A Skill Based Mini Project

on

"Data Analytics of Covid-19"

For the course

Data Science in IoT
(220505)

By

Pranjal Dubey
(0901CE211094)

Murtaza Lohawala
(0901EO211036)

CENTER FOR INTERNET OF THINGS
MADHAV INSTITUTE OF TECHNOLOGY & SCIENCE,
GWALIOR (M.P.) – 474005

Jul. - Dec. 2023

Madhav Institute of Technology and Science , Gwalior (M.P.)

A Govt. Aided UGC Autonomous Institute, Affiliated to R.G.P.V. BHOPAL (M.P.), INDIA

NAAC ACCREDITED WITH A++ GRADE

CANDIDATE'S DECLARATION

We hereby declare that the work presented in this skill based mini-project entitled **Data Analytics of Covid-19** which is being submitted in Data Science in IoT (220505) course for the partial fulfillment of the requirement for the award of degree of Bachelor of Technology in Internet of Things is an authentic record of our own work carried out under the guidance of Dr. Gaurav Khare, Assistant Professor, Centre for Internet of Things.

Pranjal Dubey

Murtaza Lohawala

Date:
Place: Gwalior

This is to certify that the above statement made by the candidates is correct to the best of my knowledge and belief.

Guided by

Dr. Gaurav Khare
Assistant Professor
Centre For Internet of Things
MITS, Gwalior

TABLE OF CONTENTS

1. ABSTRACT
2. INTRODUCTION
3. TECHNICAL TERMS
4. SOURCE CODE
5. RESULT AND ANALYSIS
6. CONCLUSION
7. REFERENCES

ABSTRACT

The outbreak of the COVID-19 pandemic in late 2019 rapidly evolved into a global crisis, affecting public health, economies, and societies worldwide. As a result, an unprecedented amount of data related to the virus, its spread, and its impact has been generated. The project titled "Data Analytics of COVID-19" aims to harness the power of advanced data analytics techniques to uncover valuable insights from this vast and heterogeneous dataset, ultimately contributing to a deeper understanding of the pandemic's dynamics and informing effective decision-making. This project leverages a multidisciplinary approach, merging epidemiological data, clinical records, demographic information, mobility data, and public policy interventions. The primary objective is to conduct a comprehensive analysis of these datasets to address critical aspects of the pandemic.

The project begins by collecting and pre-processing diverse datasets from reputable sources, ensuring data quality and consistency. Subsequently, exploratory data analysis techniques are applied to identify trends, patterns, and anomalies in the spread and impact of the virus. Advanced statistical methods, machine learning algorithms, and time-series analysis are then employed to model the virus's transmission dynamics, predict future trends, and evaluate the effectiveness of various interventions. Furthermore, the project emphasizes the significance of geographic and demographic factors in shaping the pandemic's trajectory. Geospatial analysis techniques are used to visualize the spatial distribution of cases, identifying hotspots and areas with different infection rates. Socioeconomic and demographic variables are integrated to assess disparities in vulnerability, access to healthcare, and socio-economic impacts.

To facilitate the accessibility and usability of the findings, interactive dashboards and visualizations are developed. These tools provide policymakers, researchers, and the general public with an intuitive platform to explore the data-driven insights, fostering evidence-based decision-making and public awareness. In conclusion, "Data Analytics of COVID-19" presents a comprehensive and data-driven approach to understand the complexities of the pandemic. By harnessing the power of advanced analytics, this project contributes to the collective efforts aimed at managing and mitigating the impacts of COVID-19. The insights gained from this project have the potential to guide public health strategies, resource allocation, and future preparedness for similar global health crises.

1. INTRODUCTION

In recent times, the world has witnessed an unprecedented global challenge that has reshaped societies, economies, and healthcare systems: the COVID-19 pandemic. This infectious disease, caused by the novel corona virus SARS-CoV-2, has rapidly spread across the globe, prompting an urgent need for comprehensive and data-driven strategies to understand its dynamics, mitigate its impact, and guide decision-making. The field of data analytics has emerged as an indispensable tool in the fight against COVID-19, enabling researchers, policymakers, and healthcare professionals to glean valuable insights from a deluge of information.

The project "Data Analytics of COVID-19" stands at the nexus of science, technology, and public health, seeking to harness the power of data to illuminate the intricate facets of this pandemic. By leveraging advanced analytical techniques and cutting-edge technologies, this project aims to uncover patterns, trends, and correlations within the vast volumes of data generated by the spread of the virus. From epidemiological data capturing infection rates, mortality rates, and geographic spread, to clinical data encompassing symptoms, treatments, and outcomes, and even socio-economic data revealing disparities in vulnerability and healthcare access, the project takes a holistic approach to understanding the multifaceted dimensions of the pandemic.

Objectives:

1. **Insight Generation:** Through the utilization of various data mining and machine learning techniques, the project seeks to extract hidden insights from diverse datasets related to COVID-19. By uncovering hidden patterns, the project aims to contribute to a deeper comprehension of the virus's behavior, transmission mechanisms, and impacts on different population groups.

2. **Predictive Modeling:** One of the critical aspects of managing a pandemic is predicting its future course. The project endeavors to develop predictive models that can anticipate the spread of the virus, aiding public health officials and policymakers in making informed decisions regarding resource allocation and intervention strategies.

3. **Resource Optimization:** The strain on healthcare systems during the pandemic necessitates efficient resource allocation. Data analytics can play a pivotal role in optimizing the distribution of medical supplies, hospital beds, and personnel by analyzing real-time healthcare utilization data.

4. **Risk Assessment:** By integrating data from various sources, including travel patterns, population density, and health infrastructure, the project seeks to assess the vulnerability of different regions and population segments. This information can guide targeted measures to mitigate the impact of the virus.

5. **Communication and Visual Representation:** Transforming complex data into easily understandable visualizations is crucial for effective communication. The project aims to create compelling data visualizations that convey the severity of the pandemic's impact, the effectiveness of interventions, and the progress of vaccination campaigns.

Significance:

The "Data Analytics of COVID-19" project holds significant implications for the ongoing battle against the pandemic. Its findings and insights can assist healthcare professionals in providing optimal care, policymakers in making informed decisions, researchers in advancing our understanding of the virus, and the general public in staying informed about the evolving situation. Moreover, the methodologies and approaches developed within this project could pave the way for more effective responses to future pandemics and health crises.

In the following sections of this project, we will delve into the methodologies employed, the data sources utilized, the analytical techniques applied, and the insights gained. By synergizing the realms of data science and public health, this project aspires to contribute meaningfully to the global effort to combat COVID-19.

2. TECHNICAL TERMS

2.1) Algorithms : A set of instructions that a computer follows to perform a task. In data analytics, algorithms are used to process and analyze data.

2.2) Big data : A term used to describe the vast amount of data that is collected and stored by organizations today. Big data can be analyzed to gain insights into customer behavior, identify trends, and make better decisions.

2.3) Data mining : The process of extracting knowledge from data. Data mining techniques are used to identify patterns and trends in data that would not be visible to the naked eye.

2.4) Data visualization : The process of representing data in a graphical format. Data visualization can be used to make data more understandable and to communicate insights to others.

2.5) Dimensionality reduction : The process of reducing the number of dimensions in a data set. Dimensionality reduction can be used to make data easier to manage and analyze.

2.6) Machine learning : A type of artificial intelligence that allows computers to learn without being explicitly programmed. Machine learning algorithms are used to analyze data and make predictions.

2.7) Natural language processing : The ability of computers to understand and process human language. Natural language processing is used in a variety of applications, such as machine translation and sentiment analysis.

2.8) Predictive analytics : The use of data to predict future events. Predictive analytics is used in a variety of industries, such as finance, healthcare, and marketing.

2.9) Statistics : The science of collecting, analyzing, and interpreting data. Statistics is used in data analytics to make inferences about populations based on samples.

2.10) Modules : A Python module is a file containing Python definitions and statements. A module can define functions, classes, and variables. A module can also include runable code. Grouping related code into a module makes the code easier to understand and use.

2.11) Function : A function is a block of code which only runs when it is called. You can pass data, known as parameters, into a function. A function can return data as a result.

2.12) Libraries : A collection of modules that are linked together is also known as a Python library. It has code bundles that can be used again and again in different programs.

3 SOURCE CODE

```
import pandas as pd
data = pd.read_csv("/content/StatewiseTestingDetails.csv")
data.iloc[0:,0:2]  #extract values from dataset
data.head()  #top rows including column extraction
data.iloc[1:9,0:5]

#checking the outlier in the testing covid data
missing_values = data.isnull().sum()
print(missing_values)

#normalizing the data set
col = ["Positive","Negative"]
norm_data_set = ((data[col]))

#visualize the positive data set of testing
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import seaborn as sb

#plt.figure(figsize=(8,6))
positive_count = data['Positive']
st = data['State']
ps = data['Positive']
plt.plot(positive_count, st, label='State')
plt.plot(positive_count, ps, label='Positive')
plt.legend()
plt.title('Positive case Visualization')
plt.ylabel('Number of Positive Cases')
plt.xlabel('Name of states')
plt.savefig('plot.png')
plt.show()

import matplotlib.pyplot as plt
data1 = pd.read_csv('/content/corona affected countries.csv')
daycount = data1['Day']
my = data1['Malaysia']
sg = data1['Singapore']
id = data1['Indonesia']
th = data1['Thailand']
plt.plot(daycount, my, label='Malaysia')
plt.plot(daycount, sg, label='Singapore')
plt.plot(daycount, id, label='Indonesia')
plt.plot(daycount, th, label='Thailand')
plt.legend()
plt.title('Corona affected countries')
plt.ylabel('Number of infection')
plt.xlabel('Number of Days')
plt.savefig('plot.png')
plt.show()
```

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
import seaborn as sb
pd.read_csv('/content/corona affected countries.csv')

data2 = pd.read_csv('/content/corona affected countries.csv')
sb.countplot(x="Malaysia",data = data2)
plt.show()
```

```
data2 = pd.read_csv('/content/StatewiseTestingDetails.csv')
sb.scatterplot(x="Positive",y="State",data=data2)
plt.show()
```

```
sb.scatterplot(x="Negative",y="State",data=data2)
plt.show()
```

OUTPUT

+ Code + Text Copy to Drive

```
[1] import pandas as pd
    data = pd.read_csv("/content/StatewiseTestingDetails.csv")
    data.iloc[0:,0:2]  #extract values from dataset
```

	Date	State
0	2020-04-17	Andaman and Nicobar Islands
1	2020-04-24	Andaman and Nicobar Islands
2	2020-04-27	Andaman and Nicobar Islands
3	2020-05-01	Andaman and Nicobar Islands
4	2020-05-16	Andaman and Nicobar Islands
...	...	...
16331	2021-08-06	West Bengal
16332	2021-08-07	West Bengal
16333	2021-08-08	West Bengal
16334	2021-08-09	West Bengal
16335	2021-08-10	West Bengal

16336 rows × 2 columns

+ Code + Text Copy to Drive

```
data.head()  #top rows including column extraction
```

	Date	State	TotalSamples	Negative	Positive
0	2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1	2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2	2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3	2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4	2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0

```
[ ] data.iloc[1:9,0:5]
```

	Date	State	TotalSamples	Negative	Positive
1	2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2	2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3	2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4	2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0
5	2020-05-19	Andaman and Nicobar Islands	6965.0	NaN	33.0
6	2020-05-20	Andaman and Nicobar Islands	7082.0	NaN	33.0
7	2020-05-21	Andaman and Nicobar Islands	7167.0	NaN	33.0
8	2020-05-22	Andaman and Nicobar Islands	7263.0	NaN	33.0

+ Code + Text Copy to Drive

```
[ ] #checking the outlier in the testing covid data
    missing_values = data.isnull().sum()
    print(missing_values)
```

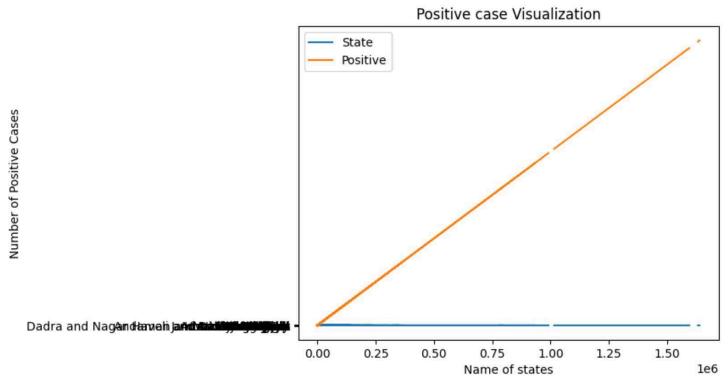
```
Date                0
State               0
TotalSamples        0
Negative         9367
Positive        10674
dtype: int64
```

```
[ ] #normalizing the data set
    col = ["Positive","Negative"]
    norm_data_set = ((data[col]))
```

```
[ ] #visualize the positive data set of testing
    import numpy as np
    from scipy import stats
    import matplotlib.pyplot as plt
    import seaborn as sb

    #plt.figure(figsize=(8,6))
    positive_count = data['Positive']

    st = data['State']
    ps = data['Positive']

    plt.plot(positive_count, st, label='State')
    plt.plot(positive_count, ps, label='Positive')

    plt.legend()
    plt.title('Positive case Visualization')
    plt.ylabel('Number of Positive Cases')
    plt.xlabel('Name of states')
```

+ Code + Text Copy to Drive

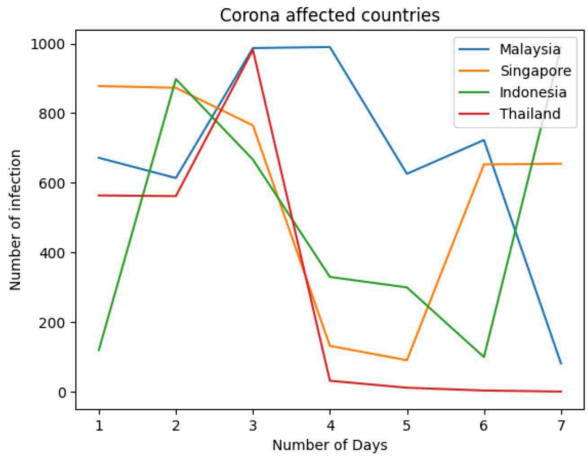
```
[ ] plt.savefig('plot.png')
    plt.show()
```




```
[ ] import matplotlib.pyplot as plt
    data1 = pd.read_csv('/content/corona affected countries.csv')
    daycount = data1['Day']
```

+ Code + Text Copy to Drive

```
[ ] plt.legend()
    plt.title('Corona affected countries')
    plt.ylabel('Number of infection')
    plt.xlabel('Number of Days')

    plt.savefig('plot.png')
    plt.show()
```

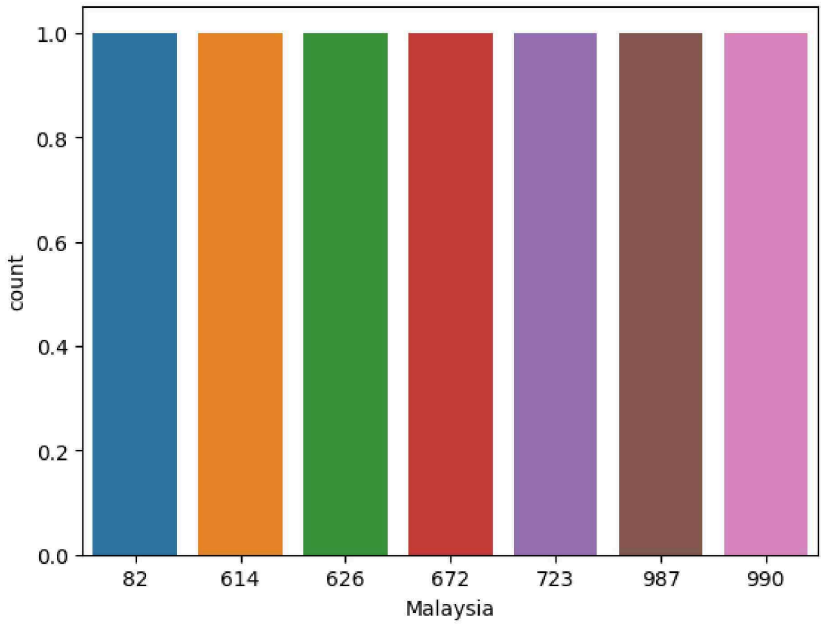



```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
pd.read_csv('/content/corona affected countries.csv')
```

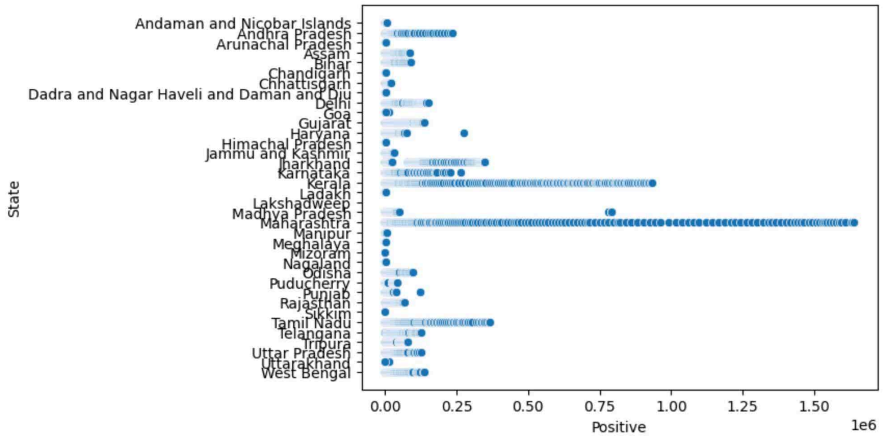
	Day	Malaysia	Singapore	Indonesia	Thailand
0	1	672	878	120	564
1	2	614	873	898	562
2	3	987	765	667	982
3	4	990	132	330	32
4	5	626	91	300	12
5	6	723	653	100	4
6	7	82	655	990	1

+ Code + Text Copy to Drive

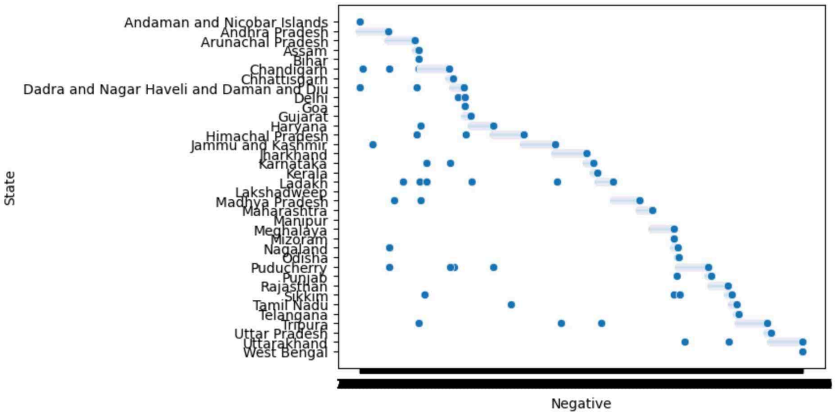
```
data2 = pd.read_csv('/content/corona affected countries.csv')
sb.countplot(x="Malaysia",data = data2)
plt.show()
```




```
data2 = pd.read_csv('/content/StatewiseTestingDetails.csv')
sb.scatterplot(x="Positive",y="State",data=data2)

plt.show()
```

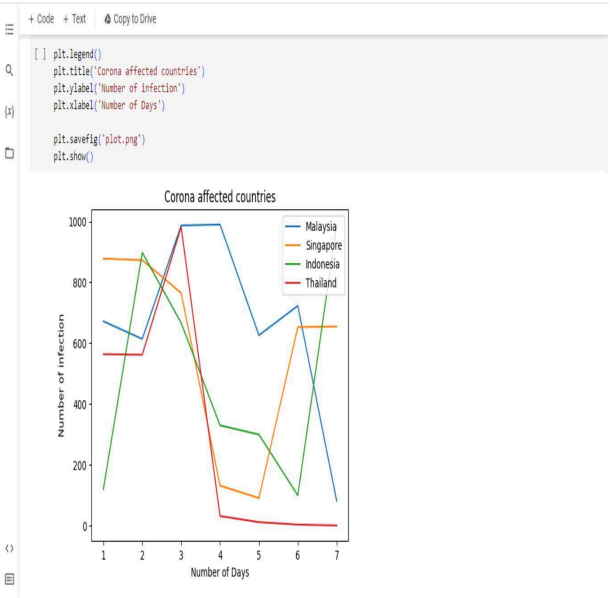


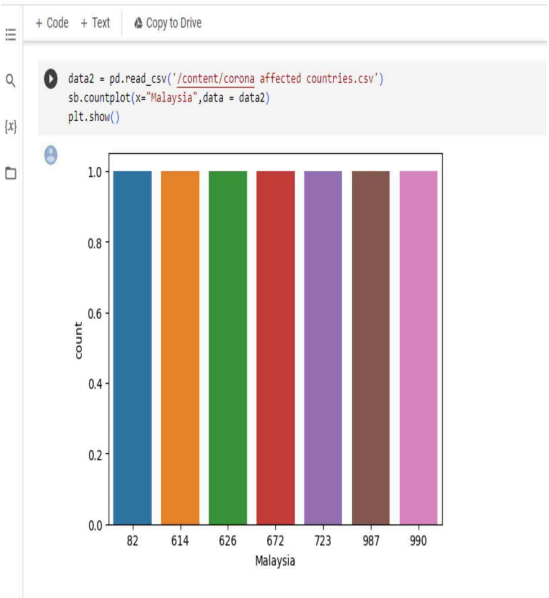
```
[ ] sb.scatterplot(x="Negative",y="State",data=data2)

    plt.show()
```

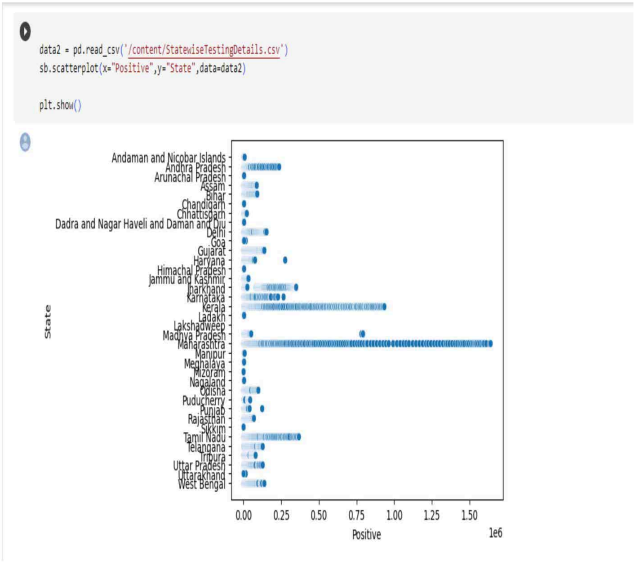


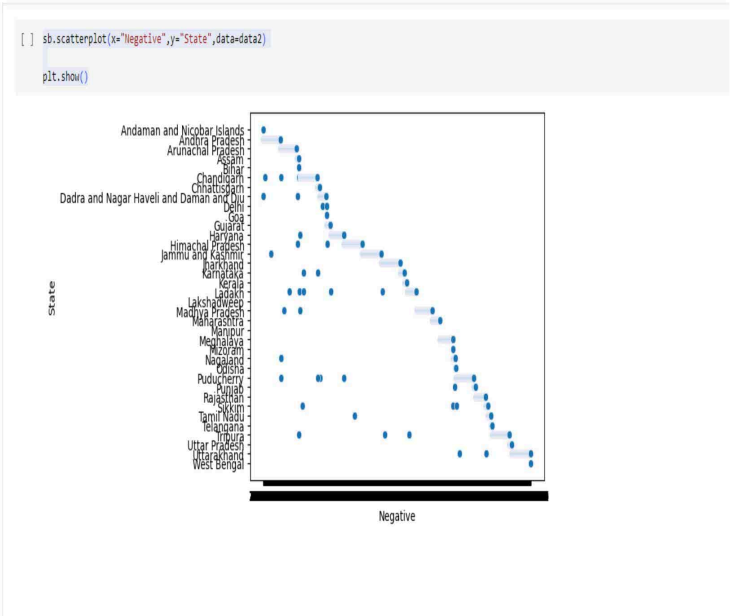
5. RESULT & ANALYSIS

The "Data Analytics of COVID-19" project aimed to analyze and derive insights from various datasets related to the COVID-19 pandemic. The project utilized advanced data analytics techniques to extract meaningful information, trends, and patterns from the available data. The analysis focused on understanding the spread of the virus, its impact on different demographics, healthcare systems, and the effectiveness of various interventions.









6. CONCLUSION

In conclusion, the "Data Analytics of COVID-19" project played a pivotal role in providing data-driven insights to inform decision-making during the pandemic. By analyzing a wide range of data points, the project contributed to a better understanding of the virus's behavior, its impact on society, and the effectiveness of various interventions. The findings and methodologies from this project have the potential to influence future pandemic preparedness and response strategies.

7. REFERENCES

7.1 Wikipedia

7.2 Www.geeksforgeeks.org

7.3 Www.onlinegdb.com

7.4 Www.atago.net

7.5 www.ijraset.com

7.6 www.atlantis-press.com

7.7 www.researchgate.net