Untitled28

January 18, 2024

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('/content/covid.csv')
df
```

[ ]:
```
     Breathing Problem Fever Dry Cough Sore throat Running Nose Asthma  \
0                  Yes   Yes       Yes         Yes          Yes     No
1                  Yes   Yes       Yes         Yes           No    Yes
2                  Yes   Yes       Yes         Yes          Yes    Yes
3                  Yes   Yes       Yes          No           No    Yes
4                  Yes   Yes       Yes         Yes          Yes     No
…                …   …       …         …          …    …
5429               Yes   Yes        No         Yes          Yes    Yes
5430               Yes   Yes       Yes          No          Yes    Yes
5431               Yes   Yes       Yes          No           No     No
5432               Yes   Yes       Yes          No          Yes     No
5433               Yes   Yes       Yes          No          Yes    Yes

     Chronic Lung Disease Headache Heart Disease Diabetes  … Fatigue   \
0                      No       No            No      Yes  …      Yes
1                     Yes      Yes            No       No  …      Yes
2                     Yes      Yes            No      Yes  …      Yes
3                      No       No           Yes      Yes  …       No
4                     Yes      Yes           Yes      Yes  …       No
…                   …      …           …      …  …      …
5429                  Yes       No            No       No  …      Yes
5430                   No      Yes            No      Yes  …      Yes
5431                   No       No           Yes       No  …       No
5432                   No      Yes           Yes       No  …       No
5433                   No      Yes            No      Yes  …      Yes

     Gastrointestinal  Abroad travel Contact with COVID Patient  \
0                  Yes            No                        Yes
1                   No            No                         No
2                  Yes           Yes                         No
```

```
3                    No            Yes                          No
4                   Yes             No                         Yes
…                …           …                         …
5429                Yes             No                          No
5430                 No             No                          No
5431                 No             No                          No
5432                 No             No                          No
5433                 No             No                          No

     Attended Large Gathering Visited Public Exposed Places  \
0                          No                           Yes
1                         Yes                           Yes
2                          No                            No
3                         Yes                           Yes
4                          No                           Yes
…                       …                           …
5429                       No                            No
5430                       No                            No
5431                       No                            No
5432                       No                            No
5433                       No                            No

     Family working in Public Exposed Places Wearing Masks  \
0                                         Yes            No
1                                          No            No
2                                          No            No
3                                          No            No
4                                          No            No
…                                      …           …
5429                                       No            No
5430                                       No            No
5431                                       No            No
5432                                       No            No
5433                                       No            No

     Sanitization from Market COVID-19
0                          No      Yes
1                          No      Yes
2                          No      Yes
3                          No      Yes
4                          No      Yes
…                       …      …
5429                       No      Yes
5430                       No      Yes
5431                       No       No
5432                       No       No
5433                       No       No
```

```
[5434 rows x 21 columns]
```

0.1 1. Info and feature

```
[ ]: df.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
```

```
[ ]: df.describe().T
```

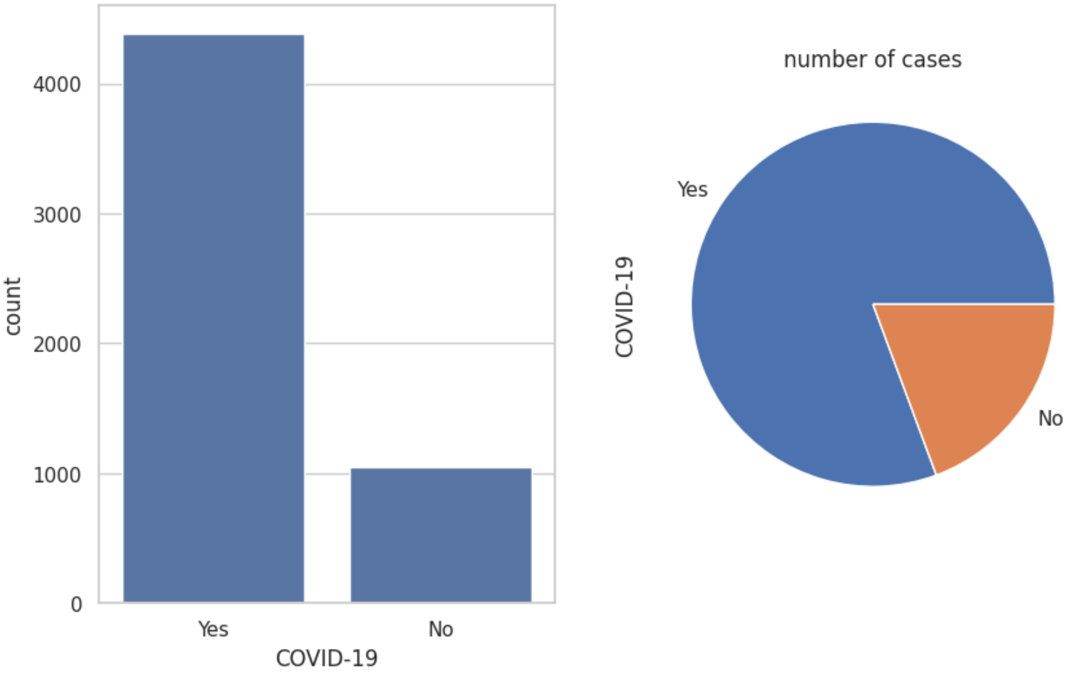
```
[ ]:
                                        count unique  top  freq
 Breathing Problem                       5434      2  Yes  3620
 Fever                                   5434      2  Yes  4273
 Dry Cough                               5434      2  Yes  4307
 Sore throat                             5434      2  Yes  3953
 Running Nose                            5434      2  Yes  2952
 Asthma                                  5434      2   No  2920
 Chronic Lung Disease                    5434      2   No  2869
 Headache                                5434      2  Yes  2736
 Heart Disease                           5434      2   No  2911
```

```
Diabetes                                  5434      2   No  2846
Hyper Tension                             5434      2   No  2771
Fatigue                                   5434      2  Yes  2821
Gastrointestinal                          5434      2   No  2883
Abroad travel                             5434      2   No  2983
Contact with COVID Patient                5434      2  Yes  2726
Attended Large Gathering                  5434      2   No  2924
Visited Public Exposed Places             5434      2  Yes  2820
Family working in Public Exposed Places   5434      2   No  3172
Wearing Masks                             5434      1   No  5434
Sanitization from Market                  5434      1   No  5434
COVID-19                                  5434      2  Yes  4383
```

0.2 2. Count

[ ]:

```
fig = plt.gcf()
# ax = f.add_subplot(111)
# ax.yaxis.tick_right()
fig.set_size_inches(10, 6)
plt.subplot(1,2,1)
'bar plot '
sns.countplot(x='COVID-19',data=df)
plt.subplot(1,2,2)
'pie plot'
covid = df["COVID-19"]

covid.value_counts().plot.pie()
plt.title('number of cases');
```



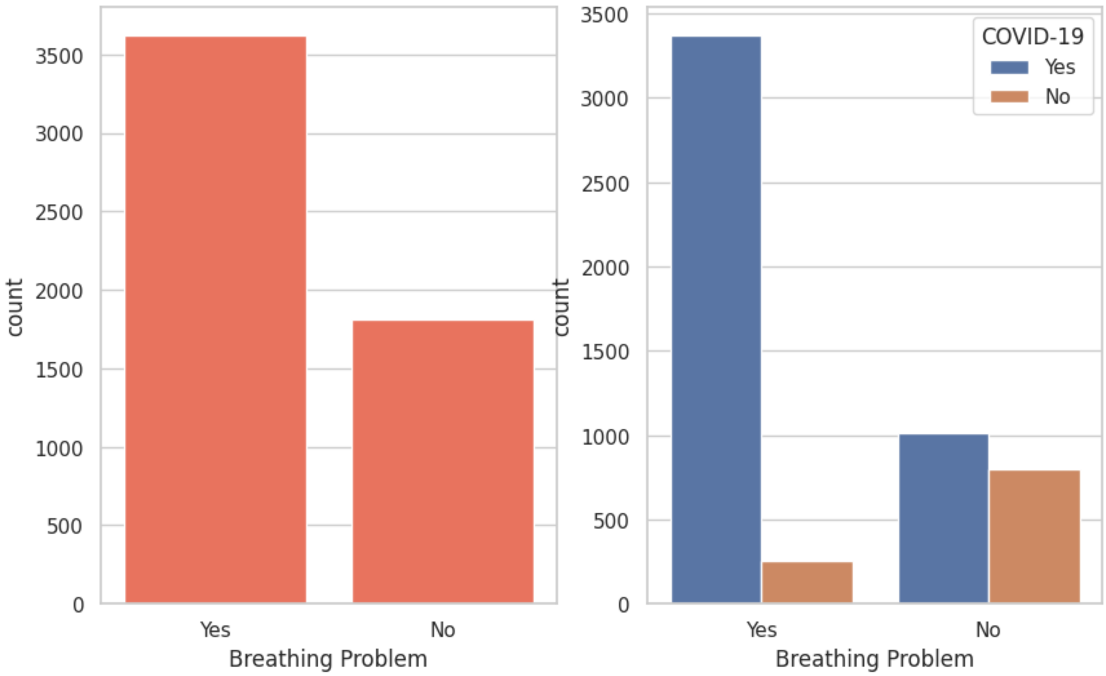
0.3 3. Count patient with breating problem

[ ]:

```
fig = plt.gcf()
# ax = f.add_subplot(111)
# ax.yaxis.tick_right()
fig.set_size_inches(10, 6)
plt.subplot(1,2,1)
'having and not having breathing problems'
sns.countplot(x='Breathing Problem',data=df, color = 'tomato')
plt.subplot(1,2,2)
sns.countplot(x='Breathing Problem',hue='COVID-19',data=df)
```

[ ]:

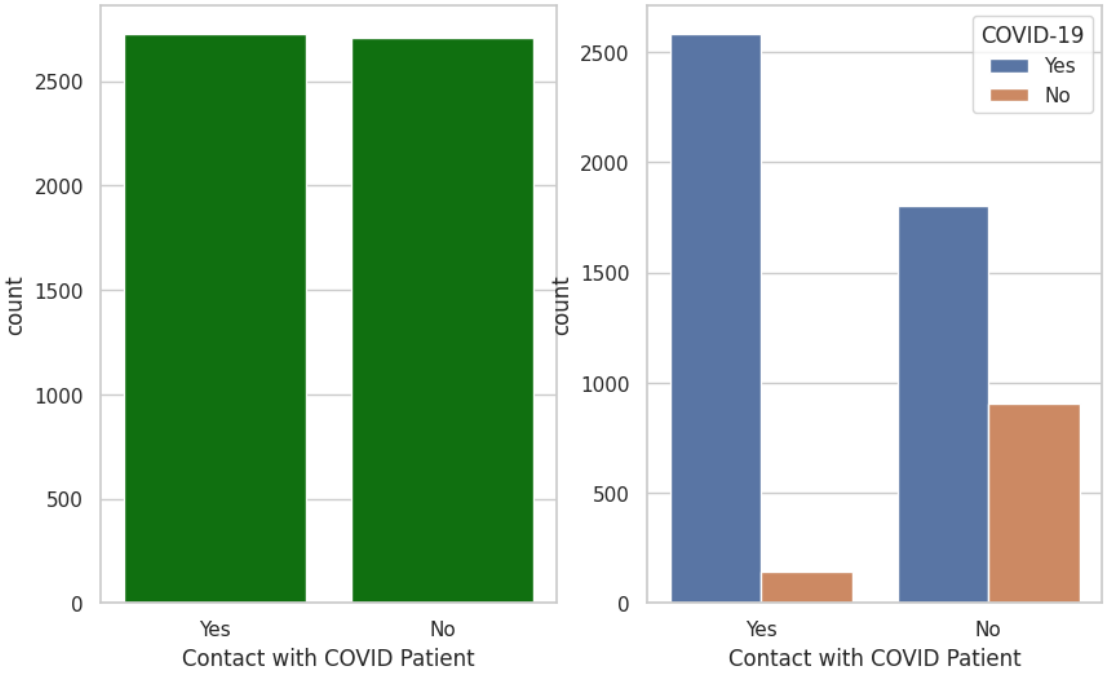
```
<Axes: xlabel='Breathing Problem', ylabel='count'>
```



0.4 4. Contact with COVID Patient

[ ]:
```
fig = plt.gcf()
# ax = f.add_subplot(111)
# ax.yaxis.tick_right()
fig.set_size_inches(10, 6)
plt.subplot(1,2,1)
sns.countplot(x='Contact with COVID Patient',data=df, color = 'green')
plt.subplot(1,2,2)

sns.countplot(x='Contact with COVID Patient',hue='COVID-19',data=df)
```

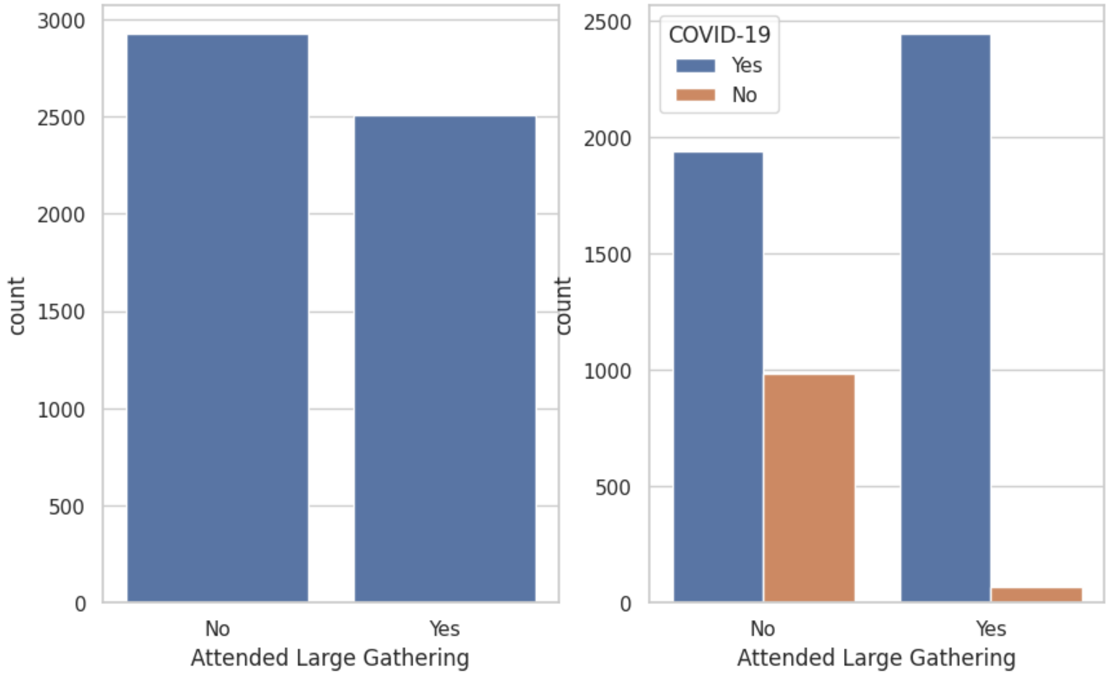
[ ]:
```
<Axes: xlabel='Contact with COVID Patient', ylabel='count'>
```



0.5 ’Attended Large Gathering

```
[ ]: fig = plt.gcf()
     # ax = f.add_subplot(111)
     # ax.yaxis.tick_right()
     fig.set_size_inches(10, 6)
     plt.subplot(1,2,1)
     sns.countplot(x='Attended Large Gathering',data=df)
     plt.subplot(1,2,2)
     sns.countplot(x='Attended Large Gathering',hue='COVID-19',data=df)
```

```
[ ]: <Axes: xlabel='Attended Large Gathering', ylabel='count'>
```



0.5.1 Processing data

[ ]:
```python
from sklearn.preprocessing import LabelEncoder
e=LabelEncoder()
df['Breathing Problem']=e.fit_transform(df['Breathing Problem'])
df['Fever']=e.fit_transform(df['Fever'])
df['Dry Cough']=e.fit_transform(df['Dry Cough'])
df['Sore throat']=e.fit_transform(df['Sore throat'])
df['Running Nose']=e.fit_transform(df['Running Nose'])
df['Asthma']=e.fit_transform(df['Asthma'])
df['Chronic Lung Disease']=e.fit_transform(df['Chronic Lung Disease'])
df['Headache']=e.fit_transform(df['Headache'])
df['Heart Disease']=e.fit_transform(df['Heart Disease'])
df['Diabetes']=e.fit_transform(df['Diabetes'])
df['Hyper Tension']=e.fit_transform(df['Hyper Tension'])
df['Abroad travel']=e.fit_transform(df['Abroad travel'])
df['Contact with COVID Patient']=e.fit_transform(df['Contact with COVID Patient'])
df['Attended Large Gathering']=e.fit_transform(df['Attended Large Gathering'])
df['Visited Public Exposed Places']=e.fit_transform(df['Visited Public Exposed Places'])
df['Family working in Public Exposed Places']=e.fit_transform(df['Family working in Public Exposed Places'])
df['Wearing Masks']=e.fit_transform(df['Wearing Masks'])
```

```
df['Sanitization from Market']=e.fit_transform(df['Sanitization from Market'])
df['COVID-19']=e.fit_transform(df['COVID-19'])
df['Dry Cough']=e.fit_transform(df['Dry Cough'])
df['Sore throat']=e.fit_transform(df['Sore throat'])
df['Gastrointestinal ']=e.fit_transform(df['Gastrointestinal '])
df['Fatigue ']=e.fit_transform(df['Fatigue '])
```

[ ]:

```
df
```

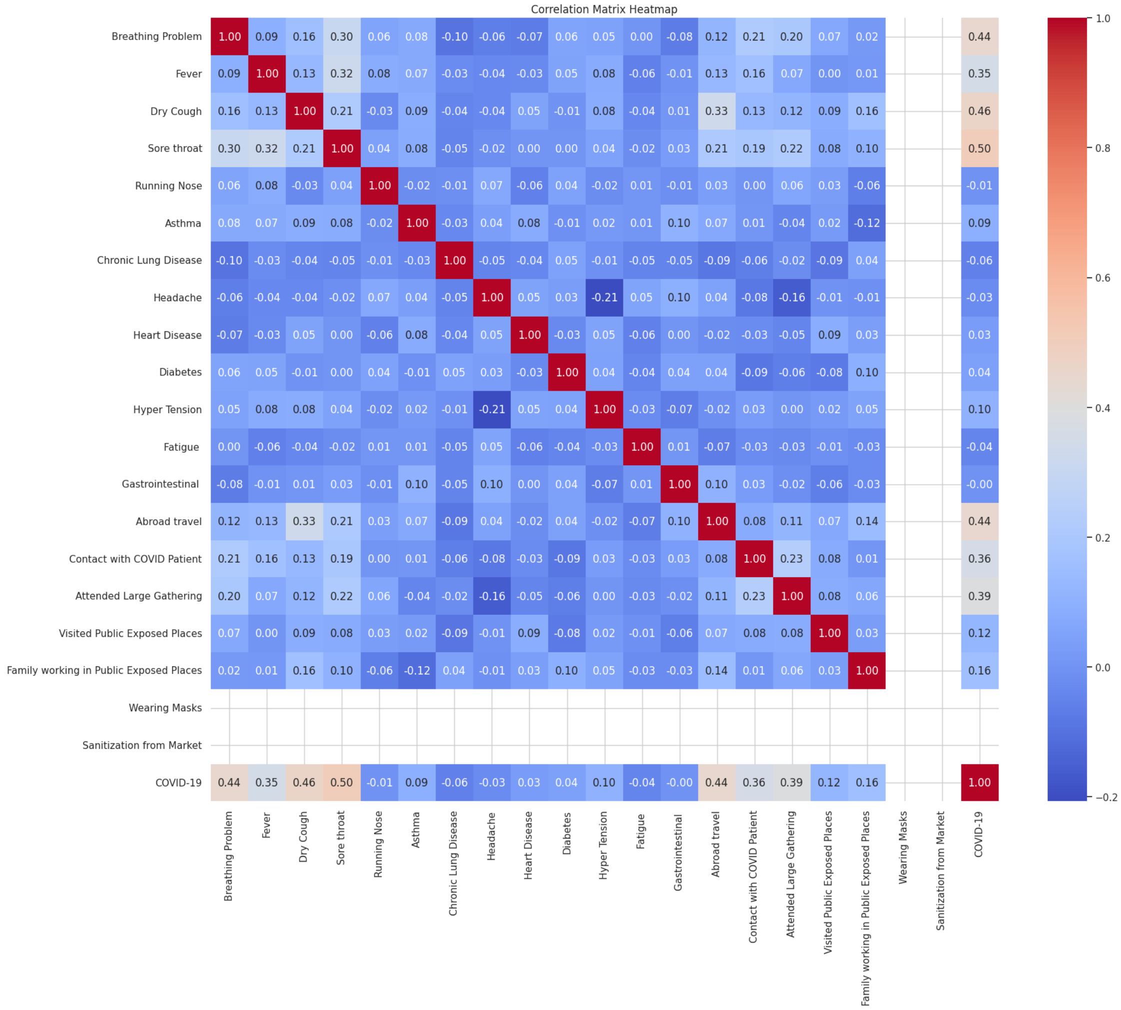
[ ]:

```
      Breathing Problem  Fever  Dry Cough  Sore throat  Running Nose  Asthma  \
0                     1      1          1            1             1       0
1                     1      1          1            1             0       1
2                     1      1          1            1             1       1
3                     1      1          1            0             0       1
4                     1      1          1            1             1       0
…                 …    …        …          …           …     …
5429                  1      1          0            1             1       1
5430                  1      1          1            0             1       1
5431                  1      1          1            0             0       0
5432                  1      1          1            0             1       0
5433                  1      1          1            0             1       1

      Chronic Lung Disease  Headache  Heart Disease  Diabetes  …  Fatigue   \
0                        0         0              0         1  …         1
1                        1         1              0         0  …         1
2                        1         1              0         1  …         1
3                        0         0              1         1  …         0
4                        1         1              1         1  …         0
…                    …       …            …       …  …       …
5429                     1         0              0         0  …         1
5430                     0         1              0         1  …         1
5431                     0         0              1         0  …         0
5432                     0         1              1         0  …         0
5433                     0         1              0         1  …         1

      Gastrointestinal   Abroad travel  Contact with COVID Patient  \
0                     1              0                           1
1                     0              0                           0
2                     1              1                           0
3                     0              1                           0
4                     1              0                           1
…                 …            …                         …
5429                  1              0                           0
5430                  0              0                           0
5431                  0              0                           0
5432                  0              0                           0
5433                  0              0                           0
```

```
      Attended Large Gathering  Visited Public Exposed Places  \
0                            0                              1
1                            1                              1
2                            0                              0
3                            1                              1
4                            0                              1
…                        …                            …
5429                         0                              0
5430                         0                              0
5431                         0                              0
5432                         0                              0
5433                         0                              0

      Family working in Public Exposed Places  Wearing Masks  \
0                                           1              0
1                                           0              0
2                                           0              0
3                                           0              0
4                                           0              0
…                                       …            …
5429                                        0              0
5430                                        0              0
5431                                        0              0
5432                                        0              0
5433                                        0              0

      Sanitization from Market  COVID-19
0                            0         1
1                            0         1
2                            0         1
3                            0         1
4                            0         1
…                        …       …
5429                         0         1
5430                         0         1
5431                         0         0
5432                         0         0
5433                         0         0

[5434 rows x 21 columns]
```

0.6 5. Correlation matrix

```
[ ]: ' Heat map'
     corr_matrix = df.corr()

     # Set up the matplotlib figure
     plt.figure(figsize=(20, 16))

     # Draw the heatmap
     sns.heatmap(corr_matrix, annot=True, fmt=".2f", cmap='coolwarm')

     # Add title
     plt.title('Correlation Matrix Heatmap')

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



0.7 6. training

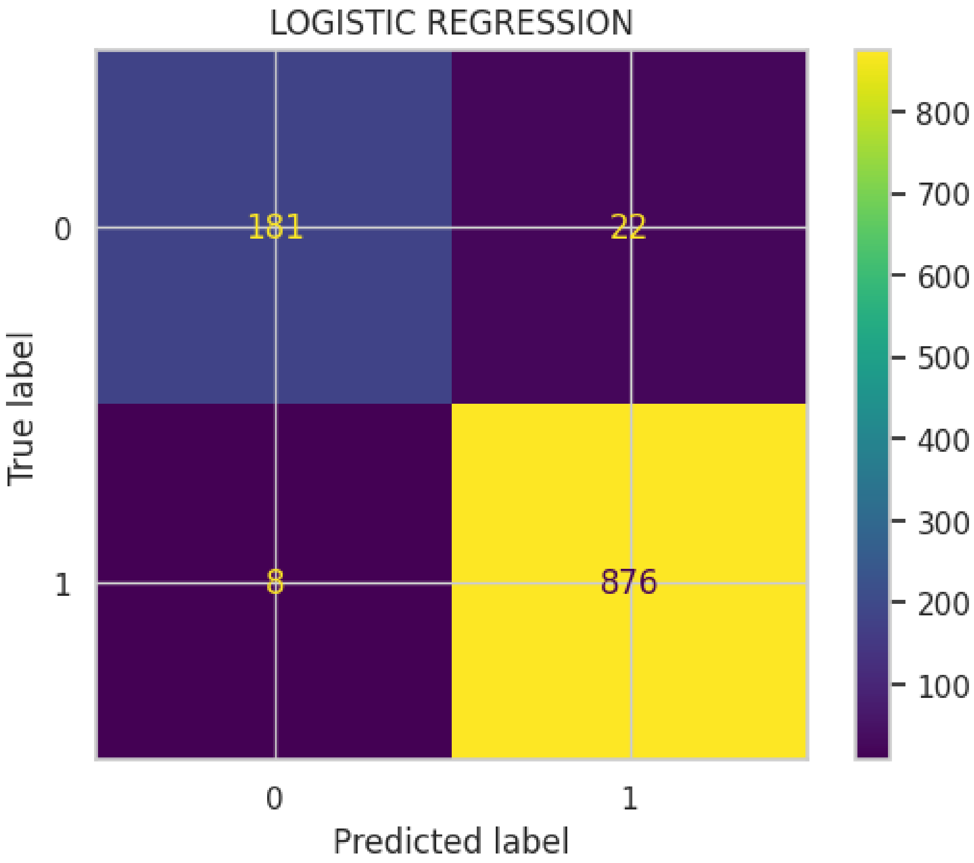
[ ]:
```
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.metrics import accuracy_score
```

[ ]:
```
X=df.drop(['COVID-19'],axis=1)
Y=df['COVID-19']

x_train, x_test, y_train, y_test = train_test_split(X, Y, test_size = 0.2, random_state=42)
```

[ ]:
```
'LOGISTIC REGRESSION'
from sklearn.linear_model import LogisticRegression
lg = LogisticRegression()
#Fit the model
lg.fit(x_train, y_train)
y_pred_lg = lg.predict(x_test)
#Score/Accuracy
acc_logreg =lg.score(x_test, y_test)*100
print ('Train : ', accuracy_score(y_train,lg.predict(x_train))*100)
print ('Test : ', acc_logreg)
```

```
Train :  96.87140556705775
Test :  97.24011039558418
```

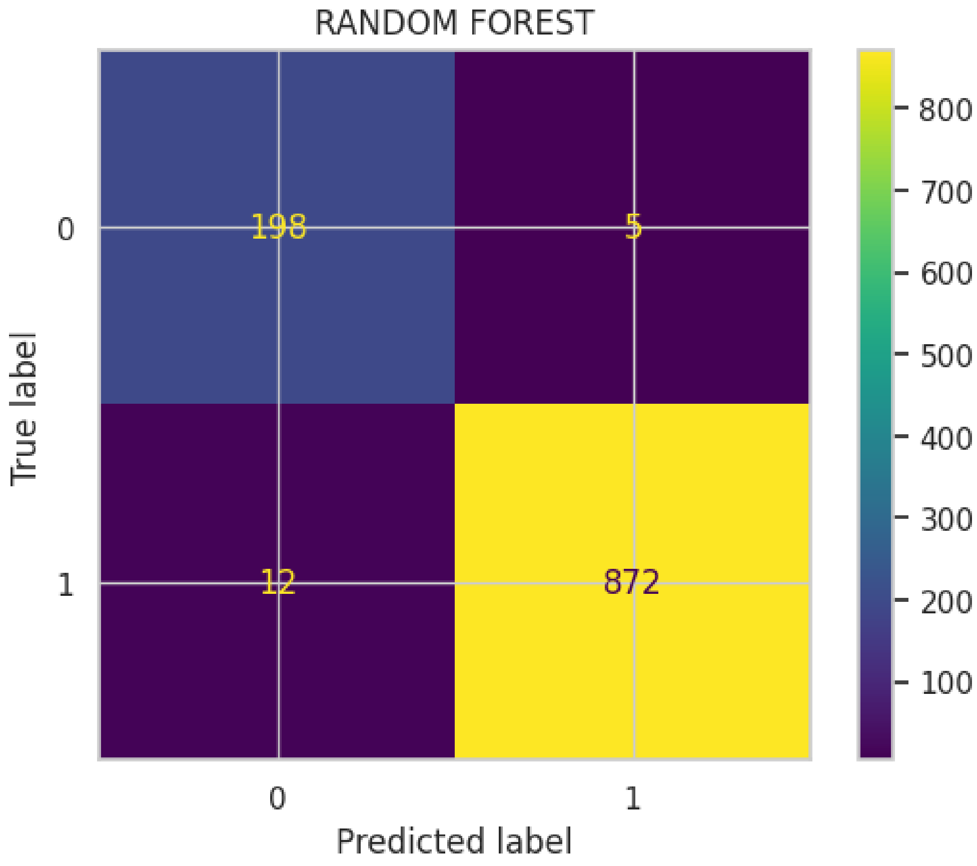
[ ]:
```
from sklearn.metrics import ConfusionMatrixDisplay,confusion_matrix, classification_report

disp = ConfusionMatrixDisplay.from_predictions(y_test, y_pred_lg)
plt.title ('LOGISTIC REGRESSION')
plt.show()
```



```
[ ]: from sklearn.ensemble import RandomForestClassifier
     rf = RandomForestClassifier(n_estimators=1000)
     #Fit
     rf.fit(x_train, y_train)
     #Score/Accuracy
     acc_randomforest=rf.score(x_test, y_test)*100
     print ('Train : ', accuracy_score(y_train,rf.predict(x_train))*100)
     print ('Test : ', acc_randomforest)
```

```
Train :  98.22866344605475
Test :  98.4360625574977
```

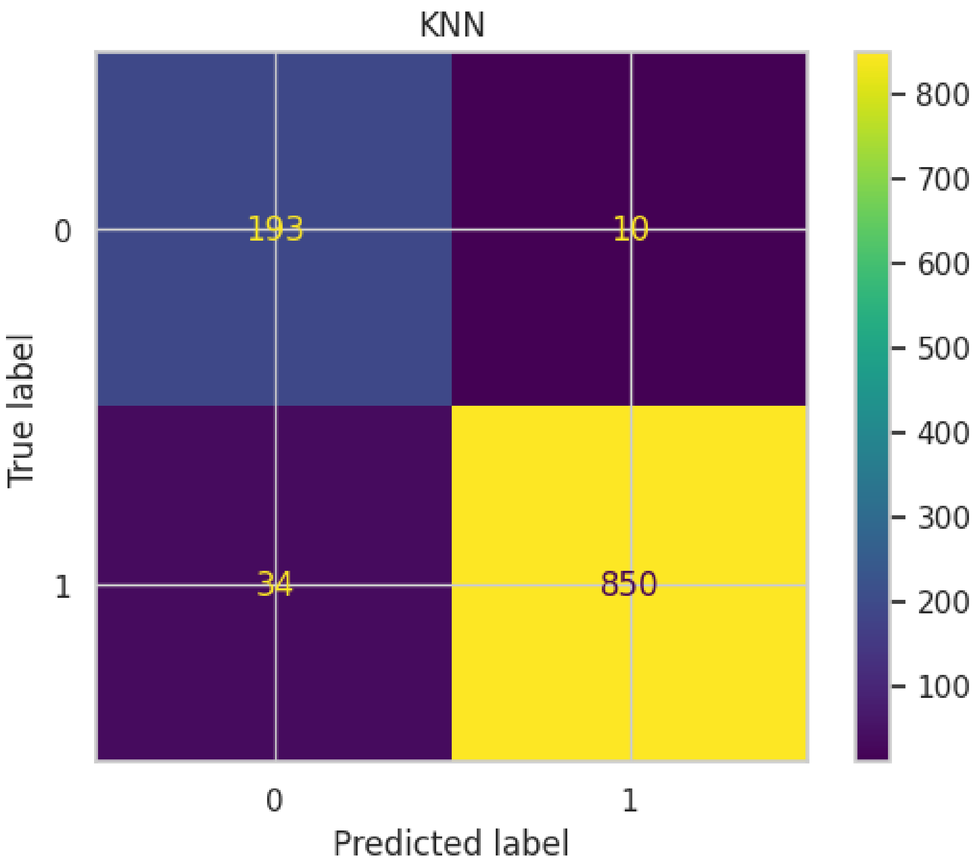
```
[ ]: from sklearn.metrics import ConfusionMatrixDisplay,confusion_matrix,␣
      ↪classification_report
     y_pred_rf = rf.predict(x_test)
     disp = ConfusionMatrixDisplay.from_predictions(y_test, y_pred_rf)
     plt.title ('RANDOM FOREST')
     plt.show()
```



```
[ ]: from sklearn.neighbors import KNeighborsClassifier
     knn = KNeighborsClassifier(n_neighbors=20)
     knn.fit(x_train, y_train)
     y_pred_knn = knn.predict(x_test)
     #Score/Accuracy
     acc_knn=knn.score(x_test, y_test)*100
     print ('Train : ', accuracy_score(y_train,knn.predict(x_train))*100)
     print ('Test : ', acc_knn)
```

```
Train :  96.84840119622729
Test :  95.95216191352345
```

```
[ ]: disp = ConfusionMatrixDisplay.from_predictions(y_test, y_pred_knn)
     plt.title ('KNN')
     plt.show()
```



0.8 7. BAr chart

```
[ ]: from sklearn.metrics import mean_squared_error ,r2_score
     compare = pd.DataFrame ({
         'model' : ['KNN', 'LogReg', 'RF'],
         'accurancy' : [acc_knn, acc_logreg, acc_randomforest],
         'r2' : [r2_score(y_test, y_pred_knn),r2_score(y_test,␣
       ↪y_pred_lg),r2_score(y_test, y_pred_rf)],
         'mse' : [mean_squared_error(y_test, y_pred_knn),mean_squared_error(y_test,␣
       ↪y_pred_lg),mean_squared_error(y_test, y_pred_rf)]
     })
     compare
```

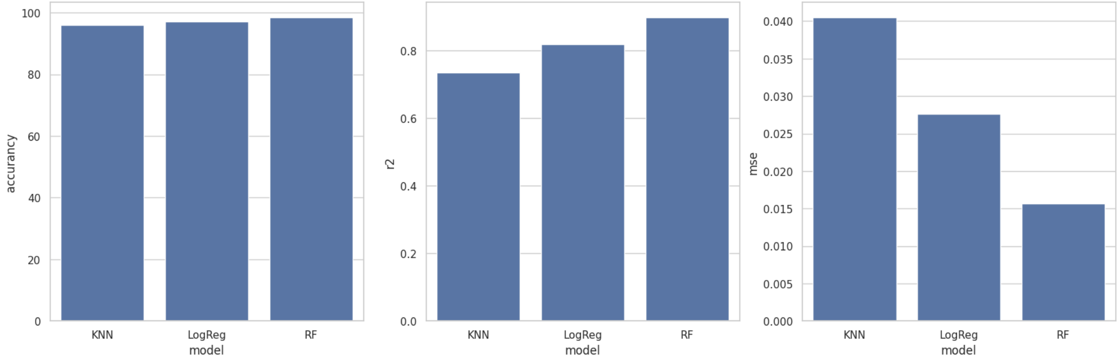
```
[ ]:     model  accurancy        r2       mse
     0     KNN  95.952162  0.733477  0.040478
     1  LogReg  97.240110  0.818280  0.027599
     2      RF  98.436063  0.897025  0.015639
```

[ ]:

```
fig = plt.gcf()
# ax = f.add_subplot(111)
# ax.yaxis.tick_right()
fig.set_size_inches(20, 6)
plt.subplot(1,3,1)
sns.barplot(x="model", y="accurancy", data=compare)
plt.subplot(1,3,2)
sns.barplot(x="model", y="r2", data=compare)
plt.subplot(1,3,3)
sns.barplot(x="model", y="mse", data=compare)
```

[ ]:

```
<Axes: xlabel='model', ylabel='mse'>
```


accurancy
KNN
LogReg
RF
model
r2
KNN
LogReg
RF
model
mse
KNN
LogReg
RF
model