Predictive Analysis of Mental Health Trends in the Tech Industry

Megha RS

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

```
data=pd.read_csv("OSMI Mental Health Dataset_final.csv",encoding='cp1252')
```

```
data.shape
```

```
(60186, 27)
```

```
data.head()
```

	ResponseID	Are you selfemployed	How many employees does your company or organization have	Is your employer primarily a tech companyorganization	Is your primary role within your company related to techIT	Do you have previous employers	h i
0	r00000	False	26-100	True	NaN	True	
1	r00000	False	26-100	True	NaN	True	
2	r00000	False	26-100	True	NaN	True	
3	r00000	False	26-100	True	NaN	True	
4	r00000	False	26-100	True	NaN	True	

5 rows × 27 columns

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60186 entries, 0 to 60185
Data columns (total 27 columns):
 #   Column                                                 Non-Null Count  Dtype
---  ------                                                 --------------  -----
 0   ResponseID                                             60186 non-null  object
 1   Are you selfemployed                                   60186 non-null  bool
 2   How many employees does your company or organization have  48132 non-null  object
 3   Is your employer primarily a tech companyorganization  48132 non-null  object
```

```
 4   Is your primary role within your company related to techIT                                 11046 non-null  object
 5   Do you have previous employers                                                             60186 non-null  bool
 6   Do you have a family history of mental illness                                             60186 non-null  object
 7   Have you had a mental health disorder in the past                                          60186 non-null  object
 8   Do you currently have a mental health disorder                                             60186 non-null  object
 9   If yes, what conditions have you been diagnosed with                                       23856 non-null  object
 10  If maybe, what conditions do you believe you have                                          13524 non-null  object
 11  Have you been diagnosed with a mental health condition by a medical professional          60186 non-null  bool
 12  If so, what conditions were you diagnosed with                                             29862 non-null  object
 13  Have you ever sought treatment for a mental health issue from a mental health professional 60186 non-null  bool
 14  What is your age                                                                           60102 non-null  float64
 15  What is your gender                                                                        60186 non-null  object
 16  Age Group                                                                                  60186 non-null  object
 17  What country do you live in                                                                60186 non-null  object
 18  What US state or territory do you live in                                                  35280 non-null  object
 19  What country do you work in                                                                60186 non-null  object
 20  What US state or territory do you work in                                                  35742 non-null  object
 21  Which of the following best describes your work position                                   60186 non-null  object
 22  Do you work remotely                                                                       60186 non-null  object
 23  Question Group                                                                             51588 non-null  object
 24  Question about speaking openly about mental health vs physical health                      60186 non-null  object
 25  Question                                                                                   60186 non-null  object
 26  Response                                                                                   43846 non-null  object
dtypes: bool(4), float64(1), object(22)
memory usage: 10.8+ MB
```

```
data.nunique()
```

```
ResponseID                                                                    1433
Are you selfemployed                                                             2
How many employees does your company or organization have                        6
Is your employer primarily a tech companyorganization                            2
Is your primary role within your company related to techIT                       2
Do you have previous employers                                                   2
Do you have a family history of mental illness                                   3
Have you had a mental health disorder in the past                                3
Do you currently have a mental health disorder                                   3
If yes, what conditions have you been diagnosed with                           128
If maybe, what conditions do you believe you have                               99
Have you been diagnosed with a mental health condition by a medical professional 2
If so, what conditions were you diagnosed with                                 116
Have you ever sought treatment for a mental health issue from a mental health professional 2
What is your age                                                                51
What is your gender                                                              3
Age Group                                                                       11
What country do you live in                                                     53
What US state or territory do you live in                                       47
What country do you work in                                                     53
What US state or territory do you work in                                       48
Which of the following best describes your work position                       264
Do you work remotely                                                             3
Question Group                                                                   6
Question about speaking openly about mental health vs physical health            2
Question                                                                        42
Response                                                                      2076
dtype: int64
```

```
data['How many employees does your company or organization have'].unique()
```

```
array(['26-100', 'Jun-25', nan, 'More than 1000', '100-500', '500-1000',
       '01-May'], dtype=object)
```

Is tech company or not

```
data['Is your employer primarily a tech companyorganization'].value_counts()
```

```
Is your employer primarily a tech companyorganization
True     37086
False    11046
Name: count, dtype: int64
```

Which country

```
data['What country do you work in'].value_counts()
```

```
What country do you work in
United States of America    35742
United Kingdom               7686
Canada                       3108
Germany                      2436
Netherlands                  1974
Australia                    1428
Sweden                        840
```

```
Ireland                      630
France                       588
Switzerland                  420
Brazil                       420
India                        378
Russia                       378
New Zealand                  378
Denmark                      294
Bulgaria                     294
Finland                      294
Belgium                      210
South Africa                 168
Poland                       168
Austria                      168
Czech Republic               126
Italy                        126
Chile                        126
Norway                       126
Romania                      126
Spain                        126
Other                         84
Bosnia and Herzegovina        84
Afghanistan                   84
Pakistan                      84
Israel                        84
Estonia                       84
Colombia                      84
Mexico                        84
Slovakia                      42
United Arab Emirates          42
Serbia                        42
Guatemala                     42
Lithuania                     42
China                         42
Venezuela                     42
Ecuador                       42
Japan                         42
Hungary                       42
Vietnam                       42
Iran                          42
Brunei                        42
Argentina                     42
Greece                        42
Costa Rica                    42
Bangladesh                    42
Turkey                        42
Name: count, dtype: int64
```

```
data.isnull().sum()
```

```
ResponseID                                                                      0
Are you selfemployed                                                            0
How many employees does your company or organization have                   12054
Is your employer primarily a tech companyorganization                       12054
Is your primary role within your company related to techIT                  49140
Do you have previous employers                                                  0
Do you have a family history of mental illness                                  0
Have you had a mental health disorder in the past                               0
Do you currently have a mental health disorder                                  0
If yes, what conditions have you been diagnosed with                        36330
If maybe, what conditions do you believe you have                           46662
Have you been diagnosed with a mental health condition by a medical professional    0
If so, what conditions were you diagnosed with                              30324
Have you ever sought treatment for a mental health issue from a mental health professional    0
What is your age                                                               84
What is your gender                                                             0
Age Group                                                                       0
What country do you live in                                                     0
What US state or territory do you live in                                   24906
What country do you work in                                                     0
What US state or territory do you work in                                   24444
Which of the following best describes your work position                        0
Do you work remotely                                                            0
Question Group                                                               8598
Question about speaking openly about mental health vs physical health           0
Question                                                                        0
Response                                                                    16340
dtype: int64
```

```
data.describe()
```

	What is your age
count	60102.000000
mean	34.106219
std	8.283055
min	15.000000
25%	28.000000
50%	33.000000
75%	39.000000
max	99.000000

Label Encoding

```
newdf=pd.DataFrame(data)
#newdf.head()
```

```
from sklearn.preprocessing import LabelEncoder
l=LabelEncoder()
for x in newdf:
    if newdf[x].dtypes=='object':
        newdf[x]=l.fit_transform(newdf[x])
```

```
newdf.head()
```

	ResponseID	Are you selfemployed	How many employees does your company or organization have	Is your employer primarily a tech companyorganization	Is your primary role within your company related to techIT	Do you have previous employers	Do you have a family history of mental illness	Have you had a mental health disorder in the past	Do you currently have a mental health disorder	If yes, what conditions have you been diagnosed with	...	co d li
0	0	False	2	1	2	True	1	2	1	128	...	
1	0	False	2	1	2	True	1	2	1	128	...	
2	0	False	2	1	2	True	1	2	1	128	...	
3	0	False	2	1	2	True	1	2	1	128	...	
4	0	False	2	1	2	True	1	2	1	128	...	

5 rows × 27 columns

original_categories = label_encoder.inverse_transform(encoded_labels) #Reverse the label encoding in python

```
newdf.isnull().sum()
```

```
ResponseID                                                                  0
Are you selfemployed                                                        0
How many employees does your company or organization have                   0
Is your employer primarily a tech companyorganization                       0
Is your primary role within your company related to techIT                  0
Do you have previous employers                                              0
Do you have a family history of mental illness                              0
Have you had a mental health disorder in the past                           0
Do you currently have a mental health disorder                              0
If yes, what conditions have you been diagnosed with                        0
If maybe, what conditions do you believe you have                           0
Have you been diagnosed with a mental health condition by a medical professional    0
If so, what conditions were you diagnosed with                              0
Have you ever sought treatment for a mental health issue from a mental health professional    0
What is your age                                                           84
What is your gender                                                         0
Age Group                                                                   0
What country do you live in                                                 0
What US state or territory do you live in                                   0
What country do you work in                                                 0
What US state or territory do you work in                                   0
Which of the following best describes your work position                    0
Do you work remotely                                                        0
Question Group                                                              0
```

```
Question about speaking openly about mental health vs physical health    0
Question                                                                 0
Response                                                                 0
dtype: int64
```

```
newdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60186 entries, 0 to 60185
Data columns (total 27 columns):
 #   Column                                                                  Non-Null Count  Dtype
---  ------                                                                  --------------  -----
 0   ResponseID                                                              60186 non-null  int32
 1   Are you selfemployed                                                    60186 non-null  bool
 2   How many employees does your company or organization have               60186 non-null  int32
 3   Is your employer primarily a tech companyorganization                   60186 non-null  int32
 4   Is your primary role within your company related to techIT              60186 non-null  int32
 5   Do you have previous employers                                          60186 non-null  bool
 6   Do you have a family history of mental illness                          60186 non-null  int32
 7   Have you had a mental health disorder in the past                       60186 non-null  int32
 8   Do you currently have a mental health disorder                          60186 non-null  int32
 9   If yes, what conditions have you been diagnosed with                    60186 non-null  int32
 10  If maybe, what conditions do you believe you have                       60186 non-null  int32
 11  Have you been diagnosed with a mental health condition by a medical professional  60186 non-null  bool
 12  If so, what conditions were you diagnosed with                          60186 non-null  int32
 13  Have you ever sought treatment for a mental health issue from a mental health professional  60186 non-null  bool
 14  What is your age                                                        60102 non-null  float64
 15  What is your gender                                                     60186 non-null  int32
 16  Age Group                                                               60186 non-null  int32
 17  What country do you live in                                             60186 non-null  int32
 18  What US state or territory do you live in                               60186 non-null  int32
 19  What country do you work in                                             60186 non-null  int32
 20  What US state or territory do you work in                               60186 non-null  int32
 21  Which of the following best describes your work position                60186 non-null  int32
 22  Do you work remotely                                                    60186 non-null  int32
 23  Question Group                                                          60186 non-null  int32
 24  Question about speaking openly about mental health vs physical health   60186 non-null  int32
 25  Question                                                                60186 non-null  int32
 26  Response                                                                60186 non-null  int32
dtypes: bool(4), float64(1), int32(22)
memory usage: 5.7 MB
```

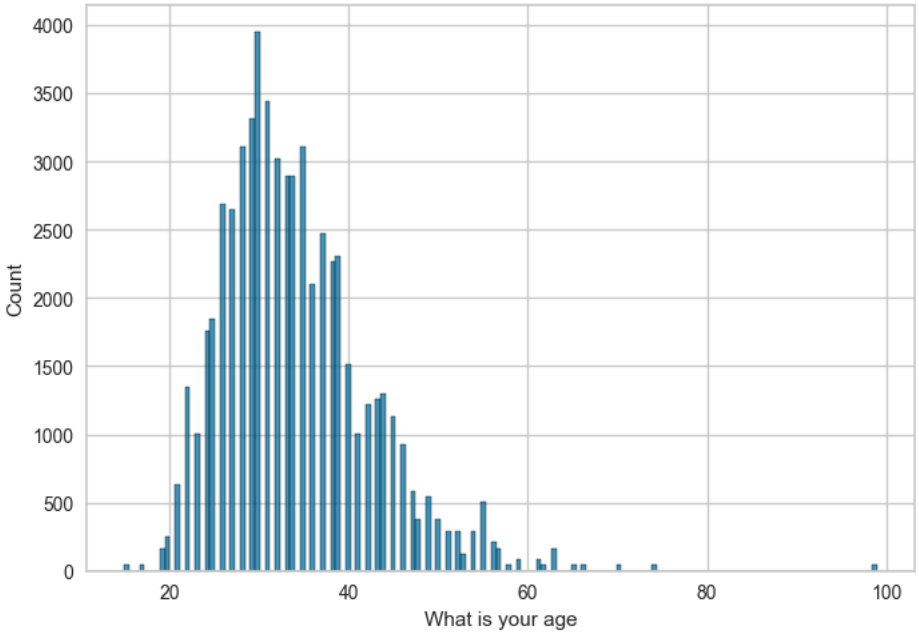
Handled Null Value

Null values will be replaced by its Mean value when the column present in the dataset is not skewed.

Null values will be replaced by its Median value when column present in the dataset is skewed.

```
sns.histplot(data["What is your age"])
```

```
<Axes: xlabel='What is your age', ylabel='Count'>
```




```
newdf['What is your age'] = newdf['What is your age'].fillna(newdf['What is your age'].median())
```

```
newdf['What is your age'].isnull().sum()
```

```
0
```

```
# Define a dictionary to map old column names to new column names
column_mapping = {'Have you ever sought treatment for a mental health issue from a mental health professional':
                  'Took treatment from professional',
                  'Have you been diagnosed with a mental health condition by a medical professional':
                  'Diagonsed or not',
                  'How many employees does your company or organization have':
                  'No. employees in company',
                  'Is your primary role within your company related to techIT': 'Role Tech or nonTech',
                  'Is your employer primarily a tech companyorganization':'Company Tech or nonTech',
                 }

# Rename the columns using the dictionary
newdf.rename(columns=column_mapping, inplace=True)
```

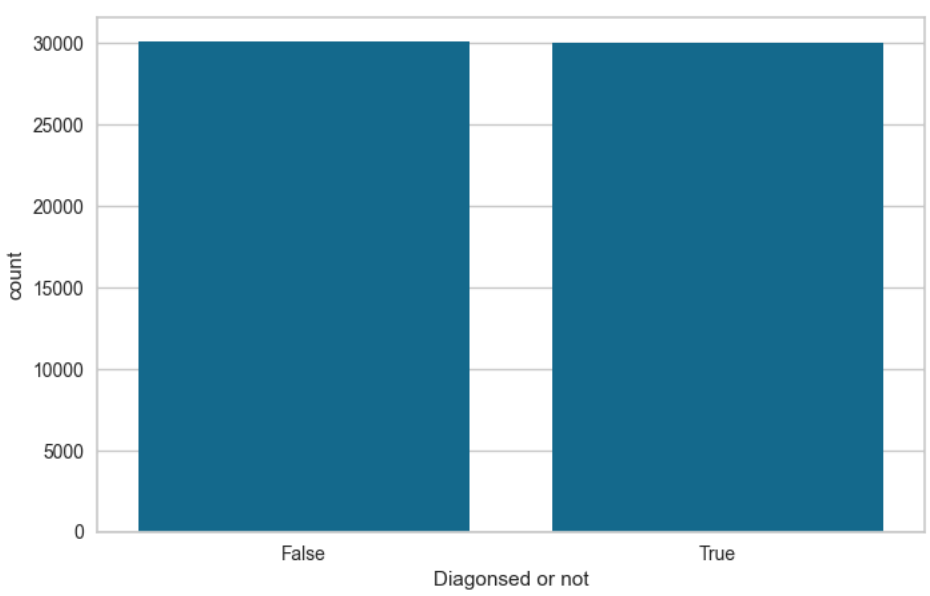
DATA is balanced or imbalanced

```
newdf.groupby('Diagonsed or not').sum()
```

Diagonsed or not	ResponseID	Are you selfemployed	No. employees in company	Company Tech or nonTech	Role Tech or nonTech	Do you have previous employers	Do you have a family history of mental illness	Have you had a mental health disorder in the past	Do you currently have a mental health disorder	If yes, what conditions have you been diagnosed with	...	What country do you live in
False	20895252	5796	101556	30450	54390	25956	32088	26460	24276	3679116	...	1229088
True	22197924	6258	104832	30744	54306	27132	44688	54306	46326	2119908	...	1294860

2 rows × 26 columns

```
plt.figure(figsize=(8,5))
sns.barplot(x=newdf['Diagonsed or not'].value_counts().index,y=newdf['Diagonsed or not'].value_counts())
plt.show()
```




```
newdf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60186 entries, 0 to 60185
Data columns (total 27 columns):
 #   Column                                                  Non-Null Count  Dtype
---  ------                                                  --------------  -----
 0   ResponseID                                              60186 non-null  int32
 1   Are you selfemployed                                    60186 non-null  bool
```

```
 2   No. employees in company                                                60186 non-null  int32
 3   Company Tech or nonTech                                                 60186 non-null  int32
 4   Role Tech or nonTech                                                    60186 non-null  int32
 5   Do you have previous employers                                          60186 non-null  bool
 6   Do you have a family history of mental illness                          60186 non-null  int32
 7   Have you had a mental health disorder in the past                       60186 non-null  int32
 8   Do you currently have a mental health disorder                          60186 non-null  int32
 9   If yes, what conditions have you been diagnosed with                    60186 non-null  int32
 10  If maybe, what conditions do you believe you have                       60186 non-null  int32
 11  Diagonsed or not                                                        60186 non-null  bool
 12  If so, what conditions were you diagnosed with                          60186 non-null  int32
 13  Took treatment from professional                                        60186 non-null  bool
 14  What is your age                                                        60186 non-null  float64
 15  What is your gender                                                     60186 non-null  int32
 16  Age Group                                                               60186 non-null  int32
 17  What country do you live in                                             60186 non-null  int32
 18  What US state or territory do you live in                               60186 non-null  int32
 19  What country do you work in                                             60186 non-null  int32
 20  What US state or territory do you work in                               60186 non-null  int32
 21  Which of the following best describes your work position                60186 non-null  int32
 22  Do you work remotely                                                    60186 non-null  int32
 23  Question Group                                                          60186 non-null  int32
 24  Question about speaking openly about mental health vs physical health   60186 non-null  int32
 25  Question                                                                60186 non-null  int32
 26  Response                                                                60186 non-null  int32
dtypes: bool(4), float64(1), int32(22)
memory usage: 5.7 MB
```

CLUSTERING

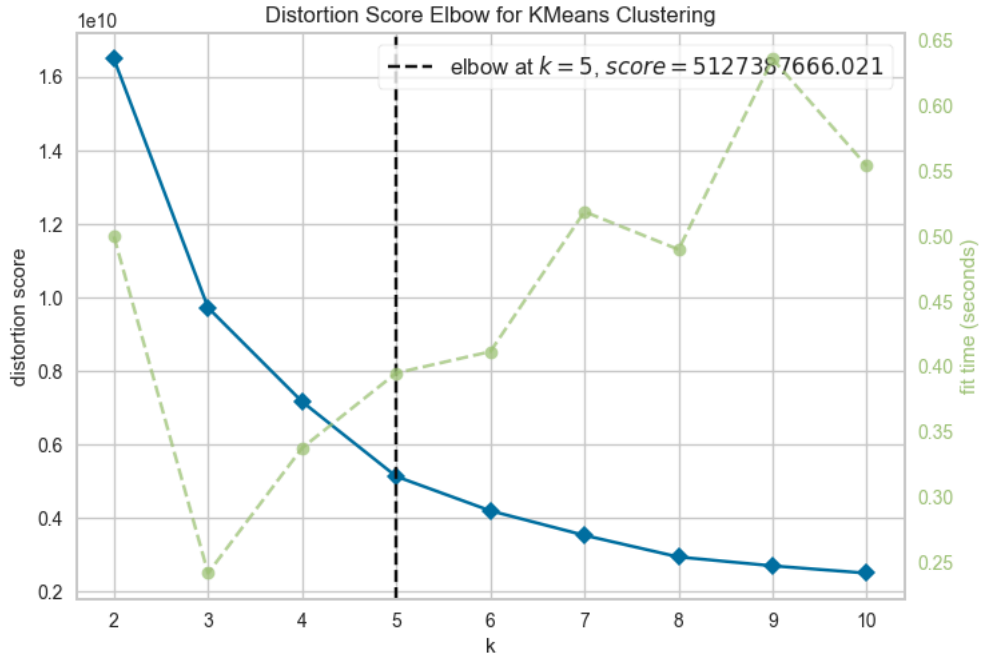
```
df1=newdf.drop('Diagonsed or not',axis=1)
```

```
! pip install yellowbrick
```

```
Defaulting to user installation because normal site-packages is not writeable
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip

Requirement already satisfied: yellowbrick in c:\users\megha\appdata\roaming\python\python311\site-packages (1.5)
Requirement already satisfied: matplotlib!=3.0.0,>=2.0.2 in c:\users\megha\appdata\roaming\python\python311\site-packages (from yell
Requirement already satisfied: scipy>=1.0.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from yellowbrick) (1.1
Requirement already satisfied: scikit-learn>=1.0.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from yellowbric
Requirement already satisfied: numpy>=1.16.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from yellowbrick) (1
Requirement already satisfied: cycler>=0.10.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from yellowbrick) (0
Requirement already satisfied: contourpy>=1.0.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=3
Requirement already satisfied: fonttools>=4.22.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=
Requirement already satisfied: packaging>=20.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=3
Requirement already satisfied: pillow>=8 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=3.0.0,>=
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib!=3
Requirement already satisfied: python-dateutil>=2.7 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotli
Requirement already satisfied: joblib>=1.1.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-learn>=1
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-le
Requirement already satisfied: six>=1.5 in c:\users\megha\appdata\roaming\python\python311\site-packages (from python-dateutil>=2.7
```

```
from sklearn.cluster import KMeans
from yellowbrick.cluster import KElbowVisualizer, SilhouetteVisualizer

model = KMeans(random_state=42)
elb_visualizer = KElbowVisualizer(model, k=(2,11))
elb_visualizer.fit(df1)
elb_visualizer.show()
#Kmeans Clustering
```

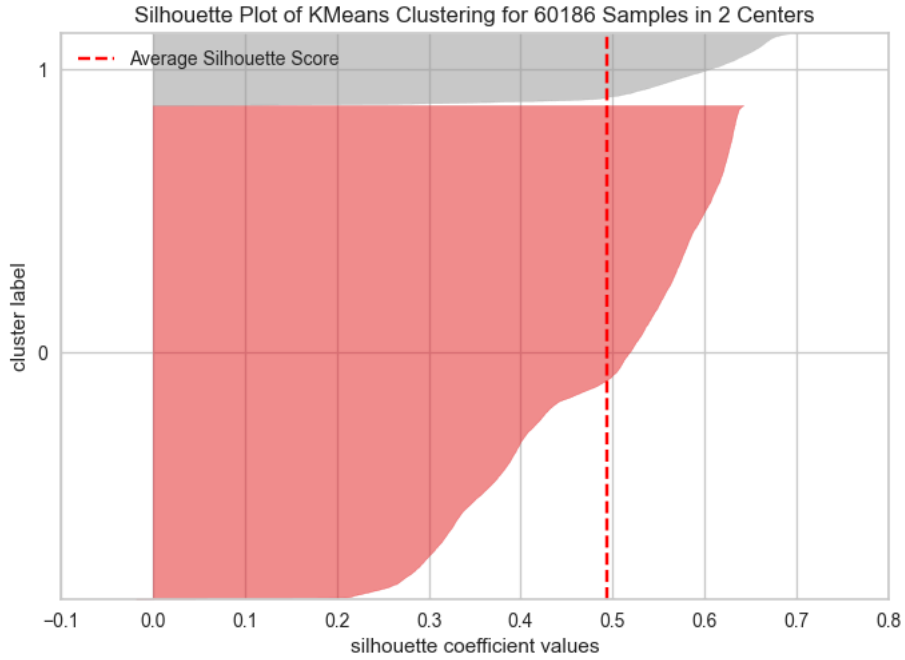
```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
```



```
<Axes: title={'center': 'Distortion Score Elbow for KMeans Clustering'}, xlabel='k', ylabel='distortion score'>
```

```
model_4clust = KMeans(n_clusters = 2, random_state=42)
sil_visualizer = SilhouetteVisualizer(model_4clust)
sil_visualizer.fit(df1)
sil_visualizer.show()
```

```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning: The default value of `
  super()._check_params_vs_input(X, default_n_init=10)
```




```
<Axes: title={'center': 'Silhouette Plot of KMeans Clustering for 60186 Samples in 2 Centers'}, xlabel='silhouette coefficient
values', ylabel='cluster label'>
```

By changing the number of clusters, the silhouette score got 0.05 higher and the clusters are more balanced.

Outlier Analysis

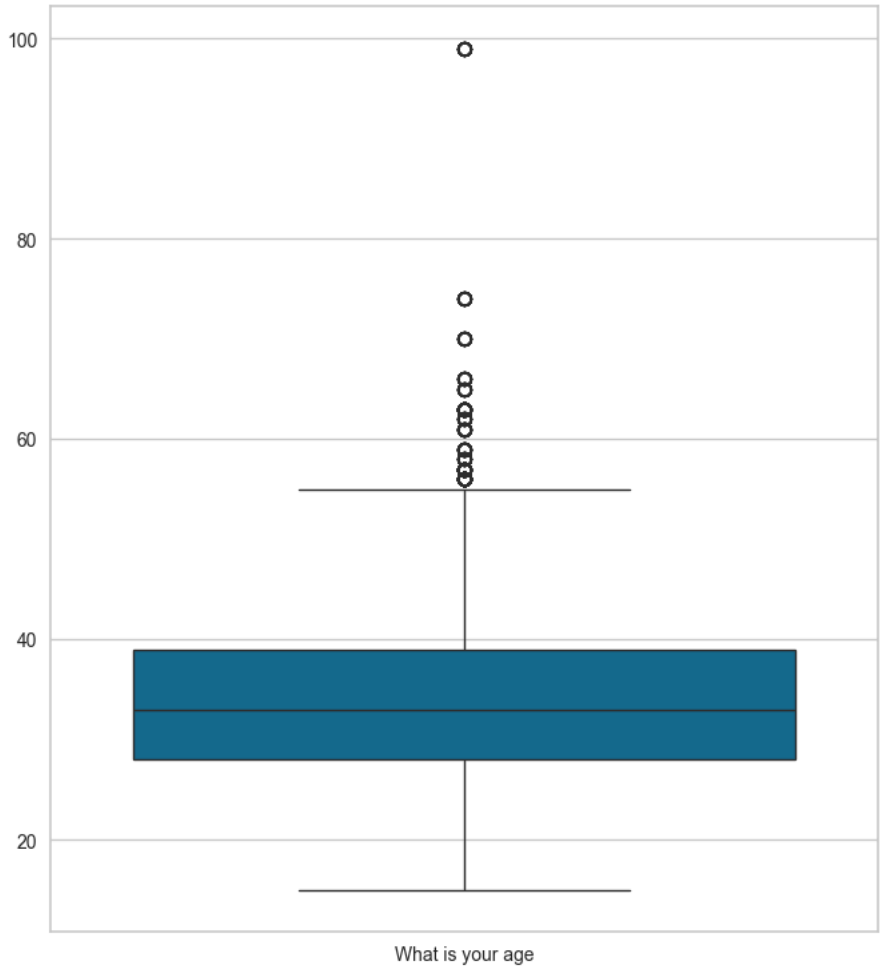
For Skewed Distributions

Use Inter-Quartile Range (IQR) proximity rule.

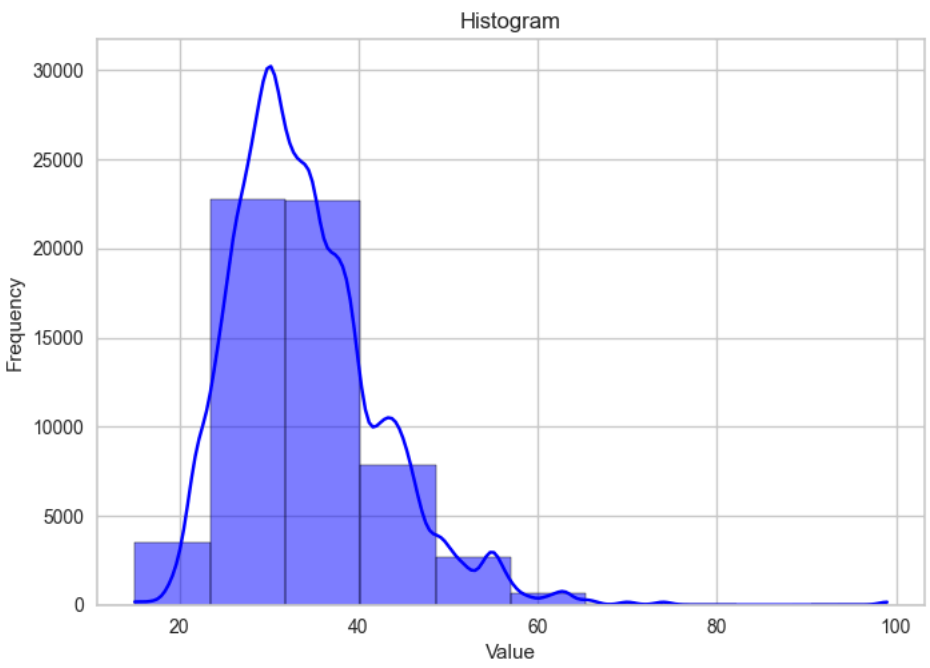
The data points that fall below Q1 − 1.5 IQR or above the third quartile Q3 + 1.5 IQR are outliers, where Q1 and Q3 are the 25th and 75th percentile of the dataset, respectively. IQR represents the inter-quartile range and is given by Q3 − Q1.

```
# Selecting only specific columns to create a new dataset
plt.figure(figsize=(8, 9))
selected_columns = ['What is your age']
columns = newdf[selected_columns]
sns.boxplot(columns)
```

<Axes: >




```
sns.histplot(newdf['What is your age'], bins=10, kde=True, color='blue', edgecolor='black')
plt.xlabel('Value')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.show()
```



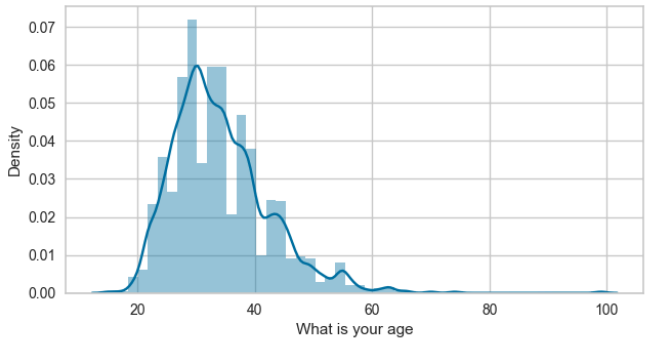

```
Q1 = newdf['What is your age'].quantile(0.25)
Q3 = newdf['What is your age'].quantile(0.75)
IQR = Q3 - Q1
```

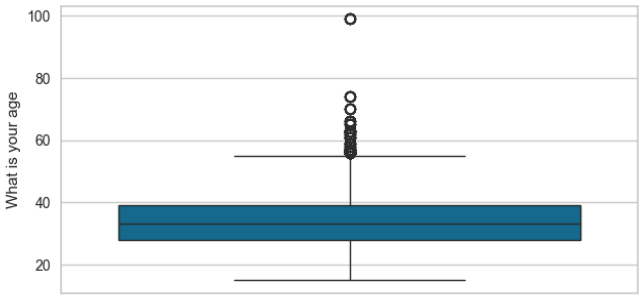
```
upper_limit = Q3 + 1.5 * IQR
lower_limit = Q1 - 1.5 * IQR
newdf[newdf['What is your age'] > upper_limit]
newdf[newdf['What is your age'] < lower_limit]
new_df2 = newdf[newdf['What is your age'] < upper_limit]
new_df2.shape
```

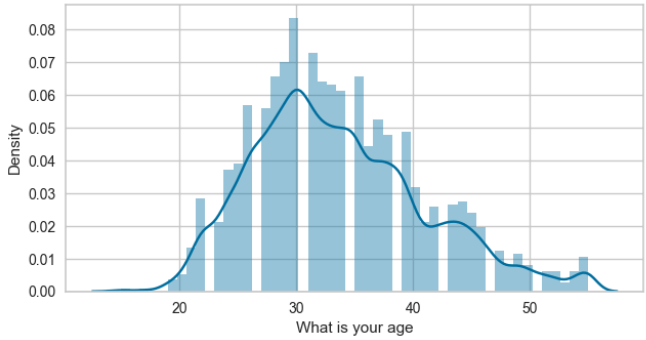
```
(59178, 27)
```

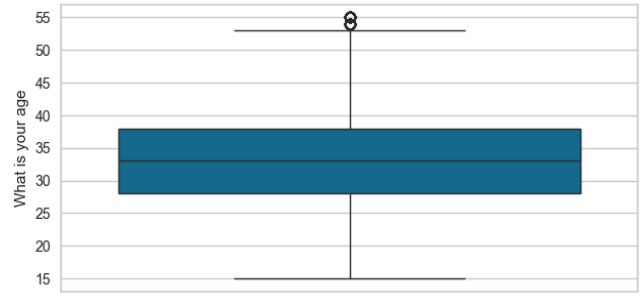
```
plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(newdf['What is your age'])
plt.subplot(2,2,2)
sns.boxplot(newdf['What is your age'])
plt.subplot(2,2,3)
sns.distplot(new_df2['What is your age'])
plt.subplot(2,2,4)
sns.boxplot(new_df2['What is your age'])
plt.show()
```

```
C:\Users\megha\AppData\Local\Temp\ipykernel_25380\1687354700.py:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(newdf['What is your age'])
C:\Users\megha\AppData\Local\Temp\ipykernel_25380\1687354700.py:7: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(new_df2['What is your age'])
```



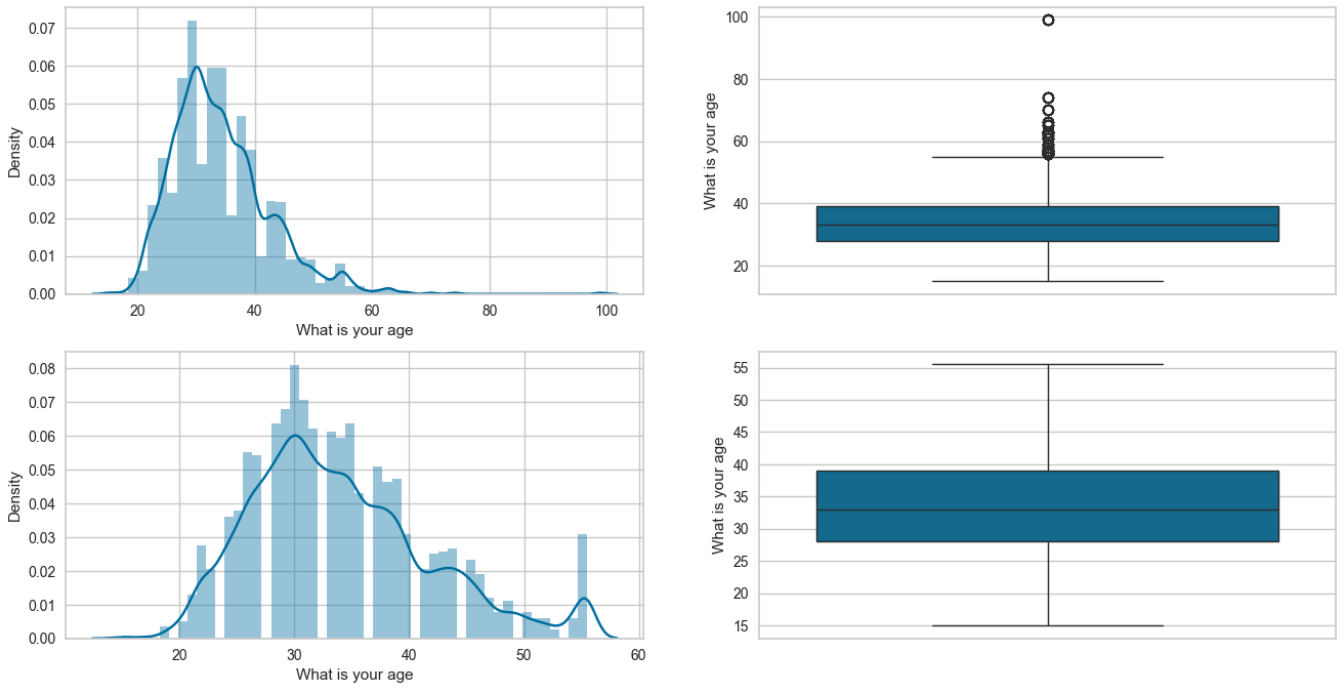






```
new_df_cap = newdf.copy()
new_df_cap['What is your age'] = np.where(
    new_df_cap['What is your age'] > upper_limit,
    upper_limit,
    np.where(
       new_df_cap['What is your age'] < lower_limit,
        lower_limit, new_df_cap['What is your age']))


plt.figure(figsize=(16,8))
plt.subplot(2,2,1)
sns.distplot(newdf['What is your age'])
plt.subplot(2,2,2)
sns.boxplot(newdf['What is your age'])
plt.subplot(2,2,3)
sns.distplot(new_df_cap['What is your age'])
plt.subplot(2,2,4)
sns.boxplot(new_df_cap['What is your age'])
plt.show()
```

```
C:\Users\megha\AppData\Local\Temp\ipykernel_25380\1179355610.py:3: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(newdf['What is your age'])
C:\Users\megha\AppData\Local\Temp\ipykernel_25380\1179355610.py:7: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(new_df_cap['What is your age'])
```




```
new_df_cap.head()
```

	ResponseID	Are you selfemployed	No. employees in company	Company Tech or nonTech	Role Tech or nonTech	Do you have previous employers	Do you have a family history of mental illness	Have you had a mental health disorder in the past	Do you currently have a mental health disorder	If yes, what conditions have you been diagnosed with	...	What country do you live in	What U state o territor do yo live i
0	0	False	2	1	2	True	1	2	1	128	...	49	4
1	0	False	2	1	2	True	1	2	1	128	...	49	4
2	0	False	2	1	2	True	1	2	1	128	...	49	4
3	0	False	2	1	2	True	1	2	1	128	...	49	4
4	0	False	2	1	2	True	1	2	1	128	...	49	4

5 rows × 27 columns

```
new_df_cap.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60186 entries, 0 to 60185
Data columns (total 27 columns):
 #   Column                                                      Non-Null Count  Dtype
---  ------                                                      --------------  -----
 0   ResponseID                                                  60186 non-null  int32
 1   Are you selfemployed                                        60186 non-null  bool
 2   No. employees in company                                    60186 non-null  int32
 3   Company Tech or nonTech                                     60186 non-null  int32
 4   Role Tech or nonTech                                        60186 non-null  int32
 5   Do you have previous employers                              60186 non-null  bool
 6   Do you have a family history of mental illness              60186 non-null  int32
 7   Have you had a mental health disorder in the past           60186 non-null  int32
 8   Do you currently have a mental health disorder              60186 non-null  int32
 9   If yes, what conditions have you been diagnosed with        60186 non-null  int32
 10  If maybe, what conditions do you believe you have           60186 non-null  int32
 11  Diagonsed or not                                            60186 non-null  bool
 12  If so, what conditions were you diagnosed with              60186 non-null  int32
 13  Took treatment from professional                            60186 non-null  bool
 14  What is your age                                            60186 non-null  float64
 15  What is your gender                                         60186 non-null  int32
 16  Age Group                                                   60186 non-null  int32
 17  What country do you live in                                 60186 non-null  int32
 18  What US state or territory do you live in                   60186 non-null  int32
 19  What country do you work in                                 60186 non-null  int32
 20  What US state or territory do you work in                   60186 non-null  int32
 21  Which of the following best describes your work position    60186 non-null  int32
 22  Do you work remotely                                        60186 non-null  int32
 23  Question Group                                              60186 non-null  int32
 24  Question about speaking openly about mental health vs physical health  60186 non-null  int32
 25  Question                                                    60186 non-null  int32
 26  Response                                                    60186 non-null  int32
dtypes: bool(4), float64(1), int32(22)
memory usage: 5.7 MB
```

MODEL

```
l=LabelEncoder()
for x in new_df_cap:
    if new_df_cap[x].dtypes=='bool':
        new_df_cap[x]=l.fit_transform(new_df_cap[x])
```

```
new_df_cap=new_df_cap.drop(["ResponseID","Response",'Question Group','Question about speaking openly about mental health vs physical hea
```

```
new_df_cap.dtypes
```

```
Are you selfemployed                                        int64
Company Tech or nonTech                                     int32
Role Tech or nonTech                                        int32
Do you have previous employers                              int64
Do you have a family history of mental illness              int32
Have you had a mental health disorder in the past           int32
Do you currently have a mental health disorder              int32
If yes, what conditions have you been diagnosed with        int32
If maybe, what conditions do you believe you have           int32
Diagonsed or not                                            int64
If so, what conditions were you diagnosed with              int32
```

```
Took treatment from professional                              int64
What is your age                                            float64
What is your gender                                           int32
Age Group                                                     int32
What country do you live in                                   int32
What US state or territory do you live in                     int32
What country do you work in                                   int32
What US state or territory do you work in                     int32
Which of the following best describes your work position      int32
Do you work remotely                                          int32
dtype: object
```

```
# saving the dataframe
#new_df_cap.to_csv('OSMI_cleaned.csv')
```

```
X=new_df_cap.drop("Diagonsed or not",axis=1)
y=new_df_cap["Diagonsed or not"]
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

```
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
from sklearn import metrics
```

KNN

```
# import KNeighbors ClaSSifier from sklearn
from sklearn.neighbors import KNeighborsClassifier

# instantiate the model
knn = KNeighborsClassifier(n_neighbors=3)
# fit the model to the training set
knn.fit(X_train, y_train)
```

```
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=3)
```

```
print(classification_report(y_test, knn_y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9075
           1       1.00      1.00      1.00      8981

    accuracy                           1.00     18056
   macro avg       1.00      1.00      1.00     18056
weighted avg       1.00      1.00      1.00     18056
```

```
knn_y_pred = knn.predict(X_test.values)
knn_y_pred
```

```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
array([0, 1, 1, ..., 0, 0, 0], dtype=int64)
```

```
print('Model accuracy score: {0:0.4f}'. format(accuracy_score(y_test, knn_y_pred)))
```

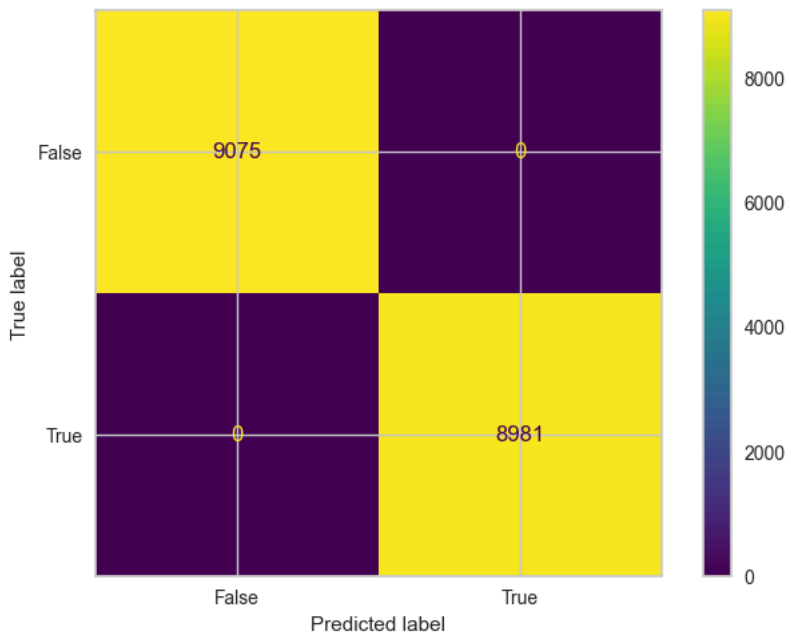
```
Model accuracy score: 1.0000
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(knn.score(X_train.values, y_train)))
print('Test set score: {:.4f}'.format(knn.score(X_test.values, y_test)))
```

```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Training set score: 1.0000
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Test set score: 1.0000
```

```
knn_cm = confusion_matrix(y_test, knn_y_pred)
```

```
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = knn_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



Logistic Regression

```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression(solver='liblinear', random_state=0)
logreg.fit(X_train, y_train)
```

```
LogisticRegression(random_state=0, solver='liblinear')
```

```
print(classification_report(y_test, logr_y_pred))
```

```
              precision    recall  f1-score   support

           0       0.98      1.00      0.99      9075
           1       1.00      0.98      0.99      8981

    accuracy                           0.99     18056
   macro avg       0.99      0.99      0.99     18056
weighted avg       0.99      0.99      0.99     18056
```

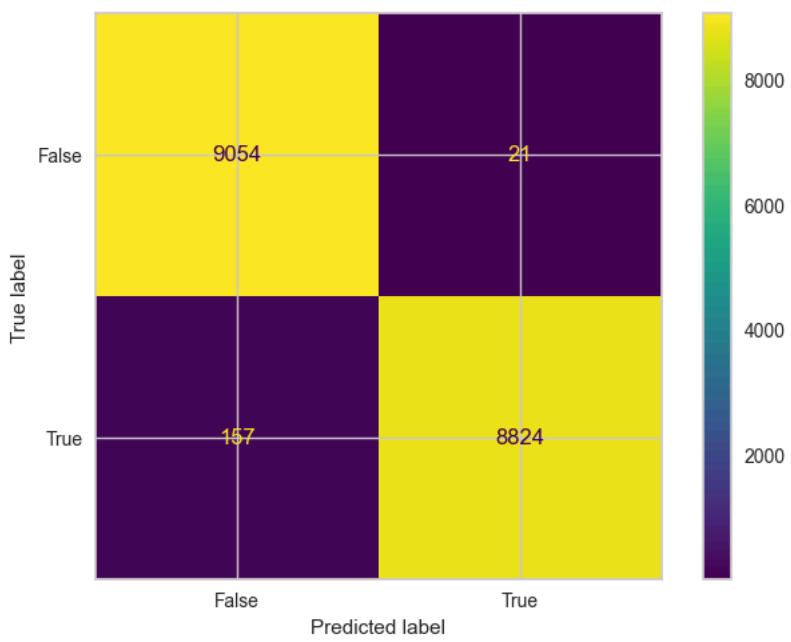
```
logr_y_pred= logreg.predict(X_test)
logr_y_pred
```

```
array([0, 1, 1, ..., 0, 0, 0], dtype=int64)
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(logreg.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(logreg.score(X_test, y_test)))
```

```
Training set score: 0.9895
Test set score: 0.9901
```

```
logr_cm = confusion_matrix(y_test, logr_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = logr_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



Decision Tree Classifier

```
# import DecisionTreeClassifier
from sklearn.tree import DecisionTreeClassifier
```

Decision Tree Classifier with criterion gini index

```
# instantiate the DecisionTreeClassifier model with criterion gini index
clf_gini = DecisionTreeClassifier(criterion='gini', max_depth=3, random_state=0)
# fit the model
clf_gini.fit(X_train, y_train)
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=3,
                       max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0,
                       random_state=0, splitter='best')
```

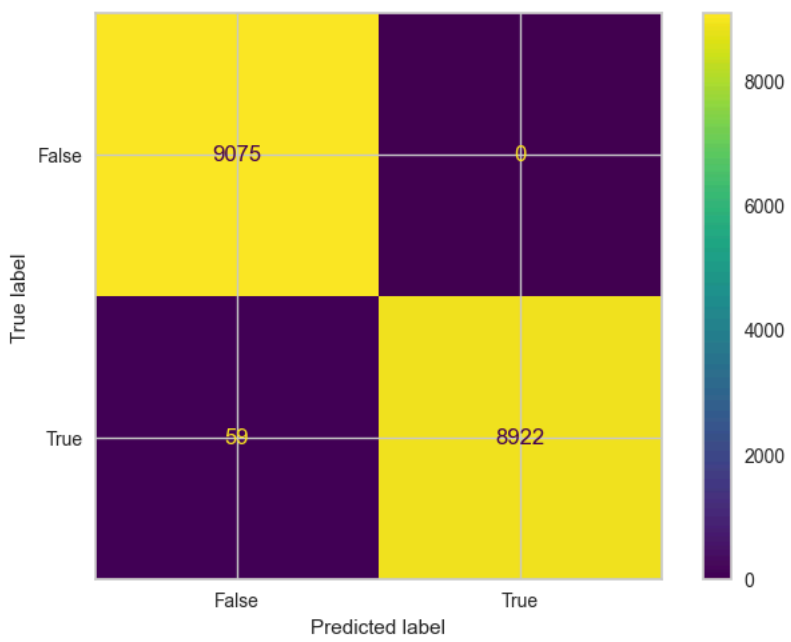
```
DecisionTreeClassifier(max_depth=3, random_state=0)
```

```
print(classification_report(y_test, y_pred_gini))
```

```
              precision    recall  f1-score   support

           0       0.99      1.00      1.00      9075
           1       1.00      0.99      1.00      8981

    accuracy                           1.00     18056
   macro avg       1.00      1.00      1.00     18056
weighted avg       1.00      1.00      1.00     18056
```

```
y_pred_gini = clf_gini.predict(X_test)
```

```
dt_cm = confusion_matrix(y_test, y_pred_gini)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = dt_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```




```
print('Model accuracy score with criterion gini index: {0:0.4f}'. format(accuracy_score(y_test, y_pred_gini)))
```

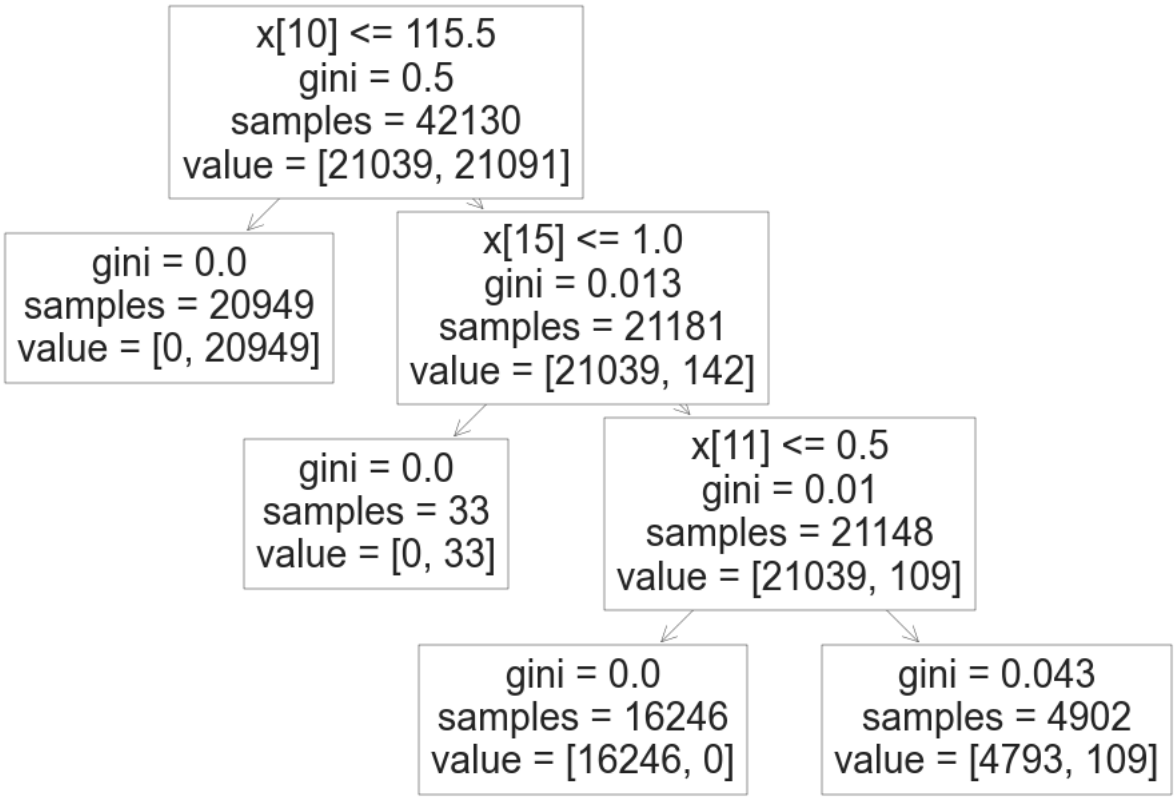
```
Model accuracy score with criterion gini index: 0.9967
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(clf_gini.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(clf_gini.score(X_test, y_test)))
```

```
Training set score: 0.9974
Test set score: 0.9967
```

```
plt.figure(figsize=(12,8))
from sklearn import tree
tree.plot_tree(clf_gini.fit(X_train, y_train))
```

```
[Text(0.3333333333333333, 0.875, 'x[10] <= 115.5\ngini = 0.5\nsamples = 42130\nvalue = [21039, 21091]'),
 Text(0.16666666666666666, 0.625, 'gini = 0.0\nsamples = 20949\nvalue = [0, 20949]'),
 Text(0.5, 0.625, 'x[15] <= 1.0\ngini = 0.013\nsamples = 21181\nvalue = [21039, 142]'),
 Text(0.3333333333333333, 0.375, 'gini = 0.0\nsamples = 33\nvalue = [0, 33]'),
 Text(0.6666666666666666, 0.375, 'x[11] <= 0.5\ngini = 0.01\nsamples = 21148\nvalue = [21039, 109]'),
 Text(0.5, 0.125, 'gini = 0.0\nsamples = 16246\nvalue = [16246, 0]'),
 Text(0.8333333333333334, 0.125, 'gini = 0.043\nsamples = 4902\nvalue = [4793, 109]')]
```




```
pip install graphviz
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: graphviz in c:\users\megha\appdata\roaming\python\python311\site-packages (0.20.1)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

Random forest

```
from sklearn.ensemble import RandomForestClassifier
rfc = RandomForestClassifier(random_state=0)
rfc.fit(X_train, y_train)
```

```
▾ RandomForestClassifier
RandomForestClassifier(random_state=0)
```

```
print(classification_report(y_test, rf_y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9075
           1       1.00      1.00      1.00      8981

    accuracy                           1.00     18056
   macro avg       1.00      1.00      1.00     18056
weighted avg       1.00      1.00      1.00     18056
```

```
rf_y_pred = rfc.predict(X_test)
```

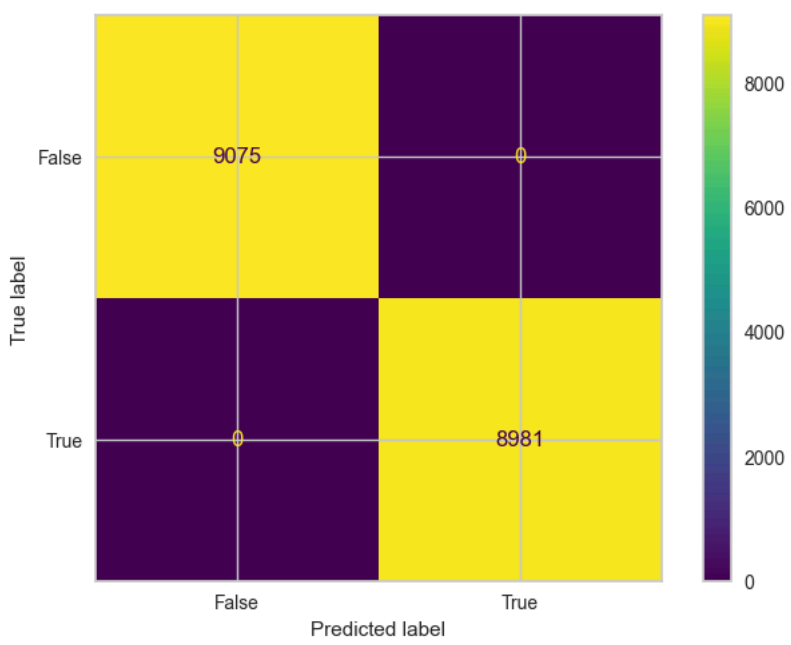
```
print('Model accuracy score with 10 decision-trees : {0:0.4f}'. format(accuracy_score(y_test, rf_y_pred)))
```

```
Model accuracy score with 10 decision-trees : 1.0000
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(rfc.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(rfc.score(X_test, y_test)))
```

```
Training set score: 1.0000
Test set score: 1.0000
```

```
rf_cm = confusion_matrix(y_test, rf_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = rf_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



Important features with Random Forest model

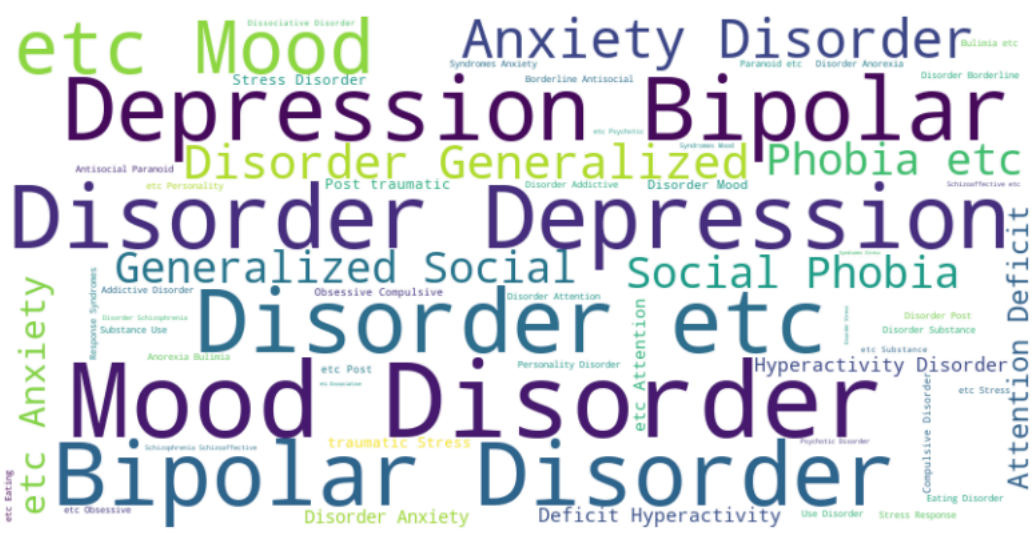
```
clf = RandomForestClassifier(n_estimators=100, random_state=0)
# fit the model to the training set
clf.fit(X_train, y_train)
RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
                       max_depth=None, max_features='auto', max_leaf_nodes=None,
                       min_impurity_decrease=0.0,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=0, verbose=0,
                       warm_start=False)
```

```
RandomForestClassifier
RandomForestClassifier(max_features='auto', random_state=0)
```

```
pip install wordcloud
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: wordcloud in c:\users\megha\appdata\roaming\python\python311\site-packages (1.9.2)
Requirement already satisfied: numpy>=1.6.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from wordcloud) (1.26
Requirement already satisfied: pillow in c:\users\megha\appdata\roaming\python\python311\site-packages (from wordcloud) (10.1.0)
Requirement already satisfied: matplotlib in c:\users\megha\appdata\roaming\python\python311\site-packages (from wordcloud) (3.8.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->w
Requirement already satisfied: cycler>=0.10 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->wordc
Requirement already satisfied: fonttools>=4.22.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib-
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib-
Requirement already satisfied: packaging>=20.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->wc
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->w
Requirement already satisfied: python-dateutil>=2.7 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotli
Requirement already satisfied: six>=1.5 in c:\users\megha\appdata\roaming\python\python311\site-packages (from python-dateutil>=2.7
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

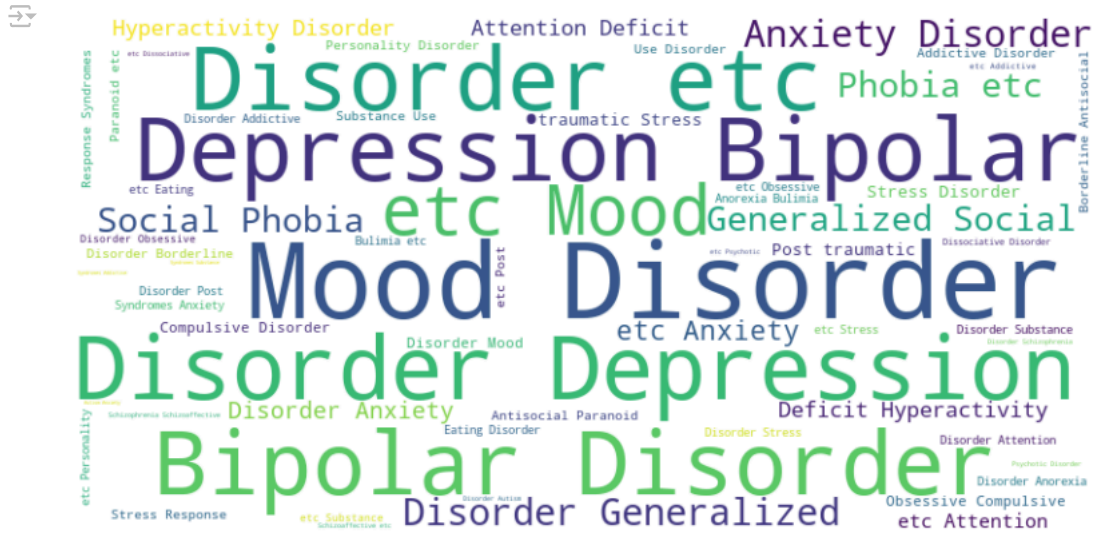
```
from wordcloud import WordCloud
import matplotlib.pyplot as plt
text_data = ' '.join(data['If so, what conditions were you diagnosed with'].dropna())
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text_data)

# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```




```
data['If so, what conditions were you diagnosed with'].value_counts()
```

```
If so, what conditions were you diagnosed with
Mood Disorder (Depression, Bipolar Disorder, etc)
7854
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)
6300
Anxiety Disorder (Generalized, Social, Phobia, etc)
2688
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit
Hyperactivity Disorder    1386
Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit Hyperactivity Disorder
1218

...
Stress Response Syndromes|Mood Disorder (Depression, Bipolar Disorder, etc)
42
Substance Use Disorder|Addictive Disorder
42
Mood Disorder (Depression, Bipolar Disorder, etc)|Suicidal Ideation
42
Anxiety Disorder (Generalized, Social, Phobia, etc)|Eating Disorder (Anorexia, Bulimia, etc)
42
autism spectrum disorder
42
Name: count, Length: 116, dtype: int64
```

```
text_data = ' '.join(data['If yes, what conditions have you been diagnosed with'].dropna())
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text_data)

# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```




```
data['If yes, what conditions have you been diagnosed with'].value_counts()
```

```
If yes, what conditions have you been diagnosed with
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)
4914
Mood Disorder (Depression, Bipolar Disorder, etc)
4284
Anxiety Disorder (Generalized, Social, Phobia, etc)
1974
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit
Hyperactivity Disorder
1092
Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit Hyperactivity Disorder
924

...
Anxiety Disorder (Generalized, Social, Phobia, etc)|Attention Deficit Hyperactivity Disorder|Obsessive-Compulsive Disorder
42
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit
Hyperactivity Disorder|Personality Disorder (Borderline, Antisocial, Paranoid, etc)|Obsessive-Compulsive Disorder|Post-traumatic
Stress Disorder      42
Addictive Disorder
42
Mood Disorder (Depression, Bipolar Disorder, etc)|Addictive Disorder
42
Obsessive-Compulsive Disorder|Eating Disorder (Anorexia, Bulimia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Anxiety
Disorder (Generalized, Social, Phobia, etc)
42
Name: count, Length: 128, dtype: int64
```

```
text_data = ' '.join(data['If maybe, what conditions do you believe you have'].dropna())
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text_data)

# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```

```
data['If maybe, what conditions do you believe you have'].value_counts()
```

```
If maybe, what conditions do you believe you have
Mood Disorder (Depression, Bipolar Disorder, etc)
2310
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)
2058
Anxiety Disorder (Generalized, Social, Phobia, etc)
1470
Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit Hyperactivity Disorder
420
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit
Hyperactivity Disorder                                                                                   420

...
Psychotic Disorder (Schizophrenia, Schizoaffective, etc)
42
Anxiety Disorder (Generalized, Social, Phobia, etc)|Post-traumatic Stress Disorder|Stress Response Syndromes
42
post-partum / anxiety
42
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Personality Disorder
(Borderline, Antisocial, Paranoid, etc)|Substance Use Disorder|Addictive Disorder                      42
Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc)|Attention Deficit
Hyperactivity Disorder|Personality Disorder (Borderline, Antisocial, Paranoid, etc)|Addictive Disorder      42
Name: count, Length: 99, dtype: int64
```

```
text_data = ' '.join(data['Which of the following best describes your work position'].dropna())
wordcloud = WordCloud(width=800, height=400, background_color='white',colormap='plasma').generate(text_data)

# Display the word cloud using matplotlib
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```


```
data['Which of the following best describes your work position'].value_counts()
```

```
Which of the following best describes your work position
Back-end Developer                                                                                    11046
Front-end Developer                                                                                    5250
Other                                                                                                  4704
Supervisor/Team Lead                                                                                   2856
Back-end Developer|Front-end Developer                                                                 2562
                                                                                                        ...
Front-end Developer|Back-end Developer|Support|Dev Evangelist/Advocate                                   42
Front-end Developer|Back-end Developer|Support|DevOps/SysAdmin|Dev Evangelist/Advocate                   42
Other|Dev Evangelist/Advocate|DevOps/SysAdmin|Support|Back-end Developer|Front-end Developer|One-person shop   42
Dev Evangelist/Advocate|Support|Back-end Developer|Front-end Developer                                   42
Supervisor/Team Lead|Front-end Developer|Back-end Developer|DevOps/SysAdmin|Dev Evangelist/Advocate      42
Name: count, Length: 264, dtype: int64
```

```
data['Question'].value_counts()
```

```
Question
Does your employer provide mental health benefits as part of healthcare coverage
1433
Would you be willing to bring up a physical health issue with a potential employer in an interview
1433
Did your previous employers provide resources to learn more about mental health issues and how to seek help
1433
```

```
Was your anonymity protected if you chose to take advantage of mental health or substance abuse treatment resources with previous employers                                          1433
Do you think that discussing a mental health disorder with previous employers would have negative consequences                                                                    1433
Do you think that discussing a physical health issue with previous employers would have negative consequences                                                                     1433
Would you have been willing to discuss a mental health issue with your previous coworkers                                                                                         1433
Would you have been willing to discuss a mental health issue with your direct supervisors                                                                                         1433
Did you feel that your previous employers took mental health as seriously as physical health                                                                                      1433
Did you hear of or observe negative consequences for coworkers with mental health issues in your previous workplaces                                                              1433
Why or why not                                                                                                                                                                    1433
Do you know the options for mental health care available under your employerprovided coverage                                                                                     1433
Would you bring up a mental health issue with a potential employer in an interview                                                                                                1433
Why or why not2                                                                                                                                                                   1433
Do you feel that being identified as a person with a mental health issue would hurt your career                                                                                   1433
Do you think that team memberscoworkers would view you more negatively if they knew you suffered from a mental health issue                                                       1433
How willing would you be to share with friends and family that you have a mental illness                                                                                          1433
Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace                                         1433
Have your observations of how another individual who discussed a mental health disorder made you less likely to reveal a mental health issue yourself in your current workplace    1433
If you have a mental health issue, do you feel that it interferes with your work when being treated effectively                                                                   1433
Did your previous employers ever formally discuss mental health as part of a wellness campaign or other official communication                                                    1433
Were you aware of the options for mental health care provided by your previous employers                                                                                          1433
Have your previous employers provided mental health benefits                                                                                                                      1433
If yes, what percentage of your work time time performing primary or secondary job functions is affected by a mental health issue                                                 1433
Has your employer ever formally discussed mental health for example, as part of a wellness campaign or other official communic                                                   1433
Does your employer offer resources to learn more about mental health concerns and options for seeking help                                                                        1433
Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by                                                   1433
If a mental health issue prompted you to request a medical leave from work, asking for that leave would be                                                                        1433
Do you think that discussing a mental health disorder with your employer would have negative consequences
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60186 entries, 0 to 60185
Data columns (total 27 columns):
 #   Column                                                                                  Non-Null Count  Dtype
---  ------                                                                                  --------------  -----
 0   ResponseID                                                                              60186 non-null  object
 1   Are you selfemployed                                                                    60186 non-null  bool
 2   How many employees does your company or organization have                               48132 non-null  object
 3   Is your employer primarily a tech companyorganization                                   48132 non-null  object
 4   Is your primary role within your company related to techIT                              11046 non-null  object
 5   Do you have previous employers                                                          60186 non-null  bool
 6   Do you have a family history of mental illness                                          60186 non-null  object
 7   Have you had a mental health disorder in the past                                       60186 non-null  object
 8   Do you currently have a mental health disorder                                          60186 non-null  object
 9   If yes, what conditions have you been diagnosed with                                    23856 non-null  object
 10  If maybe, what conditions do you believe you have                                       13524 non-null  object
 11  Have you been diagnosed with a mental health condition by a medical professional        60186 non-null  bool
 12  If so, what conditions were you diagnosed with                                          29862 non-null  object
 13  Have you ever sought treatment for a mental health issue from a mental health professional  60186 non-null  bool
 14  What is your age                                                                        60102 non-null  float64
 15  What is your gender                                                                     60186 non-null  object
 16  Age Group                                                                               60186 non-null  object
 17  What country do you live in                                                             60186 non-null  object
 18  What US state or territory do you live in                                               35280 non-null  object
 19  What country do you work in                                                             60186 non-null  object
 20  What US state or territory do you work in                                               35742 non-null  object
 21  Which of the following best describes your work position                                60186 non-null  object
 22  Do you work remotely                                                                    60186 non-null  object
 23  Question Group                                                                          51588 non-null  object
 24  Question about speaking openly about mental health vs physical health                   60186 non-null  object
 25  Question                                                                                60186 non-null  object
 26  Response                                                                                43846 non-null  object
dtypes: bool(4), float64(1), object(22)
memory usage: 10.8+ MB
```

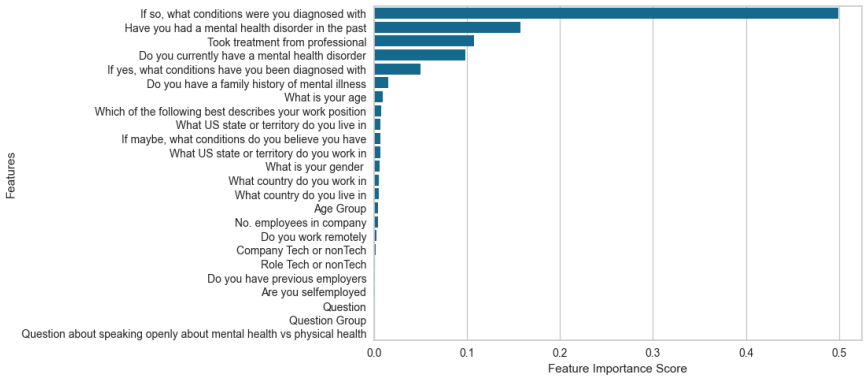
image.png

image.png

```
feature_scores = pd.Series(clf.feature_importances_, index=X_train.columns).sort_values(ascending=False)
feature_scores
```

```
If so, what conditions were you diagnosed with                          0.499669
Have you had a mental health disorder in the past                       0.157649
Took treatment from professional                                        0.107735
Do you currently have a mental health disorder                          0.098296
If yes, what conditions have you been diagnosed with                    0.050046
Do you have a family history of mental illness                          0.014947
What is your age                                                        0.009489
Which of the following best describes your work position                0.008000
What US state or territory do you live in                               0.007240
If maybe, what conditions do you believe you have                       0.006777
What US state or territory do you work in                               0.006757
What is your gender                                                     0.005707
What country do you work in                                             0.005260
What country do you live in                                             0.005146
Age Group                                                               0.004554
No. employees in company                                                0.004392
Do you work remotely                                                    0.002820
Company Tech or nonTech                                                 0.001980
Role Tech or nonTech                                                    0.001298
Do you have previous employers                                          0.001165
Are you selfemployed                                                    0.000902
Question                                                                0.000096
Question Group                                                          0.000062
Question about speaking openly about mental health vs physical health   0.000014
dtype: float64
```

```
# Creating a seaborn bar plot
sns.barplot(x=feature_scores, y=feature_scores.index)
# Add labels to the graph
plt.xlabel('Feature Importance Score')
plt.ylabel('Features')
# Visualize the graph
plt.show()
```



Ensemble Techniques

```
from sklearn.ensemble import VotingClassifier
ensemble_model = VotingClassifier(estimators=[('KNN', knn),('Random Forest',rfc),('Decision Tree',clf_gini)], voting='soft')
ensemble_model.fit(X_train,y_train)
ense_y_pred=ensemble_model.predict(X_test)
```

```
print(classification_report(y_test, ense_y_pred))
```

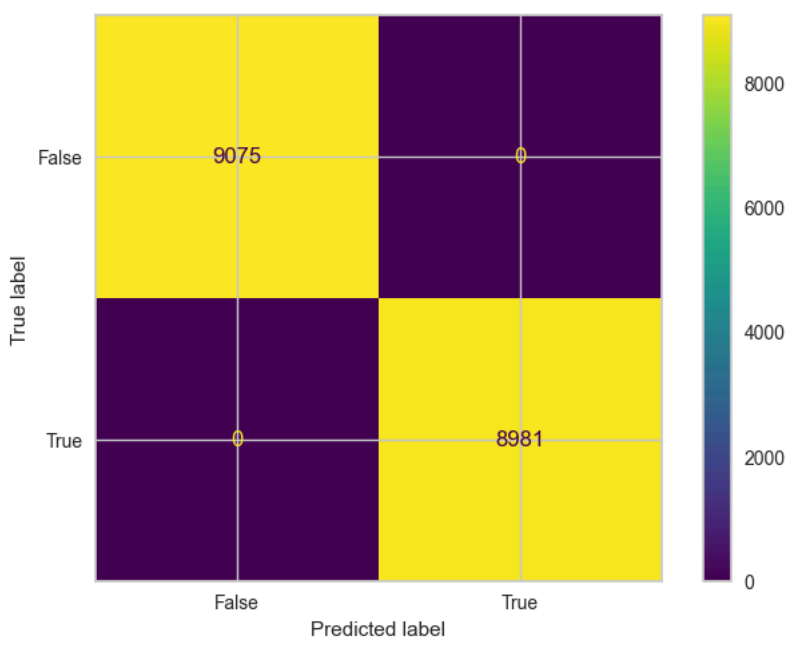
```
              precision    recall  f1-score   support
```

```
              0       1.00      1.00      1.00      9075
              1       1.00      1.00      1.00      8981

       accuracy                           1.00     18056
      macro avg       1.00      1.00      1.00     18056
   weighted avg       1.00      1.00      1.00     18056
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(ensemble_model.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(ensemble_model.score(X_test, y_test)))
```

```
Training set score: 1.0000
Test set score: 1.0000
```

```
ensemble_cm = confusion_matrix(y_test, ense_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = ensemble_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



XGBoost

```
pip install xgboost
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: xgboost in c:\users\megha\appdata\roaming\python\python311\site-packages (2.0.2)
Requirement already satisfied: numpy in c:\users\megha\appdata\roaming\python\python311\site-packages (from xgboost) (1.26.1)
Requirement already satisfied: scipy in c:\users\megha\appdata\roaming\python\python311\site-packages (from xgboost) (1.11.3)
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
from xgboost import XGBClassifier
# fit model no training data
xgb_model = XGBClassifier()
xgb_model.fit(X_train, y_train)

# make predictions for test data
xgb_y_pred = xgb_model.predict(X_test)
```

```
print(classification_report(y_test, xgb_y_pred))
```

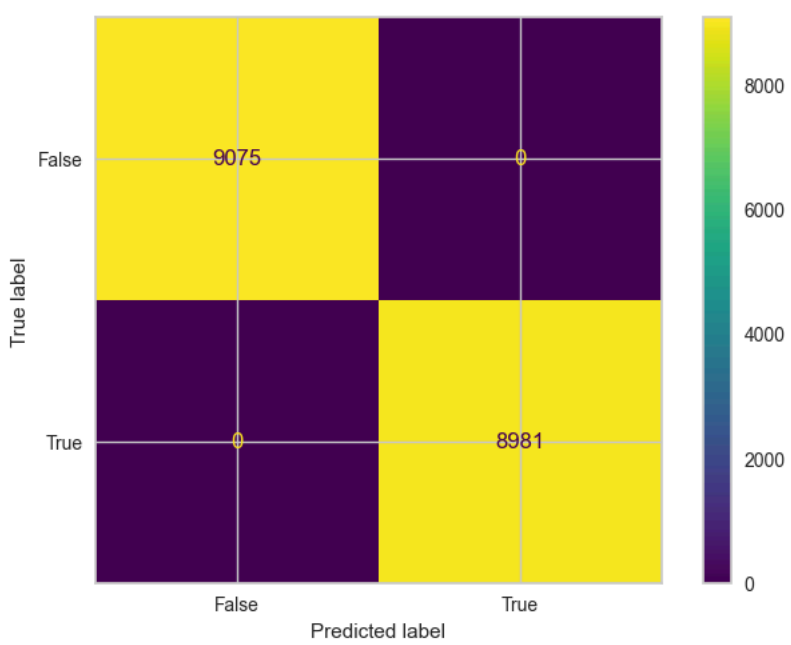
```
              precision    recall  f1-score   support

              0       1.00      1.00      1.00      9075
              1       1.00      1.00      1.00      8981

       accuracy                           1.00     18056
      macro avg       1.00      1.00      1.00     18056
   weighted avg       1.00      1.00      1.00     18056
```

```
#import pickle
#pickle.dump(xgb_model,open('osmi_xgb.pkl','wb'))
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(xgb_model.score(X_train.values, y_train)))
print('Test set score: {:.4f}'.format(xgb_model.score(X_test.values, y_test)))
```

```
Training set score: 1.0000
Test set score: 1.0000
```

```
xgb_cm = confusion_matrix(y_test, xgb_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = xgb_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



AdaBoost

```
from sklearn.ensemble import AdaBoostClassifier
# Create adaboost classifer object
adb_model = AdaBoostClassifier(n_estimators=50,learning_rate=1)
# Train Adaboost Classifer
adb_model.fit(X_train, y_train)
#Predict the response for test dataset
adb_y_pred = adb_model.predict(X_test)
```

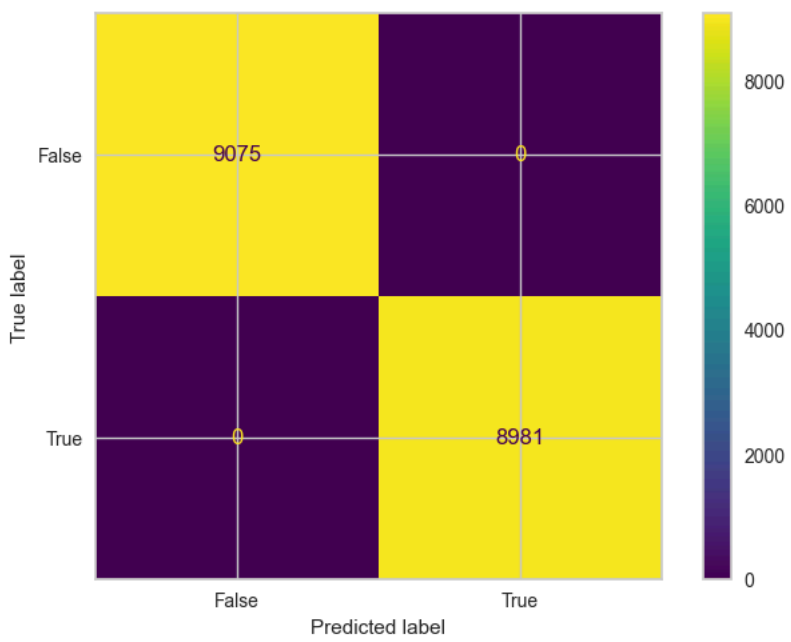
```
print(classification_report(y_test, adb_y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9075
           1       1.00      1.00      1.00      8981

    accuracy                           1.00     18056
   macro avg       1.00      1.00      1.00     18056
weighted avg       1.00      1.00      1.00     18056
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(adb_model.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(adb_model.score(X_test, y_test)))
```

```
Training set score: 1.0000
Test set score: 1.0000
```

```
adb_cm = confusion_matrix(y_test, adb_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = adb_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```



Gradient Booster Classifier

```
from sklearn.ensemble import GradientBoostingClassifier
gradient_booster = GradientBoostingClassifier(learning_rate=0.1)
gradient_booster.fit(X_train,y_train)
gb_y_pred=gradient_booster.predict(X_test)
```

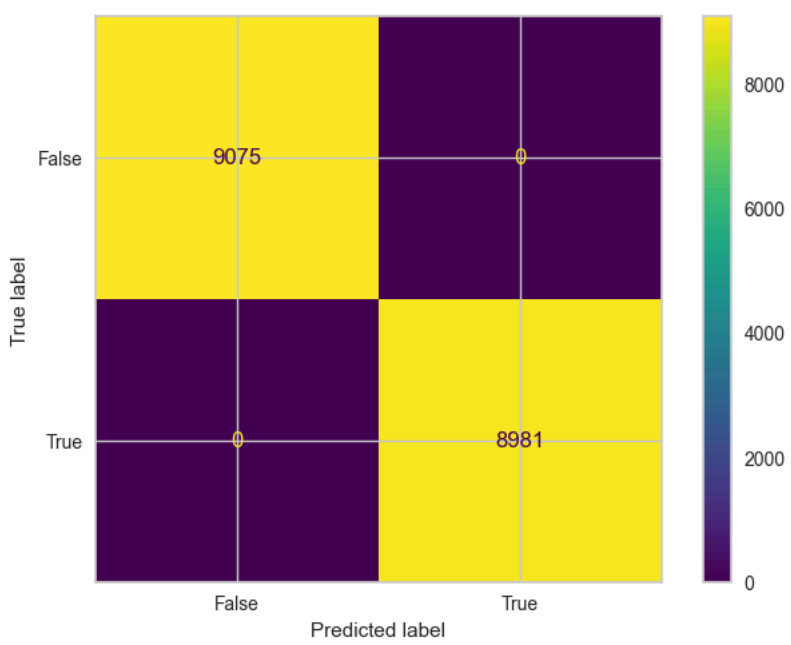
```
print(classification_report(y_test, gb_y_pred))
```

```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      9075
           1       1.00      1.00      1.00      8981

    accuracy                           1.00     18056
   macro avg       1.00      1.00      1.00     18056
weighted avg       1.00      1.00      1.00     18056
```

```
# print the scores on training and test set
print('Training set score: {:.4f}'.format(gradient_booster.score(X_train, y_train)))
print('Test set score: {:.4f}'.format(gradient_booster.score(X_test, y_test)))
```

```
Training set score: 1.0000
Test set score: 1.0000
```

```
gb_cm = confusion_matrix(y_test, gb_y_pred)
cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = gb_cm, display_labels = [False, True])
cm_display.plot()
plt.show()
```




```
accuracy = accuracy_score(y_test, ense_y_pred)
print(f'Ensemble technique Accuracy: {accuracy:.4f}')
```

```
Ensemble technique Accuracy: 1.0000
```

```
# Define the classifiers
models = {
    'K-Nearest Neighbors':KNeighborsClassifier(n_neighbors=3),
    'Logistic Regression': LogisticRegression(solver='liblinear', random_state=0),
    'Random Forest': RandomForestClassifier(random_state=0),
    'Decision Tree': DecisionTreeClassifier(criterion='gini', max_depth=3, random_state=0),
    'Gradient Booster':GradientBoostingClassifier(learning_rate=0.1),
    'AdaBoost': AdaBoostClassifier(n_estimators=50,learning_rate=1),
    'XGBoost': XGBClassifier()
}

# Train and evaluate each model
for name, clf in models.items():
    clf.fit(X_train, y_train)
    y_pred = clf.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    print(f'{name} Accuracy: {accuracy:.4f}')
```

```
K-Nearest Neighbors Accuracy: 1.0000
 Logistic Regression Accuracy: 0.9901
 Random Forest Accuracy: 1.0000
 Decision Tree Accuracy: 0.9967
 Gradient Booster Accuracy: 1.0000
 AdaBoost Accuracy: 1.0000
 XGBoost Accuracy: 1.0000
```

LIME - Explainable AI

```
pip install lime
```

```
Note: you may need to restart the kernel to use updated packages.Defaulting to user installation because normal site-packages is not
 Requirement already satisfied: lime in c:\users\megha\appdata\roaming\python\python311\site-packages (0.2.0.1)
 Requirement already satisfied: matplotlib in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (3.8.1)
 Requirement already satisfied: numpy in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (1.26.1)
 Requirement already satisfied: scipy in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (1.11.3)
 Requirement already satisfied: tqdm in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (4.66.1)
 Requirement already satisfied: scikit-learn>=0.18 in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (1.3
 Requirement already satisfied: scikit-image>=0.12 in c:\users\megha\appdata\roaming\python\python311\site-packages (from lime) (0.22
 Requirement already satisfied: networkx>=2.8 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-image>=0
 Requirement already satisfied: pillow>=9.0.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-image>=0
 Requirement already satisfied: imageio>=2.27 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-image>=0
 Requirement already satisfied: tifffile>=2022.8.12 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-ima
 Requirement already satisfied: packaging>=21 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-image>=0
 Requirement already satisfied: lazy_loader>=0.3 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-image>
 Requirement already satisfied: joblib>=1.1.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-learn>=0
 Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from scikit-le
 Requirement already satisfied: contourpy>=1.0.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->l
 Requirement already satisfied: cycler>=0.10 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->lime
 Requirement already satisfied: fonttools>=4.22.0 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->
```

```
Requirement already satisfied: kiwisolver>=1.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotlib->
Requirement already satisfied: python-dateutil>=2.7 in c:\users\megha\appdata\roaming\python\python311\site-packages (from matplotli
Requirement already satisfied: colorama in c:\users\megha\appdata\roaming\python\python311\site-packages (from tqdm->lime) (0.4.6)
Requirement already satisfied: six>=1.5 in c:\users\megha\appdata\roaming\python\python311\site-packages (from python-dateutil>=2.7


[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: python.exe -m pip install --upgrade pip
```

```
import lime
import lime.lime_tabular
```

```
explainer=lime.lime_tabular.LimeTabularExplainer(X_train.values, feature_names =
                                X.columns,
                                class_names=['Not diagnosed ','Diagnosed '],
                                verbose=True,
                                mode = 'classification')
```

```
columns=['Age Group',
'What is your age',
'Which of the following best describes your work position',
'Is your employer primarily a tech companyorganization',
'Have you ever sought treatment for a mental health issue from a mental health professional',
'Have you had a mental health disorder in the past',
'Do you work remotely',
'Do you have a family history of mental illness',
'Do you currently have a mental health disorder',
'If yes, what conditions have you been diagnosed with',
'Are you selfemployed'
]
for i in columns:
    val=data[i].value_counts()
    print(val)
```

```
Age Group
26-30       15708
31-35       15372
36-40       10668
20-25        6846
41-45        5922
46-50        2814
51-55        1512
56-60         504
Under 20      336
61-65         336
Over 65       168
Name: count, dtype: int64
What is your age
30.0    3948
31.0    3444
29.0    3318
28.0    3108
35.0    3108
32.0    3024
34.0    2898
33.0    2898
26.0    2688
27.0    2646
37.0    2478
39.0    2310
38.0    2268
36.0    2100
25.0    1848
24.0    1764
40.0    1512
22.0    1344
44.0    1302
43.0    1260
42.0    1218
45.0    1134
41.0    1008
23.0    1008
46.0     924
21.0     630
47.0     588
49.0     546
55.0     504
50.0     378
48.0     378
54.0     294
52.0     294
51.0     294
20.0     252
```

```
     56.0     210
     19.0     168
     57.0     168
     63.0     168
     53.0     126
     59.0      84
     61.0      84
```

```
columns=['Age Group',
'What is your age',
'Which of the following best describes your work position',
'Is your employer primarily a tech companyorganization',
'Have you ever sought treatment for a mental health issue from a mental health professional',
'Have you had a mental health disorder in the past',
'Do you work remotely',
'Do you have a family history of mental illness',
'Do you currently have a mental health disorder',
'If yes, what conditions have you been diagnosed with',
'Are you selfemployed'
]
for i in columns:
    val=data[i].nunique()
    print(i,":",val)
```

```
Age Group : 11
What is your age : 51
Which of the following best describes your work position : 264
Is your employer primarily a tech companyorganization : 2
Have you ever sought treatment for a mental health issue from a mental health professional : 2
Have you had a mental health disorder in the past : 3
Do you work remotely : 3
Do you have a family history of mental illness : 3
Do you currently have a mental health disorder : 3
If yes, what conditions have you been diagnosed with : 128
Are you selfemployed : 2
```

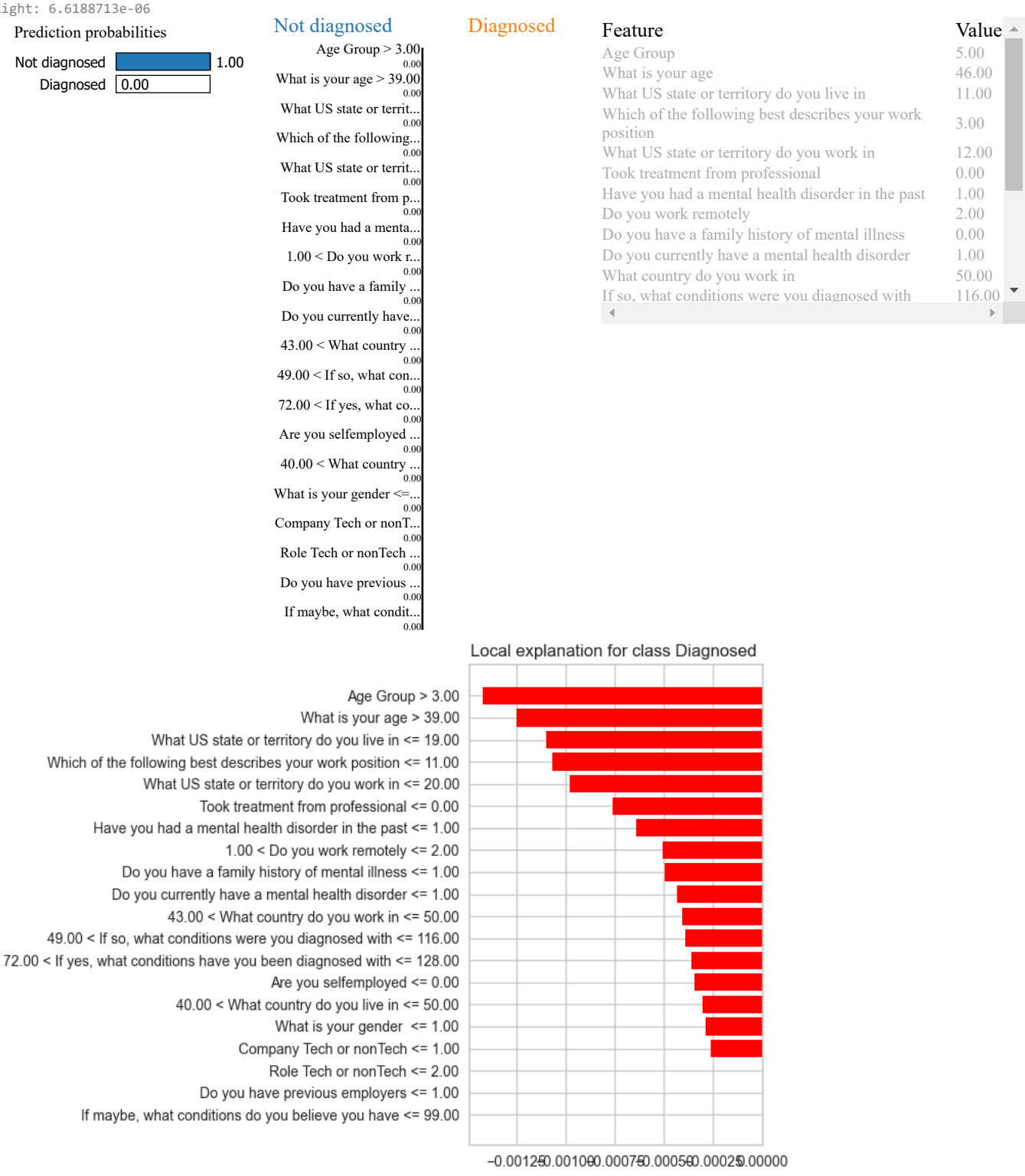
```
new_df_cap.shape
```

```
(60186, 21)
```

XGBoost with LIME

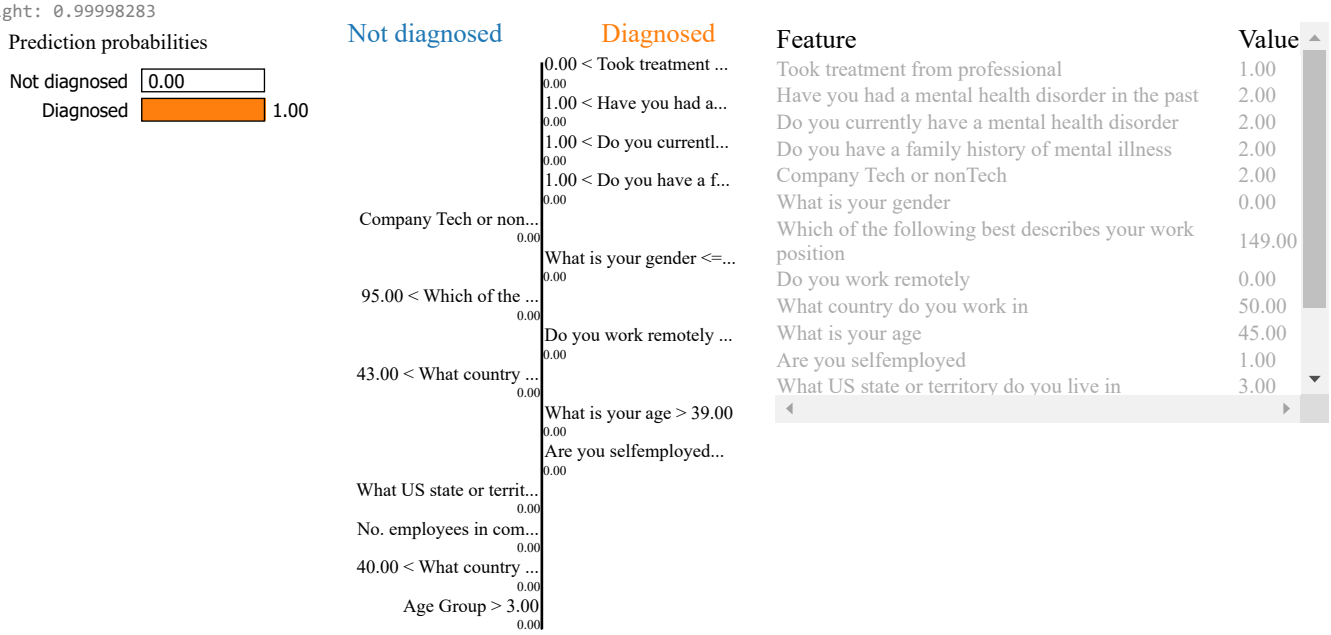
```
exp = explainer.explain_instance(X_test.values[0,:], xgb_model.predict_proba,num_features=20)
exp.show_in_notebook(show_table=True)
    #Plot local explanation
plt = exp.as_pyplot_figure()
plt.tight_layout()
```

```
Intercept 1.0043503422691562
Prediction_local [0.9933464]
Right: 6.6188713e-06
```



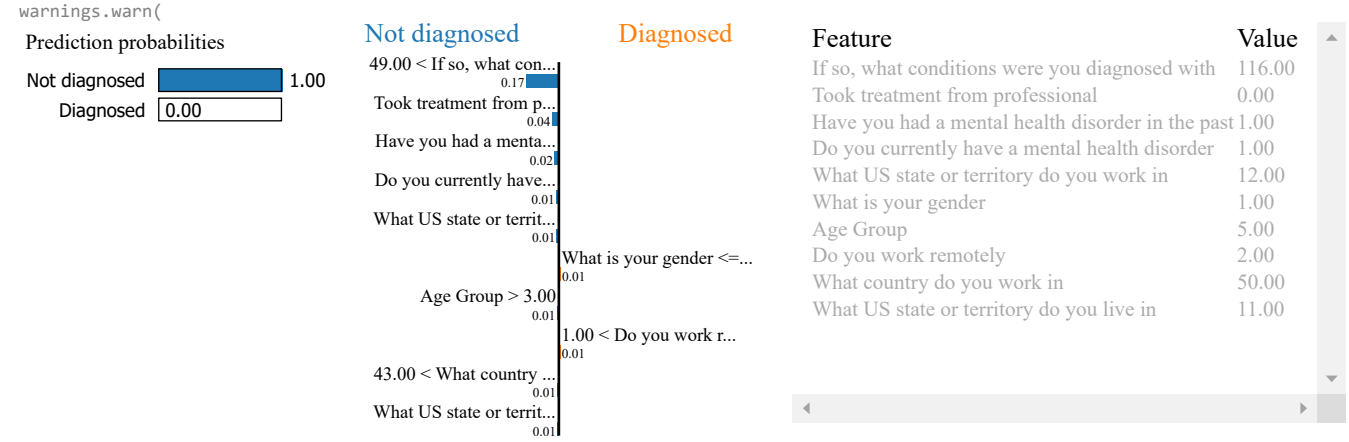

```
exp = explainer.explain_instance(X_test.values[1,:], xgb_model.predict_proba,num_features=15)
exp.show_in_notebook(show_table=True)
    #Plot local explanation
plt = exp.as_pyplot_figure()
plt.tight_layout()
```

```
Intercept 0.9998526710343509
Prediction_local [0.99999047]
Right: 0.99998283
```

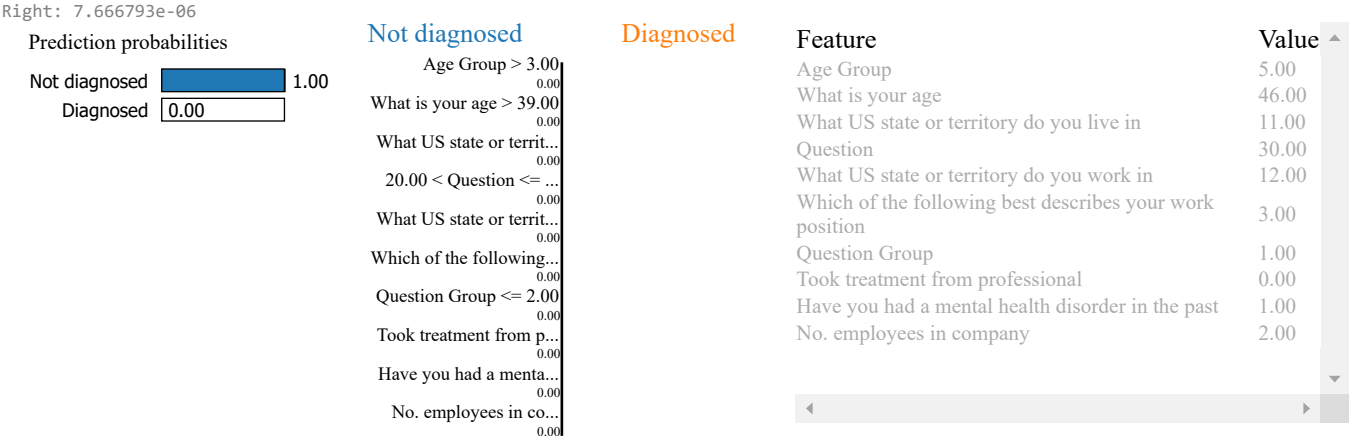



```
# List of models
models = [ensemble_model,xgb_model,gradient_booster]


for model in models:
    exp = explainer.explain_instance(X_test.values[0,:], model.predict_proba)
    exp.show_in_notebook(show_table=True)
    #Plot local explanation
    plt = exp.as_pyplot_figure()
    plt.tight_layout()
```

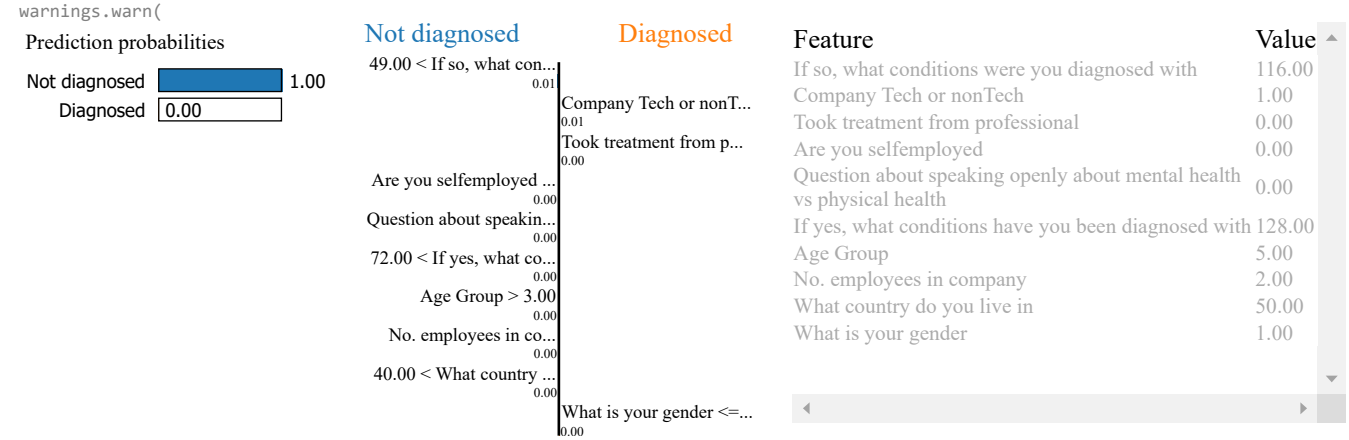
```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.9963890559712774
Prediction_local [0.73734024]
Right: 0.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

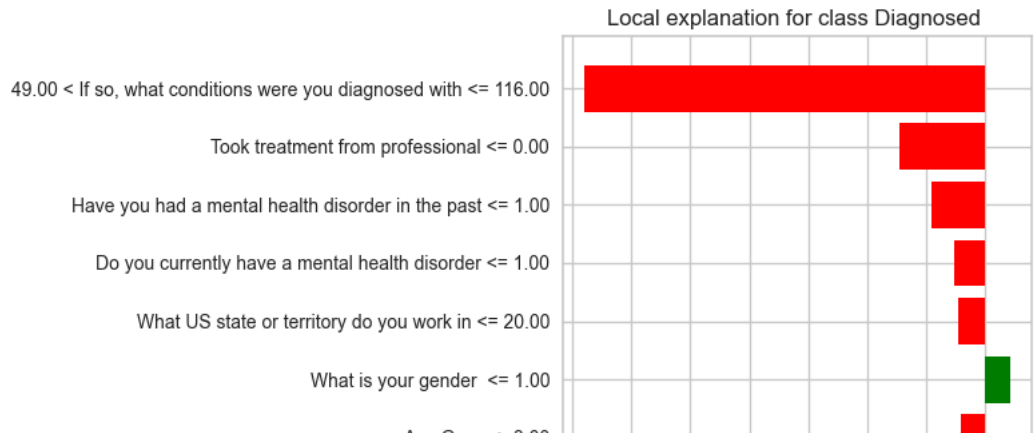


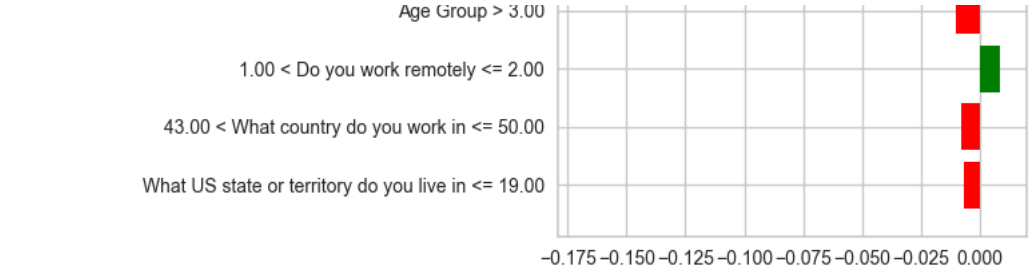

```
Intercept 1.002551749068273
Prediction_local [0.99295334]
Right: 7.666793e-06
```

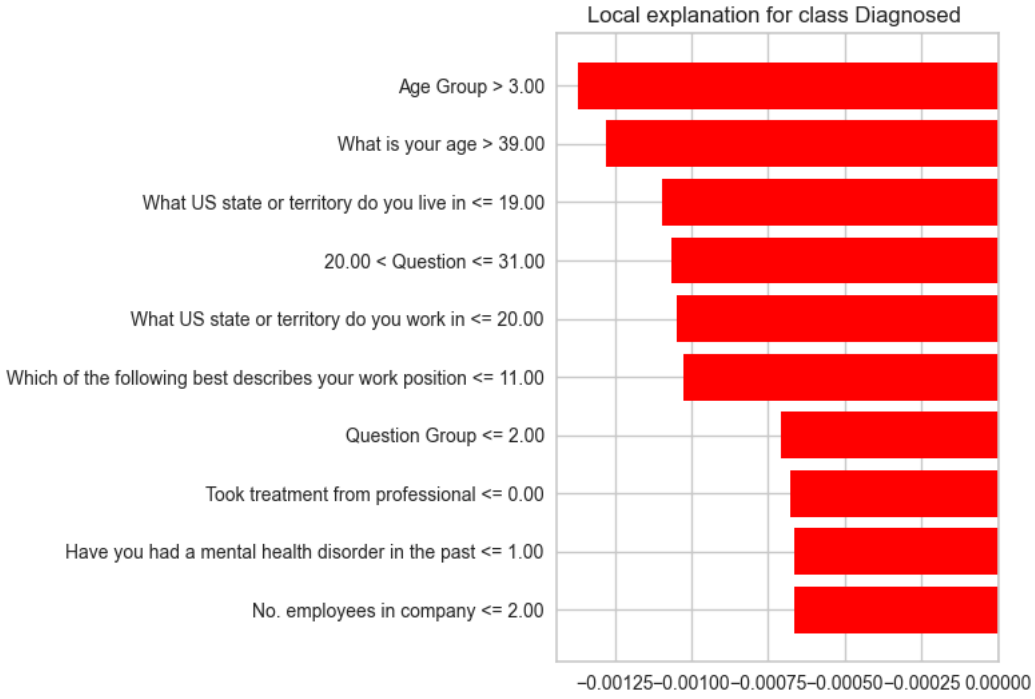


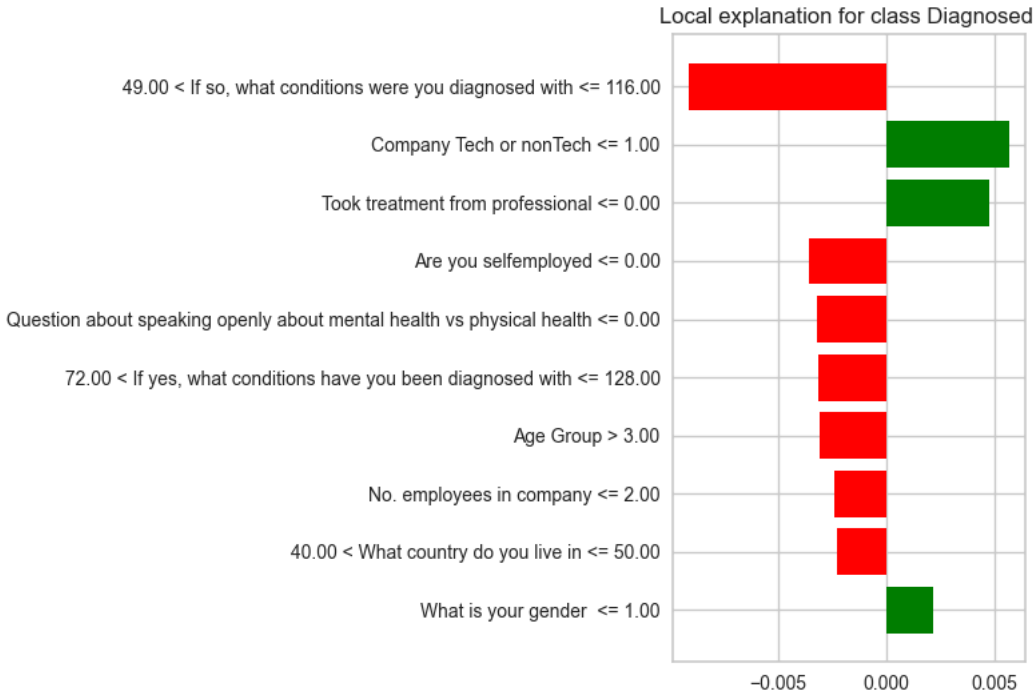

```
Intercept 1.0019850455127222
Prediction_local [0.98785155]
Right: 0.0002020675942082618
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```





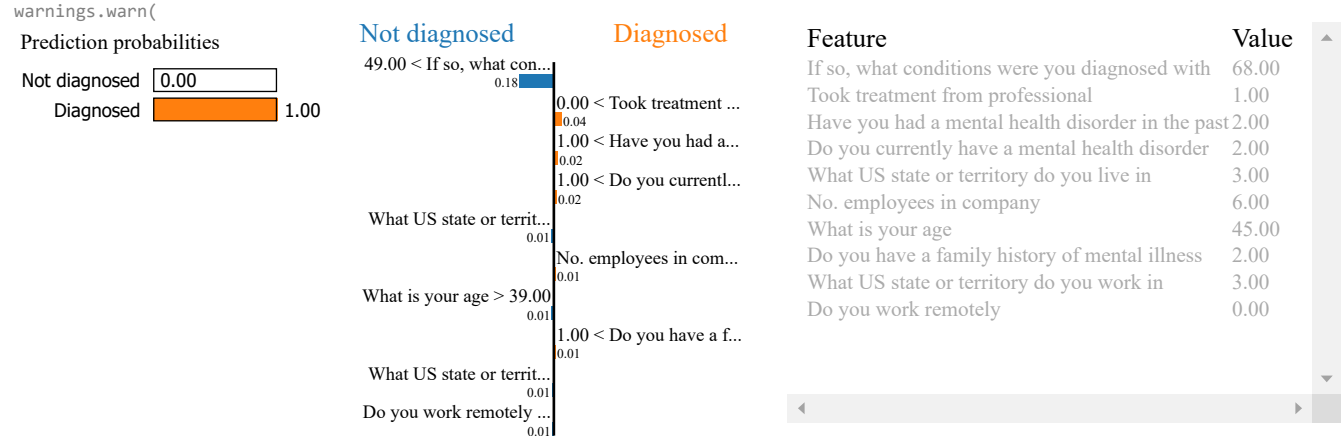




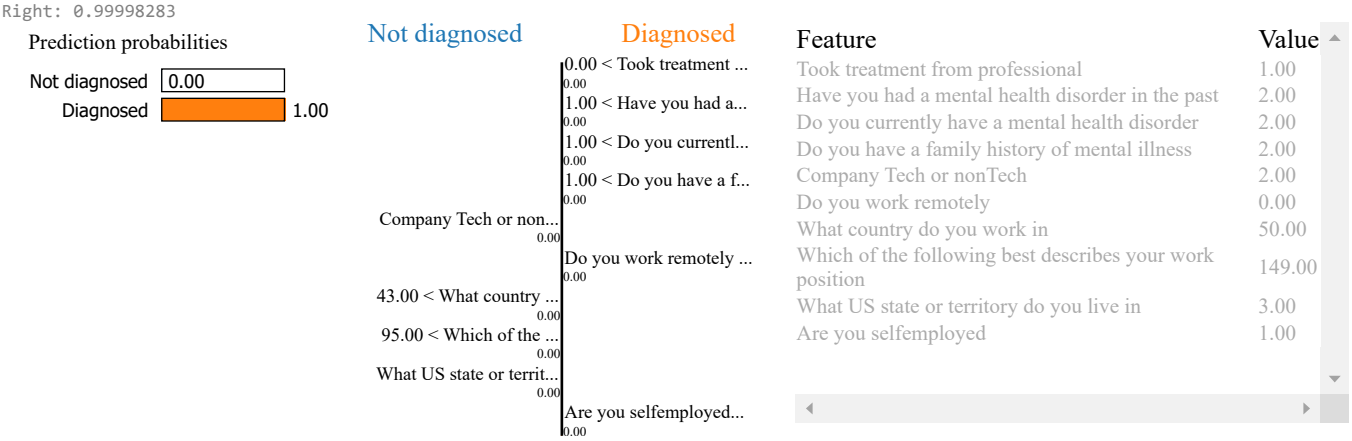



```
for model in models:
    exp = explainer.explain_instance(X_test.values[1,:], model.predict_proba)
    exp.show_in_notebook(show_table=True)
    #Plot local explanation
    plt = exp.as_pyplot_figure()
    plt.tight_layout()
```

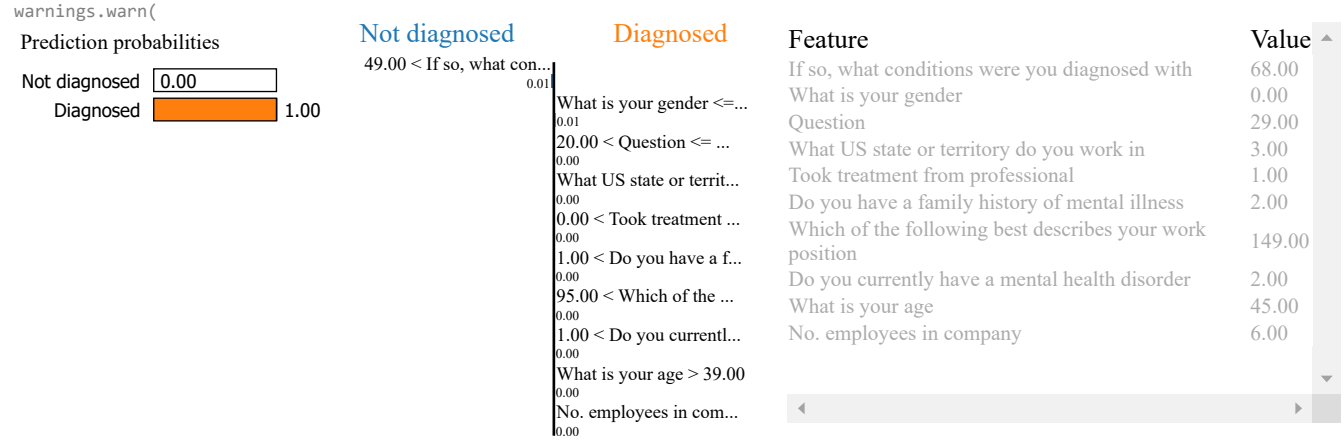
```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.9287602885525325
Prediction_local [0.80889337]
Right: 1.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

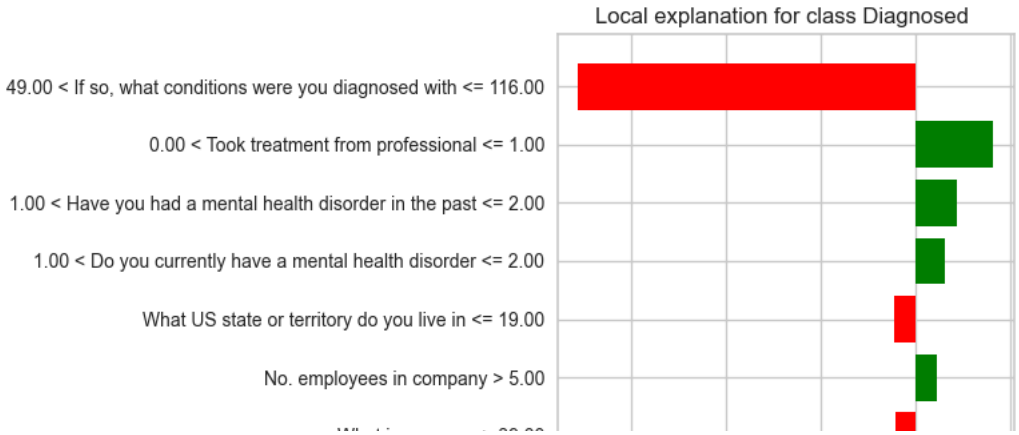


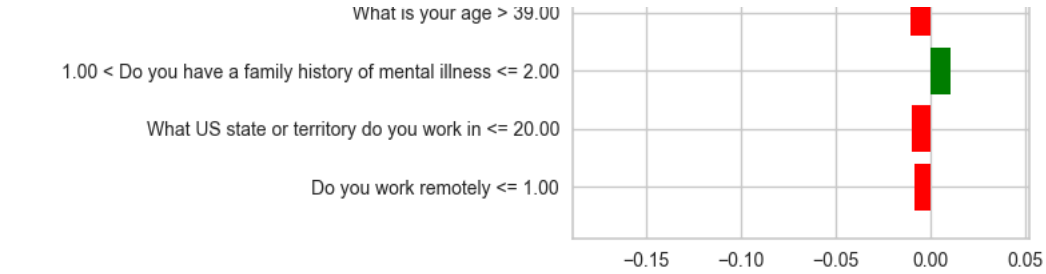

```
Intercept 0.9998614290528125
Prediction_local [0.99998832]
Right: 0.99998283
```

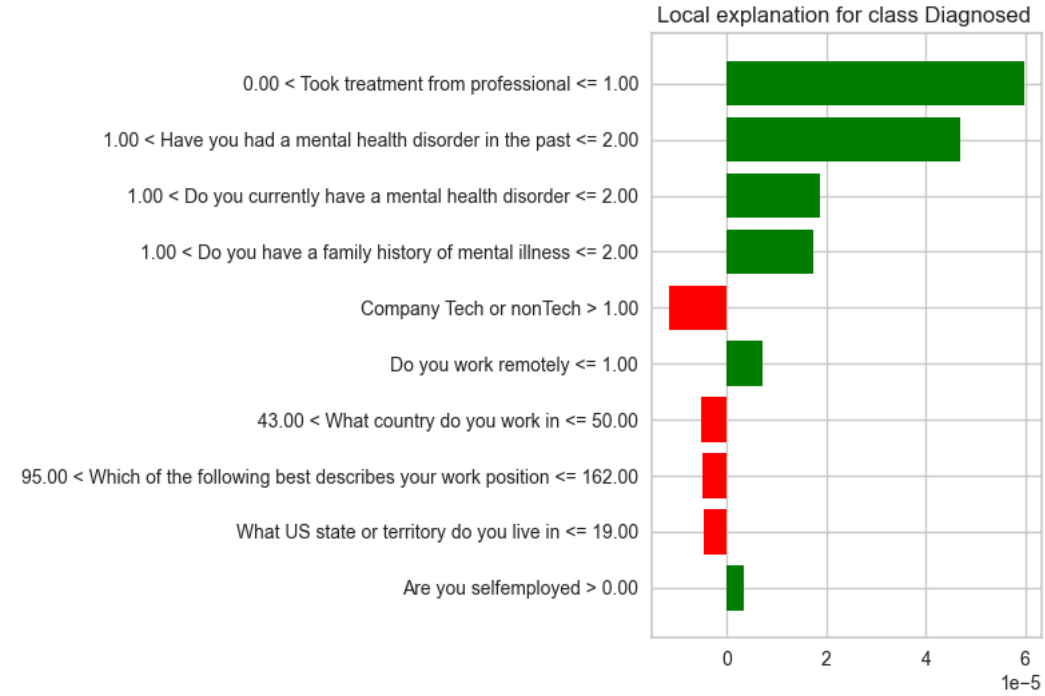


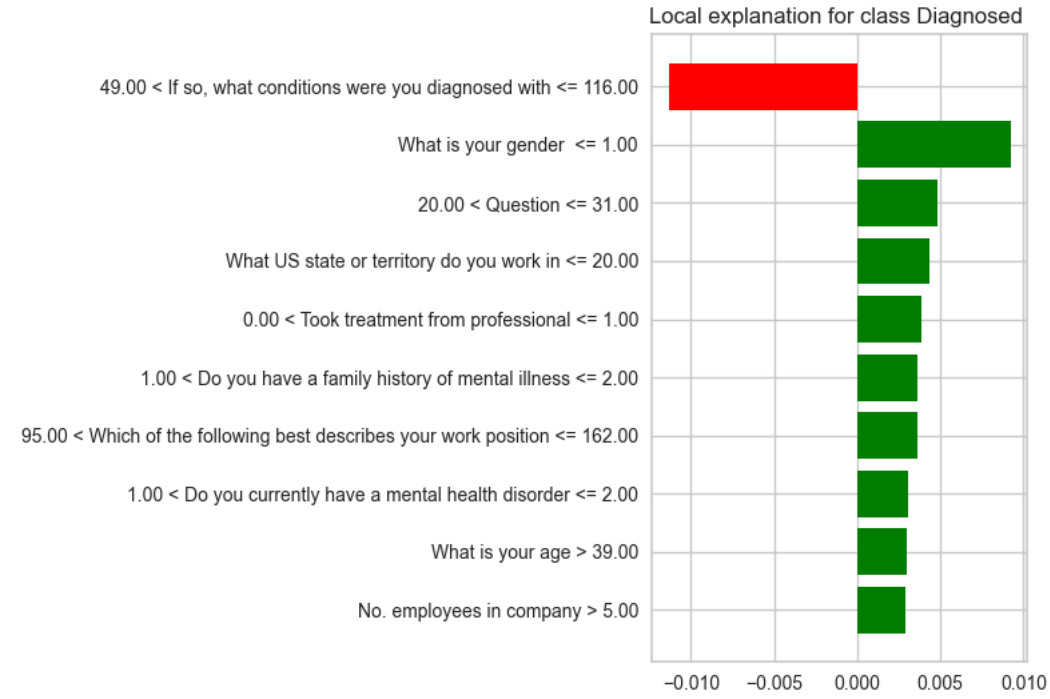

```
Intercept 0.9807405482631991
Prediction_local [1.00795955]
Right: 0.9997293265309739
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```







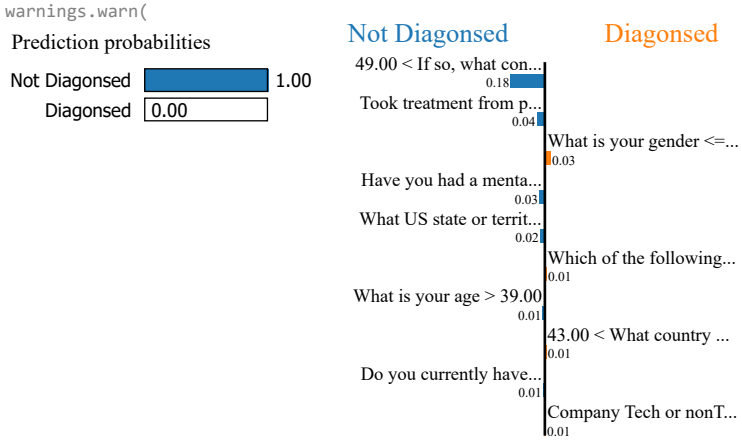





```
for model in models:
    exp = explainer.explain_instance(X_test.values[0,:], model.predict_proba)
    exp.show_in_notebook(show_table=True)
    #Plot local explanation
    plt = exp.as_pyplot_figure()
    plt.tight_layout()
```

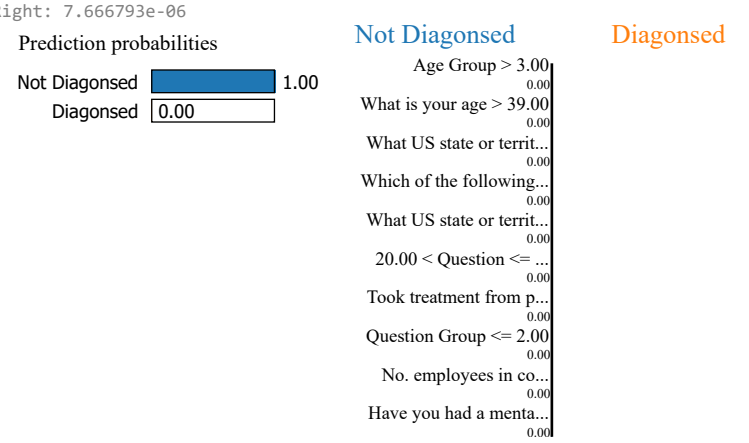
```
# Loop through each instance in the test set
for i in range(5):
    instance = X_test.iloc[i]
    for model in models:

        exp = explainer.explain_instance(instance.values, model.predict_proba)
        print(f"Explanation for instance {i + 1}:",  exp.show_in_notebook(show_table=True))
```

```
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.9666524453971248
Prediction_local [0.73585377]
Right: 0.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```



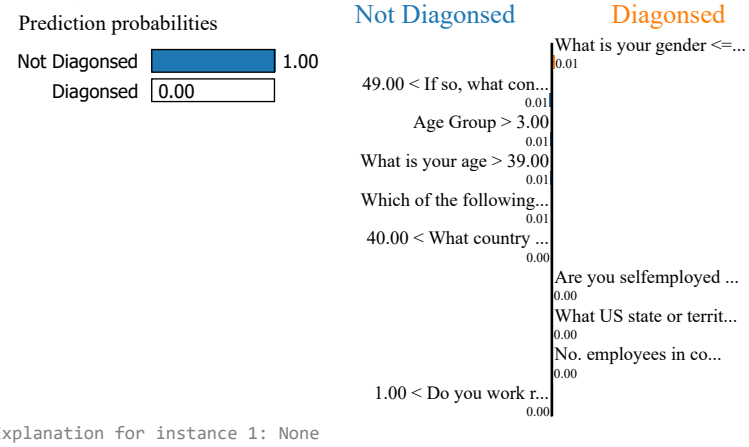
Feature	Value
If so, what conditions were you diagnosed with	116.00
Took treatment from professional	0.00
What is your gender	1.00
Have you had a mental health disorder in the past	1.00
What US state or territory do you work in	12.00
Which of the following best describes your work position	3.00
What is your age	46.00
What country do you work in	50.00
Do you currently have a mental health disorder	1.00
Company Tech or nonTech	1.00

```
Explanation for instance 1: None
Intercept 1.0026969572408468
Prediction_local [0.99239551]
Right: 7.666793e-06
```



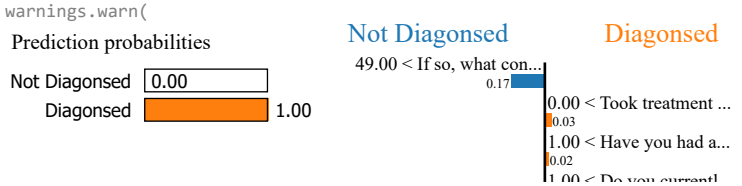
Feature	Value
Age Group	5.00
What is your age	46.00
What US state or territory do you work in	12.00
Which of the following best describes your work position	3.00
What US state or territory do you live in	11.00
Question	30.00
Took treatment from professional	0.00
Question Group	1.00
No. employees in company	2.00
Have you had a mental health disorder in the past	1.00

```
Explanation for instance 1: None
Intercept 0.990546431870831
Prediction_local [0.97787928]
Right: 0.0002020675942082618
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

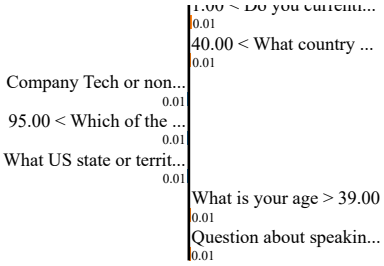


Feature	Value
What is your gender	1.00
If so, what conditions were you diagnosed with	116.00
Age Group	5.00
What is your age	46.00
Which of the following best describes your work position	3.00
What country do you live in	50.00
Are you selfemployed	0.00
What US state or territory do you work in	12.00
No. employees in company	2.00
Do you work remotely	2.00

```
Explanation for instance 1: None
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.9207240936775718
Prediction_local [0.81388507]
Right: 1.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

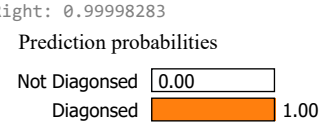


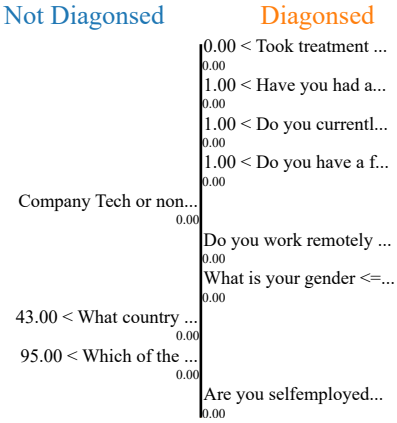
Feature	Value
If so, what conditions were you diagnosed with	68.00
Took treatment from professional	1.00
Have you had a mental health disorder in the past	2.00
Do you currently have a mental health disorder	2.00
What country do you live in	50.00



Feature	Value
Company Tech or nonTech	2.00
Which of the following best describes your work position	149.00
What US state or territory do you live in	3.00
What is your age	45.00
Question about speaking openly about mental health vs physical health	0.00

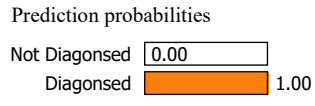
```
Explanation for instance 2: None
Intercept 0.9998493998088118
Prediction_local [0.99999334]
Right: 0.99998283
```

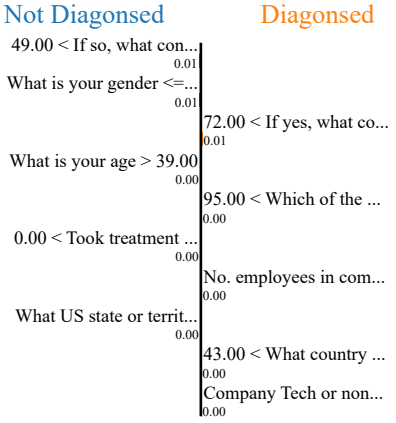




Feature	Value
Took treatment from professional	1.00
Have you had a mental health disorder in the past	2.00
Do you currently have a mental health disorder	2.00
Do you have a family history of mental illness	2.00
Company Tech or nonTech	2.00
Do you work remotely	0.00
What is your gender	0.00
What country do you work in	50.00
Which of the following best describes your work position	149.00
Are you selfemployed	1.00

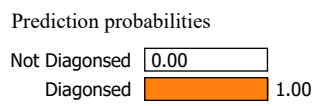
```
Explanation for instance 2: None
Intercept 1.0013282147499412
Prediction_local [0.99196278]
Right: 0.9997293265309739
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

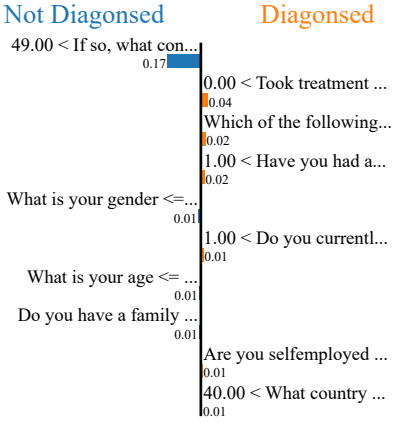




Feature	Value
If so, what conditions were you diagnosed with	68.00
What is your gender	0.00
If yes, what conditions have you been diagnosed with	83.00
What is your age	45.00
Which of the following best describes your work position	149.00
Took treatment from professional	1.00
No. employees in company	6.00
What US state or territory do you work in	3.00
What country do you work in	50.00
Company Tech or nonTech	2.00

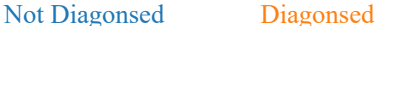
```
Explanation for instance 2: None
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.9344848007272911
Prediction_local [0.83107402]
Right: 1.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```



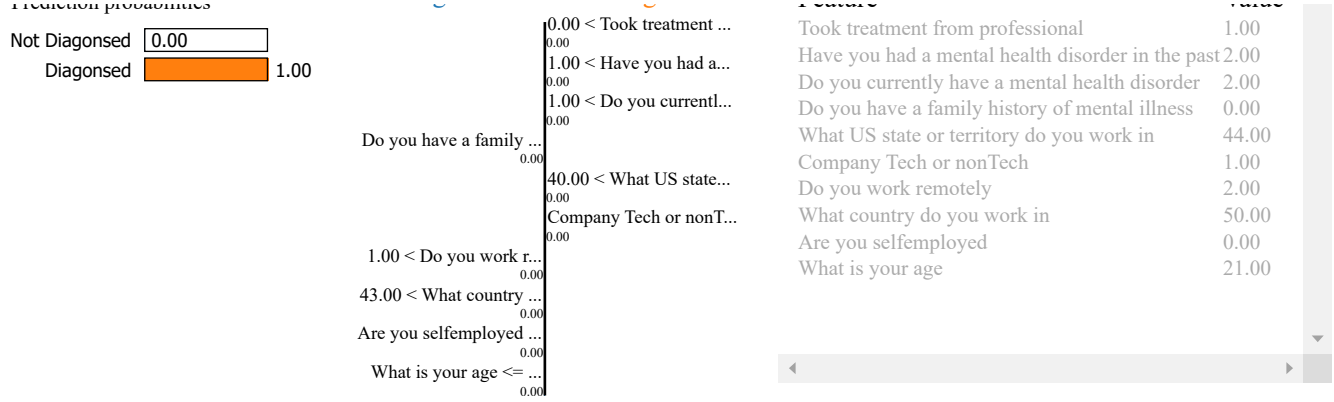


Feature	Value
If so, what conditions were you diagnosed with	68.00
Took treatment from professional	1.00
Which of the following best describes your work position	8.00
Have you had a mental health disorder in the past	2.00
What is your gender	1.00
Do you currently have a mental health disorder	2.00
What is your age	21.00
Do you have a family history of mental illness	0.00
Are you selfemployed	0.00
What country do you live in	50.00

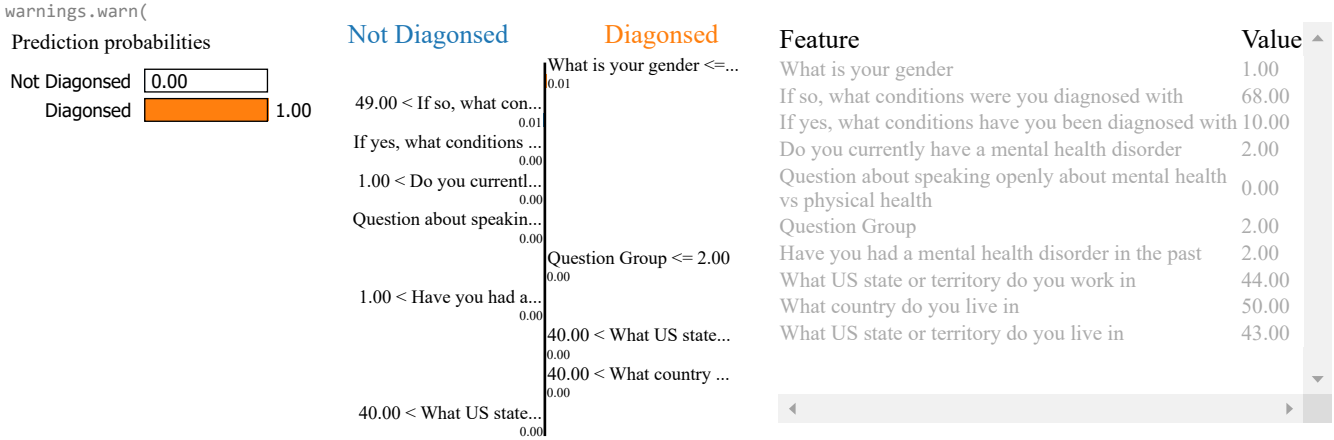
```
Explanation for instance 3: None
Intercept 0.9998655190615607
Prediction_local [0.99998621]
Right: 0.99997103
```

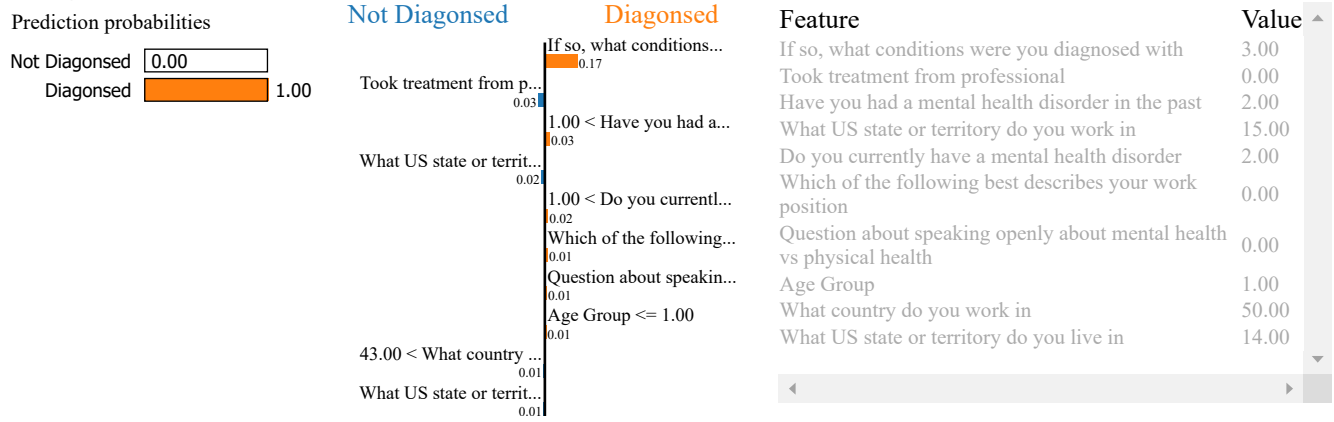
Feature Value



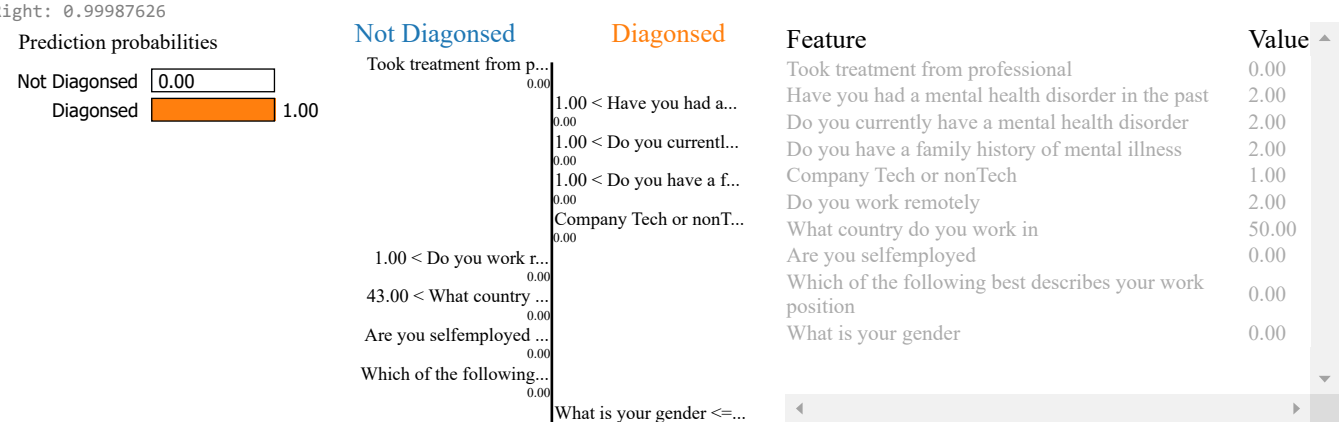

```
Explanation for instance 3: None
Intercept 0.9945819192358912
Prediction_local [0.98693082]
Right: 0.9997681672114347
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```



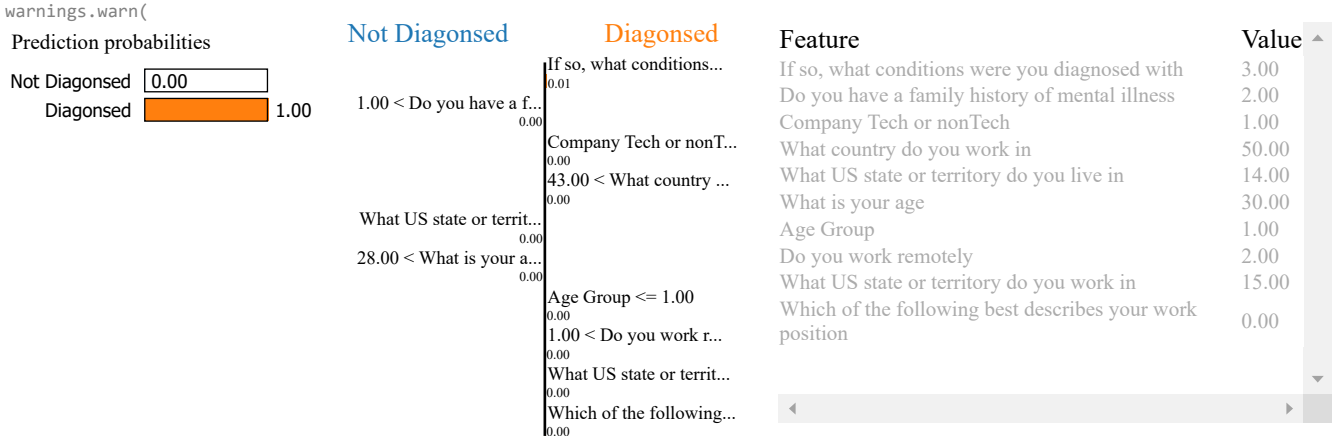

```
Explanation for instance 3: None
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.7868225105995573
Prediction_local [0.96359333]
Right: 1.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```



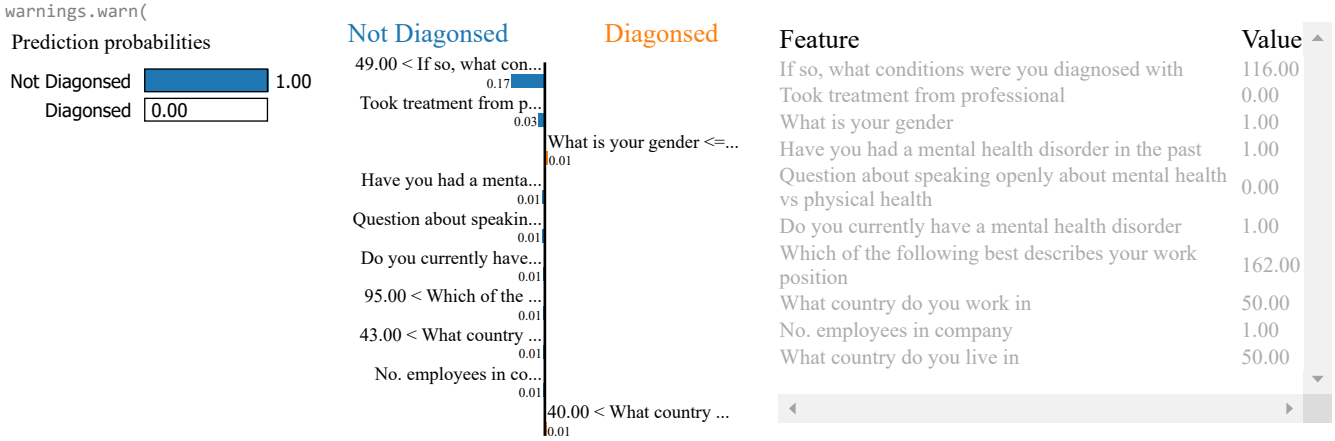

```
Explanation for instance 4: None
Intercept 0.9999144877329965
Prediction_local [0.99993343]
Right: 0.99987626
```



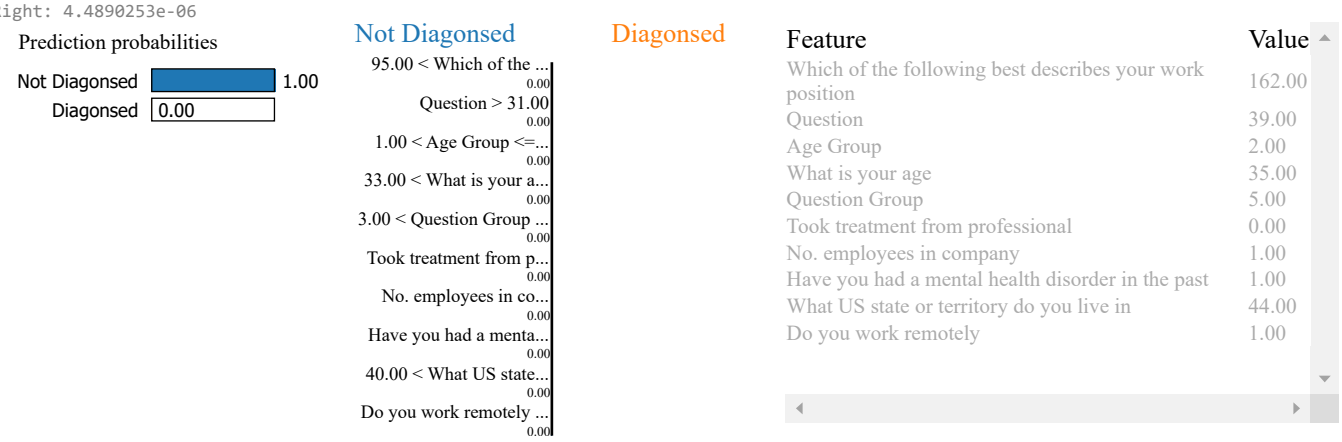
```
Explanation for instance 4: None
Intercept 0.9846110889073781
Prediction_local [1.00028861]
Right: 0.9997132044553101
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```



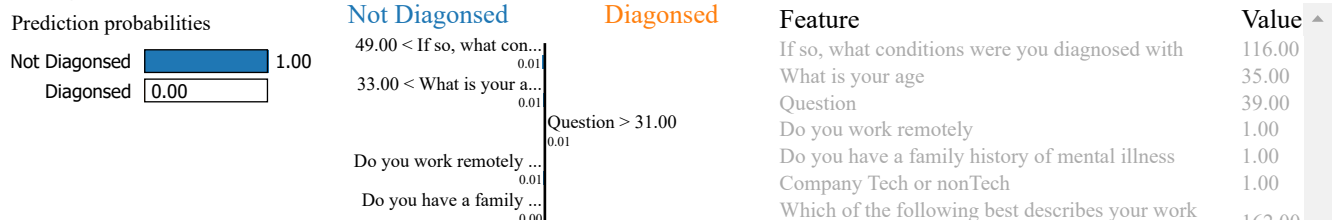

```
Explanation for instance 4: None
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
Intercept 0.990013755254641
Prediction_local [0.75066611]
Right: 0.0
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```

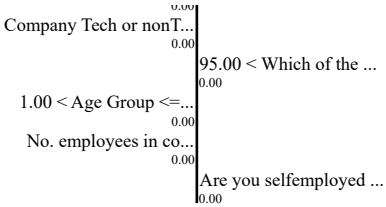



```
Explanation for instance 5: None
Intercept 1.0025889432037356
Prediction_local [0.99374435]
Right: 4.4890253e-06
```




```
Explanation for instance 5: None
Intercept 1.0086562972178035
Prediction_local [0.98011113]
Right: 0.000168592019811326
C:\Users\megha\AppData\Roaming\Python\Python311\site-packages\sklearn\base.py:465: UserWarning: X does not have valid feature names,
  warnings.warn(
```





position	162.00
Age Group	2.00
No. employees in company	1.00
Are you selfemployed	0.00

```
Explanation for instance 5: None
```

Negative correlations are shown in RED

Positive in GREEN

```
coef=pd.DataFrame(exp.as_list())[1].sum()
print('Sum of coeffients and intercept: ',coef+0.99)
pd.DataFrame(exp.as_list())
```

```
Sum of coeffients and intercept:  0.9858625699840917
```

	0	1
0	49.00 < If so, what conditions were you diagno...	-0.013247
1	No. employees in company > 5.00	0.006809
2	Are you selfemployed > 0.00	0.004740
3	95.00 < Which of the following best describes ...	-0.003783
4	Question Group <= 2.00	0.003234
5	What US state or territory do you live in <= 1...	0.003092
6	What US state or territory do you work in <= 2...	0.003075
7	1.00 < Do you currently have a mental health d...	-0.002906
8	20.00 < Question <= 31.00	-0.002879
9	43.00 < What country do you work in <= 50.00	-0.002273