COVID VACCINES ANALYSIS

Data Analytics With Cognos(PHASE 3)

INTRODUCTION

The COVID-19 pandemic has made vaccine analysis critical in understanding vaccination trends and efficacy. In this project, we use IBM Cognos to analyze and visualize COVID vaccine data. Our objective is to assess vaccine distribution, effectiveness, and their impact on the pandemic.

ANALYSIS OBJECTIVES:

1.Data Preprocessing: Clean and prepare the COVID vaccine dataset, handling missing values and outliers. Standardize data formats and ensure consistency in data types. Merge relevant datasets for a comprehensive analysis.

2.Data Exploration: Conduct exploratory data analysis (EDA) to understand the characteristics of the COVID vaccine dataset. Calculate basic statistics, such as mean, median, and standard deviation for key variables. Visualize data distributions and relationships between variables.

3.Vaccine Effectiveness Analysis: Assess the effectiveness of different COVID vaccines in terms of preventing infections, hospitalizations, and deaths. Analyze the impact of vaccination campaigns on infection rates and healthcare outcomes. Identify any variations in vaccine performance based on demographics or regions.

4.Time Series Analysis: Create time series visualizations to track the progress of vaccination campaigns over time. Identify trends and patterns in vaccine distribution and coverage. Correlate vaccination rates with changes in infection rates.

5.Geospatial Analysis: Utilize geospatial data to map vaccine distribution and coverage at regional and global levels. Analyze regional disparities in vaccine availability and uptake. Explore the relationship between vaccination rates and disease spread in specific areas.

6.IBM Cognos Integration: Use IBM Cognos to create interactive dashboards and reports for data analysis. Design visually appealing visualizations, such as charts, graphs, and maps, to communicate insights effectively. Enable users to interact with the data, filter information, and gain real-time insights.

7.Performance Monitoring: Set up key performance indicators (KPIs) to monitor the progress of vaccination campaigns. Build alerts and notifications to inform stakeholders of significant developments. Ensure data in IBM Cognos is regularly updated to provide real-time information.

8.Predictive Modeling (Optional): Develop predictive models to forecast vaccine distribution needs and potential future outbreaks. Use historical data to build models for decision support and resource allocation.

9.Share Insights: Create comprehensive reports and presentations summarizing the analysis findings. Share actionable insights with healthcare professionals, policymakers, and the public to support evidence-based decision-making.

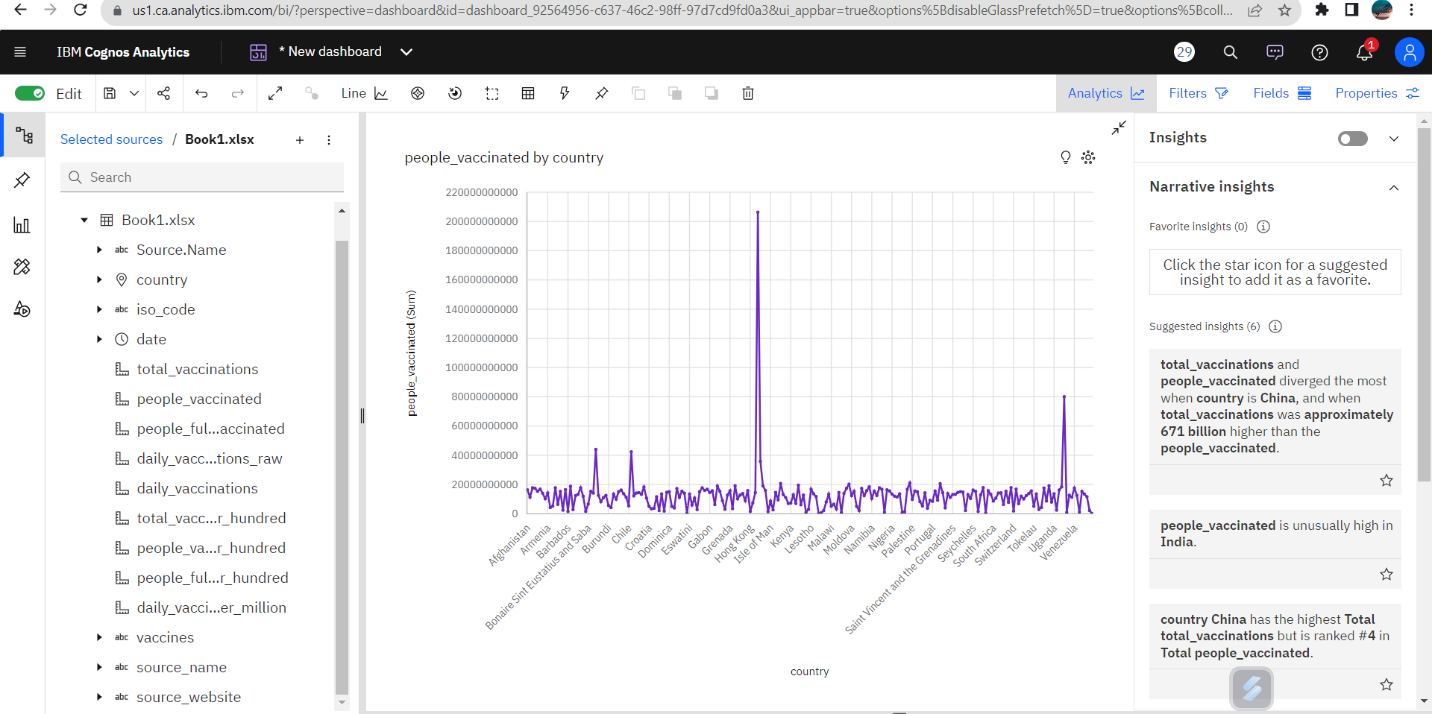
10.Continuous Improvement: Continuously assess and refine the analysis process and visualizations based on feedback and changing data. Stay updated with the latest research and findings in the field of COVID-19 vaccination.

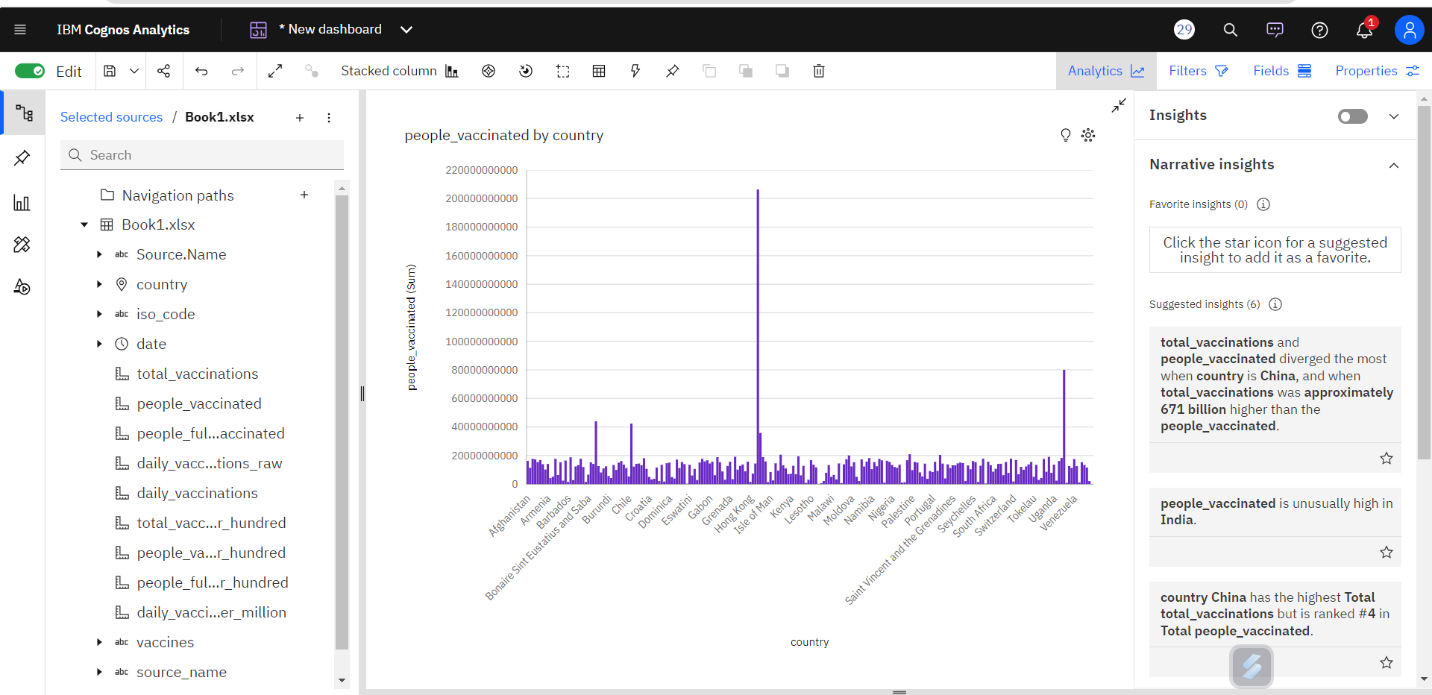
These objectives should provide a comprehensive framework for a COVID vaccines analysis project that involves data preprocessing and analysis using IBM Cognos.

Dataset Link: <https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

VISUALIZATIONS:

The process of finding trends and correlations in our data by representing it pictorially is called Data Visualization. To perform data visualization in python, we can use various python data visualization modules such as Matplotlib, Seaborn, Plotly, etc.





PREPROCESS:

Data preprocessing is essential before its actual use. Data preprocessing is the concept of changing the raw data into a clean data set. The dataset is preprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm.

CODE:

import pandas as pd

from sklearn.preprocessing import Imputer

from sklearn.cluster import KMeans

from sklearn.preprocessing import LabelEncoder

import matplotlib.pyplot as plt

# Load customer data

data = pd.read\_csv('/content/country\_vaccinations.csv')

# null value

# Load the dataset from CSV file

# Removing rows with NaN values

df\_cleaned = data.dropna()

# Print the cleaned DataFrame

print("Original DataFrame:")

print(data)  # Display the first 5 rows of the original DataFrame

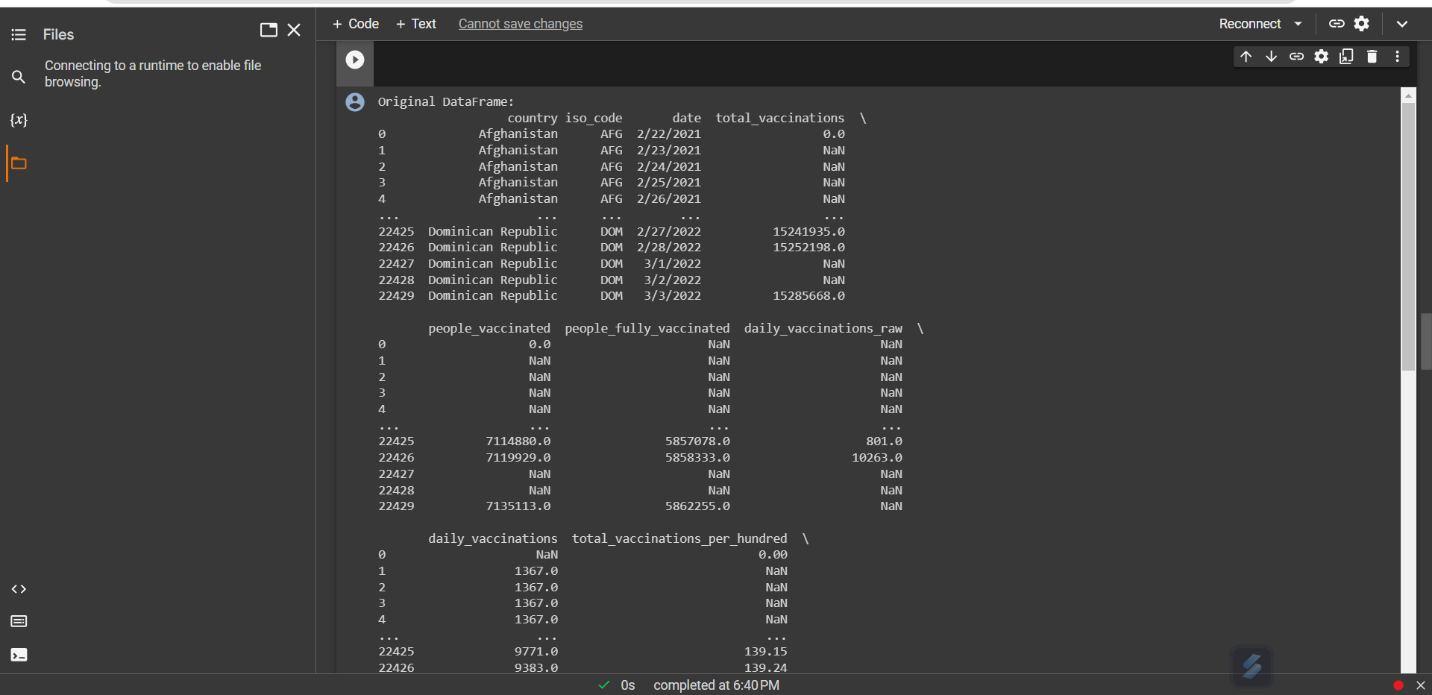
print("\nDataFrame after removing NaN values:")

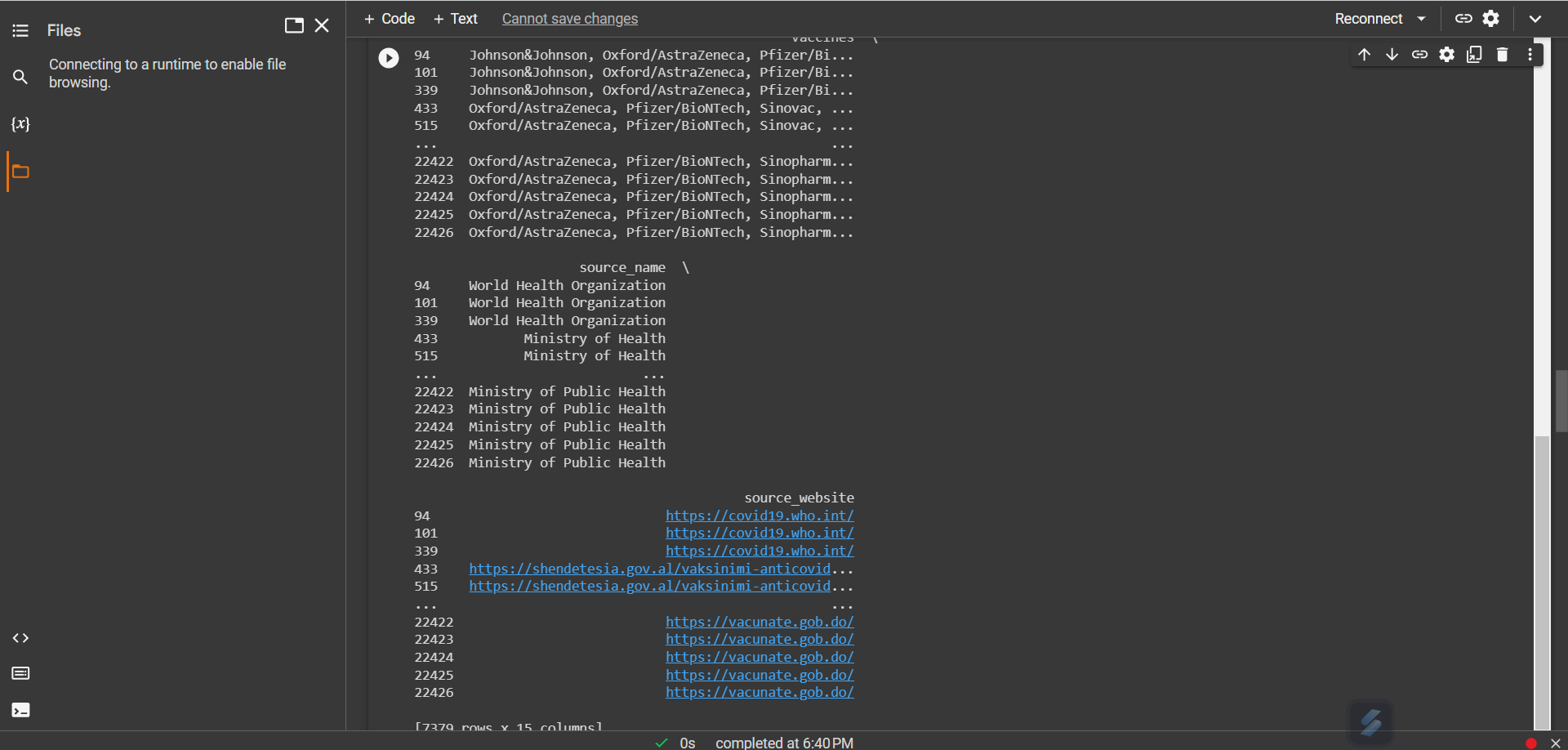
print(df\_cleaned)  # Display the first 5 rows of the cleaned DataFrame

# Optionally, you can save the cleaned DataFrame to a new CSV file

df\_cleaned.to\_csv('cleaned\_covid.csv', index=False)

OUTPUT:





PROGRAM:

import pandas as pd

from sklearn.cluster import KMeans

from sklearn.preprocessing import LabelEncoder

import matplotlib.pyplot as plt

# Load the first dataset

dataset1 = pd.read\_csv('/content/cleaned\_covid.csv') # Replace 'dataset1.csv' with the actual filename of your first dataset

# Load the second dataset

dataset2 = pd.read\_csv('/content/cleaned\_covid\_vaccines\_progress.csv') # Replace 'dataset2.csv' with the actual filename of your second dataset

# Merge the two datasets based on the common column

merged\_dataset = pd.merge(dataset1, dataset2, on='total\_vaccinations')

print(merged\_dataset)

merged\_dataset.to\_csv('merged\_dataset2.csv', index=False)

# Load customer data

data = pd.read\_csv('/content/merged\_dataset2.csv')

# Convert 'country' column to numerical values

le = LabelEncoder()

data['country'] = le.fit\_transform(data['country'])

# Select relevant features for clustering

X = data[['country', 'total\_vaccinations']]

# Determine the number of clusters (you may need to experiment with this)

n\_clusters = 3

# Apply K-Means clustering

kmeans = KMeans(n\_clusters=n\_clusters, random\_state=42)

data['Cluster'] = kmeans.fit\_predict(X)

# Visualize the clusters

plt.scatter(data['country'], data['total\_vaccinations'], c=data['Cluster'], cmap='rainbow')

plt.xlabel('country')

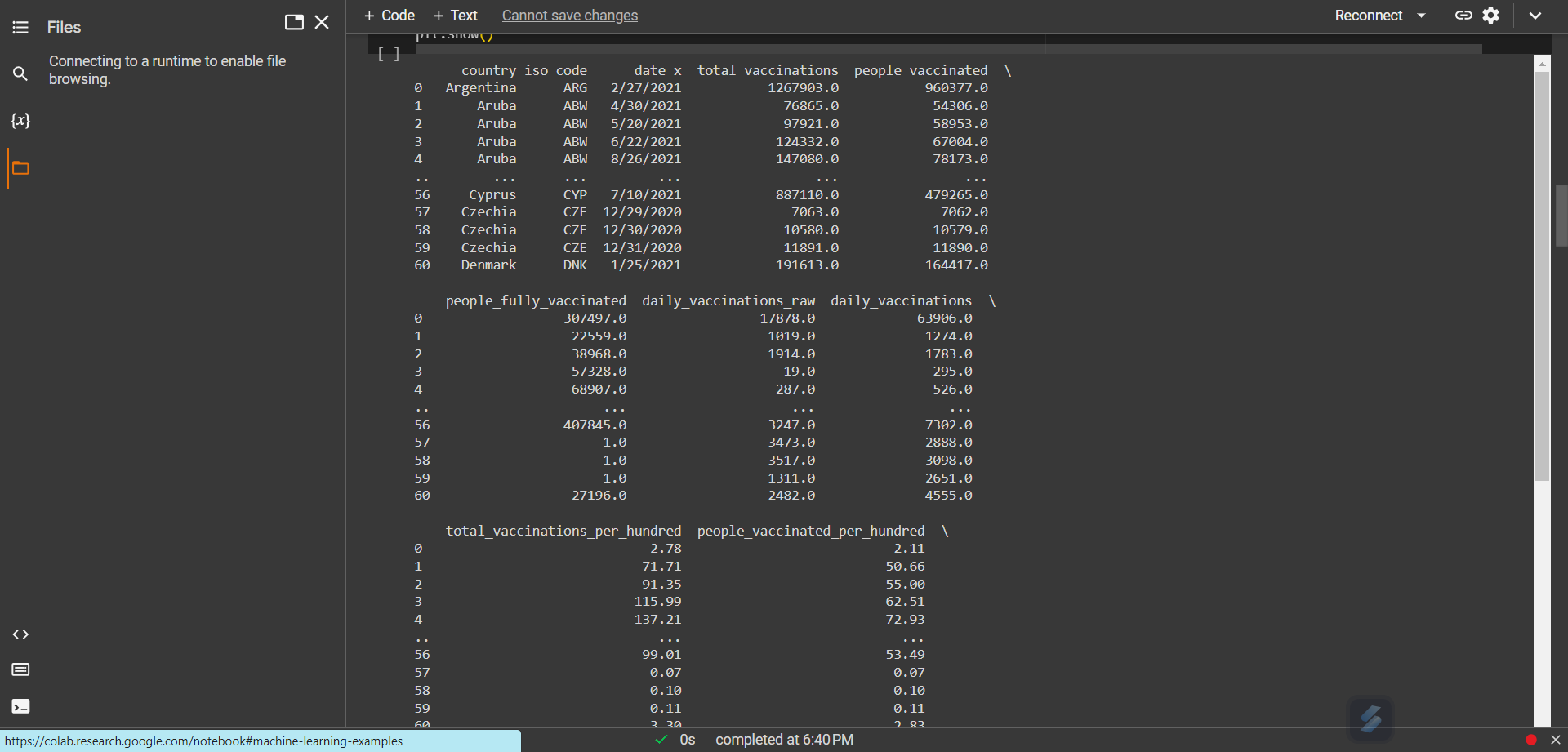
plt.ylabel('total\_vaccinations')

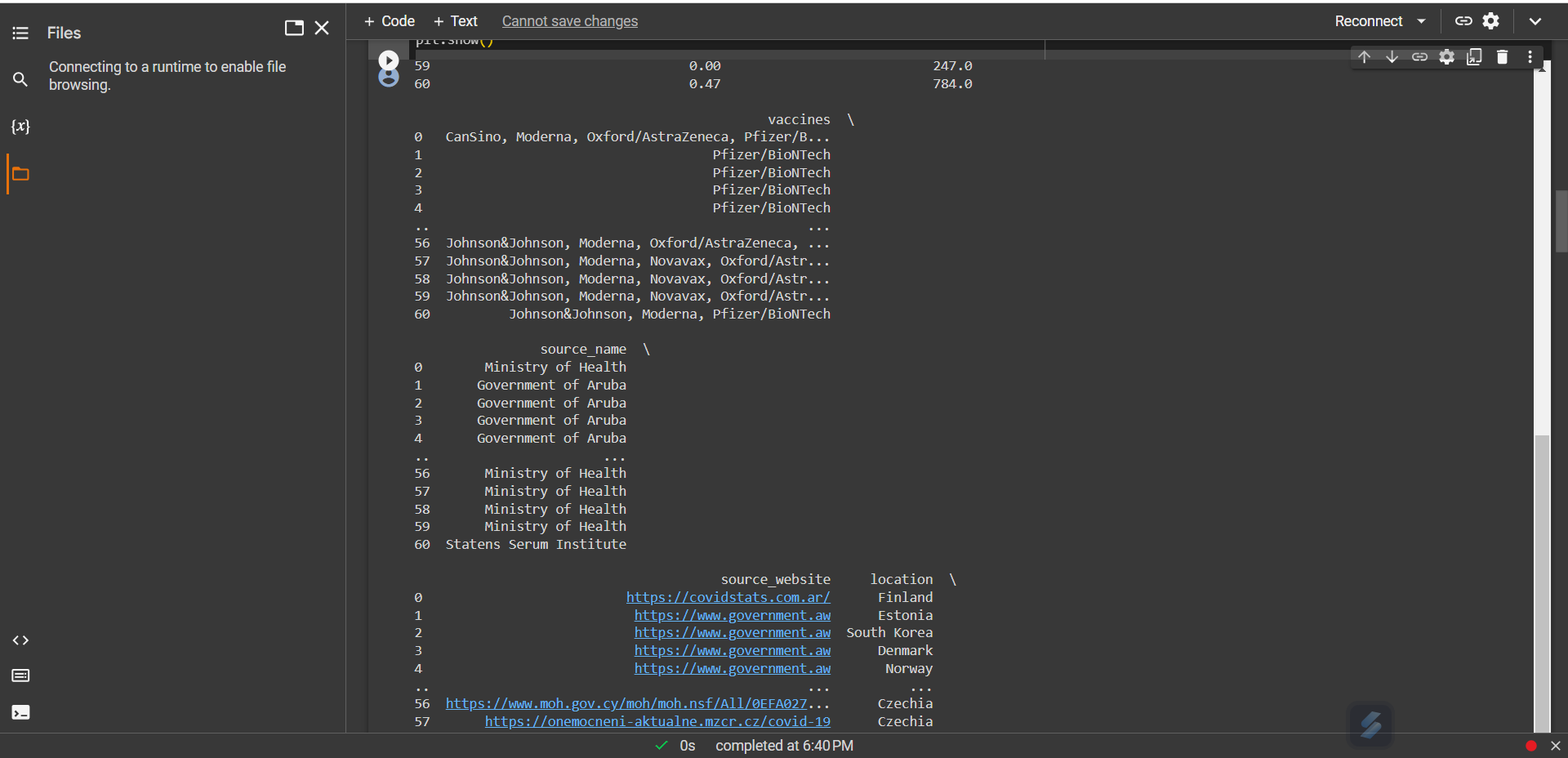
plt.title('covid vaccines')

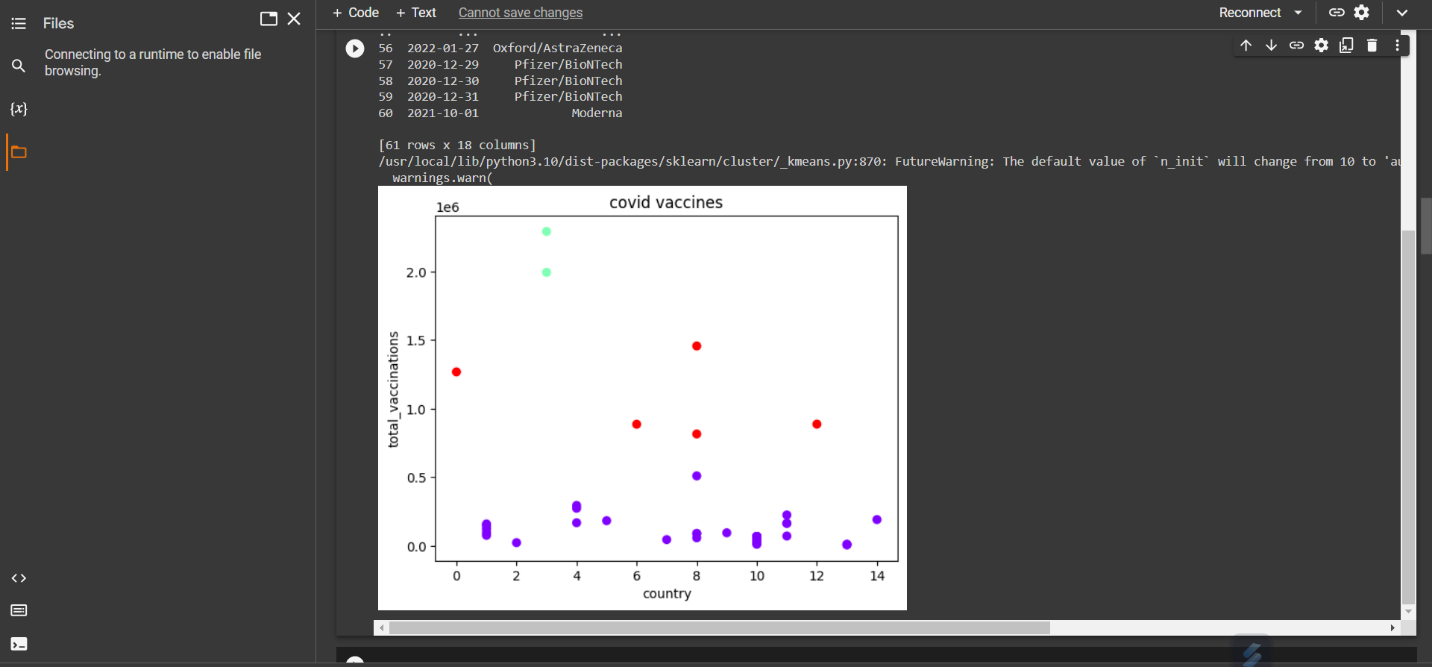
plt.show()

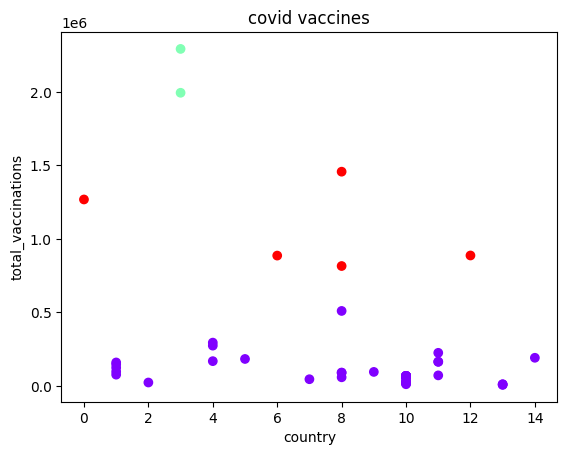
OUTPUT:

AFTER MERGED:









CONCLUSION:

Through the analysis conducted with IBM Cognos, we've gained valuable insights into COVID vaccine distribution and effectiveness. This project provides essential data-driven information to guide vaccination strategies and pandemic management.