**Project Term:**

1. **Statistical/Hypothetical Question:**

This project aimed to answer the question: **"What factors influence the likelihood of seeking mental health treatment among employees?"** Using a dataset from Kaggle (<https://www.kaggle.com/code/shreyaspj/mental-health?select=survey.csv>), we analyzed the relationship between treatment-seeking behavior and several independent variables, including **family history of mental health issues**, **age**, **gender**, **work interference**, and **company size**.

1. **Outcome of Your EDA:**

The exploratory data analysis (EDA) revealed key insights:

* **Age:** The distribution was right-skewed, with most respondents between 25 and 40 years old.
* **Gender:** Males were the most common respondents.
* **Family History:** Most respondents did not have a family history of mental health issues.
* **Work Interfere:** The most common response was "Sometimes."
* **Number of Employees:** Most respondents worked in medium-sized companies (26-100 employees).

We also found strong relationships between treatment-seeking behavior and:

* **Family History:** Employees with a family history were significantly more likely to seek treatment **(74.03%)** than those without **(35.43%).**
* **Work Interfere:** Employees who reported that mental health often interferes with work were much more likely to seek treatment **(85.00%)** compared to those who reported never experiencing interference **(14.15%)**.

1. **What Was Missed During the Analysis?**

One limitation of the analysis was the lack of additional variables that could have provided deeper insights. For example:

* **Income Level:** Income could influence access to mental health resources.
* **Job Role:** Certain roles may have higher stress levels or better access to mental health support.
* **Geographic Location:** Access to mental health services may vary by region.

1. **Variables That Could Have Helped:**

Including variables like income, job role, and geographic location could have provided a more comprehensive understanding of the factors influencing treatment-seeking behavior. For instance, employees in higher-income brackets or specific job roles might have better access to mental health resources.

1. **Assumptions Made and Their Validity:**

We made several assumptions during the analysis:

* **Encoding of Categorical Variables:** We encoded Gender as binary (Male = 0, Female = 1, Other = 0). This simplification may have overlooked nuances in gender identity.
* **Handling Missing Values:** We imputed missing values using the mode for categorical variables and the mean for numerical variables. While this is a common approach, it assumes that the missing values are random, which may not always be true.

1. **Challenges Faced:**

* **Data Cleaning:** The dataset had missing values, inconsistent entries (e.g., variations in Gender), and invalid ages, which required careful handling.
* **Understanding Relationships:** While we identified strong relationships between certain variables (e.g., **family history** and **treatment**), understanding the causal mechanisms behind these relationships was challenging.
* **Limitations of the Dataset:** The dataset did not include all potentially relevant variables (e.g., income, job role), which limited the scope of the analysis.

**Conclusion:**

The analysis revealed that family history, gender, and work interference are strong predictors of treatment-seeking behavior among employees. These findings highlight the importance of addressing mental health issues in the workplace, particularly for employees with a family history or those experiencing frequent work interference. However, the analysis could be improved by including additional variables and addressing the assumptions made during data cleaning and encoding.