In [1]:
```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings("ignore")
```

In [2]:
```
data=pd.read_csv(r"C:\Users\reshma_koduri\OneDrive\Documents\Covid Dataset.csv")
```

In [3]:
```
data
```

Out[3]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Yes | Yes | Yes | Yes | Yes | No | No | No | No | Yes | ... |
| **1** | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | No | No | ... |
| **2** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | ... |
| **3** | Yes | Yes | Yes | No | No | Yes | No | No | Yes | Yes | ... |
| **4** | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | ... |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **5429** | Yes | Yes | No | Yes | Yes | Yes | Yes | No | No | No | ... |
| **5430** | Yes | Yes | Yes | No | Yes | Yes | No | Yes | No | Yes | ... |
| **5431** | Yes | Yes | Yes | No | No | No | No | No | Yes | No | ... |
| **5432** | Yes | Yes | Yes | No | Yes | No | No | Yes | Yes | No | ... |
| **5433** | Yes | Yes | Yes | No | Yes | Yes | No | Yes | No | Yes | ... |

5434 rows × 21 columns

In [4]:
```
data.describe()
```

Out[4]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 5434 | 5434 | 5434 | 5434 | 5434 | 5434 | 5434 | 5434 | 5434 | 5434 |
| **unique** | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **top** | Yes | Yes | Yes | Yes | Yes | No | No | Yes | No | No |
| **freq** | 3620 | 4273 | 4307 | 3953 | 2952 | 2920 | 2869 | 2736 | 2911 | 2846 |

4 rows × 21 columns

In [5]:
```
data.shape
```

Out[5]:
```
(5434, 21)
```

In [6]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5434 entries, 0 to 5433
Data columns (total 21 columns):
 #   Column                                Non-Null Count  Dtype 
---  ------                                --------------  ----- 
 0   Breathing Problem                     5434 non-null   object
 1   Fever                                 5434 non-null   object
 2   Dry Cough                             5434 non-null   object
 3   Sore throat                           5434 non-null   object
 4   Running Nose                          5434 non-null   object
 5   Asthma                                5434 non-null   object
 6   Chronic Lung Disease                  5434 non-null   object
 7   Headache                              5434 non-null   object
 8   Heart Disease                         5434 non-null   object
 9   Diabetes                              5434 non-null   object
 10  Hyper Tension                         5434 non-null   object
 11  Fatigue                               5434 non-null   object
 12  Gastrointestinal                      5434 non-null   object
 13  Abroad travel                         5434 non-null   object
 14  Contact with COVID Patient            5434 non-null   object
 15  Attended Large Gathering              5434 non-null   object
 16  Visited Public Exposed Places         5434 non-null   object
 17  Family working in Public Exposed Places  5434 non-null   object
 18  Wearing Masks                         5434 non-null   object
 19  Sanitization from Market              5434 non-null   object
 20  COVID-19                              5434 non-null   object
dtypes: object(21)
memory usage: 891.6+ KB
```

In [7]:
```
data.head(10)
```

Out[7]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes | ... | Fa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Yes | Yes | Yes | Yes | Yes | No | No | No | No | Yes | ... | |
| **1** | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | No | No | ... | |
| **2** | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | ... | |
| **3** | Yes | Yes | Yes | No | No | Yes | No | No | Yes | Yes | ... | |
| **4** | Yes | Yes | Yes | Yes | Yes | No | Yes | Yes | Yes | Yes | ... | |
| **5** | Yes | Yes | Yes | No | No | No | No | No | Yes | No | ... | |
| **6** | Yes | Yes | Yes | No | No | No | Yes | No | Yes | Yes | ... | |
| **7** | Yes | Yes | Yes | No | Yes | Yes | No | No | No | Yes | ... | |

| | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes | ... | Fa |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **8** | Yes | Yes | Yes | No | Yes | No | Yes | No | No | Yes | ... | |
| **9** | Yes | Yes | Yes | No | No | Yes | No | No | No | Yes | ... | |

10 rows × 21 columns

In [8]:
```
data.tail(5)
```

Out[8]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **5429** | Yes | Yes | No | Yes | Yes | Yes | Yes | No | No | No | ... |
| **5430** | Yes | Yes | Yes | No | Yes | Yes | No | Yes | No | Yes | ... |
| **5431** | Yes | Yes | Yes | No | No | No | No | No | Yes | No | ... |
| **5432** | Yes | Yes | Yes | No | Yes | No | No | Yes | Yes | No | ... |
| **5433** | Yes | Yes | Yes | No | Yes | Yes | No | Yes | No | Yes | ... |

5 rows × 21 columns

In [9]:
```
data.isna().sum()
```

Out[9]:
```
Breathing Problem                       0
Fever                                   0
Dry Cough                               0
Sore throat                             0
Running Nose                            0
Asthma                                  0
Chronic Lung Disease                    0
Headache                                0
Heart Disease                           0
Diabetes                                0
Hyper Tension                           0
Fatigue                                 0
Gastrointestinal                        0
Abroad travel                           0
Contact with COVID Patient              0
Attended Large Gathering                0
Visited Public Exposed Places           0
Family working in Public Exposed Places 0
Wearing Masks                           0
Sanitization from Market                0
COVID-19                                0
dtype: int64
```

In [10]:
```
data.groupby(['COVID-19']).count()
```

Out[10]:

| COVID-19 | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes |
|---|---|---|---|---|---|---|---|---|---|---|
| No | 1051 | 1051 | 1051 | 1051 | 1051 | 1051 | 1051 | 1051 | 1051 | 1051 |
| Yes | 4383 | 4383 | 4383 | 4383 | 4383 | 4383 | 4383 | 4383 | 4383 | 4383 |

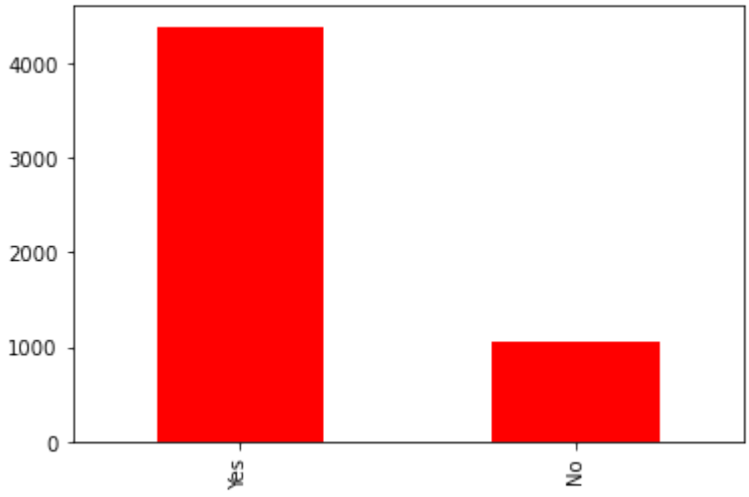
In [11]:
```
data.groupby(['Contact with COVID Patient']).count()
```

Out[11]:

| Contact with COVID Patient | Breathing Problem | Fever | Dry Cough | Sore throat | Running Nose | Asthma | Chronic Lung Disease | Headache | Heart Disease | Diabetes |
|---|---|---|---|---|---|---|---|---|---|---|
| No | 2708 | 2708 | 2708 | 2708 | 2708 | 2708 | 2708 | 2708 | 2708 | 2708 |
| Yes | 2726 | 2726 | 2726 | 2726 | 2726 | 2726 | 2726 | 2726 | 2726 | 2726 |

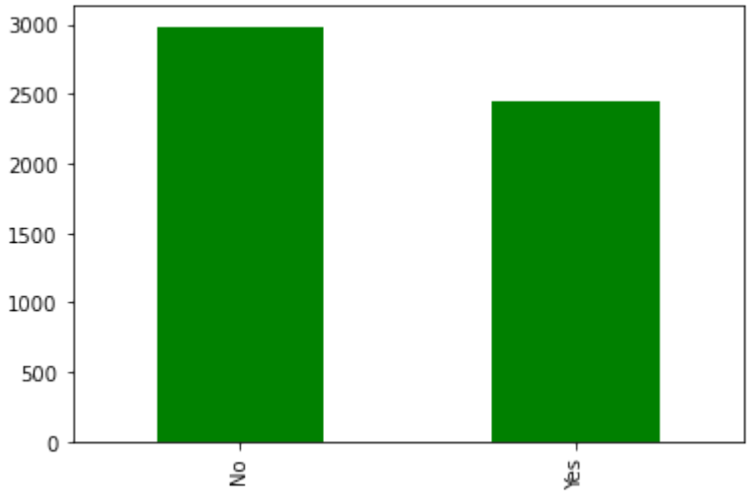
In [12]:
```
import seaborn as sns
import matplotlib.pyplot as plt
data['COVID-19'].value_counts().head(10).plot(kind='bar',color='r')
```

Out[12]:
```
<AxesSubplot:>
```



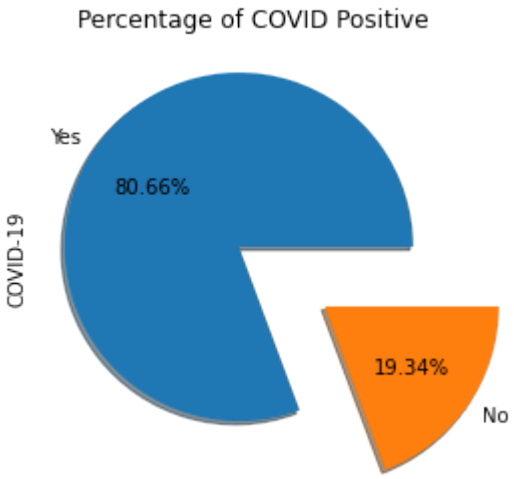
In [13]:
```
data['Abroad travel'].value_counts().head(10).plot(kind='bar',color='g')
```

Out[13]:
```
<AxesSubplot:>
```




In [14]:
```
data["COVID-19"].value_counts().plot.pie(explode=[0.1,0.5],autopct='%1.2f%%',shadow=
plt.title('Percentage of COVID Positive')
```

Out[14]:
```
Text(0.5, 1.0, 'Percentage of COVID Positive')
```




In [15]:
```
data1=data.drop(['Running Nose','Asthma','Chronic Lung Disease','Headache','Heart Di
data1
#data1=data.drop(['Asthma'],axis=1)
#data1=data.drop(['Chronic Lung Disease'],axis=1)
#data1=data.drop(['Headache'],axis=1)
#data1=data.drop(['Heart Disease'],axis=1)
#data1=data.drop(['Diabetes'],axis=1)
#data1=data.drop(['Fatigue '],axis=1)
#data1=data.drop(['Gastrointestinal '],axis=1)
#data1=data.drop(['Wearing Masks'],axis=1)
#data1=data.drop(['Sanitization from Market'],axis=1)
```

Out[15]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visited Public Exposed Places | Family working in Public Exposed Places | CC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | Yes | |
| **1** | Yes | Yes | Yes | Yes | No | No | No | Yes | Yes | No | |
| **2** | Yes | Yes | Yes | Yes | No | Yes | No | No | No | No | |
| **3** | Yes | Yes | Yes | No | No | Yes | No | Yes | Yes | No | |
| **4** | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | No | |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **5429** | Yes | Yes | No | Yes | No | No | No | No | No | No | |
| **5430** | Yes | Yes | Yes | No | Yes | No | No | No | No | No | |
| **5431** | Yes | Yes | Yes | No | Yes | No | No | No | No | No | |
| **5432** | Yes | Yes | Yes | No | No | No | No | No | No | No | |
| **5433** | Yes | Yes | Yes | No | Yes | No | No | No | No | No | |

5434 rows × 11 columns

In [16]:
```
data1.head(10)
```

Out[16]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visited Public Exposed Places | Family working in Public Exposed Places | COVID 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | Yes | Y |
| **1** | Yes | Yes | Yes | Yes | No | No | No | Yes | Yes | No | Y |
| **2** | Yes | Yes | Yes | Yes | No | Yes | No | No | No | No | Y |
| **3** | Yes | Yes | Yes | No | No | Yes | No | Yes | Yes | No | Y |
| **4** | Yes | Yes | Yes | Yes | Yes | No | Yes | No | Yes | No | Y |
| **5** | Yes | Yes | Yes | No | Yes | No | No | No | No | No | Y |
| **6** | Yes | Yes | Yes | No | Yes | No | No | Yes | Yes | Yes | Y |
| **7** | Yes | Yes | Yes | No | Yes | Yes | No | No | Yes | No | Y |
| **8** | Yes | Yes | Yes | No | No | Yes | Yes | Yes | No | No | Y |
| **9** | Yes | Yes | Yes | No | Yes | No | No | No | Yes | No | Y |

In [17]:
```
data1['Breathing Problem']=data1['Breathing Problem'].map({'Yes':1,'No':0})
data1['Fever']=data1['Fever'].map({'Yes':1,'No':0})
data1['Dry Cough']=data1['Dry Cough'].map({'Yes':1,'No':0})
```

```
data1['Sore throat']=data1['Sore throat'].map({'Yes':1,'No':0})
data1['Hyper Tension']=data1['Hyper Tension'].map({'Yes':1,'No':0})
data1['Abroad travel']=data1['Abroad travel'].map({'Yes':1,'No':0})
data1['Contact with COVID Patient']=data1['Contact with COVID Patient'].map({'Yes':1
data1['Attended Large Gathering']=data1['Attended Large Gathering'].map({'Yes':1,'No
data1['Visited Public Exposed Places']=data1['Visited Public Exposed Places'].map({'
data1['Family working in Public Exposed Places']=data1['Family working in Public Exp
data1['COVID-19']=data1['COVID-19'].map({'Yes':1,'No':0})
```

In [18]:

```
data1.head()
```

Out[18]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visited Public Exposed Places | Family working in Public Exposed Places | COVID 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | |
| **1** | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | |
| **2** | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | |
| **3** | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | |
| **4** | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | |

In [19]:

```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5434 entries, 0 to 5433
Data columns (total 11 columns):
 #   Column                                Non-Null Count  Dtype
---  ------                                --------------  -----
 0   Breathing Problem                     5434 non-null   int64
 1   Fever                                 5434 non-null   int64
 2   Dry Cough                             5434 non-null   int64
 3   Sore throat                           5434 non-null   int64
 4   Hyper Tension                         5434 non-null   int64
 5   Abroad travel                         5434 non-null   int64
 6   Contact with COVID Patient            5434 non-null   int64
 7   Attended Large Gathering              5434 non-null   int64
 8   Visited Public Exposed Places         5434 non-null   int64
 9   Family working in Public Exposed Places  5434 non-null   int64
 10  COVID-19                              5434 non-null   int64
dtypes: int64(11)
memory usage: 467.1 KB
```

In [20]:

```
cor_mat=data1.corr()
cor_mat
```

Out[20]:

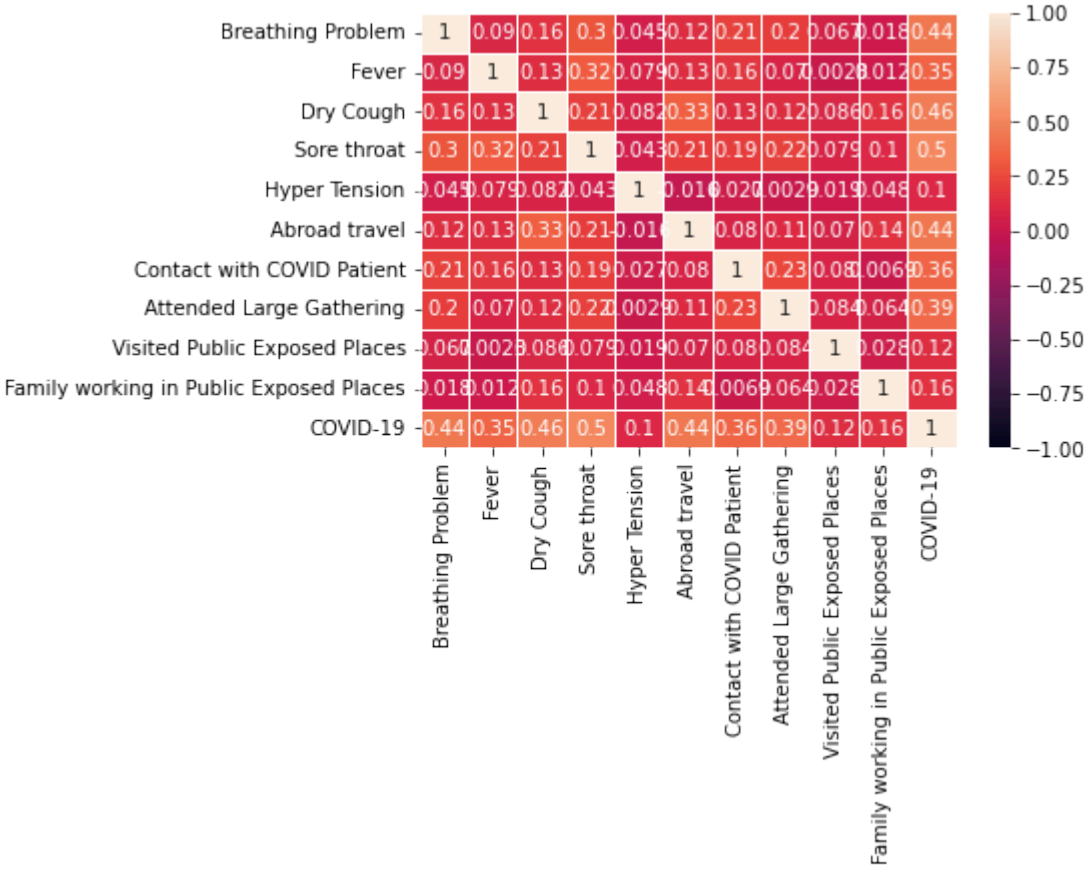
| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visit Pub Expos Plac |
|---|---|---|---|---|---|---|---|---|---|
| **Breathing Problem** | 1.000000 | 0.089903 | 0.159562 | 0.303768 | 0.045256 | 0.117795 | 0.214634 | 0.200304 | 0.0666 |
| **Fever** | 0.089903 | 1.000000 | 0.127580 | 0.322235 | 0.079001 | 0.128726 | 0.164704 | 0.070490 | 0.0022 |
| **Dry Cough** | 0.159562 | 0.127580 | 1.000000 | 0.213907 | 0.081989 | 0.331418 | 0.128330 | 0.117963 | 0.0861 |
| **Sore throat** | 0.303768 | 0.322235 | 0.213907 | 1.000000 | 0.042811 | 0.205986 | 0.189251 | 0.216438 | 0.0790 |
| **Hyper Tension** | 0.045256 | 0.079001 | 0.081989 | 0.042811 | 1.000000 | -0.016382 | 0.027307 | 0.002911 | 0.0191 |
| **Abroad travel** | 0.117795 | 0.128726 | 0.331418 | 0.205986 | -0.016382 | 1.000000 | 0.080210 | 0.113399 | 0.0696 |
| **Contact with COVID Patient** | 0.214634 | 0.164704 | 0.128330 | 0.189251 | 0.027307 | 0.080210 | 1.000000 | 0.234649 | 0.0798 |
| **Attended Large Gathering** | 0.200304 | 0.070490 | 0.117963 | 0.216438 | 0.002911 | 0.113399 | 0.234649 | 1.000000 | 0.0837 |
| **Visited Public Exposed Places** | 0.066688 | 0.002252 | 0.086176 | 0.079055 | 0.019174 | 0.069609 | 0.079800 | 0.083795 | 1.0000 |
| **Family working in Public Exposed Places** | 0.018295 | 0.012102 | 0.163102 | 0.104378 | 0.048152 | 0.143094 | 0.006909 | 0.063776 | 0.0284 |
| **COVID-19** | 0.443764 | 0.352891 | 0.464292 | 0.502848 | 0.102575 | 0.443875 | 0.357122 | 0.390145 | 0.1197 |

In [21]:

```
import seaborn as sns
sns.heatmap(cor_mat,vmax=1,vmin=-1,annot=True,linewidths=.5,cmap='rocket')
```

Out[21]:

```
<AxesSubplot:>
```




In [22]:
```
y=data1['COVID-19']
x=data1.drop(['COVID-19'],axis=1)
```

In [23]:
```
y
```

Out[23]:
```
0       1
1       1
2       1
3       1
4       1
       ..
5429    1
5430    1
5431    0
5432    0
5433    0
Name: COVID-19, Length: 5434, dtype: int64
```

In [24]:
```
x
```

Out[24]:

| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visited Public Exposed Places | Family working in Public Exposed Places |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 |
| **1** | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| **2** | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |

| | Breathing Problem | Fever | Dry Cough | Sore throat | Hyper Tension | Abroad travel | Contact with COVID Patient | Attended Large Gathering | Visited Public Exposed Places | Family working in Public Exposed Places |
|---|---|---|---|---|---|---|---|---|---|---|
| **3** | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| **4** | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **5429** | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **5430** | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **5431** | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **5432** | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **5433** | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

5434 rows × 10 columns

In [25]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.30,random_state=42)
```

In [26]:
```
print(x_train.shape, x_test.shape, y_train.shape, y_test.shape)
```

```
(3803, 10) (1631, 10) (3803,) (1631,)
```

# Logistic Regression

In [27]:
```
from sklearn.linear_model import LogisticRegression
classifier=LogisticRegression()
classifier.fit(x_train, y_train)
```

Out[27]:
```
▾ LogisticRegression

LogisticRegression()
```

In [28]:
```
y_pred=classifier.predict(x_test)
y_pred
```

Out[28]:
```
array([1, 0, 1, ..., 1, 1, 1], dtype=int64)
```

In [29]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

Out[29]:
```
array([[ 291,   25],
       [  23, 1292]], dtype=int64)
```

In [30]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[30]:
```
0.970570202329859
```

# Random Forest Classification

In [31]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.ensemble import RandomForestClassifier
cls=RandomForestClassifier()
n_estimators=[25,50,75,100,125,150,175,200]
criterion=['gini','entropy']
max_depth=[3,5,10]
parameters={'n_estimators': n_estimators,'criterion':criterion,'max_depth':max_depth
RFC_cls = GridSearchCV(cls, parameters)
RFC_cls.fit(x_train,y_train)
```

Out[31]:

GridSearchCV

estimator: RandomForestClassifier

RandomForestClassifier

In [32]:
```
RFC_cls.best_params_
```

Out[32]:
```
{'criterion': 'gini', 'max_depth': 10, 'n_estimators': 50}
```

In [33]:
```
cls=RandomForestClassifier(n_estimators=100,criterion='entropy',max_depth=10)
cls.fit(x_train,y_train)
```

Out[33]:

RandomForestClassifier

RandomForestClassifier(criterion='entropy', max_depth=10)

In [34]:
```
y_pred=cls.predict(x_test)
y_pred
```

Out[34]:
```
array([1, 0, 1, ..., 1, 1, 1], dtype=int64)
```

In [35]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[35]:
```
0.9754751686082158
```

In [36]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
```

Out[36]:
```
array([[ 292,   24],
       [  16, 1299]], dtype=int64)
```

# DecisionTree Classification

In [37]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [38]:
```
cls=DecisionTreeClassifier()
```

In [39]:
```
cls.fit(x_train,y_train)
```

Out[39]:
```
▾ DecisionTreeClassifier

DecisionTreeClassifier()
```

In [40]:
```
y_pred=cls.predict(x_test)
y_pred
```

Out[40]:
```
array([1, 0, 1, ..., 1, 1, 1], dtype=int64)
```

In [41]:
```
from sklearn.metrics import accuracy_score
accuracy_score(y_test,y_pred)
```

Out[41]:
```
0.9754751686082158
```

# TESTING

In [42]:
```
new = x_test.iloc[0]
a = np.asarray(new)
a = a.reshape(1,-1)
p = cls.predict(a)
```

In [43]:
```
if (p[0] == 1):
    print("Person is affected by Covid 19 and is at risk of dying")
else:
    print("Great! the results are negative and you don't have to worry")
```

```
Person is affected by Covid 19 and is at risk of dying
```

In [45]:
```
from prettytable import PrettyTable
table=PrettyTable()
table.field_names = ["TEST_SIZE","0.30"]
table.add_row(["Logistic",0.970570202329859])
table.add_row(["Random Forest",0.9754751686082158])
table.add_row(["DecisionTree",0.9754751686082158])
print(table)
```

```
+---------------+--------------------+
|   TEST_SIZE   |        0.30        |
+---------------+--------------------+
|    Logistic   | 0.970570202329859  |
| Random Forest | 0.9754751686082158 |
|  DecisionTree | 0.9754751686082158 |
+---------------+--------------------+
```