

▼ Author : Braham Parkash

Data Science & Business Analytics Internship

GRIP - The Spark Foundation

Task-7 Timeline Analysis_Covid-19

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import pylab as pl
import seaborn as sns
%matplotlib inline
sns.set()
```

```
!wget --header="Host: covid.ourworldindata.org" --header="User-Agent: Mozilla/5.0 (Windows
--2020-11-14 10:09:01-- https://covid.ourworldindata.org/data/owid-covid-data.csv
Resolving covid.ourworldindata.org (covid.ourworldindata.org)... 104.248.63.231, 192
Connecting to covid.ourworldindata.org (covid.ourworldindata.org)|104.248.63.231|:443:
HTTP request sent, awaiting response... 200 OK
Length: 13138568 (13M) [text/csv]
Saving to: 'owid-covid-data.csv'

owid-covid-data.csv 100%[=====>] 12.53M 3.30MB/s in 3.8s

2020-11-14 10:09:05 (3.28 MB/s) - 'owid-covid-data.csv' saved [13138568/13138568]
```



```
covid=pd.read_csv('owid-covid-data.csv')
```

```
covid.head()
```



	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed
0	ABW	North America	Aruba	2020-03-13	2.0	2.0	NaN

covid.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 56537 entries, 0 to 56536
Data columns (total 50 columns):
```

#	Column	Non-Null Count	Dtype
0	iso_code	56218 non-null	object
1	continent	55899 non-null	object
2	location	56537 non-null	object
3	date	56537 non-null	object
4	total_cases	52901 non-null	float64
5	new_cases	55612 non-null	float64
6	new_cases_smoothed	54802 non-null	float64
7	total_deaths	43595 non-null	float64
8	new_deaths	55612 non-null	float64
9	new_deaths_smoothed	54802 non-null	float64
10	total_cases_per_million	52618 non-null	float64
11	new_cases_per_million	55548 non-null	float64
12	new_cases_smoothed_per_million	54737 non-null	float64
13	total_deaths_per_million	43327 non-null	float64
14	new_deaths_per_million	55548 non-null	float64
15	new_deaths_smoothed_per_million	54737 non-null	float64
16	reproduction_rate	37696 non-null	float64
17	icu_patients	4490 non-null	float64
18	icu_patients_per_million	4490 non-null	float64
19	hosp_patients	5005 non-null	float64
20	hosp_patients_per_million	5005 non-null	float64
21	weekly_icu_admissions	357 non-null	float64
22	weekly_icu_admissions_per_million	357 non-null	float64
23	weekly_hosp_admissions	645 non-null	float64
24	weekly_hosp_admissions_per_million	645 non-null	float64
25	total_tests	21815 non-null	float64
26	new_tests	21584 non-null	float64
27	total_tests_per_thousand	21815 non-null	float64
28	new_tests_per_thousand	21584 non-null	float64
29	new_tests_smoothed	24377 non-null	float64
30	new_tests_smoothed_per_thousand	24377 non-null	float64
31	tests_per_case	22568 non-null	float64
32	positive_rate	22977 non-null	float64
33	tests_units	25333 non-null	object
34	stringency_index	46612 non-null	float64
35	population	56218 non-null	float64
36	population_density	53566 non-null	float64
37	median_age	50285 non-null	float64
38	aged_65_older	49528 non-null	float64
39	aged_70_older	50023 non-null	float64
40	gdp_per_capita	49626 non-null	float64
41	extreme_poverty	33078 non-null	float64
42	cardiovasc_death_rate	50260 non-null	float64
43	diabetes_prevalence	52100 non-null	float64
44	female_smokers	39104 non-null	float64
45	male_smokers	38599 non-null	float64
46	handwashing_facilities	23796 non-null	float64
47	hospital_beds_per_thousand	45271 non-null	float64

```

48 life_expectancy          55495 non-null float64
49 human_development_index  48522 non-null float64
dtypes: float64(45), object(5)
memory usage: 21.6+ MB

```

```
covid.describe()
```

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	n
count	5.290100e+04	55612.000000	54802.000000	4.359500e+04	55612.000000	
mean	1.622905e+05	1897.966716	1862.886770	6.738682e+03	46.502050	
std	1.634413e+06	17630.823342	17112.569358	5.474821e+04	386.132319	
min	1.000000e+00	-8261.000000	-552.000000	1.000000e+00	-1918.000000	
25%	1.750000e+02	0.000000	0.857000	1.200000e+01	0.000000	
50%	2.023000e+03	14.000000	19.000000	8.300000e+01	0.000000	
75%	2.134300e+04	229.000000	238.429000	7.095000e+02	4.000000	
max	5.277527e+07	602850.000000	566895.429000	1.293106e+06	10491.000000	

```
covid.duplicated().sum()
```

```
0
```

```
covid.isna().mean()*100
```

```

iso_code          0.564232
continent         1.128465
location          0.000000
date              0.000000
total_cases       6.431187
new_cases         1.636097
new_cases_smoothed 3.068787
total_deaths      22.891204
new_deaths        1.636097
new_deaths_smoothed 3.068787
total_cases_per_million 6.931744
new_cases_per_million 1.749297
new_cases_smoothed_per_million 3.183756
total_deaths_per_million 23.365230
new_deaths_per_million 1.749297
new_deaths_smoothed_per_million 3.183756
reproduction_rate 33.325079
icu_patients      92.058298
icu_patients_per_million 92.058298
hosp_patients     91.147390
hosp_patients_per_million 91.147390
weekly_icu_admissions 99.368555
weekly_icu_admissions_per_million 99.368555
weekly_hosp_admissions 98.859154
weekly_hosp_admissions_per_million 98.859154
total_tests       61.414649
new_tests         61.823231

```

```

total_tests_per_thousand      61.414649
new_tests_per_thousand        61.823231
new_tests_smoothed            56.883103
new_tests_smoothed_per_thousand 56.883103
tests_per_case                 60.082778
positive_rate                  59.359358
tests_units                    55.192175
stringency_index               17.554876
population                     0.564232
population_density             5.254966
median_age                     11.058245
aged_65_older                  12.397191
aged_70_older                  11.521658
gdp_per_capita                 12.223853
extreme_poverty                41.493181
cardiovasc_death_rate          11.102464
diabetes_prevalence             7.847958
female_smokers                  30.834675
male_smokers                    31.727895
handwashing_facilities          57.910749
hospital_beds_per_thousand     19.926774
life_expectancy                 1.843041
human_development_index        14.176557
dtype: float64

```

```
covid.shape
```

```
(56537, 50)
```

```
#covid.dropna(inplace=True)
```

```
covid.head(5)
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed
0	ABW	North America	Aruba	2020-03-13	2.0	2.0	NaN
1	ABW	North America	Aruba	2020-03-19	NaN	NaN	0.286
2	ABW	North America	Aruba	2020-03-20	4.0	2.0	0.286
3	ABW	North America	Aruba	2020-03-21	NaN	NaN	0.286
4	ABW	North America	Aruba	2020-03-22	NaN	NaN	0.286

```
covid1=covid[['date','total_cases','new_cases','total_deaths','new_deaths','total_tests','
```

```
covid1.head()
```

	date	total_cases	new_cases	total_deaths	new_deaths	total_tests	new_tests
0	2020-03-13	2.0	2.0	NaN	0.0	NaN	NaN
1	2020-03-19	NaN	NaN	NaN	NaN	NaN	NaN
2	2020-03-20	4.0	2.0	NaN	0.0	NaN	NaN

```
covid1['date']=pd.to_datetime(covid1['date'],infer_datetime_format=True)
```

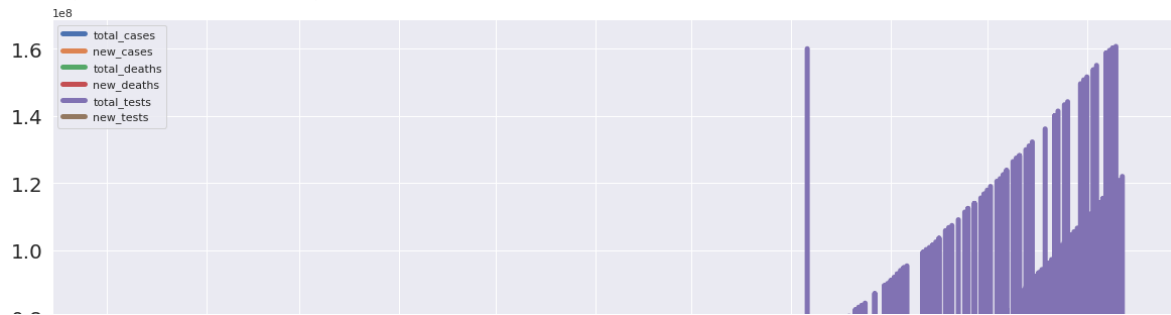
```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/10min.html>
 """Entry point for launching an IPython kernel.

```
covid1.set_index('date',inplace = True)
```

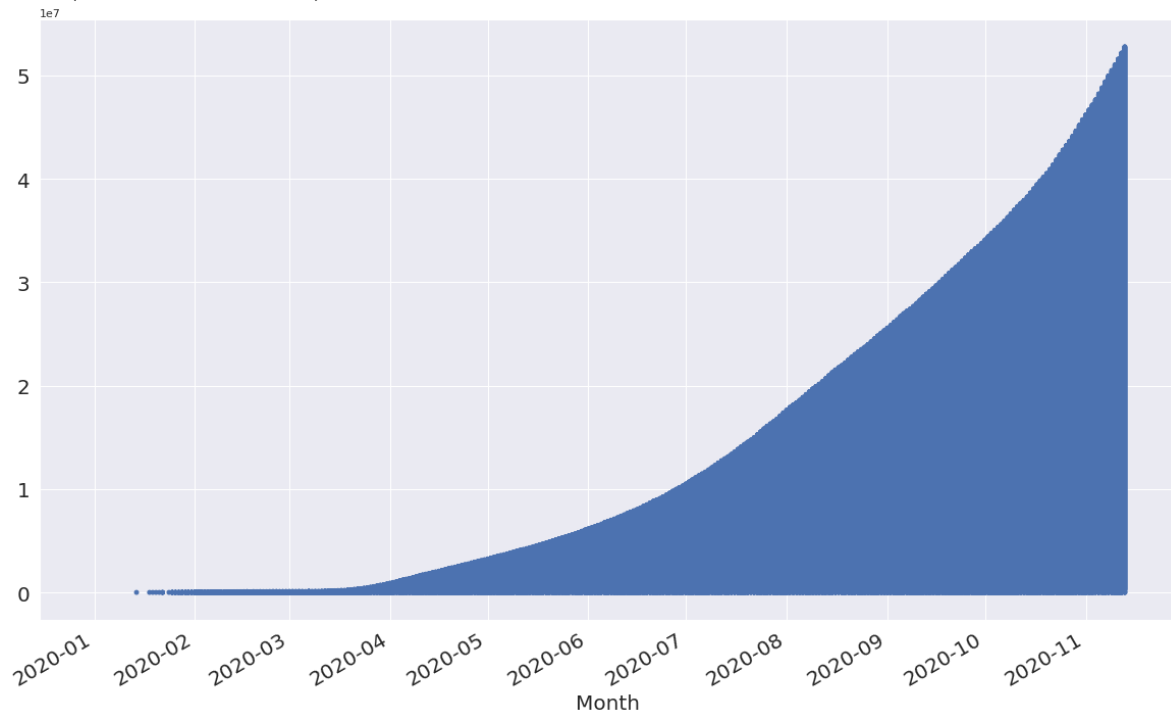
```
covid1.plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```



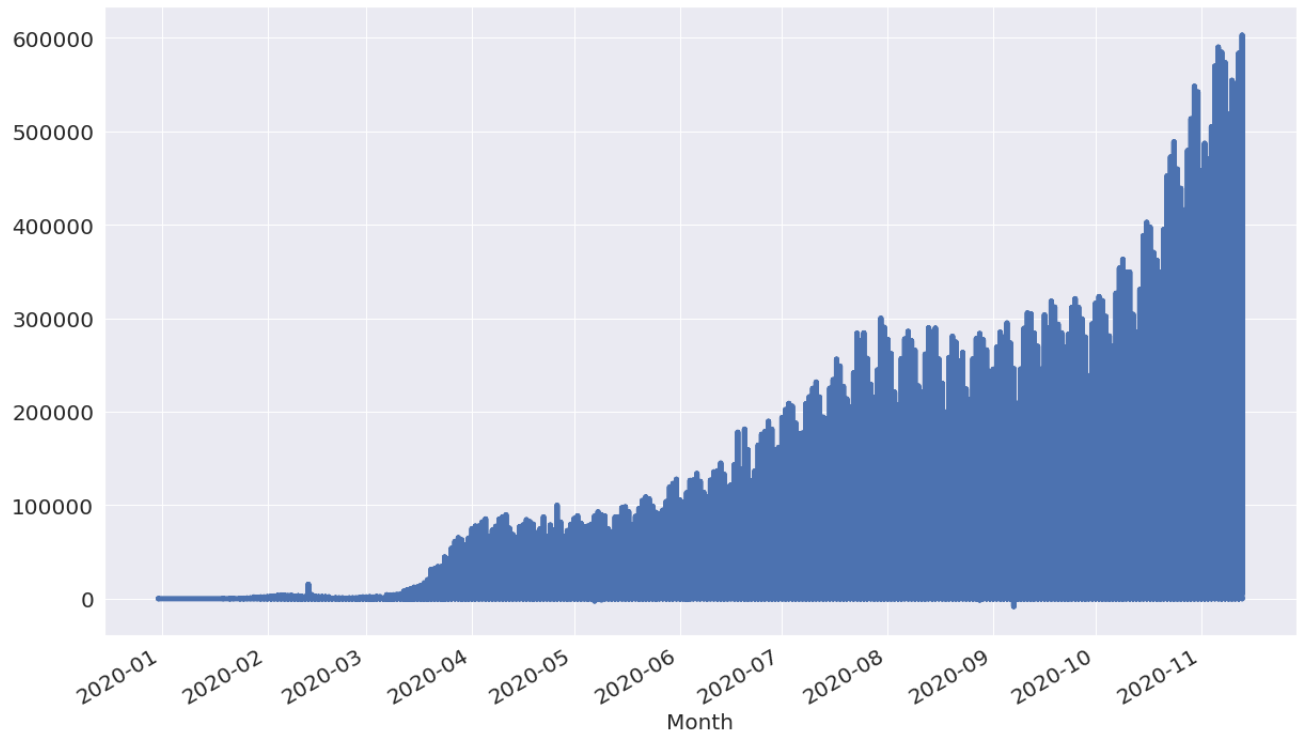
```
covid1.total_cases.plot(figsize=(20,12),linewidth=5,fontsize=20)  
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```



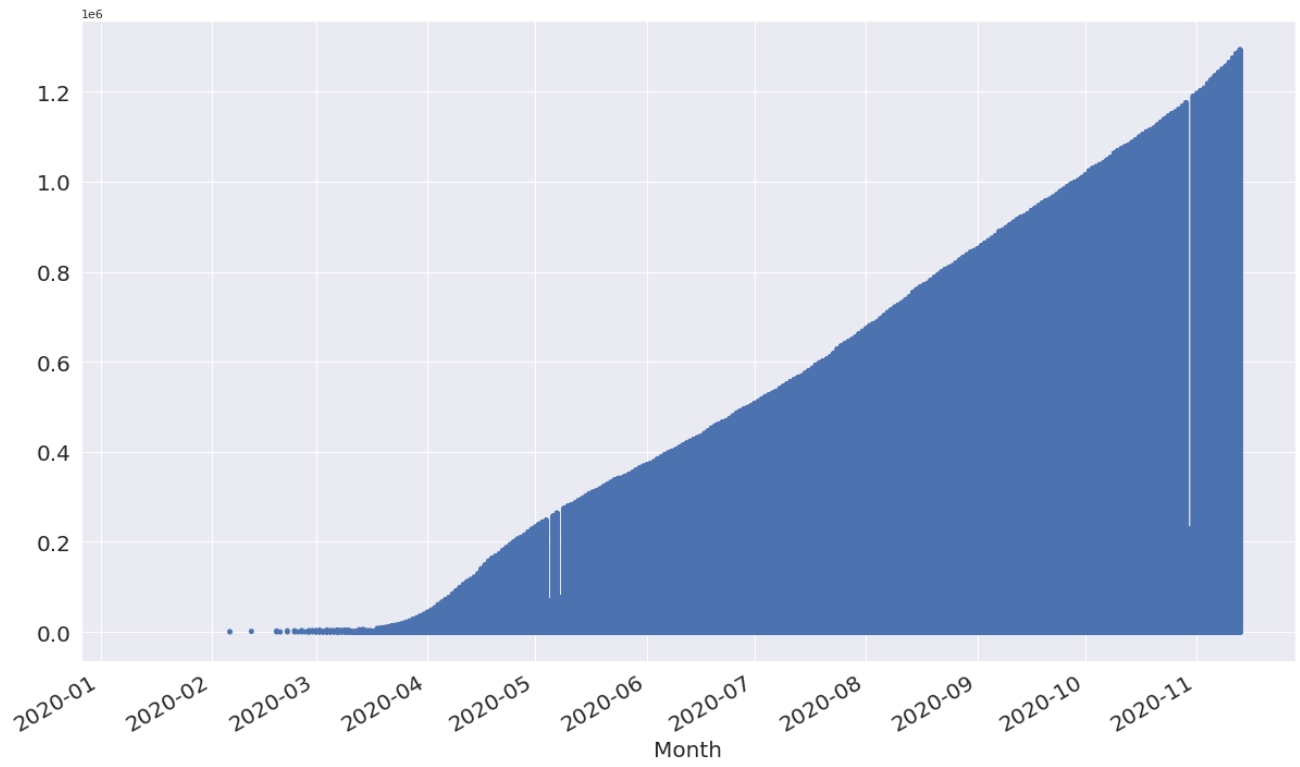
```
covid1.new_cases.plot(figsize=(20,12),linewidth=5,fontsize=20)  
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```



```
covid1.total_deaths.plot(figsize=(20,12),linewidth=5,fontsize=20)  
plt.xlabel('Month', fontsize = 20)
```

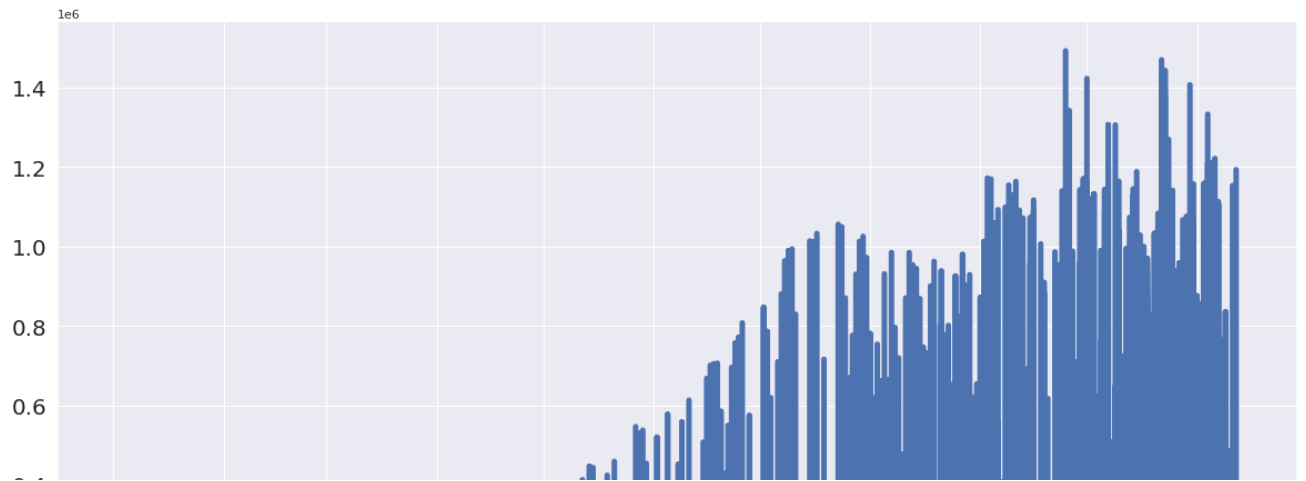
Text(0.5, 0, 'Month')



```
covid1.new_tests.plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```



```
Text(0.5, 0, 'Month')
```



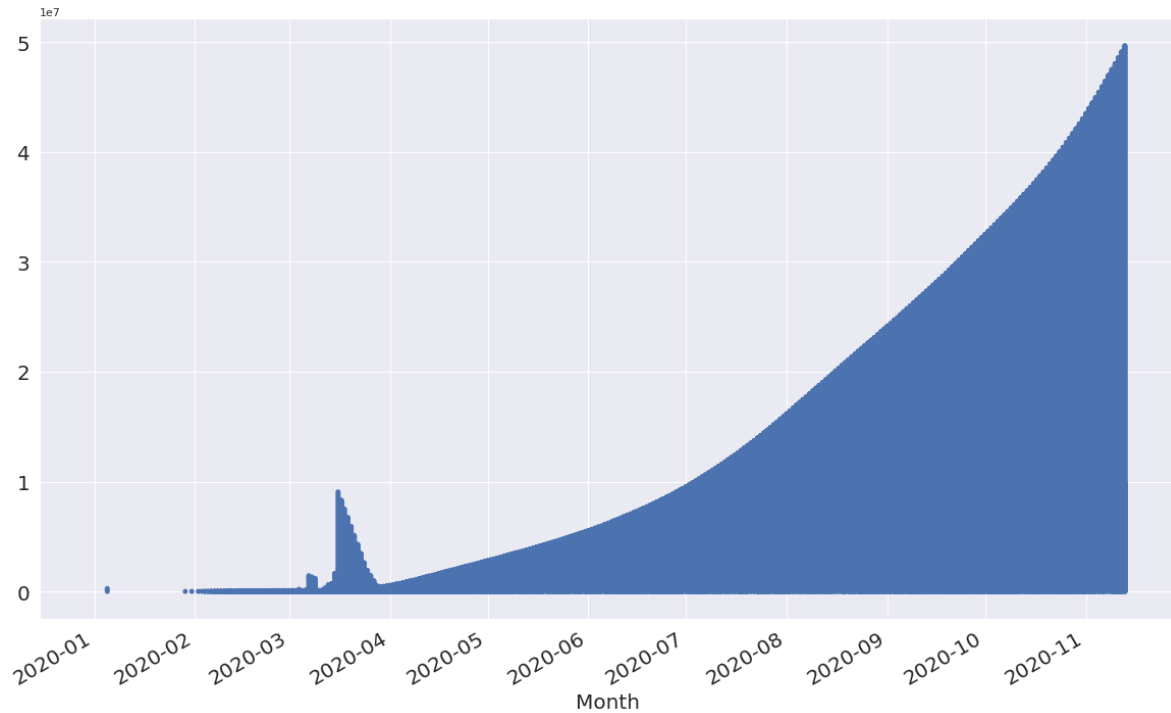
```
covid1.total_tests.plot(figsize=(20,12),linewidth=5,fontsize=20)  
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```

▼ Trends and Seasonality in TimeSeries Data

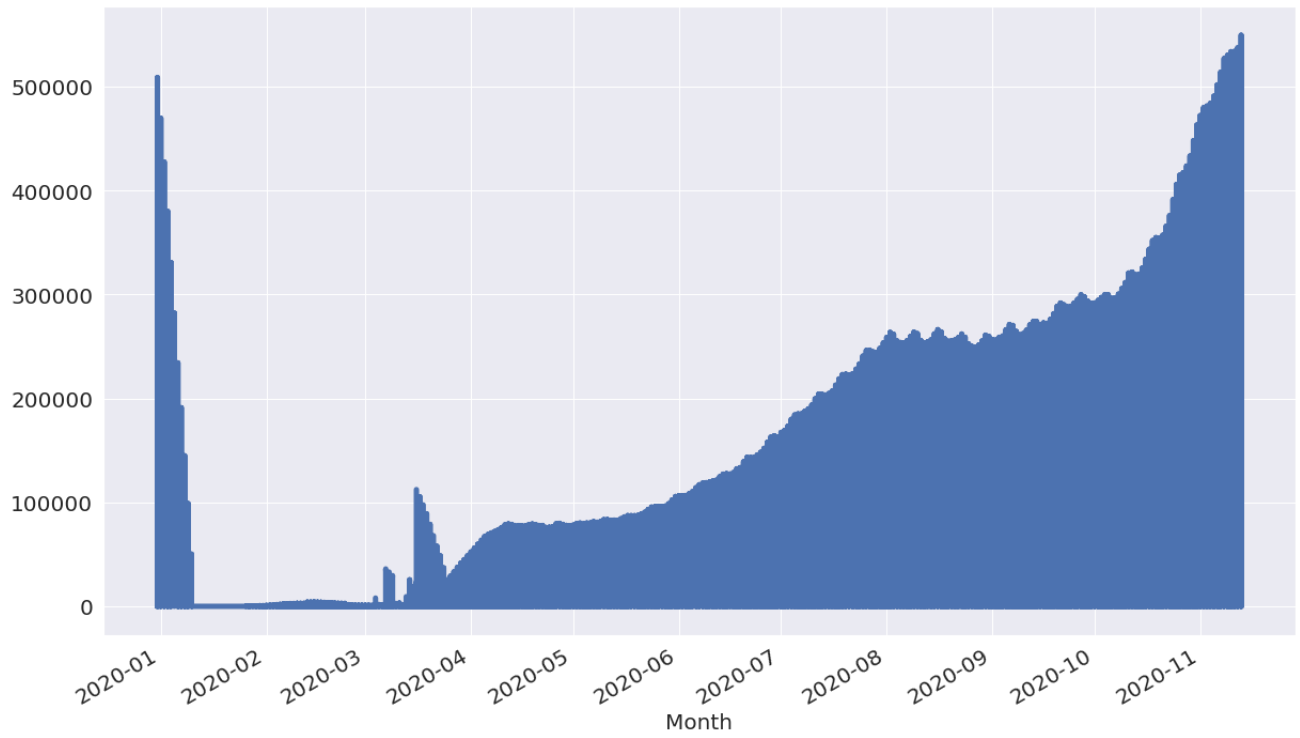
```
total_cases=covid1.total_cases
total_cases.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```



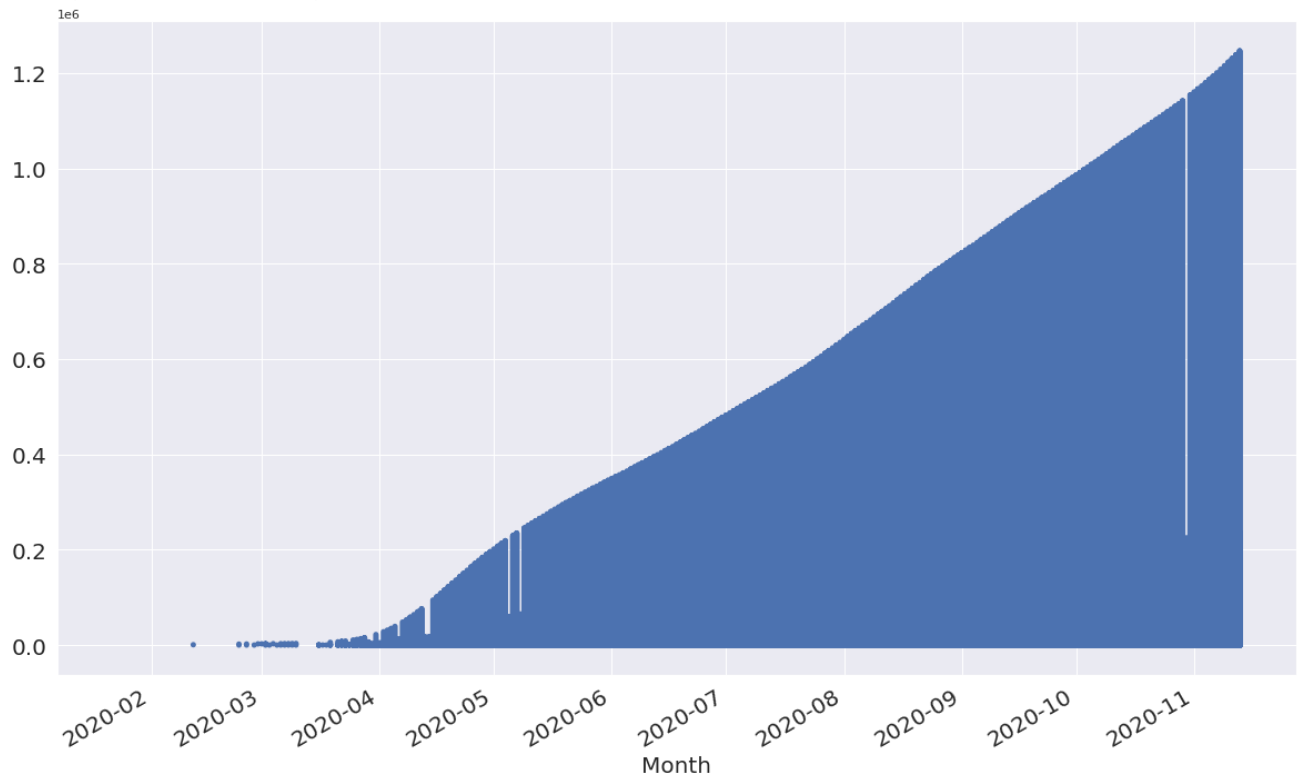
```
new_cases=covid1.new_cases
new_cases.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

Text(0.5, 0, 'Month')



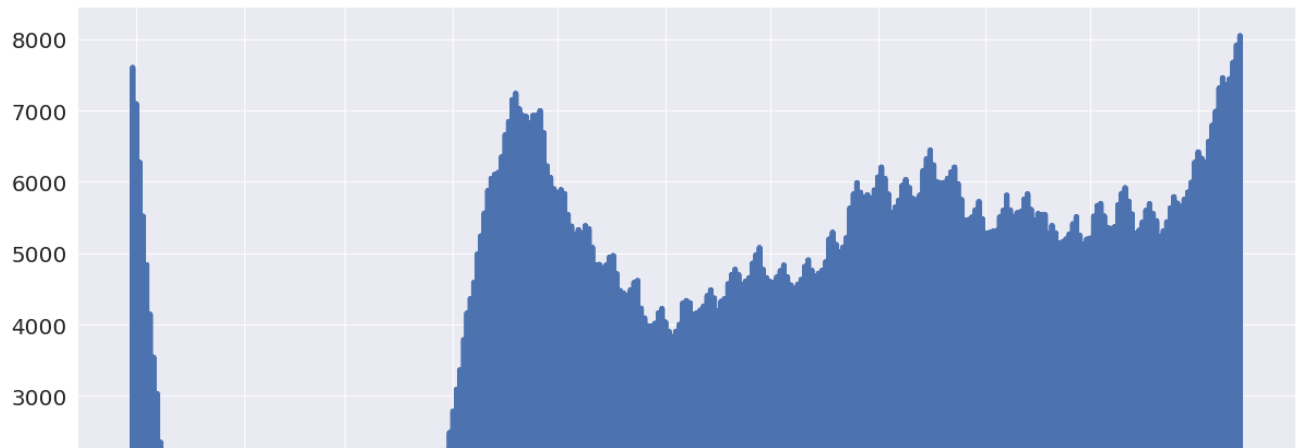
```
total_deaths=covid1.total_deaths
total_deaths.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

Text(0.5, 0, 'Month')



```
new_deaths=covid1.new_deaths
new_deaths.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

```
Text(0.5, 0, 'Month')
```



```
total_tests=covid1.total_tests  
total_tests.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)  
plt.xlabel('Month', fontsize = 20)
```

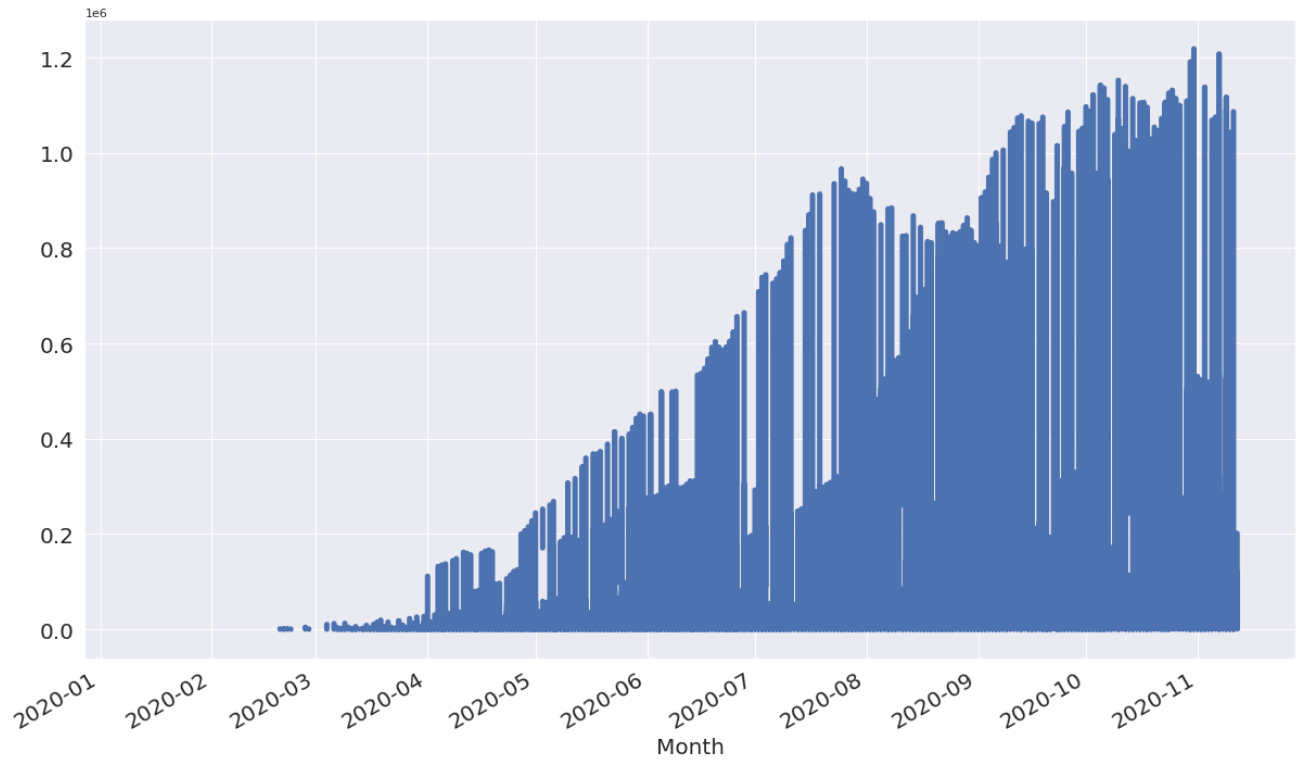
```
Text(0.5, 0, 'Month')
```

```
new_tests=covid1.new_tests
```

```
new_tests.rolling(12).mean().plot(figsize=(20,12),linewidth=5,fontsize=20)
```

```
plt.xlabel('Month', fontsize = 20)
```

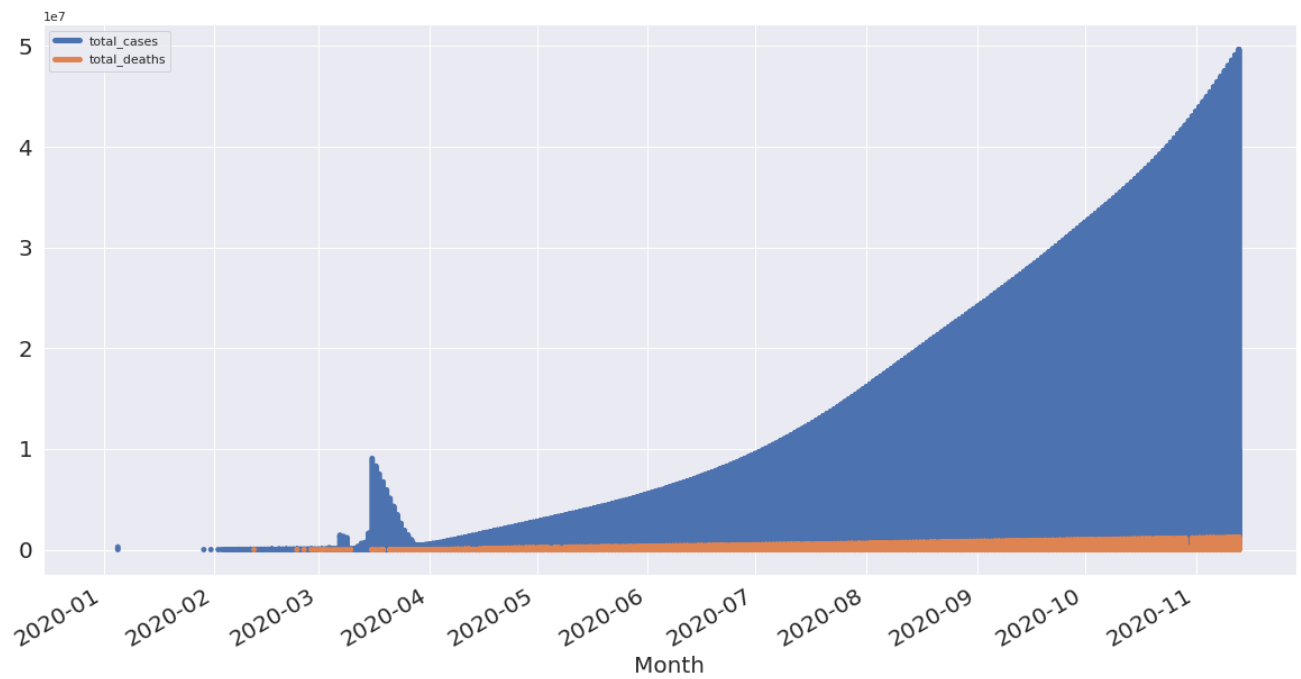
```
Text(0.5, 0, 'Month')
```



```
df_rm = pd.concat([total_cases.rolling(12).mean(), total_deaths.rolling(12).mean()], axis
```

```
df_rm.plot(figsize = (20 , 10), linewidth = 5, fontsize = 20)
```

```
plt.xlabel('Month' , fontsize = 20);
```



▼ Seasonal Patterns in Time Series Data

```
total_cases.diff().plot(figsize=(20,12),linewidth=5,fontsize=20)
plt.xlabel('Month', fontsize = 20)
```

Text(0.5, 0, 'Month')

