* Residence type
* Average glucose level
* BMI
* Smoking Status

Output

* Stroke

3. What algorithms are you planning to use or what type of models are you planning to build? (You should use more than a single algorithm to solve the same problem, this is experimentation, so you can compare the results at the end and see which algorithm works better

I’m planning to use Naïve Bayes, Random Forest classifier, and SVM. I’m kind of want to try a KNN just to see how it goes.