Final Capstone Proposal

1. **What is the problem you are attempting to solve?**
   The US Census Bureau collects income data annually in order to analyze it and determine the profiles of people that make a certain amount of money each year. This information could prove useful to them in many ways. The problem I want to solve is to use data from the US Census Bureau to predict the the income amount of a persons based on features such as education, occupation, race, sex and native-country. If they wanted to focus on a specific amount such $50,000 a year, I could then create a model to classify if a person will make below or above that specific amount.

2. **How is your solution valuable?**
   This solution is valuable because the Income Census Bureau may want to use this model for their annual income data to determine the profile of a person that makes above a certain amount and below a certain amount. They have the data, but may not be able to fully use the data to their advantage. By analyzing it and creating a classification model, we could use it to determine what are the key factors in the annual income amount.

3. **What is your data source and how will you access it?**
   I will be using a data set the US Census Bureau Website. I will load the data using pandas. Then I will look at the raw data, then use some feature engineering to clean the data if necessary.

4. **What techniques from the course do you anticipate using?**
   I will use feature engineering, data visualization, modeling to create a prediction model. I will create different models such as random forest, gradient boosting, naive bayes, and KNN. I will also use PCA if I need to reduce dimensionality if necessary. ONce I have created my models, I will then use five fold cross-validation in order to determine which model is the best for predicting the income. In order to test the accuracy of my model, I will apply it to a different but similar data set.

5. **What do you anticipate will be the biggest challenge you'll face?**
   The biggest challenge will be creating and effective model with the best accuracy. Since beginning this data science bootcamp, my python skills have improved. However, creating an effective model is something I still practicing, so I hope the models I create will help me in classifying effectively.