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

**Phase 4**

**Development Part 2**

**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**

**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.**

**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.

**Problem Statement**

In this part you will continue building your project.

Continue conducting the Covid-19 vaccines analysis by performing:

Exploratory data analysis

Statistical analysis

Visualization

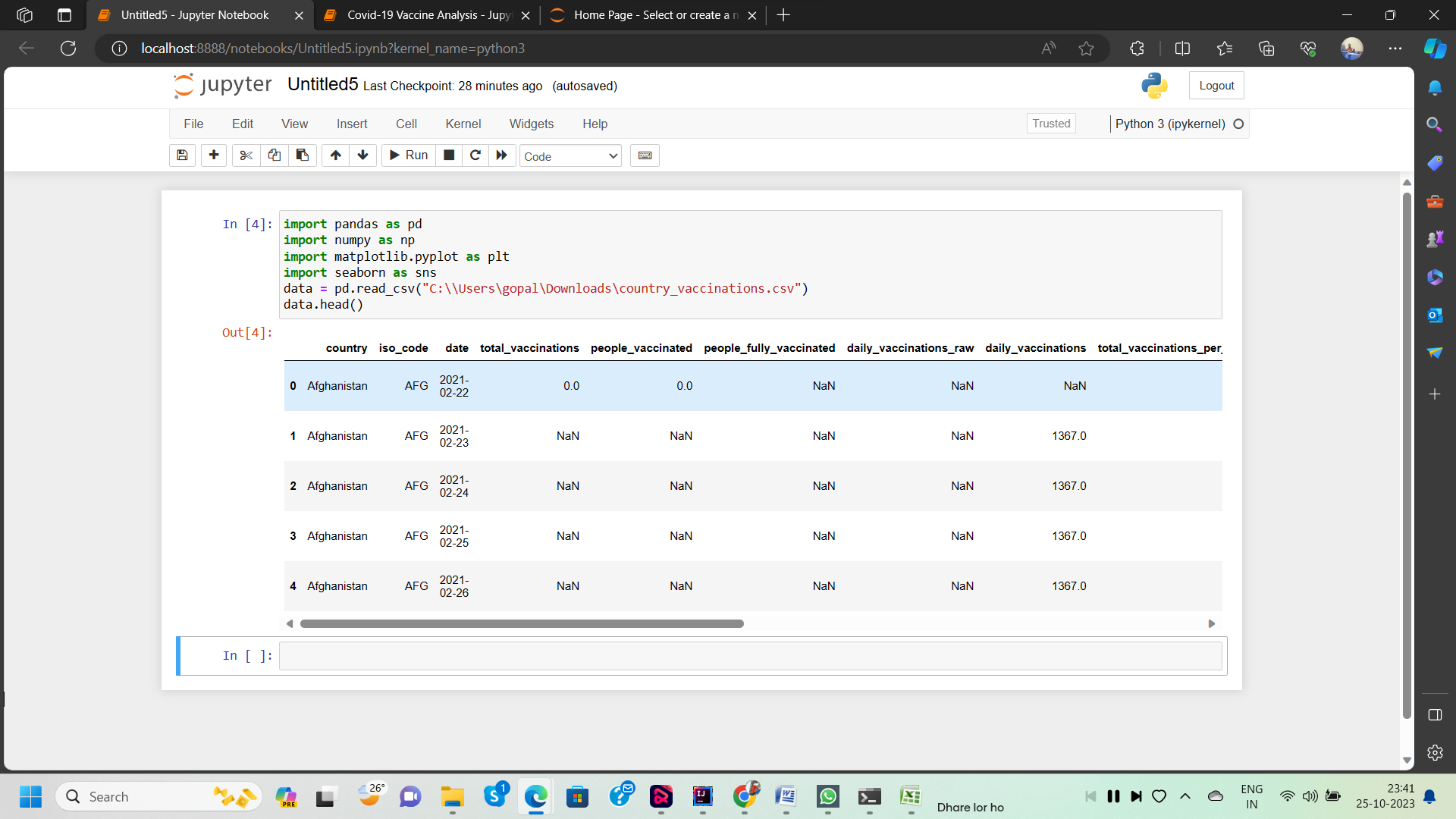
**Data source**

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 Collection**

****

**Exploratory Data Analysis (EDA)**

**1. Summary Statistics:** -

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

**2. Data Distribution:-**

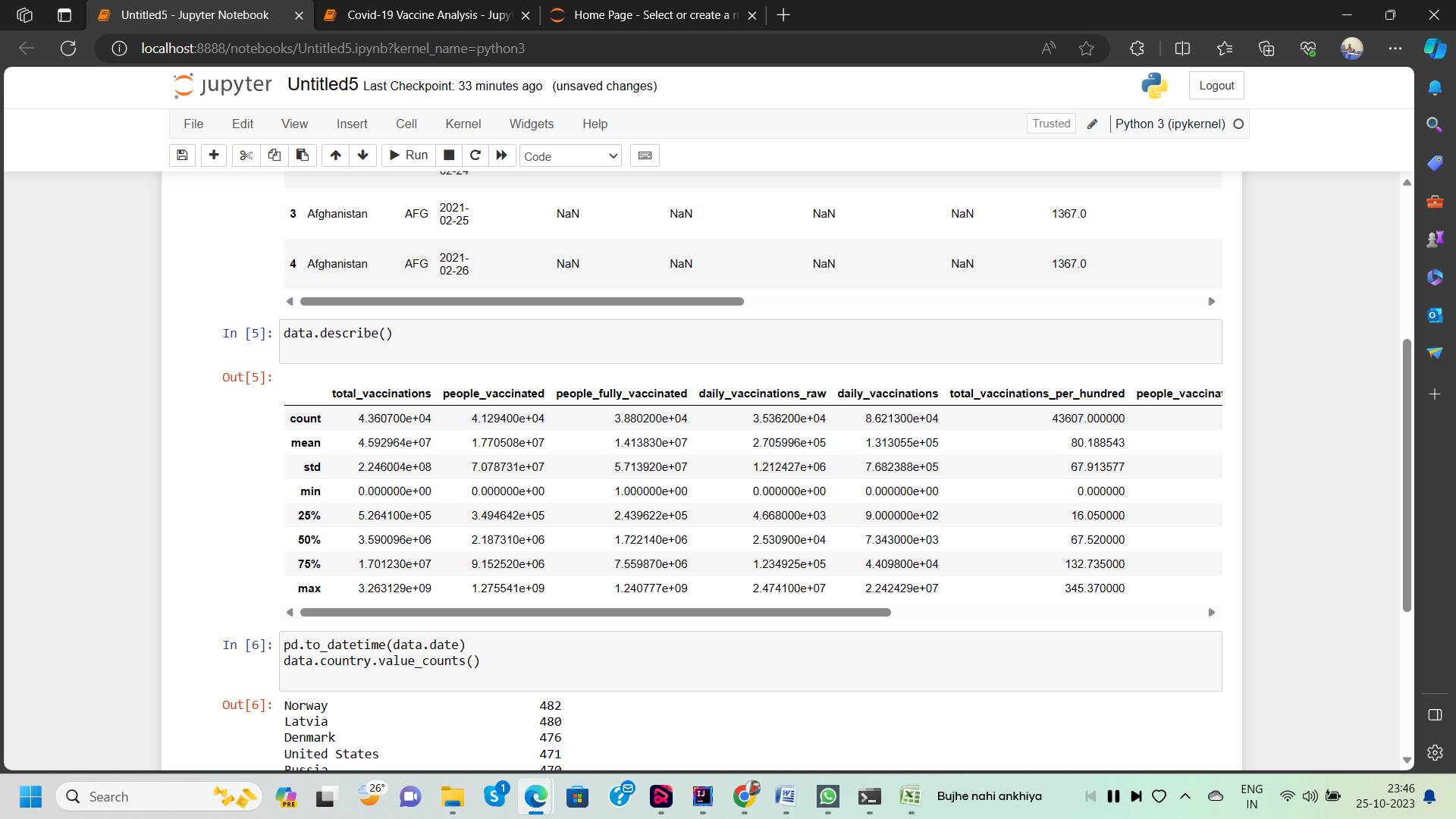
Visualize the distribution of vaccination data. Create histograms or density plots to understand how the data is distributed. Are there any outliers?

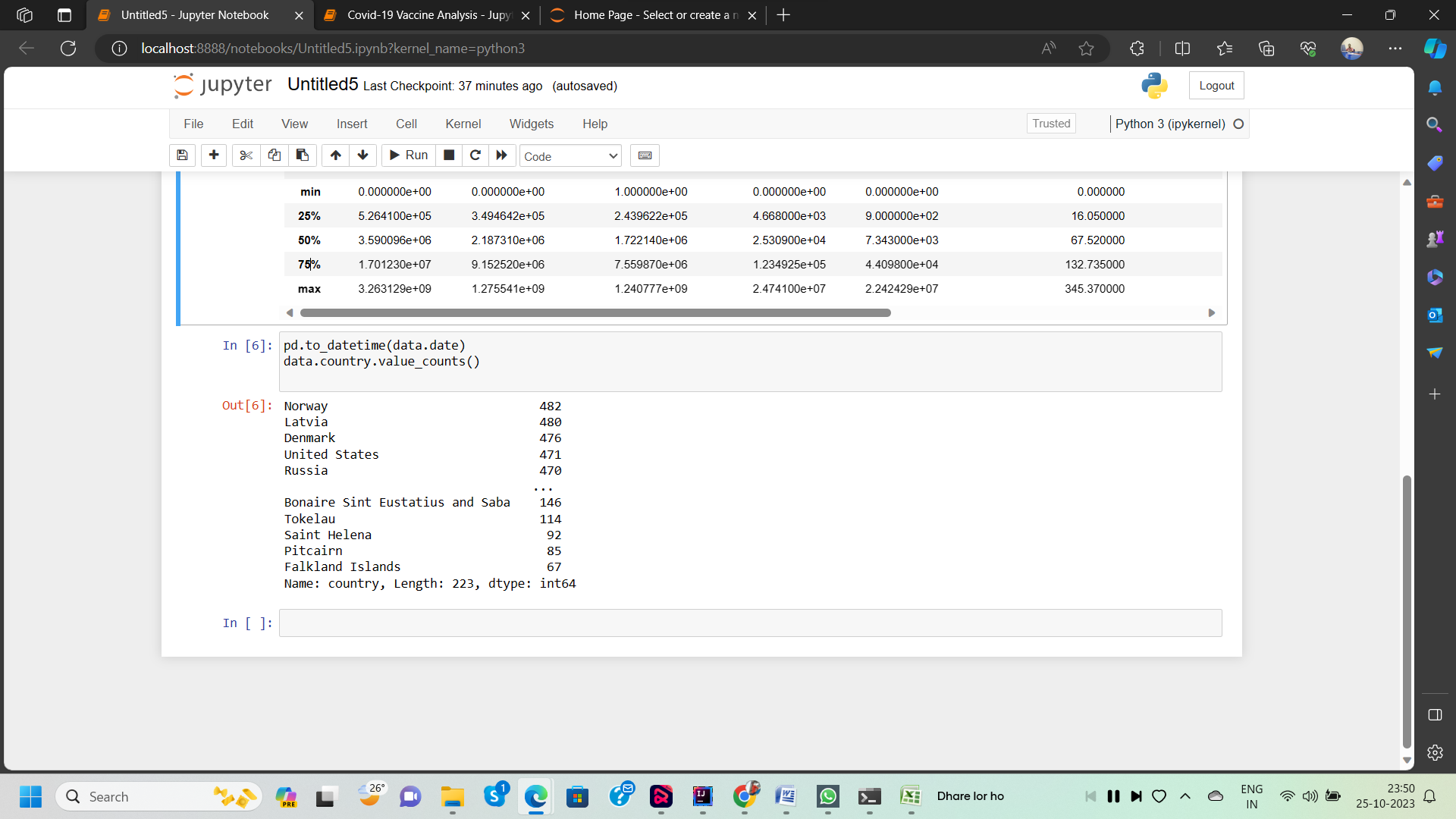
**3. Correlation Analysis:-**

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.

****

****

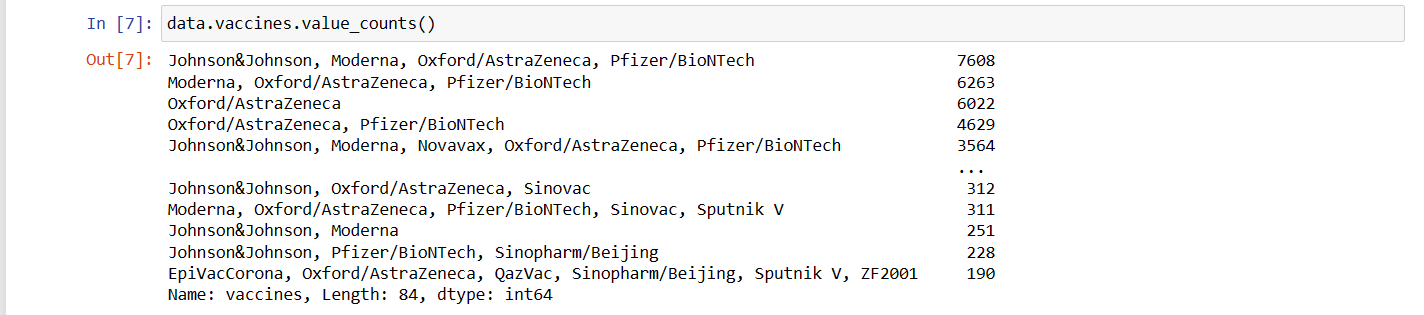
**Statistical Analysis:**

**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.

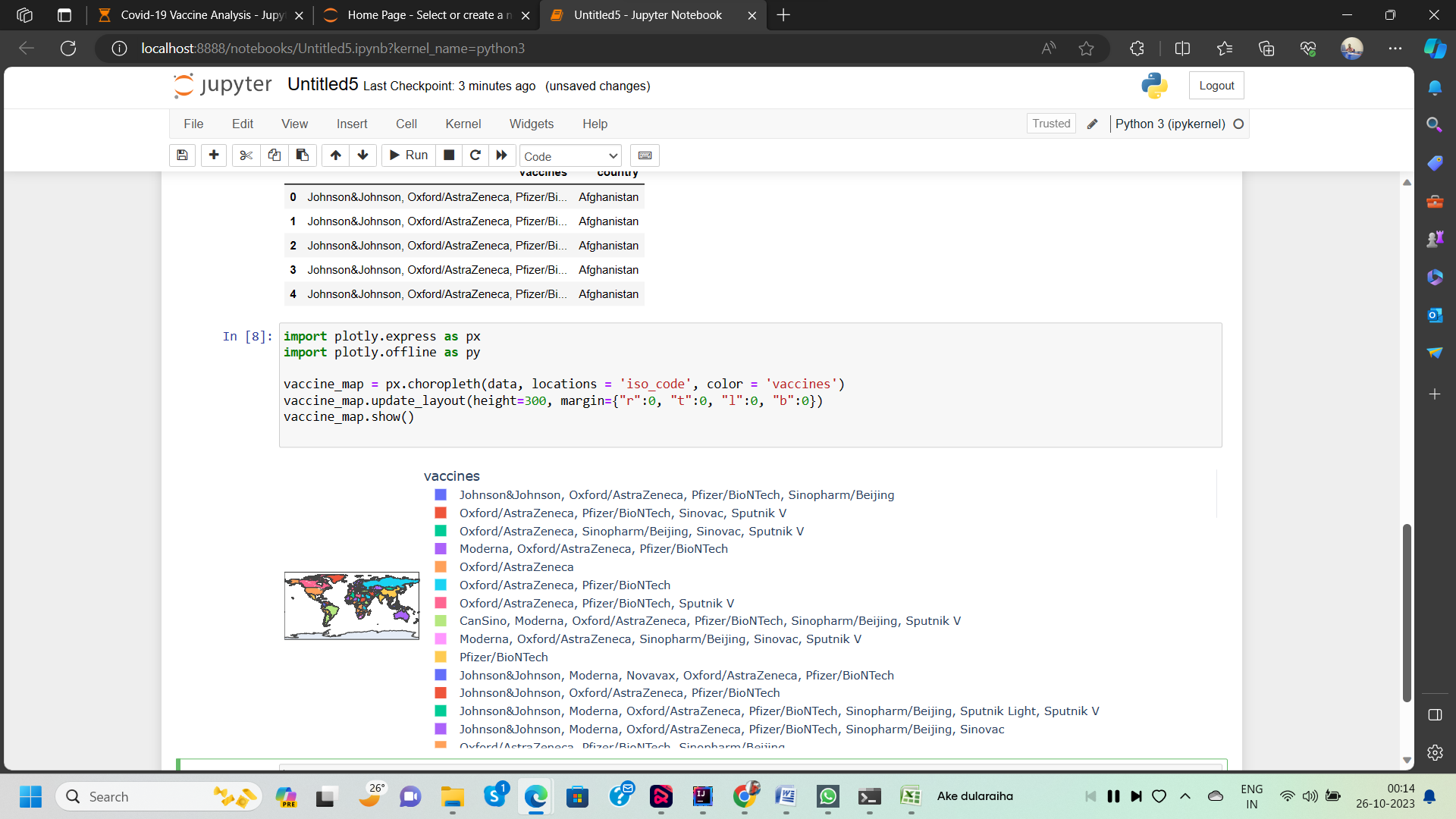
**2. Regression Analysis:-**

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.

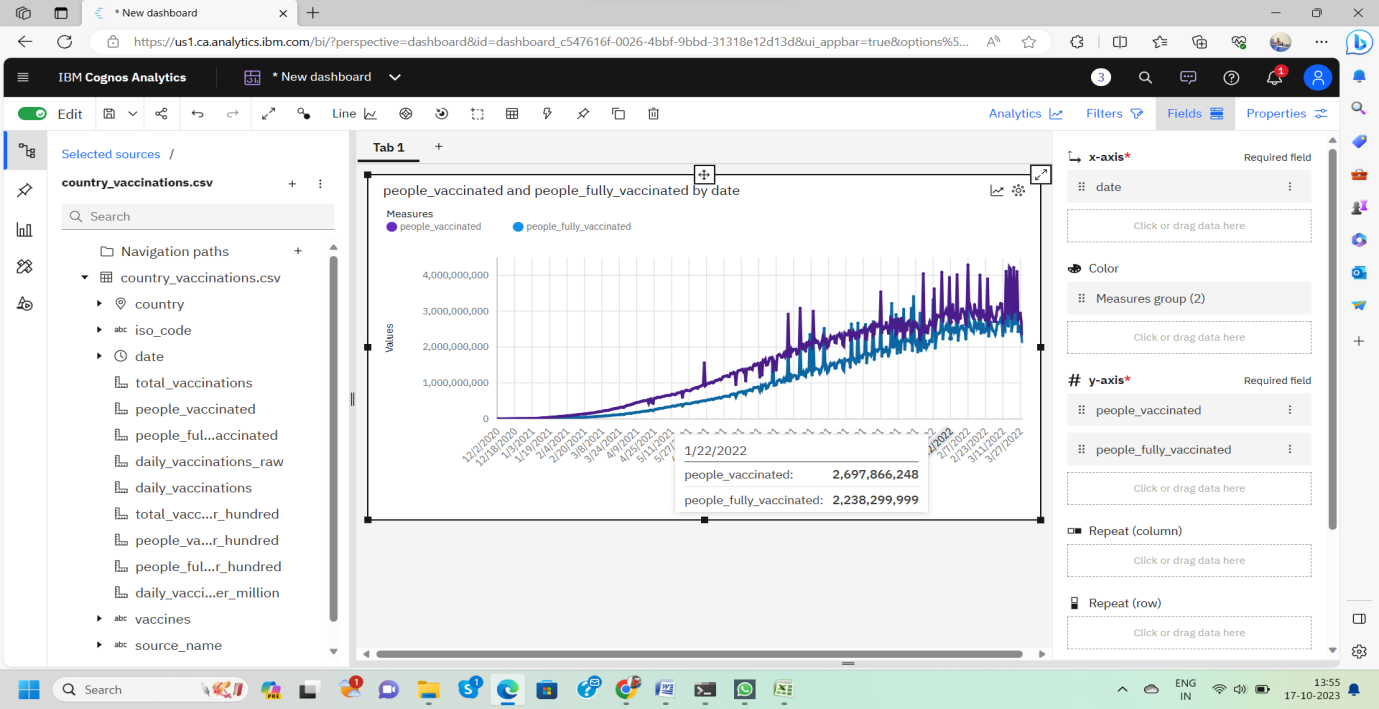
****

**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**

****

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**

**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.