

Project Name - Mental Health in Tech Survey EDA Project

Project Type - EDA

Contribution - Individual

Project Summary -

- Mental health is a crucial aspect of our overall well-being, influencing how we think, feel, and interact with the world around us. It encompasses our emotional resilience, our ability to handle stress, and our capacity to form and maintain healthy relationships. Good mental health is vital for leading a fulfilling and productive life.
- Work constitutes a significant portion of our lives, where we invest a substantial amount of our time to earn income and cultivate friendships. Having a satisfying job can positively impact our mental well-being and overall quality of life.
- Understanding the importance of mental health is essential because it affects every facet of our daily existence. Here are some key points to consider:
 - **Well-Being:** Good mental health contributes to a sense of well-being, allowing individuals to experience life's joys and challenges more effectively.
 - **Productivity:** Mental health plays a significant role in our productivity, both at work and in our personal lives. It influences our motivation, focus, and problem-solving abilities.
 - **Relationships:** Healthy mental well-being fosters positive relationships. It enables us to communicate effectively, empathize with others, and build strong connections.
 - **Resilience:** Good mental health equips individuals with the tools to cope with life's ups and downs. It helps us bounce back from adversity and adapt to change.
 - **Physical Health:** Mental health and physical health are interconnected. Neglecting one can affect the other, leading to various health issues.
- Addressing mental health issues requires a multi-faceted approach, including raising awareness, reducing stigma, providing access to mental health services, and promoting self-care and emotional well-being. It's essential for individuals to prioritize

- their mental health and seek help when needed, just as they would for physical health concerns.
- Recognizing its importance and addressing mental health issues at home as well as at work place are crucial steps toward a healthier and more resilient society.
- This project involves a comprehensive Exploratory Data Analysis (EDA) of the Mental Health in Tech Survey dataset.
- This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace.

The analysis includes:

- Cleaning and preprocessing (handling missing values, correcting outliers, and data type conversion)
- Univariate, Bivariate, and Multivariate visualizations (UBM) to identify patterns
- Analysis of mental health treatment behavior across different genders, age groups, employment types, and work environments
- Key focus on how work interference, remote work, and company support influence treatment-seeking behavior

By identifying high-risk groups and support gaps, this project helps build actionable recommendations for organizations to create a more inclusive and responsive mental health environment.

GitHub Link -

https://github.com/madhuripingle/Mental_Health_Tech_Survey_EDA

Problem Statement

- The objective of this analysis is to investigate the mental health survey data, aiming to gain deeper insights into the mental well-being of tech industry workers.
- To analyze the Mental Health in Tech dataset and extract valuable insights regarding contributing to mental health issues and to support strategic decision-making in tackling mental health.

Define Your Business Objective?

- In today's high-stress and fast-paced tech industry, mental health challenges are increasingly prevalent, yet they remain under-discussed and under-supported in many workplaces. Many employees struggle silently, unaware of or unable to access proper mental health care and support systems.
- This project aims to answer the following core questions:
 - How does the frequency of mental health illness and attitudes towards mental health vary by geographic location?
 - What are the strongest predictors of mental health illness or certain attitudes towards mental health in the workplace?
- Perform exploratory data analysis (EDA), including statistical summaries, visualization of relationships between variables, and prevalence rates of at least three mental health conditions.

Let's Begin !

1. Know Your Data

Import Libraries

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
In [3]: import warnings
warnings.filterwarnings("ignore")
```

Dataset Loading

```
In [53]: # Load Dataset  
mental_health_data = pd.read_csv('survey.csv')
```

Dataset First View

```
In [54]: mental_health_data.head()
```

```
Out[54]:
```

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...
2	2014-08-27 11:29:44	32	Male	Canada	Nan	NaN	No	No	Rarely	6-25	...
3	2014-08-27 11:29:46	31	Male	United Kingdom	Nan	NaN	Yes	Yes	Often	26-100	...
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...

5 rows × 27 columns



Dataset Rows & Columns count

```
In [55]: print("rows: ", mental_health_data.shape[0])  
print("column: ", mental_health_data.shape[1])
```

```
rows: 1259  
column: 27
```

Dataset Information

```
In [56]: # Dataset Info
```

```
mental_health_data.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

```
In [57]: mental_health_data.duplicated().sum()
```

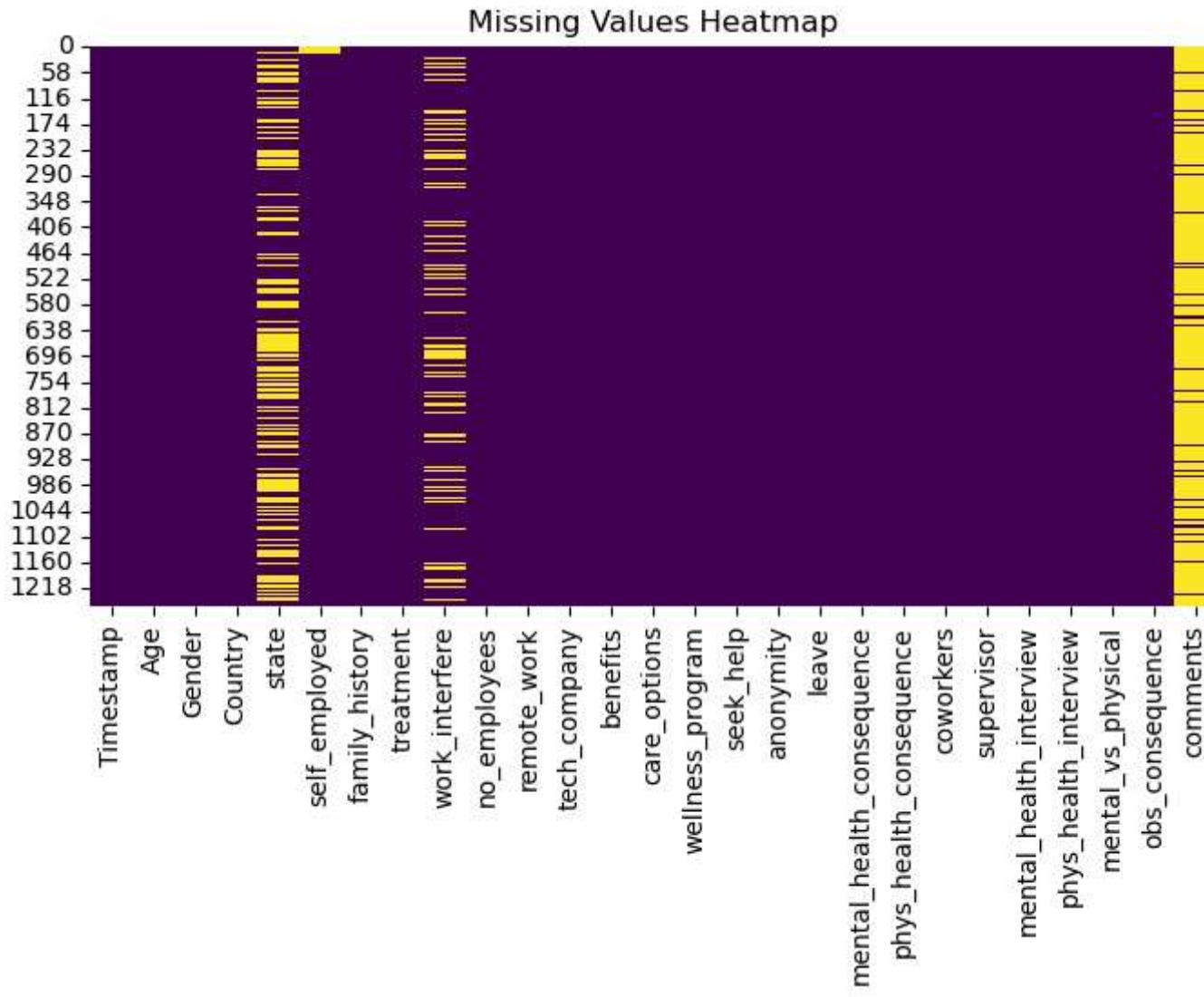
```
Out[57]: 0
```

Missing/Null values

```
In [58]: mental_health_data.isnull().sum()
```

```
Out[58]: 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
```

```
In [59]: # Visualizing the missing values
plt.figure(figsize=(8, 4))
sns.heatmap(mental_health_data.isnull(), cbar=False, cmap='viridis')
plt.title('Missing Values Heatmap')
plt.show()
```



What did you know about your dataset?

- There are 1259 rows and 27 columns in the dataset.
- Demographics (e.g., Age, Gender, Country, state)
- Work-related details (remote_work, no_employees, self_employed)

- Mental health experiences (treatment, work_interfere, seek_help, family_history, etc.)
- Out of which 4 Columns have missing Values. Column 'Comment' have most missing values of 1095 followed by 'State','work_interfere' and self_employed column.
- Out of all 'self_employed' Have least missing values of 18.

2. *Understanding Your Variables*

```
In [15]: # Dataset Columns  
mental_health_data.columns.tolist()
```

```
Out[15]: ['Timestamp',  
          'Age',  
          'Gender',  
          'Country',  
          'state',  
          'self_employed',  
          'family_history',  
          'treatment',  
          'work_interfere',  
          'no_employees',  
          'remote_work',  
          'tech_company',  
          'benefits',  
          'care_options',  
          'wellness_program',  
          'seek_help',  
          'anonymity',  
          'leave',  
          'mental_health_consequence',  
          'phys_health_consequence',  
          'coworkers',  
          'supervisor',  
          'mental_health_interview',  
          'phys_health_interview',  
          'mental_vs_physical',  
          'obs_consequence',  
          'comments']
```

```
In [16]: # Dataset describe
```

```
mental_health_data.describe(include='all')
```

Out[16]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees
count	1259	1.259000e+03	1259	1259	744	1241	1259	1259	1259	995
unique	1246	NaN	49	48	45	2	2	2	2	4
top	2014-08-27 12:44:51	NaN	Male	United States	CA	No	No	Yes	Sometimes	
freq	2	NaN	615	751	138	1095	767	637	465	
mean	NaN	7.942815e+07	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	NaN	2.818299e+09	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	NaN	-1.726000e+03	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	NaN	2.700000e+01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	NaN	3.100000e+01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	NaN	3.600000e+01	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	NaN	1.000000e+11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

11 rows × 27 columns



Variables Description

- This dataset contains the data regarding below questions asked in survey.
- **Timestamp** - Data survey was filled
- **Age** - Respondent's age (numeric).

- **Gender**- Gender identity of surveyed person
- **Country**- Country of residence
- **state**- US residents only
- **self_employed**- Are your self employeed?
- **family_history**- Do you have a family history of mental illnes?
- **treatment**- Have yuo saught treatment for a mental health condition?
- **work_interfere**- If you have a mental health condition, do you feel that it interferes with your work?
- **no_employees**- How many employees does your company or organization have?
- **remote_work**- Do you work remotely (outside of an office) at least 50% of the time?
- **tech_company**- Is your employer primarily a tech company/organization?
- **benefits**- Does your employer provide mental health benefits?
- **care_options**- Do you know the options for mental health care your employer provides?
- **wellness_program**- Has your employer ever discussed mental health as part of an employee wellness program?
- **seek_help**- Does your employer provide resources to learn more about mental health issues and how to seek help?
- **anonymity**- Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?
- **leave**- How easy is it for you to take medical leave for a mental health condition?
- **mental_health_consequence**- Do you think that discussing a mental health issue with your employer would have negative consequences?
- **phys_health_consequence**- Do you think that discussing a physical health issue with your employer would have negative consequences?

- **coworkers**- Would you be willing to discuss a mental health issue with your coworkers?
- **supervisor**- Would you be willing to discuss a mental health issue with your direct supervisor(s)?
- **mental_health_interview**- Would you bring up a mental health issue with a potential employer in an interview?
- **phys_health_interview**- Would you bring up a physical health issue with a potential employer in an interview?
- **mental_vs_physical**- Do you feel that your employer takes mental health as seriously as physical health?
- **obs_consequence**- Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?
- **comments**- Any additional notes or comments

Check Unique Values for each variable.

```
In [60]: mental_health_data.nunique()
```

```
Out[60]: Timestamp      1246  
Age             53  
Gender          49  
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

```
In [10]: # Drop dulicates  
mental_health_data.drop_duplicates(inplace=True)
```

```
In [61]: # Convert Timestamp to datetime datatype  
mental_health_data['Timestamp'] = pd.to_datetime(mental_health_data['Timestamp'])
```

```
In [62]: # Cleaning the Gender Column (As there are 46 unique genders in dataset)

def clean_gender(gender):
    gender = str(gender).strip().lower()

    if gender in ['male', 'm', 'male ', 'cis male', 'cis man', 'man', 'malr']:
        return 'Male'
    elif gender in ['female', 'f', 'femail', 'cis female', 'cis woman', 'woman']:
        return 'Female'
    else:
        return 'Other'

mental_health_data['Gender'] = mental_health_data['Gender'].apply(clean_gender)
```

```
In [63]: mental_health_data['Gender'].head(100)
```

```
Out[63]: 0      Female
1       Male
2       Male
3       Male
4       Male
...
95      Male
96      Male
97      Male
98      Male
99      Male
Name: Gender, Length: 100, dtype: object
```

```
In [64]: # Handling missing values
mental_health_data.drop(columns=['comments'], inplace=True)

mental_health_data.fillna({'state': 'Not specified'}, inplace=True)
mental_health_data.fillna({"work_interfere": mental_health_data["work_interfere"].mode()[0]}, inplace=True)
mental_health_data.fillna({'self_employed': 'No'}, inplace=True)
```

```
In [65]: # cross checking null values

mental_health_data.isnull().sum()
```

```
Out[65]: Timestamp      0  
Age            0  
Gender          0  
Country         0  
state           0  
self_employed   0  
family_history  0  
treatment       0  
work_interfere 0  
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  
dtype: int64
```

```
In [66]: mental_health_data.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1259 entries, 0 to 1258
Data columns (total 26 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Timestamp        1259 non-null    datetime64[ns]
 1   Age              1259 non-null    int64  
 2   Gender            1259 non-null    object  
 3   Country           1259 non-null    object  
 4   state             1259 non-null    object  
 5   self_employed     1259 non-null    object  
 6   family_history    1259 non-null    object  
 7   treatment          1259 non-null    object  
 8   work_interfere    1259 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  
dtypes: datetime64[ns](1), int64(1), object(24)
memory usage: 255.9+ KB

```

What all manipulations have you done and insights you found?

- 1.Null values are presented in state, self_employed, work_interfere and column. Filled NA with No in self_employed column, filled NA with mode ("Sometimes") in work_interfere and filled NA with 'Not specified' in state column.
- 2.Drop the comments column (86% missing).

- 3.Timestamp was object here.Converted it to date time format.
- 4.Converted categorical columns like type "Gender","Country","self_employed","treatment" to category data type for memory optimization and efficient analysis.
- 5.Cleaned messy Gender values (e.g., "male", "MALE ", "femail", "cis man", etc.)

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

Univariate Analysis -

Chart - 1

```
In [52]: # Chart - 1 Country Distribution

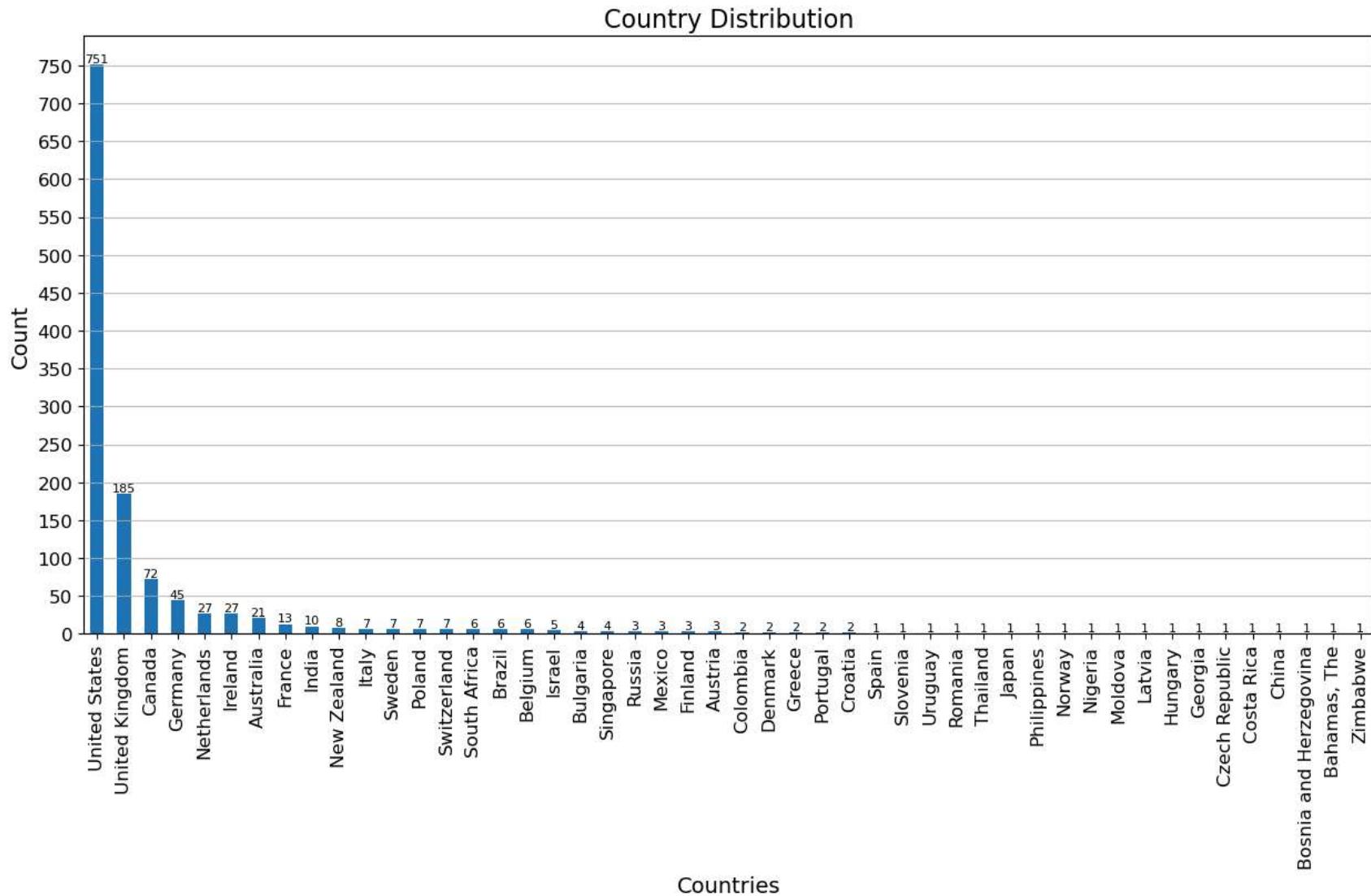
# initialize canvas of size 15 X 7
fig, ax = plt.subplots(figsize=(15, 7))

# creating a barplot of country
(mental_health_data['Country'].value_counts().plot.bar(ax=ax))

# adding some aesthetics
plt.title(label = 'Country Distribution', size = 16)
plt.xlabel(xlabel = 'Countries', size = 14)
plt.ylabel(ylabel = 'Count', size = 14)
plt.xticks(size = 12)
plt.yticks(np.arange(0, 751, 50), size = 12)
plt.grid(axis = 'y', alpha = 0.8)

# adding data Labels to enhance readability
for p in ax.patches:
    height = p.get_height()
    ax.annotate(f'{height:.0f}', (p.get_x() + p.get_width() / 2., height),
                ha='center', va='bottom', size=8)

# display the plot
plt.show()
```



1. Why did you pick the specific chart?

A vertical bar chart clearly visualizes the distribution of survey data at country level.

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

We observe that in the United States alone, the number of people who have sought treatment earlier is considerably higher than the number of people who have not.

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

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

Yes, this insight suggests that most of the responders are from US ,which means US is active country in mental health survey.On the other hand,Maximum population of responders from US and very little from other countries can not properly display the global mental in tech correctly.

Chart - 2

```
In [66]: # Chart - 2 Top 10 countries

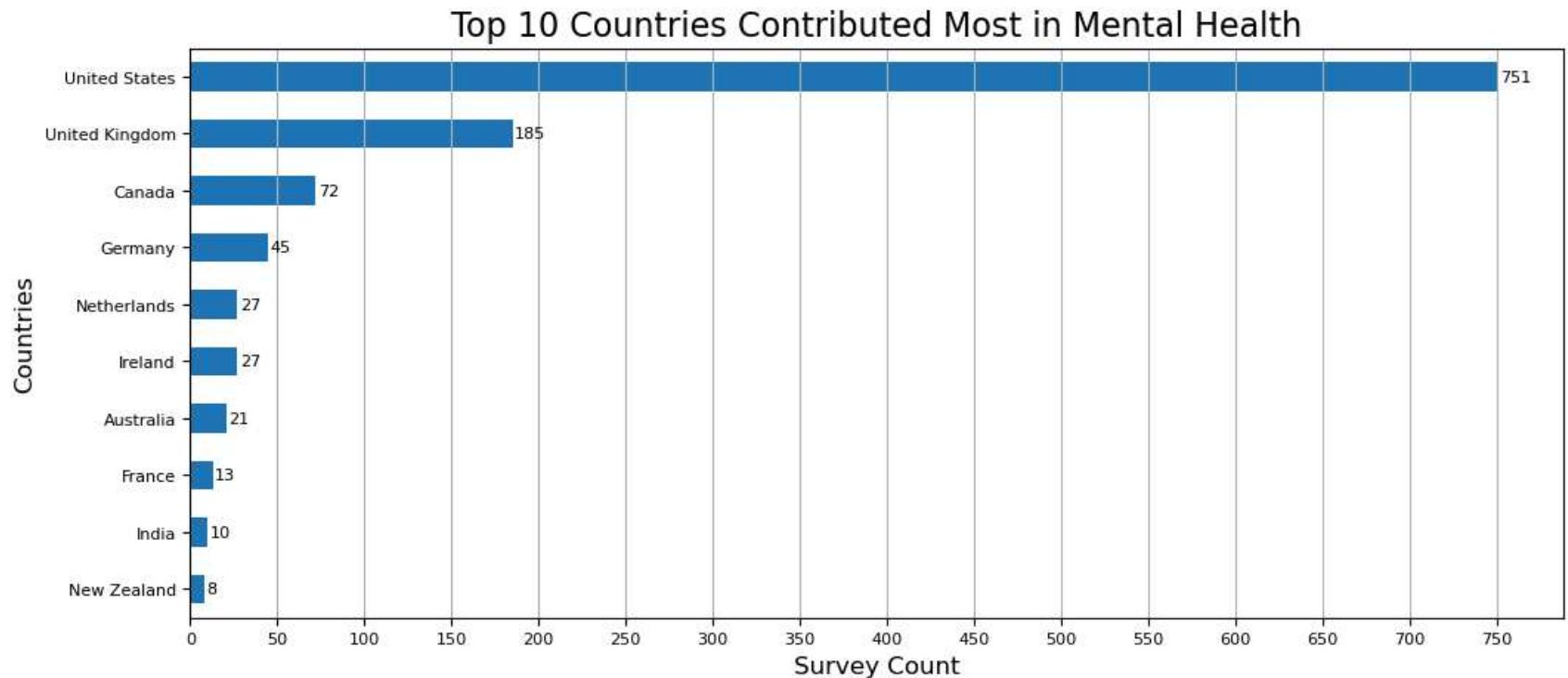
# initialize canvas of size 12 X 5
fig, ax = plt.subplots(figsize=(12, 5))

# creating a barplot of top 10 countries
mental_health_data['Country'].value_counts().sort_values(ascending = True).tail(10).plot.barh()

# adding some aesthetics
plt.yticks(size = 8)
plt.ylabel('Countries', size = 12)
plt.xticks(np.arange(0, 751, 50), size = 8)
plt.xlabel("Survey Count", size = 12)
plt.title("Top 10 Countries Contributed Most in Mental Health", size = 16)
plt.grid(axis = 'x')

# adding data labels on top of each bar
for p in ax.patches:
    width = p.get_width()
    ax.annotate(f'{width:.0f}',
                (width, p.get_y() + p.get_height() / 2),
                ha='left', va='center',
                xytext=(1, 0), # Shift the label to the right
                textcoords='offset points', size = 8)
```

```
# display plot  
plt.show()
```



1. Why did you pick the specific chart?

This horizontal bar chart was selected to clearly visualize the top 10 countries with effectively handling long category labels.

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

The data shows that respondents are overwhelmingly from the United States, with a sharp decline in numbers for other countries, most of which are Western and English-speaking.

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

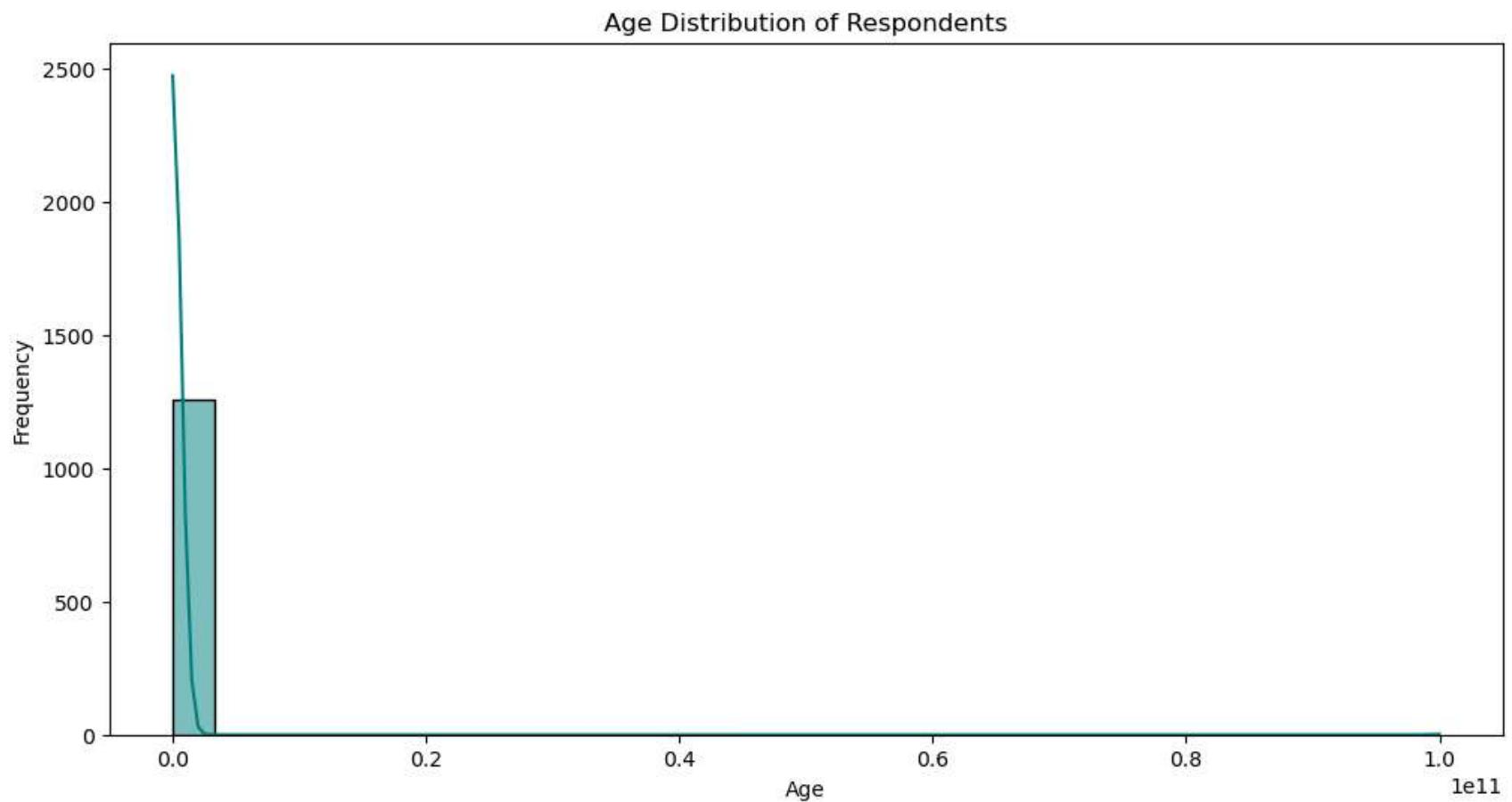
Are there any insights that lead to negative growth? Justify with specific reason.

Yes, this insight suggests that most of the responders are from US ,which means US is active country in mental health survey.On the other hand,Maximum population of responders from US and very little from other countries can not properly display the global

mental in tech correctly.

Chart - 3

```
In [55]: # Chart - 3 # Age Distribution
plt.figure(figsize=(12,6))
sns.histplot(mental_health_data['Age'], bins=30, kde=True, color='teal')
plt.title('Age Distribution of Respondents')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



1. Why did you pick the specific chart?

This histogram was chosen to effectively display the frequency distribution of the respondents' ages, which is a continuous data set.

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

The data reveals that the largest group of respondents are young adults, predominantly in their late 20s and early 30s, with a significant drop in respondents over the age of 40.

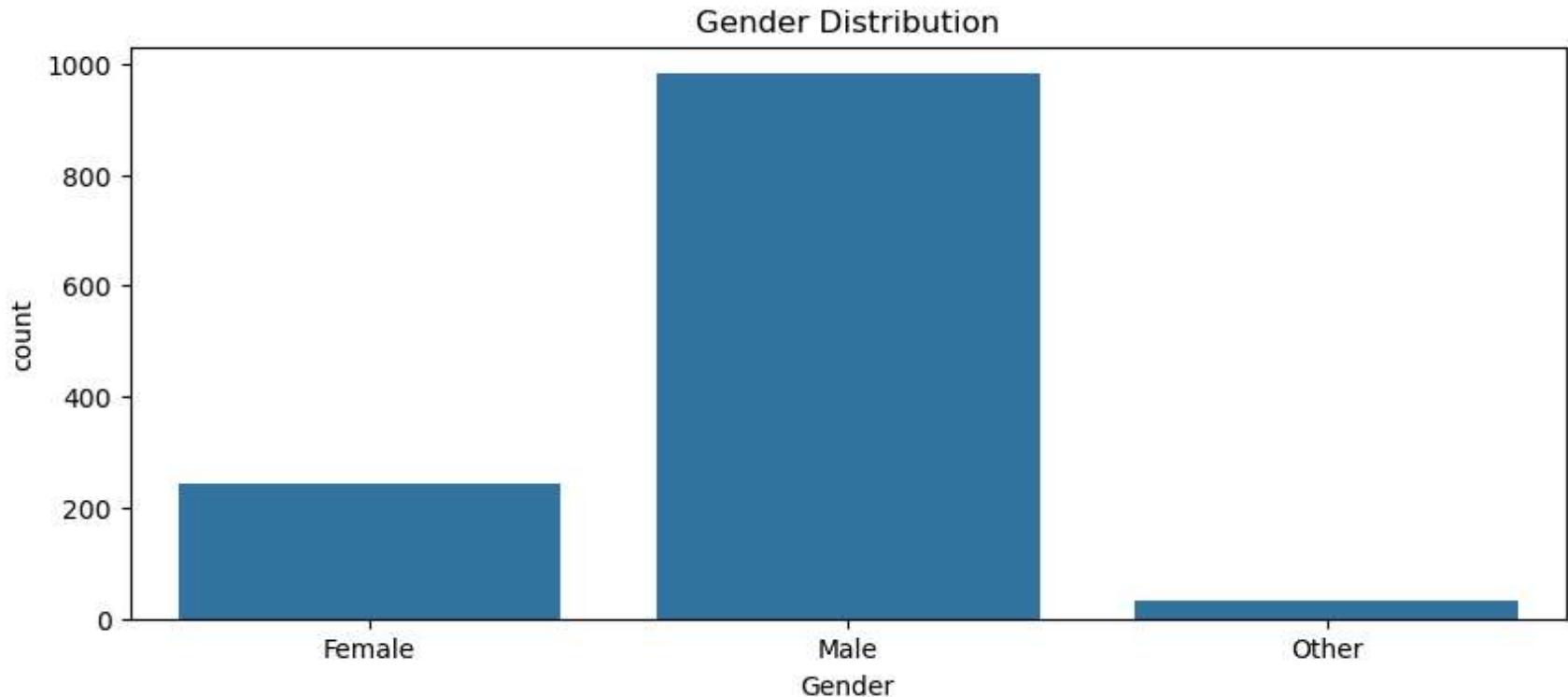
3. Will the gained insights help creating a positive business impact?

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

Yes, this insight can drive positive impact by offering healthy working environment particularly for the dominant age group, but it also reveals a potential negative factor by highlighting a major workforce age struggling with mental health issues.

Chart - 4

```
In [71]: # Chart - 4 Gender distribution/count
plt.figure(figsize=(10, 4))
sns.countplot(data=mental_health_data, x='Gender')
plt.title("Gender Distribution")
plt.show()
```



1. Why did you pick the specific chart?

I select this bargraph using sns.countplot, to observe the gender diversity in the dataset and check representation across male, female, and other categories.

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

The insight from the graph is that the respondent base is heavily skewed towards males, who make up a vast majority of the participants followed by females and others. This reflects either the tech industry imbalance or a greater willingness among men to participate in this specific survey.

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

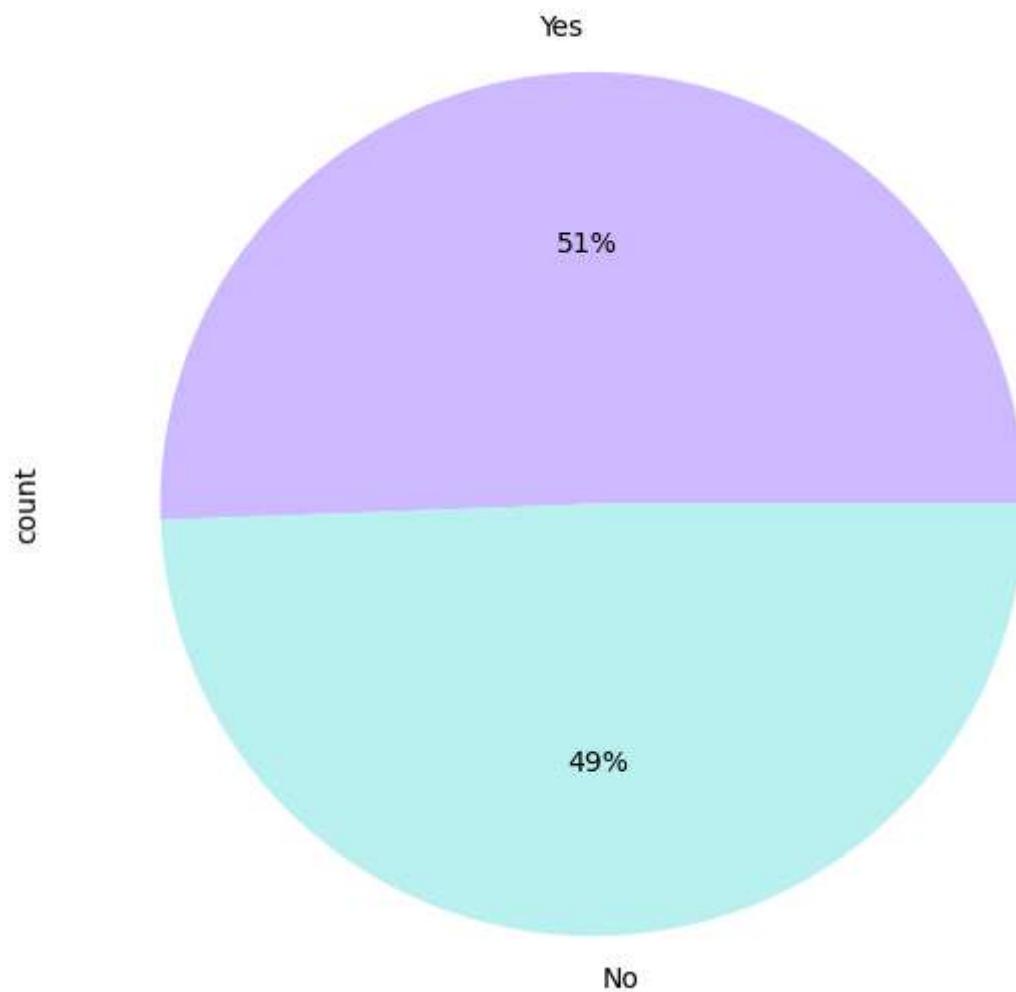
Are there any insights that lead to negative growth? Justify with specific reason.

Gender inclusion policies should ensure that mental health support is inclusive and customized for all gender identities.

Chart - 5

In [106...]

```
# Chart - 5 Treatment sought  
  
mental_health_data['treatment'].value_counts().plot.pie(figsize=(8,7), autopct='%1.0f%')  
plt.show()
```



1. Why did you pick the specific chart?

This pie chart is used to evaluate how many people have sought treatment for mental health, which indicates openness and access to care.

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

A significant number of respondents have sought mental health treatment. This may indicate rising awareness and acceptance, or it may reflect underlying stress and burnout in the tech industry.

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

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

We see a close gap between choosing to seek treatment or not. None the less most people with mental health disorders do choose to seek treatment which is a positive sign.

Bivariate analysis

Chart - 6

```
In [76]: # Chart - 6 Age Vs Treatment Analysis

# initialize canvas of size 15 X 7
fig = plt.figure(figsize = (10, 5))
sns.set_palette(['coral', 'green'])

# binning age data
age_bin = pd.cut(mental_health_data['Age'], bins = range(10,81,10))

# plot countplot of age concerning treatment
sns.countplot(x = age_bin, hue = mental_health_data['treatment'])

# add some cosmetics
plt.title(label = 'Age vs Treatment Analysis', size = 16)

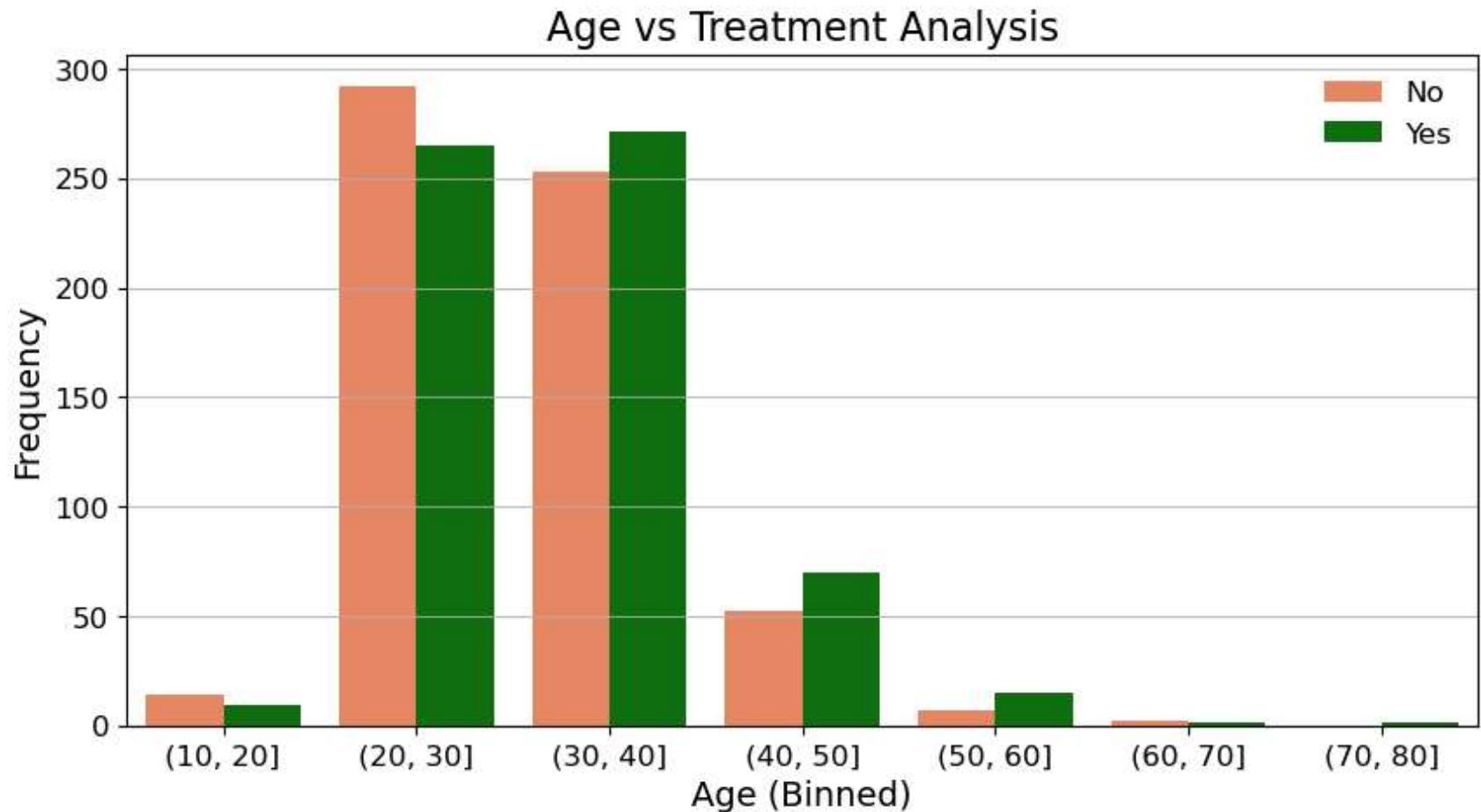
plt.xlabel(xlabel = 'Age (Binned)', size = 14)
```

```

plt.ylabel(ylabel = 'Frequency', size = 14)
plt.xticks(size = 12)
plt.yticks(size = 12)
plt.grid(axis = 'y', alpha = 0.8)
plt.legend(frameon = False, loc = 1, fontsize = 12)

# display the plot
plt.show()

```



1. Why did you pick the specific chart?

I selected this chart to analyze that, how does age relate to various behaviors and/or their awareness of their employer's attitude toward mental health.

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

Insights: People of age 14 - age 22 are less likely to take treatment while people of age 23 - age 43 are more likely to take treatment. People of age 40 and above are receptive to 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.

Mental health outreach and awareness campaigns might be more effective when targeted at younger employees (age-23 to age 43), especially in onboarding and training.

Chart - 7

```
In [79]: # Chart - 7 Gender Response on Treatment

# using pandas query method
mental_health_data.query("Gender.isin(['female', 'male', 'other'])")
mental_health_data['Receptive_Treatment'] = mental_health_data['treatment'] == 'Yes'
mental_health_data['Unresponsive_Treatment'] = mental_health_data['treatment'] == 'No'

# creating 2 subplots to visualize the graphs using groupby method
f, ax = plt.subplots(nrows=1, ncols=2, figsize=(10, 4), sharey = False)

mental_health_data.groupby('Gender')['Receptive_Treatment'].sum().plot(kind = 'pie', autopct = '%.1f%%',
                                                                    wedgeprops = dict(width = 0.15),
                                                                    startangle = 80, ax = ax[0], cmap = 'cool')

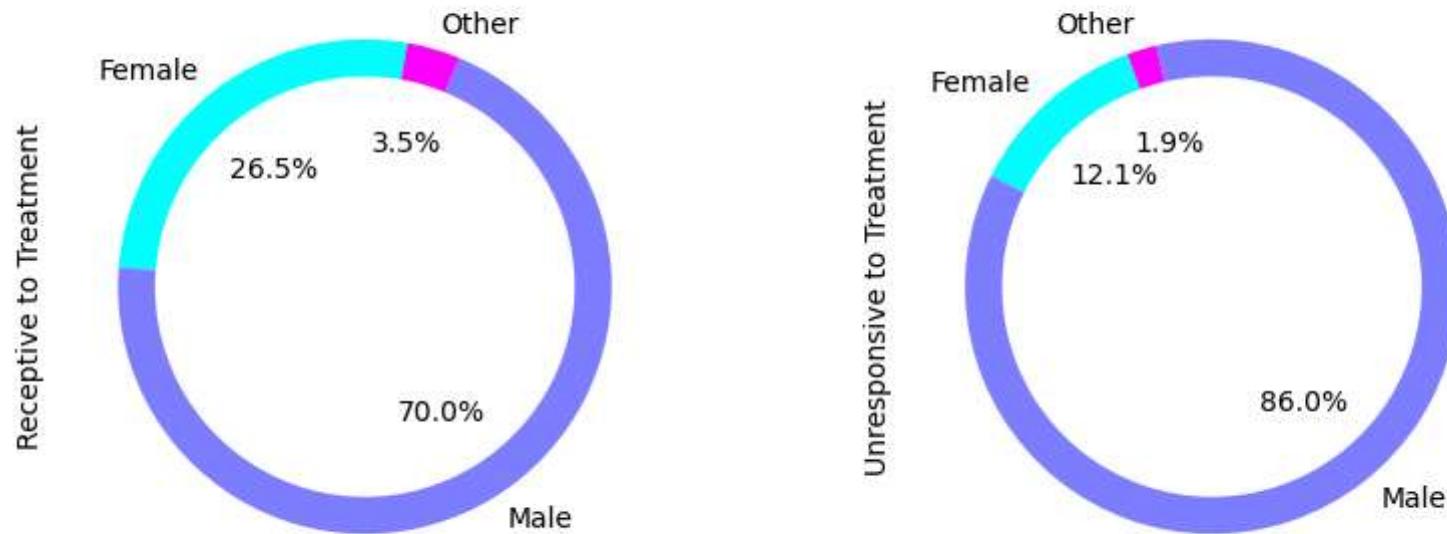
mental_health_data.groupby('Gender')['Unresponsive_Treatment'].sum().plot(kind = 'pie', autopct = '%.1f%%',
                                                                     wedgeprops = dict(width = 0.15),
                                                                     startangle = 110, ax = ax[1], cmap = 'cool')

# adding suptitle
plt.suptitle(t = 'Gender Response on Treatment', size = 18)

# adding some asthetics
ax[0].set_ylabel('Receptive to Treatment', size = 10)
ax[1].set_ylabel('Unresponsive to Treatment', size = 10)
```

```
plt.show()
```

Gender Response on Treatment



1. Why did you pick the specific chart?

I created 2 subplots charts to visualize the graphs using groupby method to analyze the association between treatment and gender in terms of ratio.

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

The chart indicates that a higher number of males are not receiving treatment compared to those who are, while for females, more are receiving treatment than not; the "Others" category shows very low counts for both treatment statuses.

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

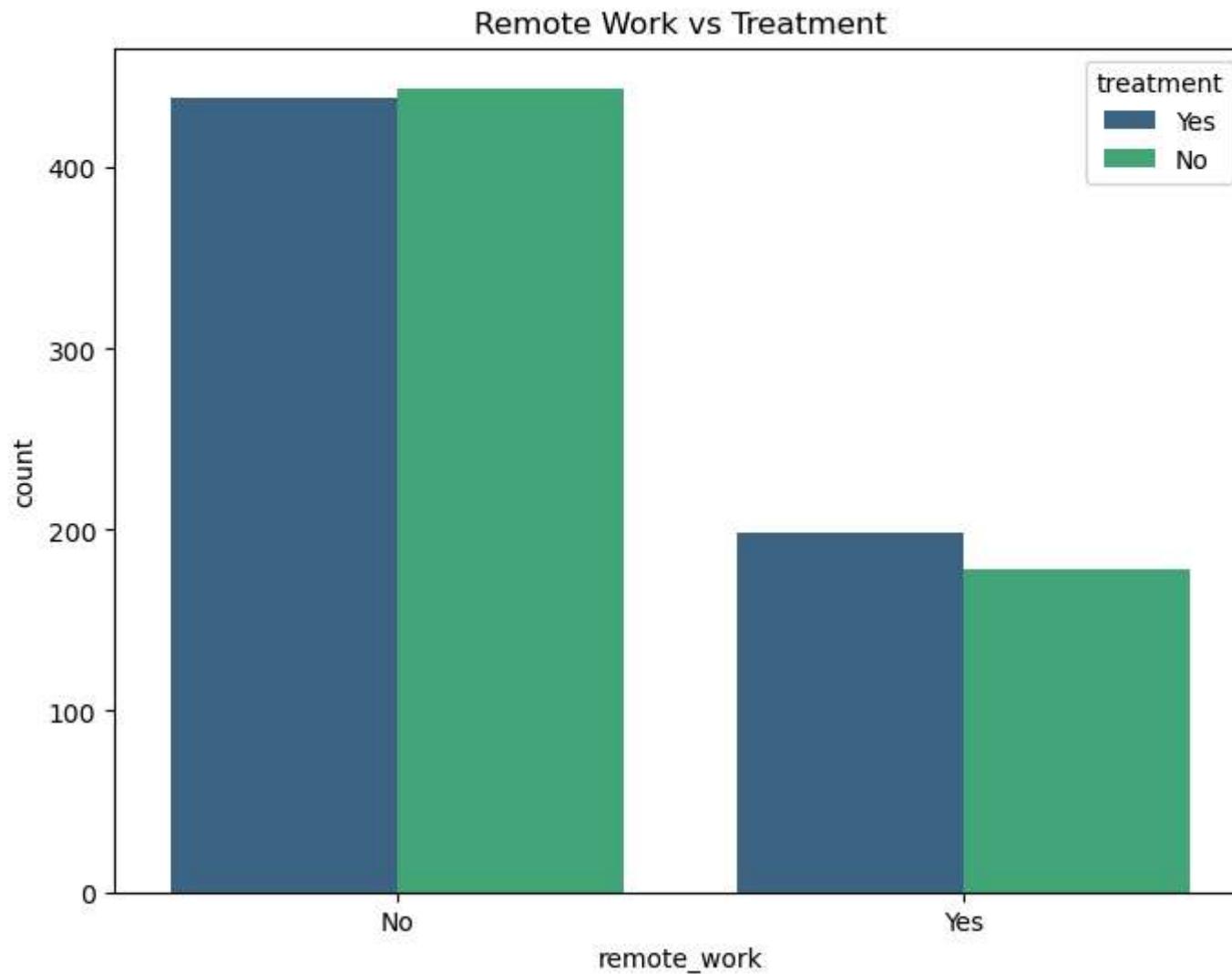
Are there any insights that lead to negative growth? Justify with specific reason.

The low treatment rates among males highlight an underserved group, offering mental health platforms a key opportunity for targeted outreach. However, neglecting the unique needs of males and the underrepresented "Others" group could worsen care disparities if not properly addressed in data and solutions.

Chart - 8

```
In [80]: # Chart - 8 Remote Work vs Treatment

plt.figure(figsize=(8,6))
sns.countplot(data=mental_health_data, x="remote_work", hue="treatment", palette='viridis')
plt.title("Remote Work vs Treatment")
plt.show()
```



1. Why did you pick the specific chart?

A grouped bar chart was chosen to effectively compare the number of individuals receiving or not receiving mental health treatment based on their remote work status.

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

The bar chart shows that among those who do not work remotely, both "Yes" and "No" responses for seeking treatment are nearly balanced, with slightly more saying "No." However, for those who work remotely, there is a slight increase in treatment-seeking.

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

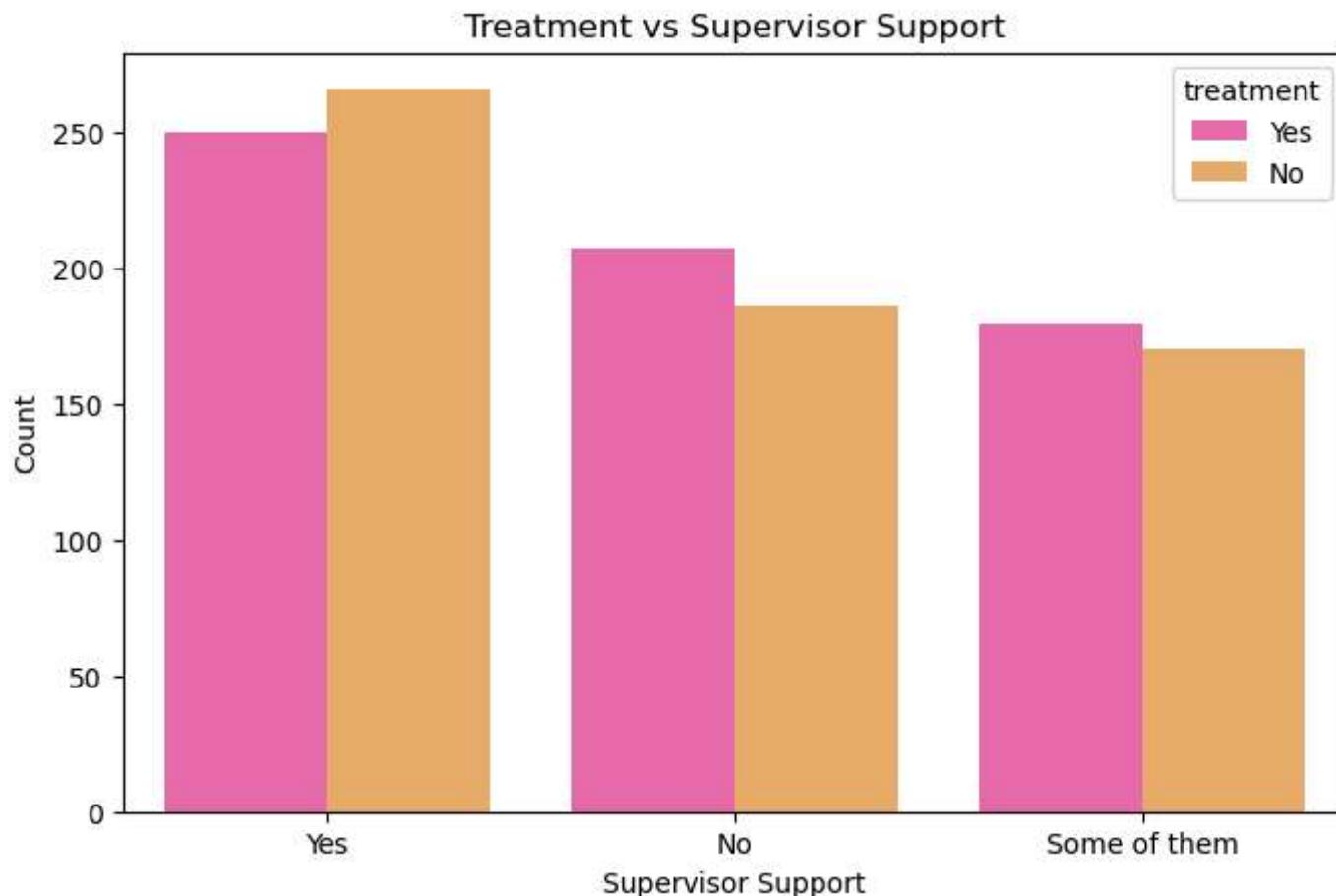
Are there any insights that lead to negative growth? Justify with specific reason.

Remote workers show a higher tendency to seek treatment, highlighting a key audience for digital mental health solutions. However, focusing too heavily on remote setups may overlook the larger non-remote workforce, limiting opportunities to address mental health needs in traditional and hybrid workplaces.

Chart - 9

In [81]: *# Chart - 9 Supervisor Support vs Treatment*

```
plt.figure(figsize=(8, 5))
sns.countplot(data=mental_health_data, x='supervisor', hue='treatment', palette='spring')
plt.title('Treatment vs Supervisor Support')
plt.xlabel('Supervisor Support')
plt.ylabel('Count')
plt.show()
```



1. Why did you pick the specific chart?

A grouped bar chart was chosen to effectively compare the number of individuals receiving or not receiving mental health treatment across different levels of supervisor support.

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

The chart indicates that individuals with supervisor support ("Yes") are slightly less likely to be in treatment, while those with no supervisor support ("No") are more likely to be in treatment. Individuals with "Some of them" support show a slightly higher likelihood of being in treatment.

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

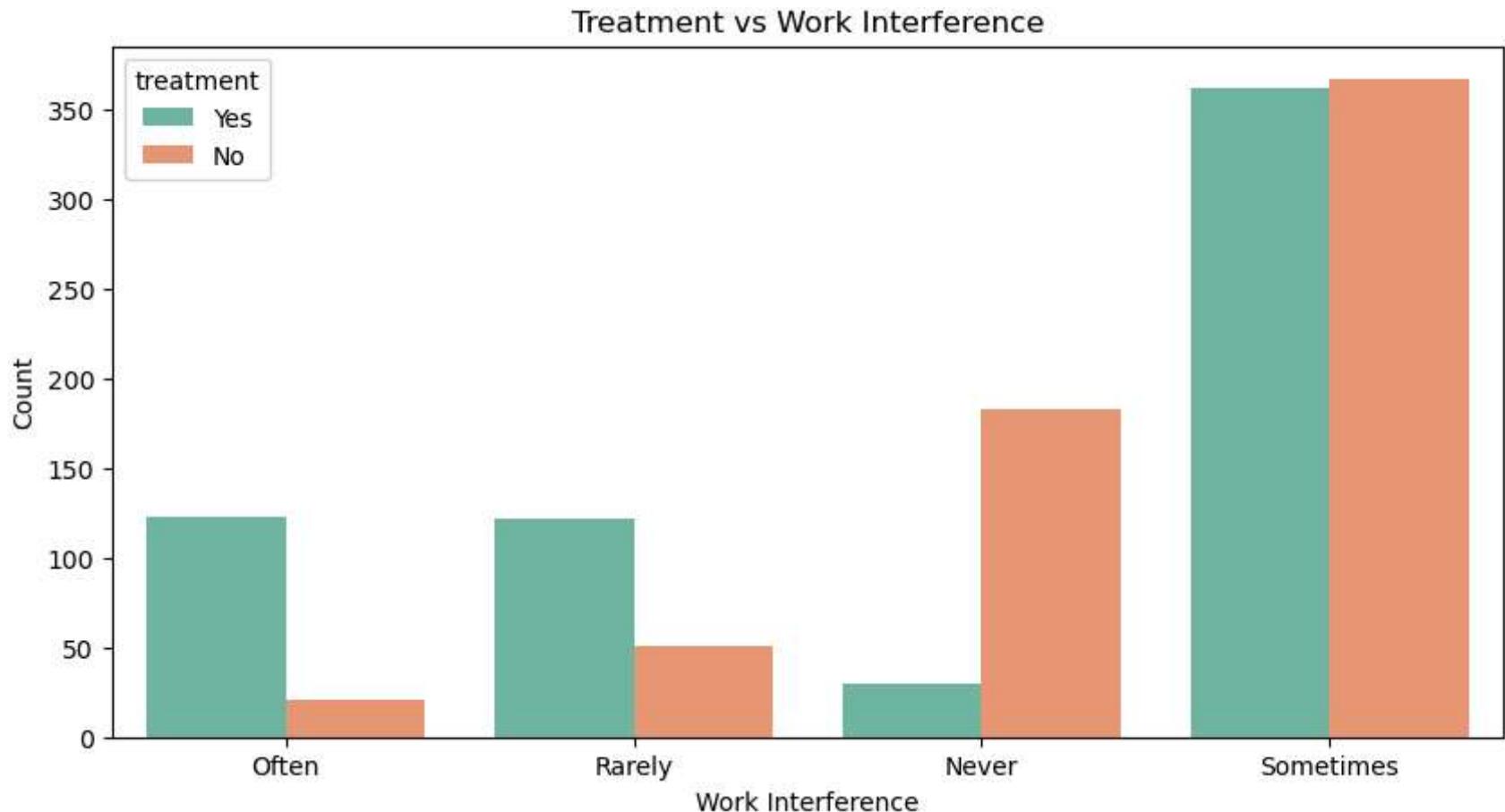
Are there any insights that lead to negative growth? Justify with specific reason.

Individuals receiving supervisor support still seek treatment, revealing demand for mental health care tools. However, those without support being more likely in treatment signals a risk: without improving managerial empathy, mental health tools alone may fall short in fostering true workplace well-being.

Chart - 10

```
In [82]: # Chart - 10 Work Interference Vs Treatment
```

```
plt.figure(figsize=(10, 5))
sns.countplot(data=mental_health_data, x='work_interfere', hue='treatment', palette='Set2')
plt.title('Treatment vs Work Interference')
plt.xlabel('Work Interference')
plt.ylabel('Count')
plt.show()
```



1. Why did you pick the specific chart?

I selected this group bar chart to compare the number of individuals receiving or not receiving mental health treatment across different levels of work interference.

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

The chart indicates that individuals who experience "Often" or "Rarely" work interference are more likely to be receiving treatment, while those who experience "Never" work interference are significantly less likely to be in treatment, and "Sometimes" shows a roughly even split.

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

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

Individuals reporting frequent work interference are actively seeking treatment, indicating strong demand for workplace mental health support. However, overlooking those with little or occasional interference risks missing a large group that could benefit from early interventions, potentially limiting long-term impact and growth.

Multiivariate analysis

Chart - 11

```
In [86]: # Chart - 8 Treatment Vs Work Interference by Gender Analysis

# create subplots with one row and three columns
f, ax = plt.subplots(nrows=1, ncols=3, figsize=(15, 6), sharey=False)
hue_order = ['Yes', 'No']

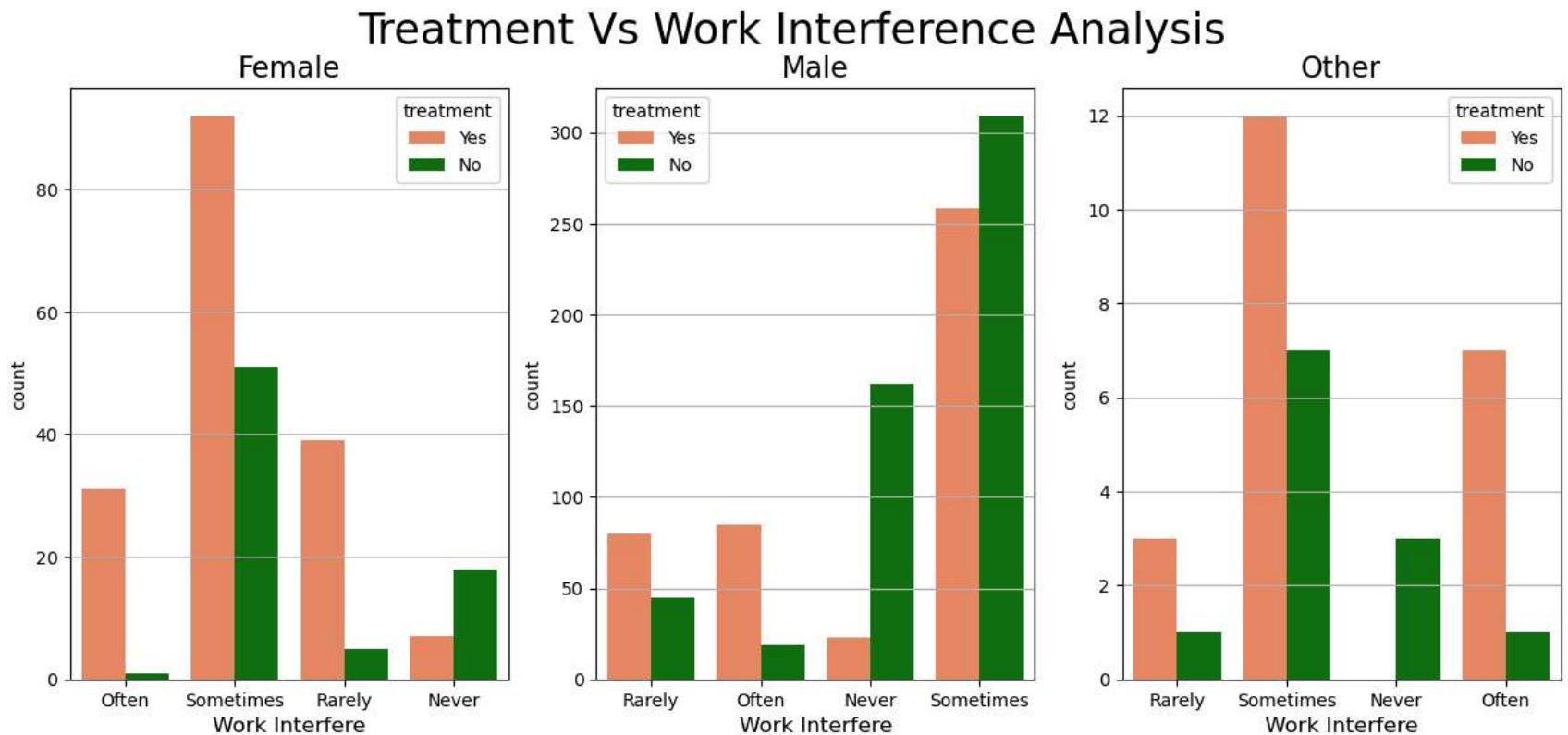
# plot count plots for each gender separately
sns.countplot(data=mental_health_data[mental_health_data['Gender'] == 'Female'], x='work_interfere', hue='treatment',
               hue_order = hue_order)
ax[0].set_title('Female', size=16)
ax[0].set_xlabel('Work Interfere', size = 12)
ax[0].grid(axis = 'y')

sns.countplot(data=mental_health_data[mental_health_data['Gender'] == 'Male'], x='work_interfere', hue='treatment',
               hue_order = hue_order)
ax[1].set_title('Male', size=16)
ax[1].set_xlabel('Work Interfere', size = 12)
ax[1].grid(axis = 'y')

sns.countplot(data=mental_health_data[mental_health_data['Gender'] == 'Other'], x='work_interfere', hue='treatment',
               hue_order = hue_order)
ax[2].set_title('Other', size=16)
ax[2].set_xlabel('Work Interfere', size = 12)
ax[2].grid(axis = 'y')
```

```
# setting a common suptitle
plt.suptitle(t = 'Treatment Vs Work Interference Analysis', size = 24)

plt.show()
```



1. Why did you pick the specific chart?

It allows us to break down how work interference due to mental health varies across gender, and how it relates to whether someone sought treatment. This multi-panel approach is excellent for spotting trends within each gender group separately, which would be hidden in a single plot.

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

Insights:

- For Male:**

- The largest group falls under "Sometimes" work interference. A majority of them did not seek treatment, even though interference exists.
- Treatment-seeking is relatively low for "Rarely" and "Never" categories.

2. For Female:

- More women seek treatment when they report work interference as "Rarely" or "Sometimes".
- Women are more proactive in seeking help when they notice mental health impacting work.

3. For Other:

- The sample size is small, but those reporting "Sometimes" interference are more likely to seek treatment.
- Despite lower numbers, the behavior is more similar to female respondents than male.

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

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

Female employees are more likely to seek help when mental health interferes with work, suggesting higher awareness or reduced stigma in this group. Male employees may need more proactive encouragement or a safer environment to seek help.

Chart - 12

```
In [7]: # Chart - 12 Family History Effects on Mental health

# creating a canvas of 11 X 7
fig = plt.figure(figsize = [11, 5])

# setting the color palette
sns.set_palette(['#d0bbff', '#b9f2f0'])

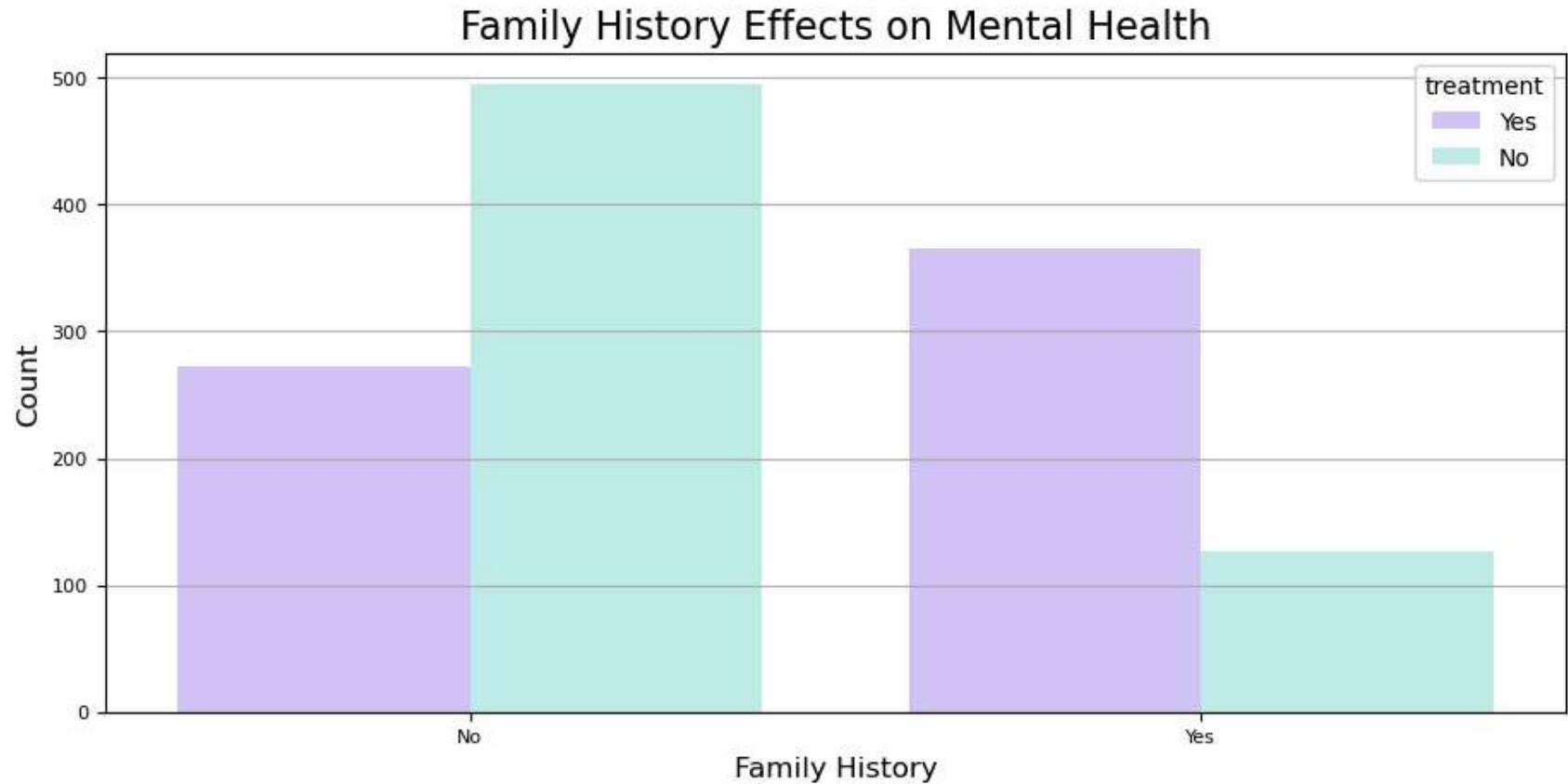
# creating a countplot
sns.countplot(data = mental_health_data, x = 'family_history', hue = 'treatment')

# adding some cosmetics
plt.xticks(size = 8)
plt.xlabel('Family History', size = 12)
plt.yticks(size = 8)
```

```
plt.ylabel('Count', size = 12)
plt.title('Family History Effects on Mental Health', size = 16)
plt.grid(axis = 'y')

plt.show()
```

<Figure size 1100x500 with 0 Axes>



1. Why did you pick the specific chart?

This chart is used to describe the association between treatment and family history of the employee. A grouped bar chart was chosen to effectively compare the number of individuals receiving or not receiving mental health treatment based on their family history of mental illness.

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

The chart indicates that individuals with no family history are more likely to not be in treatment, whereas individuals with a family history of mental illness are more likely to be receiving treatment. Employees in tech industry who have family history are more likely to take 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.

Yes, Individuals with a family history are more likely to seek treatment, offering a clear target for advanced mental health solutions. However, focusing only on high-risk groups may overlook a wider audience, limiting growth in early detection and general wellness efforts.

Chart - 13 - Correlation Heatmap

```
In [68]: # initializing empty list
cat_features = []

# appending all the categorical features to empty list
for i in mental_health_data.columns:
    if mental_health_data[i].dtype == 'object':
        cat_features.append(i)

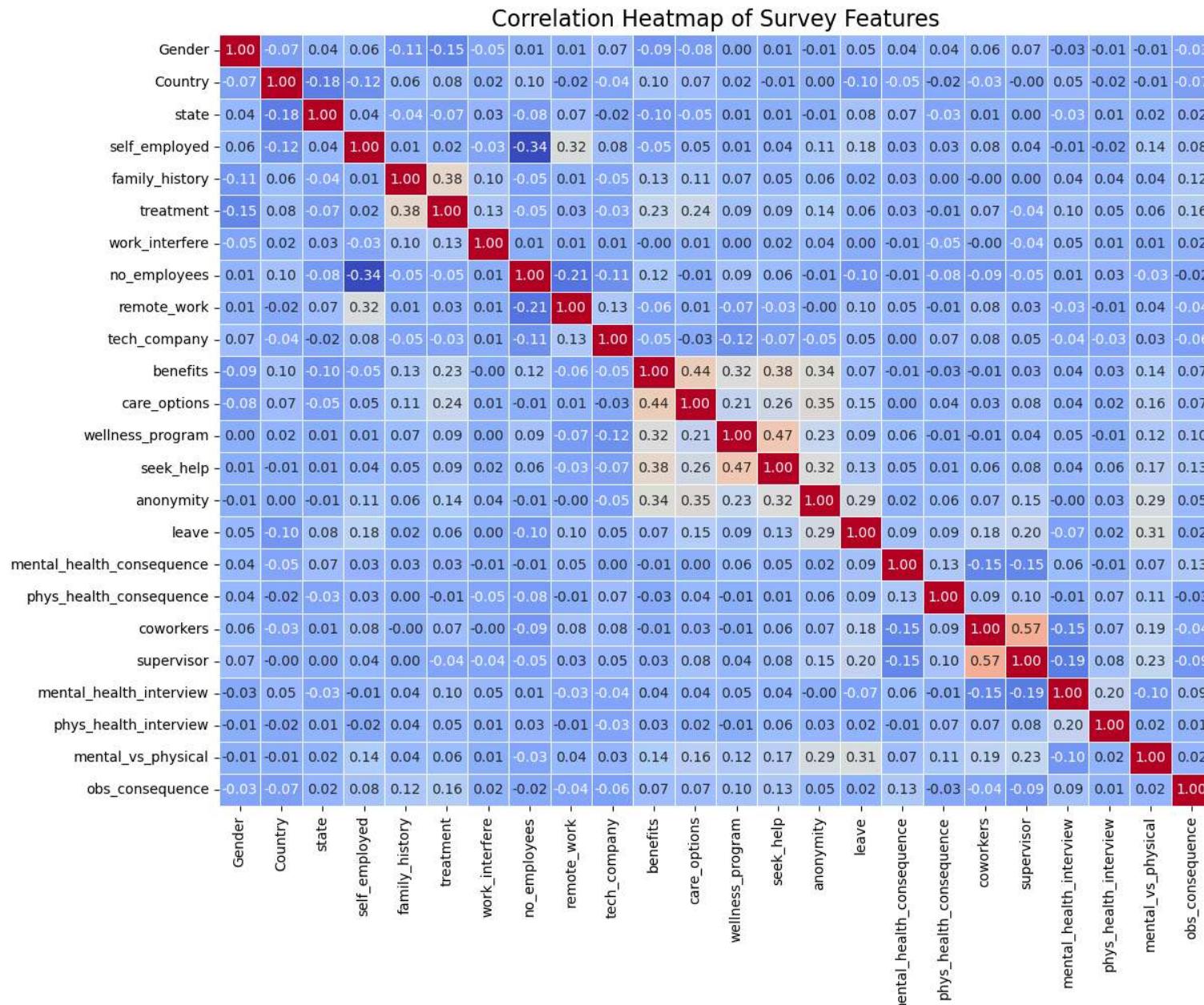
print(cat_features)

['Gender', 'Country', 'state', 'self-employed', 'family_history', 'treatment', 'work_interfere', 'no_employees', 'remote_work', 'tech_company', 'benefits', 'care_options', 'wellness_program', 'seek_help', 'anonymity', 'leave', 'mental_health_consequence', 'phys_health_consequence', 'coworkers', 'supervisor', 'mental_health_interview', 'phys_health_interview', 'mental_vs_physical', 'obs_consequence']
```

```
In [69]: from sklearn.preprocessing import LabelEncoder

encoded_data= mental_health_data[cat_features].apply(LabelEncoder().fit_transform)

# Correlation matrix
plt.figure(figsize=(16, 10))
sns.heatmap(encoded_data.corr(), cmap="coolwarm", annot=True, fmt=".2f", linewidths=0.5)
plt.title("Correlation Heatmap of Survey Features", fontsize=16)
plt.show()
```



1. Why did you pick the specific chart?

The correlation heatmap effectively visualizes relationships between multiple mental health survey features at once, helping identify key associations.

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

The Correlation Heatmap gives us the insights:

- The variables that have an effect over another variable and to what extent.
- These features can be +ve or -ve Correlated with each other signified by the sign of the correlation coefficient.
- The Magnitude of the Correlation represents the degree of the relationship.
- The correlation is drawn with Treatment and the other variables and we see, it is correlated with the following:
 - Family History, Care Options, Benefits and Obs_Consequence
- We can also judge the relationship between other features as well. Some of the highly correlated features are:
 - coworker & supervisor (0.57)
 - seek_help & wellness_program (0.46)
 - seek_help & anonymity (0.32)
 - mental_vs_physical & leave (0.31)
- The insights can drive positive business impact by improving mental health support systems through aligned wellness programs, benefits, and managerial support. No strong negative correlations were found, but weak associations may highlight missed opportunities in underutilized support services.

5. Solution to Business Objective

What do you suggest the client to achieve Business Objective ?

Explain Briefly.

- Employers should create flexible work schedule for the employees(hybrid work routine).
- Employers need to effectively communicate to their employees regarding mental health resources available at the workplace.
- Many employees either don't know or believe they don't have access to mental health resources — especially remote workers.
To solve this problem, the company should Regularly communicate available support via email, intranet, and meetings.
- Promote mental health in male-dominated teams through leaders sharing personal experiences.

- Expand Mental Health Support in Small and Remote Companies. Provide access to virtual therapy, online workshops, and HR training for small teams.
- Provide affordable, flexible programs for small to mid-sized companies
- Introduce anonymous feedback systems and safe spaces for open discussion (e.g., wellness forums or buddy programs).
- Simplify medical leave process and openly communicate leave policies, especially around mental wellness.
- Use anonymized internal surveys or HR tools to identify risk patterns and offer early support.
- Offer wellness checks and preventive support even to those not yet seeking treatment.
- Consider creating programs or resources that cater to individuals aged 40 and above, as they are more receptive to treatment.

Conclusion

- This project analyzed a 2014 survey dataset to understand mental health awareness, treatment patterns, and workplace support in the tech industry.
- The survey primarily targeted US tech employees, as 60% of the data is from the US.
- Using exploratory data analysis and visualizations, we identified key trends across gender, age, company size, and remote work setup.
- People between the ages of 23 and 43 are more likely to seek treatment, while those aged 14 to 22 are less likely.
- The gender mostly affected is the Male. Male Employees need to be enlightened about mental health, seeking help and know they can have a safe space to express themselves at the workplace. Workloads should be assigned unbiased regardless of gender.
- More than 50% of across gender believe that Sometimes mental health interferes with work, followed by Female & Male.
- Majority of the employees are not aware about resources provided by employers to seek help, get wellness program and care options. Therefore, employers need to effectively communicate to their employees regarding mental health resources available at the workplace.
- Remote employees are less likely to know or access mental health support, despite being in potentially more isolated environments.
- Family history, care options, benefits, and work consequences are correlated with seeking treatment, while other variables also show correlations among themselves. Additionally, larger companies tend to offer better mental health resources and related benefits.

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