covid_impact_analysis

April 24, 2023

0.0.1 The project is mostly about the impact of covid in different sectors. Here we also have discussed the covid cases and death percentage due to the pendemic. We have two dataset that contains crucial information regarding various issues and factors affected during covid-19. Our goal is to find out the most affected countries, death rate, the impact of covid on the economy and others.

[2]:
```
#Import of libraries
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go
```

[3]:
```
#Data import
df = pd.read_csv('covid_transformed _data.csv')
df_2 = pd.read_csv('covid_raw_data.csv')
```

0.0.2 Dataset inspection

[4]:
```
# lets have a look at the first 5 rows of transformed data
df.head()
```

[4]:

	CODE	COUNTRY	DATE	HDI	TC	TD	STI	POP	GDPCAP
0	AFG	Afghanistan	2019-12-31	0.498	0.0	0.0	0.0	17.477233	7.497754
1	AFG	Afghanistan	2020-01-01	0.498	0.0	0.0	0.0	17.477233	7.497754
2	AFG	Afghanistan	2020-01-02	0.498	0.0	0.0	0.0	17.477233	7.497754
3	AFG	Afghanistan	2020-01-03	0.498	0.0	0.0	0.0	17.477233	7.497754
4	AFG	Afghanistan	2020-01-04	0.498	0.0	0.0	0.0	17.477233	7.497754

[5]:
```
# The last 5 rows
df.tail()
```

[5]:

	CODE	COUNTRY	DATE	HDI	TC	TD	STI	\
50413	ZWE	Zimbabwe	2020-10-15	0.535	8.994048	5.442418	4.341855	
50414	ZWE	Zimbabwe	2020-10-16	0.535	8.996528	5.442418	4.341855	
50415	ZWE	Zimbabwe	2020-10-17	0.535	8.999496	5.442418	4.341855	
50416	ZWE	Zimbabwe	2020-10-18	0.535	9.000853	5.442418	4.341855	
50417	ZWE	Zimbabwe	2020-10-19	0.535	9.005405	5.442418	4.341855	

	POP	GDPCAP

```
50413  16.514381  7.549491
50414  16.514381  7.549491
50415  16.514381  7.549491
50416  16.514381  7.549491
50417  16.514381  7.549491
```

[6]:
```
# Overall view of our dataset columns and datatype and checking the null values
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50418 entries, 0 to 50417
Data columns (total 9 columns):
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   CODE     50418 non-null  object
 1   COUNTRY  50418 non-null  object
 2   DATE     50418 non-null  object
 3   HDI      44216 non-null  float64
 4   TC       50418 non-null  float64
 5   TD       50418 non-null  float64
 6   STI      50418 non-null  float64
 7   POP      50418 non-null  float64
 8   GDPCAP   50418 non-null  float64
dtypes: float64(6), object(3)
memory usage: 3.5+ MB
```

[7]:
```
df['COUNTRY'].value_counts()
```

[7]:
```
Afghanistan        294
Indonesia          294
Macedonia          294
Luxembourg         294
Lithuania          294
                  …
Tajikistan         172
Comoros            171
Lesotho            158
Hong Kong           51
Solomon Islands      4
Name: COUNTRY, Length: 210, dtype: int64
```

[8]:
```
df['COUNTRY'].value_counts().mode()
```

[8]:
```
0    294
dtype: int64
```

0.0.3 Lets also have a look the raw dataset (df_2)

[9]:
```
#Fisrt 5 rows
df_2.head()
```

[9]:
```
  iso_code     location        date  total_cases  total_deaths  \
0      AFG  Afghanistan  2019-12-31          0.0           0.0
1      AFG  Afghanistan  2020-01-01          0.0           0.0
2      AFG  Afghanistan  2020-01-02          0.0           0.0
3      AFG  Afghanistan  2020-01-03          0.0           0.0
4      AFG  Afghanistan  2020-01-04          0.0           0.0

   stringency_index  population  gdp_per_capita  human_development_index  \
0               0.0    38928341        1803.987                    0.498
1               0.0    38928341        1803.987                    0.498
2               0.0    38928341        1803.987                    0.498
3               0.0    38928341        1803.987                    0.498
4               0.0    38928341        1803.987                    0.498

  Unnamed: 9 Unnamed: 10 Unnamed: 11  Unnamed: 12  Unnamed: 13
0      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
1      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
2      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
3      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
4      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
```

[10]:
```
#last 5 rows
df_2.tail()
```

[10]:
```
      iso_code  location        date  total_cases  total_deaths  \
50413      ZWE  Zimbabwe  2020-10-15       8055.0         231.0
50414      ZWE  Zimbabwe  2020-10-16       8075.0         231.0
50415      ZWE  Zimbabwe  2020-10-17       8099.0         231.0
50416      ZWE  Zimbabwe  2020-10-18       8110.0         231.0
50417      ZWE  Zimbabwe  2020-10-19       8147.0         231.0

       stringency_index  population  gdp_per_capita  human_development_index  \
50413             76.85    14862927        1899.775                    0.535
50414             76.85    14862927        1899.775                    0.535
50415             76.85    14862927        1899.775                    0.535
50416             76.85    14862927        1899.775                    0.535
50417             76.85    14862927        1899.775                    0.535

        Unnamed: 9  Unnamed: 10 Unnamed: 11  Unnamed: 12  Unnamed: 13
50413  8.994048296  5.442417711  4.34185547    16.514381  7.549490737
50414  8.996528148  5.442417711  4.34185547    16.514381  7.549490737
50415  8.999495876  5.442417711  4.34185547    16.514381  7.549490737
50416  9.000853147  5.442417711  4.34185547    16.514381  7.549490737
```

```
50417   9.00540504  5.442417711  4.34185547    16.514381  7.549490737
```

[11]:
```
df_2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50418 entries, 0 to 50417
Data columns (total 14 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   iso_code                 50418 non-null  object
 1   location                 50418 non-null  object
 2   date                     50418 non-null  object
 3   total_cases              47324 non-null  float64
 4   total_deaths             39228 non-null  float64
 5   stringency_index         43292 non-null  float64
 6   population               50418 non-null  int64
 7   gdp_per_capita           44706 non-null  float64
 8   human_development_index  44216 non-null  float64
 9   Unnamed: 9               50418 non-null  object
 10  Unnamed: 10              50418 non-null  object
 11  Unnamed: 11              50418 non-null  object
 12  Unnamed: 12              50418 non-null  float64
 13  Unnamed: 13              50418 non-null  object
dtypes: float64(6), int64(1), object(7)
memory usage: 5.4+ MB
```

[12]:
```
df_2['location'].value_counts().mode()
```

[12]:
```
0    294
dtype: int64
```

After having an inspection of both the dataset we have found out that both dataset contains no null values and both of them contains data from 2019-12-31 to 2020-10-19. However, the county or location does not contain equal number of samples in each location/country. Hence we decided to merge them by mode value(294).

0.0.4 Aggregating the datasets

[13]:
```
#Here we are merging both the datasets
code = df["CODE"].unique().tolist()
country = df["COUNTRY"].unique().tolist()
hdi = []
tc = []
td = []
sti = []
population = df["POP"].unique().tolist()
gdp = []
```

```
for i in country:
    hdi.append((df.loc[df["COUNTRY"] == i, "HDI"]).sum()/294)
    tc.append((df_2.loc[df_2["location"] == i, "total_cases"]).sum())
    td.append((df_2.loc[df_2["location"] == i, "total_deaths"]).sum())
    sti.append((df.loc[df["COUNTRY"] == i, "STI"]).sum()/294)
    population.append((df_2.loc[df_2["location"] == i, "population"]).sum()/294)

merged_data = pd.DataFrame(list(zip(code, country, hdi, tc, td, sti,␣
 ↪population)),
                                columns = ["Country Code", "Country", "HDI",
                                           "Total Cases", "Total Deaths",
                                           "Stringency Index", "Population"])
print(merged_data.head())
```

```
  Country Code      Country       HDI  Total Cases  Total Deaths  \
0          AFG  Afghanistan  0.498000    5126433.0      165875.0
1          ALB      Albania  0.600765    1071951.0       31056.0
2          DZA      Algeria  0.754000    4893999.0      206429.0
3          AND      Andorra  0.659551     223576.0        9850.0
4          AGO       Angola  0.418952     304005.0       11820.0

   Stringency Index  Population
0          3.049673   17.477233
1          3.005624   14.872537
2          3.195168   17.596309
3          2.677654   11.254996
4          2.965560   17.307957
```

[14]:
```
df = merged_data.sort_values(by=['Total Cases'], ascending = False)
df.head()
```

[14]:
```
     Country Code        Country      HDI  Total Cases  Total Deaths  \
200          USA  United States  0.92400  746014098.0    26477574.0
27           BRA         Brazil  0.75900  425704517.0    14340567.0
90           IND          India  0.64000  407771615.0     7247327.0
157          RUS         Russia  0.81600  132888951.0     2131571.0
150          PER           Peru  0.59949   74882695.0     3020038.0

     Stringency Index  Population
200          3.350949   19.617637
27           3.136028   19.174732
90           3.610552   21.045353
157          3.380088   18.798668
150          3.430126   17.311165
```

[15]:
```
# Top 10 countries with highest number of cases
df.head(10)
```

[15]:

	Country Code	Country	HDI	Total Cases	Total Deaths	Stringency Index	Population
200	USA	United States	0.924000	746014098.0	26477574.0	3.350949	19.617637
27	BRA	Brazil	0.759000	425704517.0	14340567.0	3.136028	19.174732
90	IND	India	0.640000	407771615.0	7247327.0	3.610552	21.045353
157	RUS	Russia	0.816000	132888951.0	2131571.0	3.380088	18.798668
150	PER	Peru	0.599490	74882695.0	3020038.0	3.430126	17.311165
125	MEX	Mexico	0.774000	74347548.0	7295850.0	3.019289	18.674802
178	ESP	Spain	0.887969	73717676.0	5510624.0	3.393922	17.660427
175	ZAF	South Africa	0.608653	63027659.0	1357682.0	3.364333	17.898266
42	COL	Colombia	0.581847	60543682.0	1936134.0	3.357923	17.745037
199	GBR	United Kingdom	0.922000	59475032.0	7249573.0	3.353883	18.033340

I am considering top 10 highest countries due to the lack of sufficient or meaningful data of GDP of each country. So, to understand the economical impact of covid, it is a better idea to consider the highest affected countries.

[16]:
```
# Here two more columns are added ('GDPCPA before and After Covid')
```

[17]:
```
df=df.head(10)
df
```

[17]:

	Country Code	Country	HDI	Total Cases	Total Deaths	Stringency Index	Population
200	USA	United States	0.924000	746014098.0	26477574.0	3.350949	19.617637
27	BRA	Brazil	0.759000	425704517.0	14340567.0	3.136028	19.174732
90	IND	India	0.640000	407771615.0	7247327.0	3.610552	21.045353
157	RUS	Russia	0.816000	132888951.0	2131571.0	3.380088	18.798668
150	PER	Peru	0.599490	74882695.0	3020038.0	3.430126	17.311165
125	MEX	Mexico	0.774000	74347548.0	7295850.0	3.019289	18.674802
178	ESP	Spain	0.887969	73717676.0	5510624.0	3.393922	17.660427
175	ZAF	South Africa	0.608653	63027659.0	1357682.0	3.364333	17.898266
42	COL	Colombia	0.581847	60543682.0	1936134.0	3.357923	17.745037
199	GBR	United Kingdom	0.922000	59475032.0	7249573.0	3.353883	18.033340

```
90           3.610552   21.045353
157          3.380088   18.798668
150          3.430126   17.311165
125          3.019289   18.674802
178          3.393922   17.660427
175          3.364333   17.898266
42           3.357923   17.745037
199          3.353883   18.033340
```

[18]:
```
df["GDP Before Covid"] = [65279.53, 8897.49, 2100.75,
                          11497.65, 7027.61, 9946.03,
                          29564.74, 6001.40, 6424.98, 42354.41]
df["GDP During Covid"] = [63543.58, 6796.84, 1900.71,
                          10126.72, 6126.87, 8346.70,
                          27057.16, 5090.72, 5332.77, 40284.64]
df
```

[18]:
```
    Country Code         Country       HDI  Total Cases  Total Deaths  \
200          USA   United States  0.924000  746014098.0    26477574.0
27           BRA          Brazil  0.759000  425704517.0    14340567.0
90           IND           India  0.640000  407771615.0     7247327.0
157          RUS          Russia  0.816000  132888951.0     2131571.0
150          PER            Peru  0.599490   74882695.0     3020038.0
125          MEX          Mexico  0.774000   74347548.0     7295850.0
178          ESP           Spain  0.887969   73717676.0     5510624.0
175          ZAF    South Africa  0.608653   63027659.0     1357682.0
42           COL        Colombia  0.581847   60543682.0     1936134.0
199          GBR  United Kingdom  0.922000   59475032.0     7249573.0

     Stringency Index  Population  GDP Before Covid  GDP During Covid
200          3.350949   19.617637          65279.53          63543.58
27           3.136028   19.174732           8897.49           6796.84
90           3.610552   21.045353           2100.75           1900.71
157          3.380088   18.798668          11497.65          10126.72
150          3.430126   17.311165           7027.61           6126.87
125          3.019289   18.674802           9946.03           8346.70
178          3.393922   17.660427          29564.74          27057.16
175          3.364333   17.898266           6001.40           5090.72
42           3.357923   17.745037           6424.98           5332.77
199          3.353883   18.033340          42354.41          40284.64
```

[19]:
```
# Countries with highest number of covid cases
fig_covid_cases = px.bar(df, x='Country', y = 'Total Cases', title = 'Countries␣
 ↪with highest Covid cases')
fig_covid_cases.show()
```

US, Brazil, India are amongst the highest affected countries while UK, Colombia, SA, Mexico and Peru are comparatively similar affected areas and Russia is on the 4th position.

[20]:
```
# Countries with highest death cases
fig_death = px.bar(df, y = 'Total Deaths', x='Country', title = 'Countries with␣
 ↪highest death')
fig_death
```

In the case of death rate US, Brazil and India remains in the same place. However the death rate in Mexico and UK is higher then the other 5 most affected countries.

[21]:
```
#Comparison of total deaths and cases
fig = go.Figure()
fig.add_trace(go.Bar(
    x=df["Country"],
    y=df["Total Cases"],
    name='Total Cases',
    marker_color='blue'
))
fig.add_trace(go.Bar(
    x=df["Country"],
    y=df["Total Deaths"],
    name='Total Deaths',
    marker_color='red'
))
fig.update_layout(barmode='group', xaxis_tickangle=-45)
fig.show()
```

As we can see the death rate is quite higher in terms of covid cases in UK and US and Brazil. Whereas in India the death rate is comparatively lower than the cases.

[22]:
```
#Lets have a look at the death rate
death_rate = (df["Total Deaths"].sum() / df["Total Cases"].sum()) * 100
print("Death Rate = ", death_rate)
```

```
Death Rate =  3.6144212045653767
```

Around 3.6% people died

[23]:
```
# Percentage of Total Cases and Deaths
cases = df["Total Cases"].sum()
deceased = df["Total Deaths"].sum()

labels = ["Total Cases", "Total Deaths"]
values = [cases, deceased]

fig = px.pie(df, values=values, names=labels,
             title='Ratio of Total Cases and Deaths (%)', hole =.2)
fig.show()
```

[24]:
```
#Stringency Index
fig = px.bar(df, x='Country', y='Total Cases',
```

```
             hover_data=['Population', 'Total Deaths'],
             color='Stringency Index', height=400,
             title= "Stringency Index in different countries during covid")
fig.show()
```

India is doing very well, meaning that there are strict regulations (assumption) to control the covid situtaion while Mexico is performing the lowest

Lets inspect the economical impact of covid. We assume the GDP per capita as the initial factor for analyzing the economic decline due to covid.

[25]:
```
# Now lets us examine the GDP before and after covid
fig = px.bar(df, x='Country', y='Total Cases',
             hover_data=['Population', 'Total Deaths'],
             color='GDP Before Covid', height=500,
             title="GDP Per Capita Before Covid-19")
fig.show()
```

[26]:
```
fig = px.bar(df, x='Country', y='Total Cases',
             hover_data=['Population', 'Total Deaths'],
             color='GDP During Covid', height=500,
             title="GDP Per Capita During Covid-19")
fig.show()
```

Lets compare the GDP in these two categories

[27]:
```
#Comparison of GDP before and after covid
fig = go.Figure()
fig.add_trace(go.Bar(
    x=df["Country"],
    y=df["GDP Before Covid"],
    name='GDP Per Capita Before Covid-19',
    marker_color='blue'
))
fig.add_trace(go.Bar(
    x=df["Country"],
    y=df["GDP During Covid"],
    name='GDP Per Capita During Covid-19',
    marker_color='red'
))
fig.update_layout(barmode='group', xaxis_tickangle=-50)
fig.show()
```

The most noticable issue here is that the GDP decrease in all countries during covid-19

Another important economic performance indicator is Human Development Index that refers to the composite index of life expectancy, education, and other issues. Now, we are going to have a look at the countries who spent more or less on this sector during covid.

[28]:
```
#Spending of countries in HDI sector
fig = px.bar(df, x='Country', y='Total Cases',
             hover_data=['Population', 'Total Deaths'],
             color='HDI', height=400,
             title="Human Development Index during Covid-19")
fig.show()
```

As we can see that the USA spent highest on HDI during covid, while other countries spent less in this sector.

0.0.5 Findings:

After analyzing both the dataset we can conclude the following - The datasets contains no null values and the datatypes are correctly done. - From an initial observation we saw that USA has the highest covid cases and death rate compared to Brazil and India ranked imediately after the US. It can be assume that this might happen due to low stringency. - In case of Death rate, US having the highest number of death. Whereas, India, Russia, SA having comparatively low against the total number of covid cases. - Around 3.5% people died among the affected people, whereas 96.5% survived. - Covid impacted the economy is vast way, the GDP per capita notibely declined during covid in all the top 10 countries. - Overall we can say that covid had a serious impact in most of the sectors in all the countries.