Gabriel Reyes

December 9, 2023

CS 1030

Final Project

**Data Tools**

The module was overall easy for me because I had already learned many of the things that the module tries to teach. Although it does help to be refreshed on these tools to make computing easier. Some things that I learned from the module is that datasets can come in many different forms.

**Big Data**

The module was also easy for me to pick up even though I never really research any datasets or use large spreadsheets other than for finances. At the beginning it was a little confusing for me, but I managed to get the hang of it pretty quickly and got through the module in a couple of minutes.

**Bias in Machine Learning**

Overall, the Machine Learning module was more extensive and had more reading material than the previous two modules, so I had a harder time keeping my focus throughout the module. Some things that I have learned from the module are that with facial recognition, the darker your skin is, the less accurate it could be, causing confidence issues with facial recognition systems.

**Unit Test**

<p> When I took the unit test, it wasn’t that tedious but there was a lot of reading and quite a bit of unnecessary information in my opinion for some of the questions but overall, I felt really good taking the unit test as not only were some of the questions asked were in the previous modules but also, the format of the questions was relatively easy to understand. </p>

**Questions for Investigation**

Do men carpool more often than women?

**Process and Results of Analysis**

What I did to find the dataset was to look for data sets that involve employment on the US Census website. The way that I decide to filter down the data to get what I needed was to hide the columns and rows that were not important to answer the question. The result is that Men will drive more often than women and men will more often drive alone while women will more often carpool than men but not by more than 1.5%.

The reason why I chose the question I asked was because it seemed like it was the one that could be answered with the given data. Downloading the dataset was simple as the Census website had multiple download buttons for different format versions of the dataset. Uploading to GitHub was also fairly simple, all I did was create a new repository and upload the original dataset and the modified dataset as well as this document.

A graph of men and women

Description automatically generated