**Mental health in Tech Survey**

**Data Source:**

Mental health in Tech SurveyCSV file– [Mental Health in Tech Survey | Kaggle](https://www.kaggle.com/osmi/mental-health-in-tech-survey?select=survey.csv)

This dataset contains data of survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace details like timestamp, age, gender, country, state, self-employment status, family history (Do you have a family history of mental illness?), treatment (Have you sought treatment for a mental health condition?), remote work (do you work remotely (outside of an office) at least 50% of the time?), leave (How easy is it for you to take medical leave for a mental health condition?), supervisor (Would you be willing to discuss a mental health issue with your direct supervisor(s)?), etc.

**Overview:**

In this dataset, we are going to analyze the frequency of mental health illness and attitudes towards mental health vary by geographic location and the strongest predictors of mental health or certain attitudes towards mental health in the workplace.

In this milestone 1, we are going to perform the following graphical analysis:

1. Comparing the mental illness of employees based on the gender category.
2. Comparing the mental illness of employees based on the countries.
3. Comparing the mental illness of employees based on the remote work culture.
4. Comparing the mental illness of employees based on the co-workers.

I have taken this analysis as a personal interest to find how employees are happy and to know are they having any mental illness problems. This analysis could help some of the organizations and start-ups to restructure their organization to improve the happiness of employees.

**Data Cleaning:**

In this data cleaning phase, I have removed the following columns which are unnecessary for the analysis.

1. state - We are comparing the countries only in the graph analysis
2. self\_employed - We can avoid this because our focus is only on the corporate employees
3. family\_history - We are concerning the current status only
4. phys\_health\_consequence - We are concerned only about the mental stability
5. phys\_health\_interview - We are concerned only about the mental stability
6. Timestamp - We are dealing with a time frame so we can drop it
7. comments - It has most NaN and it's not required for the analysis
8. Also, We don't want some more columns like no\_employees, tech\_company, benefits, care\_options, wellness\_program, seek\_help, anonymity, supervisor, obs\_consequence, work\_interfere, leave, etc.

Similarly, I have converted all rows into uniform entries by following the below procedure for all the columns.

Text

Description automatically generated

**Visualization:**

### **Comparing the mental illness of employees based on the gender category.**

Chart, histogram

Description automatically generated

From the above visualization, we found that the mental illness for men is higher than the women.

### **Comparing the mental illness of employees based on the countries**

Application

Description automatically generated with low confidence

From the above graph, we found that after a certain deeper analysis, I have included all the countries and also did a graph with countries having several people with illness and without. In both, USA is the first and it shows that the number of entries is different for different countries and last week's conclusion is wrong.

1. **Comparing the mental illness of employees based on the remote work culture:**

Chart, histogram

Description automatically generated

From this visual, we can say, Remote employees also have a mental illness at the same rate as non-remote workers. So remote jobs won't give that much relief for the mental illness.

1. **Comparing the mental illness of employees based on the co-workers:**

Chart, histogram

Description automatically generated

From this one, If we have more co-workers then there is a high possibility of reducing mental illness.

**Model selection and Evaluation:**

For this scenario, I have selected the KNN classifier and perform the training as follows:

Graphical user interface, text, email

Description automatically generated

After evaluating the model, we got the accuracy as shown in the below graph,

Chart, line chart

Description automatically generated

**Conclusion:**

Based on the visualization, we can say that the factor like gender, co-workers and remote work cultures plays a vital role in a person’s mental health condition. It has been proven in the model which we have built and evaluated as above.