# DATA ANALYSIS USING PYTHON

**Komal Nagarkar[*1], Asmita Dhanawade[*2], Asmita Ulaswar[*3], Sneha Powar[*4],**

**Supriya Kagale[*5], Priyanka Bodake[*6]**

[*1,2,3,4,5,6]Department Of Computer Engineering, Sharad Institute Of Technology Polytechnic, Yadrav (Ichalkaranji) Maharashtra, India.

## ABSTRACT

Data analysis serves as the cornerstone for informed decision-making in various domains, ranging from business intelligence to scientific research. Python has emerged as a dominant force in the realm of data analysis, owing to its extensive libraries, versatile functionalities, and vibrant community support. This abstract provides an overview of the methodologies, tools, and insights involved in harnessing Python for proficient data analysis.

Python offers a plethora of libraries tailored specifically for data analysis, including but not limited to Pandas, NumPy, SciPy, and Matplotlib. These libraries empower analysts to efficiently manipulate, process, visualize, and interpret data, thereby facilitating the extraction of meaningful insights. Whether dealing with structured data, time series, or unstructured text, Python provides the necessary tools to address diverse analytical challenges.

In addition to its rich library ecosystem, Python boasts seamless integration with other data-related technologies, such as SQL databases, big data frameworks like Apache Spark, and machine learning libraries like scikit-learn and TensorFlow. This interoperability enables analysts to leverage Python as a central hub for data analysis, irrespective of data size, complexity, or source.

Furthermore, Python fosters a collaborative and reproducible analytical workflow through the use of Jupyter Notebooks, which combine code, visualizations, and explanatory text in a single document. Jupyter Notebooks promote transparency, documentation, and knowledge sharing among analysts, thereby enhancing the reproducibility and scalability of data analysis projects.

**Keywords:** Superstore Data Analysis Using Python & Tableau.

## I.     INTRODUCTION

Data analysis is a broad field that involves examining, cleaning, transforming, and interpreting data to discover meaningful insights, patterns, and trends. Python is a popular programming language for data analysis due to its rich ecosystem of libraries and tools specifically designed for this purpose. Here are some key steps and libraries commonly used in data analysis with Python

## II.     LITERATURE SURVEY

Following articles are studied during the literature survey phase of this work.

**2.1 Used Library**

- **Numpy :**

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely. NumPy stands for Numerical Python.
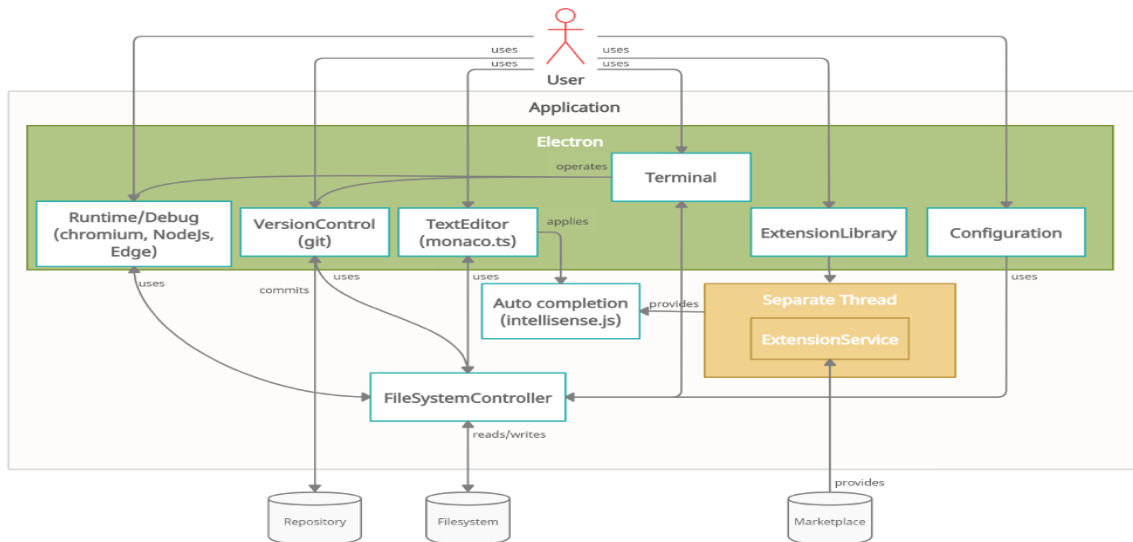
- **Pandas :**

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

## 2.2 Look Of The Data
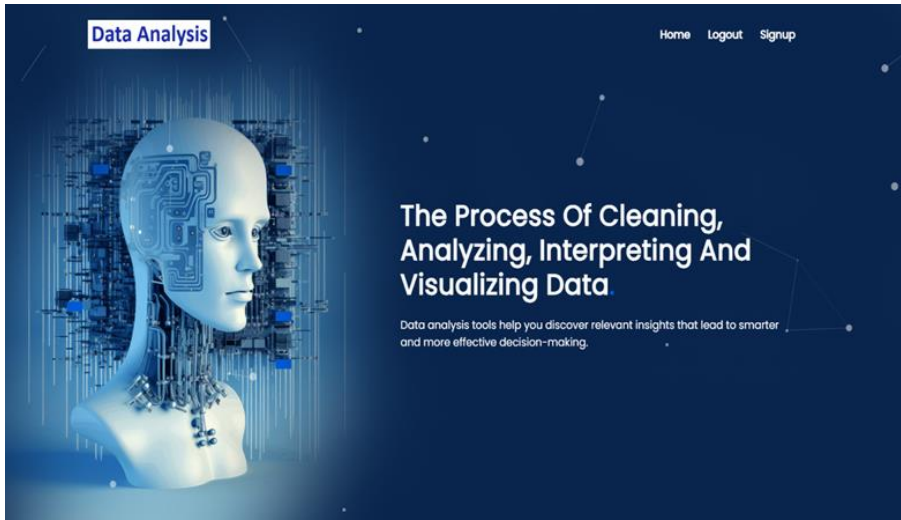


## III.    METHODOLOGY



## IV.    MULTIPLE DATASET ARE USE

### 1. Sports Youtube Channel
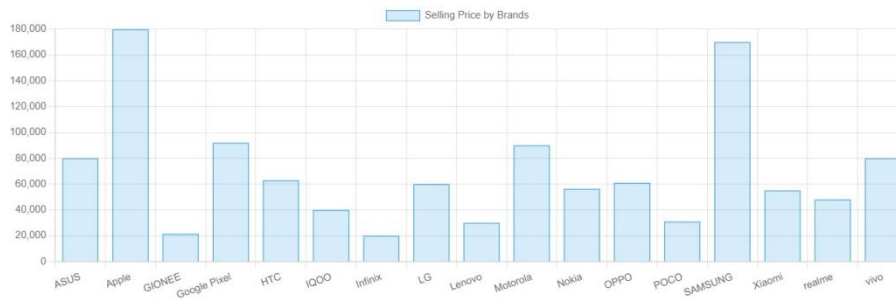
**2. All Company Mobile Sale**



**3. Delhi Metro Network**
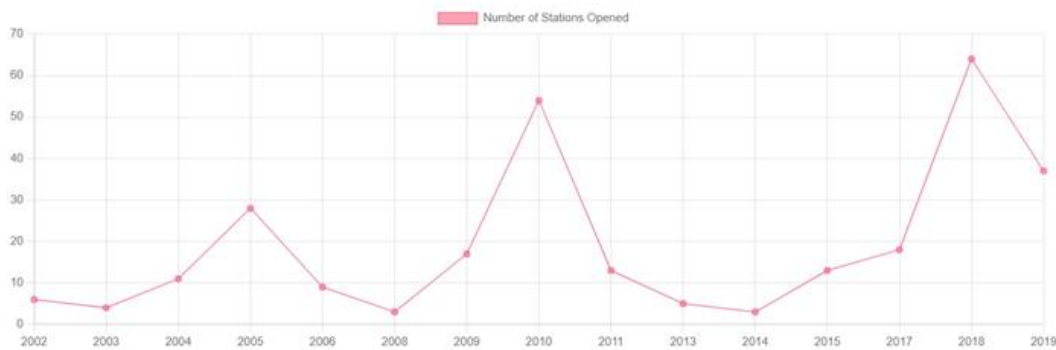


**4. T20 World Cup 2022**

**How To Analyze Data**



**Collect the data**

Set up the necessary data collection processes and assemble a team of people who can accumulate and manage the information.

**Clean The Data**

Once you've built a large data set, the next step is to clean or process the data to ensure it's accurate and free of error.

**Evaluate The Data**

Investigate the data you've gathered carefully. Use various mathematical interventions to look for connections between points.

**Establish a goal**

First, determine the purpose and key objectives of your data analysis. Think about the questions or concerns you have and the goal you want to achieve by conducting this analysis.

## V. SYSTEM ARCHITECTURE

Tableau manages many different data sources, including SQL Server, PostgreSQL, MySQL, Redshift, Snowflake, Google Big Query, and flat files across various networks. Public Cloud data connections are generally made directly. For private connections, a pool of Tableau Bridge services facilitates the live on-premises or VPC database query capabilities, and scheduled extract refreshes.

## VI. CONCLUSION

Here we can easily view that there is a huge gap Between the Central and the south region. The Sales in the Central region is the highest in all the Years under consideration. Moreover, as the Slope of the central region is upward sloping, so we can conclude that sales in the Central region is increasing overtime. But as far as the other regions are concerned, the sales is either increasing at a very slow pace or either stagnated, particularly for the Caribbean and Canada Region

## VII. REFERENCES

[1]     Data Src : https://community.tableau.com/docs/DOC-1236   Hafizhan Aliady | Data Mining | 26 Oct 2017

[2]     https://www.realcode4you.com/post/sample-superstore-sales-data-analysis-using-tableau-tableau-data analysis-help-realcode4you

[3]     Use Vs Code

[4]     Kaggle.com

[5]     Geeks for Geeks