COVID-19 DATA ANALYSIS

Introduction

- Though it has subsided by a large extent as of 2023, the COVID-19 pandemic has reshaped the world in unprecedented ways. Our 'COVID-19 Data Analysis' project delves into a comprehensive exploration of the pandemic's effects, focusing on data-driven insights to understand its impact on different nations and how they have reacted to it.

As the global community continues to navigate the complexities of the pandemic, my project embarks on a data-driven journey to uncover the multifaceted aspects of the COVID-19 crisis. Through meticulous data analysis and visualization, I seek to uncover patterns, trends, and lessons that can guide decision-making.

My mission is to harness the power of data to go beyond headlines and delve into the nuances of the pandemic's effects across countries. By examining key metrics, vaccination rates, and response strategies, I aim to provide a comprehensive overview of how nations have grappled with this global challenge.

Gathering Data

I gathered and downloaded the Covid-19 dataset from the site 'Our world in data'

(link : https://ourworldindata.org/covid-deaths).

- The original dataset contains over a 90 columns and several thousand rows of data,

- containing data of many countries in columns like 'Total cases of infection', 'change in vaccination percentage', etc. to name a few. This data is spread across several decades for each country, hence helping us to analyze each country's covid performance with more clarity

Data Cleaning

The data was stored in a .csv format. Although the data was arranged quite beautifully, in many places there were many blank spaces. These blank spaces meant that there was no particular data available for that row and the specific column.
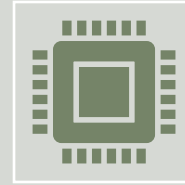
Now the problem is, since we have to connect this .csv file to our MySQL server, we cannot afford to have blank spaces, or , empty cells. Because in this case, it will read them just as a series of comas and MySQL may not understand that these are actually empty cells.
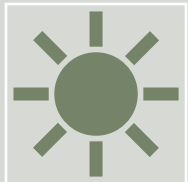
But before that, I split the big table into two halves, one table for the Covid Deaths, and the other for data related to Covid Vaccinations.

Hence to prevent any errors, I first filled all the blank cells with zeroes. After that, I used the 'load infile' command to upload both the .csv files into the MySQL server and then , we had two tables with all the data contained within them.

We had split the original table into two parts, one for storing data exclusively related to covid deaths and other for covid vaccinations.
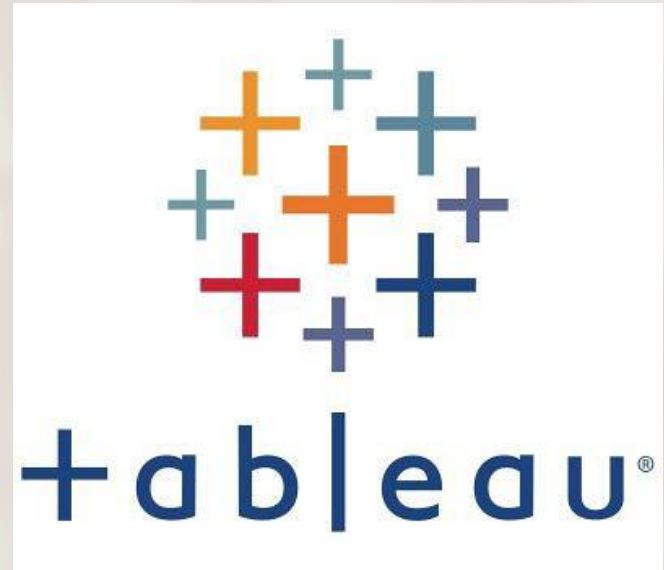
As you can see, we have two tables here, one is 'covadeaths' and the other is 'covidvaccinations'.
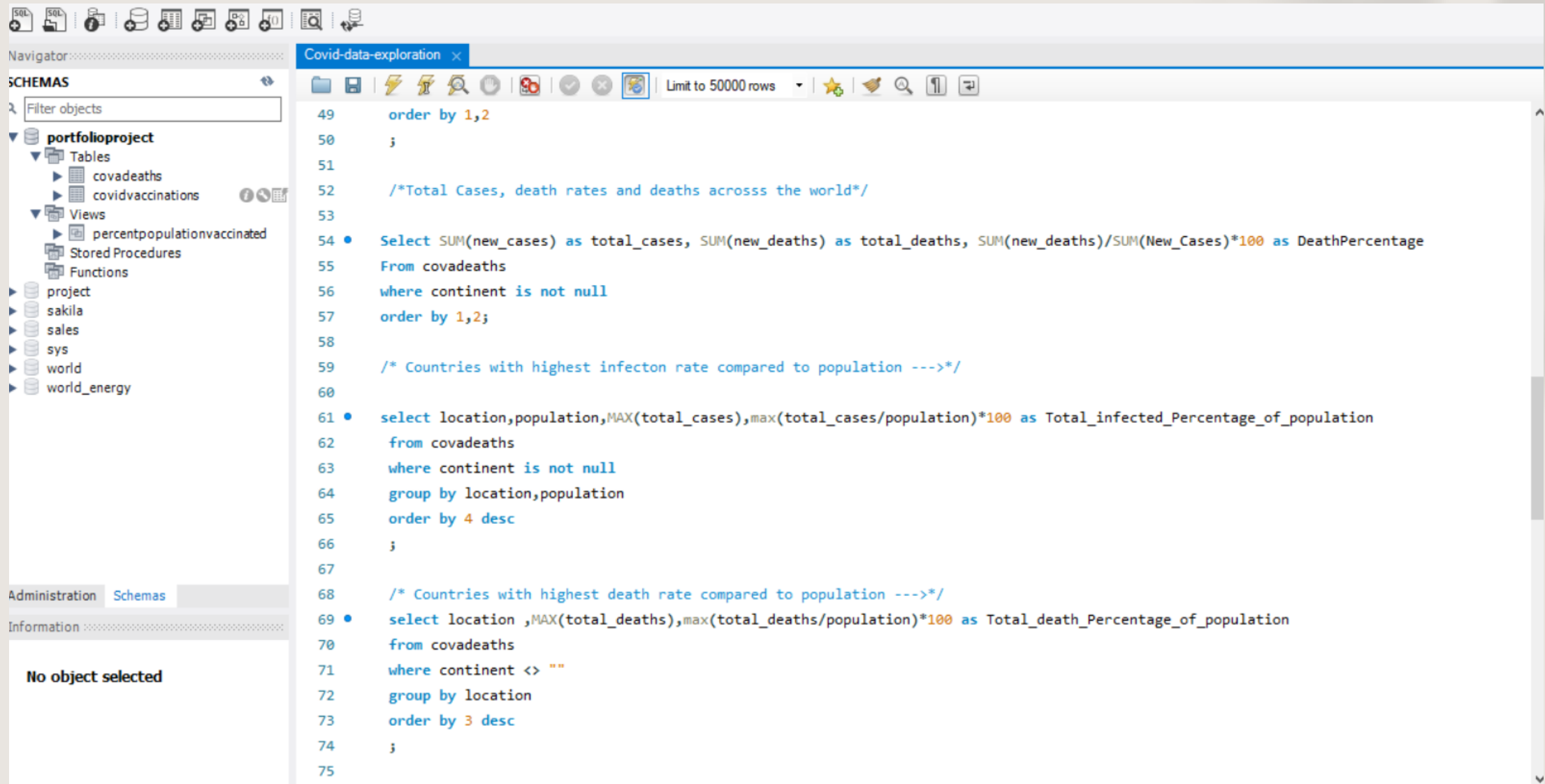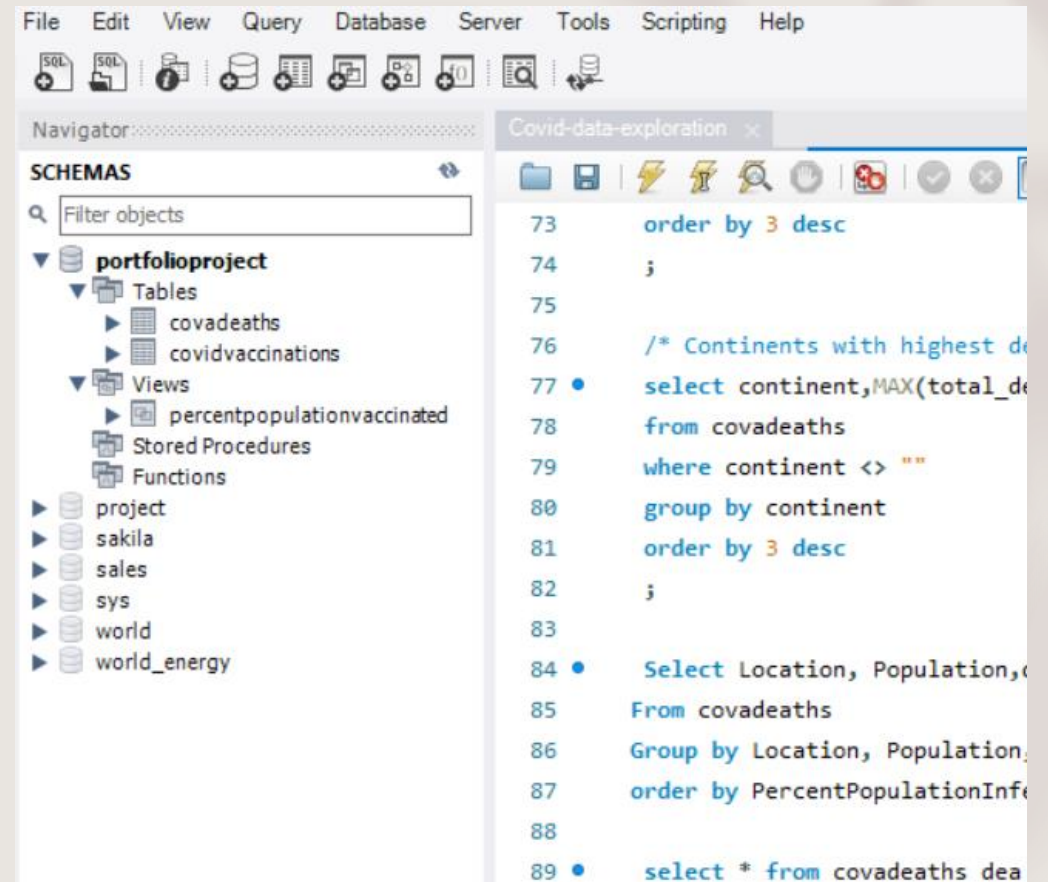
Writing Queries

- So after successfully loading our .csv files into MySQL , it's time to gain some insights on our data and answer some questions related to energy distribution and energy usage of nations.

- We answer some questions and looked into things like :

1) What are the total no. of infections , to number of deaths globally? What is the global death percentage?
2) What is the continent wise count of total deaths related to Covid? Which continent has suffered the most?
3) What is the percentage of population infected by the virus, per country?
4) What is the trend of percent of population infected over time per country, over the past one year?

- To answer these questions we first wrote relevant queries in MySQL and extracted the information. After extracting the relevant information, we put them onto excel files which were then in turn connected to Tableau, a BI tool which helps us to visualize and present our data in a more beautiful way so that we can explain it better to relevant stakeholders.

Some queries :


```
49        order by 1,2
50        ;
51
52      /*Total Cases, death rates and deaths acrosss the world*/
53
54 ●   Select SUM(new_cases) as total_cases, SUM(new_deaths) as total_deaths, SUM(new_deaths)/SUM(New_Cases)*100 as DeathPercentage
55     From covadeaths
56     where continent is not null
57     order by 1,2;
58
59     /* Countries with highest infecton rate compared to population --->*/
60
61 ●   select location,population,MAX(total_cases),max(total_cases/population)*100 as Total_infected_Percentage_of_population
62      from covadeaths
63      where continent is not null
64      group by location,population
65      order by 4 desc
66      ;
67
68     /* Countries with highest death rate compared to population --->*/
69 ●   select location ,MAX(total_deaths),max(total_deaths/population)*100 as Total_death_Percentage_of_population
70      from covadeaths
71      where continent <> ""
72      group by location
73      order by 3 desc
74      ;
75
```

Navigator

SCHEMAS

Filter objects

- portfolioproject
  - Tables
    - covadeaths
    - covidvaccinations
  - Views
    - percentpopulationvaccinated
  - Stored Procedures
  - Functions
- project
- sakila
- sales
- sys
- world
- world_energy

Administration   Schemas

Information

No object selected

Covid-data-exploration

Limit to 50000 rows

```
49          order by 1,2
50          ;
51
52          /*Total Cases, death rates and deaths acrosss the world*/
53
54 •    Select SUM(new_cases) as total_cases, SUM(new_deaths) as total_deaths, SUM(new_deaths)/SUM(New_Cases)*100 as DeathPercentage
55      From covadeaths
56      where continent is not null
57      order by 1,2;
58
59      /* Countries with highest infecton rate compared to population --->*/
60
61 •    select location,population,MAX(total_cases),max(total_cases/population)*100 as Total_infected_Percentage_of_population
62       from covadeaths
63       where continent is not null
```

	total_cases	total_deaths	DeathPercentage
▶	3253410813	29058502	0.8932

Limit to 50000 rows

```sql
55      From covadeaths
56      where continent is not null
57      order by 1,2;
58
59      /* Countries with highest infecton rate compared to population --->*/
60
61  •   select location,population,MAX(total_cases),max(total_cases/population)*100 as Total_infected_Percentage_of_population
62       from covadeaths
63       where continent is not null
64       group by location,population
65       order by 4 desc
66       ;
67
68      /* Countries with highest death rate compared to population --->*/
69  •   select location ,MAX(total_deaths),max(total_deaths/population)*100 as Total_death_Percentage_of_population
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| location	population	MAX(total_cases)	Total_infected_Percentage_of_population
Cyprus	896007	660854	73.7600
San Marino	33690	24326	72.2100
Brunei	449002	308777	68.7700
Austria	8939617	6081287	68.0300
Faeroe Islands	53117	34658	65.2500
Slovenia	2119843	1344559	63.4300
Gibraltar	32677	20550	62.8900
Martinique	367512	230354	62.6800
South Korea	51815808	32256154	62.2500
Andorra	79843	48015	60.1400
Jersey	110796	66391	59.9200
Luxembourg	647601	382785	59.1100

Result 2 ✕

# Connecting data source to Tableau

- After getting all the desried results and outputs from MySQL, we copy the output from MySql and paste them on excel files. This is done because Tableau public does not support importing data directly from sql files.
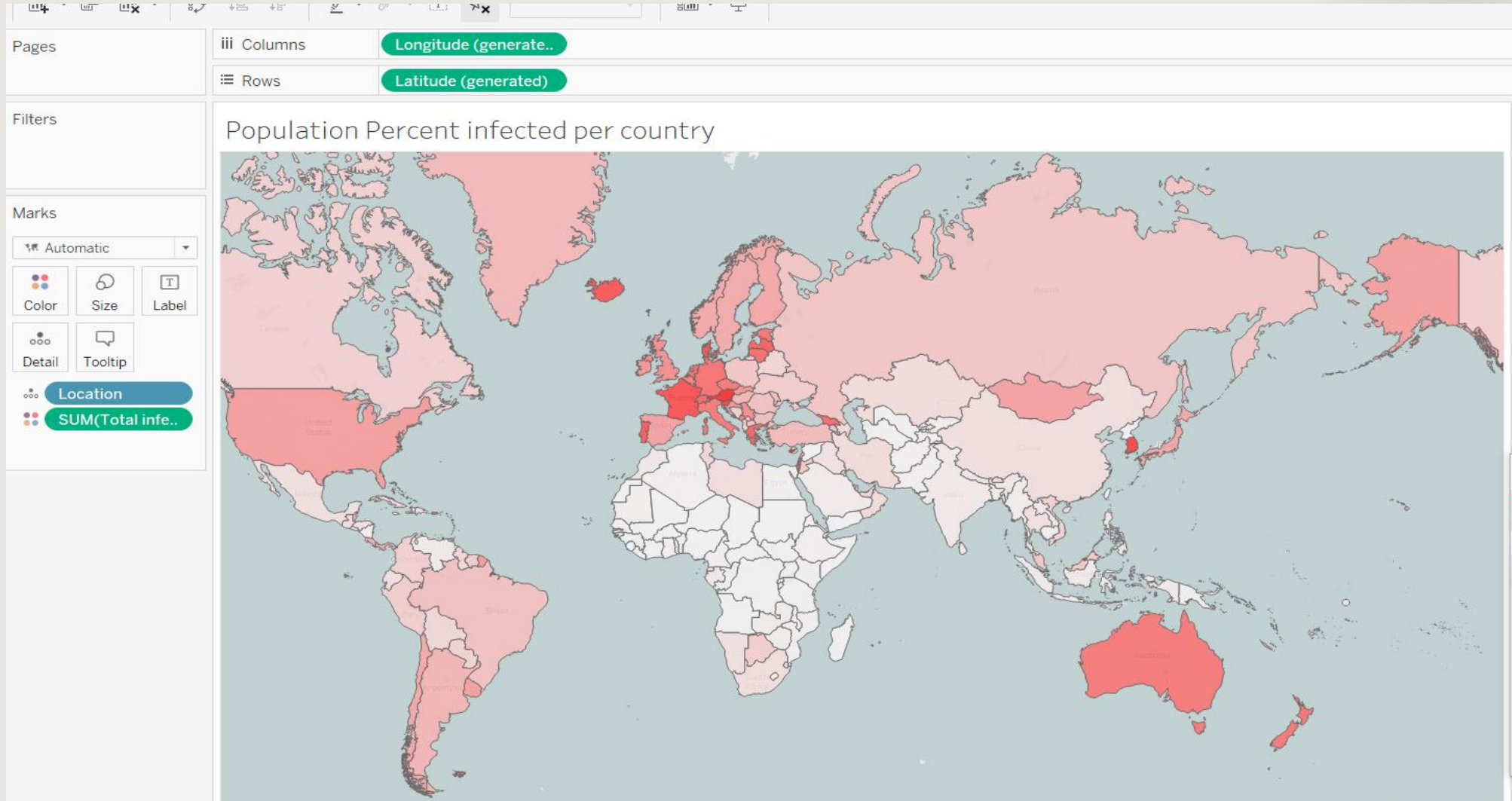
Tableau visualizations :

What are the total no. of infections, to number of deaths globally? What is the global death percentage?
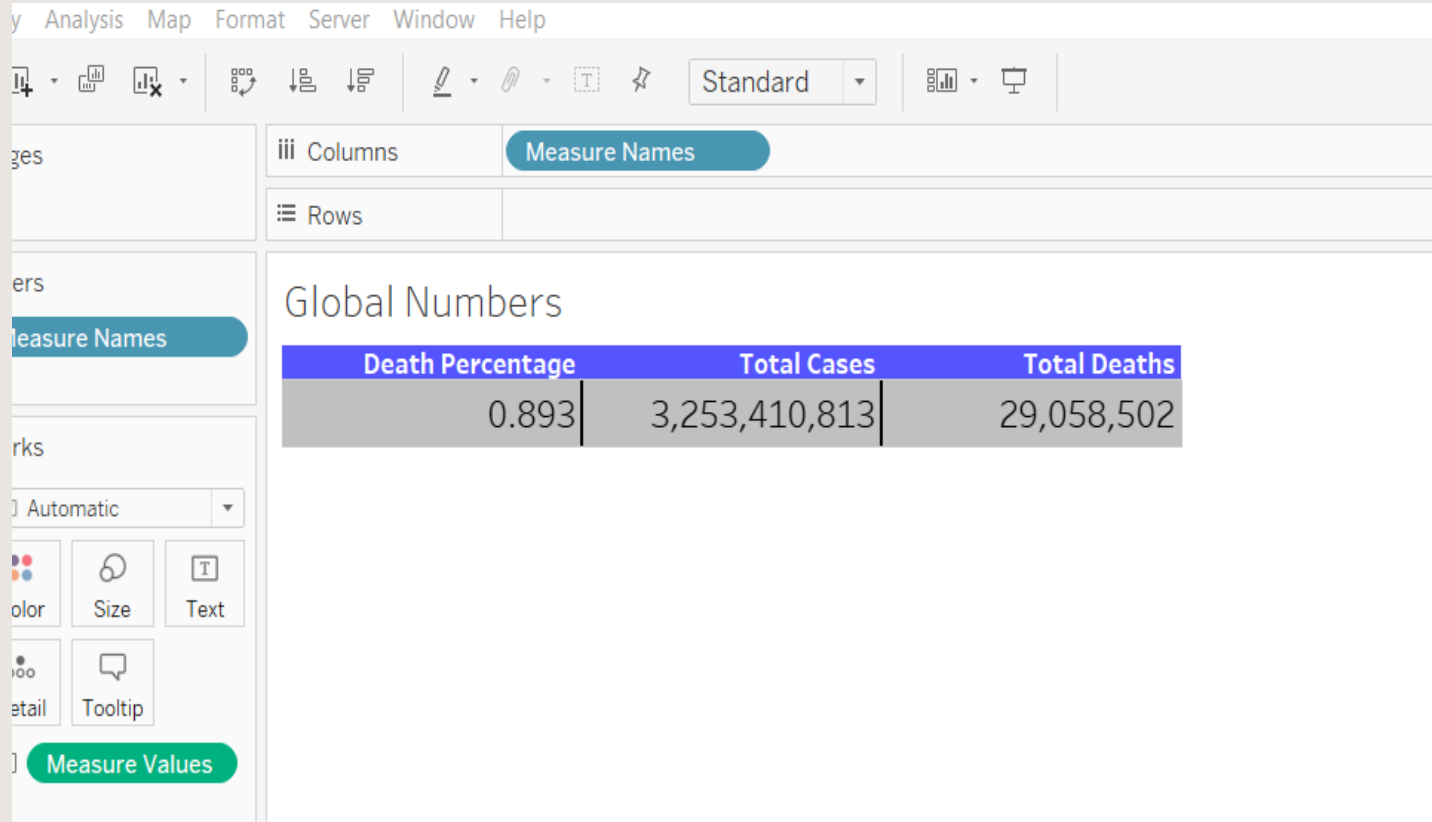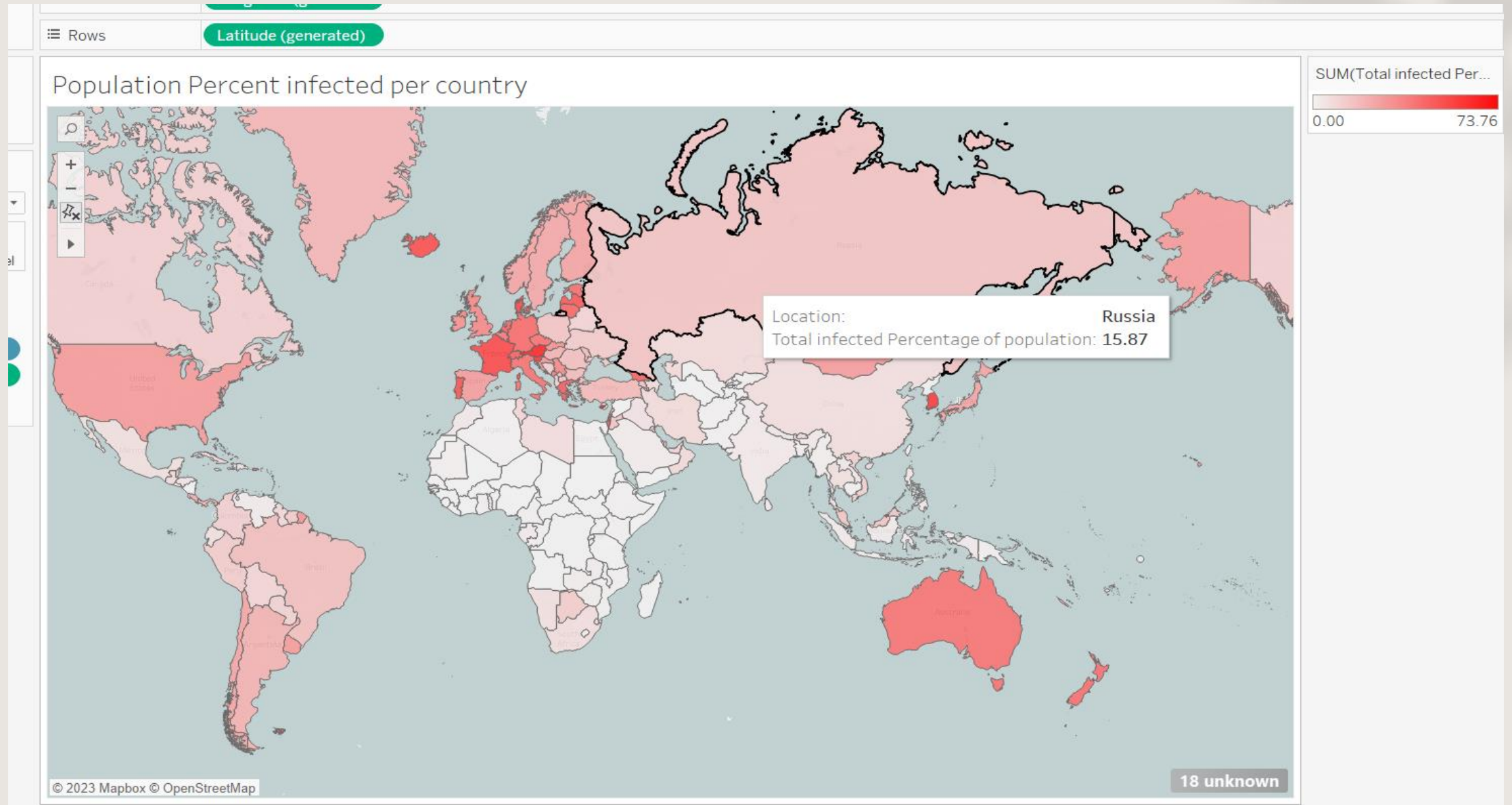
- What is the continent wise count of total deaths related to Covid? Which continent has suffered the most?
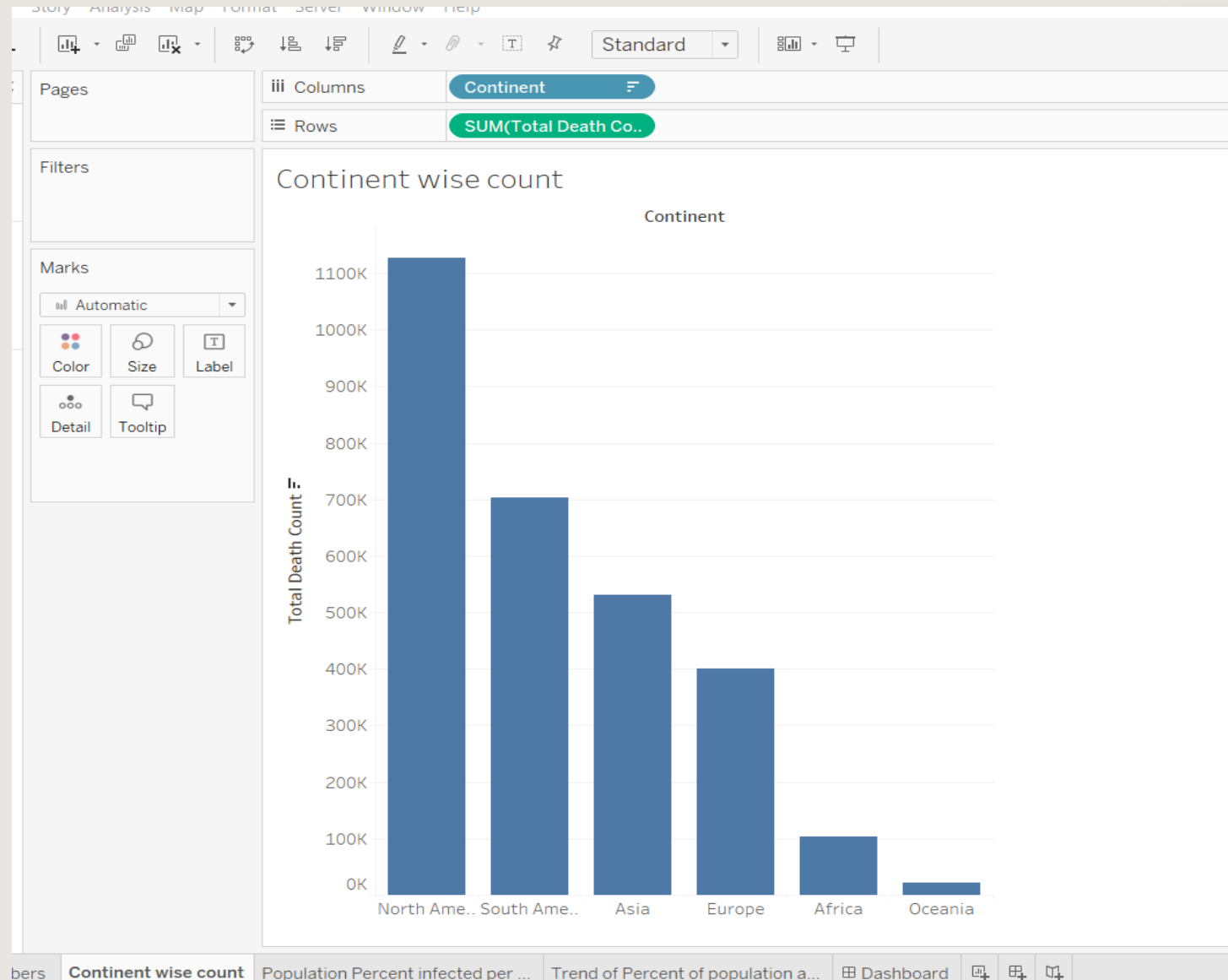
- What is the percentage of population infected by the virus, per country?
- For this viz, we have plotted the entire world map and one can view the percentage infected of each country by simply hovering/tapping on that particular country.
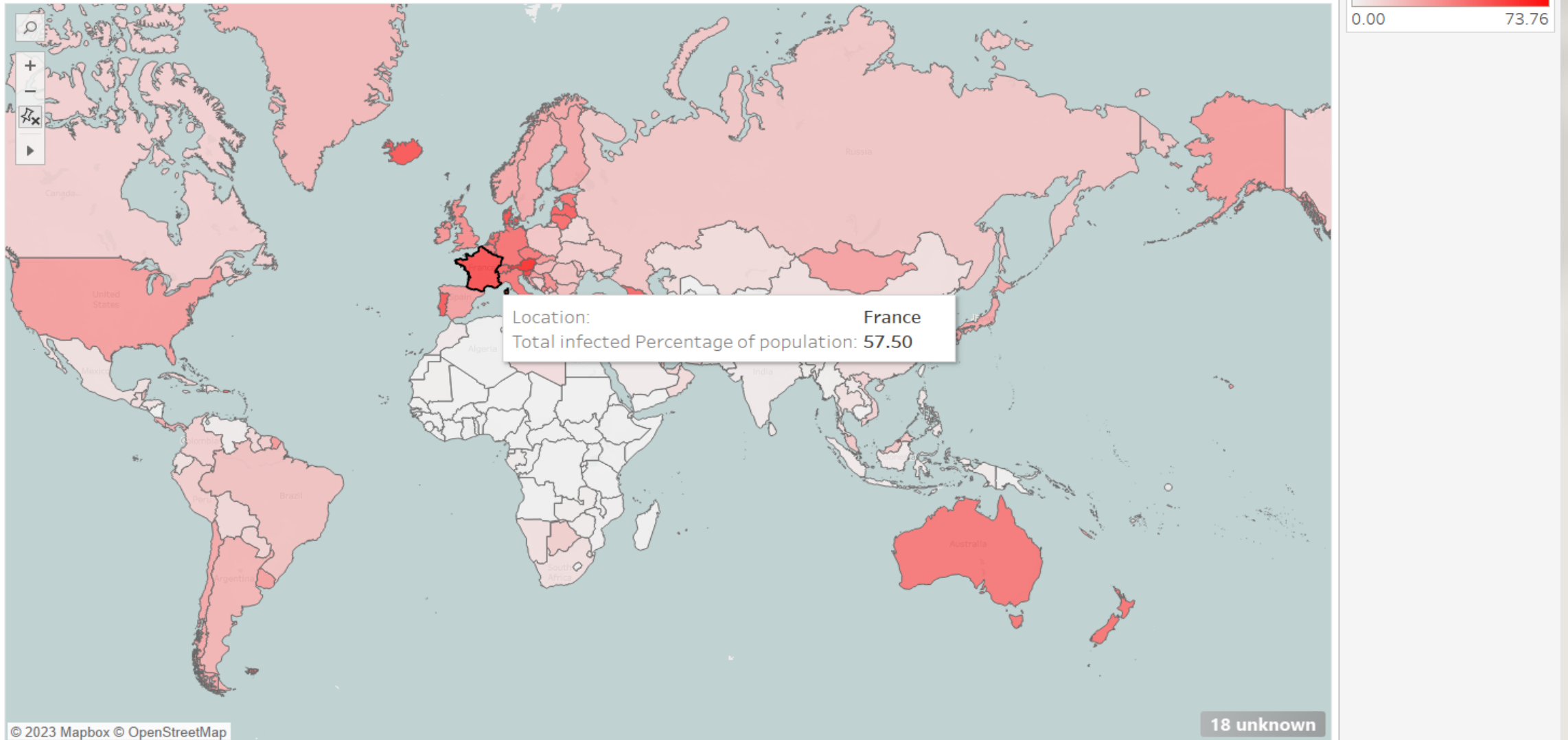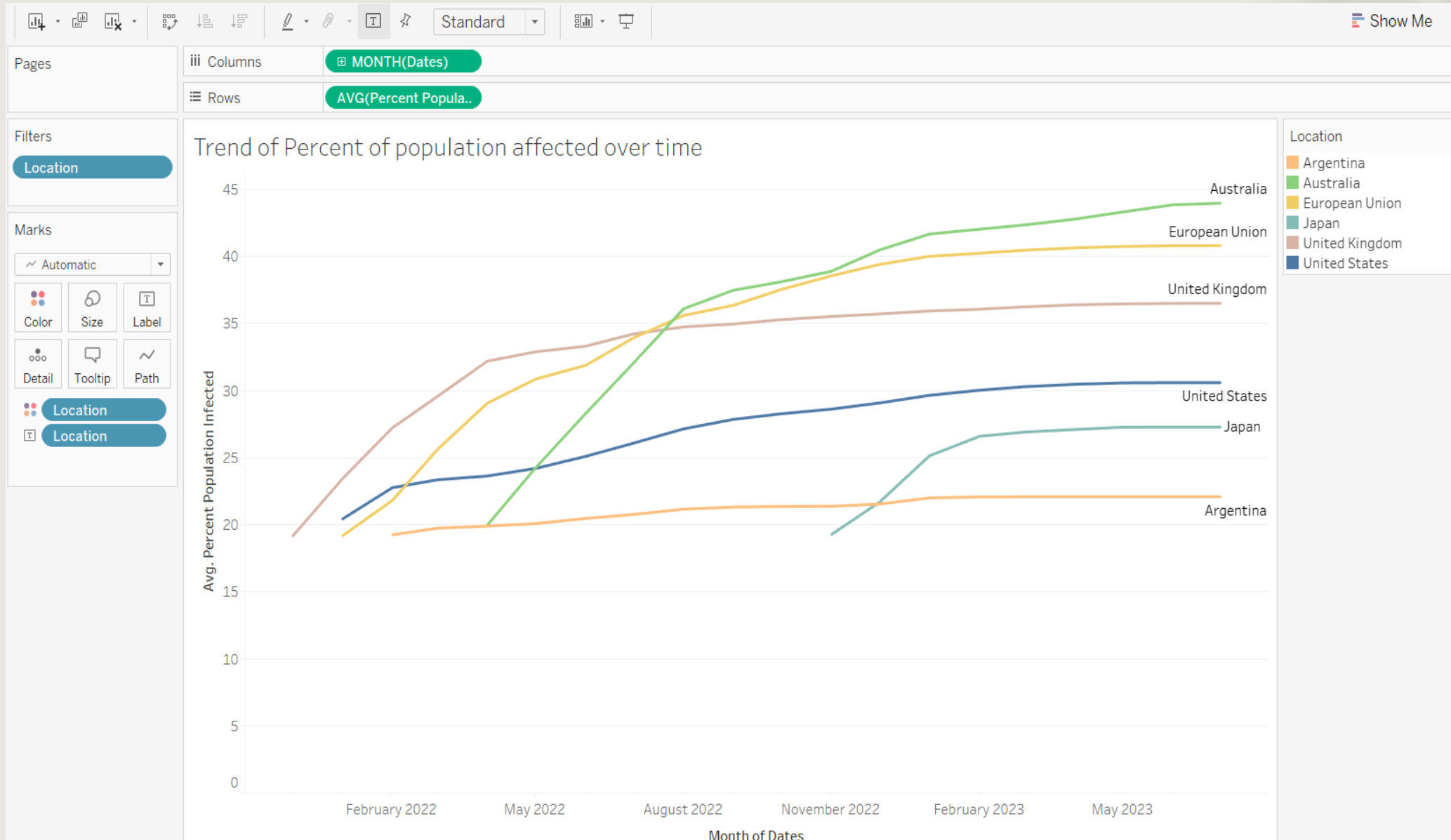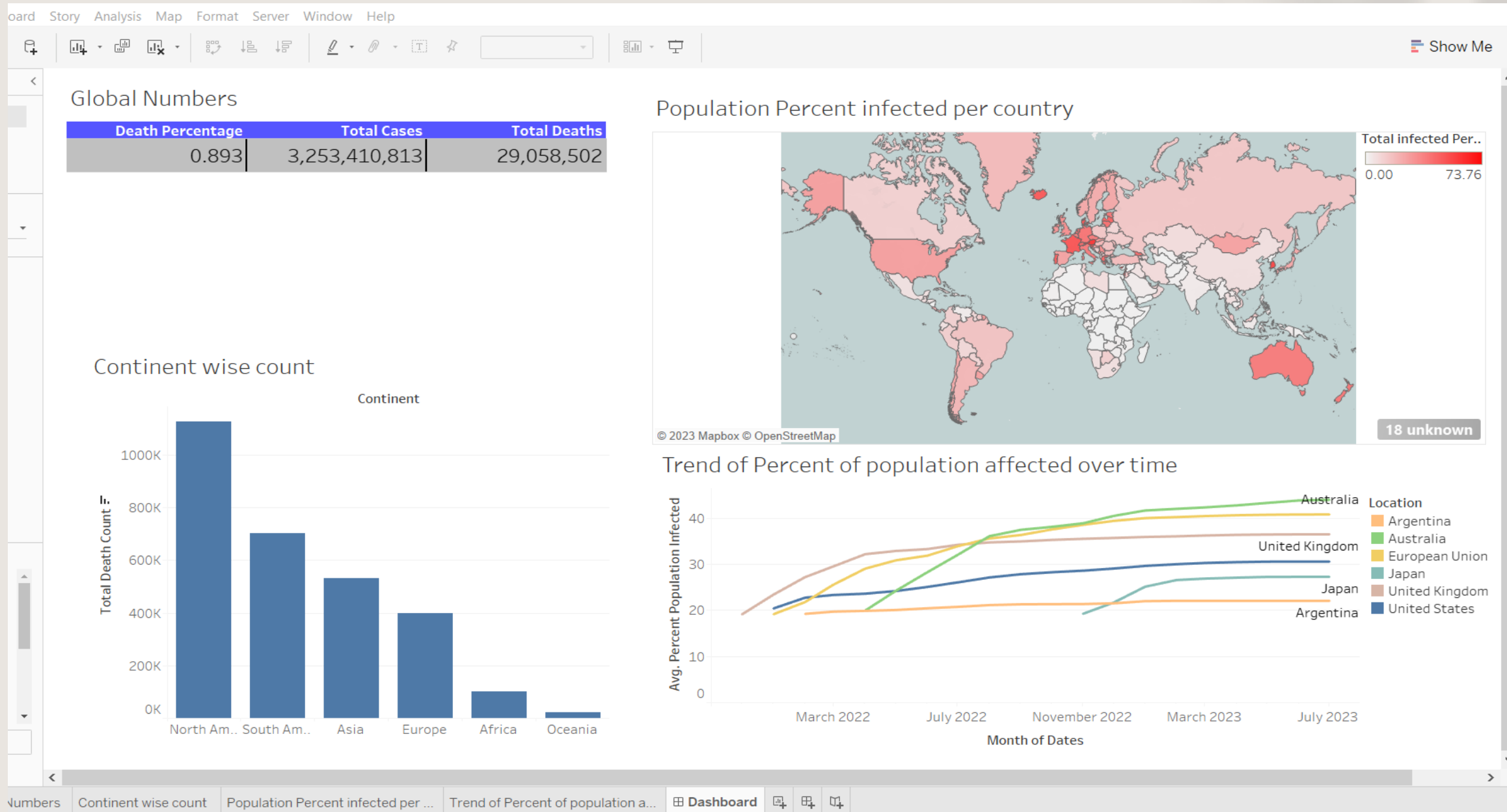
- E.g., for Russia

- For France,

Population Percent infected per country

SUM(Total infected Per...

0.00   73.76

Location: France
Total infected Percentage of population: 57.50

18 unknown

© 2023 Mapbox © OpenStreetMap

- What is the trend of percent of population infected over time per country, over the past one year?

- And finally, presenting all our viz. Together in the **Dashboard** :

Conclusion

- Hence, we saw how using Excel,SQL and Tableau can be extremely helpful in cleaning data,organising data, retreiving specific information and finally presenting them in a visually appealing manner to the stakeholders.

- We got insights of how covid has impacted different countries around the world .