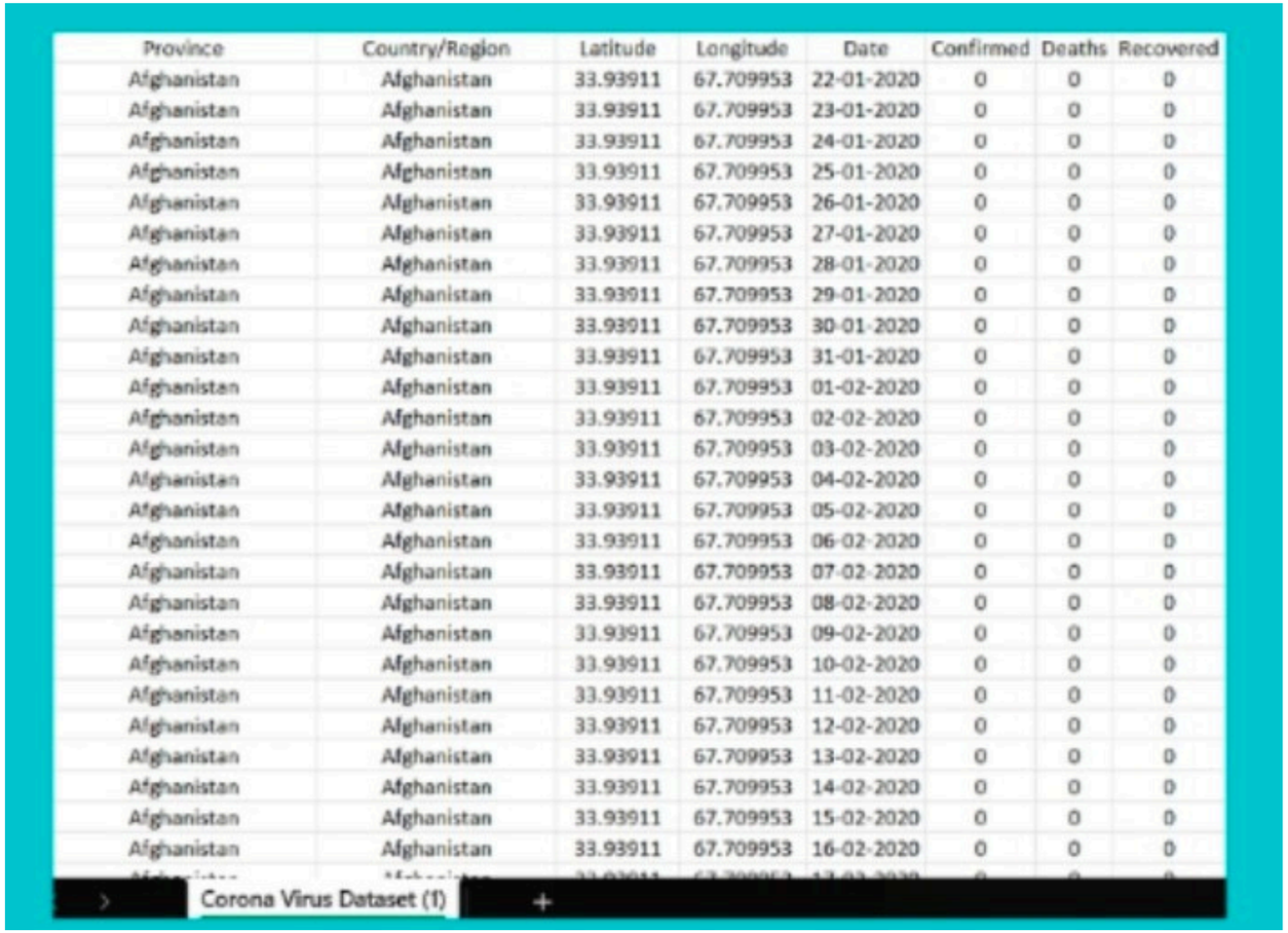
Cv analysis with SQL

Presented by:- J.shirisha
batch no :- MIP-DA-09
profile name :- Data analysis

carona virus Data set

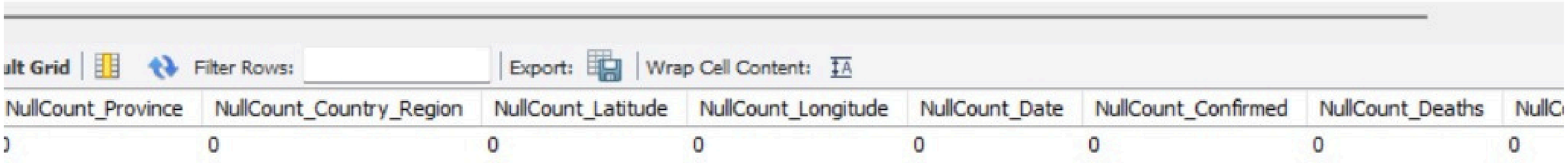


Province	Country/Region	Latitude	Longitude	Date	Confirmed	Deaths	Recovered
Afghanistan	Afghanistan	33.93911	67.709953	22-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	23-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	24-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	25-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	26-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	27-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	28-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	29-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	30-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	31-01-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	01-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	02-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	03-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	04-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	05-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	06-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	07-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	08-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	09-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	10-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	11-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	12-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	13-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	14-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	15-02-2020	0	0	0
Afghanistan	Afghanistan	33.93911	67.709953	16-02-2020	0	0	0

Corona Virus Dataset (1)

1Q:- write a code to check null values ?

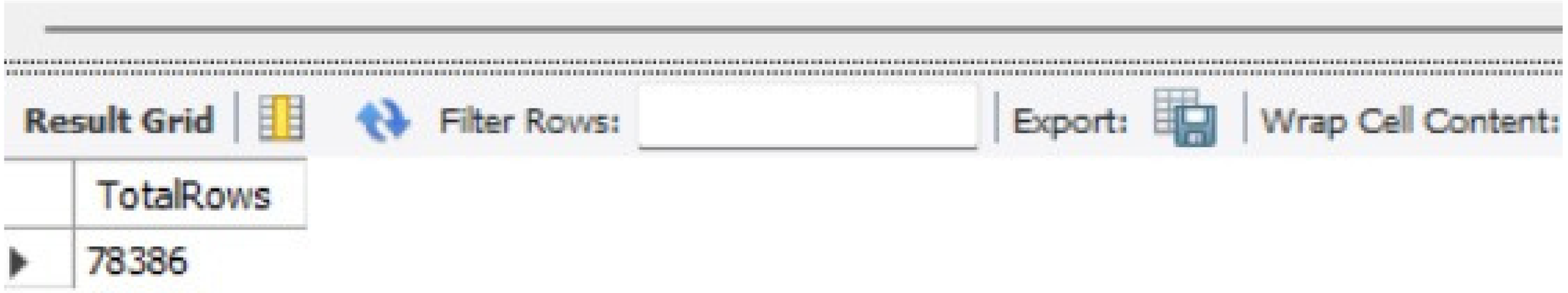
```sql
SELECT
    SUM(CASE WHEN Province IS NULL THEN 1 ELSE 0 END) AS NullCount_Province,
    SUM(CASE WHEN `Country/Region` IS NULL THEN 1 ELSE 0 END) AS NullCount_Country_Region,
    SUM(CASE WHEN Latitude IS NULL THEN 1 ELSE 0 END) AS NullCount_Latitude,
    SUM(CASE WHEN Longitude IS NULL THEN 1 ELSE 0 END) AS NullCount_Longitude,
    SUM(CASE WHEN Date IS NULL THEN 1 ELSE 0 END) AS NullCount_Date,
    SUM(CASE WHEN Confirmed IS NULL THEN 1 ELSE 0 END) AS NullCount_Confirmed,
    SUM(CASE WHEN Deaths IS NULL THEN 1 ELSE 0 END) AS NullCount_Deaths,
    SUM(CASE WHEN Recovered IS NULL THEN 1 ELSE 0 END) AS NullCount_Recovered
FROM
    covid.corona;
```



NullCount_Province	NullCount_Country_Region	NullCount_Latitude	NullCount_Longitude	NullCount_Date	NullCount_Confirmed	NullCount_Deaths	NullC
)	0	0	0	0	0	0	0

2Q:- check total number of rows ?

```
1 •    SELECT
2          COUNT(*) AS TotalRows from covid.corona
```



3Q:- check what is start, date &end time ?

```sql
SELECT
    MIN(Date) AS StartDate,
    MAX(Date) AS EndDate
FROM
    covid.corona;
```

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

StartDate	EndDate
01-01-2021	31-12-2020

4Q:- Number of month present in dataset ?

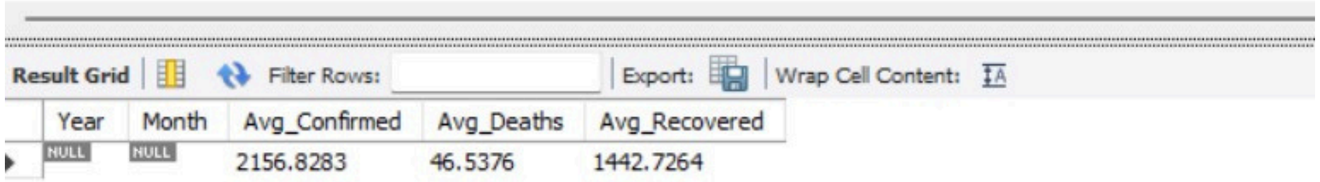
```sql
SELECT
    COUNT(DISTINCT YEAR(Date) * 100 + MONTH(Date)) AS NumberOfMonths
FROM
    covid.corona;

```

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

NumberOfMonths
0

5Q:- find monthly average for confirmed death's, recovered ?

```sql
SELECT
    YEAR(Date) AS Year,
    MONTH(Date) AS Month,
    AVG(Confirmed) AS Avg_Confirmed,
    AVG(Deaths) AS Avg_Deaths,
    AVG(Recovered) AS Avg_Recovered
FROM
    covid.corona
GROUP BY
    Year, Month
ORDER BY
    Year, Month;
```



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

Year	Month	Avg_Confirmed	Avg_Deaths	Avg_Recovered
NULL	NULL	2156.8283	46.5376	1442.7264

6Q:- find minimum values for confirmed death's, recovered per year ?

```sql
SELECT
    YEAR(Date) AS Year,
    MIN(Confirmed) AS Min_Confirmed,
    MIN(Deaths) AS Min_Deaths,
    MIN(Recovered) AS Min_Recovered
FROM
    covid.corona
GROUP BY
    Year;
```

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

Year	Min_Confirmed	Min_Deaths	Min_Recovered
NULL	0	0	0

7Q:- find maximum values of confirmed death's, recovered per year ?

```sql
SELECT
    YEAR(Date) AS Year,
    MAX(Confirmed) AS Max_Confirmed,
    MAX(Deaths) AS Max_Deaths,
    MAX(Recovered) AS Max_Recovered
FROM
    covid.corona
GROUP BY
    Year;
```

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

Year	Max_Confirmed	Max_Deaths	Max_Recovered
NULL	823225	7374	1123456

8Q:- The total number of cases confirmed death's, recovered each month?

```sql
SELECT
    YEAR(Date) AS Year,
    MONTH(Date) AS Month,
    SUM(Confirmed) AS Total_Confirmed,
    SUM(Deaths) AS Total_Deaths,
    SUM(Recovered) AS Total_Recovered
FROM
    covid.corona
GROUP BY
    Year, Month
ORDER BY
    Year, Month;
```

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

Year	Month	Total_Confirmed	Total_Deaths	Total_Recovered
NULL	NULL	169065144	3647894	113089548

9Q:-check how carona virus spread out with respect to confirmed case ?

```sql
SELECT
    SUM(Confirmed) AS Total_Confirmed,
    AVG(Confirmed) AS Avg_Confirmed,
    VARIANCE(Confirmed) AS Variance_Confirmed,
    STDDEV(Confirmed) AS Stdev_Confirmed
FROM
    covid.corona;
```

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

Total_Confirmed	Avg_Confirmed	Variance_Confirmed	Stdev_Confirmed
169065144	2156.8283	157288925.07796532	12541.488152446875

10Q:- check how carona virus spread out with respect to death cases per month ?

```sql
SELECT
    YEAR(Date) AS Year,
    MONTH(Date) AS Month,
    SUM(Deaths) AS Total_Deaths,
    AVG(Deaths) AS Avg_Deaths,
    VARIANCE(Deaths) AS Variance_Deaths,
    STDDEV(Deaths) AS Stdev_Deaths
FROM
    covid.corona
GROUP BY
    Year, Month
ORDER BY
    Year, Month;
```

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

Year	Month	Total_Deaths	Avg_Deaths	Variance_Deaths	Stdev_Deaths
NULL	NULL	3647894	46.5376	45892.01885355753	214.22422564583476

11Q:-check how carona virus spread out with respect to recovered case ?

```sql
SELECT
    SUM(Recovered) AS Total_Recovered,
    AVG(Recovered) AS Avg_Recovered,
    VARIANCE(Recovered) AS Variance_Recovered,
    STDDEV(Recovered) AS Stdev_Recovered
FROM
    covid.corona;
```

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

Total_Recovered	Avg_Recovered	Variance_Recovered	Stdev_Recovered
113089548	1442.7264	107029523.26229636	10345.507395110999

12Q:- find highest confirmed cases in country's

```sql
SELECT
    `Country/Region`,
    SUM(Confirmed) AS Total_Confirmed
FROM
    covid.corona
GROUP BY
    `Country/Region`
ORDER BY
    Total_Confirmed DESC
LIMIT 1;
```

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

Country/Region	Total_Confirmed
US	33461982

13Q:- find country having lowest number of the confirmed cases?

```sql
SELECT
    `Country/Region`,
    SUM(Deaths) AS Total_Deaths
FROM
    covid.corona
GROUP BY
    `Country/Region`
ORDER BY
    Total_Deaths ASC
LIMIT 1;
```

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

Country/Region	Total_Deaths
Dominica	0

14Q:- find top countries having highest recovered cases

```sql
SELECT
    `Country/Region`,
    SUM(Recovered) AS Total_Recovered
FROM
    covid.corona
GROUP BY
    `Country/Region`
ORDER BY
    Total_Recovered DESC
LIMIT 5;
```

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

Country/Region	Total_Recovered
India	28089649
Brazil	15400169
US	6303715
Turkey	5202251
Russia	4745756

Thank you