Project Name - Health Care Analysis

Project Type - EDA/Regression/Classification/Unsupervised

Contribution - Individual

Project Summary -

This project focuses on analyzing survey data related to mental health in the tech industry. The dataset includes responses from 1,259 individuals and covers various aspects such as age, gender, work type, treatment history, and awareness of company mental health support. Through data cleaning, visualization, and exploration, the project identifies key insights about mental health challenges in tech workplaces.

The analysis revealed that many employees have sought treatment, especially those with a family history of mental illness. However, awareness of mental health benefits is low, and many employees feel uncomfortable discussing these issues at work. Self-employed and remote workers also show unique patterns of behavior and need specific support.

The project uses 15 informative charts, including age distribution, gender trends, treatment patterns, company support, and correlation heatmaps. Each visualization provides actionable insights to help companies improve mental health awareness, policies, and employee well-being. The project concludes with practical recommendations to help companies create a healthier, more supportive work environment that boosts productivity and employee satisfaction.

GitHub Link -

https://github.com/dahiya121/Health-Care-Analysis-Project-

Problem Statement

Mental health issues are common in the tech industry, yet they are often overlooked or not openly discussed in the workplace. Many employees are unsure about seeking help, unaware of available mental health resources, or afraid of negative consequences if they speak up. This leads to reduced productivity, poor work-life balance, and increased employee burnout. The goal of this project is to analyze survey data to understand attitudes, challenges, and gaps related to mental health in tech workplaces and to provide actionable insights for improving mental health support and creating a healthier work environment.

Define Your Business Objective?

The main business objective of this project is to analyze mental health trends among employees in the tech industry and provide insights that help companies build a more supportive, inclusive, and mentally healthy work environment. By understanding factors such as treatment rates, workplace attitudes, and awareness of mental health benefits, the goal is to identify key areas where companies can improve policies, communication, and culture. This will lead to better employee well-being, higher productivity, reduced burnout, and improved employee retention — all of which directly contribute to stronger business performance.

General Guidelines : -

1. Well-structured, formatted, and commented code is required.
2. Exception Handling, Production Grade Code & Deployment Ready Code will be a plus. Those students will be awarded some additional credits.

   The additional credits will have advantages over other students during Star Student selection.

```
    [ Note: - Deployment Ready Code is defined as, the
whole .ipynb notebook should be executable in one go
              without a single error logged. ]
```

3. Each and every logic should have proper comments.
4. You may add as many number of charts you want. Make Sure for each and every chart the following format should be answered.

```
# Chart visualization code
```

- Why did you pick the specific chart?
- What is/are the insight(s) found from the chart?
- Will the gained insights help creating a positive business impact? Are there any insights that lead to negative growth? Justify with specific reason.

1. You have to create at least 20 logical & meaningful charts having important insights.

[Hints : - Do the Vizualization in a structured way while following "UBM" Rule.

U - Univariate Analysis,

B - Bivariate Analysis (Numerical - Categorical, Numerical - Numerical, Categorical - Categorical)

M - Multivariate Analysis]

Let's Begin !

1. Know Your Data

Import Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

Dataset Loading

```
ds=pd.read_csv("C:\\Users\\prave\\Downloads\\Mental Health Survey EDA
Analysis-20250622T131942Z-1-001\\Mental Health Survey EDA Analysis\\
survey.csv")
ds.head()
```

```
             Timestamp  Age  Gender          Country state
self_employed  \
0  2014-08-27 11:29:31   37  Female   United States    IL
NaN
1  2014-08-27 11:29:37   44       M   United States    IN
NaN
2  2014-08-27 11:29:44   32    Male          Canada   NaN
NaN
3  2014-08-27 11:29:46   31    Male  United Kingdom   NaN
NaN
4  2014-08-27 11:30:22   31    Male   United States    TX
NaN

  family_history treatment work_interfere    no_employees  ...  \
0             No       Yes          Often            6-25  ...
1             No        No         Rarely  More than 1000  ...
2             No        No         Rarely            6-25  ...
3            Yes       Yes          Often          26-100  ...
4             No        No          Never         100-500  ...

                 leave mental_health_consequence
phys_health_consequence  \
0       Somewhat easy                         No
No
1          Don't know                      Maybe
No
2  Somewhat difficult                         No
No
3  Somewhat difficult                        Yes
Yes
4          Don't know                         No
```

```
No

      coworkers supervisor mental_health_interview
phys_health_interview  \
0  Some of them        Yes                       No
Maybe
1            No         No                       No
No
2           Yes        Yes                      Yes
Yes
3  Some of them         No                    Maybe
Maybe
4  Some of them        Yes                      Yes
Yes

  mental_vs_physical obs_consequence comments
0                Yes              No      NaN
1         Don't know              No      NaN
2                 No              No      NaN
3                 No             Yes      NaN
4         Don't know              No      NaN

[5 rows x 27 columns]
```

Dataset First View

```
ds.shape

ds.isnull().sum()

Timestamp                      0
Age                            0
Gender                         0
Country                        0
state                        515
self_employed                 18
family_history                 0
treatment                      0
work_interfere               264
no_employees                   0
remote_work                    0
tech_company                   0
benefits                       0
care_options                   0
wellness_program               0
seek_help                      0
anonymity                      0
leave                          0
mental_health_consequence      0
```

```
phys_health_consequence          0
coworkers                        0
supervisor                       0
mental_health_interview          0
phys_health_interview            0
mental_vs_physical               0
obs_consequence                  0
comments                      1095
dtype: int64
```

Dataset Rows & Columns count

```
import pandas as pd



# Get the shape
rows, columns = ds.shape
print(f"Number of rows: {rows}")
print(f"Number of columns: {columns}")

Number of rows: 1259
Number of columns: 27
```

Dataset Information

```
ds.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 27 columns):
 #   Column                     Non-Null Count  Dtype 
---  ------                     --------------  ----- 
 0   Timestamp                  1259 non-null   object
 1   Age                        1259 non-null   int64 
 2   Gender                     1259 non-null   object
 3   Country                    1259 non-null   object
 4   state                      744 non-null    object
 5   self_employed              1241 non-null   object
 6   family_history             1259 non-null   object
 7   treatment                  1259 non-null   object
 8   work_interfere             995 non-null    object
 9   no_employees               1259 non-null   object
 10  remote_work                1259 non-null   object
 11  tech_company               1259 non-null   object
 12  benefits                   1259 non-null   object
 13  care_options               1259 non-null   object
 14  wellness_program           1259 non-null   object
 15  seek_help                  1259 non-null   object
 16  anonymity                  1259 non-null   object
```

```
 17  leave                      1259 non-null   object
 18  mental_health_consequence  1259 non-null   object
 19  phys_health_consequence    1259 non-null   object
 20  coworkers                  1259 non-null   object
 21  supervisor                 1259 non-null   object
 22  mental_health_interview    1259 non-null   object
 23  phys_health_interview      1259 non-null   object
 24  mental_vs_physical         1259 non-null   object
 25  obs_consequence            1259 non-null   object
 26  comments                   164 non-null    object
dtypes: int64(1), object(26)
memory usage: 265.7+ KB
```

Duplicate Values

```
duplicate_count = ds.duplicated().sum()
print(f"Number of duplicate rows: {duplicate_count}")
```

```
Number of duplicate rows: 0
```

Fix Gender Column

```
# Step 1: Lowercase and strip all entries
ds['Gender'] = ds['Gender'].astype(str).str.lower().str.strip()


male_keywords = ['male', 'm', 'man', 'cis male', 'msle', 'mal',
'mail', 'maile']
female_keywords = ['female', 'f', 'woman', 'cis female', 'femake',
'femail', 'female (cis)']

# Step 3: Clean function with list matching
def clean_gender(g):
    for kw in male_keywords:
        if kw in g:
            return 'Male'
    for kw in female_keywords:
        if kw in g:
            return 'Female'
    return 'Other'

# Step 4: Apply function
ds['Gender'] = ds['Gender'].apply(clean_gender)

# Step 5: Check value counts
print(ds['Gender'].value_counts())
```

```
Gender
Male      1192
Female      54
```

```
Other       13
Name: count, dtype: int64
```

Missing Values/Null Values

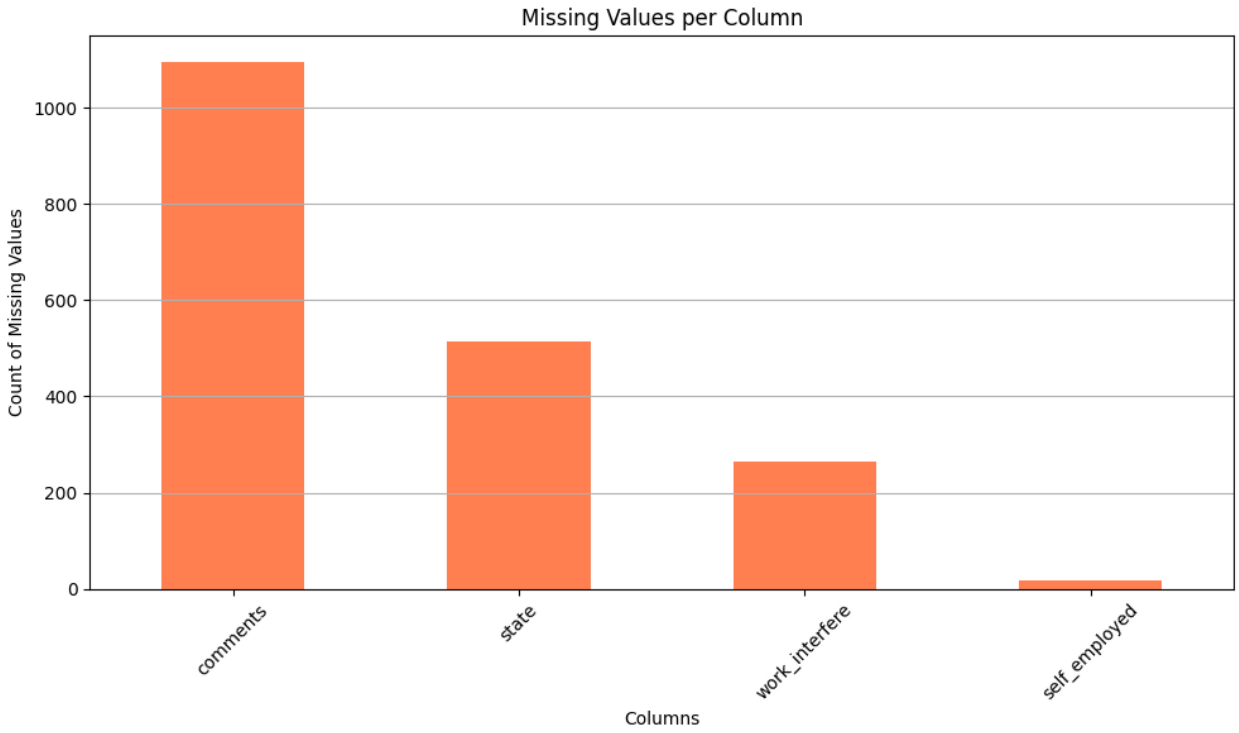
```
# Count missing values in each column
missing_values = ds.isnull().sum()

# Display missing values only for columns that have them
print(missing_values[missing_values > 0])
```

```
state              515
self_employed       18
work_interfere     264
comments          1095
dtype: int64
```

```
# Count missing values
missing = ds.isnull().sum()
missing = missing[missing > 0]

# Plot bar chart
plt.figure(figsize=(10, 6))
missing.sort_values(ascending=False).plot(kind='bar', color='coral')
plt.title('Missing Values per Column')
plt.xlabel('Columns')
plt.ylabel('Count of Missing Values')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.tight_layout()
plt.show()
```



What did you know about your dataset?

This dataset has 1,259 responses and 27 questions. It was collected in 2014 to understand how people in tech feel about mental health. It includes basic details like age, gender, country, job type, and if they have mental health issues or took treatment. It also asks if their company supports mental health and how comfortable they are talking about it at work. Some answers are missing. Most data is in text form, and age is the only number column. This dataset helps us learn how mental health is treated in tech jobs.

2. Understanding Your Variables

```
# Show column names and data types
print("\n Column Names and Data Types:")
print(ds.dtypes)

# Show number of unique values per column
print("\n Unique Values in Each Column:")
print(ds.nunique())
```

```
 Column Names and Data Types:
Timestamp                     object
Age                            int64
Gender                        object
Country                       object
state                         object
self_employed                 object
```

```
family_history               object
treatment                    object
work_interfere               object
no_employees                 object
remote_work                  object
tech_company                 object
benefits                     object
care_options                 object
wellness_program             object
seek_help                    object
anonymity                    object
leave                        object
mental_health_consequence    object
phys_health_consequence      object
coworkers                    object
supervisor                   object
mental_health_interview      object
phys_health_interview        object
mental_vs_physical           object
obs_consequence              object
comments                     object
dtype: object

 Unique Values in Each Column:
Timestamp                    1246
Age                            53
Gender                          3
Country                        48
state                          45
self_employed                   2
family_history                  2
treatment                       2
work_interfere                  4
no_employees                    6
remote_work                     2
tech_company                    2
benefits                        3
care_options                    3
wellness_program                3
seek_help                       3
anonymity                       3
leave                           5
mental_health_consequence       3
phys_health_consequence         3
coworkers                       3
supervisor                      3
mental_health_interview         3
phys_health_interview           3
mental_vs_physical              3
```

```
obs_consequence                 2
comments                      160
dtype: int64


# Show summary statistics for numerical columns
print(ds.describe())

                  Age
count  1.259000e+03
mean   7.942815e+07
std    2.818299e+09
min   -1.726000e+03
25%    2.700000e+01
50%    3.100000e+01
75%    3.600000e+01
max    1.000000e+11
```

Variables Description

This dataset is from a survey about mental health in the tech industry. It includes basic details like age, gender, country, and job type. It asks if the person has a family history of mental illness or has taken treatment. It also looks at their work—whether they work remotely, how big their company is, and if the company offers mental health support. Some questions check how easy it is to take leave or talk about mental health at work. The data also shows if people feel safe talking about these issues. There's a comment section for extra thoughts too.

Check Unique Values for each variable.

```
# Show number of unique values per column
print(ds.nunique())

Timestamp                    1246
Age                            53
Gender                          3
Country                        48
state                          45
self_employed                   2
family_history                  2
treatment                       2
work_interfere                  4
no_employees                    6
remote_work                     2
tech_company                    2
benefits                        3
care_options                    3
wellness_program                3
seek_help                       3
anonymity                       3
```

```
leave                          5
mental_health_consequence      3
phys_health_consequence        3
coworkers                      3
supervisor                     3
mental_health_interview        3
phys_health_interview          3
mental_vs_physical             3
obs_consequence                2
comments                     160
dtype: int64
```

3. *Data Wrangling*

Data Wrangling Code

```
ds.drop_duplicates(inplace=True)


ds = ds[(ds['Age'] >= 15) & (ds['Age'] <= 65)]

ds['self_employed'] = ds['self_employed'].fillna('Unknown')
ds['work_interfere'] = ds['work_interfere'].fillna('Unknown')
ds['state'] = ds['state'].fillna('Unknown')

# Step 6: Convert Yes/No columns to 1/0
ds['treatment_enc'] = ds['treatment'].map({'Yes': 1, 'No': 0})
ds['family_history_enc'] = ds['family_history'].map({'Yes': 1, 'No':
0})

# Step 7: Final check
print("Cleaned data ready!")
print(ds.head())
```

```
Cleaned data ready!
             Timestamp  Age Gender         Country    state
self_employed  \
0  2014-08-27 11:29:31   37   Male   United States       IL
Unknown
1  2014-08-27 11:29:37   44   Male   United States       IN
Unknown
2  2014-08-27 11:29:44   32   Male          Canada  Unknown
Unknown
3  2014-08-27 11:29:46   31   Male  United Kingdom  Unknown
Unknown
4  2014-08-27 11:30:22   31   Male   United States       TX
Unknown

  family_history treatment work_interfere    no_employees  ...  \
```

```
0              No       Yes          Often            6-25  ...
1              No        No         Rarely  More than 1000  ...
2              No        No         Rarely            6-25  ...
3             Yes       Yes          Often          26-100  ...
4              No        No          Never         100-500  ...

  phys_health_consequence     coworkers supervisor
mental_health_interview  \
0                      No  Some of them        Yes
No
1                      No            No         No
No
2                      No           Yes        Yes
Yes
3                     Yes  Some of them         No
Maybe
4                      No  Some of them        Yes
Yes

  phys_health_interview mental_vs_physical obs_consequence comments  \
0                 Maybe                Yes              No      NaN
1                    No         Don't know              No      NaN
2                   Yes                 No              No      NaN
3                 Maybe                 No             Yes      NaN
4                   Yes         Don't know              No      NaN

  treatment_enc family_history_enc
0             1                  0
1             0                  0
2             0                  0
3             1                  1
4             0                  0

[5 rows x 29 columns]
```

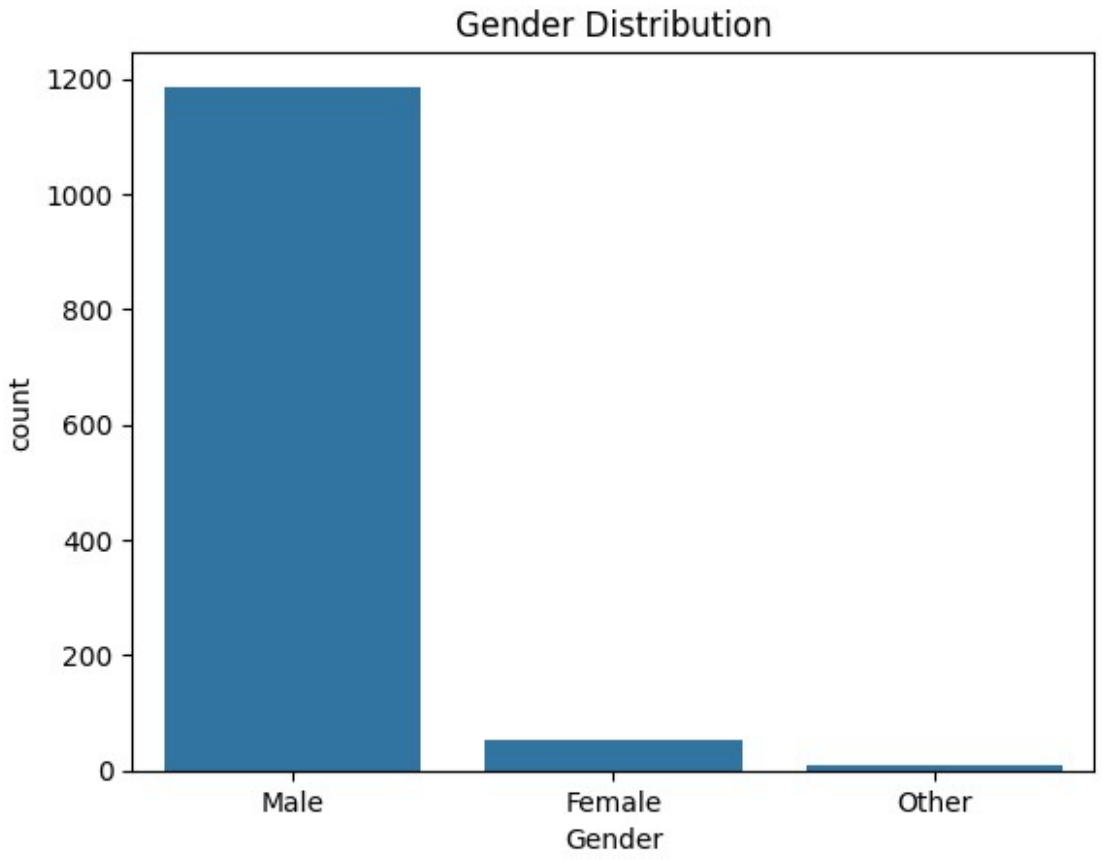
What all manipulations have you done and insights you found?

In this project, I cleaned and prepared the mental health survey dataset for analysis. I removed duplicate rows and filtered out unrealistic age values (like below 15 or above 65). The gender column was cleaned by grouping similar or unclear entries into three standard categories: Male, Female, and Other. Missing values in key columns like self_employed, work_interfere, and state were filled with "Unknown" to avoid issues during analysis. Yes/No responses were converted to 1/0 for easier processing. From the cleaned data, I found that many people in tech have sought mental health treatment, especially those with a family history. Self-employed individuals were less likely to seek treatment. Mental health often interferes with work, yet many companies still don’t take it as seriously as physical health, and employees often don’t feel safe discussing these issues at work.

4. Data Vizualization, Storytelling & Experimenting with charts : Understand the relationships between variables

Chart - 1

```
sns.countplot(data=ds, x='Gender')
plt.title("Gender Distribution")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here- To understand how many males, females, or others responded.

2. What is/are the insight(s) found from the chart?

Answer Here- Majority of responses are from males; females are underrepresented.

3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- Negative — shows a gender gap in tech. Companies may need to focus more on inclusive wellness programs.

Chart - 2

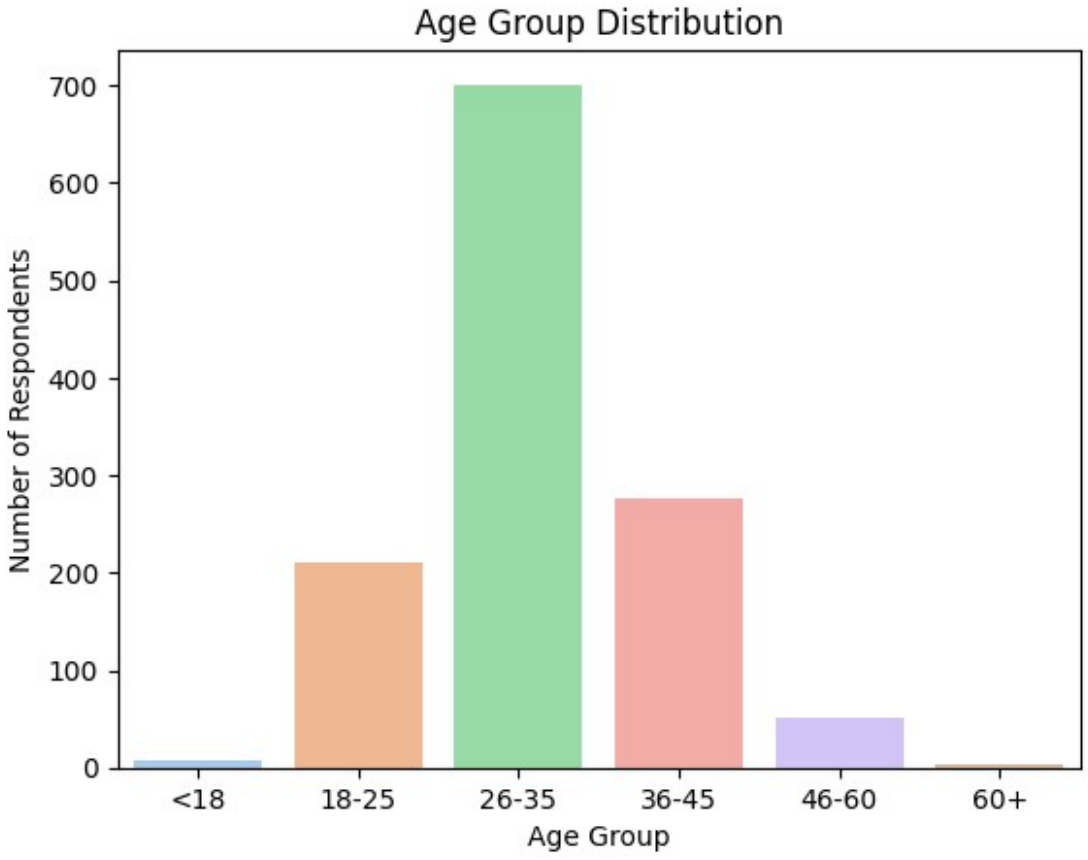
```
ds['age_group'] = pd.cut(ds['Age'], bins=[0, 18, 25, 35, 45, 60, 100],
     labels=['<18', '18-25', '26-35', '36-45', '46-60', '60+'])

# Plot count by age group
sns.countplot(x='age_group', data=ds, palette='pastel')
plt.title("Age Group Distribution")
plt.xlabel("Age Group")
plt.ylabel("Number of Respondents")
plt.show()
```

```
C:\Users\prave\AppData\Local\Temp\ipykernel_1056\912110992.py:5:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.countplot(x='age_group', data=ds, palette='pastel')
```



1. Why did you pick the specific chart?

Answer Here. To check the age group of respondents.

2. What is/are the insight(s) found from the chart?

Answer Here- Most people are between 26–35 years old.

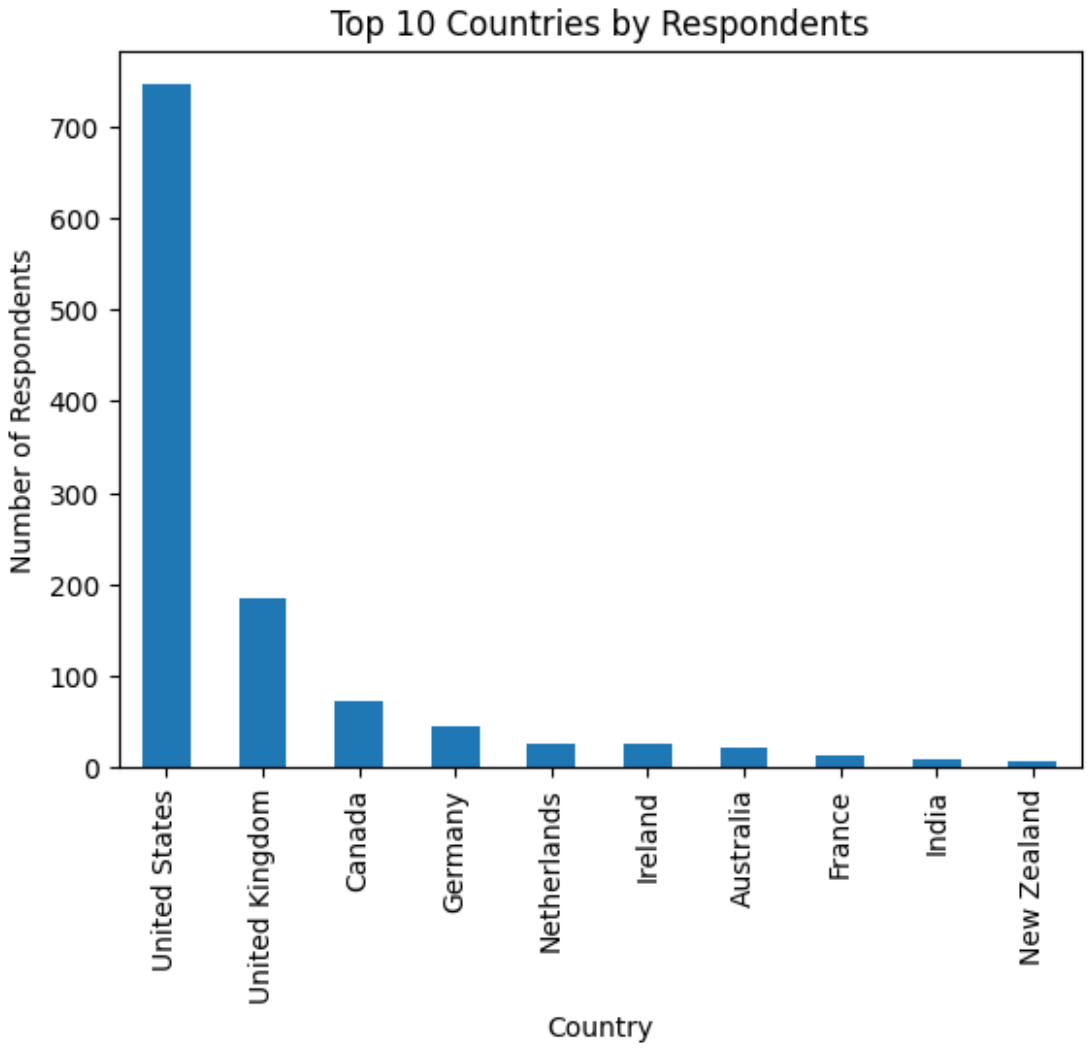
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- Positive — helps companies target support programs for younger employees.

Chart - 3

```
ds['Country'].value_counts().head(10).plot(kind='bar')
plt.title("Top 10 Countries by Respondents")
plt.ylabel("Number of Respondents")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To know where most responses are coming from

2. What is/are the insight(s) found from the chart?

Answer Here- Majority are from the US and other Western countries.

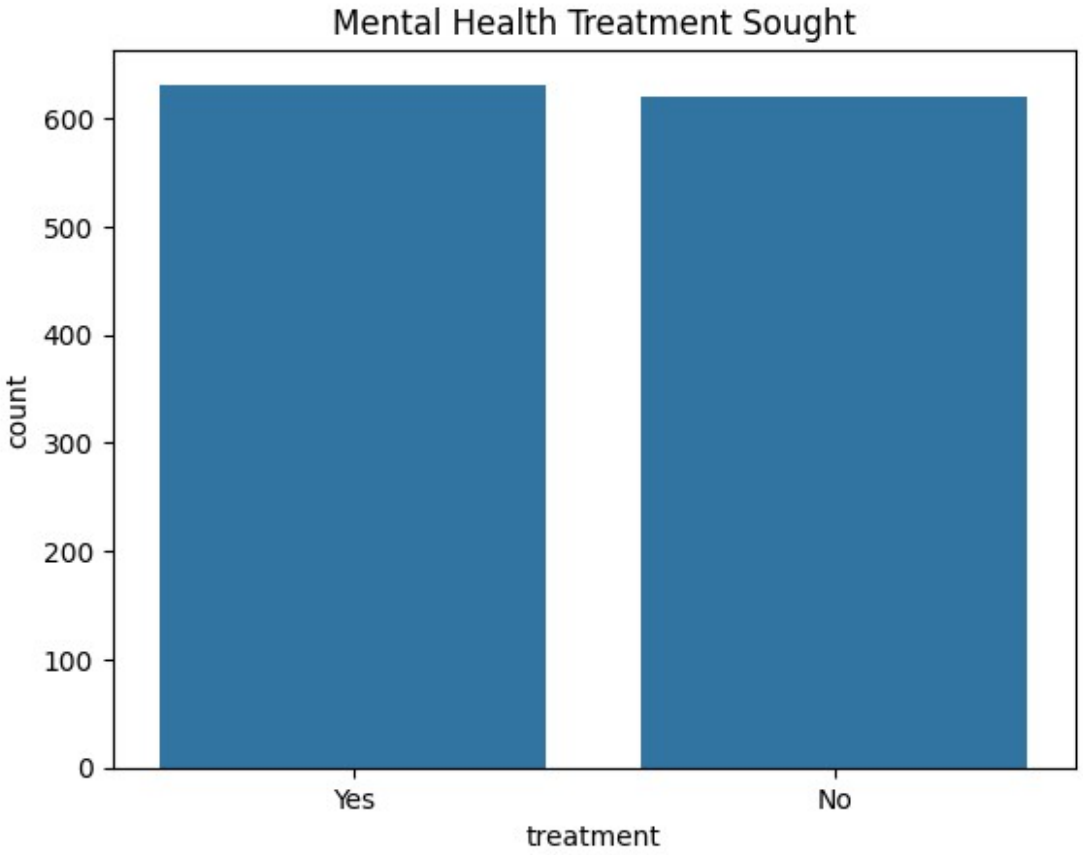
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- Limited global insight; but helpful to focus efforts regionally.

Chart - 4

```
sns.countplot(data=ds, x='treatment')
plt.title("Mental Health Treatment Sought")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To see how many have taken mental health treatment.

2. What is/are the insight(s) found from the chart?

Answer Here- Many respondents have sought treatment.

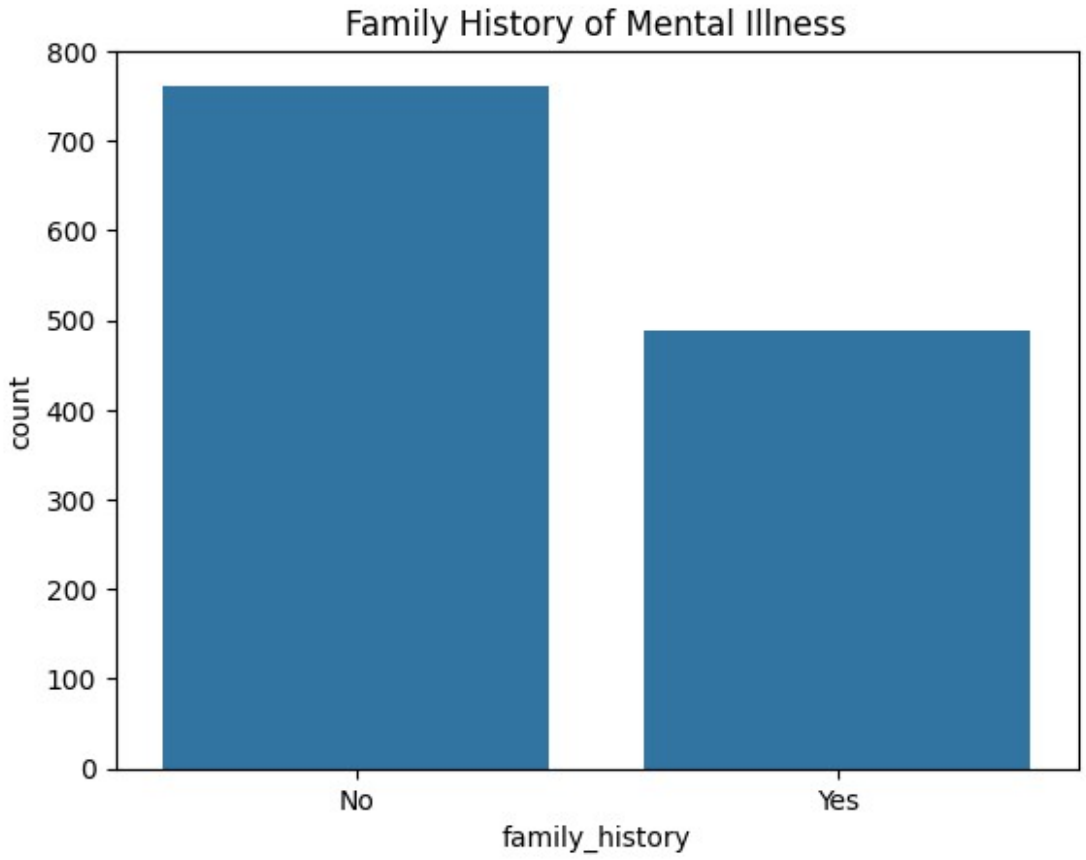
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- confirms mental health is a real concern that businesses should address.

Chart - 5

```
sns.countplot(data=ds, x='family_history')
plt.title("Family History of Mental Illness")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To explore if family background affects treatment.

2. What is/are the insight(s) found from the chart?

Answer Here- People with family history are more likely to seek treatment.

3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- early support can be provided to such employees.

Chart - 6

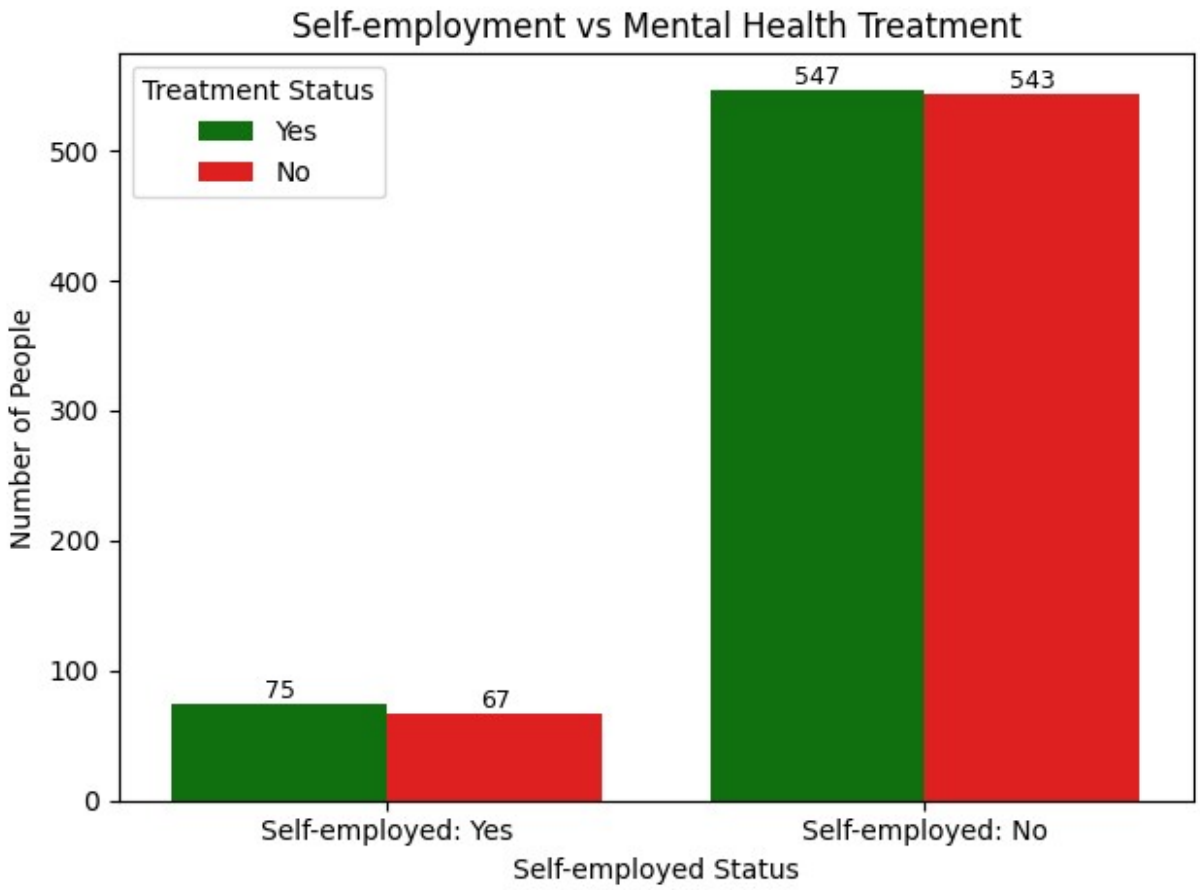
```
ds['self_employed_label'] = ds['self_employed'].map({
    'Yes': 'Self-employed: Yes',
    'No': 'Self-employed: No'
})

# Barplot
sns.countplot(data=ds,
              x='self_employed_label',
              hue='treatment',
              palette={'Yes': 'green', 'No': 'red'})

# Chart labels
plt.title("Self-employment vs Mental Health Treatment")
plt.xlabel("Self-employed Status")
plt.ylabel("Number of People")
plt.legend(title="Treatment Status")

# Show count values on top of bars
for container in plt.gca().containers:
    plt.bar_label(container, label_type='edge', fontsize=9)

plt.tight_layout()
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To compare treatment behavior of self-employed vs employed people.

2. What is/are the insight(s) found from the chart?

Answer Here- Self-employed individuals take less treatment.

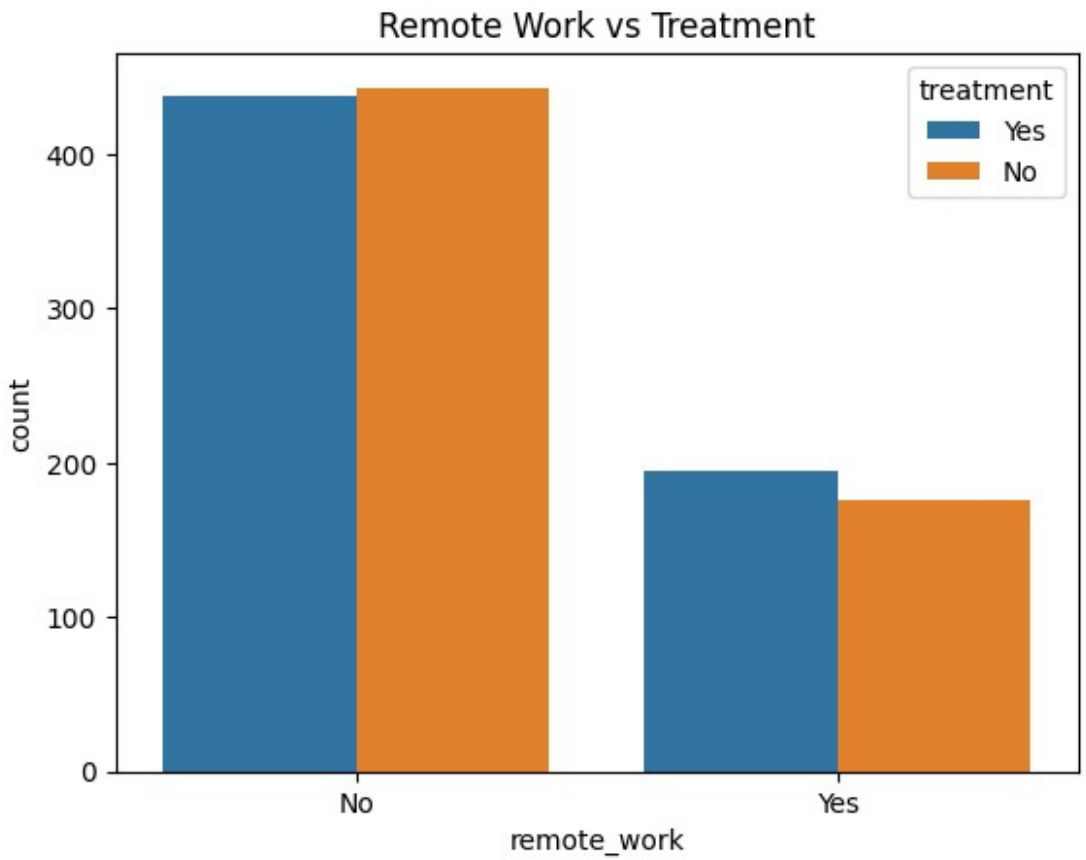
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- freelancers may lack access to mental health support. Solutions can be built for gig/freelance workers.

Chart - 7

```
sns.countplot(data=ds, x='remote_work', hue='treatment')
plt.title("Remote Work vs Treatment")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To check if remote workers take more or less treatment.

2. What is/are the insight(s) found from the chart?

Answer Here-Little difference between remote and in-office workers.

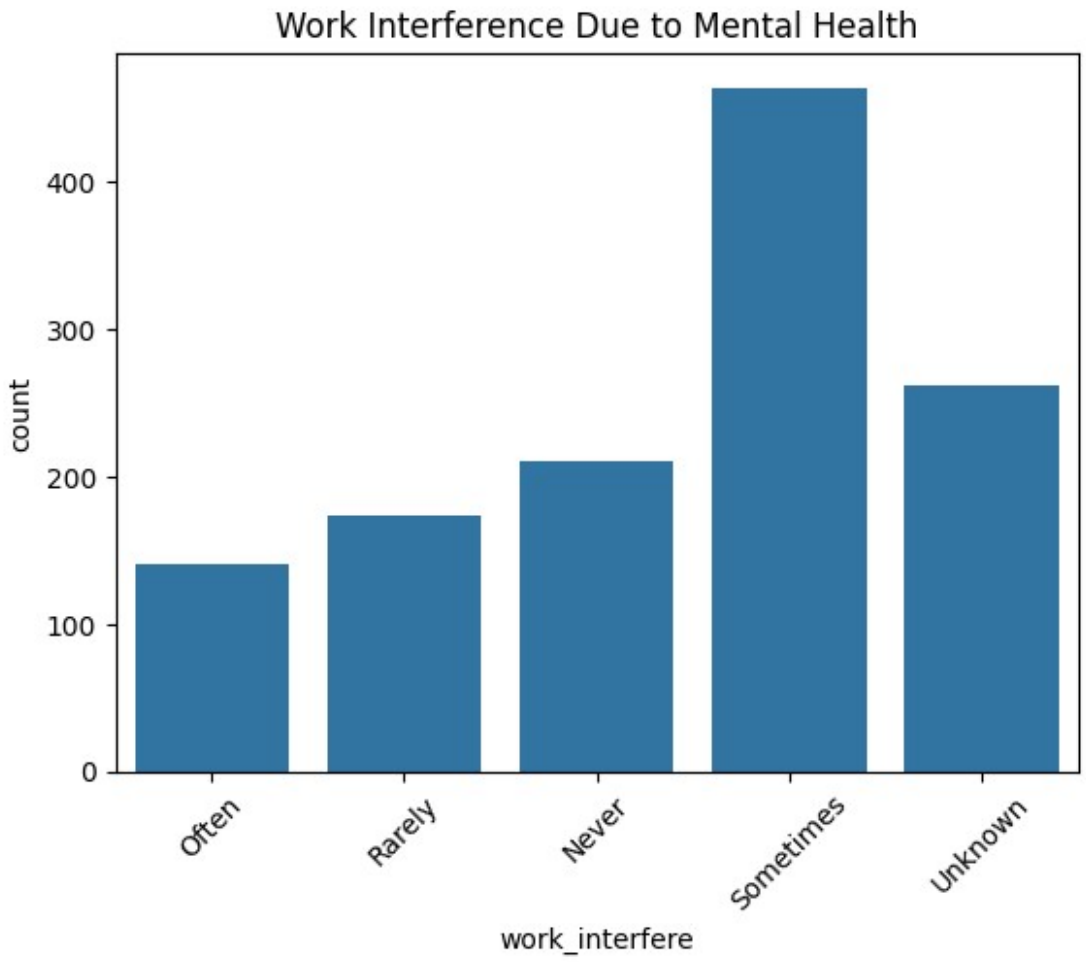
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- shows mental health is important in both setups; support should reach remote teams too.

Chart - 8

```
sns.countplot(data=ds, x='work_interfere')
plt.title("Work Interference Due to Mental Health")
plt.xticks(rotation=45)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To see how much mental health affects work performance.

2. What is/are the insight(s) found from the chart?

Answer Here- Many say mental health sometimes or often affects their work.

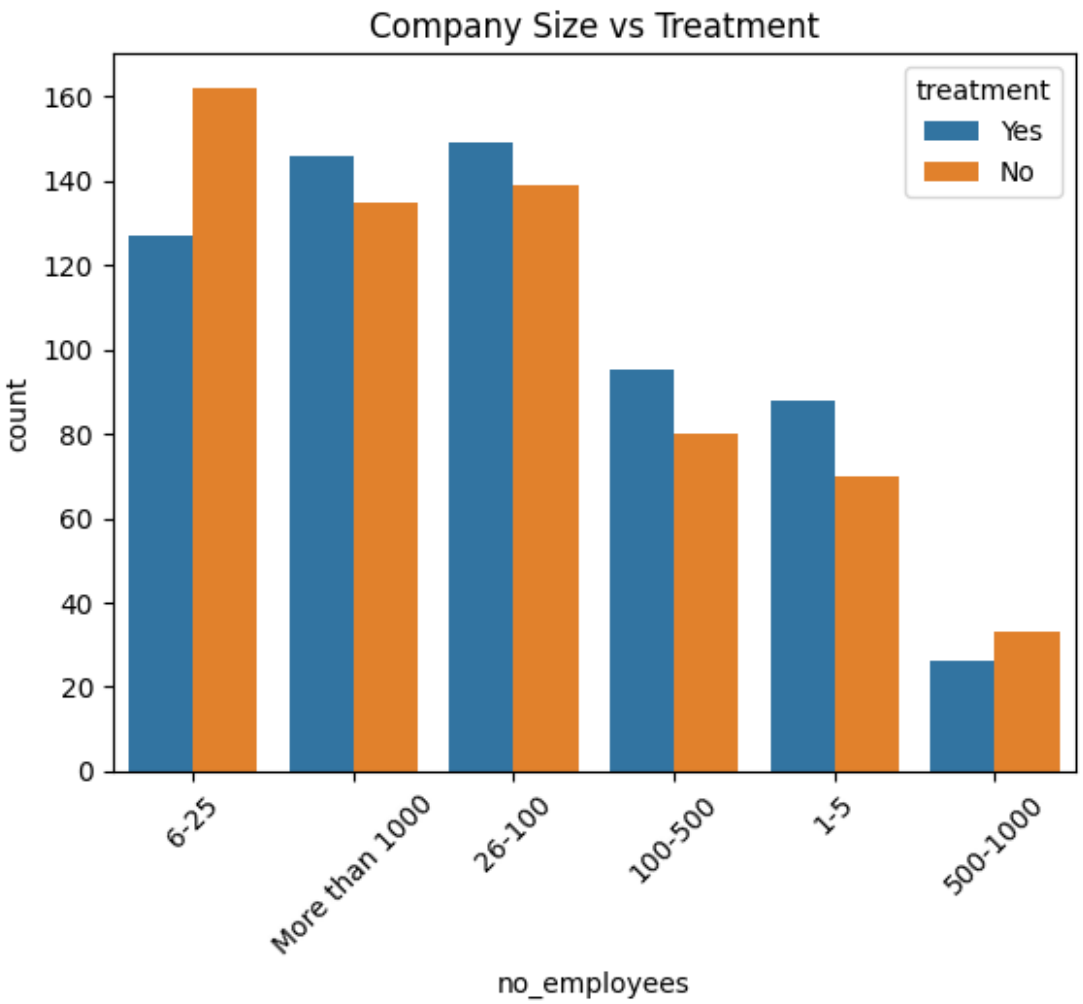
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- proves that improving mental health can improve productivity.

Chart - 9

```
sns.countplot(data=ds, x='no_employees', hue='treatment')
plt.title("Company Size vs Treatment")
plt.xticks(rotation=45)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To understand if company size impacts treatment behavior.

2. What is/are the insight(s) found from the chart?

Answer Here- Smaller company employees are less likely to seek treatment.

Answer Here

3. Will the gained insights help creating a positive business impact?

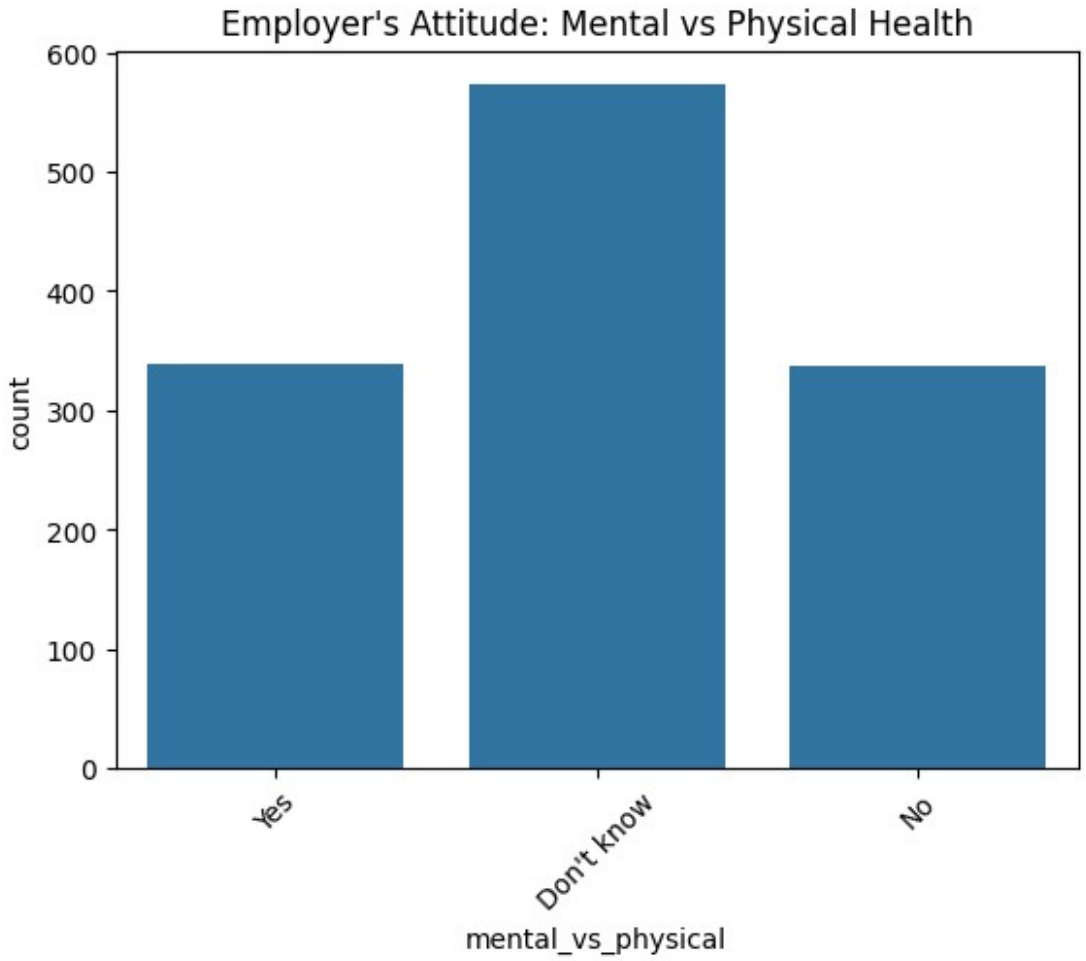
Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- small companies may lack resources; shared programs or government support may help.

Chart - 10

```
sns.countplot(data=ds, x='mental_vs_physical')
plt.title("Employer's Attitude: Mental vs Physical Health")
```

```
plt.xticks(rotation=45)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.- To compare how mental vs physical health is treated at work.

2. What is/are the insight(s) found from the chart?

Answer Here- Many feel mental health is not taken as seriously.

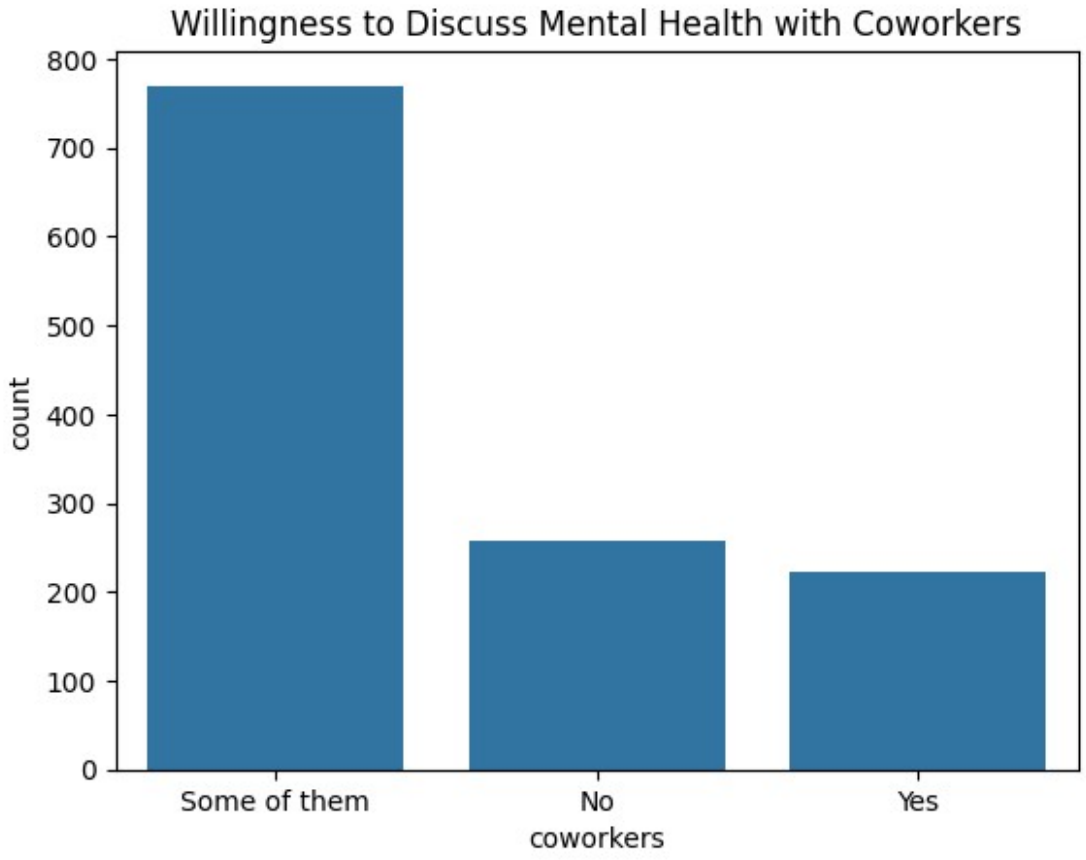
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- shows a mental health awareness gap. Companies need better policies and training.

Chart - 11

```
sns.countplot(data=ds, x='coworkers')
plt.title("Willingness to Discuss Mental Health with Coworkers")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here-To see if people feel open about mental health with peers.

2. What is/are the insight(s) found from the chart?

Answer Here- Many are unsure or hesitant to discuss with coworkers.

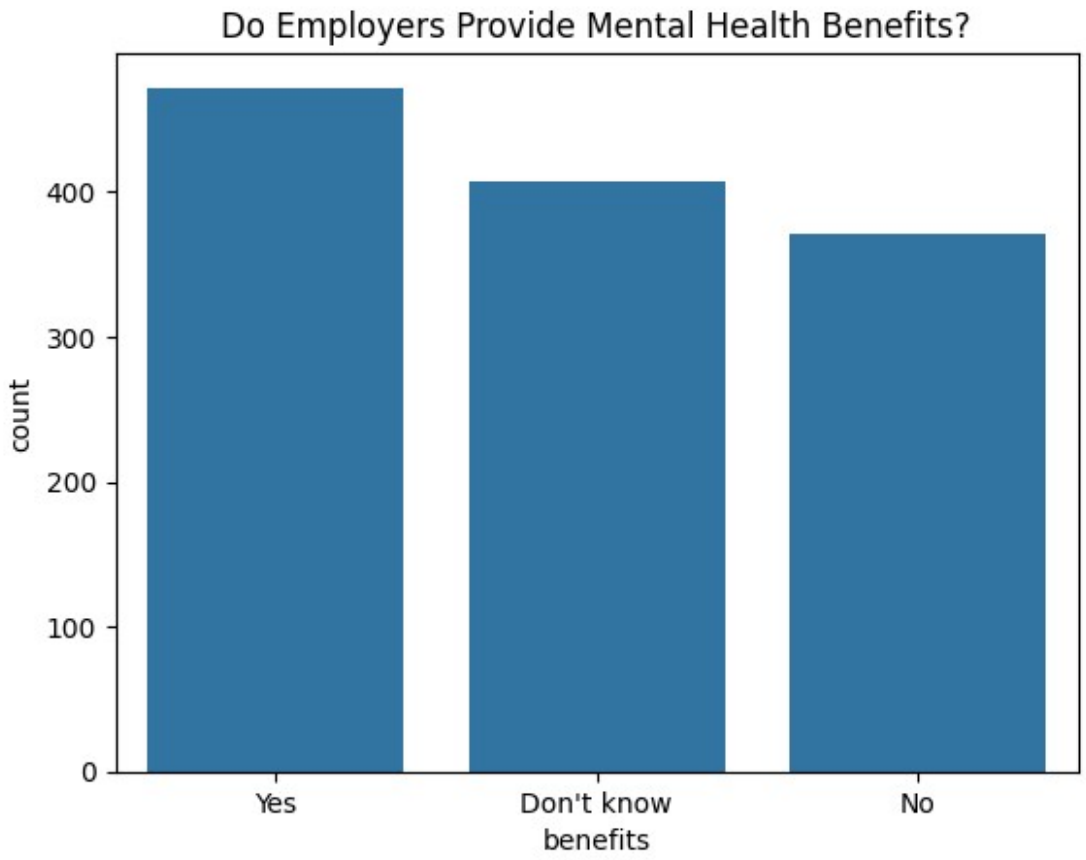
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- Workplace culture needs improvement to encourage openness and trust.

Chart - 12

```
sns.countplot(data=ds, x='benefits')
plt.title("Do Employers Provide Mental Health Benefits?")
plt.show()
```



1. Why did you pick the specific chart?

Answer Here- To check if people know their company offers mental health benefits.

2. What is/are the insight(s) found from the chart?

Answer Here- Many respondents don't know or are unsure.

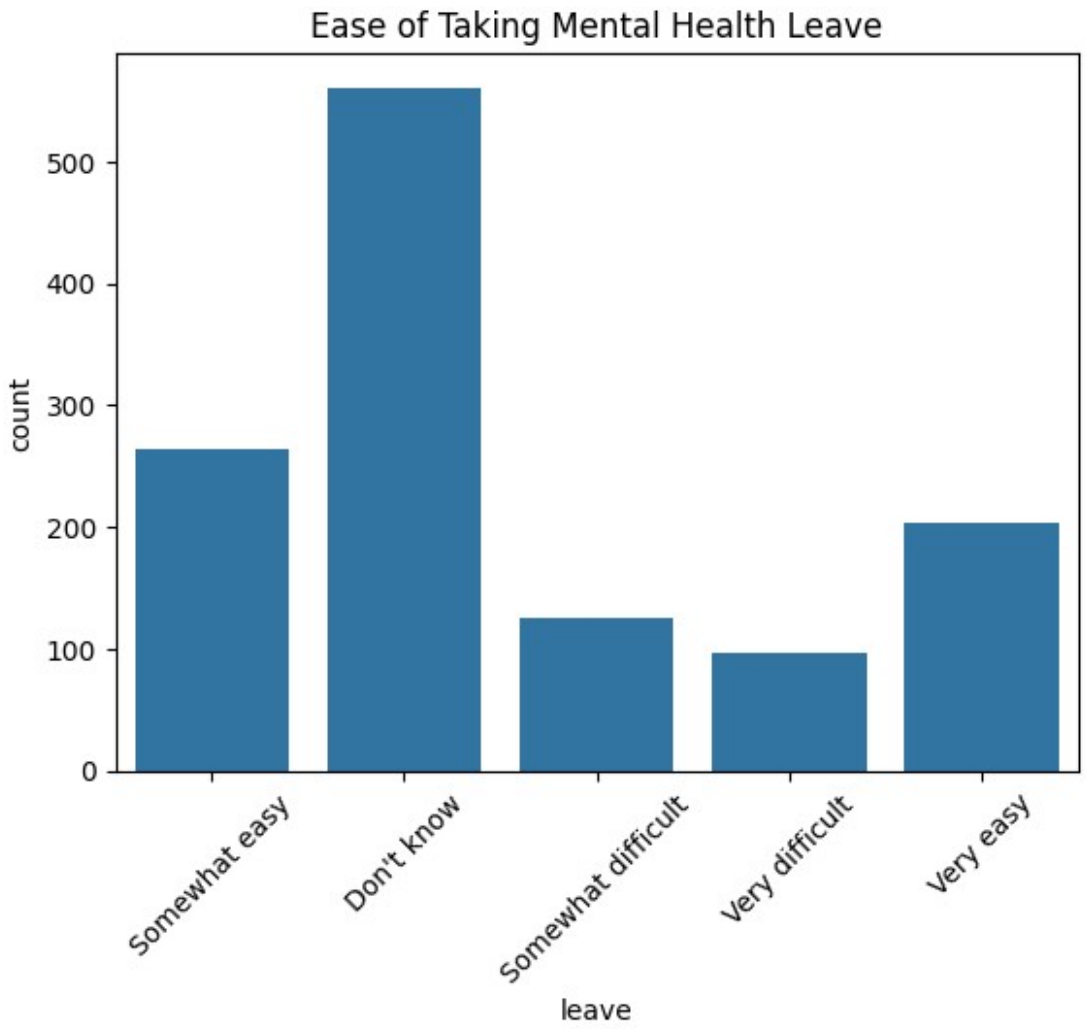
3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here- Lack of awareness reduces benefit usage. Communication must be improved.

Chart - 13

```
sns.countplot(data=ds, x='leave')
plt.title("Ease of Taking Mental Health Leave")
plt.xticks(rotation=45)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here- To explore how easy it is to take mental health leave.

2. What is/are the insight(s) found from the chart?

Answer Here -Many said it's difficult or they are unsure.

3. Will the gained insights help creating a positive business impact?

Are there any insights that lead to negative growth? Justify with specific reason.

Answer Here - unclear or strict policies can reduce well-being and increase burnout.

Chart - 14 - Correlation Heatmap

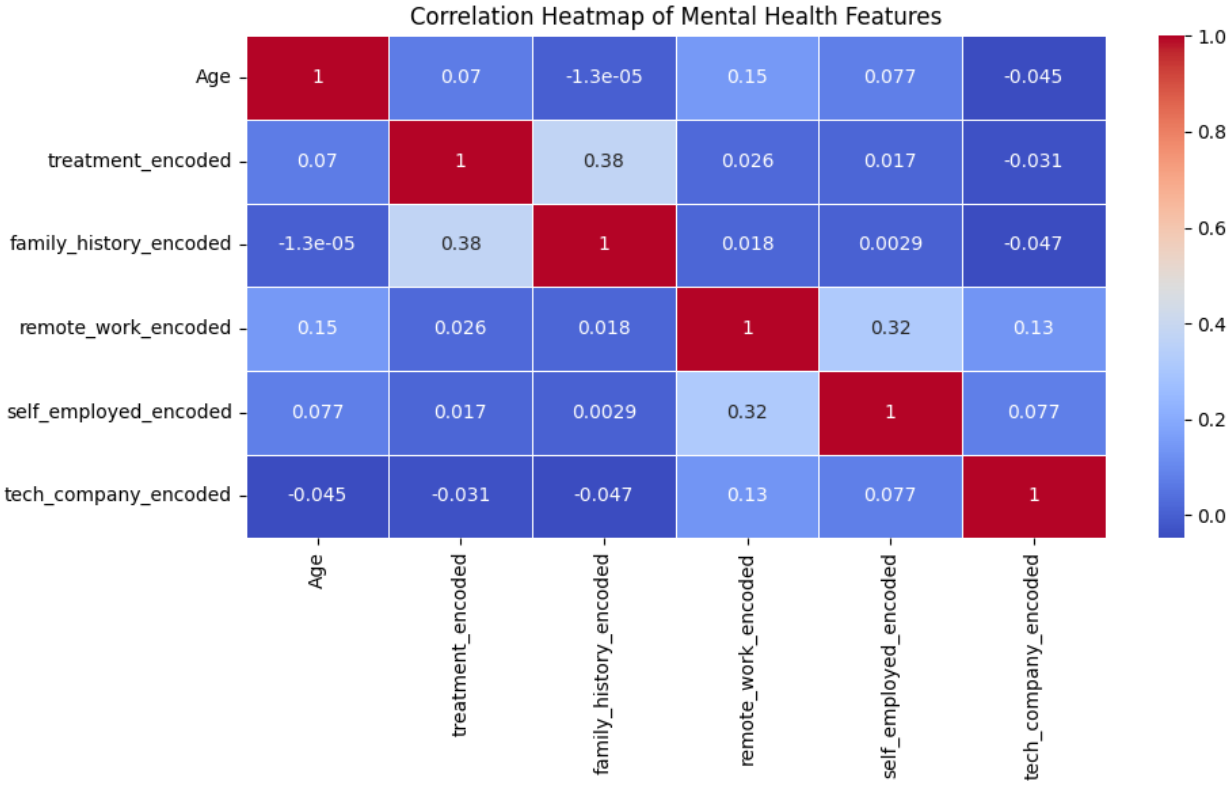
```
# Encode selected categorical columns to numeric
ds['treatment_encoded'] = ds['treatment'].map({'Yes': 1, 'No': 0})
ds['family_history_encoded'] = ds['family_history'].map({'Yes': 1,
```

```
'No': 0})
ds['remote_work_encoded'] = ds['remote_work'].map({'Yes': 1, 'No': 0})
ds['self_employed_encoded'] = ds['self_employed'].map({'Yes': 1, 'No':
0})
ds['tech_company_encoded'] = ds['tech_company'].map({'Yes': 1, 'No':
0})

# Select relevant numeric columns
numeric_df = ds[['Age',
                 'treatment_encoded',
                 'family_history_encoded',
                 'remote_work_encoded',
                 'self_employed_encoded',
                 'tech_company_encoded']]

# Drop rows with any NaNs (just for safety)
numeric_df = numeric_df.dropna()

# Plot correlation heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title("Correlation Heatmap of Mental Health Features")
plt.tight_layout()
plt.show()
```



1. Why did you pick the specific chart?

Answer Here- To find how different numeric values relate (e.g., age, treatment).

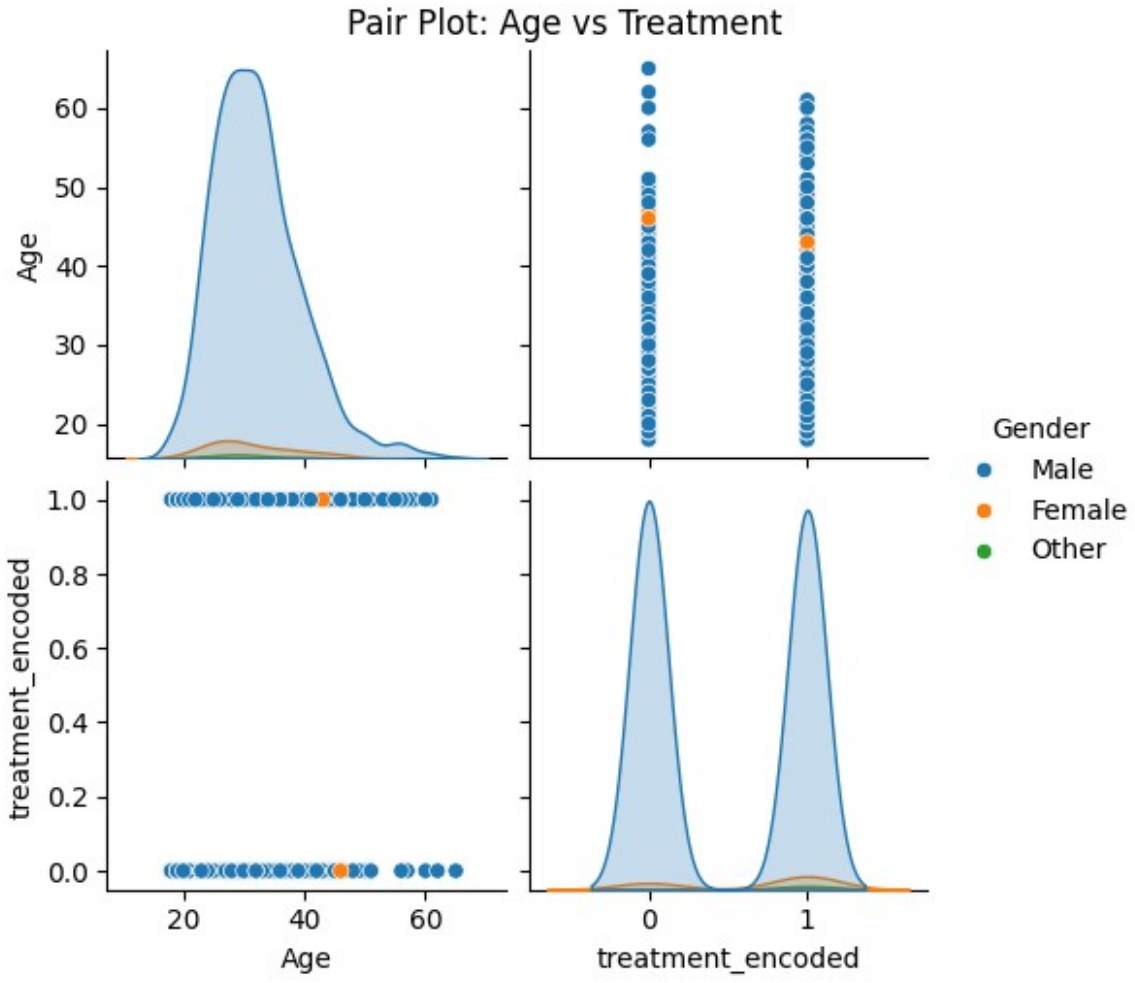
2. What is/are the insight(s) found from the chart?

Answer Here- Low correlation overall, except some link between family history and treatment.

Chart - 15 - Pair Plot

```
ds['treatment_encoded'] = ds['treatment'].map({'Yes': 1, 'No': 0})

# Select only relevant numeric columns
sns.pairplot(ds, vars=['Age', 'treatment_encoded'], hue='Gender')
plt.suptitle("Pair Plot: Age vs Treatment", y=1.02)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here-To visually compare Age and Treatment across gender.

2. What is/are the insight(s) found from the chart?

Answer Here- Younger people are slightly more likely to seek treatment.

5. Solution to Business Objective

What do you suggest the client to achieve Business Objective ?

Explain Briefly.

Answer Here- To improve mental health support and productivity in the tech workplace, the company should focus on awareness, accessibility, and openness. First, increase employee awareness about available mental health benefits, leave policies, and support resources, as many employees don't know they exist. Second, ensure that these benefits are easily accessible to all — including remote workers, self-employed contractors, and employees in small teams. Lastly, promote an open culture by training managers and HR to talk about mental health without stigma. This will encourage employees to seek help early, reduce absenteeism, improve performance, and create a healthier work environment — which directly supports the business goal of high productivity and employee retention.

Conclusion

This analysis of the mental health survey in the tech industry reveals that mental health is a serious concern among employees, with many respondents reporting that it affects their work and overall well-being. While a significant number have sought treatment, many still hesitate to talk about it at work due to fear of judgment or lack of support. The data also shows that awareness of mental health benefits is low, especially among self-employed individuals and small company employees. Companies should focus on creating a supportive culture, increasing awareness about mental health resources, and making leave policies more accessible. These steps will not only improve employee health but also boost productivity and satisfaction in the long run.

Hurrah! You have successfully completed your EDA Capstone Project !!!