**IBM DATA ANALYTICS WITH COGNOS - GROUP4 PHASE III**

**1. Team Members :**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  | | --- | --- | --- | --- | | **S.No** | **Name of the Student** | **E-Mail ID** | **Phone No.** | | **1** | Thanigavel J | thanigavellovely@gmail.com | 7708345336 | | **2** | Karthisan M | [karthim978789@gmail.com](mailto:karthim978789@gmail.com) | 8754150560 | | **3** | Aravindhan P | Dcsearavind41@gmail.com | 6382750935 | | 4 | Balamurugan E | ebalamuruganbalamurugan@gmail.com | 8870556796 | | 5 | Sanjai S | Sanjaisanjaisanjai555@gmail.com | 9087943812 | |

**2. Name of the Guide** : Ms.Suganya

**Department /**

**Designation**  : CSE/AP

**Institutional Address**  : Chettinad College of Engineering and

Technology

**3. Project Title**  : COVID Vaccines Analysis

Analyzing COVID-19 vaccine data involves various aspects, including data collection, cleaning, visualization, and statistical analysis. Here, I'll provide a general outline of the steps involved in analyzing COVID-19 vaccine data and include some example Python code snippets for illustration. Please note that the specific data sources and formats may vary, so you'll need to adapt the code to your particular dataset.

**Step 1: Data Collection**

Collect data from trusted sources like the World Health Organization (WHO), CDC, or national health agencies. Data may include vaccine coverage, adverse events, and case statistics.

**Step 2: Data Cleaning and Preprocessing**

Clean and preprocess the data to make it suitable for analysis. This may involve dealing with missing values, standardizing data formats, and merging multiple datasets. Python libraries like Pandas are helpful for this task.

**Code**

import pandas as pd

# Load your data

vaccine\_data = pd.read\_csv('vaccine\_data.csv')

# Data cleaning and preprocessing

# Example: Removing rows with missing values

vaccine\_data.dropna(inplace=True)

# Standardizing date format

vaccine\_data['date'] = pd.to\_datetime(vaccine\_data['date'])

**Step 3: Data Visualization**

Create visualizations to understand the data better. Matplotlib and Seaborn are popular libraries for data visualization in Python.

**Code**

import matplotlib.pyplot as plt

import seaborn as sns

# Example: Plotting vaccine coverage over time

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

sns.lineplot(data=vaccine\_data, x='date', y='coverage')

plt.title('Vaccine Coverage Over Time')

plt.xlabel('Date')

plt.ylabel('Coverage')

plt.show()

**Step 4: Statistical Analysis**

Perform statistical analyses to extract insights from the data. Depending on the specific questions you want to answer, this could include hypothesis testing, regression analysis, or time series analysis.

Hypothesis Testing (e.g., comparing vaccine coverage in two regions)

**Code**

from scipy.stats import ttest\_ind

region\_A = vaccine\_data[vaccine\_data['region'] == 'A']['coverage']

region\_B = vaccine\_data[vaccine\_data['region'] == 'B']['coverage']

t\_stat, p\_value = ttest\_ind(region\_A, region\_B)

if p\_value < 0.05:

print('There is a significant difference between regions A and B.')

**Regression Analysis (e.g., predicting vaccine coverage based on various factors)**

**Code**

import statsmodels.api as sm

X = vaccine\_data[['population', 'income']]

y = vaccine\_data['coverage']

X = sm.add\_constant(X) # Add a constant term

model = sm.OLS(y, X).fit()

predictions = model.predict(X)

print(model.summary())

**Step 5: Interpretation and Reporting**

Interpret the results of your analysis and prepare a report or presentation to communicate your findings.

Please note that this is a simplified example, and COVID-19 vaccine data analysis can be much more complex, depending on the specific research questions and the available data. Additionally, ensure that you have appropriate ethical and privacy considerations in place when working with health-related data.