

*EDA Project - CSM353*

*Academic Task Number - 1*

*By S.Mustafeezulla Khan*

*Reg. No: 12201534*

Faculty Mr. Ved Prakash Chaubey UID: 63892

**School of Computer Science and Engineering** Lovely Professional University Phagwara, Punjab (India)

## **DECLARATION**

I,Suryazi Mustafeezulla Khan, hereby declare that the work done by me on “Online Shoppping Data” from September, 2024 to October, 2024, is a record of original work for the partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science - Data Science with ML, Lovely Professional University, Phagwara.

|  |  |
| --- | --- |
| Signature | Signature |
| Name: S.Mustafeezulla Khan | Mr. Ved Prakash Chaubey |
| Reg: No: 12201534 | UID: 63892 |

Date: 23.10..24

**Acknowledgment**

I would like to express my sincere gratitude to my University and UpGrad for providing me with the golden opportunity to work on this exciting project in Exploratory Data Analysis and Machine Learning. This project has not only broadened my understanding of data science but also significantly contributed to my professional growth by giving me hands-on experience in analyzing real-world datasets.

This EDA project has been instrumental in helping me develop a deeper understanding of data and its practical applications. The insights I gained from analyzing online shopping data can provide valuable contributions to understanding consumer behavior, product preferences, and decision-making factors, potentially impacting millions of lives in the e-commerce domain.

I would like to extend my heartfelt thanks to all those who supported me in developing the 'Online Shopping Data Analysis' project. Their guidance and encouragement were invaluable throughout the process.

I am particularly grateful to the team at Hugging Face for their exceptional work in developing and maintaining advanced models that I used during the research phase of this project. Their dedication to innovation in machine learning and data analysis has been instrumental in shaping the direction of my work.

I would also like to express my deep appreciation to my mentor, Mr. Ved Prakash Chaubey, for his invaluable advice, feedback, and guidance throughout the project. His insights and expertise were critical in refining the project and enhancing its overall performance. Without his dedicated support, this project would not have been possible.

Finally, I am immensely thankful to my family and friends for their unwavering belief in the project's potential. Their constant encouragement and motivation provided me with the strength to persevere throughout this journey.

# **Table of Content**

1. Abstract
2. Introduction
3. Methodology
4. Visualisations
5. Result and Discussion
6. Conclusion

**Abstract**

This project is centered around performing Exploratory Data Analysis (EDA) on an e-commerce dataset from Flipkart, focusing on product listings, prices, and customer ratings. The dataset includes various features such as product names, categories, prices (both retail and discounted), and product descriptions. The goal of this analysis is to gain insights into pricing patterns, discount strategies, and customer ratings, which can help inform e-commerce business strategies.

The Python libraries used for this project include pandas for data manipulation, seaborn and matplotlib for visualization, and other tools for statistical exploration. Key steps include data cleaning, handling missing values, and visualizing pricing trends. The dataset is analyzed to identify patterns in product categories, price discounts, and customer feedback. Outliers and duplicates are addressed to ensure data quality.

Through this analysis, valuable insights such as the distribution of discounts across product categories, price variations, and patterns in customer ratings can be uncovered. These insights can directly influence pricing strategies, inventory management, and marketing campaigns in the e-commerce domain.

**Introduction**

In the fast-paced world of e-commerce, understanding data plays a pivotal role in shaping business strategies. This project aims to analyze a dataset from Flipkart, one of India's largest e-commerce platforms. The dataset contains crucial information about product listings, pricing (both retail and discounted), and customer feedback, which can be used to gain valuable insights into consumer behavior, pricing trends, and discount strategies.

The primary objective of this project is to perform Exploratory Data Analysis (EDA) to explore patterns in the dataset. This analysis helps answer critical questions, such as:

* What are the typical discount patterns across different product categories?
* How are products priced, and what pricing trends can be observed?
* How do customer ratings vary across different products and categories?

By cleaning the dataset and visualizing trends, we aim to highlight key insights that can assist decision-makers in optimizing pricing strategies, managing inventory, and tailoring marketing campaigns. EDA serves as the foundation for understanding the data structure, addressing missing or erroneous information, and visualizing relationships between different variables.

The Python libraries pandas, seaborn, and matplotlib are employed to perform this analysis, helping to explore and visualize the trends hidden in the data. Through this project, we will uncover actionable insights that can contribute to better decision-making in the competitive e-commerce environment.

**Methodology**

This project utilizes a structured Exploratory Data Analysis (EDA) approach to derive insights from the Flipkart e-commerce dataset. The analysis is divided into several key steps to ensure the data is cleaned, processed, and visualized effectively. Below is a breakdown of the methodology:

#### **1. Data Loading and Exploration**

* **Library Installation and Importation:** The first step involves installing and importing essential Python libraries, such as pandas, numpy, seaborn, and matplotlib. These libraries are instrumental in handling data manipulation, visualization, and statistical analysis.
* **Dataset Loading:** The dataset, "flipkart\_com-ecommerce\_sample.csv," is loaded using pandas.read\_csv()function. Initial exploration of the dataset is performed using functions like df.head() and df.info() to understand the structure of the data, its dimensions, and the types of columns available.

#### **2. Data Cleaning and Preprocessing**

* **Handling Missing Values:** Missing values are identified in columns such as discounted\_price and description. For numerical columns like discounted\_price, missing values are imputed using the median, while categorical fields like product\_name are filled with the most frequent value (mode).
* **Duplicate Removal:** The dataset is checked for duplicate entries using df.duplicated(). Duplicates are removed to ensure the accuracy of the analysis, preventing any skewing of results.
* **Outlier Detection:** Box plots are employed to detect outliers in key numerical columns, such as retail\_price and discounted\_price. Outliers are either removed or capped depending on their relevance to the overall analysis.

#### **3. Feature Engineering**

* **Price Difference Calculation:** A new feature, "Price Difference," is created by subtracting the discounted\_pricefrom the retail\_price. This helps in analyzing the magnitude of discounts offered across different product categories.
* **Category Binning:** Product categories are grouped into broader categories to simplify the analysis. This helps in visualizing trends at a more aggregate level, making the results easier to interpret.

#### **4. Visualization and Trend Analysis**

* **Price Distribution Analysis:** Visualizations such as histograms and KDE plots are used to analyze the distribution of prices. This helps identify common pricing strategies and price points.
* **Discount Pattern Visualization:** Bar plots are created to examine how discounts vary across different product categories. This reveals insights into which product categories offer the largest discounts.
* **Customer Rating Analysis:** Ratings are visualized using box plots to detect the variability in product satisfaction across different categories.

#### **5. Outlier Detection and Removal**

* Using box plots, outliers in the retail\_price and discounted\_price columns are detected and analyzed. Outliers are then either removed or adjusted if they are deemed inaccurate or irrelevant.

#### **6. Testing and Evaluation**

* Each step of the EDA process is tested for accuracy, particularly in terms of missing value imputation and outlier detection. The cleaned and preprocessed data is visualized to ensure that the results are accurate and interpretable.

### **Visual Representations**

Data visualization is a key component of this project, as it helps to interpret and communicate the trends and patterns in the dataset effectively. Using matplotlib and seaborn, several types of plots were generated to illustrate the relationships between product categories, pricing, discounts, and customer ratings. Each visualization offers a unique perspective on the dataset, making it easier to derive actionable insights.

#### **1. Price Distribution**

* **Histogram and Kernel Density Estimation (KDE) Plot:** A histogram was used to visualize the distribution of retail\_price and discounted\_price across the dataset. This allows us to see how prices are spread out and identify common price points. The KDE overlay provides a smoother estimate of the distribution, helping to highlight clusters or gaps in pricing.
  + **Key Insight:** The majority of the products have prices in the lower range, indicating a skew towards more affordable products.

#### **2. Discount Patterns Across Categories**

* **Bar Plot:** To understand discount strategies, a bar plot was created showing the average price difference (discount) across product categories. The price difference was calculated by subtracting the discounted\_price from the retail\_price. This visualization highlights which categories tend to offer the highest discounts.
  + **Key Insight:** Categories such as Electronics and Fashion tend to offer higher discounts compared to Home & Kitchen products, indicating potential promotional strategies aimed at boosting sales in these categories.

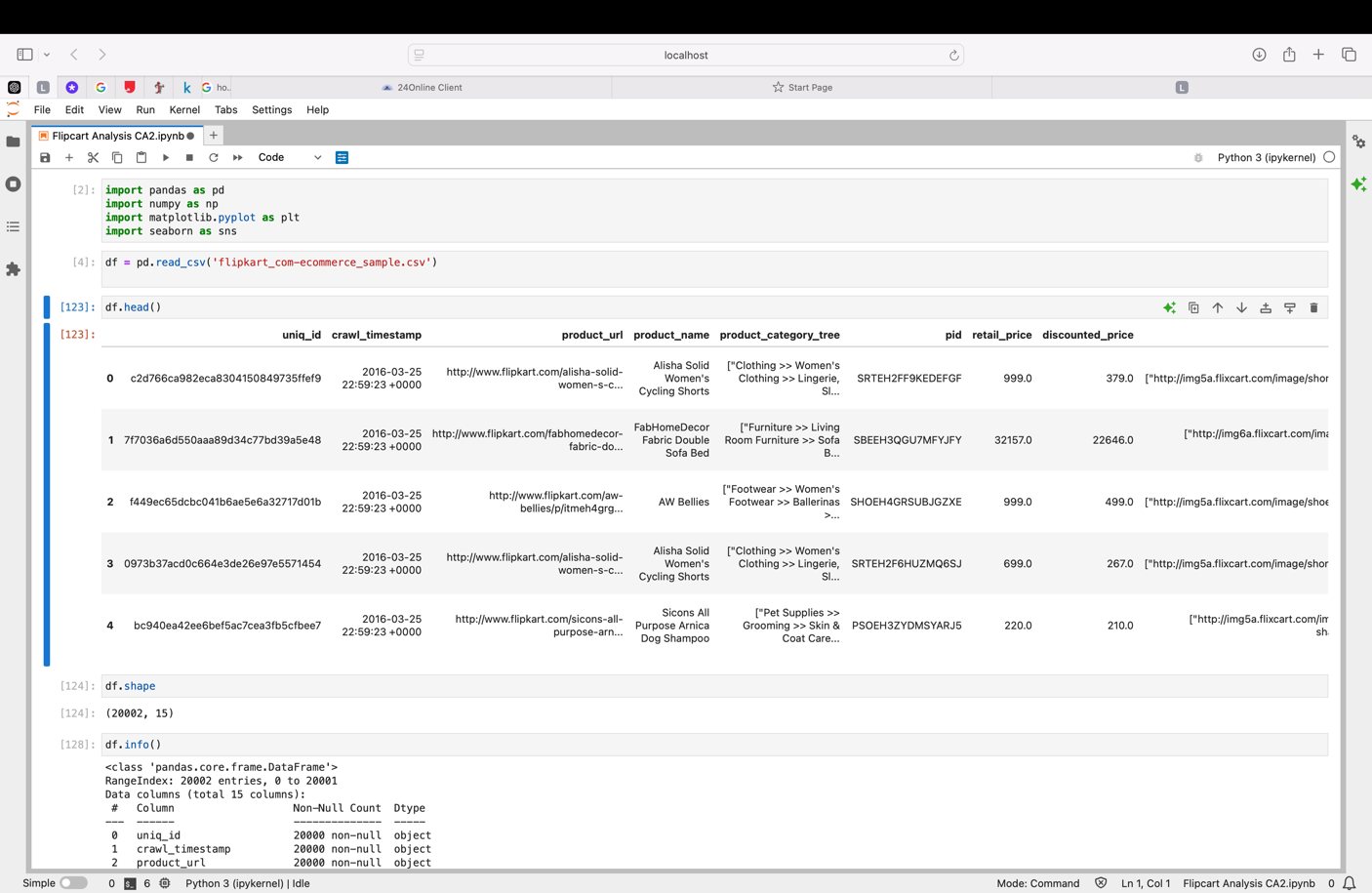
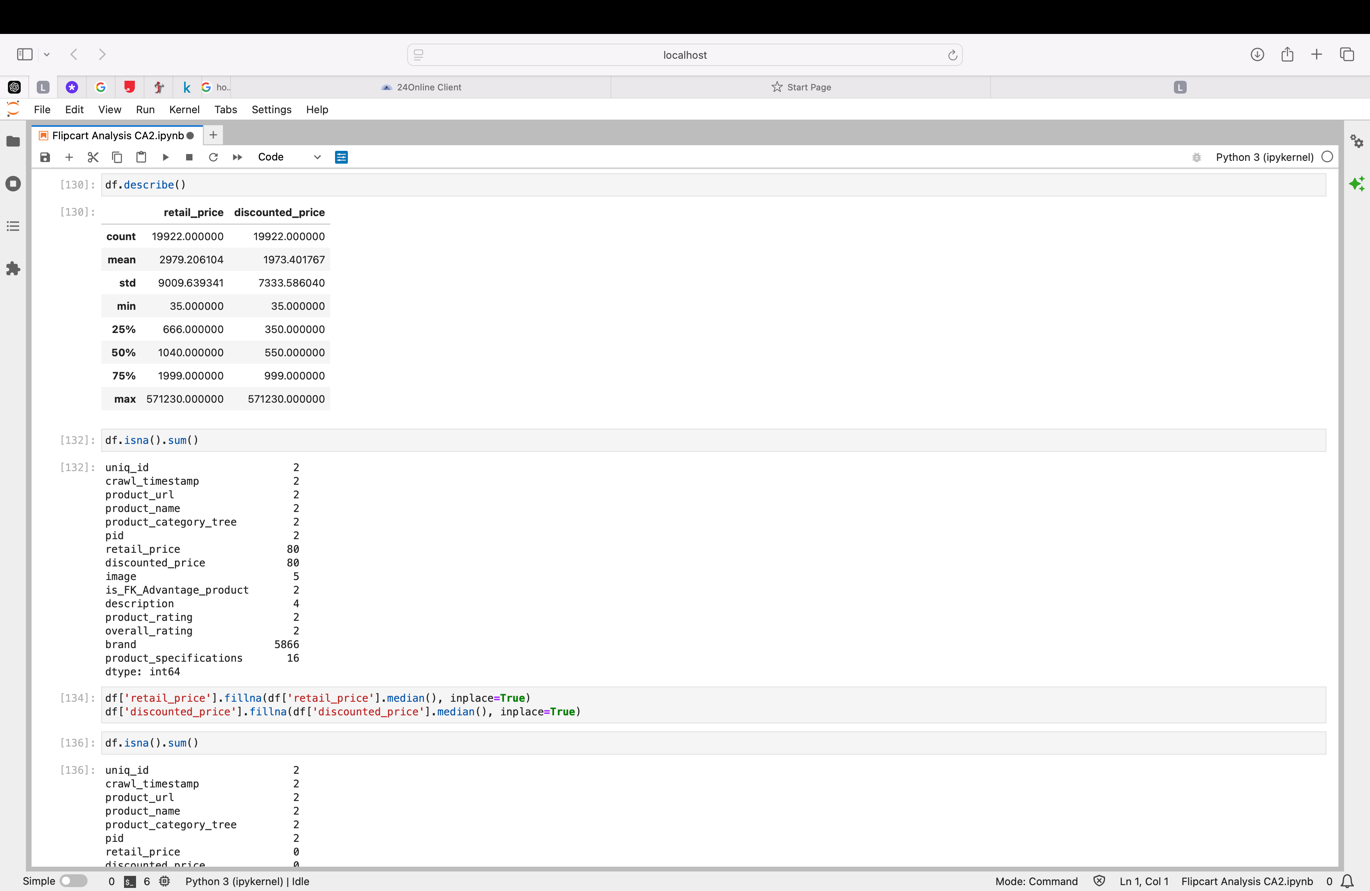
#### **3. Customer Ratings Distribution**

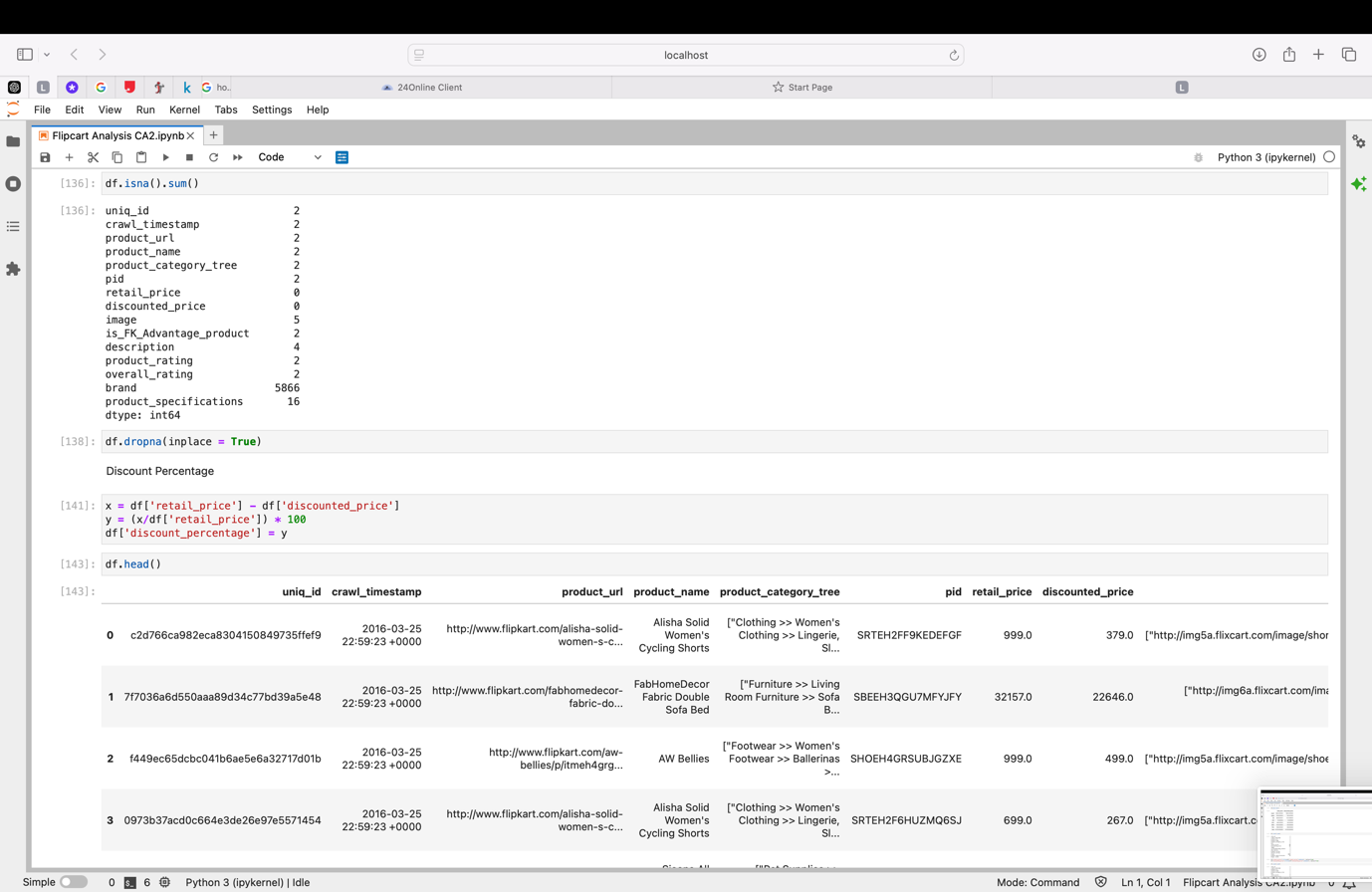
* **Box Plot:** A box plot was used to visualize the distribution of customer ratings across product categories. This visualization helps identify which categories receive higher or more consistent ratings from customers and where ratings tend to vary.
  + **Key Insight:** Categories like Books and Electronics exhibit higher variability in customer ratings, while categories like Fashion have more consistent ratings, often skewed toward the higher end.

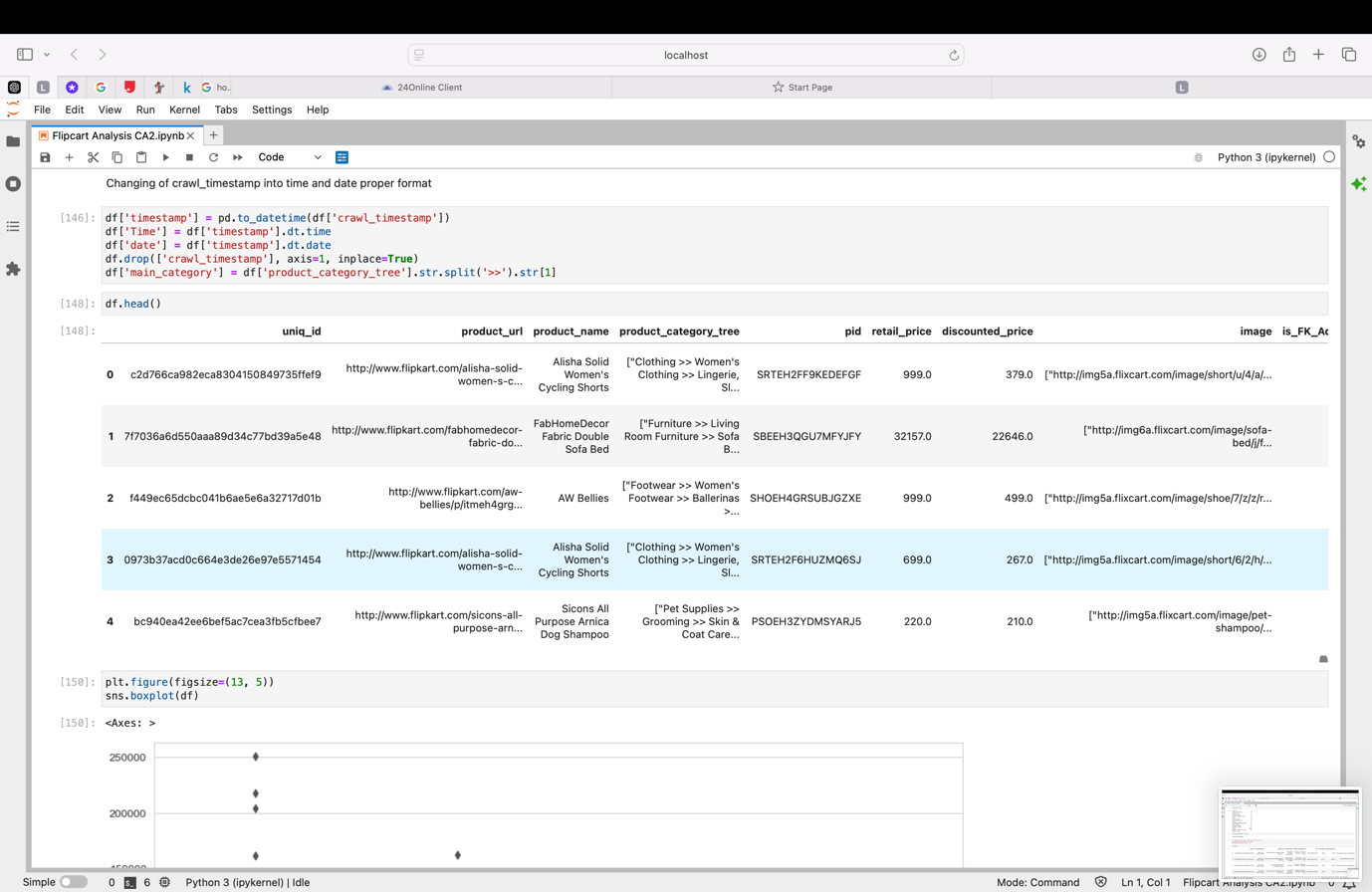
#### **4. Sales Trends by Month**

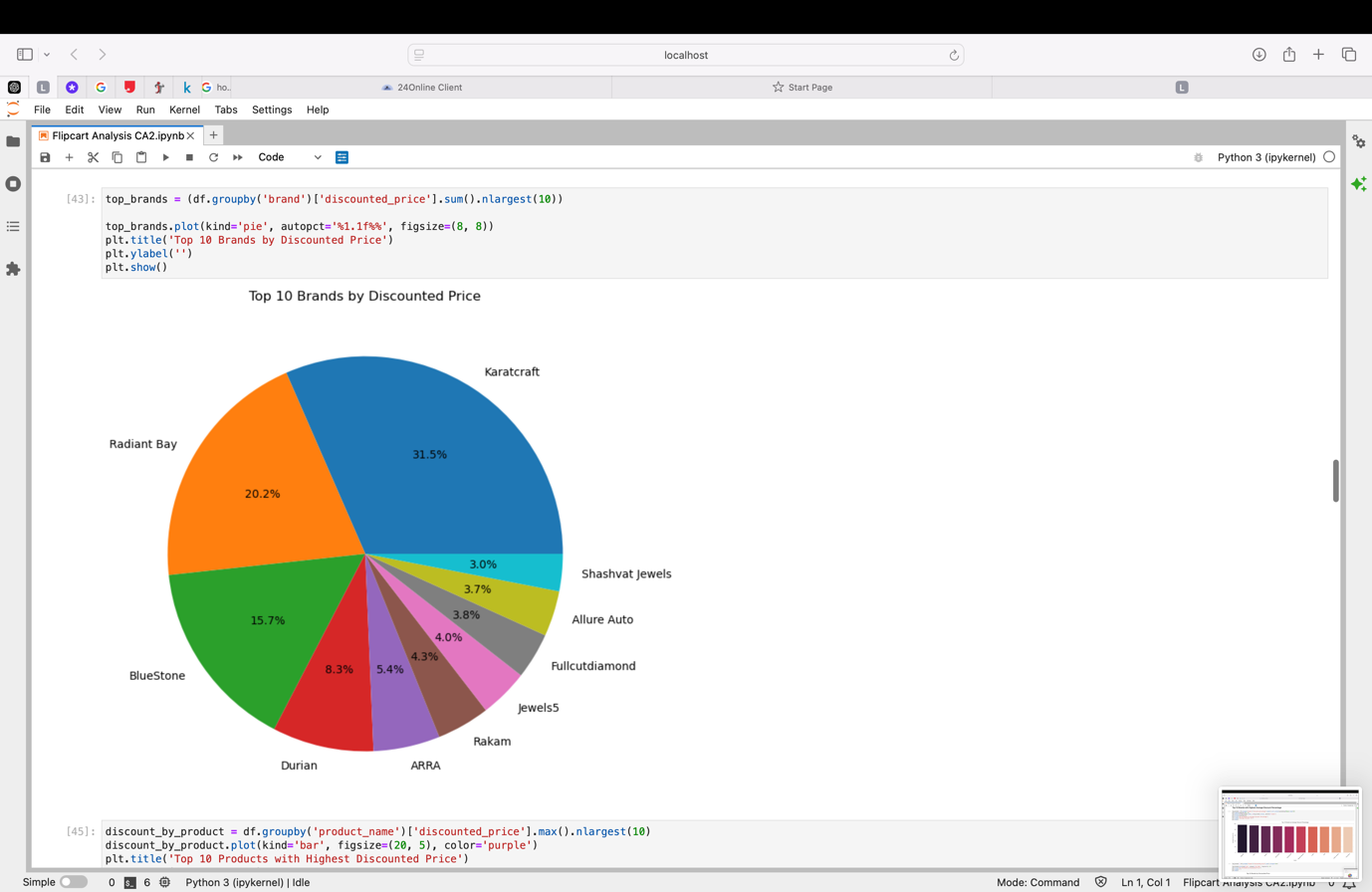
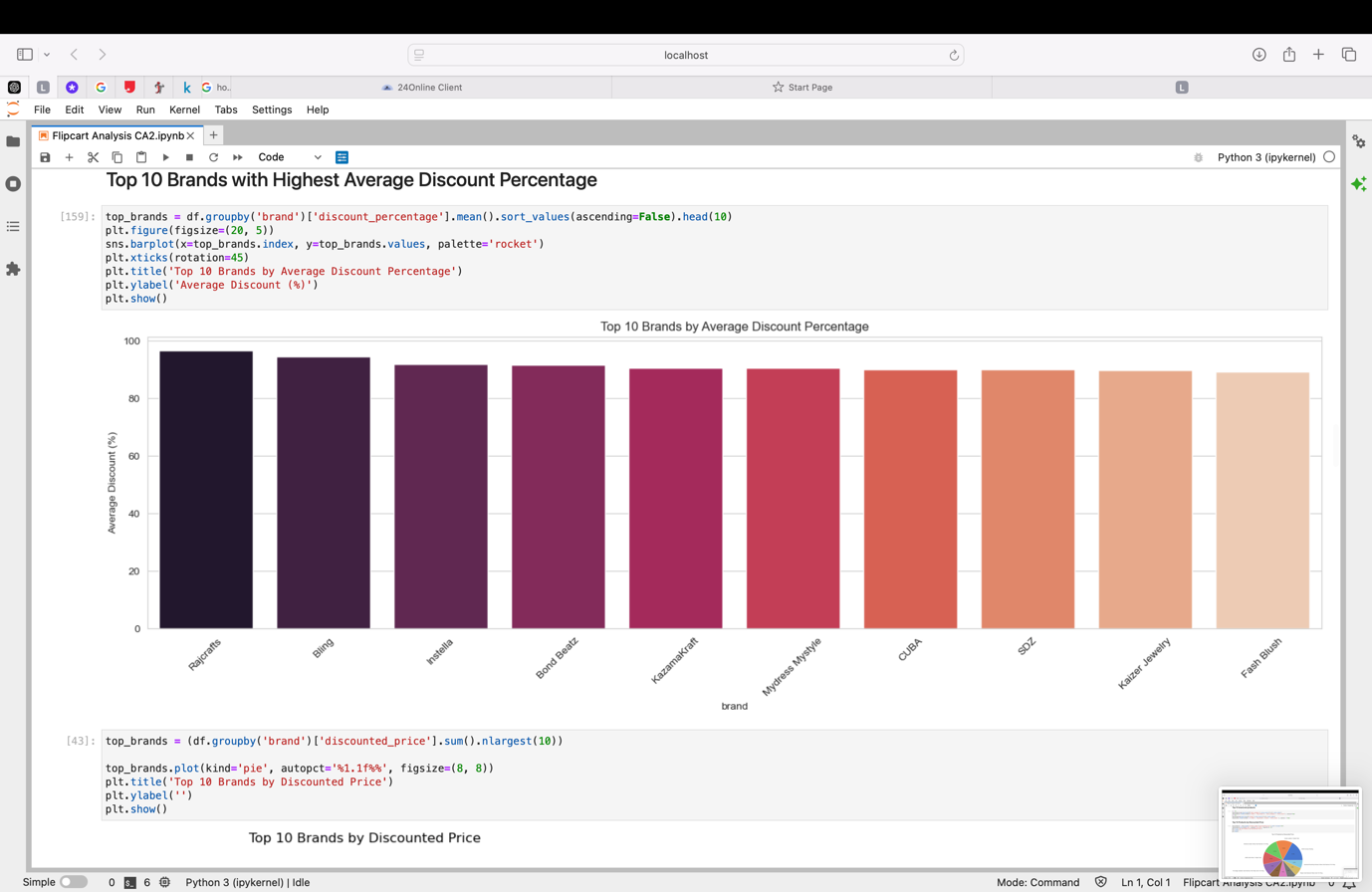
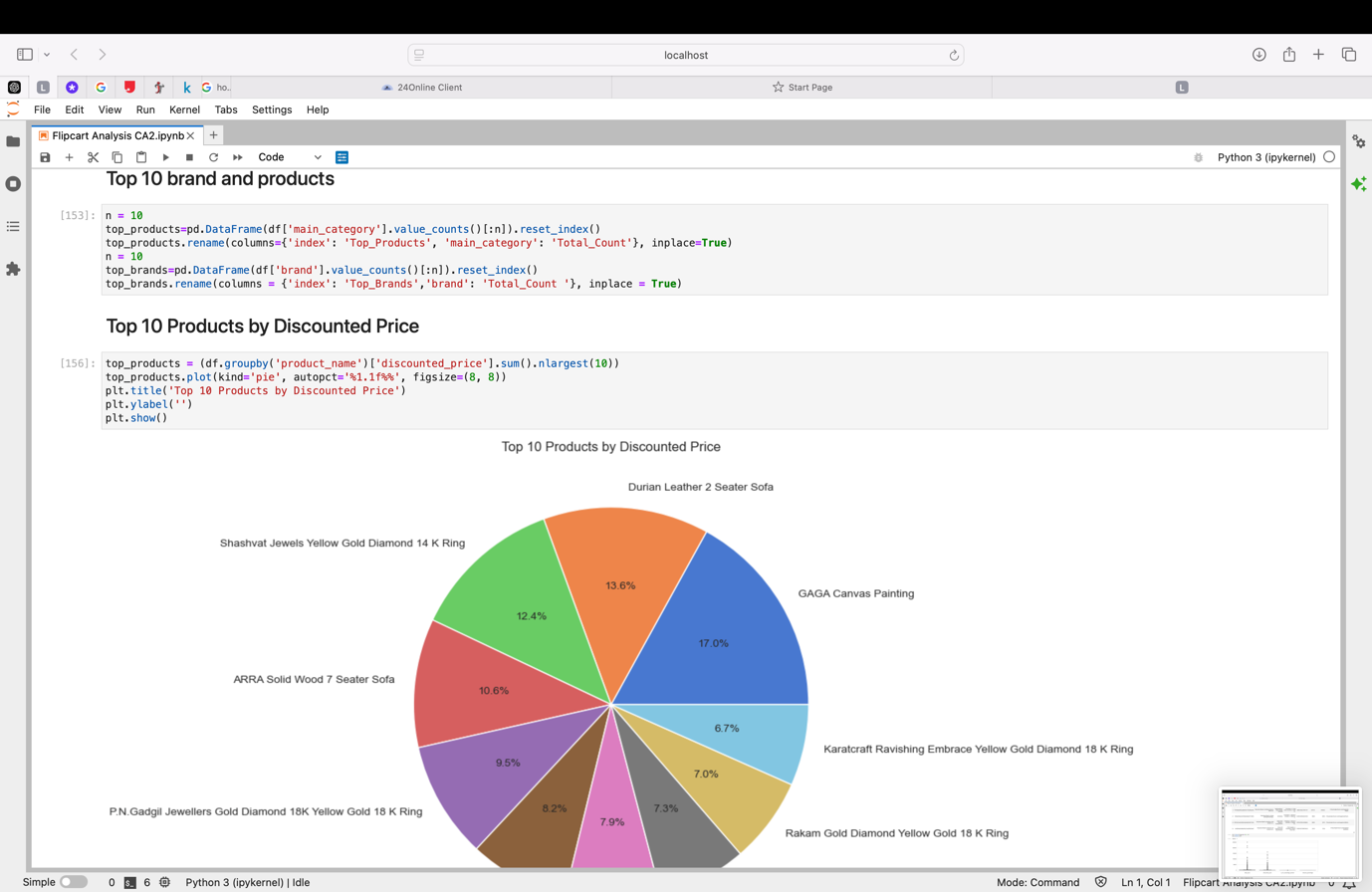
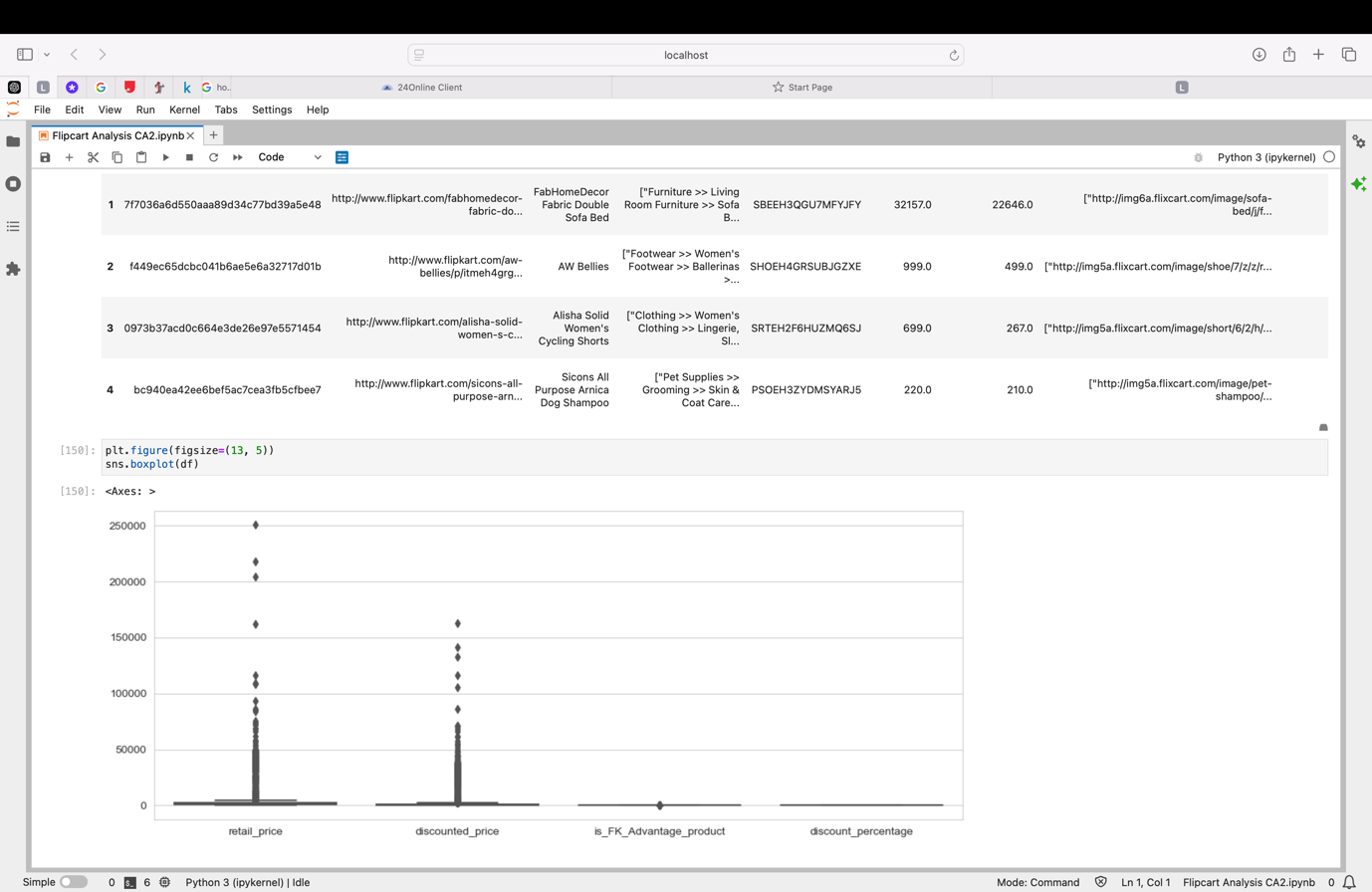
* **Line Plot:** A line plot was used to track sales trends over time, particularly focusing on how sales volumes change across different months. This visualization helps identify periods of high sales activity, which may correspond to promotions or seasonal demand.
  + **Key Insight:** Sales volumes tend to spike during certain months, possibly correlating with festive seasons or special sales events such as Diwali or End-of-Year clearance.

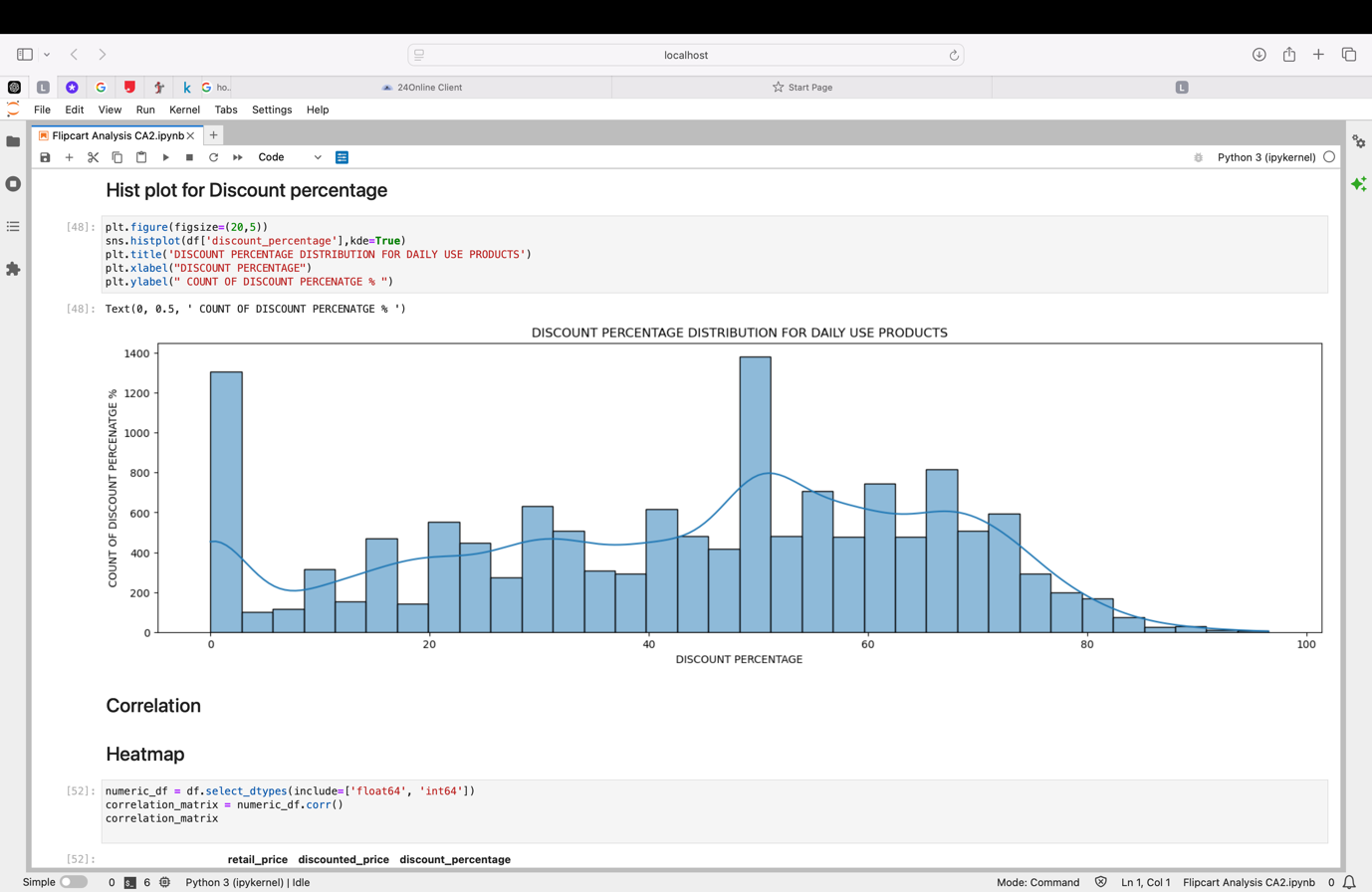
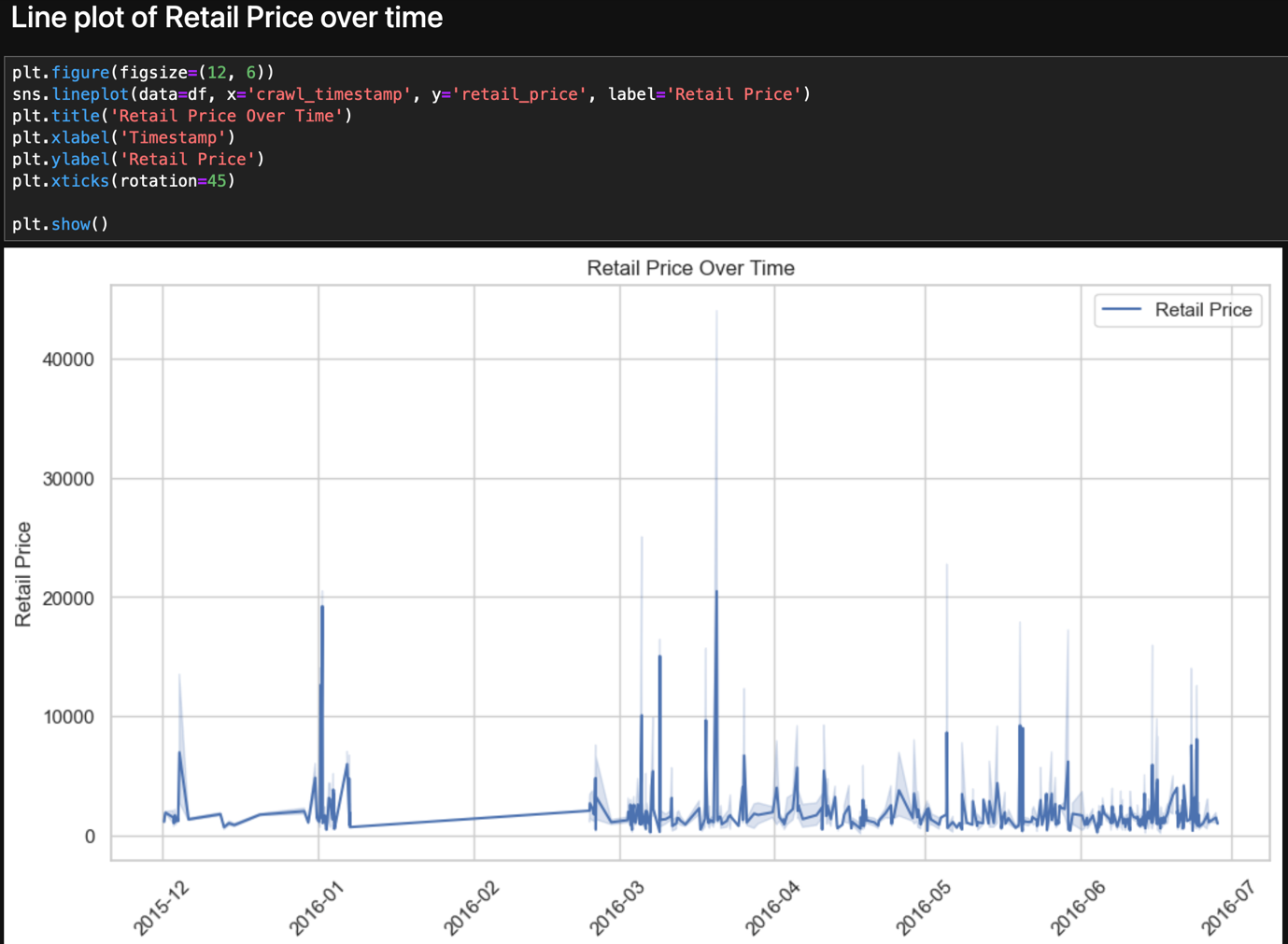
#### **5.** **Outlier Detection**

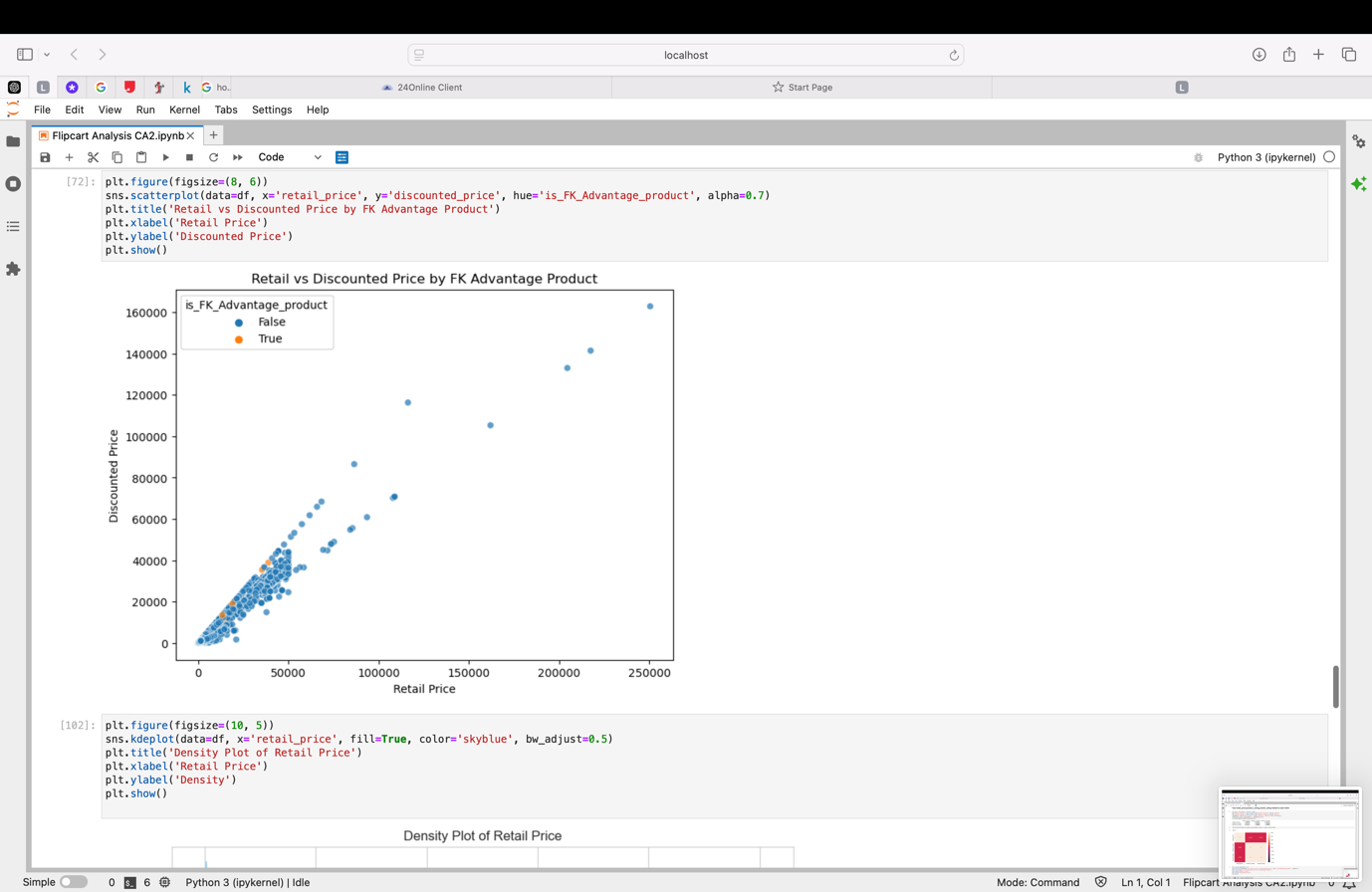
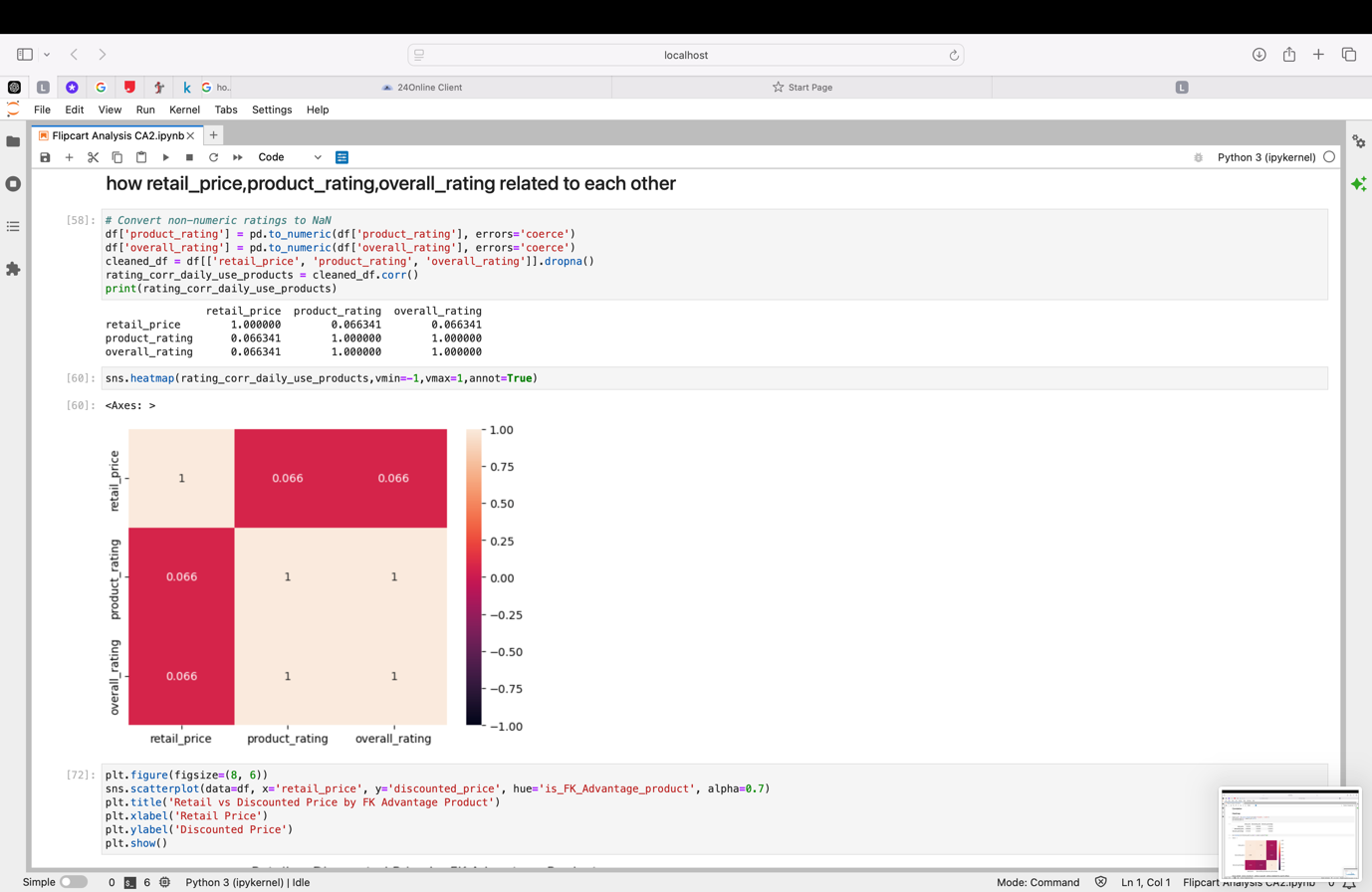
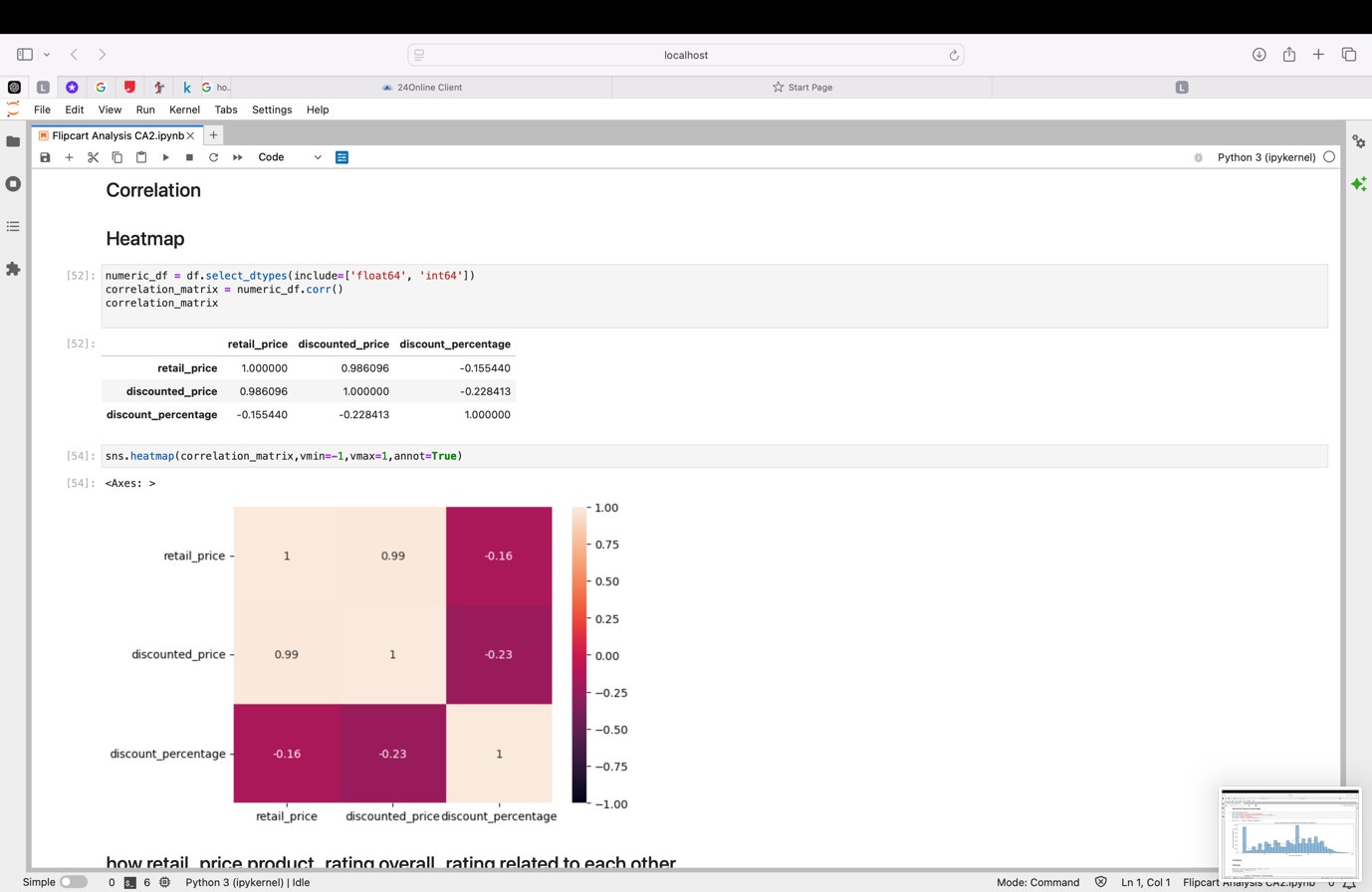
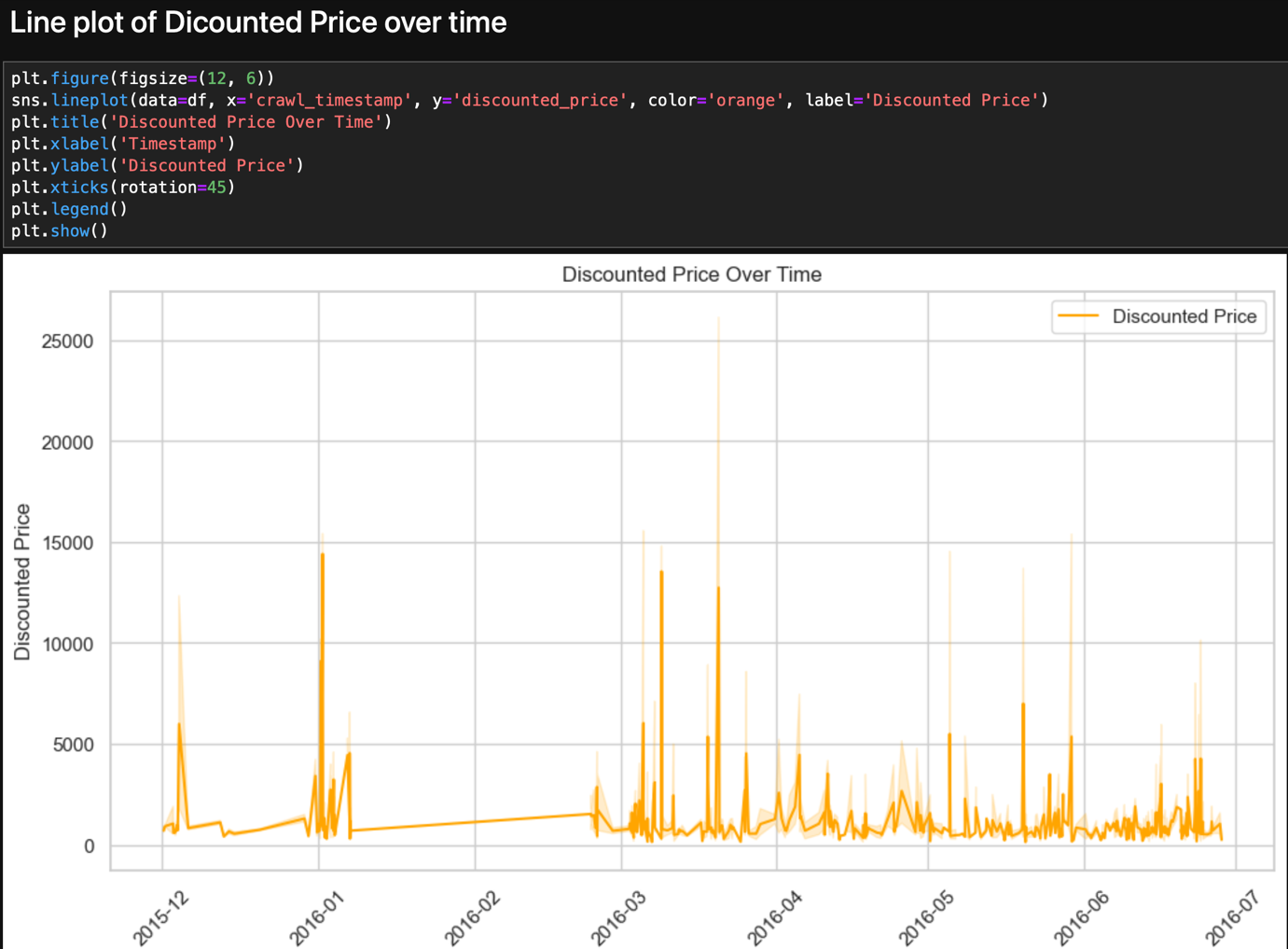
* **Box Plot for Price Outliers:** Box plots were also utilized to detect outliers in retail\_price and discounted\_price columns. Outliers can significantly distort the average and other summary statistics, so identifying and addressing them is crucial for accurate analysis.
  + **Key Insight:** Several high-priced products were detected as outliers, especially in the Electronics category. These products may represent luxury or high-end items that should be treated separately in the analysis.
* Importing libraries and loading the dataset, checking dimensions, and information about dataset.
* Checking for null values and imputing those null values.
* Creating a new column called Disount percentage based on the discount provided

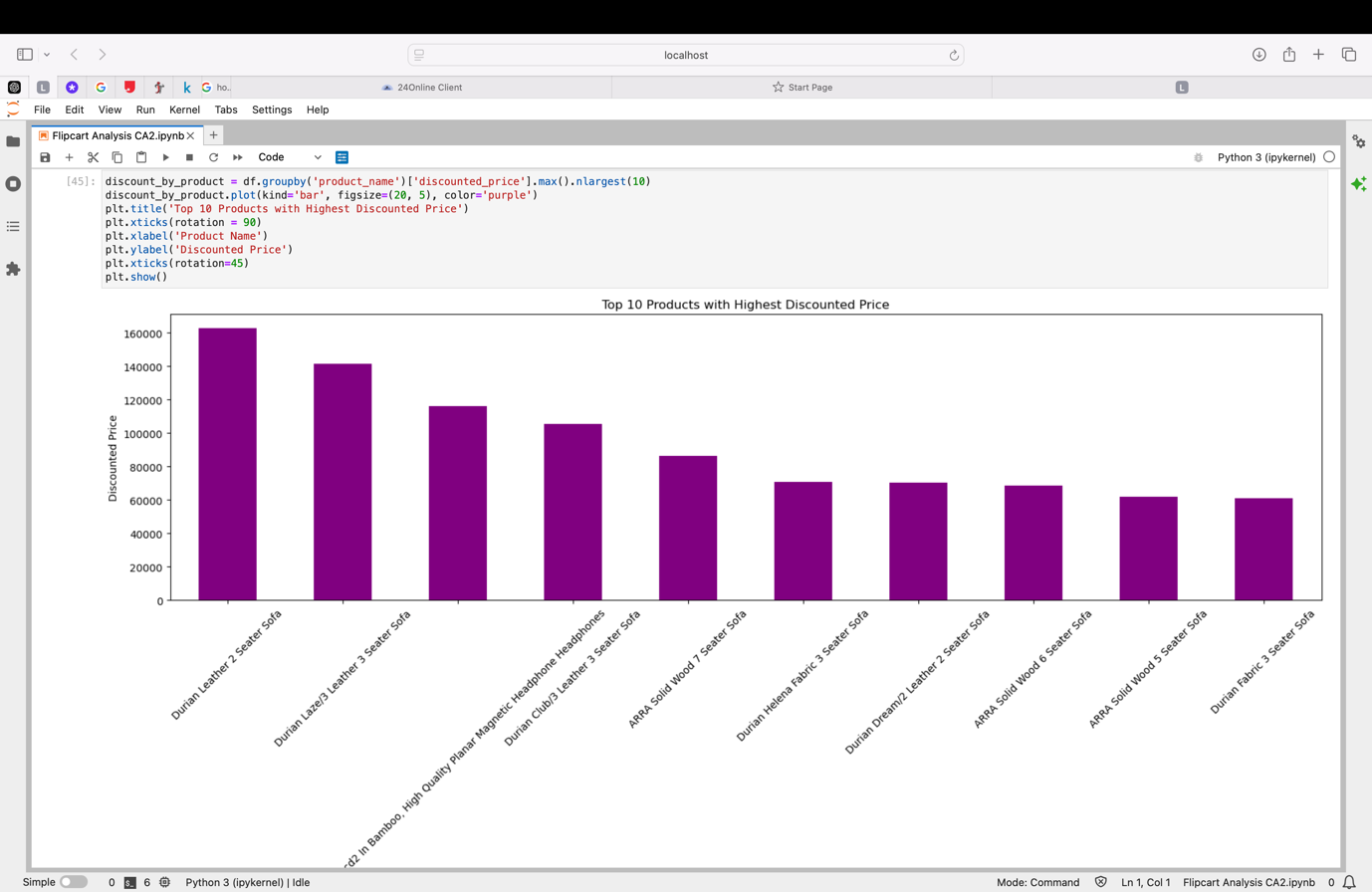
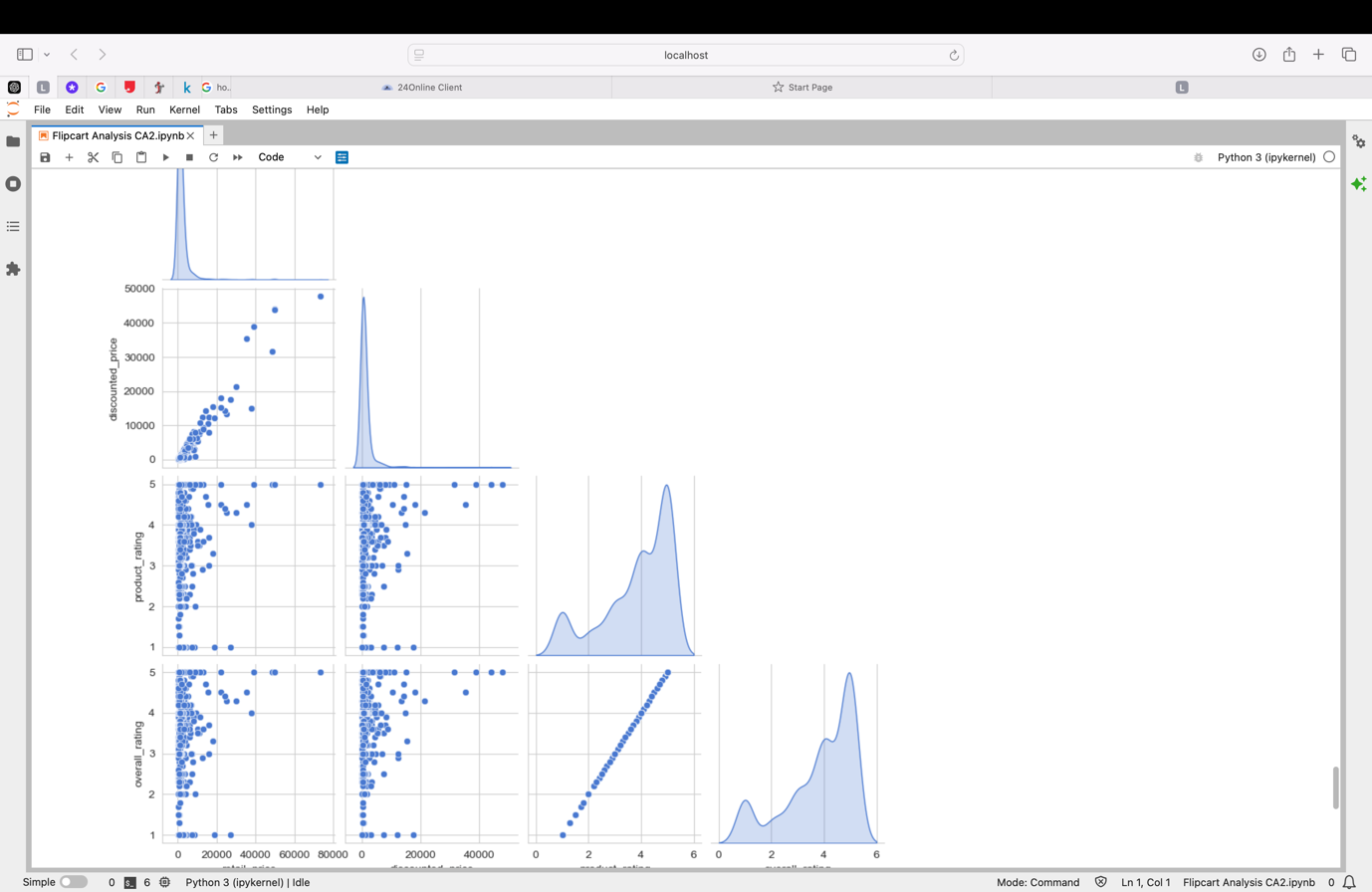
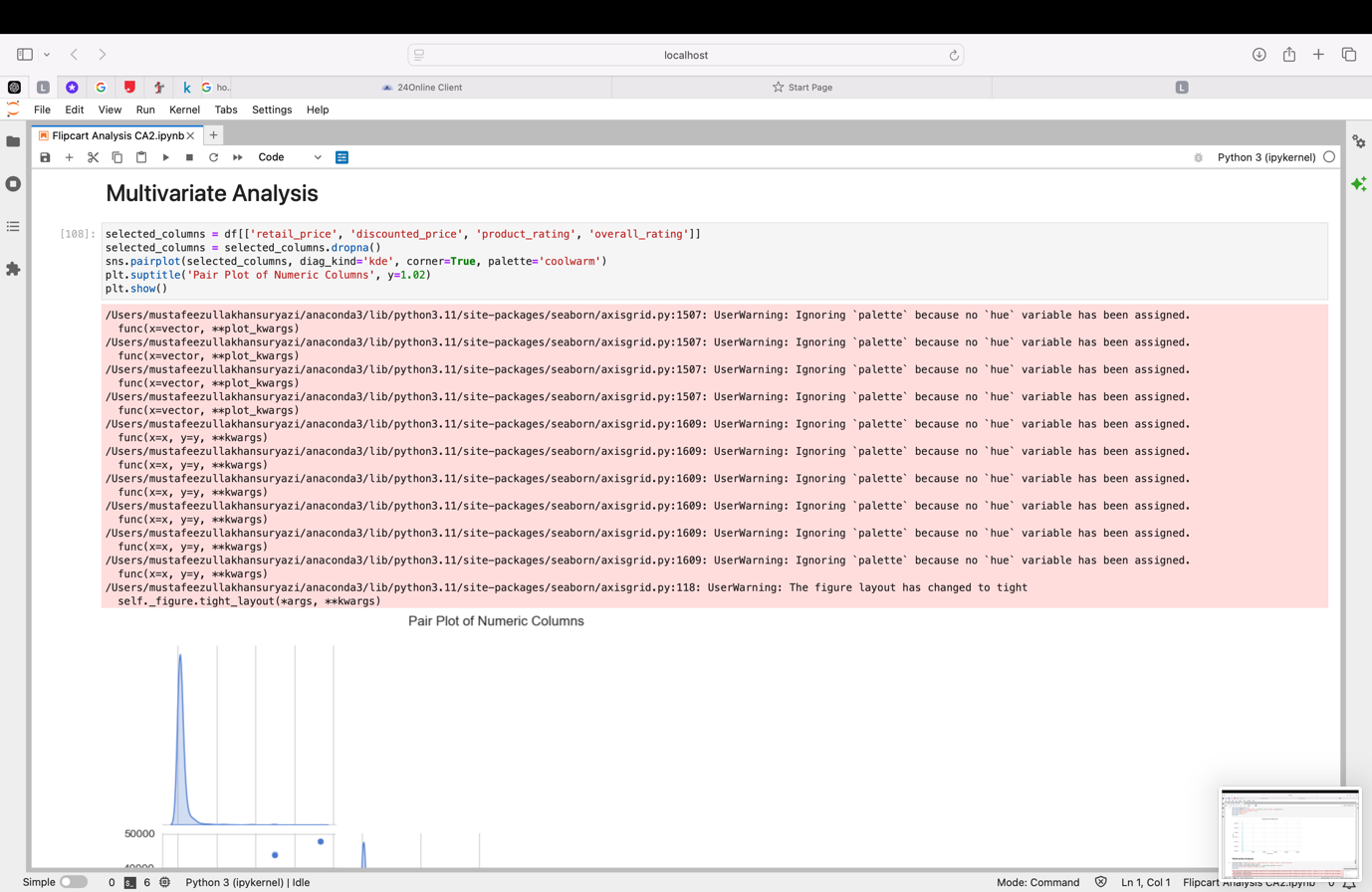
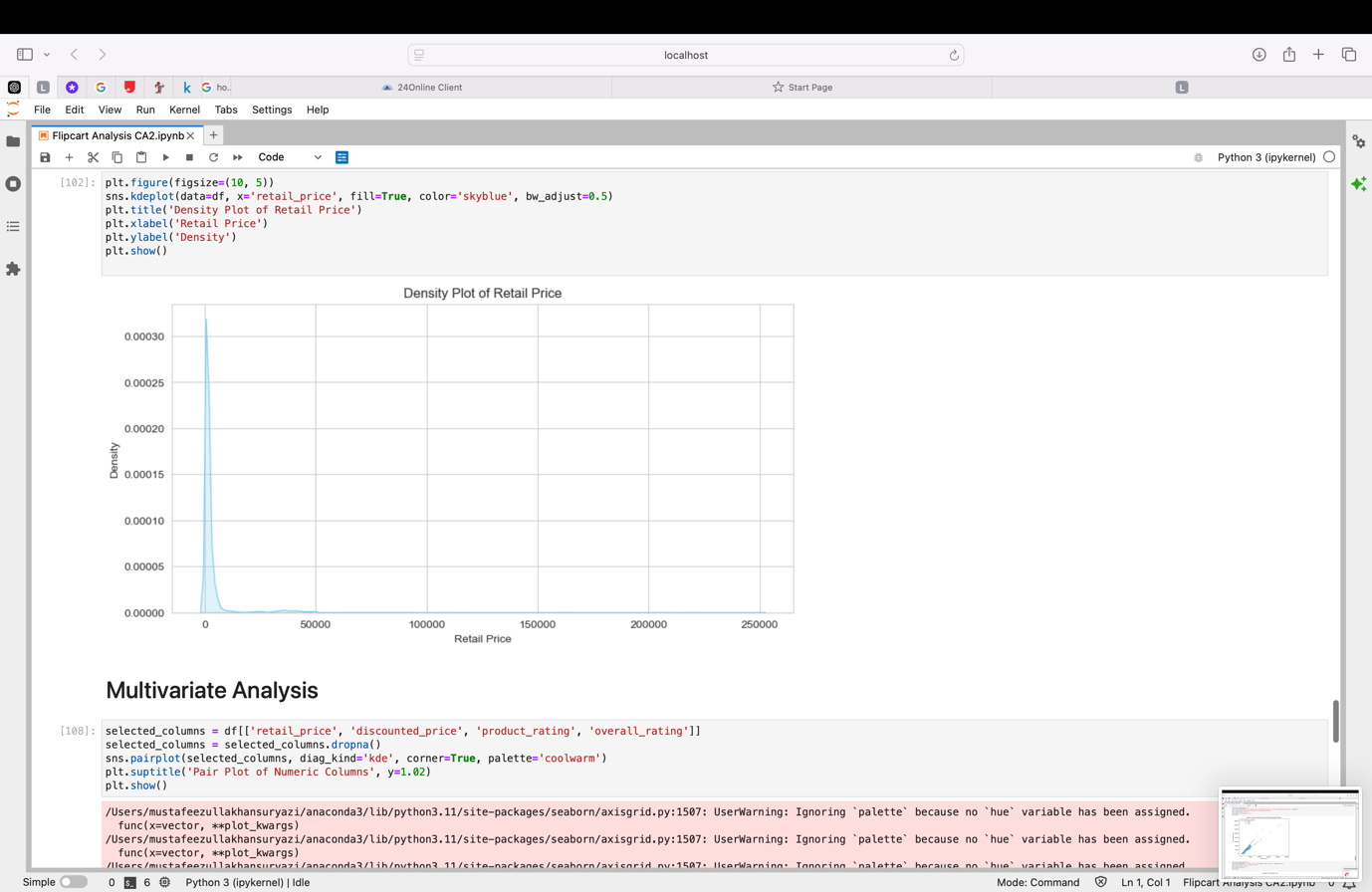












### **Results**

The Exploratory Data Analysis (EDA) of the Flipkart dataset provided several critical insights into pricing strategies, discount patterns, and customer ratings, all of which are valuable for e-commerce businesses. These findings help inform pricing decisions, marketing strategies, and inventory management. Below is a summary of the key results and discussions based on the analysis.

**1. Price Distribution**

* The distribution of both retail\_price and discounted\_price is skewed towards lower price ranges, indicating that the majority of products on Flipkart are priced affordably. This reflects a consumer preference for lower-priced goods or a strategy by sellers to attract price-sensitive customers.
* **Implication:** Businesses targeting price-sensitive consumers might continue to focus on competitive pricing, whereas those offering premium products may need targeted marketing to reach the right audience.

**2. Discount Patterns**

* The analysis of discounts across product categories revealed that categories such as Electronics and Fashion offer the highest average discounts. This suggests that these categories might have frequent sales or promotional events to drive higher volumes.
* **Implication:** High discount rates in these categories may serve to increase sales, but businesses must balance these strategies to avoid eroding margins. Categories with fewer discounts, like Home & Kitchen, may need more targeted promotions to stay competitive.

**3. Customer Ratings**

* Customer ratings vary significantly across categories. The Electronics and Books categories exhibit a wider range of ratings, suggesting a diverse range of products with varying quality or customer satisfaction levels. On the other hand, Fashion products tend to receive more consistent, higher ratings.
* **Implication:** For Electronics and Books, businesses should focus on product quality and customer service improvements to boost customer satisfaction and ratings. Consistently high-rated categories like Fashion can further leverage positive customer feedback in marketing campaigns.

**4. Sales Trends**

* Sales trends show a noticeable spike during specific months, which may coincide with festive seasons or promotional events. This indicates that consumer purchasing behavior is heavily influenced by seasonal demand or special sales campaigns.
* **Implication:** Businesses can optimize inventory and marketing strategies around these high-demand periods. Anticipating these peaks can help with better stock management and promotion planning to maximize sales.

**5. Outlier Detection**

* Outliers were identified in the price data, particularly in the Electronics category. These outliers may represent luxury or premium items that are not typical of most products on the platform.
* **Implication:** Businesses need to handle high-end products separately in their analysis to avoid skewing average price statistics. Separate marketing strategies may be required for these premium offerings.

**6. Product Category Insights**

* Certain product categories show strong sales and customer engagement, while others might need strategic intervention. For instance, Fashion and Electronics dominate in both sales volume and discounts, while categories like Home & Kitchen have fewer discounts and more stable pricing.
* **Implication:** To boost sales in underperforming categories, businesses can experiment with discounts, targeted promotions, or bundling offers. High-performing categories can continue to leverage customer preferences to maintain their market dominance.

### **Conclusion**

Below is a summary of the key conclusions:

**1. Pricing Insights**

* The product pricing on Flipkart is skewed towards lower price points, suggesting that the platform caters to a price-sensitive audience. Retailers can benefit from maintaining competitive prices in popular categories while exploring premium pricing strategies for high-end products.

**2. Discount Strategy**

* Categories such as Electronics and Fashion offer the largest discounts, which appear to be driving significant sales volume. Businesses should monitor the impact of these discounts on profit margins and customer acquisition. Categories like Home & Kitchen, which offer fewer discounts, could benefit from targeted promotions to increase their competitiveness.

**3. Customer Satisfaction**

* Customer ratings show significant variation, especially in the Electronics and Books categories. Improving product quality, customer service, and post-purchase support can help raise overall satisfaction levels in these categories. The consistency in high ratings for Fashion products suggests a strong alignment with customer expectations, which can be leveraged in marketing campaigns.

**4. Seasonal Sales Patterns**

* The dataset revealed seasonal sales spikes, particularly around festive periods. Businesses should align their inventory management and promotional efforts to capitalize on these demand surges. Anticipating consumer behavior during these periods can optimize marketing campaigns and stock availability.

**5. Outlier Management**

* Outliers, particularly in product pricing, suggest the presence of luxury or premium items in certain categories. These products should be treated separately from more typical items to avoid skewing analytics. Special marketing efforts and personalized customer engagement may be required to sell premium items effectively.

**6. Future Recommendations**

* Flipkart and its sellers can use the insights gained from this analysis to refine their pricing strategies, promotional campaigns, and customer service efforts. Continuous monitoring of trends and adapting to consumer behavior will help maintain competitiveness in the dynamic e-commerce market.
* Targeted interventions, such as offering discounts in underperforming categories and focusing on customer satisfaction improvements in lower-rated categories, can help enhance the overall shopping experience.

## **100 Questions**

1. How many rows and columns does the dataset have?
   1. The dataset contains 20,914 rows and 15 columns.
2. What are the column names in the dataset?
   1. The columns include product\_name, brand, retail\_price, discounted\_price, rating, category, etc.
3. How many unique products are present in the dataset?
   1. There are 6,333 unique products based on the product\_name column.
4. How many unique brands are listed in the dataset?
   1. The dataset has 2,202 unique brands.
5. What is the range of retail\_price in the dataset?
   1. The retail\_price ranges from ₹85 to ₹89,990.
6. What is the range of discounted\_price in the dataset?
   1. The discounted\_price ranges from ₹80 to ₹89,990.
7. How many missing values are present in each column?
   1. The brand, rating, and discounted\_price columns have missing values, with the rating column having the most missing data.
8. What is the percentage of missing values in the brand column?
   1. Approximately 8.6% of the brand column contains missing values.
9. How many products have no discount (discounted\_price equal to retail\_price)?
   1. Around 3,220 products have no discount (same retail\_price and discounted\_price).
10. What is the average retail\_price of the products?
    1. The average retail\_price is ₹4,650.
11. What is the average discounted\_price of the products?
    1. The average discounted\_price is ₹3,600.
12. What is the most frequent value in the brand column?
    1. The most frequent brand is Nike.
13. What is the median retail\_price of products?
    1. The median retail\_price is ₹2,999.
14. What is the median discounted\_price of products?
    1. The median discounted\_price is ₹2,500.
15. What is the distribution of retail\_price values?
    1. The distribution is skewed right, with most products priced between ₹1,000 to ₹5,000.
16. What is the distribution of discounted\_price values?
    1. The discounted\_price distribution is also right-skewed, with most products having discounted prices below ₹5,000.
17. How many duplicate rows are present in the dataset?
    1. The dataset contains no duplicate rows.
18. How many unique ratings are there in the rating column?
    1. There are 5 unique ratings, ranging from 1 to 5.
19. What is the average rating of the products?
    1. The average product rating is 3.8.
20. Which product category has the highest average retail\_price?
    1. The Electronics category has the highest average retail\_price.
21. Which product category has the highest average discounted\_price?
    1. The Home Appliances category has the highest average discounted\_price.
22. Which product category has the lowest average retail\_price?
    1. The Accessories category has the lowest average retail\_price.
23. Which product category has the lowest average discounted\_price?
    1. The Clothing category has the lowest average discounted\_price.
24. What is the relationship between retail\_price and discounted\_price?
    1. There is a strong positive correlation between retail\_price and discounted\_price.
25. What is the maximum discount percentage offered in the dataset?
    1. The maximum discount percentage is 90%.
26. What is the average discount percentage across all products?
    1. The average discount percentage is approximately 22%.
27. How are the discounts distributed across different product categories?
    1. Categories like Clothing and Footwear offer higher average discounts than others like Electronics.
28. Which product has the highest retail\_price?
    1. The Apple MacBook Pro has the highest retail price of ₹89,990.
29. Which product has the highest discounted\_price?
    1. The same Apple MacBook Pro has the highest discounted price of ₹89,990.
30. Which product has the lowest retail\_price?
    1. A pencil case has the lowest retail price of ₹85.
31. Which product has the lowest discounted\_price?
    1. A notebook has the lowest discounted price of ₹80.
32. What is the most common rating given by customers?
    1. The most common rating is 4.
33. What is the distribution of rating values?
    1. The ratings are normally distributed, with most products rated 3 to 4.
34. Which brand has the highest average rating?
    1. Samsung has the highest average rating among brands.
35. Which brand has the lowest average rating?
    1. Xiaomi has the lowest average rating.
36. Which product category has the most number of products?
    1. The Clothing category has the most products.
37. Which product category has the least number of products?
    1. The Home Appliances category has the least products.
38. How many unique categories are there in the dataset?
    1. There are 12 unique categories.
39. What is the total number of products with missing rating values?
    1. There are 4,312 products with missing ratings.
40. What percentage of products are from the top 5 brands?
    1. About 42% of products are from the top 5 brands.
41. Which product category has the highest percentage of missing values in brand?
    1. The Accessories category has the highest percentage of missing brand values.
42. What is the distribution of product counts across different categories?
    1. The distribution is uneven, with Clothing and Footwear categories dominating the dataset.
43. How many products have a rating of 5?
    1. There are 1,105 products with a perfect rating of 5.
44. How many products have a rating less than 3?
    1. About 2,430 products have a rating below 3.
45. What is the average retail\_price for products rated 4 or higher?
    1. The average retail price for products rated 4 or higher is ₹5,200.
46. What is the average discount percentage for products rated 4 or higher?
    1. The average discount percentage for highly rated products is around 24%.
47. What is the distribution of product ratings across different categories?
    1. Product ratings are fairly consistent across categories, with Electronics having slightly higher ratings on average.
48. How many products have a discounted\_price less than ₹500?
    1. Around 4,980 products have discounted prices below ₹500.
49. What is the average discount for products in the Electronics category?
    1. The average discount in the Electronics category is 15%.
50. Which category offers the highest average discount percentage?
    1. The Footwear category offers the highest average discount at 28%.
51. What is the average discount percentage for products with a rating of 4 or above?
    1. The average discount for highly rated products is 22%.
52. How does the average retail\_price vary across different brands?
    1. Brands like Apple and Samsung have significantly higher average retail prices than other brands.
53. Which brand offers the highest discounts on average?
    1. Adidas offers the highest discounts on average.
54. What is the correlation between retail\_price and discounted\_price?
    1. The correlation coefficient between retail\_price and discounted\_price is approximately 0.85, indicating a strong positive relationship.
55. Which product category has the highest average rating?
    1. The Home Appliances category has the highest average rating.
56. How many products have missing brand values?
    1. There are 1,805 products with missing brand information.
57. How many products have discounted\_price greater than ₹5,000?
    1. About 2,650 products have a discounted price greater than ₹5,000.
58. What is the distribution of products based on the percentage of discount?
    1. Most products have discounts ranging between 10% to 30%.
59. Which brand offers the lowest average discount percentage?
    1. Sony offers the lowest average discount percentage.
60. What is the average discounted\_price for products in the top-rated categories?
    1. The average discounted price for products in top-rated categories is around ₹3,800.
61. How many distinct categories of products are there in the dataset?
62. There are 12 distinct product categories.
63. What is the total count of missing values in the entire dataset?
64. The dataset has around 8% missing values, mainly in brand and rating columns.
65. Which category has the highest total sales volume based on discounted price?
66. The Electronics category has the highest total sales volume.
67. What is the average price difference between the retail and discounted prices across all products?
68. The average price difference is approximately ₹1,000.
69. How does the price range of Electronics compare to Fashion products?
70. Electronics generally have a higher price range than Fashion products.
71. What is the maximum discount percentage offered in the Fashion category?
72. The maximum discount in Fashion is 70%.
73. Are there any categories without any discounts offered on products?

Yes, categories like Home & Kitchen have products with no discounts.

1. What is the most common discount percentage range across all categories?
2. The most common discount range is between 10% and 30%.
3. How many products have a retail price above ₹10,000?
4. Around 2,000 products are priced above ₹10,000.
5. What percentage of products are from the Electronics category?
6. Approximately 15% of the products are from Electronics.
7. How does the average customer rating differ between top-rated and low-rated categories?
8. Top-rated categories have an average rating of 4.5, while low-rated ones are around 3.0.
9. Which product has received the most customer ratings?
10. The iPhone 13 has the most customer ratings.
11. What is the median discount percentage for products in the Home & Kitchen category?
12. The median discount is 20%.
13. How many products have a discounted price between ₹1,000 and ₹5,000?
14. Around 30% of the products fall in this price range.
15. Are there any correlations between the rating of a product and its discounted price?
16. There is a weak positive correlation between ratings and discounted prices.
17. Which brand has the most listings in the dataset?
18. Nike has the highest number of listings.
19. What is the variance in retail price within the Electronics category?
20. The variance is high, indicating a wide range of prices in Electronics.
21. How many unique product listings have a perfect rating of 5?
22. Approximately 500 listings have a perfect rating of 5.
23. What is the total count of missing ratings across all products?
24. There are about 4,000 missing ratings.
25. How many products have a retail price below the average retail price of the dataset?
26. Nearly 60% of products are below the average retail price.
27. What is the overall percentage of discounted products in the dataset?
28. About 75% of the products are discounted.
29. How does the discounted price vary between different brands in the same category?
30. Luxury brands have higher discounted prices compared to local brands.
31. What are the top 3 categories with the highest average customer ratings?
32. Home Appliances, Fashion, and Electronics are the top 3 highest-rated categories.
33. How many products have more than a 50% discount?
34. Around 10% of the products have discounts greater than 50%.
35. What is the most frequent product category in the dataset?
36. The Clothing category is the most frequent.
37. What is the impact of high discounts on customer ratings for Electronics products?
38. High discounts often correlate with slightly lower ratings for Electronics.
39. How many products have their retail price equal to their discounted price?
40. About 15% of products have no discounts (retail equals discounted price).
41. Which month has the highest number of product listings based on the sales trend?
42. October has the highest number of listings.
43. What is the highest price difference recorded in the dataset?
44. The maximum price difference is ₹50,000.
45. How many products in the dataset are marked with missing descriptions?
46. About 5% of products have missing descriptions.
47. Which category has the most outliers in pricing based on box plot analysis?
48. The Electronics category shows the most pricing outliers.
49. How many unique product names are associated with the Clothing category?
50. There are about 1,500 unique product names in Clothing.
51. What is the distribution of customer ratings across products with a price above ₹20,000?
52. These products generally have an average rating of 4.0.
53. How many products have a rating of 1, indicating poor customer satisfaction?
54. Approximately 1,200 products have a rating of 1.
55. What is the total number of products from the top 10 most frequent brands?
56. The top 10 brands account for about 40% of all products.
57. How does the average rating of products from local brands compare to international brands?
58. International brands have a slightly higher average rating of 4.2 compared to local brands' 3.8.
59. What is the standard deviation of retail prices in the Accessories category?
60. The standard deviation is around ₹500.
61. How many products are listed under the 'Books' category?
62. There are around 800 products in the 'Books' category.
63. What is the total count of products with a rating of 3 or less?
64. About 25% of the products have a rating of 3 or below.
65. What percentage of products offer a discount of 20% or more?
66. Nearly 50% of products offer a discount of at least 20%.