PHASE-5

PROJECT-5 COVID-19 VACCINESS ANALYSIS

INTRODUCTION

DATA COLLECTION

- ***Research Databases***

Acquire scientific papers, preprints, and peer-reviewed articles from reputable databases such as PubMed, IEEE Xplore, and Google Scholar. These sources offer valuable insights into vaccine development, efficacy, and safety

- ***Clinical Trial Registries***

Gather data from official clinical trial registries, including ClinicalTrials.gov and the
WHO International Clinical Trials Registry Platform (ICTRP). This data will provide information on ongoing and completed vaccine trials, including phase, participant demographics, and outcomes.

- ***News Media and Press Releases***

Collect news articles and press releases from trusted news outlets, government health agencies, and pharmaceutical companies. This data source offers real-time updates on vaccine distribution, public reactions, and regulatory approvals.

- ***Government and Health Organization Reports***

Access official reports and publications from health authorities such as the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), and the European Medicines Agency (EMA). These documents contain critical information on vaccine guidelines, safety assessments, and vaccination campaigns.

DATA PREPROCESSING

- ***Handling Missing Values***

  - Identify missing data in your dataset.

  - Decide how to handle missing values: either by removing rows with missing data, filling them with a default value (e.g., mean, median, or mode), or using more advanced techniques like interpolation.

- ***Encoding Categorical Features***

  - Convert categorical variables into numerical representations that machine learning models can understand.

  - For nominal variables (categories without a specific order), you can use one-hot encoding or label encoding.

  - For ordinal variables (categories with a meaningful order), use ordinal encoding.

- ***Feature Scaling***

  - Normalize or standardize numerical features to bring them to a similar scale.

  - Common techniques include Min-Max scaling (scaling to a specific range) or Z-score normalization (scaling to have mean=0 and standard deviation=1).

- ***Handling Outliers***

  - Identify and decide how to deal with outliers. You can choose to remove them or transform them using techniques like log transformation.

- ***Data Splitting***

  - Split your dataset into training, validation, and test sets to evaluate your model's performance properly.

EXPLORATORY DATA ANALYSIS (EDA)

- ***Data Overview***

Begin the EDA by providing a summary of the dataset, including its size, number of features, and data types. Highlight any key variables of interest that will be central to the analysis.

- ***Descriptive Statistics***

Calculate and present summary statistics for numerical features. This should include measures of central tendency (mean, median) and dispersion (standard deviation, range). For categorical variables, provide frequency counts and percentages.

- ***Data Distribution***

Visualize the distribution of key variables using appropriate plots and charts. Histograms, box plots, and density plots can reveal the underlying distribution patterns and potential outliers.

- ***Correlation Analysis***

Examine the relationships between variables by calculating correlation coefficients. Create correlation matrices and visualizations (e.g., heatmaps) to identify strong positive or negative correlations.

- ***Feature Importance***

If machine learning will be applied, use techniques such as feature importance scores to assess the relevance of different features in achieving project objectives. Present the results to guide feature selection.

- ***Exploratory Visualization***

Create visualizations that provide insights into the data. Examples include:

- Time Series Plots: If applicable, plot trends over time related to vaccine distribution, case numbers, or public sentiment.

- Word Clouds: Visualize frequently occurring words in text data to identify common themes.

- Scatterplots: Explore relationships between variables, such as vaccine coverage and disease incidence.

Statistical Analysis

- ***Hypothesis Testing***

Formulate hypotheses based on project goals and data characteristics. Common hypotheses in COVID-19 vaccine analysis include:

- "COVID-19 vaccination significantly reduces the risk of infection."

- "Vaccine efficacy varies by vaccine type."

- "Vaccine distribution is associated with reduced disease incidence."

Select appropriate statistical tests based on the nature of the hypotheses, including t-tests, chi-square tests, ANOVA, and regression analysis.

- ***Pattern Recognition***

Use descriptive statistics, data visualization, and exploratory analysis techniques to identify patterns and trends within the dataset. Key tasks include:

- Central Tendency: Analyze measures of central tendency, such as mean, median, and mode, for numerical variables.

- Data Distribution: Examine distributions and identify skewness, kurtosis, and multimodality.

- Correlation Analysis: Investigate correlations between variables to identify relationships.

- Time Series Analysis: Perform time series analysis for trends over time, especially relevant for vaccine distribution and disease incidence data.

- ***Data Validation***

Ensure data meets the assumptions of statistical methods used. Tasks include:

- Normality Tests: Verify the normality of data distribution when applicable.

- Homoscedasticity: Assess homoscedasticity, ensuring constant variance in regression analysis.

- Multicollinearity: Identify and address multicollinearity issues in regression models.

- ***Interpretation and Conclusion***

Interpret the statistical findings in the context of the project's objectives and hypotheses. Summarize key insights and implications. Address any limitations or assumptions made during the analysis.

- ***Reporting and Documentation***

Document the entire statistical analysis process, including methodologies, results, and any code or scripts used. Provide clear visualizations and tables to convey findings effectively.

VISUALIZATION

- ***Data Visualization***

Data visualization involves creating graphical representations of data to facilitate understanding. Key data visualization tasks include:

1. Descriptive Statistics: Present summary statistics using charts, histograms, and box plots to provide a snapshot of the data's distribution.

2. Time Series Plots: Visualize trends over time, such as vaccine distribution, disease incidence, and public sentiment, using line charts or time series plots.

3.Correlation Heatmaps: Create heatmaps to display correlation matrices, highlighting relationships between variables.

4. Geospatial Maps: Develop geospatial maps to visualize regional variations in vaccine distribution, disease prevalence, or public sentiment.

5. Word Clouds: Use word clouds to visualize frequently occurring words in text data, providing insights into common themes.

6. Scatterplots: Illustrate relationships between variables through scatterplots, especially when investigating vaccine efficacy, public perception, or other factors.

7. Bar and Pie Charts: Create bar and pie charts for categorical data to display distributions and proportions.

INSIGHTS

- ***Insight Generation***

-Summarize the most significant findings and trends identified during the analysis. Highlight insights related to vaccine efficacy, safety, distribution, public sentiment, and disease impact.

- ***Identifying Patterns***

- Discuss any recurring patterns or correlations uncovered in the data. Explain their implications in the context of COVID-19 vaccines.

- ***Public Perception***

- Analyze public sentiment trends and the factors influencing vaccine hesitancy or acceptance. Consider insights from social media data and surveys.

- ***Vaccine Variants***

- Assess the adaptability of existing vaccines to emerging variants of the virus and implications for future vaccination efforts.

- ***Hypothesis Validation***

- Review and validate hypotheses formulated during the analysis phase. Discuss the extent to which the data and analysis support or refute these hypotheses

RECOMMENDATIONS

- ***Public Health Recommendations***

- Develop recommendations for public health authorities, healthcare providers, and policymakers. These recommendations may include vaccination strategies, communication plans, and targeted interventions.

- ***Vaccine Distribution Strategies***

- Offer insights into optimizing vaccine distribution to ensure equitable access, especially in underserved populations or regions.

- ***Safety Monitoring***

- Recommend strategies for ongoing safety monitoring and reporting of adverse events associated with COVID-19 vaccines.

- ***Future Research Directions***

- Suggest areas for future research and analysis, such as the evaluation of booster shots, long-term vaccine impact, and vaccine adaptability to new variants.

- ***Actionable Insights***

- Emphasize actionable insights that can lead to tangible improvements in vaccination campaigns, public perception, and overall COVID-19 response efforts.

- ***Communication***

- Prepare clear and concise reports, presentations, or documents that convey the insights and recommendations effectively to various stakeholders.

INNOVATION

Innovations in the analysis of COVID-19 vaccines have been crucial for monitoring their effectiveness, safety, and adaptation to emerging variants. Here are some key innovations in this area:

1. Real-World Data Analysis: *The use of real-world data from vaccinated populations has allowed for continuous monitoring of vaccine effectiveness and safety. Large-scale data analysis has been instrumental in identifying trends and anomalies.*

2. Variant Surveillance: *Ongoing genomic sequencing and analysis of SARS-CoV-2 variants have enabled researchers to assess how vaccines perform against different strains. This has led to adjustments in vaccine strategies and booster shots.*

3. Vaccine Efficacy Models: *Mathematical models have been developed to predict vaccine efficacy under various scenarios, helping public health officials make informed decisions about vaccine distribution and vaccination strategies.*

4. Adverse Event Detection: *Advanced data analytics have been applied to rapidly detect and investigate adverse events following vaccination. This helps ensure vaccine safety and allows for swift response when potential issues arise.*

5. Immunological Assays: *Innovations in immunological assays, such as neutralization assays and T-cell response studies, have provided insights into the immune response generated by vaccines.*

6. Vaccine Effectiveness Against Variants: *Studies have assessed how well vaccines protect against specific variants, informing decisions about booster shots and updated vaccine formulations.*

7.Vaccine Heterologous Boosting: *Research into the effectiveness of mixing and matching different vaccines (heterologous boosting) has been conducted to optimize vaccination strategies.*

8. Adaptive Clinical Trials: *Trials have been designed to adapt to changing circumstances, allowing for quick assessment of new vaccine candidates and mo 3 difications to existing ones.*

9. Global Data Sharing: *International collaboration and data sharing have been critical for analyzing vaccine performance globally and ensuring equitable access to vaccines.*

10. AI and Machine Learning: *Artificial intelligence and machine learning techniques have been used to analyze vast datasets, identify trends, and predict vaccine outcomes.*

- ***DATASET :-***

I took the dataset from(www.kaggle.com/data).
The dataset is related to Covid-19 Vaccines Analysis.

MY DATASET LINK:

(https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress)

- ***DETAILS OF MY DATASET:-***

In my dataset the column names contains:

1)Country

2)Iso_code

3)Date

4)Total_vaccination

5)People_vaccinated

6)People_fully_vaccinated

7)Daily_vaccinations_raw

8)Total_vaccination

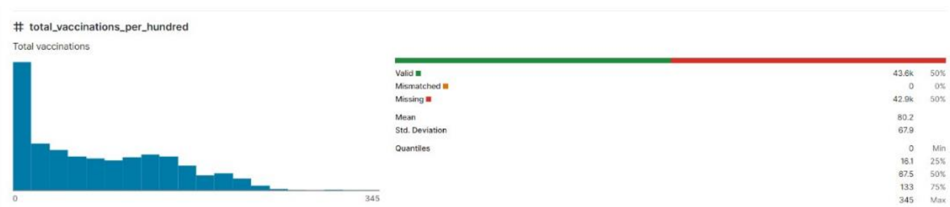
9)People_vaccinated_per_hundred

for eg: (flowchart):


country_vaccinations.csv (17.73 MB)

Detail Compact Column — 10 of 15 columns

country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vacc...	daily_vaccination...	daily_vaccin
Country	ISO Code	Date	Total vaccinations	People vaccinated	People fully vaccinated	Daily vaccinations	Daily vaccinatic
	223 unique values	2Dec20 – 29Mar22	0 – 3.26b	0 – 1.28b	1 – 1.24b	0 – 24.7m	0
Afghanistan	AFG	2021-02-22	0.0	0.0			
Afghanistan	AFG	2021-02-23					1367.0
Afghanistan	AFG	2021-02-24					1367.0
Afghanistan	AFG	2021-02-25					1367.0
Afghanistan	AFG	2021-02-26					1367.0
Afghanistan	AFG	2021-02-27					1367.0
Afghanistan	AFG	2021-02-28	8200.0	8200.0			1367.0
Afghanistan	AFG	2021-03-01					1580.0
Afghanistan	AFG	2021-03-02					1794.0
Afghanistan	AFG	2021-03-03					2008.0



METRICES USED FOR ACCURACY CHECK:-

For Classification Tasks (e.g., binary or multiclass classification):

1. ***Accuracy:*** *This is a widely used metric that calculates the ratio of correctly predicted instances to the total number of instances. It's suitable for balanced datasets. However, it may not be the best choice for imbalanced datasets, where a class is significantly more prevalent than others.*

2. ***Precision:*** *Precision measures the ratio of correctly predicted positive instances to the total predicted positive instances. It's valuable when minimizing false positives is crucial, such as in spam detection or medical diagnosis.*

3. ***F1-Score:*** *The F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is useful when there's an uneven class distribution.*

For Regression Tasks (e.g., predicting numerical values):

1. ***Mean Absolute Error (MAE):*** *MAE calculates the average absolute difference between the predicted values and the actual values. It is less sensitive to outliers compared to RMSE.*

2. ***Mean Squared Error (MSE):*** *MSE calculates the average of the squared differences between predicted and actual values. It penalizes larger errors more heavily than MAE.*

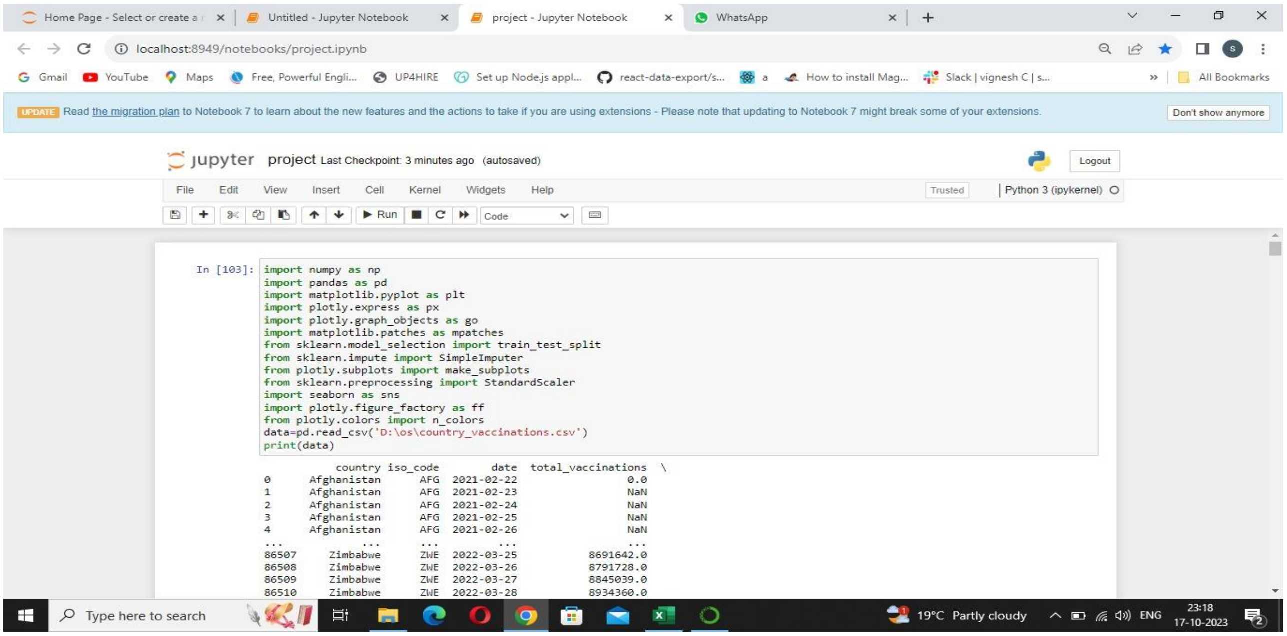
3. ***Root Mean Squared Error (RMSE):*** *RMSE is the square root of MSE and provides an interpretable measure of prediction error in the same units as the target variable. It's sensitive to outliers.*

BEGIN BUILDING THE PROJECT BY LOAD THE DATASET

To import the required libraries and read a CSV file,

```
data=pd.read_csv('D:\os\country_vaccinations.csv')

Print(data)
```



PREPROCESS DATASET

Import The Required Libraries:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import plotly.express as px

import plotly.graph_objects as go

import matplotlib.patches as mpatches

from sklearn.model_selection import train_test_split

from sklearn.impute import SimpleImputer

from plotly.subplots import make_subplots

from sklearn.preprocessing import StandardScaler *import seaborn as sns*

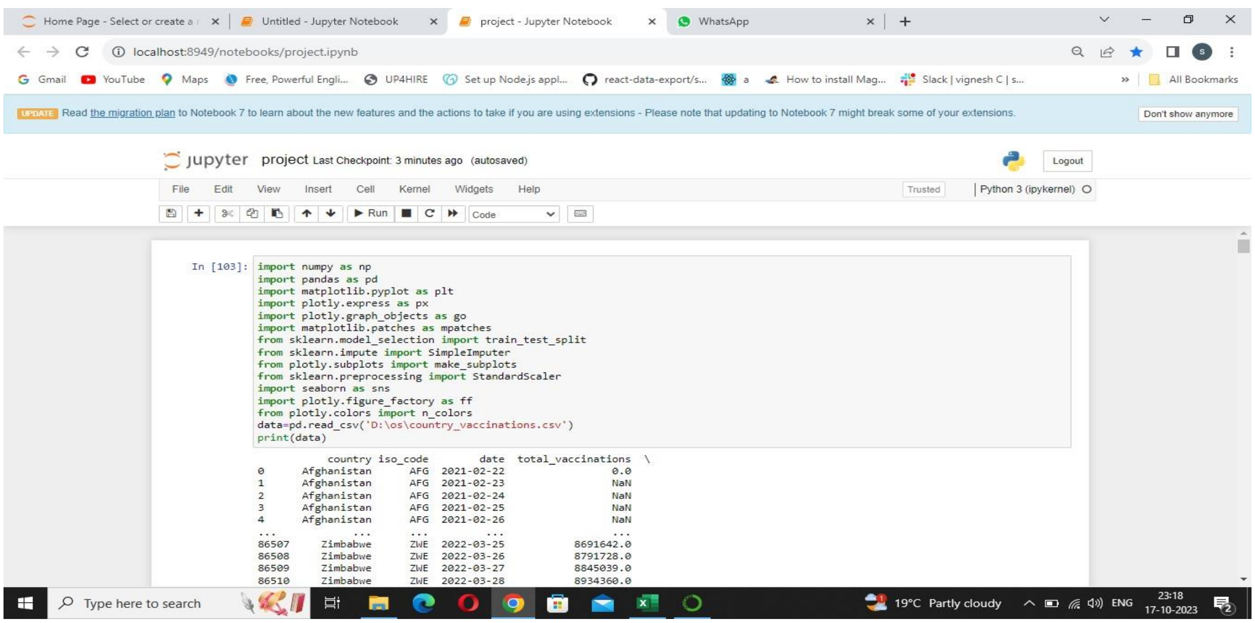
import plotly.figure_factory as ff

from plotly.colors import n_colors

Importing the Dataset:

Read Dataset,

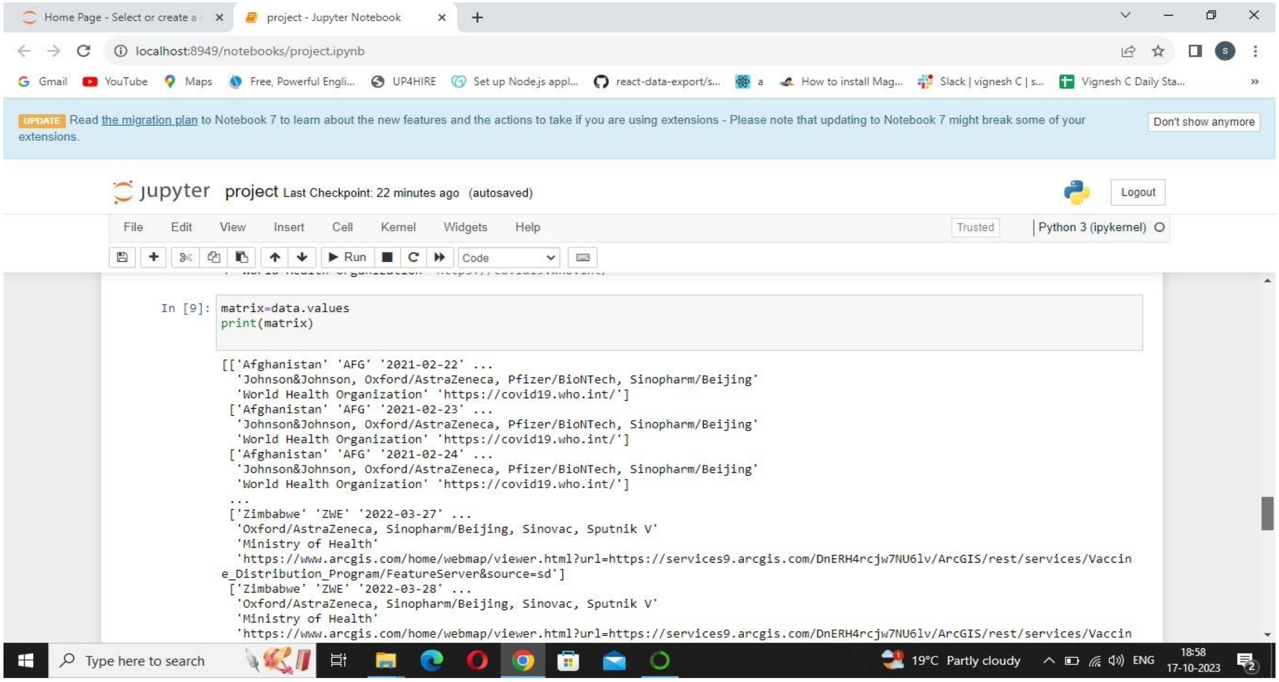
data=pd.read_csv('D:\os\country_vaccinations.csv')print(data)



- *Create Matrix,*

matrix=data.values

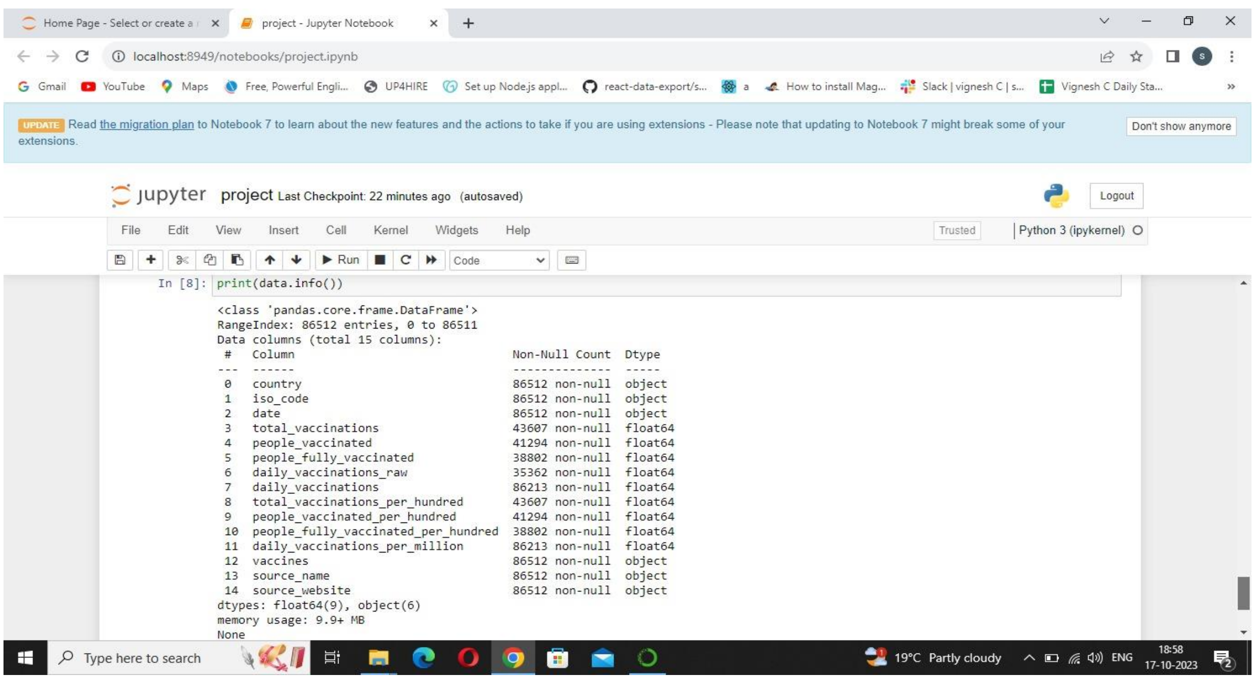
print(matrix)



```
In [9]: matrix=data.values
        print(matrix)

[['Afghanistan' 'AFG' '2021-02-22' ...
  'Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing'
  'World Health Organization' 'https://covid19.who.int/']
 ['Afghanistan' 'AFG' '2021-02-23' ...
  'Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing'
  'World Health Organization' 'https://covid19.who.int/']
 ['Afghanistan' 'AFG' '2021-02-24' ...
  'Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing'
  'World Health Organization' 'https://covid19.who.int/']
 ...
 ['Zimbabwe' 'ZWE' '2022-03-27' ...
  'Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V'
  'Ministry of Health'
  'https://www.arcgis.com/home/webmap/viewer.html?url=https://services9.arcgis.com/DnERH4rcjw7NU6lv/ArcGIS/rest/services/Vaccine_Distribution_Program/FeatureServer&source=sd']
 ['Zimbabwe' 'ZWE' '2022-03-28' ...
  'Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sputnik V'
  'Ministry of Health'
  'https://www.arcgis.com/home/webmap/viewer.html?url=https://services9.arcgis.com/DnERH4rcjw7NU6lv/ArcGIS/rest/services/Vaccin
```

- *Other Imformation about dataset,*

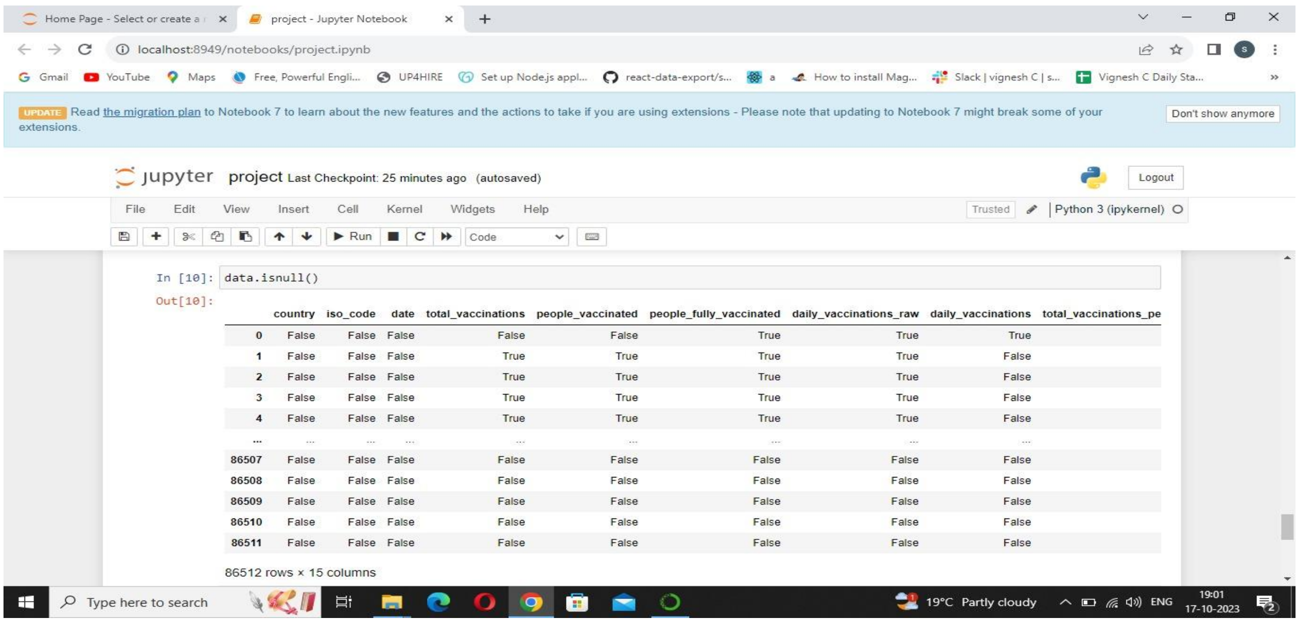


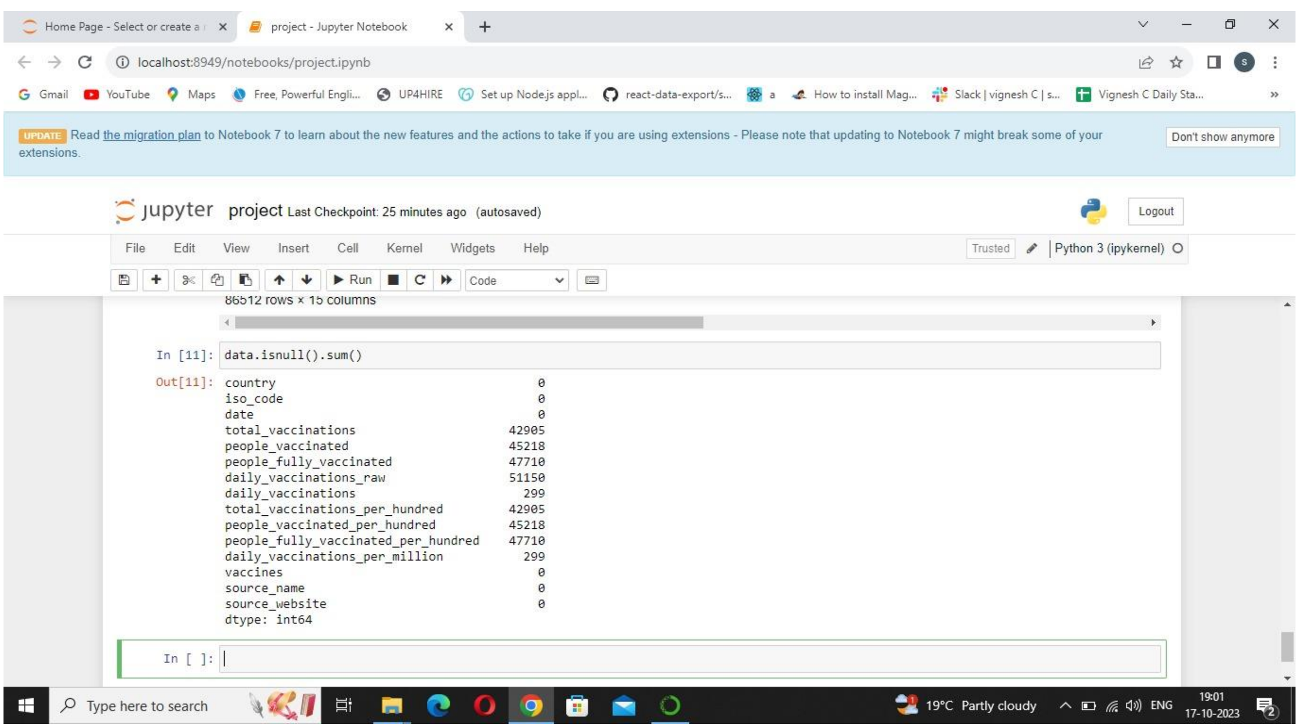
```
In [8]: print(data.info())

<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
```

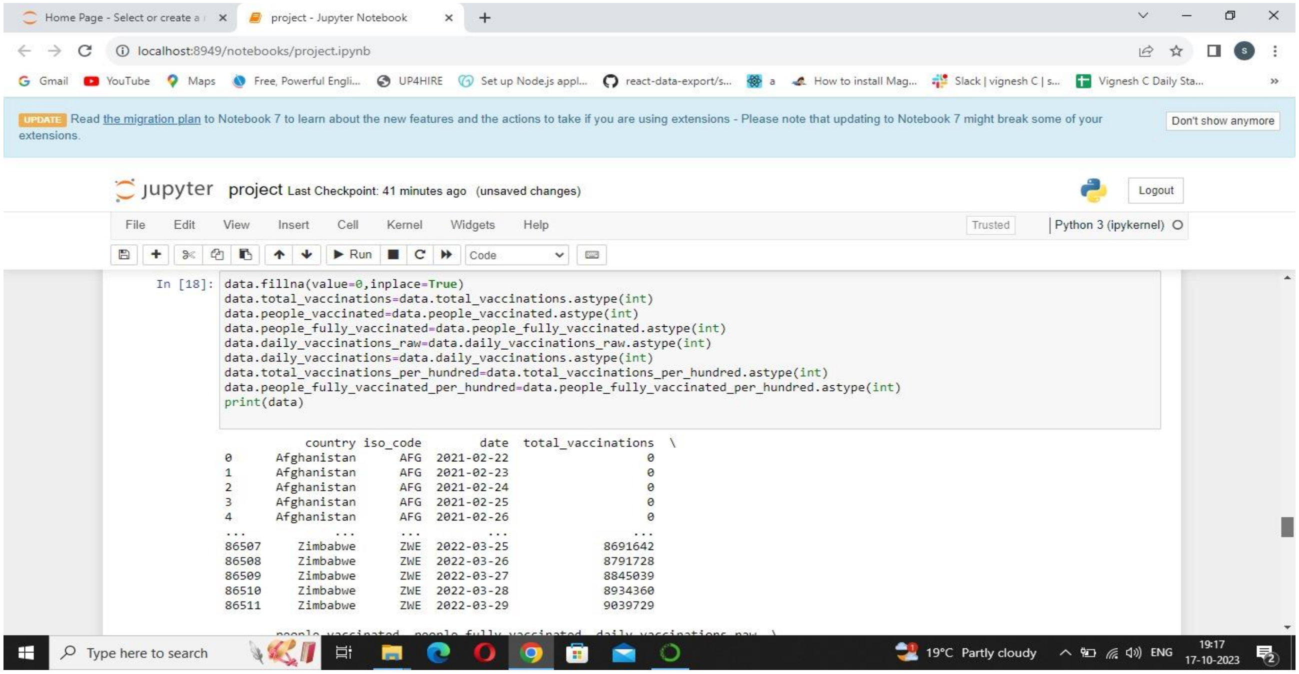
Handling The Missing Data:

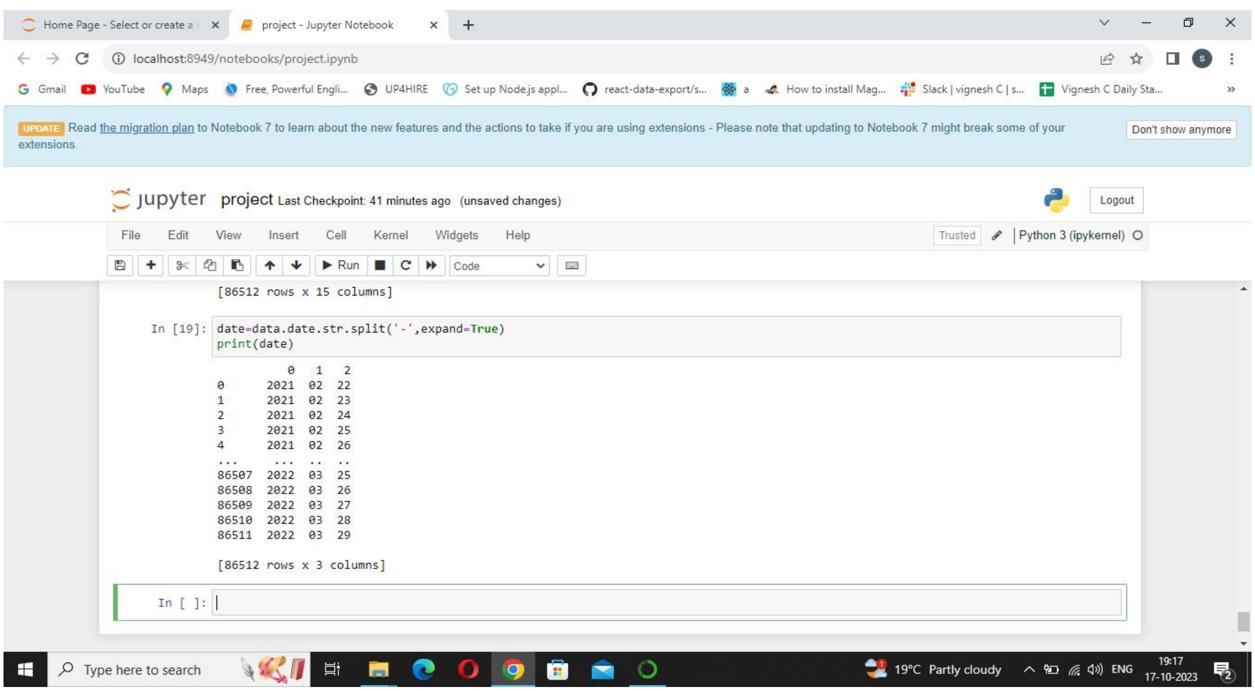
Before Handling the Missing data, we use isnull () to show the null values and using isnull().sum to get total number of null values.

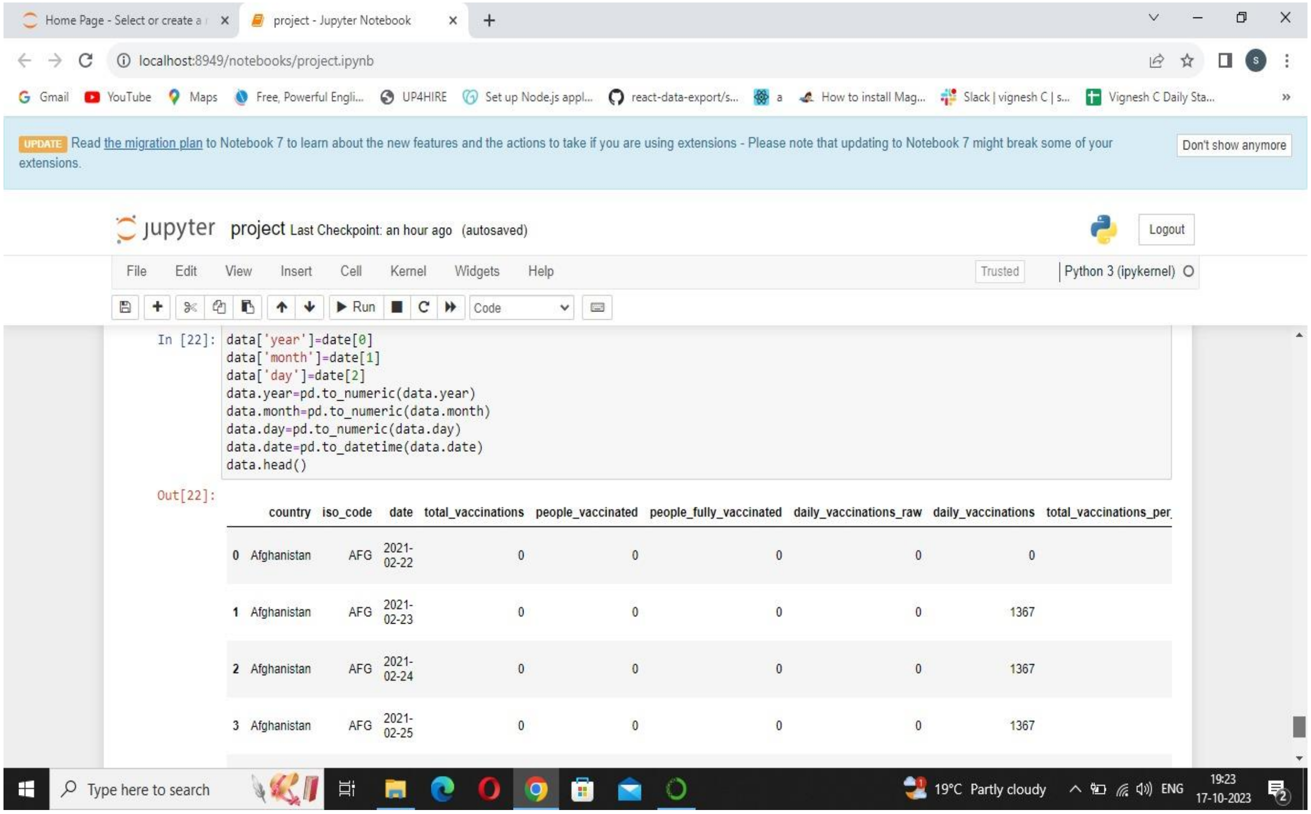




The below code in the image for all the data cleaning,



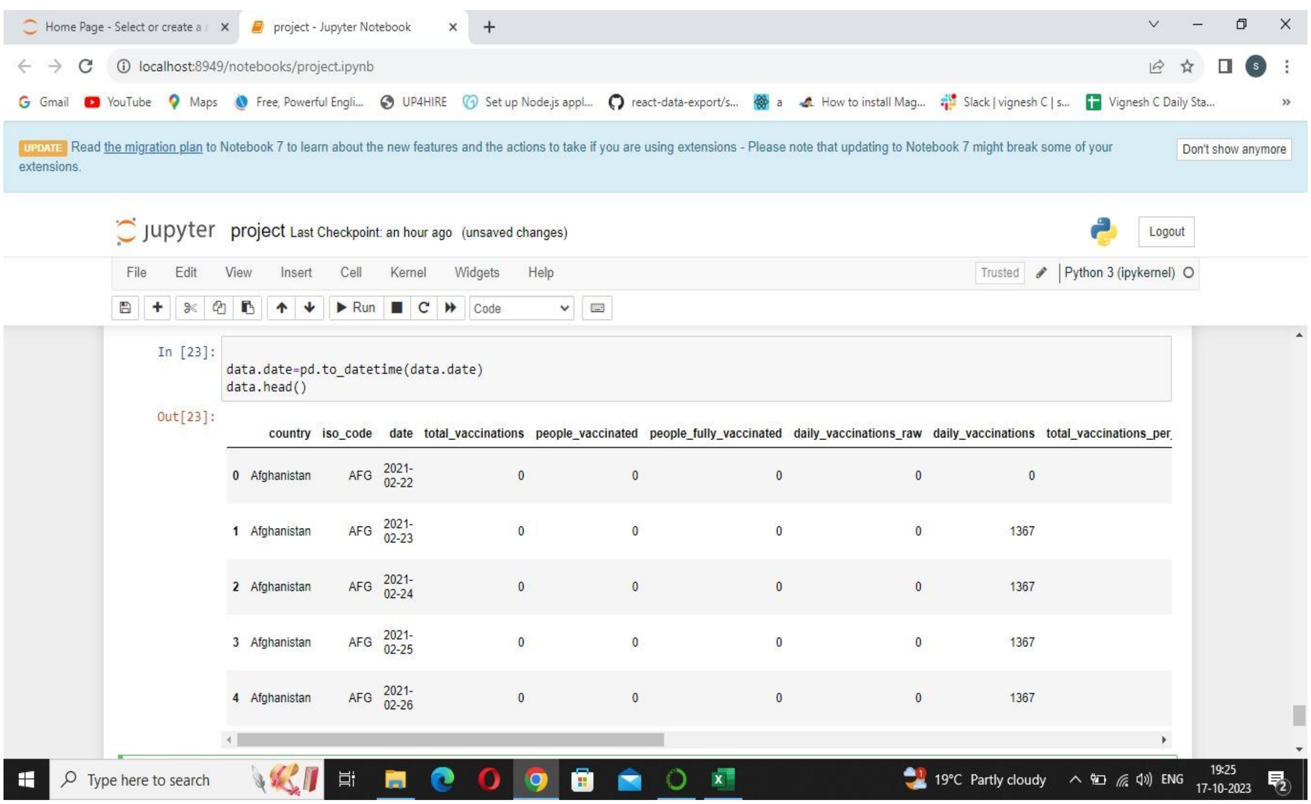




```
In [22]: data['year']=date[0]
data['month']=date[1]
data['day']=date[2]
data.year=pd.to_numeric(data.year)
data.month=pd.to_numeric(data.month)
data.day=pd.to_numeric(data.day)
data.date=pd.to_datetime(data.date)
data.head()
```

Out[22]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_
0	Afghanistan	AFG	2021-02-22	0	0	0	0	0	
1	Afghanistan	AFG	2021-02-23	0	0	0	0	1367	
2	Afghanistan	AFG	2021-02-24	0	0	0	0	1367	
3	Afghanistan	AFG	2021-02-25	0	0	0	0	1367	

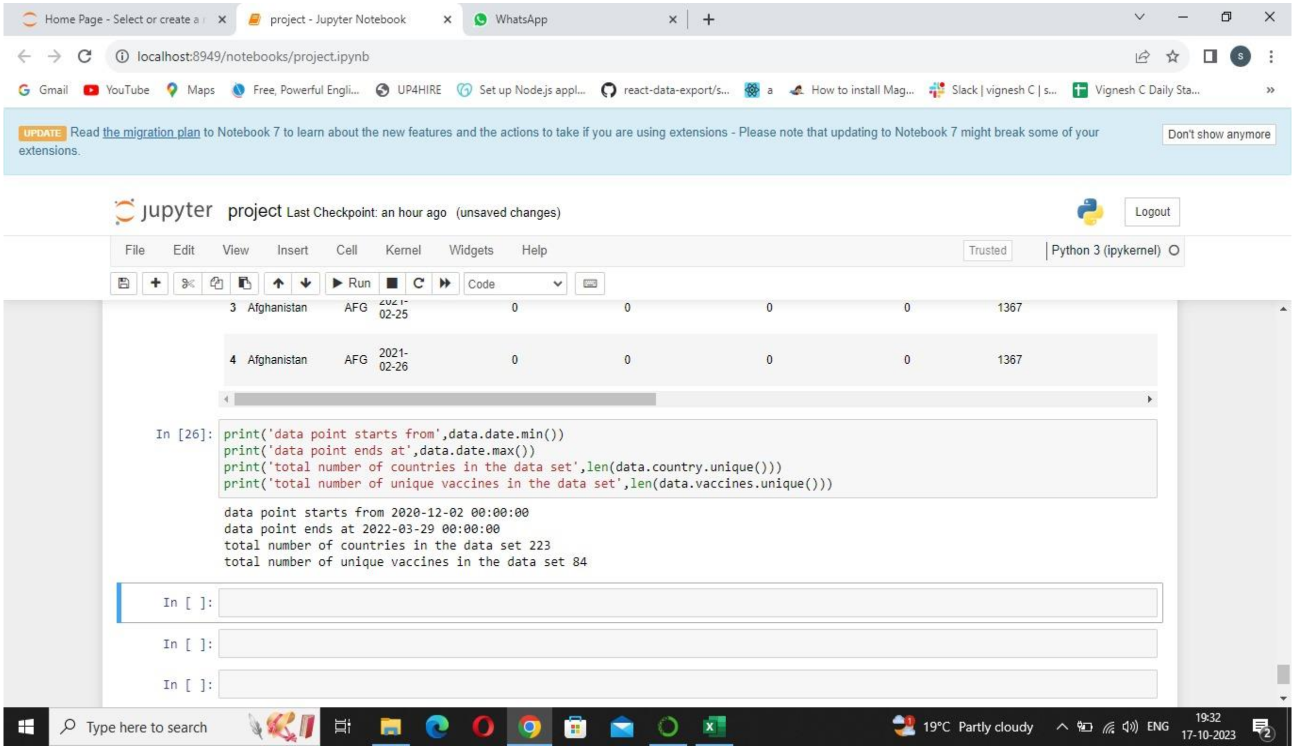


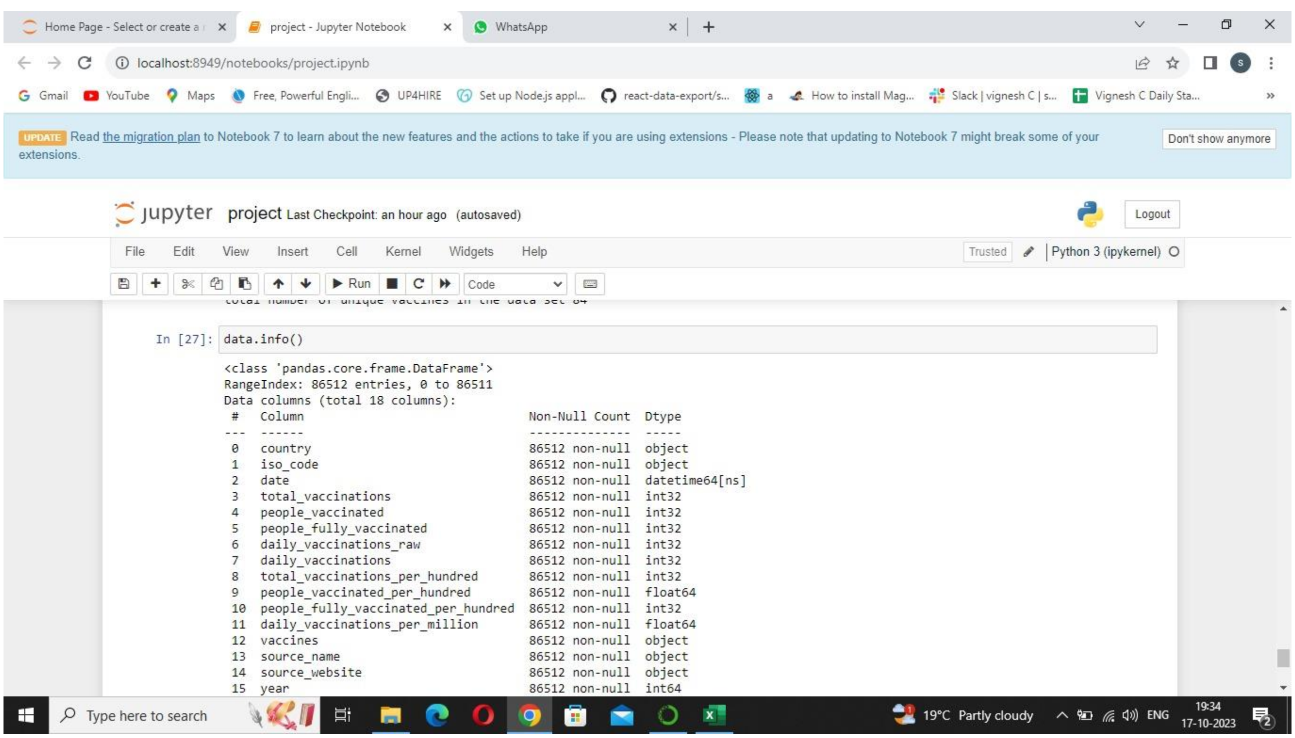
```
In [23]: 
data.date=pd.to_datetime(data.date)
data.head()
```

Out[23]:

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_
0	Afghanistan	AFG	2021-02-22	0	0	0	0	0	
1	Afghanistan	AFG	2021-02-23	0	0	0	0	1367	
2	Afghanistan	AFG	2021-02-24	0	0	0	0	1367	
3	Afghanistan	AFG	2021-02-25	0	0	0	0	1367	
4	Afghanistan	AFG	2021-02-26	0	0	0	0	1367	

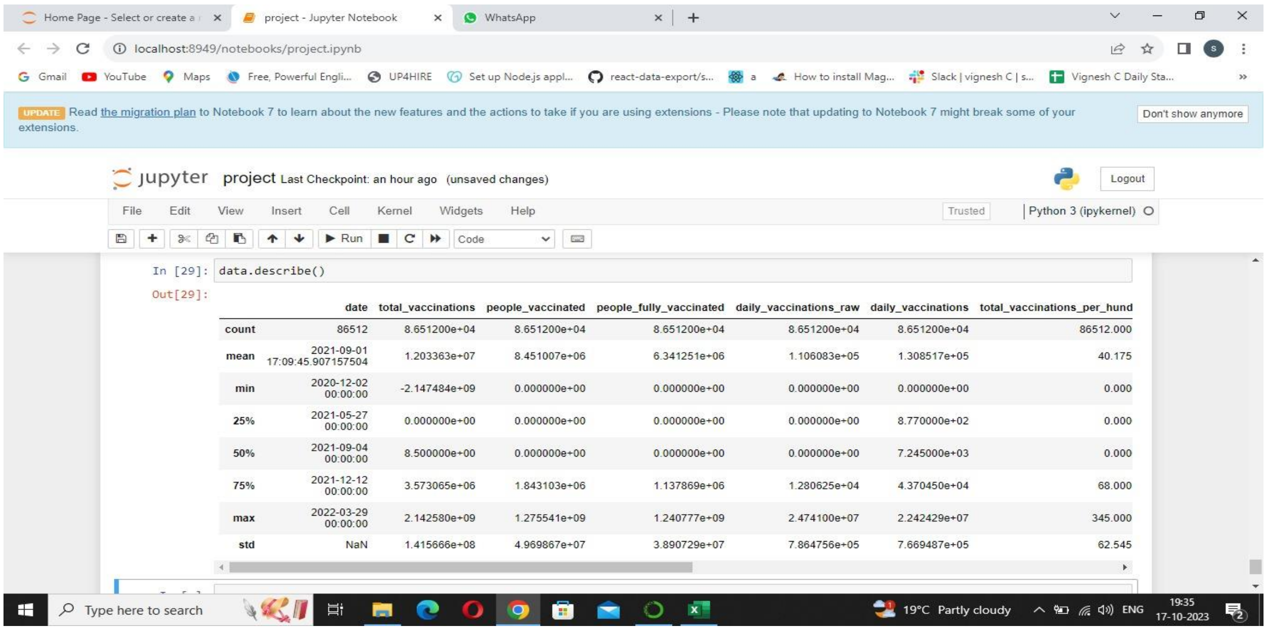
Some detailed features to specify the details using the below code,





Using Data visualization we are going to draw some visuals to get insights from dataset,

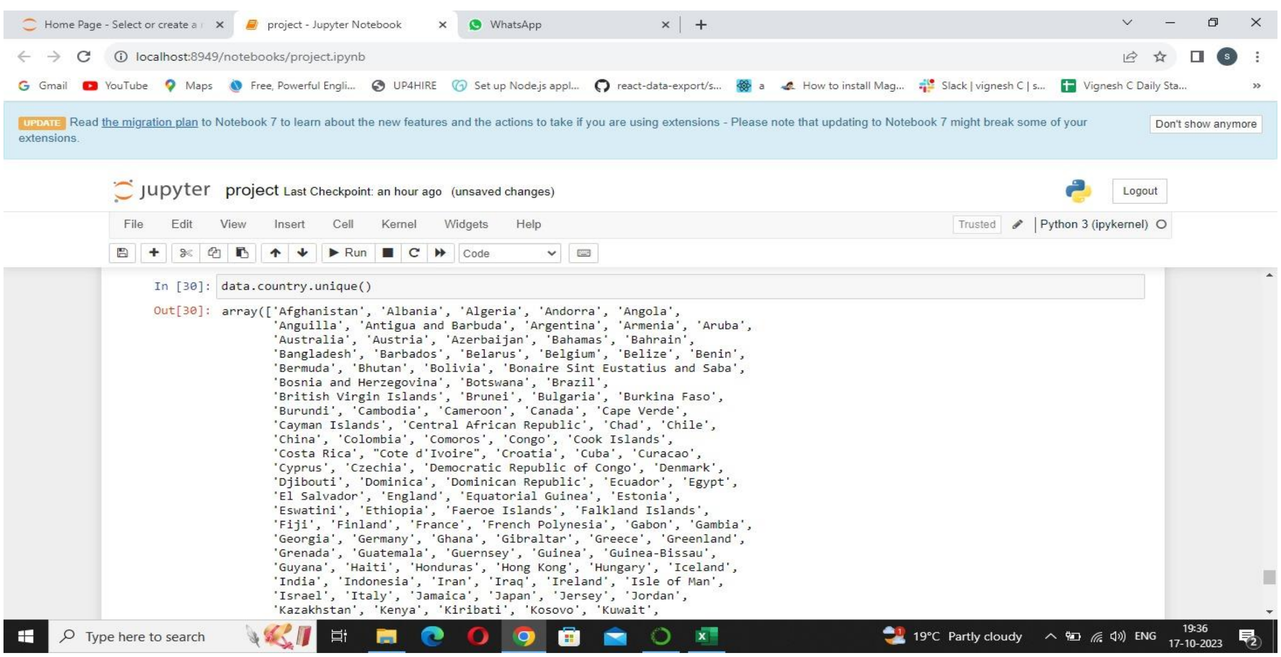
Describe () function is used to get the statistics of each feature in dataset to get count, min, max, standard deviation, median, etc.,



```
In [29]: data.describe()
Out[29]:
```

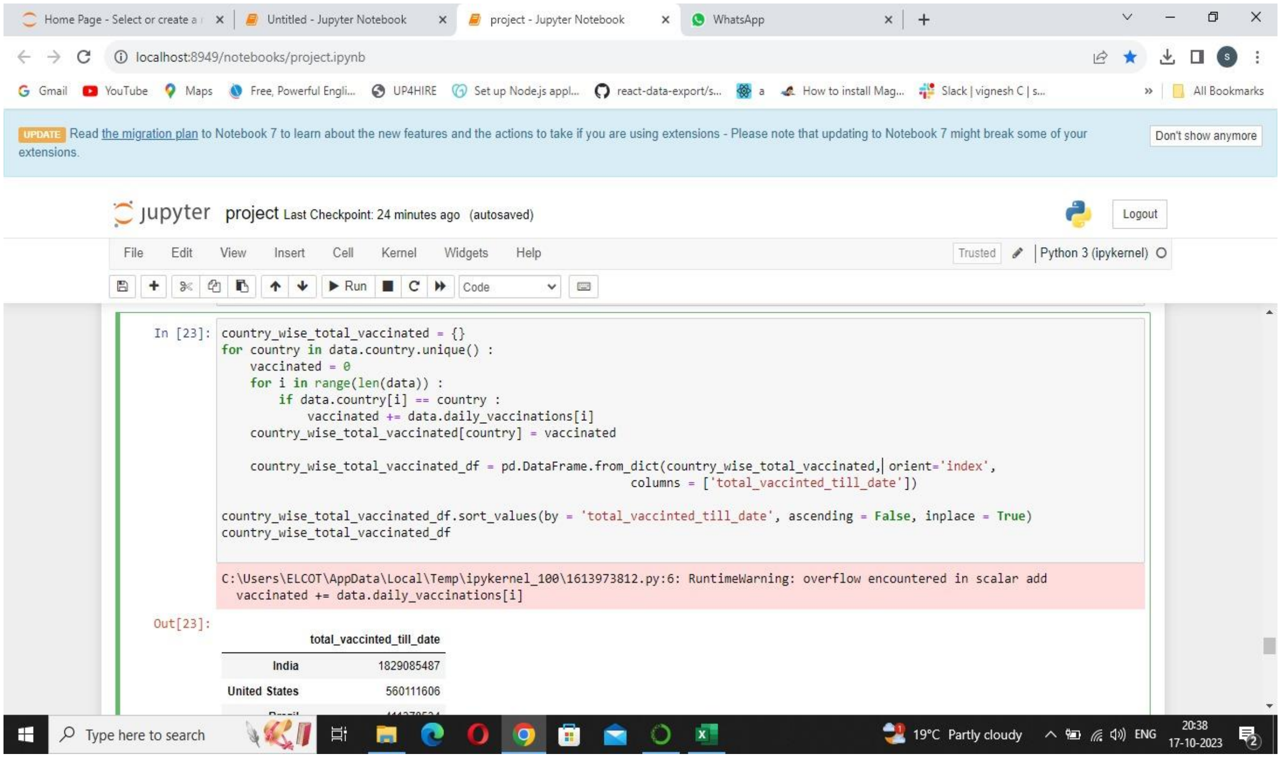
	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hund
count	86512	8.651200e+04	8.651200e+04	8.651200e+04	8.651200e+04	8.651200e+04	86512.000
mean	2021-09-01 17:09:45.907157504	1.203363e+07	8.451007e+06	6.341251e+06	1.106083e+05	1.308517e+05	40.175
min	2020-12-02 00:00:00	-2.147484e+09	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000
25%	2021-05-27 00:00:00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	8.770000e+02	0.000
50%	2021-09-04 00:00:00	8.500000e+00	0.000000e+00	0.000000e+00	0.000000e+00	7.245000e+03	0.000
75%	2021-12-12 00:00:00	3.573065e+06	1.843103e+06	1.137869e+06	1.280625e+04	4.370450e+04	68.000
max	2022-03-29 00:00:00	2.142580e+09	1.275541e+09	1.240777e+09	2.474100e+07	2.242429e+07	345.000
std	NaN	1.415666e+08	4.969867e+07	3.890729e+07	7.864756e+05	7.669487e+05	62.545

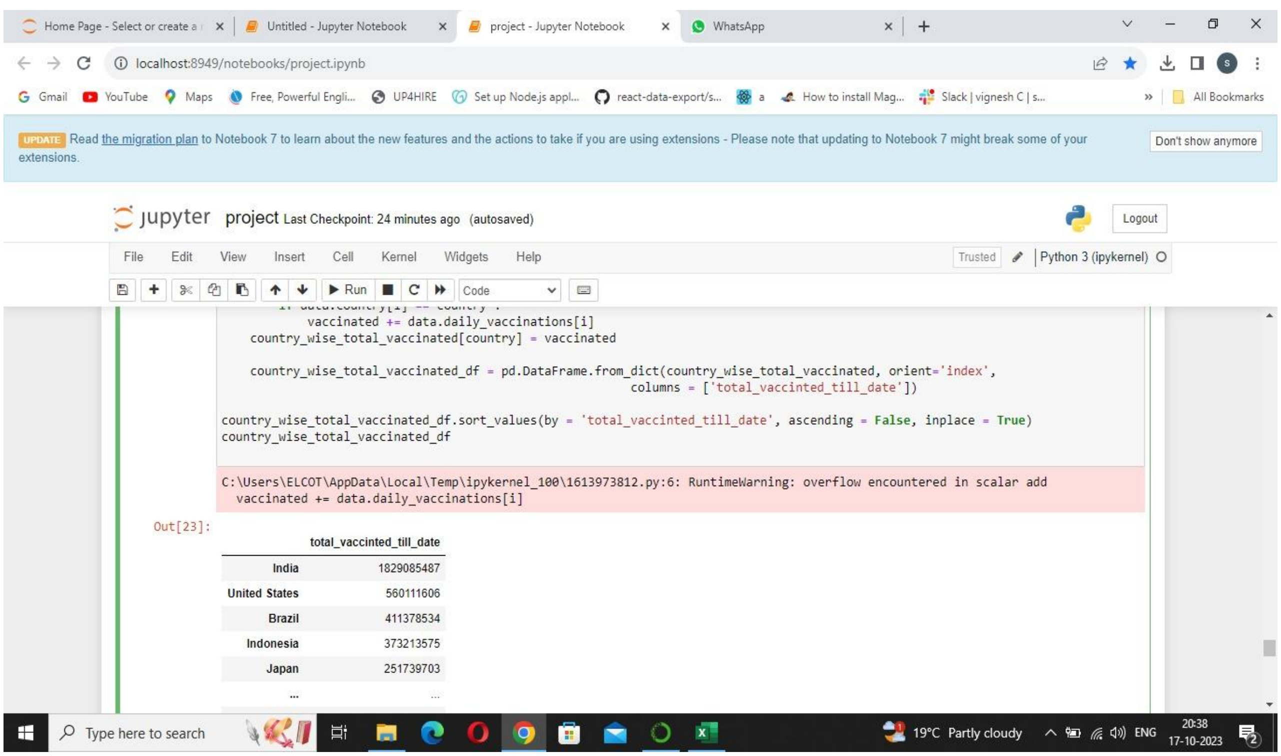
Unique () function helps to get unique values,

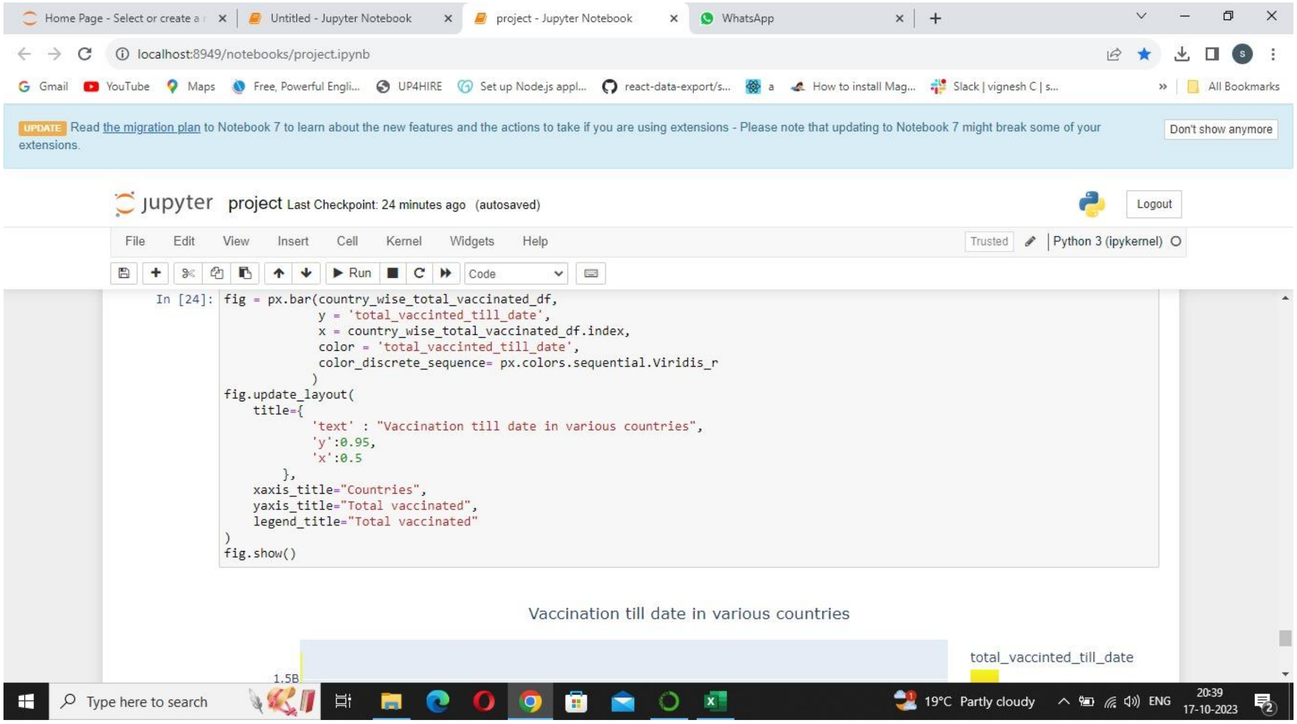


```
In [30]: data.country.unique()
Out[30]: array(['Afghanistan', 'Albania', 'Algeria', 'Andorra', 'Angola',
       'Anguilla', 'Antigua and Barbuda', 'Argentina', 'Armenia', 'Aruba',
       'Australia', 'Austria', 'Azerbaijan', 'Bahamas', 'Bahrain',
       'Bangladesh', 'Barbados', 'Belarus', 'Belgium', 'Belize', 'Benin',
       'Bermuda', 'Bhutan', 'Bolivia', 'Bonaire Sint Eustatius and Saba',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil',
       'British Virgin Islands', 'Brunei', 'Bulgaria', 'Burkina Faso',
       'Burundi', 'Cambodia', 'Cameroon', 'Canada', 'Cape Verde',
       'Cayman Islands', 'Central African Republic', 'Chad', 'Chile',
       'China', 'Colombia', 'Comoros', 'Congo', 'Cook Islands',
       'Costa Rica', "Cote d'Ivoire", 'Croatia', 'Cuba', 'Curacao',
       'Cyprus', 'Czechia', 'Democratic Republic of Congo', 'Denmark',
       'Djibouti', 'Dominica', 'Dominican Republic', 'Ecuador', 'Egypt',
       'El Salvador', 'England', 'Equatorial Guinea', 'Estonia',
       'Eswatini', 'Ethiopia', 'Faeroe Islands', 'Falkland Islands',
       'Fiji', 'Finland', 'France', 'French Polynesia', 'Gabon', 'Gambia',
       'Georgia', 'Germany', 'Ghana', 'Gibraltar', 'Greece', 'Greenland',
       'Grenada', 'Guatemala', 'Guernsey', 'Guinea', 'Guinea-Bissau',
       'Guyana', 'Haiti', 'Honduras', 'Hong Kong', 'Hungary', 'Iceland',
       'India', 'Indonesia', 'Iran', 'Iraq', 'Ireland', 'Isle of Man',
       'Israel', 'Italy', 'Jamaica', 'Japan', 'Jersey', 'Jordan',
       'Kazakhstan', 'Kenya', 'Kiribati', 'Kosovo', 'Kuwait',
```

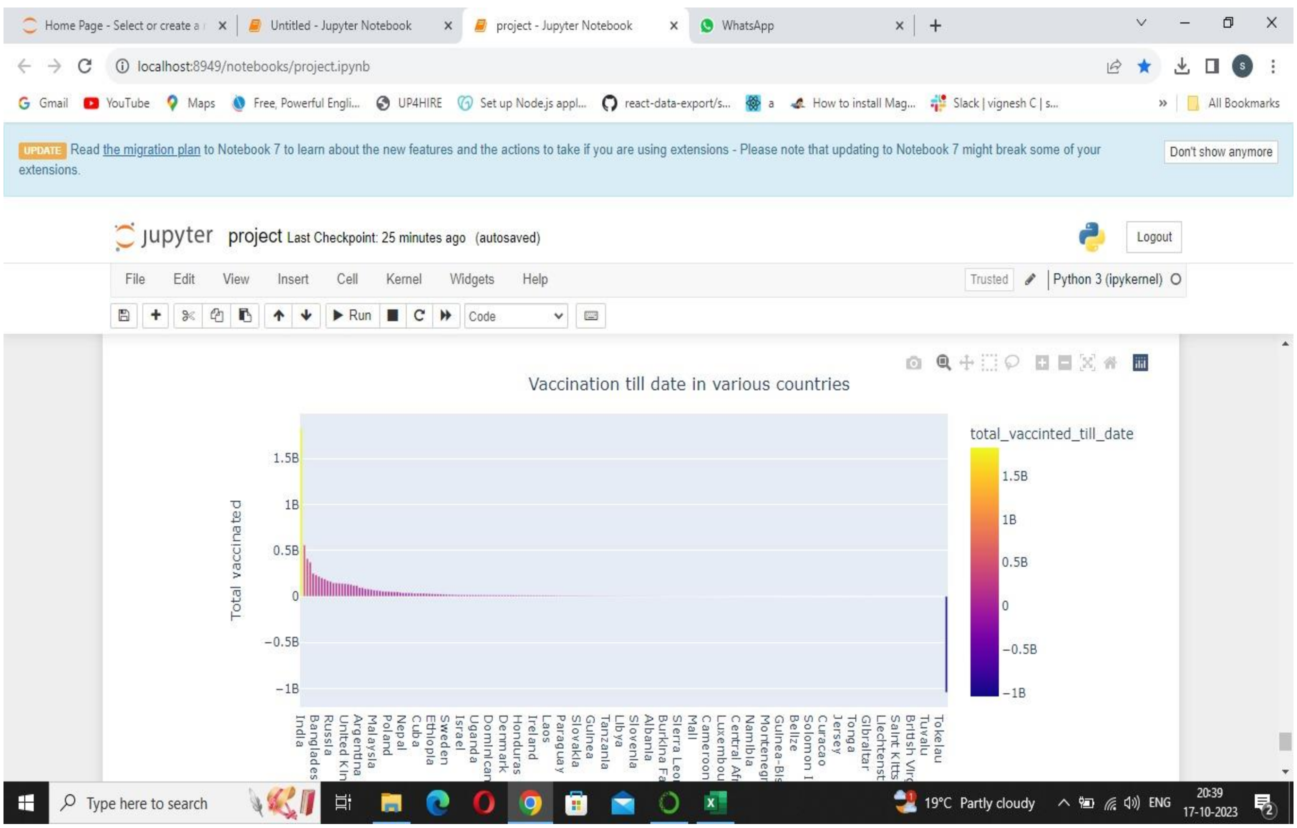
To see how many total vaccines have been used in each country using the code below,





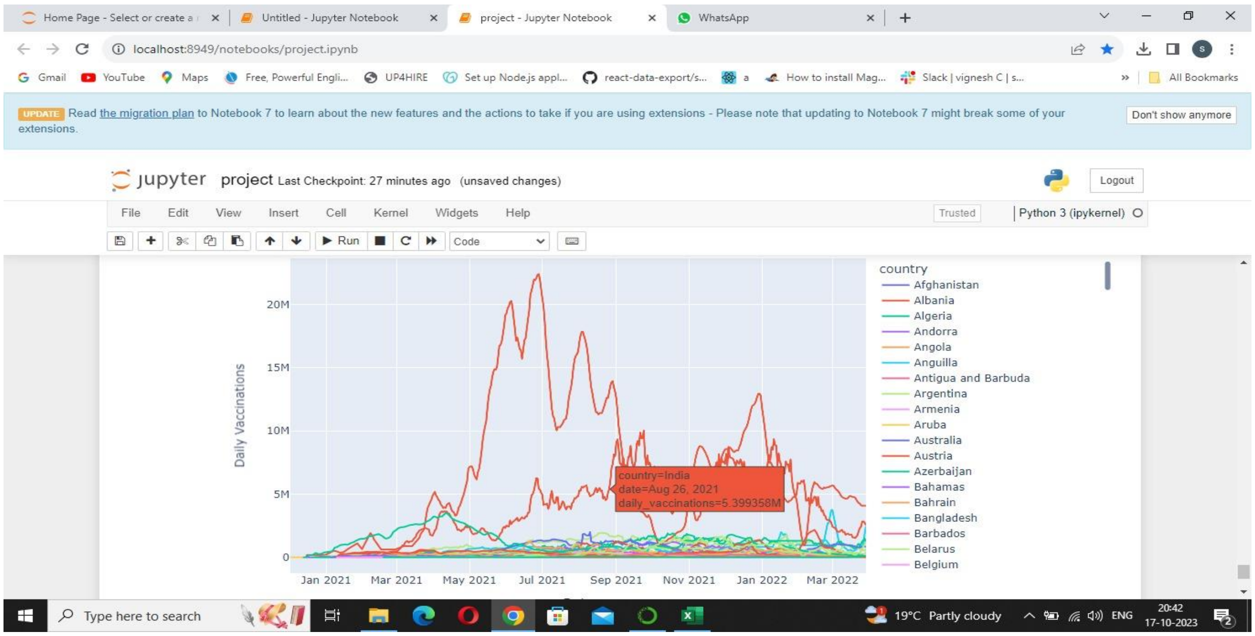


localhost:8949/notebooks/project.ipynb
jupyter project Last Checkpoint: 24 minutes ago (autosaved)
In [24]:
fig = px.bar(country_wise_total_vaccinated_df,
             y = 'total_vaccinted_till_date',
             x = country_wise_total_vaccinated_df.index,
             color = 'total_vaccinted_till_date',
             color_discrete_sequence= px.colors.sequential.Viridis_r
)
fig.update_layout(
    title={
            'text' : "Vaccination till date in various countries",
            'y':0.95,
            'x':0.5
        },
    xaxis_title="Countries",
    yaxis_title="Total vaccinated",
    legend_title="Total vaccinated"
)
fig.show()
Vaccination till date in various countries
total_vaccinted_till_date

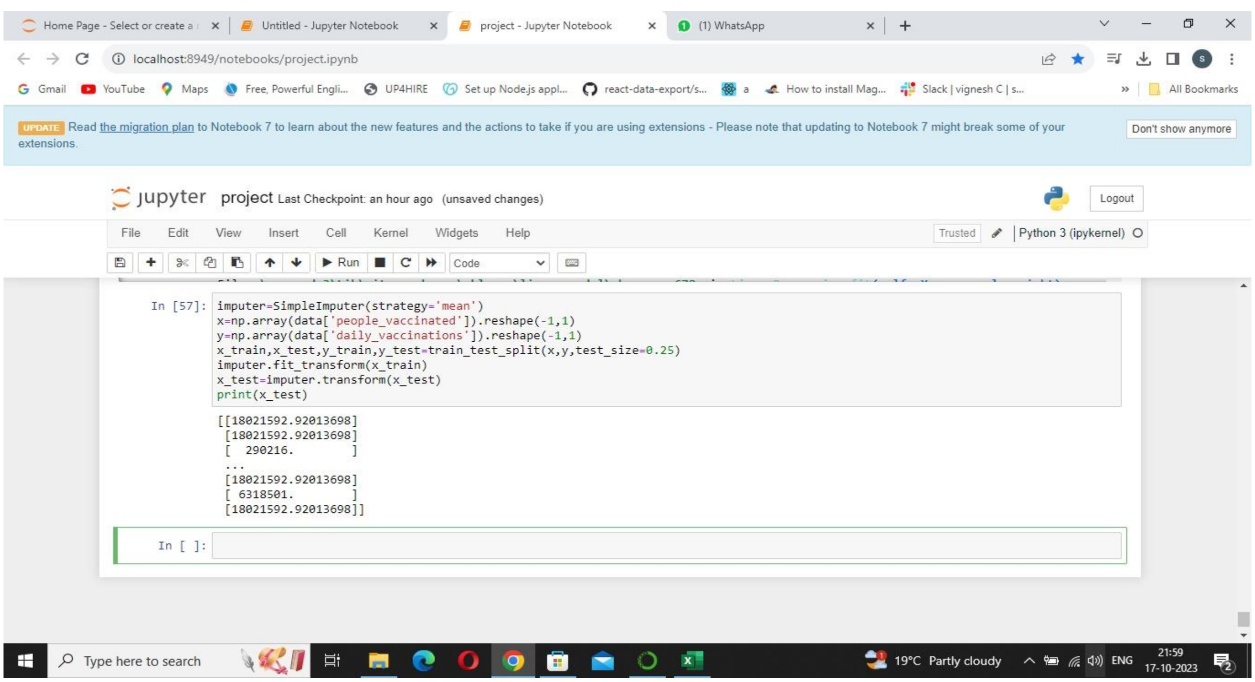


localhost:8949/notebooks/project.ipynb
jupyter project Last Checkpoint: 25 minutes ago (autosaved)
Vaccination till date in various countries
Total vaccinated
1.5B
1B
0.5B
0
−0.5B
−1B
total_vaccinted_till_date
India
Russia
Argentina
Poland
Cuba
Sweden
Uganda
Denmark
Ireland
Paraguay
Guinea
Libya
Albania
Sierra Leo
Cameroon
Central Afr
Montenegr
Belize
Curacao
Tonga
Liechtenst
British Virg
Tokelau

To draw a line plot where x-axis is Date and the y-axis is daily_vaccination using the in the image,

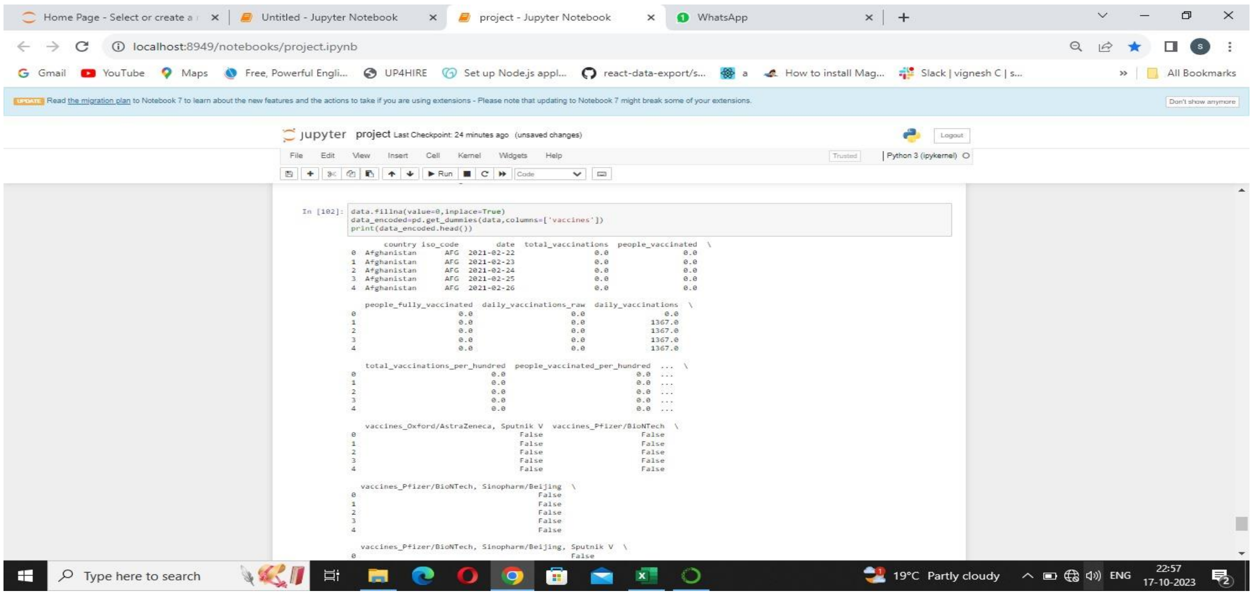


Now, using the sklearn.preprocessing library contains class called imputer, helps in missing data by using the below:



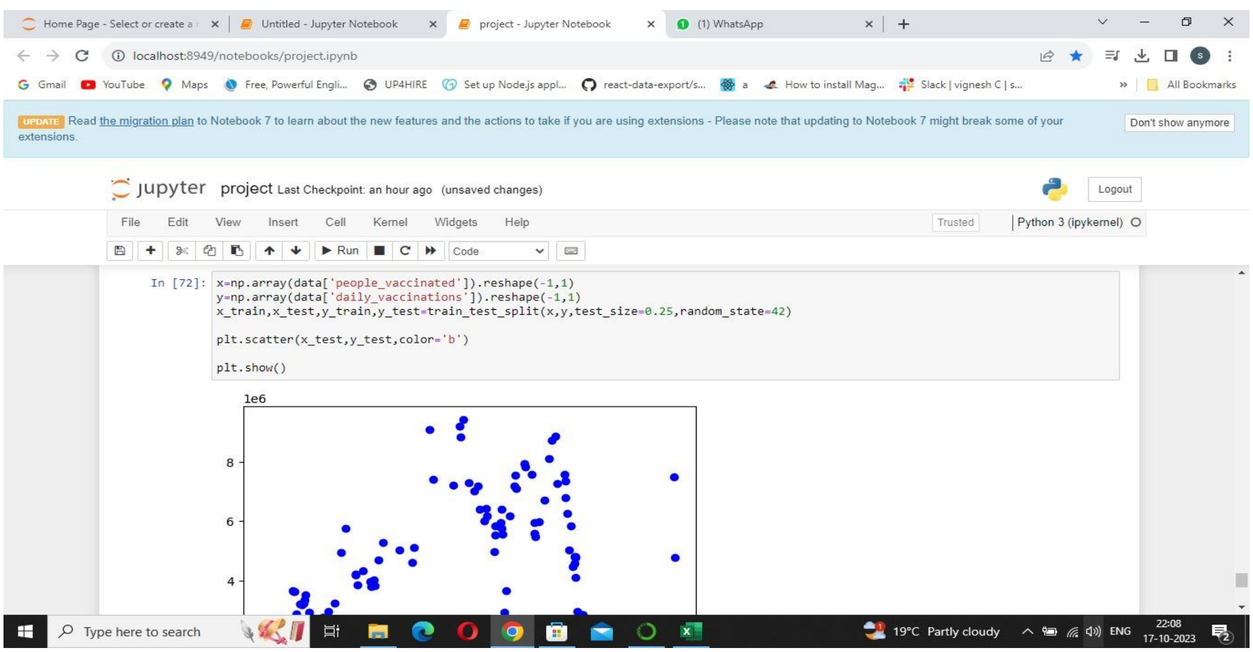
Encoding categorical data(one-hot encoding)

One-hot encoding is a technique used to convert categorical data into a numerical format that machine learning algorithms can work with. Here's how you can perform one-hot encoding in Python, assuming you have a dataset with categorical variables:

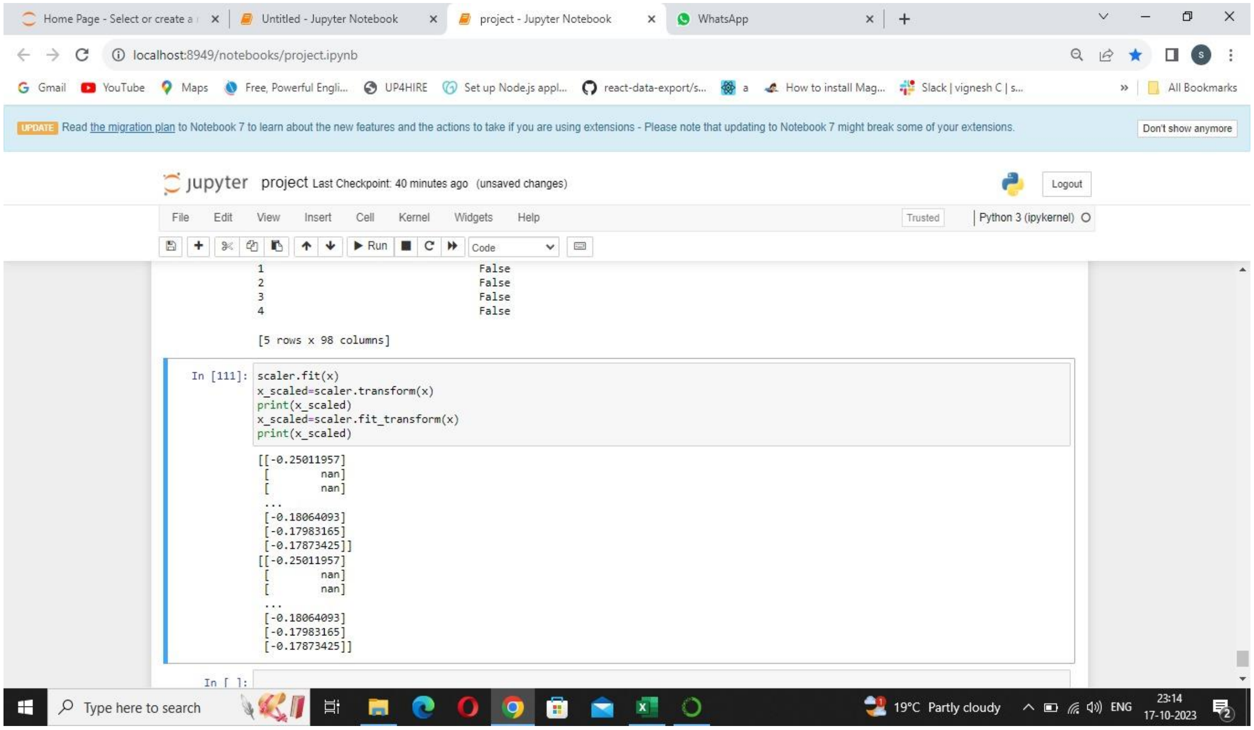


Splitting the data set into test set and training set

By using, Import train_test_split



Feature Scaling



STATISTICAL ANALYSIS

Statistical analysis for COVID-19 vaccine analysis data can provide more in-depth insights and support decision-making. Here are some statistical techniques and analyses you can perform:

1. Descriptive Statistics: *Calculate summary statistics (mean, median, standard deviation) for key vaccine-related variables, such as vaccination rates, doses administered, and adverse events.*

2. Hypothesis Testing: *Use statistical tests like t-tests or ANOVA to compare vaccination rates or outcomes between different groups (e.g., regions, age groups, vaccine types).*

- ***DESCRIPTIVE STATISTICS***
  Descriptive statistics provide a summary of your data, including measures like mean, median, and standard deviation.

mean_vaccination = data['total_vaccinations'].mean()

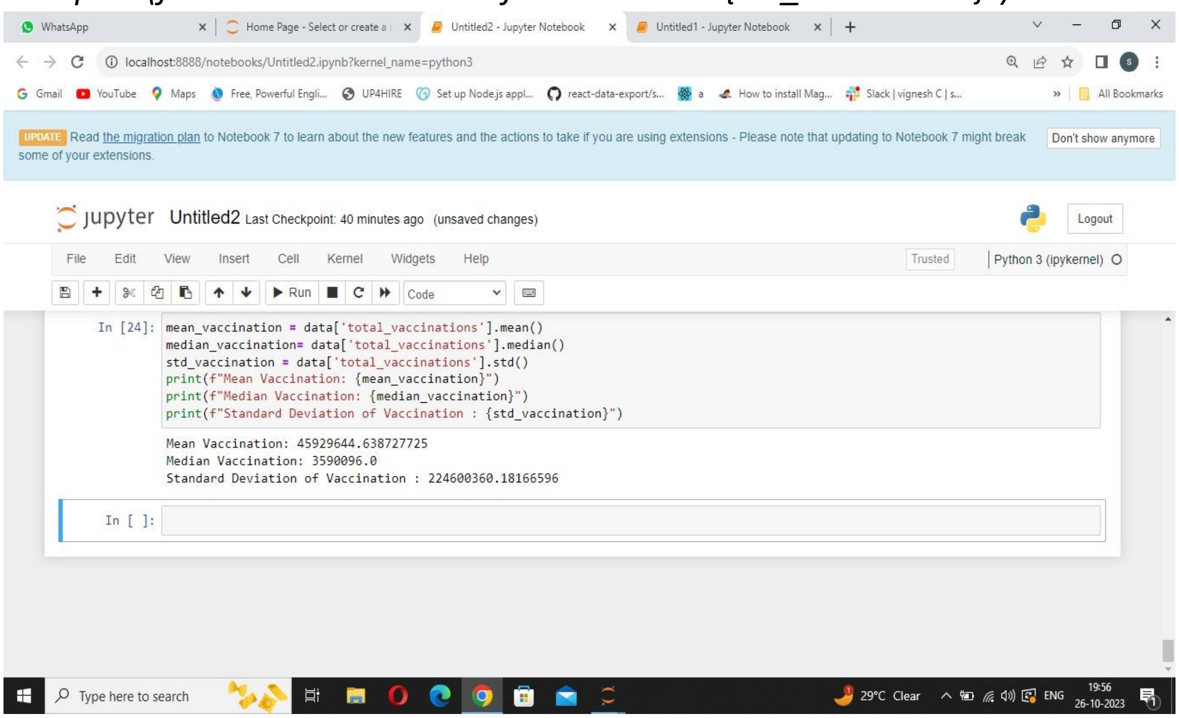
median_vaccination= data['total_vaccinations'].median()

std_vaccination = data['total_vaccinations'].std()

print(f"Mean Vaccination: {mean_vaccination}")

print(f"Median Vaccination: {median_vaccination}")

print(f"Standard Deviation of Vaccination : {std_vaccination}")



- mean_vaccination = data['people_vaccinated'].mean()

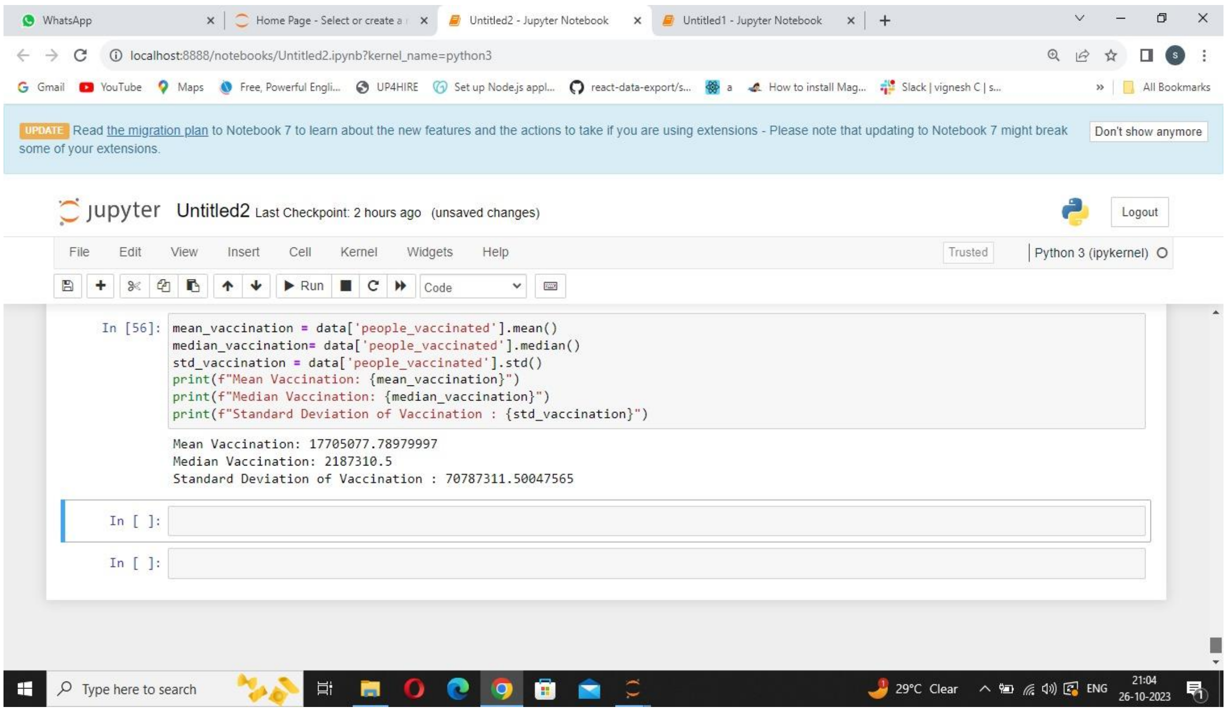
  median_vaccination= data['people_vaccinated'].median()

  std_vaccination = data['people_vaccinated'].std()

print(f"Mean Vaccination: {mean_vaccination}")

print(f"Median Vaccination: {median_vaccination}")

print(f"Standard Deviation of Vaccination : {std_vaccination}")



- *mean_vaccination = data['daily_vaccinations'].mean()*

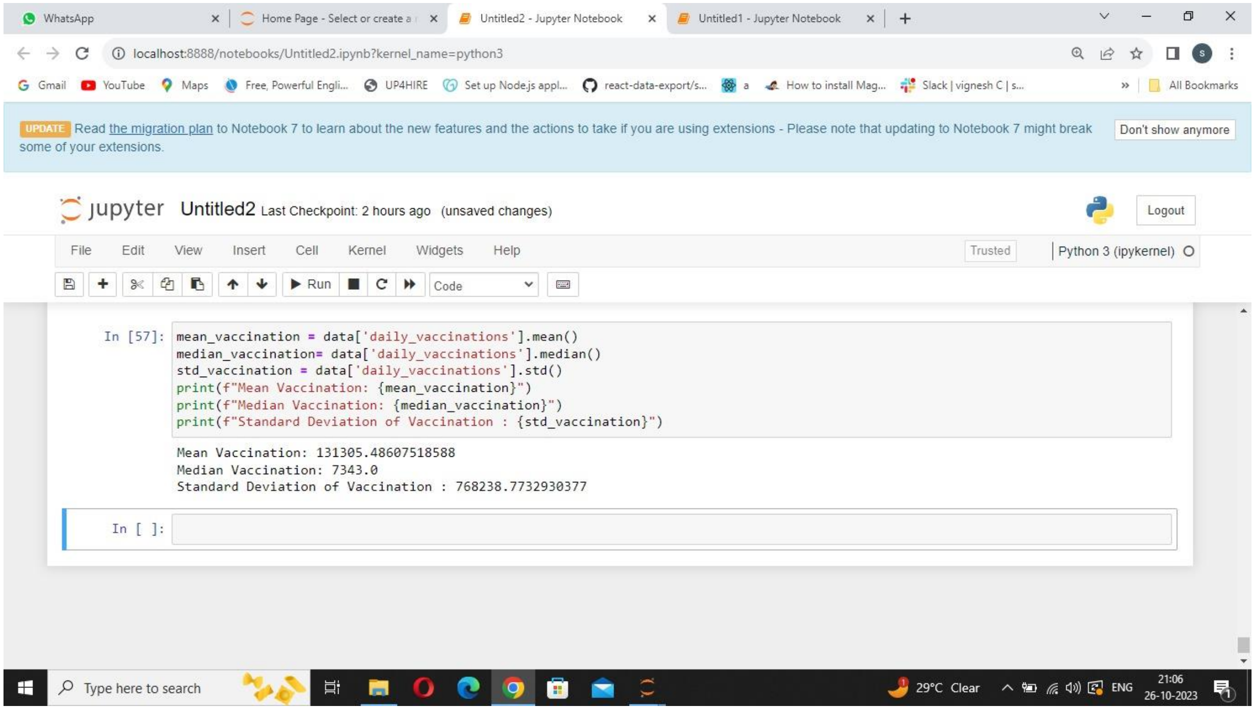
  median_vaccination= data['daily_vaccinations'].median()

  std_vaccination = data['daily_vaccinations'].std()

  print(f"Mean Vaccination: {mean_vaccination}")

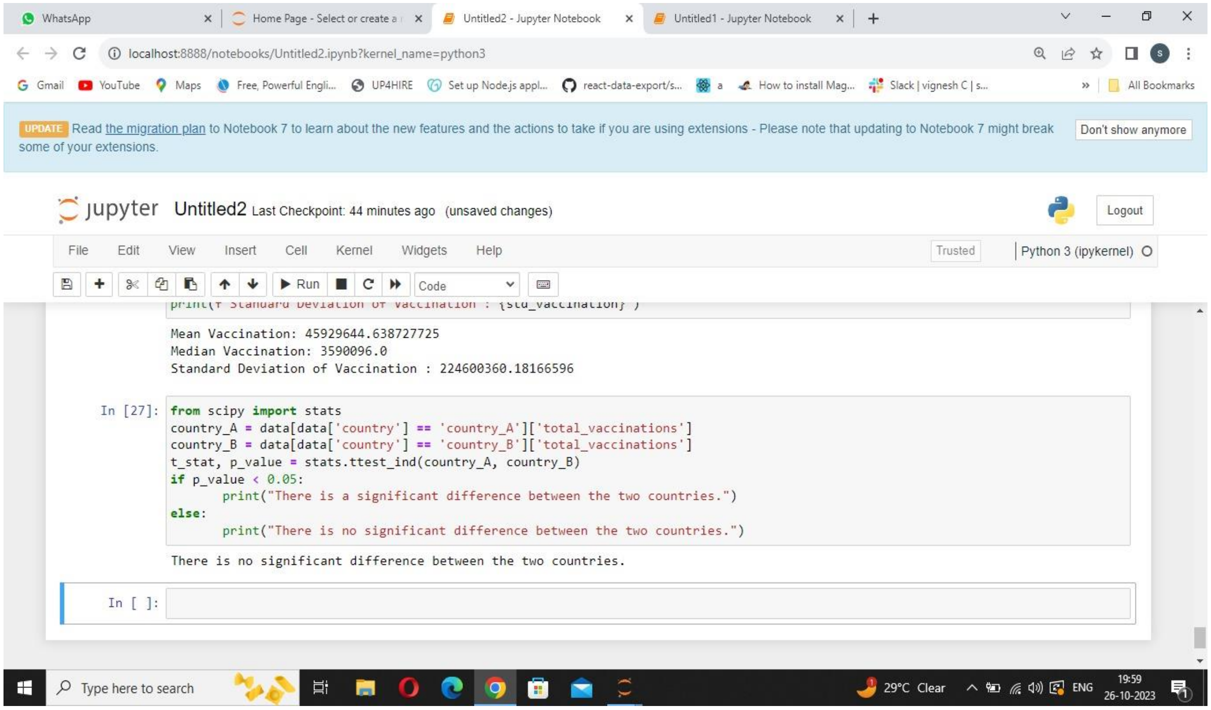
  print(f"Median Vaccination: {median_vaccination}")

  print(f"Standard Deviation of Vaccination : {std_vaccination}")



- ***HYPOTHESIS TESTING***

```
from scipy import stats
country_A = data[data['country'] == 'country_A']['total_vaccinations']
country_B = data[data['country'] == 'country_B']['total_vaccinations']
t_stat, p_value = stats.ttest_ind(country_A, country_B)
if p_value < 0.05:
    print("There is a significant difference between the two countries.")
else:
    print("There is no significant difference between the two countries.")
```



VISUALISATION

Visualizations play a crucial role in understanding and presenting COVID-19 vaccine analysis. Here are some types of visualizations you can use:

1. Bar Charts: *Show the distribution of different vaccine types administered in a region or over time.*

2. Line Charts: *Display trends in vaccination rates over time, including first and second doses administered.*

3. Area Charts: *Visualize the cumulative number of vaccines administered over time to track progress.*

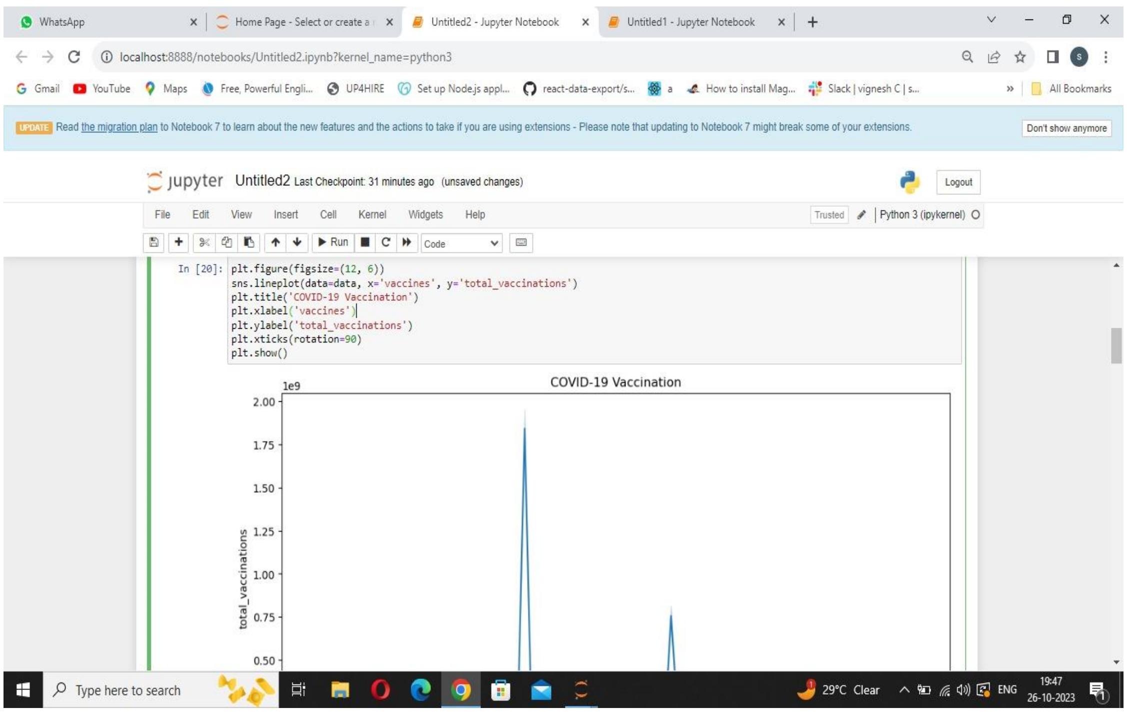
4. Stacked Bar Charts: *Illustrate the breakdown of vaccine distribution by age group or gender.*

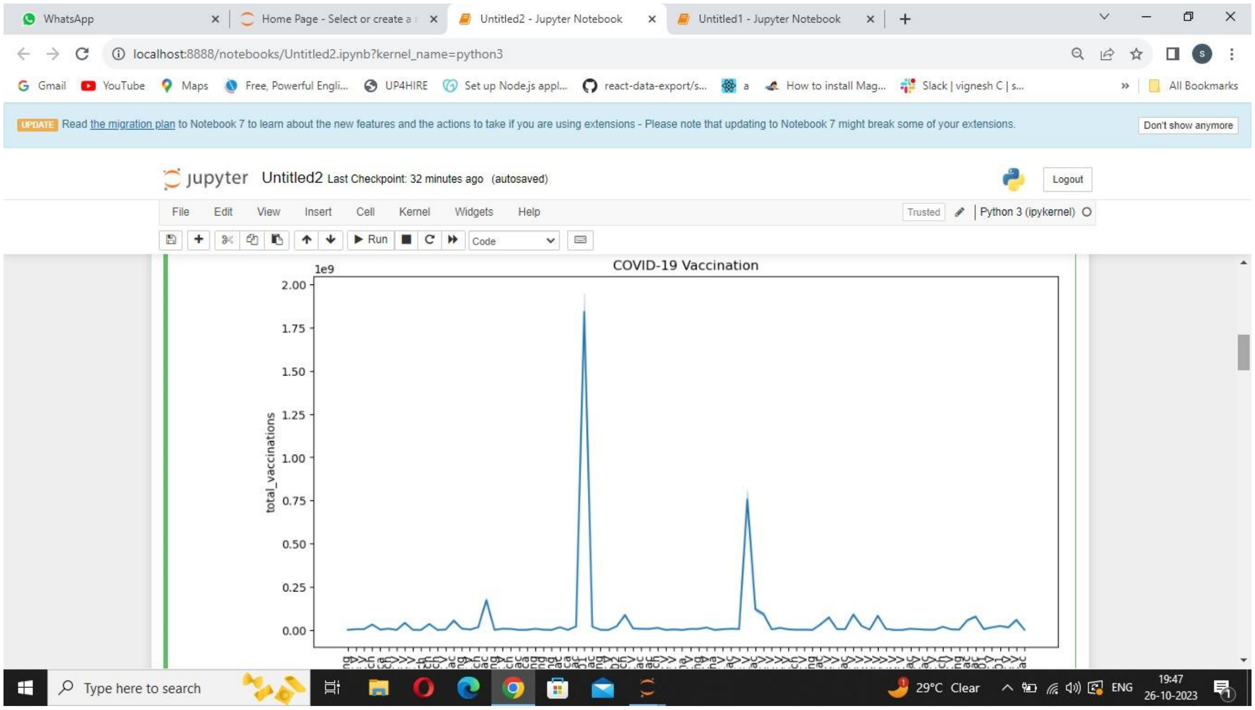
5. Heatmaps: *Depict vaccination rates across regions or countries using color gradients.*

6. Choropleth Maps: *Show vaccination coverage by shading areas on a map, with darker colors indicating higher coverage.*

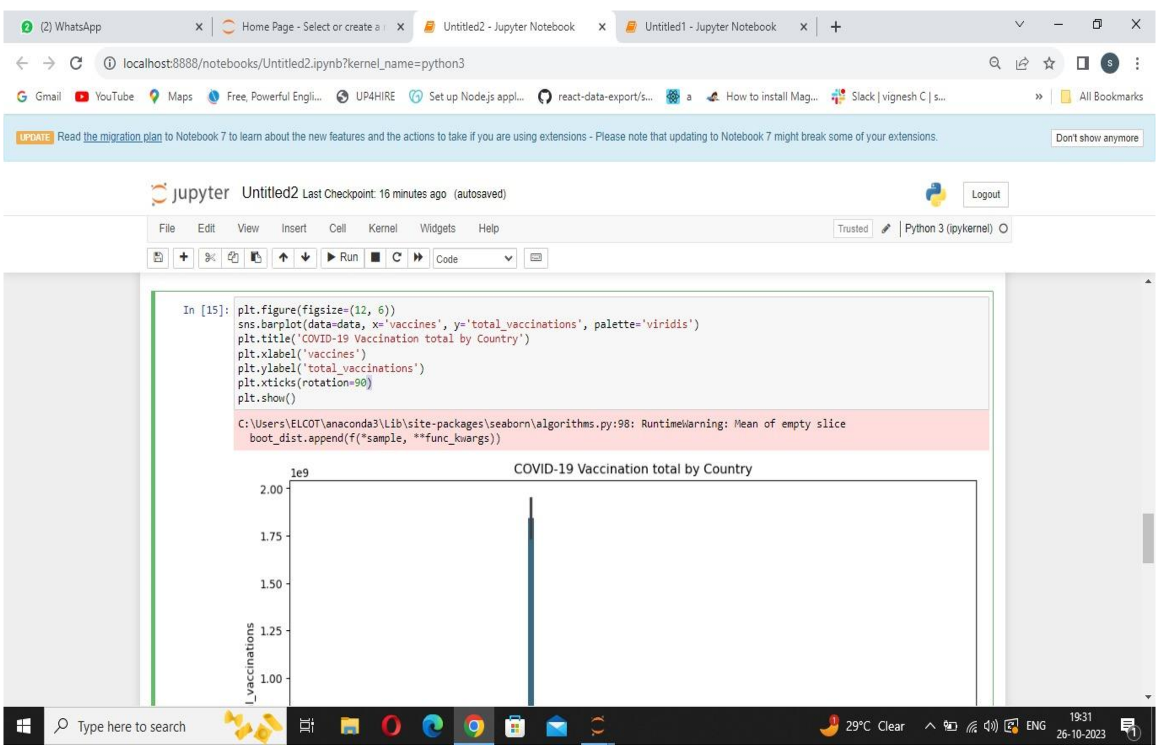
7. Scatter Plots: *Explore correlations between vaccination rates and variables like COVID-19 cases, GDP, or healthcare infrastructure.*

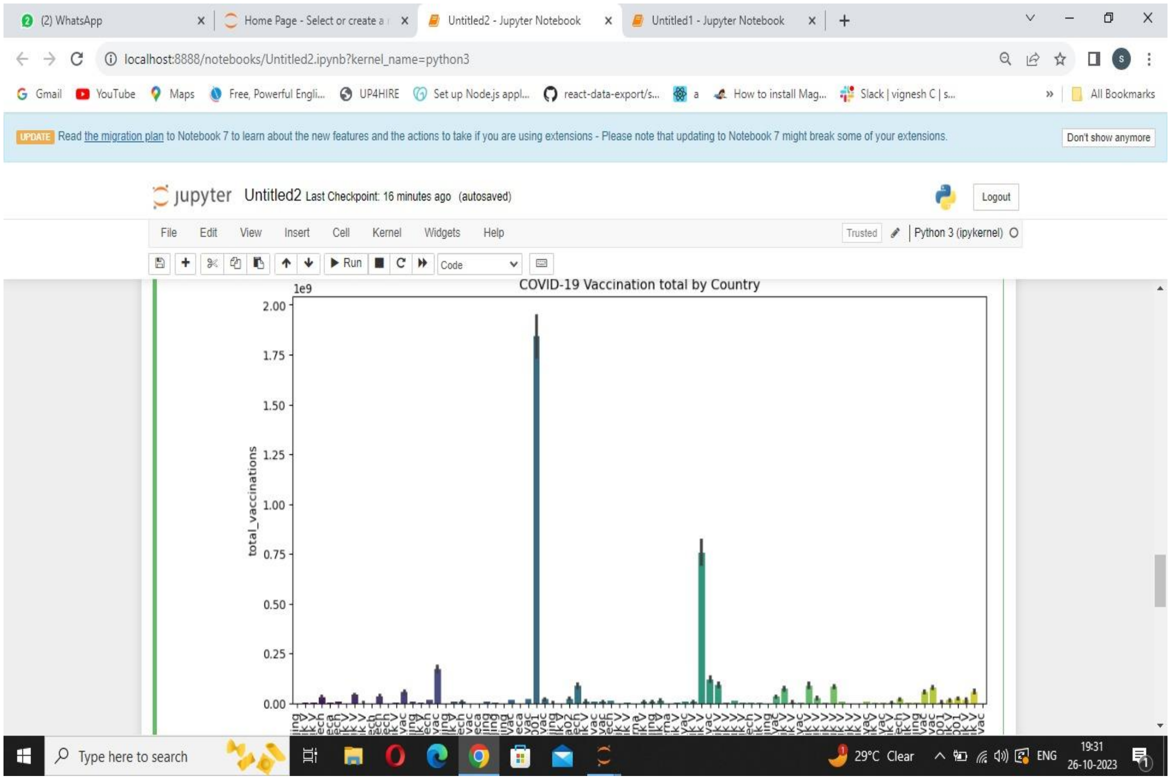
- *create a line chart to utilize the progress of vaccination*



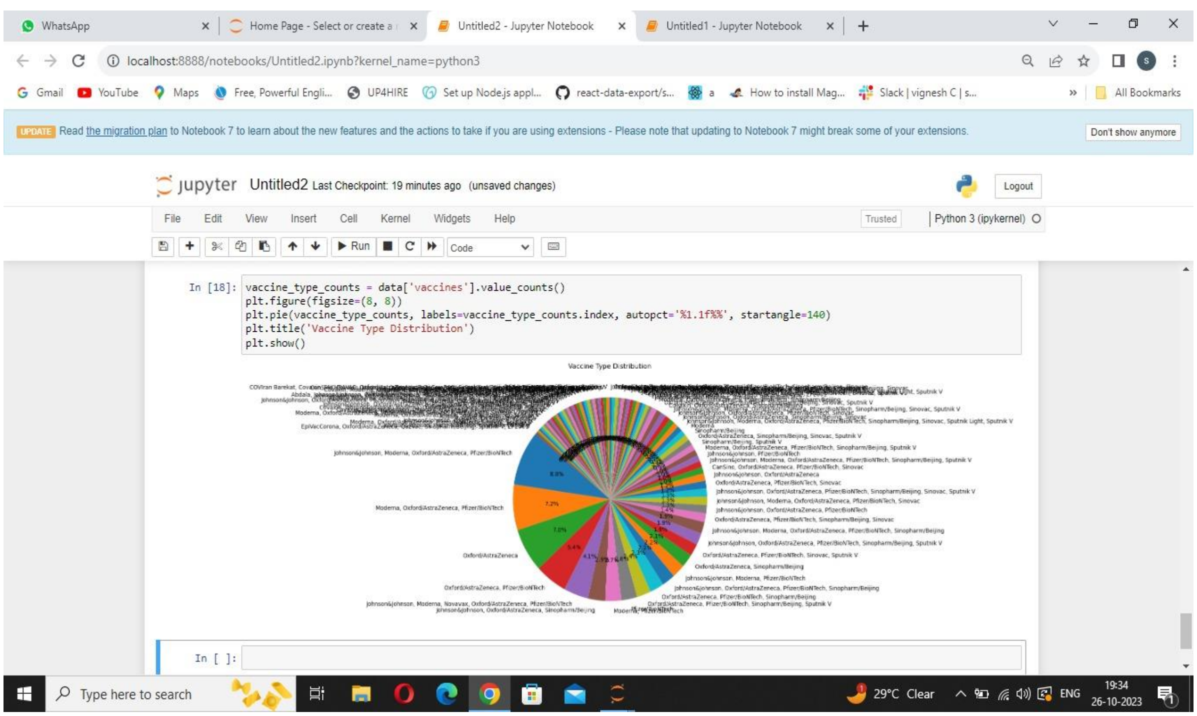


- ***create a bar chart between total by country and total_vaccination***





- ***Create a pie chart to show the distribution of vaccine types***



CONCLUSION

Thus, project covid-19 vaccines Analysis was completed successfully.