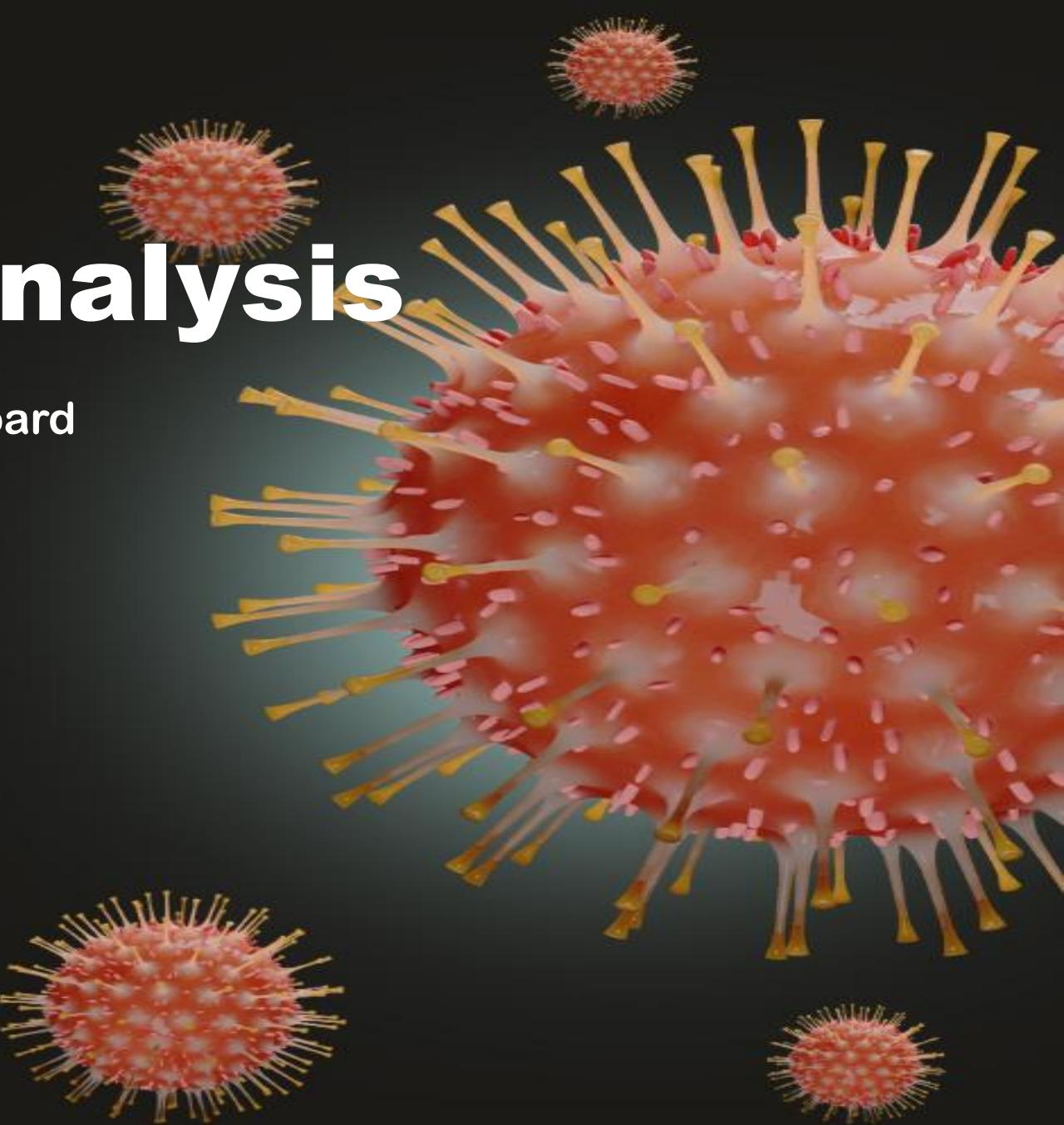


COVID-19 Data Analysis

Data Cleaning, Analysis, and Dashboard

Date: 25-04-2025

Team Members: Shikha Chaurasia
Arajita Kumari
RaviRaj



Project Goal

Our goal was to study COVID-19 data to understand how it affected people, countries, and death rate.



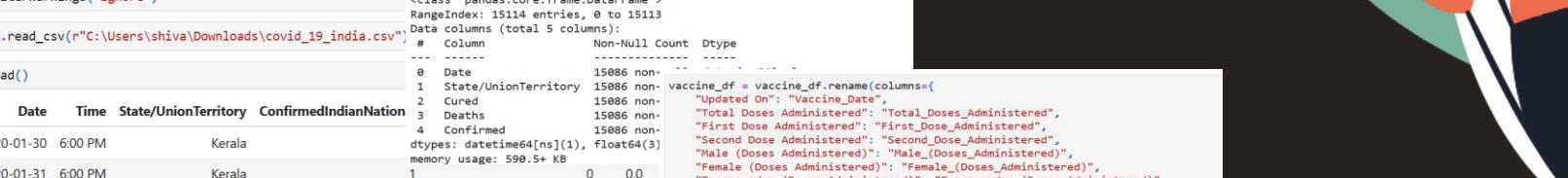
My Work- Data Collection & Cleaning

Tools Used: Python

What I Did:

- Collected raw COVID-19 data (from  <https://data.gov.in>)
 - Removed null values
 - Handled duplicates
 - Converted data types
 - Renamed columns for clarity
 - Saved clean data in **cleaned_data.csv**





```
import pandas as pd
import numpy as np

import warnings
warnings.filterwarnings("ignore")

Covid_df=pd.read_csv(r"C:\Users\shiva\Downloads\covid_19_india.csv") # Change the data type
Covid_df[\"Date\"] = pd.to_datetime(Covid_df[\"Date\"], format=\"%Y-%m-%d\")

Covid_df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15114 entries, 0 to 15113
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Date        15086 non-    object 
 1   State/UnionTerritory 15086 non-    object 
 2   Cured       15086 non-    float64
 3   Deaths      15086 non-    float64
 4   Confirmed    15086 non-    float64
dtypes: datetime64[ns](1), float64(3)
memory usage: 590.5+ KB

Covid_df.head()
Sno Date Time State/UnionTerritory Confirmed IndianNation
0 1.0 2020-01-30 6:00 PM Kerala
1 2.0 2020-01-31 6:00 PM Kerala
2 3.0 2020-02-01 6:00 PM Kerala

# Adding new column as Active cases
Covid_df[\"Active_cases\"] = Covid_df[\"Confirmed\"] - (Covid_df[\"Cured\"] + Covid_df[\"Deaths\"] + Covid_df[\"Confirmed\"] - Covid_df[\"Cured\"] - Covid_df[\"Deaths\"] - Covid_df[\"Active_cases\"]))

# Remove rows where state = India
vaccine_df[vaccine_df[\"State\"] == \"India\"] = vaccine_df[vaccine_df[\"State\"] != \"India\"]
```



[Covid Data Analysis Project.html](#)

Member 2 – Data Analysis (SQL)



Tools Used: SQL

What She Did:

- Analyzed cleaned data using SQL
- Found answers to questions like:
 - State Highest Confirmed Cases**
 - Top 5 States by Vaccine Coverage**
 - Most Used Vaccine Type**
 - Recovery vs Deaths**



covid_report.pdf

Query 1

```

22 # Maharashtra 621539441
23
24 # 2. lowest number of confirmed cases in
25 • SELECT state, SUM(confirmed) AS total_
26 FROM covid_data
27 GROUP BY state
28 ORDER BY total_cases
29 LIMIT 1;
30 # state      total_cases
31 # Daman & Diu 2
32
33 # 3. Day with the Highest Number of Va
34 • SELECT vaccine_date
35 FROM vaccine_data
36 GROUP BY vaccine_date
37 ORDER BY total DESC
38 LIMIT 1;
39 # vaccine_date
40 # 09/08/2021
41
42 # 4. Top 5 states
43 • SELECT state, SUM(total_doses_administered) AS total_vaccines
44 FROM vaccine_data
45 GROUP BY state
46 ORDER BY total_vaccines DESC
47 LIMIT 5;
48 # state      total_vaccines
49 # Maharashtra 3497245417
50 # Uttar Pradesh 3342846470
51 # Gujarat 2732064385
52 # Rajasthan 2692747175
53 # West Bengal 2382596869

```

Result Grid

state	total_cases
Daman & Diu	2
Maharashtra	621539441

Result Grid

vaccine_date	total
09/08/2021	509780506

Result Grid

state	total_vaccines
Maharashtra	3497245417
Uttar Pradesh	3342846470
Gujarat	2732064385
Rajasthan	2692747175
West Bengal	2382596869

Member 3 - Dashboard

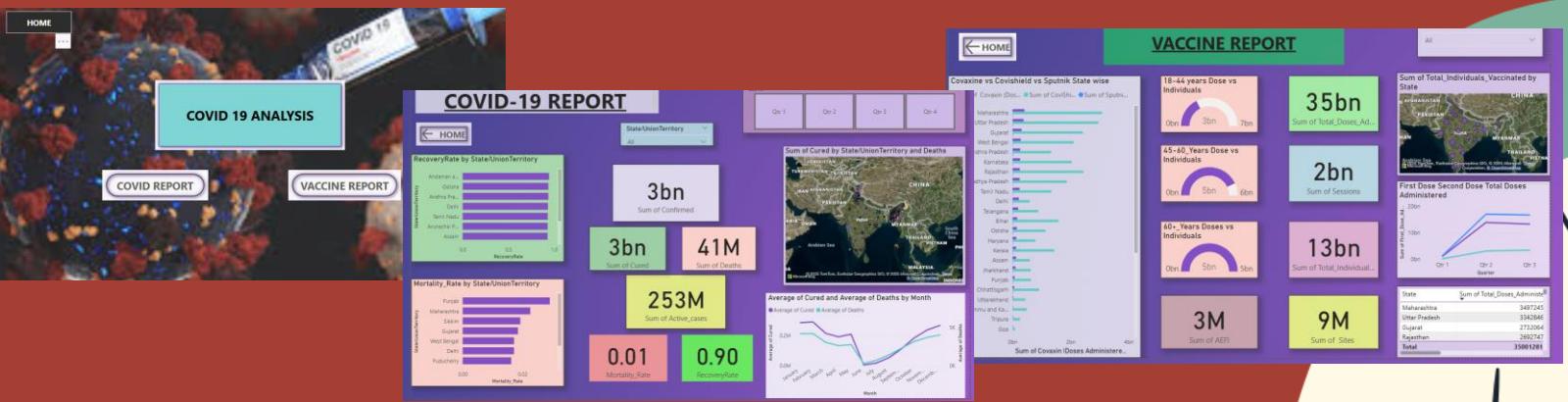
Tool Used: Power BI

What He Did:

Created dashboard to show insights

Charts used:

- Line chart – Avg cured and death by month
 - Map chart – Vaccinated by state
 - Bar chart – recovery rate by state



Conclusion

- We understood how COVID-19 changed over time
- Clean data helped to get good results
- SQL and dashboard gave useful insights
- Teamwork was smooth
- We learned data cleaning, analysis, and visualization

