**HW 10**

**Learning Experience**

**Part 1:**

The first part of the homework involved to be more analytical reasoning since we needed to find a way to relate some data with others in order to create a graph and find the right kind of graph for the relationship. As usual, we need to import pandas to read CSV files. However, there was a problem in the txt file since it did not have a row header. So, I solved this issue by passing in the read\_csv the header as a list.

**Graphical user interface, text

Description automatically generated**

After that, I was ready to plot the graphs. I addressed 9 possible questioning from the data frame: People relationship by Ethnicity, Degree by Country, Occupation by Race, Marital status by Race, Capital Gain by Age, Capital Loss By Sex, Hours Per Week By Race, Age by Sex, and Common Groups by Age.

**People relationship by Ethnicity:**

Chart, bar chart

Description automatically generatedI grouped by race and relationship to show in the graph how many people are distribute by different kind of relationship. We notice that most of the white people are husband, has own child, and also unmarried, while the minority are the less in each category.

**Degree by Country:**

Again, I grouped by education for each category (countries), and we can see that in United States, we have most people with an education. Therefore, we can conclude from this graph that people are more likely to have a degree in the US than in other countries.

Chart

Description automatically generated

**Occupation by Race:**

In this graph, I also group race for each occupation category. The white people dominate in this graph having occupied more of the jobs. Minorities is greater in private house service position having approx. 25% people working in this job VS White people occupies the 75% in this job.

**Chart, bar chart

Description automatically generated**

**Marital status by Race:**

We notice in the graph that White leads this graph having a 90% of white people married-AF-spouse. In addition, 25% of Black people are separated according to the graph. Also, Asian people are not married or spouse in the graph.

**Chart, bar chart

Description automatically generated**

**Capital Gain by Age:**

We can see that people between the age of 20 and 70 are likely to gain a capital of 100000 but they are few, while the rest of the people are most likely to gain less than 20000 of capital according to the graph.

**Chart, scatter chart

Description automatically generated**

**Capital Loss By Sex:**

This graph is a capital loss VS sex. We see that females doesn’t have a capital loss when it is a bit but when capita loss increments, we see that females dominates in the graph. Now, in the extremes of the graph we see that men dominates having more capital loss than women.

**Chart, bar chart

Description automatically generated**

**Age by Sex:**

According to our data, there are more males than females overall, but when age is 86, 100% of the people who are 86 years old are female.

**Chart, background pattern

Description automatically generated**

**Common Groups by Age:**

We see in the graph that the most common group age is the people who are from 20 to 40 years old. Showing that youth lead in the community.

**Chart, bar chart

Description automatically generated**

**Part 2:**

The first part of the assignment refreshes what we learned in the article. Basically, we need to import pandas in order to use the library along with matplotlib. First, we needed to create a dataframe with a given dictionary. In order to create a dataframe, we only need to pass a dictionary.

Text

Description automatically generated

The next part, we needed to add a new column in the data frame. We can create a new column with the following: dataframe[‘newColumn’] = (list of values for the column)

A screenshot of a computer

Description automatically generated with medium confidence

Additionally, we can also display single columns with the similar syntax but with double square brackets so the output is a table.

A picture containing graphical user interface

Description automatically generated

Moreover, we can sum the whole column only if the datatype is an **int** or **float** with the following code: dataframe[columnName].sum()

Text

Description automatically generated with medium confidence

Furthermore, we can also select some columns from out dataFrame with following code: dataFrame[[ column1, column2 ]] and store this in a new variable so that we can manipulate it without affecting the original dataframe

A screenshot of a computer

Description automatically generated with medium confidence

In my enhancement, I plotted a graph of type ‘scatter’, which shows dots between X and Y axis. X-axis is the engine power VS Y-axis the weight of the car. In the graph, we see that while the engine increases, the weight also increases. So, the graph shows a pattern, and we can predict that an engine more than 460 is likely to have a weight more than 4400

Chart

Description automatically generated