



SUPERSTORE SALES DATASET VISUALIZATION WITH STREAMLIT IN PYTHON

MISS. RAJESHREE NALBALWAR¹, MISS. SHRUTI SHINDE², MR. HARISH CHAKOTE³, MR.
DESHPANDE G.R⁴

Student^{1,2,3}, Assistant Professor⁴

Gramin Technical and Management Campus, Nanded, India
Computer Engineering

ABSTRACT:

In this project, we leverage the power of Streamlit, a user-friendly Python library, to create an interactive web application using the Streamlit framework to visualize and analyze the Superstore sales dataset. The application will allow users to explore the data and gain insights into customer behavior, product performance, and sales trends. The goal is to build an interactive dashboard that allows users to explore the data and gain insights into sales performance.

This is an analysis report which generates some insight from a dataset named "Super Store". The SuperStore dataset contains data on order details of customers for orders of a superstore in the US. This will create a Streamlit app that displays the Superstore dataset and a bar chart of sales by region. You can customize the app to add more features and visualizations.

Keywords: *Streamlit*, superstore sales dataset, data visualization, interactive dashboard, data exploration, data cleaning and preprocessing, key performance indicators (KPI), user interaction.

INTRODUCTION:

The Superstore Sales dataset contains historical sales records from supermarket branches. The dataset includes information on the product, the customer, the order, and the sales.

In the realm of data-driven decision-making, effective visualization of datasets is crucial for extracting meaningful insights. This project aims to harness the capabilities of Streamlit, a Python library designed for creating interactive web applications with minimal effort, to develop an intuitive and visually compelling dashboard for visualizing a Superstore sales dataset.

We can also use Streamlit to add filters to the dashboard. For example, we can add a filter to allow users to select a specific product or customer. This will allow users to see how the sales for that product or customer have changed over time.

The crux of our endeavor lies in leveraging Streamlit's capabilities to create an interactive and user-friendly dashboard. Through various visualizations, such as sales trends over time, product performance metrics, customer segmentation analyses, our dashboard aims to provide stakeholders with a comprehensive view of the Superstore's sales dynamics.

The ensuing sections will delve into the intricacies of data exploration, cleaning, and the step-by-step creation of the Streamlit dashboard, culminating in a holistic conclusion and potential avenues for future enhancements.

Overall, Streamlit is a powerful tool for visualizing data in Python. It is easy to use and allows users to create interactive dashboards that can be used to explore data and gain insights.

SYSTEM ARCHITECTURE:

The architecture of the system involves several components working together to enable the creation and deployment of an interactive dashboard for visualizing the Superstore sales dataset using Streamlit in Python.

1. Superstore Sales Dataset: The raw dataset containing information about sales transactions, including details on products, customers, and sales metrics.

2. Python Programming Language: Python serves as the core programming language for data manipulation, analysis, and visualization. Libraries such as Pandas, NumPy, and Matplotlib/Plotly are utilized for data processing and chart creation.

3. Streamlit Library: Streamlit is a Python library used for creating web applications with minimal effort. Streamlit's simplicity allows for quick iteration and easy integration with existing Python code.

3. Data Exploration and Preprocessing: The python libraries used in the exploratory data analysis include NumPy, Pandas, and Plotly. The dataset undergoes initial exploration and preprocessing using Pandas and other relevant Python libraries.

4. Interactive Dashboard Components:

- Charts and Visualizations: Matplotlib or Plotly can be integrated within Streamlit to create interactive charts like line plots, bar charts, scatter plots, and geospatial visualizations.
- User Interaction Elements: Streamlit provides widgets (e.g., sliders, buttons, dropdowns) to allow users to interact with the dashboard dynamically. Users can select specific time periods, product categories, or regions.

5. Key Performance Indicators (KPIs): Calculations for key metrics and KPIs (e.g., total revenue, average order value, profit margins) are integrated into the Streamlit dashboard to provide a quick overview of the Superstore's performance.

6. Web Server: Streamlit includes an embedded web server that serves the created dashboard. This server handles user requests, updates the visualizations in real-time, and ensures a responsive user experience.

7. User Interface (UI): The final output is a user-friendly web-based interface accessible through a web browser. Users can interact with the dashboard, explore visualizations, and gain insights from the Superstore sales dataset.

8. Security Considerations: Depending on the deployment platform, security measures may need to be implemented to protect sensitive data and ensure secure access to the dashboard.

METHODOLOGY:

The methodology outlines the step-by-step process for creating an interactive dashboard to visualize the Superstore sales dataset using Streamlit in Python. The methodology encompasses data exploration, cleaning, and the development of an interactive dashboard.

A. Getting a sales data:

- Import libraries and read input data
- Create a basic dashboard setup
- Design a user interface
- Add:
 - Top-level filter
 - Interactive charts
 - Page title
 - KPIs or summary cards
 - Data table

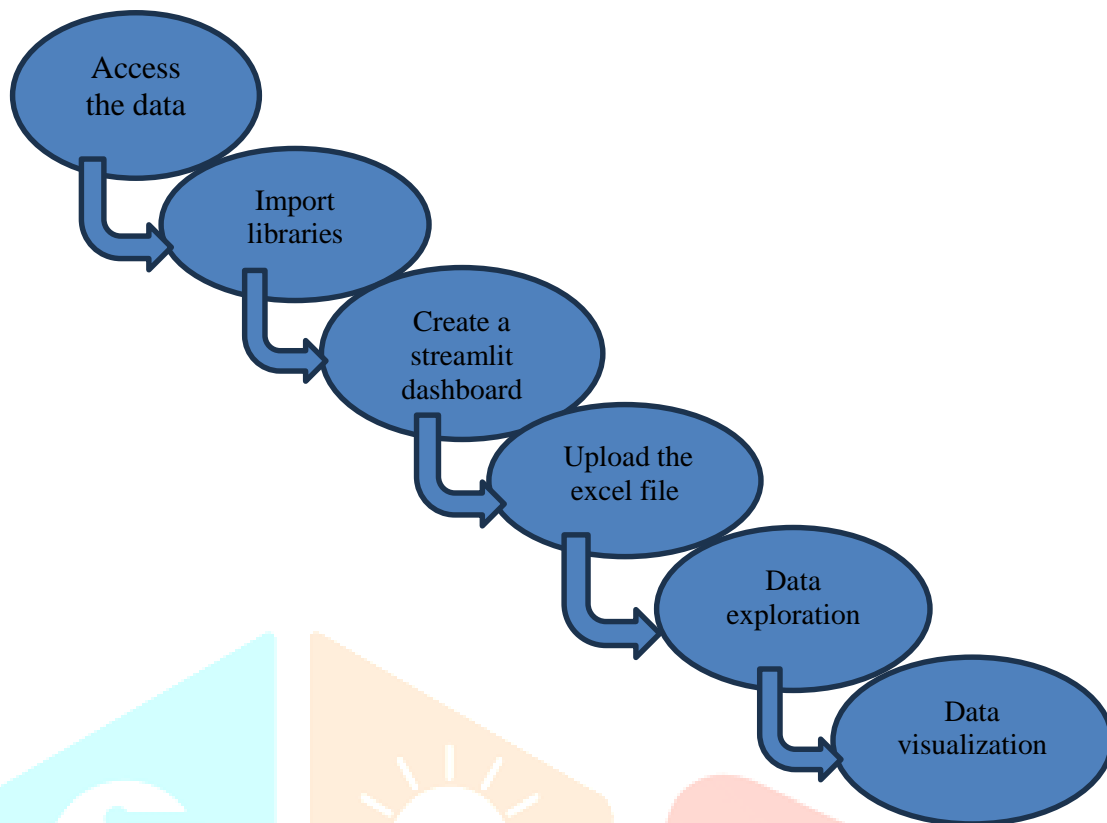
B. The project will be completed in the following steps:

- Data preparation: The data will be cleaned and prepared for visualization.
- Dashboard design: The dashboard will be designed using Streamlit.
- Data visualization: The data will be visualized using Matplotlib.
- Testing and deployment: The dashboard will be tested and deployed to a production environment.

C. Python libraries and tools:

- streamlit: Streamlit is a library that makes it easy to create interactive web applications for data science and machine learning.
- 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.
- Pandas: Pandas is a Python library used to analyze data. It is a high-performance, easy-to-use data analysis library for Python. It provides data structures and operations for manipulating numerical tables and time series.
- Matplotlib: Matplotlib is a Python library used for data visualization. It is a comprehensive plotting library that can be used to create a wide variety of plots.
- Seaborn: Seaborn is a Python library used for statistical graphics. It is a Python visualization library based on Matplotlib that provides a high-level interface for drawing attractive and informative statistical graphics.
- Plotly: Plotly is a Python library used for interactive, scientific graphs. It is a Python graph library that can be used to create interactive, scientific graphs.
- OS: The OS module is a built-in module that provides a way of using operating system dependent functionality.
- Altair: Altair is a declarative statistical visualization library for Python

D. Flow diagram:

**DISCUSSION AND RESULT:**

The Superstore dataset contains information on customer order details for a US superstore. The dataset includes columns such as: Order date, Shipping date, Product ordered, State, Region. The goal of analyzing the dataset is to gain insights and answer business questions.

The superstore sales data from 4 Jan 2012 to 31 Dec 2015.

The Following results are obtain from analysis to access and load the data from browser:

- First 5 observation

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit
0	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	261.9600	2	0.00	41.9136
1	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	731.9400	3	0.00	219.5820
2	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	14.6200	2	0.00	6.8714
3	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	957.5775	5	0.45	-383.0310
4	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	22.3680	2	0.20	2.5164

- Last 5 observation

	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub-Category	Sales	Quantity	Discount	Profit
9989	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	25.248	3	0.2	4.1028
9990	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	91.960	2	0.0	15.6332
9991	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	258.576	2	0.2	19.3932
9992	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	29.600	4	0.0	13.3200
9993	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	243.160	2	0.0	72.9480

Fig1. Metric container showing sales, profit, and orders KPI.

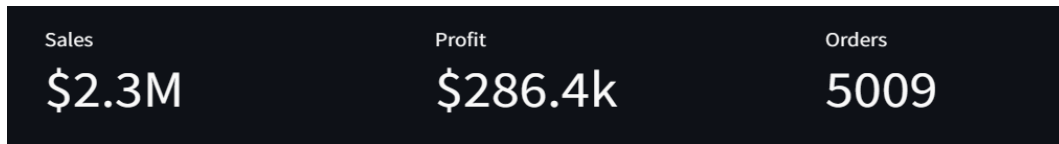
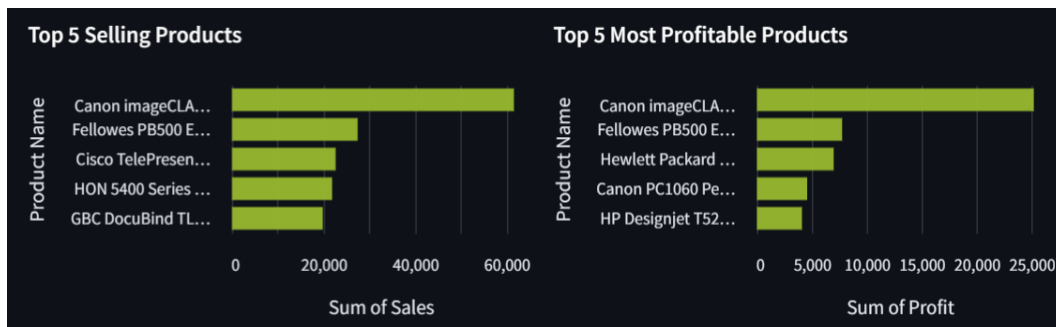


Fig2. Top 5 selling product and profitable product chart.



The create horizontal bar charts to display the metrics for the top 5 products by total sales and by total profit. In the code below we first query the data by grouping by sum of the sales and sum of profit and then get the top 5s.

Fig3. Bar chart category wise sales and Pie chart Region wise sales.

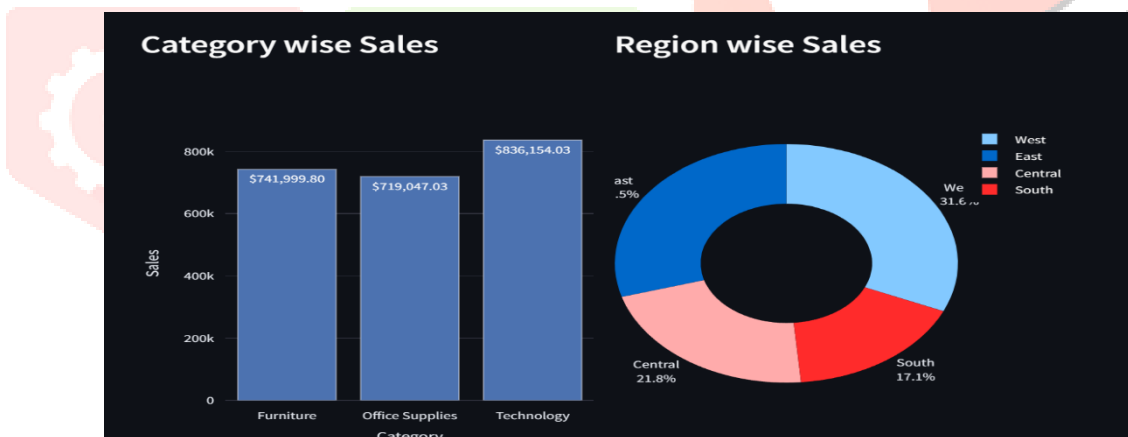
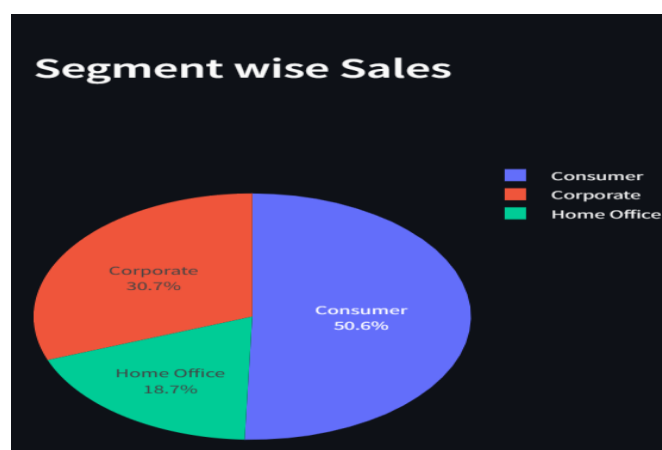


Fig4. Segment wise sales.



It made up to 50.6% of the total customers that come to the superstore. Meanwhile, corporate comes with 30.7% and home office with 18.7%.

Fig5. Relation between sales and profit using scatter plot.

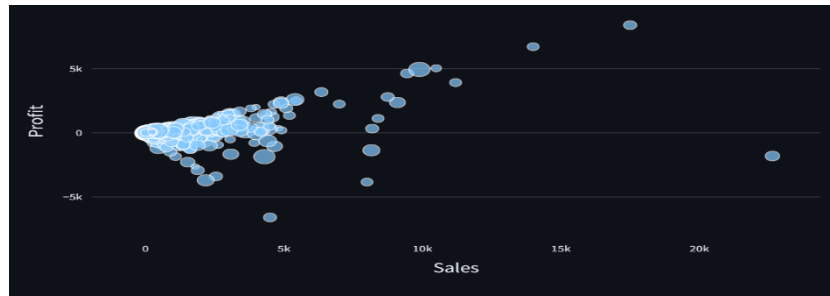
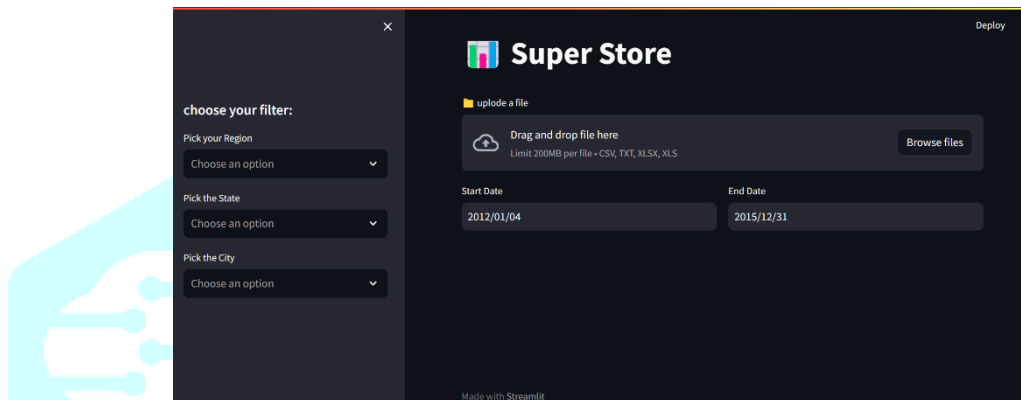


Fig6. Screenshot the home pages.



This fig are included the dashboard form the superstore sales to data will be access and the date wise ,any city ,region and state wise data filter and visualization .

CONCLUSION AND FUTURE WORK:

In conclusion, using Streamlit in Python to visualize a Superstore sales dataset provides a user-friendly and interactive way to explore and understand the data. The combination of Streamlit for building web applications and Plotly Express for creating interactive visualizations makes the process straightforward and efficient.

In the next part, I'll be focusing on the profit analysis and how the revenue and profit are in different States, Segments, Region, Categories, etc. We can predict the sales of the next 7 days from the last date of the Training dataset by using **Time Series Analysis**.

REFERENCES:

1. <https://aman2457.medium.com/superstore-sales-data-analysis-and-visualization-using-python-and-plotly-part1-fc1f71a7289a>.
2. <https://www.kaggle.com/code/daviribeirodossantos/superstore-dataset-by-davi-santos>.
3. https://www.irjmets.com/uploadedfiles/paper//issue_4_april_2023/36572/final/fin_irjmets1682186035.pdf.
4. <https://python.plainenglish.io/creating-interactive-analytics-dashboards-using-pythons-streamlit-43a770bc3194>.

5. <https://pianalytix.com/superstores-data-analysis-in-machine-learning/>
6. <youtube.com/watch?v=7yAw1nPareM&si=XPc5GPI61NczLK4z>

