

objective1

April 22, 2020

```
[1]: import numpy as np
import pandas as pd
import plotly.graph_objects as go
import pycountry_convert as pc
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn.metrics import r2_score, mean_squared_error
from scipy.optimize import curve_fit
from lmfit import Model
```

```
[3]: import pandas as pd

df = pd.read_csv('../data/novel-corona-virus-2019-dataset/covid_19_data.csv',
                  parse_dates=['Last Update'])
df.rename(columns={'ObservationDate': 'Date',
                  'Country/Region': 'Country'}, inplace=True)

df["Eradicated"] = df["Deaths"] + df["Recovered"]
df["Active"] = df["Confirmed"] - df["Eradicated"]

df["Country"].replace(["Mainland China"], ["China"], inplace=True)
df["Country"].replace(["US"], ["United States"], inplace=True)
df["Country"].replace(["UK"], ["United Kingdom"], inplace=True)
```

```
[20]: import pycountry_convert as pc

def get_continent(row):
    continents = {
        'NA': 'North America',
        'SA': 'South America',
        'AS': 'Asia',
        'OC': 'Australia',
        'AF': 'Africa',
        'EU': 'European Union'
    }
```

```

country = row["Country"]
try:
    country_code = pc.country_name_to_country_alpha2(
        country, cn_name_format="default")
    return continents[pc.country_alpha2_to_continent_code(country_code)]
except:
    return None

df["Continent"] = df.apply(lambda row: get_continent(row), axis=1)
continents = df.groupby("Continent").sum().sort_values(
    by=['Confirmed'], ascending=False).reset_index()
continents

```

```

[20]:

```

	Continent	SNo	Confirmed	Deaths	Recovered	Eradicated	\
0	European Union	23870688	14081179.0	1053364.0	2648989.0	3702353.0	
1	Asia	29540445	8789343.0	320626.0	4396963.0	4717589.0	
2	North America	34376605	7352522.0	231780.0	434879.0	666659.0	
3	South America	4269544	589768.0	21067.0	57760.0	78827.0	
4	Africa	15038283	193531.0	8745.0	24893.0	33638.0	
5	Australia	4295643	137727.0	845.0	24061.0	24906.0	

	Active
0	10378826.0
1	4071754.0
2	6685863.0
3	510941.0
4	159893.0
5	112821.0

```

[48]: import plotly.graph_objects as go
import plotly.io as pio

fig = go.Figure()

date = df.groupby("Date")["Confirmed"].sum().reset_index()["Date"].values

for c in continents["Continent"]:
    fig.add_trace(
        go.Scatter(
            x=df[df['Continent'] == c].groupby("Date")["Confirmed"].sum().
            →reset_index()["Date"],
            y=df[df['Continent'] == c].groupby("Date")["Confirmed"].sum().
            →reset_index()["Confirmed"].values,
            name=c,
        )
    )

```

```

fig.update_layout(
    title="COVID19 Spread by Continent",
    title_x=0.5,
    xaxis_title="Date",
    # xaxis= {
    #     'tickformat': '%b',
    #     # 'tickvals': pd.date_range('2020-1', '2020-4', freq='MS')
    # },
    yaxis_title="Number of Confirmed Cases",
    font={
        "family": "Courier New, monospace",
        # size=18,
        # color="#7f7f7f"
    },
    legend={
        "x": 0,
        "y": 1
    },
    shapes=[
        # 1st highlight during Feb 4 - Feb 6
        dict(
            type="rect",
            # x-reference is assigned to the x-values
            xref="x",
            # y-reference is assigned to the plot paper [0,1]
            yref="paper",
            x0="04/01/2020",
            y0=0,
            x1=max(date),
            y1=1,
            fillcolor="LightSalmon",
            opacity=0.5,
            layer="below",
            line_width=0,
        )
    ],
)

fig.show()

```

NameError

Traceback (most recent call last)

<ipython-input-48-825d605f4af1> in <module>
 30 showgrid=False,

```

31         zeroline=False,
--> 32         range=[0, max(data["Exponential"])]
33     ),
34     font={

```

NameError: name 'data' is not defined

```
[9]: df[df["Continent"] == "European Union"].groupby("Country").sum().sort_values(
      by=['Confirmed'], ascending=False).reset_index().head()
```

```
[9]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	\
0	Italy	416626	3139708.0	358875.0	515773.0	874648.0	
1	Spain	418585	2746558.0	253417.0	748696.0	1002113.0	
2	Germany	417406	2117675.0	34001.0	656844.0	690845.0	
3	France	3402269	1878998.0	167567.0	340945.0	508512.0	
4	United Kingdom	3112429	1019040.0	106910.0	6943.0	113853.0	


```

      Active
0  2265060.0
1  1744445.0
2  1426830.0
3  1370486.0
4   905187.0

```

```
[11]: df[df["Continent"] == "Asia"].groupby("Country").sum().sort_values(
      by=['Confirmed'], ascending=False).reset_index().head()
```

```
[11]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	\
0	China	13114664	5321801.0	192276.0	3462768.0	3655044.0	
1	Iran	395636	1404759.0	88961.0	591248.0	680209.0	
2	Turkey	338701	591660.0	12207.0	28220.0	40427.0	
3	South Korea	419876	401301.0	5404.0	148481.0	153885.0	
4	Israel	393256	164966.0	1108.0	14323.0	15431.0	


```

      Active
0  1666757.0
1   724550.0
2   551233.0
3   247416.0
4   149535.0

```

```
[12]: df[df["Continent"] == "North America"].groupby("Country").sum().sort_values(
      by=['Confirmed'], ascending=False).reset_index().head()
```

```
[12]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	\
0	United States	22589517	6886003.0	219107.0	353340.0	572447.0	
1	Canada	4936864	310892.0	6696.0	66652.0	73348.0	
2	Mexico	379151	46334.0	2225.0	11503.0	13728.0	
3	Panama	342148	39429.0	950.0	286.0	1236.0	
4	Dominican Republic	372904	36568.0	1759.0	954.0	2713.0	

	Active
0	6313556.0
1	237544.0
2	32606.0
3	38193.0
4	33855.0

```
[13]: df[df["Continent"] == "South America"].groupby("Country").sum().sort_values(
        by=['Confirmed'], ascending=False).reset_index().head()
```

```
[13]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	Active
0	Brazil	381395	254264.0	12113.0	5217.0	17330.0	236934.0
1	Chile	365644	93999.0	744.0	17247.0	17991.0	76008.0
2	Ecuador	371906	86784.0	3698.0	3724.0	7422.0	79362.0
3	Peru	358232	71525.0	1899.0	19575.0	21474.0	50051.0
4	Colombia	356478	33941.0	943.0	2270.0	3213.0	30728.0

```
[14]: df[df["Continent"] == "Africa"].groupby("Country").sum().sort_values(
        by=['Confirmed'], ascending=False).reset_index().head()
```

```
[14]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	Active
0	South Africa	362413	35865.0	240.0	3004.0	3244.0	32621.0
1	Egypt	404974	28397.0	1845.0	6281.0	8126.0	20271.0
2	Algeria	383742	26085.0	3124.0	4493.0	7617.0	18468.0
3	Morocco	371481	21109.0	1400.0	1618.0	3018.0	18091.0
4	Tunisia	366775	11002.0	394.0	321.0	715.0	10287.0

```
[15]: df[df["Continent"] == "Australia"].groupby("Country").sum().sort_values(
        by=['Confirmed'], ascending=False).reset_index().head()
```

```
[15]:
```

	Country	SNo	Confirmed	Deaths	Recovered	Eradicated	\
0	Australia	3297311	118101.0	809.0	19703.0	20512.0	
1	New Zealand	380113	19356.0	36.0	4358.0	4394.0	
2	Fiji	291193	229.0	0.0	0.0	0.0	
3	Papua New Guinea	285890	35.0	0.0	0.0	0.0	
4	Guam	41136	6.0	0.0	0.0	0.0	

	Active
0	97589.0
1	14962.0

```
2    229.0
3     35.0
4      6.0
```

```
[ ]: name = "South Korea"

country = df[df['Country'] == name].groupby(
    "Date")["Date", "Confirmed"].sum().reset_index()
country["Days"] = country.index
country = country.loc[(country['Confirmed'] > 0.01*country['Confirmed'].max())]
```