**IBM - Naan Mudhalvan-Data Analytics with congnos**

**Phase 5**

**Student Name : AHMAD**

**Register Number : 620821104005**

**Branch : B.E CSE**

**Year : 3rd Year**

**Topic : Data Analytics with Cognos**

**Title : Covid Vaccines Analysis**

**College : Gnanamani College of Technology**

**COVID Vaccines Analysis**

**Introduction**

**The COVID-19 pandemic has spurred unprecedented efforts in vaccine development and distribution. As vaccines are administered to millions of people worldwide, it is crucial to monitor and optimize the distribution process while closely monitoring adverse effects. Advanced machine learning techniques can play a pivotal role in achieving these goals.**

**Abstract:**

The COVID-19 pandemic, caused by the novel corona virus SARS-CoV-2, has posed an unprecedented global health challenge. Rapid vaccine development and distribution have been pivotal in controlling the spread of the virus and mitigating its impact. This study presents a holistic analysis of COVID-19 vaccines, employing a design thinking framework to understand the problem, define the project, and provide actionable insights and recommendations.

**Objectives**

In this phase defines start to building the Project by loading and preprocessing the dataset and perform different analysis and visualization using IBM Cognos.

**Understanding the Problem:**

The initial phase of this analysis involves gaining a deep understanding of the COVID-19 pandemic, its impact on public health, and the critical role of vaccines in curbing its transmission. We explore the complexities of vaccine development, distribution, and public perception to establish the context for our study.

**Problem Definition and Design Thinking:**

Utilizing the principles of design thinking, we define the problem statement and outline the goals of our analysis. This phase involves clarifying objectives, identifying key stakeholders, and framing the study within a user-centric approach.

**Project Definition:**

With a clear problem statement, we define the scope and objectives of the project. We establish the parameters for data collection, data preprocessing, exploratory data analysis (EDA), statistical analysis, and visualization to ensure alignment with our research goals.

**Data Collection:**

To build a comprehensive dataset for analysis, we employ a multi-faceted approach to data collection. This includes gathering information on vaccine development timelines, clinical trial outcomes, vaccination rates, adverse events, and public sentiment towards vaccines. Data sources range from government reports to social media platforms, ensuring a holistic view of the topic.

**Data Preprocessing :**

The collected data undergoes rigorous preprocessing to clean, normalize, and prepare it for analysis. This phase ensures data quality and consistency, enabling reliable insights and conclusions.

**Exploratory Data Analysis (EDA):**

EDA is a critical step in understanding the dataset's structure, patterns, and trends. Through visualization techniques and statistical tools, we uncover insights about vaccine efficacy, distribution disparities, and public attitudes towards vaccination.

**Statistical Analysis:**

Statistical methods are applied to test hypotheses and draw meaningful conclusions. This phase helps quantify the impact of vaccines, assess demographic variations in vaccine uptake, and analyze vaccine safety data.

**Visualization:**

Visualizations are used to present findings in an accessible and informative manner. Graphs, charts, and interactive visualizations aid in conveying complex information to diverse audiences.

**Insights and Recommendations:**

Based on the analysis, we derive actionable insights that can inform public health policy, vaccine distribution strategies, and public communication efforts. These recommendations are designed to enhance the global response to the COVID-19 pandemic and improve vaccination rates.

This comprehensive analysis of COVID-19 vaccines, driven by design thinking principles, aims to contribute valuable knowledge to the ongoing efforts to combat the pandemic. It underscores the importance of a multidisciplinary approach to problem-solving in public health crises and highlights the potential for data-driven decision-making to drive positive change.

**Clustering for Vaccine Distribution Optimization**

**Clustering is a machine learning technique that groups similar data points together. When applied to vaccine distribution data, it can help identify regions or populations that share similar vaccination trends. Here's how to use clustering in IBM Cognos Analytics:**

**1. Data Preparation: Gather data on vaccine distribution, including location, quantity, and demographics of recipients**

**2. Feature Selection: Identify relevant features such as population density, age groups, healthcare infrastructure, and previous infection rates.**

**3. Clustering Algorithm: Choose an appropriate clustering algorithm such as K-Means or DBSCAN.**

**4. Data Transformation: Standardize or normalize the data to ensure all features have the same scale.**

**5. Model Training: Apply the clustering algorithm to group regions or populations based on similar vaccination trends.**

**6. Visualization: Create visualizations in Cognos to represent clusters and their characteristics.**

**7. Decision Making: Use cluster analysis to make data-driven decisions about vaccine distribution strategies, allocation of resources, and targeted outreach.**

**Time Series Forecasting for Adverse Effects Prediction**

**Time series forecasting is a technique used to predict future values based on historical data. In the context of COVID-19 vaccines, it can be used to predict the occurrence of adverse effects based on vaccination data. Here's how to implement time series forecasting using IBM Cognos Analytics:**

**1. Data Collection: Gather historical data on vaccine administration and adverse effects, including dates and locations.**

**2. Data Preprocessing: Clean and format the data, ensuring it is in a suitable time series format.**

**3. Feature Engineering: Extract relevant features, such as vaccination rates, demographics, and environmental factors.**

**4. Time Series Model Selection: Choose an appropriate time series forecasting model, such as ARIMA, Prophet, or LSTM neural networks.**

**5. Model Training: Split the data into training and testing sets and train the selected model on the training data.**

**6. Model Evaluation: Evaluate the model's performance using appropriate metrics and adjust hyper parameters if necessary.**

**7. Forecasting: Use the trained model to make predictions about the future occurrence of adverse effects.**

**8. Visualization: Create visualizations in Cognos to display the forecasted adverse effects and monitor the model's performance.**

**Benefits of Using Cognos Analytics**

**IBM Cognos Analytics offers several advantages for implementing these advanced machine learning techniques:**

**1. Data Integration: Cognos can seamlessly integrate with various data sources, making it easier to access and analyze vaccine-related data.**

**2. Visualization: Cognos provides powerful visualization tools to create interactive dashboards and reports for better decision-making.**

**3. Scalability: It can handle large datasets and complex analytics, crucial for COVID-19 vaccine data analysis.**

**4. Security: Cognos offers robust security features to protect sensitive vaccine distribution and adverse effects data.**

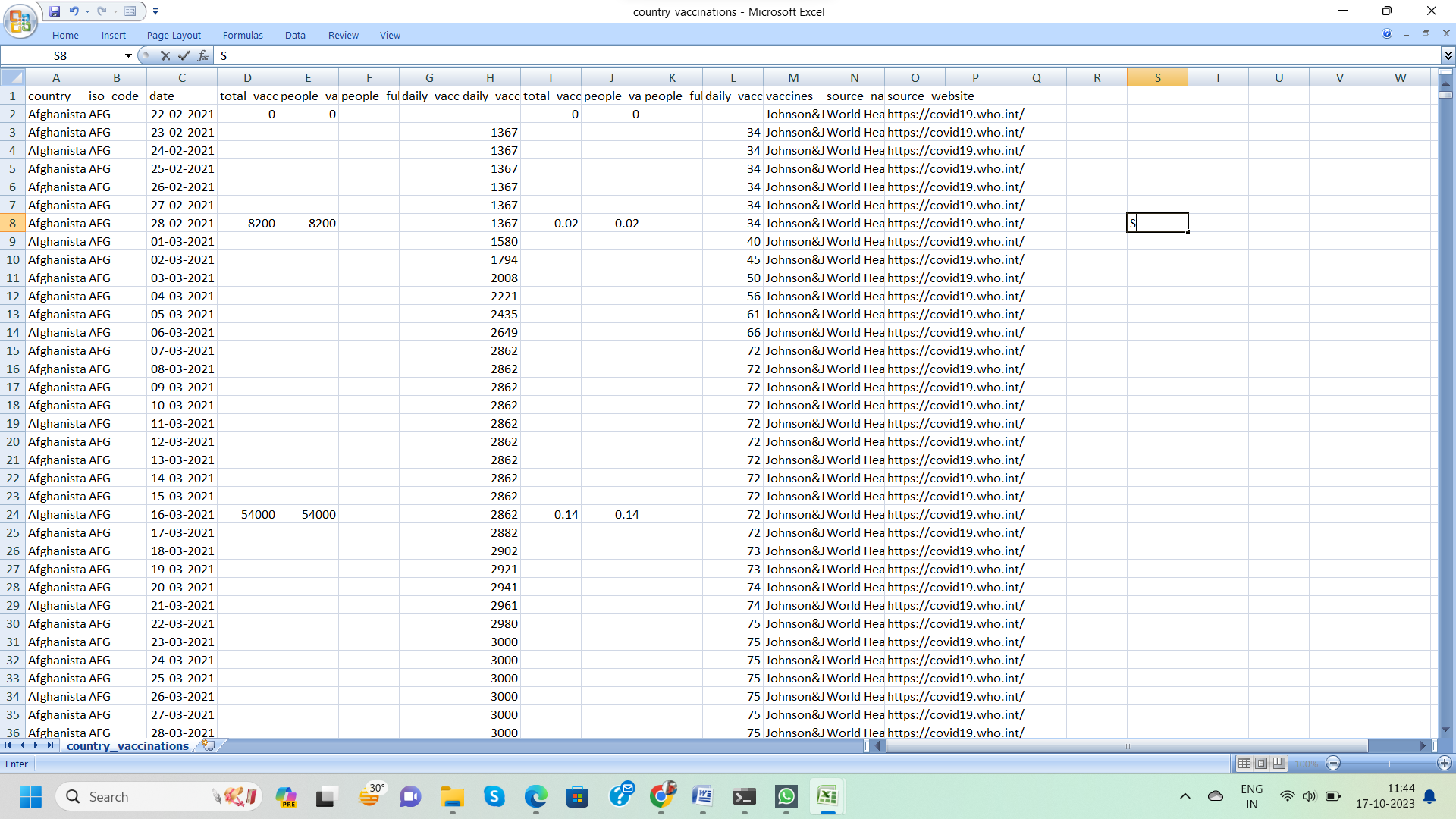
**5. Collaboration: Teams can collaborate effectively through Cognos, sharing insights and reports in real-time.**

**Data source**

Dataset is collected from the kaggle.com named “daily-website-visitors.csv” which has a data about the Days, Day of week, Date, page Loads, Unique visits, First-time visits, Returning Visits.

Dataset link:

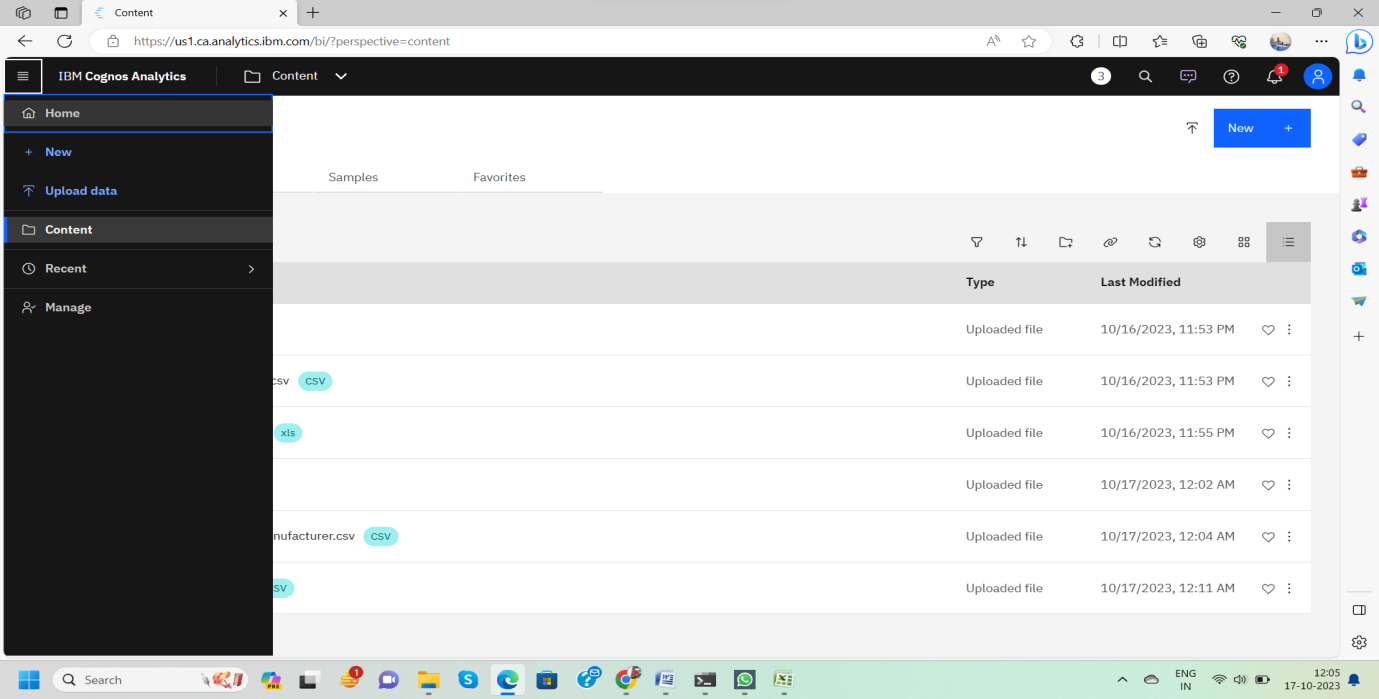
<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>

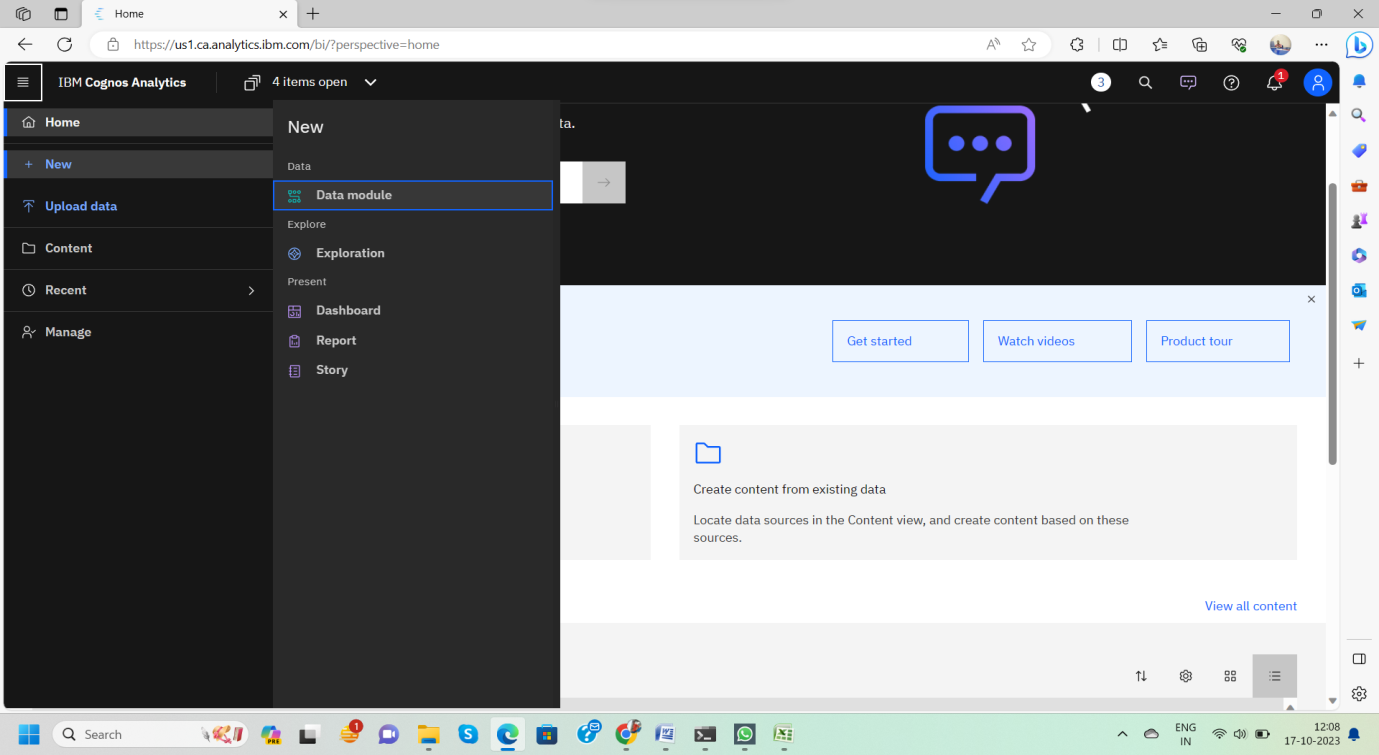
****

**Data Loading**

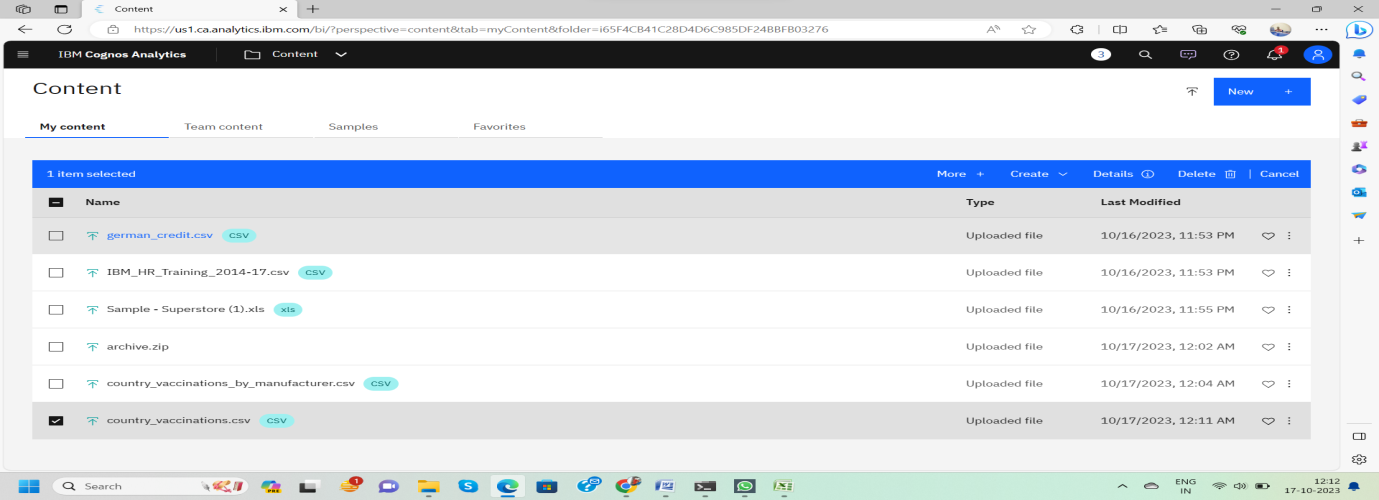
Steps Involved in data loading on IBM cognos.

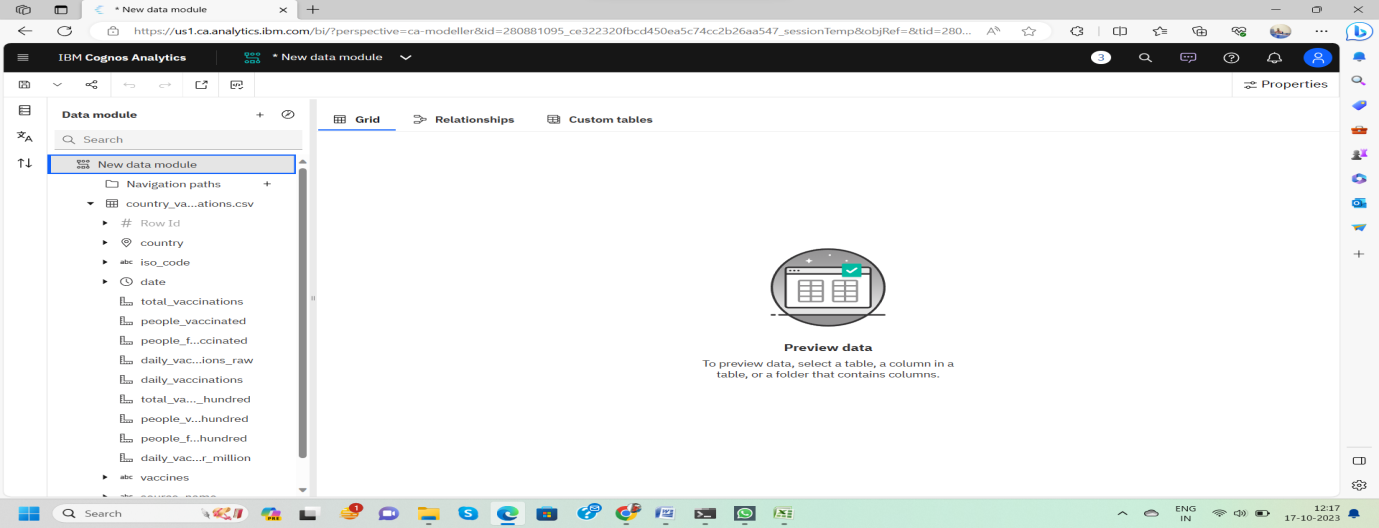
1. **Login to your IBM cognos**
2. **Click more menu from the left side**
3. **Select new tab**

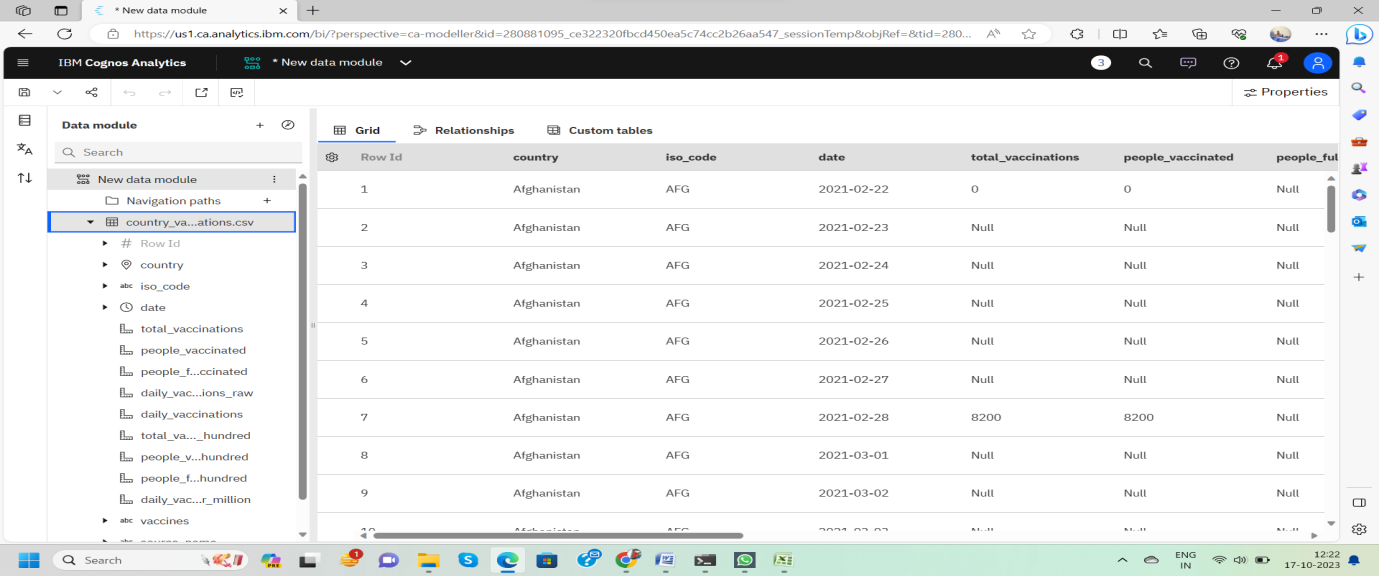
**4. Click Data module tab**

****

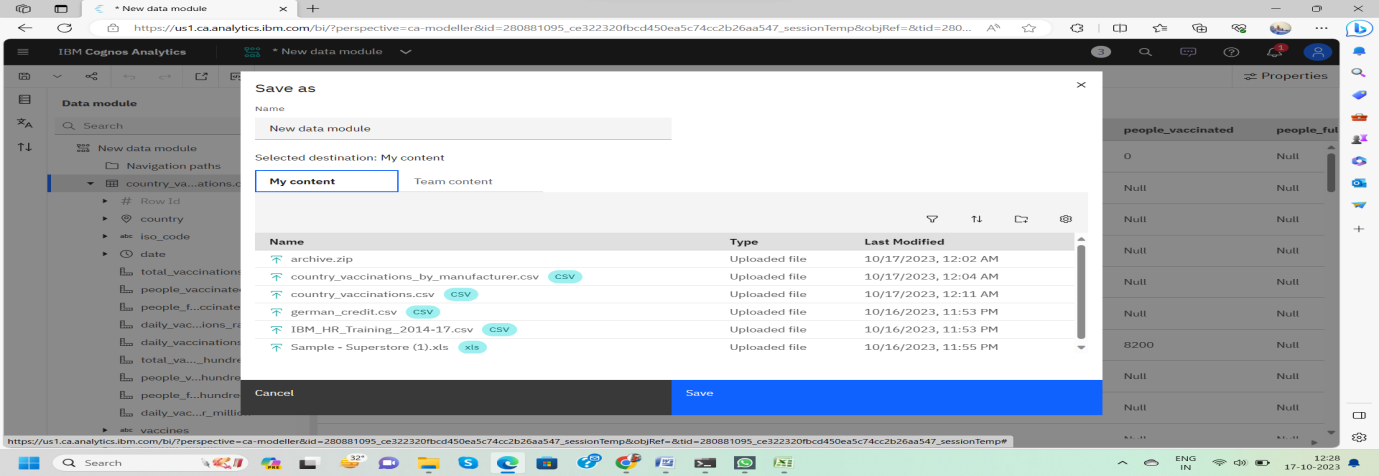
**5.Upload the dataset for your project and select the Corresponding file**

**6. preview the data**

**7.Explore the data**

****

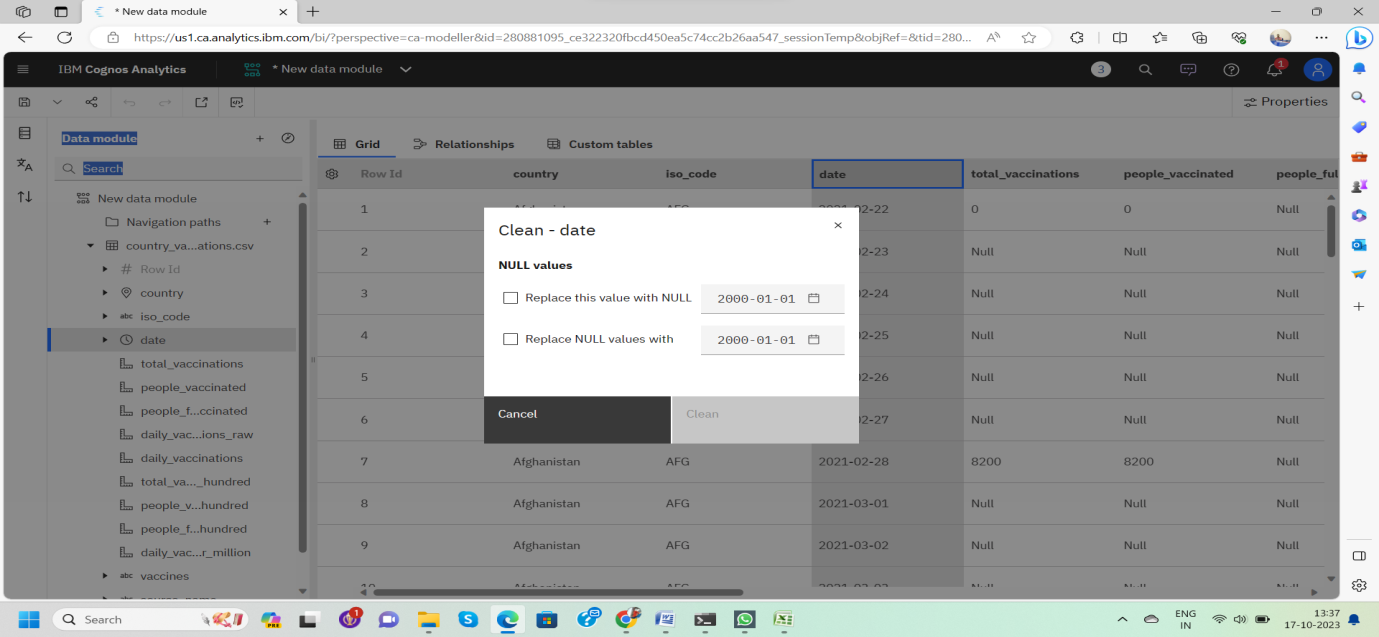
**8. save the data module**



**Data Preprocessing and Cleaning**

In this phase the following steps will taken

* Handling missing data
* Data Transformation
* Data Type Conversion
* Removing Duplicates
* Dealing OutliersOnce you saved the data module.
* Click the corresponding dataset on IBMcognos and Preview the mosuleRight
* Click the row where you want to clean the data It provides the UI to Clean the data and makes the task easy one, Now Updating and Replacing the Null values are simple

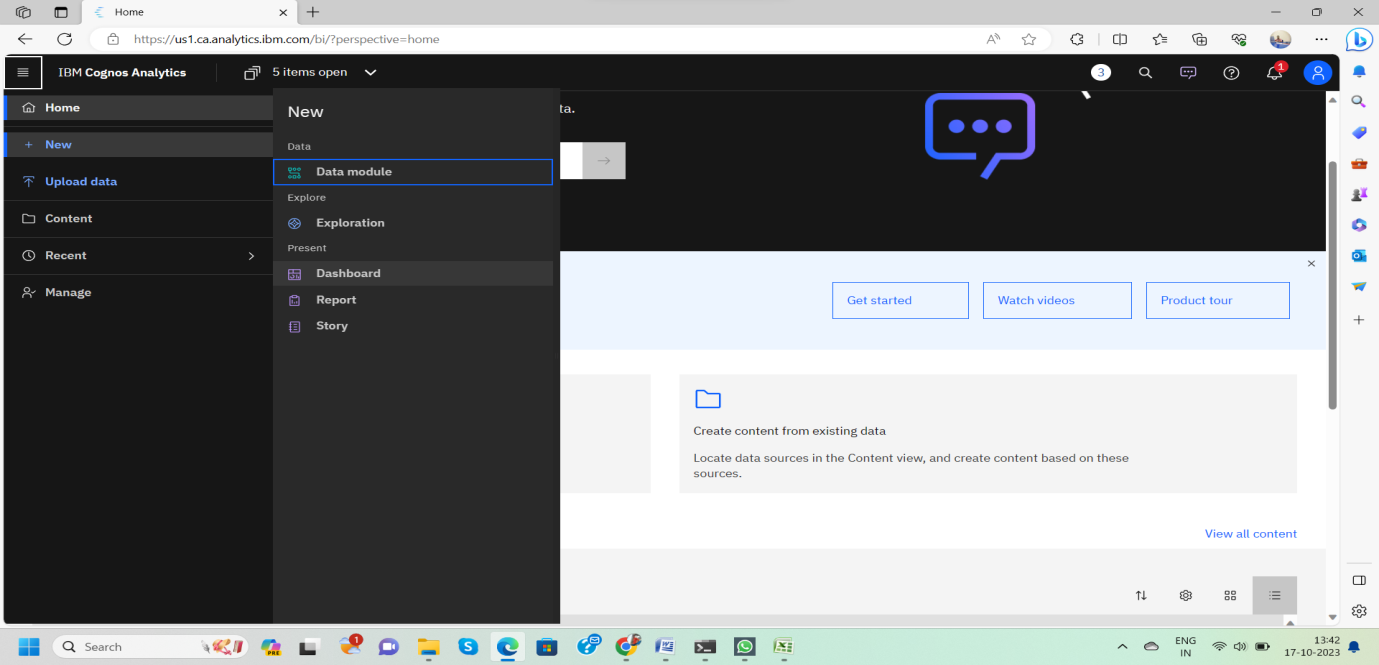
****data module will be updated by doing the above process

after the completion of process start creating the dashboard for Visualization

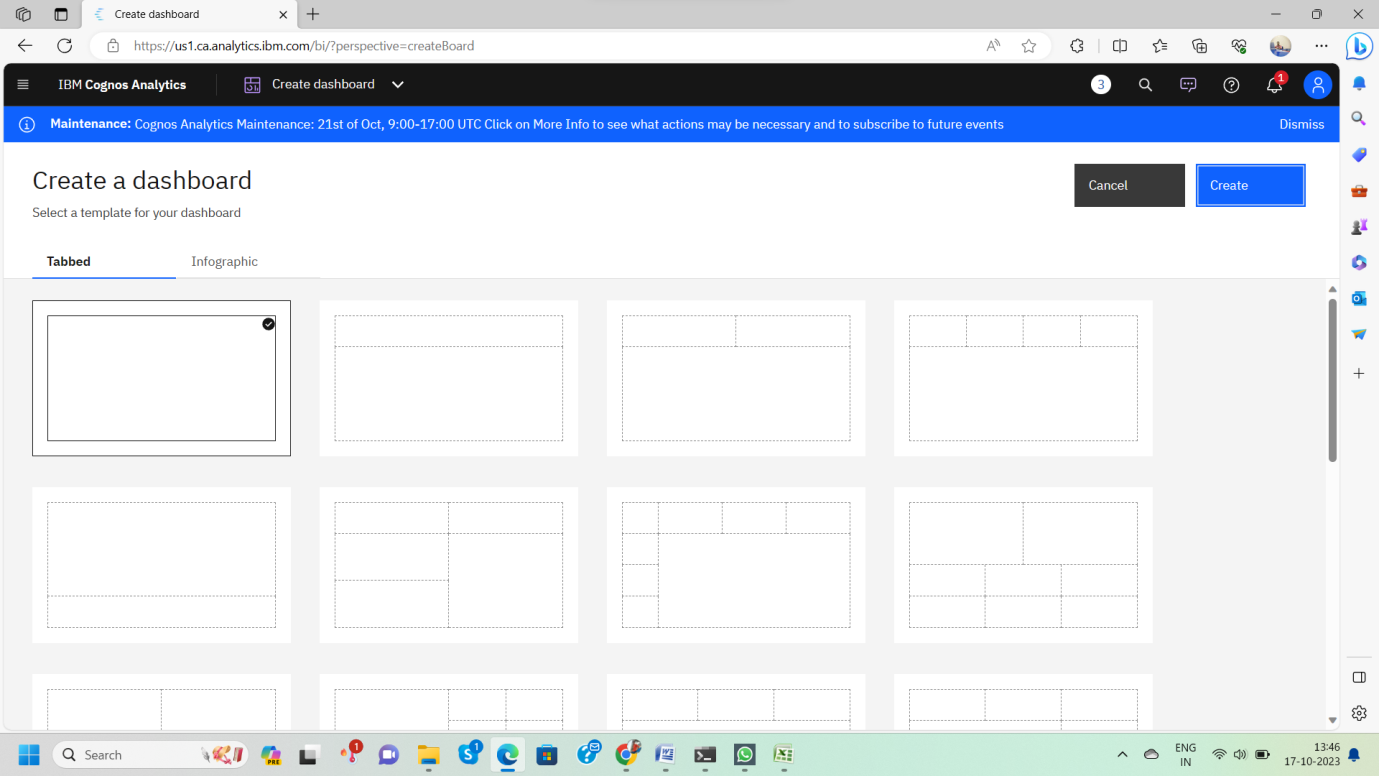
**Dashboard Creation**

**Dashboard creation are helpful to visualizing the data**

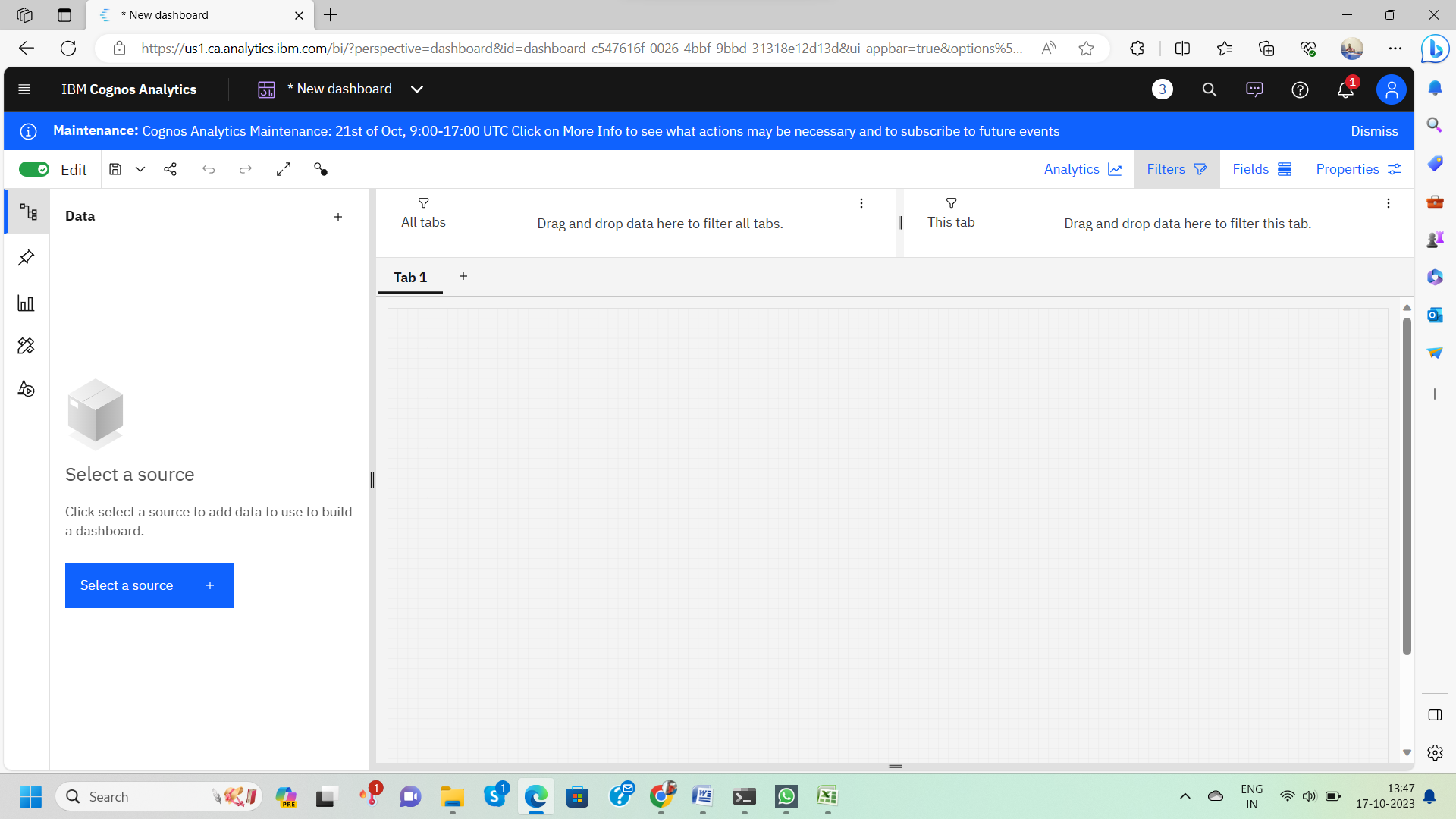
1. **Goto Home menu**
2. **Select the new tab**
3. **Click dashboard**

****

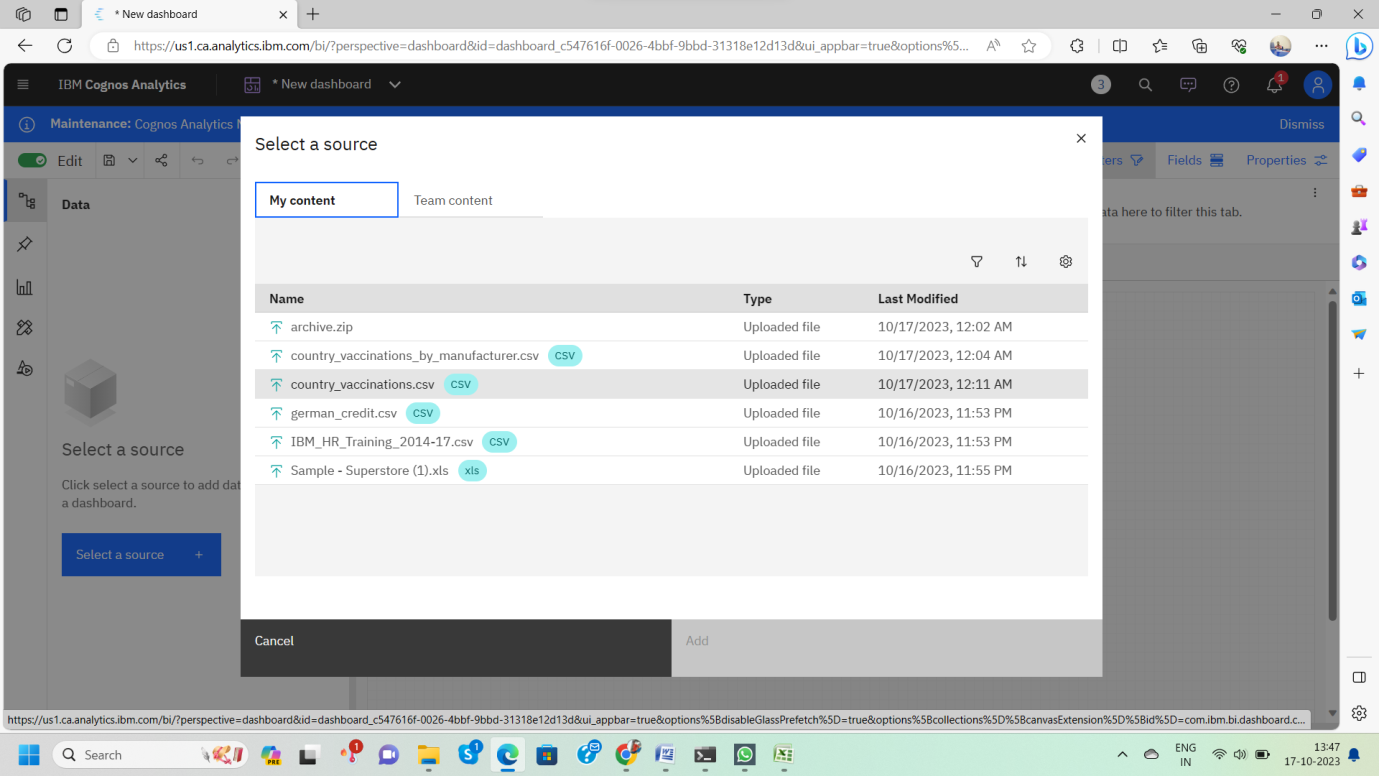
**4. Choose the template for your project and click**

****

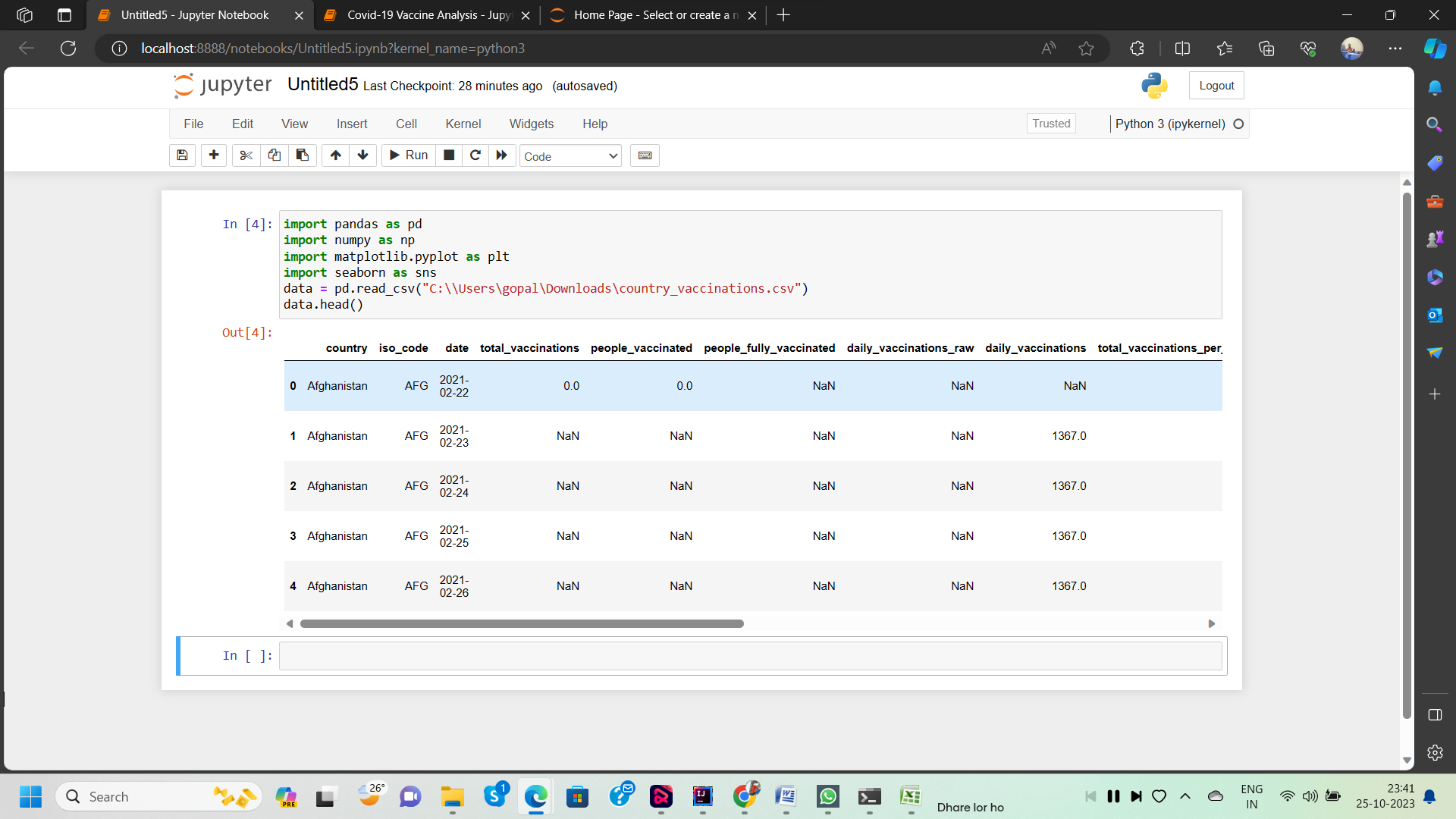
**5.Now Dashboard is created**

****

**6. Select the data source**

****

**Data Collection**

****

**Exploratory Data Analysis (EDA)**

**1. Summary Statistics:** -

Calculate basic summary statistics for your COVID-19 vaccine data.

**2. Data Distribution:-**

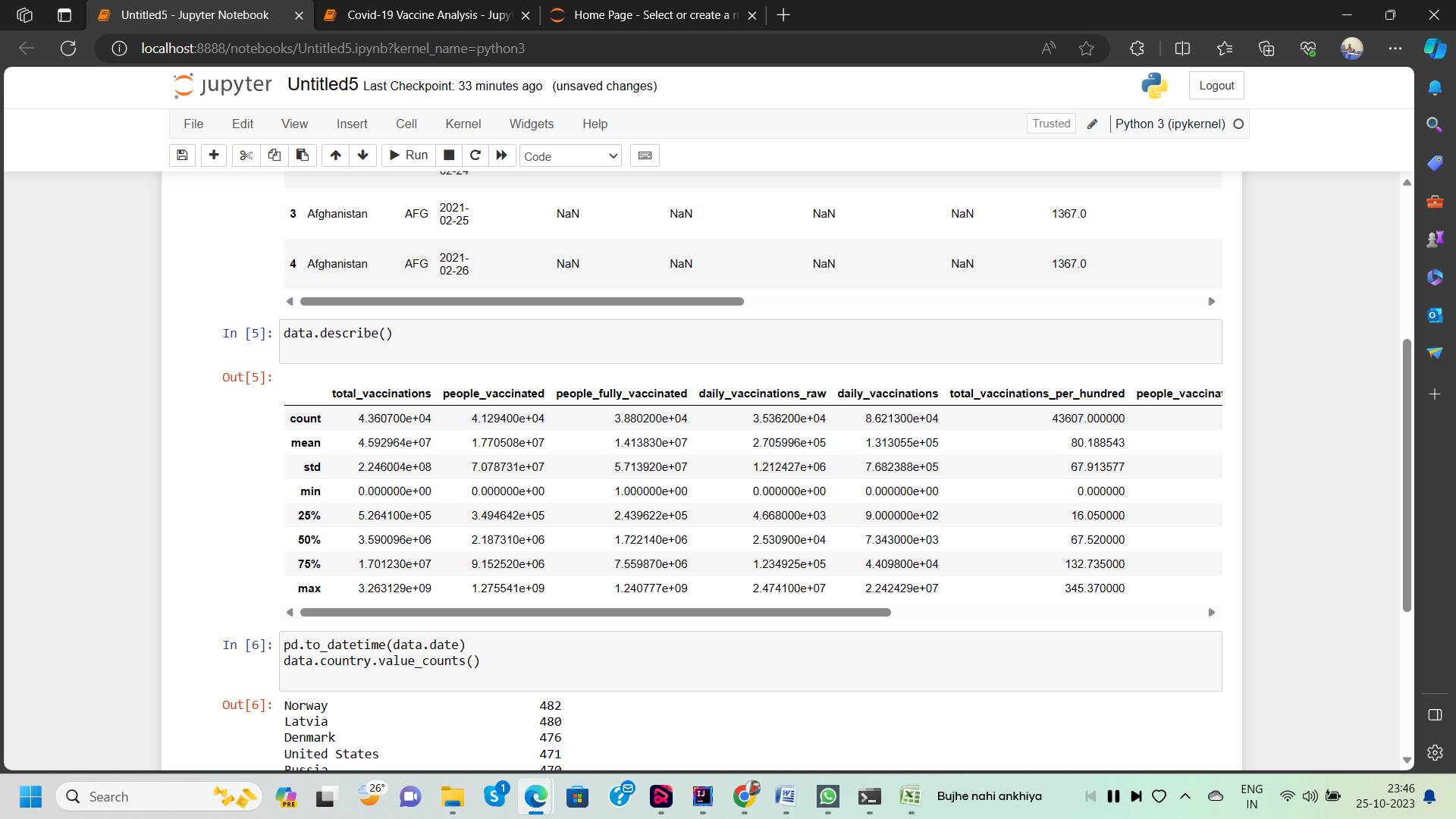
Visualize the distribution of vaccination data. Create histograms or density plots to understand how the data is distributed. Are there any outliers?

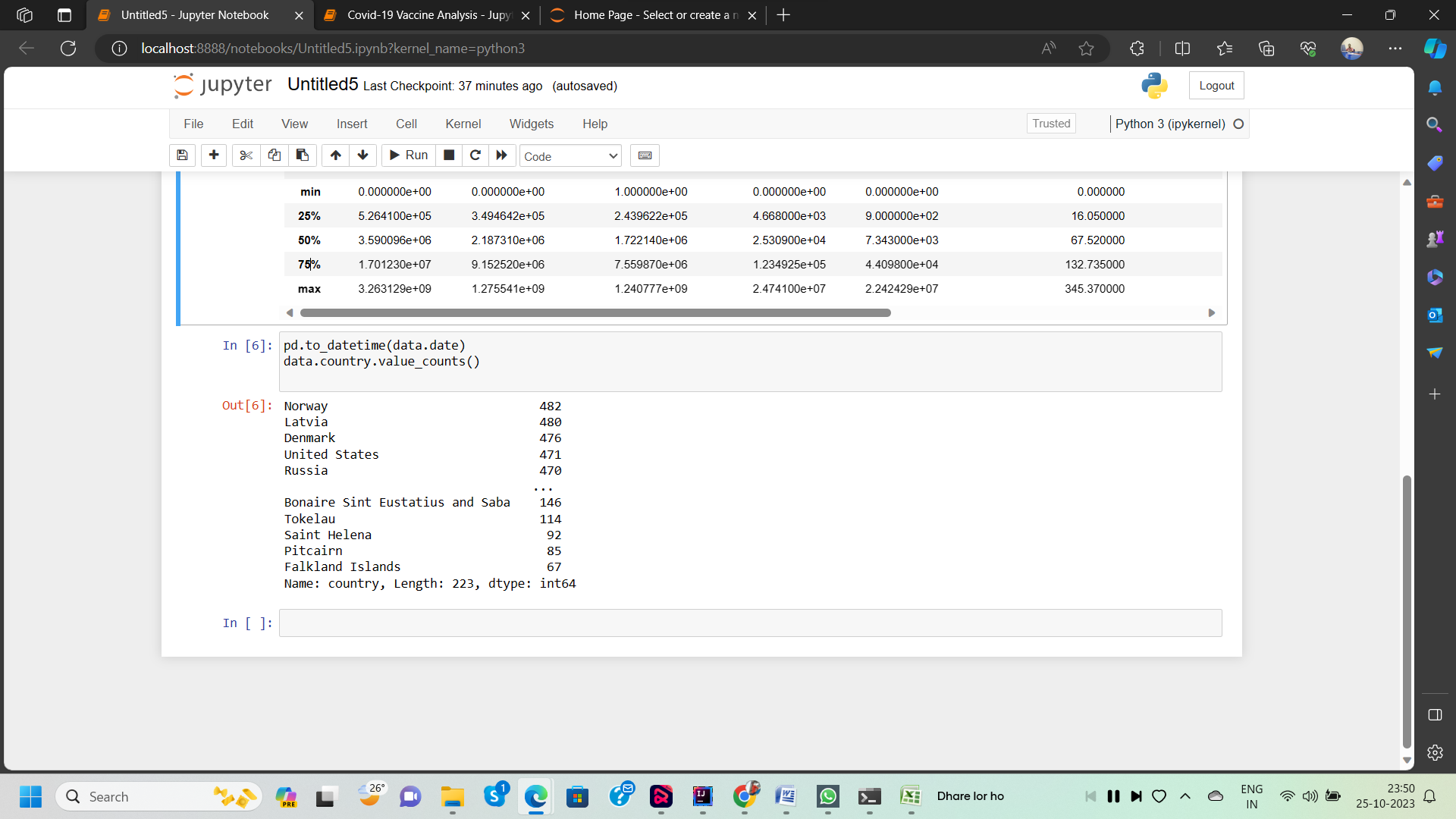
**3. Correlation Analysis:-**

Explore correlations between vaccination rates and various factors. For example, you can investigate whether vaccination rates are correlated with factors like age, gender, location, or socioeconomic factors.

**4. Time Trends:** -

If you have time-series data, examine how vaccination rates have evolved over time. Create line plots or time-series plots to observe trends.

****

****

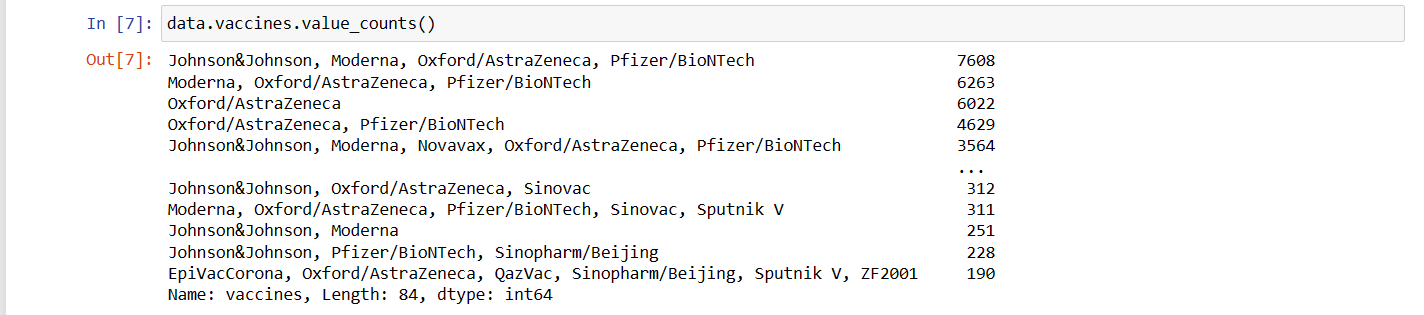
**Statistical Analysis:**

**1. Hypothesis Testing: -**

Perform hypothesis tests to determine if there are statistically significant differences in vaccination rates between different groups or regions. For example, you can use t-tests or ANOVA to compare vaccination rates by age groups or between different states.

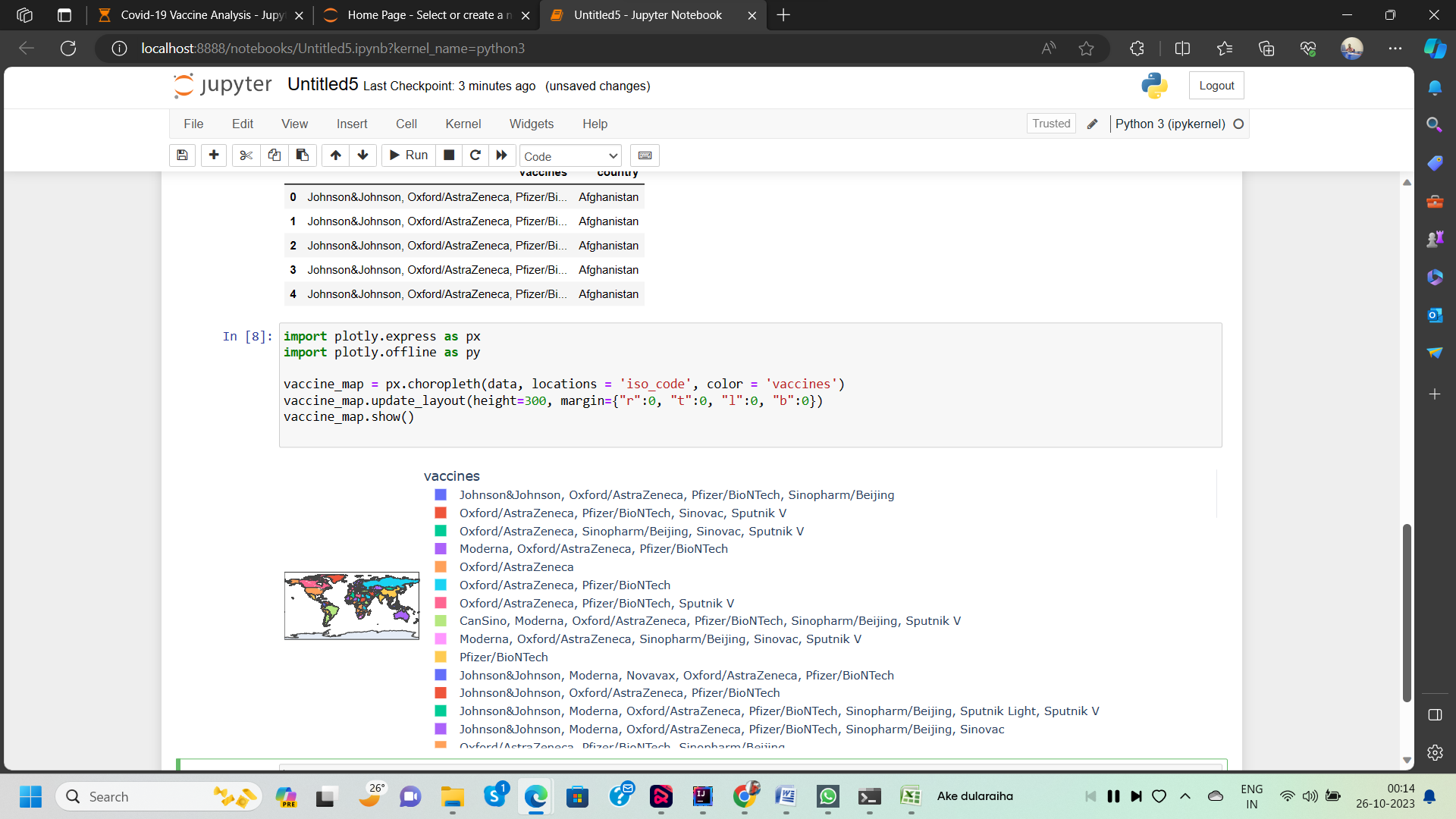
**2. Regression Analysis:-**

Perform regression analysis to model the factors that influence vaccination rates. Multiple linear regression or logistic regression can help you understand which variables have the most significant impact on vaccination rates.

****

**Stacked Area Charts:**

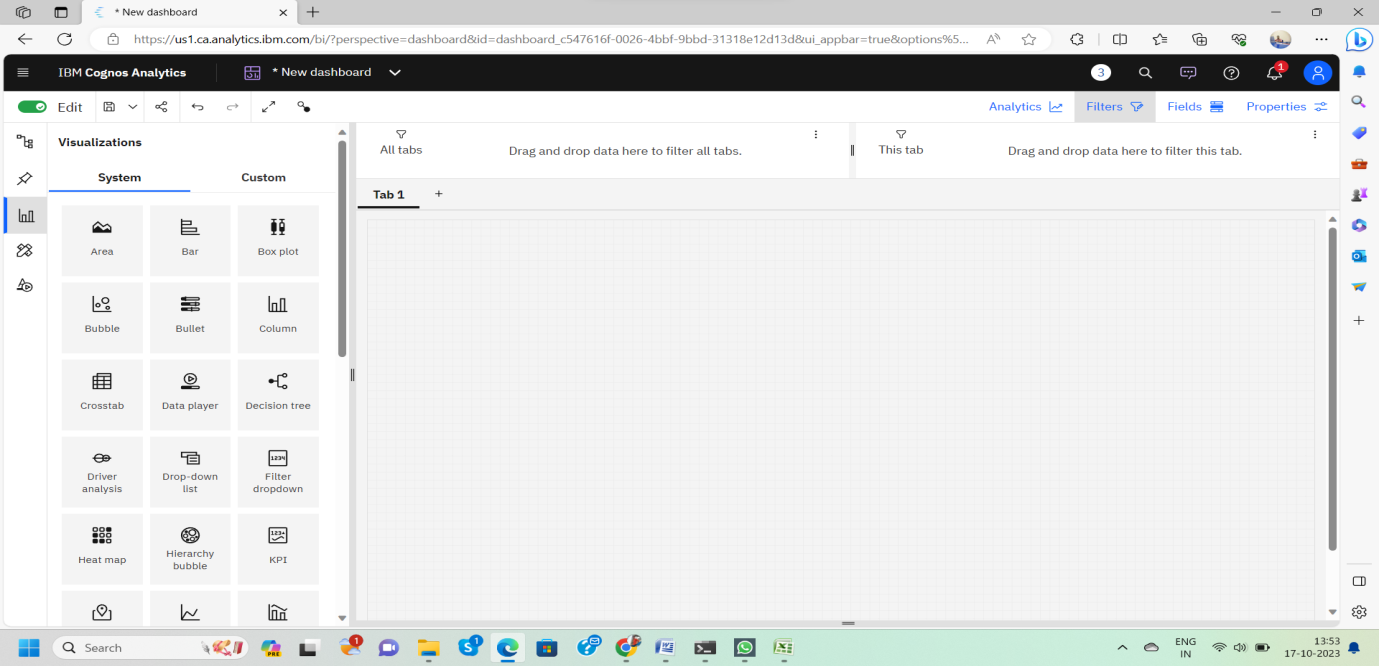
If you have data on multiple doses or vaccines, use stacked area charts to show the cumulative effect of different vaccines and doses over time.

****

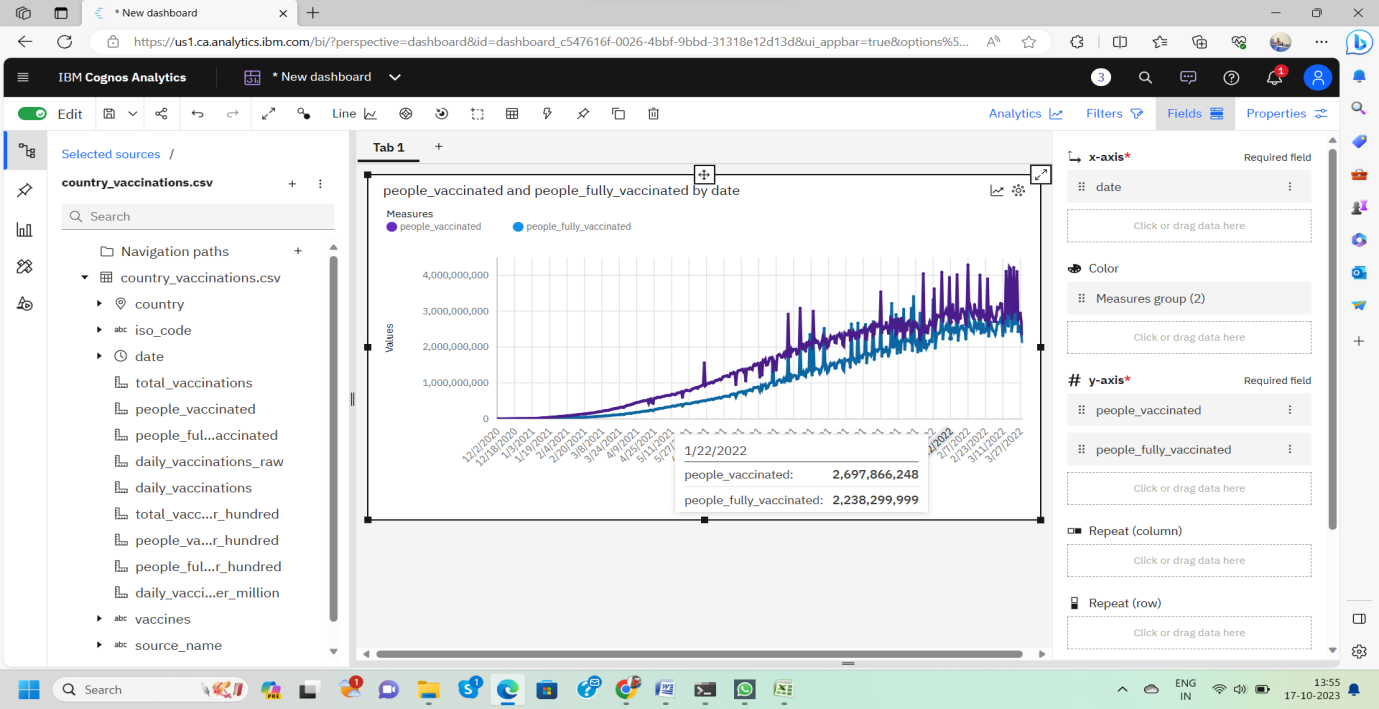
**Visualization**

**1. Goes to the Corresponding Dashboard**

**2. select the visualizations tab in the left side of title bar**

****

**3.Choose the system as you want and put the data source for the required columns**

****In the above screen shot displays the Line graph and model compares the “**people \_vaccinated**” and “**people \_fully \_vaccinated**” from the time period of 2020 to 2022

X-axis =**Dates**

Y-axis = **people \_vaccinated, people \_fully \_vaccinated**

After performing these activities a comprehensive document will be created to demonstrate the ability to Communicate and share finding.

**Conclusion**

In this initial phase of our COVID-19 vaccine analysis project, we successfully collected and preprocessed the vaccine data.

Summarize your findings, including any significant correlations or differences identified in the analysis.

Provide recommendations based on your analysis. For example, you might recommend prioritizing vaccination efforts in countries with low vaccination rates to reduce infection rates and mortality.

Highlight any limitations in the data or analysis, such as data quality issues or assumptions made during statistical testing.