

ANALYSIS ON AMAZON DATA



PERFORM EDA

BY:

HARSH THAKUR

NEHA BAIRAGI

RISHAB SEN

PRAPTI SHAH



01 - Introduction

02 - Data Cleaning

03 - Insights Drawn

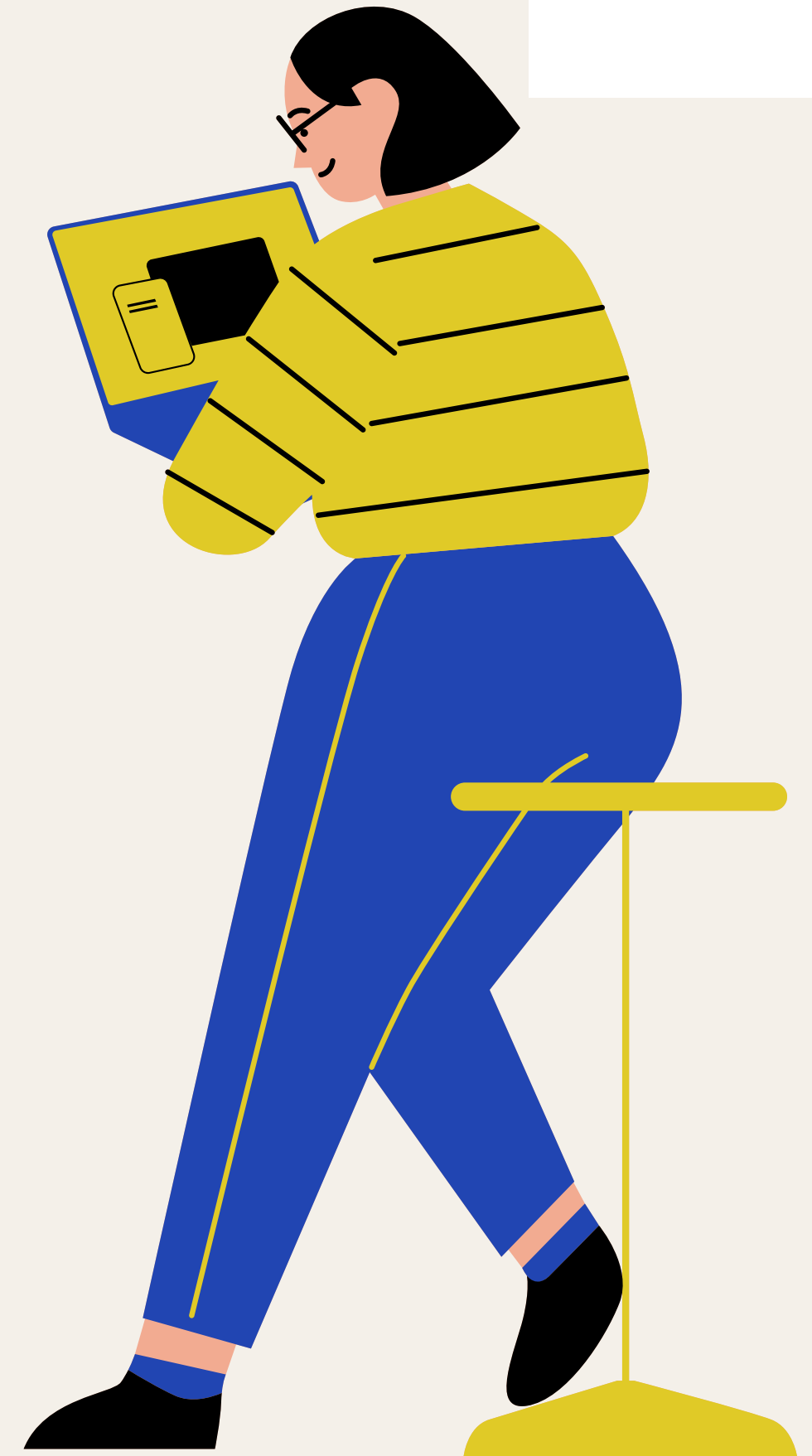
04 - Data Visualization

05 - Data Limitation

06 - Analysis Impact On Data

Data

Visualization



Introduction

- Embarking on an EDA journey with the Amazon dataset reveals a treasure trove of insights.
- Columns include product identifiers, pricing details, customer reviews, and ratings.
- Objective: Delve deep into the data, uncover patterns, and derive actionable insights.
- Aim: Drive informed decision-making and enhance user experiences.
- Let's embark on this insightful journey together!

***Analyzing data
enables informed
decision-making***

amazon



Data

Visualization

Data set :



	product_id	product_name	category	discounted_price	actual_price	discount_percentage
0	B07JW9H4J1	Wayona Nylon Braided USB to Lightning Fast Cha...	Computers&Accessories Accessories&Peripherals ...	399.0	1099.0	64.0

rating	rating_count	about_product	user_id	user_name
4.2	24269.0	High Compatibility : Compatible With iPhone 12...	AG3D6O4STAQKAY2UVGEUV46KN35Q,AHMY5CWJMMK5BJRBB...	Manav,Adarsh gupta,Sundeep,S.Sayeed Ahmed,jasp...

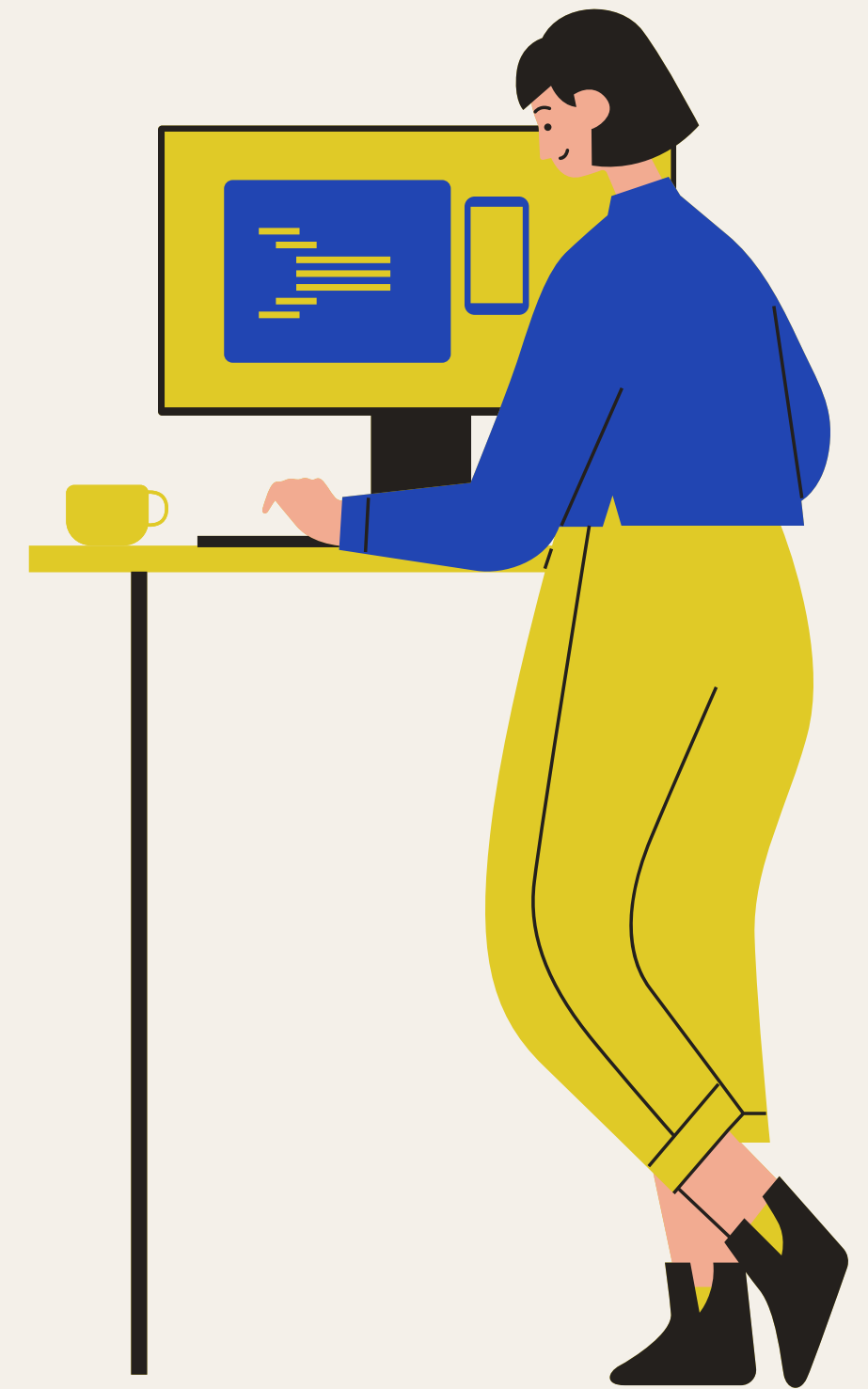
review_id	review_title	review_content	img_link	product_link
R2AJM3LFTLZHFO,R6AQJGUP6P86,R1K...	Satisfied,Charging is really fast,Value for mo...	Looks durable Charging is fine tooNo complains...	https://m.media-amazon.com/images/W/WEBP_40237...	https://www.amazon.in/Wayona-Braided-WN3LG1-Sy...

Data Cleaning

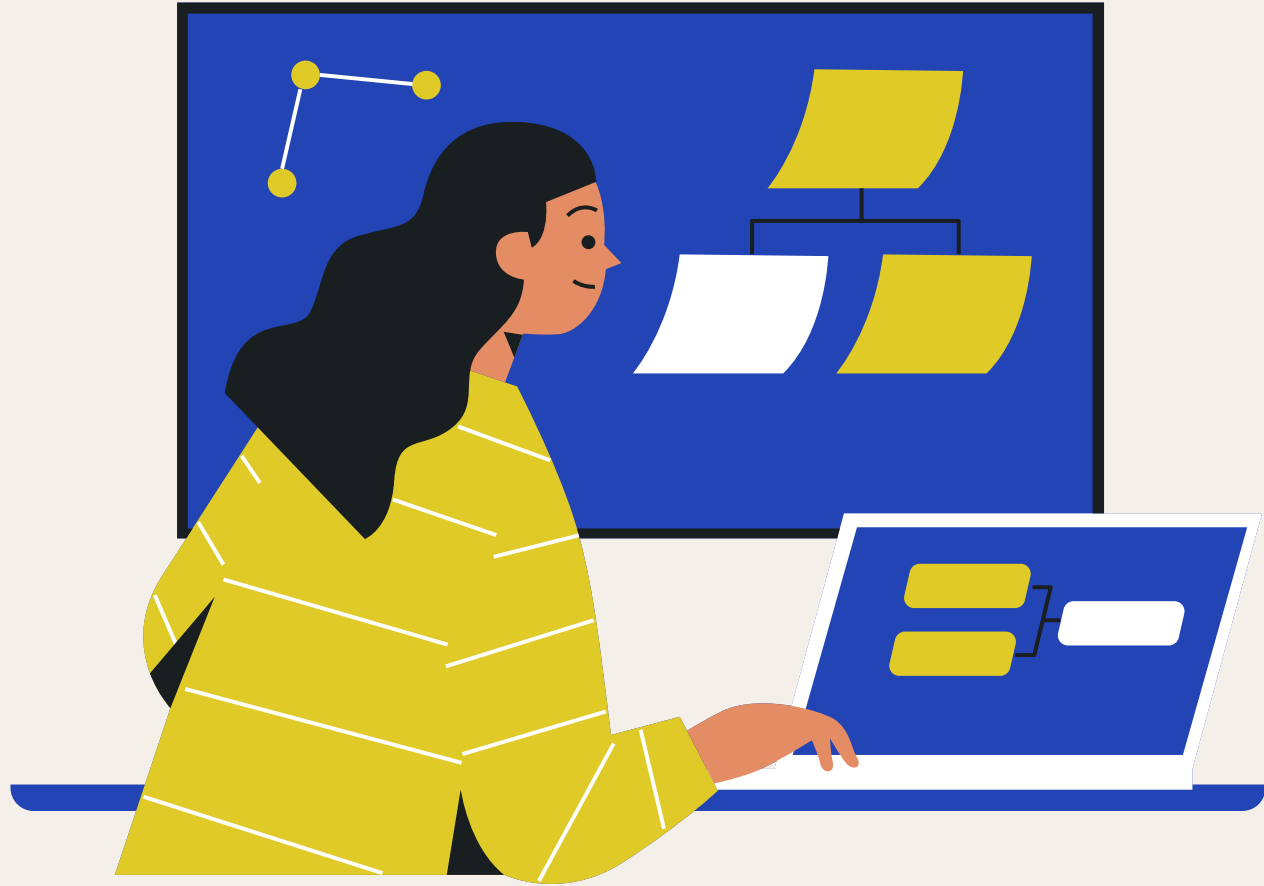
“Data cleaning is like untangling a knot – it may seem tedious at first, but once done, it reveals a clear path forward”.



- Identified formatting inconsistencies in the dataset.
- Rectified issues by removing currency symbols and commas from numerical data.
- Utilized the `isnull().sum()` function to identify null values.
- Employed the `dropna()` function to eliminate null values from the dataset.
- Ensured correct data types for columns intended as integers or floats.
- Employed descriptive functions like `describe()` and `head()` to analyze the dataset further.



Insights Drawn



Data visualization simplifies the communication of analysis findings

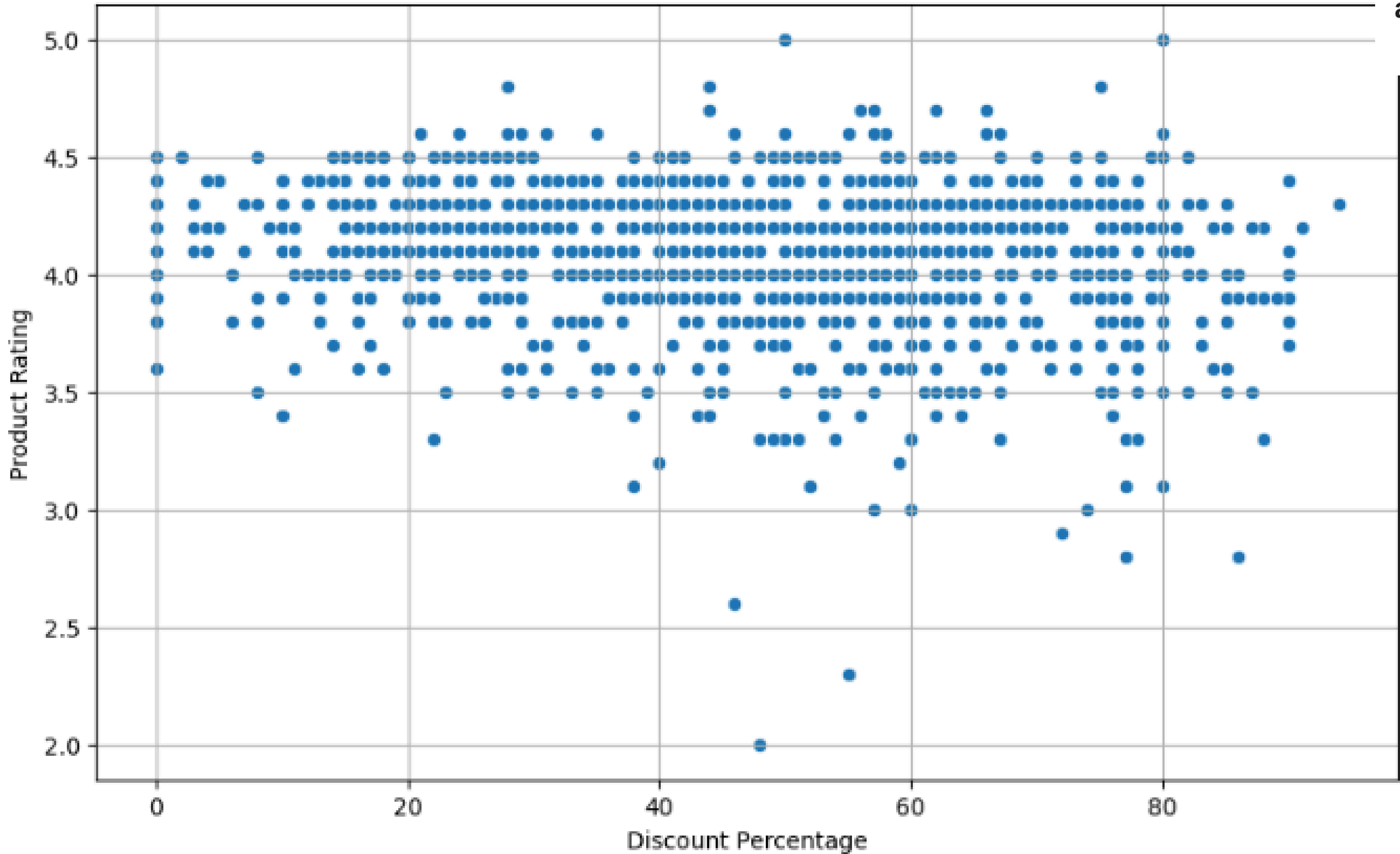
Data

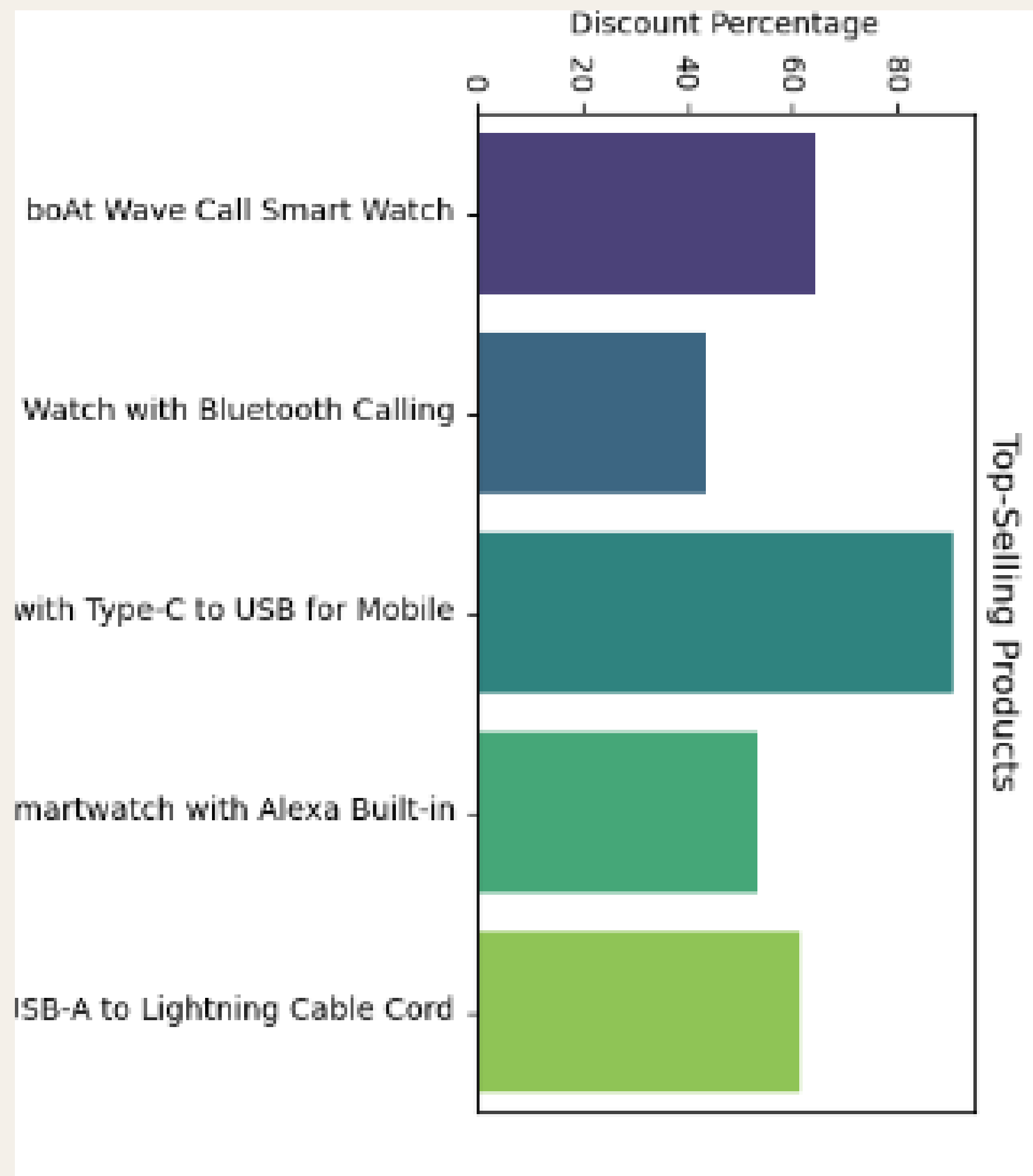
Visualization

We drawn some Following usefull Insights some of them are:

1. Correlation between discount percentage and product rating ●
2. The discount percentage vary across different brands ●
3. Insights from this analysis improve product recommendations ●
4. Rating count over the rating score. ●
5. Product prices distributed, and are there notable outliers ●

Discount Percentage vs. Product Rating





amazon

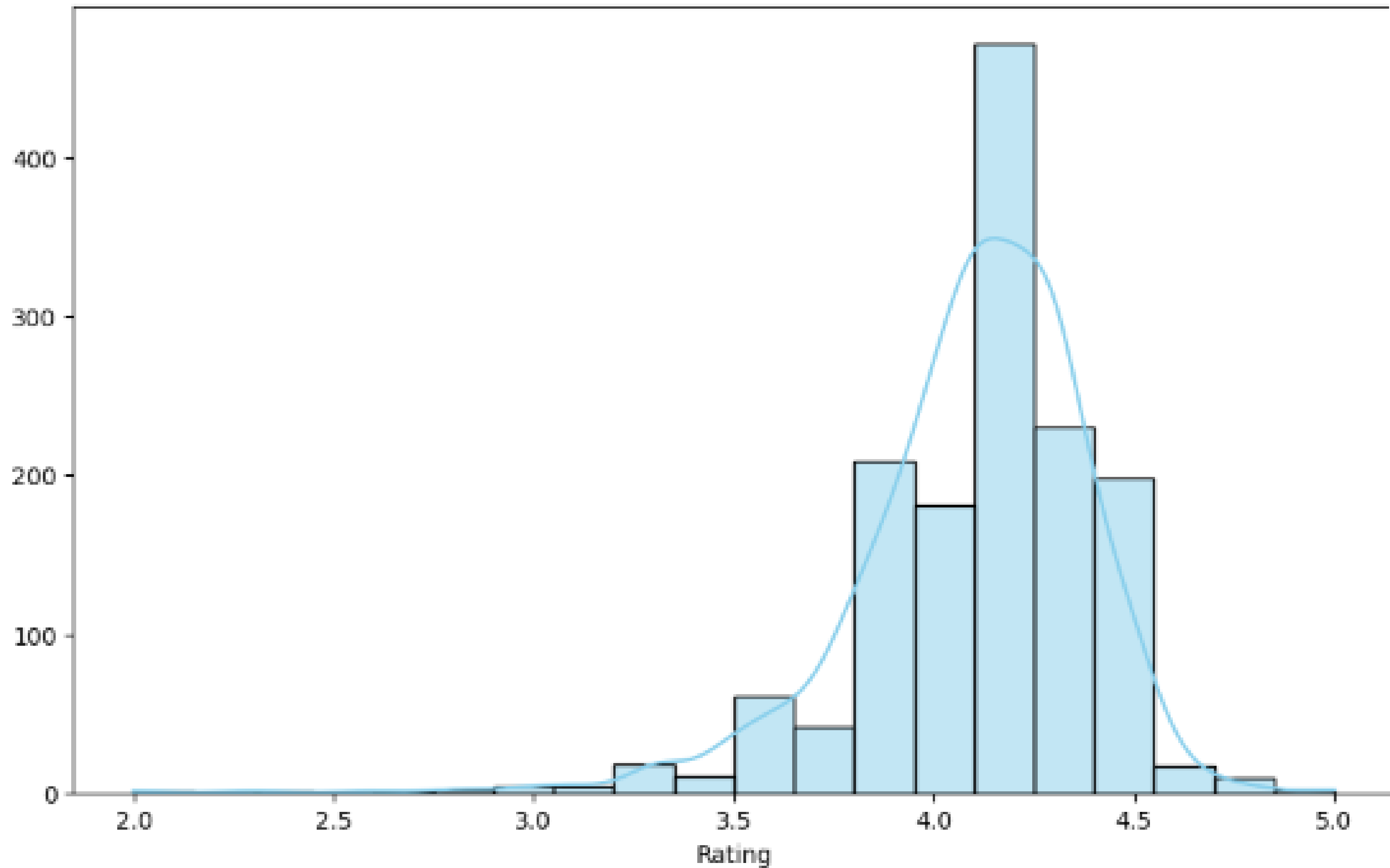
```
data['pro_name'] = data['product_name'].str.split(',').str[0]
data['pro_name']
```

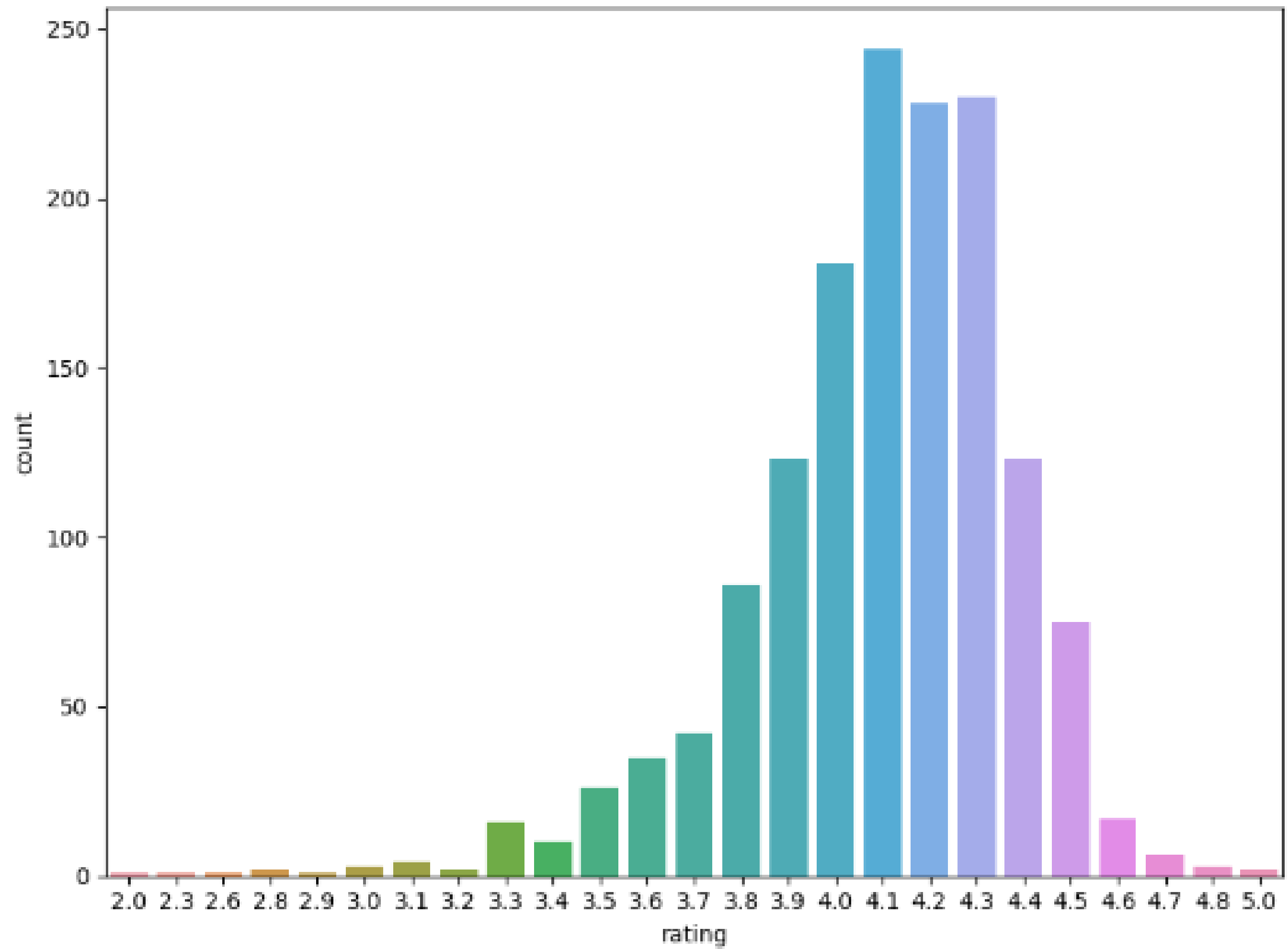
```
0    Wayona Nylon Braided USB to Lightning Fast Cha...
1    Ambrane Unbreakable 60W / 3A Fast Charging 1.5...
2    Source Fast Phone Charging Cable & Data Sync U...
3    boAt Deuce USB 300 2 in 1 Type-C & Micro USB S...
4    Portronics Konnect L 1.2M Fast Charging 3A 8 P...
```

...

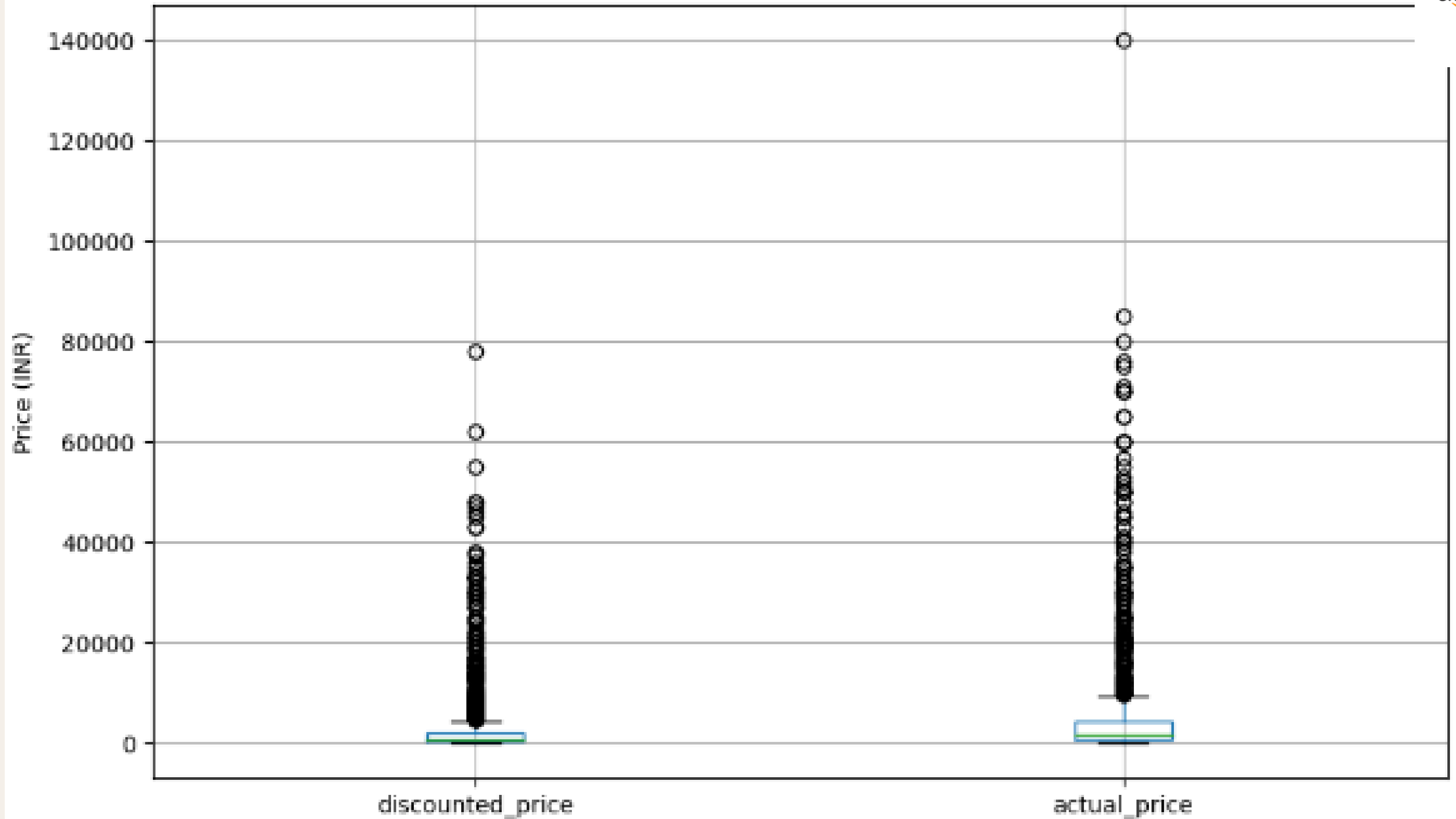
```
1460    Noir Aqua - 5pcs PP Spun Filter + 1 Spanner | ...
1461    Prestige Delight PRW0 Electric Rice Cooker (1 L
1462    Bajaj Majesty RX10 2000 Watts Heat Convector R...
1463    Havells Ventil Air DSP 230mm Exhaust Fan (Pist...
1464    Borosil Jumbo 1000-Watt Grill Sandwich Maker (...
Name: pro_name, Length: 1462, dtype: object
```


Distribution of Ratings





Distribution of Product Prices



Data Limitation

While the Amazon dataset provides valuable information for analysis, it also has several limitations that need to be considered:

- 1.Data Completeness and Quality
- 2.Sampling Bias
- 3.Temporal Bias
- 4.Limited Variables
- 5.Categorical Data Encoding
- 6.Assumptions and Simplifications
- 7.Ethical and Privacy Concerns
- 8.Limited Scope



Analysis Impact On Data



Insights from analyzing the Amazon dataset can drive enhancements in product recommendations. Understanding the correlation between price and rating enables targeted recommendations, offering higher-priced products to quality-conscious users. Moreover, analyzing the relationship between description length and price/rating highlights the importance of comprehensive product information in influencing perceived value. By delving into user preferences and behavior, personalized recommendations can be curated using collaborative and content-based filtering techniques. Continuously monitoring user feedback facilitates a feedback loop for product improvement, guiding updates to enhance quality and features.

Thanks.

