

| **TITLE :** "COVID-19 VACCINE DATA ANALYSIS"  **SUBTITLE :** "LOADING AND PREPROCESSIN ON VACCINE”  **NAME :** Arshin K H  **REG .NO :** 810721243008 |
| --- |

**INTRODUCTION :**

A dataset related to COVID-19 vaccines can be a complex task, and it involves several steps including data loading, preprocessing, and exploratory data analysis (EDA). Given the complexity and the length of the process, it's not possible to cover it in detail within a single response.

**LOADING AND PREPROCESSING :**

**Step 1: Data Collection**

1.1. Identify reliable sources for COVID-19 vaccine data.

1.2. Obtain permission to use the data if required.

1.3. Download the dataset in a suitable format in (CSV.).

**Step 2: Data Loading**

2.1. Use a suitable programming language (Python) and data manipulation libraries.

2.2. Load the dataset into your chosen environment.

2.3. Verify that the data has been loaded correctly.

**Step 3: Initial Data Exploration**

3.1. View the first few rows of the dataset to understand its structure.

3.2. Check for missing values and handle them if,

3.3. Check for duplicate rows and remove them if present.

**Step 4: Data Preprocessing**

4.1. Handle categorical variables (encoding, one-hot encoding, etc.).

4.2. Normalize or standardize numerical features if needed.

4.3. Perform feature engineering if necessary (creating new features from )

4.4. Handle time-series data if applicable (e.g., date conversion, lag features).

**Step 5: Exploratory Data Analysis (EDA)**

5.1. Visualize the dataset to gain insights (histograms, box plots, scatter plots, etc.).

5.2. Identify relationships between variables.

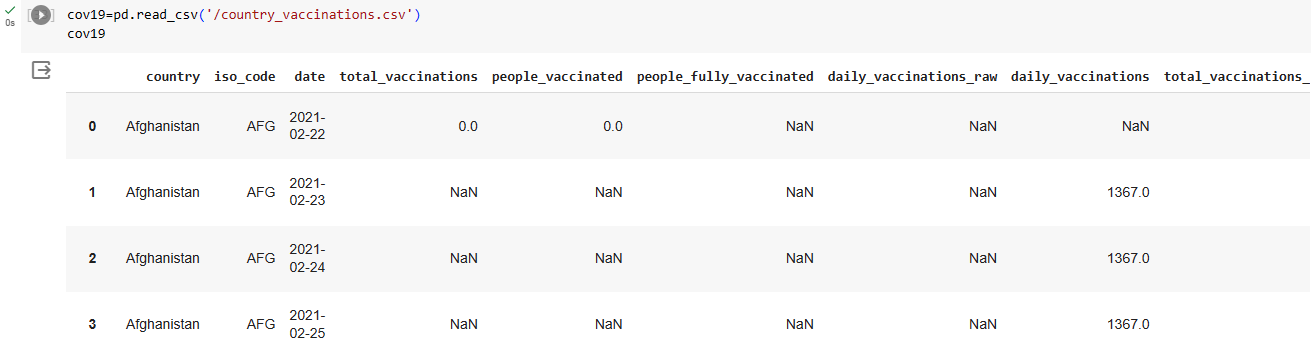
5.3. Identify outliers and decide how to handle them.

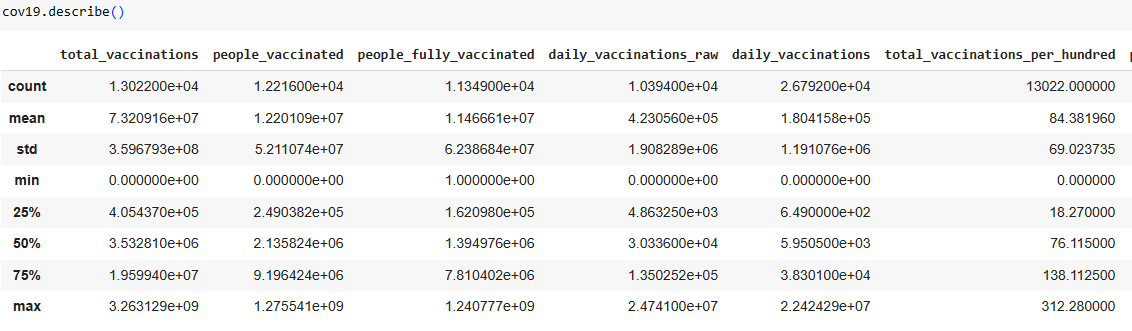
**Step 6: Save Preprocess Data**

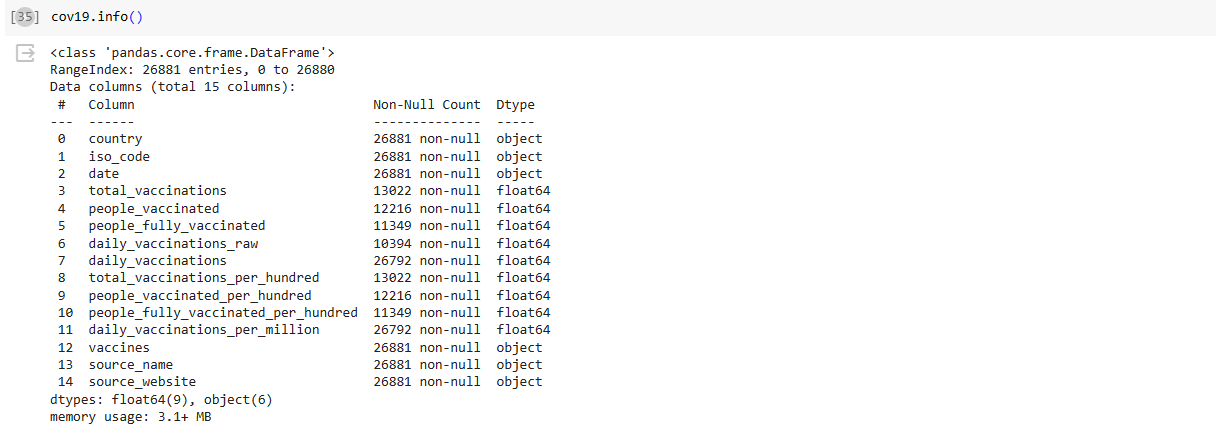
Save the data when preprocess in a suitable format for further analysis CSV.

**Importing Libraries:**

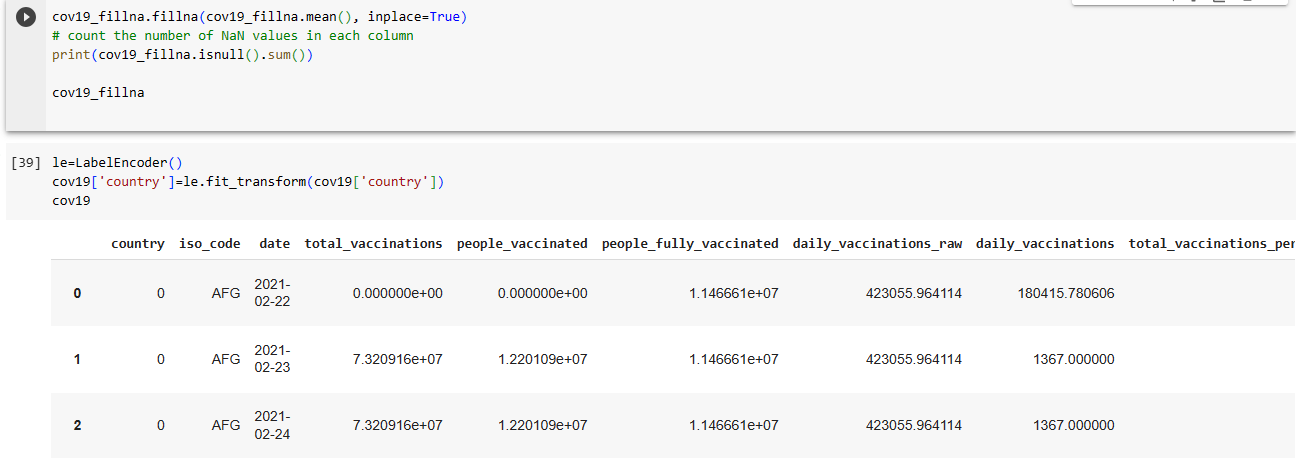


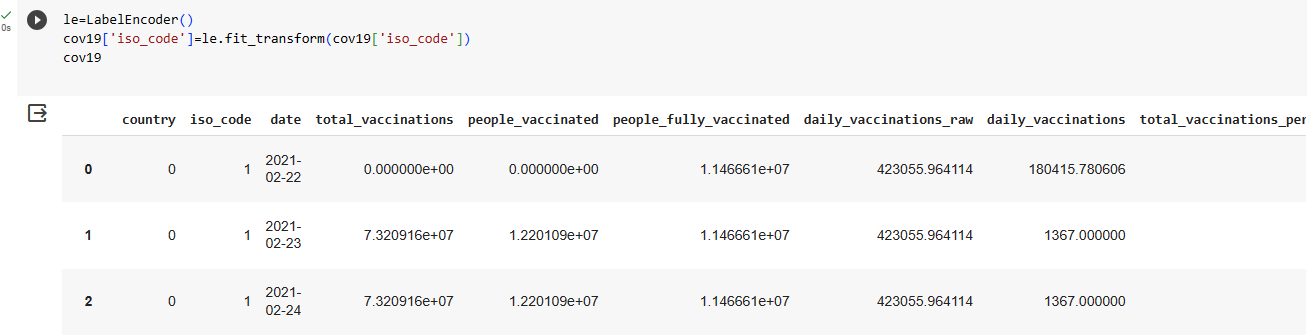
**Loading dataset:**

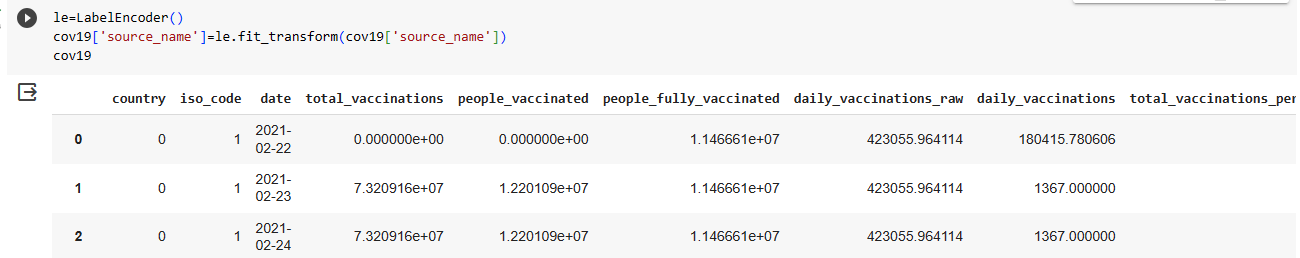


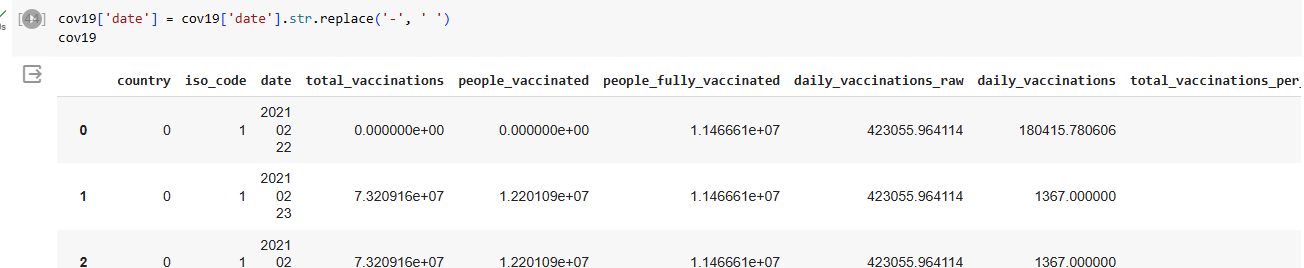


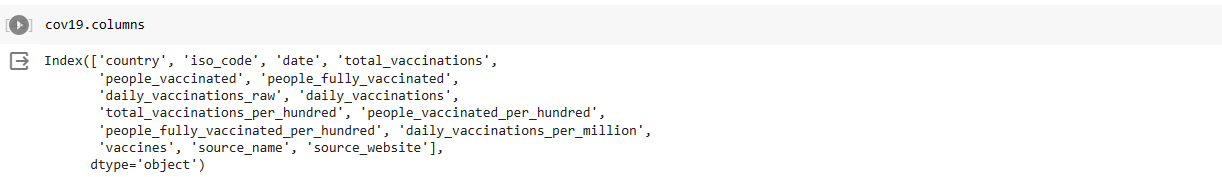
**Preprocessing:**

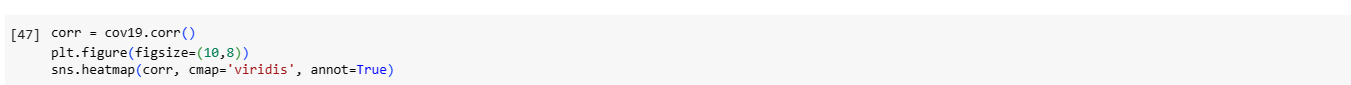


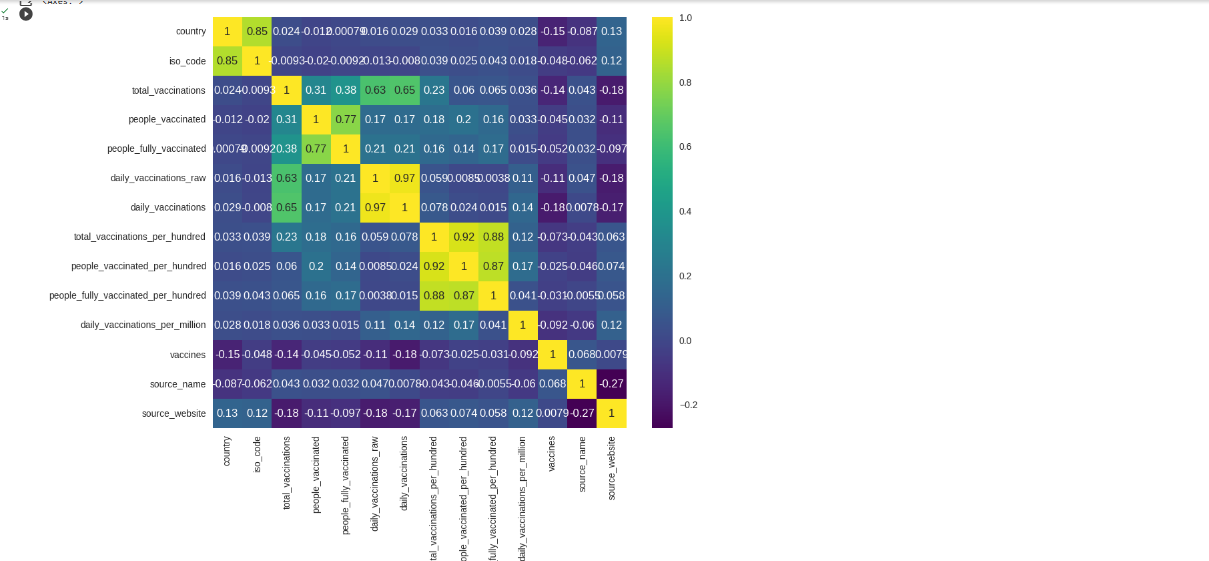




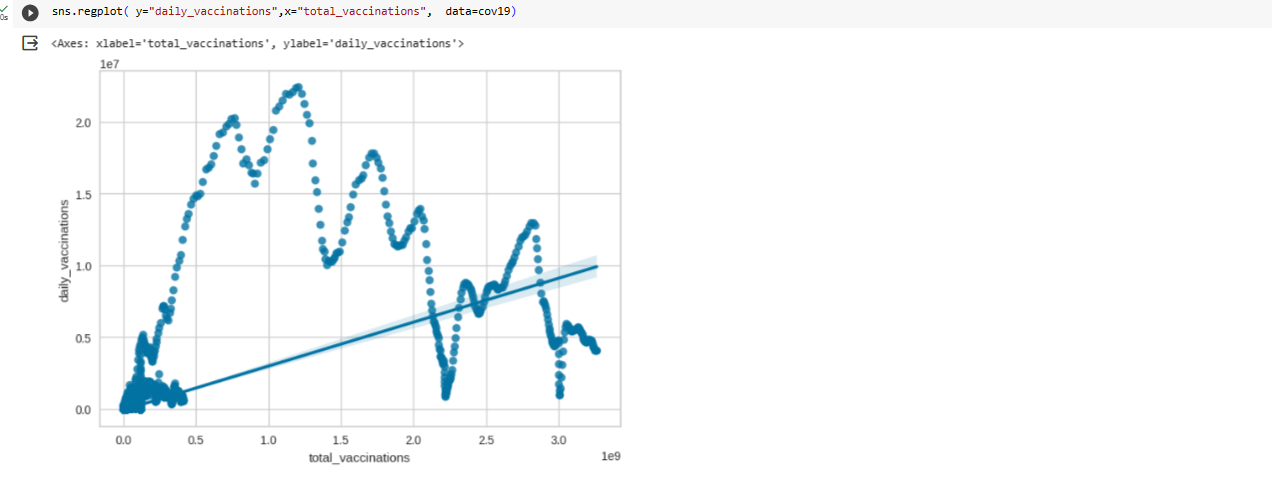


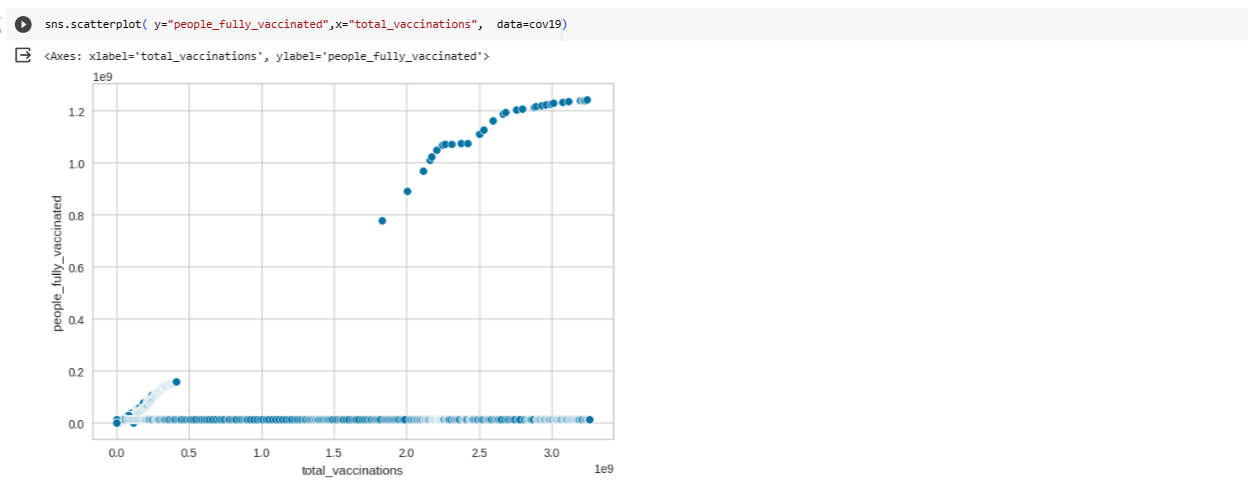


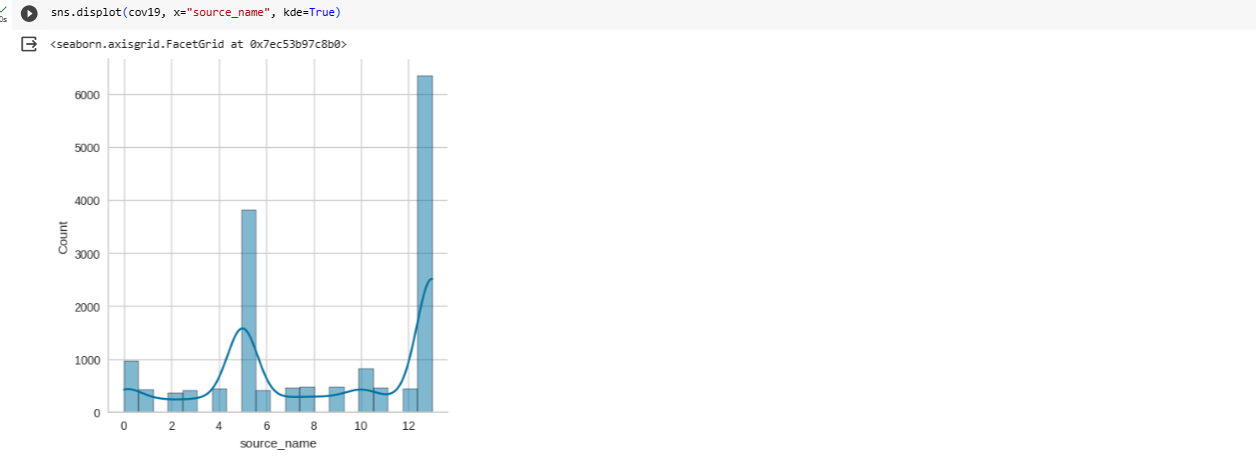


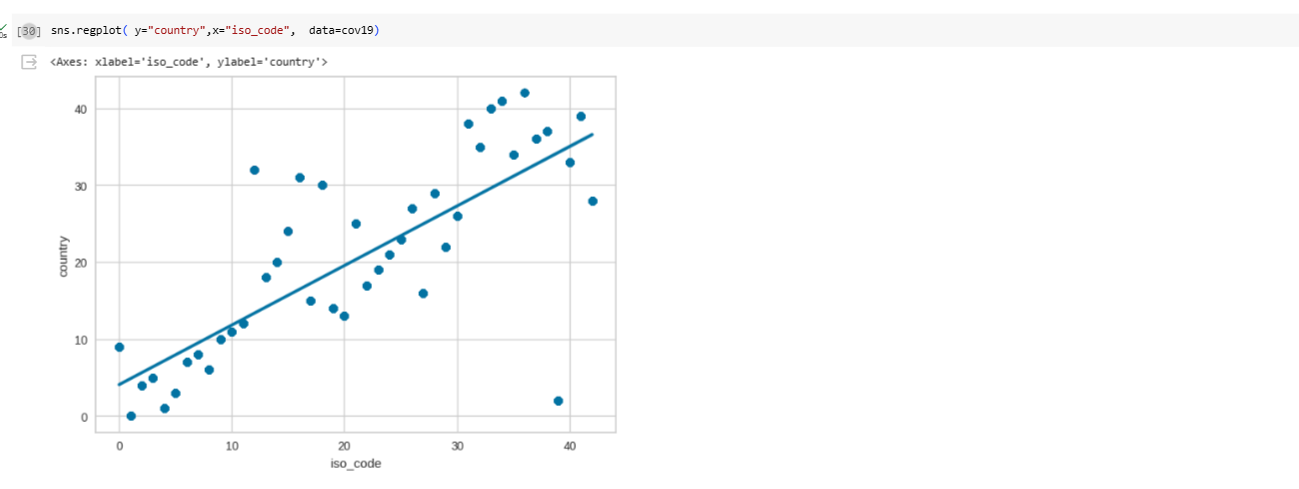


**Exploratory Data Analysis (EDA):**









**Dataset Link:**[**https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress**](https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress)

**CONCLUSION :**

Ensure that the text is stored in a suitable data structure (e.g., a string variable) for further processing .Remove any special characters, punctuation, or symbols that may not contribute to the analysis .Apply techniques like lowercasing, Stopword removal, and lemmatization or stemming to simplify and standardize the text.