

# Business and Sales Analysis

Created by Abhishek kumar



# Table of Content

Introduction

Data Collection

Data Cleaning And Preprocessing

Statistics

Ratings Analysis

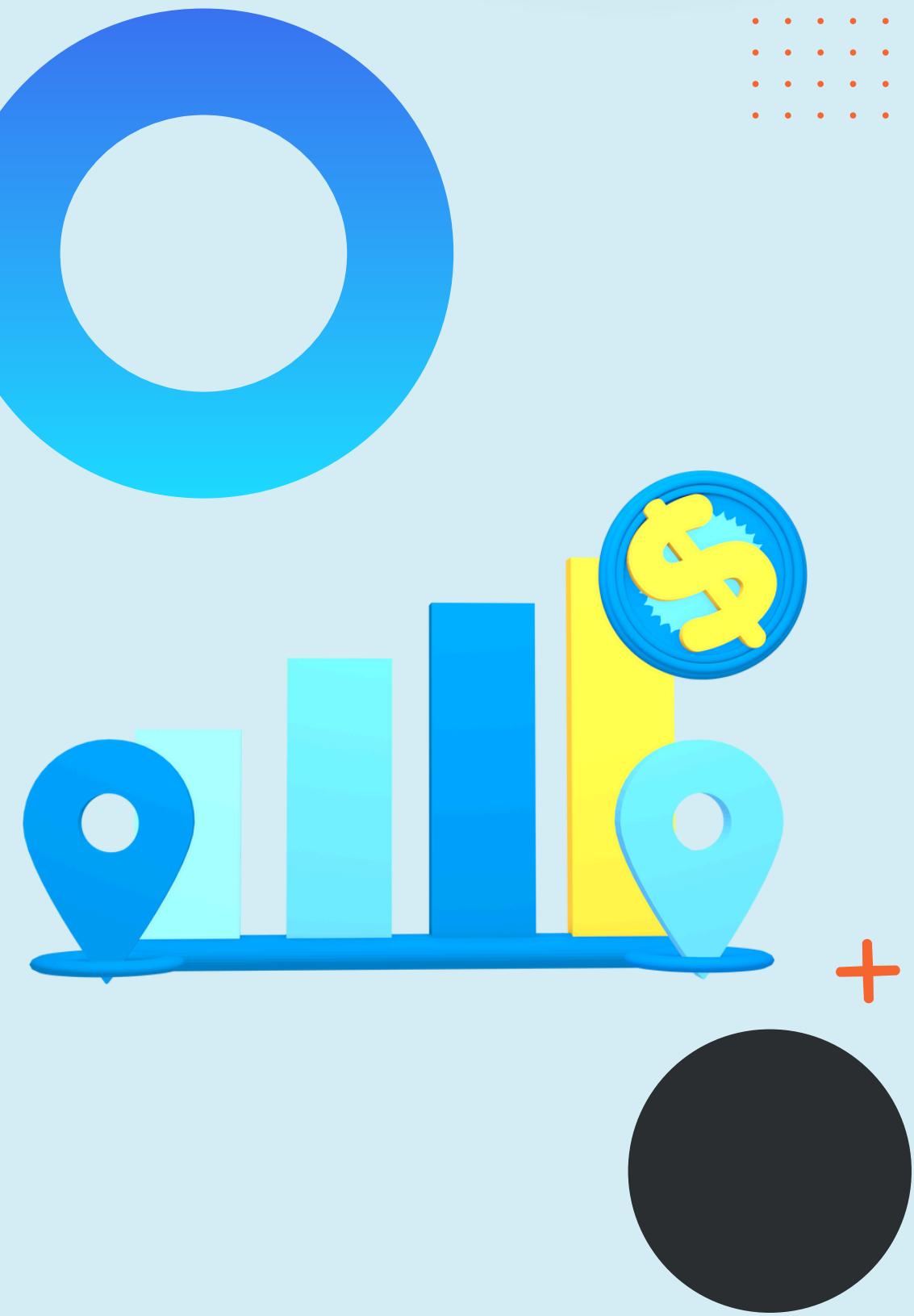
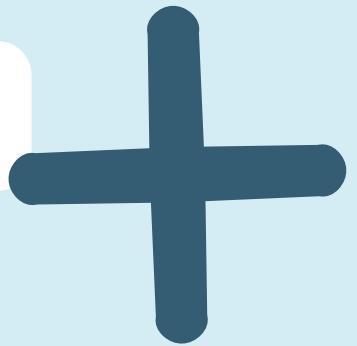
Discount Analysis

Revenue Analysis

Product Analysis

Correlation Analysis

Conclusion



# Project Overview



This project aims to analyze Amazon sales data from 2023 to uncover insights into discount strategies, product performance, and revenue generation. The analysis will help in understanding consumer behaviour and optimizing sales strategies.

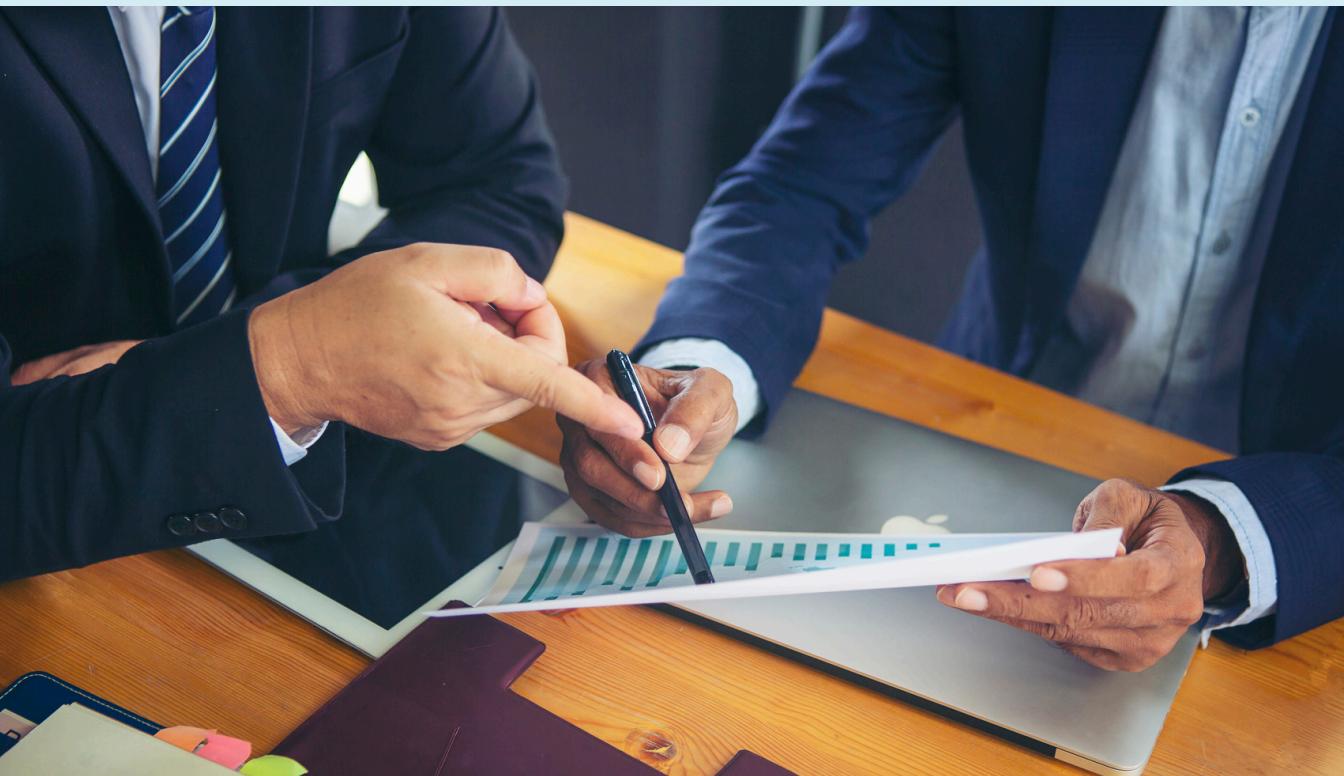


# Data Collection

## Data Source:

The data was collected from Kaggle. It is Amazon Products Sales Dataset 2023, having available product listings and sales data for the year 2023, consisting of over 550,000 rows.

[Click here to download dataset](#)



The dataset includes columns :

- *name*- The name of the product
- *main\_category*- The main category of the product belong
- *sub\_category* - The sub category to which product belong
- *image*- The image of the product look like
- *link*- The amazon website reference link of the product
- *ratings*- The ratings given by amazon customers to the product
- *no of ratings*- The number of ratings given to the product in amazon shopping
- *discount\_price*- The discount prices of the product
- *actual\_price*- The actual MRP of the product

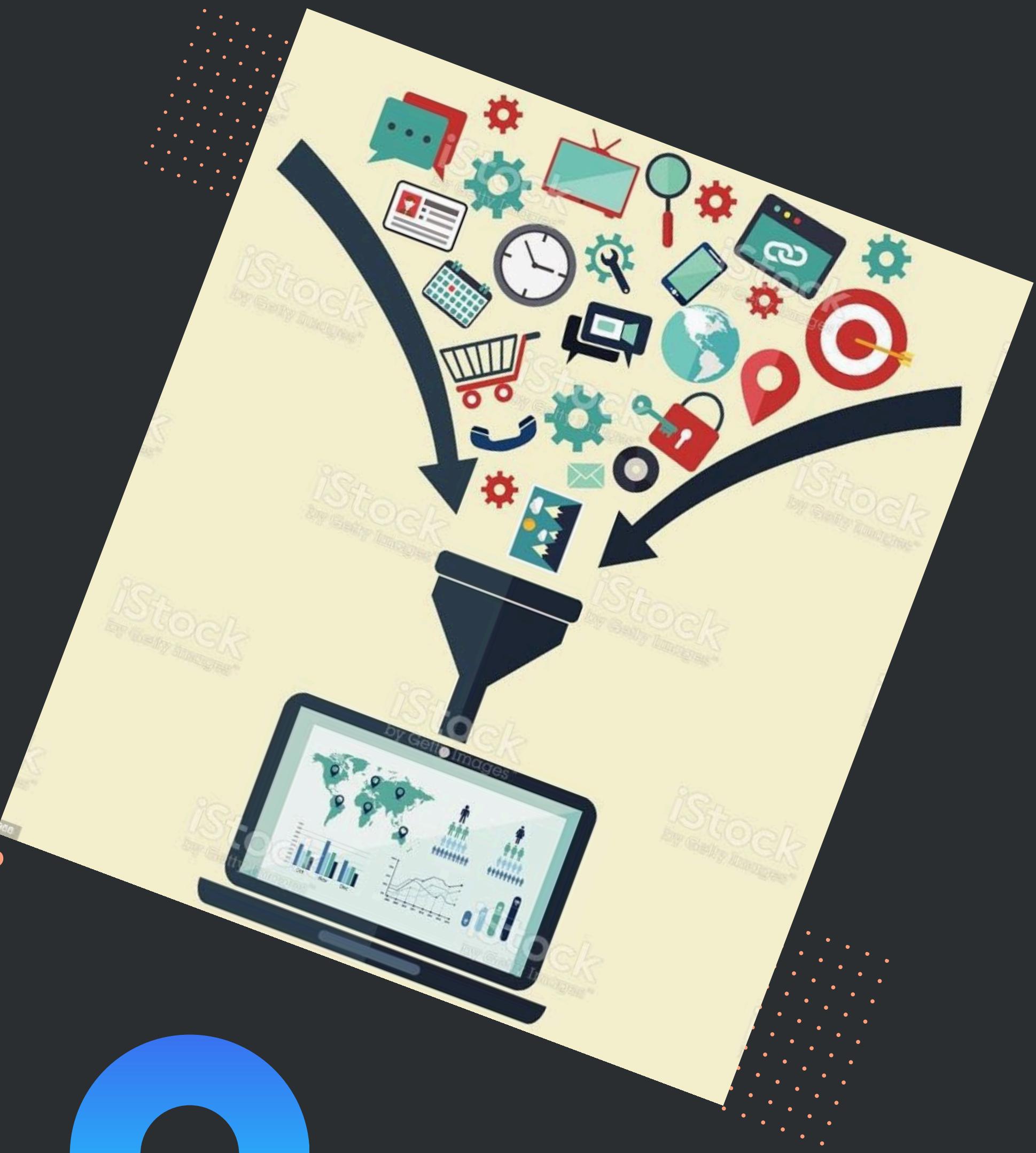
# Data Cleaning and Preprocessing

## Handling Null Values:

For numerical columns, we applied various imputation techniques and selected the one that resulted in the least change to the original statistics. This ensured minimal deviation in each column's statistical properties while effectively addressing missing values.

## Outlier Removal:

The IQR method was used to identify and remove outliers. Data points that fell below  $Q_1 - 1.5 \text{IQR}$  or above  $Q_3 + 1.5 \text{IQR}$  were considered outliers.



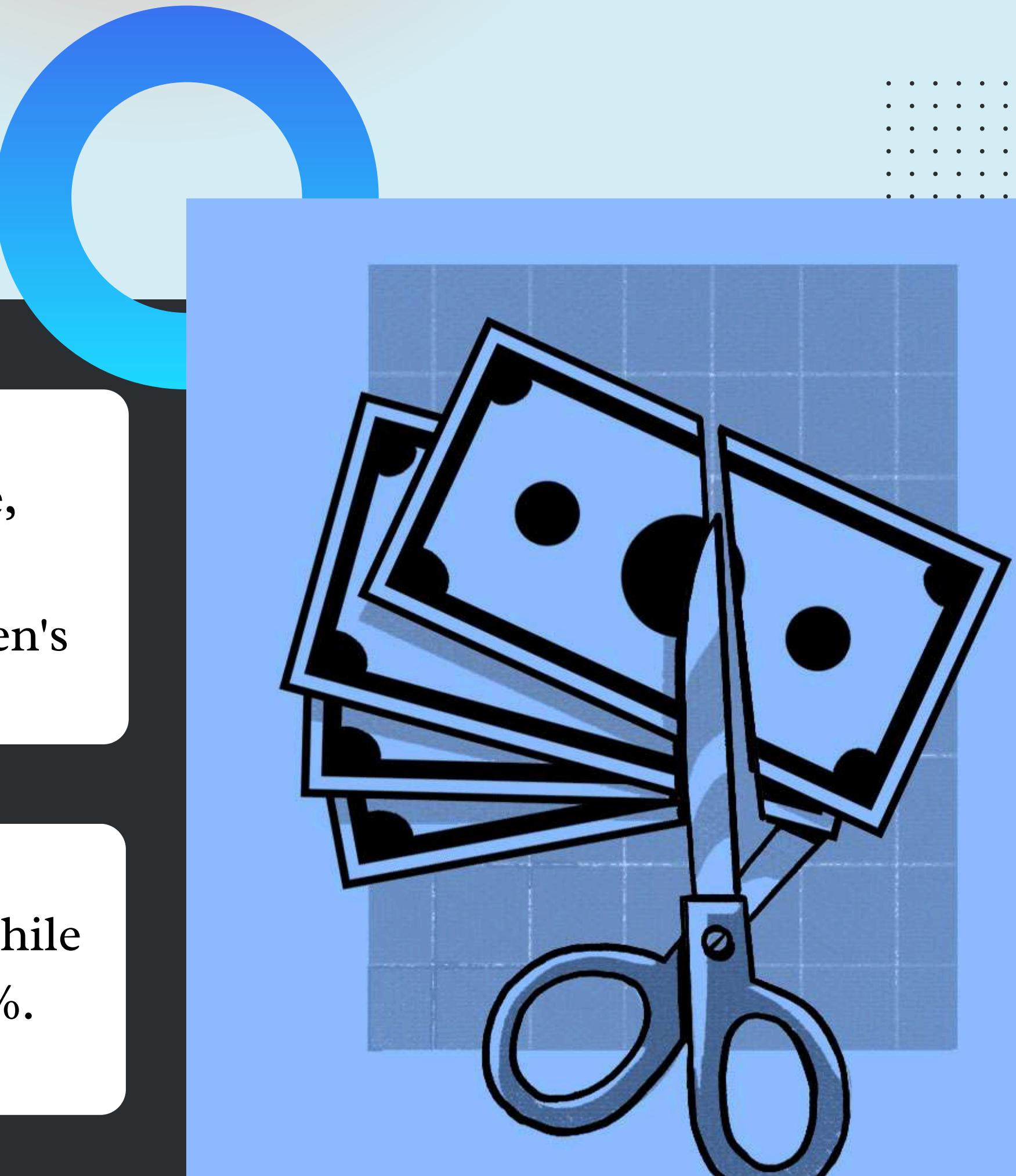
# Descriptive Statistics

	ratings	no_of_ratings	discount_price	actual_price
count	533772.00	533772.00	533772.00	533772
mean	3.88	646.44	2588.71	23111.28
std	0.76	7306.82	9209.74	13550820
min	2.15	1.00	8.00	0
25%	3.50	2.00	396.00	990
50%	4.00	12.00	693.00	1599
75%	4.40	92.00	1449.00	2999
max	5.00	589547.00	1249990.00	9900000000

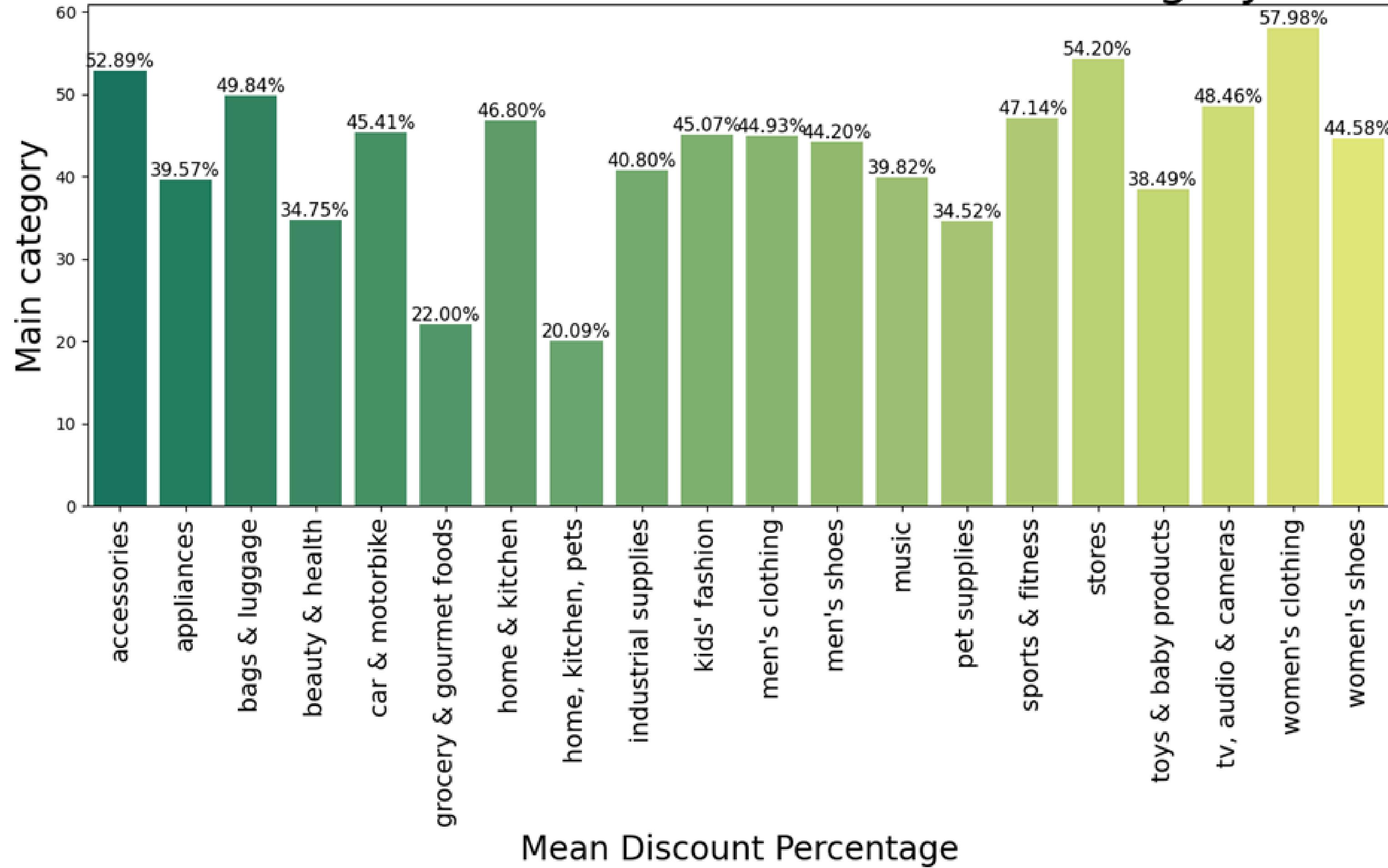
# Discount Analysis

- The overall average discount is 48.77%.
- Minimum discount of 20.09% is given to "home, kitchen pets" category.
- Maximum discount of 57.98% is given to "women's clothing"

Appliances have an average discount of 39.57 %, while Kids Fashion has an average discount of 45.07 %.



# Discount distribution across Main Category



# Product Analysis

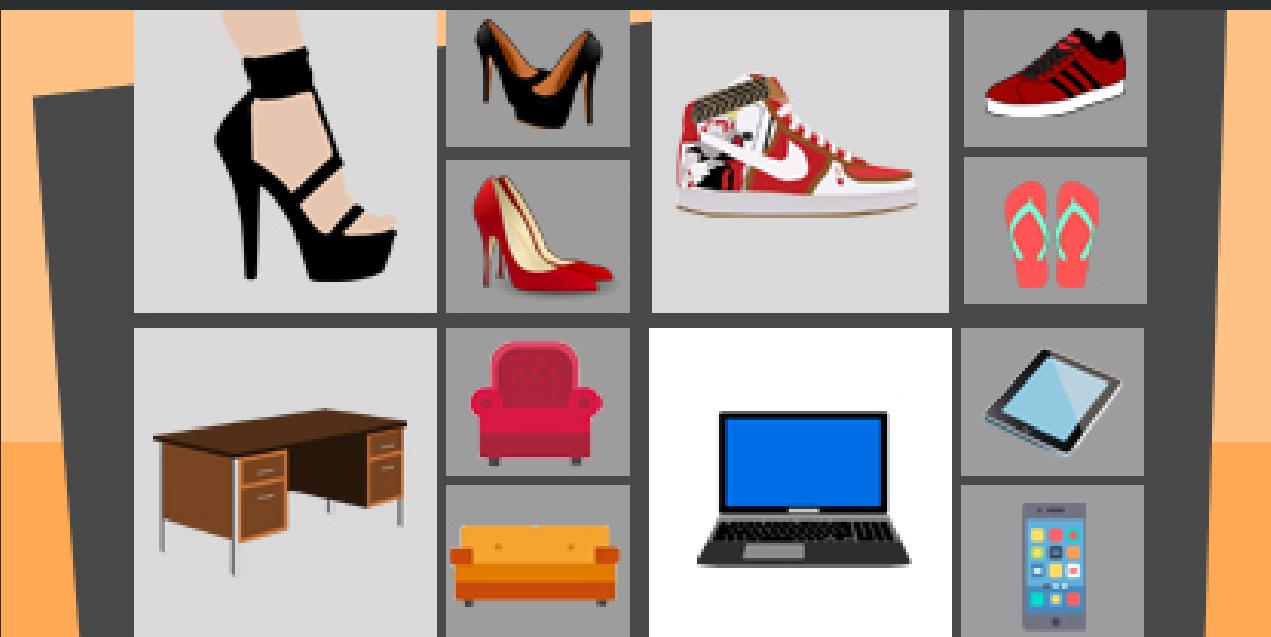


## TOP 5 CATEGORIES WITH HIGHEST NUMBER OF PRODUCTS

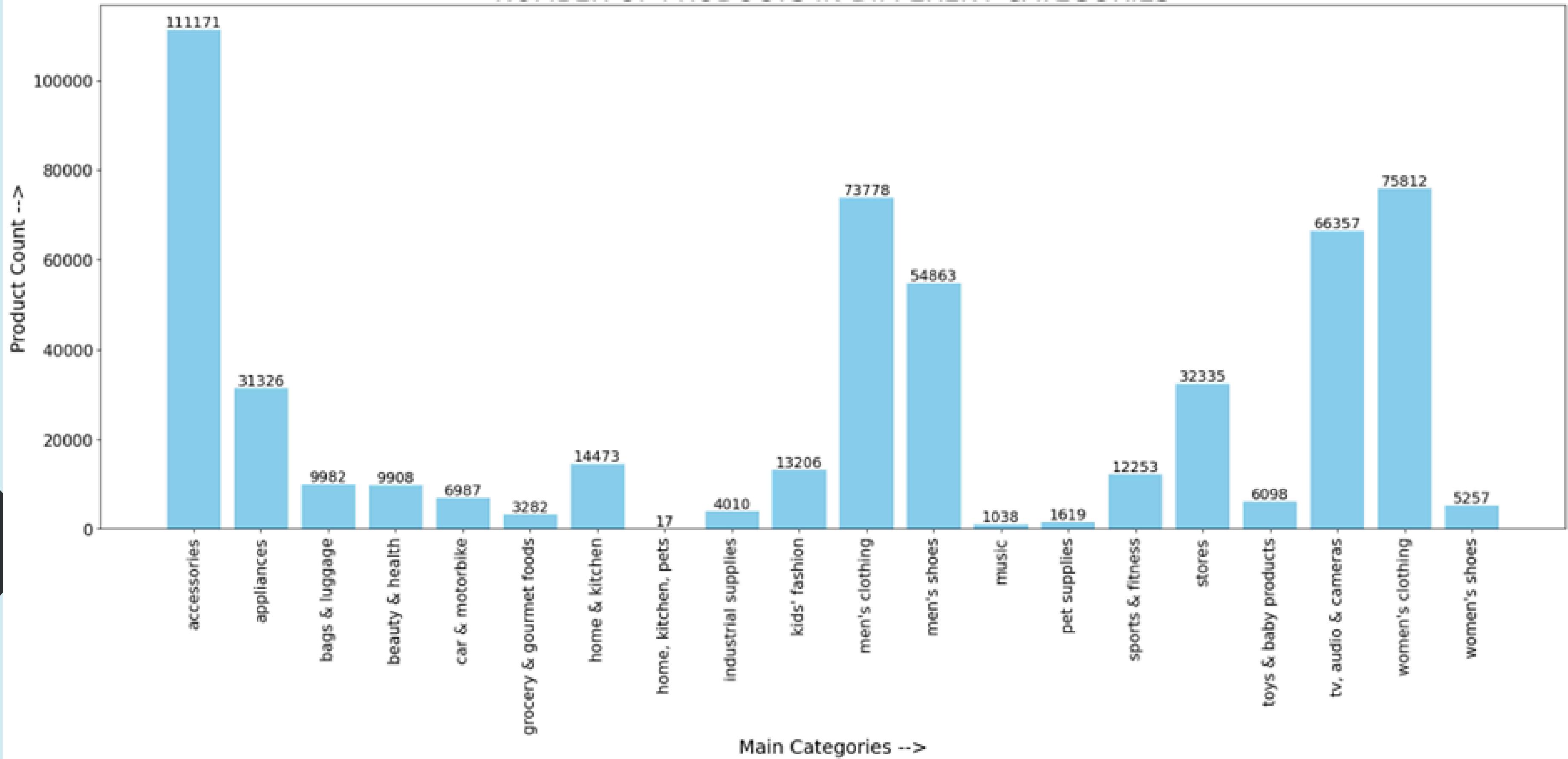
1. **Accessories** - 1111171
2. **Women's Clothing** - 75812
3. **Men's Clothing** - 73778
4. **tv, audio & cameras** - 66357
5. **men's shoes** - 54863

## TOP PRODUCTS WITH HIGHEST NUMBER OF RATINGS

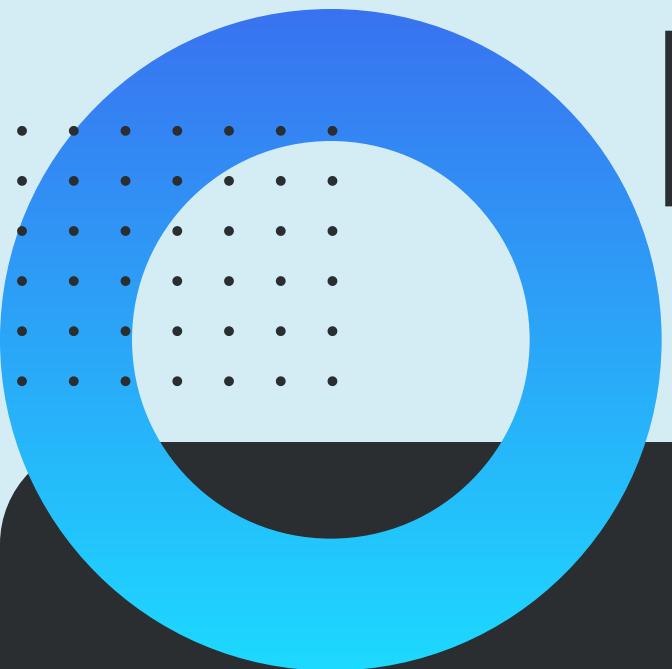
- “Amazon Basics High-Speed HDMI Cable“ of four different variety occupies the top position.
- “Amazon Basics High-Speed HDMI Cable 10 ft- 3 ft” occupies second position
- “JBL Go 3, Wireless Ultra Portable Bluetooth Speaker” holds 3rd rank.



## NUMBER OF PRODUCTS IN DIFFERENT CATEGORIES



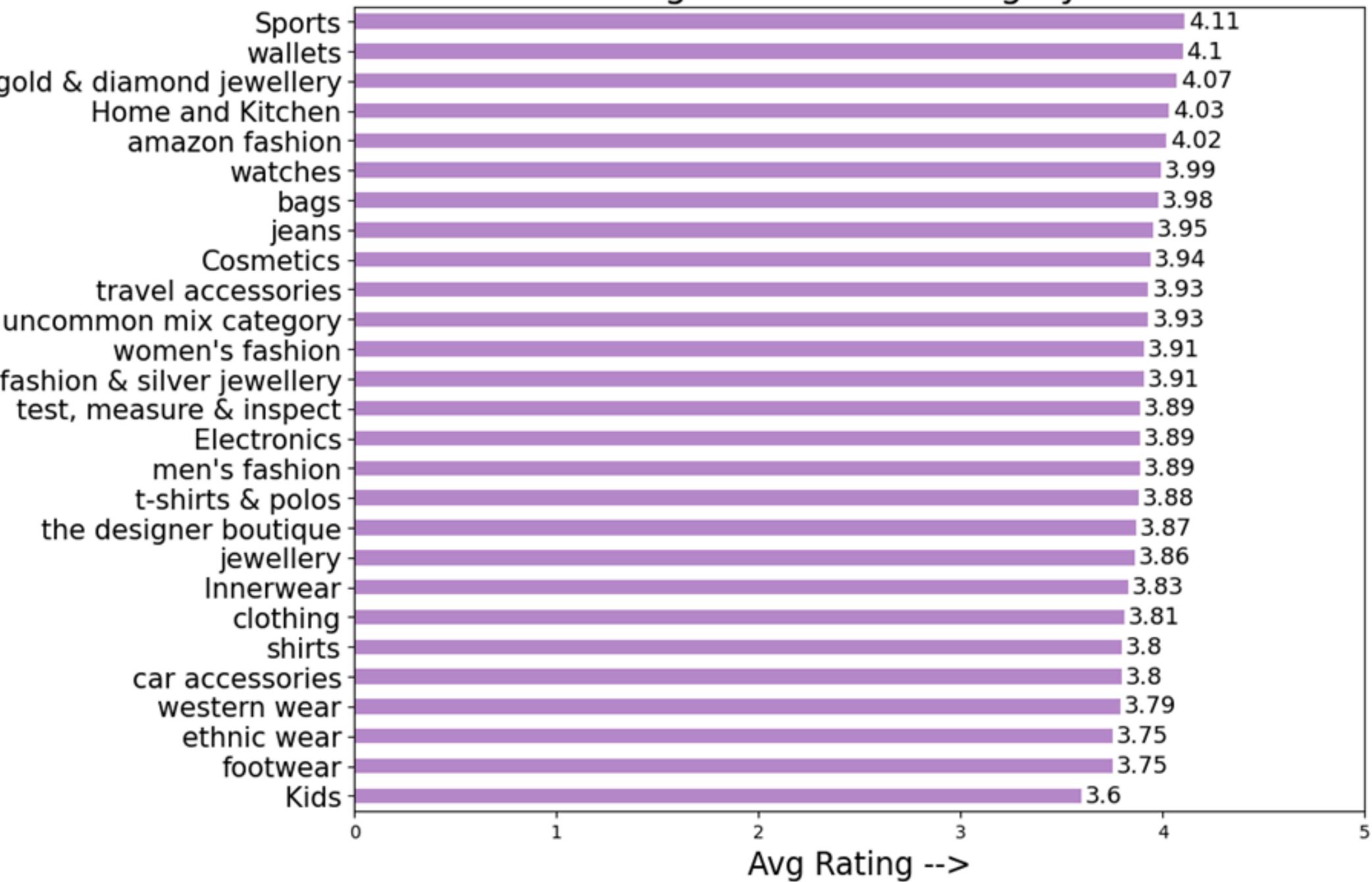
# Ratings Analysis



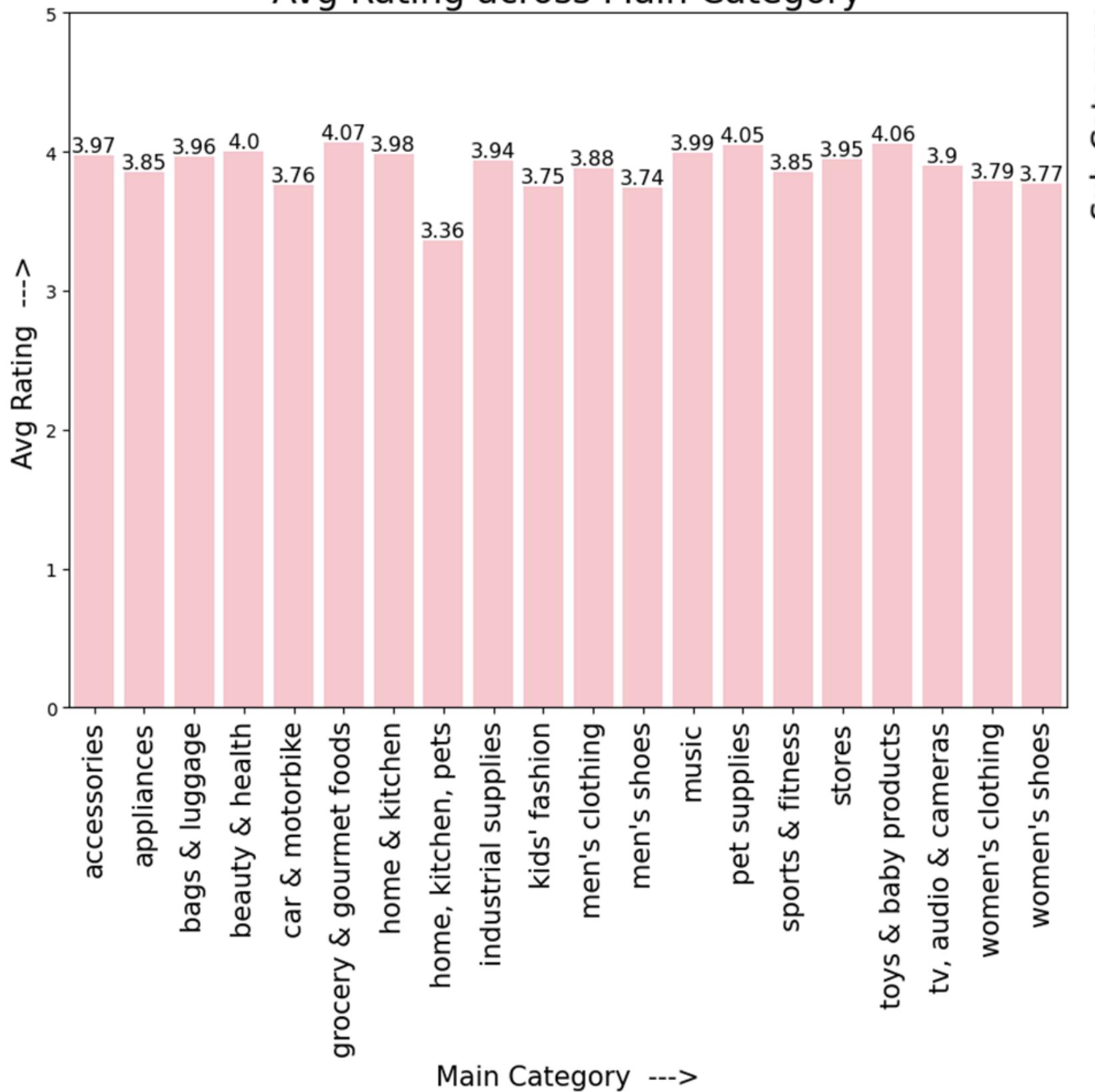
- Overall avg rating of 3.88 signifies, customers are generally satisfied with their purchases, but there is room for improvement.
- Minimum Avg Rating is 3.36 in “home, kitchen, pets” category.
- Maximum Avg Rating is 4.07 in “grocery & gourmet foods” category.

- “Sports” subcategory has got higher rating of 4.11.
- “Kids” subcategory has got least rating of 3.6.
- “Women’s shoes” has got 3.77 whereas “Men’s shoes” got 3.74 avg rating

## Ratings Across SubCategory -->



## Avg Rating across Main Category



## Sub Category



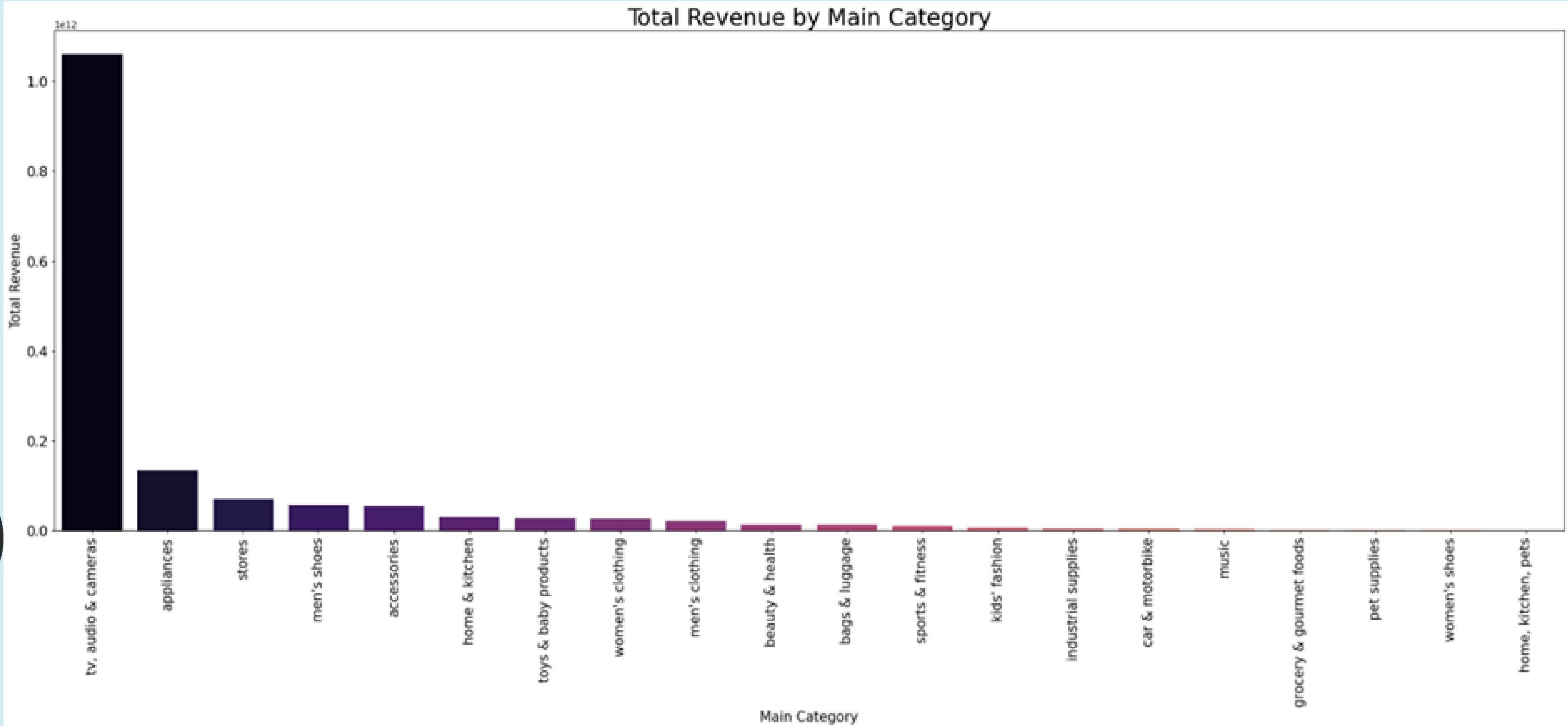
# Revenue Analysis



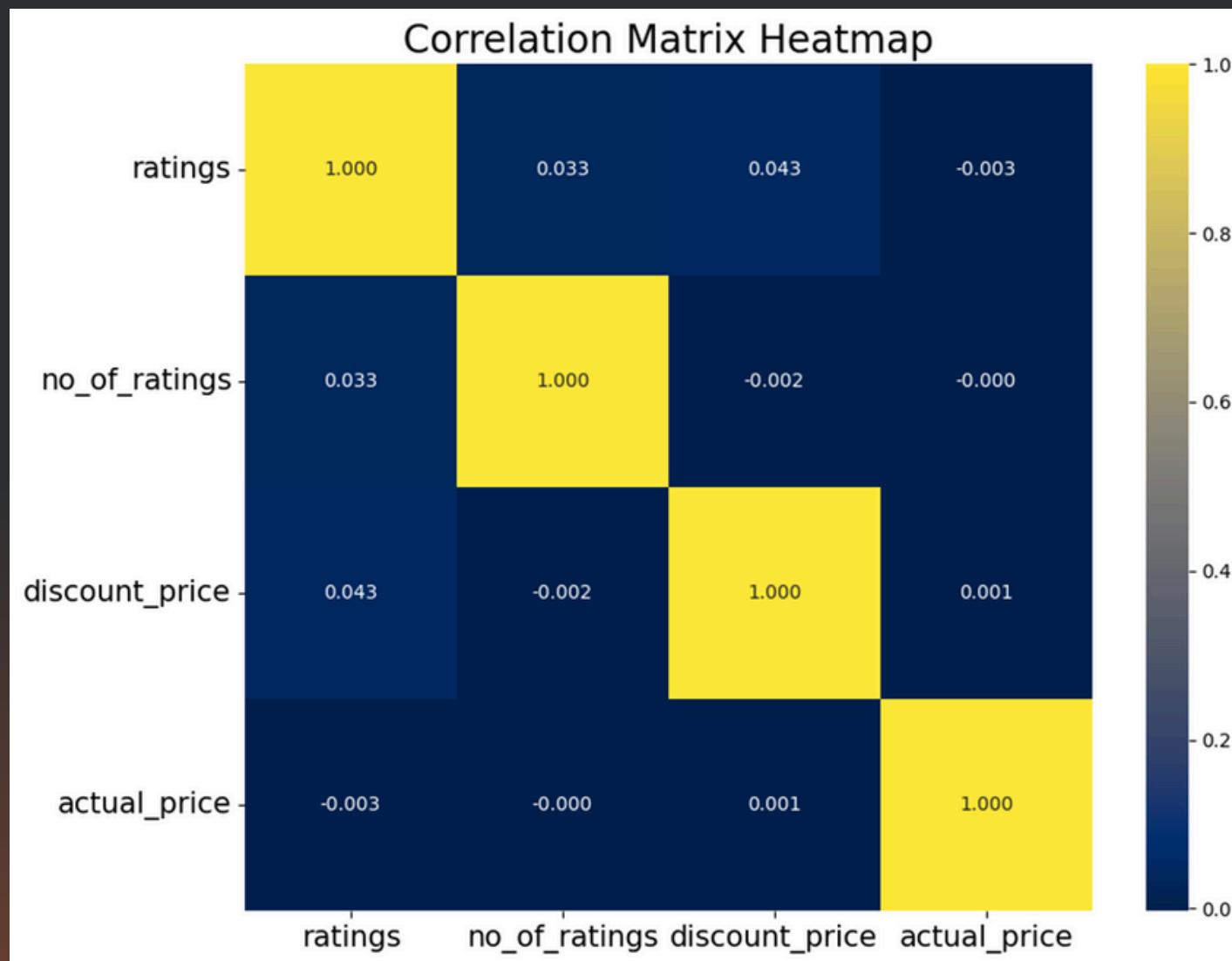
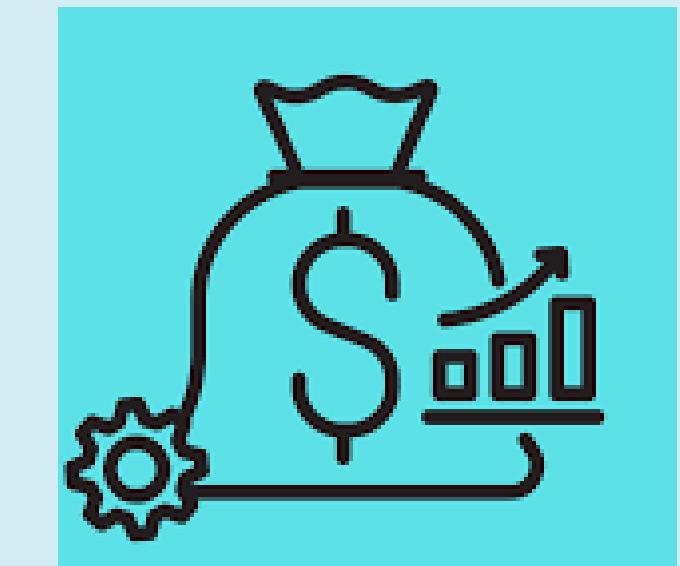
- 
1. It is “tv, audio & cameras” that generated maximum revenue,  $1.061160 \times 10^{12}$
  2. “home, Kitchen and Pets” contributed least revenue to business, may be due to minimal product variety and less discount.

Sandisk 400GB, Marutvilla Insect Killer, Samsung Evo Plus , redmi 139 are some of the products that contributed max to revenue.

## Total Revenue by Main Category



# Correlation Analysis



In a correlation matrix, light coloured regions represent strong correlation while dark regions represent minimal correlation.

Values are very close to zero thus, there is almost no relationship amongst the features.

# Conclusion

- **Key Findings:** Our comprehensive analysis of Amazon product listings provided valuable insights into various aspects, despite the near-zero correlation among features.
- **Products Analysis:** The analysis showed that product categories and specific products play a significant role in consumer preferences. Identifying top products and categories can help optimize inventory and marketing strategies.
- **Revenue Analysis:** We identified the top revenue-generating products and categories, which are crucial for targeted marketing and stock management. This analysis underscores the importance of focusing on high-performing items to maximize profits.
- **Discount Analysis:** Discounts showed varied impact on different products and categories. Understanding which products benefit most from discounts can aid in designing effective promotional strategies that drive sales without significantly eroding margins.

# Conclusion

- **Ratings Analysis:** Although the correlation between ratings, number of ratings, and pricing was minimal, the distribution analysis provided insights into common rating trends and customer satisfaction levels across different product categories. This information is vital for quality control and improving customer satisfaction.
- **Diverse Factors:** The lack of strong correlation suggests that factors beyond pricing and ratings, such as product descriptions, reviews, marketing strategies, and customer demographics, significantly influence consumer behavior and product performance.
- **Further Exploration:** The insights from this analysis highlight the need for a multifaceted approach in understanding product performance, encouraging further exploration into additional variables to fully grasp the drivers of sales and customer engagement.

**Thank You For Your  
Kind Attention and  
Valuable Time!!**

