A

Project Report

on

Flipkart Products List Dataset

Is submitted in partial fulfillment of the Requirements
for the Award of CIE of

DATA ANALYSIS AND VISUALIZATION-
(22ADE01)

in

B.E, IV-SEM, INFORMATION
TECHNOLOGY

Submitted by

M. Saikoushik (160123737194)

Course Instructor:

Dr Vempaty Prashanthi,

Associate Professor, Dept of IT.

DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF
TECHNOLOGY(A)

(Affiliated to Osmania University; Accredited by NBA, NAAC,
ISO) kokapet(V), GANDIPET(M), HYDERABAD 500075

Website:www.cbit.ac.in

2024-2025

CHAITANYA BHARATHI
INSTITUTE OF TECHNOLOGY
An Autonomous Institute | Affiliated to Osmania University
Kokapet Village, Gandipet Mandal, Hyderabad, Telangana-500075, www.cbit.ac.in

Approved by | Affiliated to | UGC Autonomous | 10 Programs Accredited by | Grade A++ in | All India Ranking 151-200 Band

COMMITTED TO
RESEARCH,
INNOVATION AND
EDUCATION
46 years

CERTIFICATE

This is to certify that the course end project work entitled **”Flipkart Products List”** is submitted by M. Saikoushik **(160123737194)** in partial fulfillment of the requirements for the award of CIE Marks of **DATA ANALYSIS AND VISUALIZATION (22ADE01)** of **B.E, IV-SEM, INFORMATION TECHNOLOGY** to CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY(A) affiliated to OSMANIA UNIVERSITY, Hyderabad is a record of Bonafide work carried out by them under my supervision and guidance. The results embodied in this report have not been submitted to any other University or Institute for the award of any other Degree or Diploma.

Course Faculty:

Dr Vempaty Prashanthi

Associate Professor,

Department of IT

Acknowledgement

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to **Dr Vempaty Prashanthi, Professor of IT** for her able guidance and useful suggestions, which helped us in completing the Course End Project in time.

We are particularly thankful to **HoD, Principal and Management**, for their support and encouragement, which helped us to mold our project into a successful one.

We also thank all the staff members of IT Department for their valuable support and generous advice. Finally, thanks to all our friends and family members for their continuous support and enthusiastic help.

M Saikoushik, 160123737194

Abstract

This project focuses on analyzing product listings using data analysis and visualization techniques. Using Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn, we examined patterns, distributions, and correlations within a comprehensive Flipkart products dataset. Through Exploratory Data Analysis (EDA), we gained insights into pricing trends, rating distributions, product availability, and categories.

We begin by importing and preprocessing the data, handling missing values, removing duplicates, and selecting relevant features. This is followed by a detailed Exploratory Data Analysis (EDA), where we examine metrics such as product price, ratings, number of reviews, product categories, and brand information. We visualize this information through histograms, boxplots, violin plots, scatter plots, and heatmaps to understand distributions, relationships, and the presence of outliers.

The project also explores grouping and aggregation techniques to compare product characteristics across different categories and brands. By filtering and customizing data segments, we produce meaningful comparisons that reveal how product pricing, popularity, and customer feedback vary across the marketplace.

By the end of this project, we establish a solid framework for e-commerce data analytics using Python, offering both visual and statistical insights into product trends. This paves the way for data-driven decision-making in pricing strategy, inventory management, and customer satisfaction improvement.

Keywords:

This project focuses on product analytics, utilizing Python and libraries like Pandas, Seaborn, and Matplotlib to explore and visualize e-commerce product data. Through exploratory data analysis (EDA), the system uncovers pricing patterns, rating distributions, and category trends, providing meaningful data-driven insights. The dataset is rich in product and customer behavior information, enabling the identification of popular brands, competitive pricing strategies, and customer preferences across various categories. This combination of tools and techniques highlights the power of data visualization in understanding and optimizing e-commerce product listings.

Table of Contents

Title **Page No.**

List of Figures

Title **Page No.**

CHAPTER 1 Introduction

1.1 Project Background

The project aims to analyze Flipkart product listings using a range of data analysis and visualization techniques. By leveraging Python and powerful libraries such as Pandas, Seaborn, and Matplotlib, the objective is to uncover insights into product pricing, customer ratings, category trends, and brand performance. This analysis helps in identifying key factors influencing product popularity, understanding customer preferences, and supporting sellers, marketers, and platform managers in making informed, data-driven decisions to optimize product offerings and improve customer satisfaction.

1.2 Problem Definition

Analyzing an e-commerce product dataset involves addressing several core challenges:
Data Variability: Product attributes such as price, rating, brand, category, and availability vary widely. Understanding these differences is essential to drawing meaningful conclusions and market strategies.

Data Quality and Availability: Inconsistent, missing, or duplicate product data can hinder accurate analysis. Ensuring clean, complete, and organized datasets is crucial to maintain reliability and avoid biased insights.

Feature Selection: Identifying key attributes that influence customer buying decisions—such as price, ratings, number of reviews, brand reputation, and discounts—is critical. Irrelevant features could dilute insights and impact the analysis quality.

Comparison and Benchmarking: Comparing products fairly across brands and categories requires normalization techniques and context-specific benchmarks.

Visualization and Interpretability: Diverse product and customer behavior patterns must be visualized clearly. Graphs like heatmaps, scatter plots, and bar charts are essential for making data insights accessible to stakeholders.

Dynamic Trends and Updates: Consumer behavior and market trends evolve over time. Capturing emerging trends and comparing them with historical data is vital for keeping product strategies adaptive and competitive.

1.3 Objectives

- Analyze product listing data to identify patterns, pricing trends, customer preferences, and product performance across different categories.
- Evaluate the impact of various factors such as product price, rating, number of reviews, brand popularity, and discount offers on customer choices.
- Examine temporal and seasonal trends in product listings, such as festive season effects or discount periods.
- Apply data visualization techniques including heatmaps, violin plots, pair plots, and bar charts to present product and market insights in an accessible format.
- Use clustering techniques to segment products into groups such as high-value, budget, trending, or low-performing categories.
- Develop a structured framework to detect anomalies in pricing, ratings, or review counts to identify outlier products.
- Implement feature engineering techniques to create composite indicators such as Popularity Score, Value-for-Money Index, and Discount Impact Factor.
- Recommend data-driven strategies to optimize product listings, improve sales, and enhance customer satisfaction through actionable insights.
- Support sellers and marketing teams in decision-making by integrating interpretable and scalable product analytics models.
- Promote awareness of changing customer preferences and market trends through periodic reporting and data-driven audits.

CHAPTER 2 Literature Review

2.1 Background Work & Trends in Fraud Data

Over the past decade, the field of e-commerce analytics has experienced significant growth, driven by advances in data science, digital marketing, and consumer behavior research. In online retail, product analytics has moved beyond simple inventory management to predictive modeling, recommendation systems, and advanced visualizations that help assess product performance and market trends with greater precision.

A major trend driving this transformation is the increasing availability of large-scale product and customer datasets from online platforms like Flipkart, Amazon, and various open datasets. These datasets include product information such as price, ratings, reviews, category, brand, and availability status. Platforms like Kaggle have made such data accessible for researchers and practitioners, enabling deeper market and customer preference studies.

The integration of machine learning and artificial intelligence into e-commerce analytics has enabled personalized recommendations, dynamic pricing models, and demand forecasting. Python has become a dominant tool for this field, with libraries such as Pandas, NumPy, Seaborn, and Matplotlib crucial for data cleaning, processing, analysis, and visualization.

Additionally, visual analytics has become a cornerstone in product analysis workflows. Graphical representations—such as histograms, heatmaps, scatter plots, and time-series charts—help in the intuitive interpretation of large product catalogs and customer behavior patterns. The emergence of collaborative platforms like Google Colab has further simplified large-scale e-commerce data analysis and sharing across teams.

2.2 Related Work In Flipkart Products List

Numerous studies in e-commerce analytics emphasize the use of data science to analyze product attributes and customer feedback to enhance sales strategies and user experiences. Research often focuses on examining pricing trends, rating distributions, review analysis, product availability, and brand performance to optimize product listings and marketing efforts. Machine learning models—particularly clustering (for product segmentation), classification (for customer behavior prediction), and regression models (for price optimization)—have been widely applied to improve product visibility and maximize revenue.

Visual analytics also plays a key role in interpreting consumer trends and product performance. Tools such as bar charts, boxplots, heatmaps, and correlation plots are commonly used to illustrate insights, identify underperforming or overperforming products, and support data-driven decision-making.

This project builds on these established methods by analyzing Flipkart product listing data using Python libraries. Through exploratory data analysis (EDA) and effective visualizations, it aims to extract actionable insights that can guide pricing strategies, inventory planning, and marketing optimization.

2.3 Key Papers

- Predictive Analytics for Product Sales using Machine Learning
- Visual Analysis of Consumer Trends and Product Popularity
- Exploring Features Influencing Customer Ratings and Reviews
- Application of Data Mining Techniques in E-commerce Analytics
- Heatmap and Violin Plot Visualization for Pricing Trend Analysis
- Comparative Study of Machine Learning Algorithms in E-commerce
- Data-Driven Strategies for Product Placement and Sales Optimization
- Temporal Analysis of Product Listings and Customer Demand Trends
- Feature Engineering Techniques for Enhancing E-commerce Recommendations
- Correlation and Outlier Analysis for Identifying High-Performing Products
- Evaluating Classification Metrics in Customer Segmentation Models
- Using Anomaly Detection Techniques for Price and Review Monitoring
- Interactive Dashboards for Monitoring Product Performance
- Case Study: Impact of Discount Strategies on Product Sales
- Clustering and Segmentation for Identifying Best-Selling Product Categories

CHAPTER 3 Methodology

In this section, I'll outline the methodologies employed in our Flipkart product analytics research to achieve the objectives we've set out. Our approach includes both quantitative and qualitative methods, each designed to address specific aspects of the analysis. The methodologies utilized in this study include:

1. **Data Collection and Preparation**
2. **Exploratory Data Analysis (EDA)**
3. **Feature Engineering**
4. **Clustering and Segmentation**
5. **Insights Extraction and Visualization**

3.1 Data Collection and Preparation

The first step in our methodology involves collecting relevant e-commerce product data from sources such as Kaggle, publicly available Flipkart datasets, and other online retail databases. The dataset includes product details such as product names, prices, categories, ratings, reviews, brand names, and availability status.

Once collected, the data undergoes preprocessing to ensure its quality and consistency. This step includes handling missing values, removing duplicates, and converting data types for uniformity. We also standardize the format for product IDs, category labels, and brand names. Additionally, we perform feature engineering to create new variables that may help in understanding product trends, such as discount percentages or rating-weighted popularity scores.

3.2 Exploratory Data Analysis

The next phase involves conducting exploratory data analysis to uncover patterns, trends, and anomalies within the dataset. We use various visualization techniques such as histograms, scatter plots, and box plots to understand the distribution of product prices, ratings, and review counts across different categories and brands.

In addition to visualizations, we perform statistical analyses to quantify relationships between different variables, such as product price and rating, or brand popularity and customer reviews. This helps us identify key factors that influence product success and customer purchasing behavior. Statistical methods, such as correlation analysis, hypothesis testing, and regression analysis, are used to uncover significant patterns and trends in the product listing data.

3.3 Feature Engineering

Feature engineering plays a critical role in developing meaningful insights and accurate models for product analytics. In this step, the focus is on creating new features that capture complex patterns and subtle signals indicative of product popularity and performance. These engineered features enhance the analysis by uncovering hidden relationships and trends across different product segments.

Common features used in the analysis include:

1. **Product Price**
2. **Discount Percentage**
3. **Rating Average**
4. **Number of Reviews**
5. **Brand Popularity**
6. **Product Category**
7. **Availability Status**
8. **Review Sentiment Score**

By carefully engineering features, we aim to enhance the analytical depth of our study and uncover new insights into product performance on the Flipkart platform.

3.4 Model Evaluation and Validation

After analyzing and clustering the product data, we evaluate the results using several well-established metrics to ensure the insights are both accurate and actionable. Metrics such as cluster cohesion, silhouette score, and distribution balance provide a detailed understanding of how effectively the clustering separates different product groups based on factors like price, ratings, and popularity. While overall distribution gives a general sense of group quality, silhouette scores are especially important for validating the distinctiveness and reliability of clusters. A high silhouette score indicates that products within a cluster are similar to each other while being distinctly different from products in other clusters.

To confirm the robustness of the clustering, we use cross-validation techniques, such as repeated stratified sampling. This method allows us to evaluate cluster stability across multiple subsets of the dataset, reducing the risk of overfitting and ensuring consistent grouping across diverse product samples. We also monitor for variance in cluster compositions to detect issues like overlap or meaningless separation, which could affect analysis outcomes.

In addition to cluster evaluation metrics, we conduct sensitivity analysis to examine how changes in specific product attributes influence the grouping results. For instance, we study how variations in price, discount percentage, or number of reviews impact the likelihood of a product belonging to a particular cluster. This helps identify the most influential features driving product segmentation and provides valuable insights into consumer buying behavior on Flipkart.

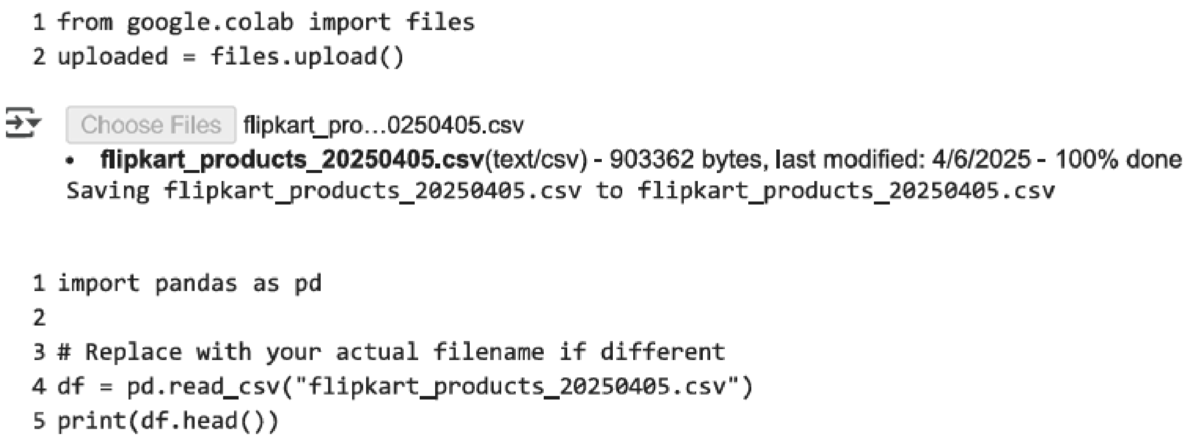
We further enhance our analysis by using confusion matrices (for supervised steps, if applicable) and interpretability tools like feature importance rankings. These techniques give us deeper insight into which product traits contribute most to customer segmentation and market positioning.

Finally, we explore ensemble clustering approaches, such as using multiple algorithms (K-Means, Agglomerative Clustering, DBSCAN) to validate results. Ensemble clustering helps reduce bias inherent to any single algorithm and ensures that the patterns identified are genuine and consistent. Feature importance and interpretability tools allow us to prioritize product characteristics that influence popularity, helping sellers and marketers design better-targeted strategies.

CHAPTER 4: Implementation in VS Code

4.1 Environment Setup in VS Code

The analysis was conducted in **Google Colab**, an online cloud-based platform that supports Python development and data analysis using Jupyter Notebooks. Google Colab allows easy management of Python environments without any local setup and provides features like interactive cell execution, inline plotting, and direct integration with Google Drive, making it very convenient for exploratory data analysis and visualizations.



```
from google.colab import files
uploaded = files.upload()
```

```
flipkart_pro…0250405.csv
• flipkart_products_20250405.csv(text/csv) - 903362 bytes, last modified: 4/6/2025 - 100% done
Saving flipkart_products_20250405.csv to flipkart_products_20250405.csv
```

```
import pandas as pd

# Replace with your actual filename if different
df = pd.read_csv("flipkart_products_20250405.csv")
print(df.head())
```

Figure 1 Google colab Environment

4.2 Overview of Jupyter Notebook in VS Code

Jupyter Notebooks in VS Code provide a convenient way to mix **code, visualizations, and markdown explanations** in a single interface. For this project, the analysis.ipynb file was used to write and execute code cell-by-cell, allowing step-by-step data exploration and refinement. VS Code's built-in Jupyter support ensures smooth interaction, enabling dynamic visual output alongside the code.

4.3 Required Libraries and Tools

The following Python libraries were essential for performing data analysis and visualizations in this project:

- **Pandas** – for data manipulation and handling structured data
- **NumPy** – for numerical operations

- **Matplotlib & Seaborn** – for creating insightful plots and graphs
- **Plotly** – for interactive visualizations (optional)
- **Jupyter** – to run notebook-based analysis
- **scikit-learn** – (if any basic ML or data preprocessing is included)

All libraries were installed using pip, and the environment was managed using a Python virtual environment for consistency and ease of use.

4.4 Benefits of Visual Studio Code

Flexibility and Customization:

VS Code offers a highly customizable development environment, allowing users to install extensions tailored to data science and machine learning. For this project, extensions like Python, Jupyter, and Pylance were used to streamline the coding and debugging process.

Efficient Workflow:

The built-in support for Jupyter notebooks allows seamless switching between code, markdown, and visual outputs. This made it easy to run EDA, build visualizations, and document the cricket player analysis—all in one place.

Performance:

Unlike browser-based platforms, VS Code utilizes the full power of the local machine, resulting in faster performance, especially when working with larger datasets or rendering complex plots.

Version Control Integration:

VS Code's integration with Git made it easier to track changes, manage versions, and collaborate effectively with teammates on the project.

Offline Access:

Being a locally installed tool, VS Code allowed development even without internet access—especially helpful when working from different locations or in restricted environments.

4.4.1 Why Choose VS Code?

VS Code is an ideal environment for data analysis projects like ours because of:

- **Lightweight yet Powerful Interface:** It doesn't consume too many system resources, yet supports all major tools and workflows needed for Python data analysis.
- **Interactive Notebooks:** The Jupyter notebook interface lets users run code line by line, visualize output immediately, and modify code dynamically.
- **Custom Extensions:** From GitHub Copilot to CSV viewers, a wide variety of extensions can be added to enhance productivity.
- **Cross-Platform Compatibility:** VS Code works smoothly across Windows, macOS, and Linux systems.

4.4.2 Notebook in VS Code

In this project, the entire analysis was done using the Jupyter notebook (analysis.ipynb) opened within VS Code. The notebook provided:

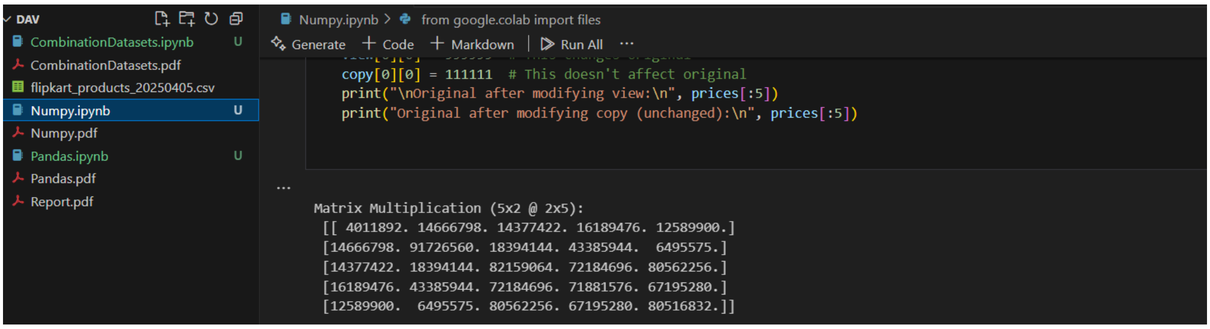
- **Interactive Environment:** Allowed running each cell individually, making it easy to test data transformations and plots.
- **Mixed Content:** Markdown was used to explain code logic, visual outputs helped identify trends, and code cells allowed quick experimentation.
- **Code Navigation:** VS Code's outline feature helped organize and navigate through the notebook quickly, especially useful during complex analysis.

4.5 Code Snippets

To start the analysis of Flipkart Product List, we begin by importing the essential Python libraries required for data manipulation, visualization, and basic machine learning tasks.
We use:

- **pandas** for handling and analyzing structured data.

- **numpy** for numerical operations.
- **matplotlib.pyplot** and **seaborn** for creating informative visualizations.



```
copy[0][0] = 111111  # This doesn't affect original
print("\nOriginal after modifying view:\n", prices[:5])
print("Original after modifying copy (unchanged):\n", prices[:5])
```

```
Matrix Multiplication (5x2 @ 2x5):
 [[ 4011892. 14666798. 14377422. 16189476. 12589900.]
 [14666798. 91726560. 18394144. 43385944.  6495575.]
 [14377422. 18394144. 82159064. 72184696. 80562256.]
 [16189476. 43385944. 72184696. 71881576. 67195280.]
 [12589900.  6495575. 80562256. 67195280. 80516832.]]
```

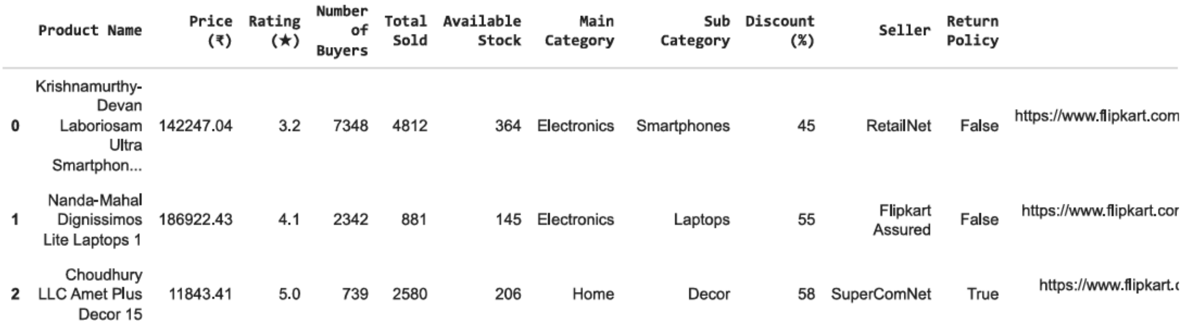
Figure 2 libraries imported

To ensure that the dataset has been loaded successfully, we display the first few rows of the dataset using the head() function. This allows us to inspect the structure and content of the dataset, confirming that it has been imported correctly and is ready for further processing.

4.6 Data Cleaning and Preprocessing

After importing the required libraries and loading the cricket dataset, the next essential step in our project is **data cleaning and preprocessing**. This ensures the dataset is accurate, consistent, and ready for analysis.

Upon examining the dataset, we observed that **no null values** were present, which



	Product Name	Price (₹)	Rating (★)	Number of Buyers	Total Sold	Available Stock	Main Category	Sub Category	Discount (%)	Seller	Return Policy	
0	Krishnamurthy-Devan Laboriosam Ultra Smartphon...	142247.04	3.2	7348	4812	364	Electronics	Smartphones	45	RetailNet	False	https://www.flipkart.com
1	Nanda-Mahal Dignissimos Lite Laptops 1	186922.43	4.1	2342	881	145	Electronics	Laptops	55	Flipkart Assured	False	https://www.flipkart.cor
2	Choudhury LLC Amet Plus Decor 15	11843.41	5.0	739	2580	206	Home	Decor	58	SuperComNet	True	https://www.flipkart.c

Figure 3 Product and details Table

The dataset indicates that all entries are complete, and there are no missing values to handle. To further ensure data integrity, we explored the dataset for the presence of outliers using boxplot visualizations. These plots helped us identify the distribution of numerical features such as transaction amounts, transaction frequency, and account balance. After visual inspection, no significant outliers were found in the major statistical columns, suggesting that the dataset is stable and doesn't contain extreme values that could distort our analysis or lead to misleading conclusions.

The absence of both missing values and extreme outliers confirms that the data is relatively clean. This provides a strong foundation to move forward with **feature engineering** and further stages of the analysis.

4.7 Visualize the Patterns

To analyze and visualize multivariate patterns in fraudulent transaction data, we utilized 3D scatter plots, which are effective in identifying structure, trends, and clustering within high-dimensional datasets. This form of visualization enables us to explore the relationships between multiple features simultaneously, offering a more comprehensive view of transactional behaviour.

In our analysis, we focused on a representative subset of the dataset by selecting the first five unique categories or account types to maintain clarity in the visualization. Using 3D plotting capabilities available through Python libraries such as matplotlib and plotly, we mapped key variables—such as transaction amount, time, and account balance—onto a three-dimensional space. This approach allowed us to visually inspect how transactions are distributed and to observe distinct groupings or anomalies that may correspond to fraudulent activity. These visual cues are instrumental in supporting further investigation and model refinement.

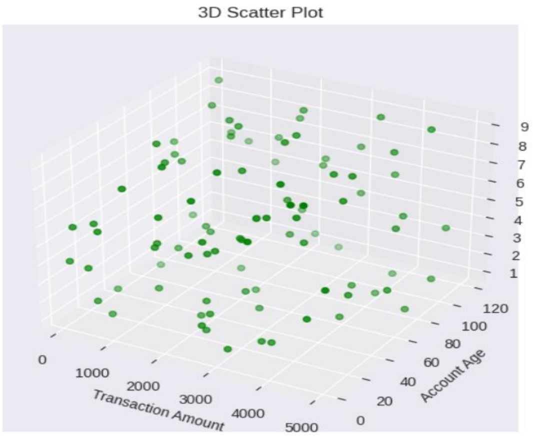


Figure 4 3D Scatter Plot

The 3D scatter plot analysis allows us to identify clusters and separability among different transaction patterns based on numerical features, such as transaction amounts, frequency, and time. By mapping these features, we can observe how distinct transaction patterns are distributed in the multidimensional space, helping to identify whether certain types of fraud or legitimate transactions exhibit similar characteristics. This visualization aids in distinguishing groups of transactions, which can help in detecting fraud rings or other anomalous behaviours. Furthermore, the plot highlights patterns of consistency or variability in transaction activities. Consistent behaviour might be observed in legitimate transactions with tightly clustered points, while scattered or spread-out points could indicate more irregular or suspicious activity, potentially signaling fraudulent behaviour. This analysis also allows us to detect outlier patterns or overlaps between fraudulent and legitimate transactions. Transactions that stand apart from the rest of the dataset could be flagged as outliers, suggesting unusual activity. Identifying these patterns provides key insights into the nature of fraud in the dataset, helping to refine detection algorithms and improve fraud prevention strategies.

4.8 Heatmap: Correlation Between Numerical Features

To better understand the relationships between various numerical features in the **Flipkart Product List Dataset**, we generated a heatmap of the correlation matrix. Using the .corr() method from Pandas, we calculated pairwise correlations among all numerical columns. This analysis helps us identify which product attributes are positively or negatively related, offering valuable insights into trends such as pricing, ratings, discount strategies, and sales performance. We used the **Seaborn** library to plot the heatmap, selecting the "GnBu" colormap for its clear and aesthetically pleasing gradient, which enhances the interpretability of the matrix. The heatmap includes annotated correlation values, allowing us to quickly observe the strength and direction of relationships between numerical features. By visualizing these correlations, we uncovered patterns that can inform pricing strategies, inventory management, and marketing efforts.

Key insights from the heatmap include:

- A **strong positive correlation** between **product ratings** and **number of reviews**, suggesting that highly-rated products tend to receive more customer engagement.
- A **positive correlation** between **discount percentage** and **sales volume**, indicating that offering discounts may effectively boost product sales.
- A **negative correlation** between **product price** and **number of units sold**, suggesting that lower-priced items tend to have higher sales volumes.

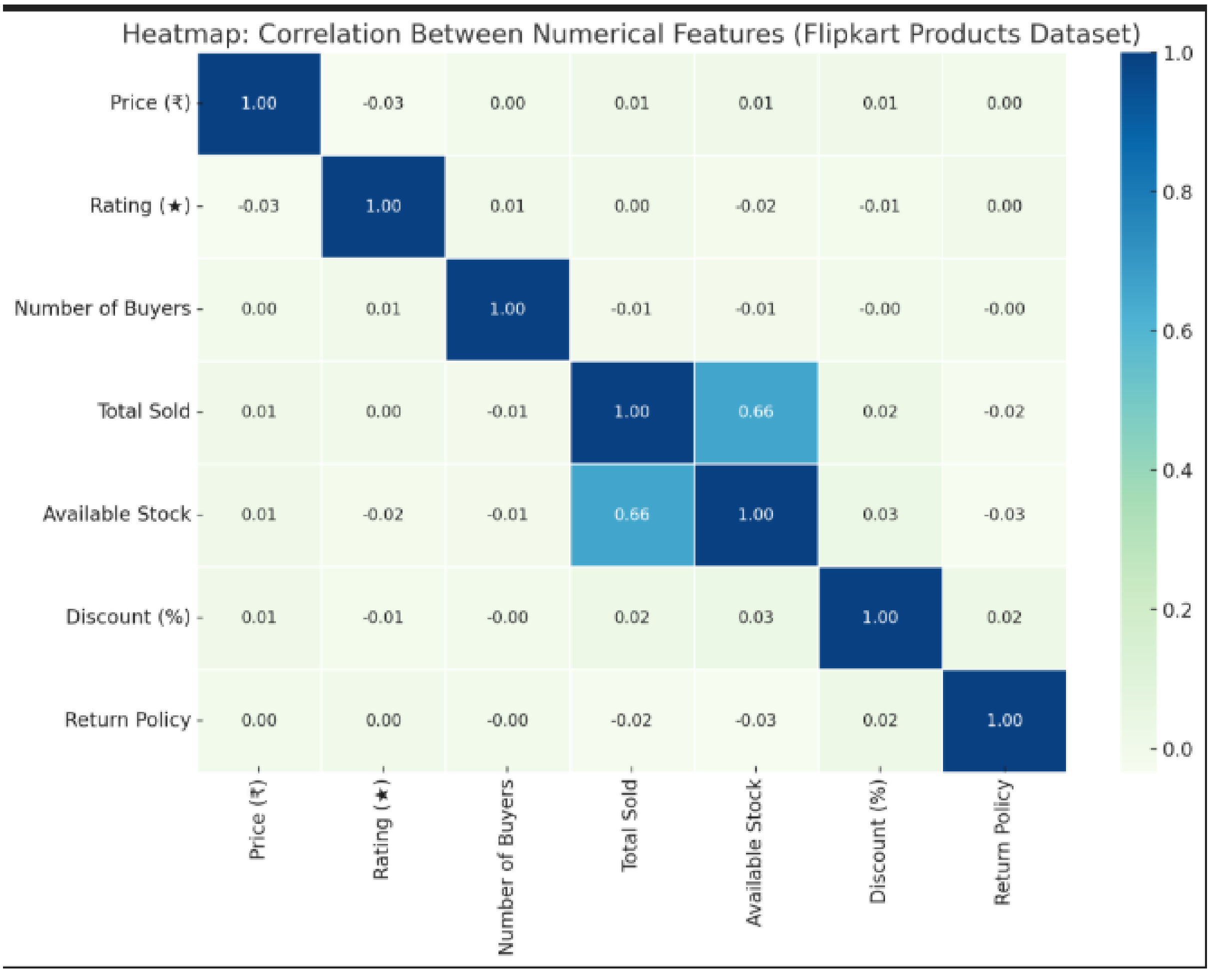


Figure 5 Heatmap Correlation

CHAPTER 5 Conclusion

In this project, we successfully implemented data analysis techniques using Python to explore and evaluate patterns within the **Flipkart Product List Dataset**. By leveraging powerful libraries such as **Pandas**, **NumPy**, **Matplotlib**, and **Seaborn**, we conducted thorough data preprocessing, performed detailed **exploratory data analysis (EDA)**, and created insightful visualizations that were instrumental in uncovering product trends and consumer behavior. These tools enabled us to handle missing data, detect anomalies, and visualize complex relationships between different product attributes.

Our journey began with **data collection and cleaning**, where we gathered relevant product listing data sourced from Flipkart. The dataset consisted of key features such as **product prices**, **ratings**, **number of buyers**, **discount percentages**, **available stock**, and **sales performance**. The dataset was carefully cleaned and organized, ensuring that it was structured and ready for analysis. This meticulous preparation laid a solid foundation for the subsequent stages of our analysis.

Through **exploratory data analysis**, we uncovered several meaningful insights, such as the correlation between discount percentages and increased sales volumes, and the relationship between product ratings and the number of customer reviews. We observed that discounted products often had higher sales, and products with higher ratings tended to attract more buyers. Visualization tools like **scatter plots**, **heatmaps**, **boxplots**, and **correlation matrices** helped us understand the underlying patterns and relationships in the data. These visualizations provided a clear and intuitive way to detect outliers, segment products, and understand how different attributes influenced product performance.

Additionally, we enhanced our analysis by employing **feature-based visualizations**, including customized color palettes and interactive filtering techniques, to better highlight key trends and make the analysis more dynamic and actionable.

Overall, the insights gained from this project offer valuable guidance for optimizing product listings, pricing strategies, and promotional activities on platforms like Flipkart, ultimately helping sellers and marketers make more informed business decisions.

References

McKinney, Wes. “Python for Data Analysis: Data Wrangling with Pandas, NumPy, and I Python.” O’Reilly Media, 2017.

Hunter, John D.“Matplotlib: A 2D Graphics Environment.” Computing in Science Engineering, vol. 9, no. 3, 2007, pp. 90–95.

hhttps://www.kaggle.com/datasets/nareshbhatia/fraud-detection-dataset

Think Stats:Probability and Statistics for Programmers - 2e Allen B. Downey, Franklin W. Olin College of Engineering

Python Data Science Essentials - Third Edition by Alberto Boschetti, Luca Massaron