**Instructions**

**Problem:**

We have been given census data about attributes of US citizens (occupation, education, gender, race). With organizations working to ensure equal pay, we want to build a model to see how accurately we can classify low income from high income citizens. Building this model will allow us to understand which attributes contribute to affluency and how we can improve policies in the US.

**Part 1:**

Use proper data cleansing techniques to ensure that you have the highest quality data to model this problem. Detail your process and discuss the decisions you made to clean the data. You should include how you cleaned the data and why you choose the method.

**Part 2:**

Explore the dataset and observe if there are significant differences between high and low income individuals. Is it job based? Education based? Use data visualization techniques to talk to key learnings that the government would be curious to learn about.

**Part 3:**

Build a nearest neighbors model with the given data, interpret the results, and convey those results to stakeholders. Highlight key learning points such as how you determined K, why you choose that final value of K, and the overall accuracy of your model and accompanying models. Build **3** models with your selection of 3 independent variables (ensure that you **aren't using** three different values of K but 3 different KNN models with 3 variables). The reader should come away with a clear outcome of whether or not we can correctly classify citizens and why matters. Overall do you recommend that (and have evidence) that citizens can be distinguished into low/high income based on their attributes.

The paper needs to be no more than five pages in APA format. Attach images in the appendix. Code should be provided in an additional document, **not in the paper**.