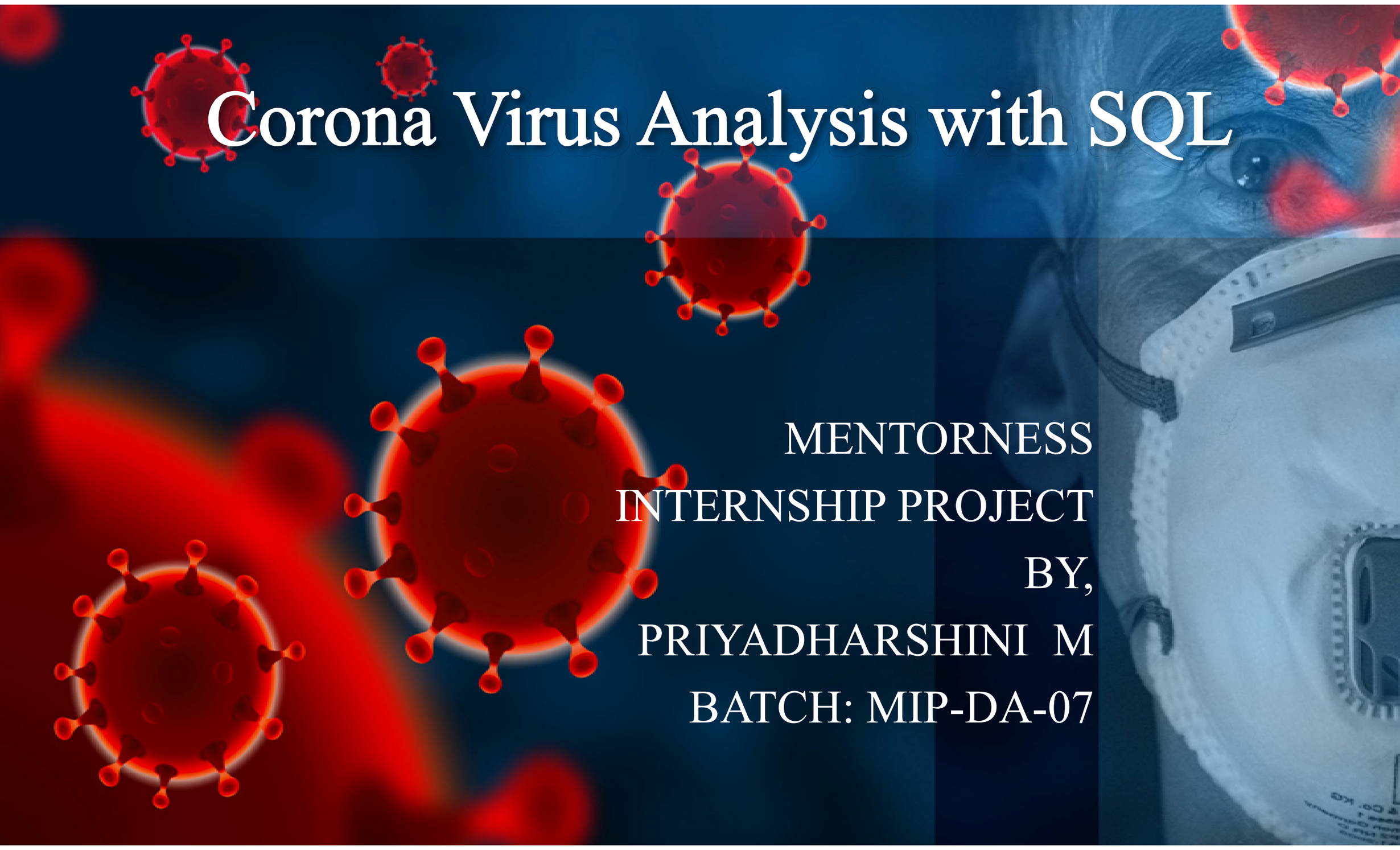
Corona Virus Analysis with SQL



MENTORNESS

INTERNSHIP PROJECT

BY,

PRIYADHARSHINI M

BATCH: MIP-DA-07

CONTENT

- Project overview
- Dataset Description
- Data Exploration and Analysis(Queries)

Project Overview

- The corona virus pandemic had a significant impact on public health and has created an urgent need for data-driven insights to understand the spread of the virus.
- As a data analyst, we are analyzing a corona virus dataset to derive meaningful insights and present our findings.

DATASET

Description of each column in dataset:

- **Province**: Geographic subdivision with a country/Region.
- **Country/Region**: Geographic entity where data is recorded.
- **Latitude**: North-South position on Earth's surface.
- **Longitude**: East-west position on Earth's surface.
- **Date**: Recorded date of Corona Virus related deaths.
- **Confirmed**: Number of diagnosed Corona Virus cases.
- **Deaths**: Number of Corona Virus related deaths.
- **Recovered**: Number of recovered Corona Virus cases.

DATASET DESCRIPTION

Query | Query History

```
1  select* from corona virus
```

Data Output | Messages | Notifications

	province character varying	country character varying	latitude numeric	longitude numeric	date timestamp without time zone	confirmed integer	deaths integer	recovered integer
1	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-22 00:00:00	0	0	0
2	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-23 00:00:00	0	0	0
3	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-24 00:00:00	0	0	0
4	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-25 00:00:00	0	0	0
5	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-26 00:00:00	0	0	0
6	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-27 00:00:00	0	0	0
7	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-28 00:00:00	0	0	0
8	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-29 00:00:00	0	0	0
9	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-30 00:00:00	0	0	0
10	Afghanistan	Afghanistan	33.93911	67.709953	2020-01-31 00:00:00	0	0	0
11	Afghanistan	Afghanistan	33.93911	67.709953	2020-02-01 00:00:00	0	0	0
12	Afghanistan	Afghanistan	33.93911	67.709953	2020-02-02 00:00:00	0	0	0
13	Afghanistan	Afghanistan	33.93911	67.709953	2020-02-03 00:00:00	0	0	0
14	Afghanistan	Afghanistan	33.93911	67.709953	2020-02-04 00:00:00	0	0	0
15	Afghanistan	Afghanistan	33.93911	67.709953	2020-02-05 00:00:00	0	0	0

DATA EXPLORATION AND ANALYSIS

QUERY-1

```sql
--Q1: Write a code to check the missing values

select * from corona_virus
where province is null or country is null
or latitude is null or longitude is null
or date is null or confirmed is null
or deaths is null or recovered is null;
```

Data Output | Messages | Notifications

	province character varying	country character varying	latitude numeric	longitude numeric	date timestamp without time zone	confirmed integer	deaths integer	recovered integer

DATA EXPLORATION AND ANALYSIS

QUERY-2

```sql
--Q2: If null values are present, update them with zeros for all columns

update corona_virus set province= coalesce(province,''),
country= coalesce(country,''),
latitude = coalesce(latitude,0),
longitude= coalesce(longitude,0),
confirmed= coalesce(confirmed,0),
deaths= coalesce(deaths,0),
recovered= coalesce(recovered,0);
```

DATA EXPLORATION AND ANALYSIS

QUERY-3

```
--Q3: Check total number of rows

select count(*) from corona_virus;
```

Data Output | Messages | Notifications

	count bigint
1	78386

DATA EXPLORATION AND ANALYSIS

QUERY-4

```sql
--Q4: Check what is start date

select date from corona_virus
order by date
limit 1;
```

Data Output | Messages | Notifications

	date timestamp without time zone
1	2020-01-22 00:00:00

```sql
--Check what is end date

Select date from corona_virus
order by date desc
limit 1;
```

Data Output | Messages | Notifications

	date timestamp without time zone
1	2021-06-13 00:00:00

DATA EXPLORATION AND ANALYSIS

QUERY-5

```
--Q5: Number of months present in Dataset

Select Extract(month from date) as Number_of_month
from corona_virus
group by number_of_month
order by number_of_month;
```

Data Output | Messages | Notifications

	number_of_month numeric
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12

DATA EXPLORATION AND ANALYSIS

QUERY-6

```
--Q6: Find monthly average for confirmed,deaths,recovered

select Extract(month from date) as month,
avg(confirmed) as avg_confirmed,
avg(deaths) as avg_deaths,
avg(recovered) as avg_recovered
from corona_virus
group by month
order by month
```

Data Output | Messages | Notifications

	month numeric	avg_confirmed numeric	avg_deaths numeric	avg_recovered numeric
1	1	2958.2814380741210010	63.6811846689895470	1451.4554957237884067
2	2	1203.1187058555479608	34.2777398040555935	769.1034404192298929
3	3	1538.9637620444072057	33.9302471721826561	840.0799120234604106
4	4	2602.5778138528138528	59.9805194805194805	1623.2136363636363636
5	5	2290.0519480519480519	53.5305823209049016	2162.9020737327188940
6	6	1357.8852310480217457	40.8356991845363938	1220.1532769556025370
7	7	1432.3611227482195224	35.1095517385839966	983.0582320904901550
8	8	1611.8428990364474235	37.5366568914956012	1299.2947214076246334
9	9	1784.5874458874458874	34.7772727272727273	1438.9067099567099567
10	10	2412.1996229576874738	36.7582739840804357	1420.6430666108085463
11	11	3592.1943722943722944	56.7634199134199134	1985.3445887445887446
12	12	4050.4396732299958106	71.2182656053623796	2497.8850020946795140

DATA EXPLORATION AND ANALYSIS

QUERY-7

```
--Q7: Find most frequent value for confirmed,deaths,recovered each month

select extract(month from date) as month,
max(confirmed) as most_frequent_confirmed,
max(deaths) as most_frequent_deaths,
max(recovered) as most_frequent_recovered
from corona_virus
group by month
order by month;
```

Data Output | Messages | Notifications

	month numeric	most_frequent_confirmed integer	most_frequent_deaths integer	most_frequent_recovered integer
1	1	300462	4475	87090
2	2	134975	3907	98389
3	3	100158	3869	102138
4	4	401993	4249	299988
5	5	414188	4529	422436
6	6	134154	7374	231456
7	7	75866	1595	140050
8	8	85687	1505	95881
9	9	97894	1703	101468
10	10	99264	3351	388340
11	11	207933	2259	139292
12	12	823225	3752	1123456

DATA EXPLORATION AND ANALYSIS

QUERY-8

```
--Q8: Find minimum values for confirmed,deaths,recovered per year

select extract(year from date) as year,
min(confirmed) as min_confirmed,
min(deaths) as min_deaths,
min(recovered) as min_recovered
from corona_virus
group by year
order by year;
```

Data Output | Messages | Notifications

	year numeric	min_confirmed integer	min_deaths integer	min_recovered integer
1	2020	0	0	0
2	2021	0	0	0

DATA EXPLORATION AND ANALYSIS

QUERY-9

```
--Q9: Find maximum values for confirmed,deaths,recovered per year

select extract(year from date) as year,
max(confirmed) as max_confirmed,
max(deaths) as max_deaths,
max(recovered) as max_recovered
from corona_virus
group by year
order by year;
```

Data Output | Messages | Notifications

	year numeric	max_confirmed integer	max_deaths integer	max_recovered integer
1	2020	823225	3752	1123456
2	2021	414188	7374	422436

DATA EXPLORATION AND ANALYSIS

QUERY-10

```
--Q10: The total number of case of confirmed,deaths,recovered each month

select extract(month from date) as month,
count(confirmed) as total_confirmed,
count(deaths) as total_deaths,
count(recovered) as total_recovered
from corona_virus
group by month
order by month;
```

Data Output | Messages | Notifications

	month numeric	total_confirmed bigint	total_deaths bigint	total_recovered bigint
1	1	6314	6314	6314
2	2	8778	8778	8778
3	3	9548	9548	9548
4	4	9240	9240	9240
5	5	9548	9548	9548
6	6	6622	6622	6622
7	7	4774	4774	4774
8	8	4774	4774	4774
9	9	4620	4620	4620
10	10	4774	4774	4774
11	11	4620	4620	4620
12	12	4774	4774	4774

DATA EXPLORATION AND ANALYSIS

QUERY-11

```
--Q11: Check how the corona virus spread out with respect to confirmed case

select count(confirmed) as total_confirmed_cases,
avg(confirmed) as avg_confirmed,
variance(confirmed) as variance_confirmed,
STDDEV(confirmed) as std_confirmed
from corona_virus;
```

Data Output | Messages | Notifications

	total_confirmed_cases bigint	avg_confirmed numeric	variance_confirmed numeric	std_confirmed numeric
1	78386	2156.8283111780164825	157290931.69817455	12541.56815148

DATA EXPLORATION AND ANALYSIS

QUERY-12

```
--Q12: Check how corona virus spread out with respect to death case per month

select extract(month from date) as month,
count(deaths) as total_deaths_per_month,
avg(deaths) as avg_deaths_per_months,
variance(deaths) as var_deaths_per_months,
stddev(deaths) as std_deaths_per_months
from corona_virus
group by month
order by month;
```

Data Output | Messages | Notifications

	month numeric	total_deaths_per_month bigint	avg_deaths_per_months numeric	var_deaths_per_months numeric	std_deaths_per_months numeric
1	1	6314	63.6811846689895470	79012.044546925182	281.090811921922
2	2	8778	34.2777398040555935	34852.618305840004	186.688559654415
3	3	9548	33.9302471721826561	29785.052429518872	172.583465110418
4	4	9240	59.9805194805194805	67905.924720587346	260.587652663336
5	5	9548	53.5305823209049016	76775.779414471802	277.084426510174
6	6	6622	40.8356991845363938	46250.187470278323	215.058567535168
7	7	4774	35.1095517385839966	21144.584057079556	145.411774134970
8	8	4774	37.5366568914956012	23277.872425108734	152.570876726552
9	9	4620	34.7772727272727273	20107.121414513177	141.799581855918
10	10	4774	36.7582739840804357	17583.754252708491	132.603749014530
11	11	4620	56.7634199134199134	27779.806542101184	166.672752848512
12	12	4774	71.2182656053623796	65359.059829716994	255.654180153028

DATA EXPLORATION AND ANALYSIS

QUERY-13

```
--Q13: Check how corona virus spread out with respect to recovered case

Select count(recovered) as total_recovered,
avg(recovered) as avg_recovered,
variance(recovered) as var_recovered,
stddev(recovered) as std_recovered
from corona_virus;
```

Data Output | Messages | Notifications

	total_recovered bigint	avg_recovered numeric	var_recovered numeric	std_recovered numeric
1	78386	1442.7263541959023295	107030888.69602982	10345.57338653

DATA EXPLORATION AND ANALYSIS

QUERY-14

```
--Q14: Find country having highest number of the confirmed case

Select country,max(confirmed) as highest_confirmed_case from corona_virus
group by country
order by highest_confirmed_case desc
limit 1;
```

Data Output | Messages | Notifications

	country character varying	highest_confirmed_case integer
1	Turkey	823225

DATA EXPLORATION AND ANALYSIS

QUERY-15

```
--Q15: Find Country having lowest number of the death case

select country, min(deaths) as lowest_deaths_case from corona_virus
group by country
order by lowest_deaths_case
```

Data Output | Messages | Notifications

	country character varying	lowest_deaths_case integer
104	Panama	0
105	Yemen	0
106	South Sudan	0
107	Lithuania	0
108	Bulgaria	0
109	Croatia	0
110	Tunisia	0
111	North Macedonia	0
112	Morocco	0
113	Mexico	0
114	Nepal	0
115	Tanzania	0
116	Poland	0
117	Lebanon	0
118	Costa Rica	0
119	Haiti	0
120	Samoa	0
121	Somalia	0

DATA EXPLORATION AND ANALYSIS

QUERY-16

```
--Q16: Find top 5 countries having highest recovered case

select country,max(recovered) as highest_recovered_case from corona_virus
group by country
order by highest_recovered_case desc
limit 5;
```

Data Output | Messages | Notifications

	country character varying	highest_recovered_case integer
1	Turkey	1123456
2	India	422436
3	Brazil	388340
4	US	150267
5	Colombia	89557

THANK YOU!