**ASSIGNMENT 5**

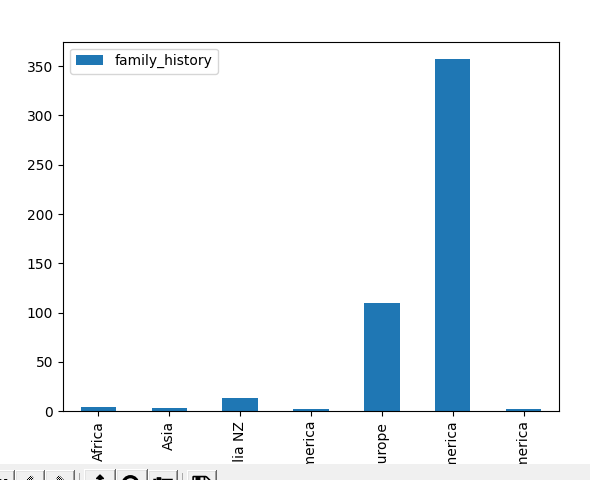
In this assignment, we had been given a dataset called survey.csv which was an unclean dataset that was a stored response from a large audience from around the world related to Mental Illness.

Notable Impediments –

1. The data was not clean at all and we had to clean the data.
2. For this, I have used a tool called OpenRefine to a certain extent and later, to try my hand on how I could clean the rest of the data using pandas, I cleaned certain columns using pandas.
3. There is a separate python file in my project, that deals with only functions related to data cleaning.
4. I have also dropped certain columns that were not required for my analysis.

Analysis –

1. For the initial analysis, I wanted to check how well or bad people are doing with their mental health in different regions of the globe.
2. To facilitate that, I have created a new column called ‘geographical\_region’ which encompasses several countries from similar geographic locations into one umbrella category.
3. The graph for the first analysis is as follows –



1. From this graph, we can see that America has the most cases of family history in terms of mental illness.
2. In the next graphs, we try to understand if there are people in our dataset who experience interference in work due to their mental illness and try to understand if the reason could be their tech jobs.

|  |  |
| --- | --- |
|  |  |

1. From the above charts, we can see that there are way too many cases of people experiencing interference with work due to mental illness often in North America rather than other countries.
2. But, from the data that we have, we cannot conclude that the reason for their mental illness is not because of tech jobs.

GitHub –

1. I have made a GitHub repository called ‘[mental\_health\_in\_tech](https://github.com/sairamgunner/mental_health_in_tech)’. ()
2. From the ‘main’ branch, I have created a branch called ‘dev\_branch’.
3. All the work has been done in the ‘dev\_branch’ and later, this branch has been merged into the ‘main’ branch.
4. Below is a screenshot showing the activity on the branches –

