covid-19

July 6, 2024

[1]:
```
import pandas as pd
```

[2]:
```
covid = pd.read_csv('covid.csv')
```

[3]:
```
#Having a glance at some of the records
covid.head()
```

[3]:

	iso_code	location	date	total_cases	new_cases	total_deaths	\
0	ABW	Aruba	2020-03-13	2	2	0	
1	ABW	Aruba	2020-03-20	4	2	0	
2	ABW	Aruba	2020-03-24	12	8	0	
3	ABW	Aruba	2020-03-25	17	5	0	
4	ABW	Aruba	2020-03-26	19	2	0	

	new_deaths	total_cases_per_million	new_cases_per_million	\
0	0	18.733	18.733	
1	0	37.465	18.733	
2	0	112.395	74.930	
3	0	159.227	46.831	
4	0	177.959	18.733	

	total_deaths_per_million	…	aged_65_older	aged_70_older	\
0	0.0	…	13.085	7.452	
1	0.0	…	13.085	7.452	
2	0.0	…	13.085	7.452	
3	0.0	…	13.085	7.452	
4	0.0	…	13.085	7.452	

	gdp_per_capita	extreme_poverty	cvd_death_rate	diabetes_prevalence	\
0	35973.781	NaN	NaN	11.62	
1	35973.781	NaN	NaN	11.62	
2	35973.781	NaN	NaN	11.62	
3	35973.781	NaN	NaN	11.62	
4	35973.781	NaN	NaN	11.62	

	female_smokers	male_smokers	handwashing_facilities	hospital_beds_per_100k
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN

```
2             NaN          NaN                    NaN                    NaN
3             NaN          NaN                    NaN                    NaN
4             NaN          NaN                    NaN                    NaN

[5 rows x 32 columns]
```

[4]:
```
#Looking at the shape
covid.shape
```

[4]:
```
(19496, 32)
```

[5]:
```
covid.columns
```

[5]:
```
Index(['iso_code', 'location', 'date', 'total_cases', 'new_cases',
       'total_deaths', 'new_deaths', 'total_cases_per_million',
       'new_cases_per_million', 'total_deaths_per_million',
       'new_deaths_per_million', 'total_tests', 'new_tests',
       'total_tests_per_thousand', 'new_tests_per_thousand',
       'new_tests_smoothed', 'new_tests_smoothed_per_thousand', 'tests_units',
       'stringency_index', 'population', 'population_density', 'median_age',
       'aged_65_older', 'aged_70_older', 'gdp_per_capita', 'extreme_poverty',
       'cvd_death_rate', 'diabetes_prevalence', 'female_smokers',
       'male_smokers', 'handwashing_facilities', 'hospital_beds_per_100k'],
      dtype='object')
```

[6]:
```
#Looking at the different locations
covid["location"].value_counts()
```

[6]:
```
location
Sweden            146
Canada            146
United States     146
Lithuania         146
Brazil            146
                 ...
Yemen              45
Western Sahara     29
Tajikistan         24
Comoros            23
Lesotho            10
Name: count, Length: 212, dtype: int64
```

[7]:
```
#Checking if columns have null values
covid.isna().any()
```

[7]:
```
iso_code                         True
location                        False
date                            False
```

```
total_cases                          False
new_cases                            False
total_deaths                         False
new_deaths                           False
total_cases_per_million               True
new_cases_per_million                 True
total_deaths_per_million              True
new_deaths_per_million                True
total_tests                           True
new_tests                             True
total_tests_per_thousand              True
new_tests_per_thousand                True
new_tests_smoothed                    True
new_tests_smoothed_per_thousand       True
tests_units                           True
stringency_index                      True
population                            True
population_density                    True
median_age                            True
aged_65_older                         True
aged_70_older                         True
gdp_per_capita                        True
extreme_poverty                       True
cvd_death_rate                        True
diabetes_prevalence                   True
female_smokers                        True
male_smokers                          True
handwashing_facilities                True
hospital_beds_per_100k                True
dtype: bool
```

[8]:
```
#Getting the sum of null values across each column
covid.isna().sum()
```

[8]:
```
iso_code                          64
location                           0
date                               0
total_cases                        0
new_cases                          0
total_deaths                       0
new_deaths                         0
total_cases_per_million          377
new_cases_per_million            377
total_deaths_per_million         377
new_deaths_per_million           377
total_tests                    14332
new_tests                      14904
```

```
total_tests_per_thousand           14332
new_tests_per_thousand             14904
new_tests_smoothed                 13866
new_tests_smoothed_per_thousand    13866
tests_units                        13267
stringency_index                    4500
population                            64
population_density                   850
median_age                          1743
aged_65_older                       1980
aged_70_older                       1832
gdp_per_capita                      1982
extreme_poverty                     7878
cvd_death_rate                      1817
diabetes_prevalence                 1174
female_smokers                      5052
male_smokers                        5206
handwashing_facilities             11822
hospital_beds_per_100k              3160
dtype: int64
```

[9]:
```
#Getting the cases in India
india_case=covid[covid["location"]=="India"]
```

[10]:
```
india_case.head()
```

[10]:

	iso_code	location	date	total_cases	new_cases	total_deaths	\
8379	IND	India	2019-12-31	0	0	0	
8380	IND	India	2020-01-01	0	0	0	
8381	IND	India	2020-01-02	0	0	0	
8382	IND	India	2020-01-03	0	0	0	
8383	IND	India	2020-01-04	0	0	0	

	new_deaths	total_cases_per_million	new_cases_per_million	\
8379	0	0.0	0.0	
8380	0	0.0	0.0	
8381	0	0.0	0.0	
8382	0	0.0	0.0	
8383	0	0.0	0.0	

	total_deaths_per_million	…	aged_65_older	aged_70_older	\
8379	0.0	…	5.989	3.414	
8380	0.0	…	5.989	3.414	
8381	0.0	…	5.989	3.414	
8382	0.0	…	5.989	3.414	
8383	0.0	…	5.989	3.414	

```
      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
8379        6426.674             21.2          282.28                10.39
8380        6426.674             21.2          282.28                10.39
8381        6426.674             21.2          282.28                10.39
8382        6426.674             21.2          282.28                10.39
8383        6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8379             1.9         20.6                   59.55
8380             1.9         20.6                   59.55
8381             1.9         20.6                   59.55
8382             1.9         20.6                   59.55
8383             1.9         20.6                   59.55

      hospital_beds_per_100k
8379                    0.53
8380                    0.53
8381                    0.53
8382                    0.53
8383                    0.53

[5 rows x 32 columns]
```

[11]:
```
india_case.tail()
```

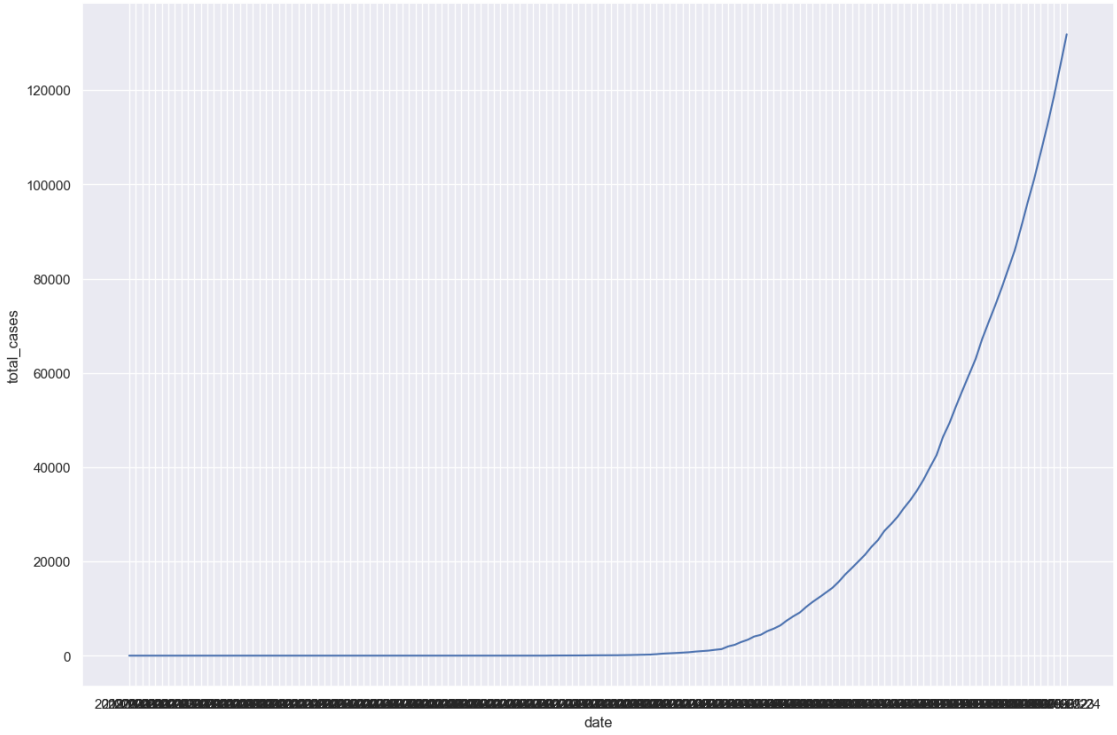
[11]:
```
     iso_code location        date  total_cases  new_cases  total_deaths  \
8519      IND    India  2020-05-20       106750       5611          3303
8520      IND    India  2020-05-21       112359       5609          3435
8521      IND    India  2020-05-22       118447       6088          3583
8522      IND    India  2020-05-23       125101       6654          3720
8523      IND    India  2020-05-24       131868       6767          3867

      new_deaths  total_cases_per_million  new_cases_per_million  \
8519         140                   77.355                  4.066
8520         132                   81.419                  4.064
8521         148                   85.831                  4.412
8522         137                   90.653                  4.822
8523         147                   95.556                  4.904

      total_deaths_per_million  …  aged_65_older  aged_70_older  \
8519                     2.393  …          5.989          3.414
8520                     2.489  …          5.989          3.414
8521                     2.596  …          5.989          3.414
8522                     2.696  …          5.989          3.414
8523                     2.802  …          5.989          3.414

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
```

```
8519        6426.674             21.2          282.28                10.39
8520        6426.674             21.2          282.28                10.39
8521        6426.674             21.2          282.28                10.39
8522        6426.674             21.2          282.28                10.39
8523        6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8519             1.9         20.6                   59.55
8520             1.9         20.6                   59.55
8521             1.9         20.6                   59.55
8522             1.9         20.6                   59.55
8523             1.9         20.6                   59.55

      hospital_beds_per_100k
8519                    0.53
8520                    0.53
8521                    0.53
8522                    0.53
8523                    0.53

[5 rows x 32 columns]
```

[12]:
```
import seaborn as sns
from matplotlib import pyplot as plt
```

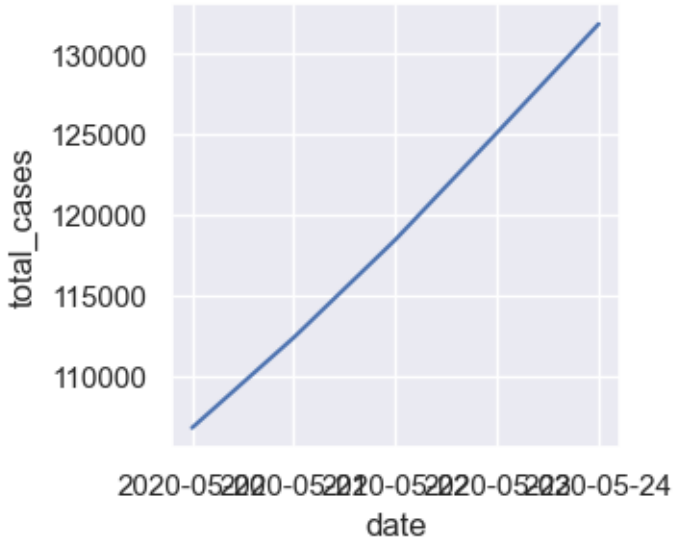
[13]:
```
#Total cases per day
sns.set(rc={'figure.figsize':(15,10)})
sns.lineplot(x="date",y="total_cases",data=india_case)
plt.show()
import warnings
warnings.filterwarnings('ignore')
```

```
D:\SOFTWARES\Anaconda\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
D:\SOFTWARES\Anaconda\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

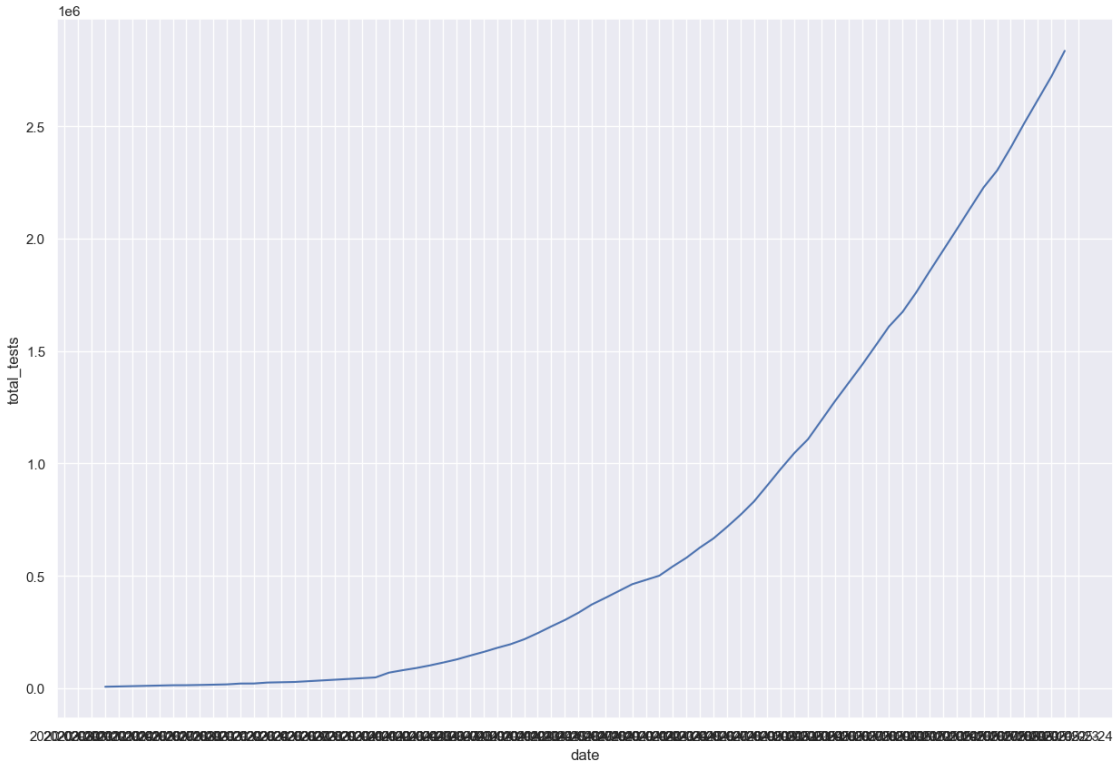


[14]:
```
#Making a dataframe for last 5 days
india_last_5_days=india_case.tail()
```

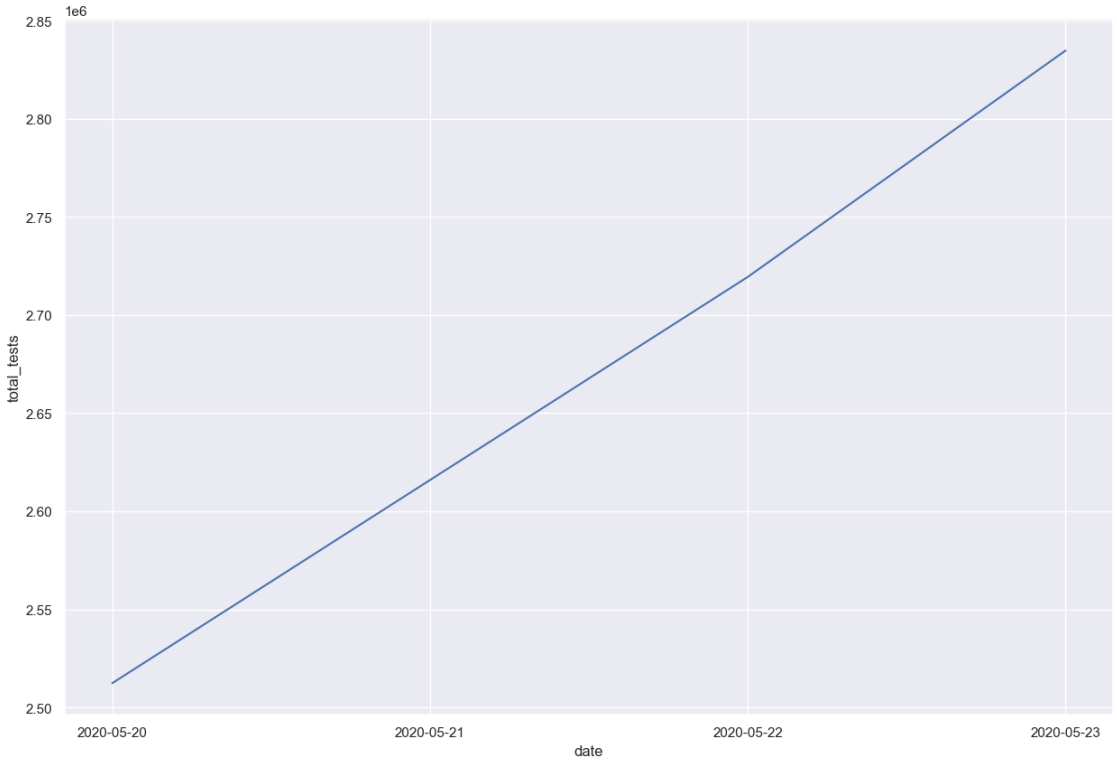
[15]:
```
#Total cases in last 5 days
sns.set(rc={'figure.figsize':(3,3)})
sns.lineplot(x="date",y="total_cases",data=india_last_5_days)
plt.show()
```



[16]:
```
#Total tests per day
sns.set(rc={'figure.figsize':(15,10)})
sns.lineplot(x="date",y="total_tests",data=india_case)
plt.show()
```



[17]:
```
#Total tests in last 5 days
sns.set(rc={'figure.figsize':(15,10)})
sns.lineplot(x="date",y="total_tests",data=india_last_5_days)
plt.show()
```



[18]:
```
#Brazil Case
brazil_case=covid[covid["location"]=="Brazil"]
brazil_case.head()
```

[18]:

	iso_code	location	date	total_cases	new_cases	total_deaths	\
2510	BRA	Brazil	2019-12-31	0	0	0	
2511	BRA	Brazil	2020-01-01	0	0	0	
2512	BRA	Brazil	2020-01-02	0	0	0	
2513	BRA	Brazil	2020-01-03	0	0	0	
2514	BRA	Brazil	2020-01-04	0	0	0	

	new_deaths	total_cases_per_million	new_cases_per_million	\
2510	0	0.0	0.0	
2511	0	0.0	0.0	
2512	0	0.0	0.0	
2513	0	0.0	0.0	
2514	0	0.0	0.0	

	total_deaths_per_million	…	aged_65_older	aged_70_older	\
2510	0.0	…	8.552	5.06	
2511	0.0	…	8.552	5.06	
2512	0.0	…	8.552	5.06	

```
2513                    0.0  …          8.552           5.06
2514                    0.0  …          8.552           5.06

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
2510       14103.452              3.4         177.961                 8.11
2511       14103.452              3.4         177.961                 8.11
2512       14103.452              3.4         177.961                 8.11
2513       14103.452              3.4         177.961                 8.11
2514       14103.452              3.4         177.961                 8.11

      female_smokers male_smokers  handwashing_facilities  \
2510            10.1         17.9                     NaN
2511            10.1         17.9                     NaN
2512            10.1         17.9                     NaN
2513            10.1         17.9                     NaN
2514            10.1         17.9                     NaN

      hospital_beds_per_100k
2510                     2.2
2511                     2.2
2512                     2.2
2513                     2.2
2514                     2.2

[5 rows x 32 columns]
```

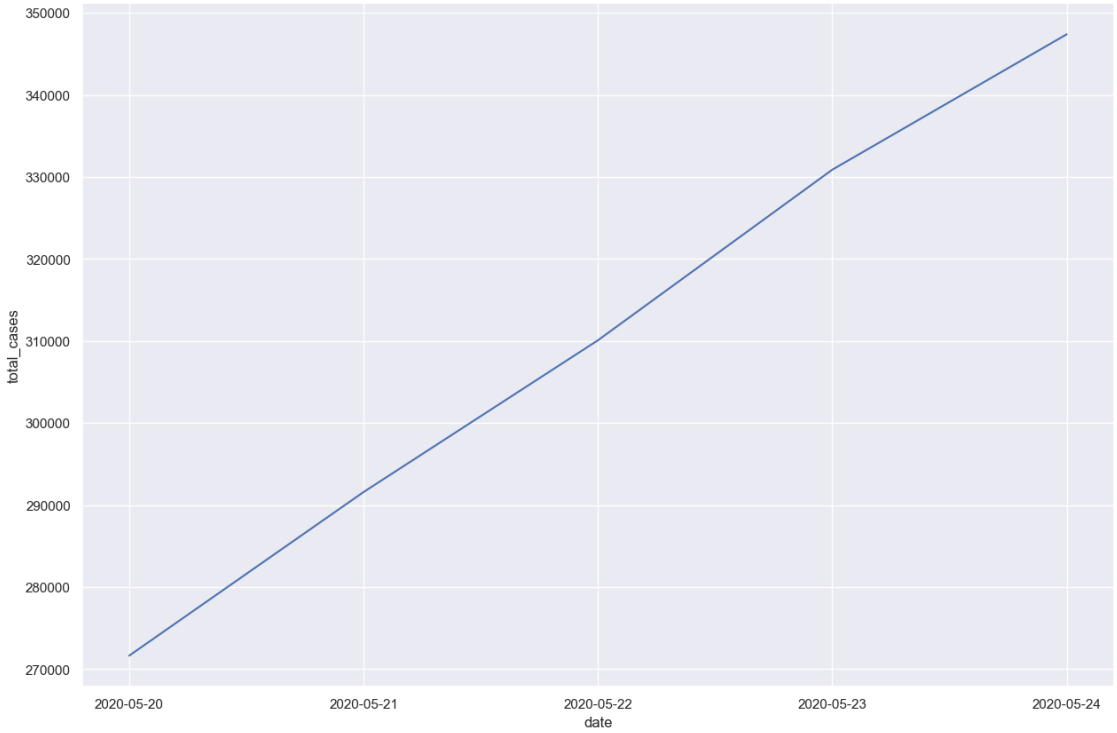
[19]:
```
brazil_case.tail()
```

[19]:
```
     iso_code location        date  total_cases  new_cases  total_deaths  \
2651      BRA   Brazil  2020-05-20       271628      17408         17971
2652      BRA   Brazil  2020-05-21       291579      19951         18859
2653      BRA   Brazil  2020-05-22       310087      18508         20047
2654      BRA   Brazil  2020-05-23       330890      20803         21048
2655      BRA   Brazil  2020-05-24       347398      16508         22013

      new_deaths  total_cases_per_million  new_cases_per_million  \
2651        1179                 1277.892                 81.897
2652         888                 1371.753                 93.861
2653        1188                 1458.825                 87.072
2654        1001                 1556.694                 97.869
2655         965                 1634.357                 77.663

      total_deaths_per_million  …  aged_65_older  aged_70_older  \
2651                    84.546  …          8.552           5.06
2652                    88.723  …          8.552           5.06
2653                    94.312  …          8.552           5.06
2654                    99.022  …          8.552           5.06
```

```
2655                  103.562  …          8.552           5.06

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
2651       14103.452              3.4         177.961                 8.11
2652       14103.452              3.4         177.961                 8.11
2653       14103.452              3.4         177.961                 8.11
2654       14103.452              3.4         177.961                 8.11
2655       14103.452              3.4         177.961                 8.11

      female_smokers male_smokers  handwashing_facilities  \
2651            10.1         17.9                     NaN
2652            10.1         17.9                     NaN
2653            10.1         17.9                     NaN
2654            10.1         17.9                     NaN
2655            10.1         17.9                     NaN

      hospital_beds_per_100k
2651                     2.2
2652                     2.2
2653                     2.2
2654                     2.2
2655                     2.2

[5 rows x 32 columns]
```

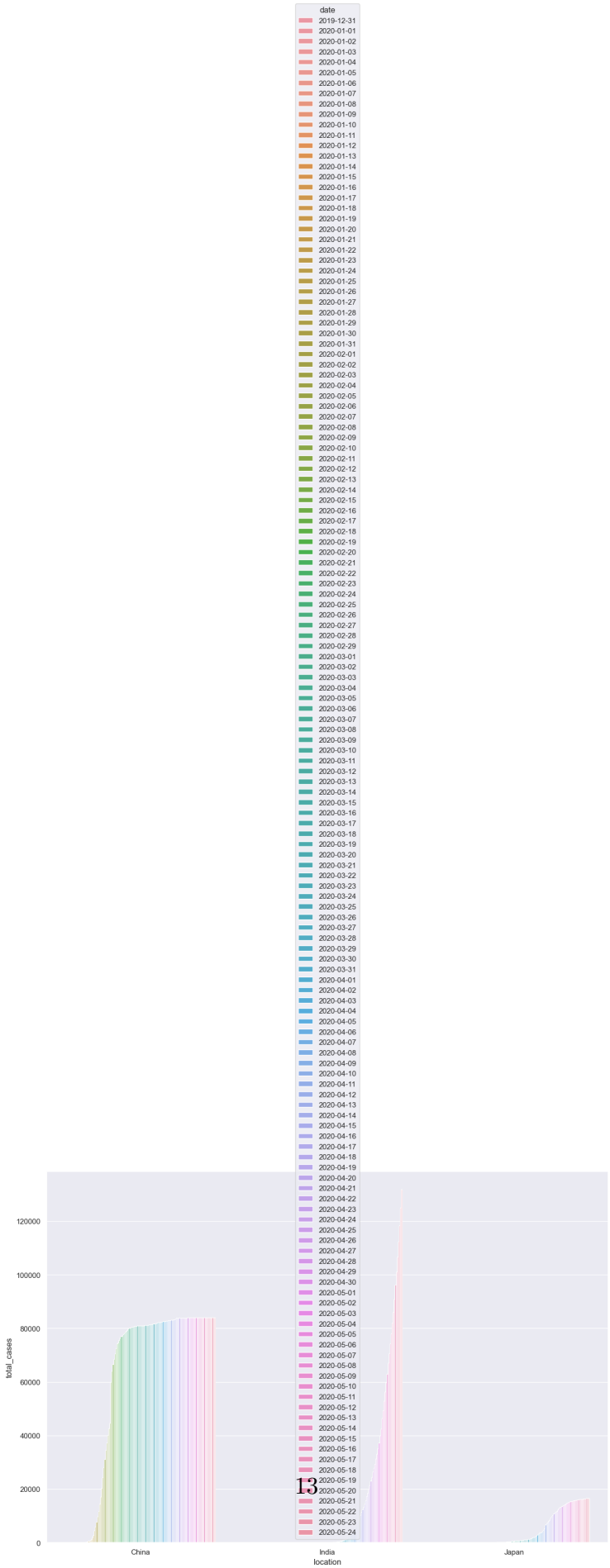
[20]:
```
#Making a dataframe for brazil for last 5 days
brazil_last_5_days=brazil_case.tail()
```

[21]:
```
#Total cases in last 5 days
sns.set(rc={'figure.figsize':(15,10)})
sns.lineplot(x="date",y="total_cases",data=brazil_last_5_days)
plt.show()
```



[22]:
```
#Understanding cases of India, China and Japan
india_japan_china=covid[(covid["location"] =="India") | (covid["location"]␣
↪=="China") | (covid["location"]=="Japan")]
```

[23]:
```
#Plotting growth of cases across China, India and Japan
sns.set(rc={'figure.figsize':(15,10)})
sns.barplot(x="location",y="total_cases",data=india_japan_china,hue="date")
plt.show()
```


date
2019-12-31
2020-01-01
2020-01-02
2020-01-03
2020-01-04
2020-01-05
2020-01-06
2020-01-07
2020-01-08
2020-01-09
2020-01-10
2020-01-11
2020-01-12
2020-01-13
2020-01-14
2020-01-15
2020-01-16
2020-01-17
2020-01-18
2020-01-19
2020-01-20
2020-01-21
2020-01-22
2020-01-23
2020-01-24
2020-01-25
2020-01-26
2020-01-27
2020-01-28
2020-01-29
2020-01-30
2020-01-31
2020-02-01
2020-02-02
2020-02-03
2020-02-04
2020-02-05
2020-02-06
2020-02-07
2020-02-08
2020-02-09
2020-02-10
2020-02-11
2020-02-12
2020-02-13
2020-02-14
2020-02-15
2020-02-16
2020-02-17
2020-02-18
2020-02-19
2020-02-20
2020-02-21
2020-02-22
2020-02-23
2020-02-24
2020-02-25
2020-02-26
2020-02-27
2020-02-28
2020-02-29
2020-03-01
2020-03-02
2020-03-03
2020-03-04
2020-03-05
2020-03-06
2020-03-07
2020-03-08
2020-03-09
2020-03-10
2020-03-11
2020-03-12
2020-03-13
2020-03-14
2020-03-15
2020-03-16
2020-03-17
2020-03-18
2020-03-19
2020-03-20
2020-03-21
2020-03-22
2020-03-23
2020-03-24
2020-03-25
2020-03-26
2020-03-27
2020-03-28
2020-03-29
2020-03-30
2020-03-31
2020-04-01
2020-04-02
2020-04-03
2020-04-04
2020-04-05
2020-04-06
2020-04-07
2020-04-08
2020-04-09
2020-04-10
2020-04-11
2020-04-12
2020-04-13
2020-04-14
2020-04-15
2020-04-16
2020-04-17
2020-04-18
2020-04-19
2020-04-20
2020-04-21
2020-04-22
2020-04-23
2020-04-24
2020-04-25
2020-04-26
2020-04-27
2020-04-28
2020-04-29
2020-04-30
2020-05-01
2020-05-02
2020-05-03
2020-05-04
2020-05-05
2020-05-06
2020-05-07
2020-05-08
2020-05-09
2020-05-10
2020-05-11
2020-05-12
2020-05-13
2020-05-14
2020-05-15
2020-05-16
2020-05-17
2020-05-18
2020-05-19
2020-05-20
2020-05-21
2020-05-22
2020-05-23
2020-05-24
total_cases
0
20000
40000
60000
80000
100000
120000
China
India
Japan
location

[24]:
```
#Understanding cases of germany and spain
germany_spain=covid[(covid["location"] =="Germany") | (covid["location"]␣
 ↪=="Spain")]
```

[25]:
```
#Plotting growth of cases across Germany and Spain
sns.set(rc={'figure.figsize':(15,10)})
sns.barplot(x="location",y="total_cases",data=germany_spain,hue="date")
plt.show()
```


date
2019-12-31
2020-01-01
2020-01-02
2020-01-03
2020-01-04
2020-01-05
2020-01-06
2020-01-07
2020-01-08
2020-01-09
2020-01-10
2020-01-11
2020-01-12
2020-01-13
2020-01-14
2020-01-15
2020-01-16
2020-01-17
2020-01-18
2020-01-19
2020-01-20
2020-01-21
2020-01-22
2020-01-23
2020-01-24
2020-01-25
2020-01-26
2020-01-27
2020-01-28
2020-01-29
2020-01-30
2020-01-31
2020-02-01
2020-02-02
2020-02-03
2020-02-04
2020-02-05
2020-02-06
2020-02-07
2020-02-08
2020-02-09
2020-02-10
2020-02-11
2020-02-12
2020-02-13
2020-02-14
2020-02-15
2020-02-16
2020-02-17
2020-02-18
2020-02-19
2020-02-20
2020-02-21
2020-02-22
2020-02-23
2020-02-24
2020-02-25
2020-02-26
2020-02-27
2020-02-28
2020-02-29
2020-03-01
2020-03-02
2020-03-03
2020-03-04
2020-03-05
2020-03-06
2020-03-07
2020-03-08
2020-03-09
2020-03-10
2020-03-11
2020-03-12
2020-03-13
2020-03-14
2020-03-15
2020-03-16
2020-03-17
2020-03-18
2020-03-19
2020-03-20
2020-03-21
2020-03-22
2020-03-23
2020-03-24
2020-03-25
2020-03-26
2020-03-27
2020-03-28
2020-03-29
2020-03-30
2020-03-31
2020-04-01
2020-04-02
2020-04-03
2020-04-04
2020-04-05
2020-04-06
2020-04-07
2020-04-08
2020-04-09
2020-04-10
2020-04-11
2020-04-12
2020-04-13
2020-04-14
2020-04-15
2020-04-16
2020-04-17
2020-04-18
2020-04-19
2020-04-20
2020-04-21
2020-04-22
2020-04-23
2020-04-24
2020-04-25
2020-04-26
2020-04-27
2020-04-28
2020-04-29
2020-04-30
2020-05-01
2020-05-02
2020-05-03
2020-05-04
2020-05-05
2020-05-06
2020-05-07
2020-05-08
2020-05-09
2020-05-10
2020-05-11
2020-05-12
2020-05-13
2020-05-14
2020-05-15
2020-05-16
2020-05-17
2020-05-18
2020-05-19
2020-05-20
2020-05-21
2020-05-22
2020-05-23
2020-05-24
total_cases
0
50000
100000
150000
200000
Germany
Spain
location

[26]:
```
#Getting latest data
last_day_cases=covid[covid["date"]=="2020-05-24"]
last_day_cases
```

[26]:

```
       iso_code      location        date  total_cases  new_cases  \
62          ABW         Aruba  2020-05-24          101          0
198         AFG   Afghanistan  2020-05-24         9998        782
262         AGO        Angola  2020-05-24           60          0
321         AIA      Anguilla  2020-05-24            3          0
398         ALB       Albania  2020-05-24          989          8
…           …            …         …           …        …
19045       YEM         Yemen  2020-05-24          212          7
19153       ZAF  South Africa  2020-05-24        21343       1218
19220       ZMB        Zambia  2020-05-24          920          0
19285       ZWE      Zimbabwe  2020-05-24           56          0
19431  OWID_WRL         World  2020-05-24      5273572      97636

       total_deaths  new_deaths  total_cases_per_million  \
62                3           0                  945.994
198             216          11                  256.831
262               3           0                    1.826
321               0           0                  199.973
398              31           0                  343.665
…               …           …                        …
19045            39           6                    7.108
19153           407          10                  359.863
19220             7           0                   50.044
19285             4           0                    3.768
19431        341722        3633                  676.550

       new_cases_per_million  total_deaths_per_million  …  aged_65_older  \
62                     0.000                    28.099  …         13.085
198                   20.088                     5.549  …          2.581
262                    0.000                     0.091  …          2.405
321                    0.000                     0.000  …            NaN
398                    2.780                    10.772  …         13.188
…                      …                        …  …             …
19045                  0.235                     1.308  …          2.922
19153                 20.537                     6.862  …          5.344
19220                  0.000                     0.381  …          2.480
19285                  0.000                     0.269  …          2.822
19431                 12.526                    43.840  …          8.696

       aged_70_older  gdp_per_capita  extreme_poverty  cvd_death_rate  \
```

```
62              7.452       35973.781              NaN             NaN
198             1.337        1803.987              NaN         597.029
262             1.362        5819.495              NaN         276.045
321               NaN             NaN              NaN             NaN
398             8.643       11803.431              1.1         304.195
…                 …             …                …               …
19045           1.583        1479.147             18.8         495.003
19153           3.053       12294.876             18.9         200.380
19220           1.542        3689.251             57.5         234.499
19285           1.882        1899.775             21.4         307.846
19431           5.355       15469.207             10.0         233.070

       diabetes_prevalence  female_smokers male_smokers  \
62                   11.62             NaN          NaN
198                   9.59             NaN          NaN
262                   3.94             NaN          NaN
321                    NaN             NaN          NaN
398                  10.08           7.100       51.200
…                    …             …           …
19045                 5.35           7.600       29.200
19153                 5.52           8.100       33.200
19220                 3.94           3.100       24.700
19285                 1.82           1.600       30.700
19431                 8.51           6.434       34.635

       handwashing_facilities  hospital_beds_per_100k
62                        NaN                     NaN
198                    37.746                   0.500
262                    26.664                     NaN
321                       NaN                     NaN
398                       NaN                   2.890
…                          …                   …
19045                  49.542                   0.700
19153                  43.993                   2.320
19220                  13.938                   2.000
19285                  36.791                   1.700
19431                  60.130                   2.705

[207 rows x 32 columns]
```

[27]:
```
#Sorting data w.r.t total_cases
max_cases_country=last_day_cases.sort_values(by="total_cases",ascending=False)
max_cases_country
```

[27]:
```
       iso_code                    location        date  total_cases  \
19431  OWID_WRL                       World  2020-05-24      5273572
18391       USA               United States  2020-05-24      1622670
```

```
2655        BRA                            Brazil  2020-05-24       347398
15569       RUS                            Russia  2020-05-24       335882
9396        ITA                             Italy  2020-05-24       229327
…           …                                 …           …           …
18723       VGB            British Virgin Islands  2020-05-24            8
1645        BES  Bonaire Sint Eustatius and Saba  2020-05-24            6
5543        ESH                    Western Sahara  2020-05-24            6
321         AIA                          Anguilla  2020-05-24            3
11086       LSO                           Lesotho  2020-05-24            2

       new_cases  total_deaths  new_deaths  total_cases_per_million  \
19431      97636        341722        3633                  676.550
18391      21236         97087        1080                 4902.287
2655       16508         22013         965                 1634.357
15569       9434          3388         139                 2301.595
9396         669         32735         119                 3792.922
…            …             …           …                        …
18723          0             1           0                  264.577
1645           0             0           0                  228.824
5543           0             0           0                   10.045
321            0             0           0                  199.973
11086          1             0           0                    0.934

       new_cases_per_million  total_deaths_per_million  …  aged_65_older  \
19431                 12.526                    43.840  …          8.696
18391                 64.157                   293.312  …         15.413
2655                  77.663                   103.562  …          8.552
15569                 64.645                    23.216  …         14.178
9396                  11.065                   541.416  …         23.021
…                        …                         …  …            …
18723                  0.000                    33.072  …            NaN
1645                   0.000                     0.000  …            NaN
5543                   0.000                     0.000  …            NaN
321                    0.000                     0.000  …            NaN
11086                  0.467                     0.000  …          4.506

       aged_70_older  gdp_per_capita  extreme_poverty  cvd_death_rate  \
19431          5.355       15469.207             10.0         233.070
18391          9.732       54225.446              1.2         151.089
2655           5.060       14103.452              3.4         177.961
15569          9.393       24765.954              0.1         431.297
9396          16.240       35220.084              2.0         113.151
…                …             …                …              …
18723            NaN             NaN              NaN             NaN
1645             NaN             NaN              NaN             NaN
5543           1.380             NaN              NaN             NaN
321              NaN             NaN              NaN             NaN
```

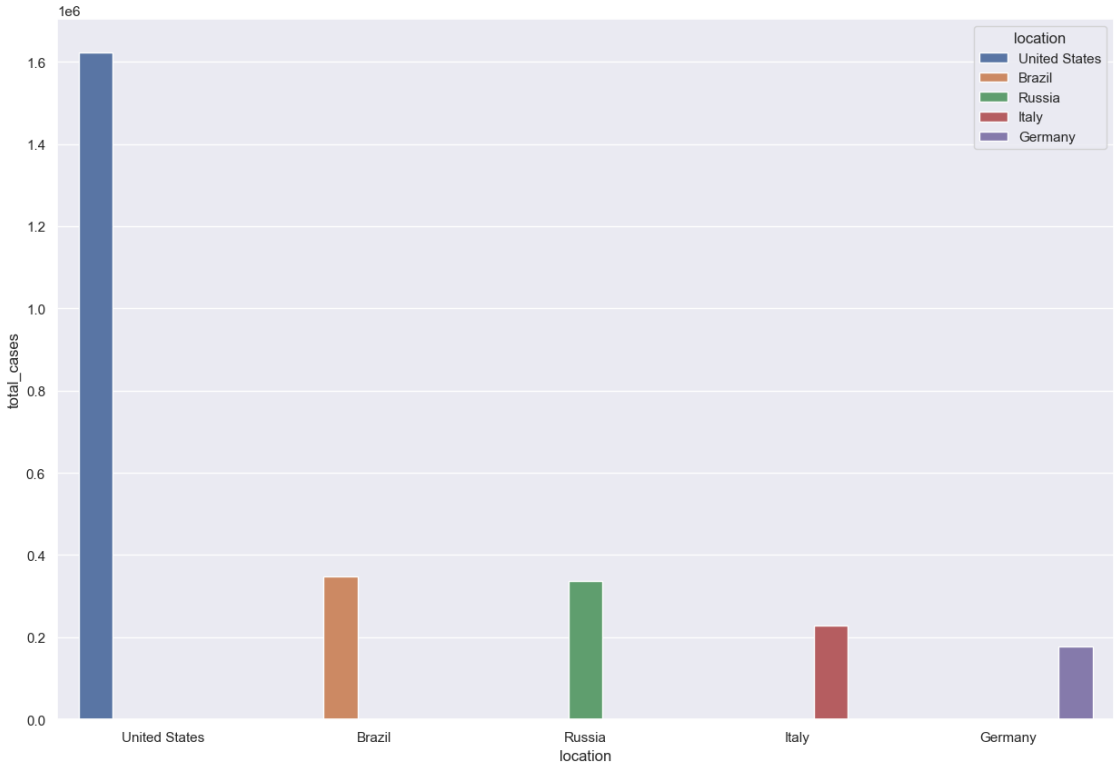
```
11086          2.647        2851.153             59.6         405.126

       diabetes_prevalence  female_smokers male_smokers  \
19431                 8.51           6.434       34.635
18391                10.79          19.100       24.600
2655                  8.11          10.100       17.900
15569                 6.18          23.400       58.300
9396                  4.78          19.800       27.800
…                        …               …            …
18723                13.67             NaN          NaN
1645                   NaN             NaN          NaN
5543                   NaN             NaN          NaN
321                    NaN             NaN          NaN
11086                 3.94           0.400       53.900

       handwashing_facilities  hospital_beds_per_100k
19431                  60.130                   2.705
18391                     NaN                   2.770
2655                      NaN                   2.200
15569                     NaN                   8.050
9396                      NaN                   3.180
…                         …                       …
18723                     NaN                     NaN
1645                      NaN                     NaN
5543                      NaN                     NaN
321                       NaN                     NaN
11086                   2.117                     NaN

[207 rows x 32 columns]
```

[28]:
```
#Top 5 countries with maximum cases
max_cases_country[1:6]
```

[28]:
```
      iso_code       location        date  total_cases  new_cases  \
18391      USA  United States  2020-05-24      1622670      21236
2655       BRA         Brazil  2020-05-24       347398      16508
15569      RUS         Russia  2020-05-24       335882       9434
9396       ITA          Italy  2020-05-24       229327        669
4613       DEU        Germany  2020-05-24       178281        431

       total_deaths  new_deaths  total_cases_per_million  \
18391         97087        1080                 4902.287
2655          22013         965                 1634.357
15569          3388         139                 2301.595
9396          32735         119                 3792.922
4613           8247          31                 2127.866
```

```
       new_cases_per_million  total_deaths_per_million  …  aged_65_older  \
18391                 64.157                   293.312  …         15.413
2655                  77.663                   103.562  …          8.552
15569                 64.645                    23.216  …         14.178
9396                  11.065                   541.416  …         23.021
4613                   5.144                    98.432  …         21.453

       aged_70_older  gdp_per_capita  extreme_poverty  cvd_death_rate  \
18391          9.732       54225.446              1.2         151.089
2655           5.060       14103.452              3.4         177.961
15569          9.393       24765.954              0.1         431.297
9396          16.240       35220.084              2.0         113.151
4613          15.957       45229.245              NaN         156.139

       diabetes_prevalence  female_smokers male_smokers  \
18391                10.79            19.1         24.6
2655                  8.11            10.1         17.9
15569                 6.18            23.4         58.3
9396                  4.78            19.8         27.8
4613                  8.31            28.2         33.1

       handwashing_facilities  hospital_beds_per_100k
18391                     NaN                    2.77
2655                      NaN                    2.20
15569                     NaN                    8.05
9396                      NaN                    3.18
4613                      NaN                    8.00

[5 rows x 32 columns]
```

[29]:
```
#Making bar-plot for countries with top cases
sns.barplot(x="location",y="total_cases",data=max_cases_country[1:
 ↪6],hue="location")
plt.show()
```



[30]:
```
india_case.head()
```

[30]:

	iso_code	location	date	total_cases	new_cases	total_deaths	\
8379	IND	India	2019-12-31	0	0	0	
8380	IND	India	2020-01-01	0	0	0	
8381	IND	India	2020-01-02	0	0	0	
8382	IND	India	2020-01-03	0	0	0	
8383	IND	India	2020-01-04	0	0	0	

	new_deaths	total_cases_per_million	new_cases_per_million	\
8379	0	0.0	0.0	
8380	0	0.0	0.0	
8381	0	0.0	0.0	
8382	0	0.0	0.0	
8383	0	0.0	0.0	

	total_deaths_per_million	…	aged_65_older	aged_70_older	\
8379	0.0	…	5.989	3.414	
8380	0.0	…	5.989	3.414	
8381	0.0	…	5.989	3.414	
8382	0.0	…	5.989	3.414	
8383	0.0	…	5.989	3.414	

```
      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
8379         6426.674             21.2          282.28                10.39
8380         6426.674             21.2          282.28                10.39
8381         6426.674             21.2          282.28                10.39
8382         6426.674             21.2          282.28                10.39
8383         6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8379             1.9         20.6                   59.55
8380             1.9         20.6                   59.55
8381             1.9         20.6                   59.55
8382             1.9         20.6                   59.55
8383             1.9         20.6                   59.55

      hospital_beds_per_100k
8379                    0.53
8380                    0.53
8381                    0.53
8382                    0.53
8383                    0.53

[5 rows x 32 columns]
```

[31]:
```
#Linear regression
from sklearn.model_selection import train_test_split
```

[32]:
```
#converting string date to date-time
import datetime as dt
india_case['date'] = pd.to_datetime(india_case['date'])
india_case.head()
```

[32]:
```
     iso_code location       date  total_cases  new_cases  total_deaths  \
8379      IND    India 2019-12-31            0          0             0
8380      IND    India 2020-01-01            0          0             0
8381      IND    India 2020-01-02            0          0             0
8382      IND    India 2020-01-03            0          0             0
8383      IND    India 2020-01-04            0          0             0

      new_deaths  total_cases_per_million  new_cases_per_million  \
8379           0                      0.0                    0.0
8380           0                      0.0                    0.0
8381           0                      0.0                    0.0
8382           0                      0.0                    0.0
8383           0                      0.0                    0.0

      total_deaths_per_million  …  aged_65_older  aged_70_older  \
```

```
8379                    0.0  …          5.989          3.414
8380                    0.0  …          5.989          3.414
8381                    0.0  …          5.989          3.414
8382                    0.0  …          5.989          3.414
8383                    0.0  …          5.989          3.414

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
8379        6426.674             21.2          282.28                10.39
8380        6426.674             21.2          282.28                10.39
8381        6426.674             21.2          282.28                10.39
8382        6426.674             21.2          282.28                10.39
8383        6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8379             1.9         20.6                   59.55
8380             1.9         20.6                   59.55
8381             1.9         20.6                   59.55
8382             1.9         20.6                   59.55
8383             1.9         20.6                   59.55

      hospital_beds_per_100k
8379                    0.53
8380                    0.53
8381                    0.53
8382                    0.53
8383                    0.53

[5 rows x 32 columns]
```

[33]:
```
india_case.head()
```

[33]:
```
     iso_code location       date  total_cases  new_cases  total_deaths  \
8379      IND    India 2019-12-31            0          0             0
8380      IND    India 2020-01-01            0          0             0
8381      IND    India 2020-01-02            0          0             0
8382      IND    India 2020-01-03            0          0             0
8383      IND    India 2020-01-04            0          0             0

      new_deaths  total_cases_per_million  new_cases_per_million  \
8379           0                      0.0                    0.0
8380           0                      0.0                    0.0
8381           0                      0.0                    0.0
8382           0                      0.0                    0.0
8383           0                      0.0                    0.0

      total_deaths_per_million  …  aged_65_older  aged_70_older  \
8379                       0.0  …          5.989          3.414
```

```
8380                    0.0  …          5.989          3.414
8381                    0.0  …          5.989          3.414
8382                    0.0  …          5.989          3.414
8383                    0.0  …          5.989          3.414

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
8379        6426.674             21.2          282.28                10.39
8380        6426.674             21.2          282.28                10.39
8381        6426.674             21.2          282.28                10.39
8382        6426.674             21.2          282.28                10.39
8383        6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8379             1.9         20.6                   59.55
8380             1.9         20.6                   59.55
8381             1.9         20.6                   59.55
8382             1.9         20.6                   59.55
8383             1.9         20.6                   59.55

      hospital_beds_per_100k
8379                    0.53
8380                    0.53
8381                    0.53
8382                    0.53
8383                    0.53

[5 rows x 32 columns]
```

[34]:
```
#converting date-time to ordinal
india_case['date']=india_case['date'].map(dt.datetime.toordinal)
india_case.head()
```

[34]:
```
     iso_code location    date  total_cases  new_cases  total_deaths  \
8379      IND    India  737424            0          0             0
8380      IND    India  737425            0          0             0
8381      IND    India  737426            0          0             0
8382      IND    India  737427            0          0             0
8383      IND    India  737428            0          0             0

      new_deaths  total_cases_per_million  new_cases_per_million  \
8379           0                      0.0                    0.0
8380           0                      0.0                    0.0
8381           0                      0.0                    0.0
8382           0                      0.0                    0.0
8383           0                      0.0                    0.0

      total_deaths_per_million  …  aged_65_older  aged_70_older  \
```

```
8379                    0.0  …          5.989          3.414
8380                    0.0  …          5.989          3.414
8381                    0.0  …          5.989          3.414
8382                    0.0  …          5.989          3.414
8383                    0.0  …          5.989          3.414

      gdp_per_capita  extreme_poverty  cvd_death_rate  diabetes_prevalence  \
8379        6426.674             21.2          282.28                10.39
8380        6426.674             21.2          282.28                10.39
8381        6426.674             21.2          282.28                10.39
8382        6426.674             21.2          282.28                10.39
8383        6426.674             21.2          282.28                10.39

      female_smokers male_smokers  handwashing_facilities  \
8379             1.9         20.6                   59.55
8380             1.9         20.6                   59.55
8381             1.9         20.6                   59.55
8382             1.9         20.6                   59.55
8383             1.9         20.6                   59.55

      hospital_beds_per_100k
8379                    0.53
8380                    0.53
8381                    0.53
8382                    0.53
8383                    0.53

[5 rows x 32 columns]
```

[35]:
```
#getting dependent variable and inpedent variable
x=india_case['date']
y=india_case['total_cases']
```

[36]:
```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3)
```

[37]:
```
from sklearn.linear_model import LinearRegression
```

[38]:
```
lr = LinearRegression()
```

[39]:
```
import numpy as np
lr.fit(np.array(x_train).reshape(-1,1),np.array(y_train).reshape(-1,1))
```

[39]:
```
LinearRegression()
```

[40]:
```
india_case.tail()
```

[40]:

	iso_code	location	date	total_cases	new_cases	total_deaths	\
8519	IND	India	737565	106750	5611	3303	
8520	IND	India	737566	112359	5609	3435	
8521	IND	India	737567	118447	6088	3583	
8522	IND	India	737568	125101	6654	3720	
8523	IND	India	737569	131868	6767	3867	

	new_deaths	total_cases_per_million	new_cases_per_million	\
8519	140	77.355	4.066	
8520	132	81.419	4.064	
8521	148	85.831	4.412	
8522	137	90.653	4.822	
8523	147	95.556	4.904	

	total_deaths_per_million	…	aged_65_older	aged_70_older	\
8519	2.393	…	5.989	3.414	
8520	2.489	…	5.989	3.414	
8521	2.596	…	5.989	3.414	
8522	2.696	…	5.989	3.414	
8523	2.802	…	5.989	3.414	

	gdp_per_capita	extreme_poverty	cvd_death_rate	diabetes_prevalence	\
8519	6426.674	21.2	282.28	10.39	
8520	6426.674	21.2	282.28	10.39	
8521	6426.674	21.2	282.28	10.39	
8522	6426.674	21.2	282.28	10.39	
8523	6426.674	21.2	282.28	10.39	

	female_smokers	male_smokers	handwashing_facilities	\
8519	1.9	20.6	59.55	
8520	1.9	20.6	59.55	
8521	1.9	20.6	59.55	
8522	1.9	20.6	59.55	
8523	1.9	20.6	59.55	

	hospital_beds_per_100k
8519	0.53
8520	0.53
8521	0.53
8522	0.53
8523	0.53

[5 rows x 32 columns]

[41]:
```
y_pred=lr.predict(np.array(x_test).reshape(-1,1))
```

[42]:
```
from sklearn.metrics import mean_squared_error
```

[43]:
```
mean_squared_error(x_test,y_pred)
```

[43]:
```
517987319887.33966
```

[44]:
```
lr.predict(np.array([[737573]]))
```

[44]:
```
array([[54969.10209972]])
```

[ ]:
```

```