COVID-19 OCEANIA

Libraries

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
from sklearn import linear_model
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/tools/
_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use
the functions in the public API at pandas.testing instead.
  import pandas.util.testing as tm
```

```
df= pd.read_csv('oceania_covid.csv')
df.head(10)
```

```
       Country/Other  Total Cases  Total Deaths  Total Recovered  \
0               Fiji        63999           834          62008.0
1   French Polynesia        67651           641              NaN
2           Kiribati         2953            11           2241.0
3      New Caledonia        55502           301          47969.0
4        New Zealand       166098            56          19263.0
5              Palau         3823             6           3313.0
6   Papua New Guinea        41335           638          39714.0
7    Solomon Islands         7258           106           1761.0
8  Wallis and Futuna          454             7            438.0

   Active Cases  Tot Cases/ 1M pop  Deaths/ 1M pop  Total Tests  \
0        1157.0              70542             919     497559.0
1           NaN             238536            2260          NaN
2         701.0              24107              90          NaN
3        7232.0             191345            1038      98964.0
4      146779.0              33206              11    6768479.0
5         504.0             209629             329      37269.0
6         983.0               4479              69     249149.0
7        5391.0              10153             148       5117.0
8           9.0              41606             641      20508.0

   Tests/ 1M pop  Population
0       548425.0      907250
1            NaN      283609
2            NaN      122496
3       341181.0      290063
4      1353127.0     5002100
5      2043593.0       18237
6        26997.0     9228750
```

```
7         7158.0      714874
8      1879399.0       10912
```

Data source

https://www.kaggle.com/anandhuh/covid-in-oceania-latest-data

Description of the data base and variables

This data base shows the behaviour of covid-19 on Oceania

- Country- It's a categorical variable that shows the names of the countries
- Total Cases- It's a numerical variable that shows the total of cases on Oceania
- Total deaths - It's a numerical variable that shows the total of deaths on Oceania
- Total recovered - It's a numerical variable that shows the total of pacients who recovered from covid-19
- Total Tests - It's a numerical variable that shows the total number of tests taken

```
df.columns
```

```
Index(['Country/Other', 'Total Cases', 'Total Deaths', 'Total
Recovered',
       'Active Cases', 'Tot Cases/ 1M pop', 'Deaths/ 1M pop', 'Total
Tests',
       'Tests/ 1M pop', 'Population'],
      dtype='object')
```

```
df.drop(['Tot Cases/ 1M pop','Deaths/ 1M pop','Tests/ 1M pop',],
        axis= 1,
        inplace= True)
df.columns
```

```
Index(['Country/Other', 'Total Cases', 'Total Deaths', 'Total
Recovered',
       'Active Cases', 'Total Tests', 'Population'],
      dtype='object')
```

```
df.rename(columns = { 'Country/Other':'Country',},
          inplace= True
         )
df.head(10)
```

```
             Country  Total Cases  Total Deaths  Total Recovered  \
0               Fiji        63999           834          62008.0
1   French Polynesia        67651           641              NaN
2           Kiribati         2953            11           2241.0
3      New Caledonia        55502           301          47969.0
4        New Zealand       166098            56          19263.0
```

```
5              Palau         3823            6          3313.0
6   Papua New Guinea        41335          638         39714.0
7    Solomon Islands         7258          106          1761.0
8  Wallis and Futuna          454            7           438.0
```

```
   Active Cases  Total Tests  Population
0        1157.0     497559.0      907250
1           NaN          NaN      283609
2         701.0          NaN      122496
3        7232.0      98964.0      290063
4      146779.0    6768479.0     5002100
5         504.0      37269.0       18237
6         983.0     249149.0     9228750
7        5391.0       5117.0      714874
8           9.0      20508.0       10912
```

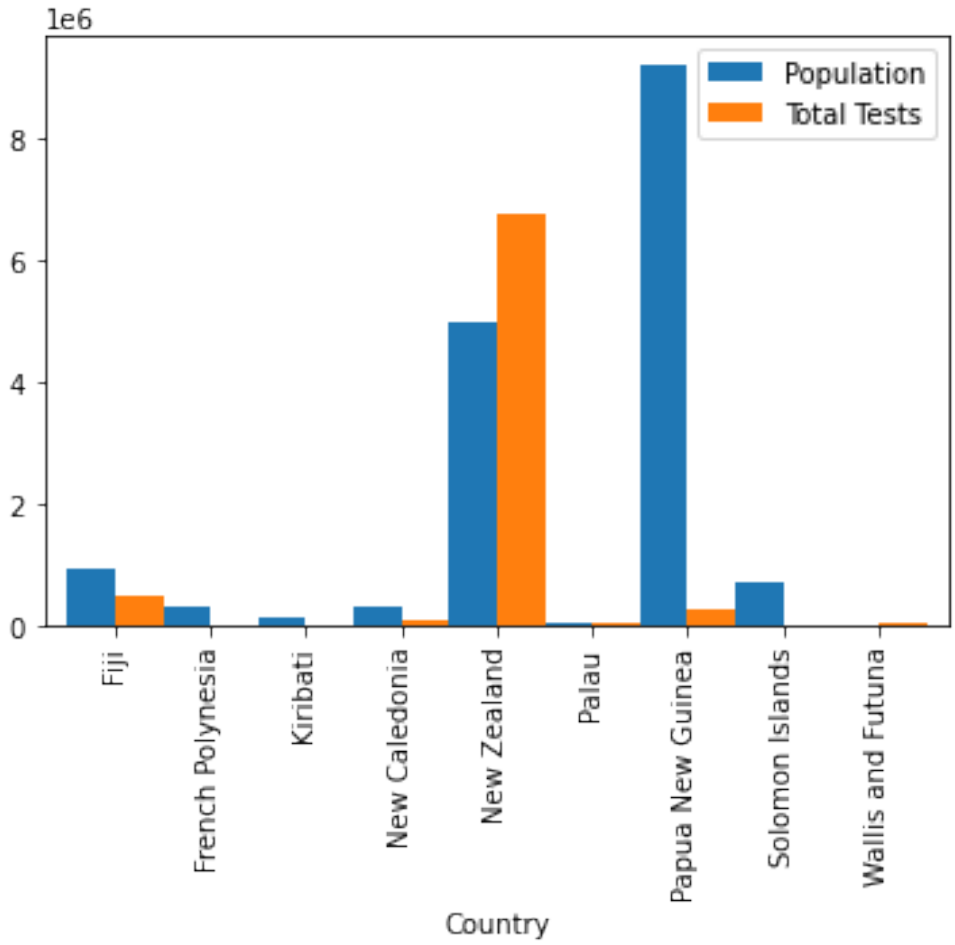
Questions

1. Wich country has more population and more tests taken?

```
myPivot = pd.pivot_table(df,
                         index = ['Country'],
                         values= ['Population','Total Tests'],
                        )
myPivot
```

```
                  Population  Total Tests
Country
Fiji                  907250     497559.0
French Polynesia      283609          NaN
Kiribati              122496          NaN
New Caledonia         290063      98964.0
New Zealand          5002100    6768479.0
Palau                  18237      37269.0
Papua New Guinea      9228750     249149.0
Solomon Islands        714874       5117.0
Wallis and Futuna       10912      20508.0
```

```
my_plot = df.plot(x="Country", y=["Population","Total Tests"],
kind="bar",width=1)
plt.show()
```



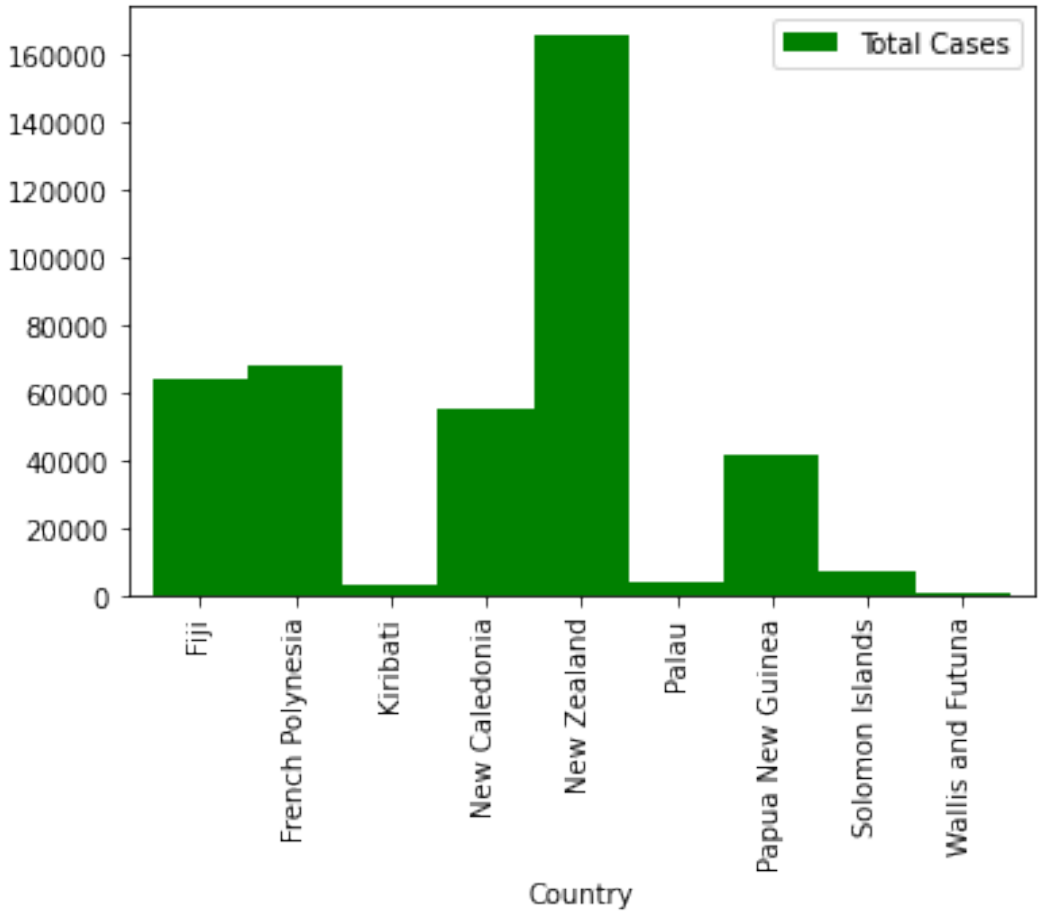
As we can see the country with more population is Papua New Guinea and the country with more Tests taken is New Zealand.

2. What is the country with the highest number of positive cases on the continent of Oceania?

```
myPivot = pd.pivot_table(df,
                         index = ['Country'],
                         values= ['Total Cases'],
                        )
myPivot
```

Country	Total Cases
Fiji	63999
French Polynesia	67651
Kiribati	2953
New Caledonia	55502
New Zealand	166098
Palau	3823
Papua New Guinea	41335
Solomon Islands	7258
Wallis and Futuna	454

```
my_plot = df.plot("Country", "Total Cases",
kind="bar",width=1,color='green')
plt.show()
```



In conclusion we can say that New Zealand far exceeds other countries in terms of the number of covid-19 cases

3. Does the quantity of tests taken affects on the total of cases?

```
df_new=df.dropna()
df_new
```

```
             Country  Total Cases  Total Deaths  Total Recovered  \
0               Fiji        63999           834          62008.0
3      New Caledonia        55502           301          47969.0
4        New Zealand       166098            56          19263.0
5              Palau         3823             6           3313.0
6   Papua New Guinea        41335           638          39714.0
7    Solomon Islands         7258           106           1761.0
8  Wallis and Futuna          454             7            438.0

   Active Cases  Total Tests  Population
0        1157.0     497559.0      907250
```

```
3        7232.0      98964.0      290063
4      146779.0    6768479.0     5002100
5         504.0      37269.0       18237
6         983.0     249149.0     9228750
7        5391.0       5117.0      714874
8           9.0      20508.0       10912
```

```
y=pd.DataFrame(df_new['Total Cases'])
x=pd.DataFrame(df_new['Total Tests'])
lm= linear_model.LinearRegression()
model_lm= lm.fit(x,y)
model_lm
```

```
LinearRegression()
```

```
print('The y intercep, b0 is= ', model_lm.intercept_)
print('The coef., b1 is= ', model_lm.coef_)
print('The R^2 is = ',model_lm.score(x,y))
```

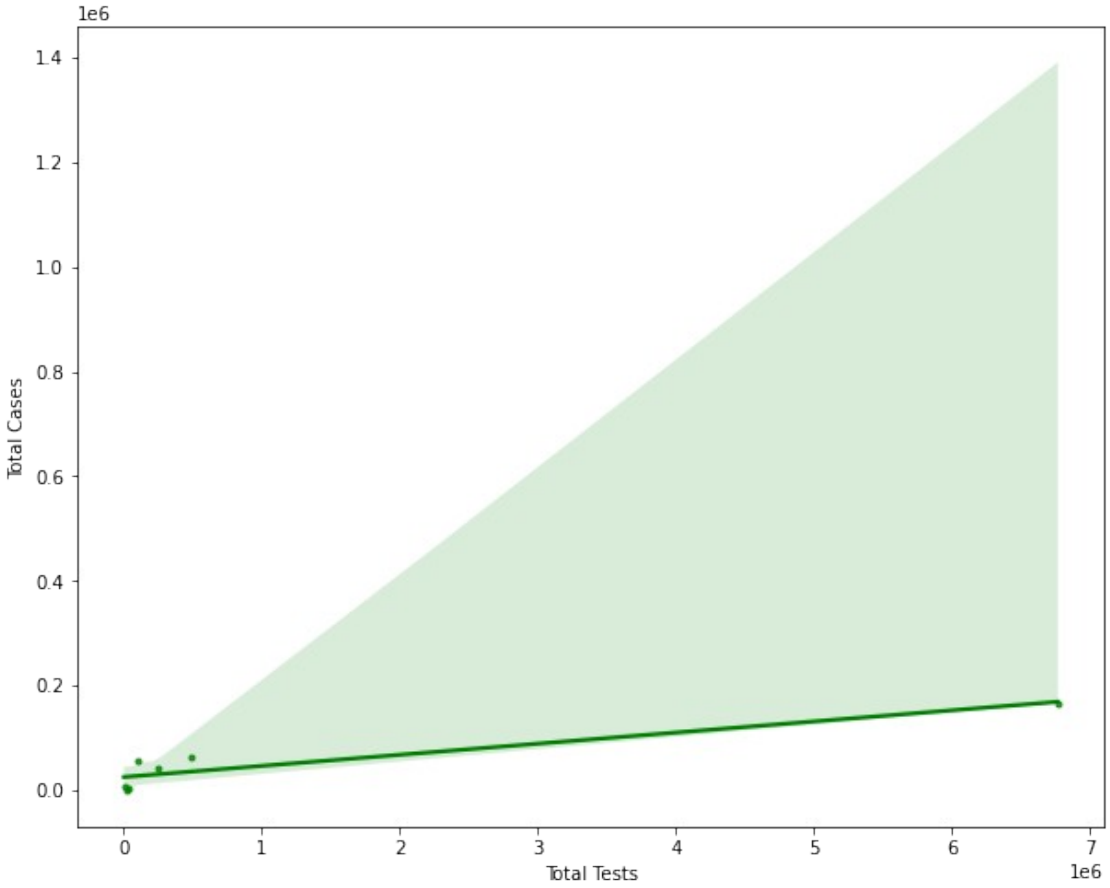
```
The y intercep, b0 is=  [25064.09771372]
The coef., b1 is=  [[0.02123477]]
The R^2 is =  0.8429129477352586
```

Ecuation

- Total Tests= b0+b1*Total Cases
- Total Tests= 25064.09771372+0.02123477*Total Cases

```
plt.figure(figsize= (10,8))
ax= sns.regplot(x= 'Total Tests',
                y='Total Cases',
                data=df,
                color= 'green',
                marker='.',

               )
```



The quantity of tests taken does affect on the number on the total of cases.

4. Does the number of population afects on the cuantity of deaths?

```
y=pd.DataFrame(df['Total Deaths'])
x=pd.DataFrame(df['Population'])
lm= linear_model.LinearRegression()
model_lm= lm.fit(x,y)
model_lm
```

```
LinearRegression()
```

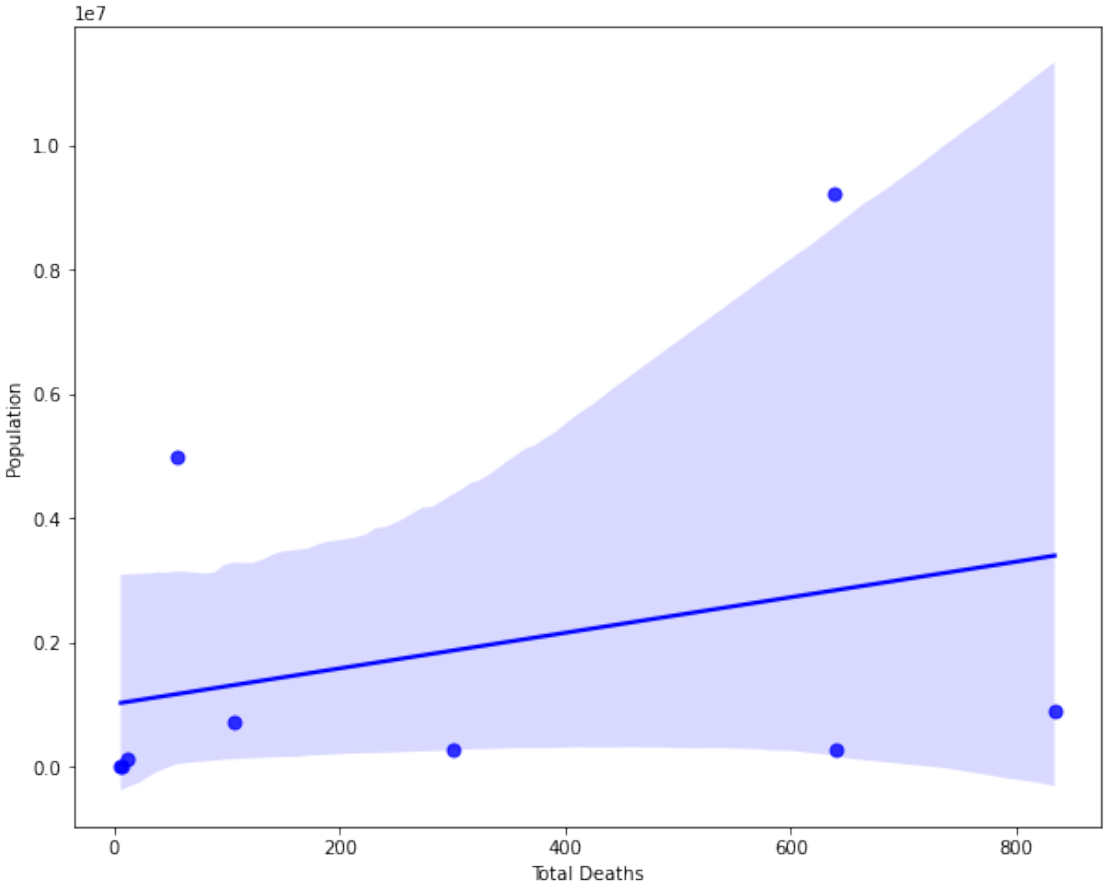
```
print('The y intercep, b0 is= ', model_lm.intercept_)
print('The coef., b1 is= ', model_lm.coef_)
print('The R^2 is = ',model_lm.score(x,y))
```

```
The y intercep, b0 is=  [232.56544006]
The coef., b1 is=  [[3.05767971e-05]]
The R^2 is =  0.08747237084823867
```

Ecuation

- Population=b0+b1* Total Deaths
- Population=232.56544006+3.05767971e-05* Total deaths

```
plt.figure(figsize= (10,8))
ax= sns.regplot(y= 'Population',
                x='Total Deaths',
                data=df,
                color= 'blue',
                marker='.',
                scatter_kws={'s':200},
               )
```



The quantity of deaths does not only depends on population, but also, some other vaiables as well.

5. What is the difference of percentaje between deaths and total recovered in Oceania?

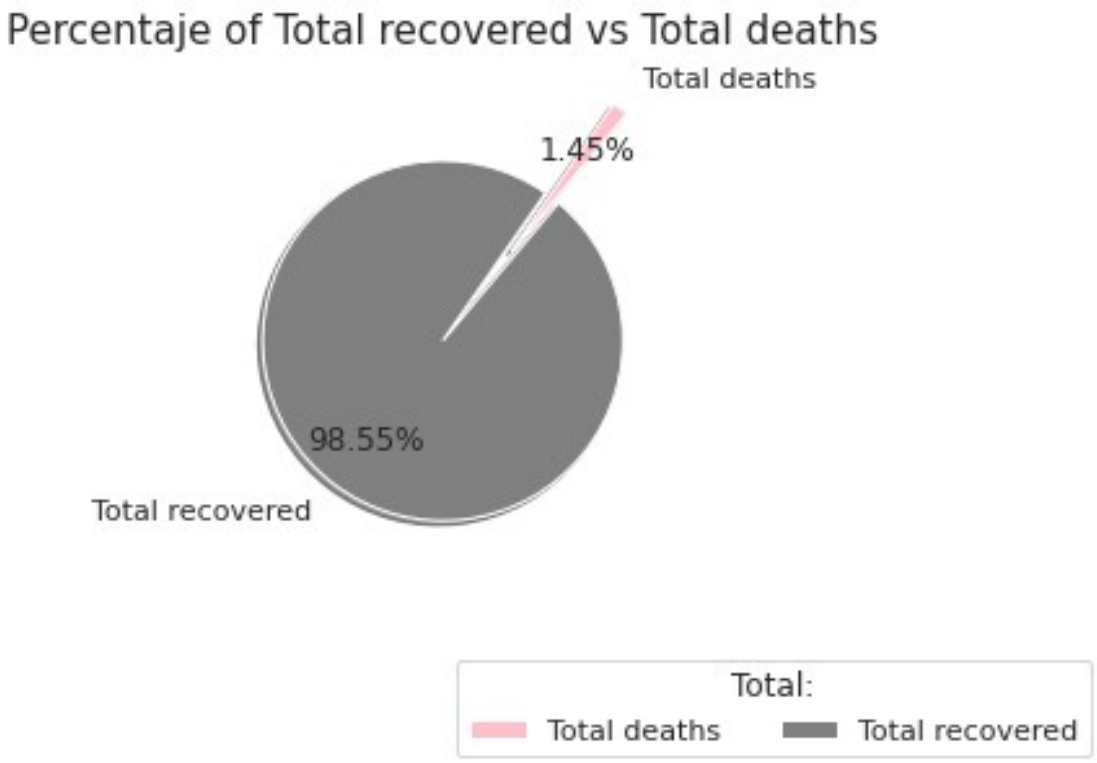
```
deaths=df['Total Deaths'].sum()
deaths
```

```
2600
```

```
recuperados=df['Total Recovered'].sum()
recuperados
```

```
176707.0
```

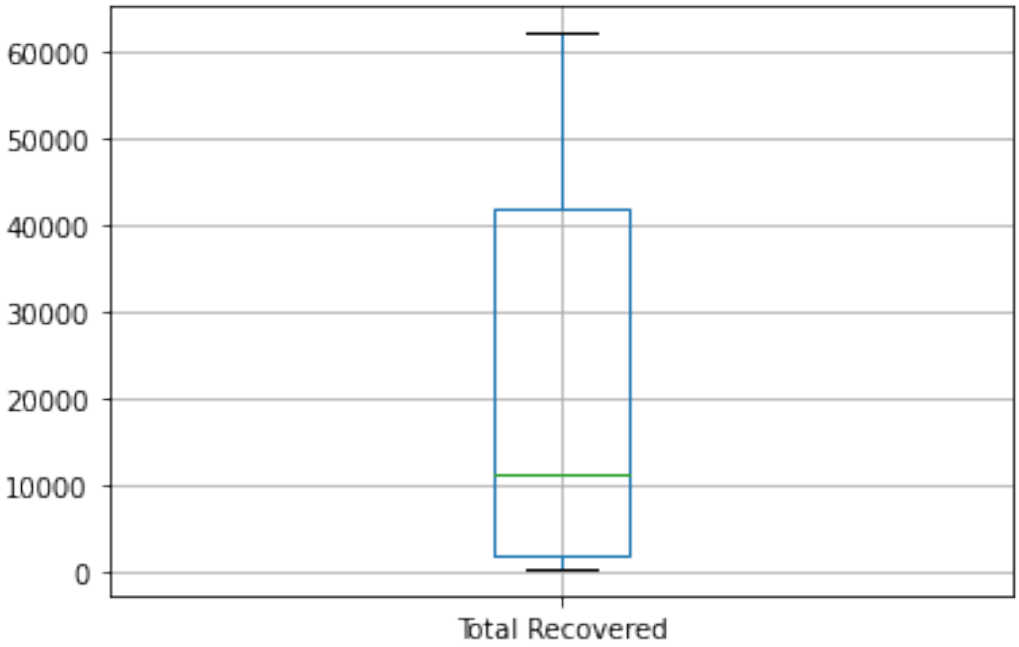
```
y = np.array([deaths,recuperados])
mylabels = ["Total deaths", "Total recovered"]
explode = (0.5, 0)
colors = ( "pink", "grey")
plt.pie(y, labels = mylabels,autopct='%1.2f%
%',explode=explode,shadow=True,startangle=50,colors=colors,radius=0.8,
pctdistance=0.7,
        labeldistance=1.2)
plt.legend(title = "Total:",loc="upper left",bbox_to_anchor=(0.5, -
0.04), ncol=2)
plt.title('Percentaje of Total recovered vs Total
deaths',fontsize=15,loc='center')
plt.show()
```



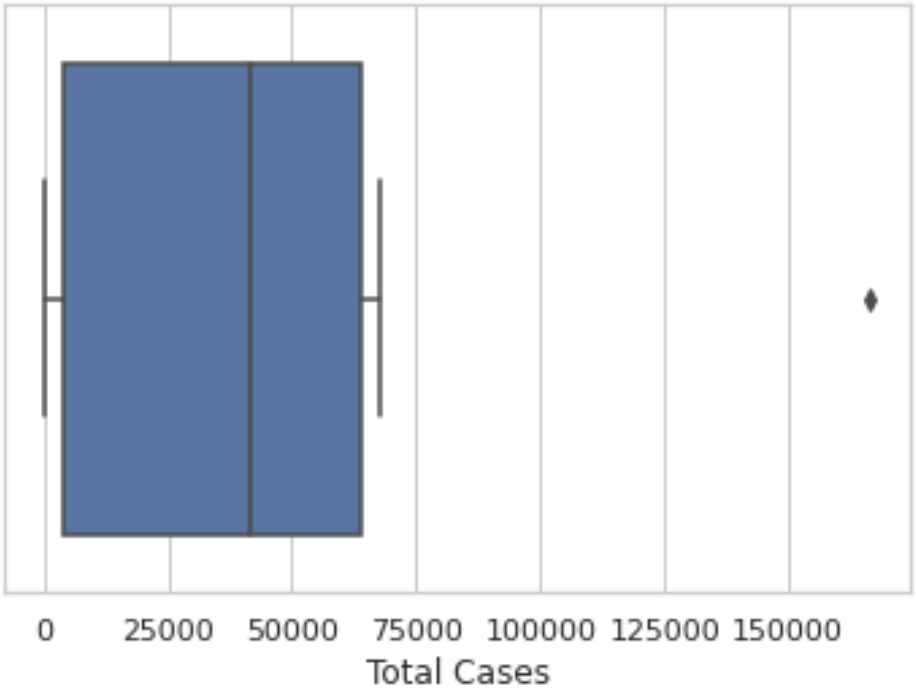
In conclusion we can say that more than 98% of the total of cases have recovered and only a 1,45% have died of covid, so we can say the mortality rate is very low.

6. What is the behaviour of the variables Total Recovered and Total cases, Do we see any observations?

```
boxplot = df.boxplot(column=['Total Recovered'])
```




```
sns.set_theme(style="whitegrid")
ax = sns.boxplot(x=df["Total Cases"])
```



As we can see the only boxplot who have an attypical value is the boxplot of total cases, so we have to see that value.