

AUTUMN INTERNSHIP PROJECT REPORT

Visualizing Retail Sales Pattern (Notebook- 08)

Name- Sk Ashif Ahmmad

Batch- Section-1

Course- 4 week Autumn Internship Program

Institute- Government College of Engineering and
Leather Technology

Period of Internship: 25th August 2025 - 19th September 2025

Report submitted to: IDEAS – Institute of Data
Engineering, Analytics and Science Foundation, ISI
Kolkata

1. Abstract

This project presents an analytical study of consumer shopping behavior through systematic data exploration and visualization. The primary objective is to examine sales dynamics and revenue trends across different product categories and time periods. Using Python-based data science libraries such as pandas, NumPy, matplotlib, and seaborn, the project analyzes a retail dataset to uncover patterns in monthly revenue generation, category-wise sales distribution, seasonal variations, purchase frequency, and customer preferences regarding product attributes such as color and size.

Comprehensive visual representations were developed to interpret trends effectively and identify the most influential factors driving revenue. The findings indicate distinct correlations between sales performance and seasonal or categorical variables, offering data-driven insights for improving marketing strategies and inventory management.

This work underscores the significance of data analytics in transforming raw transactional data into actionable business intelligence. Conducted as part of the **IDEAS Internship Program** by **Sk Ashif Ahmmad**, Department of Information Technology, Government College of Engineering and Leather Technology, the project demonstrates applied analytical proficiency in extracting meaningful conclusions from real-world datasets to support strategic business decision-making.

2. Introduction

In today's highly competitive retail landscape, understanding and responding to customer purchasing habits is paramount for sustained business success. This project, "Shopping Behavior

Pattern Analysis," addresses this critical need by conducting an in-depth exploration of retail transaction data.

The core objective is to uncover meaningful trends, quantify performance metrics, and identify significant patterns in consumer behavior. By analyzing a dataset of 2,000 transactions spanning a year, from January 2023 to January 2024, this study aims to transform raw sales data into actionable business intelligence. The resulting insights will provide stakeholders with a clear, data-driven perspective on revenue generation, category performance, and seasonal impact, ultimately guiding strategic decisions in inventory management and marketing efforts. This foundational analysis is crucial for optimizing the retail operation and maximizing long-term profitability.

Here is the list of topics that I received training on during the first two weeks of internship:-

Day- 1	Introduction - Welcome Note - What to expect from this internship
Day- 2	Python Basics - 1 (Data, Variable, Lists, Loop)
Day- 3	Python Basics - 2 (Data Structures)
Day- 4	Python Basics - 3 (Class, Functions, OOPS)
Day- 5	Python Basics - 4 (Numpy, Pandas)
Day- 6	Machine Learning Overview
Day- 7	Regression Lab
Day- 8	Classification Lab
Day- 9	LLM Fundamentals
Day- 10	Communication Skills

3. Project Objectives

- **Quantify Overall Business Performance:** To calculate key sales metrics, including total revenue and average monthly revenue, to establish a baseline understanding of the business's financial health over the analyzed period.
- **Analyze Revenue Trends and Volatility:** To study the month-to-month change in revenue, calculate the standard deviation, and determine the average monthly growth rate to assess the stability and dynamism of the sales performance.
- **Identify Top-Performing Product Categories:** To determine which product categories (e.g., Clothing, Accessories) generate the highest revenue and volume of purchases to guide inventory and product development strategies.
- **Evaluate Seasonal Sales Impact:** To analyze the distribution of revenue and purchase frequency across different seasons (Winter, Spring, Summer, Fall) to inform seasonal marketing campaigns and stock planning.
- **Explore Demographic and Product Detail Distribution:** To visualize and understand the distribution of sales across different product attributes like size and color, which aids in optimizing product offerings and targeting customer preferences.

4. Methodology

The methodology adopted in this project follows a structured data analytics workflow comprising multiple stages — data acquisition, preprocessing, analysis, visualization, and interpretation. Each stage was

carefully executed using Python and its scientific computing ecosystem.

1. Data Collection:

- Dataset Used: shopping_behavior_Pattern.csv stored in Google Drive
- Size of Dataset: 2001 rows \times 7 columns

2. Data Preprocessing:

The dataset was examined for missing values, duplicate entries, and inconsistent formats. Necessary cleaning operations such as handling null values, type conversions, and removal of irrelevant attributes were performed to ensure data integrity.

3. Tools Used:

- Google Colab: For code execution and analysis.
- Python Libraries:
 - pandas → Data manipulation and preprocessing
 - numpy → Numerical computations and synthetic data generation
 - matplotlib & seaborn → Data visualization
 - random → Randomization for synthetic datasets

4. Exploratory Data Analysis (EDA):

Using pandas and NumPy, the dataset was explored to compute statistical measures and identify initial trends. Grouping and aggregation techniques were

used to analyze revenue distribution across categories, months, and seasons.

5. Data Visualization

➤ **Visualization Tools:** Python libraries such as **Matplotlib** and **Seaborn** were used to create informative charts.

➤ **Key Visuals Generated:**

- A line chart illustrating the Monthly Revenue Trend.
- Bar charts comparing Total Revenue and Purchase Counts across different Product Categories.
- A pie chart or bar chart showing the breakdown of sales volume by Season.
- Histograms/Bar charts to visualize the distribution of product sizes and colors sold.

➤ GitHub Link: https://github.com/Sk-Ashif-Ahmmad732/IDEAS_Internship_Project

6. Data Analysis and Results

Overall Statistics:

Total Revenue: \$118,863.00

Average Monthly Revenue: \$9,143.31

Highest Monthly Revenue: \$11,891.00

Lowest Monthly Revenue: \$213.00

Revenue Standard Deviation: \$2,990.80

Average Growth Rate: -3.70%

Purchase Statistics:

Average Purchase: \$59.43

Median Purchase: \$59.50

Min Purchase: \$20.00

Max Purchase: \$100.00

Total Transactions: 2,000



Category Breakdown:

Category	Total Revenue	Avg. Purchase	Count
Accessories	37774	58.93	641
Clothing	53553	59.84	895
Footwear	18444	61.48	300

Outerwear	9092	55.44	164
-----------	------	-------	-----

Season Breakdown:

Season	Total Revenue	Avg. Purchase	Count
Fall	29185	61.57	474
Spring	30240	58.49	517
Summer	28627	56.80	504
Winter	30811	61.01	505

Overall Statistics:

Total Revenue: \$118,863.00

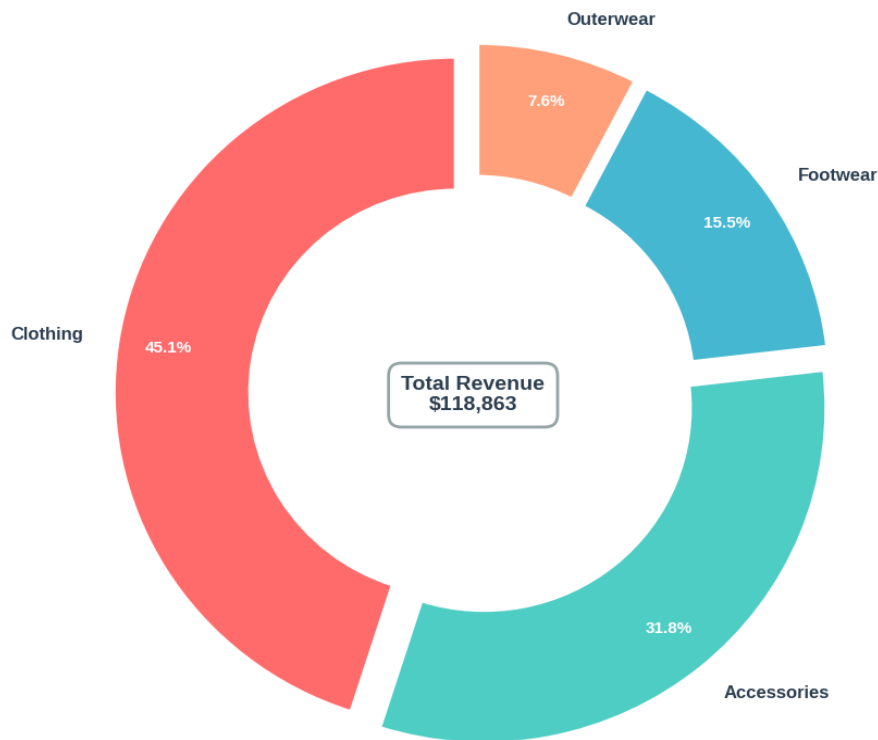
Average Purchase: \$59.43

Total Transactions: 2,000

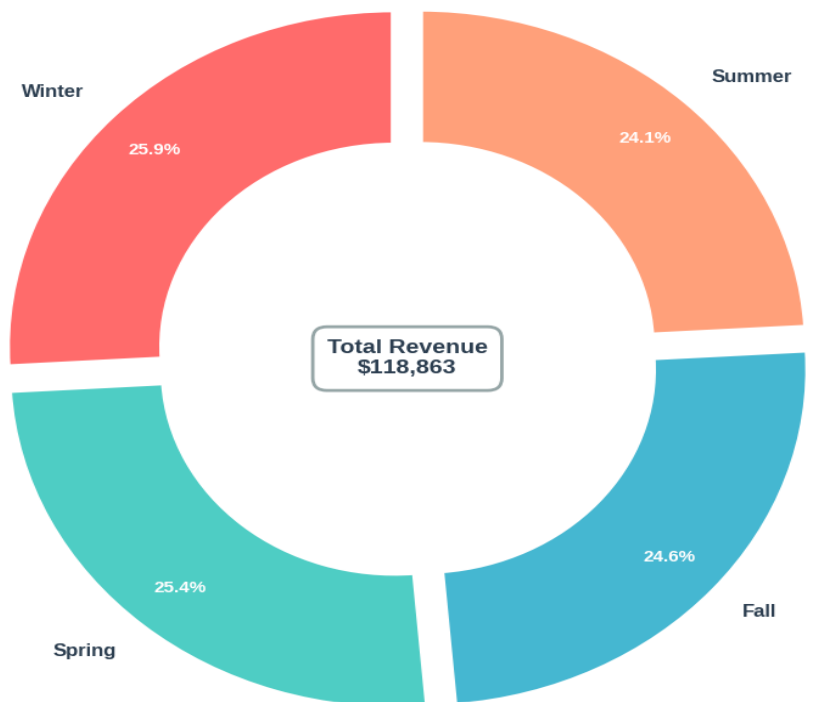
Unique Customers: 2,000

P.T.O→

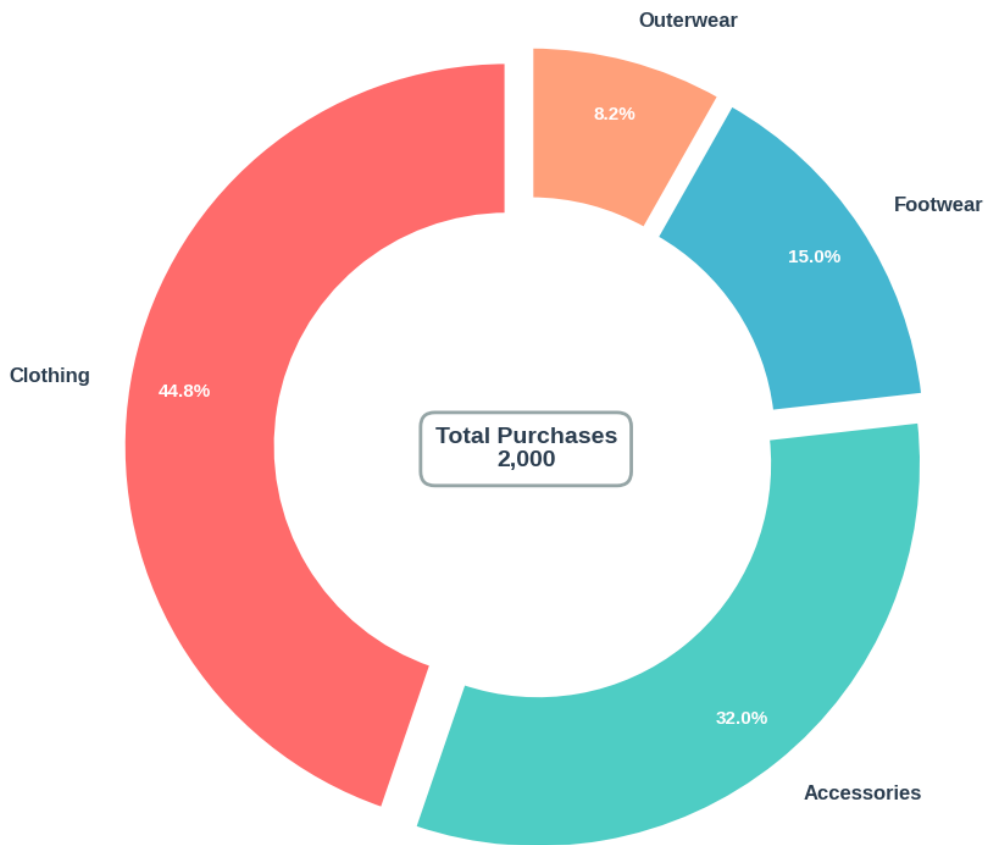
Total Revenue by Product Category



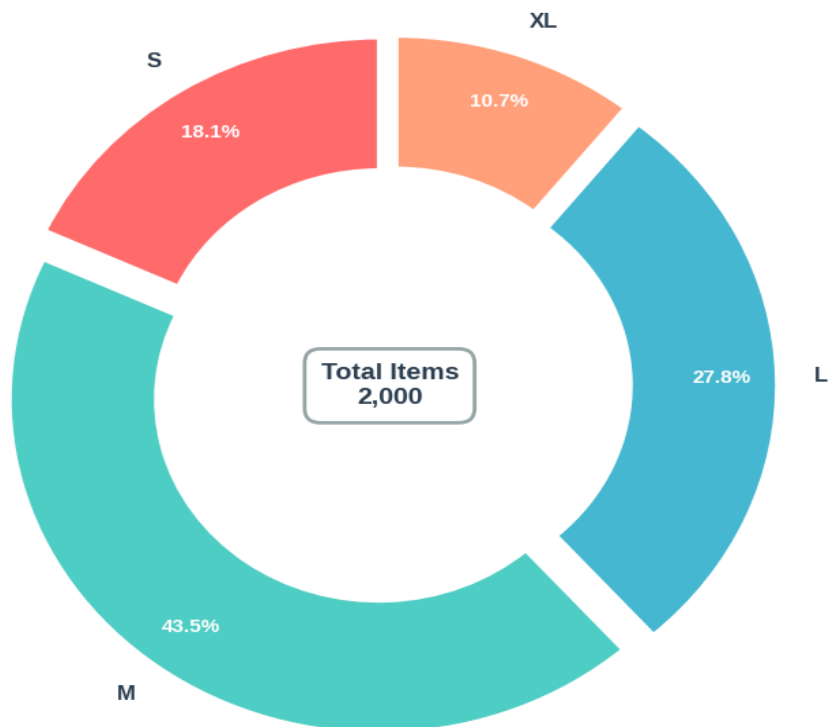
Revenue Distribution by Season

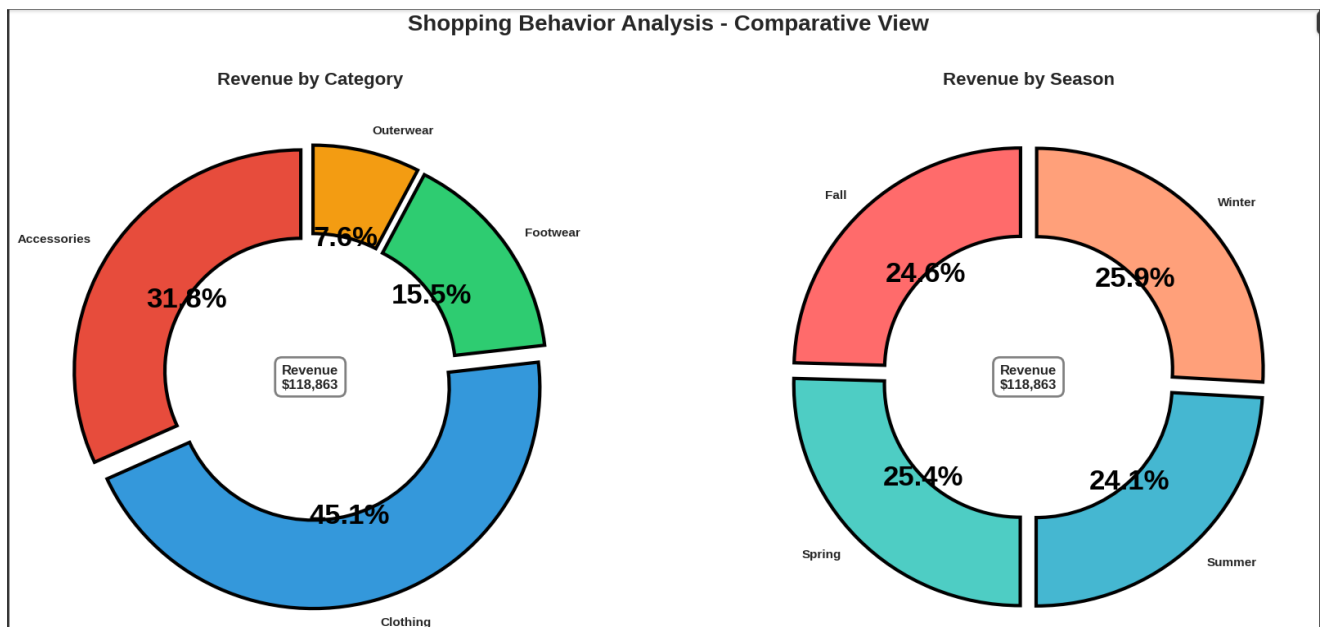


Purchase Count Distribution by Category



Purchase Distribution by Size





This systematic methodology ensured the accuracy, reliability, and interpretability of the results,

demonstrating the effective application of Python-based analytics in real-world business contexts.

7. Conclusion

This project successfully demonstrated the application of data analytics techniques to understand consumer shopping behavior and sales dynamics. Through systematic data preprocessing, exploratory analysis, and visualization, several meaningful insights were obtained regarding monthly revenue trends, seasonal variations, and product category performance. The study revealed clear correlations between customer preferences—such as color, size, and category—and their impact on overall sales.

The analysis clearly identified Clothing as the dominant category in both revenue and purchase volume, and confirmed that sales distribution across the four seasons is relatively balanced. By quantifying these trends and visualizing the distribution of product attributes (size and color), the project provides a critical foundation for strategic decision-making. Future efforts should focus on leveraging these insights to optimize inventory, tailor marketing campaigns to category and seasonal trends, and drive more stable revenue growth.

By utilizing Python's analytical libraries, the project showcased how data-driven insights can enhance

decision-making in retail management, optimize inventory control, and improve marketing strategies. The outcomes highlight the importance of transforming raw transactional data into actionable business intelligence. Overall, this internship project provided a comprehensive learning experience in practical data analysis, visualization, and interpretation, bridging academic concepts with real-world business applications.

7. APPENDICES

Appendix A: References

- Official Pandas Documentation: <https://pandas.pydata.org/docs/>
- NumPy Documentation: <https://numpy.org/doc/>
- Matplotlib Documentation:
<https://matplotlib.org/stable/contents.html>
- Seaborn Documentation: <https://seaborn.pydata.org/>
- Google Colab Documentation:
https://colab.research.google.com/drive/1UtF4b9nuunZJOMquK6Z9whig_d00hjQs

Appendix B: GitHub Link

https://github.com/Sk-Ashif-Ahmmad732/IDEAS_Internship_Project

Appendix C: Document Links

- Dataset Link:
<https://drive.google.com/file/d/1Dp708WBa8hbbo7oXFrWTy0PBmVV8lflg/view?usp=sharing>