amazon

Amazon Sales Analysis

DETAILED PROJECT REPORT

Shubham Tembhurne

PROJECT DETAIL

| Project Title	Amazon Sales Analysis
Technology	Data Analysis
Domain	E-commerce
Project Difficulty level	Advance
Programming Language Used	Python
Tools Used	Jupyter Notebook, Ms-Excel, Tableau

OBJECTIVE

- The goal of this project is to analyze the Amazon

Sales occurrence, based on a combination of features that describes product category, product name, rating, rating count, actual price and discounted price.

PROBLEM STATEMENT

- This dataset consists more than 1000 of real products with their identification number listed in the Amazon marketplace specifically from the region India. I noticed the region due to the currency used in the dataset is Rupee India. My objective is to clean and prepare the data due to the raw data being very unorganized. I will then move on to finding insights about the data and try to elaborate in the form of visualization.

Architecture
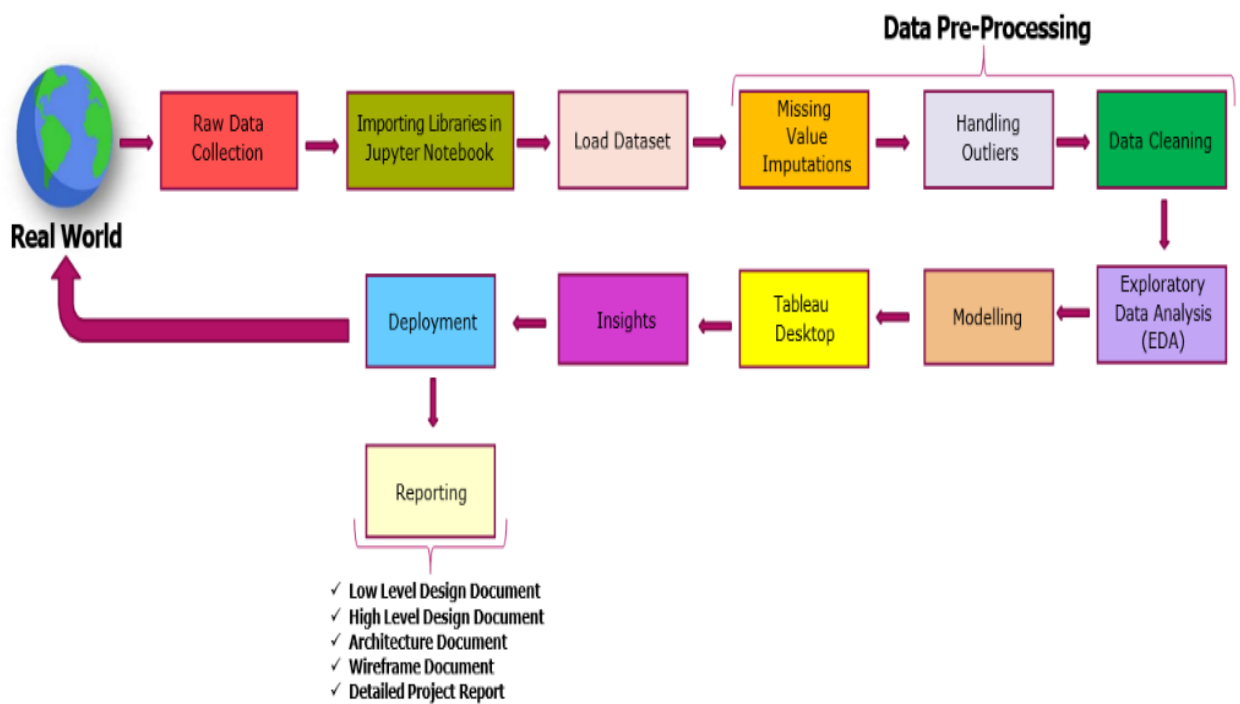
DATASET INFORMATION

- **PRODUCT_ID:** A product ID is a unique identifier assigned to a specific product or item for the purpose of tracking and managing sales transactions. It helps to distinguish one product from another and enables efficient inventory management, order processing, and sales analysis.

- **PRODUCT_NAME:** A product name refers to the specific name or designation given to a product or item being sold. It is used to identify and describe the product to customers, sales representatives, and other stakeholders involved in the sales process. The product name plays a crucial role in marketing, branding, and communicating the features and benefits of the product to potential buyers

- **CATEGORY:** A category refers to a grouping or classification system used to categorize products based on shared characteristics, attributes, or functions. Categories help organize and manage inventory, facilitate browsing and search for customers, and provide insights into sales performance by product type. Common examples of sales categories include electronics, clothing, home appliances, beauty products, and automotive. Categorizing products allows for better organization, targeted marketing, and more effective sales analysis.

- **DISCOUNTED_PRICE:** A discounted price refers to a reduced price offered to customers as an incentive or promotion. It is typically lower than the original or regular price of a product. Discounts can be in the form of a percentage off, a fixed amount off, or a combination of both. The discounted price is communicated to customers to encourage them to make a purchase and create a sense of value or savings. Discounted prices can be temporary, such as during a sale or promotion, or they can be available for specific customer segments, such as loyal customers or bulk buyers.

- **ACTUAL_PRICE:** the actual price refers to the original or regular price of a product before any discounts or promotional offers are applied. It represents the standard price at which the product is typically sold. The actual price is often used as a reference point to calculate discounts or determine the savings a customer can receive through promotional pricing. It is important for customers to know the actual price to understand the value they are receiving when comparing it to discounted prices or competing offers.

- **DISCOUNT_PERCENTAGE:** The discount percentage in a sale refers to the percentage by which the original price of a product is reduced to calculate the discounted price.

- **RATING:** Ratings refer to the numerical or qualitative assessments provided by customers or experts to evaluate the quality, performance, or satisfaction level of a product or service.

- **RATING_COUNT:** Rating count in sales refers to the total number of ratings or reviews received for a product or service

- **DIFFERENCE_PRICE:** The difference price refers to the variance or discrepancy between two prices.

INSIGHTS

Product Category

- Below are the list of Main Category and Sub-Category to help determine which sub-category belongs to which main category:

| Main Category	Sub-Category	Producd ID
Car & Motorbike	CarAccessories	1
Computers & Accessories	Accessories&Peripherals	381
	Components	5
	ExternalDevices&DataStorage	18
	Laptops	1
	Monitors	2
	NetworkingDevices	34
	Printers,Inks&Accessories	11
	Tablets	1
Electronics	Accessories	14
	Cameras&Photography	16
	GeneralPurposeBatteries&BatteryChargers	14
	Headphones,Earbuds&Accessories	66
	HomeAudio	16
	HomeTheater,TV&Video	162
	Mobiles&Accessories	161
	PowerAccessories	1
	WearableTechnology	76
Health & PersonalCare	HomeMedicalSupplies&Equipment	1
Home & Kitchen	CraftMaterials	7
	Heating,Cooling&AirQuality	116
	HomeStorage&Organization	16
	Kitchen&Dining	1
	Kitchen&HomeAppliances	308
Home Improvement	Electrical	2
Musical Instruments	Microphones	2
Office Products	OfficeElectronics	4
	OfficePaperProducts	27
Toys & Games	Arts&Crafts	1

- ## Most Amount of Product by Category:
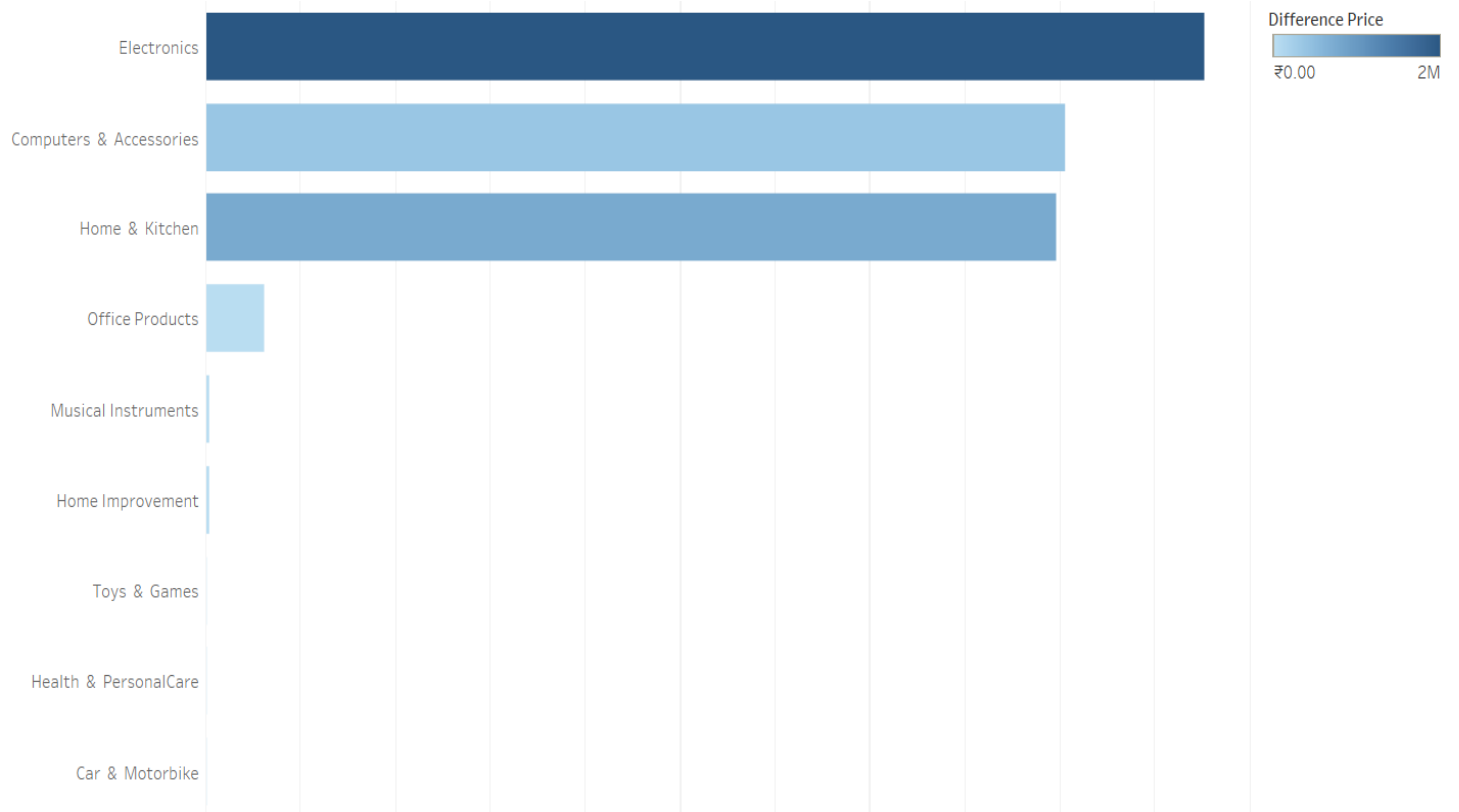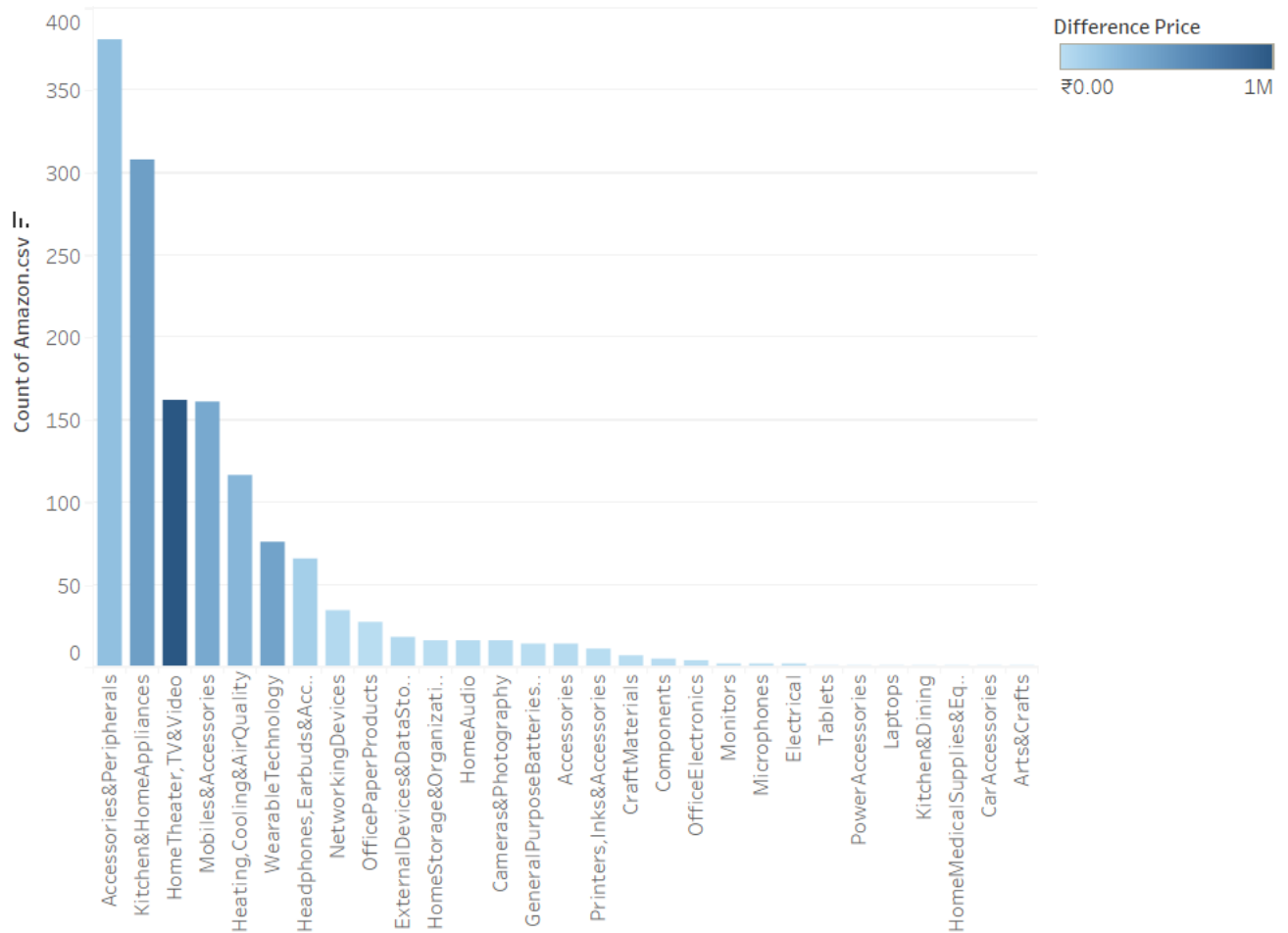
Electronics especially accessories & peripherals and Kitchen & home appliance contain most of the products in this data set. In general, most products are related to the electric devices in this dataset.

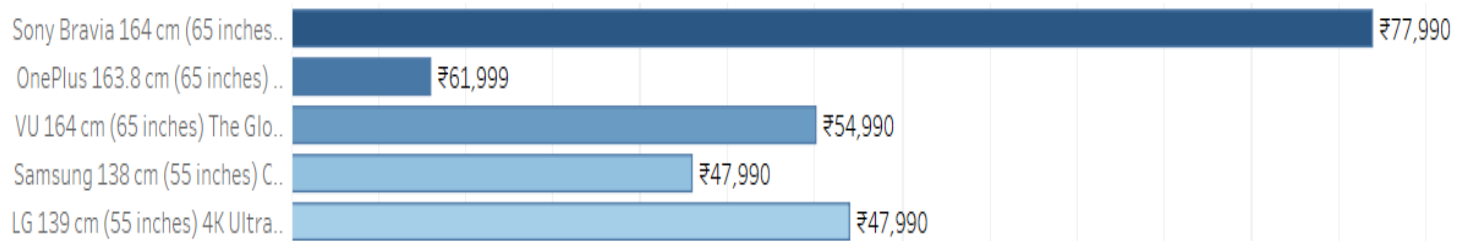- Top 5 Most Expensive Products After Discount:
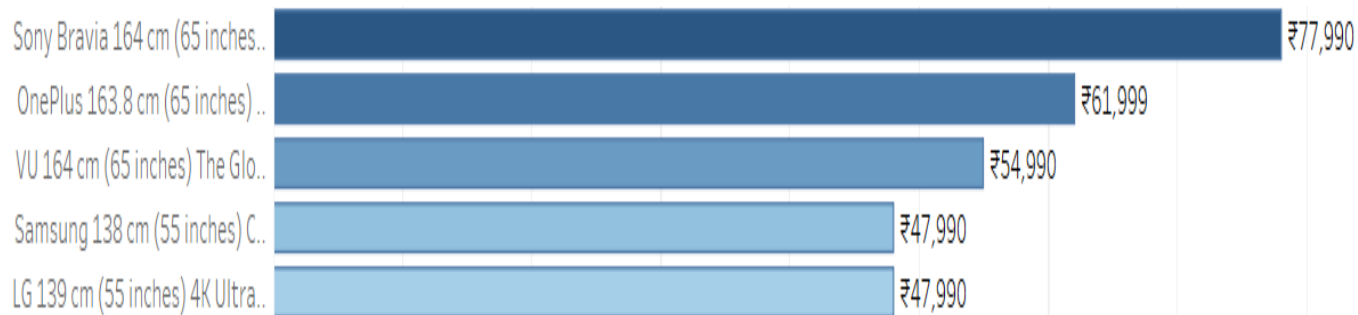
Sony Bravia 164 cm (65 inches) is the most expensive product after discount.

- Top 5 Cheapest Products after Discount:
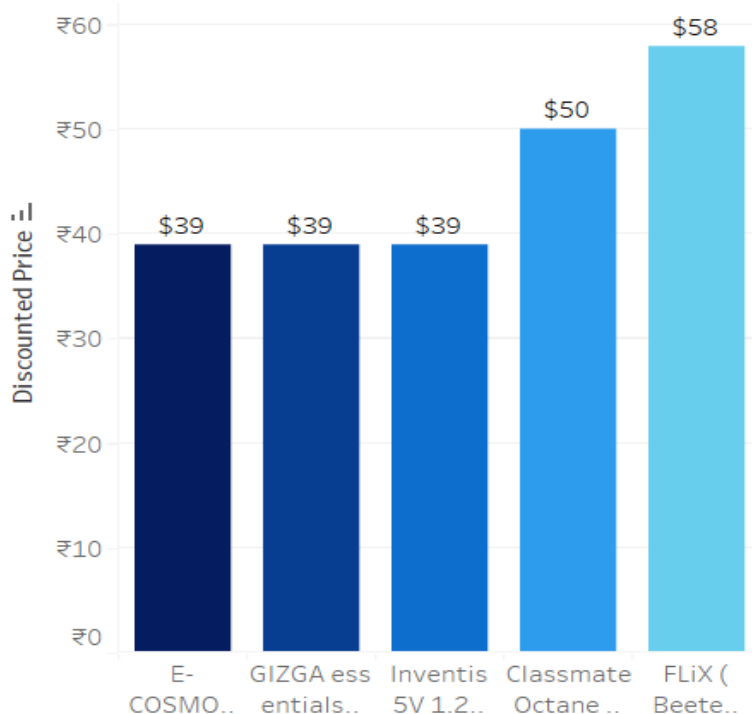
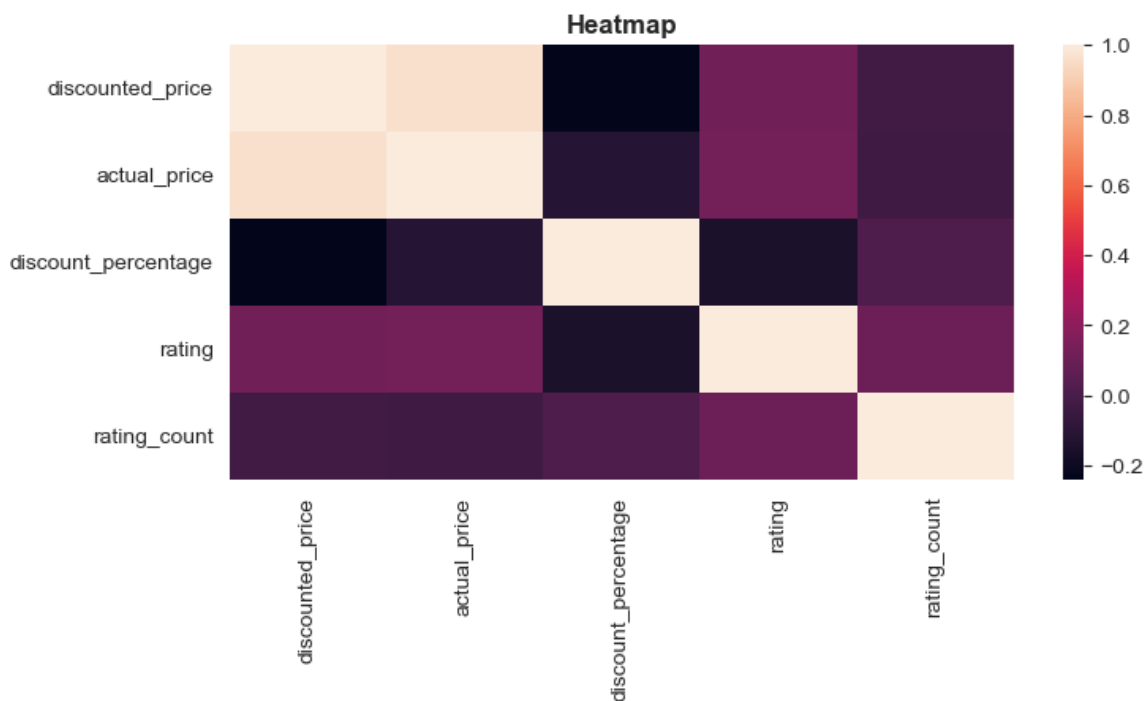**E-cosmos 5V 1.2W Portable Flexible is the cheapest product after discount.**

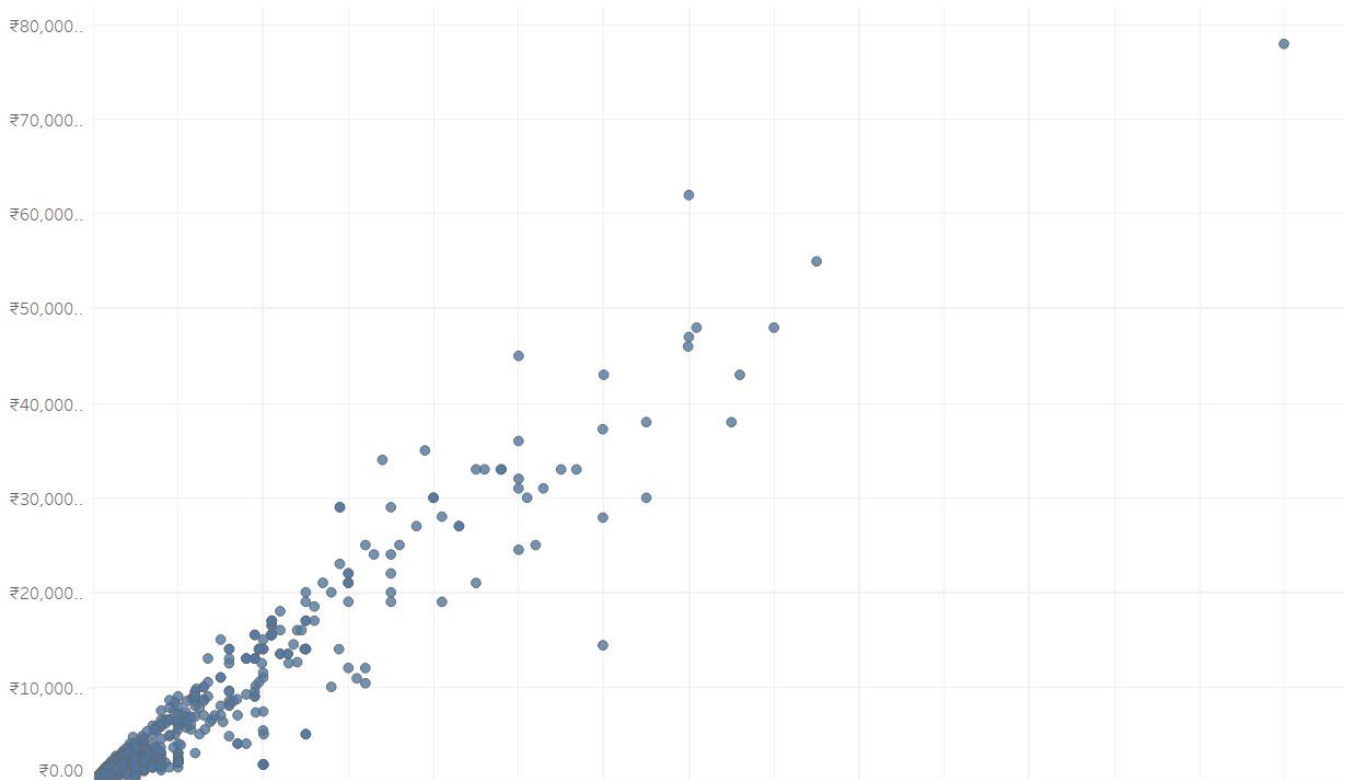- Price difference due to discount in Product :

Sony Bravia 164cm having the largest price difference due to discount.
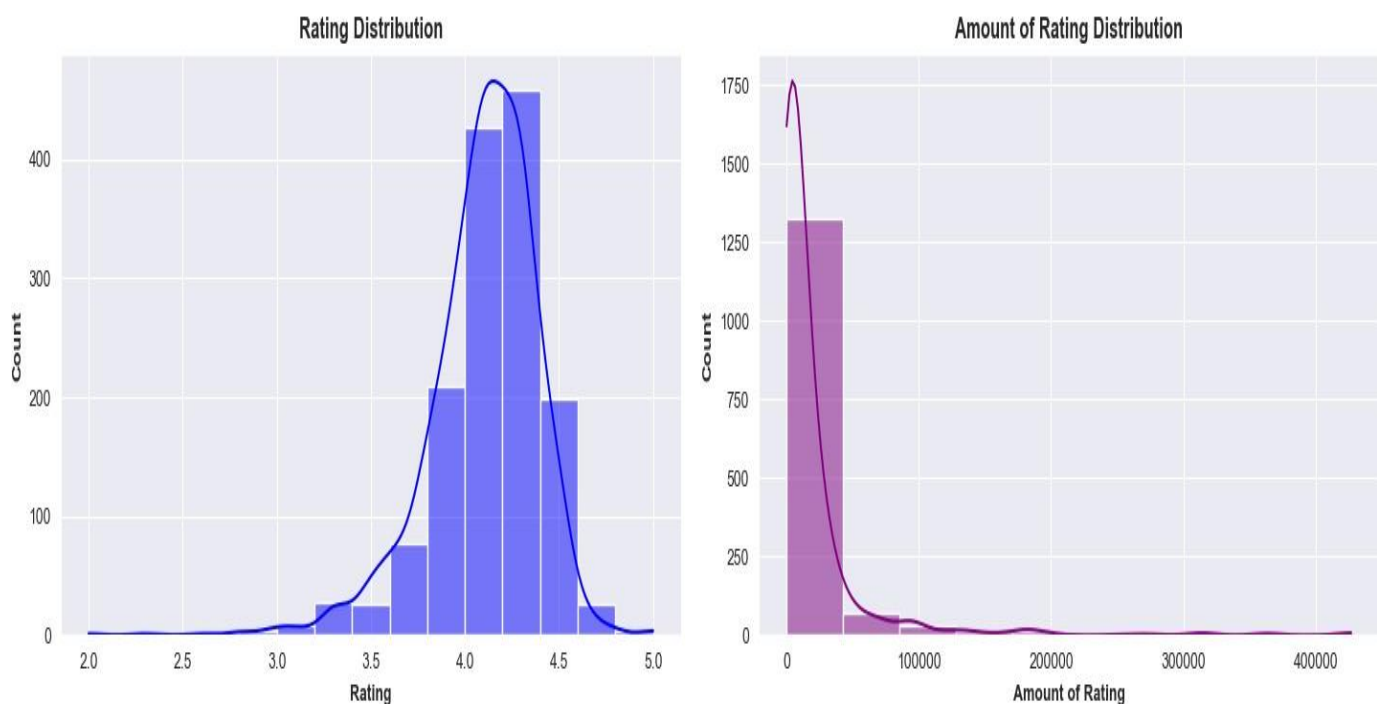
- Correlation between features:

There is almost no correlation between the dataset but there is positive correlation betweenthe discounted price of product and actual price of product.
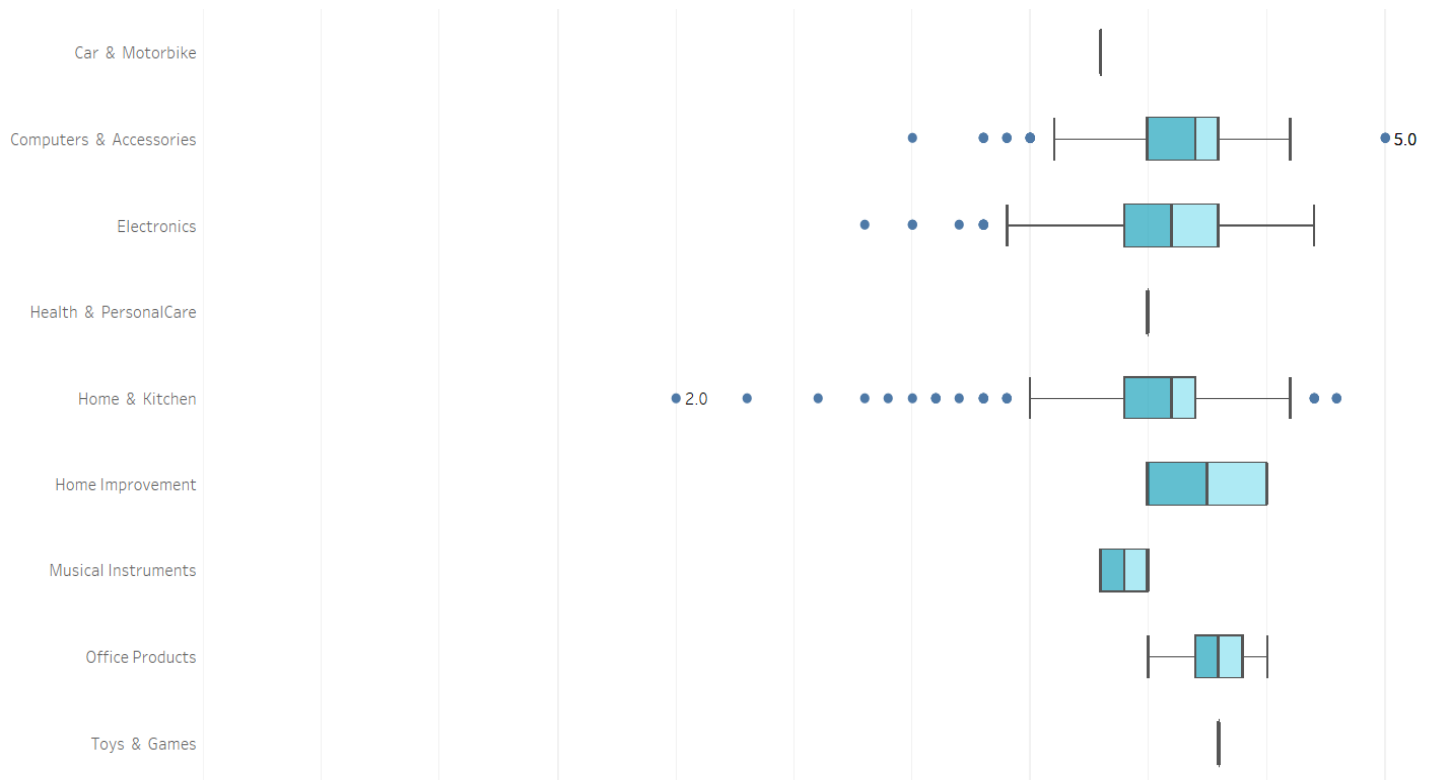
Product Rating

- Rating Amount and of Rating Distribution

Rating and Amount of Rating Distribution

Most of the product range around 4.0 to 4.37 with no products under the score of 2.0. The Rating Distribution is Slightly Left-Skewed. The amount of ratings given to a product is very widespread. Most of the products that have been rated, have around 0 - 5000 amount of ratingfor each product. Interestingly there are products that have more than 40,000 ratings. The amount of ratings distribution is highly right skewed.

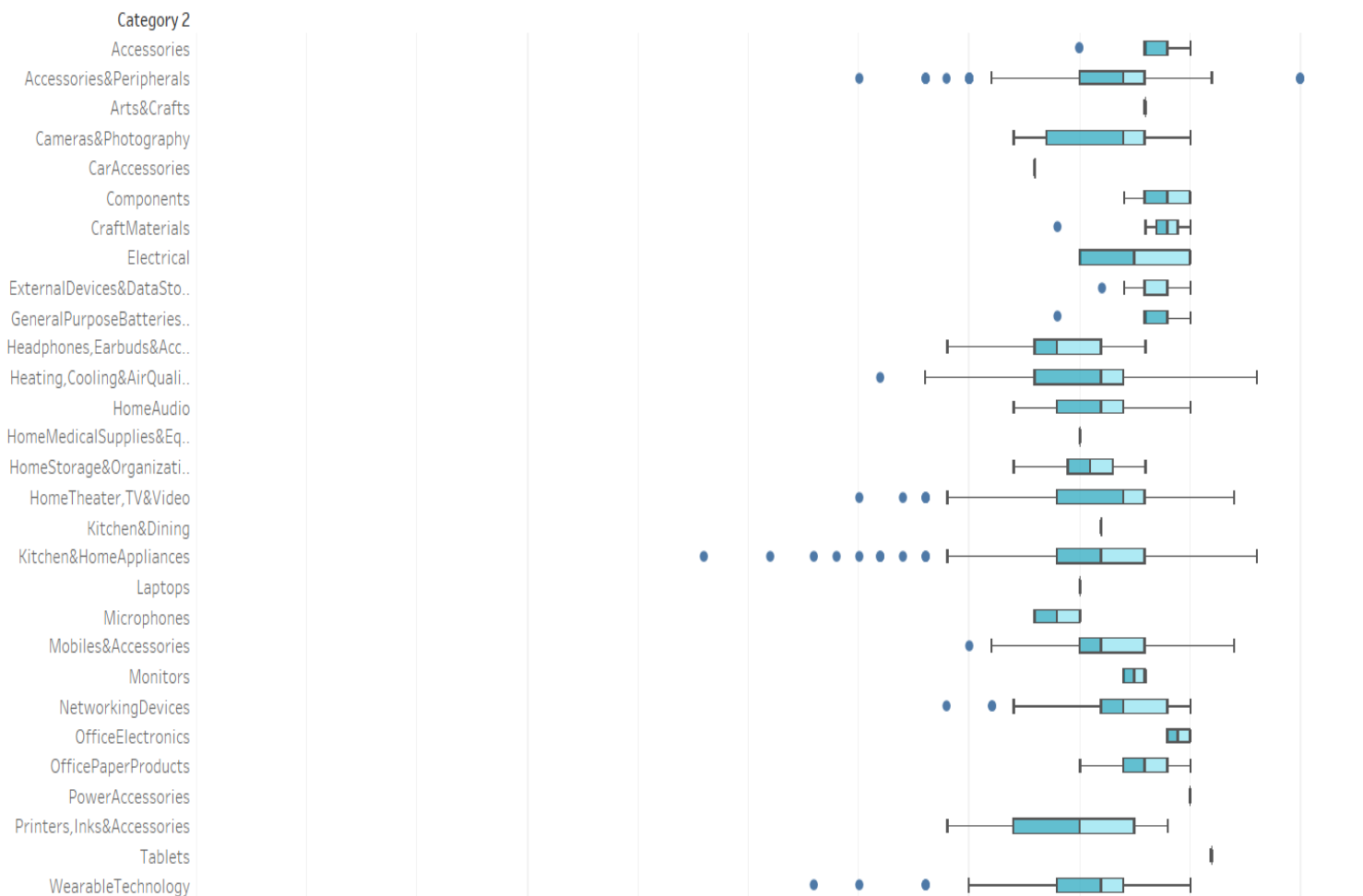- Rating Distribution by Product Main Category

Toys&Games, Car&Motorbike and health&PersonalCare product rating around 3.7 to 4.6. All homeImpprovement and officeProduct have the minimal rating of 4.0. Many of the Computer &Accessories, and Electronics products have ratings in the range of 3.6 - 4.6. Though these categories do have products that have a high rating such as 5.0 and low rating, going down to 2.75.

- Rating of Products Based on Rating category

	Main category	Rating Category	Amount
0	Car & Motorbike	Below Average	0
1	Car & Motorbike	Average	1
2	Car & Motorbike	Above Average	0
3	Car & Motorbike	Excellent	0
4	Computers & Accessories	Below Average	0
5	Computers & Accessories	Average	75
6	Computers & Accessories	Above Average	375
7	Computers & Accessories	Excellent	3
8	Electronics	Below Average	1
9	Electronics	Average	132
10	Electronics	Above Average	393
11	Electronics	Excellent	0
12	Health & PersonalCare	Below Average	0
13	Health & PersonalCare	Average	0
14	Health & PersonalCare	Above Average	1
15	Health & PersonalCare	Excellent	0
16	Home & Kitchen	Below Average	5
17	Home & Kitchen	Average	139
18	Home & Kitchen	Above Average	304
19	Home & Kitchen	Excellent	0

#	Main category	Rating Category	Amount
20	Home Improvement	Below Average	0
21	Home Improvement	Average	0
22	Home Improvement	Above Average	2
23	Home Improvement	Excellent	0
24	Musical Instruments	Below Average	0
25	Musical Instruments	Average	1
26	Musical Instruments	Above Average	1
27	Musical Instruments	Excellent	0
28	Office Products	Below Average	0
29	Office Products	Average	0
30	Office Products	Above Average	31
31	Office Products	Excellent	0
32	Toys & Games	Below Average	0
33	Toys & Games	Average	0
34	Toys & Games	Above Average	1
35	Toys & Games	Excellent	0

- Rating Distribution by Product Sub-Category

What I observed in this Graph of Rating Distribution by Product Sub Category is that Accessories & Peripherals is highly rated product. The Lowest rated product came from the subcategory of heating, cooling & air quality.
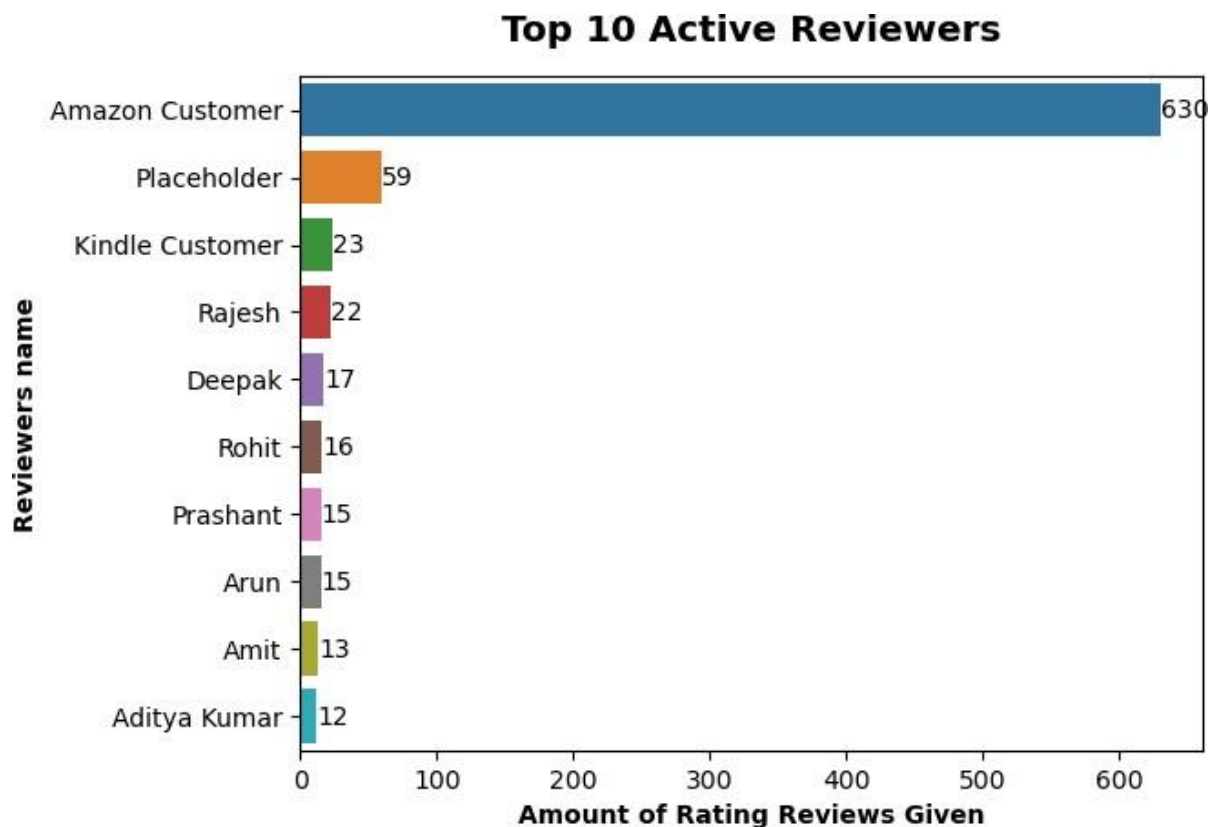
- Rating of all Product in Percentage
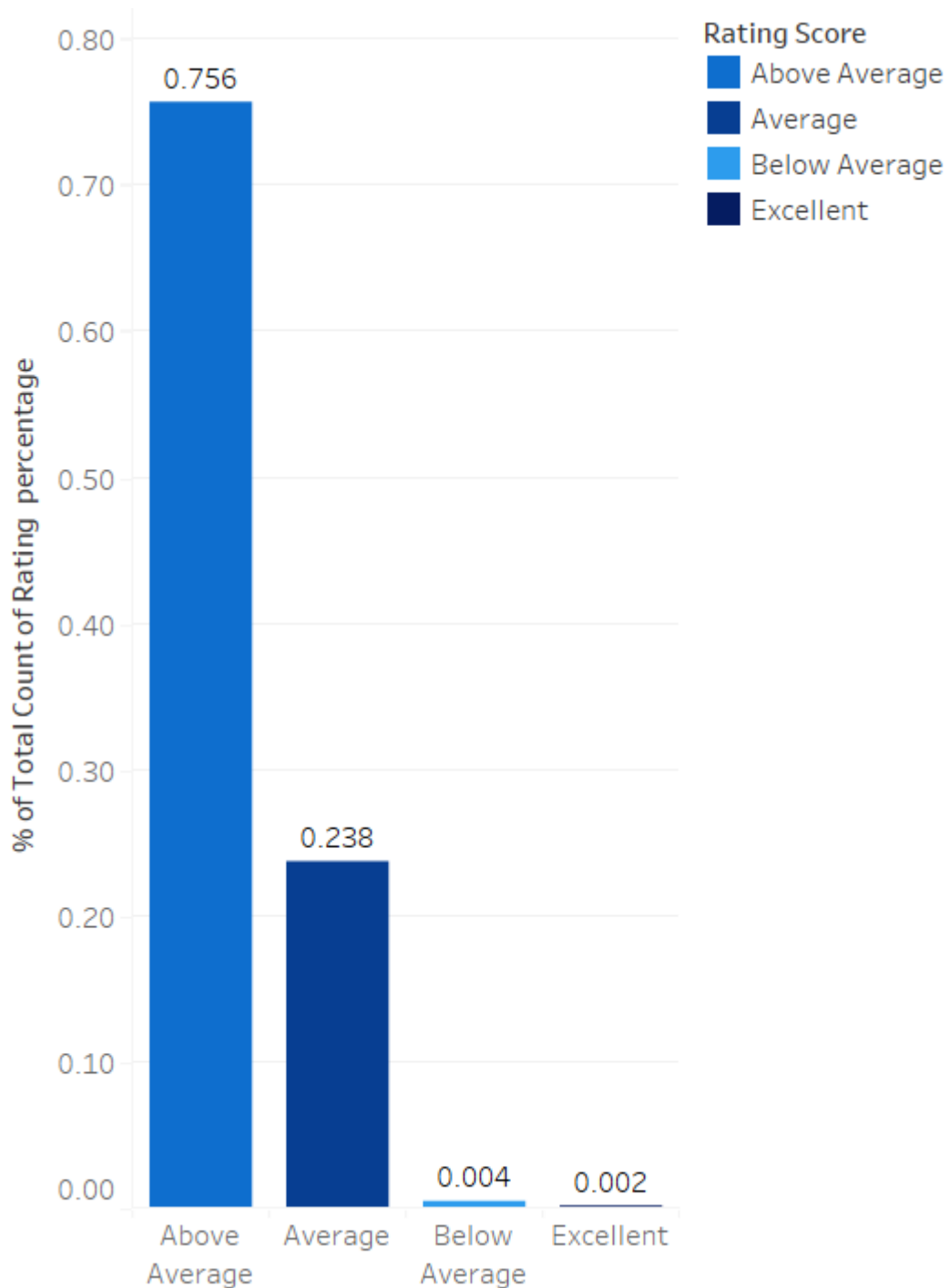
Most of the product in the dataset have been rated Above average. There are extremely fewproducts are rated below Average and Excellent. No Products are rated poor in this dataset.

Reviewers

- Reviewers who gave rating and reviews for more than one product

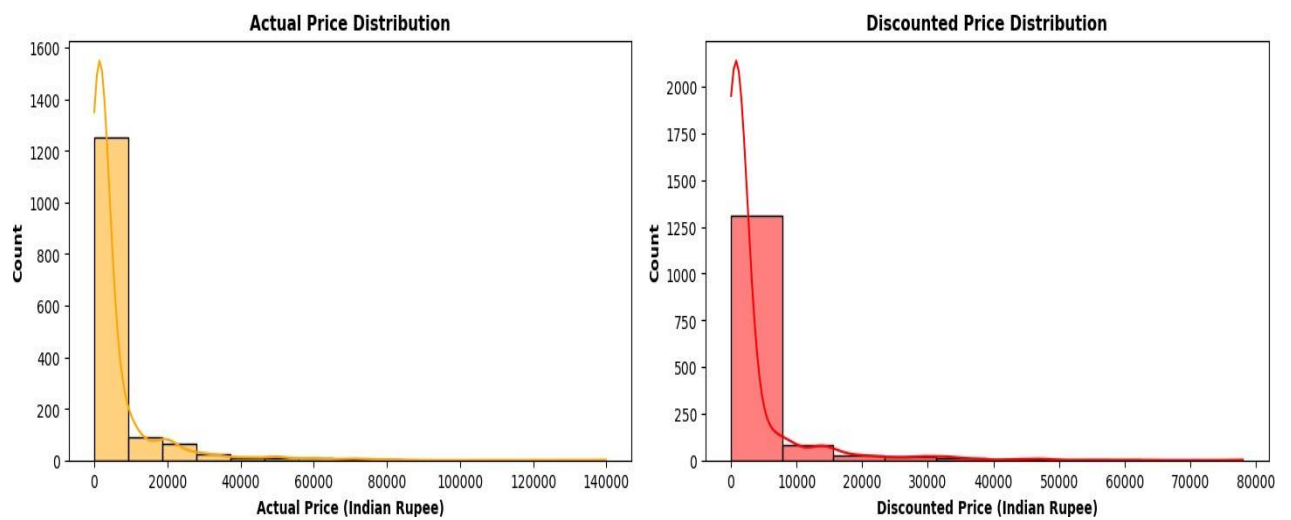
There are more than 500 active reviewers who review the product anonymously under the alias of Amazon customer, Placeholder, kindle customer There are more than 8 people who have given ratings and reviews to more than 10 products on this dataset.
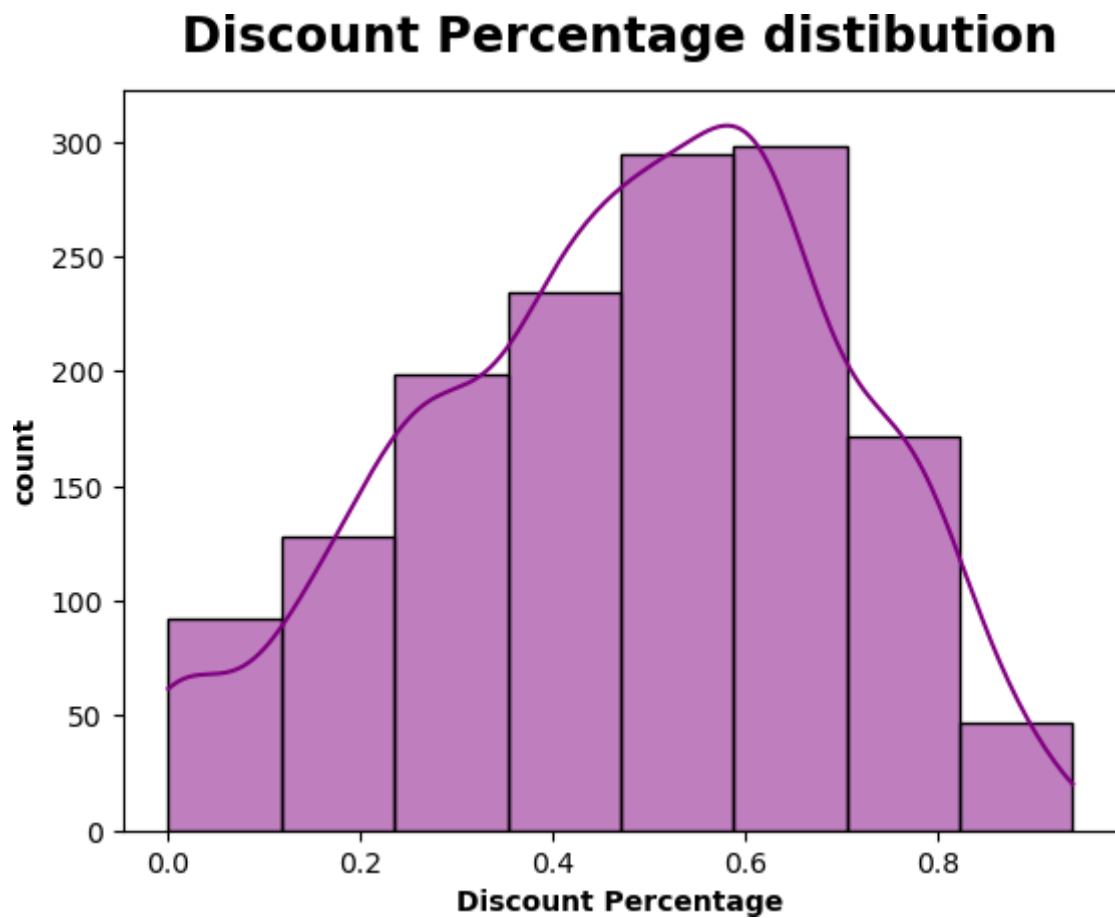
Product Pricing

- Actual price and discounted Price distribution
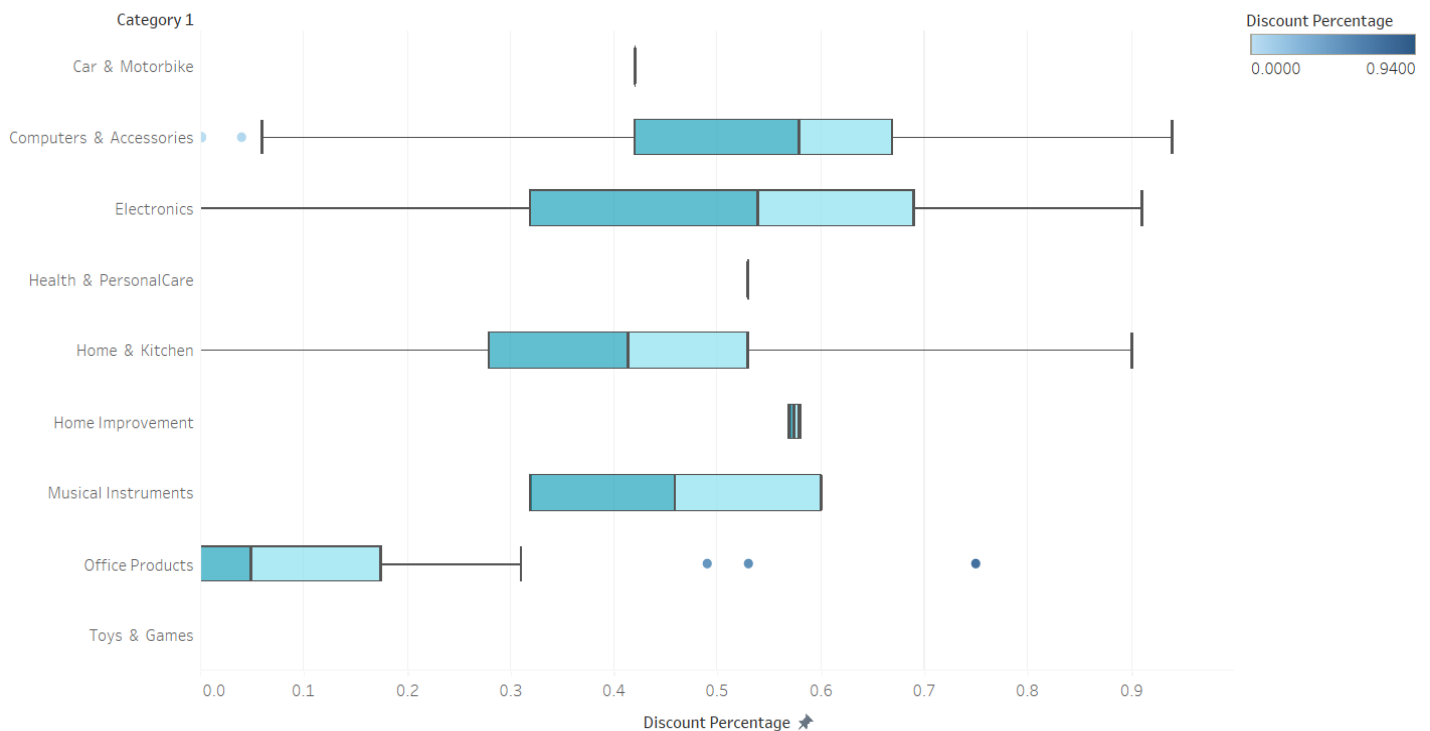

Actual Price and Distcounted Distribution

Both of the Graph shows the same results which is positive Skewed to right.

- Discount Percentage distribution

## Discount Percentage distibution

Most of the Product have the discount of more than 50% to 80%.

- The Discount range by Product Main Category

Computers & Accessories, Electronics, Home & Kitchen have a large widely spread discount ranging from minimal 10% to 90%. Toys & game, Car & Motorbike, Health & Personal Care, HomeImprovement are the least spread discount. office product does not give a large amount of discount as compared to product main category.

- ## The Discount range by Product Sub Category
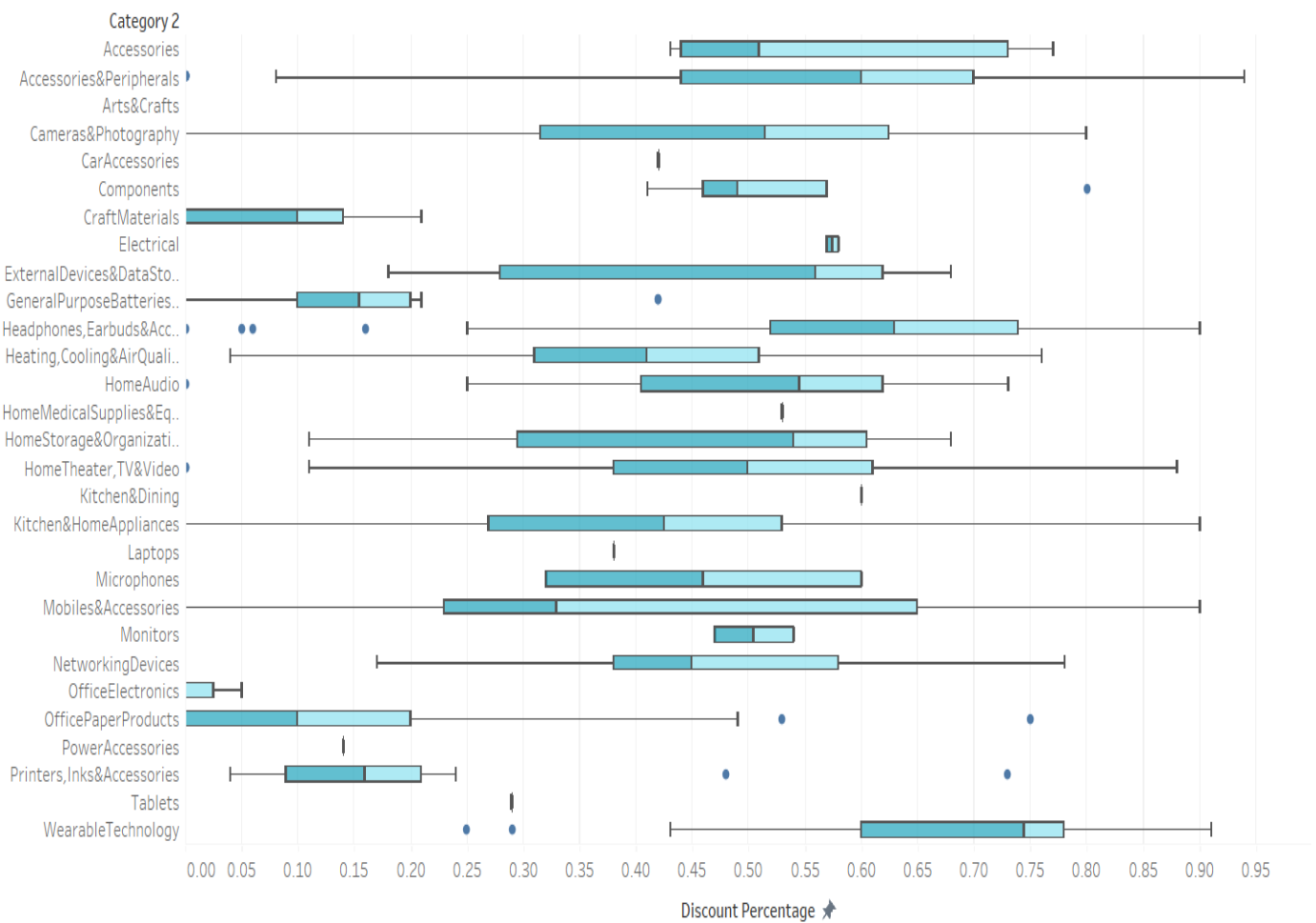
- Actual Price range and discounted Price range by product Main Category
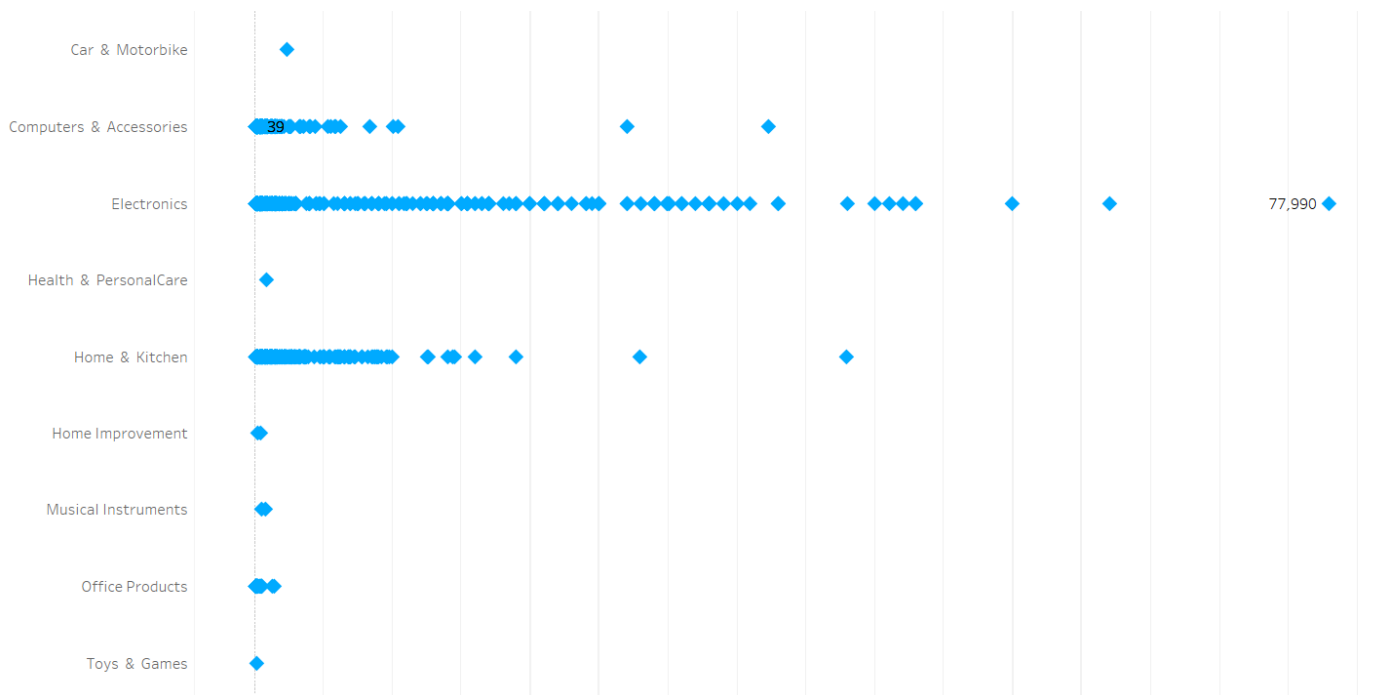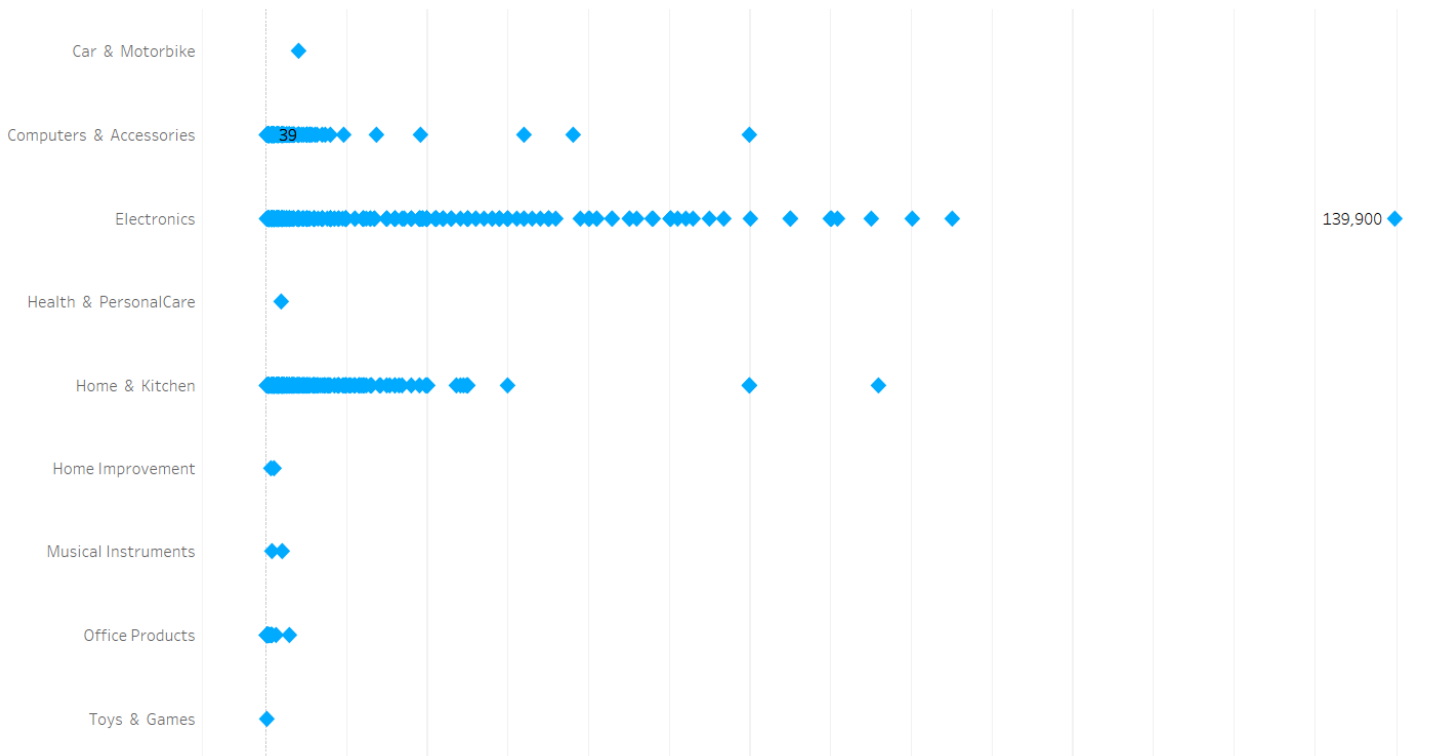
There is the decrease in the product category of electronic after applying Discount. Most of theproduct's actual price falls below 20,000 Rupee. For the discounted price, most of the products fall under 10,000 Rupee.

- Actual Price range and discounted Price range by product Sub Category

Category 2

KEY PERFORMANCE INDICATOR(KPI)

- **Top 5 Most Expensive Products After Discount**
- **Top 5 Cheapest Products after Discount**
- **Top 5 Cheapest Products after Discount**
- **Correlation between features**
- **Rating Amount and of Rating Distribution**
- **Actual price and discounted Price distribution**
- **Discount Percentage distribution**
- **Actual Price range and discounted Price**

CONCLUSION

- **Electronics especially accessories & peripherals and Kitchen & home appliance contain most of the products in this data set.**
- **Sony Bravia 164 cm (65 inches) is the most expensive product after discount.**
- **E-cosmos 5V 1.2W Portable Flexible is the cheapest product after discount.**
- **There are more than 500 active reviewers who review the product anonymously under the alias of Amazon customer, Placeholder, kindle customer There are more than 8 people who have given ratings and reviews to more than 10 products on this dataset.**
- **Sony Bravia 164cm having the largest price difference due to discount**.
- **There is almost no correlation between the dataset but there is positive correlation between the discounted price of product and actual price of product.**
- **Most of the product range around 4.0 to 4.37 with no products under the score of 2.0. The Rating Distribution is Slightly Left-Skewed. The amount of ratings given to a product is very widespread. Most of the products that have been rated, have around 0 - 5000 amount of rating for each product. Interestingly there are products that have more than 40,000 ratings. The amount of ratings distribution is highly right skewed.**
- **Toys&Games, Car&Motorbike and health&PersonalCare product rating around 3.7 to 4.6. All homeImpprovement and officeProduct have the minimal rating of 4.0. Many of the Computer &Accessories, and Electronics products have ratings in**

the range of 3.6 - 4.6. Though these categories do have products that have a high rating such as 5.0 and low rating, going down to 2.75.

- What I observed in this Graph of Rating Distribution by Product Sub Category is that Accessories & Peripherals is highly rated product. The Lowest rated product came from the subcategory of heating, cooling & air quality.

- Most of the product in the dataset have been rated Above average. There are extremely few products are rated below Average and Excellent. No Products are rated poor in this dataset.

- Actual price and discounted Price distribution both of the Graph shows the same results which is positive Skewed to right.

- Most of the Product have the discount of more than 50% to 80%.

- Computers & Accessories, Electronics, Home & Kitchen have a large widely spread discount ranging from minimal 10% to 90%. Toys & game, Car & Motorbike, Health & Personal Care, Home Improvement are the least spread discount. office product does not give a large amount of discount as compared to product main category.

- There is the decrease in the product category of electronic after applying Discount. Most of the product's actual price falls below 20,000 Rupee. For the discounted price, most of the products fall under 10,000 Rupee.

Q & A

Q1) What's the source of data?

Ans) The Dataset was taken from Kaggle

Q2) What was the type of data?

Ans) The data was a combination of numerical and Categorical values.

Q3) What's the complete flow you followed in this Project?

Ans) Refer slide 5th for better Understanding

Q4) What techniques were you using for data?

Ans) -Removing unwanted attributes

-Visualizing relation of independent variables with each other and output variables

-Removing outliers

-Cleaning data and imputing if null values are present.

-Converting Numerical data into Categorical values.

Q6) What were the libraries that you

used in Python?

Ans) I used Pandas, NumPy, Matplotlib,

  and Seaborn libraries in Pandas.

THANK YOU