Kristie Kooken

Project 2, Diabetes

For my final project, I am interested to better understand if certain characteristics can predict whether or not a woman might develop diabetes. I have a close friend who became diabetic during her pregnancy and now, she and her child both have diabetes. It made me wonder if it would have been possible to predict diabetes based on certain factors, such as glucose levels. Once factors are established, it would be possible to then look into interventions by understanding if changes in the predicator variables can influence the outcome.

When considering what could have an impact on my research question, I think that glucose levels, amount of exercise and heredity (having a parent or grandparent who had diabetes) can change course for whether an individual can become diabetic. While there are many factors that can influence becoming a diabetic, my grandmother had diabetes in her old age and these were factors that greatly influenced her outcomes (and likely mine as well). Diabetes is a very serious health condition and important area of medical research to help the population of people who become diabetic whether born with diabetes or become diabetic in older age.

In order to investigate the factors that influence diabetes, I began to explore different data sources available to me from the internet.

Data Source: Earlier in the term, I reviewed kaggle.com site to explore the available data sources for this project. As I am learning both coding in Python and statistical methods, I wanted to have a dataset that was not as complex or complicated as some of the government data site. For example, I explored using hospital emergency visit survey data but found the data to be very complex and wanted to be able to manage my understanding of the process so settled on a data source that is a subset of a very large dataset from National Institute of Diabetes and