

## Question2

\*Since some part of the questions was not clear for me, I have supposed some circumstances. This report involves these different explanations.

1. Each row in data represents the traffic stop. Here we want the proportion of male drivers with respect to both genders. I used Pandas to find male drivers. I have ignored 'Nan' values using 'dropna()'; This proportion is: **0.67**.
  2. Factor increase is:  $\frac{Value_2 - Value_1}{Value_1}$  I found the number of traffic stop arrests and traffic stop arrests that occurred out of state, using: **stop\_outcome** and **out\_of\_state** columns. The factor increase is **2.53**. \*I calculated chi-squared for driver age of arrest stops using Scipy package.
  3. Due to part 1, the proportion is **0.658**.
  4. Due to part 2, I found the factor increase. Here I used the violation column and using contain method to find all DUI. The factor increase is **10.9**.
  5. \*There was no 2020 stop date; therefore, I changed it to 2010. First, I selected all rows that had a stop date equal to 2010. Then, I calculated the average of vehicle year column. The result is **2001**. \*I supposed x as vehicle year and y as stop year. Then I used the 'Linregress' method in Scipy to find p\_value, and it was **0.0**.
  6. \*Min hour is 00:00 and Max hour is 23:59. So I counted the number of Min and Max and then the difference, for both MT and VT difference was **121**.
  7. \*I didn't find any table with the surface area on the Internet. I did this part of the question concerning the population of countries. First, I found a webpage including a table of all U.S. countries, and then I used BeautifulSoup to scraping the web page. Then I split the particular content from the web page and put it in a list of tuples. Then by sorting the list, the most populated country was found. The most crowded county in MT: **Jefferson County**.
- Use Jupyter Notebook or Google Colab (Import data from Google Drive) to run the script.