Project: Covid-19 Vaccine Analysis

Phase3: Development (part 2)

Question:

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 set and its details:

The dataset "COVID-19 World Vaccination Progress" on Kaggle is a collection of data related to the COVID-19 vaccination efforts worldwide. It provides information about the progress of COVID-19 vaccinations in various countries and regions. This dataset is designed to help researchers, data scientists, and analysts understand and analyze the progress of COVID-19 vaccination campaigns across different countries. A second file, with manufacturers information, is included. Below is a detailed overview of the dataset:

Title: COVID-19 World Vaccination Progress

Dataset ID: gpreda/covid-world-vaccination-progress

Source: The dataset was created by a Kaggle user named Gabriel Preda, collected from various sources, including government health agencies, international organizations, and research institutions.

Description:

1. The dataset provides information about the COVID-19 vaccination progress from various countries around the world.
2. It includes data on vaccine distribution, vaccination coverage, and other related statistics.
3. The dataset may include information about the types of vaccines used, vaccination rates over time, and population demographics.

Columns/Attributes:

1. The dataset typically contains columns such as country, iso_code, date, total_vaccinations, people_vaccinated, people_fully_vaccinated, daily_vaccinations_raw, daily_vaccinations, and more.

2.These columns provide information about the total number of vaccinations, daily vaccination rates, and other vaccination-related metrics for each country.

Usage:

1. Analyzing vaccination progress over time for different countries.
2. Identifying countries with high vaccination rates or disparities.
3. Forecasting future vaccination trends.
4. Studying the impact of different vaccines on vaccination rates.

5. Correlating vaccination progress with COVID-19 infection and mortality rates.

Data Format:

The data is usually structured as a CSV (Comma-Separated Values) file, with rows representing different countries or regions and columns representing various attributes related to vaccination progress and population.

Updates:

The dataset may be updated regularly to reflect the latest vaccination data, making it useful for tracking changes and trends over time.

Columns:

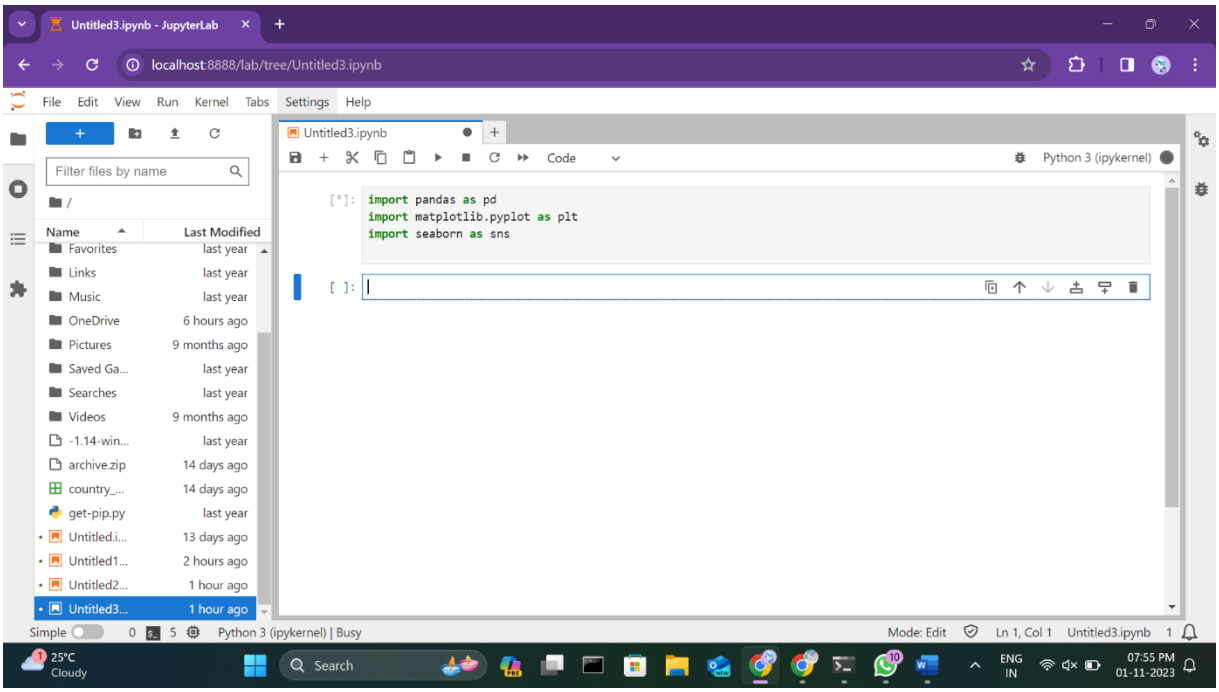
- Country- this is the country for which the vaccination information is provided.
- Country ISO Code - ISO code for the country.
- Date - date for the data entry; for some of the dates we have only the daily vaccinations, for others, only the (cumulative) total.
- Total number of vaccinations - this is the absolute number of total immunizations in the country. Total number of people vaccinated - a person, depending on the immunization scheme, will receive one or more (typically 2) vaccines; at a certain moment, the number of vaccinations might be larger than the number of people.
- Total number of people fully vaccinated - this is the number of people that received the entire set of immunization according to the immunization scheme (typically 2); at a certain moment in time, there might be a certain number of people that received one vaccine and another number (smaller) of people that received all vaccines in the scheme.
- Daily vaccinations (raw) - for a certain data entry, the number of vaccinations for that date/country.
- Daily vaccinations - for a certain data entry, the number of vaccinations for that date/country.
- Total vaccinations per hundred - ratio (in percent) between vaccination number and total population up to the date in the country.
- Total number of people vaccinated per hour- ratio (in percent) between population immunized and total population up to the date in the country.
- Total number of people fully vaccinated per hundred - ratio (in percent) between population fully immunized and total population up to the date in the country.
- Number of vaccinations per day - number of daily vaccinations for that day and country.
- Daily vaccinations per million - ratio (in ppm) between vaccination number and total population for the current date in the country.
- Vaccines used in the country - total number of vaccines used in the country (up to date).
- Source name - source of the information (national authority, international organization, local organization etc.).
- Source website - website of the source of information.

There is a second file added (country vaccinations by manufacturer), with the following columns:

- Location - country.
- Date - date.
- Vaccine - vaccine type.
- Total number of vaccinations - total number of vaccinations / current time and vaccine type.

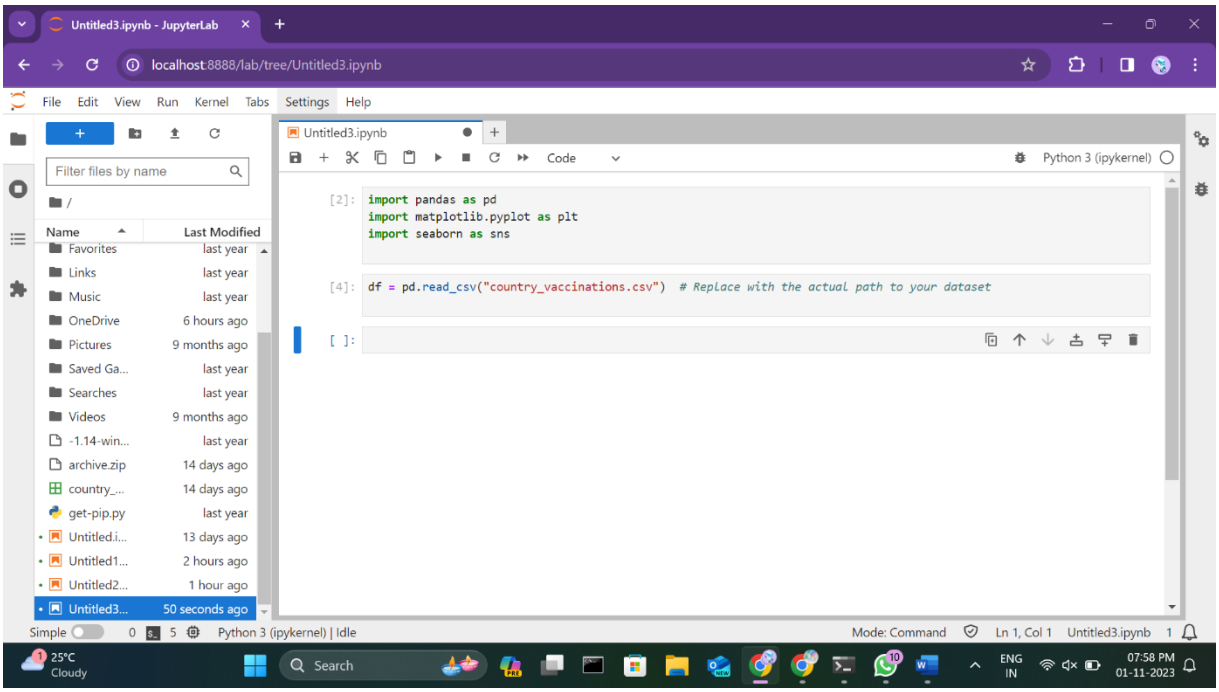
Importing the required libraries:

First, import the necessary Python libraries for data analysis and visualization:



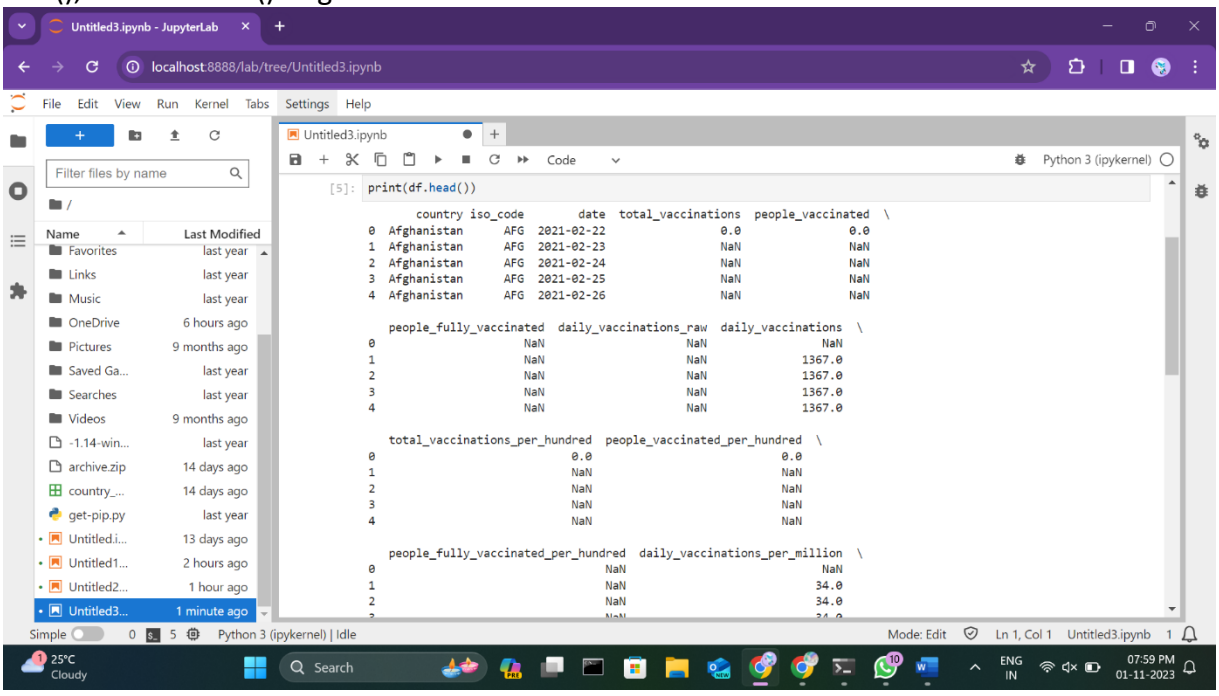
Importing the dataset:

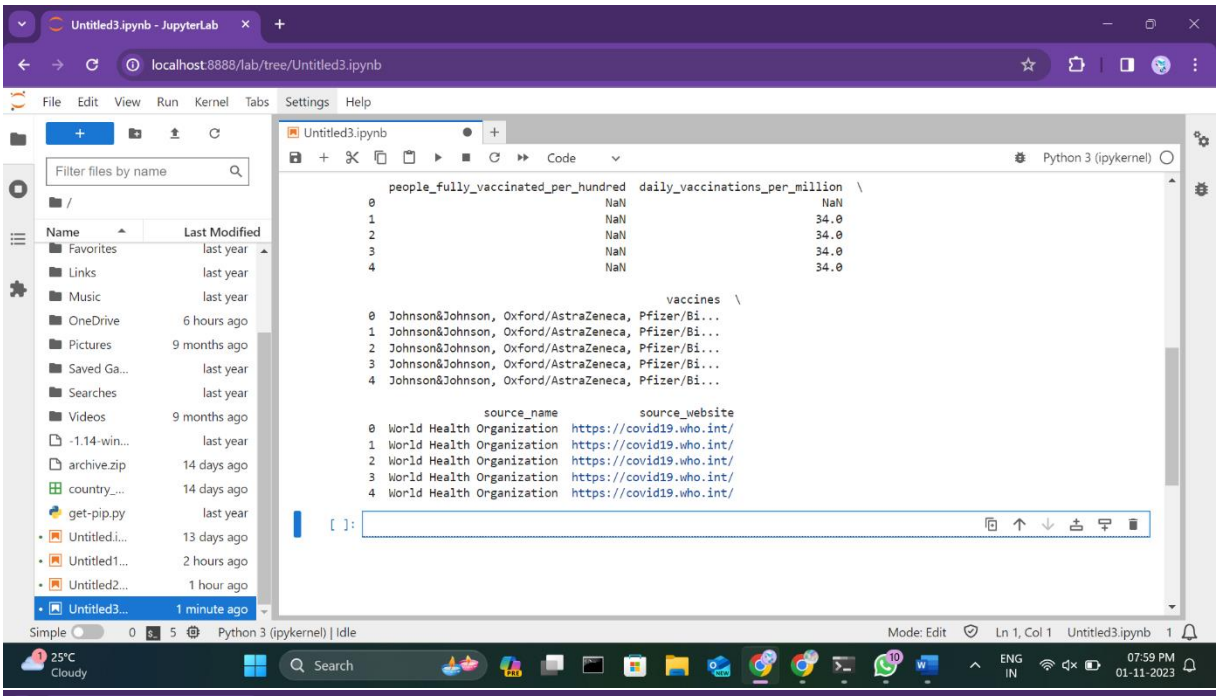
Load COVID-19 vaccination dataset into a Pandas Data Frame:

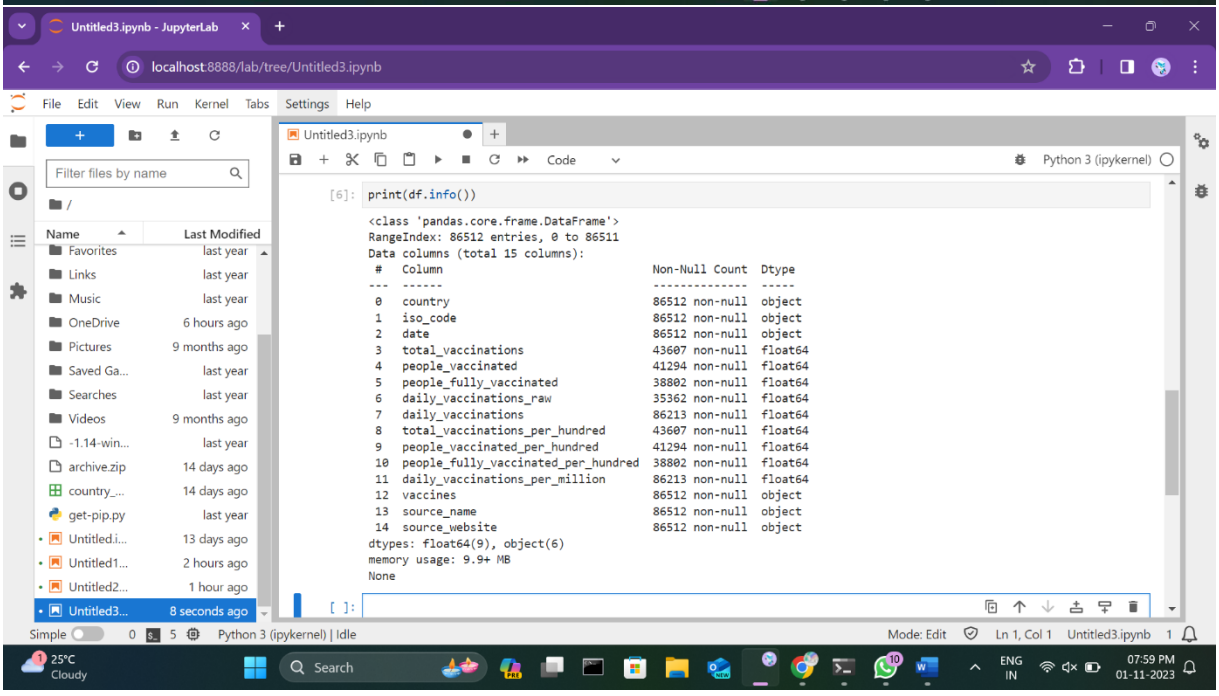


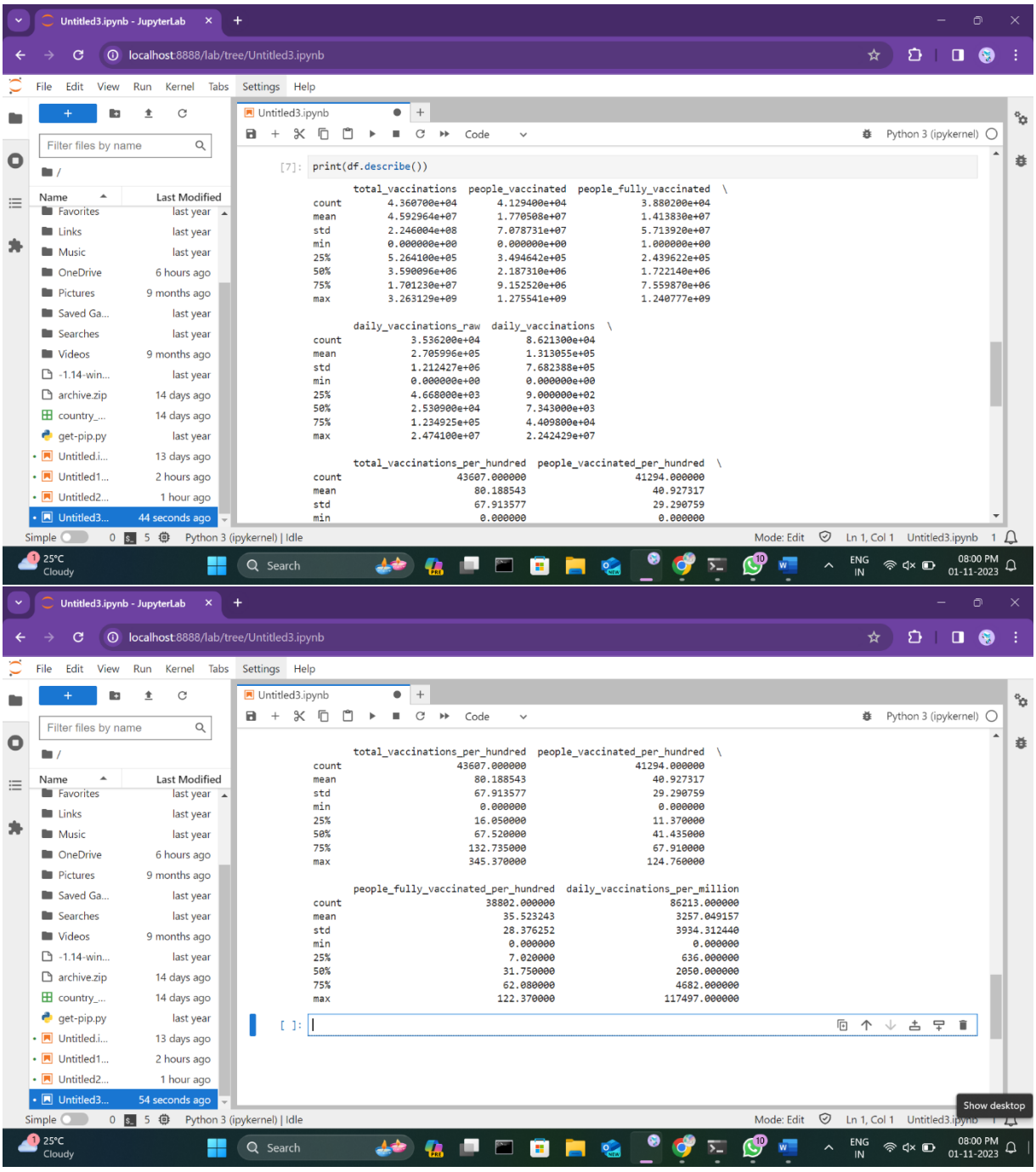
Exploratory Data Analysis:

Data Overview: Start by examining the structure of the dataset using functions like head(), info(), and describe() to get a sense of the data.

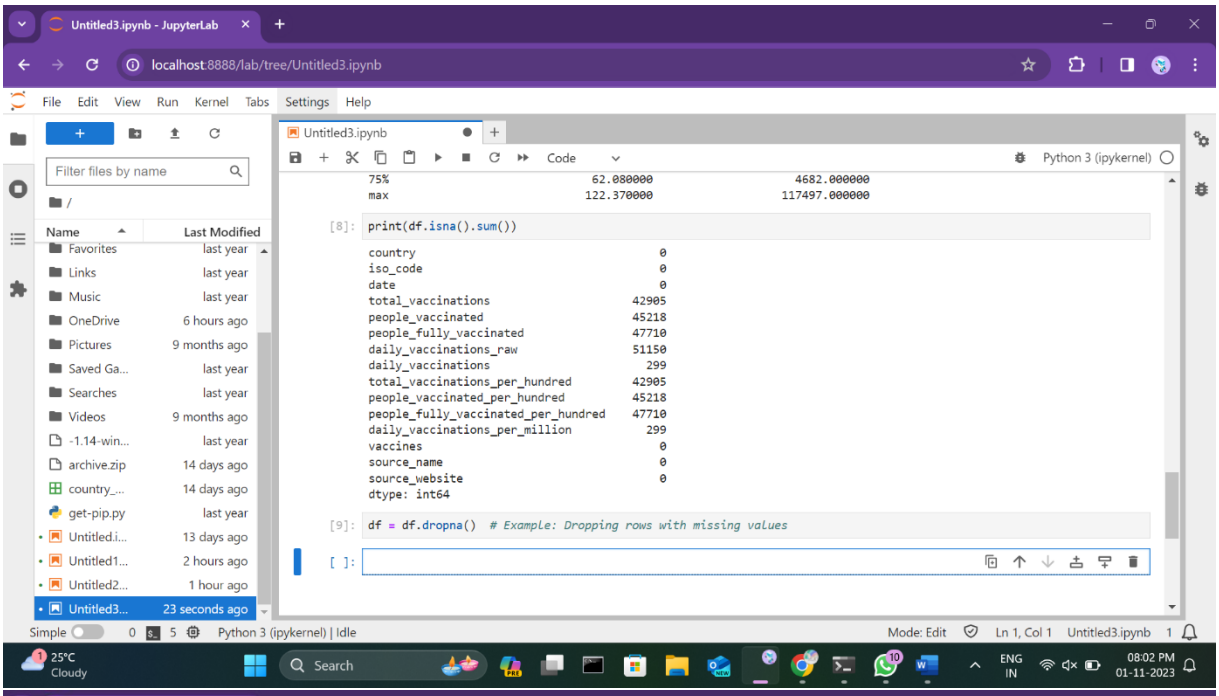


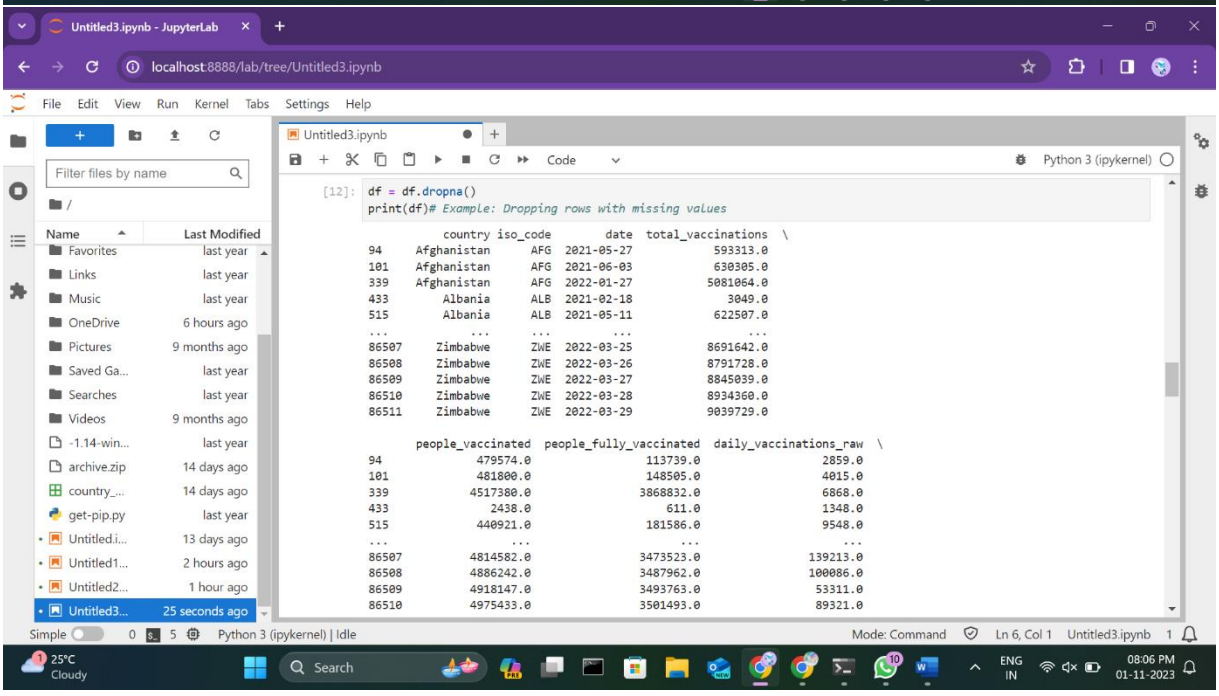

Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
people_fully_vaccinated_per_hundred daily_vaccinations_per_million \
0 NaN NaN
1 NaN 34.0
2 NaN 34.0
3 NaN 34.0
4 NaN 34.0
vaccines \
0 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
1 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
2 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
3 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
4 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
source_name source_website
0 World Health Organization https://covid19.who.int/
1 World Health Organization https://covid19.who.int/
2 World Health Organization https://covid19.who.int/
3 World Health Organization https://covid19.who.int/
4 World Health Organization https://covid19.who.int/
[ ]:
Simple
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 1, Col 1
Untitled3.ipynb
25°C
Cloudy
Search
ENG
IN
07:59 PM
01-11-2023

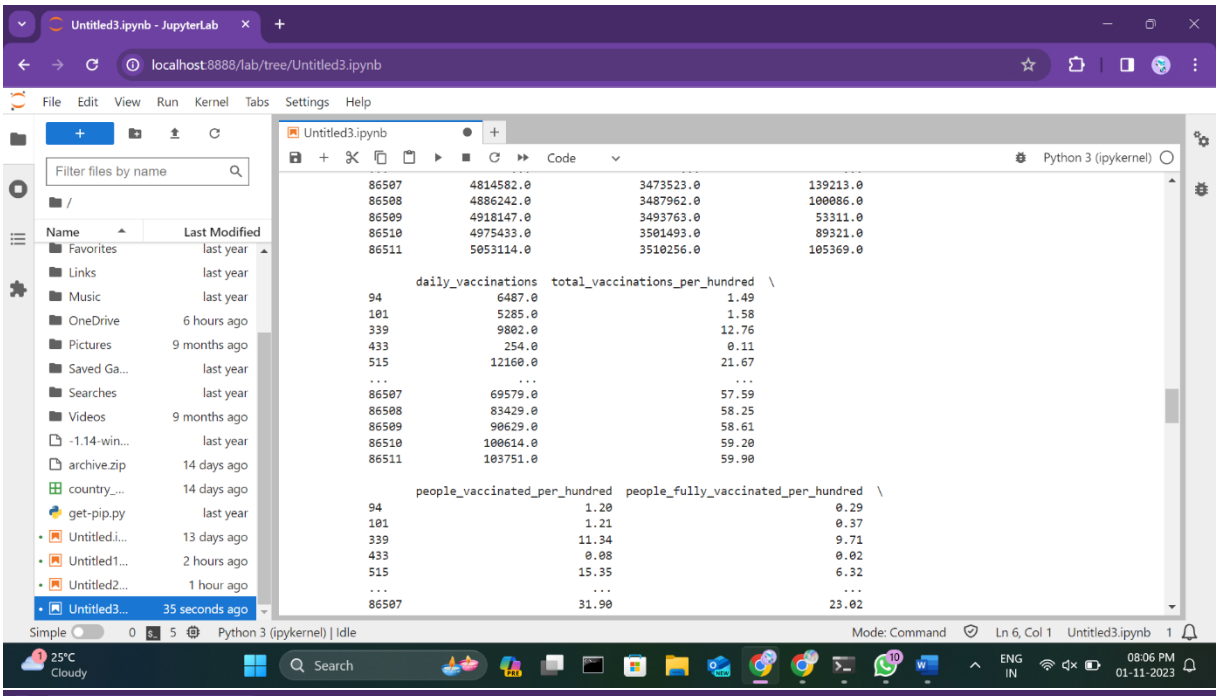

Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
[6]: print(df.info())
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 86512 entries, 0 to 86511
Data columns (total 15 columns):
Column Non-Null Count Dtype
--- ------ -------------- -----
0 country 86512 non-null object
1 iso_code 86512 non-null object
2 date 86512 non-null object
3 total_vaccinations 43607 non-null float64
4 people_vaccinated 41294 non-null float64
5 people_fully_vaccinated 38802 non-null float64
6 daily_vaccinations_raw 35362 non-null float64
7 daily_vaccinations 86213 non-null float64
8 total_vaccinations_per_hundred 43607 non-null float64
9 people_vaccinated_per_hundred 41294 non-null float64
10 people_fully_vaccinated_per_hundred 38802 non-null float64
11 daily_vaccinations_per_million 86213 non-null float64
12 vaccines 86512 non-null object
13 source_name 86512 non-null object
14 source_website 86512 non-null object
dtypes: float64(9), object(6)
memory usage: 9.9+ MB
None
[ ]:
Simple
Python 3 (ipykernel) | Idle
Mode: Edit
Ln 1, Col 1
Untitled3.ipynb
25°C
Cloudy
Search
ENG
IN
07:59 PM
01-11-2023

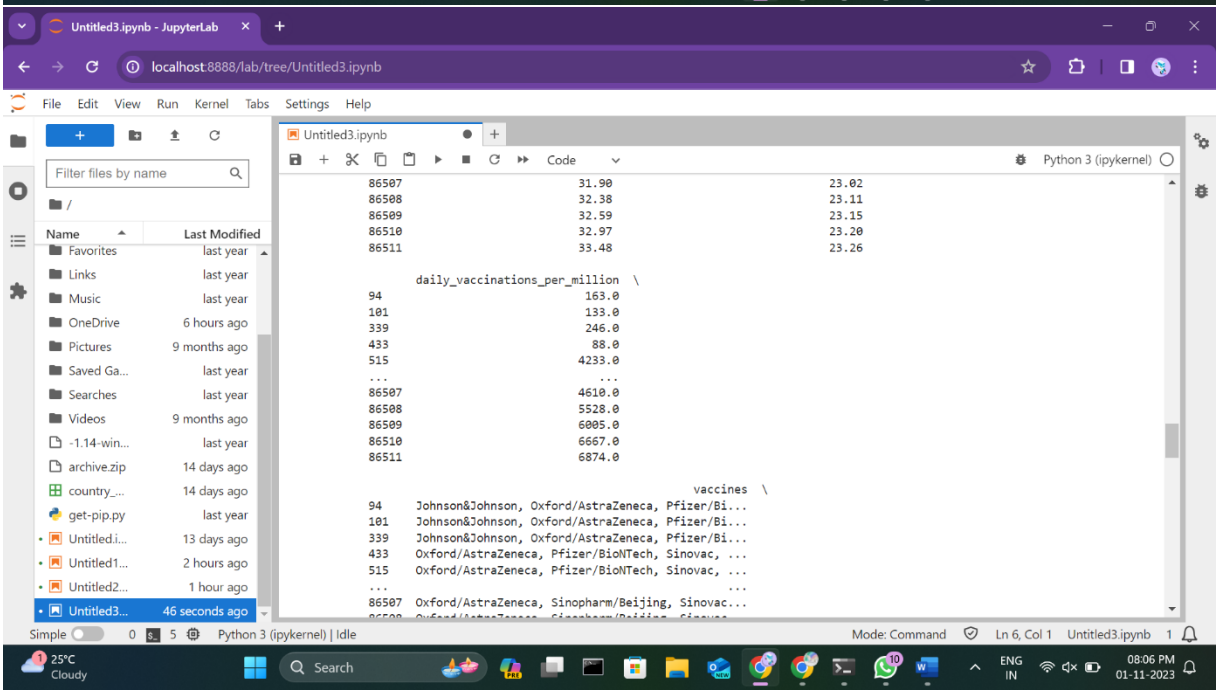


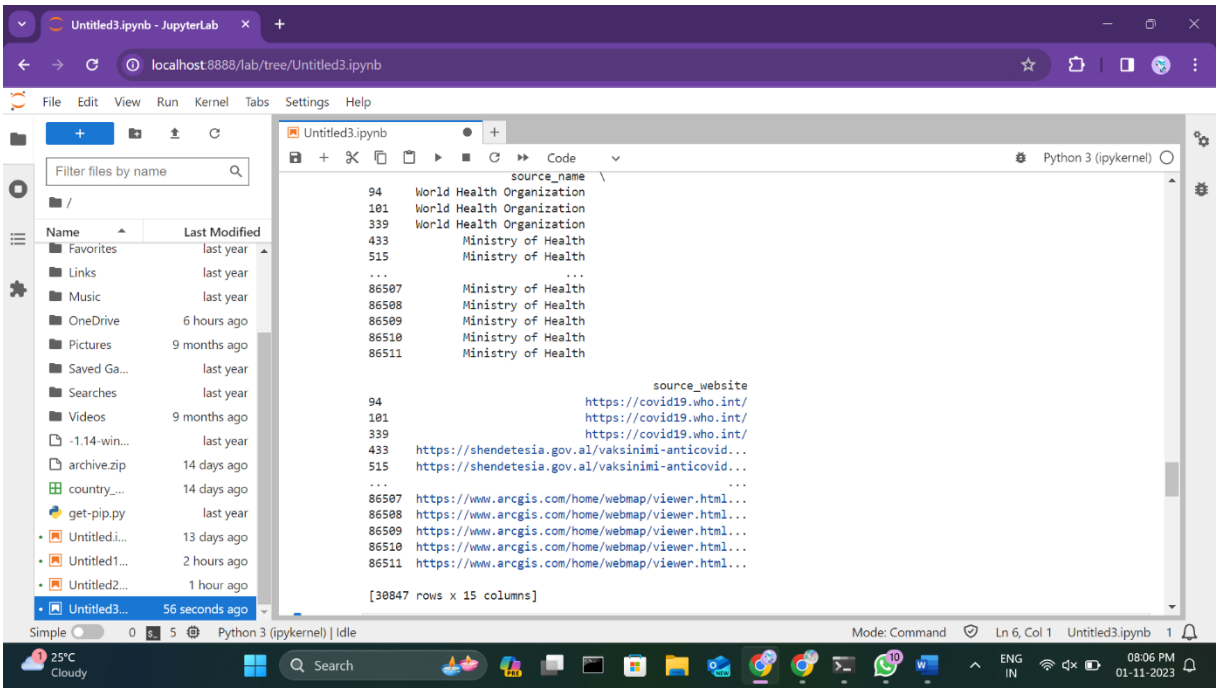
Data Cleaning: Check for missing values and handle them as needed. Use isna(), fillna(), or dropna() to manage missing data.


Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
75% 62.080000 4682.000000
max 122.370000 117497.000000
[8]: print(df.isna().sum())
country 0
iso_code 0
date 0
total_vaccinations 42905
people_vaccinated 45218
people_fully_vaccinated 47710
daily_vaccinations_raw 51150
daily_vaccinations 299
total_vaccinations_per_hundred 42905
people_vaccinated_per_hundred 45218
people_fully_vaccinated_per_hundred 47710
daily_vaccinations_per_million 299
vaccines 0
source_name 0
source_website 0
dtype: int64
[9]: df = df.dropna() # Example: Dropping rows with missing values
[ ]:
Simple 0 5 Python 3 (ipykernel) | Idle
Mode: Edit Ln 1, Col 1 Untitled3.ipynb 1
25°C Cloudy
Search
ENG IN
08:02 PM 01-11-2023

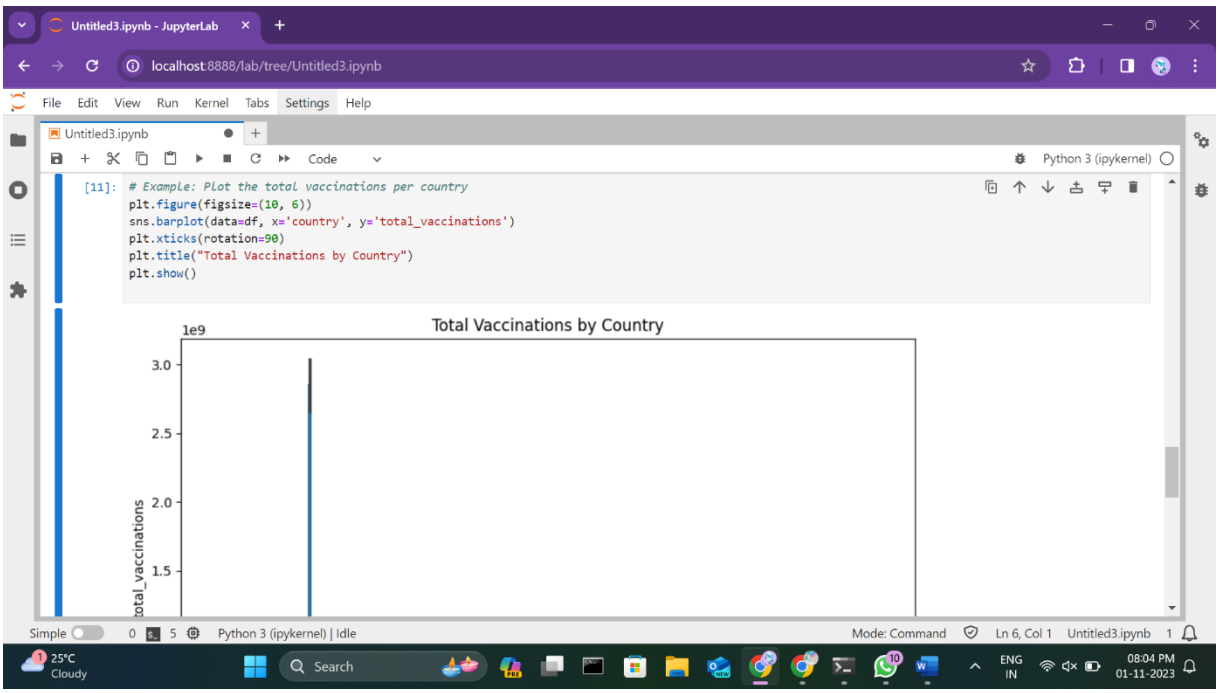

Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
[12]: df = df.dropna()
print(df)# Example: Dropping rows with missing values
country iso_code date total_vaccinations \
94 Afghanistan AFG 2021-05-27 593313.0
101 Afghanistan AFG 2021-06-03 630305.0
339 Afghanistan AFG 2022-01-27 5081064.0
433 Albania ALB 2021-02-18 3049.0
515 Albania ALB 2021-05-11 622507.0
... ...
86507 Zimbabwe ZWE 2022-03-25 8691642.0
86508 Zimbabwe ZWE 2022-03-26 8791728.0
86509 Zimbabwe ZWE 2022-03-27 8845039.0
86510 Zimbabwe ZWE 2022-03-28 8934360.0
86511 Zimbabwe ZWE 2022-03-29 9039729.0
people_vaccinated people_fully_vaccinated daily_vaccinations_raw \
94 479574.0 113739.0 2859.0
101 481800.0 148505.0 4015.0
339 4517380.0 3868832.0 6868.0
433 2438.0 611.0 1348.0
515 440921.0 181586.0 9548.0
... ...
86507 4814582.0 3473523.0 139213.0
86508 4886242.0 3487962.0 100086.0
86509 4918147.0 3493763.0 53311.0
86510 4975433.0 3501493.0 89321.0
Simple 0 5 Python 3 (ipykernel) | Idle
Mode: Command Ln 6, Col 1 Untitled3.ipynb 1
25°C Cloudy
Search
ENG IN
08:06 PM 01-11-2023

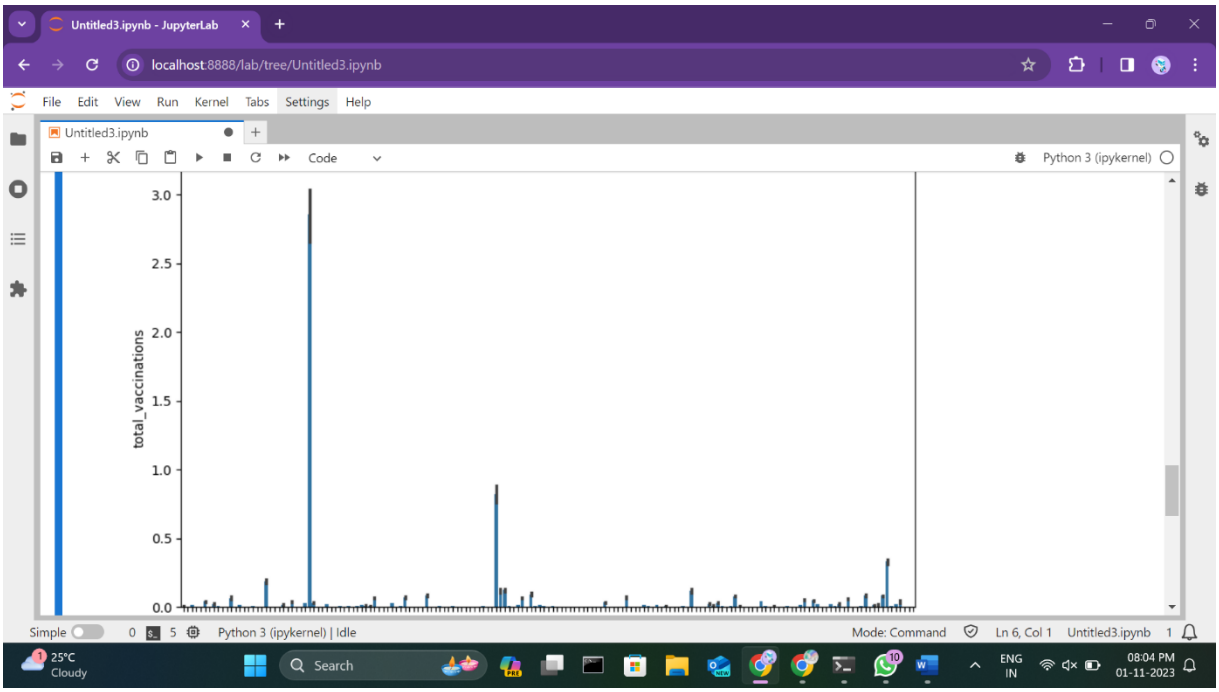

Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
86507 4814582.0 3473523.0 139213.0
86508 4886242.0 3487962.0 100086.0
86509 4918147.0 3493763.0 53311.0
86510 4975433.0 3501493.0 89321.0
86511 5053114.0 3510256.0 105369.0
daily_vaccinations total_vaccinations_per_hundred \
94 6487.0 1.49
101 5285.0 1.58
339 9802.0 12.76
433 254.0 0.11
515 12160.0 21.67
86507 69579.0 57.59
86508 83429.0 58.25
86509 90629.0 58.61
86510 100614.0 59.20
86511 103751.0 59.90
people_vaccinated_per_hundred people_fully_vaccinated_per_hundred \
94 1.20 0.29
101 1.21 0.37
339 11.34 9.71
433 0.08 0.02
515 15.35 6.32
86507 31.90 23.02
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 6, Col 1
Untitled3.ipynb
25°C Cloudy
Search
ENG IN
08:06 PM 01-11-2023


Untitled3.ipynb - JupyterLab
localhost:8888/lab/tree/Untitled3.ipynb
File Edit View Run Kernel Tabs Settings Help
Filter files by name
Name
Last Modified
Untitled3.ipynb
Code
Python 3 (ipykernel)
86507 31.90 23.02
86508 32.38 23.11
86509 32.59 23.15
86510 32.97 23.20
86511 33.48 23.26
daily_vaccinations_per_million \
94 163.0
101 133.0
339 246.0
433 88.0
515 4233.0
86507 4610.0
86508 5528.0
86509 6005.0
86510 6667.0
86511 6874.0
vaccines \
94 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
101 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
339 Johnson&Johnson, Oxford/AstraZeneca, Pfizer/Bi...
433 Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, ...
515 Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, ...
86507 Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac...
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 6, Col 1
Untitled3.ipynb
25°C Cloudy
Search
ENG IN
08:06 PM 01-11-2023



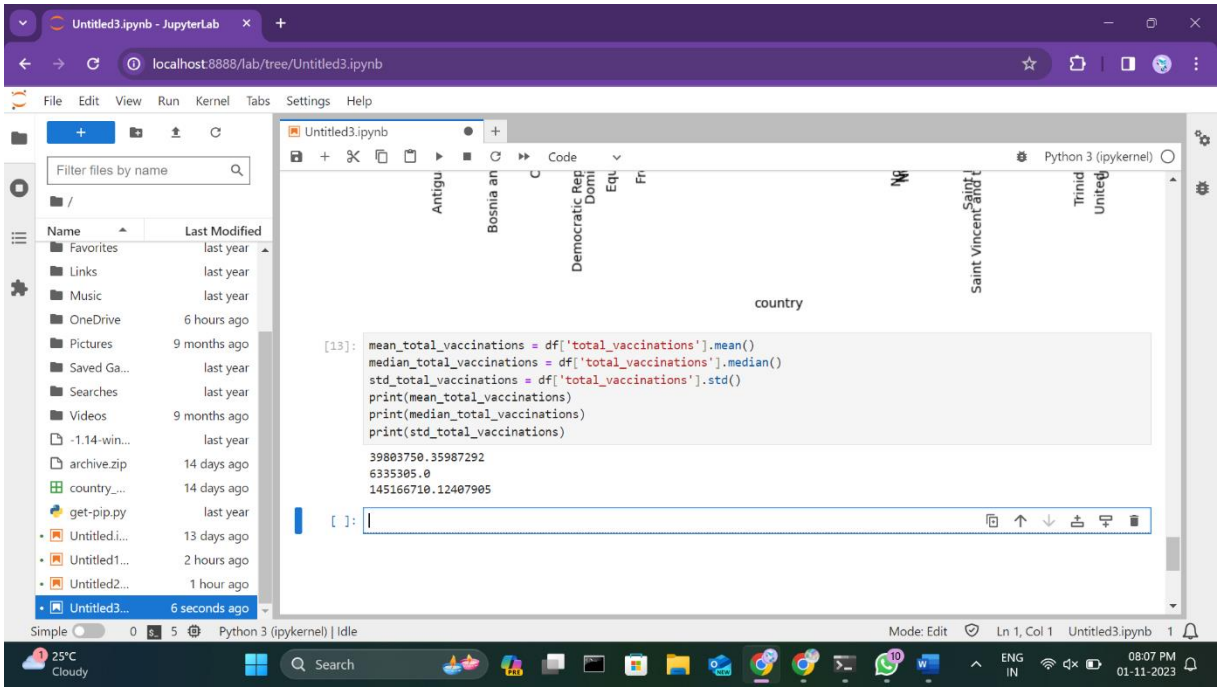
Data Visualization: Create visualizations to explore the data. Use tools like Matplotlib and Seaborn for this.





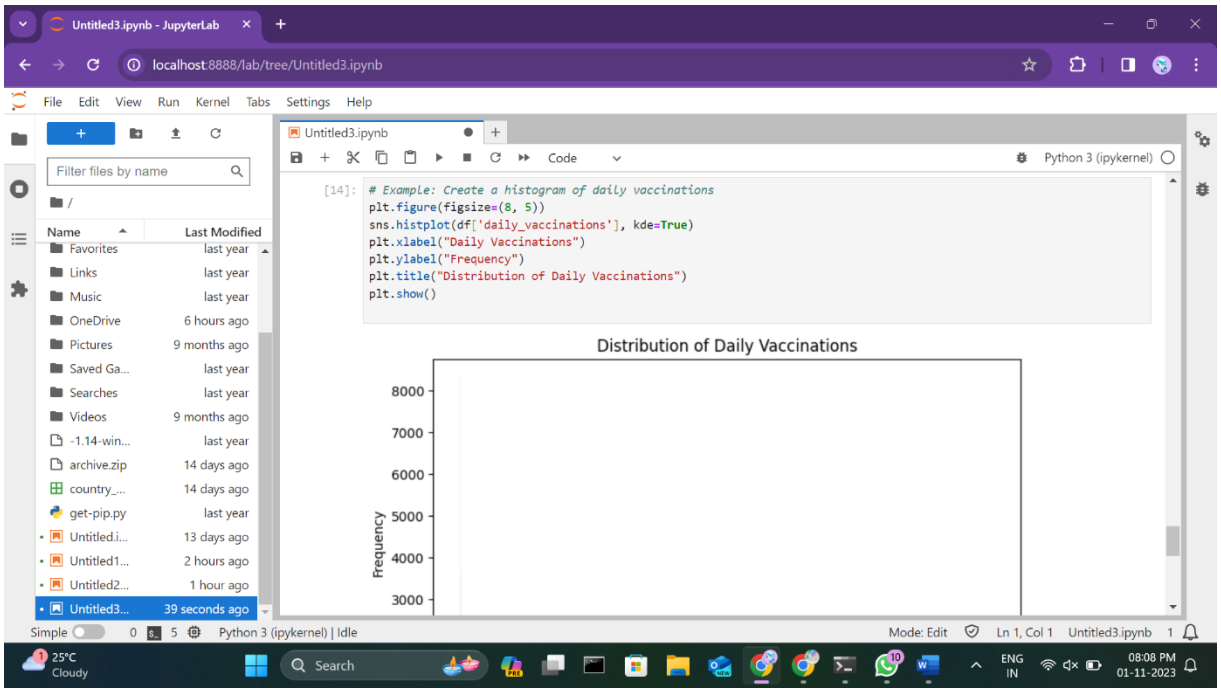
Statistical Analysis:

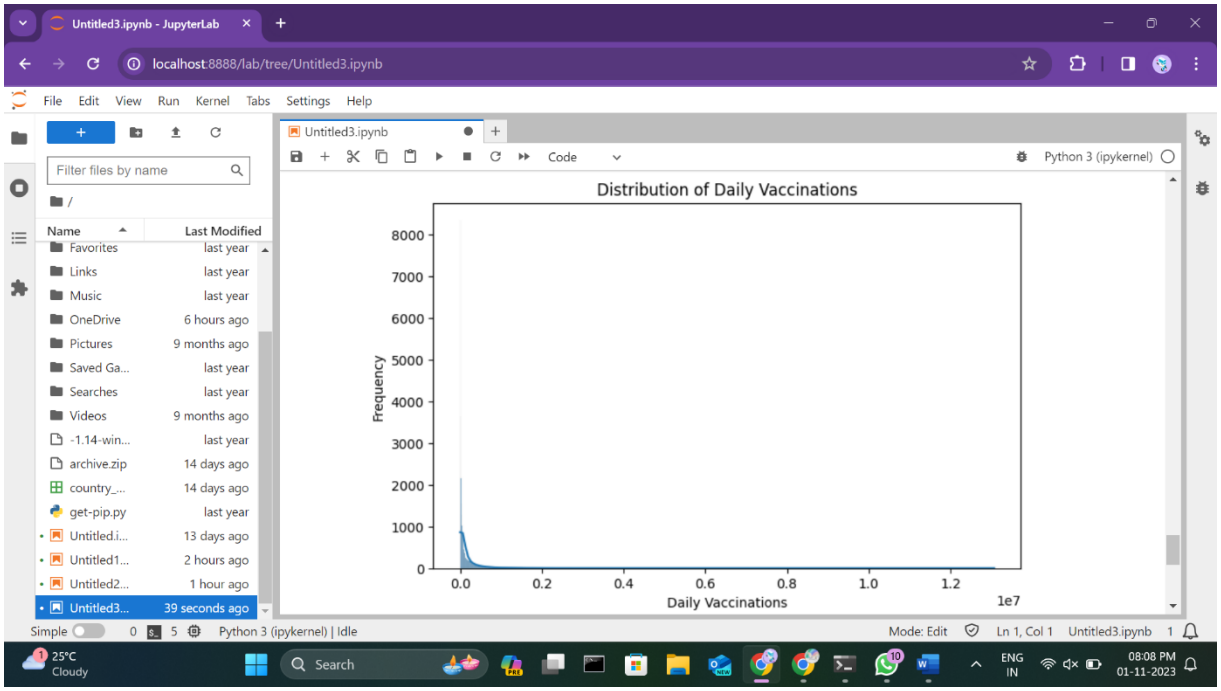
Calculate summary statistics, such as mean, median, and standard deviation.



Data Visualisation:

Visualize the results of statistical analysis.





TEAM-MATES: RITHIKA B

SOWMIYA G