

به نام خدا



موضوع

پروژه مصورسازی

نام دانشجو

سارا معصومی

شماره دانشجویی : 40164041327

اساتید

دکتر موسی گلعلی زاده – دکتر علیرضا حسن زاده

مقدمه

در این پروژه قصد داریم همه گیری ویروس کرونا را در سراسر جهان به صورت بصری بررسی کنیم. بنابر جدیدترین آمار جهانی کرونا، مجموع مبتلایان به بیماری کرونا در دنیا تاکنون به ۶۶۴ میلیون و ۷۲۶ هزار و ۳۴۶ نفر رسیده و مرگ شش میلیون و ۶۹۶ هزار و ۷۲۲ بیمار نیز تایید شده است.

ابتدا مجموعه داده های مورد نیاز را از آدرس <https://covid19.who.int/data> دانلود میکنیم. این مجموعه داده شامل اطلاعات ثبت شده در مورد تعداد مبتلایان جدید و تعداد فوتی های جدید در هر روز برای تمامی کشورها میباشد. در یکی از ستون های داده های خود ستونی به نام WHO_region وجود دارد که مقادیر این ستون تقسیم بندی نقشه جغرافیایی بر اساس سازمان بهداشت جهانی میباشد و مقادیر آن به شرح زیر میباشد:

❖ دفتر منطقه ای آفریقایی: AFRO

❖ دفتر منطقه ای برای کشورهای آمریکایی: AMRO

❖ دفتر منطقه ای برای کشورهای مدیترانه شرقی: EMRO

❖ دفتر منطقه ای کشورهای اروپایی: EURO

❖ دفتر منطقه ای کشورهای آسیای جنوب شرقی: SEARO

❖ دفتر منطقه ای کشورهای پاسیفیک غربی (غرب اقیانوس کبیر): WPRO

همچنین جهت ترسیم نمودارها از زبان پایتون و کتابخانه های pandas seaborn numpy و plotly.express استفاده خواهیم کرد.

آماده سازی داده

1. ابتدا لازم است کتابخانه هایی که در طول کار به آنها نیاز میشود را فراخوانی کنیم:

```
import pandas as pd
```

```
import numpy as np
```

```
import seaborn as sbn
```

```
import plotly.express as px
```

2. حال لازم است داده ها را دانلود کنیم و سپس به کمک Pandas آنها را فراخوانی کنیم و ستون تاریخ را به عنوان index معرفی میکنیم:

```
data3 = pd.read_csv("ProjectData/WHO-COVID-19-global-data.csv", parse_dates=["Date_reported"],
index_col="Date_reported")
```

```
data3.head(10)
```

Out[307]:

Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
2020-01-08	AF	Afghanistan	EMRO	0	0	0	0
2020-01-09	AF	Afghanistan	EMRO	0	0	0	0
2020-01-10	AF	Afghanistan	EMRO	0	0	0	0
2020-01-11	AF	Afghanistan	EMRO	0	0	0	0
2020-01-12	AF	Afghanistan	EMRO	0	0	0	0

تفسیر: داده های فوق به ترتیب از سمت چپ شامل ستون های تاریخ ثبت اطلاعات ، نام اختصاری کشور ها ، نام کامل کشور، نام منطقه بهداشتی متناظر آنها ، تعداد مبتلایان جدید، فراوانی تجمعی مبتلایان جدید، تعدا فوتی های جدید و فراوانی تجمعی فوتی های جدید میباشد. که ستون تاریخ ثبت به عنوان index در نظر گرفته شده است.

3. برای مرتب سازی داده ها آنها را بر اساس تاریخ ثبت مرتب میکنیم:

```
data3.sort_values(by=['Date_reported'],inplace=True)
```

```
data3.head()
```

Out[398]:

Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
2020-01-03	ME	Montenegro	EURO	0	0	0	0
2020-01-03	TG	Togo	AFRO	0	0	0	0
2020-01-03	BO	Bolivia (Plurinational State of)	AMRO	0	0	0	0
2020-01-03	IQ	Iraq	EMRO	0	0	0	0

تفسیر: همانطور که در خروجی بالا میبینیم داده ها بر اساس تاریخ ثبت مرتب شده اند اما برخی از ردیف ها تمام مقادیرشان برابر صفر میباشد. برای اینکه حجم داده ها کمتر شود و کار با آنها ساده تر شود ردیف هایی را که شامل هیچ مقدار عددی نمیشاند حذف میکنیم.

4. به کمک دستور زیر میتوانیم ردیف هایی را که تمام مقادیر عددی آنها برابر صفر است از مجموعه داده های خود حذف کنیم :

```
data3 = data3.loc[(data3["New_cases"]!= 0) | (data3["New_deaths"]!= 0) |  
(data3["Cumulative_cases"]!= 0) | (data3["Cumulative_deaths"]!= 0) ]
```

```
data3.head()
```

Out[405]:

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
99295	2020-01-04	DE	Germany	EURO	1	1	0	0
53467	2020-01-04	CN	China	WPRO	1	1	0	0
90384	2020-01-04	FI	Finland	EURO	1	1	0	0
53468	2020-01-05	CN	China	WPRO	0	1	0	0
99296	2020-01-05	DE	Germany	EURO	0	1	3	3
90385	2020-01-05	FI	Finland	EURO	0	1	0	0
53469	2020-01-06	CN	China	WPRO	3	4	0	0
90386	2020-01-06	FI	Finland	EURO	0	1	0	0
99297	2020-01-06	DE	Germany	EURO	0	1	0	3
99298	2020-01-07	DE	Germany	EURO	0	1	0	3

تفسیر: همانطور که در جدول صفحه قبل میبینیم دیگر هیچ ردیفی وجود ندارد که شامل اطلاعات مفیدی نباشد. و همه ی ردیف ها حداقل یکی از ستون های متناظرشان دارای مقدار عددی میباشد. اما تاریخ های ثبت شده بسیار جزئی هستند و حجم داده ها بسیار بالاست برای راحتی کار میتوانیم تاریخ ها را در مقیاس کلی تر مانند مقیاس ماهانه بررسی کنیم.

5. برای تبدیل تاریخ از روزانه به ماهانه از دستورات زیر کمک میگیریم :

```
data4 = data3.resample('M').sum()
```

```
data4.head(10)
```

Out[407]:

	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
Date_reported				
2020-01-31	9846	37542	219	1010
2020-02-29	75362	1633808	2717	47134
2020-03-31	696282	8087962	40577	374902
2020-04-30	2223610	57137169	193612	4269307
2020-05-31	2887315	136685188	164398	10082180
2020-06-30	4268490	237586653	153671	14389302
2020-07-31	6973772	419917707	184438	20040235
2020-08-31	8287664	662789587	193832	26086537
2020-09-30	8623887	893625296	168305	30679343
2020-10-31	11771167	1230304436	177318	36907448

تفسیر: تاریخ ثبت اطلاعات به صورت ماهانه برای تمام کشور ها در نظر گرفته شده است. به عنوان مثال در ردیف اول در یک ماه از تاریخ 2020.1.1 تا 2020.1.31 در سراسر جهان چه تعداد بیمار و فوتی ناشی از کرونا ثبت شده است.

مصور سازی

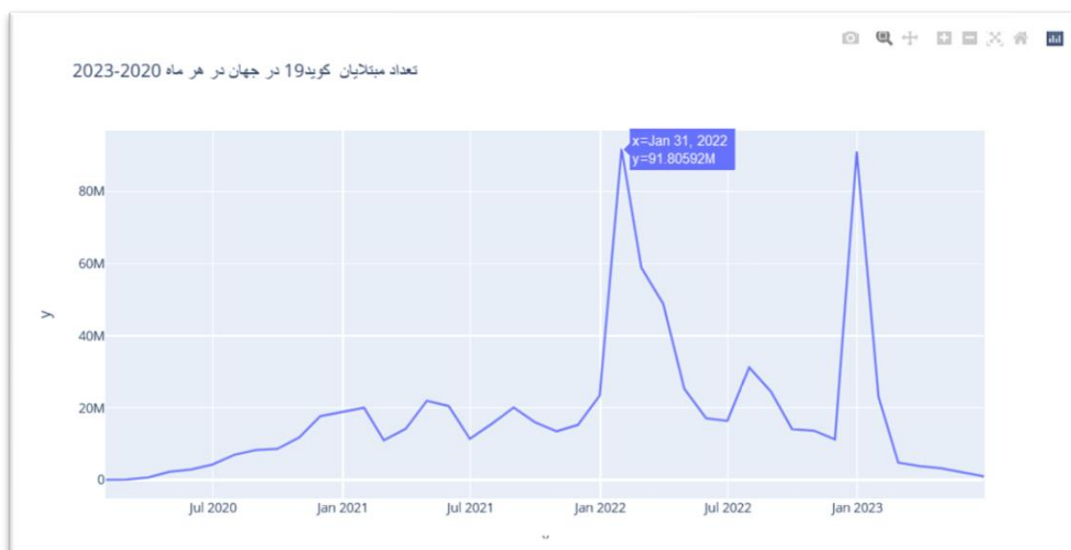
1. می‌خواهیم نموداری را جهت بررسی روند ابتلای کرونا در جهان از سال 2020-2023 رسم کنیم:

```
x= data4.index
```

```
y= data4["New_cases"]
```

```
fig = px.line( x = x , y = y, title = 'تعداد مبتلایان کوید 19 در جهان در هر ماه 2020-2023')
```

```
fig.show()
```



تفسیر: همانطور که در تصویر بالا می‌بینیم تعداد مبتلایان جهان در طول زمان ترسیم شده اند نمودار رسم شده دارای 6 موج است که بیشترین ابتلا در این 3 سال مربوط به 2022 jan 31 و پس از آن اوایل سال 2023 می‌باشد. برای بررسی بهتر می‌توانیم این نمودار را برای تعداد فوتی های ناشی از کرونا نیز رسم کنیم.

این نمودار قابلیت های **hover** و **zoom** را دارا می‌باشد.

```
x= data4.index
```

```
y= data4["New_deaths"]
```

```
fig2 = px.line( x = x , y = y, title = 'تعداد فوتی های کوید19 در جهان در هر ماه 2020-2023')
```

```
fig2.show()
```



تفسیر: همانطور که در نمودار فوق میبینیم بیشترین فوتی های ناشی از کرونا در تاریخ 2021 jan 31 ثبت شده است در واقع یعنی در زمانی که تازه موج اول کرونا جهان را بر گرفته بوده است. در تاریخ 2023 jan که تعداد مبتلایان بسیار بیشتر از 2021 jan است اما تعداد فوتی ها بسیار کمتر میباشد. دلیل این اتفاق آشنایی مردم با شیوه های مراقبتی در طول بیماری و همچنین استفاده از واکسن کرونا میباشد.

2. می‌خواهیم همه گیری ویروس کرونا و فوتی های ناشی از آن را به تفکیک مناطق بهداشتی بر روی نمودار ببینیم:

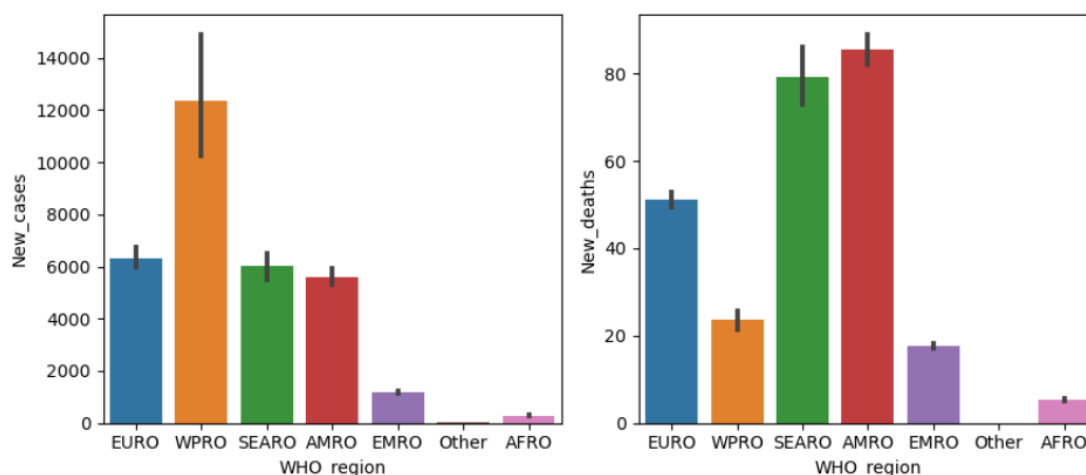
```
plt.rcParams["figure.figsize"] = [9.00, 4.00]
```

```
plt.rcParams["figure.autolayout"] = True
```

```
fig, axs = plt.subplots(1,2)
```

```
sbn.barplot(x="WHO_region", y="New_cases", data=data3, ax=axs[0])
```

```
sbn.barplot(x="WHO_region", y="New_deaths", data=data3, ax=axs[1])
```



تفسیر: در تصویر فوق نمودار سمت چپ نمودار تعداد مبتلایان جدید به تفکیک منطقه بهداشتی است و نمودار سمت راست تعداد فوتی های جدید به تفکیک مناطق بهداشتیست. همانطور که در نمودار سمت چپ می‌بینیم بیشترین ابتلا مربوط به منطقه WPRO یعنی غرب اقیانوس کبیر میباشد و کمترین ابتلا مربوط به منطقه AFRO یعنی کشور های افریقایی میباشد. اما امار فوتی ها مطابق امار مبتلایان نیست. و بیشترین فوتی ها مربوط به منطقه ی AMRO یعنی کشورهای آمریکایی میباشد بنابراین با توجه به تعدا مبتلایان و تعداد فوتی های کشور های آمریکایی میتوان نتیجه گرفت کشور های آمریکایی در بهبود حال بیماران کرونایی عملکرد ضعیفی داشته اند اما در مقابل کشور های منطقه ی WPRO با توجه به این که بیشترین امار مبتلایان ثبت شده مربوط به این منطقه است اما در بهبود حال بیماران عملکرد خوبی داشته اند چرا که امار فوتی های این منطقه نسبت به سایر مناطق کمتر است.

3. می‌خواهیم برای متغیر تعداد مبتلایان جدید و متغیر تعداد فوتی‌های جدید نمودار جعبه‌ای رسم کنیم تا ببینیم آیا مشاهده‌ی پرت برایشان وجود دارد یا خیر. همچنین ارتباط آنها را نیز روی نمودار ببینیم.

اما برای این کار ابتدا لازم است داده‌های خود را یکبار دیگر آماده‌سازی کنیم چرا که نباید ستون تاریخ به عنوان index تعریف شده باشد :

```
data2 = pd.read_csv("ProjectData/WHO-COVID-19-global-data.csv")
```

```
data2.sort_values(by=['Date_reported'],inplace=True)
```

```
data2 = data2.loc[(data2["New_cases"]!= 0) | (data2["New_deaths"]!= 0) |  
(data2["Cumulative_cases"]!= 0) | (data2["Cumulative_deaths"]!= 0) ]
```

```
data2.head()
```

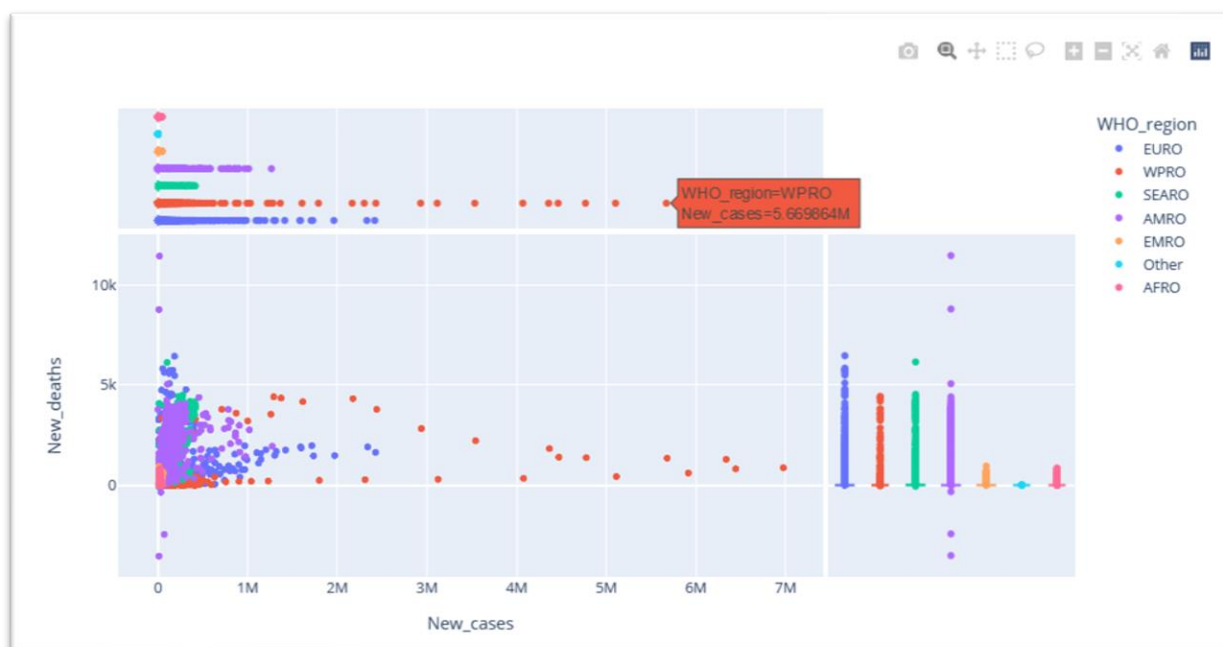
Out[405]:

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
99295	2020-01-04	DE	Germany	EURO	1	1	0	0
53467	2020-01-04	CN	China	WPRO	1	1	0	0
90384	2020-01-04	FI	Finland	EURO	1	1	0	0
53468	2020-01-05	CN	China	WPRO	0	1	0	0
99296	2020-01-05	DE	Germany	EURO	0	1	3	3
90385	2020-01-05	FI	Finland	EURO	0	1	0	0
53469	2020-01-06	CN	China	WPRO	3	4	0	0
90386	2020-01-06	FI	Finland	EURO	0	1	0	0
99297	2020-01-06	DE	Germany	EURO	0	1	0	3
99298	2020-01-07	DE	Germany	EURO	0	1	0	3

ادامه صفحه بعد

```
fig2 = px.scatter(data2, x = "New_cases", y = "New_deaths", color="WHO_region",
marginal_y="box",marginal_x="box")
```

```
fig2.show()
```



تفسیر: همانطور که در نمودار فوق میبینیم تعداد مبتلایان برای منطقه WPRO شامل داده های پرت میباشد یعنی در بعضی ماه ها تعداد مبتلایان بسیار زیاد تر از ماه های دیگر بوده است اما در منطقه WPRO برای تعداد فوتی ها هیچ داده ی پرتی مشاهده نمیشود وقتی به نمودار پراکنش برای این منطقه نگاه کنیم میبینیم که درست است که تعداد مبتلایان این منطقه زیاد است اما در مدیریت حال بیماران برای جلوگیری از فوت آنها عملکرد بسیار خوبی داشته اند.

4. میخواهیم روند گسترش ویروس کرونا را در سراسر جهان بر روی نقشه ی جغرافیایی در طول زمان ببینیم:

```
fig = px.choropleth(data2, locations="Country",
```

```
locationmode="country names",
```

```
color="New_cases",
```

```
hover_name="Country",
```

```
animation_frame="Date_reported",
```

```
animation_group = "Date_reported",
```

```
color_continuous_scale='Greens')
```

```
fig.update_layout(
```

```
title_text = 'Renewable energy production across the world (% of electricity production)', geo = dict(projection={'type':'equiarectangular'}))
```

```
fig.show()
```



تفسیر: اگر نمودار فوق را از روز اول داده ها play کنیم میبینیم که شروع ویروس کرونا از کشور چین بوده است و سپس در برخی از کشورهای قاره ی آمریکا گسترش یافته است پساز آنها به تریج تمام کشور ها دچار این بیماری میشوند..اما با گذر زمان میبینیم که کشور چین به سرعت ابتلای ویروس کرونا را در کشور خود مهار میکند اما همچنان سایر کشور ها درگیر این بیماری هستند.

1. <https://covid19.who.int/data>
2. <https://seaborn.pydata.org/generated/seaborn.barplot.html>
3. <https://www.khabaronline.ir/news/1713571/%D8%A2%D8%AE%D8%B1%DB%8C%D9%86-%D8%A2%D9%85%D8%A7%D8%B1-%D9%85%D8%A8%D8%AA%D9%84%D8%A7%DB%8C%D8%A7%D9%86-%D9%88-%D9%82%D8%B1%D8%A8%D8%A7%D9%86%DB%8C%D8%A7%D9%86-%DA%A9%D8%B1%D9%88%D9%86%D8%A7-%D8%AF%D8%B1-%D8%AC%D9%87%D8%A7%D9%86-%DB%B1%DB%B0-%DA%A9%D8%B4%D9%88%D8%B1%DB%8C-%DA%A9%D9%87-%D8%A8%D8%A7%D9%84%D8%A7%D8%AA%D8%B1%DB%8C%D9%86>
4. <https://www.tutorialspoint.com/how-to-plot-two-seaborn-lmplots-side-by-side-matplotlib>