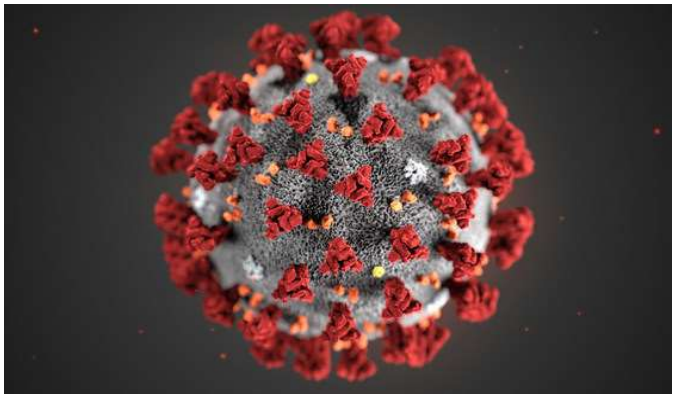
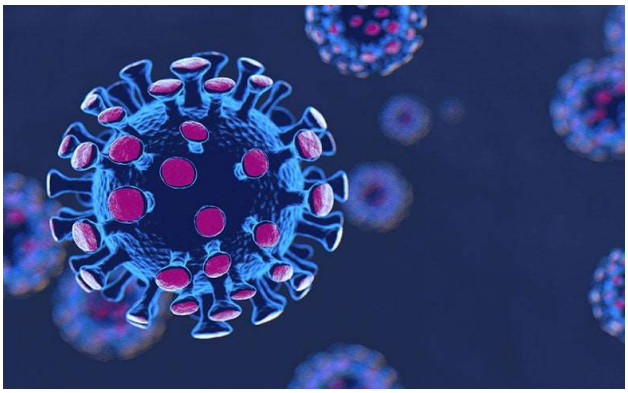
ProjectCoronavirus





TEAM ABC:

Aditya Jain

INDEX

Executive Summary

This comprehensive database schema is designed to equip businesses with the critical insights needed to maximize revenue in the context of the COVID-19 pandemic. By integrating extensive data across multiple dimensions, businesses can strategically navigate the challenges posed by the pandemic, identify new opportunities, and optimize their revenue streams. The schema is organized into four main categories:

1. **Basic Information and Location**: This section includes essential details such as ISO code, continent, location, and date, enabling businesses to accurately track and contextualize COVID-19 data for different countries or regions. This foundational data is crucial for market segmentation and targeted revenue strategies, ensuring businesses can tailor their approaches based on regional variations in pandemic impact.
2. **Testing and Positive Cases**: These tables capture critical information on COVID-19 testing efforts and outcomes. Key metrics include the number of new and total tests conducted, testing rates per thousand people, smoothed averages, positive test rates, and the number of tests per confirmed case. This data helps businesses understand the level of pandemic control in different regions, informing decisions on market entry, product launches, and customer engagement strategies to maximize revenue potential in safer, more stable markets.
3. **Vaccination Information**: This section focuses on vaccination efforts, detailing total and new vaccinations, the number of people vaccinated with at least one dose, fully vaccinated individuals, and vaccination rates per hundred people. Smoothed averages provide a clearer picture of vaccination trends over time. Understanding vaccination progress is essential for businesses planning to reopen or expand operations, ensuring a safe environment for employees and customers. High vaccination rates can correlate with increased consumer confidence and spending, driving revenue growth.
4. **Demographic and Health Indicators**: This category encompasses a range of demographic and health-related factors that influence the spread and impact of COVID-19. Metrics include population density, median age, the percentage of elderly population, GDP per capita, extreme poverty rates, and various health indicators (e.g., cardiovascular death rate, diabetes prevalence, smoking rates, handwashing facilities availability, hospital beds per thousand people, life expectancy, and the Human Development Index). Additionally, the government response stringency index is included to evaluate the rigor of policy measures in different regions. These indicators help businesses assess market stability and consumer behaviour, allowing for more accurate revenue forecasting and strategy adjustments.

Industry Picture

The COVID-19 pandemic significantly reshaped global industries, leading to a surge in the healthcare sector, with vaccine sales reaching $35 billion and PPE market value hitting $77 billion. Technology adoption accelerated, exemplified by Zoom's 326% revenue increase in 2021. E-commerce boomed, with Amazon's sales growing 38% to $386 billion in 2020. The hospitality and travel sectors faced severe losses but began gradual recovery, while streaming services like Netflix added millions of subscribers. The e-learning market grew to $250.8 billion in 2020, and EdTech funding soared to $16.1 billion. Supply chain disruptions cost the automotive industry $210 billion, driving investments in local production. Digital banking expanded, supported by $5 trillion in U.S. government stimulus. The real estate market saw residential property prices surge, while commercial spaces struggled with high vacancy rates. Environmental impacts included a temporary 5.8% drop in global CO_2 emissions due to reduced activity during lockdowns.

Business and Revenue Overview

Objective: The SQL ProjectCoronavirus aims to analyse the impact of the COVID-19 pandemic on business sectors, focusing on revenue, market trends, and recovery patterns using SQL databases.

Key Areas of Focus:

1. **Revenue Impact Analysis:** Examining how different sectors overall experienced revenue changes due to the pandemic. Comparing revenue impacts across different regions and countries to understand geographical disparities in business performance.
2. **Market Trends:** Study changes in consumer behaviour, such as increased shopping, shifts in product demand, and changes in spending habits.
3. **Recovery Patterns:** Analyse the effect of government support measures like stringency helps on business recovery.

Methodology:

- **Data Collection:** Gather data from "**ourworldindata.com"**, Our World in Data has been cited in academic scientific journals, medicine and global health journals, and social science journals. The Washington Post, The New York Times, and The Economist have used Our World in Data as a source.
- **SQL Analysis:** Utilize SQL to query and analyse large dataset, extract meaningful insights, and generate reports.
- **Visualization:** Create visual representations of data using tools like Tableau or Power BI to highlight key findings and trends.

Expected Outcomes:

- Provide a detailed report on the impact of COVID-19 in different ways, supported by data-driven insights. Offer strategic recommendations for businesses to enhance resilience and adapt to post-pandemic market conditions. Highlight the implications for policymakers in supporting economic recovery and future-proofing businesses against similar crises.

Table Structure

❖ CovidDeaths Table:

- **Basic Information and Location**
  - iso_code: ISO code for the country or region
  - continent: Continent where the country or region is located
  - location: Name of the country or region
  - date: Date of the recorded data

- **Population and General Statistics**
  - population: Total population of the country or region
  - total_cases: Total confirmed COVID-19 cases reported
  - new_cases: New confirmed COVID-19 cases reported on the given date
  - new_cases_smoothed: Smoothed new confirmed COVID-19 cases (rolling average)

- **Fatalities and Mortality Rates**
  - total_deaths: Total deaths due to COVID-19 reported
  - new_deaths: New deaths due to COVID-19 reported on the given date
  - new_deaths_smoothed: Smoothed new deaths due to COVID-19 (rolling average)
  - total_cases_per_million: Total confirmed cases per million people
  - new_cases_per_million: New confirmed cases per million people
  - new_cases_smoothed_per_million: Smoothed new confirmed cases per million people
  - total_deaths_per_million: Total deaths per million people
  - new_deaths_per_million: New deaths per million people
  - new_deaths_smoothed_per_million: Smoothed new deaths per million people

- **Healthcare and Pandemic Management**
  - reproduction_rate: Effective reproduction rate of COVID-19 transmission
  - icu_patients: Number of COVID-19 patients in intensive care units
  - icu_patients_per_million: Number of COVID-19 patients in ICU per million people
  - hosp_patients: Number of COVID-19 patients hospitalized
  - hosp_patients_per_million: Number of COVID-19 patients hospitalized per million people
  - weekly_icu_admissions: Weekly number of COVID-19 patients admitted to ICU
  - weekly_icu_admissions_per_million: Weekly number of COVID-19 patients admitted to ICU per million people
  - weekly_hosp_admissions: Weekly number of COVID-19 patients admitted to hospitals
  - weekly_hosp_admissions_per_million: Weekly number of COVID-19 patients admitted to hospitals per million people

❖ CovidVaccinations Table:

- **Basic Information and Location**
  - iso_code: ISO code for the country or region
  - continent: Continent where the country or region is located
  - location: Name of the country or region
  - date: Date of the recorded data

- **Testing and Positive Cases**
  - new_tests: New COVID-19 tests conducted on the given date
  - total_tests: Total number of COVID-19 tests conducted
  - total_tests_per_thousand: Total number of COVID-19 tests per thousand people
  - new_tests_per_thousand: New COVID-19 tests per thousand people
  - new_tests_smoothed: Smoothed new COVID-19 tests (rolling average)
  - new_tests_smoothed_per_thousand: Smoothed new COVID-19 tests per thousand people
  - positive_rate: Percentage of positive COVID-19 tests
  - tests_per_case: Number of tests conducted per confirmed case
  - tests_units: Units used for reporting COVID-19 tests

- **Vaccination Information**
  - total_vaccinations: Total number of COVID-19 vaccine doses administered
  - people_vaccinated: Number of people vaccinated with at least one dose
  - people_fully_vaccinated: Number of people fully vaccinated (received all recommended doses)
  - new_vaccinations: New COVID-19 vaccine doses administered on the given date

- new_vaccinations_smoothed: Smoothed new COVID-19 vaccine doses administered (rolling average)
- total_vaccinations_per_hundred: Total COVID-19 vaccine doses administered per hundred people
- people_vaccinated_per_hundred: Number of people vaccinated with at least one dose per hundred people
- people_fully_vaccinated_per_hundred: Number of people fully vaccinated per hundred people
- new_vaccinations_smoothed_per_million: Smoothed new COVID-19 vaccine doses administered per million people

- **Demographic and Health Indicators**
  - stringency_index: Government response stringency index
  - population_density: Population density (people per square kilometre)
  - median_age: Median age of the population
  - aged_65_older: Percentage of the population aged 65 years and older
  - aged_70_older: Percentage of the population aged 70 years and older
  - gdp_per_capita: Gross domestic product (GDP) per capita
  - extreme_poverty: Percentage of the population living in extreme poverty
  - cardiovasc_death_rate: Cardiovascular disease death rate (annual deaths per 100,000 people)
  - diabetes_prevalence: Diabetes prevalence (% of population aged 20-79)
  - female_smokers: Percentage of females who smoke
  - male_smokers: Percentage of males who smoke
  - handwashing_facilities: Percentage of population with basic handwashing facilities
  - hospital_beds_per_thousand: Hospital beds per thousand people
  - life_expectancy: Life expectancy at birth

Public Visualization's Link

https://public.tableau.com/app/profile/aditya.jain5502/vizzes

Exploratory Data Analysis

a) How many lines does the dataset have?

- The dataset has 85171 rows in both COVID deaths and vaccinations table.

	DeathsRows
1	85171

	VaccRows
1	85171

b) Basic queries - Selecting everything from a table.

	iso_code	continent	location	date	population	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_millic
1	AFG	Asia	Afghanistan	2020-02-24 00:00:00.000	38928341	1	1	NULL	NULL	NULL	NULL	0.026
2	AFG	Asia	Afghanistan	2020-02-25 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026
3	AFG	Asia	Afghanistan	2020-02-26 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026
4	AFG	Asia	Afghanistan	2020-02-27 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026
5	AFG	Asia	Afghanistan	2020-02-28 00:00:00.000	38928341	1	0	NULL	NULL	NULL	NULL	0.026

c) Selecting Data that we will be using (6 columns and all rows)

	location	date	total_cases	new_cases	total_deaths	population
1	Afghanistan	2020-02-24 00:00:00.000	1	1	NULL	38928341
2	Afghanistan	2020-02-25 00:00:00.000	1	0	NULL	38928341
3	Afghanistan	2020-02-26 00:00:00.000	1	0	NULL	38928341
4	Afghanistan	2020-02-27 00:00:00.000	1	0	NULL	38928341
5	Afghanistan	2020-02-28 00:00:00.000	1	0	NULL	38928341
6	Afghanistan	2020-02-29 00:00:00.000	1	0	NULL	38928341

d) Checking for duplicate values

- The dataset's do not have any duplicate values.

	date	continent	location	Checking_Dup

	date	continent	location	Checking_Dup

e) Checking the number of continents and countries.

- There are 6 continents and 210 countries. The countries list also include countries such as Micronesia, Timor and more.

	Continent	Count
1	Africa	1
2	Asia	1
3	Europe	1
4	North America	1
5	Oceania	1
6	South America	1
7	Total	6
8	Total	0

	No_Of_Countries
1	210

Result output of the queries built in the project

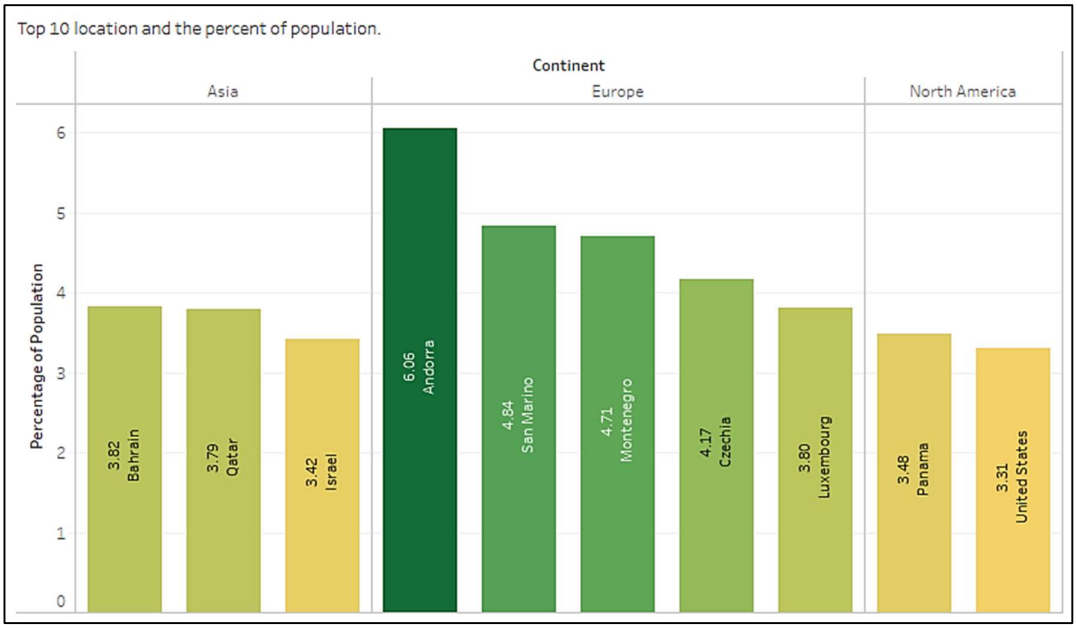
f) Top 10 location, the continents and the percent of population affected on average.

Considering top 10 locations only, the average percentage of the population affected for each continent is

- Europe: (6.06 + 4.84 + 4.71 + 4.17 + 3.8) / 5 = 4.716
- Asia: (3.82 + 3.79 + 3.42) / 3 = 3.6767
- North America: (3.48 + 3.31) / 2 = 3.395

Observations:

- Europe has the highest average percentage of population affected.
- North America has the second-highest average percentage.
- Asia has the lowest average percentage of population affected among these three continents.



	continent	location	Percentage_Population
1	Europe	Andorra	6.06
2	Europe	San Marino	4.84
3	Europe	Montenegro	4.71
4	Europe	Czechia	4.17
5	Asia	Bahrain	3.82
6	Europe	Luxembourg	3.8
7	Asia	Qatar	3.79
8	North America	Panama	3.48
9	Asia	Israel	3.42
10	North America	United States	3.31

g) Likelihood of dying if you contract covid in one's country.

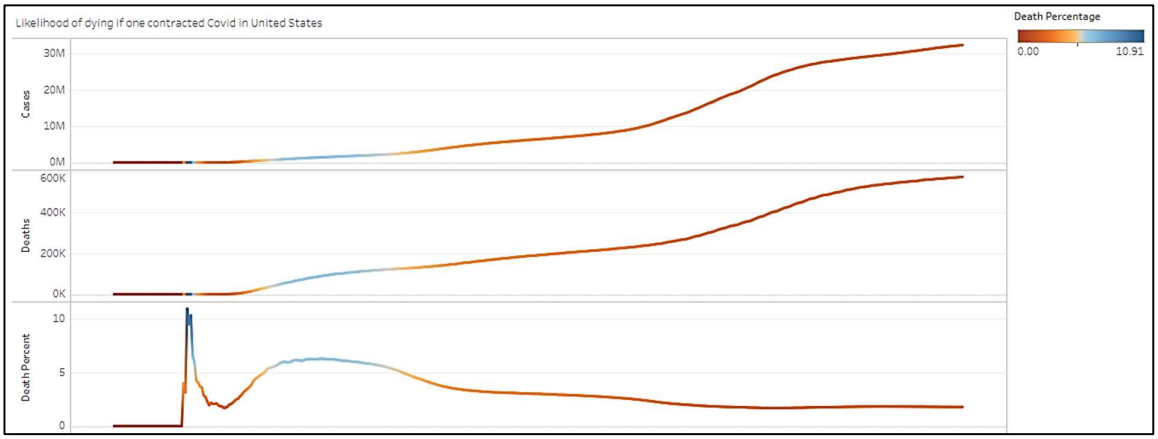
- There is a 1.78% possibility that one could die if a person catches COVID and around or on 30 Apr, 2021 in the United States. This percentage peaked at 6.25% on 15 May, 2020 in the United States.

Observations:

- **Cases:** The top line chart shows the cumulative number of COVID-19 cases and exhibits a sharp initial increase followed by a more gradual rise.
- **Deaths:** The middle line chart displays the cumulative number of COVID-19 deaths over time. It mirrors the trend of cases but with a delayed and less steep increase.
- **Death Percentage:** The bottom-line chart illustrates the percentage of COVID-19 cases that resulted in death. It shows a significant spike during the early months of the pandemic, followed by a gradual decline.

Inferences:

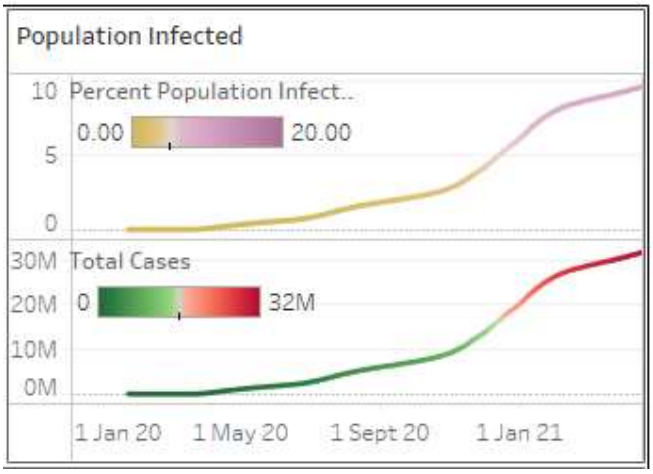
- The data suggests that the overall mortality rate (death percentage) from COVID-19 in the United States has decreased over time. This could be attributed to various factors, such as improved medical treatments, increased awareness and preventative measures, and the impact of vaccinations.
- The initial surge in both cases and deaths highlights the rapid spread and severe impact of the virus during the early stages of the pandemic.
- The gradual decline in the death percentage indicates that the healthcare system has become more adept at managing COVID-19 cases, leading to a lower likelihood of fatal outcomes.



(The data is for North America - United States in the Dashboard.)

h) Show what percentage of population infected with Covid.

- In the United States, 9.77% of the population or 32 million people were infected with COVID on 30 Apr, 2021.



	Location	date	Population	total_cases	PercentPopulationInfected
80799	United States	2020-03-17 00:00:00.000	331002647	6512	0.00196735586830519
80800	United States	2020-03-18 00:00:00.000	331002647	9169	0.00277006848226202
80801	United States	2020-03-19 00:00:00.000	331002647	13663	0.0041277615523117
80802	United States	2020-03-20 00:00:00.000	331002647	20030	0.00605131112440923
80803	United States	2020-03-21 00:00:00.000	331002647	26025	0.00786247488830505
80804	United States	2020-03-22 00:00:00.000	331002647	34898	0.0105431181038259
80805	United States	2020-03-23 00:00:00.000	331002647	46136	0.013938257116113

(The data is for North America - United States in the Dashboard.)

i) Countries with Highest Infection Rate compared to Population

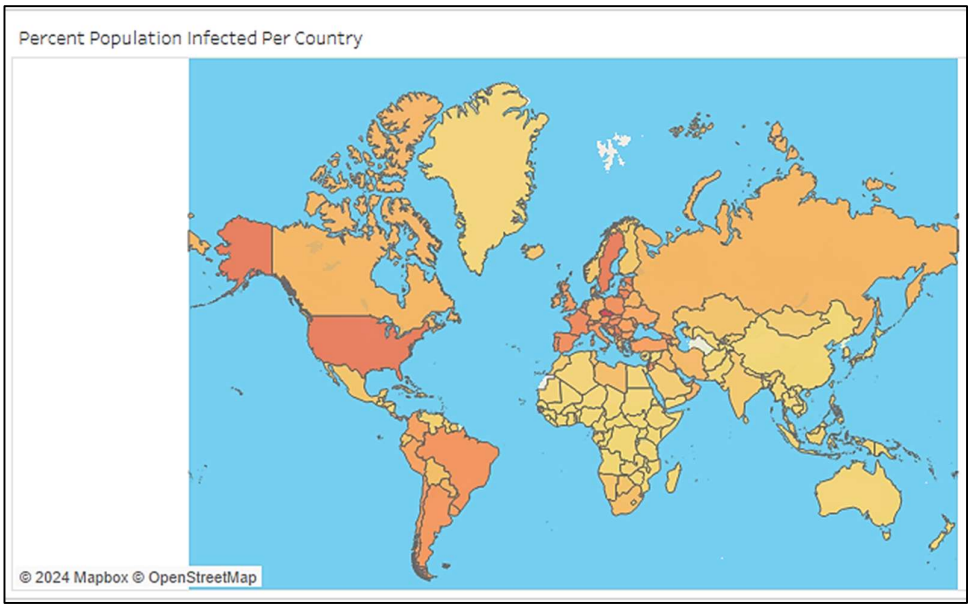
Top 5 Countries with Highest Infection Rates:

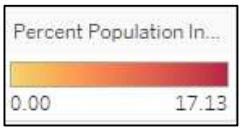
- Andorra: With a population of 77,265 and 13,232 infections, Andorra has the highest infection rate at 17.13%.
- Montenegro: Following closely behind, Montenegro has a population of 628,062 and 97,389 infections, resulting in a 15.51% infection rate.
- Czechia: With a population of 10,708,982 and 1,630,758 infections, Czechia has a 15.23% infection rate.
- San Marino: A smaller country with a population of 33,938 and 5,066 infections, San Marino has a 14.93% infection rate.
- Slovenia: With a population of 2,078,932 and 240,292 infections, Slovenia has an 11.56% infection rate.

Observations:

- Smaller countries like Andorra and San Marino appear to have a higher percentage of the population infected compared to larger countries. This could be due to various factors, including population density, healthcare systems, and testing rates.

	Location	Population	HighestInfectionCount	PercentPopulationInfected
1	Andorra	77265	13232	17.1254772536077
2	Montenegro	628062	97389	15.5062716738156
3	Czechia	10708982	1630758	15.2279460363273
4	San Marino	33938	5066	14.9272202251164
5	Slovenia	2078932	240292	11.5584348117206
6	Luxembourg	625976	67205	10.7360346083556
7	Bahrain	1701583	176934	10.3981997939566
8	Serbia	6804596	689557	10.1336949320724
9	United States	331002647	32346971	9.77242064169958
10	Israel	8655541	838481	9.68721654718059





j) Countries with Highest Death Count per Population.

United States has the highest death count:

- The United States has recorded the highest number of COVID-19 deaths at **576,232**. This significantly surpasses the figures of other countries on the list.

Brazil and Mexico follow at a distance:

- Brazil holds the second-highest death count at **403,781**, followed by Mexico with **216,907** deaths. These numbers, while substantial, are still considerably lower than the United States.

India and the United Kingdom have comparable figures:

- India and the United Kingdom have reported similar death tolls, with **211,853** and **127,775** deaths, respectively. These figures are notably higher than those of Italy, Russia, and France.

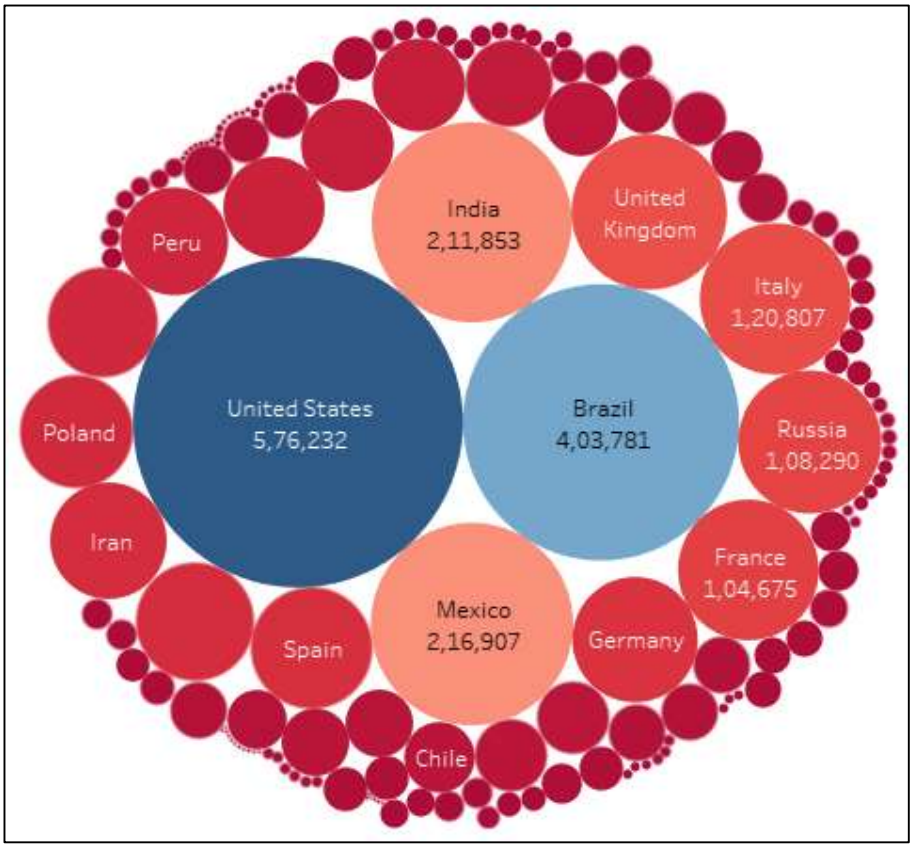
European countries show varying levels of impact:

- Among the European nations listed, Italy has the highest death count at **120,807**, followed by the United Kingdom, France, Germany, and Spain. These countries have experienced significant COVID-19 mortality, but their numbers are lower than those of the United States, Brazil, and Mexico.

Russia's figures are lower than some European nations:

- Russia, despite being a large country, has reported a lower death count (**108,290**) compared to several European nations. This could be attributed to various factors such as population density, healthcare systems, and data reporting practices.

	Location	TotalDeathCount
1	United States	576232
2	Brazil	403781
3	Mexico	216907
4	India	211853
5	United Kingdom	127775
6	Italy	120807
7	Russia	108290
8	France	104675
9	Germany	83097
10	Spain	78216



k) Showing continents with the highest death count per population.

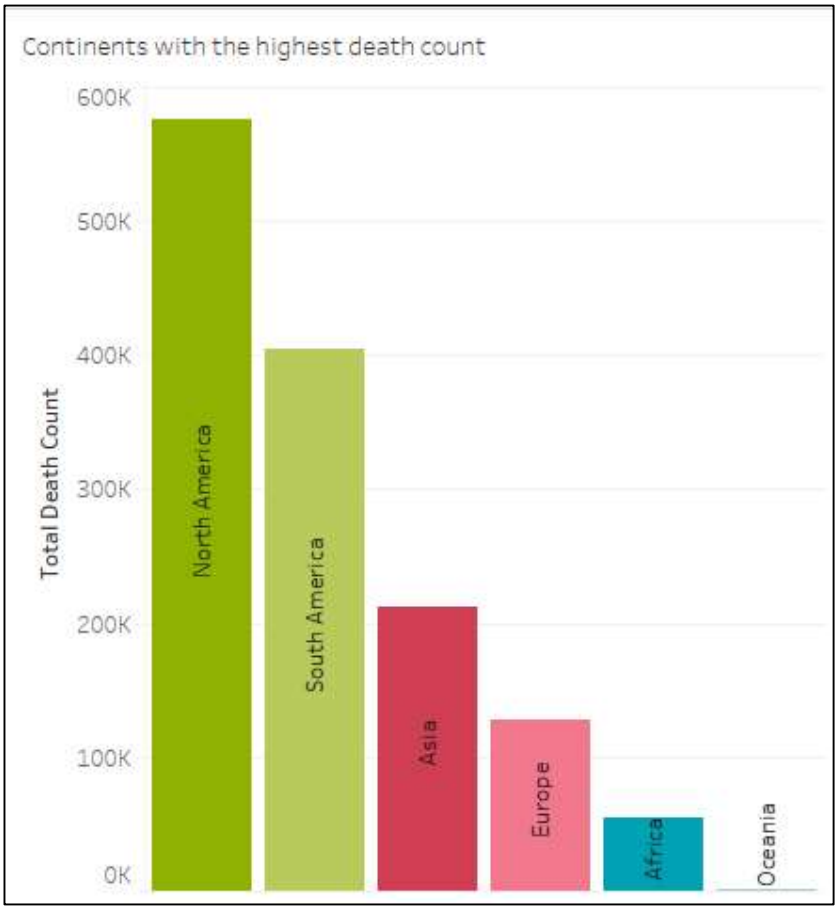
Top 3 Continents with Highest Death Counts:

- Europe: With a staggering 1,016,750 deaths, Europe has the highest recorded death count.
- North America: Following closely behind, North America has recorded 847,942 deaths.
- South America: In third place, South America has reported 672,415 deaths.

Observations:

- These three continents account for the vast majority of global COVID-19 deaths, highlighting their significant impact.

	location	TotalDeathCount
1	World	3180238
2	Europe	1016750
3	North America	847942
4	South America	672415
5	Asia	520286
6	Africa	121784
7	Oceania	1046



l) GLOBAL NUMBERS

The data indicates a significant global impact of the COVID-19 pandemic. And a mortality rate of 2%.

	total_cases	total_deaths	DeathPercentage
1	150574977	3180206	2.11204149810363

Global Numbers

Total Cases	150,574,977
Total Deaths	3,180,206
Death Perce..	2

--- VACCINATION ANALYSIS ---

m) What is number of people who has received at least one COVID vaccine and the percentage vaccinated globally? (Rolling Basis)

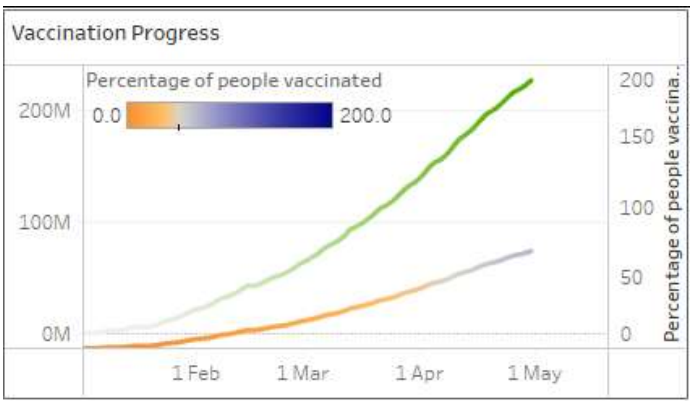
Number of people who have received at least one COVID vaccine:

- The green line represents the cumulative number of people vaccinated.
- By the end of the timeframe shown (around May 1st), the green line appears to be slightly above 200 million.
- Therefore, approximately 200 million people had received at least one COVID vaccine by that point.

Percentage of people vaccinated:

- The blueish line represents the percentage of the population vaccinated.
- At the end of the timeframe, this line seems to be close to 60%.
- This suggests that roughly 60% of the population had received at least one COVID vaccine dose by that point.

	continent	location	date	population	new_vaccinations	RollingPeopleVaccinated
756	Europe	Albania	2021-01-13 00:00:00.000	2877800	60	60
757	Europe	Albania	2021-01-14 00:00:00.000	2877800	78	138
758	Europe	Albania	2021-01-15 00:00:00.000	2877800	42	180
759	Europe	Albania	2021-01-16 00:00:00.000	2877800	61	241
760	Europe	Albania	2021-01-17 00:00:00.000	2877800	36	277
761	Europe	Albania	2021-01-18 00:00:00.000	2877800	42	319
762	Europe	Albania	2021-01-19 00:00:00.000	2877800	36	355



(The data is for North America - United States in the Dashboard.)

n) Creating View to store data for later visualizations.

	continent	location	date	population	new_vaccinations	RollingPeopleVaccinated
1	Asia	Afghanistan	2020-02-24 00:00:00.000	38928341	NULL	NULL
2	Asia	Afghanistan	2020-02-25 00:00:00.000	38928341	NULL	NULL
3	Asia	Afghanistan	2020-02-26 00:00:00.000	38928341	NULL	NULL
4	Asia	Afghanistan	2020-02-27 00:00:00.000	38928341	NULL	NULL
5	Asia	Afghanistan	2020-02-28 00:00:00.000	38928341	NULL	NULL
6	Asia	Afghanistan	2020-02-29 00:00:00.000	38928341	NULL	NULL
7	Asia	Afghanistan	2020-03-01 00:00:00.000	38928341	NULL	NULL
8	Asia	Afghanistan	2020-03-02 00:00:00.000	38928341	NULL	NULL
9	Asia	Afghanistan	2020-03-03 00:00:00.000	38928341	NULL	NULL
10	Asia	Afghanistan	2020-03-04 00:00:00.000	38928341	NULL	NULL
11	Asia	Afghanistan	2020-03-05 00:00:00.000	38928341	NULL	NULL

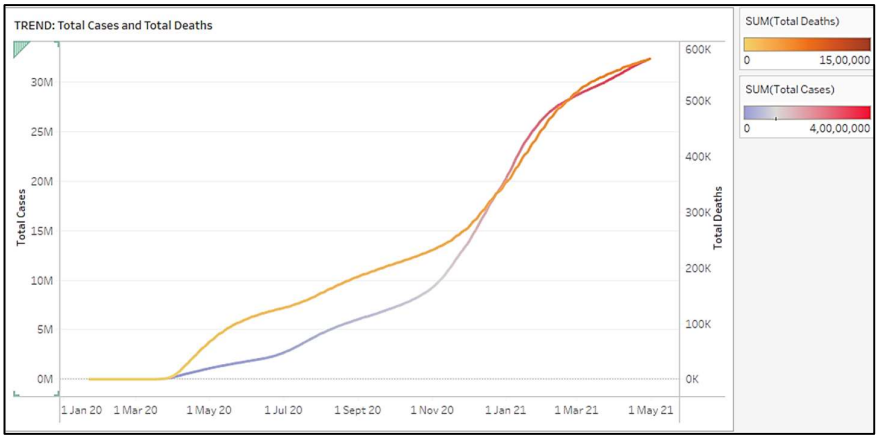
--- TREND ANALYSIS ---

o)1) How have total cases and total deaths trended over time globally?

Both total cases and total deaths have exhibited a steady upward trend over the period. The rate of increase for cases appears to have accelerated at some point, and there's a noticeable correlation and lag between the trends of cases and deaths.

It seems that approximately before the start of 2021 the total deaths were more than total cases and around the start of 2021 the vaccinations came and since then graph of total deaths has gone below total cases. The vaccination attempts of the government seems to be working.

	Location	date	total_cases	total_deaths
77403	United States	2020-11-22 00:00:00.000	12360235	258828
77404	United States	2020-11-23 00:00:00.000	12534684	259928
77405	United States	2020-11-24 00:00:00.000	12710198	262057
77406	United States	2020-11-25 00:00:00.000	12893485	264320
77407	United States	2020-11-26 00:00:00.000	13005807	265705
77408	United States	2020-11-27 00:00:00.000	13213995	267254
77409	United States	2020-11-28 00:00:00.000	13369528	268614



(The data is for North America - United States in the Dashboard.)

o)2) What are the trends in NEW cases and NEW deaths over time globally?

	date	TotalNewCases	TotalNewDeaths
22	2020-01-22 00:00:00.000	0	0
23	2020-01-23 00:00:00.000	294	3
24	2020-01-24 00:00:00.000	860	24
25	2020-01-25 00:00:00.000	1477	48
26	2020-01-26 00:00:00.000	2055	42
27	2020-01-27 00:00:00.000	2428	78
28	2020-01-28 00:00:00.000	7957	147

New Cases

- Rapid Increase: There was a steep increase in new cases from late 2020 to early 2021.
- Peak: The peak of new cases appears to have occurred around January 2021.
- Decline and Fluctuations: Following the peak, there was a general downward trend with significant fluctuations. The number of new cases has remained relatively high throughout the period.

*Equation: Total New Cases = 5154.2*Day of Date - 2.26171e+08*
Slope: Positive (5154.2)
Intercept: -2.26171e+08
p-value: < 0.0001 (highly significant)

Interpretation:

- Similar to the new death's regression, the positive slope here indicates a statistically significant positive relationship between 'Day of Date' and 'Total New Cases'. This means that there's a strong tendency for the number of new cases to increase as the days progress.
- The intercept of -2.26171e+08 is likely a negative value due to the starting point of the data and shouldn't be interpreted as an actual number of cases on Day 0.
- The p-value of less than 0.0001 suggests a very strong statistical relationship between the 'Day of Date' and 'Total New Cases'. This strengthens the conclusion that the increase in new cases is likely not random.

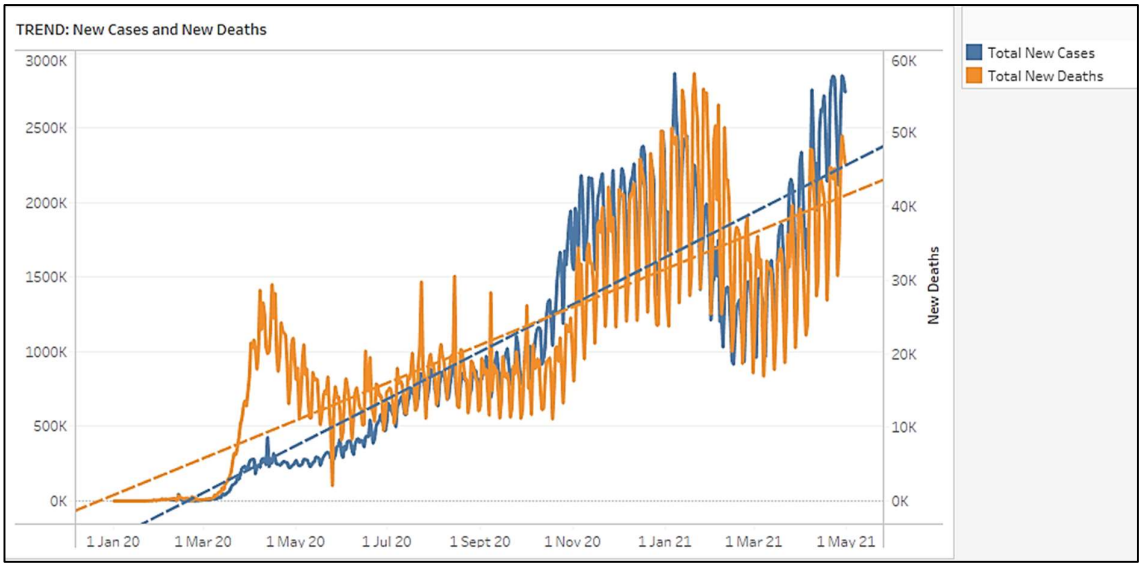
New Deaths

- Steady Increase: New deaths showed a steady increase throughout the entire period, though the rate of increase was not as dramatic as for new cases.
- Correlation with Cases: The trend in new deaths generally follows the trend in new cases, with a slight delay.
- Continued Occurrence: Similar to new cases, new deaths remained at a relatively high level throughout the period.

*Equation: Total New Deaths = 83.8639*Day of Date - 3.67513e+06*
Slope: Positive (83.8639)
Intercept: -3,675,130.00
p-value: < 0.0001 (highly significant)

Interpretation:

- The positive slope indicates a statistically significant positive relationship between the 'Day of Date' and 'Total New Deaths'. This means that as the days progress, the number of new deaths tends to increase.
- The intercept of -3,675,130.00 represents the estimated value of new deaths at Day 0. It's important to note that this is likely a negative value due to the starting point of the data and shouldn't be interpreted as an actual number of deaths on Day 0.
- The p-value of less than 0.0001 suggests a very strong statistical relationship between the 'Day of Date' and 'Total New Deaths'. This reinforces the conclusion that the observed increase in new deaths is likely not random.



(The data is for North America - United States in the Dashboard.)

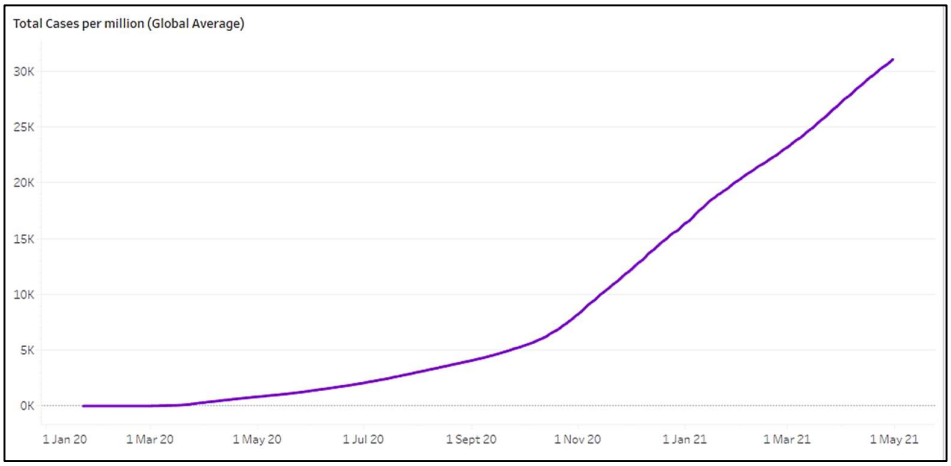
--- PER CAPITA ANALYSIS ---

p)1) What is the highest total cases per million by date globally?

The highest max total cases per million were in Andorra at 171,254 cases per million people on 30 Apr, 2021.

"Note: The graph is not made to show location, it shows the max total cases per million on a particular date."

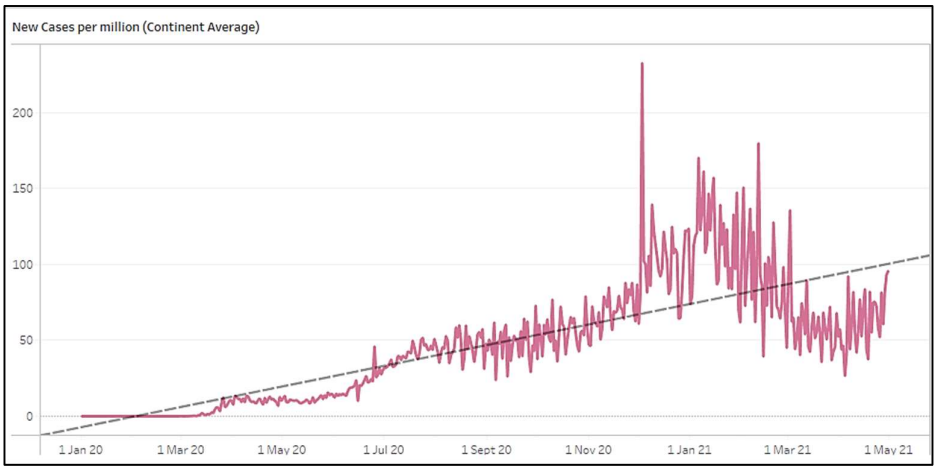
	Country	date	MaxTotalCasesPerMillion
1	Andorra	2021-04-30 00:00:00.000	171254.773
2	Andorra	2021-04-29 00:00:00.000	170814.729
3	Andorra	2021-04-28 00:00:00.000	170167.605
4	Andorra	2021-04-27 00:00:00.000	169818.158
5	Andorra	2021-04-26 00:00:00.000	169326.344
6	Andorra	2021-04-25 00:00:00.000	169028.668
7	Andorra	2021-04-24 00:00:00.000	168562.739



p)2) How do new cases per million vary across different continents?

- Sharp Increase: The number of new cases per million starts low in early 2020 but experiences a sharp increase around March 2020. This likely coincides with the initial wave of the COVID-19 pandemic in the U.S.
- Fluctuations: Throughout 2020 and 2021, the number of cases fluctuates with periods of peaks and troughs. These fluctuations might correspond to various factors such as new variants, changes in public health measures, and seasonal effects.
- Overall Trend: While there are fluctuations, there's a general downward trend in new cases per million from the peak in early 2021 to the end of the data in May 2021. This suggests that the impact of the pandemic was gradually decreasing during this period.

	continent	date	AverageNewCasesPerMillion
537	Africa	2020-04-14 00:00:00.000	2.07290384615385
538	Asia	2020-04-14 00:00:00.000	10.8197555555556
539	Europe	2020-04-14 00:00:00.000	42.7416739130435
540	North America	2020-04-14 00:00:00.000	8.92552173913043
541	Oceania	2020-04-14 00:00:00.000	1.50875
542	South America	2020-04-14 00:00:00.000	5.16291666666667



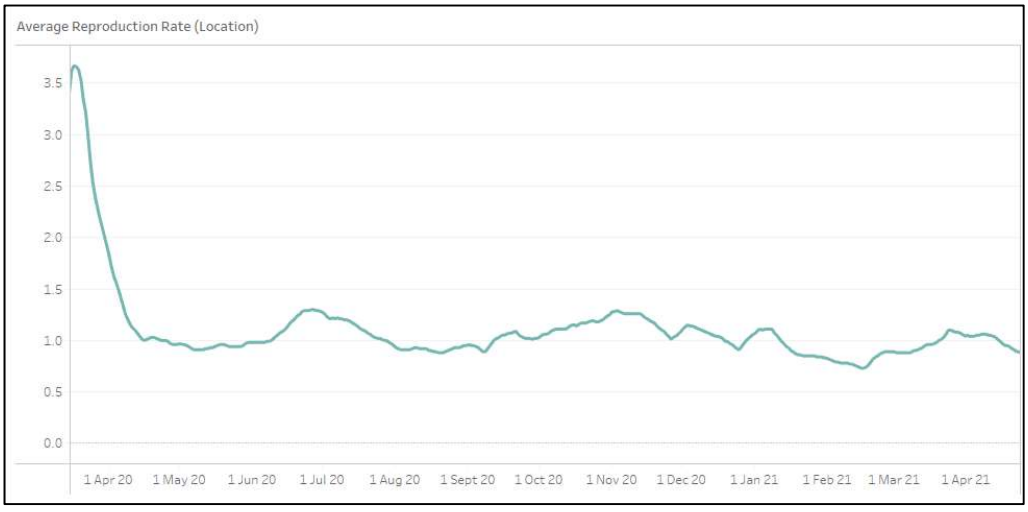
(The data is for North America since we selected United States in the Dashboard.)

--- CASE SEVERITY ---

q)1) What is the reproduction rate over time in different regions?

- Initial Peak: The graph shows a sharp increase in the average reproduction rate (R) at the beginning of the recorded period, reaching a peak of around 3.5 in April 2020. This indicates rapid virus transmission during the early stages of the pandemic.
- Sharp Decline: Following the initial peak, there's a steep decline in R, falling below 1.0 by June 2020. This suggests that the initial wave of the pandemic was brought under control through interventions like lockdowns and social distancing measures.
- Fluctuations and Plateaus: Throughout the rest of the recorded period (June 2020 to April 2021), the average reproduction rate fluctuates between 0.5 and 1.5. There are periods of slight increases, possibly due to factors like seasonal variations or the emergence of new variants, followed by declines. Overall, R remains below 1.0 for most of this period, indicating that the virus was generally not spreading exponentially.

	location	date	AverageReproductionRate
77157	United States	2020-03-21 00:00:00.000	3.34
77158	United States	2020-03-22 00:00:00.000	3.22
77159	United States	2020-03-23 00:00:00.000	2.99
77160	United States	2020-03-24 00:00:00.000	2.74
77161	United States	2020-03-25 00:00:00.000	2.54
77162	United States	2020-03-26 00:00:00.000	2.4
77163	United States	2020-03-27 00:00:00.000	2.29



(The data is for America since we selected United States in the Dashboard.)

q)2) How has the number of ICU patients and hospital patients changed over time?

ICU Patients:

- Initial Peak: The number of ICU patients starts increasing rapidly in March 2020, reaching a peak in April 2020.
- Fluctuations: After the initial peak, the number of ICU patients fluctuates with several additional peaks throughout 2020 and early 2021.

- Overall Trend: While there are fluctuations, there's a general downward trend in the number of ICU patients from the peak in April 2020 to April 2021.

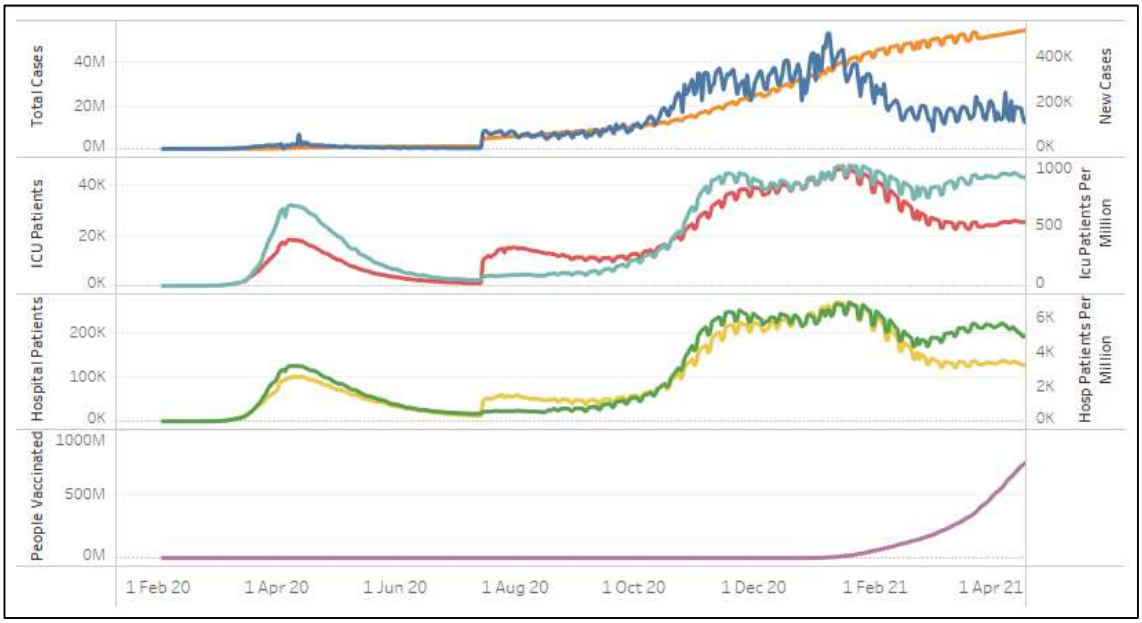
Hospital Patients:

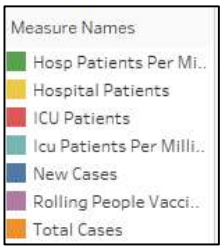
- Initial Peak: Similar to ICU patients, the number of hospital patients starts increasing rapidly in March 2020, reaching a peak in April 2020.
- Fluctuations: After the initial peak, the number of hospital patients also fluctuates with several additional peaks throughout 2020 and early 2021. These fluctuations generally align with the trends in ICU patients.
- Overall Trend: Like ICU patients, there's a general downward trend in the number of hospital patients from the peak in April 2020 to April 2021.

Conclusion:

- As we see around early April, 2020, as the new cases started to grow the ICU Patients and Hospital Patients started to grow. This trend also follows during early Aug, 2020. But as the vaccinations starts coming out in early Jan, 2021 with a increasing trend, the New cases, hospital patients and ICU patients started falling down, which would conclude the effective working of vaccinations.

	location	date	population	total_cases	new_cases	ICUPatients	icu_patients_per_million	HospitalPatients	hosp_patients_per_million
7858	United States	2021-04-24 00:00:00.000	331002647	32045113	53363	9580	28.942	37273	112.606
7859	United States	2021-04-25 00:00:00.000	331002647	32077178	32065	9545	28.837	37200	112.386
7860	United States	2021-04-26 00:00:00.000	331002647	32124869	47691	9756	29.474	37445	113.126
7861	United States	2021-04-27 00:00:00.000	331002647	32175725	50856	9682	29.251	37193	112.365
7862	United States	2021-04-28 00:00:00.000	331002647	32230850	55125	9540	28.822	36790	111.147
7863	United States	2021-04-29 00:00:00.000	331002647	32289049	58199	9315	28.142	35721	107.918
7864	United States	2021-04-30 00:00:00.000	331002647	32346971	57922	9229	27.882	35420	107.008





(The data is for North America - United States in the Dashboard.)

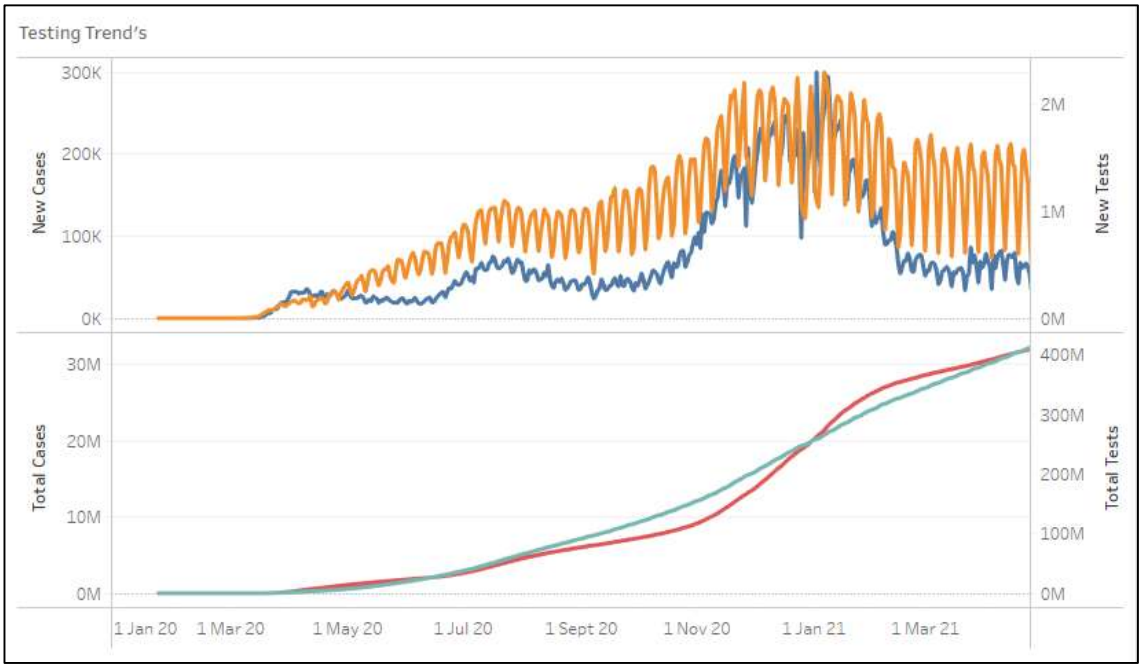
--- TESTING ANALYSIS ---

r)1) TESTING TRENDS - How has the number of new tests, new cases, total tests, total cases changed over time?

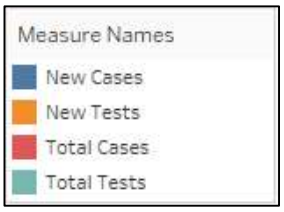
The trends in new tests and new cases appear to be somewhat correlated, suggesting that increased testing can lead to a higher detection of new cases.

The total tests and total cases steadily increase over time, reflecting the ongoing nature of the pandemic and the continuous testing efforts.

	location	date	new_tests	new_cases	total_tests	total_cases
81089	United States	2021-01-01 00:00:00.000	1236413	153628	257914...	20252991
81090	United States	2021-01-02 00:00:00.000	1102864	300310	259017...	20553301
81091	United States	2021-01-03 00:00:00.000	1036425	208746	260054...	20762047
81092	United States	2021-01-04 00:00:00.000	1386042	184282	261440...	20946329
81093	United States	2021-01-05 00:00:00.000	1944957	235111	263385...	21181440
81094	United States	2021-01-06 00:00:00.000	2309578	255444	265694...	21436884
81095	United States	2021-01-07 00:00:00.000	2242906	278290	267937...	21715174

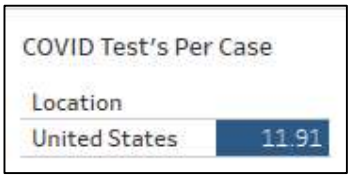


(The data is for North America since we selected United States in the Dashboard.)



r)2) TEST EFFECTIVENESS - How does the number of tests per case vary across different countries?

For every positive COVID-19 case in the US, an average of 11.91 tests were conducted. This could indicate that the US has a robust testing infrastructure and is able to identify a high proportion of cases.



(The data is for North America since we selected United States in the Dashboard.)

--- DEMOGRAPHIC AND SOCIOECONOMIC ANALYSIS ---

--- POPULATION AND DENSITY ---

s)1) How do total cases and deaths relate to population density in different countries?

Population Density vs. Total Cases:

- The total cases are distributed more evenly across different population densities.
- The trend line for total cases is relatively flat with a slight downward slope, suggesting a minimal decrease in total cases as population density increases.

Population Density vs. Total Deaths:

- There is a noticeable spread of total deaths across different population densities.
- The trend line shows a slight downward curve as population density increases, indicating that higher population densities may have a lower rate of total deaths, but the relationship is weak.

Important Considerations:

> *"This analysis is based solely on the visual representation of the data. A more rigorous statistical analysis would be required to draw definitive conclusions."*
>
> *"Other factors such as healthcare systems, government policies, and socioeconomic conditions likely play a significant role in the spread of the virus and the resulting number of cases and deaths."*

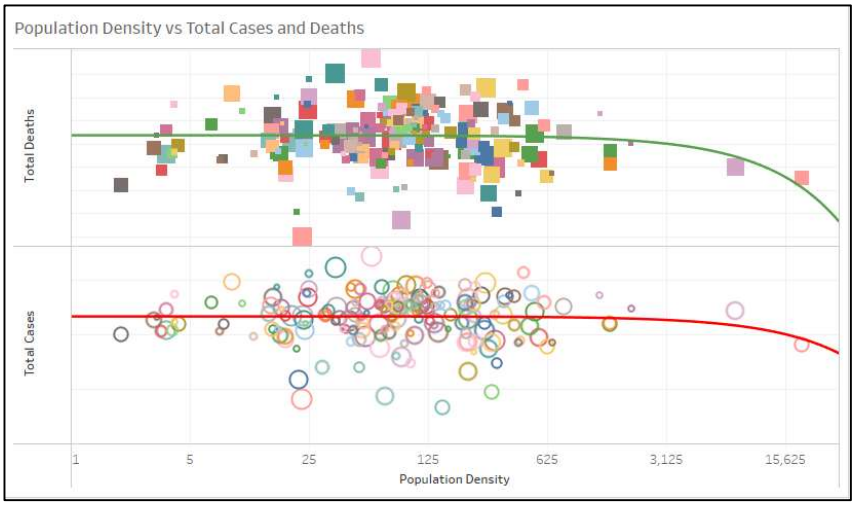
Insights:

Weak Relationship:

1. Both trend lines suggest a weak relationship between population density and the total number of cases and deaths. This indicates that population density alone may not be a strong predictor of total cases or deaths due to COVID-19.

Outliers and Spread:

2. The scatter plots show significant spread and variation, with several outliers that do not follow the general trend, highlighting the complexity and multifactorial nature of COVID-19 impacts.



	location	population_density	total_cases	total_deaths
153	United States	35.608	5094206089	106611210
154	Faeroe Islands	35.308	NULL	NULL
155	Kyrgyzstan	32.333	19789176	380028
156	Latvia	31.212	11590516	207922
157	Estonia	31.033	10166358	99611
158	Laos	29.715	15439	NULL
159	Peru	25.129	306695063	11292517

s)2) What is the impact of median age on the number of cases and deaths?

1. **Total Deaths vs. Median Age** (Top Plot):

- The scatterplot shows a positive trend, indicating that as the median age increases, the total number of deaths also tends to increase. The trendline (green) suggests a slight upward slope, further supporting this positive correlation.
- Larger squares represent higher values, and there's a noticeable concentration of larger squares at higher median ages, emphasizing the impact of an older population on the death toll.

2. **Total Cases vs. Median Age** (Bottom Plot):

- The scatterplot shows a less pronounced trend compared to deaths. However, there is still a slight positive correlation, as indicated by the red trendline.

- The circles represent the total cases, with larger circles indicating higher numbers of cases. There's a spread across various median ages, suggesting that total cases are not as strongly correlated with median age as deaths are.

Analysis:

1. **Correlation Between Median Age and Total Deaths**:

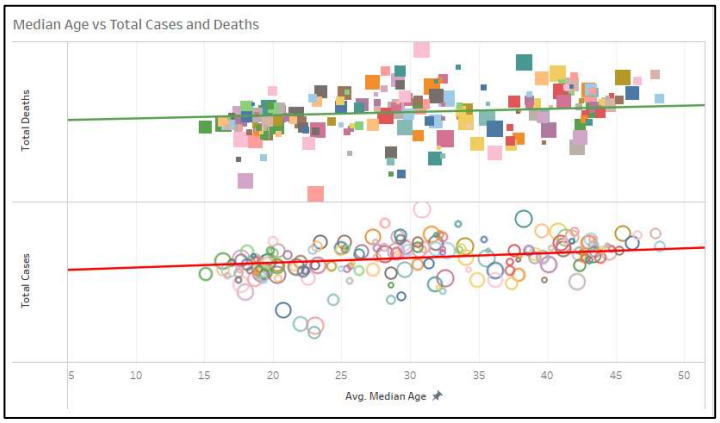
   The stronger positive correlation between median age and total deaths suggests that older populations are more susceptible to higher mortality rates from COVID-19. This is likely due to factors such as increased vulnerability to severe disease and pre-existing health conditions in older age groups.

2. **Correlation Between Median Age and Total Cases**:

   The weaker correlation between median age and total cases suggests that the spread of COVID-19 is less dependent on the age distribution of the population. Factors like population density, testing rates, and social behaviours might play more significant roles in determining the total number of cases.

3. **Implications for Public Health**:

   These insights can help in tailoring public health interventions. For instance, regions with older populations might need more robust healthcare support and targeted measures to protect vulnerable groups.



	location	median_age	total_cases	total_deaths
16	Czechia	43.3	190015352	3124948
17	Netherlan...	43.2	188195013	3655650
18	Switzerland	43.1	96217836	1877246
19	Cuba	43.1	7531481	66469
20	Romania	43	142066438	3775113
21	Finland	42.8	10498175	179019
22	Estonia	42.7	10166358	99611

--- ECONOMIC IMPACT ---

s)3) Is there a relationship between GDP per capita and the number of COVID-19 cases or deaths?

1. Total Deaths vs. GDP per capita (Top Plot):

- As GDP per capita increases, the total deaths seem to increase slightly, indicated by the positive slope of the green trend line.
- There is a significant spread in the total deaths among countries with higher GDP per capita, suggesting variability in how different countries managed the pandemic despite higher economic resources.
- Green Trend Line: For total deaths. This line shows a slight positive correlation between GDP per capita and total deaths.

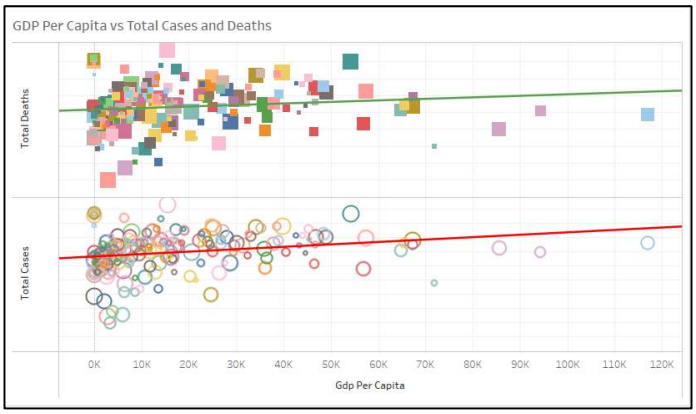
2. Total Cases vs. GDP per capita (Top Plot):
    - Similar to total deaths, there is a slight increase in total cases as GDP per capita increases, indicated by the red trend line.
    - The number of total cases shows a broader distribution for countries with lower GDP per capita, with some outliers having a high number of cases despite lower GDP.
    - Red Trend Line: For total cases. This line also shows a slight positive correlation between GDP per capita and total cases.

Outliers:

"There are several outliers where countries with high GDP per capita have either significantly high or low numbers of total cases and deaths. This suggests that factors other than economic strength also play critical roles in pandemic outcomes."

Conclusion:

"The scatter plot indicates a slight positive correlation between GDP per capita and the number of COVID-19 cases and deaths. This could imply that higher GDP per capita is associated with a greater reporting of cases and deaths, possibly due to better testing and healthcare infrastructure."

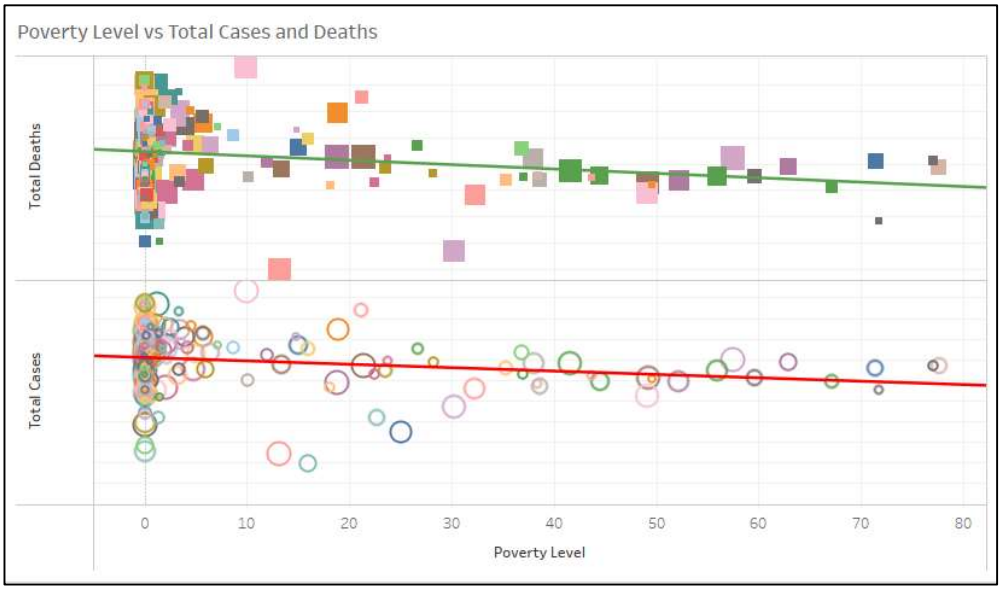


	location	gdp_per_capita	total_cases	total_deaths
22	Germany	45229.245	420309755	10869835
23	Australia	44648.71	8578785	242043
24	Canada	44017.591	155125952	4971886
25	Bahrain	43290.705	28074935	102401
26	Belgium	42658.576	149178238	5577510
27	Finland	40585.721	10498175	179019
28	United Kingdom	39753.244	653726018	25219149

s)4) How does extreme poverty influence the spread and mortality of COVID-19?

1. Correlation (Total Cases): The trend line for "Total Cases" shows a slight upward slope, suggesting a weak positive correlation between poverty level and the number of cases. This implies that areas with higher poverty levels tend to have a slightly higher number of cases, but the relationship is not very strong.
2. Correlation (Total Deaths): The trend line for "Total Deaths" also shows a slight upward slope, indicating a similarly weak positive correlation between poverty level and the number of deaths. This suggests that areas with higher poverty levels may experience a slightly higher number of deaths, but the relationship is again not very strong.

   Weak Correlation: The analysis reveals a weak positive correlation between poverty level and both the number of cases and deaths. This means that while there's a tendency for higher poverty levels to be associated with more cases and deaths, the relationship is not substantial enough to draw definitive conclusions.



	location	(No column name)	total_cases	total_deaths
1	Madagascar	77.6	5427538	79215
2	Democratic Republic of Congo	77.1	5261539	142432
3	Burundi	71.7	367466	724
4	Malawi	71.4	4243005	133874
5	Guinea-Bissau	67.1	865085	13757
6	Mozambique	62.9	7938299	81198
7	Lesotho	59.6	1392455	34944

Final conclusion for overall (S) group:

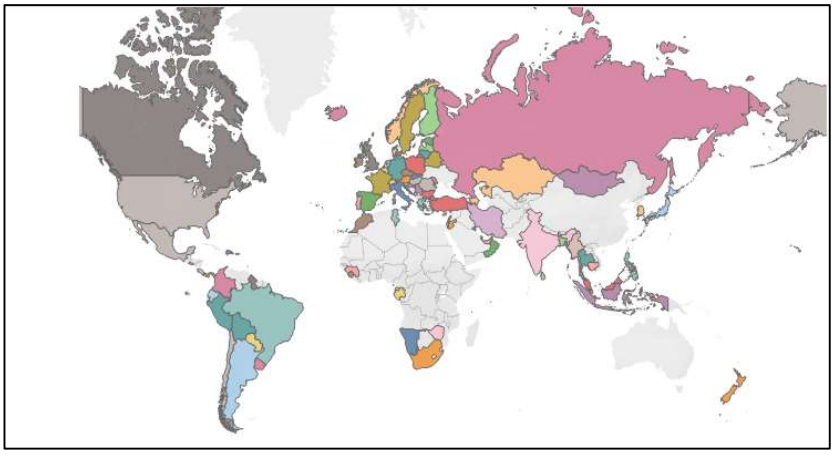
- **Weak Relationships with Single Variables:** The analyses show that no single factor—whether population density, median age, GDP per capita, or extreme poverty—strongly predicts COVID-19 cases or deaths on its own. Each factor exhibits only weak correlations with the outcomes of total cases and deaths, indicating that COVID-19 impacts are influenced by a complex interplay of multiple variables.
- **Variability and Multifactorial Nature:** There is significant variability and spread in the data for all factors examined. This highlights the multifactorial nature of the pandemic, suggesting that other elements like healthcare quality, government policies, public compliance with health measures, and social behaviours play crucial roles in determining the spread and severity of COVID-19.
- **Specific Insights:**
  - Population Density: There's a weak inverse relationship, indicating that higher population densities may slightly correlate with fewer cases and deaths, though this relationship is weak.
  - Median Age: Older populations show a stronger correlation with higher death rates, underscoring the vulnerability of elderly individuals to severe outcomes.
  - GDP per Capita: Wealthier countries tend to report more cases and deaths, possibly due to better testing and reporting infrastructure, though there are outliers.
  - Extreme Poverty: Higher poverty levels show a slight tendency towards more cases and deaths, but again, this correlation is weak.

--- FINAL IMPACT ---

u)1) Countries with more vaccinations per million than deaths per million on 30 Apr, 2021.

There was a successful vaccination effort across a diverse set of countries, highlighting the importance of vaccines in combating the COVID-19 pandemic. This positive trend underscores the need for continued international collaboration, efficient healthcare logistics, and public health strategies to ensure that vaccination rates outpace mortality rates globally.

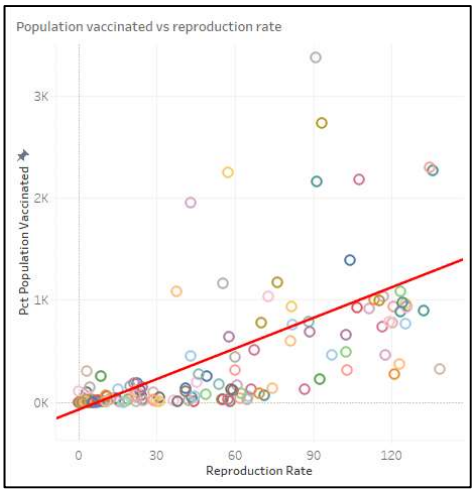
There seems to be robust efforts in all the continents, except Africa to cope up with more vaccination than deaths.



	location
1	Andorra
2	Argentina
3	Austria
4	Azerbaijan
5	Bahrain
6	Bangladesh

-- u)2) How does the percentage of the population vaccinated affect the reproduction rate?

Positive Correlation: The upward slope of the regression line suggests a positive correlation between the percentage of the population vaccinated and the reproduction rate. This indicates that as the vaccination rate increases, the reproduction rate tends to increase.



	location	date	pct_population_vaccinated	reproduction_rate
5703	United States	2021-01-12 00:00:00.000	2.81784393101847	0.99
5704	United States	2021-01-15 00:00:00.000	3.20114237636293	0.93
5705	United States	2021-01-25 00:00:00.000	5.81635197618223	0.85
5706	United States	2021-01-28 00:00:00.000	6.55541766709799	0.84
5707	United States	2021-01-31 00:00:00.000	7.61357748296194	0.83
5708	United States	2021-02-03 00:00:00.000	8.20384859339207	0.8

--- VACCINATION IMPACT ---

v)1) Is there a correlation between the number of people vaccinated and the number of new cases or new deaths?

Step 1: Calculate means (average values) for people_vaccinated and new_cases

	avg_people_vaccinated	avg_new_cases
1	8974135	5808

Step 2: Calculate Pearson correlation coefficient

	pearson_correlation_coefficient
1	0.392523059126471

The **Pearson correlation coefficient** (r) measures the strength and direction of the linear relationship between two variables. It ranges from -1 to 1, where:

- r=1: Perfect positive linear relationship
- r=−1: Perfect negative linear relationship
- r=0: No linear relationship

In the case, r=0.3925 suggests a moderate positive correlation between the number of people vaccinated and the number of new cases. This implies that, on average, as the number of people vaccinated increases, the number of new cases also increases, albeit not strongly. This suggests that there might be some other factors influencing the relationship.

Potential Factors Affecting the Correlation

1) **Vaccination Coverage:**
   - The extent of vaccination coverage within different regions or countries.
   - Vaccination might not immediately reduce new cases due to the time lag for immunity to develop.
2) **Variant Prevalence:**
   - The emergence of new variants that may affect the infection rates despite vaccination.
3) **Public Health Measures:**
   - Other measures like social distancing, mask mandates, and lockdowns that can influence new cases independent of vaccination rates.
4) **Reporting Delays:**
   - Delays in reporting new cases or vaccinations can introduce inaccuracies in the correlation analysis.

v)2) Calculate covariance and standard deviations

Step1: Calculate means (average values) for people_vaccinated and new_cases

	avg_people_vaccinated	avg_new_cases
1	8974135	5808

Step2: Calculate covariance

	covariance
1	3126857072452

Step 3: Calculate standard deviation for people_vaccinated and new_cases

	stddev_people_vaccinated	stddev_new_cases
1	43112619.7227983	64567.0981695812

Covariance measures the joint variability of two random variables. If the covariance is positive, the two variables tend to increase or decrease together. If it's negative, one variable tends to increase when the other decreases.

A **positive covariance** indicates that, on average, the number of people vaccinated and the number of new cases tend to increase together.

Interpretation:

Moderate Positive Correlation

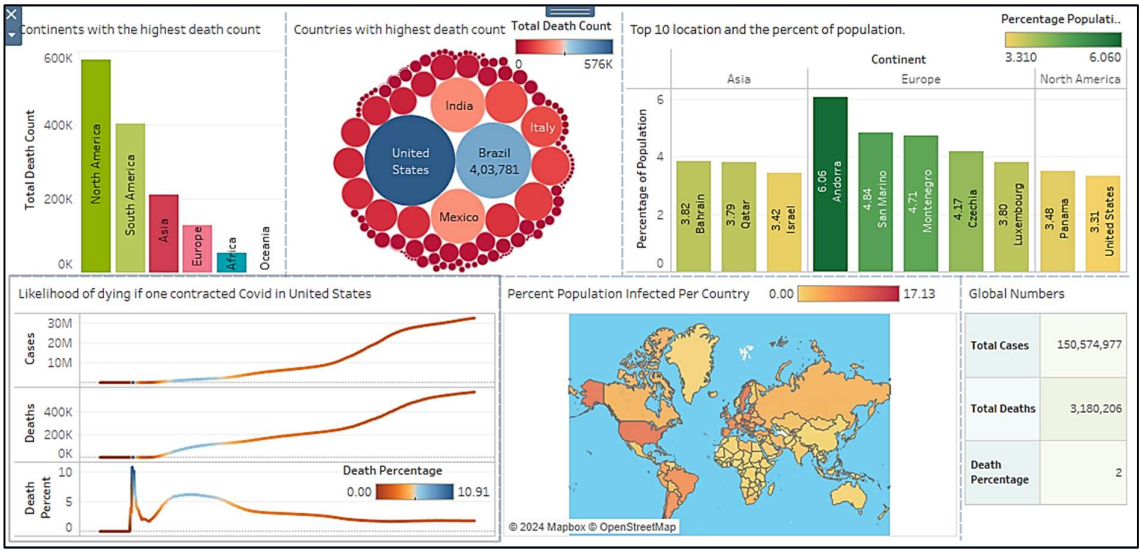
The moderate positive correlation suggests that while there is some positive relationship between the number of people vaccinated and the number of new cases, this relationship is not very strong.

Various factors could influence this correlation:

- Lag Effect: The impact of vaccinations on new cases may not be immediate. It takes time for vaccines to induce immunity.
- Variants: The emergence of new variants may influence the number of new cases despite high vaccination rates.
- Behavioural Changes: Changes in public behaviour following vaccination (e.g., reduced social distancing) might lead to an increase in new cases temporarily.
- Other Measures: Other public health measures like mask mandates, lockdowns, and travel restrictions also play a significant role in the number of new cases.

Dashboard and conclusion

1) Global Picture



Conclusion

1. **Average Percentage of Population Affected by Continent (Based on Top 10 Locations):**
    - **Europe**: 4.716%
    - **Asia**: 3.6767%
    - **North America**: 3.395%

    Observations:
    - Europe has the highest average percentage of population affected.
    - North America has the second-highest average percentage.
    - Asia has the lowest average percentage of population affected among these three continents.
2. **Mortality Rate in the United States:**
    - On April 30, 2021, the possibility of death if a person catches COVID-19 was 1.78%.
    - This percentage peaked at 6.25% on May 15, 2020.

    Observations:
    - **Cases**: The cumulative number of COVID-19 cases shows a sharp initial increase followed by a more gradual rise.
    - **Deaths**: The cumulative number of COVID-19 deaths over time mirrors the trend of cases but with a delayed and less steep increase.
    - **Death Percentage**: The percentage of COVID-19 cases resulting in death spiked during the early months of the pandemic, followed by a gradual decline.

Inferences:

- The overall mortality rate from COVID-19 in the United States has decreased over time, likely due to improved medical treatments, increased awareness and preventative measures, and the impact of vaccinations.
- The initial surge in both cases and deaths highlights the rapid spread and severe impact of the virus during the early stages of the pandemic.
- The gradual decline in the death percentage indicates that the healthcare system has become more adept at managing COVID-19 cases, leading to a lower likelihood of fatal outcomes.

3. **Top 5 Countries with Highest Infection Rates:**
   - **Andorra**: 17.13%
   - **Montenegro**: 15.51%
   - **Czechia**: 15.23%
   - **San Marino**: 14.93%
   - **Slovenia**: 11.56%

   Observations:

   - Smaller countries like Andorra and San Marino have higher infection rates compared to larger countries. This could be due to factors such as population density, healthcare systems, and testing rates.

4. **Countries with Highest Death Counts:**
   - **United States**: 576,232 deaths
   - **Brazil**: 403,781 deaths
   - **Mexico**: 216,907 deaths

   Observations:

   - The United States has the highest number of COVID-19 deaths, significantly surpassing other countries.
   - Brazil and Mexico follow at a distance with substantial, but lower, death counts.
   - India and the United Kingdom have comparable death tolls, with 211,853 and 127,775 deaths, respectively.

5. **European Countries' Impact:**
   - Among European nations, Italy has the highest death count at 120,807, followed by the United Kingdom, France, Germany, and Spain.
   - Russia has a lower death count (108,290) compared to several European nations, which could be attributed to various factors such as population density, healthcare systems, and data reporting practices.

6. **Top 3 Continents with Highest Death Counts:**

   - **Europe**: 1,016,750 deaths
   - **North America**: 847,942 deaths
   - **South America**: 672,415 deaths

   Observations:

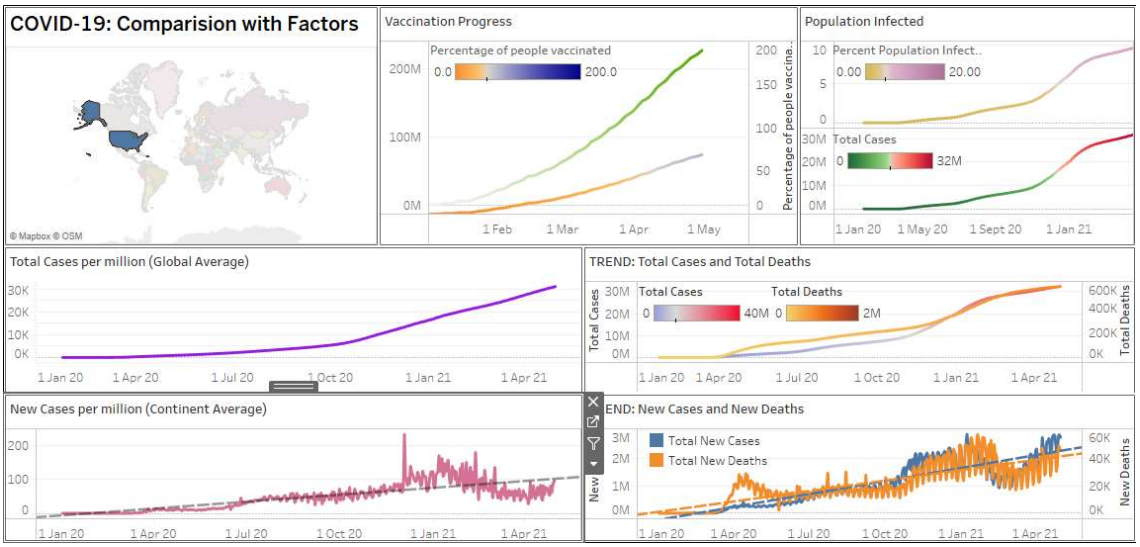
   - These three continents account for the vast majority of global COVID-19 deaths, highlighting their significant impact.

Overall Inference:

The information indicates a significant global impact of the COVID-19 pandemic with a mortality rate of 2%. Europe has been the hardest hit in terms of the percentage of the population affected and the death count, while the United States has recorded the highest number of deaths among individual countries. The decline in mortality rates over time reflects improvements in medical treatments and the effects of widespread vaccination efforts.

---

2) COVID-19 Comparison - Case Severity and Testing – 1:



Conclusion (United States)

1. **Vaccination Progress:**

   - By the end of the timeframe around May 1st, approximately 200 million people had received at least one COVID vaccine.
   - Roughly 60% of the population had received at least one COVID vaccine dose by that point.

2. **Infection Rate in the United States:**
    - On April 30, 2021, 9.77% of the population, or 32 million people, were infected with COVID-19.
3. **Trends in Total Cases and Deaths:**
    - Both total cases and total deaths exhibited a steady upward trend over the period.
    - The rate of increase for cases accelerated at some point, showing a noticeable correlation and lag between the trends of cases and deaths.
    - Before the start of 2021, total deaths were more than total cases. With the introduction of vaccinations around early 2021, the graph of total deaths fell below the graph of total cases, indicating the effectiveness of the vaccination efforts.
4. **New Cases and Deaths Trends:**

    New Cases:
    - Rapid increase from late 2020 to early 2021, peaking around January 2021.
    - General downward trend with significant fluctuations post-peak.

    New Deaths:
    - Steady increase throughout the period, with a trend following new cases but with a slight delay.
    - Continued occurrence at relatively high levels.

    Regression Analysis:

    New Cases:
    - Equation: Total New Cases = 5154.2 * Day of Date - 2.26171e+08
    - Positive slope (5154.2) indicates a strong tendency for the number of new cases to increase over time.
    - P-value < 0.0001, indicating a highly significant statistical relationship.

    New Deaths:
    - Equation: Total New Deaths = 83.8639 * Day of Date - 3.67513e+06
    - Positive slope (83.8639) indicates a strong tendency for the number of new deaths to increase over time.
    - P-value < 0.0001, indicating a highly significant statistical relationship.
5. **New Cases per Million:**
    - Sharp increase around March 2020, coinciding with the initial wave of the pandemic in the U.S.
    - Fluctuations throughout 2020 and 2021 with periods of peaks and troughs.

- General downward trend in new cases per million from early 2021 to May 2021, suggesting a gradual decrease in the pandemic's impact.

6. **Top Infection Rates:**
    - **Andorra**: 171,254 cases per million people as of April 30, 2021.
7. **Highest Death Counts by Country:**
    - **United States**: 576,232 deaths
    - **Brazil**: 403,781 deaths
    - **Mexico**: 216,907 deaths
    - **India**: 211,853 deaths
    - **United Kingdom**: 127,775 deaths

    Observations:
    - The United States has the highest number of COVID-19 deaths, significantly surpassing other countries.
    - Brazil and Mexico follow at a distance with substantial death counts.
    - India and the United Kingdom have comparable death tolls, significantly higher than other European nations like Italy, Russia, and France.
8. **Death Counts by Continent:**
    - **Europe**: 1,016,750 deaths
    - **North America**: 847,942 deaths
    - **South America**: 672,415 deaths

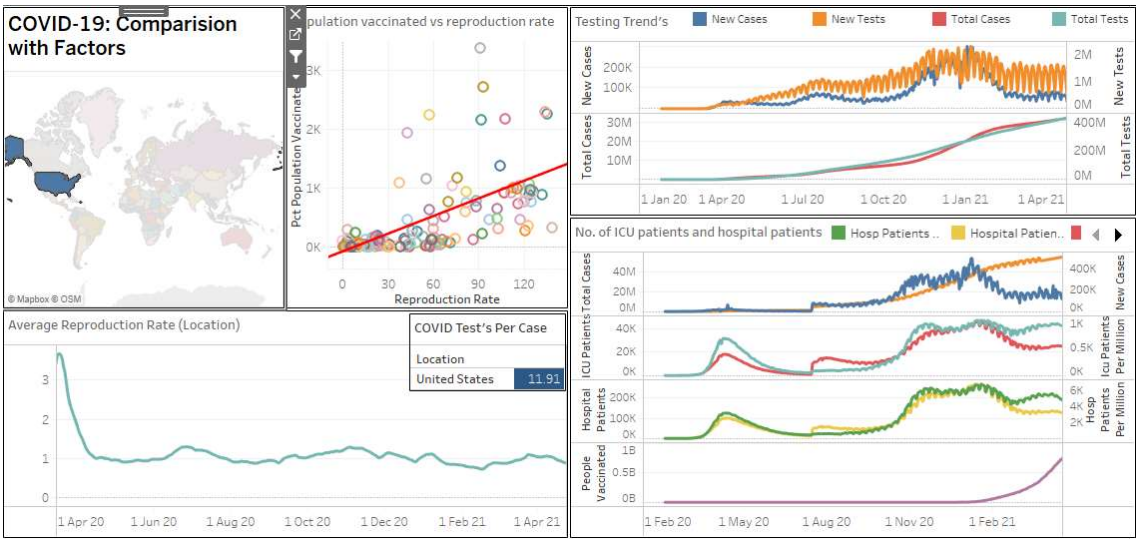
    Observations:
    - These three continents account for the vast majority of global COVID-19 deaths, highlighting their significant impact.

Overall Inference:

The information indicates a significant global impact of the COVID-19 pandemic with a mortality rate of 2%. Europe has been the hardest hit in terms of the percentage of the population affected and the death count, while the United States has recorded the highest number of deaths among individual countries. The decline in mortality rates over time reflects improvements in medical treatments and the effects of widespread vaccination efforts.

---

3) COVID-19 Comparison - Case Severity and Testing – 2:



Conclusion

1. **ICU and Hospital Patients:**

   ICU Patients:

   - Initial Peak: Rapid increase starting in March 2020, peaking in April 2020.
   - Fluctuations: Several additional peaks throughout 2020 and early 2021.
   - Overall Trend: General downward trend from the peak in April 2020 to April 2021.

   Hospital Patients:

   - Initial Peak: Rapid increase starting in March 2020, peaking in April 2020.
   - Fluctuations: Several additional peaks throughout 2020 and early 2021, aligning with ICU patient trends.
   - Overall Trend: General downward trend from the peak in April 2020 to April 2021.

   Conclusion:

   - The increase in ICU and hospital patients aligns with the rise in new cases around early April 2020 and August 2020. The introduction of vaccinations in early January 2021 coincides with a decline in new cases, hospital patients, and ICU patients, indicating the effectiveness of vaccinations.

2. **New Tests and New Cases:**

   - The trends in new tests and new cases appear correlated, suggesting increased testing leads to higher detection of new cases.
   - Total tests and total cases steadily increase over time, reflecting ongoing pandemic and continuous testing efforts.

3. **Testing Efficiency in the US:**

- For every positive COVID-19 case, an average of 11.91 tests were conducted, indicating a robust testing infrastructure and high case detection rate.

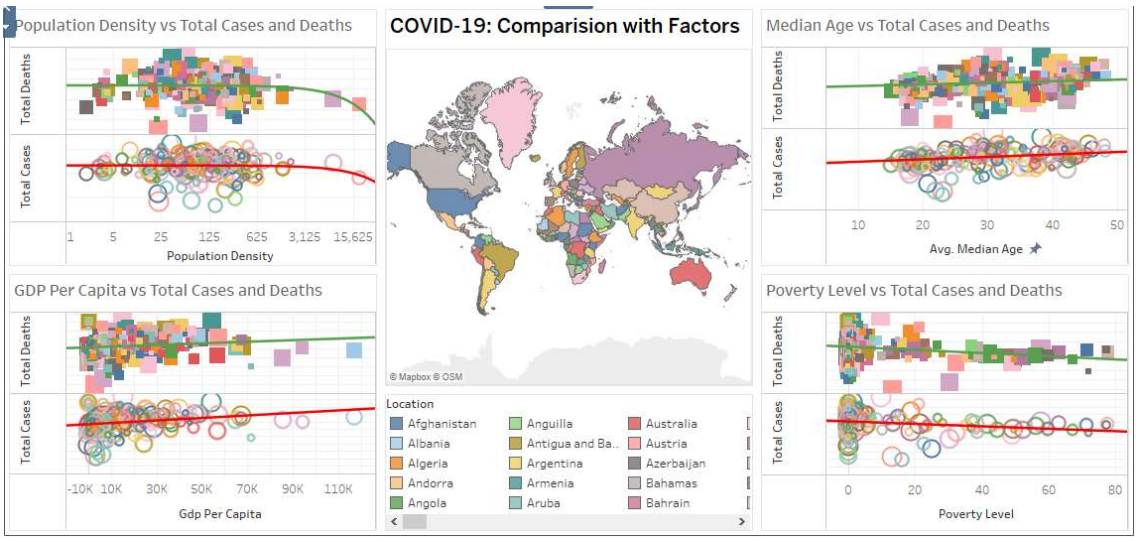
4. **Reproduction Rate (R):**
    - **Initial:** Increase at around 3.5 normal in April 2020, indicating normal reproduction rates.
    - **Sharp Decline:** Steep decline below 1.0 by June 2020, suggesting by control measures like lockdowns and social distancing.
    - **Fluctuations and Plateaus:** From June 2020 to April 2021, R fluctuates between 0.5 and 1.5, with periods of slight increases possibly due to released lockdowns, but generally remains below 1.0, indicating controlled people meetings.
5. **Vaccination and Reproduction Rate:**
    - Positive correlation between the percentage of the population vaccinated and the reproduction rate. As the vaccination rate increases, the reproduction rate tends to increase.

Overall Inference:

The data indicates that the initial rapid spread of COVID-19 was brought under control through interventions, and the introduction of vaccinations has significantly reduced the number of new cases, ICU patients, and hospital patients. The robust testing infrastructure in the US has enabled effective detection and control of the virus. The fluctuations in the reproduction rate highlight the continued need for monitoring and responding to new variants and seasonal effects.

---

4) COVID-19 – Factor comparison with total cases and deaths



Conclusion

1. **Population Density vs. Total Cases and Deaths:**

   Total Cases:

   - The total cases are distributed more evenly across different population densities.
   - The trend line for total cases shows a slight downward slope, indicating a minimal decrease in total cases as population density increases.

   Total Deaths:

   - There is a noticeable spread of total deaths across different population densities.
   - The trend line shows a slight downward curve, suggesting higher population densities may have a lower rate of total deaths, but the relationship is weak.

   ***"Important Considerations:** This analysis is based on visual representation, and more rigorous statistical analysis is needed for definitive conclusions. Other factors affect the results."*

   Insights:

   - Both trend lines suggest a weak relationship between population density and total cases and deaths, indicating that population density alone may not be a strong predictor of COVID-19 impact.
   - Significant spread and variation in the data, with several outliers, highlight the complexity and multifactorial nature of COVID-19 impacts.

2. **Impact of Median Age on Cases and Deaths:**

   Total Deaths vs. Median Age:

   - Positive trend, with total deaths increasing as median age increases.
   - Larger values concentrated at higher median ages, emphasizing the impact of an older population on the death toll.

   Total Cases vs. Median Age:

   - Slight positive correlation, with total cases increasing with median age, but less pronounced than for deaths.

   Analysis:

   - The stronger correlation between median age and total deaths suggests older populations are more susceptible to higher mortality rates from COVID-19.
   - The weaker correlation between median age and total cases suggests that the spread of COVID-19 is less dependent on the age distribution of the population.

   Implications for Public Health:

   - Regions with older populations may need more robust healthcare support and targeted measures to protect vulnerable groups.

3. **GDP per Capita and COVID-19 Cases and Deaths:**

Total Deaths vs. GDP per Capita:

- Positive slope of the trend line indicates a slight increase in total deaths with higher GDP per capita.
- Significant spread among countries with higher GDP per capita, suggesting variability in pandemic management.

Total Cases vs. GDP per Capita:

- Positive slope of the trend line indicates a slight increase in total cases with higher GDP per capita.
- Broader distribution for countries with lower GDP per capita, with some outliers.

Outliers:

- Some countries with high GDP per capita have either significantly high or low numbers of total cases and deaths, indicating other critical factors influencing pandemic outcomes.

Conclusion:

- Slight positive correlation between GDP per capita and the number of COVID-19 cases and deaths, possibly due to better testing and healthcare infrastructure in wealthier countries.

4. **Extreme Poverty and COVID-19 Spread and Mortality:**

Total Cases:

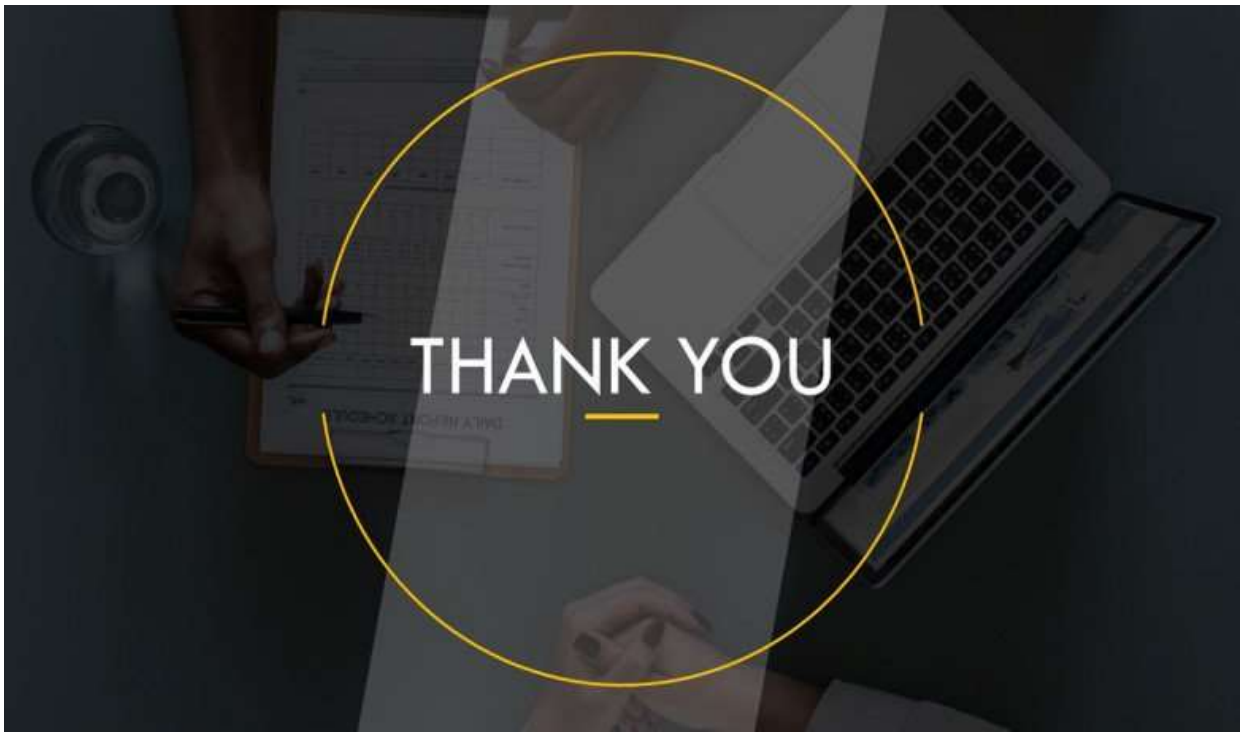
- Weak positive correlation, with areas of higher poverty levels tending to have slightly higher numbers of cases**.**

Total Deaths:

- Weak positive correlation, with areas of higher poverty levels tending to have slightly higher numbers of deaths.

Conclusion:

- The analysis reveals a weak positive correlation between poverty levels and both the number of cases and deaths, indicating that higher poverty levels may be associated with more cases and deaths, but the relationship is not substantial enough to draw definitive conclusions.


THANK YOU