

# COVID DATA ANALYSIS

By: Omar Nasr

A Comprehensive Overview of Covid-19 Project

# PROJECT OVERVIEW

- The Dataset Is about Covid-19.The Analysis Process Was Done On 190 Country In All Continent.And It's In Time Range between 1-1-2020 and 30-4-2021
- **Objective:** "Analyzing COVID-19 Information, Deaths, And Vaccinations To Extract Insights And Provide Recommendations For Public Health Policies.

# DATA SOURCES

- **First: Covid Deaths** Contain Information About Total Deaths, Total Cases, New Deaths, New Cases And All Of This Per Million.
- **Second: Covid Vaccinations** Contain Information About Total And New Tests, Total Vaccination, People Vaccinated , Test Per Case And So on.
- **Third: Covid Information** This Table I Created To Make Understand Data Easily All Info Related To Population Like Different Diseases, Economic And Health Situation.
- Finally All Data Sources Cointain Info About Date, Country And Continent.



# EXPLORE DATA

- I Write Some Queries Using Sql Server To Understand Column And Relation Between Them So I Can Work In Dataset In Easy Way .
- Some Part Of Codes That Explain What I Say :**

```
--COVID DEATHS TABLE--  
select location,date,population,total_cases,new_cases,new_cases_smoothed,total_cases_per_million,new_cases_per_million,new_cases_smoothed_per_million  
from CovidDeaths  
--where location='cambodia'  
order by date  
  
select location,date,population,total_deaths,new_deaths,new_deaths_smoothed,total_deaths_per_million,new_deaths_per_million,new_deaths_smoothed_per_million  
from CovidDeaths  
--where location='united states'  
order by location,date  
  
select location,date,reproduction_rate  
from CovidDeaths  
--where location='United States'  
order by date
```

# Continue Exploring

## ● Exploring In General:

```
--COVID VACCINATIONS TABLE--  
  
select location,date,population,new_tests,new_tests_per_thousand,total_tests,total_tests_per_thousan  
,new_tests_smoothed_per_thousand  
from CovidVaccinations  
where location ='united states'  
order by date  
  
select location,date,tests_per_case,tests_units  
from CovidVaccinations  
where location ='united states'  
order by date  
  
select *  
from CovidVaccinations  
where tests_units is null  
  
select location,date,new_tests,new_tests_per_thousand,total_tests,total_tests_per_thousand,new_tests  
,new_tests_smoothed_per_thousand,tests_per_case,tests_units  
from CovidVaccinations  
where new_tests_smoothed is not null and  
new_tests_smoothed_per_thousand is not null  
order by date  
  
select location,date,  
total_vaccinations,total_vaccinations_per_hundred,  
people_vaccinated,people_vaccinated_per_hundred,  
people_fully_vaccinated,people_fully_vaccinated_per_hundred  
from CovidVaccinations  
--where location='United States'
```

## ● Creating Info Table To Understand Data Easy:

```
Created A New Table To Make The Data More Easy To Understand  
  
create table CovidInformation  
(  
iso_code nvarchar(255),  
continent nvarchar(255),  
location nvarchar(255),  
date date,  
population float,  
median_age float,  
aged_65_older float,  
gdp_per_person float,  
extreme_poverty float,  
stringency_index float,  
cardiovasc_death_rate float,  
diabets_prevalence float,  
life_expectancy float  
)  
  
insert into CovidInformation(iso_code,continent,location,date,population,  
median_age,aged_65_older,gdp_per_person,extreme_poverty,stringency_index,cardiovasc_death_rate,  
diabets_prevalence,life_expectancy)  
select  
iso_code,continent,location,date,population,  
median_age,aged_65_older,gdp_per_capita,extreme_poverty,stringency_index,cardiovasc_death_rate,  
diabetes_prevalence,life_expectancy  
from CovidDeaths  
  
select distinct *from CovidInformation  
where location ='United States'  
order by location,date
```

# CONTINUE

## Exploring In More Advanced Way By Asking Question

--1-What Is Country That Has Max Total Cases And Total Deaths?--

```
select location,max(total_cases)as "Max Num Of Cases",max(total_deaths)as "Max Num Of Deaths"
from CovidDeaths
group by location
order by "Max Num Of Cases"desc , "Max Num Of Deaths"desc
```

--What Is The Death Percentage If You Contract Covid In Your Country?--

```
select location,sum(total_cases)as "Total Cases",sum(total_deaths)as "Total Deaths",
(sum(total_deaths)/sum(total_cases))*100 as "Death Percentage"
from CovidDeaths
--where location='United States'
group by location
order by "Total Cases" desc
```

--What Percentage Of Population Got Covid Through The Daily History?--

```
select location,date,population,total_cases,(total_cases/population)*100 as "Percent Of Population Infected "
from CovidDeaths
order by location,date
```

--Total cases & New cases & New deaths & Total deaths Over Date--

```
select sum(total_cases)as "Sum of Total Cases",sum(new_cases)as "Sum Of New Cases"
from CovidDeaths
--group by date
--order by date desc
```

```
select sum(total_deaths)as "Sum of Total Deaths",sum(new_deaths)as "Sum Of New Deaths"
from CovidDeaths
--group by date
--order by date desc
```

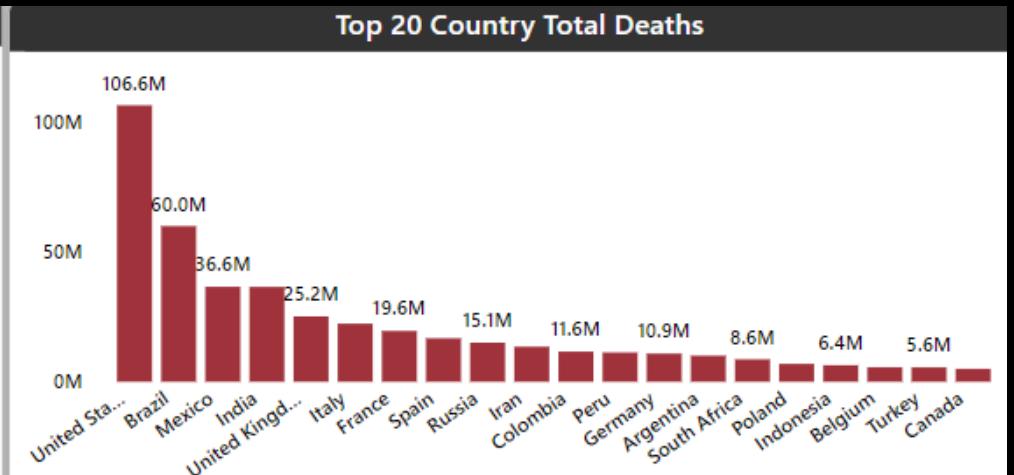
--Different Diseases--

```
select distinct location,cardiovasc_death_rate,diabetes_prevalence
from CovidInformation
where cardiovasc_death_rate is not null and diabetes_prevalence is not null
order by location
```

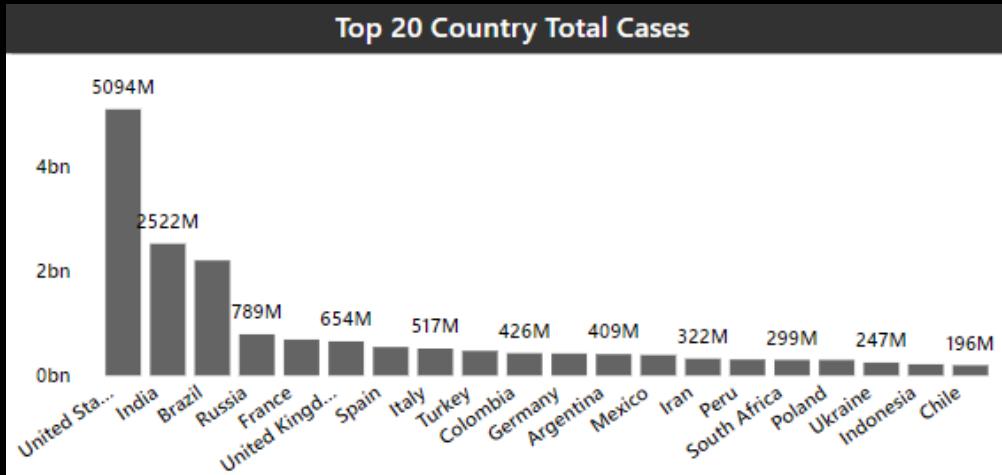
--Some Age Information Related To Population In Each Country--

```
select distinct location,population,median_age,aged_65_older,life_expectancy
from CovidInformation
order by population desc
```

Top 20 Country Total Deaths



Top 20 Country Total Cases



Total Deaths

**526.44M**

Total Cases

**21,556.75M**

# DATA CLEANING

- When I Exploring The Data I found Some Obstacle In Data That Needed To Cleaned In The Column Of Continent And Location And I Found Also A Lot Missing Values That Needed To Cleaned Also.
- **Some Part Of Codes That Explain What I Say :**

```
order by continent desc

select continent,location from coviddeaths
where continent = 'world'
order by continent desc

select *from CovidDeaths
where location= 'European Union'
order by 2,4

update CovidVaccinations
set continent=Location,Location=null
where continent is null and
Location in ('Africa','Asia','Europe','Oceania','South America','World','North America','International')

select *from CovidVaccinations
where location='European Union'

update CovidVaccinations
set continent='Europe'
where Location='European Union'

update CovidVaccinations
```

# CONTINUE CLEANING

## ● Check For Duplicate Values:

```
--Check If There Is Any Duplicate Values--  
select location, date, COUNT(*)  
FROM CovidInformation  
group by location, date  
having COUNT(*) > 1  
  
select location, date, COUNT(*)  
FROM CovidDeaths  
group by location, date  
having COUNT(*) > 1  
  
select location, date, COUNT(*)  
FROM CovidVaccinations  
group by location, date  
having COUNT(*) > 1  
  
select * from CovidInformation  
--where extreme_poverty is not null  
where median_age is null and aged_65_older is null and gdp_per_person is null and extreme_poverty is null  
and stringency_index is null and cardiovasc_death_rate is null and diabetes_prevalence is null and life_expectancy is null  
order by location,date  
  
select * from CovidDeaths d join CovidVaccinations v  
on d.iso_code=v.iso_code  
where d.location in ('European Union','Guernsey','Jersey')  
and  
v.location in ('European Union','Guernsey','Jersey')  
order by d.location,d.date
```

## ● Cleaning Using Select To Understand And Update:

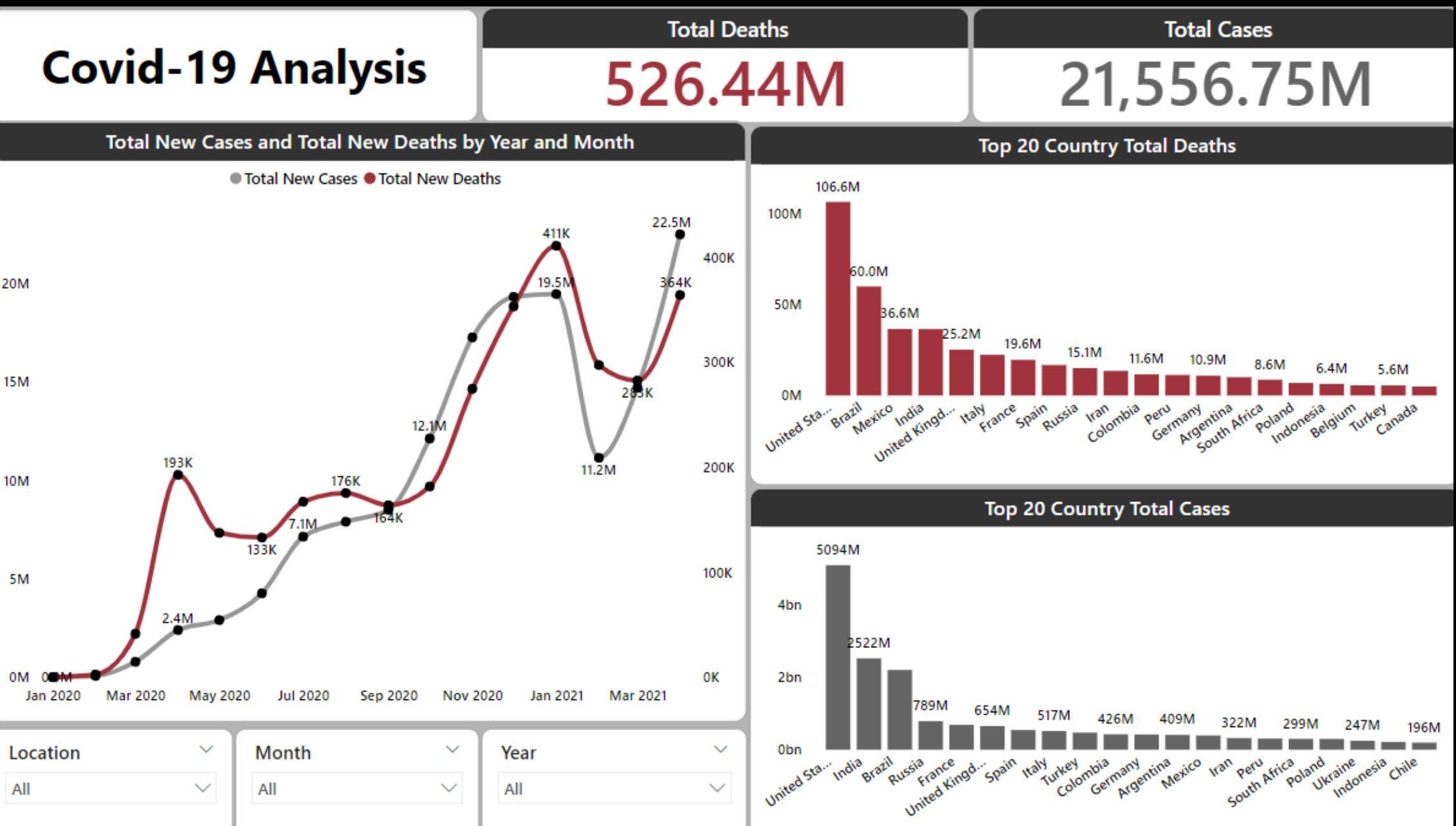
```
UPDATE CovidVaccinations  
SET CovidVaccinations.population= CovidDeaths.population  
FROM CovidDeaths  
JOIN CovidVaccinations  
ON CovidDeaths.location = CovidVaccinations.location;  
  
select d.continent,d.location,v.continent,v.location,i.continent,i.location,d.population,v.population,i.population  
from CovidDeaths d join CovidVaccinations v on d.iso_code=v.iso_code join CovidInformation i on v.iso_code=i.iso_code  
where d.population is null and v.population is null and i.population is null  
  
select *from CovidDeaths  
where population is null  
  
delete from CovidInformation  
where population is null  
  
delete from CovidVaccinations  
where population is null  
  
select reproduction_rate from CovidDeaths  
where reproduction_rate is null  
order by location,date  
  
select continent,location,date  
from CovidVaccinations  
order by location,date
```

# ASKING QUESTION AND EXTRACT INSIGHTS

- This Stage Is Considered A Complement To The Exploration Stage But In More Advanced Way By Asking More Advanced Question That Will Helping In Extract Insights And Do Recommendation For This Data.
- I Will Do This Using SQL,POWER BI.

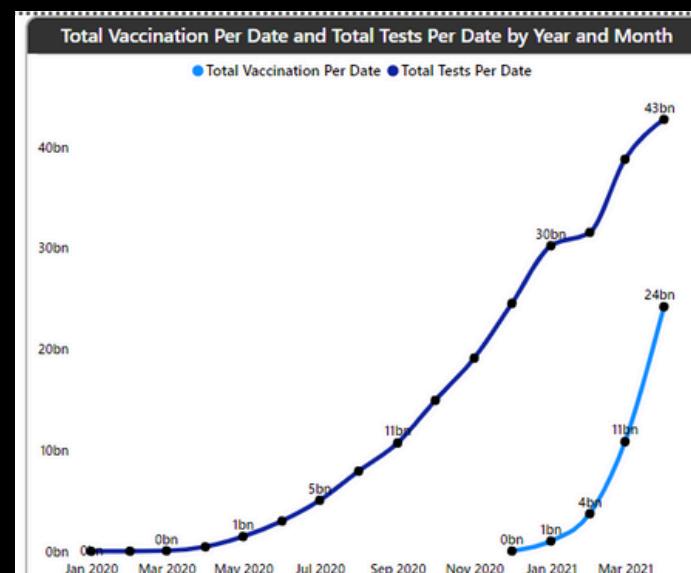


# FIRST DASHBOARD USING POWER BI



- From This Dashboard We Can Extract Some Of Valuable Insights Like:

- The Total New Cases And Death In A Significant Rise Over Time Specially The Interval Between 10-2020 And 1-2021.



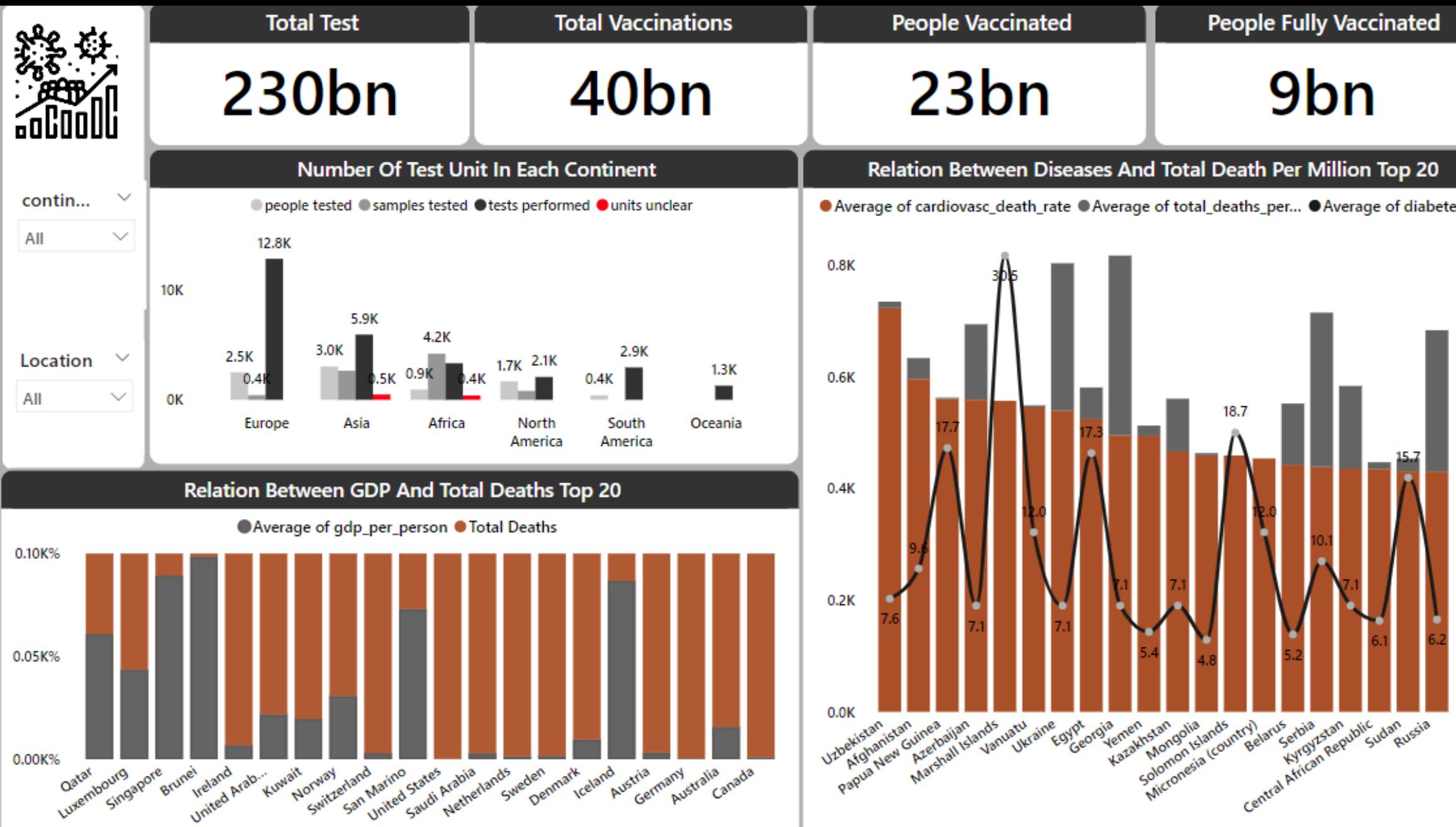
- I Think Reasons For That Is The Emergence Of New Mutants With Greater Ability To Spread, Maybe Easing Restrictions And Precautionary Measures Somewhat.

- This Interval Shows That There Was Second Wave More Powerful From First With Weaknesses Of Prevention Procedures At That Time.

- In 12-2020 The Curve Of Total Vaccination And Tests Become Increasing Which Led To Total New Cases And Deaths Decreasing But We Notice After That The Curve Become To increase Another Time And This Is Due To Several Reason I Think Because The Data Is Not To Show That.

- Most Countries That Effected By Covid-19 Is United States, Brazil, India And So On. This Show That These Countries Was Too Late To Take Preventive Measures.

# SECOND DASHBOARD USING POWER BI



- From This Dashboard We Can Extract More And More Insights:

- First:** Covid-19 Was Resisted In Very Large Way As This Is Represented By Total Tests And Total Vaccinations And The Following Query Will Show That By Results.

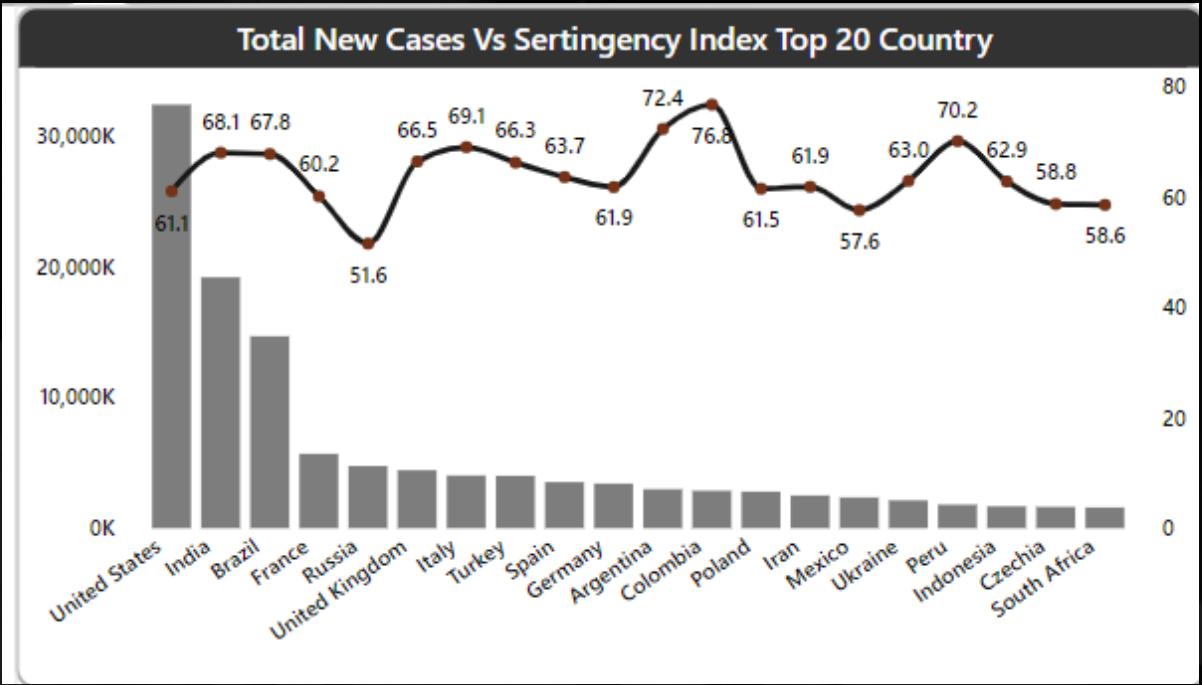
location	Population	Total Cases	Total Tests	Total Vaccinations
1 China	143022774	40520977	295000000	694545700
2 India	138004348	252229196	4076590619	4470222325
3 United States	30102647	509420689	65318143024	1124991759
4 Pakistan	205000000	100000000	100000000	100000000
5 Pakistan	205000000	140227955	162927054	6753261
6 Brazil	21259409	2202143086	50049409	1322292828
7 Nigeria	20619597	28603059	11593064	19118307
8 Bangladesh	164695000	138800023	88064120	2196170
9 Russia	144600000	147700000	100000000	205000000
10 Mexico	12932753	39149940	955790411	55590542
11 Japan	126474585	69818138	131278848	49398827
12 Ethiopia	114963503	20000000	477191661	43000
13 Philippines	114400000	100000000	100000000	240000000
14 Egypt	102334403	43077116	NULL	825549
15 Vietnam	9733583	490772	41462651	4229719
16 Democratic	89561404	5261539	NULL	1700
17 Turkey	84339000	471791753	633066189	799375102
18 Spain	83000000	321791715	7090751429	3765348
19 Germany	8379345	420309715	1305469340	1074497309
20 Thailand	69799678	4125390	545674030	21349721
21 France	68147687	687236610	NULL	8317880
22 United Kingdom	653100000	2023188091	264597240	240000000
23 Luxembourg	60461238	51601020	7640575089	441792
24 Tanzania	5973213	190320	NULL	NULL

We Will Find That The Total Number Are Different From One Country To Another But That Show How Much Resistance.

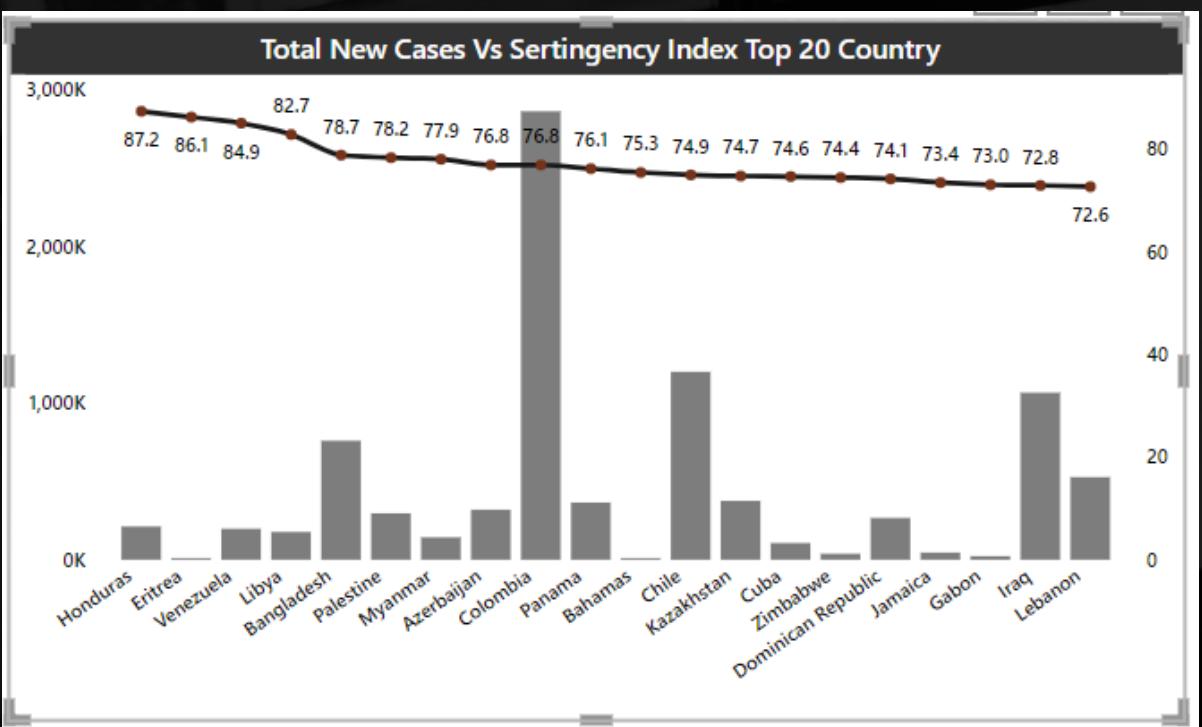
location	Average GDP	Sum Of Deaths
1 United States	54225.446	106611210
2 Brazil	14103.452	60047815
3 Mexico	17336.469	36622615
4 India	6426.674	36578144
5 United Kingdom	39753.244	25219149
6 Italy	35220.084	22409670
7 France	38605.671	19603382
8 Spain	34272.36	16767799
9 Russia	24765.954	15069275
10 Iran	19082.62	13528093
11 Colombia	13254.949	11636649
12 Peru	12236.706	11292517
13 Germany	45229.245000001	10869835
14 Argentina	18933.907	10071947
15 South Africa	12294.876	8648025
16 Poland	27216.445	6889123

location	Average GDP	Sum Of Deaths
1 Qatar	116935.6	75484
2 Luxembourg	94277.965	122327
3 Singapore	85535.383	10324
4 Brunei	71809.251000001	1057
5 Ireland	67335.293	937846
6 United Arab...	67293.483	241083
7 Kuwait	65530.537	270044
8 Norway	64800.057	146129
9 Switzerland	57410.166	1877246
10 San Marino	56861.47	21048
11 United States	54225.446	106611210
12 Saudi Arabia	49045.411	1653691
13 Netherlands	48472.545	3655650
14 Sweden	46949.283	292602
15 Denmark	46682.515	442575
16 Iceland	46482.958	7146
17 Austria	45436.686	1375401
18 Germany	45229.245000001	10869835

# MORE INSIGHTS



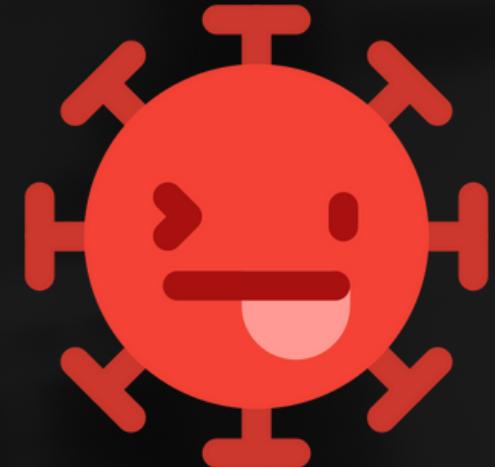
- This Two Visual Explain The Relation Between Stringency Index And Total New Cases.
- First One In Case That Countries Have Most Total Cases And In Second In Case Countries That Have Most Stringency Index And Based On Those Cases.
- In My Opinion That Stringency Index Play A Slight Role For Total New Cases ,As There Are A Lot Of Country Has High Rate Of This Index And Yet There Are Many Cases In It And Vice Versa



# RECOMMENDATION

After Exploring,Cleaning,Asking And Extract Some Insights I will Write Some Recommendation In My Opinion Based On The Data I Worked On And I See That It Is Summarized In:

- Making Disease-Related Data Files In General Has More Details About Everything That Helping All People In Data Field Extract The Largest Amount Of Information In Order To Help In The Future Tf The World Faces Any Other Types Of Deases.
- Develop Pedictive Models To Predict The Spread Of The Virus In The Future, Which Helps In Making Early Decisions On Preventive Measures Especially In Countries With Large Number Of Deaths And Cases(ex:United States).
- Work To Raise The Level Of Health Care In General In Most Countries
- International Cooperation Must Be Strengthened To Exchange Data, Experiences And Knowledge In The Field Of Epidemic Control.
- Raising Awareness Of Societies In General Of Th Importance Of Commitment To Government, Health Instructions To Help Resist Epidemics.
- Developing Epidemiological Surveillance Systems To Follow Up Mutations For Any Epidemic In General In Order Not To Give The Epidemic A Chance To spread Out Quickly.





See You Next

# THANK YOU

OMAR NASR