COVID 19 SPATIAL DATA ANALYSIS

NARMADHA.S

NARMADHA DEVI.V

PAVITHRA.N

ABSTRACT

In this paper, we present a spatial data science system for analysing big covid 19 epidemiological data, with focus on the spatial data analytics among different geographical location. The system helps user to get a better understanding of information about the confirmed cases of covid 19, evaluation results shows the benefit of our system in spatial data analytics of big covid 19 data.

PROBLEM STATEMENT

DATA VISUALIZATION ANALYSIS BASED ON

- Specific geographic region
- Episode week
- Gender
- Age group
- Set of symptoms [dry cough ,high fever, sore throat, difficulty in breathing, infected with covid-19]
- Hospitalization status
- Age group distribution over region, gender, hospitalization status.
- Clinical outcome distribution [death and recovered]
- Absolute and relative cumulative covid 19 cases in a region of population

DATASET AND FEATURE SELECTION

Dataset Overview

- The dataset was collected and integrated from a provincial and territorial public health authorities by the Public Health Agency of Canada [PHAC].
- The dataset contains information such as Case Identifier Number (COV_ID), Region (COV_REG), Episode week (COV_EW), Episode year (COV_EY), Gender (COV_GDR), Age group (COV_AGR), Hospital status (COV_HSP), and Death (COV_DTH) for individual cases.

Feature Selection Rationale:

- **Region (COV_REG):** Essential for understanding geographical variations in COVID-19 cases and outcomes.
- **Episode week and Episode year (COV_EW, COV_EY):** Important for temporal analysis, tracking the progression of cases over time.
- **Gender (COV_GDR):** Vital for gender-based analysis to identify potential disparities.
- **Age group (COV_AGR):** Significant for understanding the impact of age on COVID-19 outcomes.
- **Hospital status and Death (COV_HSP, COV_DTH):** Crucial outcome variables for predictive modeling or survival analysis.

DATA PREPROCESSING AND CLEANING

reads three datasets, checks for missing values, removes duplicates, handles outliers, and transforms the data to prepare it for analysis, including features selection and cleaning steps.

DATA MODELLING

Spatial Analysis:

- If geospatial data is available, spatial analysis techniques may be used to study the geographic spread of COVID-19 cases. Spatial regression models or spatial autocorrelation analysis could be applied.

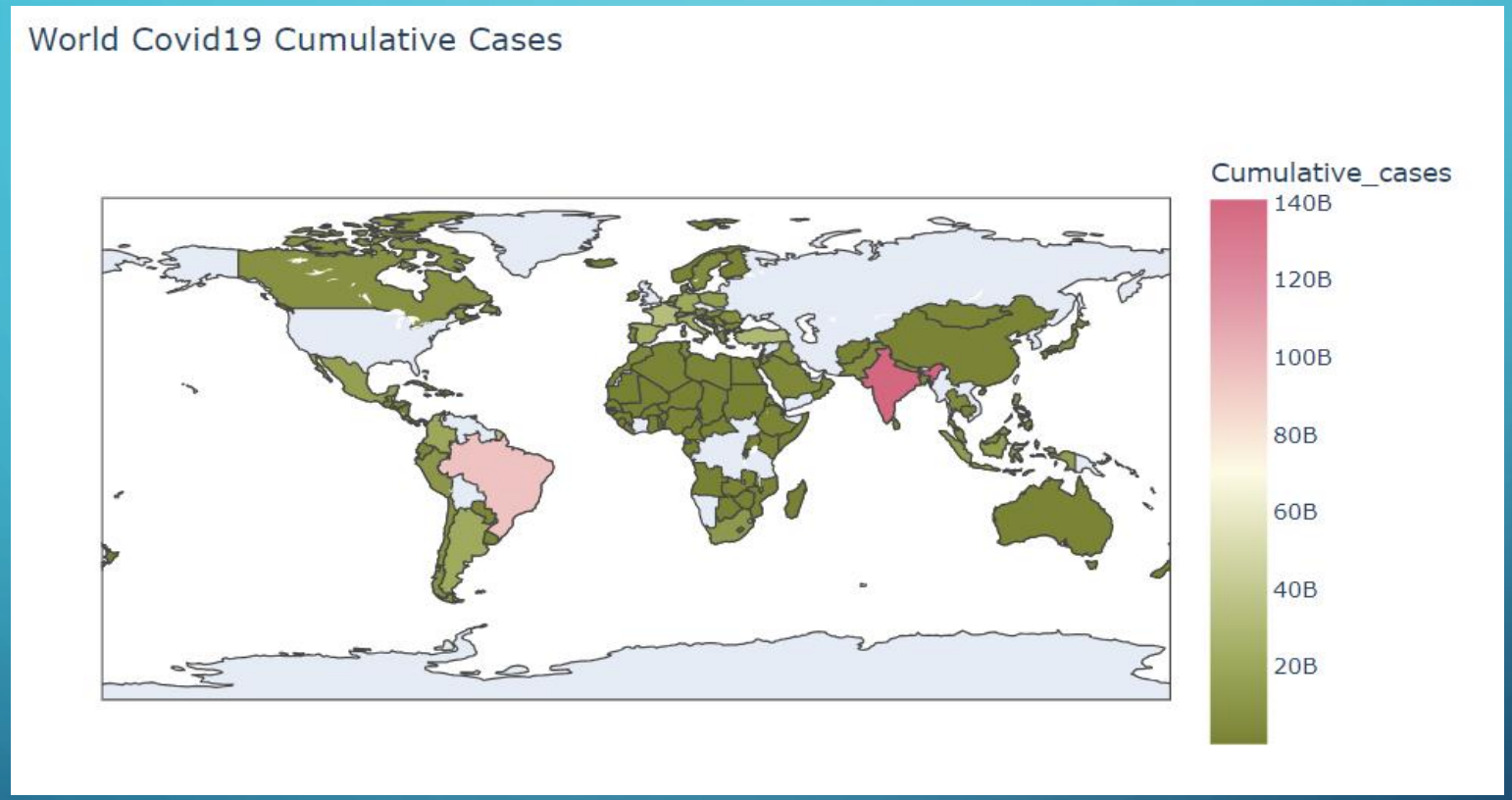
Descriptive Analytics:

- Descriptive statistics and visualizations are essential for summarizing and presenting key characteristics of the dataset. This might involve creating bar charts, line plots, heatmaps, or choropleth maps to convey information.

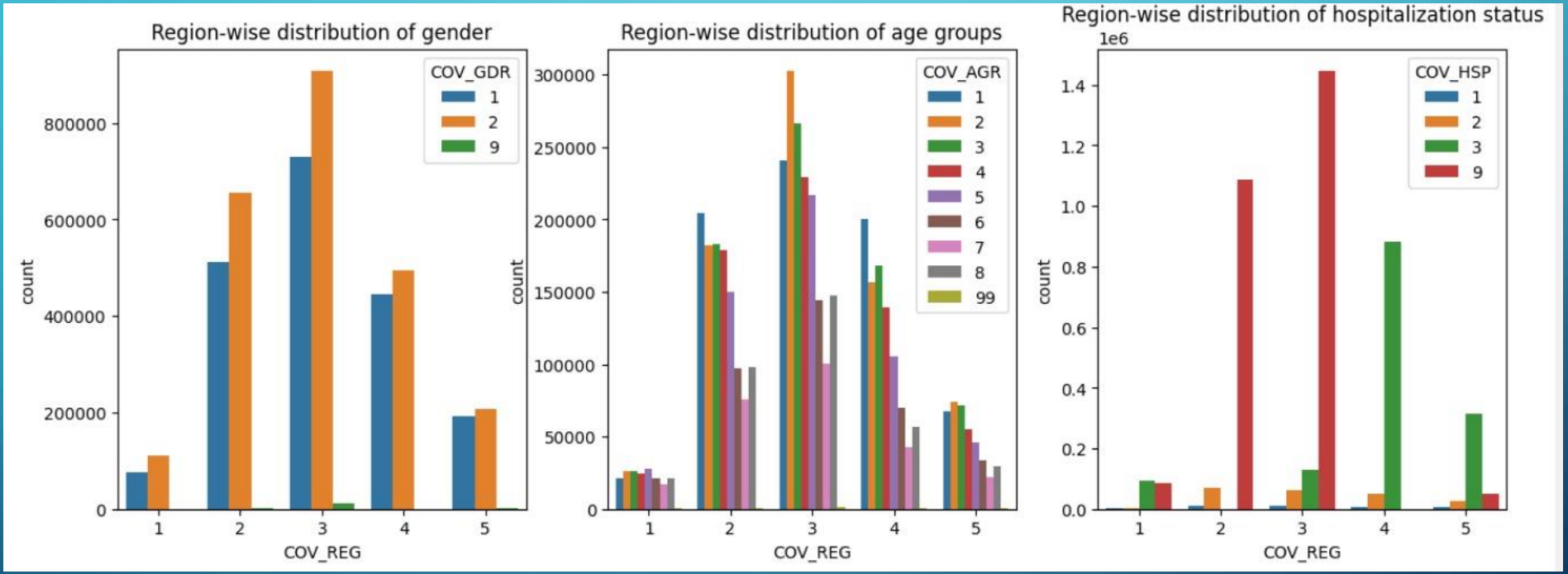
Predictive Modeling:

- This involves building models to predict outcomes such as hospitalization or mortality based on various features in the dataset. Common algorithms include logistic regression.

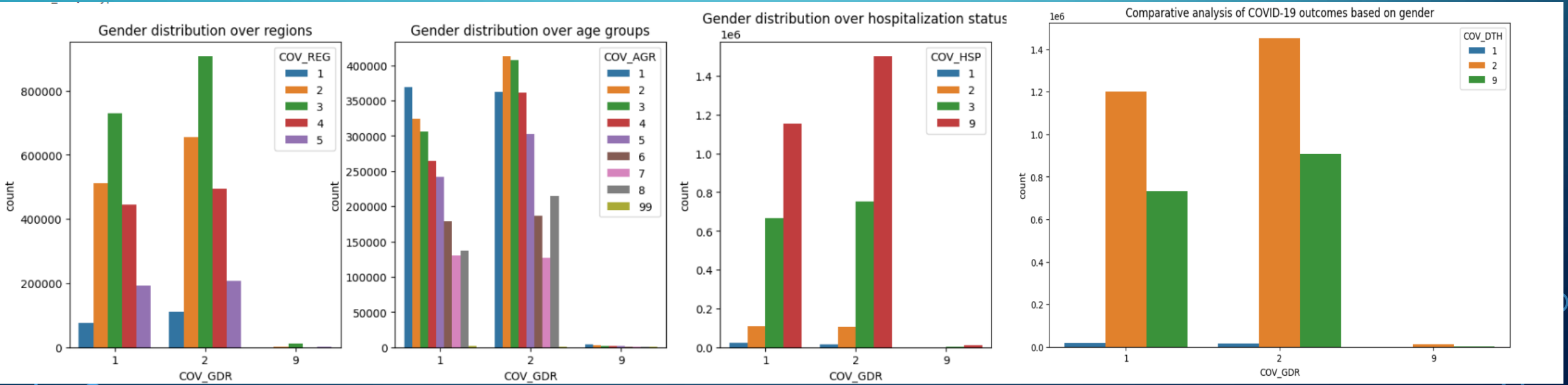
DATA VISUALIZATION AND RESULT



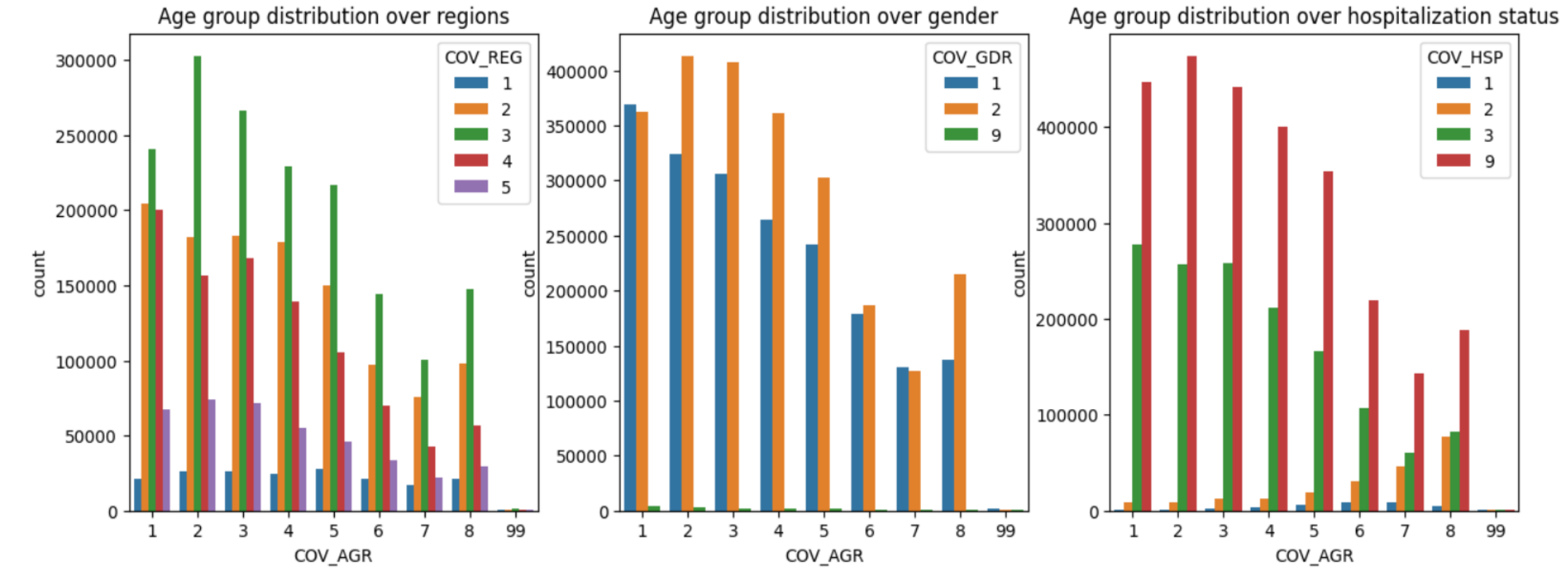
SPECIFIC GEOGRAPHIC REGION



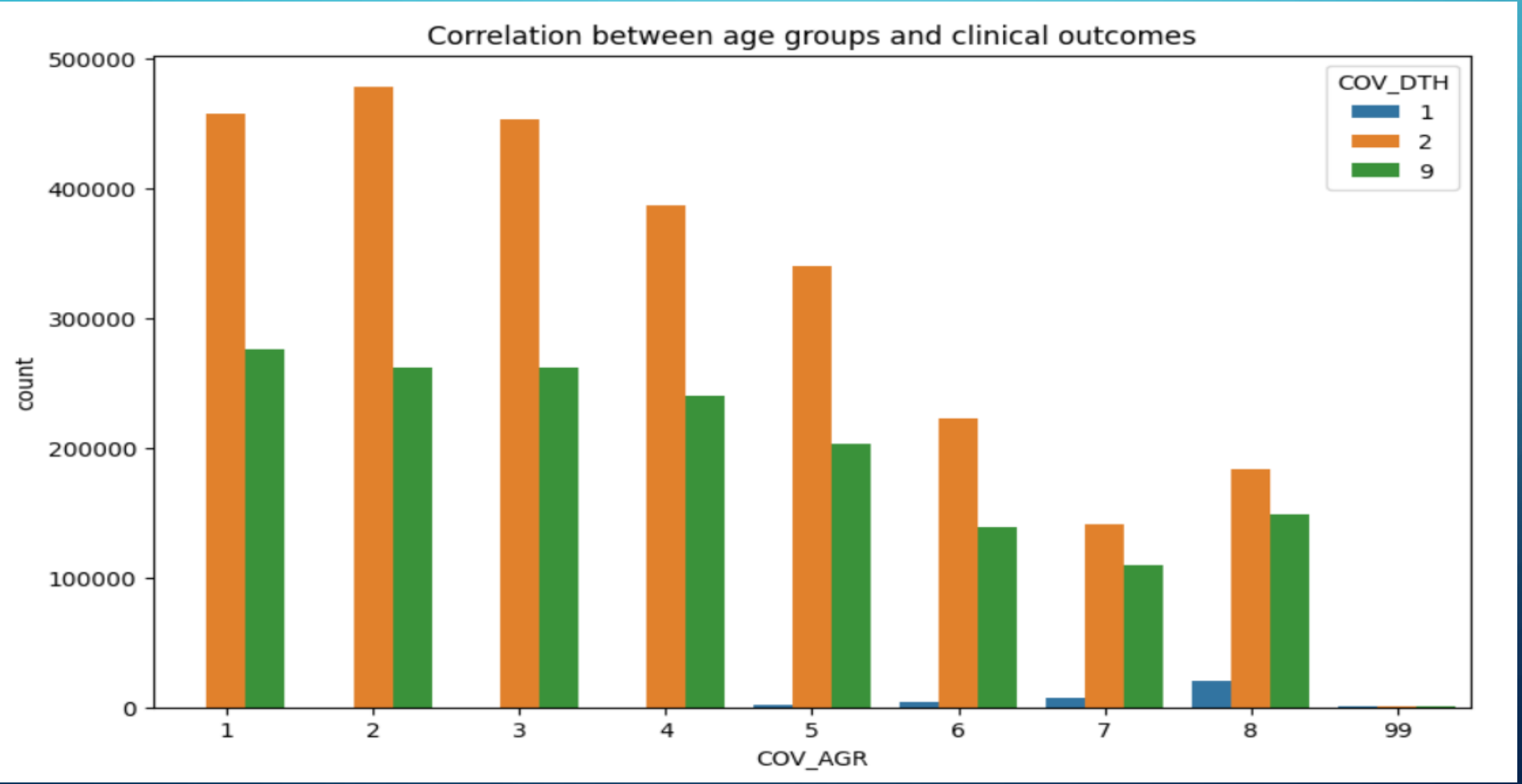
ANALYSIS BASED ON GENDER DISTRIBUTION



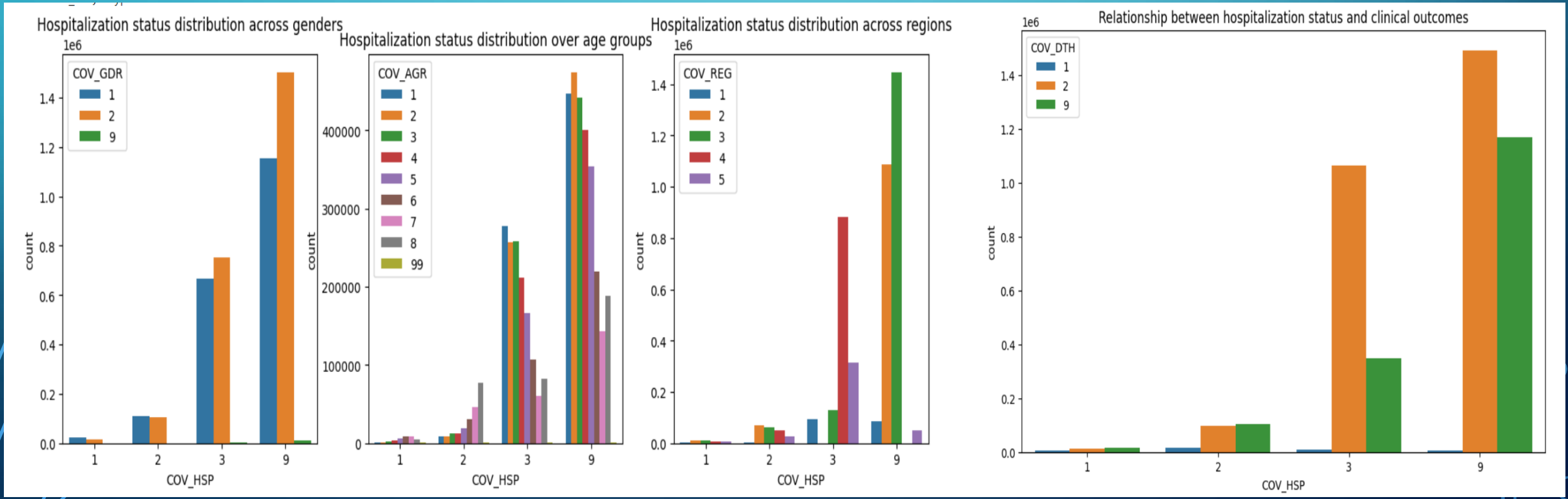
ANALYSIS BASED ON AGE GROUP



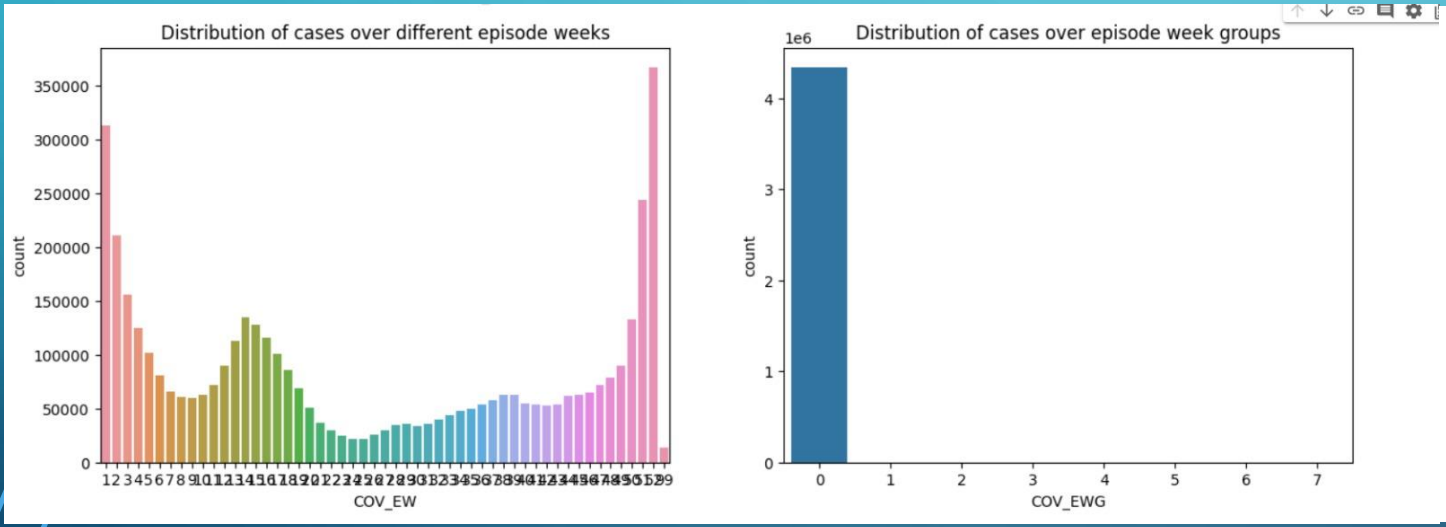
ANALYSIS BASED ON AGE GROUP

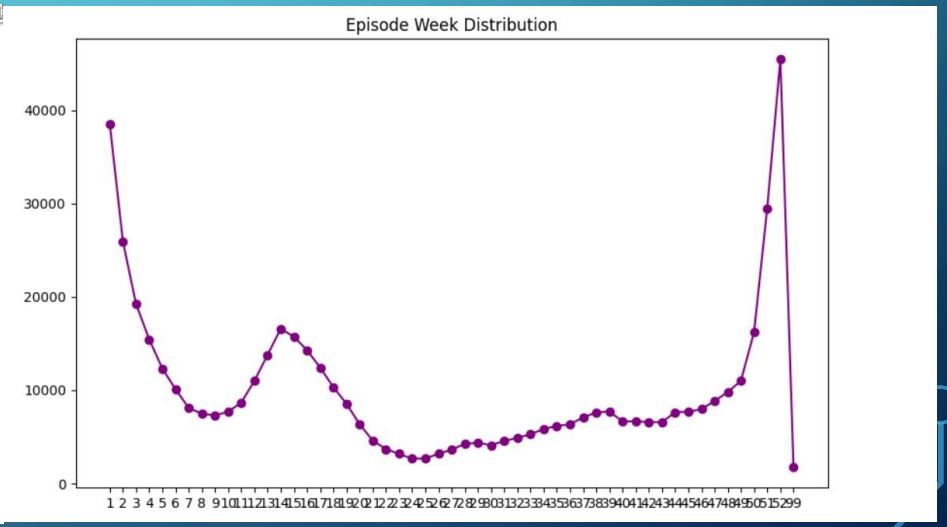


HOSPITALIZATION STATUS ANALYSIS

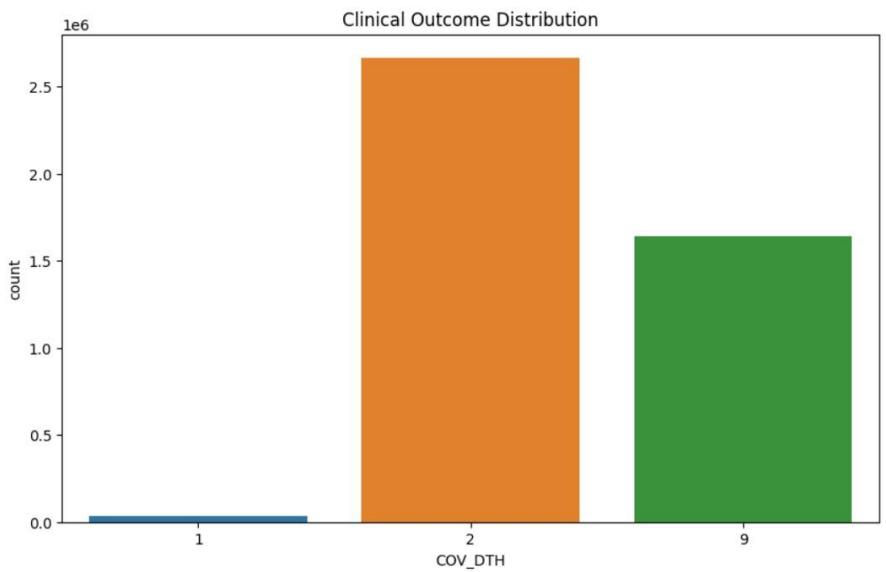


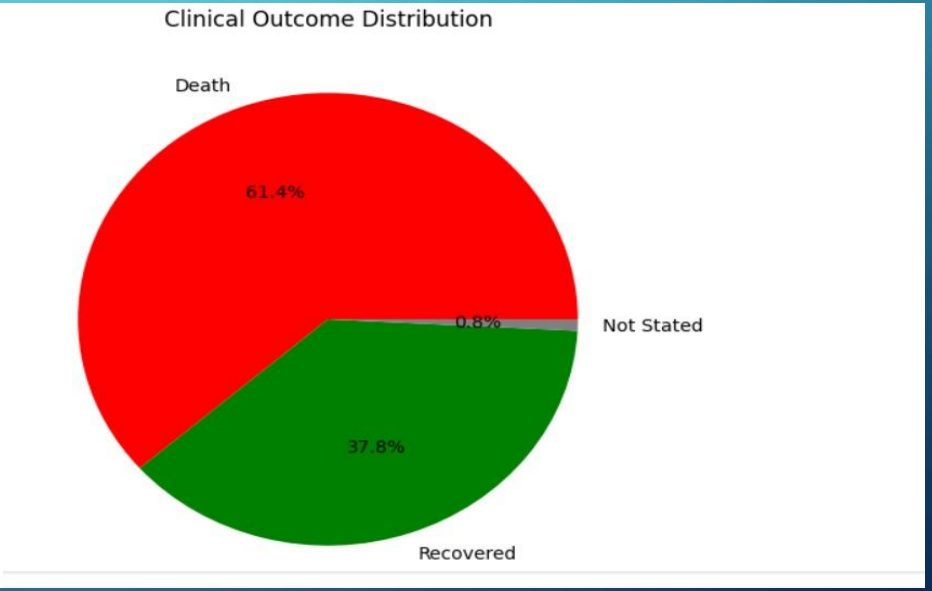
ANALYTICS BASED ON EPISODE WEEK



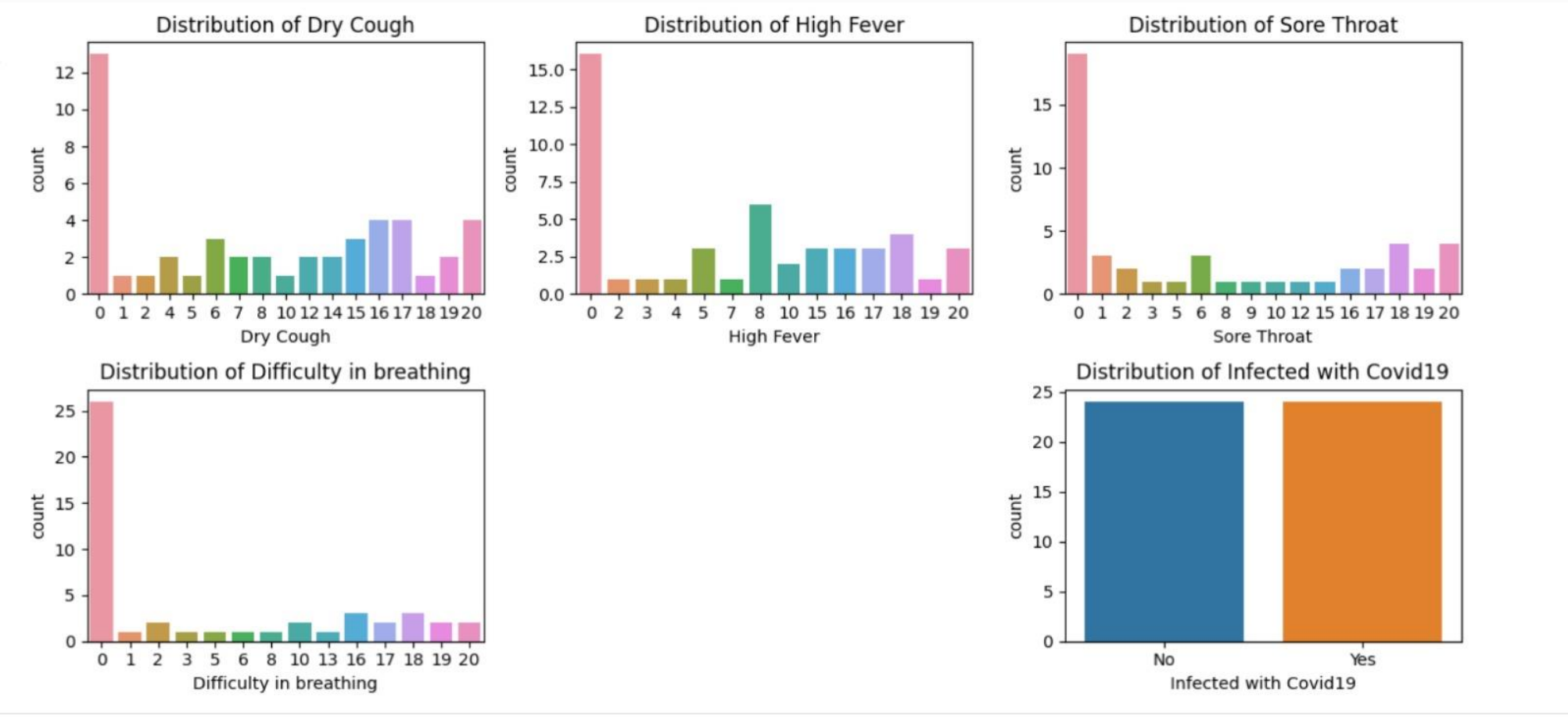


CLINICAL OUTCOME DISTRIBUTION [DEATH AND RECOVERED]

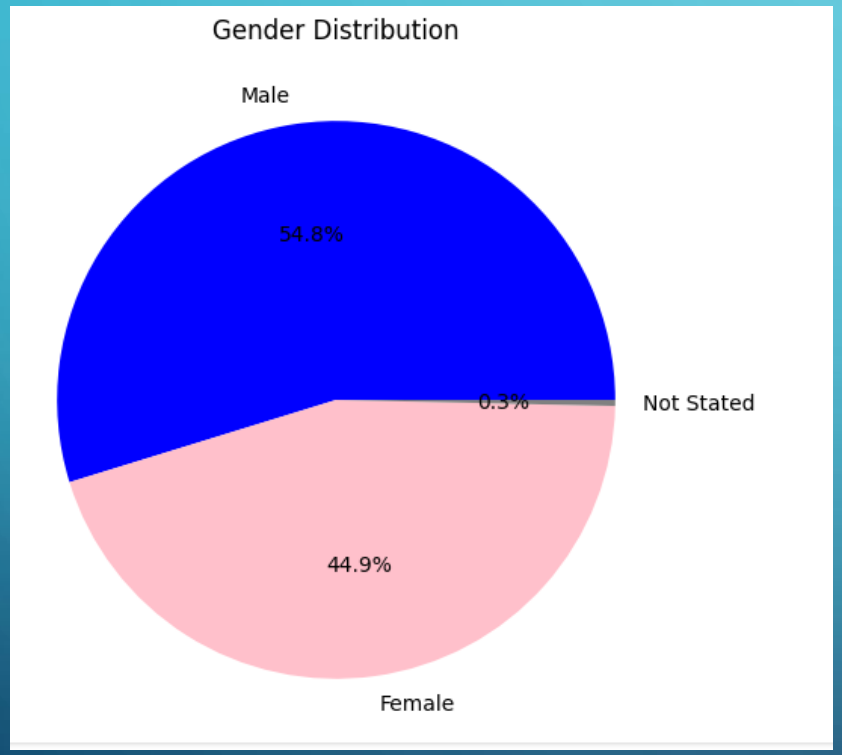


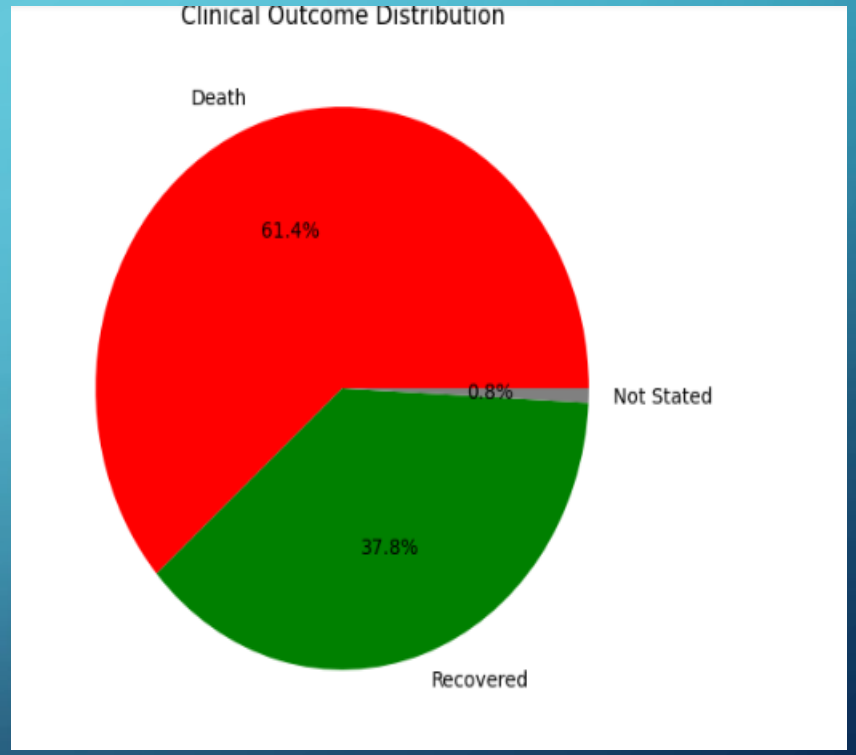


SET OF SYMPTOMS [DRY COUGH ,HIGH FEVER, SORE THROAT, DIFFICULTY IN BREATHING, INFECTED WITH COVID-19]

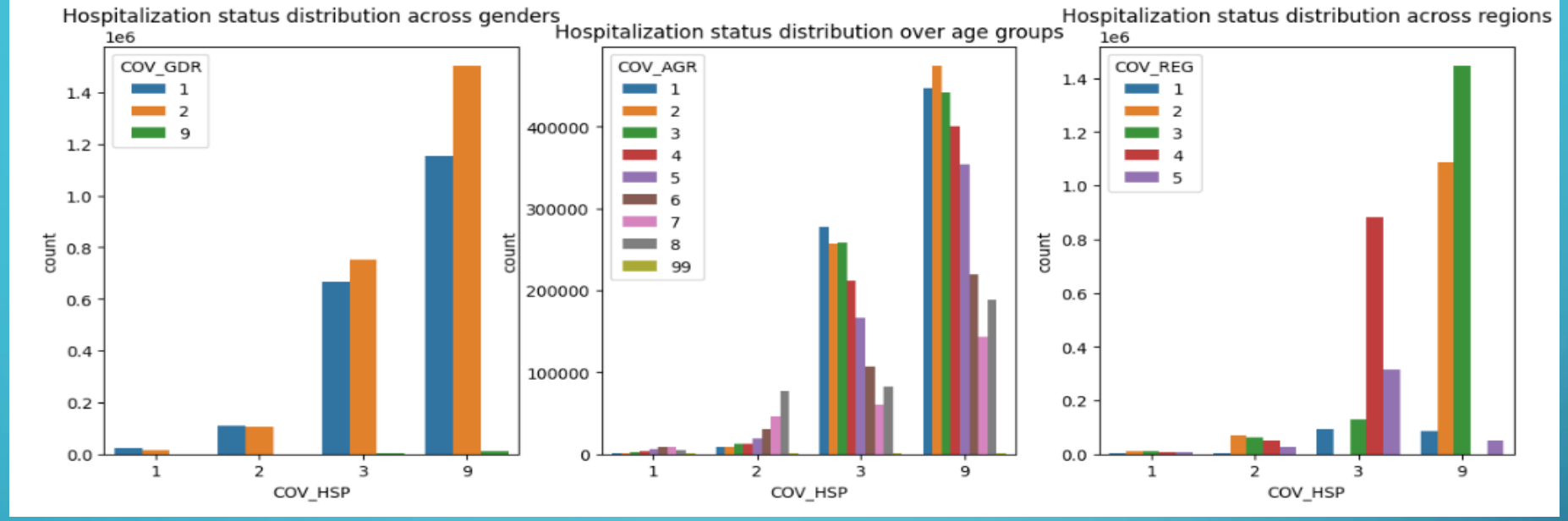


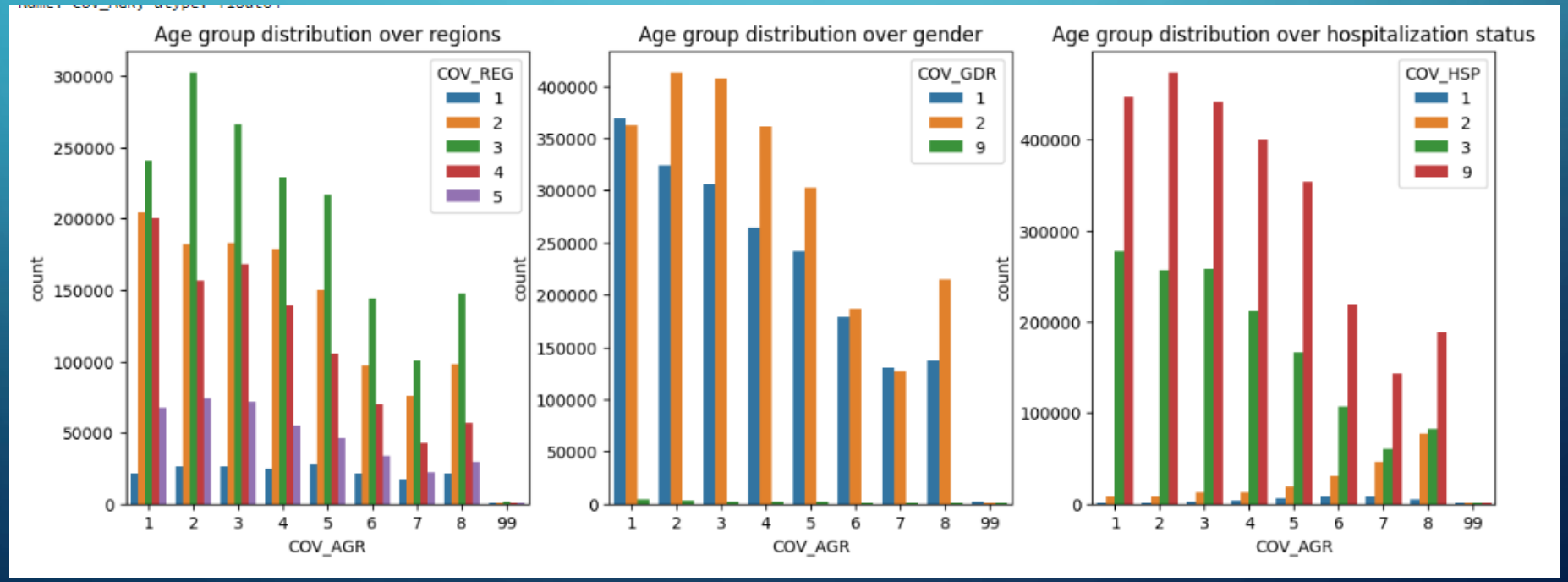
ANALYSIS BASED ON GENDER DISTRIBUTION AND CLINICAL OUTCOME DISTRIBUTION





Name: COV_HSP, dtype: float64





PROJECT EXECUTION

https://colab.research.google.com/drive/1xSIY5uj84YRkaMQJKupKHZo13WHDt7Jc#scrollTo=_UtUOiRc2Y9P&line=1&uniqifier=1

https://colab.research.google.com/drive/1yGSGrROdPzLx07FH1GW8b2Zk9-Pn9Rck#scrollTo=fzZZJlzLhB2O&line=1&uniqifier=1

ABSOLUTE FREQUENCY

Absolute Frequency:

	State/UTs	Total Cases	Active	Discharged	\
0	Andaman and Nicobar	10766	0	10637	
1	Andhra Pradesh	2340676	0	2325943	
2	Arunachal Pradesh	67049	0	66753	
3	Assam	746159	5	738119	
4	Bihar	855267	1	842952	
5	Chandigarh	100693	0	99508	
6	Chhattisgarh	1187695	0	1173505	
7	Dadra and Nagar Haveli and Daman and Diu	11592	0	11588	
8	Delhi	2040910	14	2014230	
9	Goa	263346	3	259329	
10	Gujarat	1291383	5	1280299	
11	Haryana	1078903	27	1068121	
12	Himachal Pradesh	322905	4	318660	
13	Jammu and Kashmir	482023	0	477231	
14	Jharkhand	443826	1	438491	
15	Karnataka	4088769	12	4048399	
16	Kerala	6907241	18	6835181	
17	Ladakh	29602	0	29371	
18	Lakshadweep	11415	0	11363	
19	Madhya Pradesh	1056351	0	1045565	
20	Maharashtra	8171048	214	8022276	
21	Manipur	140034	0	137885	
22	Meghalaya	96983	3	95352	
23	Mizoram	239560	1	238825	
24	Nagaland	36033	0	35251	
25	Odisha	1348409	59	1339135	
26	Puducherry	177547	0	175566	
27	Punjab	793644	1233	773073	
28	Rajasthan	1326465	2	1316727	
29	Sikkim	44927	4	44422	
30	Tamil Nadu	3610655	5	3572569	
31	Telengana	844432	8	840313	
32	Tripura	108493	1	107550	
33	Uttar Pradesh	2145431	57	2121662	
34	Uttarakhand	452571	0	444803	
35	West Bengal	2126282	135	2104592	

	Deaths	Active Ratio	Discharge Ratio	Death Ratio	Population
0	129	0.00	98.80	1.20	100896618
1	14733	0.00	99.37	0.63	128500364
2	296	0.00	99.56	0.44	658019
3	8035	0.00	98.92	1.08	290492
4	12314	0.00	98.56	1.44	40100376
5	1185	0.00	98.82	1.18	30501026
6	14190	0.00	98.81	1.19	28900667
7	4	0.00	99.97	0.03	231502578
8	26666	0.00	98.69	1.31	773997
9	4014	0.00	98.47	1.52	3772103
10	11079	0.00	99.14	0.86	70400153
11	10755	0.00	99.00	1.00	7503010
12	4241	0.00	98.69	1.31	3436948
13	4792	0.00	99.01	0.99	66001
14	5334	0.00	98.80	1.20	124904071
15	40358	0.00	99.01	0.99	1711947
16	72042	0.00	98.96	1.04	91702478
17	231	0.00	99.22	0.78	4184959
18	52	0.00	99.54	0.46	11700099
19	10786	0.00	98.98	1.02	14999397
20	148558	0.00	98.18	1.82	399001
21	2149	0.00	98.47	1.53	47099270
22	1628	0.00	98.32	1.68	79502477
23	734	0.00	99.69	0.31	1308967
24	782	0.00	97.83	2.17	38157311
25	9215	0.00	99.31	0.68	19301096
26	1981	0.00	98.88	1.12	2073074
27	19338	0.16	97.41	2.44	34698876
28	9736	0.00	99.27	0.73	1521992
29	501	0.01	98.88	1.12	83697770
30	38081	0.00	98.95	1.05	35998752
31	4111	0.00	99.51	0.49	69599762
32	942	0.00	99.13	0.87	1646050
33	23712	0.00	98.89	1.11	1158040
34	7768	0.00	98.28	1.72	85002417
35	21555	0.01	98.98	1.01	32199722

RELATIVE FREQUENCY

Relative Frequency (Percentage):

	State/UTs	Active Ratio	Discharge Ratio	\
0	Andaman and Nicobar	0.0	9880.0	
1	Andhra Pradesh	0.0	9937.0	
2	Arunachal Pradesh	0.0	9956.0	
3	Assam	0.0	9892.0	
4	Bihar	0.0	9856.0	
5	Chandigarh	0.0	9882.0	
6	Chhattisgarh	0.0	9881.0	
7	Dadra and Nagar Haveli and Daman and Diu	0.0	9997.0	
8	Delhi	0.0	9869.0	
9	Goa	0.0	9847.0	
10	Gujarat	0.0	9914.0	
11	Haryana	0.0	9900.0	
12	Himachal Pradesh	0.0	9869.0	
13	Jammu and Kashmir	0.0	9901.0	
14	Jharkhand	0.0	9880.0	
15	Karnataka	0.0	9901.0	
16	Kerala	0.0	9896.0	
17	Ladakh	0.0	9922.0	
18	Lakshadweep	0.0	9954.0	
19	Madhya Pradesh	0.0	9898.0	
20	Maharashtra	0.0	9818.0	
21	Manipur	0.0	9847.0	
22	Meghalaya	0.0	9832.0	
23	Mizoram	0.0	9969.0	
24	Nagaland	0.0	9783.0	
25	Odisha	0.0	9931.0	
26	Puducherry	0.0	9888.0	
27	Punjab	16.0	9741.0	
28	Rajasthan	0.0	9927.0	
29	Sikkim	1.0	9888.0	
30	Tamil Nadu	0.0	9895.0	
31	Telengana	0.0	9951.0	
32	Tripura	0.0	9913.0	
33	Uttar Pradesh	0.0	9889.0	
34	Uttarakhand	0.0	9828.0	
35	West Bengal	1.0	9898.0	

	Death Ratio	Population
0	120.0	100896618
1	63.0	128500364
2	44.0	658019
3	108.0	290492
4	144.0	40100376
5	118.0	30501026
6	119.0	28900667
7	3.0	231502578
8	131.0	773997
9	152.0	3772103
10	86.0	70400153
11	100.0	7503010
12	131.0	3436948
13	99.0	66001
14	120.0	124904071
15	99.0	1711947
16	104.0	91702478
17	78.0	4184959
18	46.0	11700099
19	102.0	14999397
20	182.0	399001
21	153.0	47099270
22	168.0	79502477
23	31.0	1308967
24	217.0	38157311
25	68.0	19301096
26	112.0	2073074
27	244.0	34698876
28	73.0	1521992
29	112.0	83697770
30	105.0	35998752
31	49.0	69599762
32	87.0	1646050
33	111.0	1158040
34	172.0	85002417
35	101.0	32199722

CONCLUSION

In this project, we conducted a comprehensive spatial data analysis of COVID-19 cases, leveraging information from the provided dataset. The dataset included essential variables such as case identifier numbers, geographical regions, episode weeks, gender, age groups, hospitalization status, and clinical outcomes.

Key Findings:

1.Geographical Distribution:

1. Our spatial analysis revealed the geographic distribution of COVID-19 cases across different regions, allowing us to identify areas with higher prevalence.

2.Temporal Trends:

1. Analysis of episode weeks highlighted temporal trends, aiding in understanding the progression of COVID-19 over time.

3.Demographic Patterns:

1. Examination based on gender and age groups provided insights into how different demographics were affected, contributing to a better understanding of vulnerability.

4.Hospitalization and Clinical Outcomes:

1. Exploration of hospitalization status and clinical outcomes (recovered and death) facilitated an assessment of the severity and impact of cases.

5.Symptomatic Analysis:

1. Spatial visualization of COVID-19 cases with specific symptoms like dry cough, high fever, sore throat, and difficulty in breathing provided additional layers of information.

Implications:

•The analysis can inform public health strategies, resource allocation, and targeted interventions based on geographic and demographic patterns.
•Insights into temporal trends are crucial for predicting healthcare demands and planning for future outbreaks.
•Understanding the spatial distribution of symptoms can aid in identifying potential hotspots and areas requiring focused public health campaigns.

Limitations and Further Research:

Future research could delve into more granular spatial analyses, considering factors like population density, socio-economic status, and healthcare infrastructure.

THANK YOU