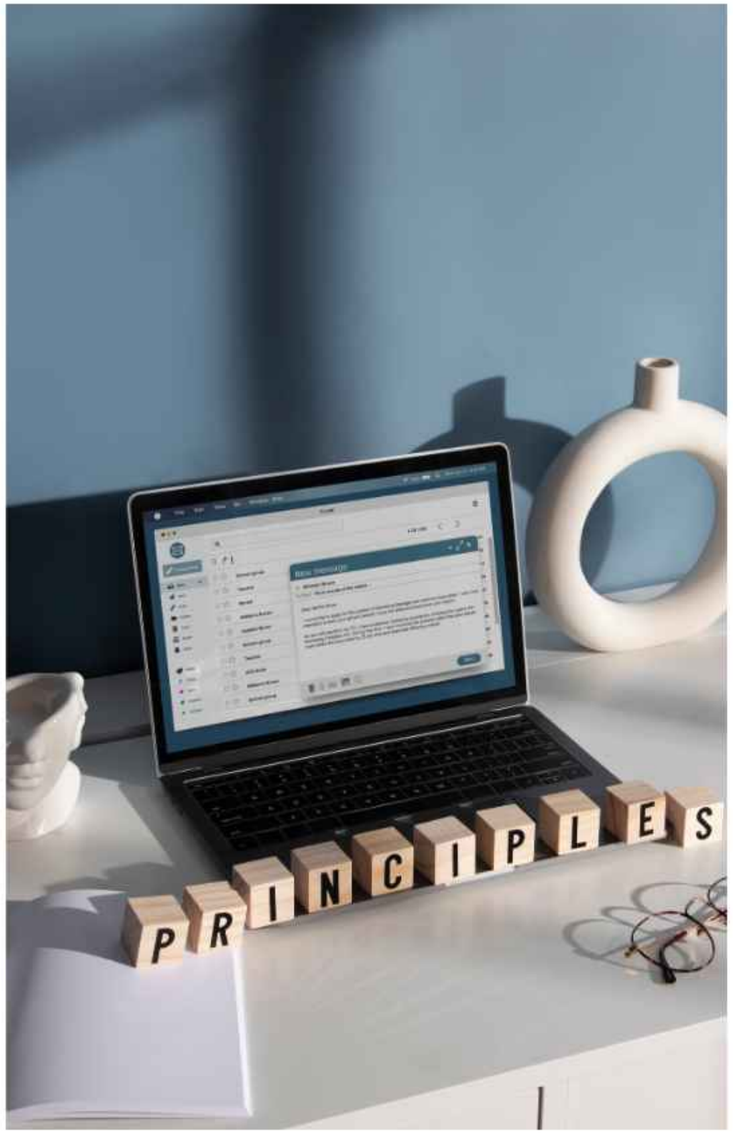
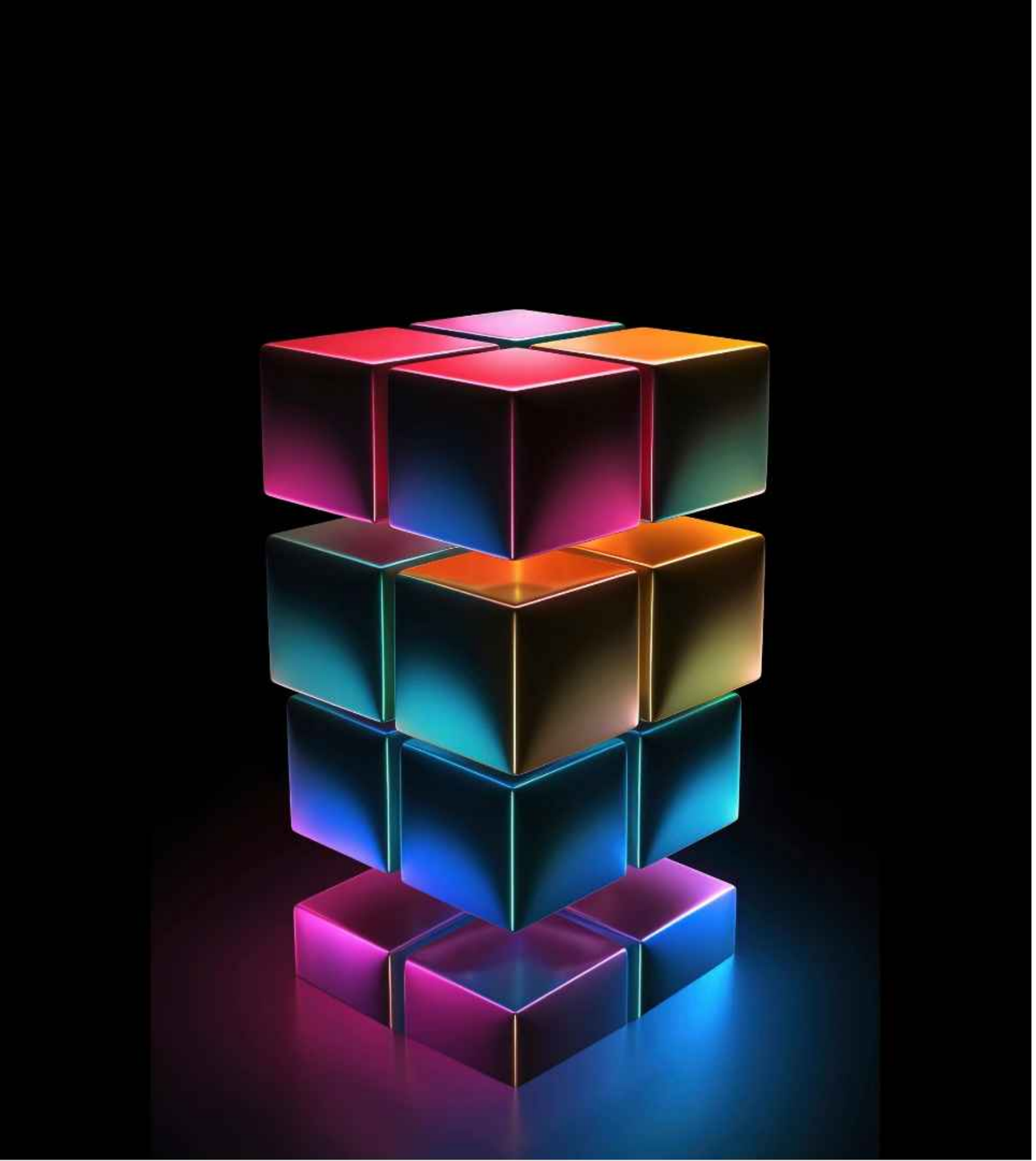
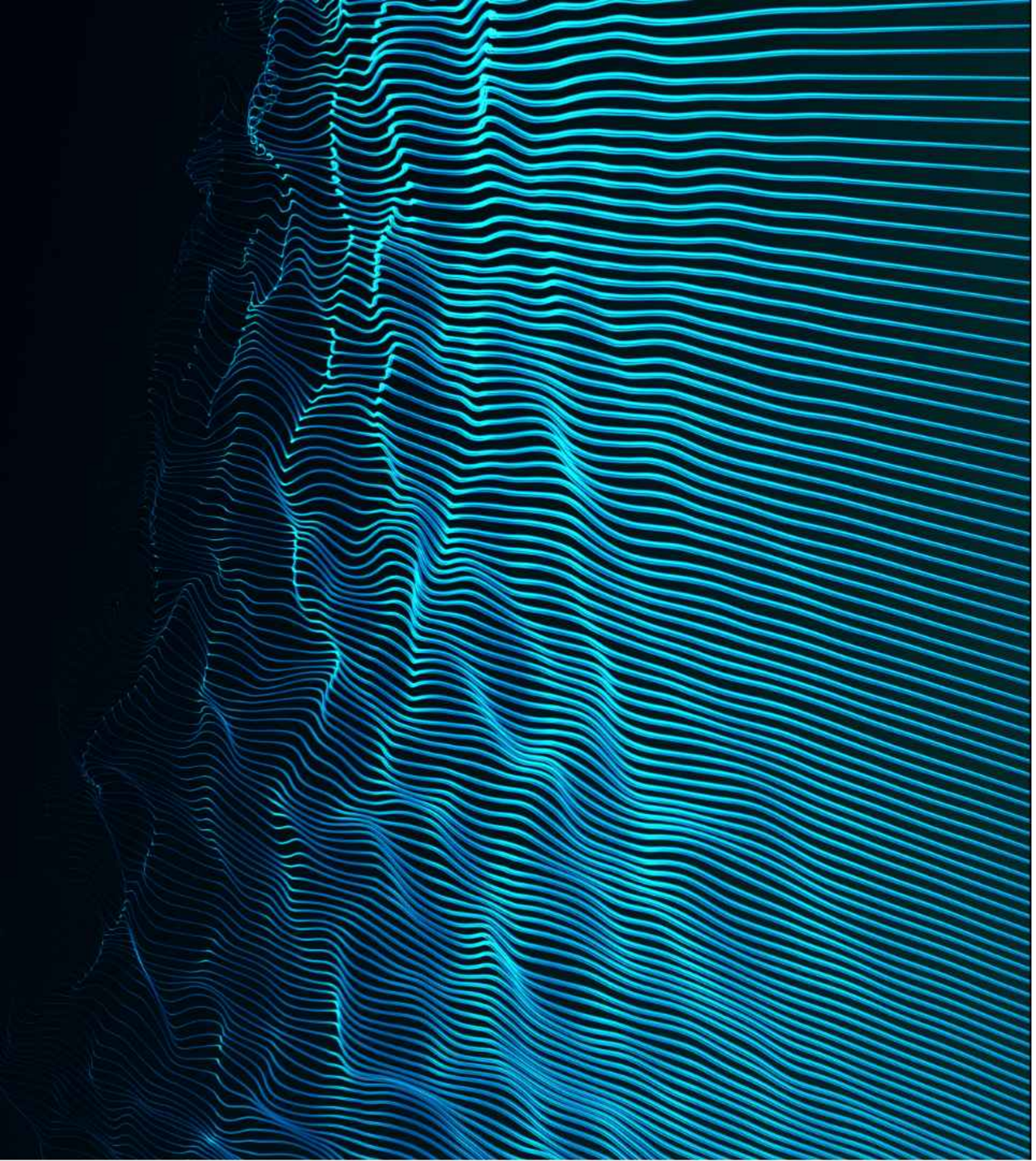


Data analysis using Python

by NALINI AGARWAL



INDEX



Introduction

In today's digital age, analyzing sales data is vital for businesses to gauge performance and improve strategies. They need to understand how well they're doing and what they can do better. To do this, they use special computer tools called Python libraries like NumPy, Pandas, Matplotlib, and Seaborn. These tools help them make sense of all the data they have and find important information that can help them improve.

Problem Statement

In this project we study about the sales of Amazon company and analyze properties like sizes, sales numbers, states having more sales, buyers and sellers percentage, delivery statuses, etc.

Breakdown of project

1. Introduction to the project.
2. Setting up the development environment with Python and the necessary libraries.
3. Loading the raw sales data into Pandas Data Frames for efficient manipulation and analysis.
4. Cleaning and preprocessing the data to ensure accuracy and reliability.
5. Performing descriptive statistics using NumPy to gain insights into the dataset.
6. Visualizing sales trends and patterns using Matplotlib and Seaborn.
7. Conducting correlation analysis to identify relationships between different variables.
8. Generating meaningful reports and visualizations to communicate findings effectively.

Tools and Libraries

Libraries-

NumPy: A fundamental package for numerical computing in Python, enabling efficient operations on large arrays and matrices.

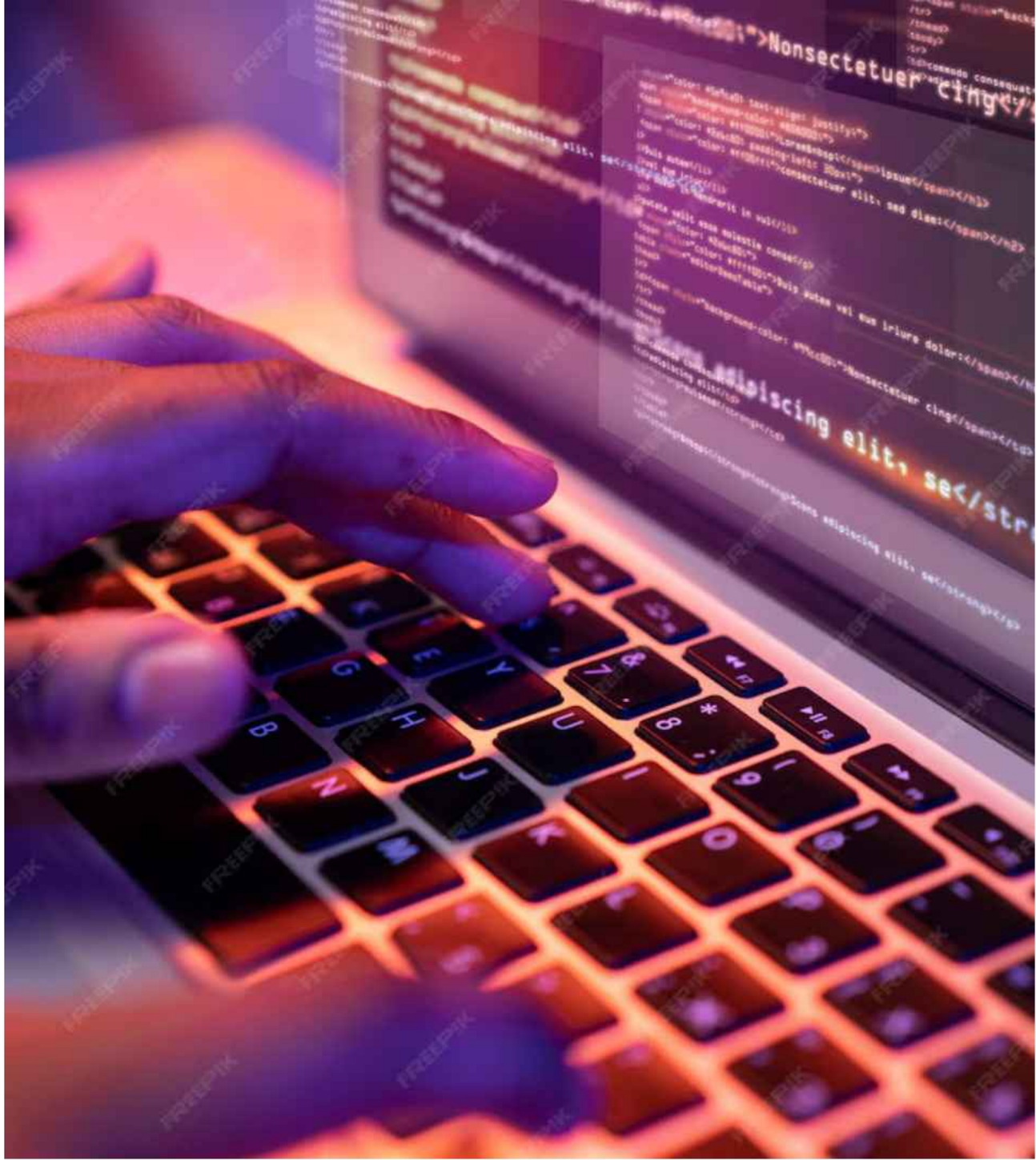
Pandas: A powerful library for data manipulation and analysis, particularly for working with structured data like tables.

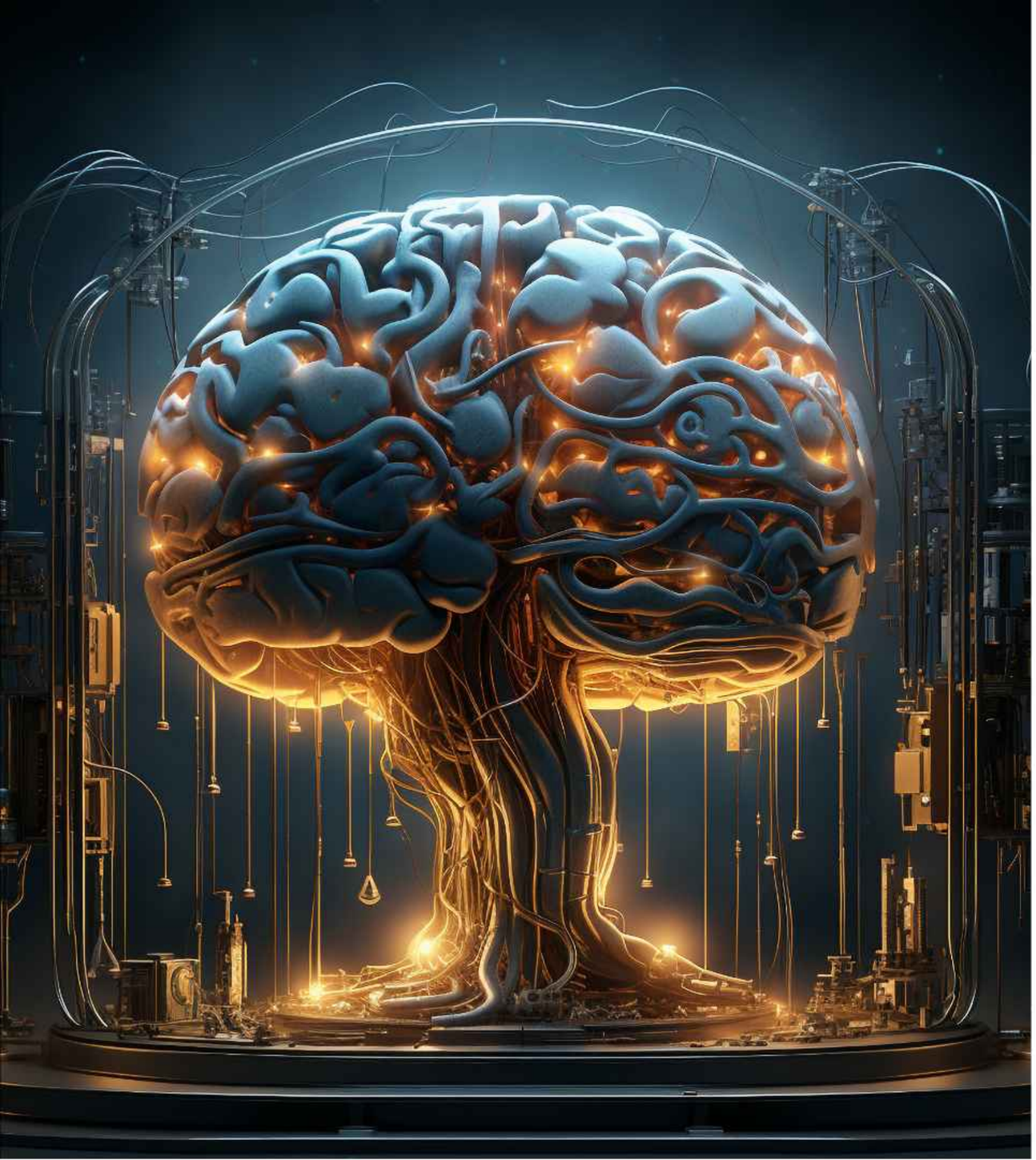
Matplotlib: A comprehensive library for creating static, interactive, and animated visualizations in Python.

Seaborn: Built on top of Matplotlib, Seaborn provides a high-level interface for creating attractive and informative statistical graphics.

Tools-

Jupyter Notebook: An open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text.



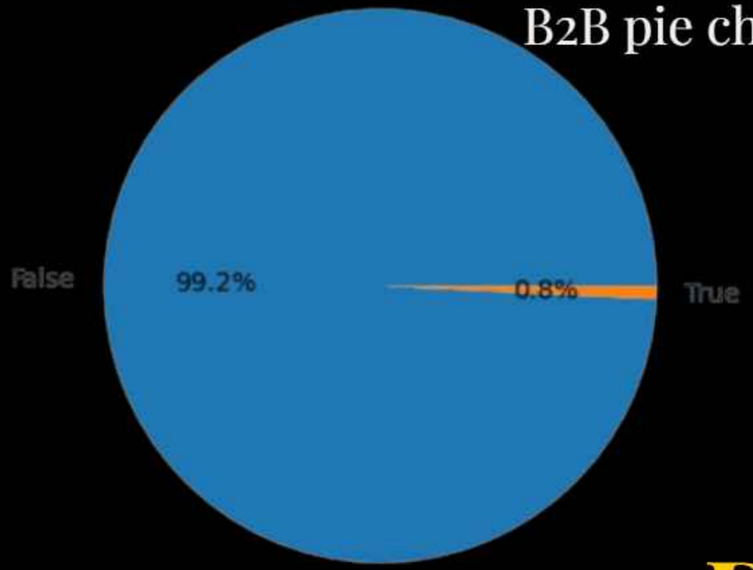


WOW Factor

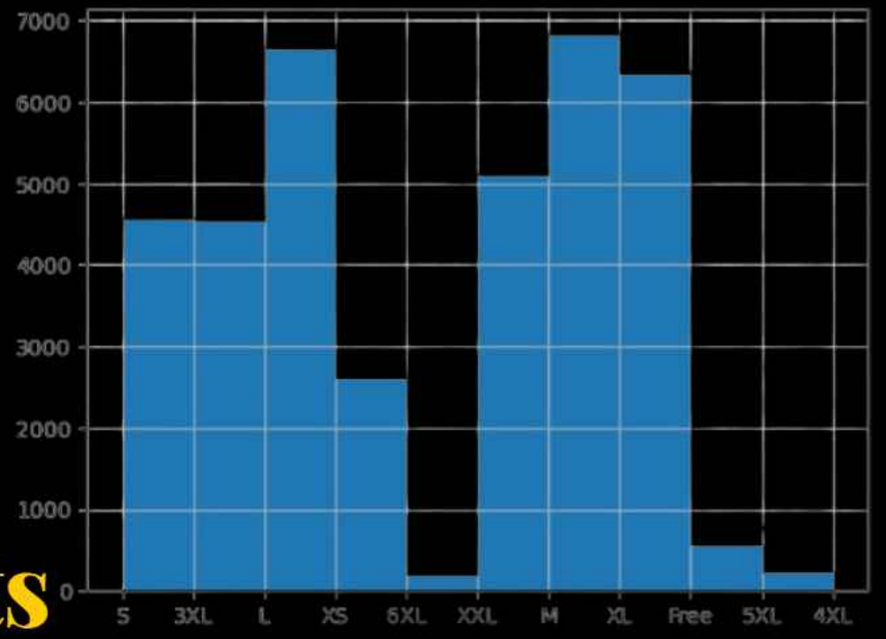
The "wow" factor of this project lies in its ability to transform raw Amazon sales data into actionable insights using advanced data analysis and visualization techniques. By leveraging Python libraries, the project can uncover hidden patterns, trends, and correlations within the sales data.
The interactive and informative visualizations created through Matplotlib and Seaborn add another layer of sophistication, enabling stakeholders to grasp complex information intuitively.

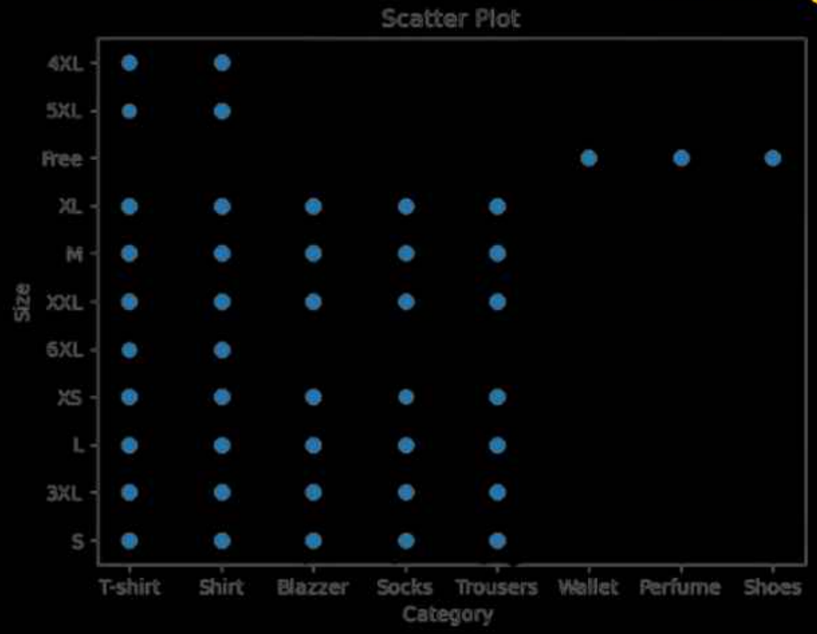
Diagrams and charts

B2B pie chart

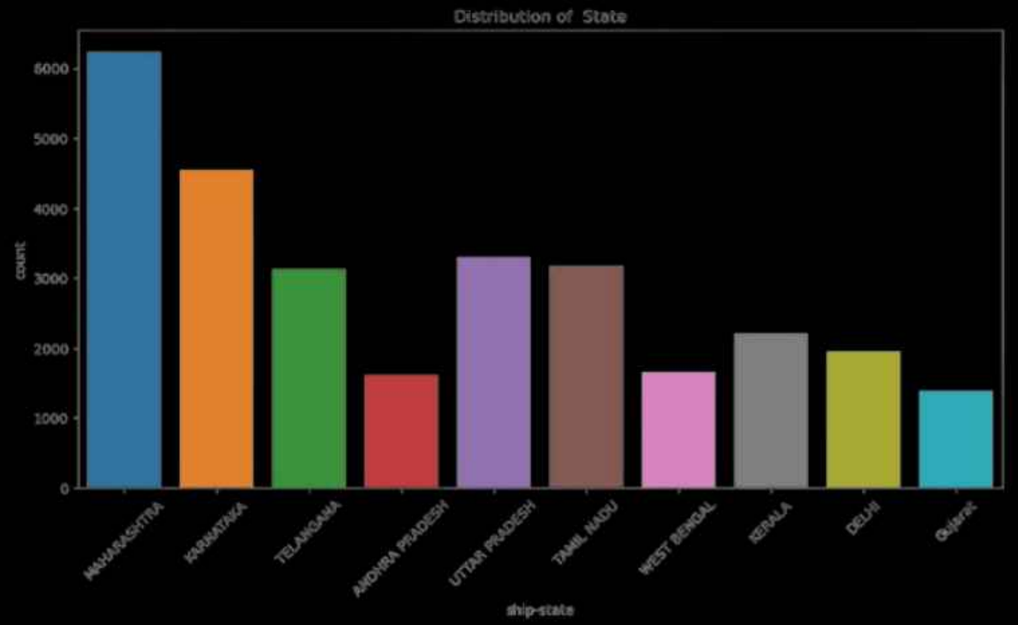


Histogram of size





Scatter plot of categories and sizes



Plot count of top 10 cities

Conclusion and result

The data analysis done in this project reveals that the business has a significant customer base in Maharashtra state, mainly serves retailers, fulfills orders through Amazon, experiences high demand for T-shirts, and it is seen that M-Size is the preferred choice among buyers.

In conclusion, this project showcases the effectiveness of Python and data analysis tools in deriving valuable insights.

Future scope

In the future, expanding the project to incorporate natural language processing (NLP) techniques could enable sentiment analysis of customer reviews, offering insights into product satisfaction and potential areas for improvement. Moreover, integrating with external data sources such as social media trends or economic indicators could provide a more comprehensive understanding of market dynamics for strategic decision-making.

Thank You!!