SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Analysis Preparation on COVID-19 datasets

Asma Alshahrani

Introduction

In this project, we want to specify the most affected countries in one day, different features, to calculation maximum, minmum , mean for Confirmed,Deaths, Recovered and Active, check missing values,check missing values, Seperate dates to see how many days, how many months, how many years.
The data contains 10 column and 490,69 row.
The data contains the following columns:
1)Province/State
2) Country/Region
3) Lat
4) Long
5) Date
6) Confirmed
7) Deaths
8) Recovered
9) Active
10) WHO Region

The question will disscuse about it in the data is:

- What is the Maximum,Minimum,median for the Confirmed,Deaths,Recovered, Active?
- Comaper between Deaths and Active?
-How many the Total Corona Virus Active vs Recovered؟

General Properties

In [30]:
```
1 df = pd.read_csv('covid_19_clean_complete.csv') # Load dataset anto a dataframe
```

In [31]:
```
1 df.head() # Shows the first 5 rows of data
```

Out[31]:

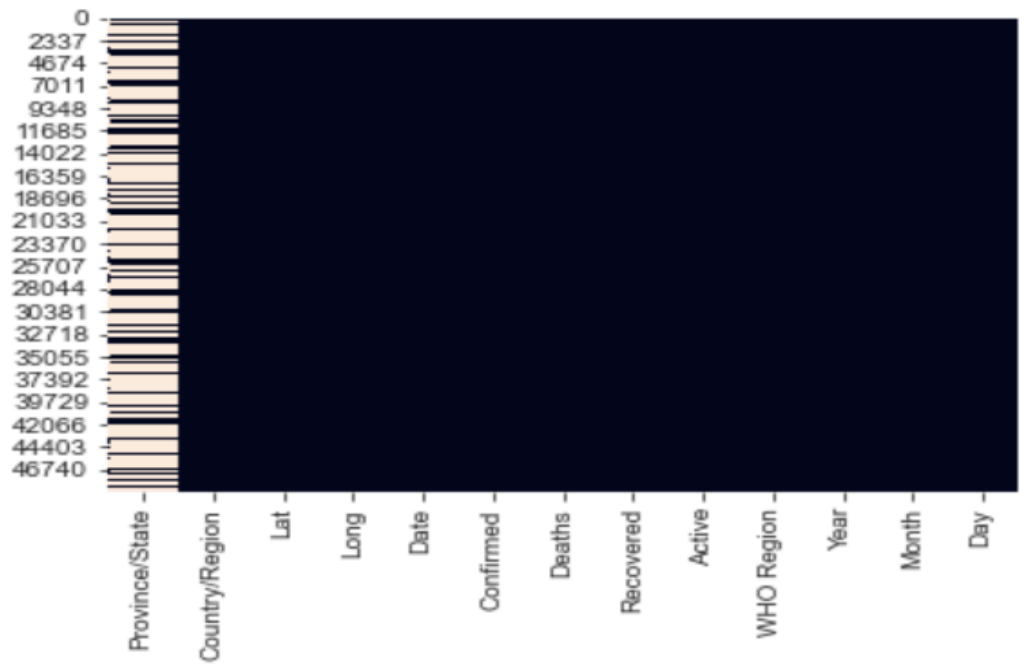
	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO Region
0	NaN	Afghanistan	33.93911	67.709953	2020-01-22	0	0	0	0	Eastern Mediterranean
1	NaN	Albania	41.15330	20.168300	2020-01-22	0	0	0	0	Europe
2	NaN	Algeria	28.03390	1.659600	2020-01-22	0	0	0	0	Africa
3	NaN	Andorra	42.50630	1.521800	2020-01-22	0	0	0	0	Europe
4	NaN	Angola	-11.20270	17.873900	2020-01-22	0	0	0	0	Africa

Visualize the missingness issue in the dataset

Data visualization ¶

```
:
1  sns.set_style("ticks") # darkgrid, whitegrid, dark, white, ticks
2  # Visualize the missingness issue in the dataset
3  sns.heatmap(df.isnull(), cbar=False)
```

```
23]: <AxesSubplot:>
```



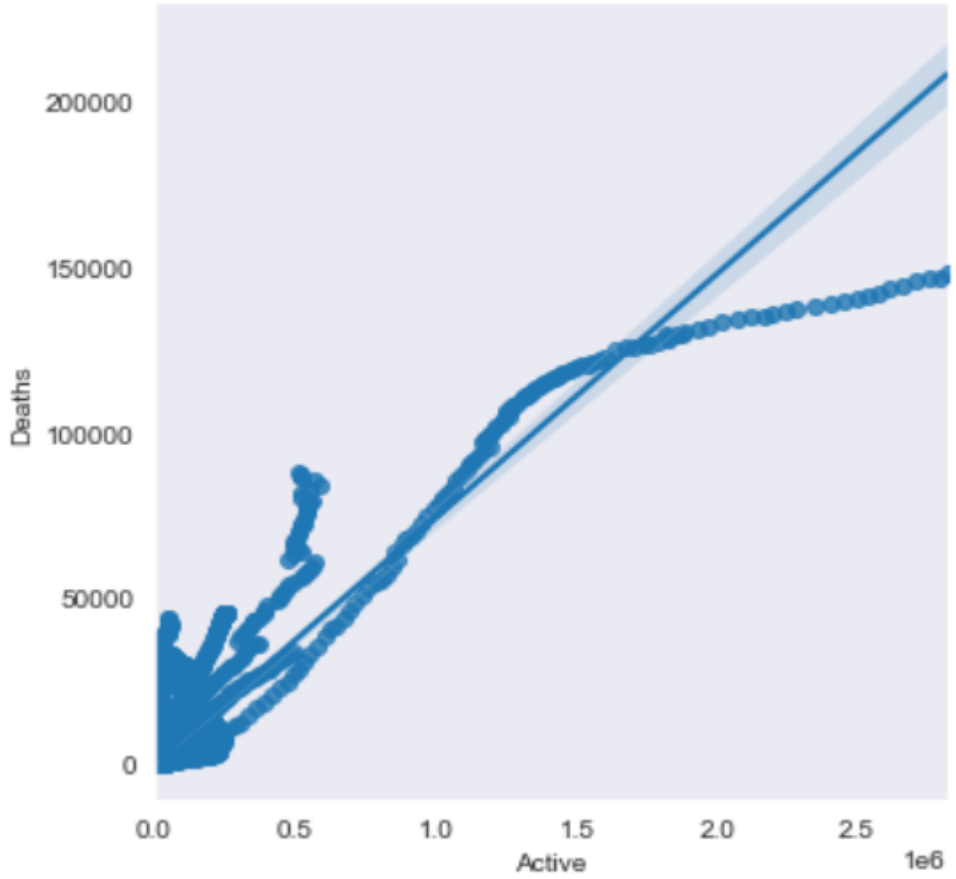
What is the Maximum,Minimum,median for the Confirmed,Deaths,Recovered, Active?

```
The Maximum number of Confirmed case= 4290259
The Minimum number of Confirmed case= 0
The median number of Confirmed case= 168.0
-------------------------------------------------
The Maximum number of Deaths case= 148011
The Minimum number of of Deaths case= 0
The median number of Deaths case= 2.0
-------------------------------------------------
The Maximum number of Recovered case= 1846641
The Minimum number of Recovered case= 0
The median number of Recovered case= 29.0
-------------------------------------------------
The Maximum number of Active case= 2816444
The Minimum number of Active case= 0
The median number of Active case= 29.0
```

Comaper between Deaths and Active?

```
1 sns.lmplot(x = 'Active', y = 'Deaths', data = df)
```

```
<seaborn.axisgrid.FacetGrid at 0x1a6152ddcd0>
```



Import LinearRegression from sklearn.linear_model

This model will help to predict active case of the COVID-١٩ per month

```
<matplotlib.legend.Legend at 0x1a6729987c0>
```