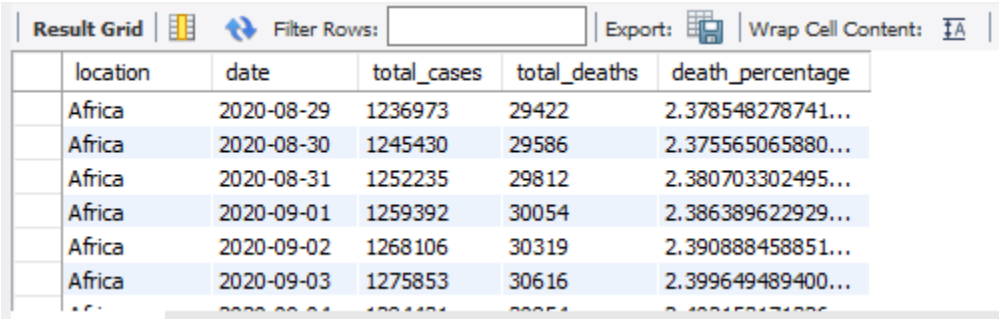
# **EDA** using SQL

## Covid dataset

1. Death percentage query:

```
Select
location,date,total_cases,total_deaths,(total_deaths/total_cases)*
100 as death_percentage
from coviddeaths
order by 1,2;
```



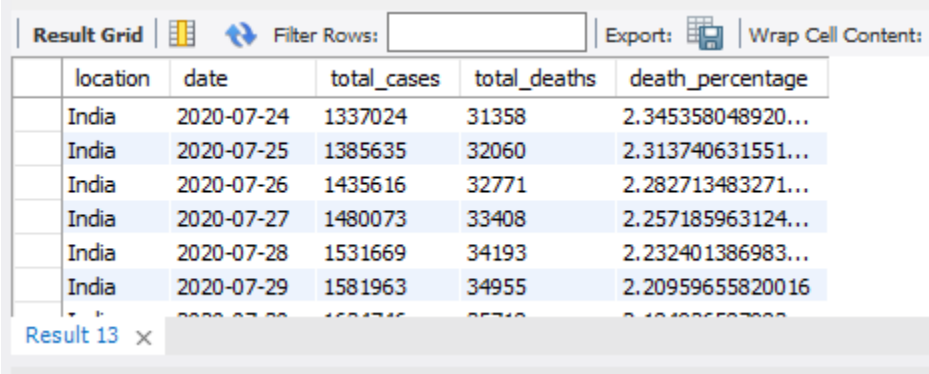
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| location | date | total_cases | total_deaths | death_percentage |
|---|---|---|---|---|
| Africa | 2020-08-29 | 1236973 | 29422 | 2.378548278741... |
| Africa | 2020-08-30 | 1245430 | 29586 | 2.375565065880... |
| Africa | 2020-08-31 | 1252235 | 29812 | 2.380703302495... |
| Africa | 2020-09-01 | 1259392 | 30054 | 2.386389622929... |
| Africa | 2020-09-02 | 1268106 | 30319 | 2.390888458851... |
| Africa | 2020-09-03 | 1275853 | 30616 | 2.399649489400... |

2. Death percentage in India

```
select
location,date,total_cases,total_deaths,(total_deaths/total_cases)*
100 as death_percentage
from coviddeaths
where location='India'
```

order by 1,2;



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

| location | date | total_cases | total_deaths | death_percentage |
|---|---|---|---|---|
| India | 2020-07-24 | 1337024 | 31358 | 2.345358048920... |
| India | 2020-07-25 | 1385635 | 32060 | 2.313740631551... |
| India | 2020-07-26 | 1435616 | 32771 | 2.282713483271... |
| India | 2020-07-27 | 1480073 | 33408 | 2.257185963124... |
| India | 2020-07-28 | 1531669 | 34193 | 2.232401386983... |
| India | 2020-07-29 | 1581963 | 34955 | 2.20959655820016 |

Result 13

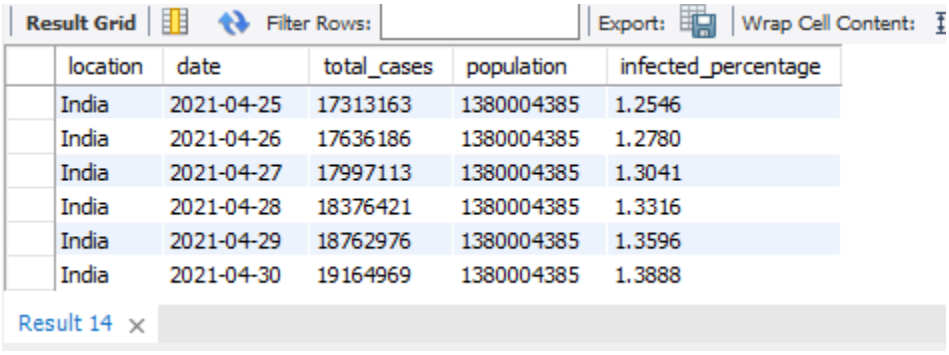
3. Percentage of people infected with covid

```
select
location,date,total_cases,population,(total_cases/population)*100
as infected_percentage

from coviddeaths

where location='India'

order by 1,2;
```



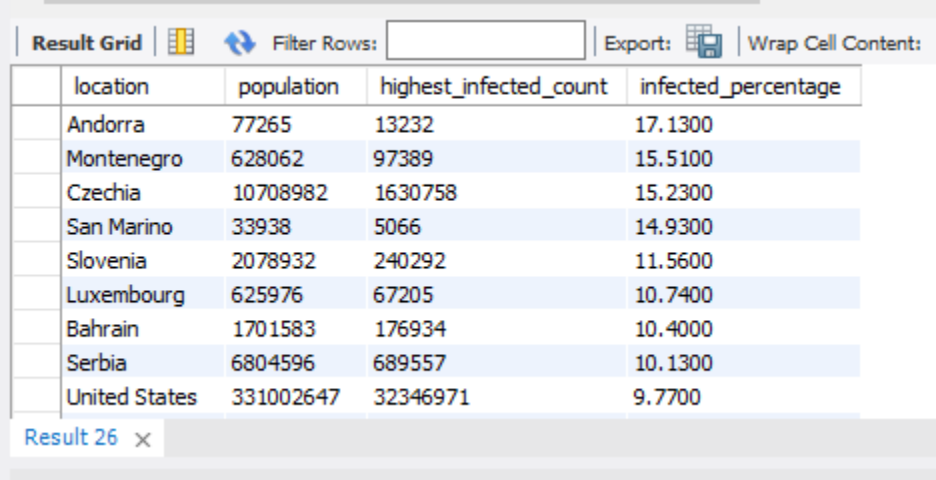
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| location | date | total_cases | population | infected_percentage |
|---|---|---|---|---|
| India | 2021-04-25 | 17313163 | 1380004385 | 1.2546 |
| India | 2021-04-26 | 17636186 | 1380004385 | 1.2780 |
| India | 2021-04-27 | 17997113 | 1380004385 | 1.3041 |
| India | 2021-04-28 | 18376421 | 1380004385 | 1.3316 |
| India | 2021-04-29 | 18762976 | 1380004385 | 1.3596 |
| India | 2021-04-30 | 19164969 | 1380004385 | 1.3888 |

Result 14

4. Countries with highest infection rate based on population

```
select location,population,max(total_cases) as highest_infected_count,
max((total_cases/population))*100 as infected_percentage
from coviddeaths
group by location,population
order by infected_percentage desc
```



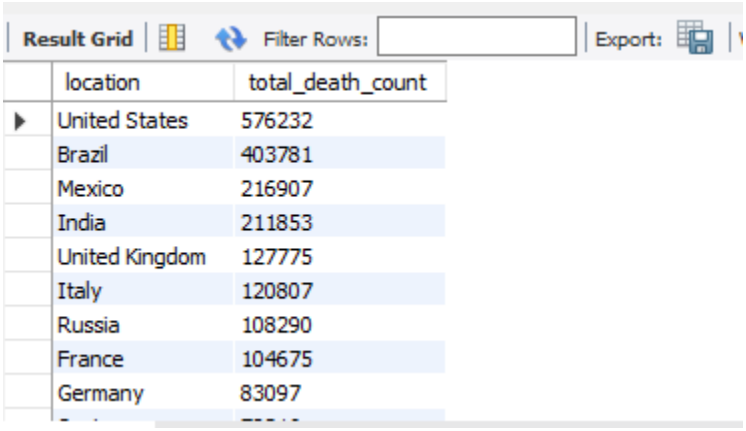
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| location | population | highest_infected_count | infected_percentage |
|---|---|---|---|
| Andorra | 77265 | 13232 | 17.1300 |
| Montenegro | 628062 | 97389 | 15.5100 |
| Czechia | 10708982 | 1630758 | 15.2300 |
| San Marino | 33938 | 5066 | 14.9300 |
| Slovenia | 2078932 | 240292 | 11.5600 |
| Luxembourg | 625976 | 67205 | 10.7400 |
| Bahrain | 1701583 | 176934 | 10.4000 |
| Serbia | 6804596 | 689557 | 10.1300 |
| United States | 331002647 | 32346971 | 9.7700 |

Result 26

5. Countries with highest death count per population

```
select location,max(cast(total_deaths as signed)) as total_death_count
from coviddeaths
where continent != ""
group by location
order by total_death_count desc
```

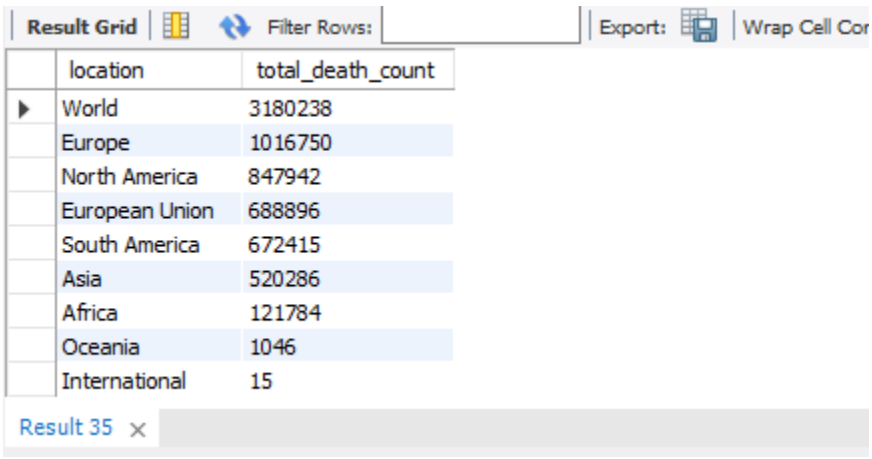


Result Grid | Filter Rows: | Export:

| location | total_death_count |
|---|---|
| United States | 576232 |
| Brazil | 403781 |
| Mexico | 216907 |
| India | 211853 |
| United Kingdom | 127775 |
| Italy | 120807 |
| Russia | 108290 |
| France | 104675 |
| Germany | 83097 |

6. Continents with highest deaths per population

```
select location,max(cast(total_deaths as signed)) as total_death_count
from coviddeaths
where continent = ""
group by location
order by total_death_count desc;
```



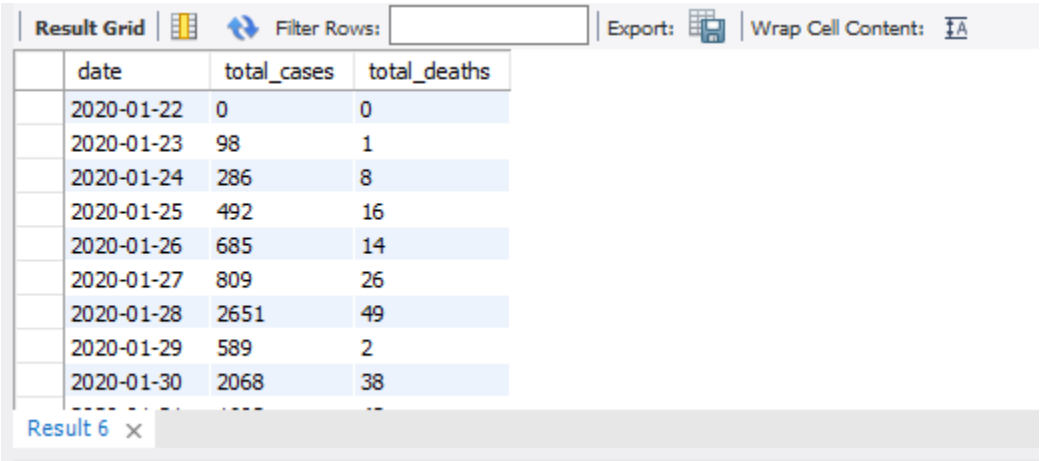
Result Grid | Filter Rows: | Export: | Wrap Cell Cor

| location | total_death_count |
|---|---|
| World | 3180238 |
| Europe | 1016750 |
| North America | 847942 |
| European Union | 688896 |
| South America | 672415 |
| Asia | 520286 |
| Africa | 121784 |
| Oceania | 1046 |
| International | 15 |

Result 35

7. Total number of cases and deaths globally per day

```
select date,sum(new_cases) as total_cases,sum(new_deaths) as
total_deaths
from coviddeaths
where continent!=""
group by date
order by date;
```



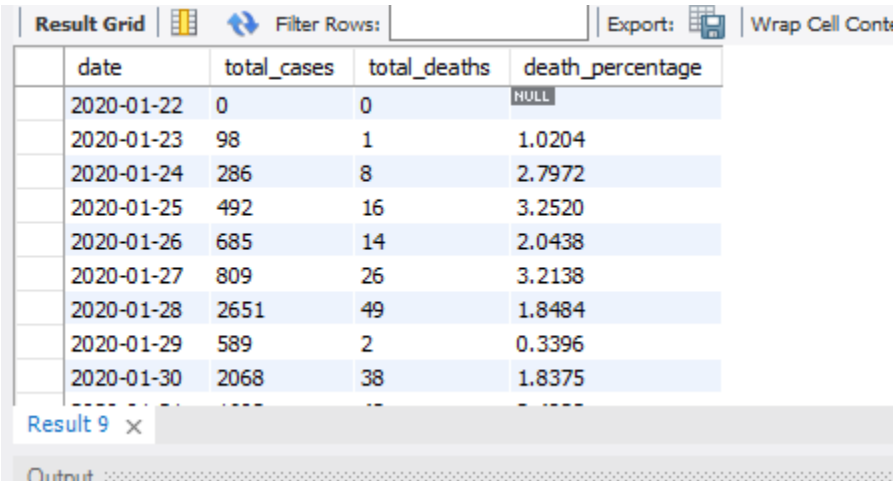
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

| date | total_cases | total_deaths |
|---|---|---|
| 2020-01-22 | 0 | 0 |
| 2020-01-23 | 98 | 1 |
| 2020-01-24 | 286 | 8 |
| 2020-01-25 | 492 | 16 |
| 2020-01-26 | 685 | 14 |
| 2020-01-27 | 809 | 26 |
| 2020-01-28 | 2651 | 49 |
| 2020-01-29 | 589 | 2 |
| 2020-01-30 | 2068 | 38 |

Result 6

8. Percentage of deaths globally

```
select date,sum(new_cases) as total_cases,sum(new_deaths) as
total_deaths, (sum(cast(new_deaths as
signed))/sum(new_cases))*100 as death_percentage
from coviddeaths
where continent!=""
group by date
order by date;
```



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

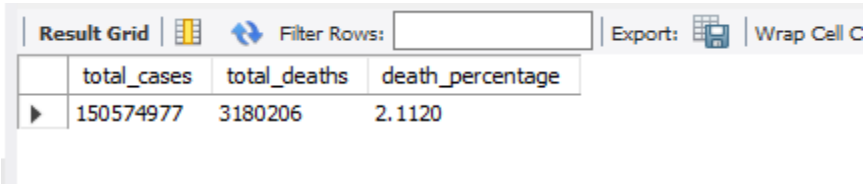
| date | total_cases | total_deaths | death_percentage |
|---|---|---|---|
| 2020-01-22 | 0 | 0 | NULL |
| 2020-01-23 | 98 | 1 | 1.0204 |
| 2020-01-24 | 286 | 8 | 2.7972 |
| 2020-01-25 | 492 | 16 | 3.2520 |
| 2020-01-26 | 685 | 14 | 2.0438 |
| 2020-01-27 | 809 | 26 | 3.2138 |
| 2020-01-28 | 2651 | 49 | 1.8484 |
| 2020-01-29 | 589 | 2 | 0.3396 |
| 2020-01-30 | 2068 | 38 | 1.8375 |

Result 9 ×

Output

9. Total number of cases, deaths and death percentage globally

```
select sum(new_cases) as total_cases,sum(new_deaths) as
total_deaths, (sum(cast(new_deaths as
signed))/sum(new_cases))*100 as death_percentage
from coviddeaths
where continent!=""
order by date;
```

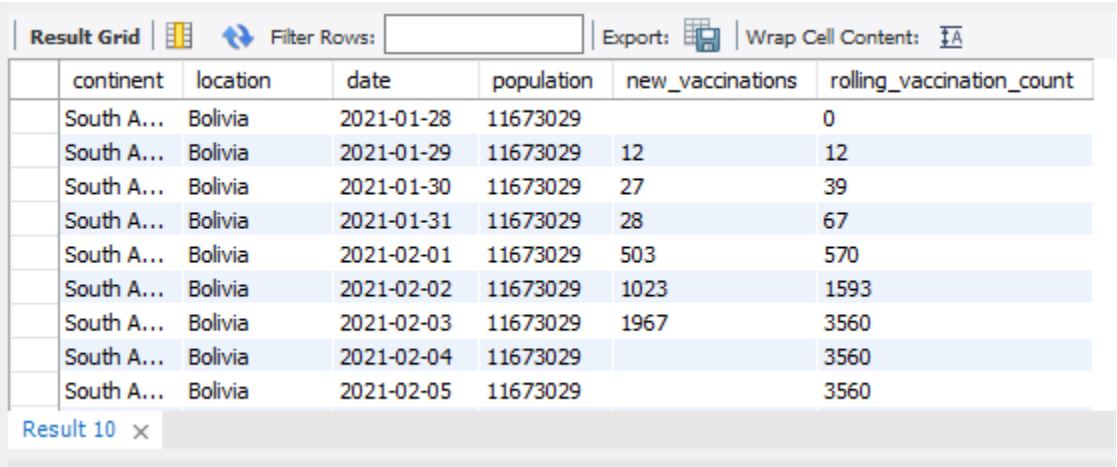


Result Grid | Filter Rows: | Export: | Wrap Cell C

| total_cases | total_deaths | death_percentage |
|---|---|---|
| 150574977 | 3180206 | 2.1120 |

10. Rolling count of new vaccination

```
select
dea.continent,dea.location,dea.date,dea.population,vac.new_vaccinations,sum(vac.new_vaccinations) over (partition by dea.location order by dea.location,dea.date) as rolling_vaccination_count
from coviddeaths dea join covidvaccinations vac
on dea.date = vac.date and dea.location = vac.location
where dea.continent !=""
order by 2,3;
```



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

| continent | location | date | population | new_vaccinations | rolling_vaccination_count |
|---|---|---|---|---|---|
| South A... | Bolivia | 2021-01-28 | 11673029 | | 0 |
| South A... | Bolivia | 2021-01-29 | 11673029 | 12 | 12 |
| South A... | Bolivia | 2021-01-30 | 11673029 | 27 | 39 |
| South A... | Bolivia | 2021-01-31 | 11673029 | 28 | 67 |
| South A... | Bolivia | 2021-02-01 | 11673029 | 503 | 570 |
| South A... | Bolivia | 2021-02-02 | 11673029 | 1023 | 1593 |
| South A... | Bolivia | 2021-02-03 | 11673029 | 1967 | 3560 |
| South A... | Bolivia | 2021-02-04 | 11673029 | | 3560 |
| South A... | Bolivia | 2021-02-05 | 11673029 | | 3560 |

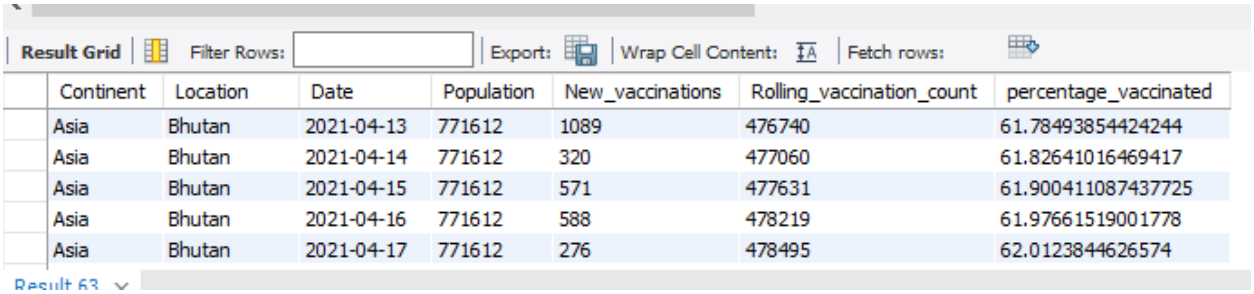
Result 10

11. CTE for percentage of vaccinated people based on population

```
with
pop_vs_vac(Continent,Location,Date,Population,New_vaccinations,Rolling_vaccination_count)
as
(
```

```
select
dea.continent,dea.location,dea.date,dea.population,vac.new_vacc
inations,sum(vac.new_vaccinations) over (partition by
dea.location order by dea.location,dea.date) as
rolling_vaccination_count
from coviddeaths dea join covidvaccinations vac
on dea.date = vac.date and dea.location = vac.location
where dea.continent !=""
)

select *,(Rolling_vaccination_count/population)*100 as
percentage_vaccinated from pop_vs_vac;
```



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

| Continent | Location | Date | Population | New_vaccinations | Rolling_vaccination_count | percentage_vaccinated |
|---|---|---|---|---|---|---|
| Asia | Bhutan | 2021-04-13 | 771612 | 1089 | 476740 | 61.78493854424244 |
| Asia | Bhutan | 2021-04-14 | 771612 | 320 | 477060 | 61.82641016469417 |
| Asia | Bhutan | 2021-04-15 | 771612 | 571 | 477631 | 61.900411087437725 |
| Asia | Bhutan | 2021-04-16 | 771612 | 588 | 478219 | 61.97661519001778 |
| Asia | Bhutan | 2021-04-17 | 771612 | 276 | 478495 | 62.0123844626574 |

Result 63

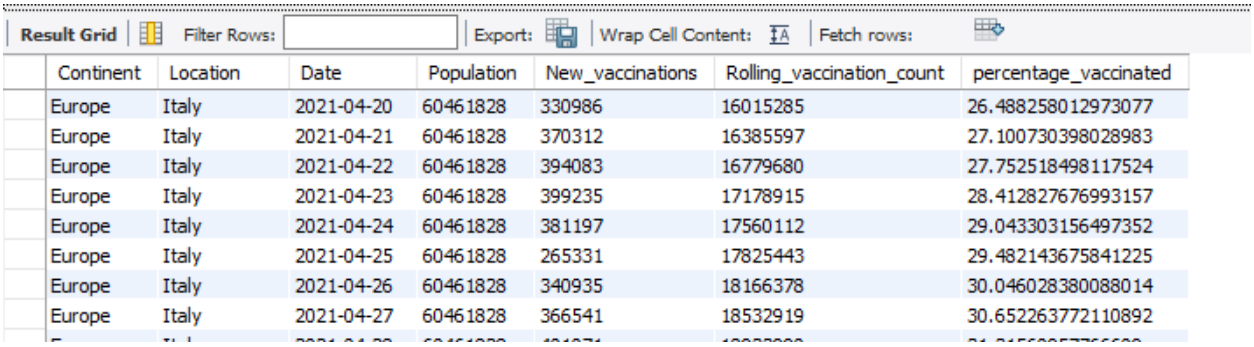
12. Temp table for percentage of vaccinated people based on population

```
drop table if exists percent_pop_vac_temp_table;
create table percent_pop_vac_temp_table(
Continent char(255),
Location char(255),
Date datetime,
```

```
Population numeric,
New_vaccinations numeric null,
Rolling_vaccination_count numeric
);

Insert into percent_pop_vac_temp_table
select
dea.continent,dea.location,dea.date,dea.population,vac.new_vacc
inations,sum(vac.new_vaccinations) over (partition by
dea.location order by dea.location,dea.date) as
rolling_vaccination_count
from coviddeaths dea join covidvaccinations vac
on dea.date = vac.date and dea.location = vac.location
where dea.continent !="";

select *,(Rolling_vaccination_count/population)*100 as
percentage_vaccinated from percent_pop_vac_temp_table;
```



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

| Continent | Location | Date | Population | New_vaccinations | Rolling_vaccination_count | percentage_vaccinated |
|---|---|---|---|---|---|---|
| Europe | Italy | 2021-04-20 | 60461828 | 330986 | 16015285 | 26.488258012973077 |
| Europe | Italy | 2021-04-21 | 60461828 | 370312 | 16385597 | 27.100730398028983 |
| Europe | Italy | 2021-04-22 | 60461828 | 394083 | 16779680 | 27.752518498117524 |
| Europe | Italy | 2021-04-23 | 60461828 | 399235 | 17178915 | 28.412827676993157 |
| Europe | Italy | 2021-04-24 | 60461828 | 381197 | 17560112 | 29.043303156497352 |
| Europe | Italy | 2021-04-25 | 60461828 | 265331 | 17825443 | 29.482143675841225 |
| Europe | Italy | 2021-04-26 | 60461828 | 340935 | 18166378 | 30.046028380088014 |
| Europe | Italy | 2021-04-27 | 60461828 | 366541 | 18532919 | 30.652263772110892 |

13. Creating view

```
create view percentage_infected_view as
select location,population,max(total_cases) as
highest_infected_count, max((total_cases/population))*100 as
infected_percentage
from coviddeaths
group by location
order by infected_percentage desc;
```