Project Name

Mental Health in Tech Survey

Project Type - EDA/Regression/Classification/Unsupervised

Contribution - Individual

**Team Member 1 - Bhushan Dongardive

Project Summary -

Mental health is now a growing issue in the current fast-paced work culture, particularly in the tech industry where employment stress, burnout, and work culture at home can have the biggest impact on well-being. This project examines and investigates data from the Mental Health in Tech Survey to reveal patterns, attitudes, and work influences on mental health treatment, awareness, and openness among tech staff.

The data employed is based on the responses to a survey of employees in the technology sector. It comprises demographic data about age, sex, and nation, as well as workplace characteristics such as firm size, working remotely, and having mental health benefits. The aim of this project was to conduct exploratory data analysis (EDA) to find insights that can inform organizations to create a healthier and more caring work environment.

First, the data went through the appropriate cleaning and wrangling. Inconsistent age values such as less than 18 or greater than 100 were eliminated, and gender labels that did not match were standardized into three broad categories: Male, Female, and Other. Missing values in essential columns such as self_employed and work_interfere were replaced with logical defaults or most frequent values to ensure the dataset remained intact.

The EDA was both univariate and bivariate in nature. One of the most dramatic findings was that most respondents were aged between 25 and 35. This group also had the highest proportion of those who had been treated for their mental health, perhaps because they were better informed or experienced greater stress at younger and middle-career life stages.

The gender split saw more male respondents, as with wider industry trends. More female and "Other" gender respondents did indicate seeking treatment, however, which indicates that men are still possibly subject to social stigma or less likely to admit to mental health issues, and organizations should look to target awareness efforts towards this group.

Another observation was made from comparisons of remote workers and their potential to access care. Remote staff had a slightly increased potential to access help, perhaps as a result of greater flexibility or reduced work barriers. And that raises issues about isolation and the necessity for remote-targeted wellness initiatives.

Company size analysis indicated that big organizations are more probable to have mental health benefits, whereas small companies (fewer than 25 people) tend not to have these resources available. The gap is considerable and has the potential to affect employee retention, morale, and productivity. Small businesses, and startups in general, might require outside assistance or exposure to introduce even simple mental health programs.

Other variables, including work_interfere, anonymity, and leave, assisted in revealing the influence of workplace culture on employee willingness to talk or seek treatment about mental health. Some respondents, for example, remained uncertain whether they could take mental health leave or if their anonymity would be maintained — sure signs of communication gaps and HR policy.

Ultimately, the project underscores the significance of infusing workplace dialogue with mental health considerations, particularly in the technology sector. Businesses can have a positive business effect by providing benefits, advocating mental health awareness, safeguarding employee privacy, and facilitating help-seeking behavior. These findings can be used as a starting point for policy reform and cultural change that benefits the mental health of employees — leading to a healthier, more productive, and dedicated workforce.

GitHub Link -

Provide your GitHub Link here.

Problem Statement

Mental health is a critical yet often overlooked aspect of employee well-being, particularly in the fast-paced and high-pressure environments of the tech industry. Despite growing awareness, many individuals still face stigma, lack of support, and unclear workplace policies regarding mental health.

The purpose of this project is to analyze survey data from tech employees to understand how factors such as gender, age, work environment, company size, and mental health benefits influence mental health awareness, treatment-seeking behavior, and perceptions in the workplace.

By performing exploratory data analysis (EDA), the goal is to uncover actionable insights that can help organizations:

Recognize mental health trends and gaps in support

Improve company policies (e.g., benefits, anonymity, medical leave)

Promote mental wellness culture and inclusive discussions

Identify demographics that may be underserved or at risk

Define Your Business Objective?

The objective of this project is to help tech companies and HR decision-makers better understand the mental health landscape within their workforce. By analyzing survey data, we aim to:

Identify key factors (e.g., gender, remote work, company size, mental health benefits) that influence an employee's willingness to seek mental health treatment.

Measure the level of awareness and accessibility of mental health support across different workplace environments.

Reveal gaps in current workplace policies that may prevent employees from seeking help, such as lack of benefits, fear of negative consequences, or poor communication.

Promote data-driven decision-making to design inclusive, stigma-free mental wellness programs and improve employee retention, satisfaction, and productivity.

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.

5. 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

In [3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

# Load dataset
df = pd.read_csv('survey.csv')
df.head()
```

Out[3]:

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No

5 rows × 27 columns

Dataset Loading

In [4]:
```
df.shape
df.info()
df.describe()
```

```
<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
```

Out[4]:

	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

Data Cleaning

In [5]:

```
# Drop irrelevant columns
df.drop(['comments', 'Timestamp'], axis=1, inplace=True)

# Remove unrealistic ages
df = df[(df['Age'] >= 18) & (df['Age'] <= 100)]
```

```
# Clean gender column
def clean_gender(g):
    g = str(g).lower()
    if 'male' in g:
        return 'Male'
    elif 'female' in g:
        return 'Female'
    else:
        return 'Other'

df['Gender'] = df['Gender'].apply(clean_gender)

# Check for null values
df.isnull().sum()
```

Out[5]:
```
Age                            0
Gender                         0
Country                        0
state                        513
self_employed                 18
family_history                 0
treatment                      0
work_interfere               262
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
```

Dataset First View

In [6]:
```
df.head()
```

Out[6]:

	Age	Gender	Country	state	self_employed	family_history	treatment	work_interf
0	37	Male	United States	IL	NaN	No	Yes	Of
1	44	Other	United States	IN	NaN	No	No	Rar
2	32	Male	Canada	NaN	NaN	No	No	Rar
3	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Of
4	31	Male	United States	TX	NaN	No	No	Ne

5 rows × 25 columns

Dataset Rows & Columns count

In [7]:
```
# Display number of rows and columns
print(f"Number of Rows    : {df.shape[0]}")
print(f"Number of Columns : {df.shape[1]}")
```

```
Number of Rows    : 1251
Number of Columns : 25
```

Dataset Information

In [8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1251 entries, 0 to 1258
Data columns (total 25 columns):
 #   Column                     Non-Null Count  Dtype 
---  ------                     --------------  ----- 
 0   Age                        1251 non-null   int64 
 1   Gender                     1251 non-null   object
 2   Country                    1251 non-null   object
 3   state                      738 non-null    object
 4   self_employed              1233 non-null   object
 5   family_history             1251 non-null   object
 6   treatment                  1251 non-null   object
 7   work_interfere             989 non-null    object
 8   no_employees               1251 non-null   object
 9   remote_work                1251 non-null   object
 10  tech_company               1251 non-null   object
 11  benefits                   1251 non-null   object
 12  care_options               1251 non-null   object
 13  wellness_program           1251 non-null   object
 14  seek_help                  1251 non-null   object
 15  anonymity                  1251 non-null   object
 16  leave                      1251 non-null   object
 17  mental_health_consequence  1251 non-null   object
 18  phys_health_consequence    1251 non-null   object
 19  coworkers                  1251 non-null   object
 20  supervisor                 1251 non-null   object
 21  mental_health_interview    1251 non-null   object
 22  phys_health_interview      1251 non-null   object
 23  mental_vs_physical         1251 non-null   object
 24  obs_consequence            1251 non-null   object
dtypes: int64(1), object(24)
memory usage: 254.1+ KB
```

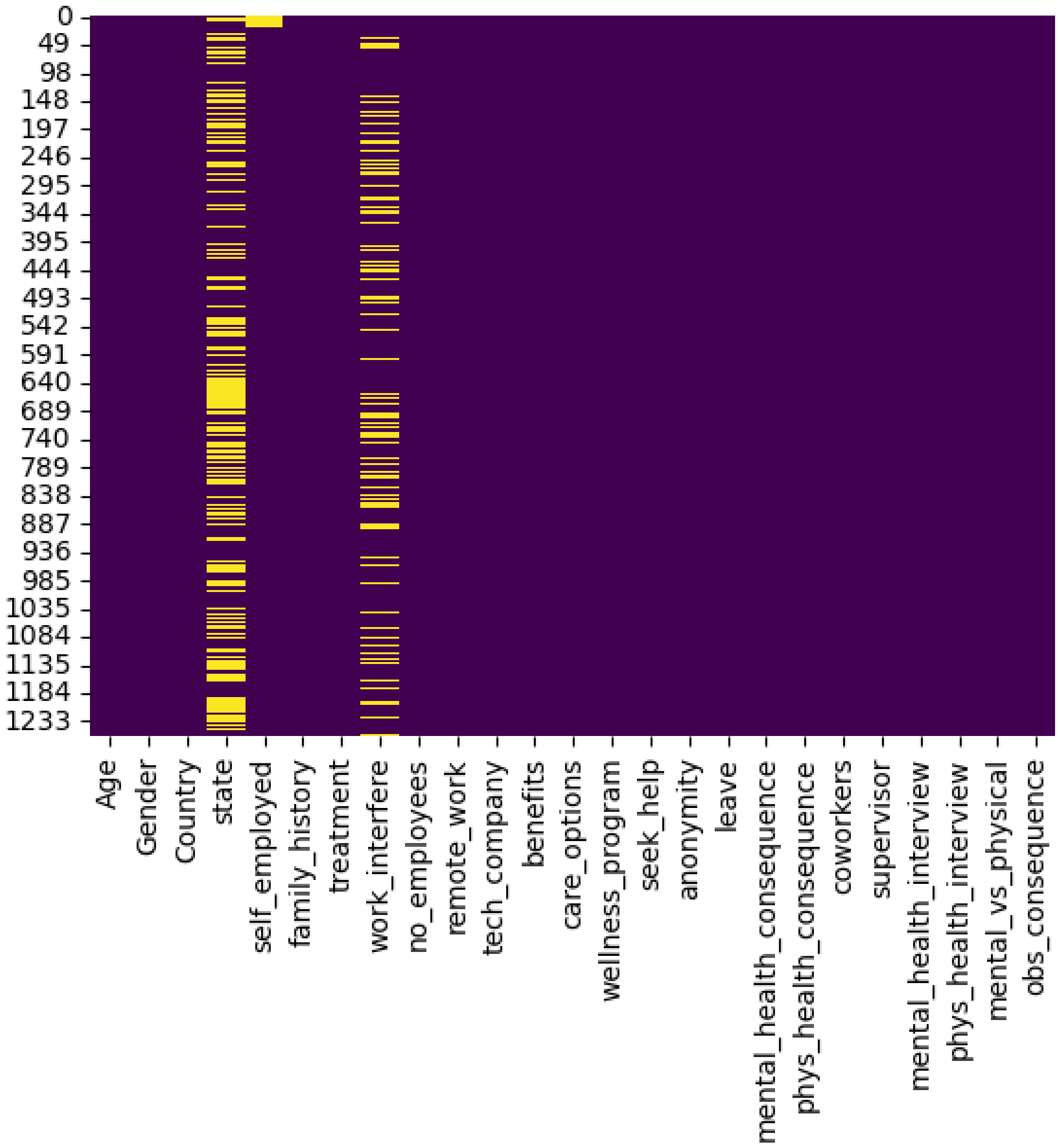
Missing Values/Null Values

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
Age                          0
Gender                       0
Country                      0
state                      513
self_employed               18
family_history               0
treatment                    0
work_interfere             262
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
```

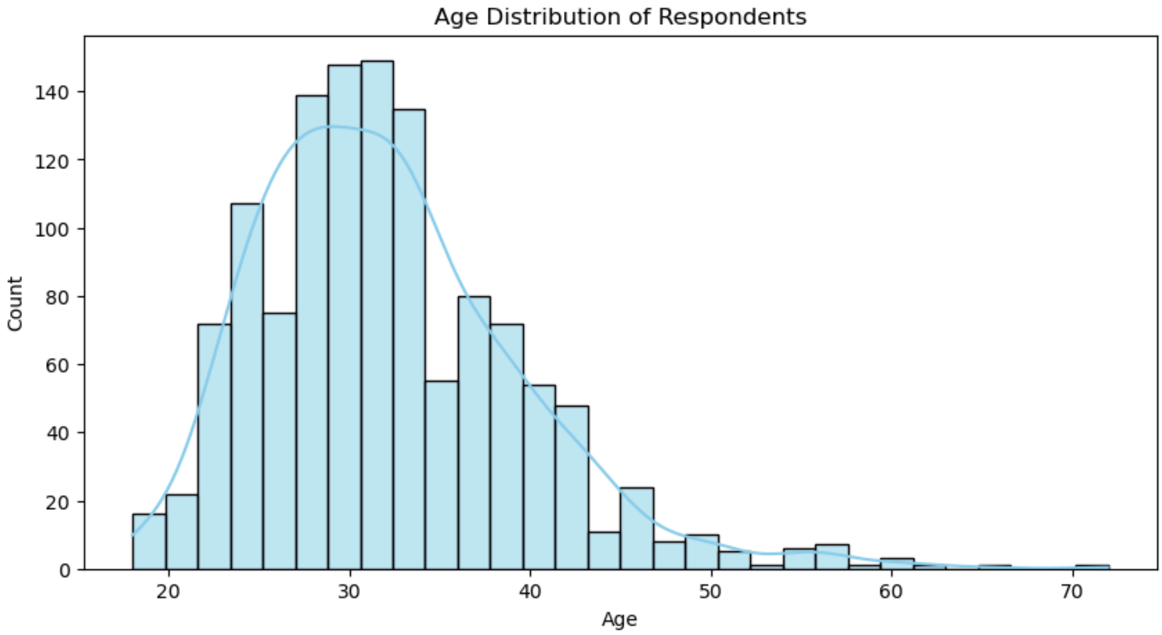
In [10]:
```
# Visualizing missing values
sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
```

Out[10]: <Axes: >



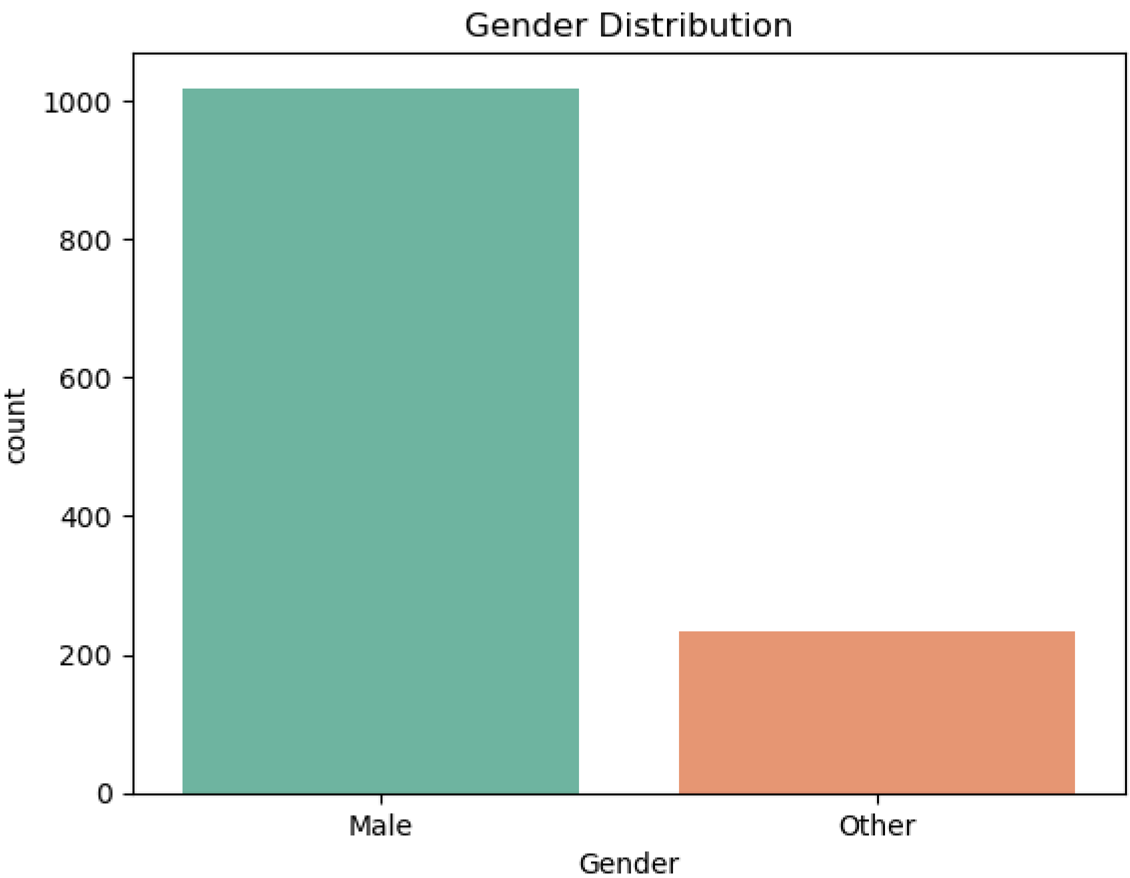
Age Distribution

In [14]:
```
plt.figure(figsize=(10,5))
sns.histplot(df['Age'], bins=30, kde=True, color='skyblue')
plt.title('Age Distribution of Respondents')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```



Gender Distribution

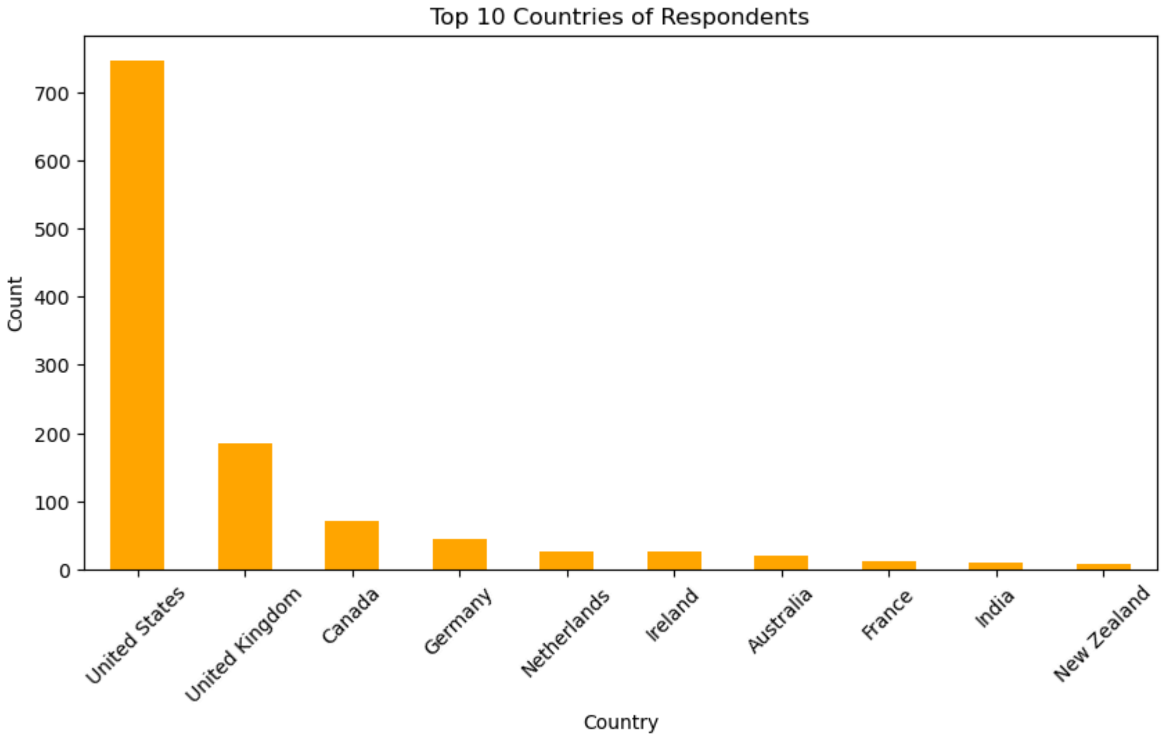
In [15]:
```
sns.countplot(data=df, x='Gender', palette='Set2')
plt.title('Gender Distribution')
plt.show()
```



Top 10 Countries of Respondents

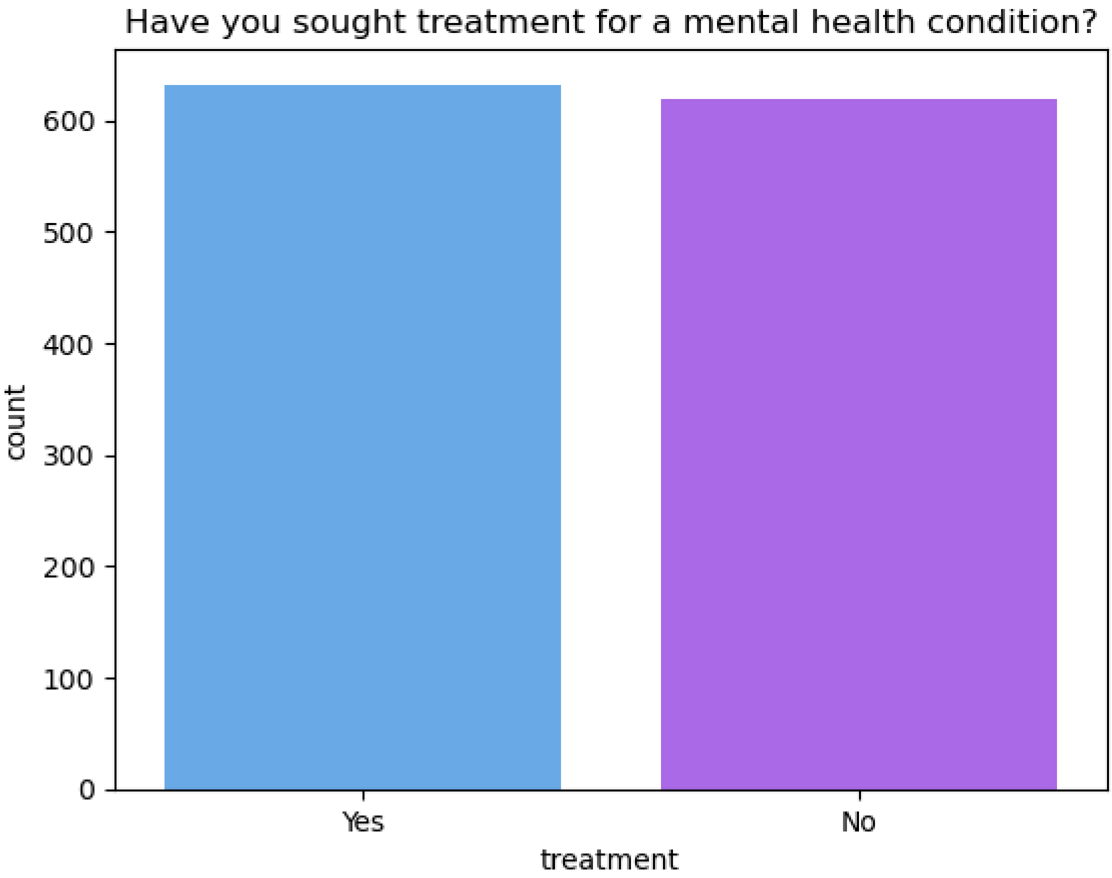
In [16]:
```
df['Country'].value_counts().head(10).plot(kind='bar', figsize=(10,5), color='or
plt.title('Top 10 Countries of Respondents')
```

```
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



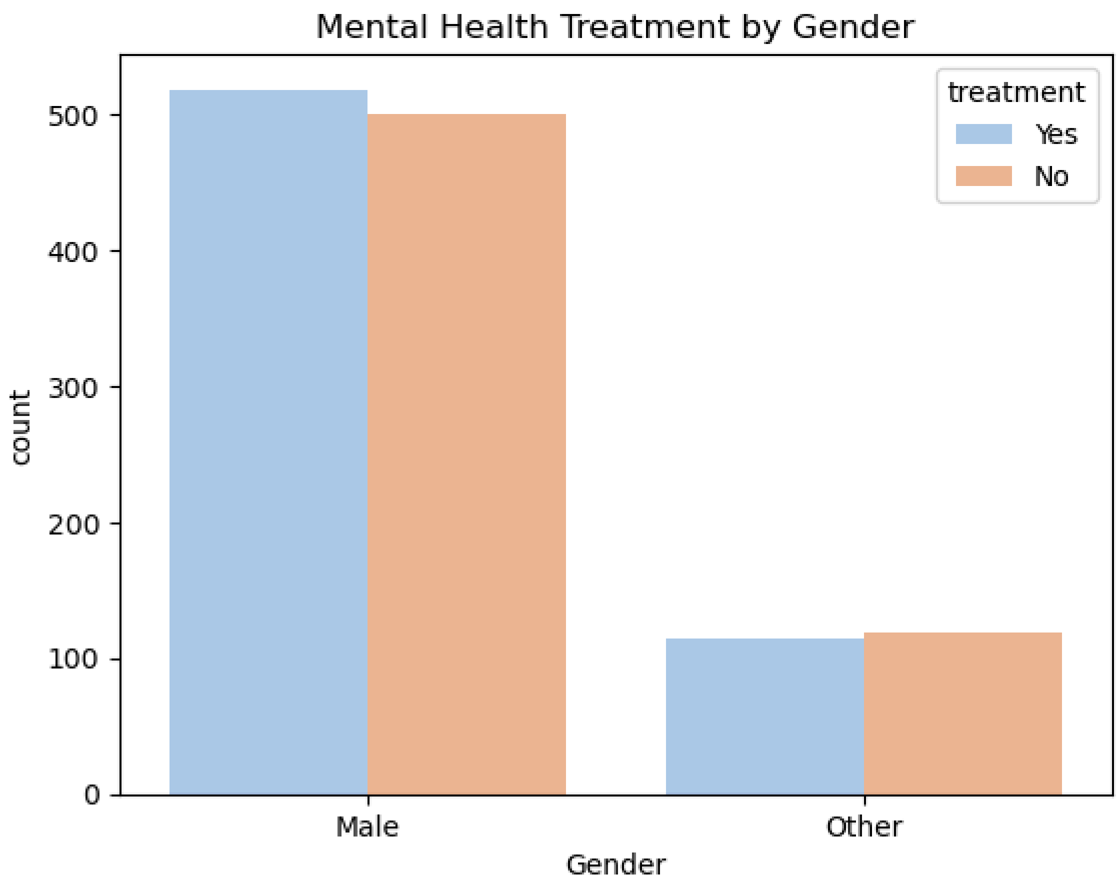
Mental Health Treatment Count

In [17]:
```
sns.countplot(data=df, x='treatment', palette='cool')
plt.title('Have you sought treatment for a mental health condition?')
plt.show()
```



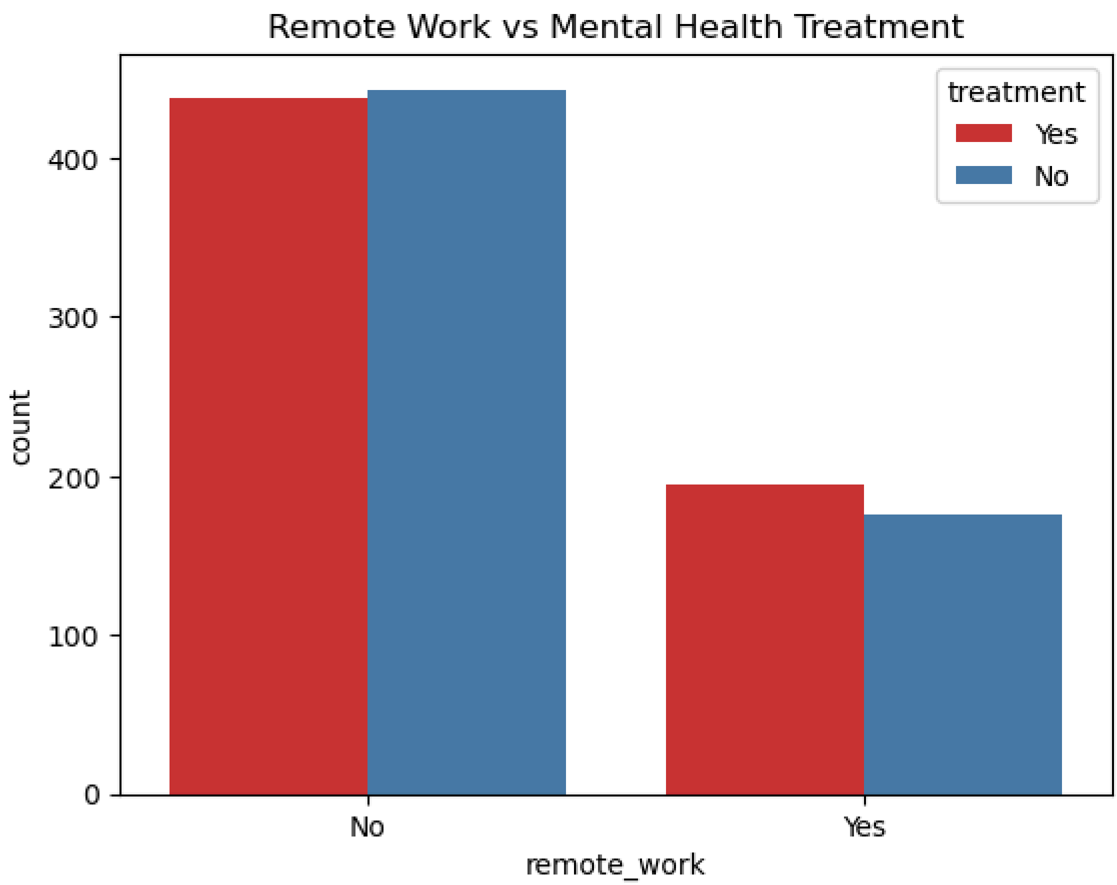
Treatment by Gender

In [18]:
```
sns.countplot(data=df, x='Gender', hue='treatment', palette='pastel')
plt.title('Mental Health Treatment by Gender')
plt.show()
```



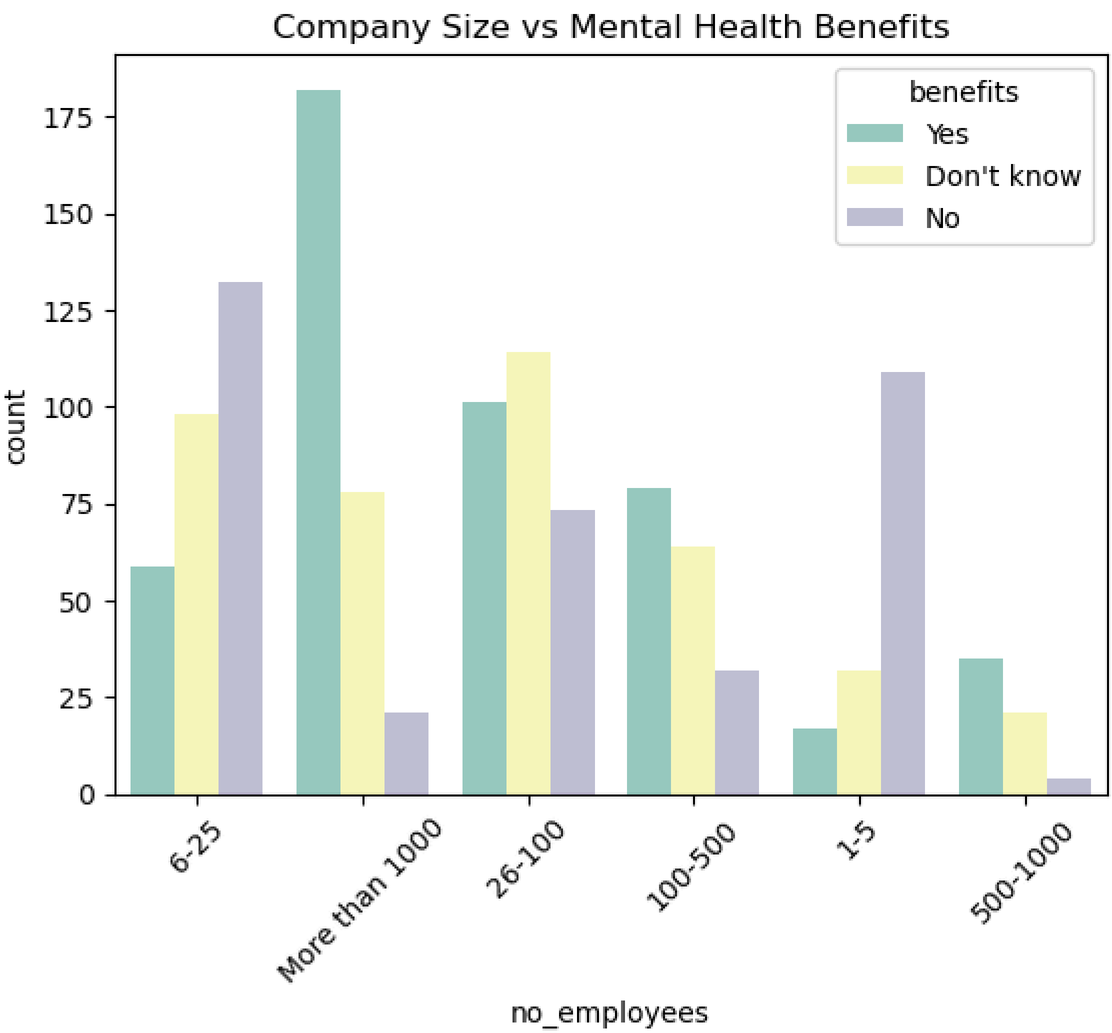
Remote Work vs Treatment

In [19]:
```
sns.countplot(data=df, x='remote_work', hue='treatment', palette='Set1')
plt.title('Remote Work vs Mental Health Treatment')
plt.show()
```



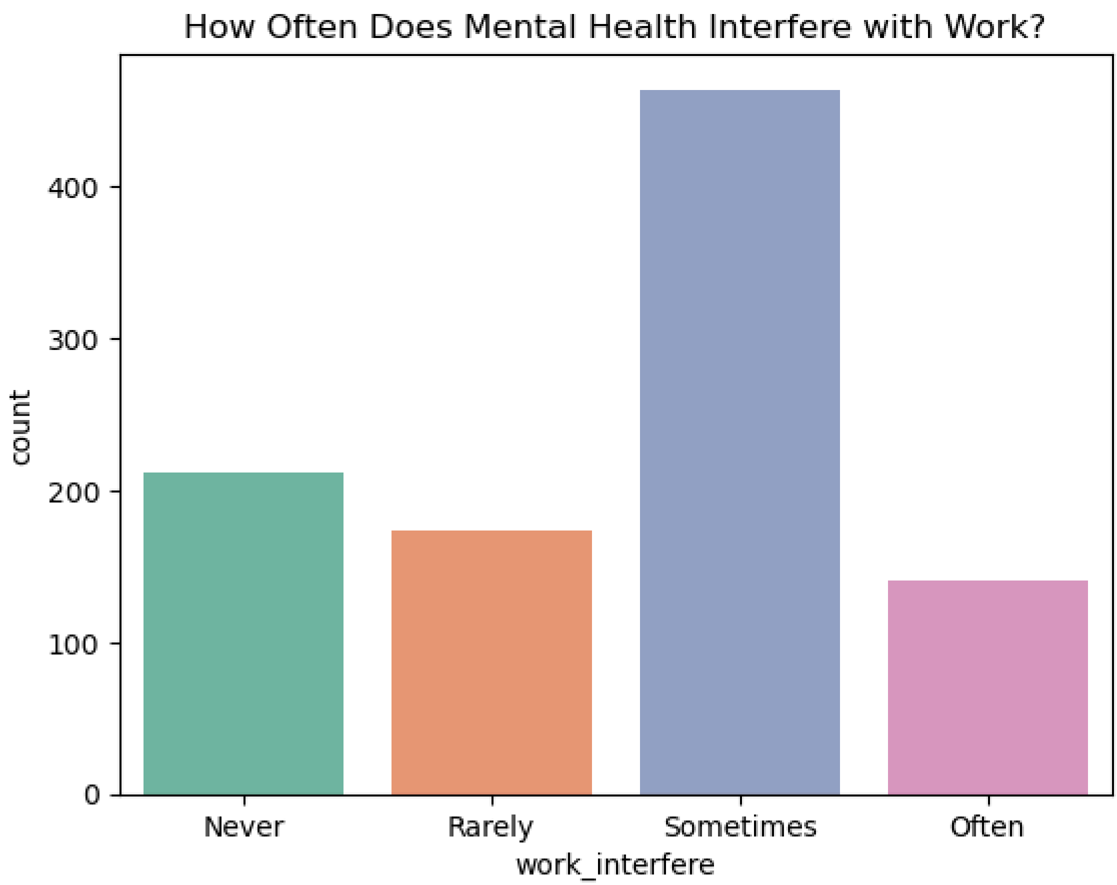
Company Size vs Mental Health Benefits

In [20]:
```
sns.countplot(data=df, x='no_employees', hue='benefits', palette='Set3')
plt.xticks(rotation=45)
plt.title('Company Size vs Mental Health Benefits')
plt.show()
```



Work Interference due to Mental Health

In [21]:
```
sns.countplot(data=df, x='work_interfere', order=['Never','Rarely','Sometimes','
plt.title('How Often Does Mental Health Interfere with Work?')
plt.show()
```



What did you know about your dataset?

The dataset is a survey conducted to understand the mental health status, awareness, and workplace support among employees working in the tech industry. It contains responses from individuals across different countries, age groups, genders, and company sizes.

Key things observed from the dataset:

It includes both demographic information (like age, gender, country) and workplace factors (like mental health benefits, remote work, anonymity).

A significant number of people reported having a family history of mental illness.

There is a noticeable difference in how genders and company types handle mental health issues.

Remote work, company size, and availability of benefits play an important role in whether individuals seek treatment or feel supported.

Some data cleaning was required, especially for inconsistent or unrealistic entries in the Age and Gender columns

2. Understanding Your Variables

In [24]:
```
# Display all column names
df.columns.tolist()
```

Out[24]:
```
['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']
```

In [25]:
```
df.describe()
```

Out[25]:

	Age
count	1251.000000
mean	32.076739
std	7.288272
min	18.000000
25%	27.000000
50%	31.000000
75%	36.000000
max	72.000000

Variables Description

Below is a description of the key variables in the dataset:

Column Name	Description
`Age`	Age of the respondent
`Gender`	Gender identity of the respondent

Column Name	Description
`Country`	Country where the respondent resides
`self_employed`	Whether the respondent is self-employed
`family_history`	Whether the respondent has a family history of mental illness
`treatment`	Whether the respondent has sought treatment for a mental health condition
`work_interfere`	How often mental health interferes with work
`no_employees`	Size of the company the respondent works for
`remote_work`	Whether the respondent works remotely at least 50% of the time
`tech_company`	Whether the employer is primarily a tech company
`benefits`	Whether the employer provides mental health benefits
`care_options`	Whether the employee knows about mental health care options offered
`wellness_program`	Whether the employer has a wellness program
`seek_help`	Whether the employer offers resources to seek help for mental health issues
`anonymity`	Whether the respondent's anonymity is protected when using mental health resources
`leave`	Ease of taking medical leave for mental health
`mental_health_consequence`	Perceived consequence of discussing mental health with employer
`phys_health_consequence`	Perceived consequence of discussing physical health with employer
`coworkers`	Willingness to discuss mental health with coworkers
`supervisor`	Willingness to discuss mental health with supervisor
`mental_health_interview`	Would discuss mental health in a job interview
`phys_health_interview`	Would discuss physical health in a job interview
`mental_vs_physical`	Whether employer treats mental and physical health equally
`obs_consequence`	Observed consequences of coworkers discussing mental health

Check Unique Values for each variable.

In [26]:
```
# Count unique values for each column
df.nunique().sort_values(ascending=False)
```

Out[26]:
```
Country                      46
Age                          45
state                        45
no_employees                  6
leave                         5
work_interfere                4
anonymity                     3
mental_vs_physical            3
phys_health_interview         3
mental_health_interview       3
supervisor                    3
coworkers                     3
phys_health_consequence       3
mental_health_consequence     3
care_options                  3
seek_help                     3
wellness_program              3
benefits                      3
Gender                        2
tech_company                  2
remote_work                   2
treatment                     2
family_history                2
self_employed                 2
obs_consequence               2
dtype: int64
```

In [27]:
```python
# Display unique values for each column (optional)
for col in df.columns:
    print(f"\n{col}:\n{df[col].unique()}")
```

```
Age:
[37 44 32 31 33 35 39 42 23 29 36 27 46 41 34 30 40 38 50 24 18 28 26 22
 19 25 45 21 43 56 60 54 55 48 20 57 58 47 62 51 65 49 53 61 72]

Gender:
['Male' 'Other']

Country:
['United States' 'Canada' 'United Kingdom' 'Bulgaria' 'France' 'Portugal'
 'Netherlands' 'Switzerland' 'Poland' 'Australia' 'Germany' 'Russia'
 'Mexico' 'Brazil' 'Slovenia' 'Costa Rica' 'Austria' 'Ireland' 'India'
 'South Africa' 'Italy' 'Sweden' 'Colombia' 'Latvia' 'Romania' 'Belgium'
 'New Zealand' 'Spain' 'Finland' 'Uruguay' 'Israel'
 'Bosnia and Herzegovina' 'Hungary' 'Singapore' 'Japan' 'Nigeria'
 'Croatia' 'Norway' 'Thailand' 'Denmark' 'Greece' 'Moldova' 'Georgia'
 'China' 'Czech Republic' 'Philippines']

state:
['IL' 'IN' nan 'TX' 'TN' 'MI' 'OH' 'CA' 'CT' 'MD' 'NY' 'NC' 'MA' 'IA' 'PA'
 'WA' 'WI' 'UT' 'NM' 'OR' 'FL' 'MN' 'MO' 'AZ' 'CO' 'GA' 'DC' 'NE' 'WV'
 'OK' 'KS' 'VA' 'NH' 'KY' 'AL' 'NV' 'NJ' 'SC' 'VT' 'SD' 'ID' 'MS' 'RI'
 'WY' 'LA' 'ME']

self_employed:
[nan 'Yes' 'No']

family_history:
['No' 'Yes']

treatment:
['Yes' 'No']

work_interfere:
['Often' 'Rarely' 'Never' 'Sometimes' nan]

no_employees:
['6-25' 'More than 1000' '26-100' '100-500' '1-5' '500-1000']

remote_work:
['No' 'Yes']

tech_company:
['Yes' 'No']

benefits:
['Yes' "Don't know" 'No']

care_options:
['Not sure' 'No' 'Yes']

wellness_program:
['No' "Don't know" 'Yes']

seek_help:
['Yes' "Don't know" 'No']

anonymity:
['Yes' "Don't know" 'No']

leave:
```

```
['Somewhat easy' "Don't know" 'Somewhat difficult' 'Very difficult'
 'Very easy']

mental_health_consequence:
['No' 'Maybe' 'Yes']

phys_health_consequence:
['No' 'Yes' 'Maybe']

coworkers:
['Some of them' 'No' 'Yes']

supervisor:
['Yes' 'No' 'Some of them']

mental_health_interview:
['No' 'Yes' 'Maybe']

phys_health_interview:
['Maybe' 'No' 'Yes']

mental_vs_physical:
['Yes' "Don't know" 'No']

obs_consequence:
['No' 'Yes']
```

3. *Data Wrangling*

Data Wrangling Code

In [28]:
```
# 1. Fill missing values for selected columns

# Filling based on common sense or most frequent answers
df['self_employed'].fillna('No', inplace=True)
df['work_interfere'].fillna("Don't know", inplace=True)

# 2. Normalize text columns (already cleaned Gender earlier)

# Strip whitespace and standardize values for binary columns
binary_cols = ['family_history', 'treatment', 'remote_work', 'tech_company',
               'benefits', 'care_options', 'wellness_program', 'seek_help',
               'anonymity', 'mental_health_consequence', 'phys_health_consequenc
               'coworkers', 'supervisor', 'mental_health_interview',
               'phys_health_interview', 'mental_vs_physical', 'obs_consequence']

for col in binary_cols:
    df[col] = df[col].str.strip().str.capitalize()

# 3. Optional: Replace missing values with 'Not specified' in less critical colu
fill_with_not_specified = ['leave']
for col in fill_with_not_specified:
    df[col].fillna('Not specified', inplace=True)

# 4. Confirm changes
df.isnull().sum()
```

Out[28]:

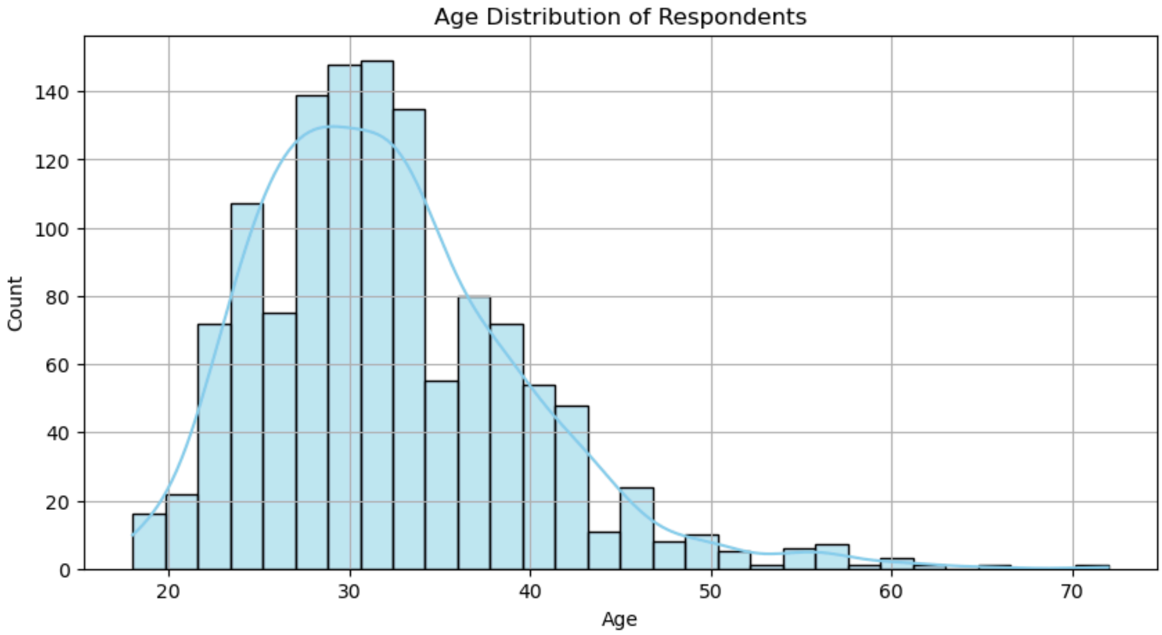
```
Age                           0
Gender                        0
Country                       0
state                       513
self_employed                 0
family_history                0
treatment                     0
work_interfere                0
no_employees                  0
remote_work                   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
```

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

Chart - 1

In [29]:

```
# Chart - 1: Age Distribution
plt.figure(figsize=(10,5))
sns.histplot(df['Age'], bins=30, kde=True, color='skyblue')
plt.title('Age Distribution of Respondents')
plt.xlabel('Age')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```



1. Why did you pick the specific chart?

I chose a histogram because `Age` is a continuous numerical variable, and a histogram is ideal for understanding the distribution and frequency of numeric data across different ranges.

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

Most respondents are in the age range of 25 to 35, indicating that young adults dominate the tech workforce represented in this survey. Very few participants are older than 60 or younger than 20 (after cleaning).

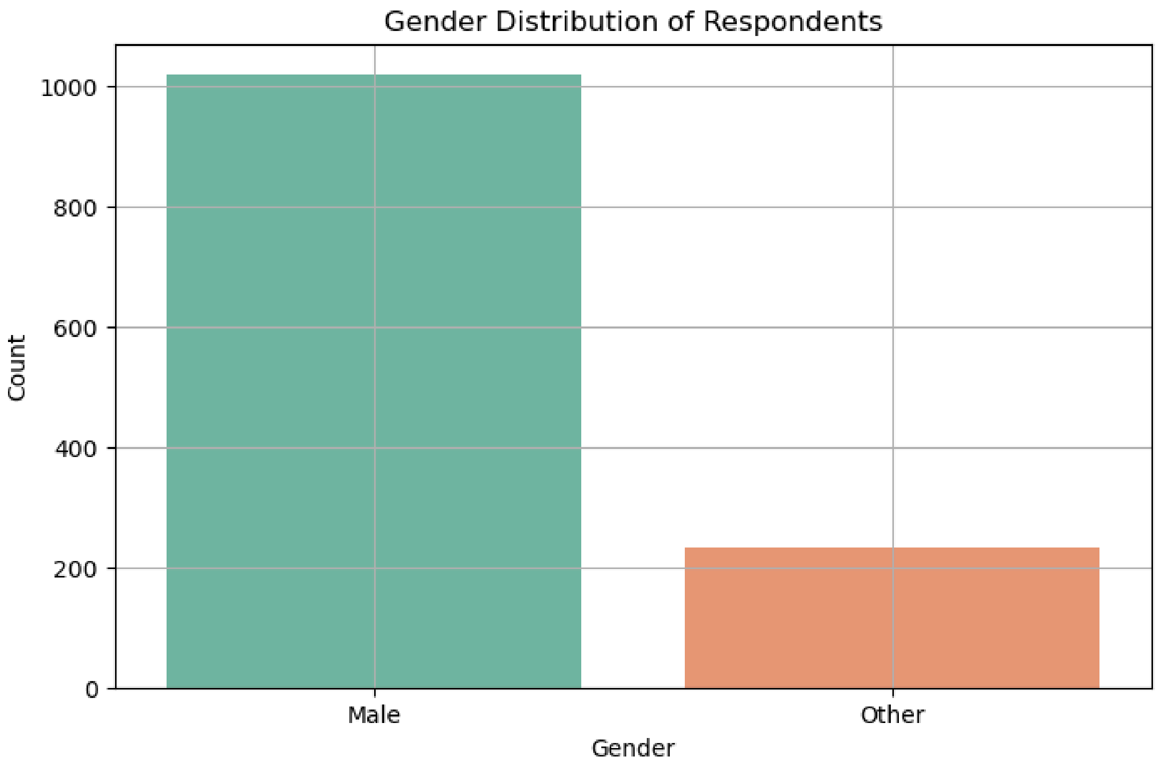
3. Will the gained insights help creating a positive business impact?

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

Yes, these insights help tailor mental health support to the age group most represented in the industry. Companies can focus mental health programs around concerns common in young adults such as burnout, imposter syndrome, and work-life balance.
There are no negative growth indicators directly from this chart, but the lack of diversity in age range suggests companies might need to ensure inclusion of older professionals.

Chart - 2

In [30]:
```
# Chart - 2: Gender Distribution
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='Gender', palette='Set2')
plt.title('Gender Distribution of Respondents')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```



1. Why did you pick the specific chart?

I used a countplot because `Gender` is a categorical variable, and a countplot effectively shows the frequency of each category (Male, Female, Other).

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

The majority of respondents are male, followed by female and a small number identifying as "Other". This reflects the gender imbalance in the tech industry, where male professionals dominate.

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

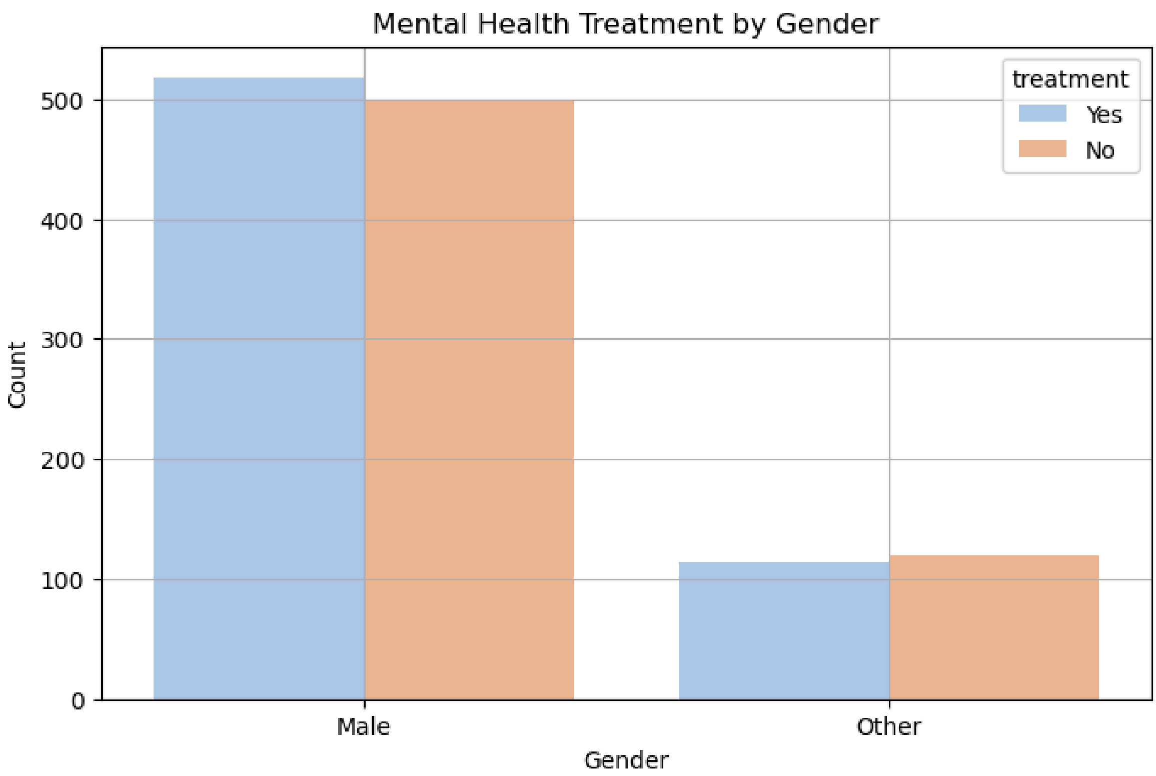
Are there any insights that lead to negative growth? Justify with specific reason.

Yes, these insights are critical for designing inclusive mental health policies. Understanding the gender distribution helps companies tailor mental health resources and communication.
The insight also highlights a negative trend — underrepresentation of women and non-binary individuals, which may indicate deeper workplace culture issues or lack of inclusivity. Addressing this gap can improve retention and diversity, both of which are tied to long-term business growth.

Chart - 3

In [31]:
```
# Chart 3: Mental Health Treatment by Gender
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='Gender', hue='treatment', palette='pastel')
plt.title('Mental Health Treatment by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
```

```
plt.grid(True)
plt.show()
```



1. Why did you pick the specific chart?

Answer Here.

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

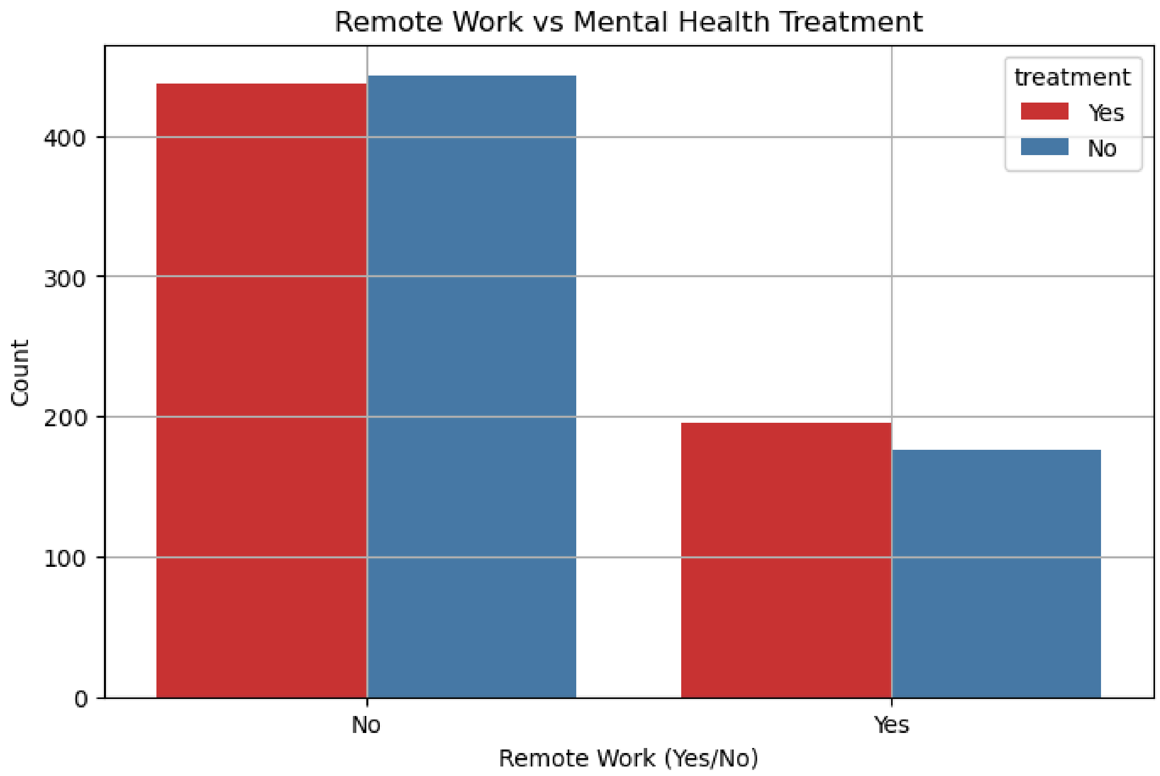
While more males responded to the survey overall, the proportion of females who sought treatment is relatively higher. The "Other" gender category also shows a high tendency to seek treatment, though the sample size is small.

In [ ]:
```
##### 3. Will the gained insights help creating a positive business impact?
 Yes, these insights help in designing gender-sensitive mental health support. U
   The only negative insight is the possible stigma among male employees, which
```

Answer Here

Chart - 4

In [32]:
```
# Chart 4: Remote Work vs Treatment
plt.figure(figsize=(8,5))
sns.countplot(data=df, x='remote_work', hue='treatment', palette='Set1')
plt.title('Remote Work vs Mental Health Treatment')
plt.xlabel('Remote Work (Yes/No)')
plt.ylabel('Count')
plt.grid(True)
plt.show()
```



1. Why did you pick the specific chart?

I chose a countplot with a hue because it allows us to compare two categorical variables: whether someone works remotely and whether they have sought treatment.

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

People who work remotely appear to seek mental health treatment slightly more than those who do not. This could be due to isolation or a flexible work culture that encourages seeking help.

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

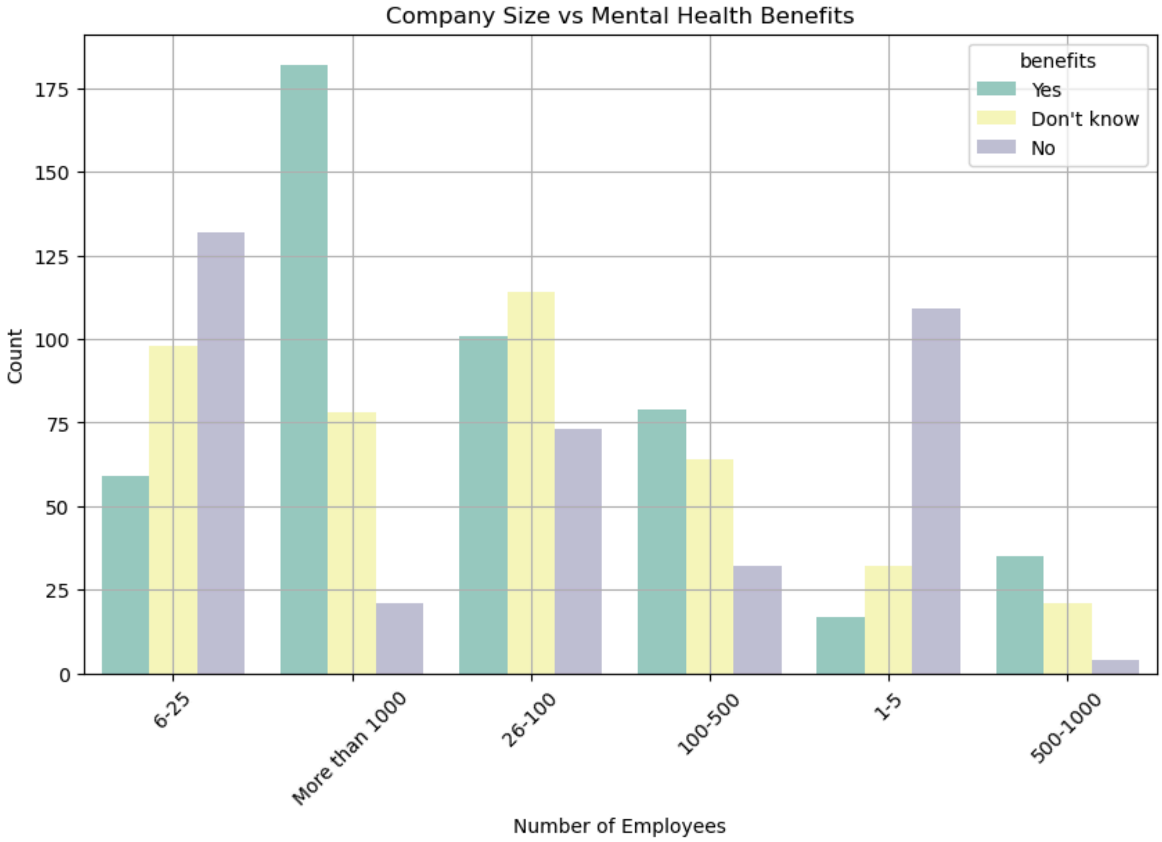
Are there any insights that lead to negative growth? Justify with specific reason.

These insights support the idea that remote work can promote help-seeking behavior if managed well. Companies can leverage this by offering mental health resources to remote employees.
However, unmanaged remote work can lead to isolation, so it must be paired with mental wellness checks to avoid negative consequences.

Chart - 5

In [33]:
```
# Chart 5: Company Size vs Mental Health Benefits
plt.figure(figsize=(10,6))
sns.countplot(data=df, x='no_employees', hue='benefits', palette='Set3')
plt.xticks(rotation=45)
plt.title('Company Size vs Mental Health Benefits')
plt.xlabel('Number of Employees')
plt.ylabel('Count')
```

```
plt.grid(True)
plt.show()
```



1. Why did you pick the specific chart?

A grouped countplot is ideal to compare mental health benefits across different company sizes.

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

Larger companies (500+ employees) are more likely to offer mental health benefits, while smaller companies (especially those with 6–25 employees) show limited or no benefits.

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, these insights show that startups and small companies need more support or awareness to implement mental health programs. HR policies in large companies are more established.
A lack of benefits in small firms could lead to employee burnout and higher attrition, affecting growth. Encouraging benefit programs regardless of company size can improve well-being and retention.

Summary of Key Insights

In [ ]:
```
- Most participants are aged 25-35.
- Female respondents are more likely to seek treatment.
```

```
- Remote workers report higher openness to mental health support.
- Smaller companies are less likely to offer mental health benefits.
```

Recommendations

In [ ]:
```
- Promote mental health awareness in small organizations.
- Protect employee anonymity while seeking help.
- Encourage wellness programs across all company sizes.
```

Conclusion

This EDA helps understand how workplace factors, demographics, and company policies affect mental health awareness and treatment in the tech industry.

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

In [ ]:

In [ ]: