SQL Exploratory Data Analysis (EDA) Results

This pdf contains pictures which are supplementary to SQL_EDA query script. Please refer to each query's picture number in context. For creating temp tables or procedures and such, I won't display the screenshot as it all contains 'successful' only. Thank you for reading!

Pic 1. Query: `select * from COVID_Deaths`

What a mess, but it helps to see the overall picture first

Pic 2. Query: `select location, max((total_deaths/total_cases)*100) as Mortality_Percentage`
`from COVID_Deaths`
`group by location`

	location	Mortality_Percentage
1	Saint Kitts and Nevis	1.01302460202605
2	Antigua and Barbuda	13.0434782608696
3	Puerto Rico	NULL
4	Finland	5.21381202835582
5	Lower middle income	14.9425287356322
6	Sao Tome and Principe	13.0434782608696
7	North America	6.45216878783641
8	Falkland Islands	NULL
9	Rwanda	1.41926477431363
10	South Korea	2.39643652561247
11	Bahrain	0.954653937947494
12	Eswatini	6.25
13	Northern Mariana Islands	NULL
14	Gambia	50
15	Liberia	23.0769230769231
16	Guatemala	50
17	Andorra	6.70170827858081
18	Vietnam	3.34608030592734
19	Asia	4.30655475619504
20	World	7.72829944639044
21	New Zealand	1.52925531914894
22	Low income	5.55555555555556
23	Angola	28.5714285714286
24	Uganda	3.00760366535663
25	Nauru	NULL
26	Africa	5.46014682747771

Pic 3. Query: `select location, date, ((total_deaths/total_cases)*100) as`
`Mortality_Percentage`
`from COVID_Deaths`
`where date in (select max(date) from COVID_Deaths)`
`order by 1`

	location	date	Mortality_Percentage
1	Afghanistan	2022-09-08 00:00:00.000	3.98672797468484
2	Africa	2022-09-08 00:00:00.000	2.08025830370482
3	Albania	2022-09-08 00:00:00.000	1.08380432251646
4	Algeria	2022-09-08 00:00:00.000	2.54300258403664
5	Andorra	2022-09-08 00:00:00.000	0.336130809099386
6	Angola	2022-09-08 00:00:00.000	1.86776569624693
7	Anguilla	2022-09-08 00:00:00.000	0.311607374707868
8	Antigua and Barbuda	2022-09-08 00:00:00.000	1.61577891687096
9	Argentina	2022-09-08 00:00:00.000	1.3392245771121
10	Armenia	2022-09-08 00:00:00.000	1.98339008121779
11	Aruba	2022-09-08 00:00:00.000	0.530602746101932

Pic 4.1 Query: `create procedure asia_iso as`
`select distinct iso_code, location from COVID_Deaths`
`where continent = 'Asia'`

'Command Completed Successfully!' No screenshot needed :)

Pic 4.2 Query: `exec asia_iso`

	iso_code	location
1	AFG	Afghanistan
2	ARE	United Arab Emirates
3	ARM	Armenia
4	AZE	Azerbaijan
5	BGD	Bangladesh
6	BHR	Bahrain
7	BRN	Brunei
8	BTN	Bhutan
9	CHN	China
10	GEO	Georgia
11	HKG	Hong Kong
12	IDN	Indonesia
13	IND	India

Pic 4.3 Query: `select location, date, ((total_deaths/total_cases)*100) as`
`Mortality_Percentage`
`from COVID_Deaths`
`where date in (select max(date) from COVID_Deaths) and (iso_code = 'IDN' or`
`iso_code = 'SGP' or iso_code = 'MYS')`
`order by 3`

	location	date	Mortality_Percentage
1	Singapore	2022-09-08 00:00:00.000	0.0863326422048526
2	Malaysia	2022-09-08 00:00:00.000	0.755678054896771
3	Indonesia	2022-09-08 00:00:00.000	2.47025123959694

Pic 5 Query: `select location, population, max(total_cases) as infection_to_date,`

```
max((total_cases/population))*100 as infected_rate_population
from COVID_Deaths
group by location, population
order by infected_rate_population desc
```

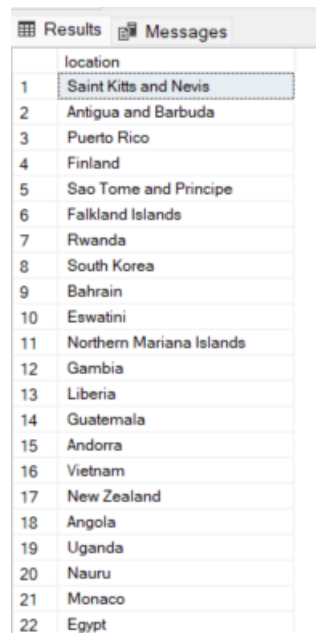|	location	population	infection_to_date	infected_rate_population
1	Faeroe Islands	52888	34658	65.5309332929965
2	Cyprus	896007	578030	64.5117727874894
3	Gibraltar	32670	20069	61.4294459749005
4	San Marino	33746	20456	60.6175546731464
5	Andorra	79034	46113	58.3457752359744
6	Denmark	5854240	3279222	56.0144783951461
7	Austria	8922082	4984809	55.8704683503245
8	Iceland	370335	205009	55.3577166619412
9	Slovenia	2119410	1138346	53.71051377506
10	Saint Pierre and Miquelon	5883	3131	53.2211456739759
11	Portugal	10290103	5438351	52.8503067462007
12	France	67422000	34725285	51.5043828423957
13	Falkland Islands	3764	1886	50.106269925611

Pic 6.1 Query:
```
select location, max(total_deaths) as death_count
from COVID_Deaths
group by location
order by death_count desc
```

	location	death_count
1	United States	999999
2	European Union	999980
3	South Africa	99976
4	Iran	9996
5	Belgium	9996
6	Bosnia and Herzegovina	9995
7	Switzerland	9994
8	Egypt	9994
9	Nepal	9994
10	Colombia	99934
11	Tunisia	9993

Pic 6.2 Query:
```
select location, max(cast(total_deaths as int)) as death_count
from COVID_Deaths
group by location
order by death_count desc
```

	location	death_count
1	World	6510110
2	High income	2571235
3	Upper middle income	2569457
4	Europe	1930100
5	North America	1501520
6	Asia	1476702
7	South America	1326081
8	Lower middle income	1325801
9	European Union	1140714
10	United States	1049749
11	Brazil	684425

Pic 6.3 Query: `select distinct location`
            `from COVID_Deaths`
            `where continent is not null`

	location
1	Saint Kitts and Nevis
2	Antigua and Barbuda
3	Puerto Rico
4	Finland
5	Sao Tome and Principe
6	Falkland Islands
7	Rwanda
8	South Korea
9	Bahrain
10	Eswatini
11	Northern Mariana Islands
12	Gambia
13	Liberia
14	Guatemala
15	Andorra
16	Vietnam
17	New Zealand
18	Angola
19	Uganda
20	Nauru
21	Monaco
22	Egypt

Pic 6.4 Query: `select location, max(cast(total_deaths as int)) as death_count`
            `from COVID_Deaths`
            `where continent is not null`
            `group by location`
            `order by death_count desc`

	location	death_count
1	United States	1049749
2	Brazil	684425
3	India	528121
4	Russia	377273
5	Mexico	329652
6	Peru	215982
7	United Kingdom	205718
8	Italy	176098
9	Indonesia	157729
10	France	154506
11	Germany	148217
12	Iran	144108
13	Colombia	141708
14	Argentina	129769
15	Poland	117219
16	Ukraine	116681
17	Spain	112931
18	South Africa	102129
19	Turkey	100840
20	Romania	66825
21	Philippines	62206
22	Chile	60696

Pic 7. Query: `select location, population, max(cast(total_deaths as int)) as death_count,`
`(max(cast(total_deaths as int))/population)*100 as death_rate_to_population`
`from COVID_Deaths`
`where continent is not null`
`group by location, population`
`order by death_rate_to_population desc`

	location	population	death_count	death_rate_to_population
1	Peru	33715472	215982	0.640602035765657
2	Bulgaria	6885868	37638	0.546597756448425
3	Bosnia and Herzegovina	3270943	16086	0.491784784999311
4	Hungary	9709786	47367	0.487827435125759
5	North Macedonia	2103330	9503	0.451807372119449
6	Georgia	3757980	16889	0.449416974012634
7	Montenegro	627859	2778	0.442456029140301
8	Croatia	4060135	16772	0.413089712534189
9	Czechia	10510750	40898	0.389106391075803
10	Moldova	3061506	11783	0.384875940141878
11	Slovakia	5447622	20236	0.371464833646681
12	San Marino	33746	118	0.349671072127067
13	Romania	19328560	66825	0.345731911740968
14	Lithuania	2786651	9297	0.333626277564001
15	Gibraltar	32670	108	0.330578512396694
16	Slovenia	2119410	6789	0.320324996107407

Pic 8. Query: `select date, sum(new_cases) as total_cases, sum(cast(new_deaths as int)) as`
`total_deaths,`
`sum(cast(new_deaths as int))/sum(new_cases)*100 as death_percentage`
`from COVID_Deaths`
`where continent is not null`
`group by date`
`order by 1,2`

	date	total_cases	total_deaths	death_percentage
31	2020-01-31 00:00:00.000	1690	42	2.48520710059172
32	2020-02-01 00:00:00.000	2111	46	2.17906205589768
33	2020-02-02 00:00:00.000	4749	104	2.1899347230996
34	2020-02-03 00:00:00.000	3100	64	2.06451612903226
35	2020-02-04 00:00:00.000	4012	66	1.64506480558325
36	2020-02-05 00:00:00.000	3745	72	1.92256341789052
37	2020-02-06 00:00:00.000	3162	70	2.21378874130297
38	2020-02-07 00:00:00.000	3533	85	2.40588734786301
39	2020-02-08 00:00:00.000	2731	87	3.18564628341267
40	2020-02-09 00:00:00.000	3028	100	3.30250990752972
41	2020-02-10 00:00:00.000	2538	107	4.21591804570528
42	2020-02-11 00:00:00.000	2043	100	4.89476260401371
43	2020-02-12 00:00:00.000	378	5	1.32275132275132
44	2020-02-13 00:00:00.000	15152	253	1.66974656810982
45	2020-02-14 00:00:00.000	6485	152	2.34387047031611
46	2020-02-15 00:00:00.000	2076	143	6.88824662813102
47	2020-02-16 00:00:00.000	2113	104	4.9219119734974
48	2020-02-17 00:00:00.000	1936	98	5.06198347107438

Pic 9. Query: `select dead.continent, dead.location, dead.date, dead.population,`
        `vacc.new_vaccinations`
        `from COVID_Deaths dead`
        `join COVID_Vaccinations vacc`
        `on dead.location = vacc.location and`
        `dead.date = vacc.date`
        `where dead.continent is not null`
        `order by 2,3`

	continent	location	date	population	new_vaccinations
71549	Europe	Greece	2020-12-21 00:00:00.000	10445365	NULL
71550	Europe	Greece	2020-12-22 00:00:00.000	10445365	NULL
71551	Europe	Greece	2020-12-23 00:00:00.000	10445365	NULL
71552	Europe	Greece	2020-12-24 00:00:00.000	10445365	NULL
71553	Europe	Greece	2020-12-25 00:00:00.000	10445365	NULL
71554	Europe	Greece	2020-12-26 00:00:00.000	10445365	NULL
71555	Europe	Greece	2020-12-27 00:00:00.000	10445365	NULL
71556	Europe	Greece	2020-12-28 00:00:00.000	10445365	NULL
71557	Europe	Greece	2020-12-29 00:00:00.000	10445365	630
71558	Europe	Greece	2020-12-30 00:00:00.000	10445365	645
71559	Europe	Greece	2020-12-31 00:00:00.000	10445365	617
71560	Europe	Greece	2021-01-01 00:00:00.000	10445365	264
71561	Europe	Greece	2021-01-02 00:00:00.000	10445365	546
71562	Europe	Greece	2021-01-03 00:00:00.000	10445365	720
71563	Europe	Greece	2021-01-04 00:00:00.000	10445365	6458
71564	Europe	Greece	2021-01-05 00:00:00.000	10445365	6304
71565	Europe	Greece	2021-01-06 00:00:00.000	10445365	3985

Pic 10.1 Query: `select location, min(date) from COVID_Vaccinations`
        `where new_vaccinations is not null and date < '2021-01-01 00:00:00.000' and`
        `continent is not null`
        `group by location`
        `order by 2`

	location	(No column name)
1	Norway	2020-12-09 00:00:00.000
2	Canada	2020-12-15 00:00:00.000
3	Denmark	2020-12-16 00:00:00.000
4	Latvia	2020-12-16 00:00:00.000
5	Israel	2020-12-20 00:00:00.000
6	Switzerland	2020-12-22 00:00:00.000
7	Bahrain	2020-12-24 00:00:00.000
8	Chile	2020-12-25 00:00:00.000
9	Romania	2020-12-28 00:00:00.000
10	France	2020-12-28 00:00:00.000
11	Italy	2020-12-28 00:00:00.000
12	Germany	2020-12-28 00:00:00.000
13	Lithuania	2020-12-28 00:00:00.000
14	Estonia	2020-12-28 00:00:00.000
15	Mexico	2020-12-28 00:00:00.000
16	Czechia	2020-12-28 00:00:00.000
17	Austria	2020-12-28 00:00:00.000
18	Slovenia	2020-12-28 00:00:00.000
19	Ireland	2020-12-29 00:00:00.000
20	Greece	2020-12-29 00:00:00.000

Pic 10.2 Query:
```sql
select dead.continent, dead.location, dead.date, dead.population,
vacc.new_vaccinations, sum(cast(vacc.new_vaccinations as bigint))
over (partition by dead.location order by dead.location, dead.date) as
total_vaccinations_over_time
from COVID_Deaths dead
join COVID_Vaccinations vacc
on dead.location = vacc.location and
dead.date = vacc.date
where dead.continent is not null and dead.date between '2020-12-09
00:00:00.000' and (select max(date) from COVID_Deaths)
order by 2, 3
```

	continent	location	date	population	new_vaccinations	total_vaccinations_over_time
95...	Oceania	Northern ...	2021-12-03 00:00:00.000	49481	NULL	NULL
95...	Oceania	Northern ...	2021-12-04 00:00:00.000	49481	NULL	NULL
95...	Europe	Norway	2020-12-09 00:00:00.000	5403021	1	1
95...	Europe	Norway	2020-12-10 00:00:00.000	5403021	1	2
95...	Europe	Norway	2020-12-11 00:00:00.000	5403021	4	6
95...	Europe	Norway	2020-12-12 00:00:00.000	5403021	1	7
95...	Europe	Norway	2020-12-13 00:00:00.000	5403021	NULL	7
95...	Europe	Norway	2020-12-14 00:00:00.000	5403021	NULL	7
95...	Europe	Norway	2020-12-15 00:00:00.000	5403021	NULL	7
95...	Europe	Norway	2020-12-16 00:00:00.000	5403021	5	12
95...	Europe	Norway	2020-12-17 00:00:00.000	5403021	9	21
95...	Europe	Norway	2020-12-18 00:00:00.000	5403021	15	36
95...	Europe	Norway	2020-12-19 00:00:00.000	5403021	5	41
95...	Europe	Norway	2020-12-20 00:00:00.000	5403021	7	48
95...	Europe	Norway	2020-12-21 00:00:00.000	5403021	14	62

Pic 11. Query:
```sql
select sum(convert(bigint ,vacc.new_vaccinations)) as sum_of_vaccs
from COVID_Deaths dead
join COVID_Vaccinations vacc
on dead.location = vacc.location and dead.date = vacc.date
where dead.continent is not null
```

	sum_of_vaccs
1	9964525692

Pic 12. Query:
```sql
with Pops_vs_Vacc (Continent, Location, Date, Population, New_Vaccinations,
Total_Vaccinations) as
(
select dead.continent, dead.location, dead.date, dead.population,
vacc.new_vaccinations, sum(cast(vacc.new_vaccinations as bigint)) over
(partition by dead.location order by
dead.location, dead.date) as total_vaccinations_over_time
from COVID_Deaths dead
join COVID_Vaccinations vacc
        on dead.location = vacc.location and
dead.date = vacc.date
where dead.continent is not null and dead.date between '2020-12-09
00:00:00.000' and (select max(date) from COVID_Deaths)
)
select *, (Total_Vaccinations/Population)*100 as Vaccinations_Percentage
from Pops_vs_Vacc
```

	Continent	Location	Date	Population	New_Vaccinations	Total_Vaccinations	Vaccinations_Percentage
18861	Asia	Brunei	2021-11-03 00:00:00.000	445373	NULL	108180	24.2897526343088
18862	Asia	Brunei	2021-11-04 00:00:00.000	445373	NULL	108180	24.2897526343088
18863	Asia	Brunei	2021-11-05 00:00:00.000	445373	NULL	108180	24.2897526343088
18864	Asia	Brunei	2021-11-06 00:00:00.000	445373	568	108748	24.4172861848383
18865	Asia	Brunei	2021-11-07 00:00:00.000	445373	7257	116005	26.0467069175725
18866	Asia	Brunei	2021-11-08 00:00:00.000	445373	2510	118515	26.6102794736098
18867	Asia	Brunei	2021-11-09 00:00:00.000	445373	5935	124450	27.9428703581043
18868	Asia	Brunei	2021-11-10 00:00:00.000	445373	5321	129771	29.1375992707236
18869	Asia	Brunei	2021-11-11 00:00:00.000	445373	6729	136500	30.648467688881
18870	Asia	Brunei	2021-11-12 00:00:00.000	445373	7035	143535	32.2280425620772
18871	Asia	Brunei	2021-11-13 00:00:00.000	445373	5327	148862	33.4241186600894
18872	Asia	Brunei	2021-11-14 00:00:00.000	445373	7179	156041	35.0360259827156
18873	Asia	Brunei	2021-11-15 00:00:00.000	445373	3941	159982	35.9209022549638
18874	Asia	Brunei	2021-11-16 00:00:00.000	445373	7617	167599	37.631154111273
18875	Asia	Brunei	2021-11-17 00:00:00.000	445373	7307	174906	39.2718013889481
18876	Asia	Brunei	2021-11-18 00:00:00.000	445373	7656	182562	40.9908099503113
18877	Asia	Brunei	2021-11-19 00:00:00.000	445373	8701	191263	42.944453300941
18878	Asia	Brunei	2021-11-20 00:00:00.000	445373	3121	194384	43.6452142361571
18879	Asia	Brunei	2021-11-21 00:00:00.000	445373	7807	202191	45.398126963242
18880	Asia	Brunei	2021-11-22 00:00:00.000	445373	1719	203910	45.784095578313
18881	Asia	Brunei	2021-11-23 00:00:00.000	445373	5609	209519	47.0434893897924
18882	Asia	Brunei	2021-11-24 00:00:00.000	445373	4047	213566	47.9521659373155
18883	Asia	Brunei	2021-11-25 00:00:00.000	445373	NULL	213566	47.9521659373155
18884	Asia	Brunei	2021-11-26 00:00:00.000	445373	NULL	213566	47.9521659373155
18885	Asia	Brunei	2021-11-27 00:00:00.000	445373	453	214019	48.0538784344808

Pic 13.1 Query:
```
drop table if exists #Vaccinated_Population_Percentage
create table #Vaccinated_Population_Percentage
(
Continent nvarchar(255),
Location nvarchar(255),
Date datetime,
Population numeric,
New_Vaccinations numeric,
Total_Vaccinations numeric
)
```

'Command Completed Successfully!' No screenshot needed :)

Pic 13.2 Query:
```
insert into #Vaccinated_Population_Percentage
select dead.continent, dead.location, dead.date, dead.population,
vacc.new_vaccinations,
sum(cast(vacc.new_vaccinations as bigint)) over (partition by dead.location
order by
dead.location, dead.date) as total_vaccinations_over_time
from COVID_Deaths dead
join COVID_Vaccinations vacc
on dead.location = vacc.location and
dead.date = vacc.date
where dead.continent is not null and dead.date between '2020-12-09
00:00:00.000' and (select max(date) from COVID_Deaths)
order by 2, 3
```

'Command Completed Successfully!' No screenshot needed :)

Pic 13.3 Query:
```
select *, (Total_Vaccinations/Population)*100 as Vaccinations_Percentage
from #Vaccinated_Population_Percentage
```

The exploration can go on and on for a long time, but it is time to prepare some data for our Tableau visualizations. 4 queries were made, in which the results are saved in also 4 excel files. I'm sure Tableau can connect with various popular databases but unfortunately, that feature isn't available with Tableau Public. Thus, manual data extraction and input is the only choice.

Queries for Tableau:

Q1. :
```
select sum(new_cases) as total_cases, sum(cast(new_deaths as int)) as
total_deaths,
sum(cast(new_deaths as bigint))/sum(new_cases)*100 as Death_Percentage
from COVID_Deaths
where continent is not null
order by 1, 2
```

Q2. :
```
select location, sum(cast(new_deaths as bigint)) as total_death_count
from COVID_Deaths
where continent is null
and location not in ('World', 'European Union', 'International') and location not
like '%income'
group by location
order by total_death_count desc
```

Q3. :
```
select location, population, max(total_cases) as max_infection_count,
max((total_cases/population))*100 as percent_population_infected
from COVID_Deaths
group by location, population
```

Q4. :
```
select location, population, date, max(total_cases) as max_infection_count,
max((total_cases/population))*100 as percent_population_infected
from COVID_Deaths
group by location, population, date
order by 1, 3 desc
```