



# DATA ANALYSIS FOR AMAZON



X



# TEAM MEMBERS :

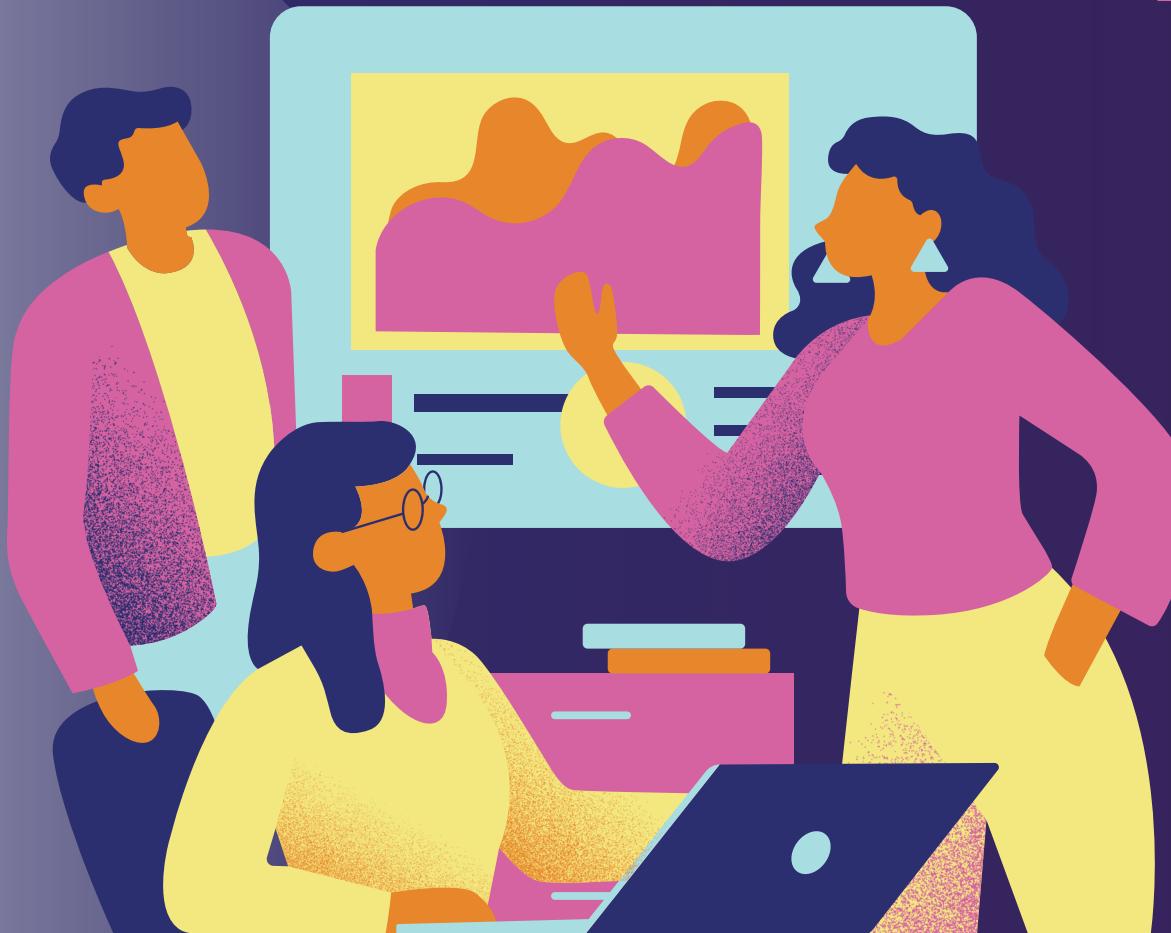
**AMANY IBRAHIM MAHMOUD MAHMOUD**  
**RESPONSIBLE FOR THE PRESENTATION,**  
**INSIGHTS, ORGANIZING THE CODE, WRITING SOME**  
**STEPS, AND A SMALL PART OF THE CODE**

**Mostafa Gamal**  
**Responsible for writing most of the code and**  
**creating the README.md file**



# WHAT IS THIS PROJECT ABOUT? THIS DATASET CONTAINS DATA FOR +1000 AMAZON PRODUCTS

## PROBLEMS



- Data size
- The presence of columns in an incorrect data format
- Column name length
- The presence of outliers and missing data

# SOLVING

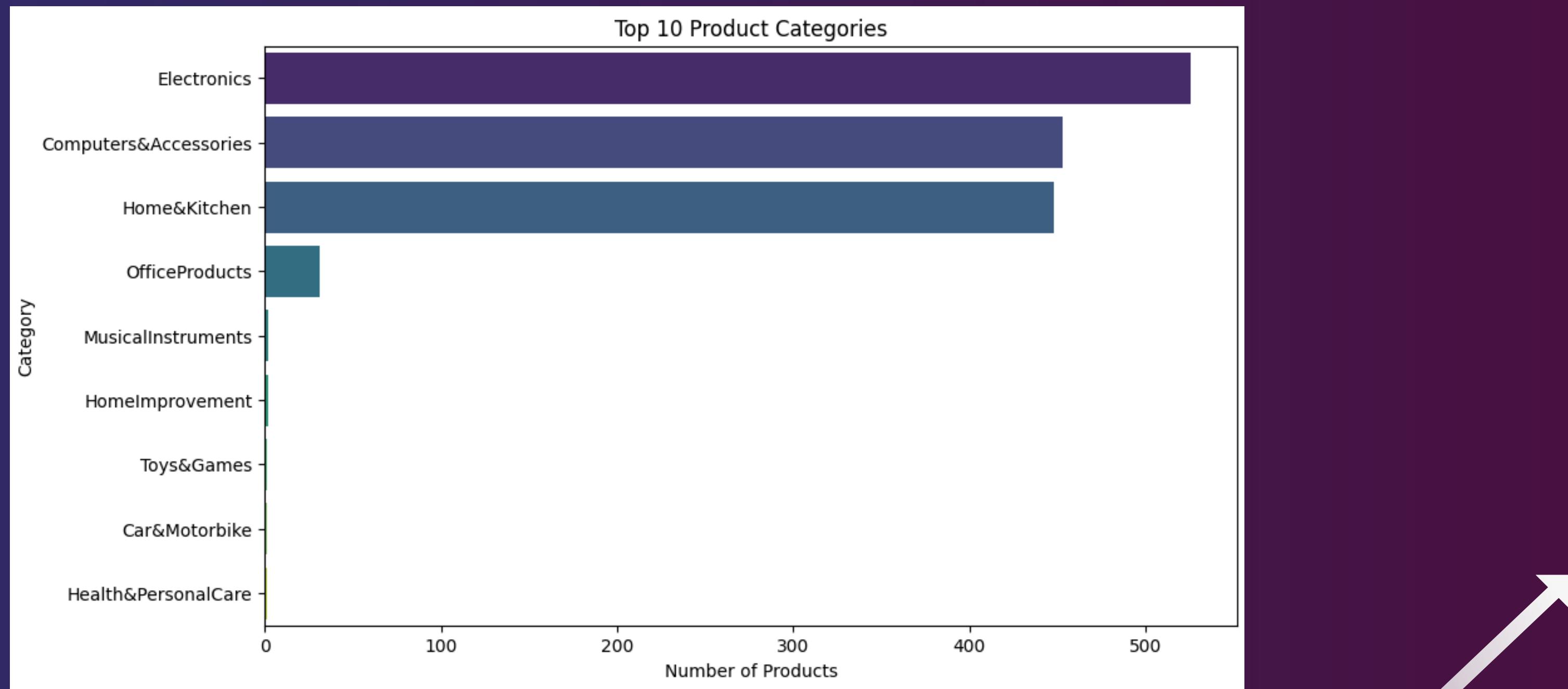
WORKING ON A PART OF THE DATA

REMOVING ANY SYMBOLS AND CONVERTING THE DATA TYPE  
 TO NUMERIC USING TO\_NUMERIC

باستخدام ال feature Engineering

CALCULATING THE NUMBER OF  
 MISSING VALUES AND IMPUTING THEM  
 USING THE MEAN, MEDIAN, OR 0





# DATA ANALYSIS

## Insights

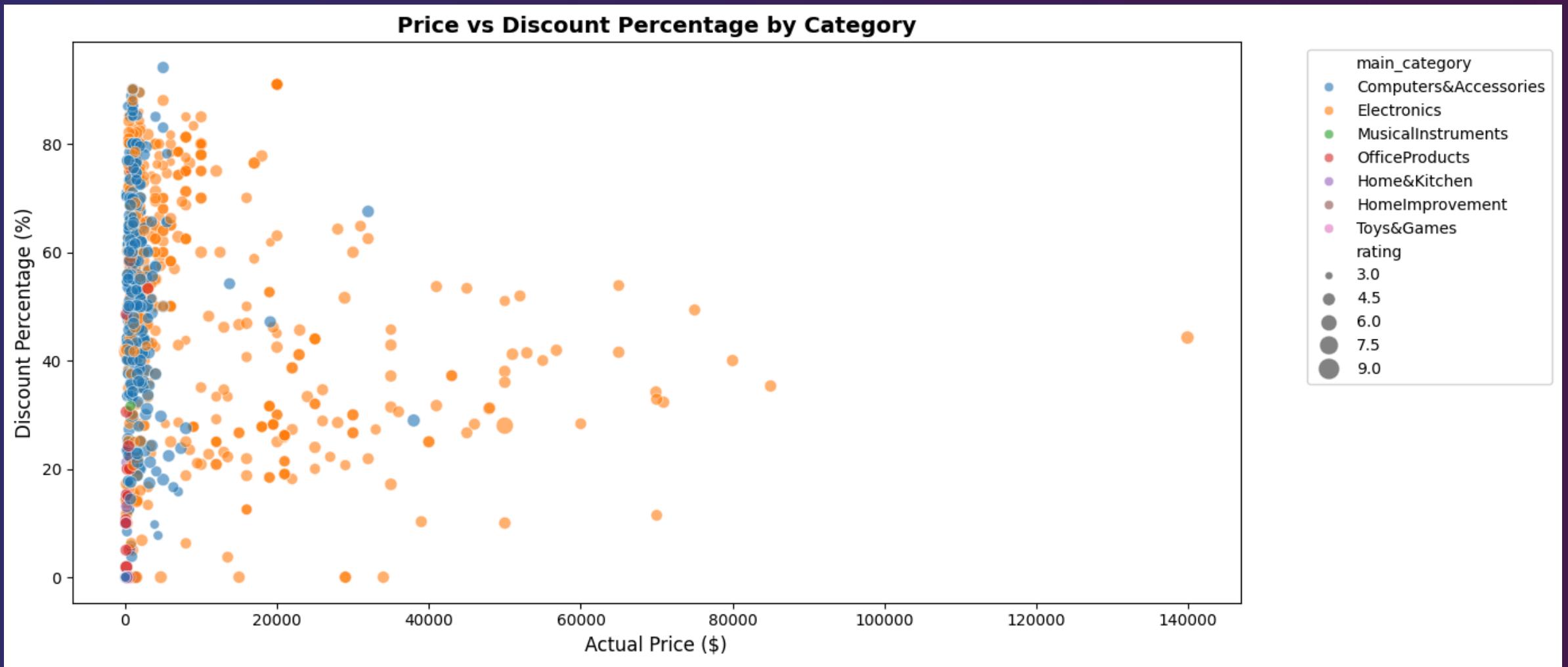
- I- Electronics is the highest category
  - Extreme values were detected.

## Key Observations

- There are extreme outliers.
  - The other categories are few in number compared to the first 3.

## business recommendations

- Increase advertising and marketing campaigns targeting higher income brackets
  - offer discounts to lower income brackets
  - promote products in lower income brackets.



## \*\*Insights\*\*

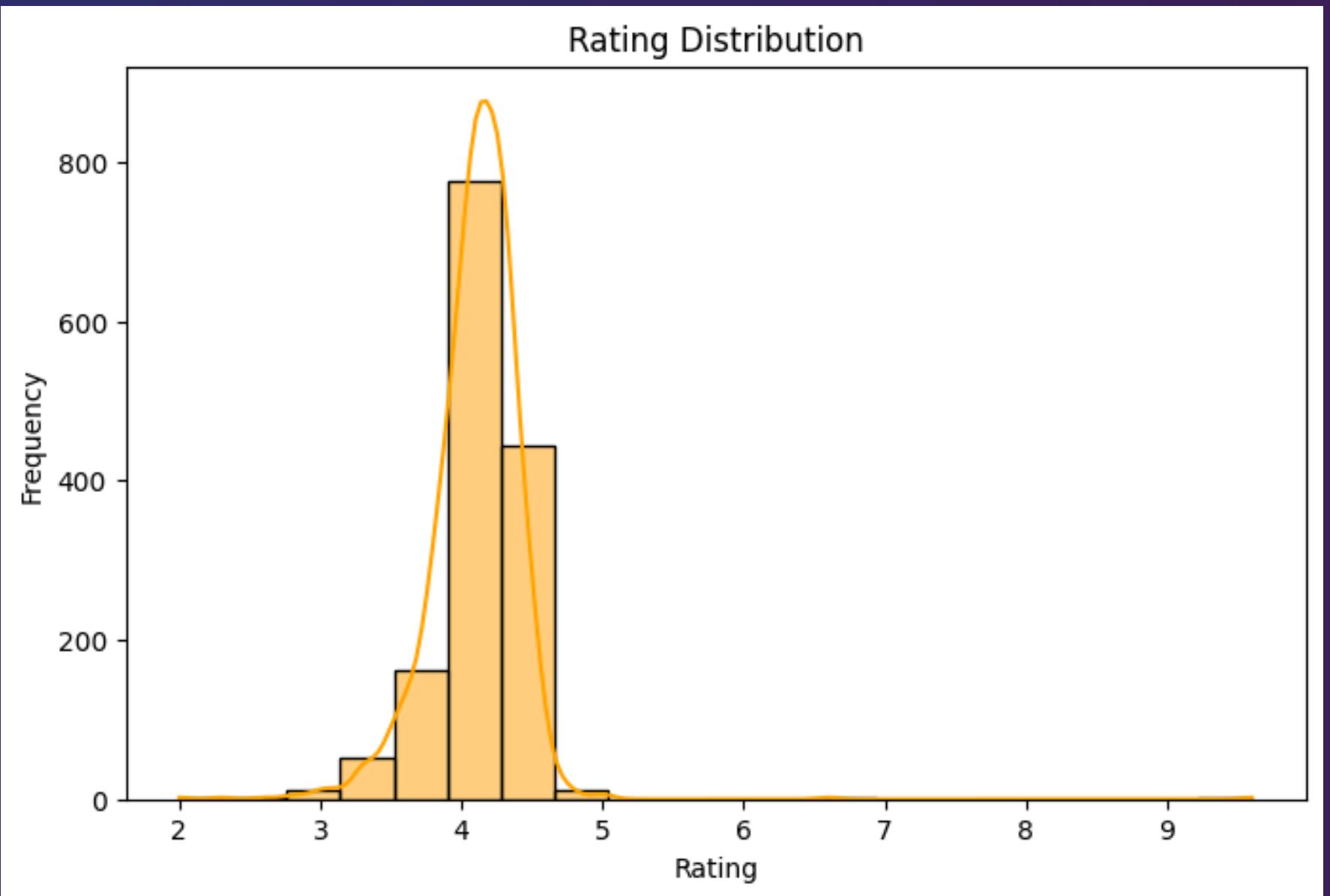
**Discounts increase on lower-priced products**

- discounts are high lower-priced products
- ranging from 15% to 95%, which is a high percentage.

## business recommendation

**Offering discounts on higher-priced products**

- studying whether high-discounted products generate profits
- increasing discounts on lower-demand products to attract customers



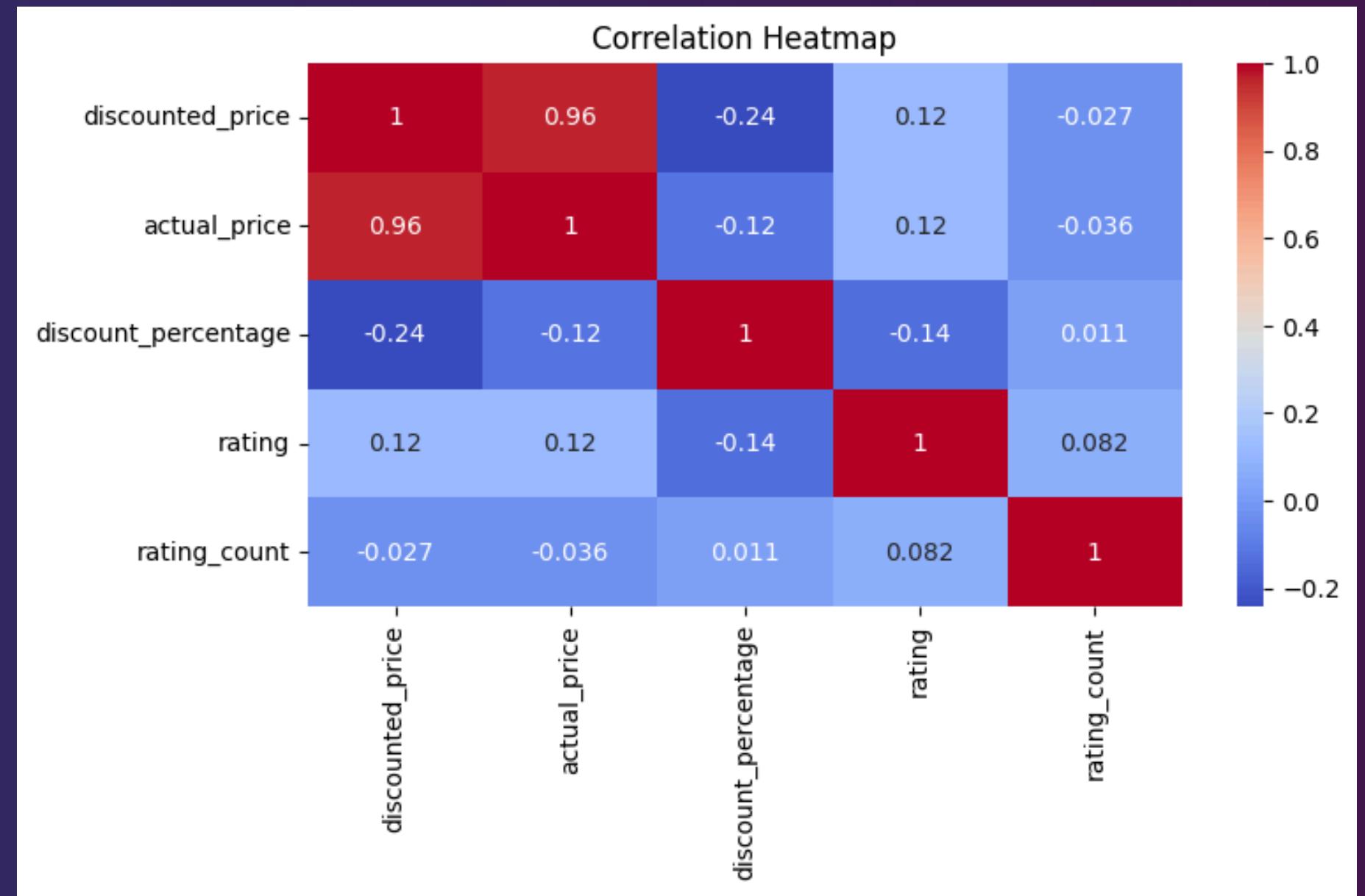
# INSIGHTS

THE RATING FALLS BETWEEN 4 AND 4.3, MEANING THE SERVICE IS VERY GOOD AND CONSISTENT

# RECOMMENDATIONS

The study aims to identify and improve the poor evaluation of low-rated products.

- Increase the number of products with high values
- Promote the highest-rated products and display a “top-rated” badge to encourage purchases



# BUSINESS RECOMMENDATIONS



## IMPROVE PRODUCT QUALITY



Encourage customers to leave reviews.  
For example, offer discount coupons to customers who review the product they purchased

### \*\*Key Observations\*\*

- علاقة طردية قوية → 1

- علاقة عكسية قوية → -1

- لا يوجد ارتباط → 0

- The strongest relationship between the actual price and the discounted price is 0.96.
- The weakest relationship is between the discounted price and the discount percentage(-.24).
- The relationship between rating and rating count is weak.

### \*\*INSIGHTS\*\*

- THE ORIGINAL PRICE IS CORRELATED WITH THE DISCOUNTED PRICE.
- A HIGHER PRICE DOES NOT IMPLY A HIGHER RATING.
- A SMALL NUMBER OF CUSTOMERS PROVIDE FEEDBACK ON THE PRODUCTS.

# INSIGHTS

NEAR-PERFECT LEADERS

AMAZON BASICS WIRELESS AND REDTECH USB-C TO LIGHTNING  
PRODUCT CATEGORY -

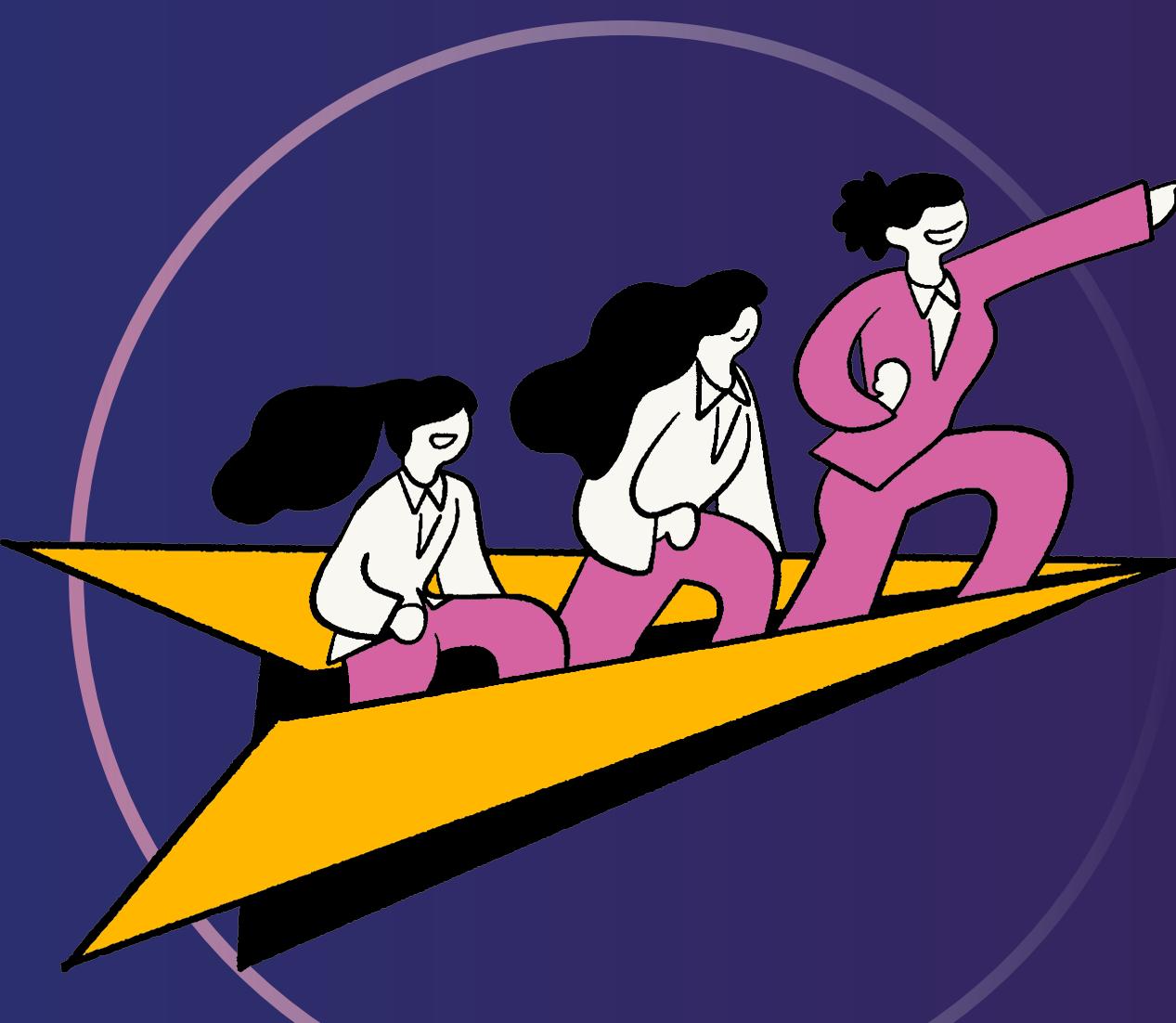
:MOST PRODUCTS FALL INTO

ELECTRONICS ACCESSORIES (USB CABLES, SCREEN PROTECTORS)

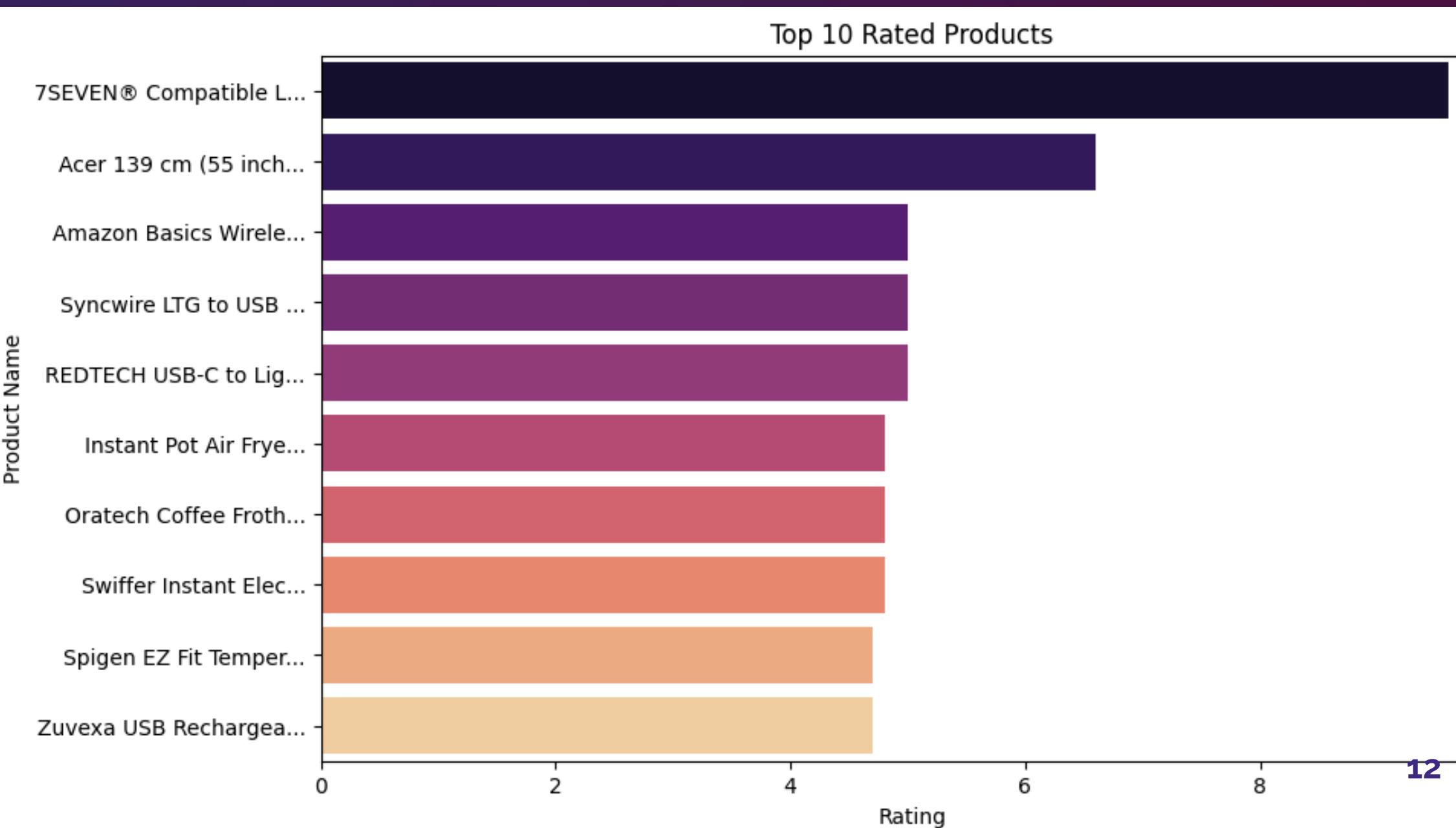
SMALL HOME APPLIANCES

UTILITY TOOLS

.THE RATINGS ARE HIGH -



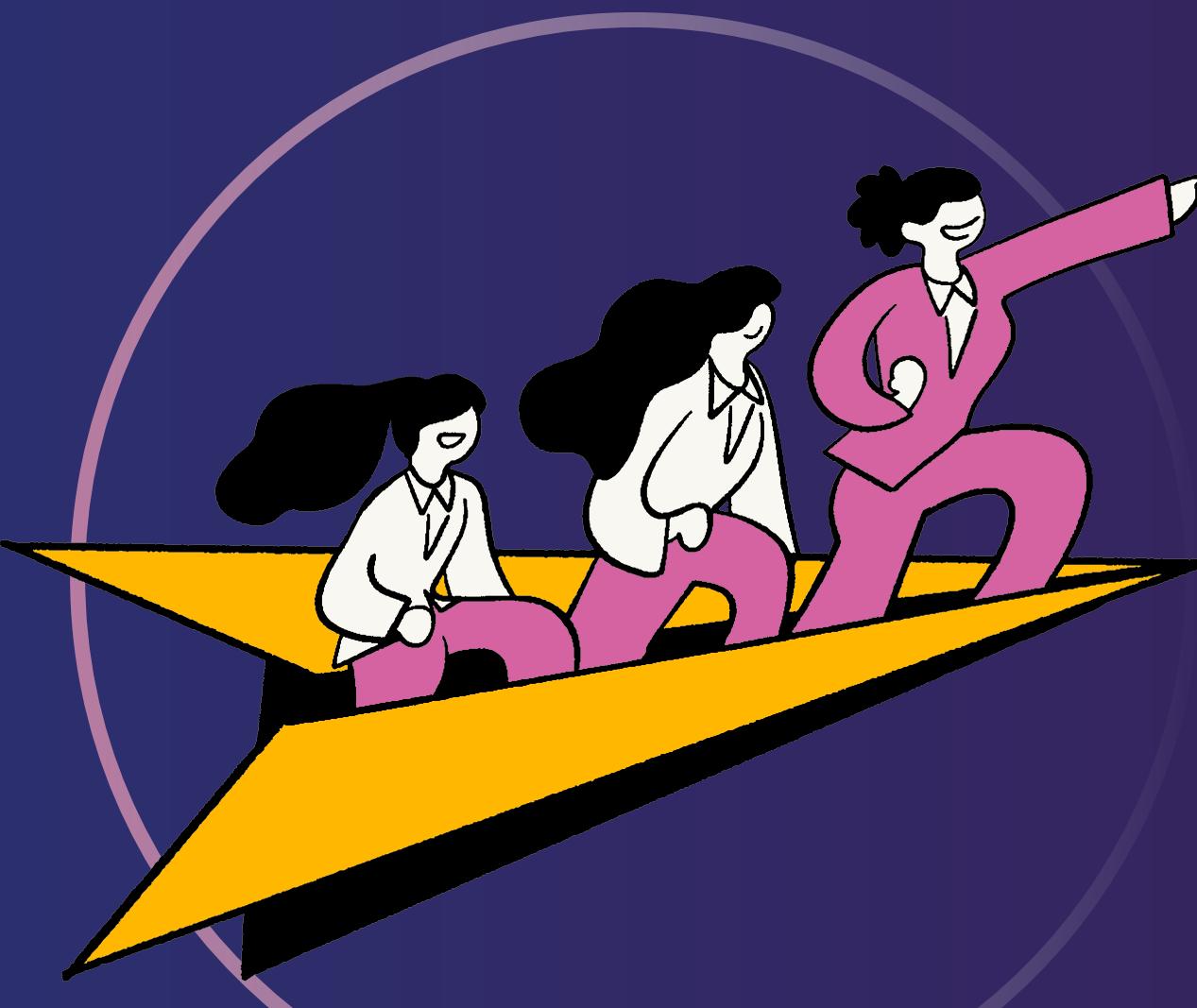
Top 10 Rated Products



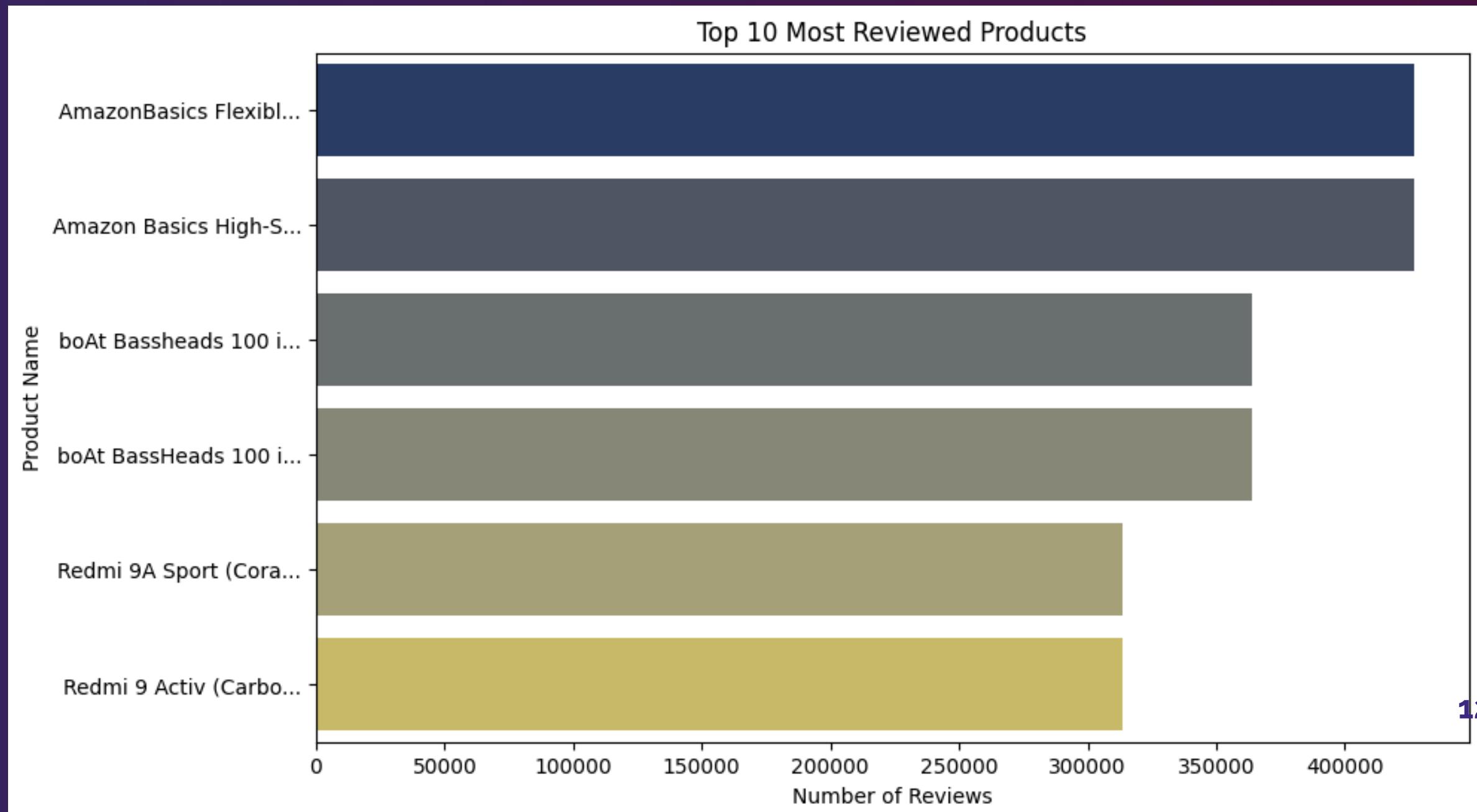
## INSIGHTS

### INSIGHTS\*\*

- THE \*\*AMAZON BASICS HIGH-SPEED HDMI CABLE\*\* HAS THE HIGHEST NUMBER OF REVIEWS.
- THE HIGHEST-RATED PRODUCT DOES NOT NECESSARILY HAVE THE MOST REVIEWS.

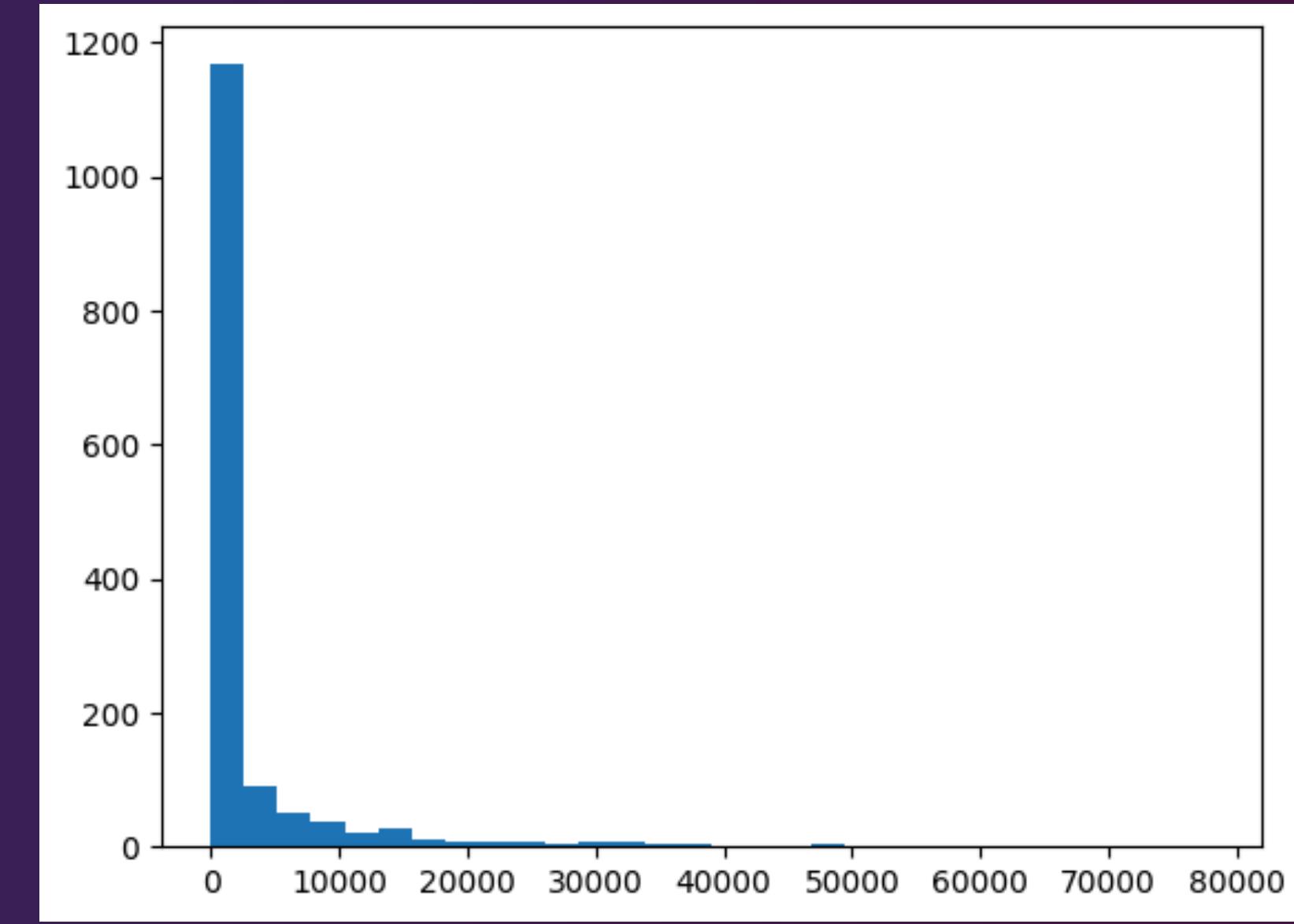
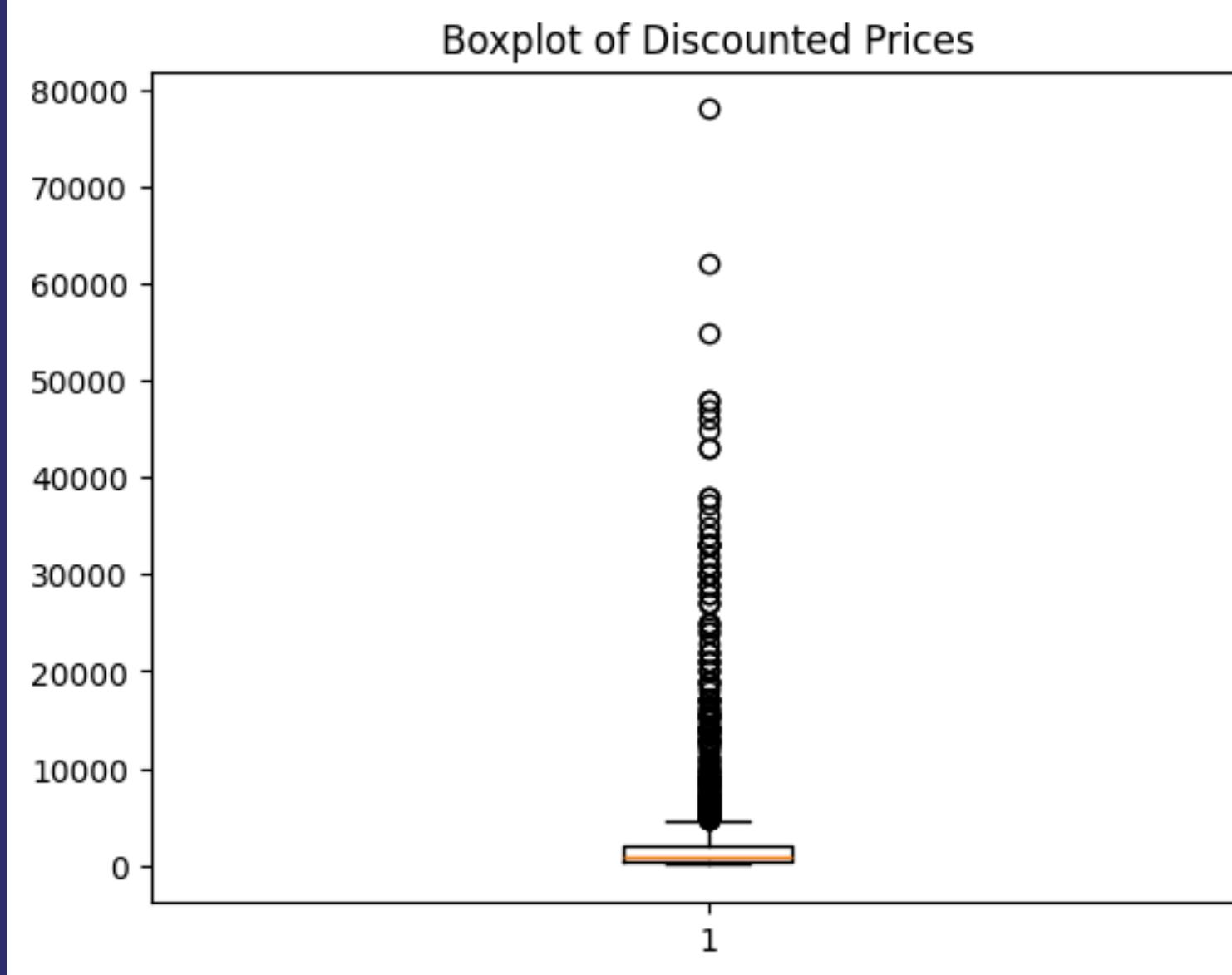


Top 10 Most Reviewed Products

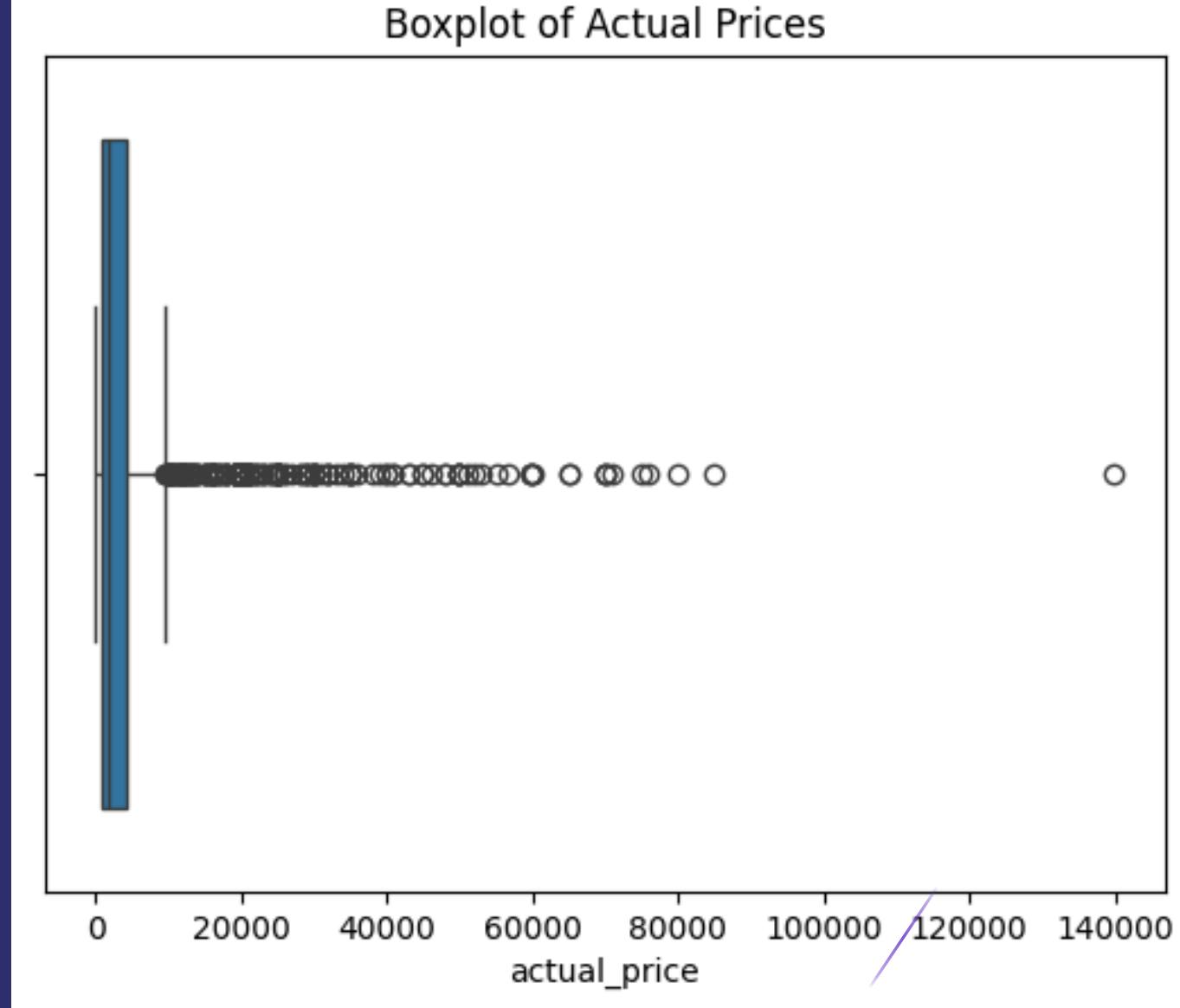


## insights

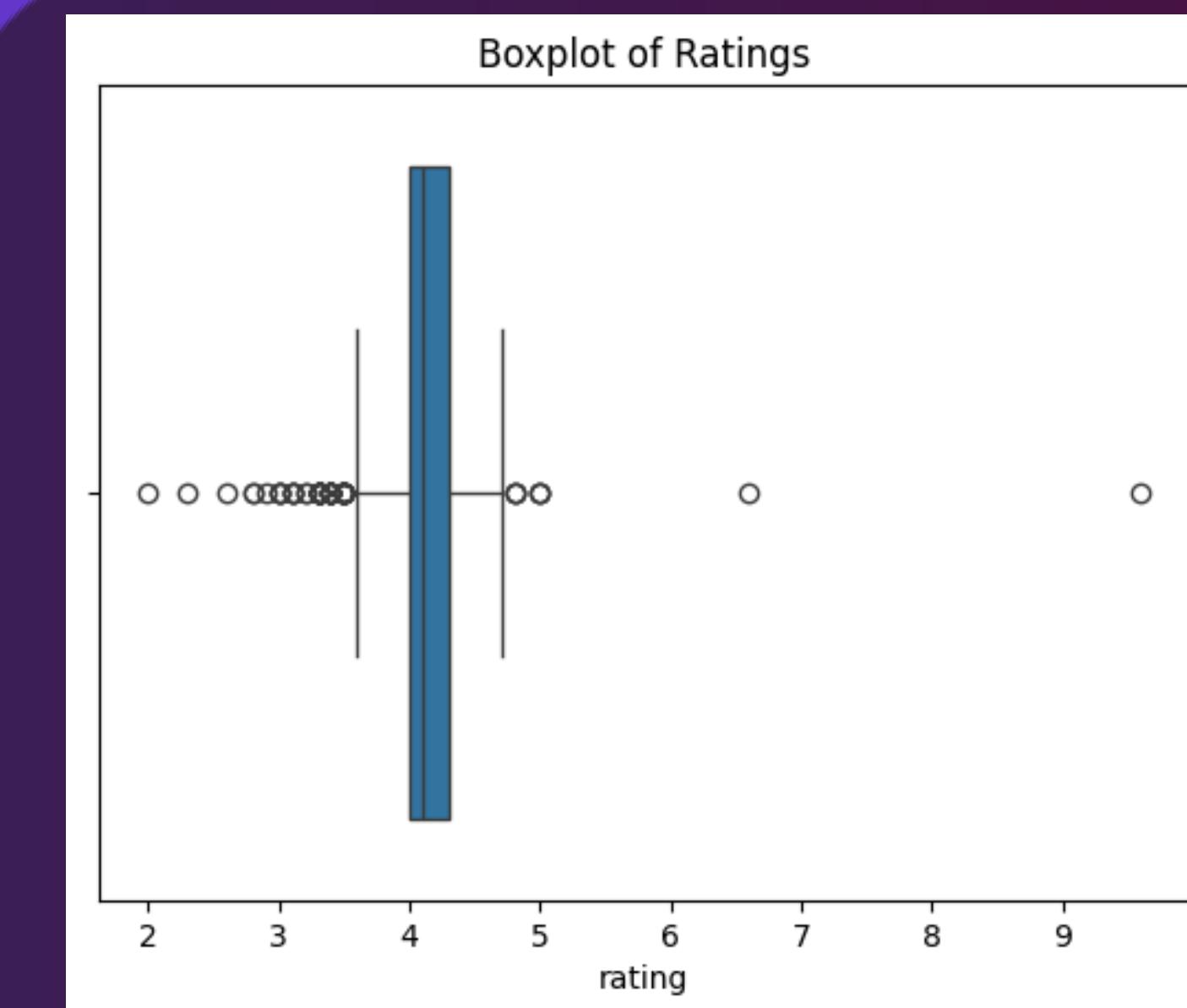
- Highly right-skewed distribution
- Large concentration of prices in the lower range
- Presence of many extreme high-price outliers



Boxplot of Actual Prices

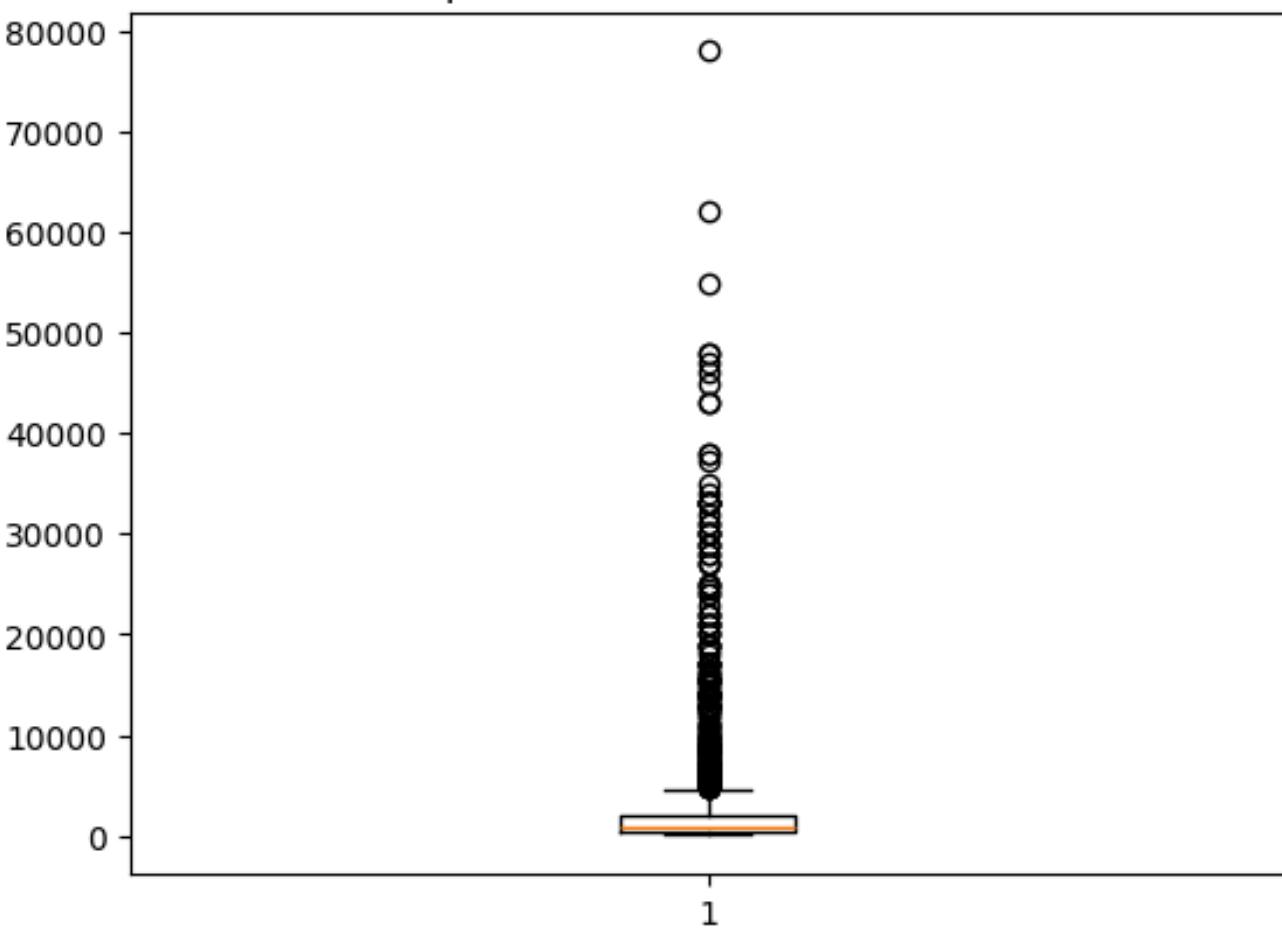


Boxplot of Ratings

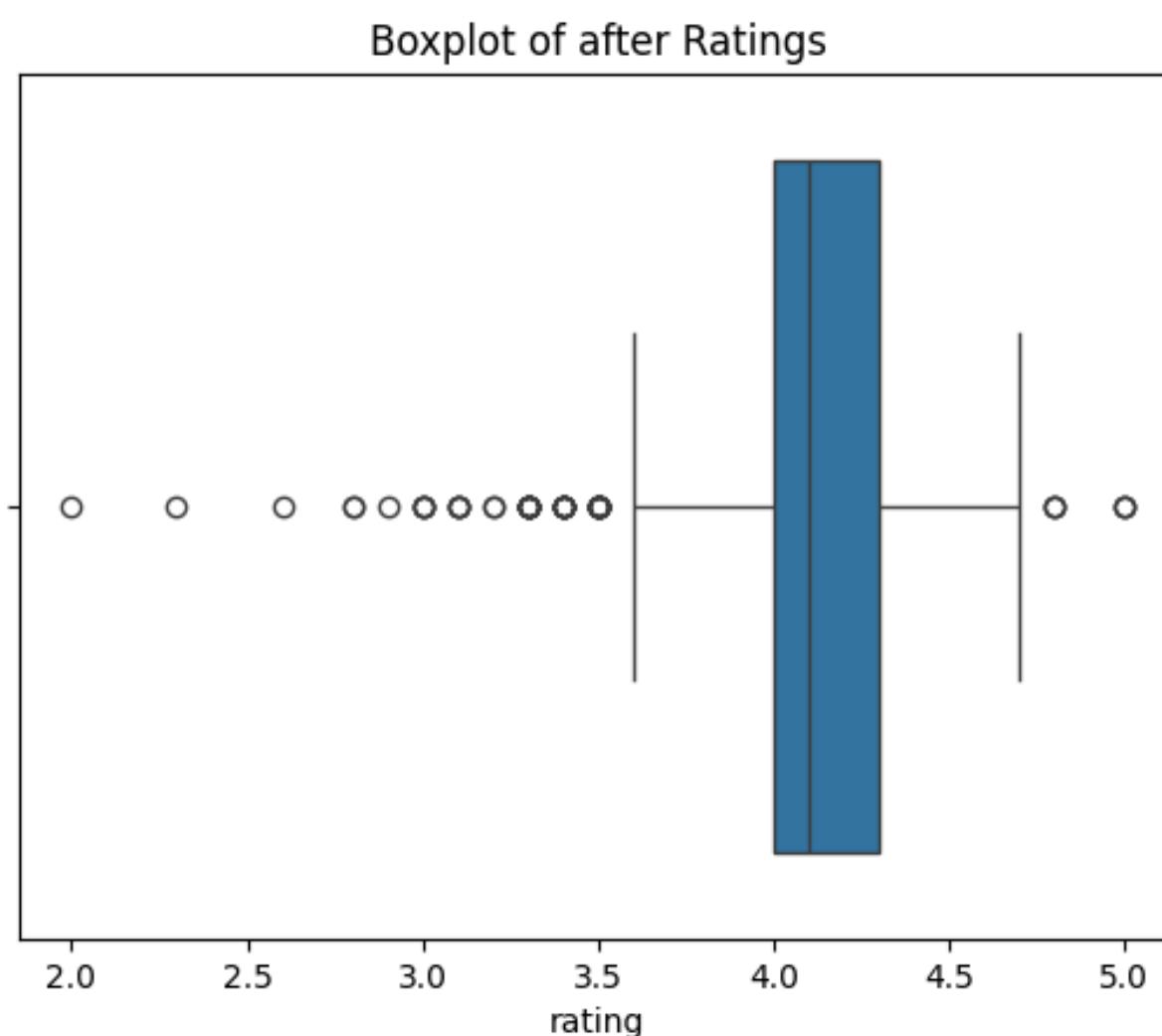


INSIGHTS ##  
THERE ARE RATING VALUES GREATER THAN 5, WHILE THE RATING SCALE IS -  
.FROM 0 TO 5

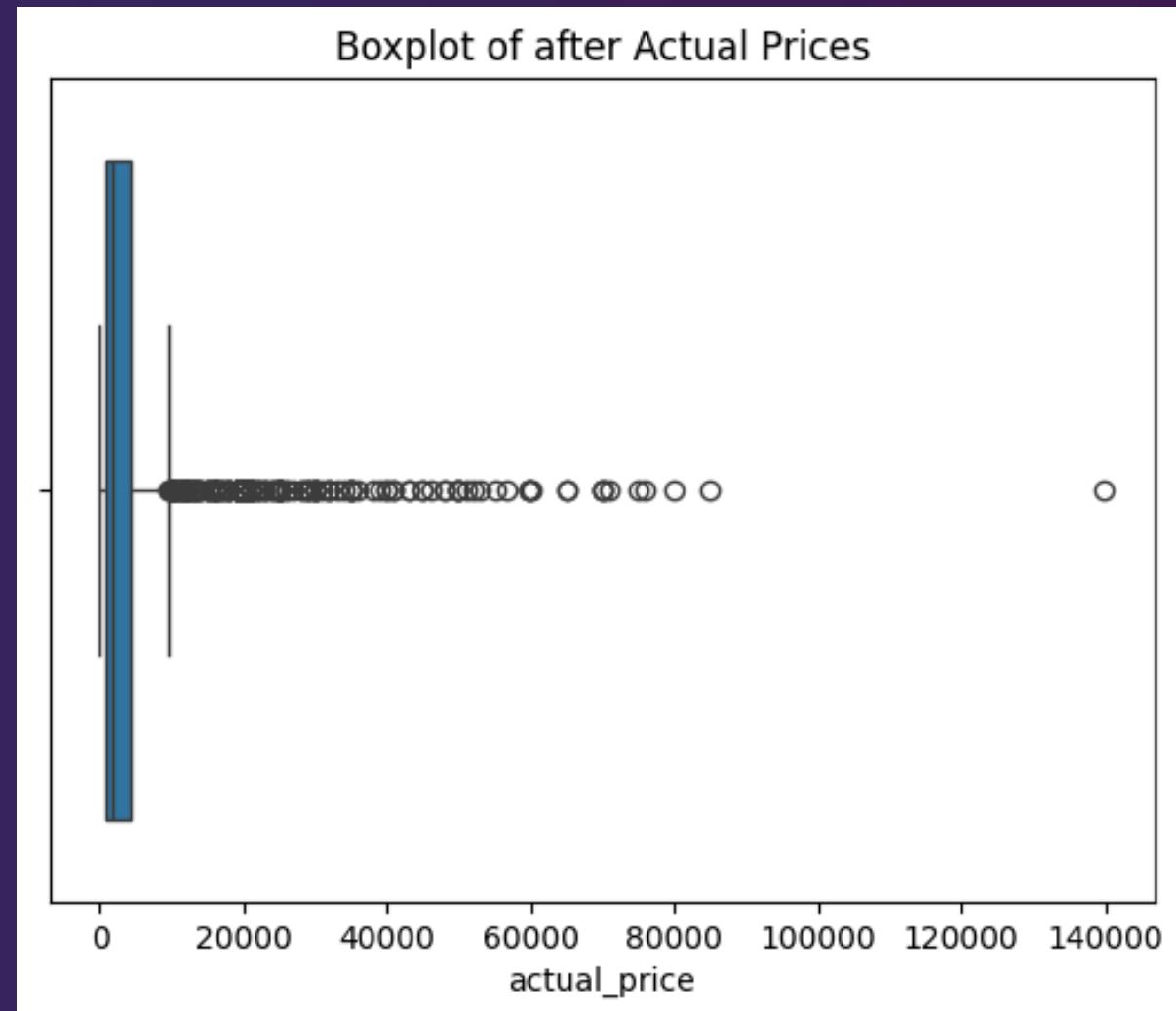
Boxplot of after Discounted Prices



Boxplot of after Ratings



Boxplot of after Actual Prices



# QUESTIONS

## WHICH CATEGORY PERFORMS BEST?

- \*\*  
عندما \*\*Electronics\*\* و \*\*Home&Kitchen\*\* فئة -  
أعلى عدد من المنتجات المباعة وعدد المراجعات  
المنتجات في هذه الفئات هي الأكثر \*\*Insight:\*\* -  
شعبية وربحًا، وتركز عليها الاستراتيجيات التسويقية

## ARE EXPENSIVE PRODUCTS RATED BETTER?

- \*\*  
لا يوجد ارتباط قوي بين السعر المرتفع والتقييم الأعلى -  
السعر وحده لا يضمن جودة التقييم، \*\*Insight:\*\* -  
المنتجات متوسطة السعر يمكن أن تحقق تقييمات  
ممتازة

## ?DO HIGHER DISCOUNTS INCREASE RATINGS

- \*  
الخصومات تجذب المشترين لكن التقييم \*\*Insight:\*\* -  
يعتمد على جودة المنتج، لا على السعر فقط

## WHICH FEATURES MAKE PRODUCTS MORE POPULAR

- لمنتجات ذات \*\*تقييمات عالية\*\* و \*\*عدد مراجعات كبير\*\* هي الأكثر مبيعاً

# THANK

# YOU!

ZAG-ENG

DATA ANALYSIS IS KEY TO BUSINESS  
GROWTH AND SUCCESS!

/ENG

AHMED N DIAB  
BASMALA

15

BY:  
AMANY IBRAHIM