**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in this Minor thesis titled

**“VACCINE USAGE (INDIA) DATA ANALYSIS”** in fulfilment of the requirement for the degree of BACHELOR OF COMPUTER APPLICATION (specialization in Data Science) and submitted to

“**SATYUG DARSHAN INSTITUTE OF ENGINEERING AND TECHNOLOGY**”*,* is an authentic record of my own work carried out under the supervision of **Mr.**

**ANKIT MISHRA.**

The work contained in this thesis has not been submitted to any other University or Institute for the award of any other degree or diploma by me.

MANU SINGHAL SWARNIM PRATEEK

BCA(DS-23/026) BCA (DS-23/052)

# 

# ACKNOWLEDGEMENT

It is with deep sense of gratitude and reverence that I express my sincere thanks to my supervisor **Mr. ANKIT MISHRA** for her guidance, encouragement, help and useful suggestions throughout. Her untiring and painstaking efforts, methodical approach and individual help made it possible for me to complete this work in time. I consider myself very fortunate for having been associated with the scholar like her. Her affection, guidance and scientific approach served a veritable incentive for completion of this work.

I shall ever remain indebted to the faculty members of **Satyug Darshan Institute of Engineering and Technology, Faridabad**.

Although it is not possible to name individually, I cannot forget my well-wishers at Satyug Darshan Institute of Engineering and Technology, Faridabad and outsiders for their persistent support and cooperation.

This acknowledgement will remain incomplete if I fail to express my deep sense of obligation to my parents and God for their consistent blessings and encouragement

(SWARNIM PRATEEK)

# 

# CERTIFICATE

**LIST OF ABBRIVATIONS**

* from datetime import datetime
* import matplotlib.pyplot as plt
* import plotly.express as px
* import numpy as np
* import pandas as pd
* import seaborn as sns

# CHAPTER 1: INTRODUCTION TO TOPIC

### Development of Vaccines

1. **Predictive Modeling**: ML algorithms can analyze vast amounts of biological data to predict which viral proteins are most likely to induce an immune response. This accelerates the identification of potential vaccine candidates.
2. **Simulation of Virus Behaviour**: ML models can simulate how viruses mutate and spread, helping researchers design vaccines that are effective against multiple strains.
3. **Optimization of Vaccine Formulations**: Machine learning can help optimize the components of a vaccine to ensure maximum efficacy and minimal side effects.

### Clinical Trials

1. **Patient Selection**: ML can be used to analyze patient data and select suitable candidates for clinical trials, ensuring diverse and representative samples.
2. **Monitoring and Analysis**: During clinical trials, ML can monitor patient responses in real time and analyze large datasets to identify patterns and outcomes more quickly and accurately than traditional methods.

### Manufacturing and Distribution

1. **Supply Chain Optimization**: Machine learning models can predict demand for vaccines and optimize supply chains to ensure vaccines are delivered where they are needed most efficiently.
2. **Quality Control**: ML algorithms can be used to monitor the manufacturing process and detect anomalies or defects in vaccine batches, ensuring high-quality products.

### Post-Distribution

1. **Surveillance and Reporting**: Machine learning can analyze data from various sources, such as electronic health records and social media, to detect and respond to adverse effects or outbreaks related to vaccines.
2. **Effectiveness Monitoring**: ML can help track the effectiveness of vaccines in real-world settings by analyzing health data to see how well vaccines prevent disease over time.

### Case Studies and Examples

1. **COVID-19 Vaccine Development**: During the COVID-19 pandemic, machine learning played a crucial role in the rapid development of vaccines. ML algorithms were used to predict the structure of the virus and identify potential vaccine targets.
2. **Flu Vaccine Optimization**: ML models are used annually to predict the most likely flu strains and inform the composition of the seasonal flu vaccine.

# CHAPTER 2: PYTHON FOR DATA SCIENCE

### 1.1. Introduction to Python

“Python is an interpreted, object-oriented, high-level programming language with dynamic semantics”. This language consists of mainly data structures which make it very easy for the data scientists to analyse the data very effectively. It does not only help in forecasting and analysis it also helps in connecting the two different languages. Two best features of this programming language is that it does not have any compilation step as compared to the other programming language in which compilation is done before the program is being executed and other one is the reuse of the code, it consist of modules and packages due to which we can use the previously written code anywhere in between the program whenever is required. There are multiple languages for example R, Java, SQL, MATLAB available in market which can be used to analyse and evaluate the data, but due to some outstanding features python is the most famous language used in the field of data science.

Python is mostly used and easy among all other programming languages.

**1.2 Understanding Standard Libraries Pandas, Numpy…..**

Libraries in Python

Python library is vast. There are built in functions in the library which are written in C language. This library provide access to system functionality such as file input output and that is not accessible to Python programmers. This modules and library provide solution to the many problems in programming.

Following are some Python libraries.

Matplotlib

Pandas

Numpy

### Matplotlib

”Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy”[11]. MATLAB provides an application that is used in graphical user interface tool kits. Another such library is pyplot which is almost same as MATLAB.

It is a library for 2D graphics, it finds its application in web application servers, graphical user interface toolkit and shell. Below is the example of a basic plot in python.

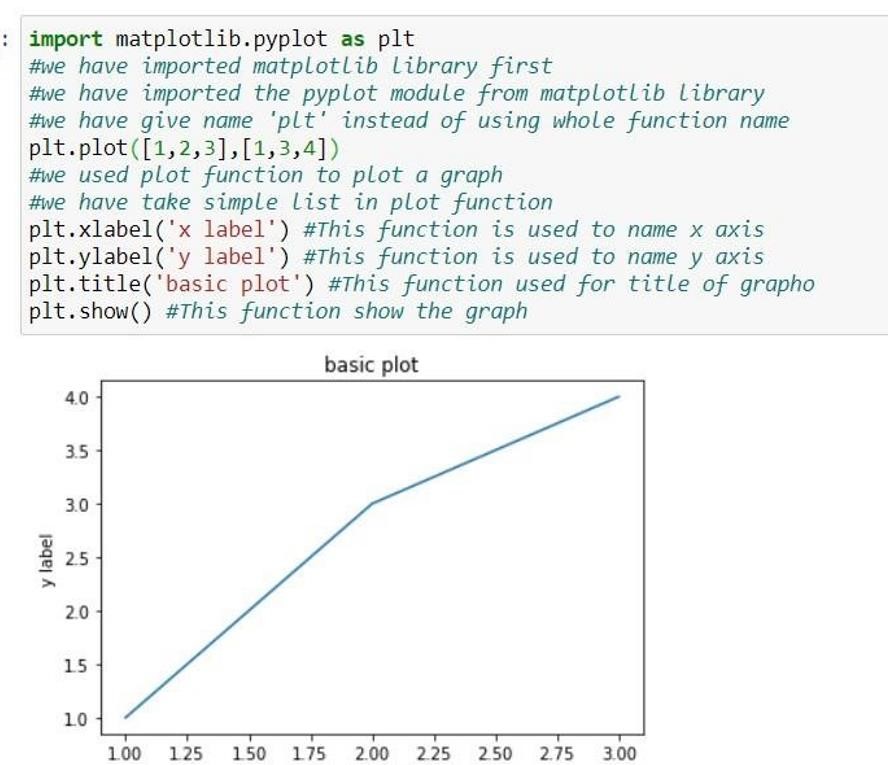


Figure 1.3.1: Matplotlib basic example

#### Pandas

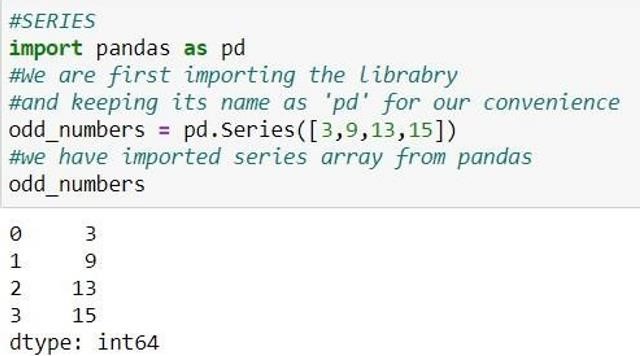
Pandas is also a library or a data analysis tool in python which is written in python programming language. It is mostly used for data analysis and data manipulation. It is also used for data structures and time series.

We can see the application of python in many fields such as - Economics, Recommendation Systems - Spotify, Netflix and Amazon, Stock Prediction,

Neuro science, Statistics, Advertising, Analytics, Natural Language Processing. Data can be analysed in pandas in two ways -

**Data frames -** In this data is two dimensional and consist of multiple series. Data is always represented in rectangular table.

**Series -** In this data is one dimensional and consist of single list with index.



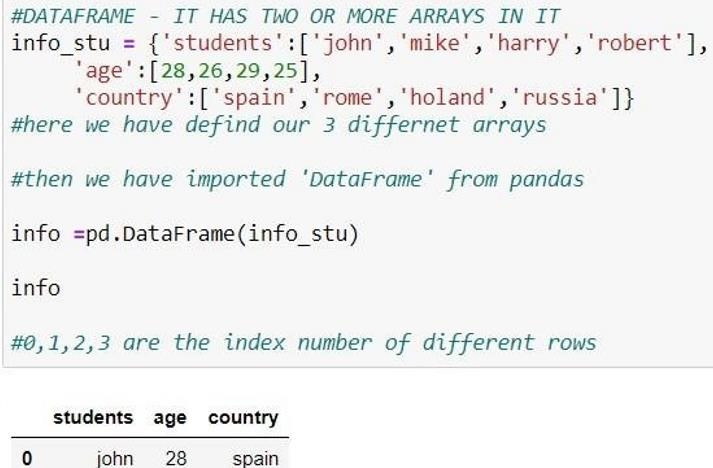


Figure 1.3.2: series and data frame in pandas

### NumPy

”NumPy is a library for the Python programming language, adding support for large, multi- dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays”. The previous similar programming of NumPy is Numeric, and this language was originally created by Jim Hugunin with contributions from several other developers. In 2005, Travis Oliphant created NumPy by incorporating features of the competing Num-array into Numeric, with extensive modifications. [12] It is an opensource library and free of cost.

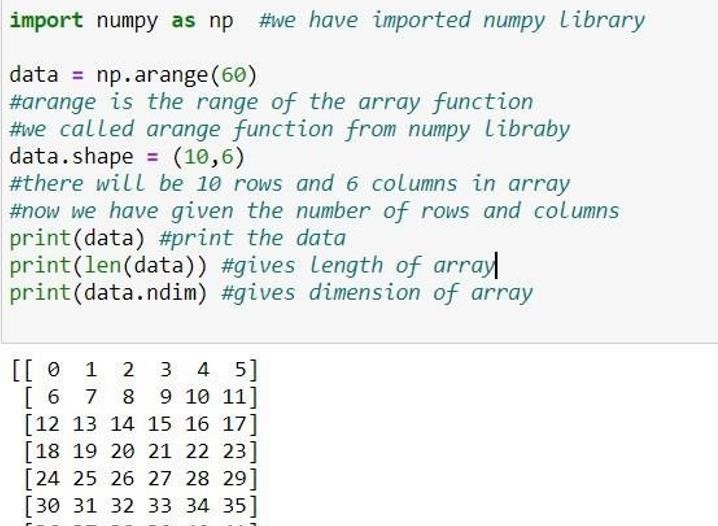


Figure 1.3.3: NumPy basic example

# CHAPTER 3: APPROACH USED (REQUIRED TOOLS)

* **Decision Tree:** A decision tree is a type of supervised machine learning used to categorize or make predictions based on how a previous set of questions were answered.
* **KNN Algorithm:** K-NN algorithm stores all the available data and classifies a new data point based on the similarity.
* **Logistic Regression:** Logistic regression is an example of supervised learning. It is used to calculate or predict the probability of a binary (yes/no) event occurring.

**REQUIRED TOOLS:**

For application development, the following Software Requirements are:

Operating System: Windows 11

Language: python

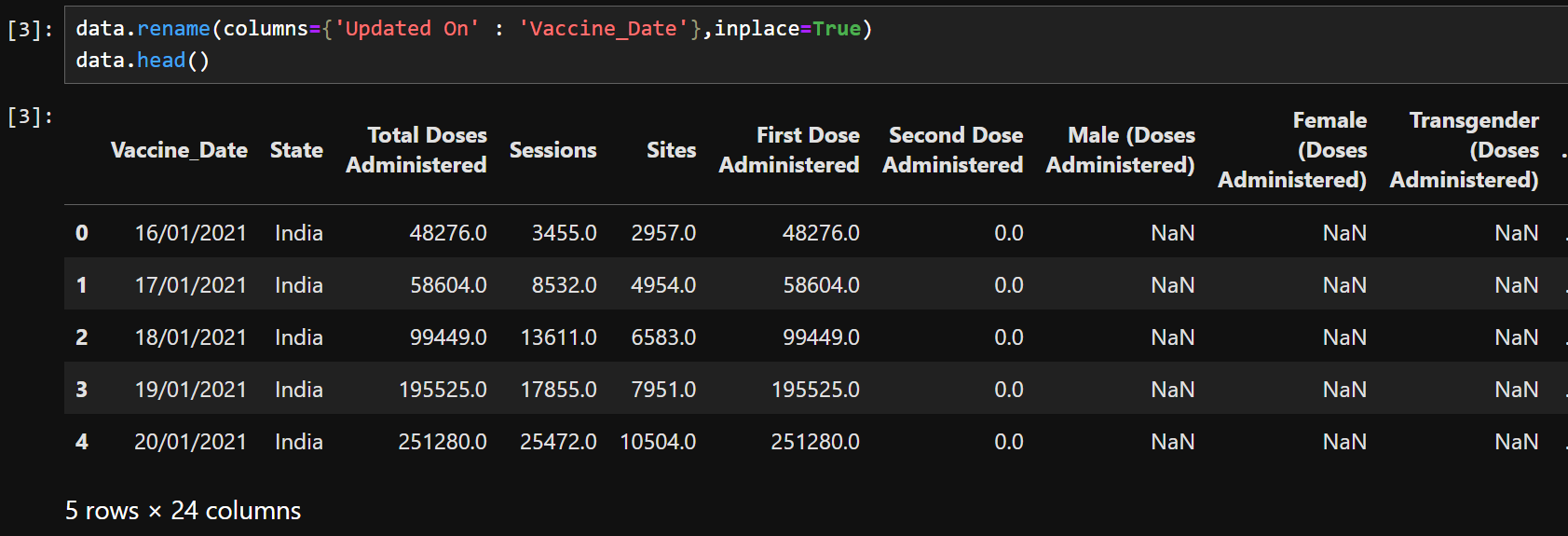
Tools: JUPYTER notebook or COLAB, Microsoft Excel (Optional). Technologies used: python.

# CHAPTER 4: RESULTS

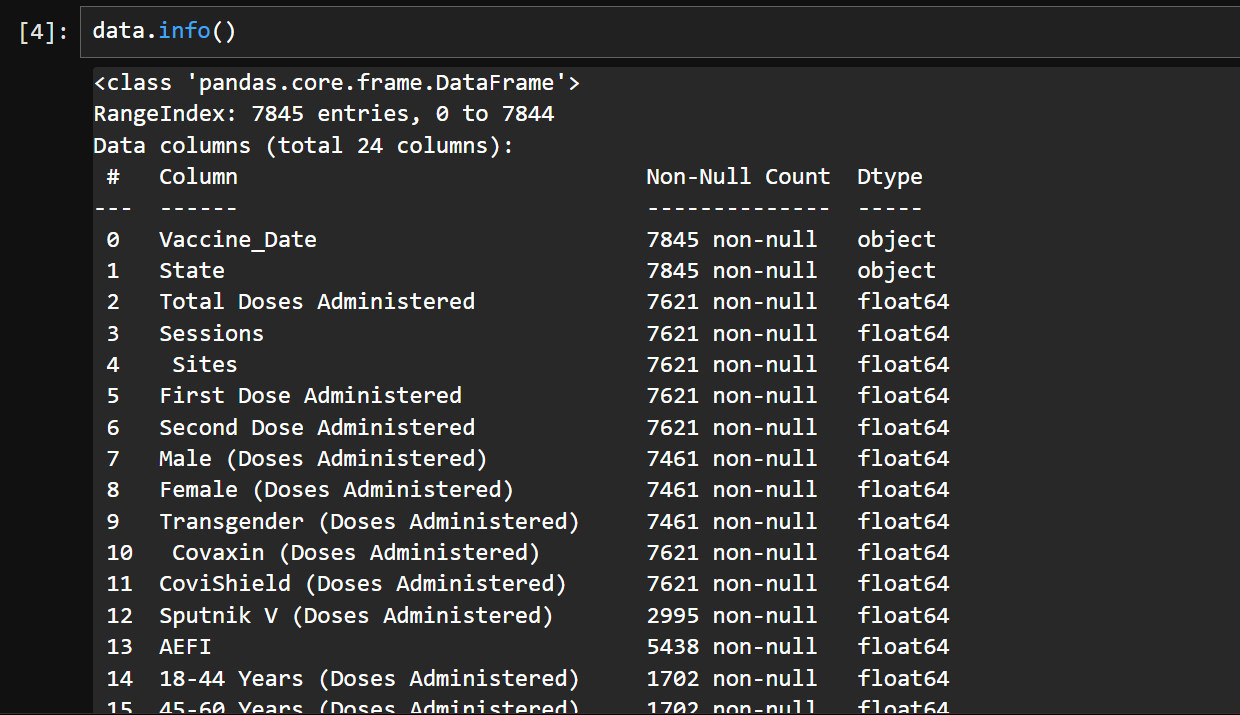
# DATA COLLECTION

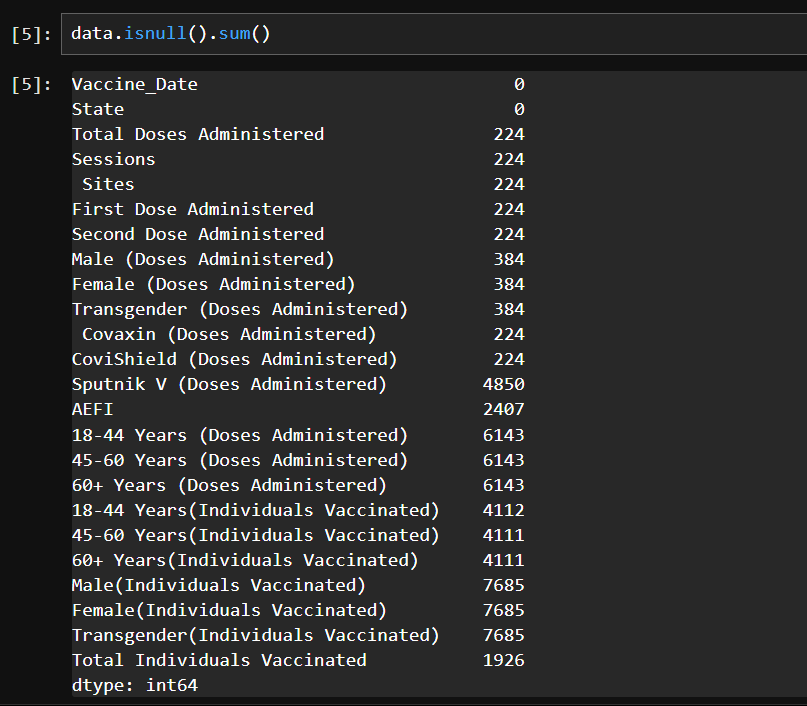
# 

# DATA RENAME

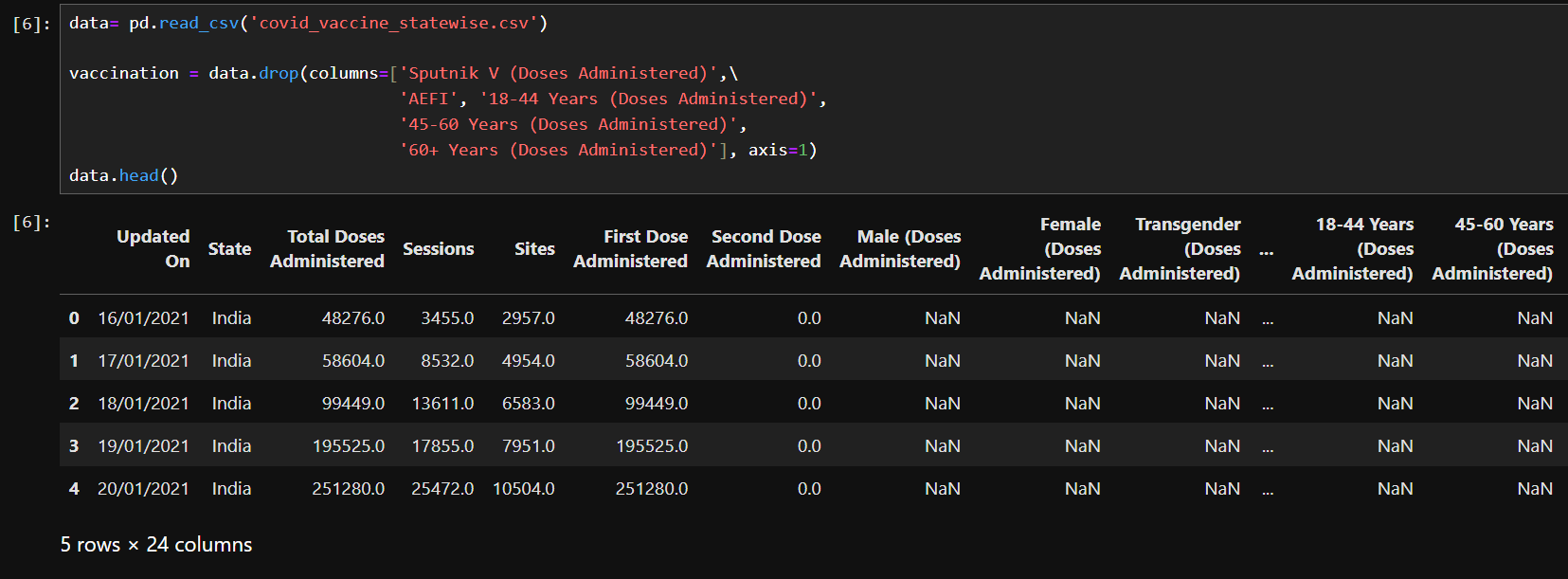


# GATHERING INFO OF DATA

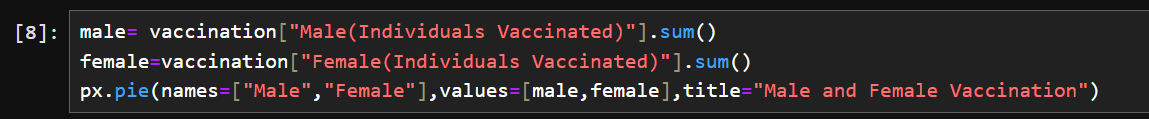


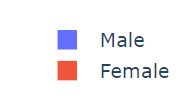
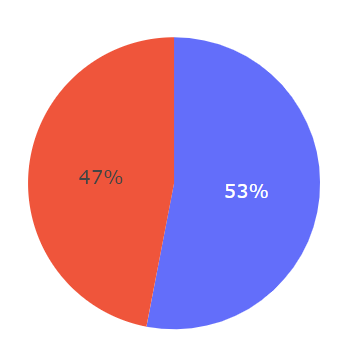


# TO DROP A VALUE

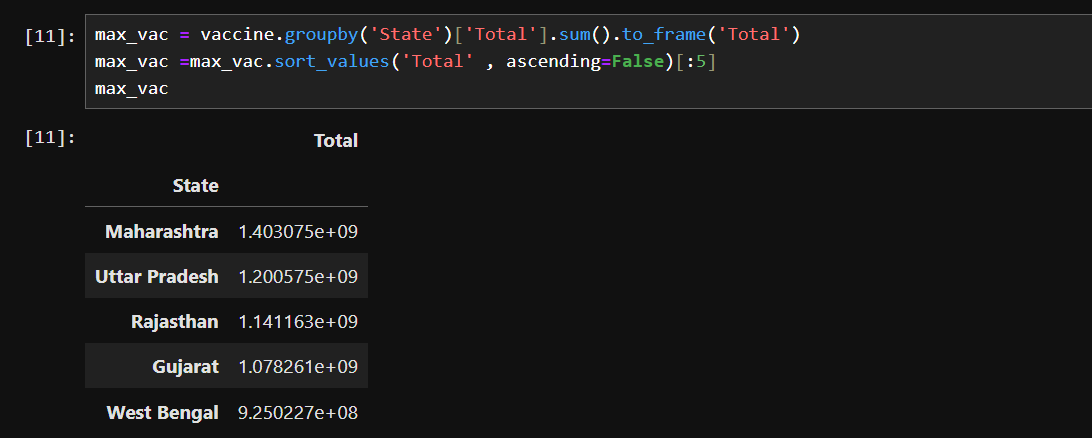


# GRAPH OF MALE AND FEMALE VACCINATION

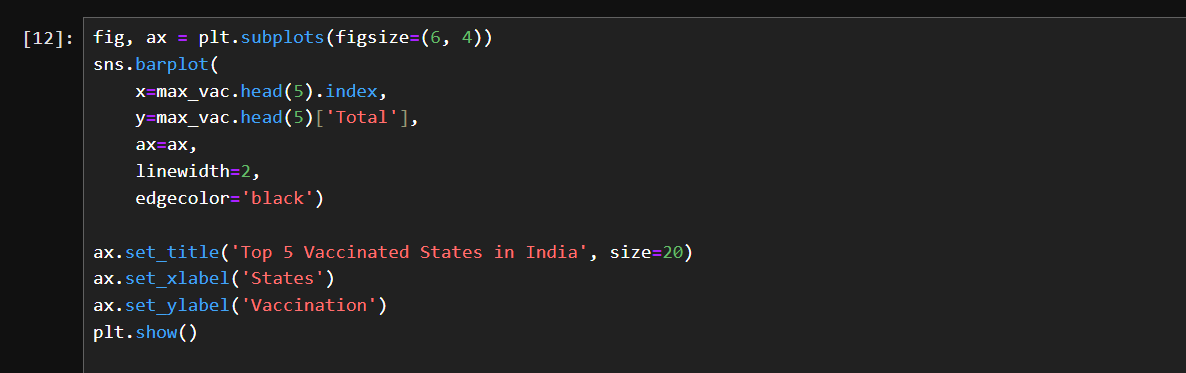


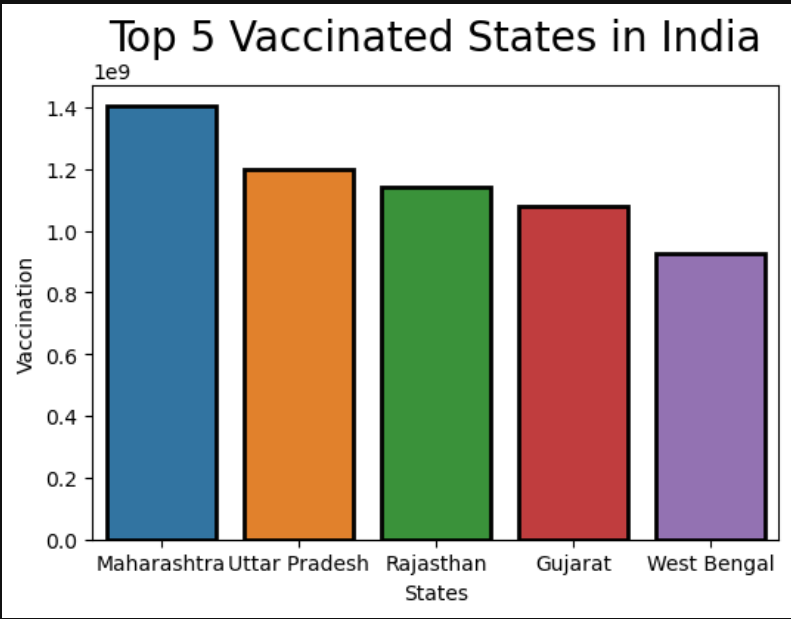


# MAXIMUM 5 VACCINATED STATES

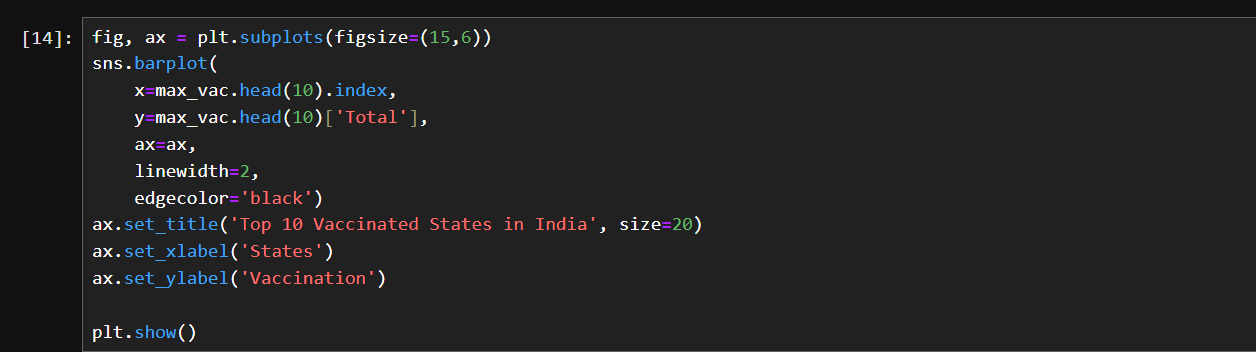


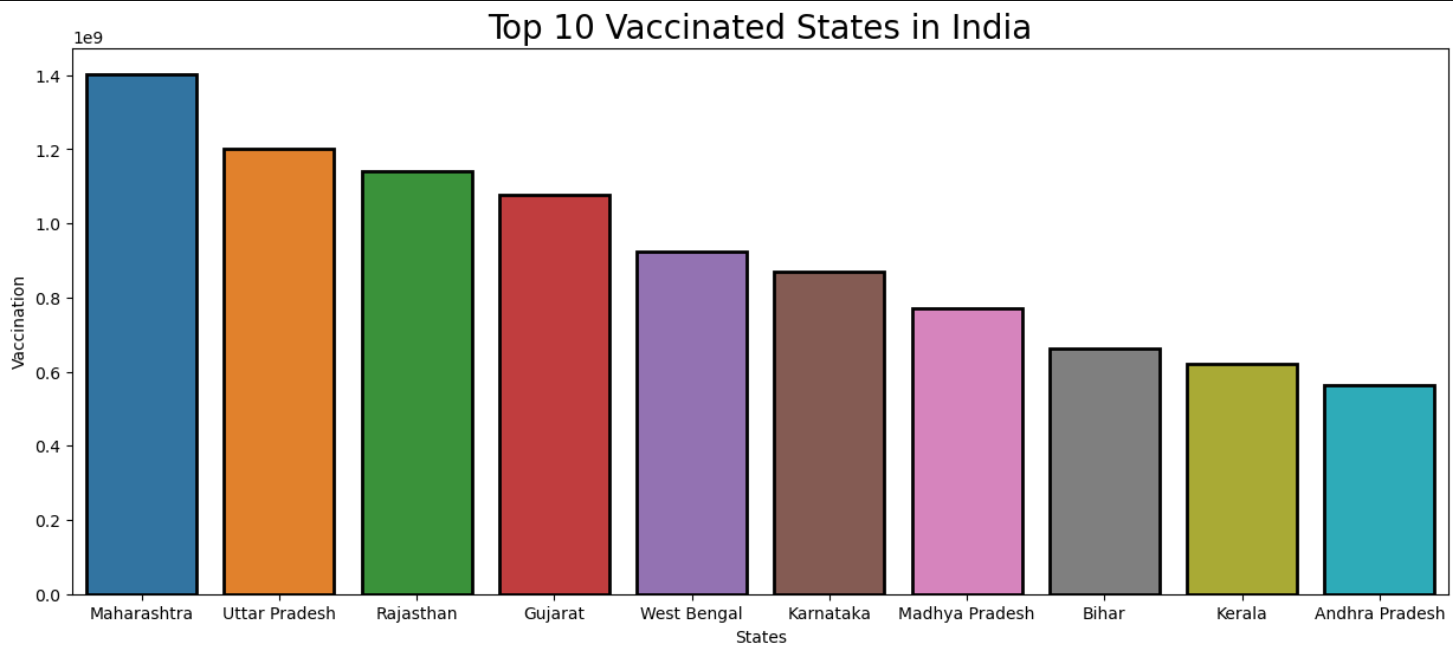
# GRAPH OF TOP 5 VACCINATED STATES



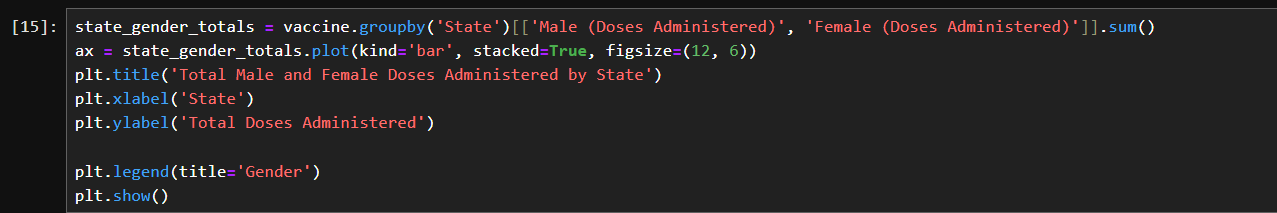


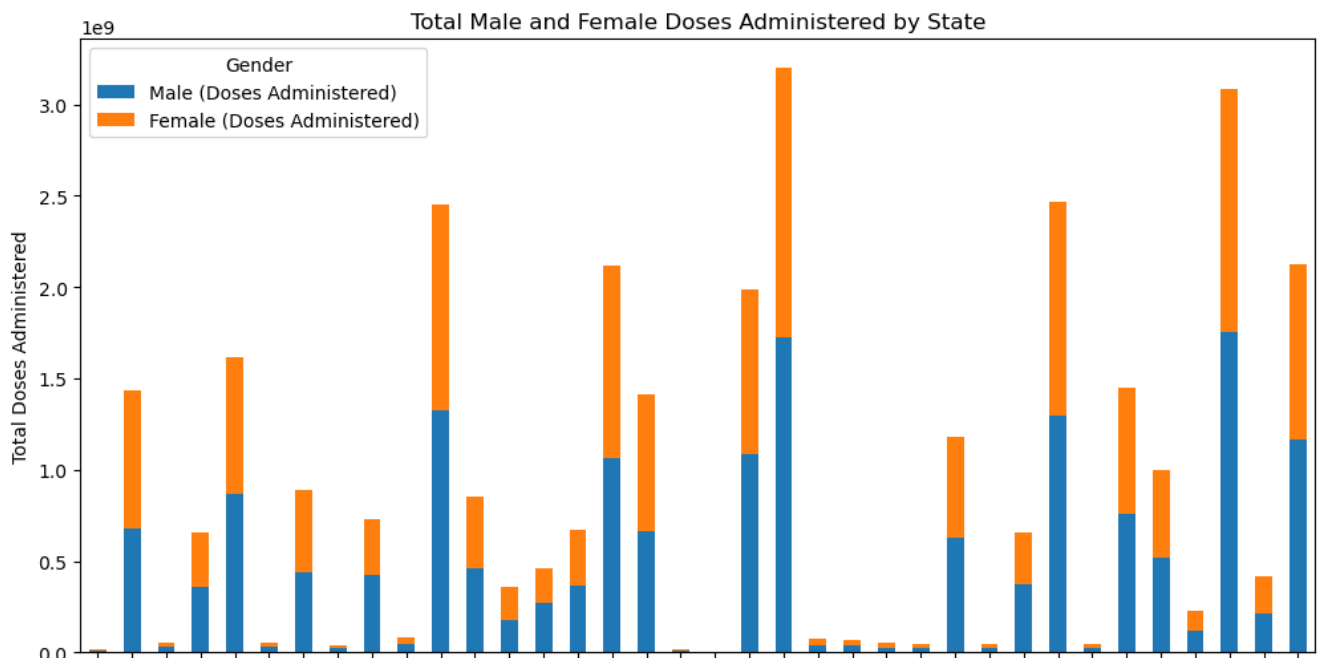
# TOP 10 VACCINATED STATES

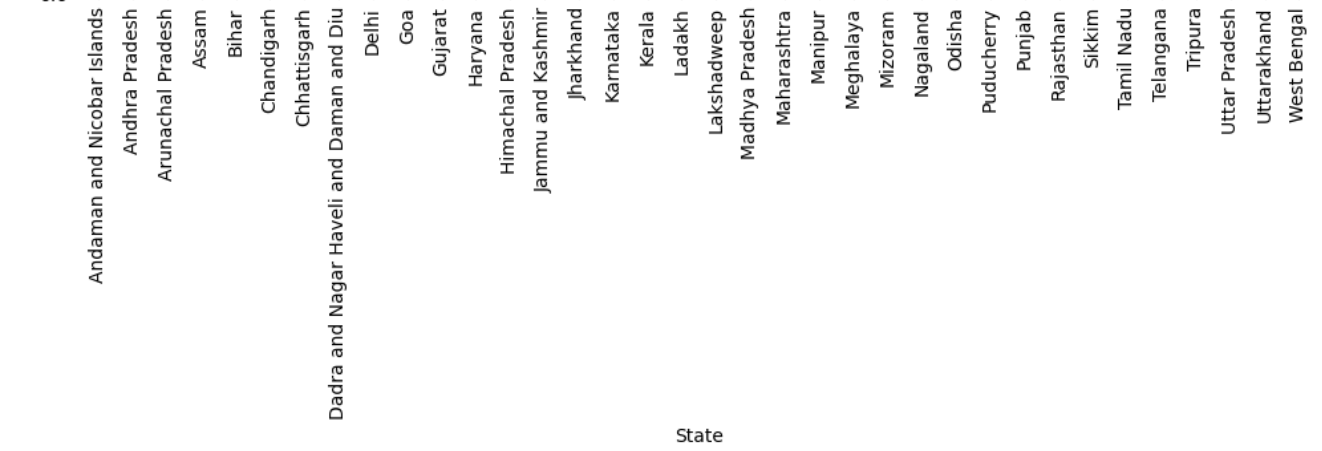




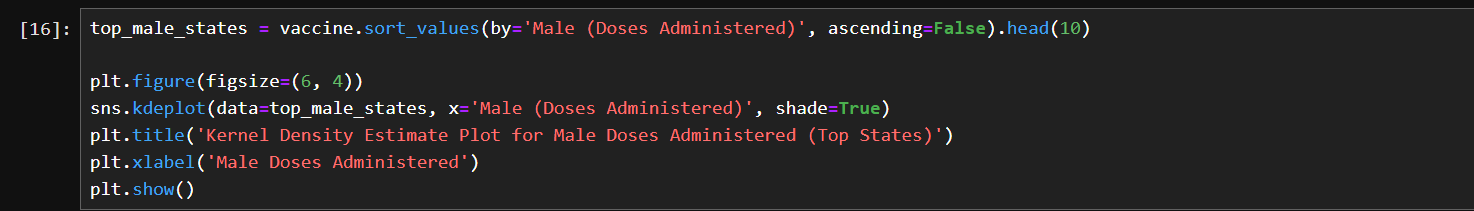
# TOTAL MALE AND FEMALE ADMINISTRATED BY STATE

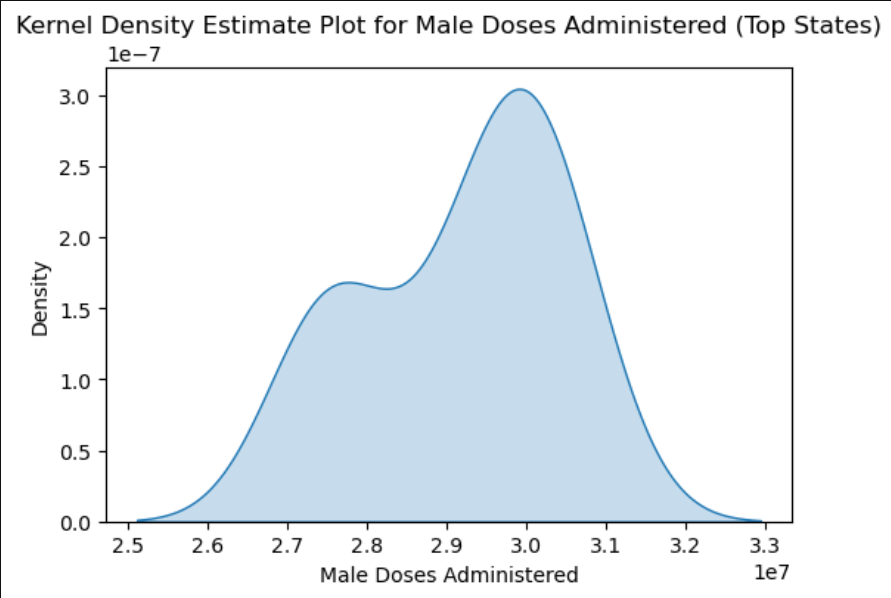






# DENSITY ESTIMATE PLOT





# CHAPTER 5: CONCLUSIONS

The application of machine learning in predicting vaccine usage presents a significant advancement in public health management. Through the integration of diverse data sources and sophisticated predictive models, this project successfully demonstrates how data-driven approaches can optimize vaccine distribution and ensure higher vaccination rates. By accurately forecasting vaccine demand, health authorities can better allocate resources, reduce wastage, and respond proactively to potential outbreaks. The insights gained from this project underscore the potential of machine learning to address complex public health challenges, ultimately contributing to more effective disease prevention and healthier communities. Continued research and refinement of these models will further enhance their accuracy and applicability, paving the way for smarter, data-informed public health strategies

**REFERENCES & BIBLIOGRAPHY**

[**www.kaggle.com**](http://www.kaggle.com/)  [**www.wikipedia.com**](http://www.wikipedia.com/) [**www.geeksofgeeks.com**](http://www.geeksofgeeks.com/) [**www.javatpoint.com**](http://www.javatpoint.com/)