# Richard Kabiru Waithera

Student Number: 150684

# Mental Health Survey Analysis

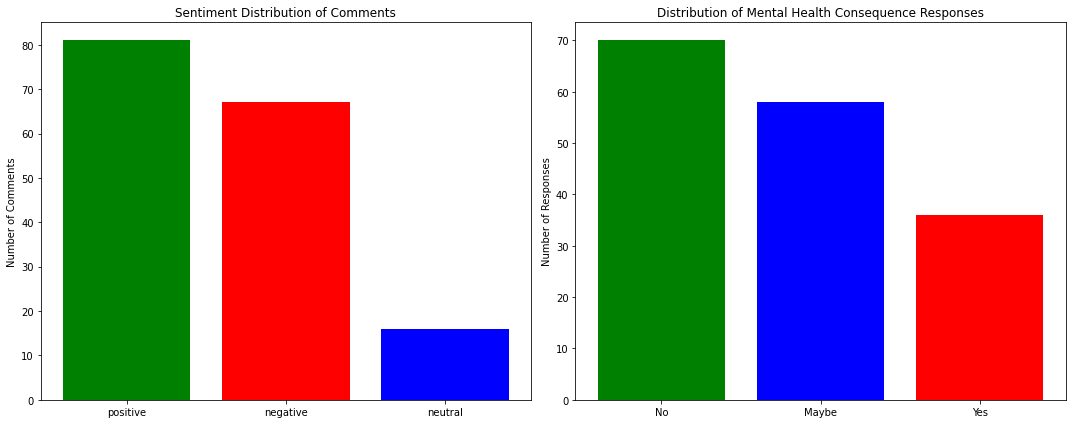
## Data Loading and Exploration

The dataset seems to be a survey related to mental health in the tech workplace. It has multiple columns, including demographic information (like Age, Gender, Country, etc.), workplace-related questions, and a comments column.

## Sentiment Analysis of Comments

The sentiment analysis has been applied to the comments. Here are a few examples:   
- Comment: 'I'm not on my company's health insurance which would cover therapy, but I don't want to use drugs for my mental issues. I want to use therapy.' Sentiment: Positive   
- Comment: 'I have chronic low-level neurological issues that have mental health side effects. My employer has been very supportive, but I did not feel comfortable sharing with my previous employer, who was also aware of the issues.' Sentiment: Positive   
- Comment: 'My company does provide healthcare but not to part-time employees.' Sentiment: Negative   
... and so on.

## Visualization and Comparison



The visualizations display the distribution of sentiments for the comments and the responses to the mental\_health\_consequence question. While the majority of the comments lean towards positive sentiment, the responses to the mental\_health\_consequence question show a more varied distribution. This indicates that while the general sentiment in the comments might be positive, many respondents still have reservations or concerns about discussing mental health at work.

## Summary

To summarize:  
- The comments in the dataset mostly have a positive sentiment.  
- However, when it comes to discussing mental health issues at work, there's a significant portion of respondents who believe there might be negative consequences or are unsure.  
- Next steps could involve a deeper dive into the comments to understand the themes or topics discussed or further comparisons with other survey questions to glean more insights.

**Analysis and Visualization of Mental Health Survey Data**

The dataset at hand provided insights into mental health conditions among individuals across various demographic and professional backgrounds. The goal was to derive meaningful interpretations from the data, particularly focusing on the sentiment of the comments and comparing it with other categorical variables present in the dataset.

**1. Data Exploration:** The first step was a preliminary analysis of the dataset to understand its structure, columns, and the types of data it contained. We also identified any missing or null values that could potentially affect the subsequent analysis.

**2. Sentiment Analysis:** A significant portion of the analysis was dedicated to the 'comments' column. Using the TextBlob library, each comment was analyzed to determine its sentiment - whether it was positive, negative, or neutral. This analysis provided an understanding of the feelings and perspectives of the respondents.

**3. Data Visualization with Plotly:** Utilizing Plotly, a powerful visualization library, several plots were crafted:

* A bar chart illustrating the distribution of sentiments among the comments.
* A histogram showcasing the age distribution of respondents.
* A bar chart depicting the responses to the question about the consequences of mental health conditions in the workplace.

These visualizations provided a graphical representation of the data, making it easier to discern patterns and insights.

**4. Data Visualization with Bokeh:** To diversify the visualization techniques, Bokeh, another prominent visualization tool, was employed for developing the dashboard. With Bokeh, the following charts were produced:

* Sentiment Distribution: A bar chart that provides a clear view of how sentiments are spread among the comments.
* Mental Health Consequence Distribution: This bar chart illustrated how respondents felt about discussing mental health conditions in a professional setting.
* Age Distribution: A histogram to understand the age groups of the respondents.
* Gender Distribution: Given the multitude of gender responses, a bar chart gave a clearer understanding of the gender breakdown.
* Work Interference Distribution: Highlighted how mental health conditions interfered with the respondents' work.

Enhancements were also made to improve the legibility of the x-axis labels on the charts by adjusting their orientation and font size.

**5. Documentation and Reporting:** Throughout the process, emphasis was placed on proper documentation. The analysis was not only presented here but also encapsulated in an IPython Notebook (ipynb) and a DOCX file. This ensures that the work is reproducible and can be easily shared.

**Conclusion:** The analysis of the mental health survey data was comprehensive, encompassing data preprocessing, sentiment analysis, and visualization. The use of the Plotly and Bokeh library offered diverse perspectives on the data, each bringing its strengths to the table. The sentiment analysis was particularly revealing, providing a deeper layer of understanding beyond the categorical responses. With mental health being a crucial subject, especially in professional settings, such insights are invaluable for organizations aiming to foster a supportive environment for their employees.