COVID-19 EXPLORATORY DATA ANALYSIS

BY FOLASHADE OLAITAN

LinkedIn: https://www.linkedin.com/in/folaitan/

Data Source: https://ourworldindata.org/covid-deaths

Introduction:

The COVID-19 pandemic has swept across the globe with unprecedented speed, causing widespread disruption to daily life and significant loss of life. Since the first cases were reported in December 2019, the virus has spread to nearly every country in the world, with millions of confirmed cases and hundreds of thousands of deaths. Governments and public health officials have implemented a range of measures to slow the spread of the virus, including lockdowns, social distancing guidelines, and widespread testing and vaccination campaigns.

Exploratory data analysis (EDA) can help us to better understand the patterns and trends of COVID-19 cases, deaths, and vaccination rates around the world. By analysing data from different countries and regions, we can identify factors that may be contributing to the spread of the virus, such as population density, age distribution, and underlying health conditions. EDA can also help us to identify effective strategies for controlling the spread of the virus, such as increased testing and contact tracing, targeted vaccination campaigns, and public health messaging to encourage people to adopt behaviours that reduce transmission.

I have made use of two datasets for this exploratory analysis. The first is the deaths dataset containing COVID-19 cases and mortality data, as well as hospitalization and healthcare-related metrics for different countries.

The second is the vaccination dataset containing data related to COVID-19 testing and vaccination, as well as various demographic, economic, and health-related factors for the different countries. The datasets contain data from January 2020 -March 15, 2023 (see appendix for the data dictionary).

Caveats:

- Data collection, cleaning and data quality control is continuous, please interpret with caution.
- Completeness for indicators vary between countries, region and overtime.
- Case identification is subject to detection, testing and reporting strategies which vary from country to country.
- It is important to be cautious when interpreting and comparing COVID-19 data across different countries, as testing and reporting strategies may vary widely. It is recommended to consult multiple sources and experts in the field to gain a more nuanced understanding of the data.

GLOBAL ANALYSIS

1. ----What is the world's total confirmed cases, total deaths and what proportion of the infected cases DIED?

```
SELECT SUM(new_cases) AS GlobalConfirmedCases,
              SUM(CAST(new_deaths AS INT)) AS GlobalConfirmedDeaths,
              SUM(new_cases)- SUM(CAST(new_deaths AS INT)) AS ConfirmedSurvivors,
              SUM(CAST(new_deaths AS INT))/SUM(new_cases) * 100 AS DeathRate
FROM CovidAnalysis..Deaths$
WHERE continent is not null
ORDER BY DeathRate
```

	GlobalConfirmedCases	GlobalConfirmedDeaths	ConfirmedSurvivors	DeathRate
1	760384896	6876500	753508396	0.904

The total **confirmed cases worldwide** as of March 15, 2023 is about **760million** cases with almost **6.9million deaths** which **is 0.90%** of the total confirmed cases.

2. ----What percentage of the world's population had covid?

```
SELECT SUM(population)AS World_Population, SUM(new_cases) AS Total_Cases, SUM(new_cases)/SUM(population)*100 AS
Percentage_Infected
FROM CovidAnalysis..Deaths$
WHERE continent is not null
ORDER BY Percentage_Infected DESC
```

	World_Population	Total_Cases	Percentage_Infected
1	9398206703706	760384896	0.008

Approximately **0.01%** of the total world population has gotten infected by the COVID-19 virus from January 2020 till March 15, 2023.

3. ----How many people in the whole world have been vaccinated? Fully vaccinated? and gotten a booster dose?

```
SELECT
       SUM(CAST(new_vaccinations AS BIGINT)) AS Total_Vaccinations,
       MAX(CAST(people_vaccinated AS BIGINT)) AS People_Vaccinated_Atleast_Once,
       MAX(CAST(people_fully_vaccinated AS BIGINT)) AS FullyVaccinated,
       MAX(CAST(total_boosters AS BIGINT)) AS People_Boosted
FROM CovidAnalysis..Vaccinations$
WHERE continent is not null
```

	Total_Vaccinations	People_Vaccinated_Atleast_Once	FullyVaccinated	People_Boosted
1	10849394955	1310292000	1276760000	826913000

As of March 15, 2023 **10,849,394,955** total **vaccines** have been **administered** worldwide out of which only **1,276,760,000** have been **fully vaccinated** and **826,913,000** have gotten a **boosted** dose.

CONTINENTAL ANALYSIS

1. ----What continent has the highest death count and rate?

```
SELECT continent AS Continent, SUM(new_cases) AS ConfirmedCases,
       SUM(CAST(new_deaths AS INT)) AS ConfirmedDeaths,
       ROUND(SUM(CAST(new_deaths AS INT))/SUM(new_cases) * 100,2) AS DeathRate
FROM CovidAnalysis..Deaths$
WHERE continent is not null
GROUP BY continent
ORDER BY DeathRate DESC
```

	Continent	ConfirmedCases	ConfirmedDeaths	DeathRate
1	South America	68040323	1351566	1.99
2	Africa	13064302	258788	1.98
3	North America	122926681	1588153	1.29
4	Europe	247189266	2026663	0.82
5	Asia	295110299	1625978	0.55
6	Oceania	14054025	25352	0.18

Asia recorded the **highest confirmed COVID-19 cases** from January 2020 up until March 15, 2023 recording a little over **295million** cases. **Europe** has the **second highest** confirmed cases with approximately **247million** recorded cases.

However, the highest death rate was recorded in South America with approximately 2% of the people infected died. **Africa** has the **second highest** confirmed **deaths rate** at **1.98%** shamefully, as it had the **lowest number of confirmed cases**.

Oceania has the **lowest death rate** even though their recorded confirmed cases was a little higher than Africa's.

2. ----Which continents have the highest vaccination count?

```
SELECT continent AS Continent,
       SUM(CAST(new_vaccinations AS BIGINT)) AS Total_Vaccinations_Administered,
       MAX(CAST(people_fully_vaccinated AS BIGINT)) AS Fully_Vaccinated
FROM CovidAnalysis..Vaccinations$
WHERE continent is not null
GROUP BY continent
ORDER BY Total_Vaccinations_Administered DESC
```

	Continent	Total_Vaccinations_Administered	Fully_Vaccinated
1	Asia	7546077630	1276760000
2	Europe	1277379995	79582290
3	North America	992924109	230142115
4	South America	841170588	176110664
5	Africa	117367074	66717081
6	Oceania	74475559	21654488

Asia has the highest vaccination counts with total vaccinations of **7,546,077,630** as of March 15, 2023. This is not surprising because they have recorded a low death rate even though they have the highest number of confirmed cases.

Interestingly, the continents with the highest number of confirmed cases also have the highest vaccination counts respectively in the same order. This is seen to have a positive correlation on their death rates as well.

3. ----See death counts and change overtime (2020-2022)

```
SELECT  CD2020.Continent, CD2020.TotalDeaths_2020, CD2021.TotalDeaths_2021,
       ROUND(((CD2021.TotalDeaths_2021 - CD2020.TotalDeaths_2020)/CD2020.TotalDeaths_2020)*100,2) AS
Percentage_Change_Btw_2020_and_2021,
       CD2022.TotalDeaths_2022,
       ROUND(((CD2022.TotalDeaths_2022 - CD2021.TotalDeaths_2021)/CD2021.TotalDeaths_2021)*100,2) AS
Percentage_Change_Btw_2021_and_2022
FROM #ContinentalDeaths2020 CD2020
JOIN #ContinentalDeaths2021 CD2021
       ON CD2020.Continent = CD2021.Continent
JOIN #ContinentalDeaths2022 CD2022
       ON CD2020.Continent = CD2022.Continent
ORDER BY CD2020.Continent
```

	Continent	TotalDeaths_2020	TotalDeaths_2021	Percentage_Change_Btw_2020_and_2021	TotalDeaths_2022	Percentage_Change_Btw_2021_and_2022
1	Africa	65057	163826	151.82	29629	-81.91
2	Asia	337924	920940	172.53	271304	-70.54
3	Europe	568361	972508	71.11	452666	-53.45
4	North America	536069	688836	28.5	327600	-52.44
5	Oceania	1193	3791	217.77	18485	387.6
6	South America	420438	772492	83.74	149226	-80.68

4. ----See vaccination counts and change overtime (2020-2022)

```
SELECT  CV2020.Continent, CV2020.Total_Administered_Vaccines_2020, CV2021.Total_Administered_Vaccines_2021,
             ROUND(((CV2021.Total_Administered_Vaccines_2021 -
CV2020.Total_Administered_Vaccines_2020)/CV2020.Total_Administered_Vaccines_2020)*100,5) AS
Percentage_Change_Btw_2020_and_2021,
             CV2022.Total_Administered_Vaccines_2022,
```

```
            ROUND(((CV2022.Total_Administered_Vaccines_2022 -
CV2021.Total_Administered_Vaccines_2021)/CV2021.Total_Administered_Vaccines_2021)*100,5) AS
Percentage_Change_Btw_2021_and_2022
FROM #ContinentalVaccines2020 CV2020
JOIN #ContinentalVaccines2021 CV2021
       ON CV2020.Continent = CV2021.Continent
JOIN #ContinentalVaccines2022 CV2022
       ON CV2020.Continent = CV2022.Continent
ORDER BY CV2020.Continent
```

	Continent	Total_Administered_Vaccines_2020	Total_Administered_Vaccines_2021	Percentage_Change_Btw_2020_and_2021	Total_Administered_Vaccines_2022	Percentage_Change_Btw_2021_
1	Africa	NULL	82606042	NULL	34759190	0
2	Asia	1016759	5545983589	545300	1963185034	0
3	Europe	345764	1050008034	303500	224195489	0
4	North America	5780392	763574155	13100	216013250	0
5	Oceania	NULL	50651364	NULL	23794705	0
6	South America	31135	592763655	1903700	241599665	0

Looking at the death counts per continent overtime, we can see that ALL continents have recorded an increase in the number of deaths between the year 2020 and year 2021.

However, between the year 2021 and year 2022, there has been a massive decline in the recorded death cases. This is clearly due to the introduction and massive adoption of the COVID vaccines (as seen in the second result screenshot). There was an impressive adoption of the vaccines in 2021 when it was new introduced and this has helped to reduce the negative health impact of the COVID-19 virus on those infected, thereby resulting in lower death rates since then. As we have continue to have more people get vaccinated, we will continue to see a decline in the number of deaths recorded world-wide.

DEEP DIVE INTO AFRICA

I am an African and I am very proud of my heritage 😍😍😜😜😜, so in this next section, I am going to have a deep dive into the African continent to gain insights from the data of each country that makes up the continent.

1. Top 5 countries with the highest recorded cases and death rate. What % do they make of the total Africa confirmed cases?

```
--- Top 5 African countries by confirmed cases
	SELECT TOP 5 location AS Country,
		SUM(new_cases) AS Total_Confirmed_Cases,
		SUM(new_deaths) AS Confirmed_Deaths,
		ROUND((SUM(new_deaths)/SUM(new_cases)) * 100, 2) AS Death_Rate
	FROM CovidAnalysis..Deaths$ dea
	WHERE continent ='Africa'
	GROUP BY location
	ORDER BY Total_Confirmed_Cases DESC
```

	Country	Total_Confirmed_Cases	Confirmed_Deaths	Death_Rate
1	South Africa	4068224	102595	2.52
2	Morocco	1272526	16296	1.28
3	Tunisia	1151333	29345	2.55
4	Egypt	515792	24815	4.81
5	Libya	507201	6437	1.27

```
--- Top 5 African countries by death rate
	SELECT TOP 5 location AS Country,
		SUM(new_cases) AS Total_Confirmed_Cases,
		SUM(new_deaths) AS Confirmed_Deaths,
		ROUND((SUM(new_deaths)/SUM(new_cases)) * 100, 2) AS Death_Rate
	FROM CovidAnalysis..Deaths$ dea
	WHERE continent ='Africa'
	GROUP BY location
	ORDER BY Death_Rate DESC
```

	Country	Total_Confirmed_Cases	Confirmed_Deaths	Death_Rate
1	Sudan	63853	5023	7.87
2	Somalia	27324	1361	4.98
3	Egypt	515792	24815	4.81
4	Liberia	8091	295	3.65
5	Niger	9513	315	3.31

2. Top 5 major contributing countries to Africa's number of confirmed cases

```
SELECT TOP 5 dea.location AS Country,
       SUM(dea.new_cases) AS Total_Confirmed_Cases,
       SUM(dea.new_deaths) AS Confirmed_Deaths,
       ROUND((SUM(dea.new_deaths)/SUM(dea.new_cases)) * 100, 2) AS Death_Rate,
       Africa_Confirmed_Cases,
       ROUND((SUM(dea.new_cases) / afr.Africa_Confirmed_Cases) *100, 2) AS Percentage_Case_Contribution_to_Africa
FROM CovidAnalysis..Deaths$ dea
JOIN #Africa afr
       ON dea.continent = afr.continent
WHERE dea.continent is not null
GROUP BY location, Africa_Confirmed_Cases
ORDER BY Percentage_Case_Contribution_to_Africa DESC
```

	Country	Total_Confirmed_Cases	Confirmed_Deaths	Death_Rate	Africa_Confirmed_Cases	Percentage_Case_Contribution_to_Africa
1	South Africa	4068224	102595	2.52	13064302	31.14
2	Morocco	1272526	16296	1.28	13064302	9.74
3	Tunisia	1151333	29345	2.55	13064302	8.81
4	Egypt	515792	24815	4.81	13064302	3.95
5	Libya	507201	6437	1.27	13064302	3.88

The **5 countries** with the **highest** number of **confirmed cases** (**South Africa, Morocco, Tunisia, Egypt and** Libya in that order) are the major contributor to the total African confirmed number of cases, with **South Africa** contributing **31.14%,** followed by **Morocco** with **9.74%** and **Tunisia** with **8.81%** confirmed cases.

Though they have the highest number of cases, they are NOT the ones with the highest **death rates**. **Sudan (7.87%), Somalia (4.98%), Egypt (4.81%), Liberia (3.65%)** and **Niger (3.31%)** have the highest death rates but little contributions to the total African cases as seen in the below result.

3. What is the vaccination rate of each African country? This is the percentage of the population that has gotten ATLEAST one dose of vaccine.

```
WITH Vaccinations AS (
	SELECT
		location AS Country,
		SUM(CAST(new_vaccinations AS BIGINT)) AS Total_Vaccinations
	FROM
		CovidAnalysis..Vaccinations$
	WHERE
		continent = 'Africa'
	GROUP BY
		location
),
Populations AS (
	SELECT
		location AS Country,
		population
	FROM
		CovidAnalysis..Deaths$
	WHERE
		continent = 'Africa'
)
SELECT
	v.Country,
	v.Total_Vaccinations,
	ROUND((CAST(v.Total_Vaccinations AS FLOAT) / p.population) * 100, 2) AS Vaccination_Rate
FROM
	Vaccinations v
	JOIN Populations p ON v.Country = p.Country
GROUP BY v.Country, v.Total_Vaccinations, p.population
ORDER BY Vaccination_Rate DESC
```

Country	Total_Vaccinations	Vaccination_Rate
Morocco	27409359	73.17
Tunisia	8546870	69.17
Zimbabwe	10801392	66.18
Seychelles	42863	40.01
South Africa	20014831	33.42
Kenya	11113785	20.57
Ethiopia	23047615	18.68
Zambia	2517495	12.58
Guinea	1586407	11.45
Eswatini	110765	9.22
Namibia	234709	9.14
Cote d'Ivoire	2278710	8.09
Rwanda	920644	6.68
Malawi	1256993	6.16
Libya	200655	2.95
Mauritius	35153	2.71
Egypt	2157499	1.94
Senegal	330137	1.91
Uganda	704976	1.49
Nigeria	3160328	1.45
Cape Verde	7462	1.26
Djibouti	12179	1.09
Mozambique	295812	0.9
Ghana	180950	0.54
Mauritania	24659	0.52
Botswana	10435	0.4
Algeria	170786	0.38
Central African Republic	12887	0.23
Somalia	37292	0.21

Liberia	10089	0.19
Sudan	45572	0.1
Sierra Leone	7981	0.09
Guinea-Bissau	1658	0.08
South Sudan	7489	0.07
Gabon	1486	0.06
Madagascar	13544	0.05
Equatorial Guinea	639	0.04
Democratic Republic of Congo	42303	0.04
Cameroon	10290	0.04
Congo	1486	0.02
Gambia	429	0.02
Burundi	460	0
Eritrea	NULL	NULL
Chad	NULL	NULL
Comoros	NULL	NULL
Angola	NULL	NULL
Benin	NULL	NULL
Burkina Faso	NULL	NULL
Lesotho	NULL	NULL
Mali	NULL	NULL
Mayotte	NULL	NULL
Tanzania	NULL	NULL
Togo	NULL	NULL
Niger	NULL	NULL
Reunion	NULL	NULL
Saint Helena	NULL	NULL
Sao Tome and Principe	NULL	NULL
Western Sahara	NULL	NULL

Morocco, Tunisia, Zimbabwe are the 3 countries having over **50%** of their populations vaccinated with at least a dose of vaccine as off March 15, 2023. While **Seychelle** and **South Africa** have over **30%** already vaccinated. This is evident in their death rates (it's not high in relation to each country's population) as they have been able to manage the impact of the virus to an extent

4. What % of each countries population had COVID? What % died of COVID and what % have been vaccinated?

```
WITH Vaccinations AS (
	SELECT
		location AS Country,
		SUM(CAST(new_vaccinations AS BIGINT)) AS Total_Vaccinations
	FROM
		CovidAnalysis..Vaccinations$
	WHERE
		continent = 'Africa'
	GROUP BY
		location
),
Populations AS (
	SELECT
		location AS Country,
		SUM(new_cases) AS Total_Cases,
		SUM(new_deaths) AS Total_Deaths,
		MAX(population) AS population
	FROM
		CovidAnalysis..Deaths$
	WHERE
		continent = 'Africa'
	GROUP BY
		location
)
SELECT
	v.Country,
```

```
        p.population,
        p.Total_Cases,
        v.Total_Vaccinations,
        p.Total_Deaths,
        ROUND((CAST(p.Total_Cases AS FLOAT) / p.population) * 100, 2) AS had_COVID,
        ROUND((CAST(p.Total_Deaths AS FLOAT) / p.population) * 100, 2) AS Died_of_COVID,
        ROUND((CAST(v.Total_Vaccinations AS FLOAT) / p.population) * 100, 2) AS Pop_Vaccinated
FROM
        Vaccinations v
        JOIN Populations p ON v.Country = p.Country
WHERE
        p.population > 0 AND v.Total_Vaccinations > 0
ORDER BY
        had_COVID DESC, Died_of_COVID DESC, Pop_Vaccinated DESC
```

Country	Population	Total_Cases	Total_Vaccinations	Total_Deaths	had_COVID	Died_of_COVID	Pop_Vaccinated
Seychelles	107135	50937	42863	172	47.54	0.16	40.01
Mauritius	1299478	296042	35153	1044	22.78	0.08	2.71
Botswana	2630300	329769	10435	2795	12.54	0.11	0.4
Cape Verde	593162	63245	7462	414	10.66	0.07	1.26
Tunisia	12356116	1151333	8546870	29345	9.32	0.24	69.17
Libya	6812344	507201	200655	6437	7.45	0.09	2.95
South Africa	59893884	4068224	20014831	102595	6.79	0.17	33.42
Namibia	2567024	171208	234709	4090	6.67	0.16	9.14
Eswatini	1201680	74323	110765	1425	6.18	0.12	9.22
Morocco	37457976	1272526	27409359	16296	3.4	0.04	73.17
Gabon	2388997	48981	1486	306	2.05	0.01	0.06
Zambia	20017670	343250	2517495	4057	1.71	0.02	12.58
Zimbabwe	16320539	264391	10801392	5672	1.62	0.03	66.18
Djibouti	1120851	15690	12179	189	1.4	0.02	1.09
Mauritania	4736146	63494	24659	997	1.34	0.02	0.52
Equatorial Guinea	1674916	17130	639	183	1.02	0.01	0.04
Rwanda	13776702	133194	920644	1468	0.97	0.01	6.68
Mozambique	32969520	233214	295812	2242	0.71	0.01	0.9

Kenya	54027484	342943	11113785	5688	0.63	0.01	20.57
Algeria	44903228	271522	170786	6881	0.6	0.02	0.38
Senegal	17316452	88933	330137	1971	0.51	0.01	1.91
Ghana	33475870	171281	180950	1462	0.51	0	0.54
Gambia	2705995	12598	429	372	0.47	0.01	0.02
Egypt	110990096	515792	2157499	24815	0.46	0.02	1.94
Cameroon	27914542	124605	10290	1966	0.45	0.01	0.04
Malawi	20405318	88710	1256993	2686	0.43	0.01	6.16
Guinea-Bissau	2105580	8960	1658	176	0.43	0.01	0.08
Congo	5970430	25110	1486	388	0.42	0.01	0.02
Burundi	12889583	53661	460	15	0.42	0	0
Ethiopia	123379928	500169	23047615	7572	0.41	0.01	18.68
Uganda	47249588	170553	704976	3630	0.36	0.01	1.49
Cote d'Ivoire	28160548	88277	2278710	834	0.31	0	8.09
Guinea	13859349	38280	1586407	467	0.28	0	11.45
Central African Republic	5579148	15367	12887	113	0.28	0	0.23
Madagascar	29611718	67941	13544	1423	0.23	0	0.05
South Sudan	10913172	18368	7489	138	0.17	0	0.07
Somalia	17597508	27324	37292	1361	0.16	0.01	0.21
Liberia	5302690	8091	10089	295	0.15	0.01	0.19
Sudan	46874200	63853	45572	5023	0.14	0.01	0.1
Nigeria	218541216	266641	3160328	3155	0.12	0	1.45
Democratic Republic of Congo	99010216	95814	42303	1464	0.1	0	0.04
Sierra Leone	8605723	7760	7981	126	0.09	0	0.09

Seychelles is on the top ladder having **47.54%** of their entire population **contracting COVID-19.** This means that there is almost a 50% chance that anyone in the country at that time will contract covid. However, a big applause to them for having just **0.16% died of Covid**, this may mean that they had effective testing and recovery strategies in place. **About 40% of their population is also currently vaccinated.**

Mauritius, Botswana and Cape Verde have all done greatly as well in being **able to manage deaths due to covid**, they however need to **improve on the number of or percentage of their citizens that have been vaccinated**.

5. What is the average life expectancy of each country? Which 3 countries have the highest life expectancy and how does this compare to their death rate?

```
SELECT
    d.location AS Country,
    AVG(CAST(v.life_expectancy AS FLOAT)) AS Average_Life_Expectancy,
    SUM(CAST(d.new_deaths AS BIGINT)) AS Total_Deaths,
    ROUND(SUM(CAST(d.new_deaths AS FLOAT)) / SUM(CAST(d.new_cases AS FLOAT)) * 100, 2) AS Death_Rate
FROM
    CovidAnalysis..Deaths$ d
    JOIN CovidAnalysis..Vaccinations$ v ON d.location = v.location
WHERE
    d.continent = 'Africa'
GROUP BY
    d.location
ORDER BY
    Average_Life_Expectancy DESC, Death_Rate DESC
```

Country	Average_Life_Expectancy	Total_Deaths	Death_Rate
Saint Helena	80.56	0	0
Reunion	80.48	1081325	0.19
Mayotte	79.46	219772	0.44
Algeria	76.88	8043889	2.53
Tunisia	76.7	34304305	2.55
Morocco	76.68	19050024	1.28
Mauritius	74.99	1220436	0.35
Seychelles	73.4	201068	0.34
Cape Verde	72.98	483966	0.65
Libya	72.91	7524853	1.27

Egypt	71.99	29008735	4.81
Sao Tome and Principe	70.39	90013	1.23
Western Sahara	70.26	NULL	NULL
Botswana	69.59	3267355	0.85
Rwanda	69.02	1716092	1.1
Senegal	67.94	2304099	2.22
Djibouti	67.11	220941	1.2
Madagascar	67.04	1663487	2.09
Kenya	66.7	6649272	1.66
Ethiopia	66.6	8851668	1.51
Gabon	66.47	357714	0.62
Eritrea	66.32	120407	1.01
Tanzania	65.46	988974	1.97
Sudan	65.31	5871887	7.87
Mauritania	64.92	1165493	1.57
Congo	64.57	453572	1.55
Comoros	64.32	188209	1.78
Malawi	64.26	3139934	3.03
South Africa	64.13	119933555	2.52
Liberia	64.1	344855	3.65
Ghana	64.07	1709078	0.85
Zambia	63.89	4742633	1.18
Namibia	63.71	4781210	2.39
Uganda	63.37	4243470	2.13
Niger	62.42	368235	3.31
Gambia	62.05	434868	2.95
Benin	61.77	190547	0.58
Guinea	61.6	545923	1.22
Burkina Faso	61.58	462924	1.8
Burundi	61.58	17535	0.03

Zimbabwe	61.49	6630568	2.15
Angola	61.15	2259677	1.84
Togo	61.04	339010	0.74
Mozambique	60.85	2620898	0.96
Democratic Republic of Congo	60.68	1711416	1.53
Eswatini	60.19	1665825	1.92
Mali	59.31	868567	2.25
Cameroon	59.29	2298254	1.58
Equatorial Guinea	58.74	213927	1.07
Guinea-Bissau	58.32	205744	1.96
South Sudan	57.85	161322	0.75
Cote d'Ivoire	57.78	974946	0.94
Somalia	57.4	1591009	4.98
Sierra Leone	54.7	147294	1.62
Nigeria	54.69	3688195	1.18
Lesotho	54.33	825314	2.05
Chad	54.24	226786	2.53
Central African Republic	53.28	132097	0.74

Mortality rate and life expectancy are related, as they both provide information about health outcomes in a population. Life expectancy is a measure of the average number of years a person is expected to live in a given population, and it is influenced by many factors including genetics, lifestyle, and access to healthcare. Mortality rate, on the other hand, is a measure of the number of deaths in a population, usually expressed as a rate per 1,000 or 100,000 people. Mortality rate can be influenced by factors such as age, sex, underlying health conditions, and access to healthcare.

In general, countries with higher life expectancies tend to have lower mortality rates, as these countries typically have better healthcare systems, higher standards of living, and lower rates of disease and injury and vice versa.

However, it's important to note that mortality rates and life expectancy are complex measures that are influenced by many factors. Overall, mortality rate and life expectancy are important measures of population health and can provide insight into the overall well-being and healthcare needs of a population.

Assuming the general rule holds and all things being equal, **Saint Helena**, **Reunion** and **Mayotte** are the 3 countries with the **highest life expectancies** and they have also recorded **low death rates** on their COVID-19 cases.

Looking through, we can see that countries with average **life expectancy > 70** all have a **considerably low death rate**. The only exception to this is **Egypt** with **death rate of 4.81%** and **average life expectancy of 71.99.**

6. What demographic and health related factors have contributed to a country's COVID-19 outcomes?

```
SELECT
    d.location AS Country,
    v.median_age AS Median_Age,
    v.diabetes_prevalence AS Diabetes_Prevalence,
    d.total_cases AS Total_Cases,
    d.total_deaths AS Total_Deaths,
    v.total_tests AS Total_Tests,
    v.total_vaccinations AS Total_Vaccinations
FROM
    CovidAnalysis..Deaths$ d
JOIN
    CovidAnalysis..Vaccinations$ v
ON
    d.location = v.location
WHERE d.continent = 'Africa'
GROUP BY d.location,  v.median_age, v.diabetes_prevalence,d.total_cases,d.total_deaths,v.total_tests,v.total_vaccinations
ORDER BY Median_Age DESC, Diabetes_Prevalence DESC
```

Countries with higher median ages or higher rates of diabetes may be at higher risk for severe outcomes.

7. The reproduction rate data can provide insights into the rate at which the virus is spreading in different countries. A reproduction rate greater than 1 indicates that the virus is spreading rapidly, while a rate less 1 suggests that the virus is being contained.

```
SELECT
       location AS Country, population AS Population,
       ROUND((SUM(new_cases)/population) * 100,2) AS Perc_Pop_With_COVID,
       MAX(CAST(reproduction_rate AS FLOAT)) AS Reproduction_Rate,
       SUM(new_cases) AS Confirmed_Cases
FROM
       CovidAnalysis..Deaths$
WHERE continent= 'Africa'
GROUP BY location,population
ORDER BY Reproduction_Rate DESC,
         Perc_Pop_With_COVID DESC,
         Confirmed_Cases DESC;
```

Country	Population	Perc_Pop_With_COVID	Reproduction_Rate	Confirmed_Cases
Eswatini	1201680	6.18	4	74323
Cape Verde	593162	10.66	2.95	63245
Angola	35588996	0.3	2.94	105298
Ethiopia	123379928	0.41	2.91	500169
Kenya	54027484	0.63	2.87	342943
Mauritania	4736146	1.34	2.86	63494
Zimbabwe	16320539	1.62	2.63	264391
Mozambique	32969520	0.71	2.62	233214
Malawi	20405318	0.43	2.53	88710
Zambia	20017670	1.71	2.43	343250
Rwanda	13776702	0.97	2.41	133194
Morocco	37457976	3.4	2.28	1272526
Uganda	47249588	0.36	2.26	170553
Cote d'Ivoire	28160548	0.31	2.25	88277
Sao Tome and Principe	227393	2.76	2.15	6281

Senegal	17316452	0.51	2.13	88933
South Africa	59893884	6.79	2.08	4068224
Djibouti	1120851	1.4	2.06	15690
Togo	8848700	0.45	2.02	39407
Mali	22593598	0.15	2.02	33067
Sierra Leone	8605723	0.09	1.86	7760
Algeria	44903228	0.6	1.79	271522
Liberia	5302690	0.15	1.74	8091
Tunisia	12356116	9.32	1.68	1151333
Egypt	110990096	0.46	1.62	515792
Libya	6812344	7.45	1.61	507201
South Sudan	10913172	0.17	1.59	18368
Mauritius	1299478	22.78	1.56	296042
Madagascar	29611718	0.23	1.55	67941
Namibia	2567024	6.67	1.54	171208
Burkina Faso	22673764	0.1	1.54	22056
Ghana	33475870	0.51	1.52	171281
Nigeria	218541216	0.12	1.52	266641
Niger	26207982	0.04	1.52	9513
Lesotho	2305826	1.5	1.5	34490
Gabon	2388997	2.05	1.46	48981
Botswana	2630300	12.54	1.43	329769
Sudan	46874200	0.14	1.42	63853
Cameroon	27914542	0.45	1.41	124605
Guinea	13859349	0.28	1.41	38280
Guinea-Bissau	2105580	0.43	1.39	8960
Gambia	2705995	0.47	1.37	12598
Eritrea	3684041	0.28	1.35	10189
Somalia	17597508	0.16	1.34	27324
Chad	17723312	0.04	1.31	7682

Equatorial Guinea	1674916	1.02	1.3	17130
Seychelles	107135	47.54	1.25	50937
Democratic Republic of Congo	99010216	0.1	1.24	95814
Burundi	12889583	0.42	1.21	53661
Central African Republic	5579148	0.28	1.18	15367
Comoros	836783	1.08	1.11	9048
Benin	13352864	0.21	1.09	27999
Congo	5970430	0.42	0.93	25110
Tanzania	65497752	0.07	0.5	42927
Reunion	974062	50.78	NULL	494595
Saint Helena	5401	40.1	NULL	2166
Mayotte	326113	13.15	NULL	42879
Western Sahara	576005	NULL	NULL	NULL

LIMITATIONS AND CHALLENGES

Listed below are some of the limitations and challenges of the dataset used for this project.

I would have loved to gain insights into the following, but the data is limited does not provide information on them.

- People who got vaccinated but still died of COVID virus.
- People who did not get vaccinated, contracted COVID and recovered.
- The health and demographics profile of those who died of COVID

APPENDIX

Below are the meta data for the two data tables used in this exploratory data analysis.

Deaths Table Metadata:

Column Name	Description
iso_code	ISO 3166-1 alpha-3 code of the country
continent	Continent of the country
location	Name of the country
date	Date of the observation
population	Total population of the country
total_cases	Total number of confirmed cases of COVID-19
new_cases	New confirmed cases of COVID-19 on the given date
new_cases_smoothed	New confirmed cases of COVID-19 (7-day smoothed) on the given date
total_deaths	Total number of deaths due to COVID-19
new_deaths	New deaths due to COVID-19 on the given date
new_deaths_smoothed	New deaths due to COVID-19 (7-day smoothed) on the given date

total_cases_per_million	Total confirmed cases of COVID-19 per million population
new_cases_per_million	New confirmed cases of COVID-19 per million population on the given date
new_cases_smoothed_per_million	New confirmed cases of COVID-19 (7-day smoothed) per million population on the given date
total_deaths_per_million	Total deaths due to COVID-19 per million population
new_deaths_per_million	New deaths due to COVID-19 per million population on the given date
new_deaths_smoothed_per_million	New deaths due to COVID-19 (7-day smoothed) per million population on the given date
reproduction_rate	The estimated average number of people who will be infected by one infected person
icu_patients	Number of COVID-19 patients in ICU on the given date
icu_patients_per_million	Number of COVID-19 patients in ICU per million population on the given date
hosp_patients	Number of COVID-19 patients in hospital on the given date
hosp_patients_per_million	Number of COVID-19 patients in hospital per million population on the given date
weekly_icu_admissions	Number of COVID-19 patients admitted to the ICU in the week preceding the given date
weekly_icu_admissions_per_million	Number of COVID-19 patients admitted to the ICU per million population in the week preceding the given date
weekly_hosp_admissions	Number of COVID-19 patients admitted to hospitals in the week preceding the given date
weekly_hosp_admissions_per_million	Number of COVID-19 patients admitted to hospitals per million population in the week preceding the given date

Vaccinations Table Metadata:

Column Name	Description
iso_code	ISO 3166-1 alpha-3 country code
continent	Continent of the location
location	Location name
date	Date of observation
new_tests	New tests administered
total_tests	Total tests administered
total_tests_per_thousand	Total tests per thousand people
new_tests_per_thousand	New tests per thousand people
new_tests_smoothed	New tests (7-day smoothed)
new_tests_smoothed_per_thousand	New tests (7-day smoothed) per thousand people
positive_rate	The share of COVID-19 tests that are positive, given as a decimal fraction
tests_per_case	The number of COVID-19 tests conducted per confirmed case, given as a decimal fraction
tests_units	Units used by the location to report COVID-19 test data
total_vaccinations	Total number of COVID-19 vaccination doses administered
people_vaccinated	Total number of people who received at least one COVID-19 vaccination dose
people_fully_vaccinated	Total number of people who received all doses prescribed by the vaccination protocol
new_vaccinations	New COVID-19 vaccination doses administered (only for dates after the country began reporting this data)
new_vaccinations_smoothed	New COVID-19 vaccination doses administered (7-day smoothed) (only for dates after the country began reporting this data)
total_vaccinations_per_hundred	Total number of COVID-19 vaccination doses administered per 100 people in the total population
people_vaccinated_per_hundred	Total number of people who received at least one COVID-19 vaccination dose per 100 people in the total population

people_fully_vaccinated_per_hundred	Total number of people who received all doses prescribed by the vaccination protocol per 100 people in the total population
new_vaccinations_smoothed_per_million	New COVID-19 vaccination doses administered (7-day smoothed) per 1,000,000 people (only for dates after the country began reporting this data)
stringency_index	Government response stringency index (scale of 0 to 100, with 100 being the strictest response)
population_density	Number of people divided by land area, measured in square kilometers
median_age	Median age of the population
aged_65_older	Share of the population that is 65 years and older, given as a percentage
aged_70_older	Share of the population that is 70 years and older, given as a percentage
gdp_per_capita	Gross domestic product per capita
extreme_poverty	Share of the population living in extreme poverty, given as a percentage
cardiovasc_death_rate	Cardiovascular disease death rate, given as annual number of deaths per 100,000 people
diabetes_prevalence	Diabetes prevalence (% of population aged 20 to 79)
female_smokers	Share of women who smoke, given as a percentage
male_smokers	Share of men who smoke, given as a percentage
handwashing_facilities	Share of the population with basic handwashing facilities, given