Project Report - Group 17

AUTHOR
Sarath Kumar, Yash, Saee , Mayuresh

Acknowledgement

We are grateful to our professor, Dr. Dootika Vats, for giving us the opportunity to work on the Data Science project on **"DECODING BOAT LIFESTYLE"**. *With Dr. Dootika Vats guidance, we were able to complete the project on time and with our best efforts. This experience has been incredible for our team, allowing us to learn and work together collaborately.*

1. Introduction

1.1 About Consumer Electronics and Accessories

When the words "Consumer Electronics and Accessories" are mentioned, our minds often conjure images of a vast array of cutting-edge gadgets and technological innovations that have become an integral part of our daily lives. Consumer electronics and accessories encompass a diverse range of products, including home audio systems, earbuds, smartwatches, chargers, timers, and car accessories, among others. This industry has experienced significant growth in recent years, driven by the increasing demand for sophisticated, feature-rich devices that cater to the evolving needs and preferences of modern consumers.

The market for consumer electronics and accessories is characterized by:

- T*echnological innovation and advancement*
- *Dynamic consumer preferences and behavior*
- *Intense competition among industry players*
- *Evolving trends and patterns in product design and functionality*

1.2 Problem Statement

In this project, the focus is on developing a streamlined approach to assist consumers in finding suitable products on the Boat website. By leveraging data science techniques, the project aims to curate a comprehensive final dataset encompassing essential product attributes, including price, reviews, ratings, and other relevant factors. This initiative is aimed at reducing the time and effort consumers spend on the website, allowing them to access a consolidated dataset that expedites their decision-making process and enhances their overall shopping experience.

1.3 project Motivation and Objective

Our project is driven by the growing demand for data-driven insights in the consumer electronics and accessories industry. By analyzing data from Boat's official website and Kaggle, we aim to uncover crucial market trends and consumer preferences across various product categories. Our objective is to provide actionable recommendations for businesses, enabling them to make informed decisions

regarding product development, marketing strategies, and long-term growth in this competitive market landscape.

2. About our dataset

2.1 Source of our data collection

1.Boat Lifestyle's Official Website:

Data was collected directly from Boat Lifestyle's official website, focusing on various consumer electronics and accessories such as home audio systems, smartwatches, chargers, timers, and other relevant products.

2. Kaggle Datasets (*Amazon earbuds*):

Information on Amazon earbuds was collected from relevant datasets available on Kaggle, contributing valuable insights and data specific to the earbuds segment within the consumer electronics and accessories market.

2.2 Data Set Variables

There are two final datasets: one for Boat Lifestyle and another for Amazon earbuds.

- **boat dataset:** *The dataset we collected is "Big_Frame.csv" which is a CSV file that contains complete details about 585 observations (or products) with 13 variables. NA's have been introduced in some instances where the value is not provided.*
- **Amazon earbuds:** *We collected detailed information about the Amazon earbuds, including various product features and have complete details of 72 observations (or products) with 8 varaibles . The dataset has information on several earbud products, with specific details about each product.*

2.2.1 Categorical Variables

1. **ProductName:** The specific name or title of the product as listed on the Boat website, aiding in its identification and differentiation from other items.
2. **Country_of_Origin:** The country where the product is manufactured or produced, indicating its source or place of origin.
3. **Category:** denotes the product's classification as either "Wireless" or "Wired," distinguishing between the two types based on their connectivity features. This categorical text variable allows consumers to identify and select products based on their preferred connectivity type.
4. **Type:** Type represents the specific classification or group to which the product belongs, aiding in the organization and segmentation of products based on their shared characteristics or functionalities. This categorical variable assists consumers in easily navigating and identifying products within their preferred product segments.
5. **Model_Name*(in Amazon earbuds dataset)***:Model_Name represents the specific name or identifier assigned to the product by the manufacturer, allowing for easy identification and differentiation of product variations or versions.

6. **AiredDate*(in Amazon earbuds dataset)***:AiredDate signifies the date on which the product was released or made available for purchase, providing information about the product's launch or introduction to the market.

2.2.2 Numerical Variables

1. **ProductRating*(common in both Datasets)***: ProductRating is the average customer satisfaction score, indicating the overall quality of the product based on feedback.
2. **ReviewCount*(common in both Datasets)***: ReviewCount represents customer engagement and popularity, determined by the total number of customer reviews or feedback received.
3. **Sale_price**: Sale_price displays the current selling price, incorporating any ongoing promotions or discounts, providing the present cost of the product.
4. **Regular_Price**: Regular_Price denotes the original retail value before any discounts or special offers, serving as the standard market price for the product.
5. **Discount_Percentage*(common in both Datasets)***: Discount_Percentage reveals the percentage by which the regular price has been reduced, demonstrating the amount of discount provided to customers.
6. **ColourOptions**: ColourOptions represents the total number of color choices or options available for the product, indicating the range of visual variants that consumers can choose from. This numerical variable provides insights into the diversity of color selections offered for the product, allowing consumers to understand the available options and make informed decisions based on their preferences.
7. **Warranty:**Warranty signifies the duration, in years, of the warranty offered for the product, providing assurance to customers regarding product quality and support. This numerical variable denotes the specific length of time for which the product is covered under the warranty.
8. **Charging_Time:** The duration required for the product to be fully charged, indicating the time needed for the charging process, in hours.
9. **Charge_Duration:** The duration for which the product can remain operational or provide power after being fully charged, indicating its battery life or usage duration, in days.
10. **Amazon_Price*(in Amazon earbuds dataset)***: Amazon_Price signifies the price at which the product is listed for sale on the Amazon website, representing the cost of the product for potential buyers.
11. **MRP*(in Amazon earbuds dataset)***: MRP (Maximum Retail Price) denotes the highest price at which the product can be sold to customers, providing the upper limit for product pricing as determined by the manufacturer or retailer.
12. **Batteries*(in Amazon earbuds dataset)***:Batteries represent the number of batteries required for the earbuds, indicating the power source needed for their operation.

2.3 Dataset Details

2.3.1 Boat Dataset

- **Name:** Big_Frame.csv
- **Observations:** 585 products
- **Variables:** 13

2.3.2 Amazon Earbuds Dataset

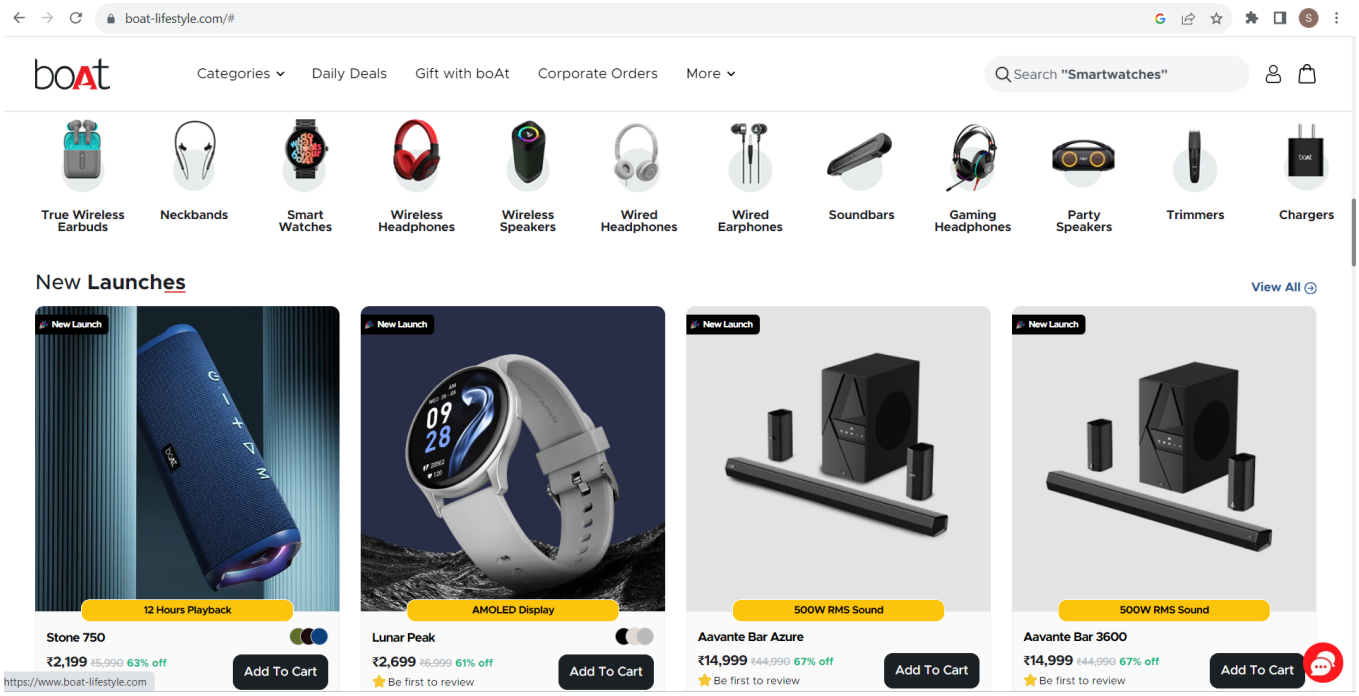
- **Observations:** 72 products
- **Variables:** 8

3. Obtaining The Data

3.1. Complete details of the Data collection sources

3.1.1. Boat Lifestyle's Official Website (boAt official site):

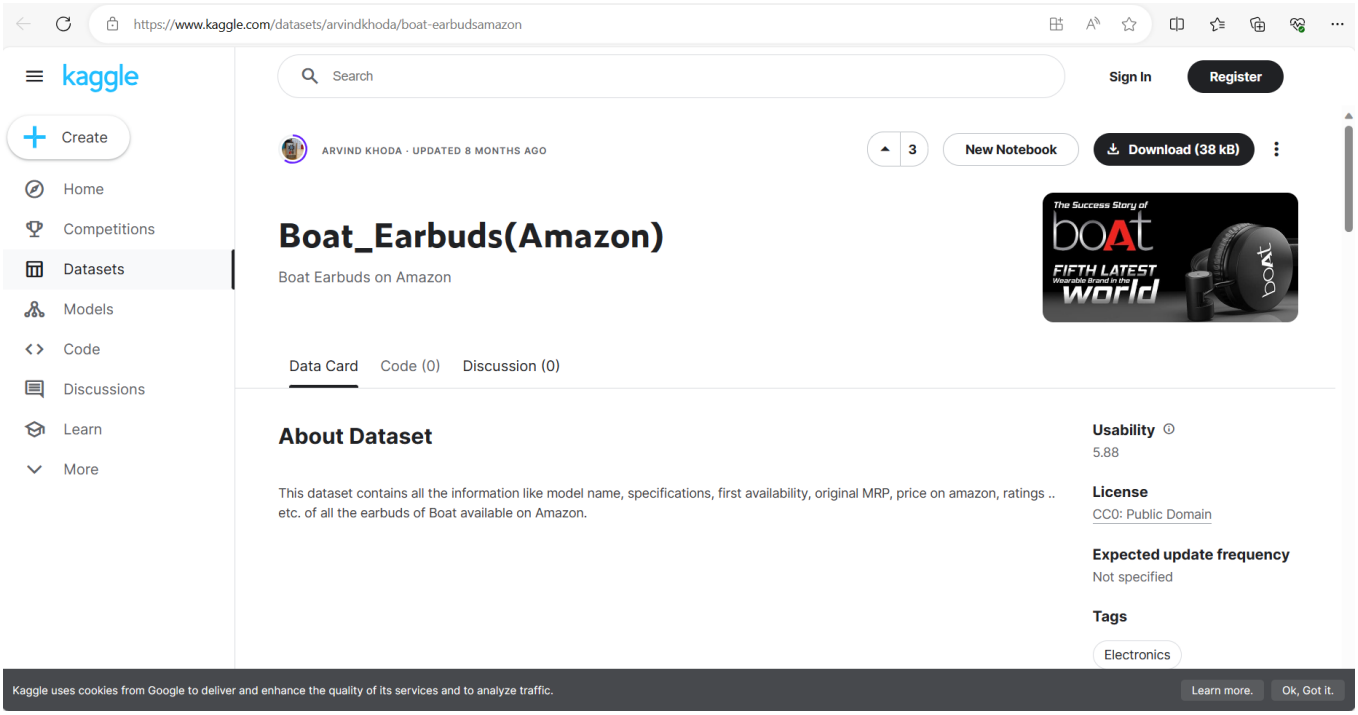
Data was collected directly from Boat Lifestyle's official website, focusing on various consumer electronics and accessories such as home audio systems, smartwatches, chargers, timers, and other relevant products. The website provided comprehensive information crucial for our analysis.



We scraped all 18 categories that are given on this website.

3.1.2. Kaggle Datasets (amazon earbuds):

Information on Amazon earbuds was collected from relevant datasets available on Kaggle, contributing valuable insights and data specific to the earbuds segment within the consumer electronics and accessories market.



3.2. Libraries used during the Data collection Process

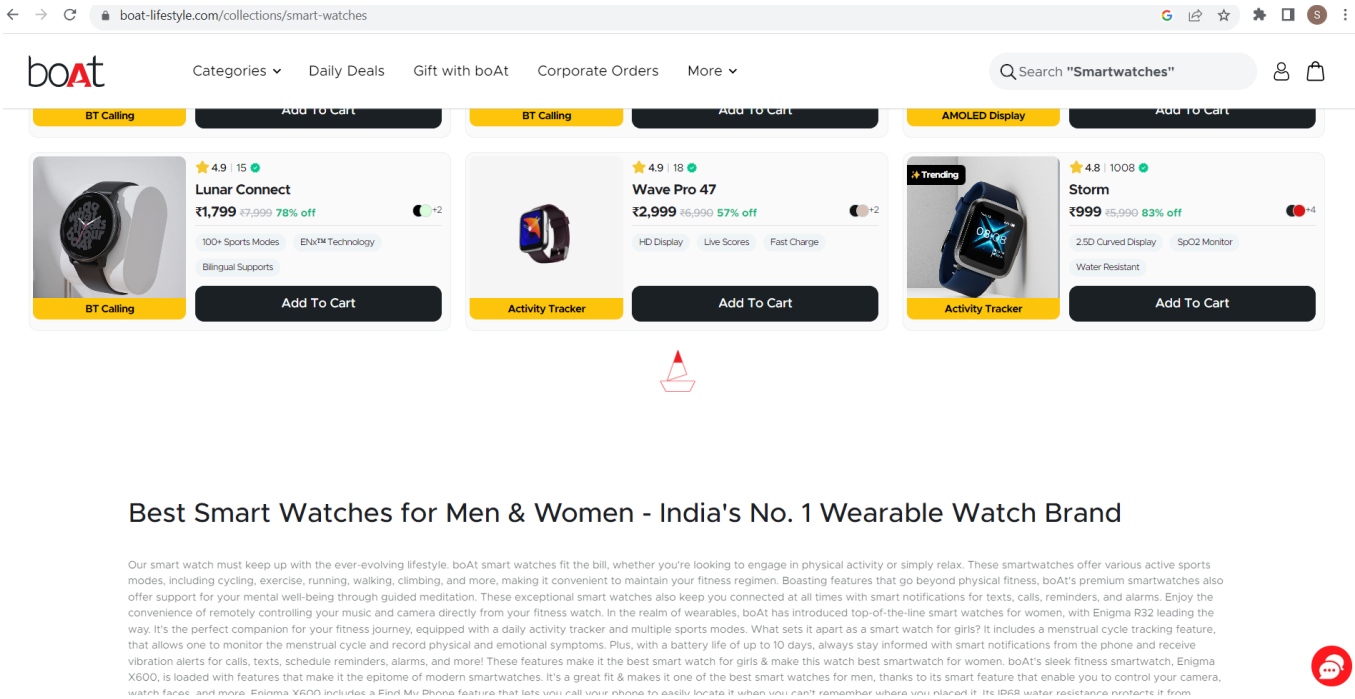
- tidyverse
- rvest
- selenium: *This library is used for automating web browser interaction from Python. It allows you to simulate a user's interaction with the browser, including navigating to webpages, interacting with web elements, and scraping data.*
- time: *This library is used to add delays or pauses in the execution of the code. It is used in this script to provide time for the page to load more content.*

3.3. Problems Faced during this process

During the scraping process of the Boat website, our group encountered limitations due to the website's dynamic nature. We found that we were unable to extract more than 18 products from each web page, which hindered our data collection efforts.

dynamic nature of the website(webpage for a particular category looks similar to this)

3.4. Web Scraping Approach



Utilization of Selenium

The decision to employ Selenium for web scraping was motivated by the dynamic nature of the target website. Traditional scraping methods were insufficient due to the website's dependence on JavaScript for content rendering. Additionally, RSelenium was considered but found to be outdated and less suitable for the task.

Dynamic Content Loading

During the scraping process, a notable observation was made regarding the website's dynamic content loading mechanism. Specifically, a loader was present at approximately 3.5 times the browser's height from the bottom of the page. This loader introduced a delay of 3-5 seconds to load additional products. Selenium provided the necessary functions to effectively handle these dynamic loading behaviors.

Data Collection and Cleaning

After scraping the URLs for each category, the data was subsequently collected and cleaned using R codes. These R codes are available in this repository. The cleaned data is organized and made available for further analysis.

Amazon Dataset Comparison

In addition to scraping and cleaning data from the target website, an Amazon dataset from Kaggle was also scraped and cleaned. This dataset will be used for comparisons and analysis, as per the project's requirements.

4. Cleaning the Data

The cleaning of data is a crucial step in ensuring the accuracy and reliability of the information for analysis. Below are the steps taken for cleaning the collected datasets from Boat Lifestyle's official website and Kaggle (Amazon earbuds).

4.1 Cleaning Boat Lifestyle Dataset

- **The Libraries we used:** dplyr
- **Handling Missing Values:** Some data was missing in the dataset, so we left those parts as 'NA' without putting any specific value. This helps to keep things clear, showing that we don't have information for those parts. We did this to keep the data accurate and make it clear where info was missing
- **Data Type Conversion:**

  To bolster the precision of our analysis, a thorough examination and adjustment of data types were carried out for key variables. The following variables, including those related to product identification and categorization, underwent conversion to the numeric data type:

  - **ProductRating:** Now represented as a numeric variable, allowing for a more nuanced analysis of customer satisfaction scores.
  - **ReviewCount:** Transformed into a numeric format to enable comprehensive quantitative evaluations of customer engagement and product popularity.
  - **Sale_Price:** Converted to a numeric type, providing a solid foundation for precise calculations involving the current selling prices.
  - **Regular_Price:** Transitioned to a numeric format to maintain consistency in handling the original retail values before any discounts.
  - **Discount_Percentage:** Now stored as a numeric variable, facilitating numerical computations related to discount percentages.
  - **ColourOptions:** Converted to a numeric format, streamlining the analysis of the total number of color choices available for each product.
  - **Warranty:** Now represented numerically, indicating the duration of warranty coverage in years for each product.
  - **Charging_Time:** Converted to a numeric format for an accurate measurement of the duration required for product charging, expressed in hours.
  - **Charging_Duration:** Now stored as a numeric variable, signifying the operational duration or battery life after a full charge, measured in days.
  - Only the data types of the variables mentioned earlier, including ProductRating, ReviewCount, Sale_Price, Regular_Price, Discount_Percentage, ColourOptions, Warranty, Charging_Time, and Charging_Duration, were changed during this process.
- **Biases found in the Initial Dataset:**

Some values in the 'Charging_Time' variable were originally obtained in hours, while others were in minutes. We standardized all these values to be uniformly represented in hours. Similarly, for the 'Charging_Duration' variable, values were initially obtained in hours and days. We converted all these values to be consistently expressed in days.

4.2 Cleaning Amazon Earbuds Dataset

The dataset obtained from Kaggle, containing information on Amazon earbuds, underwent similar cleaning processes:

- **The Libraries we used:** dplyr
- **Handling Missing Values:** Just like the Boat Lifestyle dataset, we managed missing values by introducing NA's. In this step, we dropped observations where 'MRP' or 'rating' was zero and also if Batteries is not a number since they are crucial for our analysis.

Data Type Conversion: During the process of converting data types, I specifically changed the following variables to numeric:

- **Discount_Percentage:** Converted to numeric for precise calculations related to discount percentages.
- **Amazon_Price:** Now stored as a numeric variable, facilitating accurate numerical analysis of the product's price on Amazon.
- **ProductRating:** Converted to numeric for a more detailed analysis of customer satisfaction scores.
- **ReviewCount:** Transformed into a numeric format to enable comprehensive quantitative evaluations of customer engagement and product popularity.
- **MRP:** Now represented as a numeric variable for consistent handling of maximum retail prices.
- **Batteries:** Converted to numeric, allowing for numerical computations related to the number of batteries required for the earbuds.
- Only the data types of the variables mentioned earlier, including ProductRating, ReviewCount, Sale_Price, Regular_Price, Discount_Percentage, ColourOptions, Warranty, Charging_Time, and Charging_Duration, were changed during this process.

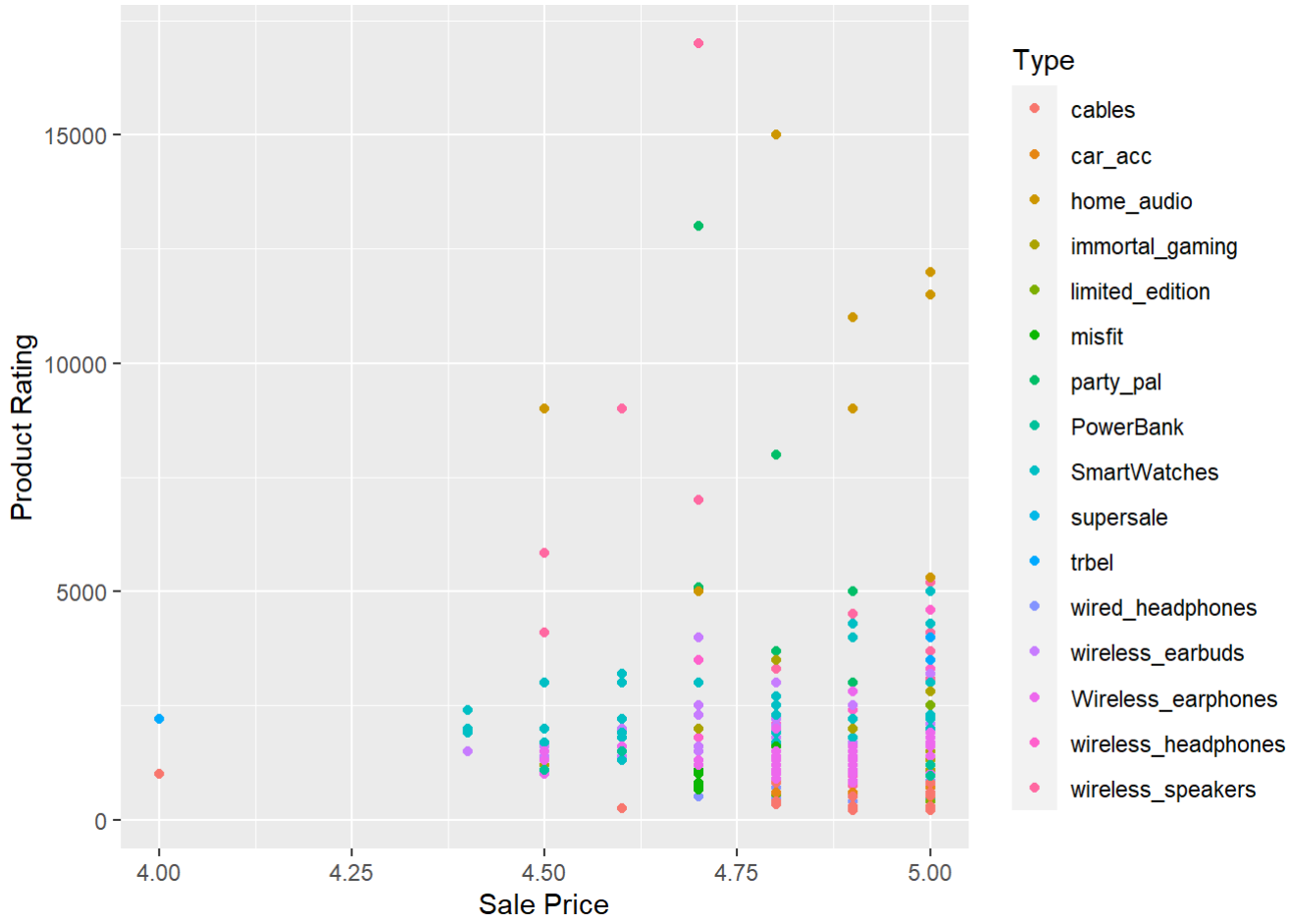
After fixing and changing data, we made the final table. It's all neat now, ready for looking deep, doing math, and making pictures. We can now dig in, do number stuff, and draw to find important things from the data.

5. Interesting Questions

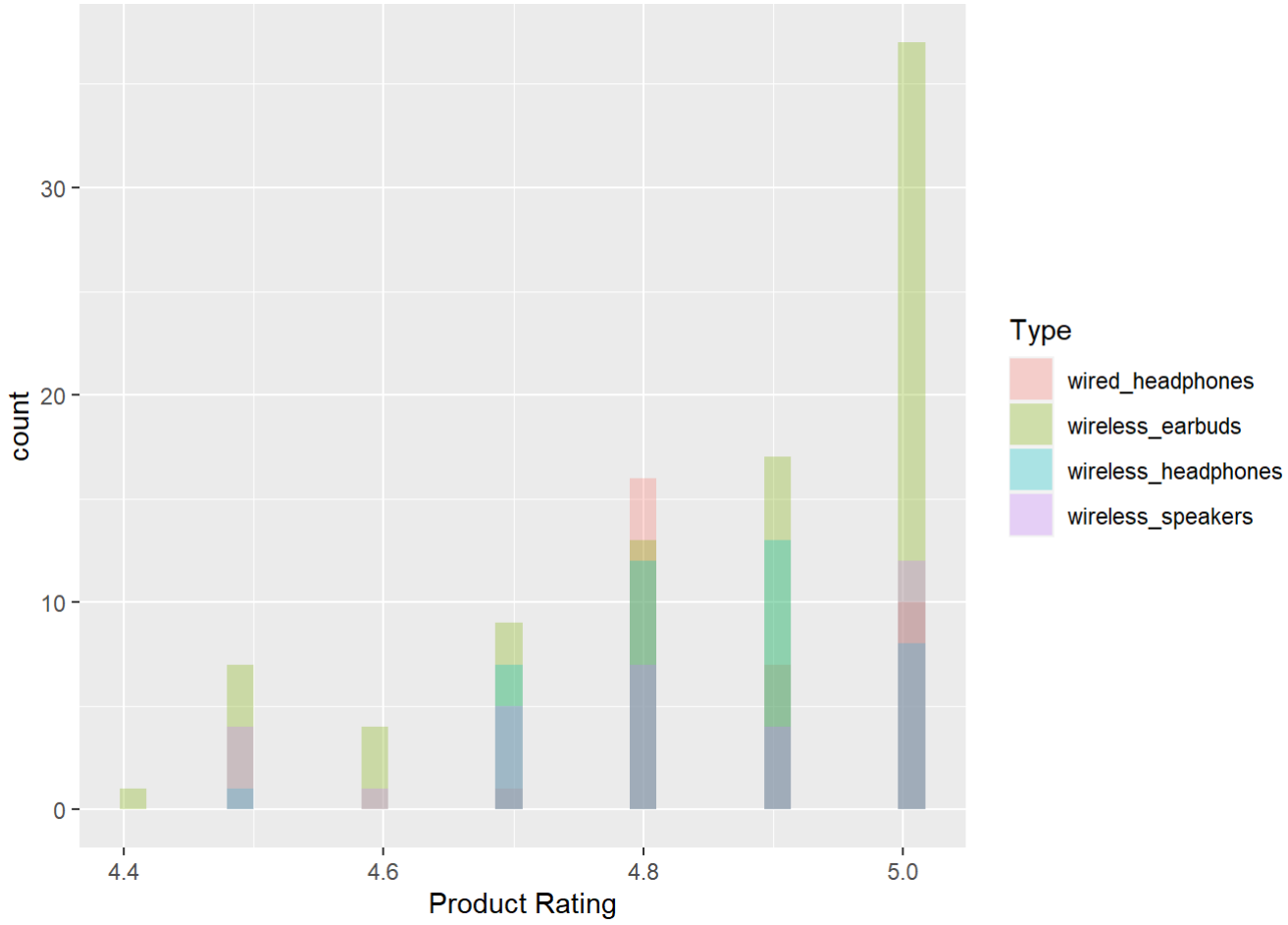
- Is there any relation between rating of product and price?
- which product has highest Product rating and why?
- Which is most popular product of Boat?
- If someone wants to buy earbuds and having budget below 5k then which is good option among wireless and wired category?
- which price segment is mainly focused by boat?

6. Important Visualizations

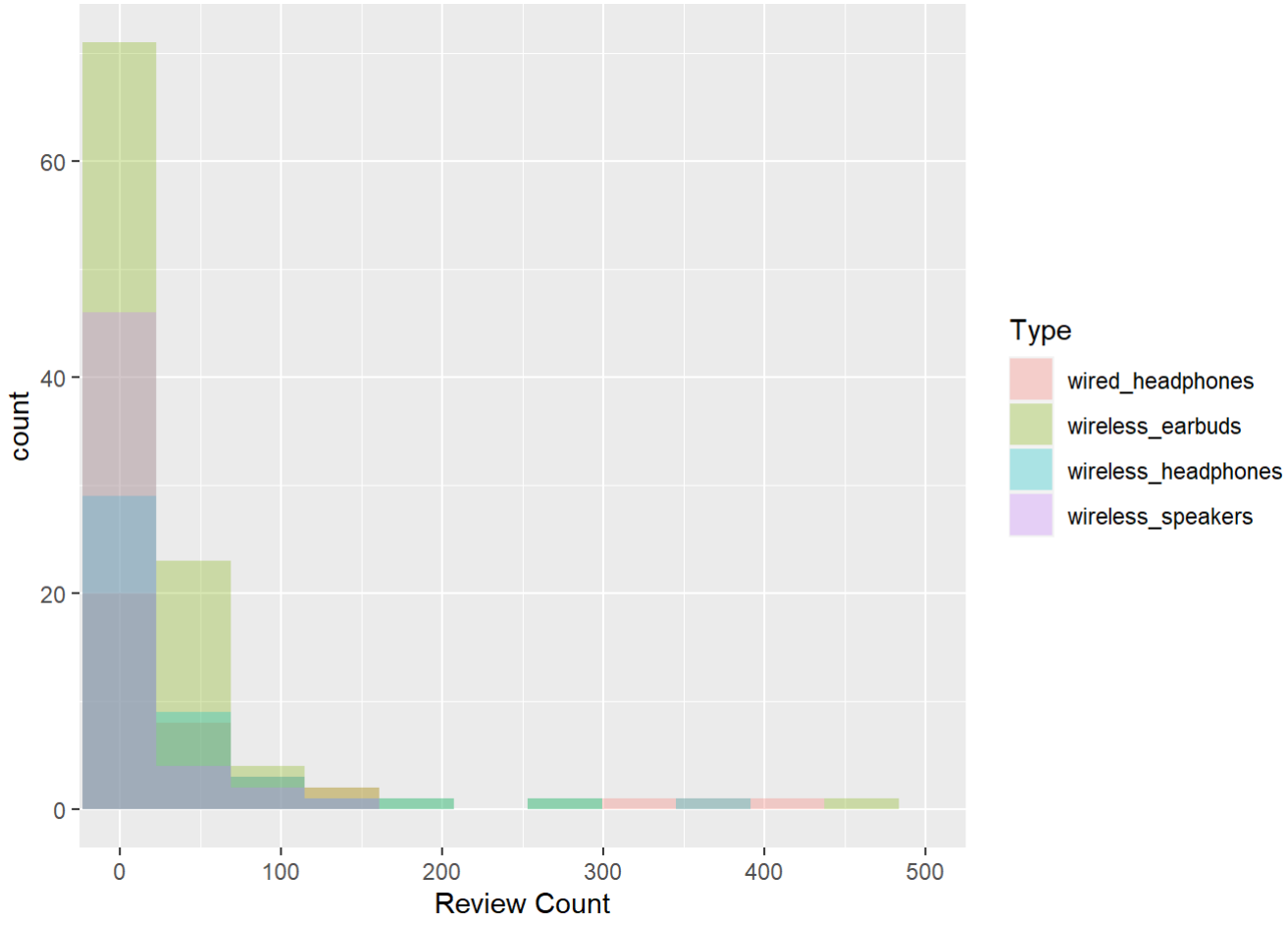
6.1. Scatterplot of Product Rating vs Sale_Price



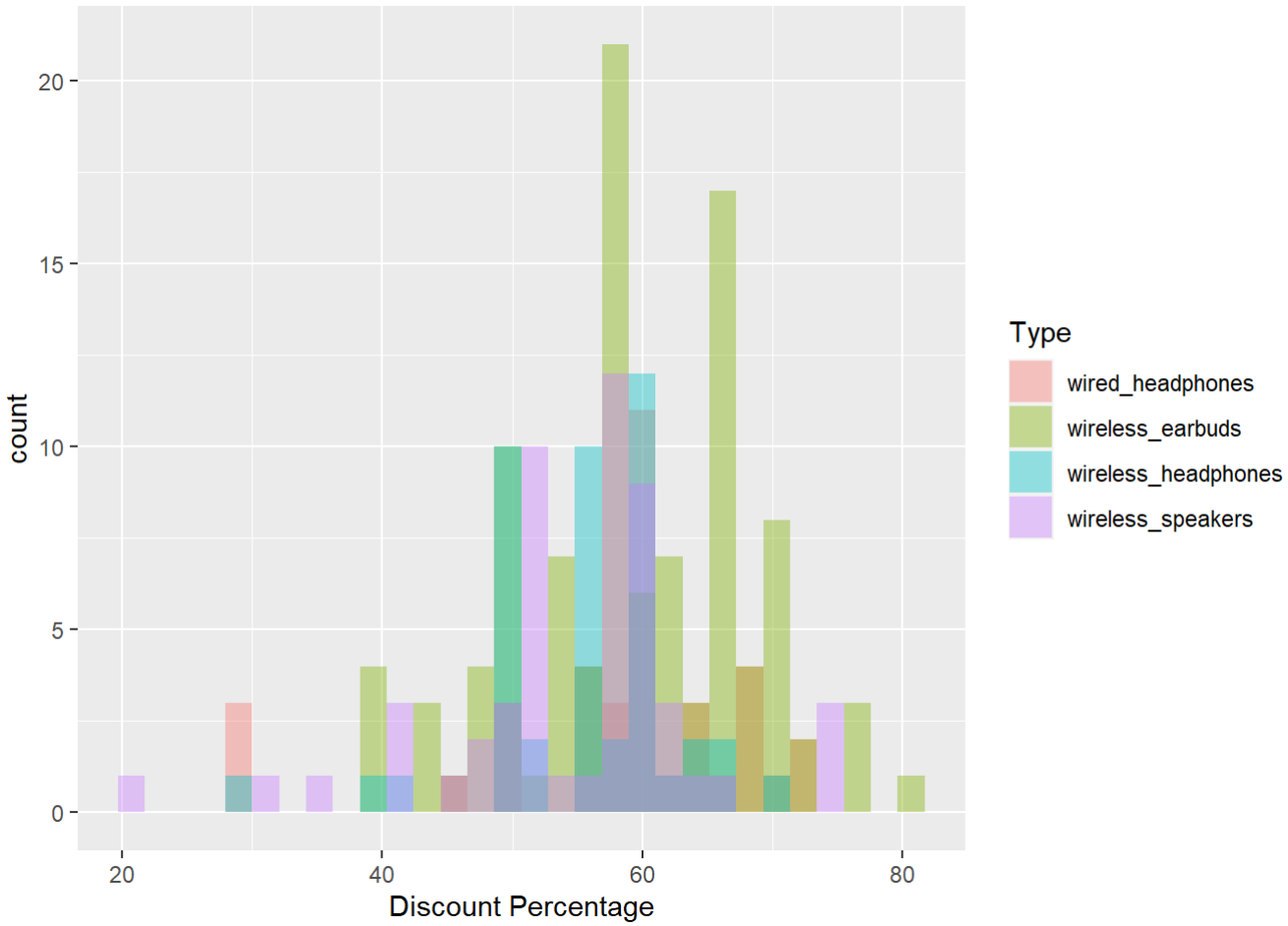
6.2. Histogram of Product Ratings of some Products



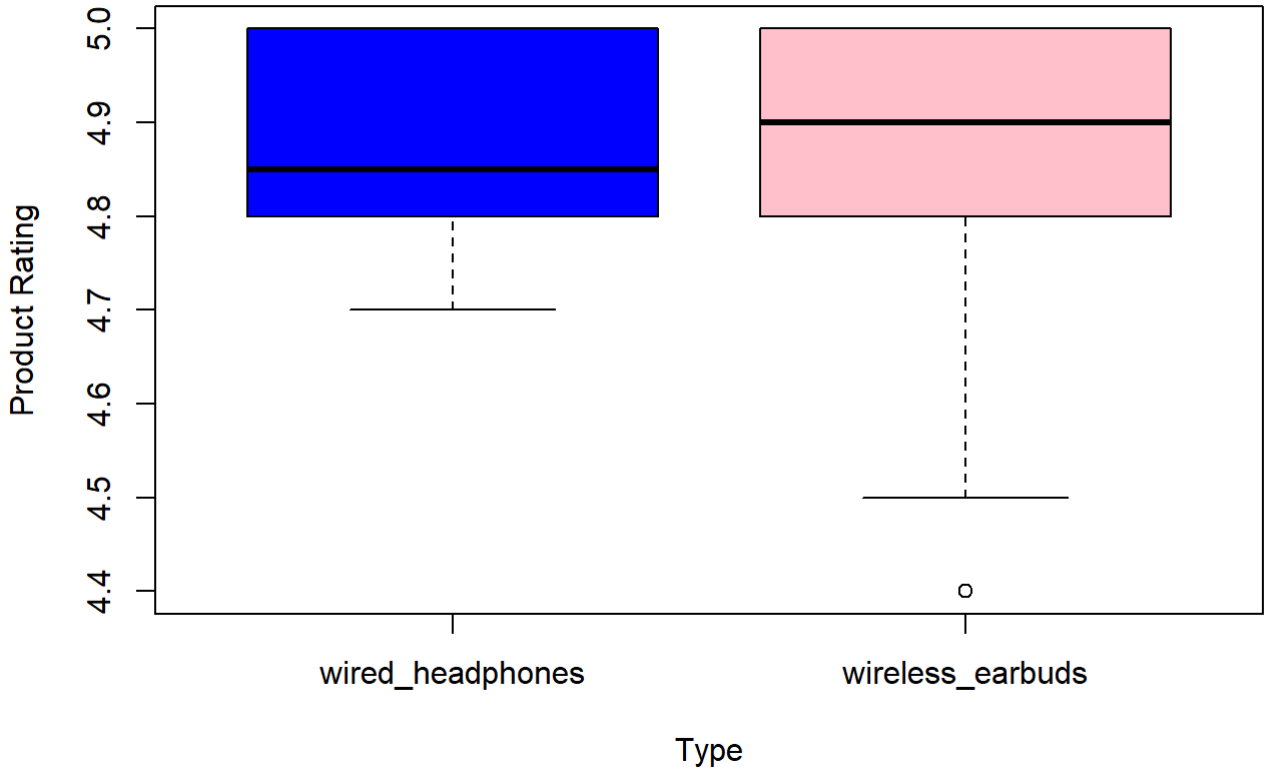
6.3. Histogram of Review Counts of some Products



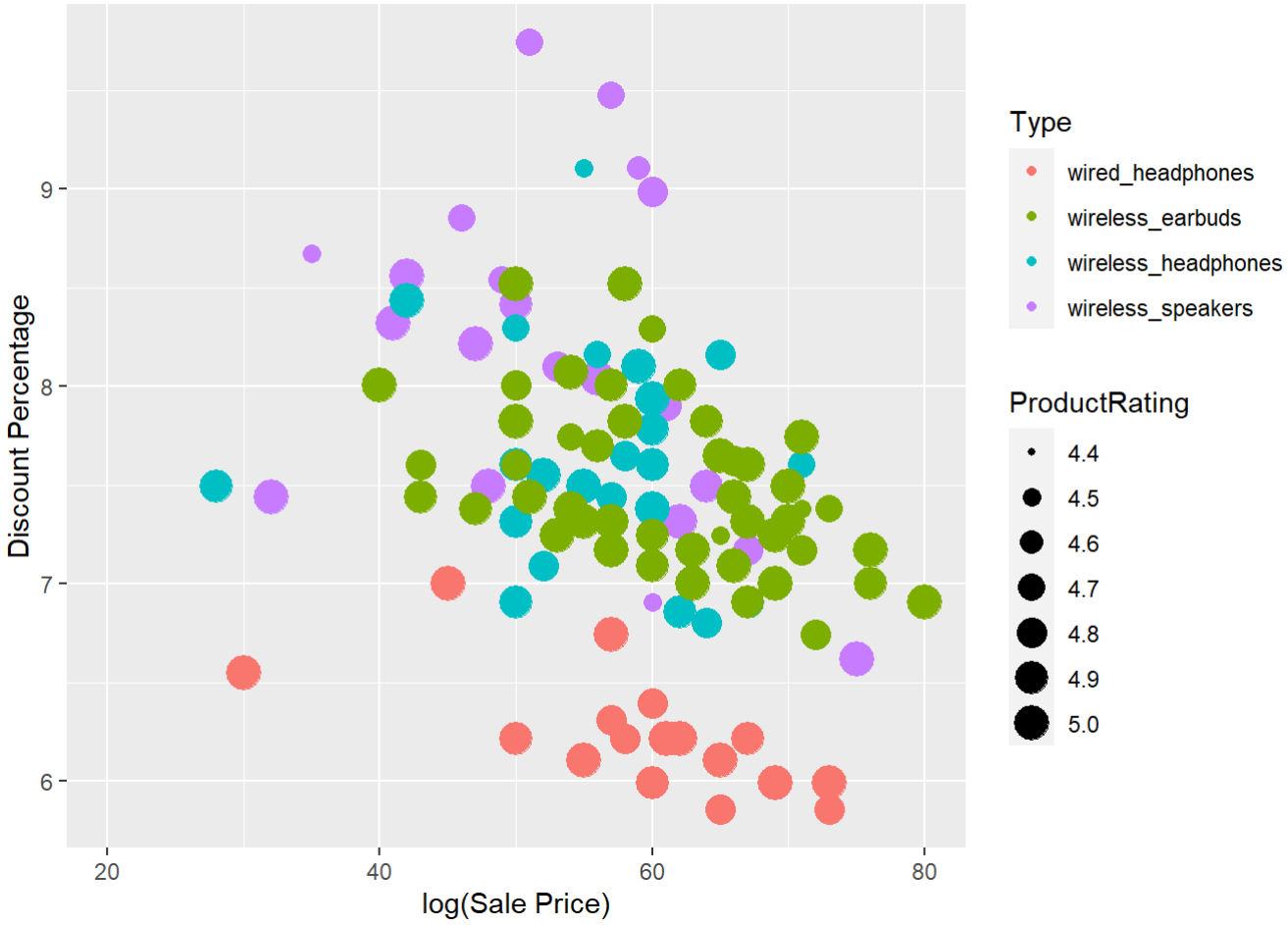
6.4. Histogram of Discount percentages



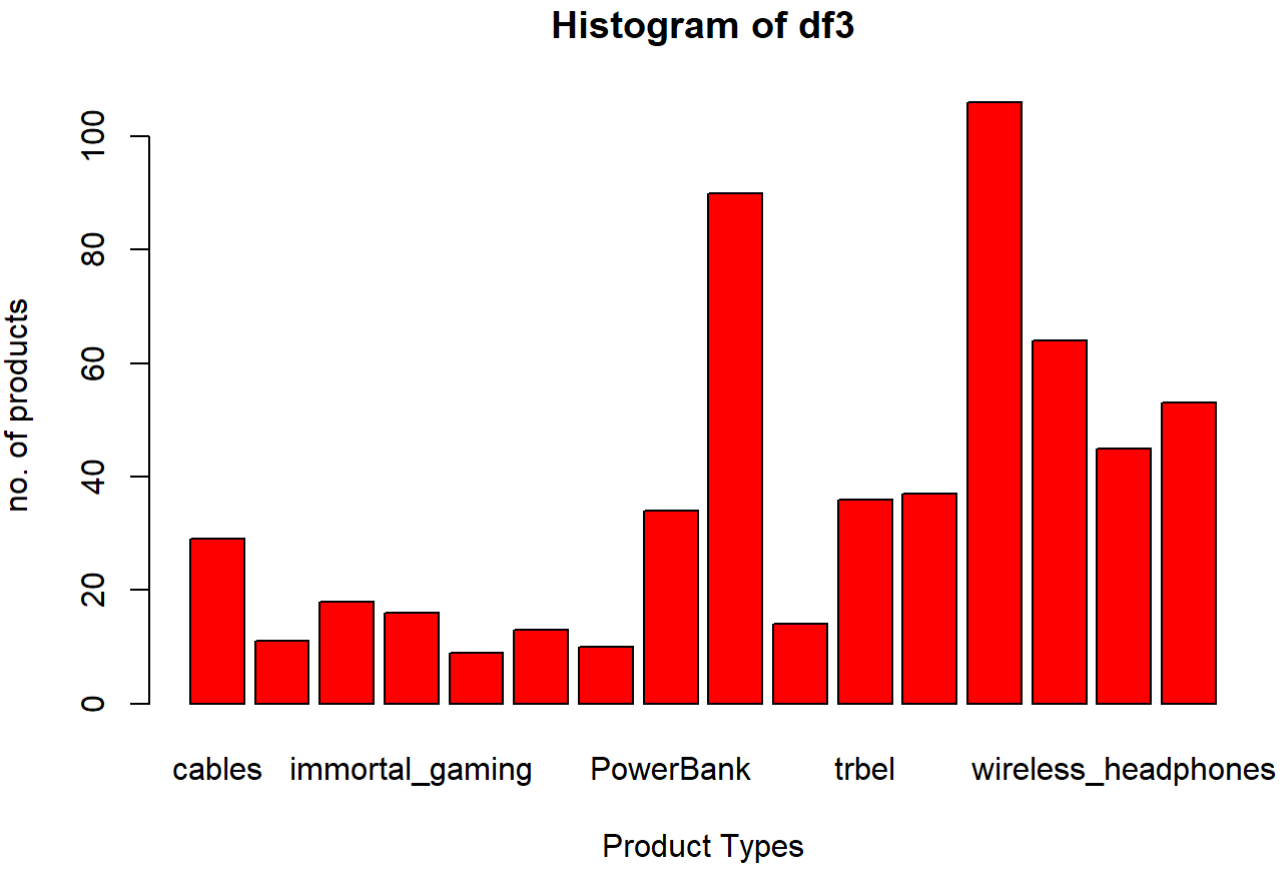
6.5. box plot of Product Ratings for wired and wireless earbuds



6.6. Scatterplot for Sale_Price vs Discount percentage to find the relation between Sale_price and Discount Percentage



6.7. Histogram of Various Product Types



7. Conclusions

The comprehensive analysis of Boat Lifestyle's product portfolio, supported by the Shiny app and insights from the Amazon dataset, has unveiled strategic marketing patterns and shed light on the company's pricing strategies.

1. **Main Products and Pricing Strategy:**
    - Our investigation highlighted that audio systems and smartwatches are the primary products sold by Boat Lifestyle. A striking observation from the analysis is the inverse relationship between sale price and discount percentage. Notably, wired headphones receive the highest discounts, suggesting a strategic approach to pricing.
2. **Discount Allocation by Product Type:**
    - The graph reveals that wireless earbuds, despite being priced lower than other products, receive higher discount percentages. This aligns with a potential marketing strategy where Boat aims to attract consumers with competitive pricing, coupled with substantial discounts, possibly to gain market share in this product category.
3. **Marketing Strategy of High Discount Percentages:**
    - An intriguing observation is the concentration of discount percentages around 50-60%. This consistent discount range suggests a deliberate and prominent market strategy employed by Boat. The company may be strategically balancing product pricing to appeal to a wide range of consumers without compromising perceived value.
4. **Dynamic Pricing of Earbuds:**
    - The notably lower prices of earbuds compared to other products, coupled with higher discount percentages, indicate a distinct marketing strategy. Boat appears to position earbuds as an affordable entry point, strategically increasing the sale price while offering significant discounts to attract cost-conscious consumers.
5. **Target Audience and Product Display:**
    - Boat's strategic product display aligns with targeting an audience with budgets between 0 and 5000. By showcasing products with ratings between 4.5 and 5, Boat is likely aiming to appeal to discerning consumers seeking high-rated products within budget constraints.
6. **Balancing Discounts for Perceived Value:**
    - The observed marketing strategy of offering moderate discounts suggests that Boat aims to strike a balance between perceived value and affordability. This approach can contribute to maintaining a premium brand image while making products accessible to a broader consumer base.

In conclusion, Boat Lifestyle's marketing and pricing strategies, as revealed through our analysis, showcase a deliberate and calculated approach. The nuanced balance between sale price, discount percentages, and product ratings aligns with a strategic effort to capture a diverse consumer market.

These findings not only provide insights into Boat's current positioning but also offer valuable considerations for future marketing and pricing decisions.

8. References

Sources of the data:

- Boat Lifestyle's Official Website (boAt official site)
- Kaggle Datasets (amazon earbuds)

resources for Python selenium library:

- learn selenium

To know more about boat

- boat marketing strategy