PUBLIC HEALTH AWARNESS AND CAMPAGIN ANALYSIS... ...

INTRODUCTION:

A health awareness campaign is marketing designed to educate people about diseases and health hazards, with goals of prevention and encouraging people to take control of their health... .

The scope of public health and hygiene is broad and practices aimed at promoting and protecting the health and well-being of communities and populations...

ABSTRACT:

The science underlying the nation' s system of response to public health emergencies is seriously deficient, hampering the nation' s ability to respond to emergencies most effectively to save lives and preserve well-being humanity..

OBJECTIVES:

The goal of public health are to save money, improve the quality of life, help children thrive, and reduce human sufferings... .

```
In [1]: # I am using Pandas for data manipulation.
        # Matplotlib, Seaborn and hvPlot for Data Visualization

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

In [9]: df=pd.read_csv(r'C:\Users\Machines\Downloads\survay.csv')

In [10]: print(df)

          Timestamp   Age                                        Gender  \
0   8/27/2014 11:29  37.0                                        Female
1   8/27/2014 11:29  44.0                                             M
2   8/27/2014 11:29  32.0                                          Male
3   8/27/2014 11:29  31.0                                          Male
4   8/27/2014 11:30  31.0                                          Male
5   8/27/2014 11:31  33.0                                          Male
6   8/27/2014 11:31  35.0                                        Female
7   8/27/2014 11:32  39.0                                             M
8   8/27/2014 11:32  42.0                                        Female
9   8/27/2014 11:32  23.0                                          Male
10  8/27/2014 11:32  31.0                                          Male
11  8/27/2014 11:32  29.0                                          male
12  8/27/2014 11:33  42.0                                        female
13  8/27/2014 11:33  36.0                                          Male
14  8/27/2014 11:33  27.0                                          Male
15  8/27/2014 11:34  29.0                                        female
16  8/27/2014 11:34  23.0                                          Male
17  8/27/2014 11:34  32.0                                          Male
18  8/27/2014 11:34  46.0                                          male
19  8/27/2014 11:35  36.0                                          Male
20  8/27/2014 11:35  29.0                                          Male
21  8/27/2014 11:35  31.0                                          male
22  8/27/2014 11:35  46.0                                          Male
23  8/27/2014 11:36  41.0                                          Male
```

```
24     8/27/2014 11:36  33.0                                       male
25     8/27/2014 11:37  35.0                                       male
26     8/27/2014 11:37  33.0                                       male
27     8/27/2014 11:37  35.0                                     Female
28     8/27/2014 11:38  34.0                                       male
29     8/27/2014 11:38  37.0                                       Male
...                ...   ...                                        ...
1230    2/22/2015 2:40  39.0                                       Male
1231    2/24/2015 8:54  23.0                                     Female
1232    2/24/2015 8:58  24.0                                    Cis Man
1233    2/24/2015 9:00  25.0                                       Male
1234    2/24/2015 9:13  23.0                                       Male
1235    2/24/2015 9:15  24.0  ostensibly male, unsure what that really means
1236    2/24/2015 9:18  23.0                                       Male
1237   2/24/2015 10:32  60.0                                       Male
1238    2/26/2015 5:44  28.0                                       male
1239    4/2/2015 15:47  28.0                                       Male
1240    4/4/2015 11:22  30.0                                       male
1241    4/6/2015 14:58  31.0                                       male
1242   4/11/2015 14:35  31.0                                       Male
1243   4/23/2015 14:03  28.0                                       Male
1244    5/5/2015 14:22  43.0                                          f
1245    5/5/2015 15:16  32.0                                     female
1246    5/6/2015 10:14  22.0                                       Male
1247    5/6/2015 16:55  32.0                                       Male
1248    5/7/2015 10:08  36.0                                       male
1249   6/25/2015 12:24  41.0                                     Female
1250   7/22/2015 18:57  30.0                                          M
1251   7/27/2015 23:25  30.0                                       Male
1252    8/17/2015 9:38  36.0                                       Male
1253   8/20/2015 16:52  29.0                                       male
1254   8/25/2015 19:59  36.0                                       Male
1255   9/12/2015 11:17  26.0                                       male
1256    9/26/2015 1:07  32.0                                       Male
1257   11/7/2015 12:36  34.0                                       male
1258  11/30/2015 21:25  46.0                                          f
1259    2/1/2016 23:04  25.0                                       Male

             Country state self_employed family_history treatment  \
0      United States    IL           NaN             No       Yes
1      United States    IN           NaN             No        No
2             Canada   NaN           NaN             No        No
3     United Kingdom   NaN           NaN            Yes       Yes
4      United States    TX           NaN             No        No
5      United States    TN           NaN            Yes        No
6      United States    MI           NaN            Yes       Yes
7             Canada   NaN           NaN             No        No
8      United States    IL           NaN            Yes       Yes
```

9	Canada	NaN	NaN	No	No
10	United States	OH	NaN	No	Yes
11	Bulgaria	NaN	NaN	No	No
12	United States	CA	NaN	Yes	Yes
13	United States	CT	NaN	Yes	No
14	Canada	NaN	NaN	No	No
15	United States	IL	NaN	Yes	Yes
16	United Kingdom	NaN	NaN	No	Yes
17	United States	TN	NaN	No	Yes
18	United States	MD	Yes	Yes	No
19	France	NaN	Yes	Yes	No
20	United States	NY	No	Yes	Yes
21	United States	NC	Yes	No	No
22	United States	MA	No	No	Yes
23	United States	IA	No	No	Yes
24	United States	CA	No	Yes	Yes
25	United States	TN	No	Yes	Yes
26	United States	TN	No	No	No
27	United States	CA	No	Yes	Yes
28	United States	OH	No	No	Yes
29	United Kingdom	NaN	No	No	No
...	...	...	...	...	...
1230	Greece	NaN	No	No	No
1231	United Kingdom	NaN	No	Yes	Yes
1232	United Kingdom	NaN	Yes	No	Yes
1233	United Kingdom	NaN	No	Yes	Yes
1234	United Kingdom	NaN	No	No	Yes
1235	United Kingdom	NaN	No	No	Yes
1236	Canada	NaN	No	No	Yes
1237	United States	CA	No	No	Yes
1238	Ireland	NaN	No	No	No
1239	United States	TN	No	Yes	Yes
1240	Netherlands	NaN	No	No	No
1241	Germany	NaN	No	Yes	Yes
1242	Poland	NaN	Yes	No	Yes
1243	Ireland	NaN	No	No	Yes
1244	United States	FL	No	Yes	Yes
1245	United Kingdom	NaN	No	No	No
1246	Australia	NaN	No	Yes	Yes
1247	United States	OR	No	No	No
1248	Finland	NaN	No	No	Yes
1249	United States	WA	No	Yes	Yes
1250	United States	CA	No	Yes	Yes
1251	United States	CA	Yes	Yes	Yes
1252	South Africa	NaN	No	Yes	Yes
1253	United States	NC	No	Yes	Yes
1254	United States	UT	No	Yes	No
1255	United Kingdom	NaN	No	No	Yes

```
1256  United States    IL           No           Yes        Yes
1257  United States    CA           No           Yes        Yes
1258  United States    NC           No            No         No
1259  United States    IL           No           Yes        Yes

     work_interfere    no_employees  \
0             Often          25-Jun
1            Rarely  More than 1000
2            Rarely          25-Jun
3             Often          26-100
4             Never         100-500
5         Sometimes          25-Jun
6         Sometimes           5-Jan
7             Never           5-Jan
8         Sometimes         100-500
9             Never          26-100
10        Sometimes          25-Jun
11            Never         100-500
12        Sometimes          26-100
13            Never        500-1000
14            Never          25-Jun
15           Rarely          26-100
16        Sometimes          26-100
17        Sometimes          25-Jun
18        Sometimes           5-Jan
19              NaN          25-Jun
20        Sometimes         100-500
21            Never           5-Jan
22            Often          26-100
23            Never  More than 1000
24           Rarely          26-100
25        Sometimes  More than 1000
26              NaN           5-Jan
27           Rarely          25-Jun
28        Sometimes          26-100
29        Sometimes          25-Jun
...             ...             ...
1230            NaN          25-Jun
1231      Sometimes          25-Jun
1232      Sometimes          25-Jun
1233      Sometimes  More than 1000
1234         Rarely          25-Jun
1235      Sometimes          25-Jun
1236          Often          26-100
1237          Often  More than 1000
1238      Sometimes          26-100
1239          Often  More than 1000
1240      Sometimes        500-1000
```

```
1241      Sometimes          100-500
1242          Often           25-Jun
1243         Rarely           26-100
1244         Rarely   More than 1000
1245            NaN   More than 1000
1246          Often          100-500
1247          Never          100-500
1248          Often           25-Jun
1249      Sometimes           26-100
1250      Sometimes           26-100
1251          Often           26-100
1252          Often          100-500
1253      Sometimes          100-500
1254         Rarely   More than 1000
1255            NaN           26-100
1256          Often           26-100
1257      Sometimes   More than 1000
1258            NaN          100-500
1259      Sometimes           26-100

                              ...                              leave  \
0                             ...                      Somewhat easy
1                             ...                         Don't know
2                             ...                 Somewhat difficult
3                             ...                 Somewhat difficult
4                             ...                         Don't know
5                             ...                         Don't know
6                             ...                 Somewhat difficult
7                             ...                         Don't know
8                             ...                     Very difficult
9                             ...                         Don't know
10                            ...                         Don't know
11                            ...                         Don't know
12                            ...                 Somewhat difficult
13                            ...                         Don't know
14                            ...                      Somewhat easy
15                            ...                      Somewhat easy
16                            ...                          Very easy
17                            ...                         Don't know
18                            ...                          Very easy
19                            ...                      Somewhat easy
20                            ...                 Somewhat difficult
21                            ...                 Somewhat difficult
22                            ...                         Don't know
23                            ...                         Don't know
24                            ...                         Don't know
25                            ...                          Very easy
26                            ...                         Don't know
```

```
27                                 ...                    Don't know
28                                 ...            Somewhat difficult
29                                 ...                Very difficult
...                                ...                           ...
1230                               ...                    Don't know
1231                               ...                     Very easy
1232                               ...                    Don't know
1233                               ...                     Very easy
1234                               ...                    Don't know
1235                               ...                    Don't know
1236                               ...                    Don't know
1237                               ...                 Somewhat easy
1238                               ...                    Don't know
1239                               ...                 Somewhat easy
1240                               ...                    Don't know
1241                               ...                 Somewhat easy
1242                               ...                 Somewhat easy
1243                               ...                    Don't know
1244                               ...                    Don't know
1245                               ...                    Don't know
1246                               ...                    Don't know
1247                               ...                 Somewhat easy
1248                               ...                Very difficult
1249                               ...                    Don't know
1250                               ...                     Very easy
1251                               ...                    Don't know
1252                               ...                 Somewhat easy
1253                               ...                    Don't know
1254                               ...                 Somewhat easy
1255                               ...                 Somewhat easy
1256                               ...            Somewhat difficult
1257                               ...            Somewhat difficult
1258                               ...                    Don't know
1259                               ...                    Don't know

     mental_health_consequence phys_health_consequence     coworkers  \
0                           No                      No  Some of them
1                        Maybe                      No            No
2                           No                      No           Yes
3                          Yes                     Yes  Some of them
4                           No                      No  Some of them
5                           No                      No           Yes
6                        Maybe                   Maybe  Some of them
7                           No                      No            No
8                        Maybe                      No           Yes
9                           No                      No           Yes
10                          No                      No  Some of them
11                          No                      No           Yes
```

12	Yes	Yes	Yes
13	No	No	Yes
14	No	No	Some of them
15	No	No	Yes
16	Maybe	No	Some of them
17	Maybe	No	Some of them
18	No	No	Yes
19	No	No	Some of them
20	Maybe	No	Some of them
21	No	No	Some of them
22	Maybe	No	Some of them
23	Maybe	No	No
24	No	No	Yes
25	Yes	No	Some of them
26	Maybe	Maybe	Some of them
27	No	No	Yes
28	No	No	Some of them
29	Yes	Maybe	Some of them
...	...	...	...
1230	Yes	No	No
1231	No	No	Yes
1232	Maybe	Maybe	Some of them
1233	No	No	Yes
1234	No	No	Yes
1235	Yes	Maybe	No
1236	Maybe	No	Yes
1237	Maybe	Maybe	Some of them
1238	Yes	Maybe	No
1239	Yes	Maybe	Some of them
1240	Maybe	No	Yes
1241	No	No	Some of them
1242	Maybe	No	Some of them
1243	Maybe	No	No
1244	No	No	Some of them
1245	Maybe	No	Some of them
1246	Maybe	Maybe	No
1247	No	No	No
1248	Yes	No	Some of them
1249	Yes	Maybe	No
1250	No	No	Yes
1251	No	No	Some of them
1252	No	No	Some of them
1253	Yes	No	Some of them
1254	Maybe	Maybe	Some of them
1255	No	No	Some of them
1256	No	No	Some of them
1257	Yes	Yes	No
1258	Yes	No	No

```
1259                     Maybe                      No  Some of them

       supervisor mental_health_interview phys_health_interview  \
0             Yes                      No                 Maybe
1              No                      No                    No
2             Yes                     Yes                   Yes
3              No                   Maybe                 Maybe
4             Yes                     Yes                   Yes
5             Yes                      No                 Maybe
6              No                      No                    No
7              No                      No                    No
8             Yes                      No                 Maybe
9             Yes                   Maybe                 Maybe
10            Yes                      No                    No
11            Yes                     Yes                   Yes
12            Yes                   Maybe                 Maybe
13            Yes                      No                    No
14   Some of them                   Maybe                   Yes
15   Some of them                   Maybe                 Maybe
16             No                   Maybe                 Maybe
17            Yes                      No                    No
18            Yes                      No                   Yes
19   Some of them                   Maybe                 Maybe
20   Some of them                      No                    No
21   Some of them                      No                 Maybe
22            Yes                      No                 Maybe
23             No                      No                   Yes
24            Yes                      No                   Yes
25            Yes                      No                   Yes
26             No                      No                    No
27            Yes                   Maybe                 Maybe
28             No                      No                    No
29             No                      No                 Maybe
...           ...                     ...                   ...
1230           No                      No                    No
1231          Yes                   Maybe                 Maybe
1232 Some of them                      No                 Maybe
1233          Yes                     Yes                   Yes
1234          Yes                      No                 Maybe
1235           No                      No                 Maybe
1236 Some of them                      No                    No
1237           No                      No                 Maybe
1238           No                      No                    No
1239          Yes                      No                    No
1240          Yes                   Maybe                   Yes
1241           No                      No                    No
1242           No                      No                    No
1243           No                      No                 Maybe
```

```
1244           Yes                      No                      No
1245           Yes                      No                     Yes
1246           Yes                      No                   Maybe
1247  Some of them                   Maybe                     Yes
1248            No                      No                   Maybe
1249            No                      No                      No
1250           Yes                   Maybe                   Maybe
1251           Yes                   Maybe                   Maybe
1252           Yes                      No                     Yes
1253            No                      No                   Maybe
1254  Some of them                      No                      No
1255  Some of them                      No                      No
1256           Yes                      No                      No
1257            No                      No                      No
1258            No                      No                      No
1259            No                      No                      No

     mental_vs_physical obs_consequence  \
0                   Yes              No
1            Don't know              No
2                    No              No
3                    No             Yes
4            Don't know              No
5            Don't know              No
6            Don't know              No
7                    No              No
8                    No              No
9                   Yes              No
10           Don't know              No
11           Don't know              No
12                   No             Yes
13           Don't know              No
14                  Yes              No
15           Don't know              No
16                   No              No
17                   No              No
18                  Yes             Yes
19           Don't know              No
20                   No              No
21                  Yes              No
22                   No              No
23           Don't know              No
24           Don't know              No
25                   No              No
26           Don't know              No
27                  Yes              No
28                   No              No
29                   No              No
```

```
...              ...          ...
1230       Don't know          No
1231              Yes          No
1232       Don't know         Yes
1233              Yes          No
1234       Don't know          No
1235       Don't know          No
1236       Don't know          No
1237       Don't know          No
1238               No         Yes
1239               No         Yes
1240       Don't know          No
1241       Don't know          No
1242       Don't know          No
1243       Don't know          No
1244       Don't know          No
1245               No          No
1246       Don't know         Yes
1247       Don't know          No
1248       Don't know         Yes
1249       Don't know          No
1250              Yes          No
1251              Yes          No
1252              Yes          No
1253               No          No
1254       Don't know          No
1255       Don't know          No
1256              Yes          No
1257               No          No
1258               No          No
1259       Don't know          No

                                               comments
0                                                   NaN
1                                                   NaN
2                                                   NaN
3                                                   NaN
4                                                   NaN
5                                                   NaN
6                                                   NaN
7                                                   NaN
8                                                   NaN
9                                                   NaN
10                                                  NaN
11                                                  NaN
12                                                  NaN
13    I'm not on my company's health insurance which...
14                                                  NaN
```

```
15      I have chronic low-level neurological issues t...
16      My company does provide healthcare but not to ...
17                                                    NaN
18                                                    NaN
19                                                    NaN
20                                                    NaN
21                                                    NaN
22                                                    NaN
23                                                    NaN
24                    Relatively new job. Ask again later
25      Sometimes I think  about using drugs for my me...
26                                                    NaN
27                                                    NaN
28                                                    NaN
29                                                    NaN
...                                                   ...
1230                                                  NaN
1231                                                  NaN
1232                                                  NaN
1233    I work at a large university with a track reco...
1234                                                  NaN
1235    i'm in a country with social health care so my...
1236                                                  NaN
1237                                                  NaN
1238                                                  NaN
1239                                                  NaN
1240                                                  NaN
1241                                                  NaN
1242                                                  NaN
1243                                                  NaN
1244                                                  NaN
1245                                                  NaN
1246    In australia all organisations of a certain si...
1247                                                  NaN
1248                                                  NaN
1249                                                  NaN
1250                                     Bipolar disorder
1251                                                  NaN
1252                                                  NaN
1253                                                  NaN
1254                                                  NaN
1255                                                  NaN
1256                                                  NaN
1257                                                  NaN
1258                                                  NaN
1259                                                  NaN

[1260 rows x 27 columns]
```

In [11]:
```
# We seek for information about our data using .info()
# It will help us see both the missing values & data types for each attribute
# The index, columns, and number of missing values

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1260 entries, 0 to 1259
Data columns (total 27 columns):
Timestamp                    1260 non-null object
Age                          1259 non-null float64
Gender                       1259 non-null object
Country                      1259 non-null object
state                        744 non-null object
self_employed                1241 non-null object
family_history               1259 non-null object
treatment                    1259 non-null object
work_interfere               995 non-null object
no_employees                 1259 non-null object
remote_work                  1259 non-null object
tech_company                 1259 non-null object
benefits                     1259 non-null object
care_options                 1259 non-null object
wellness_program             1259 non-null object
seek_help                    1259 non-null object
anonymity                    1259 non-null object
leave                        1259 non-null object
mental_health_consequence    1259 non-null object
phys_health_consequence      1259 non-null object
coworkers                    1259 non-null object
supervisor                   1259 non-null object
mental_health_interview      1259 non-null object
phys_health_interview        1259 non-null object
mental_vs_physical           1259 non-null object
obs_consequence              1259 non-null object
comments                     164 non-null object
dtypes: float64(1), object(26)
memory usage: 265.9+ KB
```

In [12]:
```
# I can see the exact number of missing values there is using the isna().sum()
```

In [13]:
```
df.isna().sum()
```

Out[13]:
```
Timestamp                    0
Age                          1
Gender                       1
```

```
Country                          1
state                          516
self_employed                   19
family_history                   1
treatment                        1
work_interfere                 265
no_employees                     1
remote_work                      1
tech_company                     1
benefits                         1
care_options                     1
wellness_program                 1
seek_help                        1
anonymity                        1
leave                            1
mental_health_consequence        1
phys_health_consequence          1
coworkers                        1
supervisor                       1
mental_health_interview          1
phys_health_interview            1
mental_vs_physical               1
obs_consequence                  1
comments                      1096
dtype: int64
```

In [14]:
```
# From the above analyses, we see that;

# "state",
# "self_employed" and
# "work_interfere"
# "comments"

# all have missing values
```

In [15]:
```
# In the future, I will drop "state" and "comments' columns because we won't be needing
# Our analysis will be done at "country" level, since our dataset covers different coun
# So I will be filling up the missing values for just "self_employed" and "work_interfe
```

In [16]:
```
# I want to see the percentage missing values in "self_employed" and "work_interfere" c
# I will use the round() function to round them to 2 decimal places

self_employed_percent = (df["self_employed"].isnull().sum()/len(df["self_employed"]))*1
work_interfere_percent = (df["work_interfere"].isnull().sum()/len(df["work_interfere"])

print(f"The percentage of missing values in self_employed column is {round(self_employe
print(f"The percentage of missing values in work_interfere column is {round(work_interf
```

```
The percentage of missing values in self_employed column is 1.51%
```

```
The percentage of missing values in work_interfere column is 21.03%
```

```
In [17]: # Because I have approx. 80% of both, I will fill up both columns with the mode for eac

In [18]: df["self_employed"] = df["self_employed"].fillna(df["self_employed"].mode()[0])

         df["work_interfere"] = df["work_interfere"].fillna(df["work_interfere"].mode()[0])

In [19]: # I can now take a peek at our dataframe to see how it looks like, using the .head() me

In [20]: df.head()
```

```
Out[20]:          Timestamp   Age  Gender         Country state self_employed  \
         0  8/27/2014 11:29  37.0  Female   United States    IL            No
         1  8/27/2014 11:29  44.0       M   United States    IN            No
         2  8/27/2014 11:29  32.0    Male          Canada   NaN            No
         3  8/27/2014 11:29  31.0    Male  United Kingdom   NaN            No
         4  8/27/2014 11:30  31.0    Male   United States    TX            No

           family_history treatment work_interfere    no_employees   ...     \
         0             No       Yes          Often          25-Jun   ...
         1             No        No         Rarely  More than 1000   ...
         2             No        No         Rarely          25-Jun   ...
         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]
```

In [21]:
```
# Final step is to drop the "state" and "comments" columns
# And then, I will confirm that all missing values have been fixed
```

In [22]:
```
df.drop(["state", "comments"], axis=1, inplace=True)
```

In [23]:
```
df.isna().sum()
```

Out[23]:
```
Timestamp                    0
Age                          1
Gender                       1
Country                      1
self_employed                0
family_history               1
treatment                    1
work_interfere               0
no_employees                 1
remote_work                  1
tech_company                 1
benefits                     1
care_options                 1
wellness_program             1
seek_help                    1
anonymity                    1
leave                        1
mental_health_consequence    1
phys_health_consequence      1
coworkers                    1
supervisor                   1
mental_health_interview      1
phys_health_interview        1
mental_vs_physical           1
obs_consequence              1
dtype: int64
```

In [24]:
```
# Our dataframe is now free of any missing values.
```

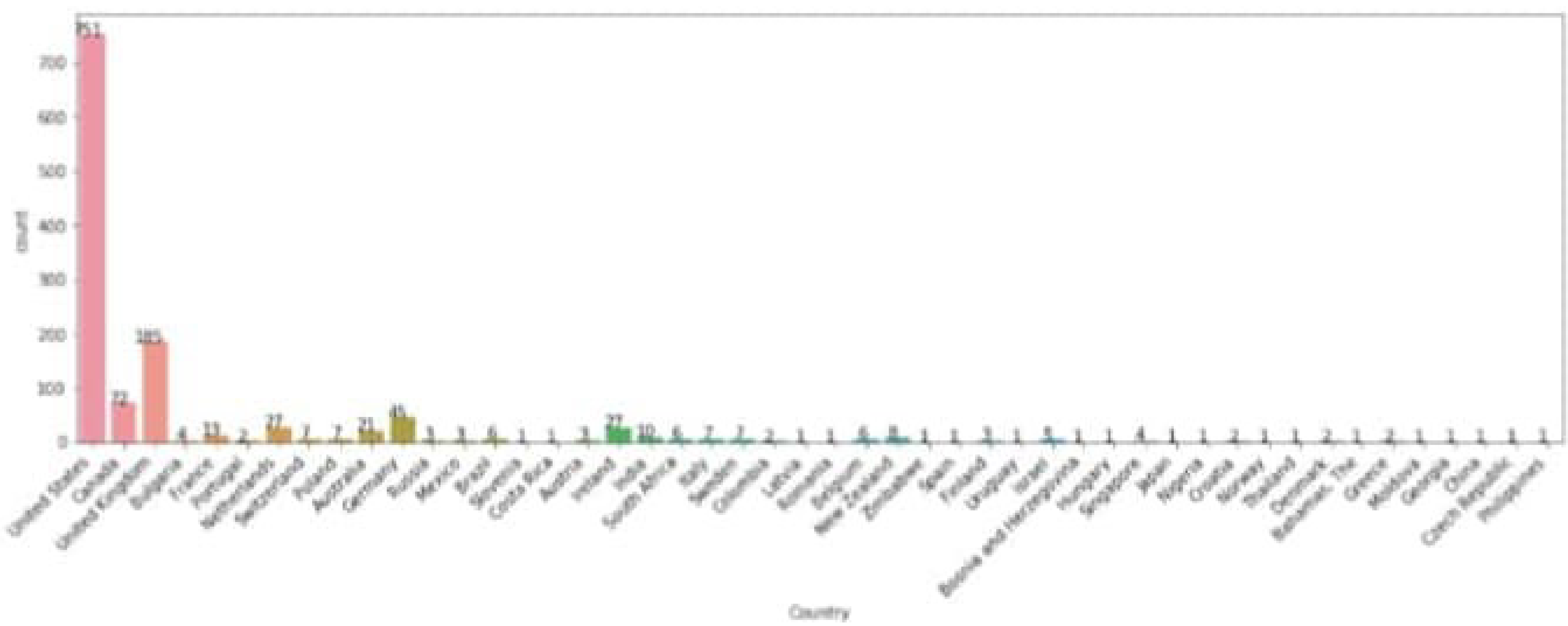
In [25]:
```
# First, let's see all our columns
```

In [26]:
```
df.columns
```

Out[26]:
```
Index(['Timestamp', 'Age', 'Gender', 'Country', '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'],
      dtype='object')
```

In [27]:
```
# First, I will look at some of the interesting attributes and analyse them
```

```
In [28]: plt.figure(figsize=(17,5))
         ax = sns.countplot(x='Country', data=df)

         ax.set_xticklabels(
             ax.get_xticklabels(),
             rotation=45,
             horizontalalignment='right'
         )
         None # So it won't show the label objects

         # Then we also display the values for each bar above it;

         for p in ax.patches:
            ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```




```
In [29]: # We clearly see above that the countries with the highest number of mental health issu

         # United States
         # United Kingdom
         # Canada

In [30]: # But why is this so?
         # I can proceed to look at other factors like "Age", "remote_work", "family_history", '

In [31]: # To analyze this, it's best I group them into categories

         # To help me get a good enough group size, I will see the min, max, median and mean of

In [32]: min_age=df["Age"].min()
         max_age=df["Age"].max()
         mean_age=df["Age"].mean()
         median_age=df["Age"].median()

         print(f"Min:  {min_age},  \nMax: {max_age}, \nMean: {mean_age}, \nMedian: {median_age}'
```

```
Min:  -1726.0,
Max: 99999999999.0,
Mean: 79428148.31135821,
Median: 31.0
```

```
In [33]: # We see from above, that even though all Age collumns are filled, some of the values c
```

```
In [34]: # By my own personal preference, I will replace all invalid values by the median, 31.
         # I chose this because I believe for people in tech today, 31 seems appropariate to use
         # I will also exclude those above 80, and treat them as invalid.
         # Then I will see if there are other outliers worth looking at.
```

```
In [35]: df["Age"].unique()
```

```
Out[35]: array([ 3.700e+01,  4.400e+01,  3.200e+01,  3.100e+01,  3.300e+01,
                 3.500e+01,  3.900e+01,  4.200e+01,  2.300e+01,  2.900e+01,
                 3.600e+01,  2.700e+01,  4.600e+01,  4.100e+01,  3.400e+01,
                 3.000e+01,  4.000e+01,  3.800e+01,  5.000e+01,  2.400e+01,
                 1.800e+01,  2.800e+01,  2.600e+01,  2.200e+01,  1.900e+01,
                 2.500e+01,  4.500e+01,  2.100e+01, -2.900e+01,  4.300e+01,
                 5.600e+01,  6.000e+01,  5.400e+01,  3.290e+02,  5.500e+01,
                 1.000e+11,  4.800e+01,  2.000e+01,  5.700e+01,  5.800e+01,
                 4.700e+01,  6.200e+01,  5.100e+01,  6.500e+01,        nan,
                 4.900e+01, -1.726e+03,  5.000e+00,  5.300e+01,  6.100e+01,
                 8.000e+00,  1.100e+01, -1.000e+00,  7.200e+01])
```

```
In [36]: # I would like to see the number of values with negative or above 80.

         negative_age = (df["Age"]<0).sum()
         over_age = (df["Age"]>80).sum()

         print(f"Number of negative age entries: {negative_age}\nNumber of overage: {over_age}")
```

```
Number of negative age entries: 3
Number of overage: 2
```

```
In [37]: # Setting them to the median age

         df.loc[df.Age<0, ["Age"]] = df["Age"].median()
         df.loc[df.Age>80, ["Age"]] = df["Age"].median()
```

```
In [38]: df["Age"].unique()
```

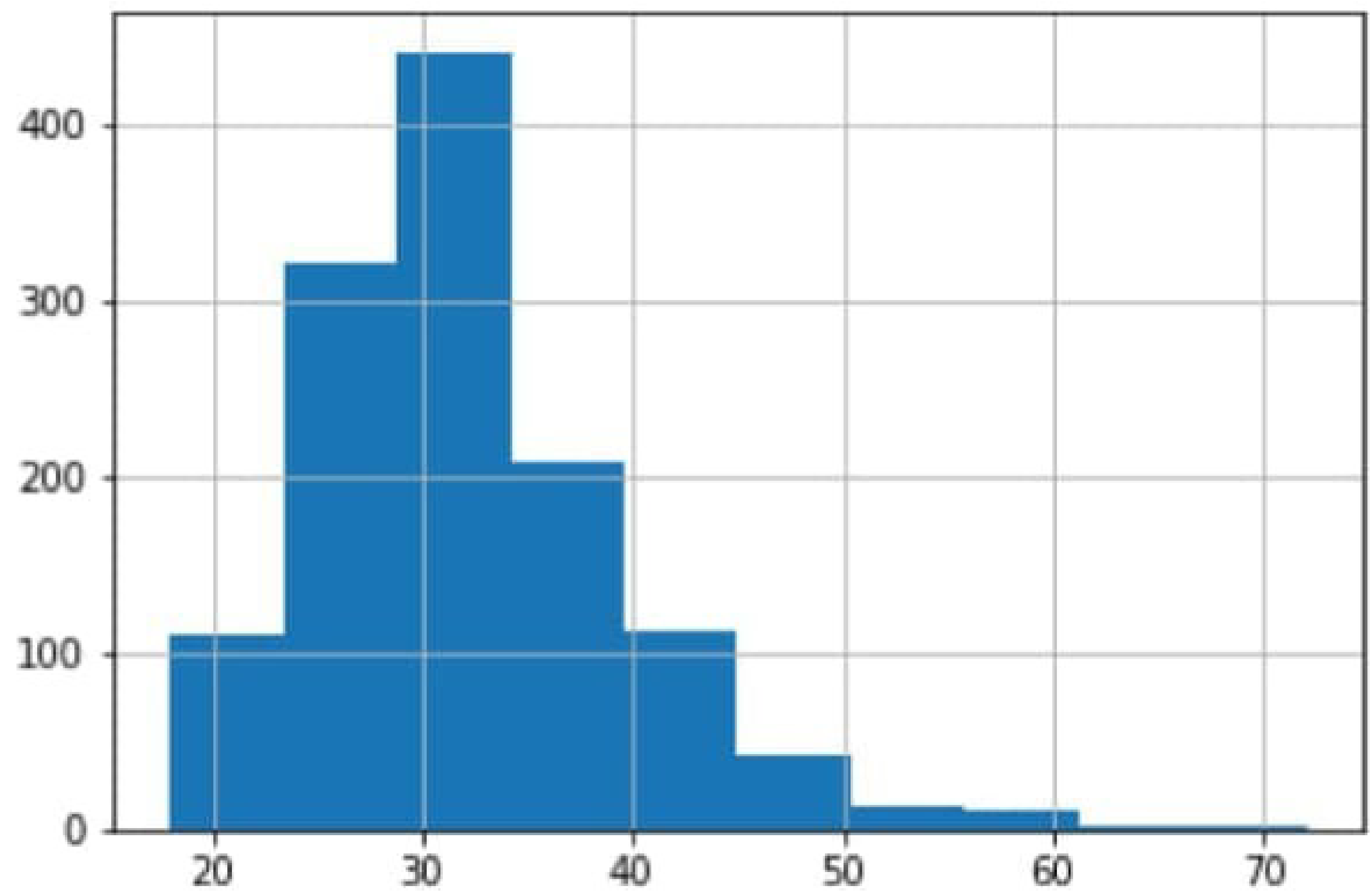
```
Out[38]: array([37., 44., 32., 31., 33., 35., 39., 42., 23., 29., 36., 27., 46.,
                41., 34., 30., 40., 38., 50., 24., 18., 28., 26., 22., 19., 25.,
                45., 21., 43., 56., 60., 54., 55., 48., 20., 57., 58., 47., 62.,
                51., 65., nan, 49.,  5., 53., 61.,  8., 11., 72.])
```

In [39]:
```
# We have outliers of 5years, 8years
# With our domain knowledge, we can confidently go further to treat the
# Only 18 and above is accepted to be a legal tech employee.
# I will replace them with the median.
```

In [40]:
```
df.loc[df.Age<18, ["Age"]] = df["Age"].median()
```

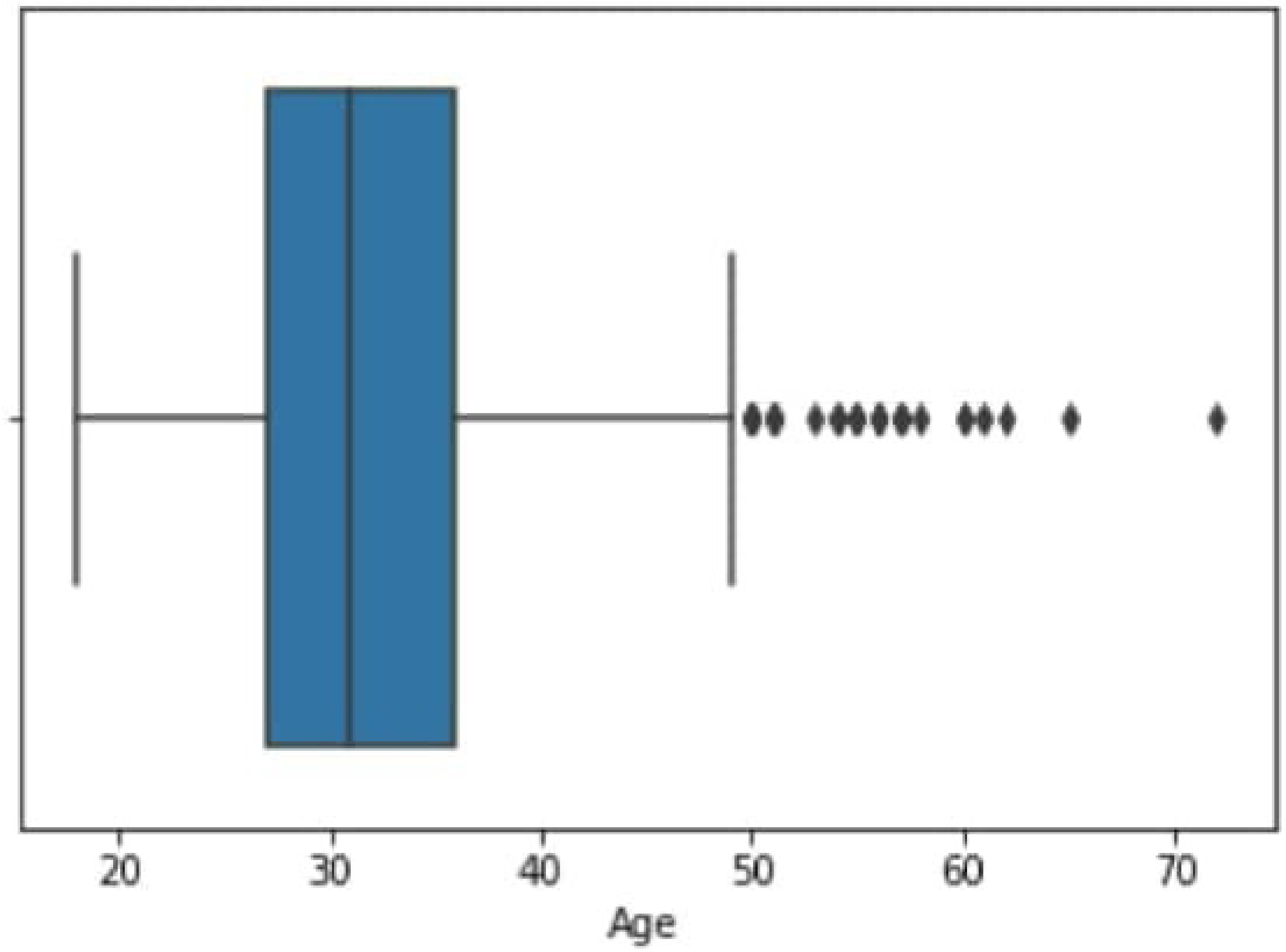
In [41]:
```
df["Age"].hist()
```

Out[41]: <matplotlib.axes._subplots.AxesSubplot at 0x1d203ddba58>

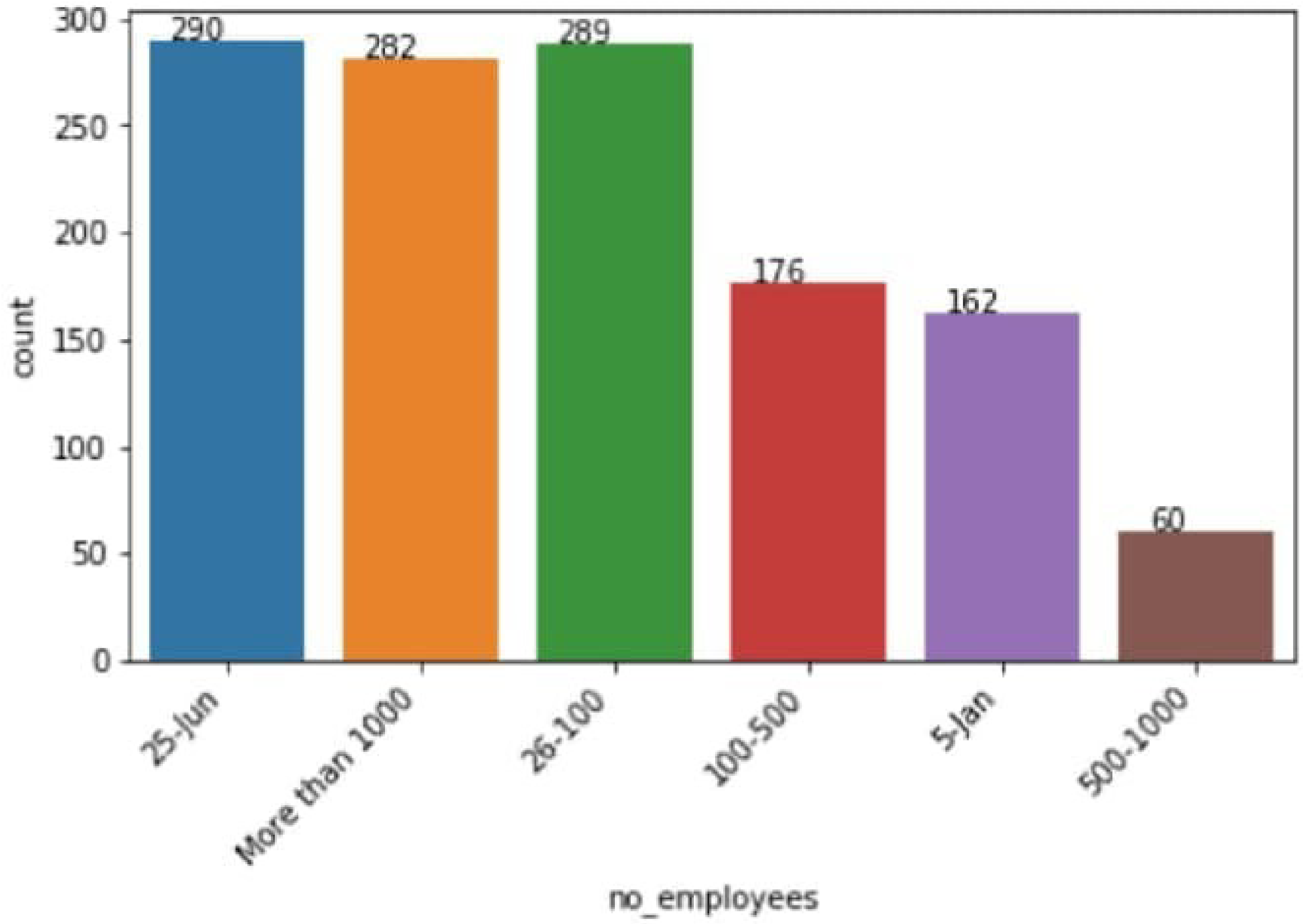


In [42]:
```
sns.boxplot(x=df["Age"])
```

Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x1d2052d04e0>

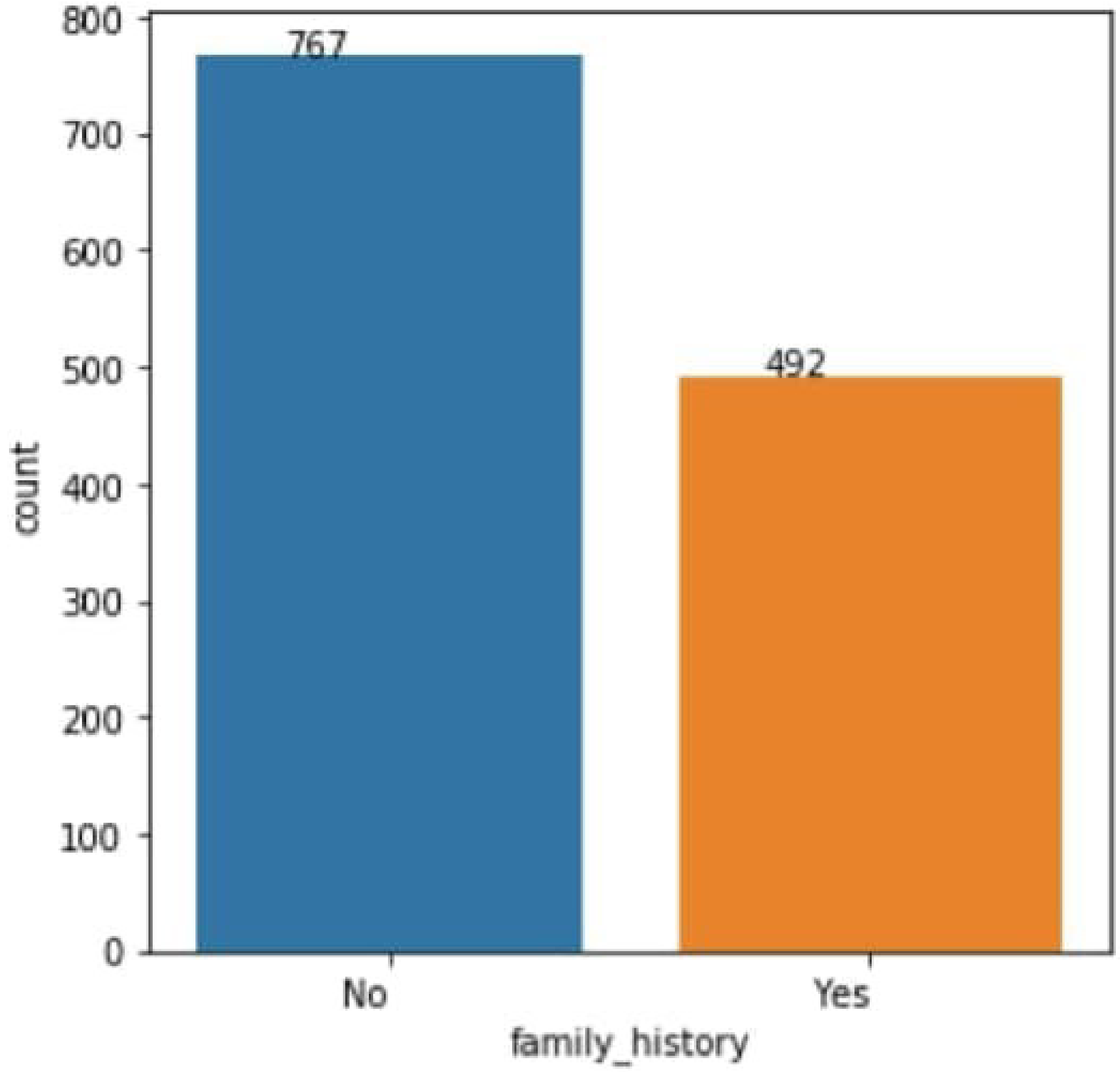



```
In [43]: # Outliers: Age above 50
         # These can be men and women who entered tech in the 80's

In [44]: import statistics

         variance_age = df["Age"].var()
         standard_dev_age = statistics.stdev(df["Age"])

         print(f"Mean: {round(mean_age, 2)}"
               f"\nVariance: {round(variance_age, 2)}"
               f"\nStandard Deviation: {round(standard_dev_age, 2)}")

Mean: 79428148.31
Variance: 52.79
Standard Deviation: nan


In [45]: plt.figure(figsize=(7,4))
         ax = sns.countplot(x='no_employees', data=df)

         ax.set_xticklabels(ax.get_xticklabels(),
                            rotation=45,
                            horizontalalignment='right')
         # Then we also display the values for each bar above it;
         for p in ax.patches:
            ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```

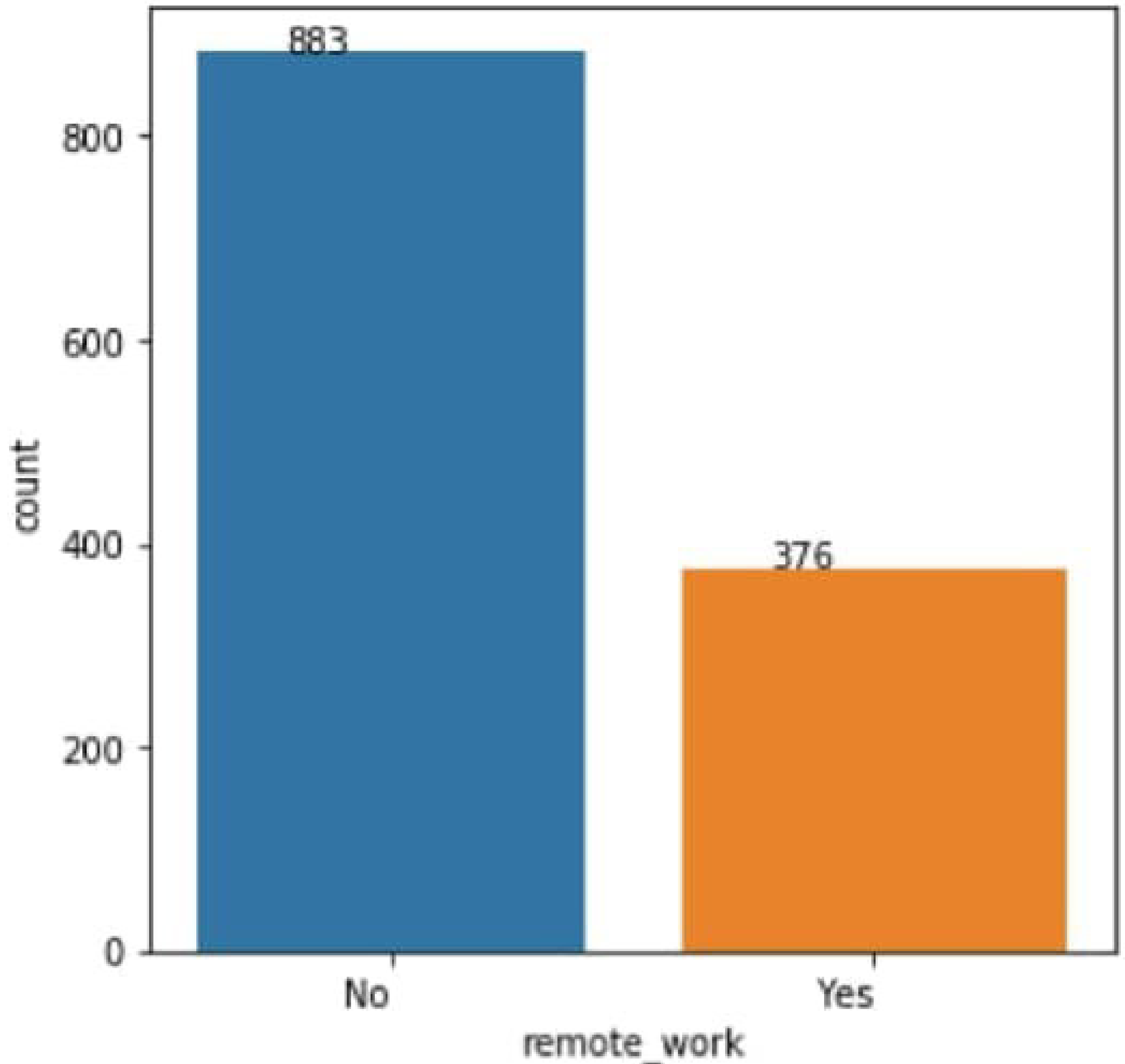



```
In [46]: # This doesn't help us much, because number of employees isn't directly proportional w
         # In fact, there are identical mental health issues with employees working in companies
             # 6 - 25
             # 25 - 100
             # More than 1000
```

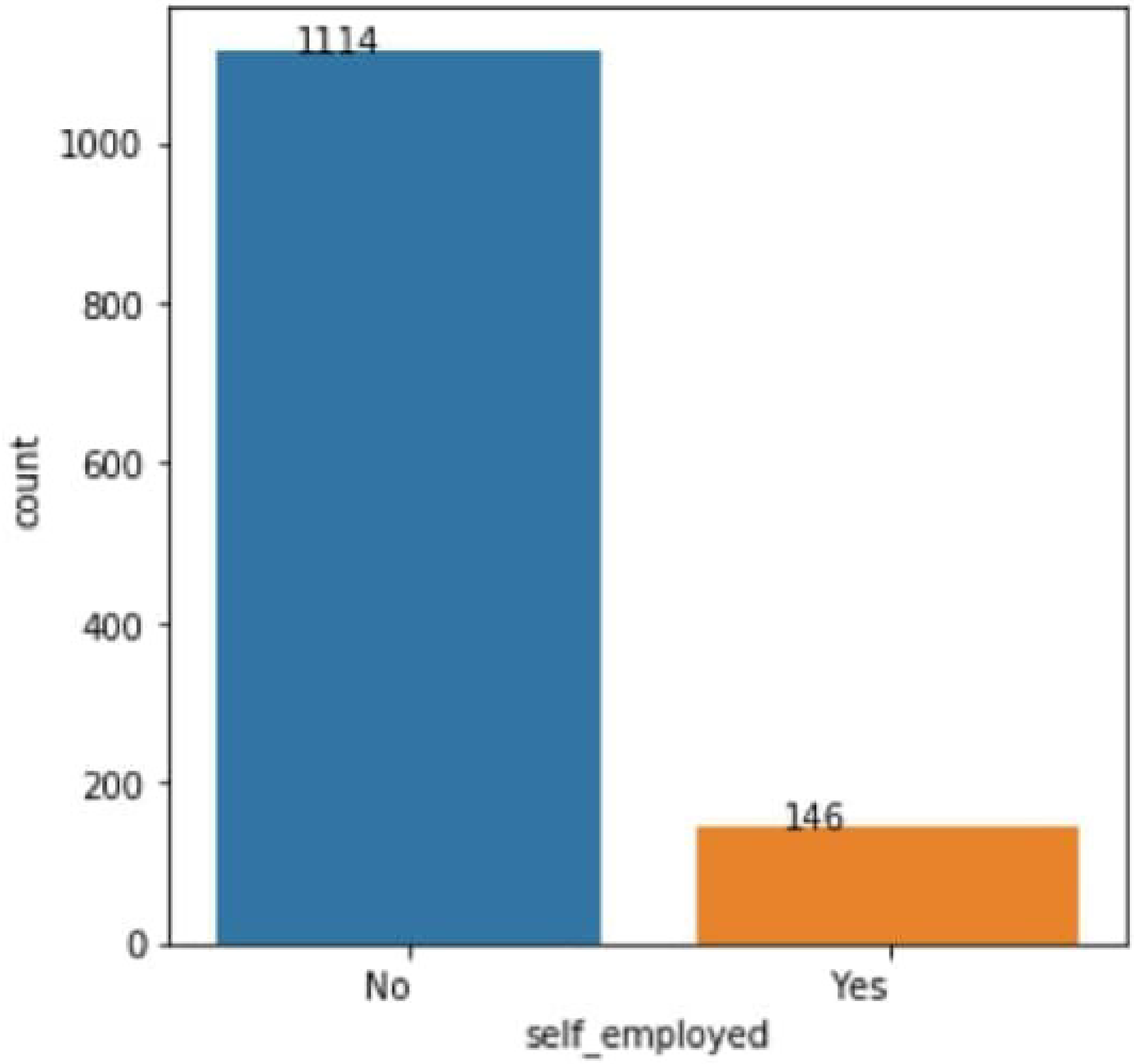
```
In [47]: plt.figure(figsize=(5,5))
         ax = sns.countplot(x='family_history', data=df)
         ax.set_xticklabels(ax.get_xticklabels(),
                            horizontalalignment='right')
         for p in ax.patches:
            ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```



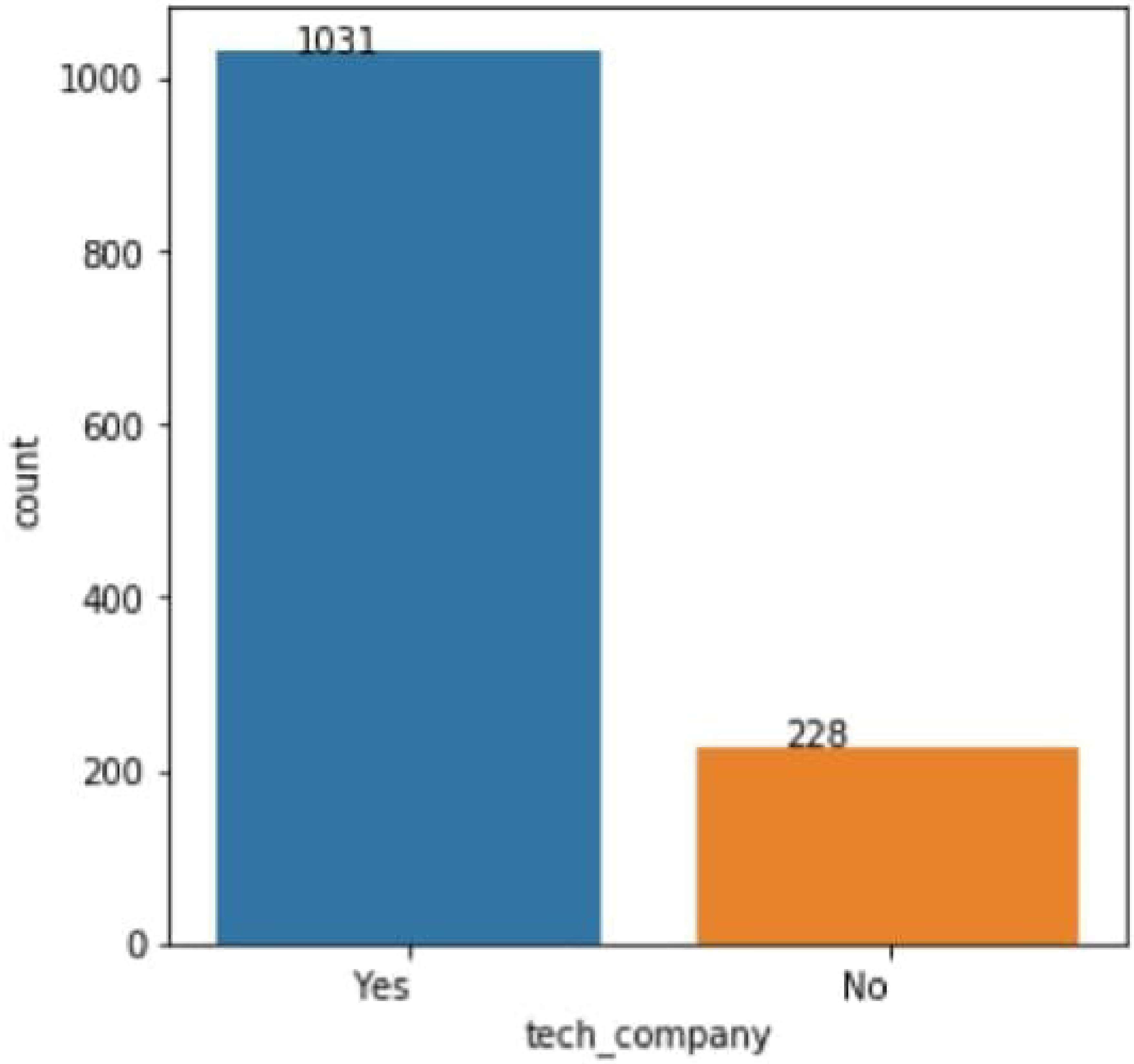

```
In [48]: plt.figure(figsize=(5,5))
         ax = sns.countplot(x='remote_work', data=df)
         ax.set_xticklabels(ax.get_xticklabels(),
                            horizontalalignment='right')
         for p in ax.patches:
            ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```




```
In [49]: plt.figure(figsize=(5,5))
         ax = sns.countplot(x='self_employed', data=df)
         ax.set_xticklabels(ax.get_xticklabels(),
                            horizontalalignment='right')
         for p in ax.patches:
            ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```



In [50]:
```
plt.figure(figsize=(5,5))
ax = sns.countplot(x='tech_company', data=df)
ax.set_xticklabels(ax.get_xticklabels(),
                   horizontalalignment='right')
for p in ax.patches:
   ax.annotate(p.get_height(), (p.get_x()+0.25, p.get_height()+0.01), ha='center')
```




```
In [51]: df["Timestamp"].head()

Out[51]: 0    8/27/2014 11:29
         1    8/27/2014 11:29
         2    8/27/2014 11:29
         3    8/27/2014 11:29
         4    8/27/2014 11:30
         Name: Timestamp, dtype: object

In [52]: # For me to work with just the Years, excluding the Time, Months and Days, I will do th

         df_year = pd.to_datetime(df["Timestamp"]).dt.year
         df_year.head()


        ---------------------------------------------------------------------------

        TypeError                                 Traceback (most recent call last)

        ~\Anaconda3\lib\site-packages\pandas\core\tools\datetimes.py in _convert_listlike(arg, b
        376             try:
```

CONCLUSION:

The potential for substantial adverse or beneficial health effects or irreversible or catastrophic effects, even if the effects have a low likelihood.

Social relationships have an impact on our mental health, physical health and mortality risk. Over the years, sociologists have created a link between social relationships and health outcomes.

Studies are showing that social relationships both quality and quantity are having short and long-term effects on our health.