**COVID 19 CASE ANALYSIS USING COGNOS**

**BATCH MEMBER**

**611721205020: PREETHI S**

**Phase 5 submission document**

**Project Title: Covid 19 case Analysis**

**Phase 5: PROJECT DOCUMENTATION AND SUBMISSION**

**Topic:**

Outline the project's objective, design thinking process, and development phases.

Describe the analysis objectives, data collection process, data visualization using IBM Cognos, and insights generated from the comparison.

Explain how the insights from the analysis can aid in understanding COVID-19 trends and impacts.

**OBJECTIVE:**

**Explore Data Trends:**

Analyzing COVID-19 data trends over time, including infection rates, recovery rates, and mortality rates in different regions.

**Identify Patterns:**

Using Cognos to identify patterns in the data, such as high-risk areas, demographic groups most affected, and common symptoms among patients.

**Predictive Modeling:**

Implementing predictive models to forecast potential outbreaks based on historical data, aiding in proactive measures.

**Resource Allocation:**

Analyzing the data to optimize resource allocation, helping healthcare facilities manage equipment, staff, and other resources efficiently.

**Public Awareness:**

Creating visualizations and reports to communicate the findings effectively, raising public awareness about the importance of preventive measures.

**Policy Recommendations:**

Providing data-driven recommendations to policymakers for implementing targeted interventions and public health policies.

**Comparative Analysis:**

Comparing COVID-19 data across different countries or regions, understanding variations in response strategies and their impact on case numbers.

**DESIGN THINKING PROCESS:**

**1)Empathize:**

**Understand Stakeholders:**

Identifying the key stakeholders, such as healthcare professionals, policymakers, and the general public, and understand their needs and concerns regarding COVID-19 data analysis.

**User Research:**

Gathering insights from potential users to understand what specific data analysis and visualizations would be most useful for them.

**2)Define:**

**Problem Definition:**

Clearly defining the problem you aim to address through your analysis, such as identifying trends, predicting outbreaks, or optimizing resource allocation.

**User Needs:**

Based on research, articulating the specific needs and requirements of your stakeholders regarding COVID-19 data analysis.

**3)Ideate:**

**Brainstorm Solutions:**

Encouraging creative brainstorming sessions to generate ideas for data analysis techniques, visualization methods, and predictive models.

**Prototyping:**

Developing rough prototypes of visualizations and models to visualize how the data will be presented and interpreted.

**4)Prototype:**

**Develop Prototypes:**

Using Cognos to create prototypes of your data visualizations and analytical models. Test these prototypes with a small group of users to gather feedback.

**Iterate:**

Based on feedback, iterating on your prototypes to refine the design, ensuring that the final output is user-friendly and meets stakeholders' needs.

**5)Test:**

**User Testing:**

Conducting extensive user testing sessions with various stakeholders to validate the effectiveness of your data analysis tools and visualizations.

**Feedback Analysis:**

Analysing the feedback received during testing and make necessary adjustments to improve the user experience and the accuracy of your analysis.

**6)Implement:**

**Finalize Tools:**

Based on the feedback and testing results, finalizing data analysis tools, predictive models, and visualization techniques.

**Deployment:**

Deploy the final solution, ensuring that it is accessible to the intended users and meets their requirements effectively.

**7)Evaluate:**

**Impact Assessment:**

Evaluating the impact of your data analysis by assessing how it has been utilized by stakeholders, the decisions influenced, and the improvements made in addressing the problem defined.

**Iterative Improvement:**

Using feedback and evaluation results to make continuous improvements to your analysis methods and tools, ensuring they remain relevant and effective.

**DEVELOPMENT PHASES OF THE PROJECT:**

**PHASE 1: PROJECT OBJECTIVE AND DEFINITION**

**PHASE 2: INNOVATION**

**PHASE 3: DEVELOPMENT PART 1**

**PHASE 4: DEVELOPMENT PART 2**

**PHASE 5:DOCUMENTING THE PROJECT**

**1)Planning:**

**Project Scope**

Defining the scope of your analysis, specifying the data sources, analysis techniques, and the outcomes you aim to achieve.

**Resource Allocation:**

Allocating necessary resources, including data sets, tools (such as Cognos), and a skilled team for data analysis and development.

**2) Data Collection and Preparation:**

**Data Gathering:**

Collecting relevant COVID-19 data from reliable sources, ensuring the data is accurate, comprehensive, and up-to-date.

**Data Cleaning:**

Cleansing the data to remove inconsistencies, missing values, and errors that could affect the accuracy of your analysis.

**Data Integration:**

Integrating data from different sources if necessary, creating a unified dataset for analysis.

**3)Data Analysis and Modeling:**

**Exploratory Data Analysis (EDA):**

Performing EDA to understand the data's characteristics, identify patterns, and gain insights into the COVID-19 cases.

**Statistical Analysis:**

Using statistical techniques to analyze trends, correlations, and other relevant metrics within the data.

**Predictive Modeling:**

Implementing predictive models to forecast future COVID-19 cases based on historical data, using techniques like regression, time series analysis, or machine learning algorithms.

**4)Visualization and Reporting:**

**Data Visualization:**

Utilizing Cognos to create interactive and informative data visualizations, such as charts, graphs, and maps, to represent COVID-19 trends visually.

**Report Generation:**

Generating detailed reports summarizing the analysis findings, insights, and recommendations for stakeholders.

**5)Implementation:**

**Cognos Development:**

Developing the required dashboards, reports, and data models within Cognos, ensuring they align with the project objectives and stakeholder requirements.

**Integration:**

Integrating the developed Cognos solutions with other tools or systems if necessary, ensuring seamless data flow and accessibility.

**6)Testing:**

**Quality Assurance:**

Conducting rigorous testing to validate the accuracy of your data analysis, ensuring that the visualizations and models provide reliable results.

**User Acceptance Testing (UAT):**

Allowing stakeholders to perform UAT to validate that the developed solutions meet their needs and expectations.

**7)Deployment:**

**Deployment Planning:**

Planing the deployment strategy, considering factors such as server requirements, user access, and security protocols.

**Rollout:**

Deploy the finalized Cognos solutions to the intended users, ensuring they are trained to use the tools effectively.

**8)Monitoring and Maintenance:**

**Performance Monitoring:**

Monitoring the performance of the deployed solutions, ensuring they operate efficiently and provide real-time data analysis.

**Maintenance:**

Addressing any issues, update data sources, and modify visualizations as needed to ensure the continued relevance and accuracy of your analysis.

**DESCRIBING THE ANALYSIS OBJECTIVE:**

**1)Epidemiological Analysis:**

**Identify Hotspots:**

Determine regions or areas with high infection rates to aid targeted interventions.

**Demographic Patterns:**

Analyze how different demographics are affected to understand vulnerable populations better.

**2)Trend Analysis:**

**Temporal Trends:**

Analyzing how COVID-19 cases have evolved over time, identifying peaks and trends.

**Comparative Analysis:**

Comparing trends across different regions, countries, or continents to identify global patterns.

**3)Healthcare Resource Allocation:**

**Hospital Bed Utilization:**

Analyzing hospitalization rates to predict and optimize bed allocations in healthcare facilities.

**Medical Supplies:**

Predicting demand for medical supplies like ventilators and PPE based on case data.

**4) Public Health Interventions:**

**Effectiveness of Measures:**

Evaluating the impact of interventions like lockdowns or mask mandates on case numbers.

**Social Distancing Adherence:**

Analyzing mobility data to assess adherence to social distancing guidelines.

**5)Predictive Modeling:**

**Outbreak Prediction:**

Developing models to predict potential outbreaks, enabling proactive response measures.

**Vaccine Rollout Planning:**

Using predictive models to forecast vaccination needs and plan distribution strategies.

**6) Public Awareness and Communication:**

**Information Dissemination:**

Identifying the most effective channels for disseminating information about preventive measures.

**Myth vs. Reality:**

Analyzing public sentiment and common misconceptions to tailor awareness campaigns.

**7)Impact Assessment:**

**Economic Impact:**

Analyzing the correlation between COVID-19 cases and economic indicators to understand the pandemic's economic impact.

**Mental Health Impact:**

Studying the relationship between the pandemic and mental health indicators in affected populations.

**8) Long-term Effects and Recovery:**

**Post-Recovery Analysis:**

Studying the long-term effects of COVID-19 on recovered patients.

**Economic Recovery:**

Analyzing strategies and policies that facilitate economic recovery in the aftermath of the pandemic.

**DATA COLLECTION PROCESS:**

**1)Define Data Requirements:**

**Specify Data Variables:**

Determining what specific COVID-19 data points you need (e.g., confirmed cases, deaths, recoveries, demographics, testing rates, hospitalizations).

**Temporal Scope:**

Defining the time period for your analysis (daily, weekly, monthly) and consider historical data for trend analysis.

**2) Identify Reliable Sources:**

**Government Health Departments:**

Utilizing data provided by official health departments, both national and local, as they often publish accurate and up-to-date COVID-19 statistics.

**World Health Organization (WHO):**

Accessing global data and guidelines from WHO, which provides standardized information across countries.

**Research Institutions:**

Considering data from reputable research institutions and universities conducting COVID-19 studies.

**3) Data Collection Methods:**

**Web Scraping:**

Automating the collection process using web scraping tools to extract data from official health department websites or reliable online databases.

**APIs:**

Utilizing APIs provided by organizations like WHO or government health agencies to fetch structured and real-time data.

**Data Aggregators:**

Using data aggregators that compile information from various sources into standardized datasets.

**4) Data Validation and Cleaning:**

**Data Validation:**

Cross-verifying data from multiple sources to ensure accuracy and consistency.

**Cleaning:**

Cleansing the data by removing duplicates, correcting errors, and addressing missing or inconsistent values.

**Survey and Research Studies:**

Considering survey data and research studies related to COVID-19, especially for demographic and behavioral insights.

**5) Data Storage and Management:**

**Database:**

Storing the collected data in a structured database system, ensuring it is well-organized and easily accessible for analysis.

**Versioning:**

Implementing version control to track changes and updates made to the dataset over time.

**6) Ethical Considerations:**

**Privacy:**

Ensuring that the collected data is anonymized and does not violate individuals' privacy rights.

**Compliance:**

Adhering to data protection laws and regulations, obtaining necessary permissions if dealing with sensitive personal data,

**7) Documentation:**

**Metadata:**

Documenting metadata such as data source, collection methods, and any data transformations applied.

**Data Dictionary:**

Creating a data dictionary detailing the meaning and format of each variable in your dataset.

**8) Data Update and Maintenance:**

**Regular Updates:**

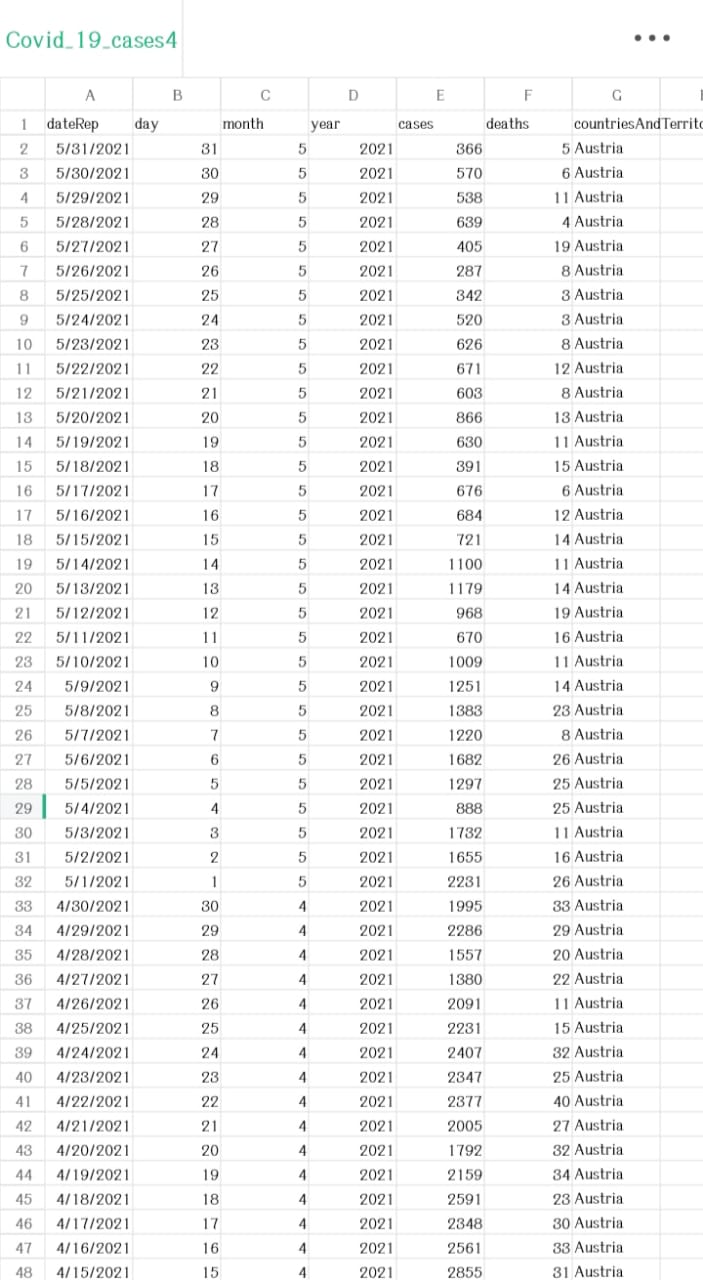
Establishing a schedule for regular data updates to maintain the relevance of your analysis.

**Data Maintenance:**

Continuously monitoring data sources for changes, update collection methods accordingly, and perform periodic data audits.

**DATA VISUALIZATION:**

**Given Dataset:**



2731 rows x 5 columns

**CODE:**

* #Import necessary libraries

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler, LabelEncoder

# Load your COVID-19 dataset into a Pandas DataFrame

data = pd.read\_csv(‘https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases')

# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Standardize features by removing the mean and scaling to unit variance

scaler = StandardScaler()

# Handle missing values (if any)

data = data.dropna()

# Encode categorical variables (if any)

label\_encoders = {}

categorical\_columns = ['column1', 'column2'] # Specify the categorical columns in your dataset

for column in categorical\_columns:

label\_encoders[column] = LabelEncoder()

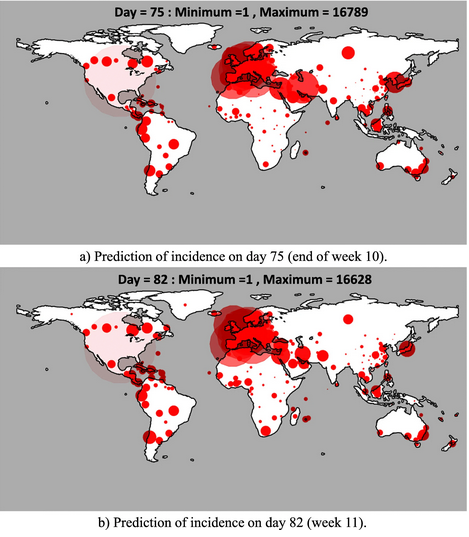
data[column] = label\_encoders[column].fit\_transform(data[column])

# Split the data into features (X) and target variable (y)

X = data.drop(columns=['target\_column']) # Replace 'target\_column' with the name of your target variable

y = data['target\_column']

**OUTPUT:**

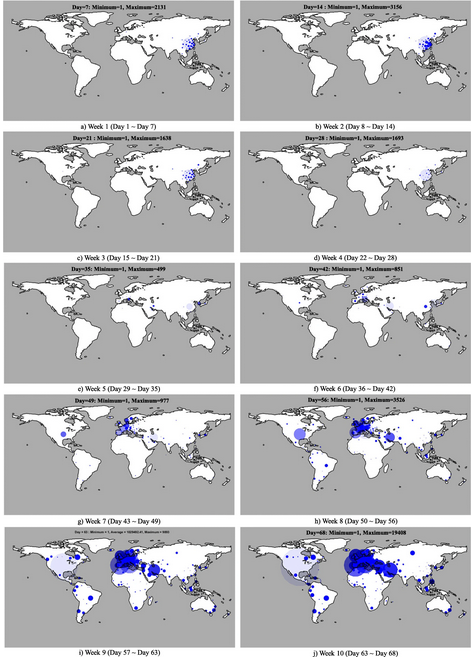


# Split the data into training and testing sets

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Standardize features by removing the mean and scaling to unit variance

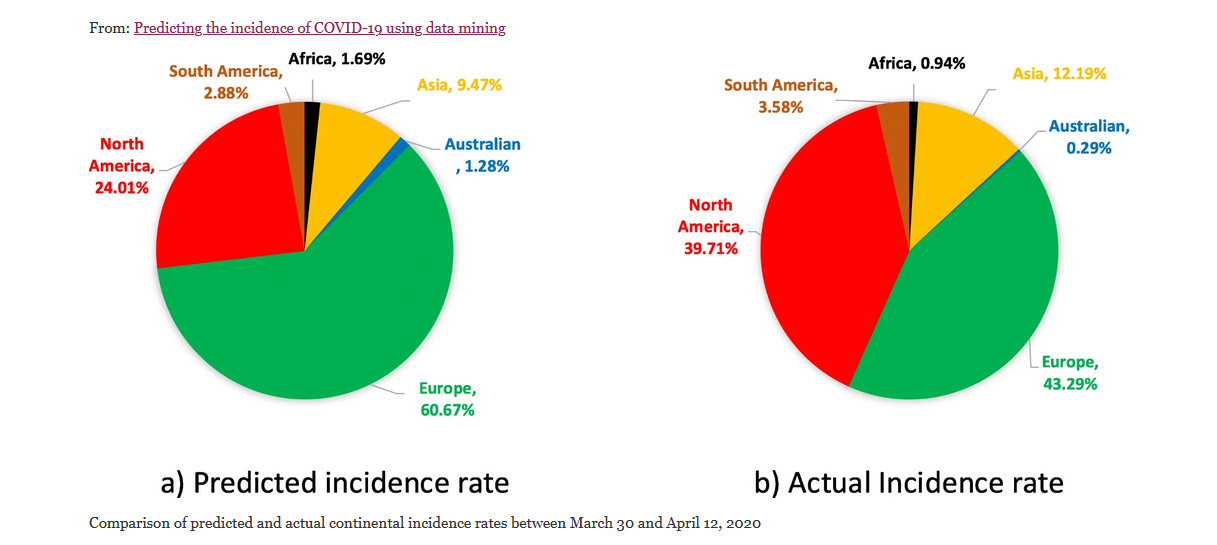
scaler = StandardScaler()



X\_train\_scaled = scaler.fit\_transform(X\_train)

X\_test\_scaled = scaler.transform(X\_test)

**OUTPUT:**



import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load your preprocessed COVID-19 dataset into a Pandas DataFrame

# (Assuming you already have a preprocessed DataFrame named 'data')

# data = pd.read\_csv(‘https://www.kaggle.com/datasets/chakradharmattapalli/covid-19-cases’)

sns.lineplot(x='Date', y='Cases', data=data)

plt.xlabel('Date')

plt.ylabel('Number of Cases')

plt.title('Daily COVID-19 Cases Over Time')

plt.xtic')

Bar plot - COVID-19 cases by country

plt.figure(figsize=(12, 6))

sns.barplot(x='Country', y='Cases', data=data)

plt.xticks(rotation=90)

plt.xlabel('Country')

plt.ylabel('Number of Cases')

plt.title('COVID-19 Cases by Country')

plt.show()

Line plot - Daily cases over time

plt.figure(figsize=(12, 6))

sns.lineplot(x='Date', y='Cases', data=data)

plt.xlabel('Date')

plt.ylabel('Number of Cases')

plt.title('Daily COVID-19 Cases Over Time')

plt.xticks(rotation=45)

plt.show()

Box plot - Distribution of cases by continent

plt.figure(figsize=(10, 6))

sns.boxplot(x='Continent', y='Cases', data=data)

plt.xlabel('Continent')

plt.ylabel('Number of Cases')

plt.title('Distribution of COVID-19 Cases by Continent')

plt.show()

Heatmap - Correlation between numerical variables

correlation\_matrix = data.corr()

plt.figure(figsize=(10, 8))

sns.heatmap(correlation\_matrix, annot=True, cmap='coolwarm')

plt.title('Correlation Heatmap')

plt.show()

Pairplot - Relationship between multiple variables

sns.pairplot(data[['Cases', 'Deaths', 'Recovered', 'Population']])

plt.suptitle('Pairplot of COVID-19 Data', y=1.02)

plt.show()

**VISUALIZATIONS:**

PLOTTING SIMPLE PLOT:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data = pd.read\_csv('case\_time\_series.csv')

Y = data.iloc[61:,1].values

R = data.iloc[61:,3].values

D = data.iloc[61:,5].values

X = data.iloc[61:,0]

plt.plot(X,Y)

**OUTPUT:**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**data = pd.read\_csv('case\_time\_series.csv')**

**Y = data.iloc[61:,1].values**

**R = data.iloc[61:,3].values**

**D = data.iloc[61:,5].values**

**X = data.iloc[61:,0]**

**plt.figure(figsize=(25,8))**

**ax = plt.axes()**

**ax.grid(linewidth=0.4, color='#8f8f8f')**

**ax.set\_facecolor("black")**

**ax.set\_xlabel('\nDate',size=25,color='#4bb4f2')**

**ax.set\_ylabel('Number of Confirmed Cases\n',**

**size=25,color='#4bb4f2')**

**ax.plot(X,Y,**

**color='#1F77B4',**

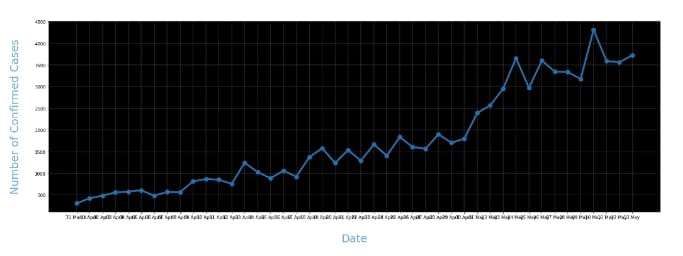
**marker='o',**

**linewidth=4,**

**markersize=15,**

**markeredgecolor='#035E9B**

**OUTPUT:**

****

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**data = pd.read\_csv('case\_time\_series.csv')**

**Y = data.iloc[61:,1].values**

**R = data.iloc[61:,3].values**

**D = data.iloc[61:,5].values**

**X = data.iloc[61:,0]**

**plt.figure(figsize=(25,8))**

**ax = plt.axes()**

**ax.grid(linewidth=0.4, color='#8f8f8f')**

**ax.set\_facecolor("black")**

**ax.set\_xlabel('\nDate',size=25,color='#4bb4f2')**

**ax.set\_ylabel('Number of Confirmed Cases\n',**

**size=25,color='#4bb4f2')**

**plt.xticks(rotation='vertical',size='20',color='white')**

**plt.yticks(size=20,color='white')**

**plt.tick\_params(size=20,color='white')**

**for i,j in zip(X,Y):**

**ax.annotate(str(j),xy=(i,j+100),color='white',size='13')**

**ax.annotate('Second Lockdown 15th April',**

**xy=(15.2, 860),**

**xytext=(19.9,500),**

**color='white',**

**size='25',**

**arrowprops=dict(color='white',**

**linewidth=0.025))**

**plt.title("COVID-19 IN : Daily Confirmed\n",**

**size=50,color='#28a9ff')**

**ax.plot(X,Y,**

**color='#1F77B4',**

**marker='o',**

**linewidth=4,**

**markersize=15,**

**markeredgecolor='#035E9B**

**OUTPUT:**

****

**data = pd.read\_csv('district.csv')**

**data.head()**

**re=data.iloc[:30,5].values**

**de=data.iloc[:30,4].values**

**co=data.iloc[:30,3].values**

**x=list(data.iloc[:30,0])**

**plt.figure(figsize=(25,10))**

**ax=plt.axes()**

**ax.set\_facecolor('black')**

**ax.grid(linewidth=0.4, color='#8f8f8f')**

**plt.xticks(rotation='vertical',**

**size='20',**

**color='white')#ticks of X**

**plt.yticks(size='20',color='white')**

**ax.set\_xlabel('\nDistrict',size=25,**

**color='#4bb4f2')**

**ax.set\_ylabel('No. of cases\n',size=25,**

**color='#4bb4f2')**

**plt.tick\_params(size=20,color='white')**

**ax.set\_title('Maharashtra District wise breakdown\n',**

**size=50,color='#28a9ff')**

**plt.bar(x,co,label='re')**

**plt.bar(x,re,label='re',color='green')**

**plt.bar(x,de,label='re',color='red')**

**for i,j in zip(x,co):**

**ax.annotate(str(int(j)),**

**xy=(i,j+3),**

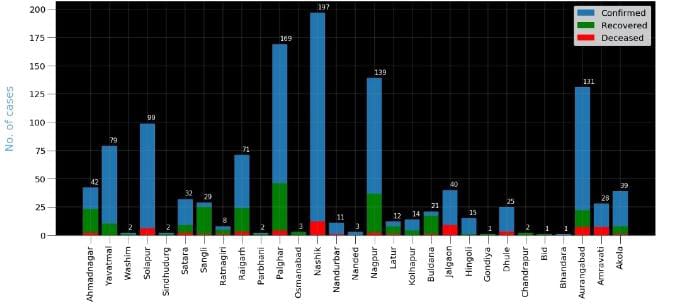
**color='white',**

**size='15')**

**plt.legend(['Confirmed','Recovered','Deceased'],**

**fontsize=20)**

**OUTPUT:**

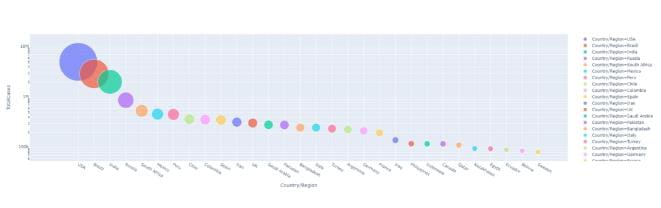
****

**px.scatter(dataset1.head(30), x='Country/Region', y='TotalCases',**

**hover\_data=['Country/Region', 'Continent'],**

**color='Country/Region**

**OUTPUT:**

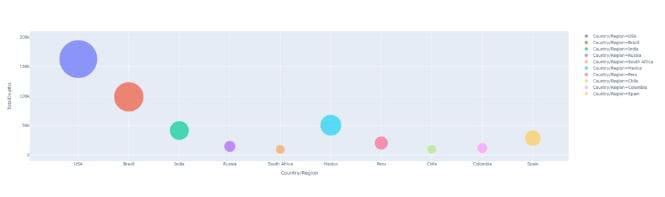
****

**px.scatter(dataset1.head(10), x='Country/Region', y= 'TotalDeaths',**

**hover\_data=['Country/Region', 'Continent'],**

**color='Country/Region', size= 'TotalDeaths', size\_max=80)**

**OUTPUT:**

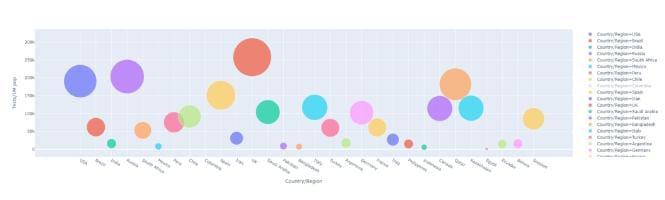
****

**px.scatter(dataset1.head(30), x='Country/Region', y= 'Tests/1M pop',**

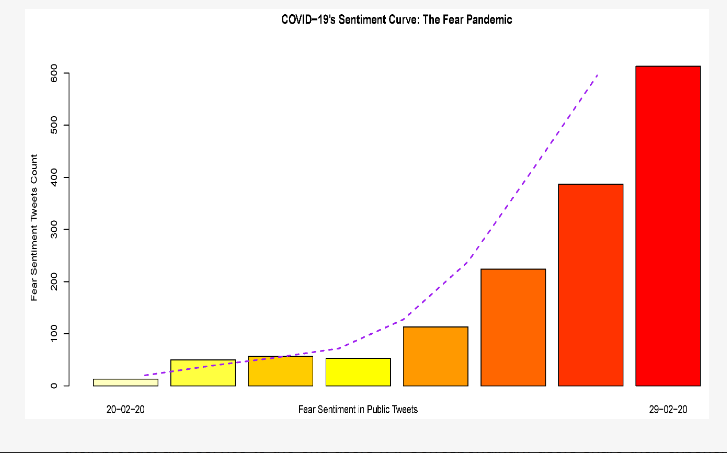
**hover\_data=['Country/Region', 'Continent'],**

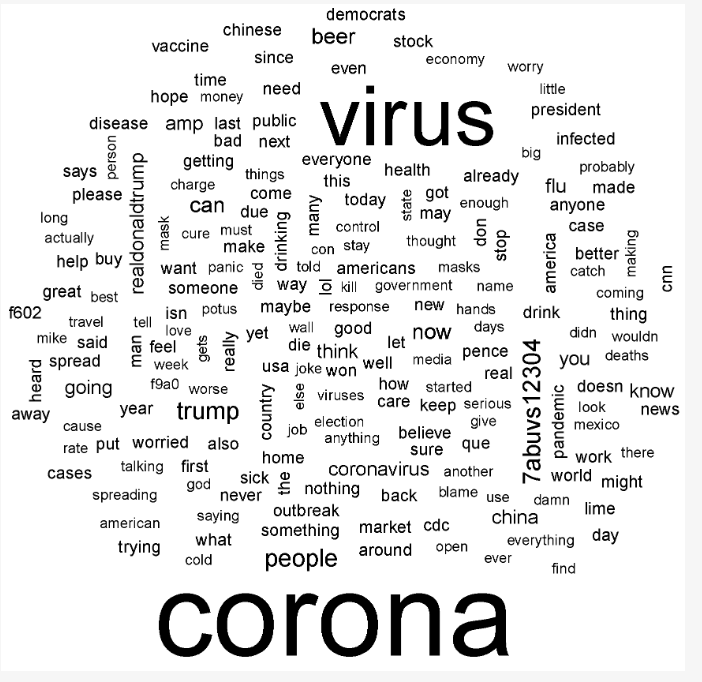
**color='Country/Region', size= 'Tests/1M pop', size\_max=80)**

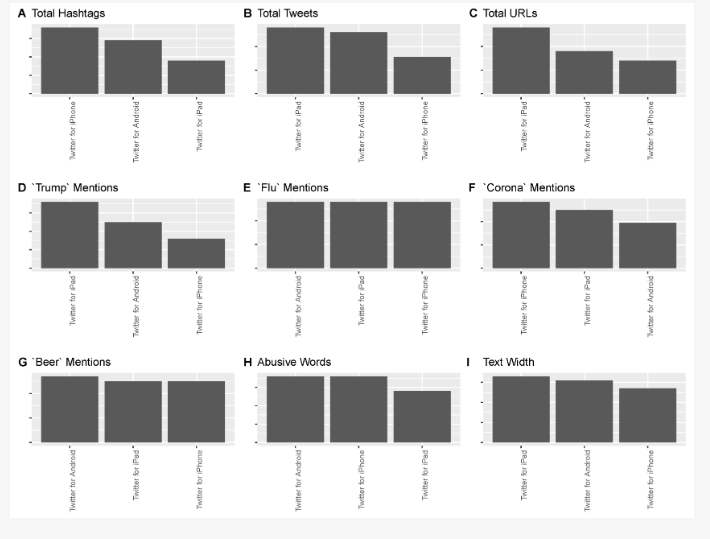
**OUTPUT:**

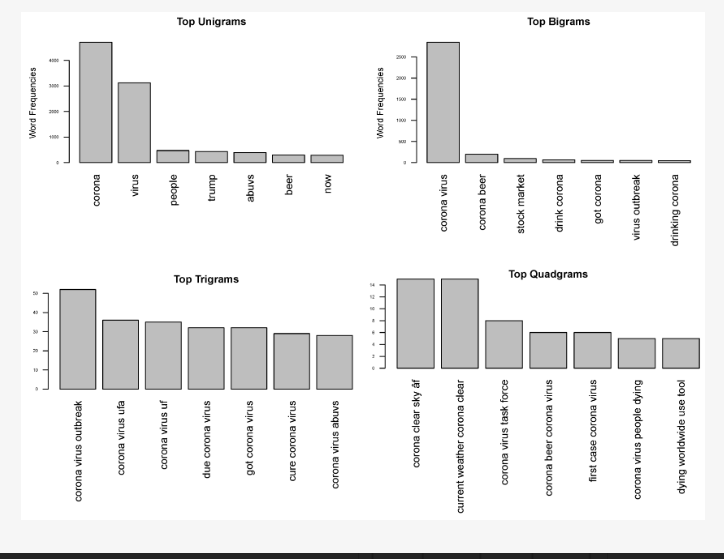
****

**INSIGHTS GENERATED BY THE COMPARISON:**

****

****

****

****

**CONCLUSION OF PHASE 5:**

In the phase 5 conclusion of documentation of the project was completed. we performed the covid 19 cases analysis and create visualizations, death cases , recovered, decreased, maximum causes. then we created some visualizations using data visualization libraries(e.g., matplotlib, seaborn) for **covid 19 cases analysis** by using the given dataset and it was documented.