



Amazon Sales Analysis

DETAILED PROJECT REPORT

Shubham Tembhurne

PROJECT DETAIL

Project Title	Amazon Sales Analysis
Technology	Business Intelligence
Domain	E-commerce
Project Difficulty level	Advance
Programming Language Used	Python
Tools Used	Jupyter Notebook, Tableau

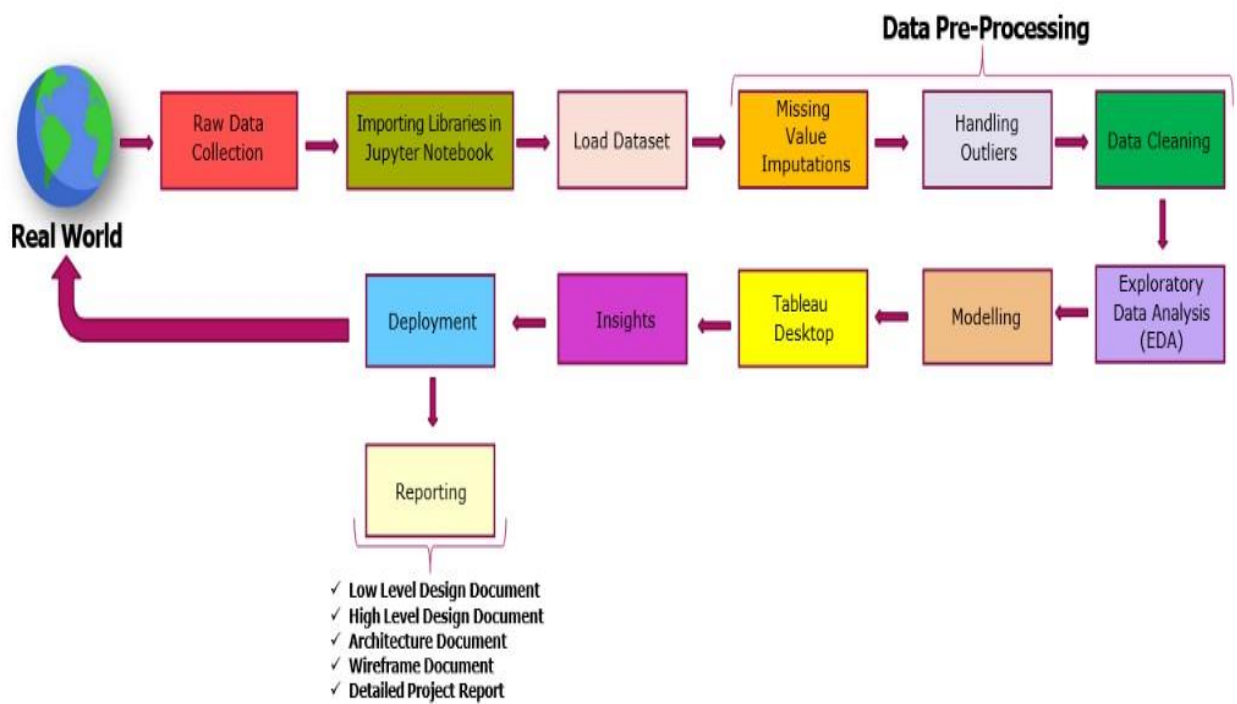
OBJECTIVE

- 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

- Sales management has gained importance to meet increasing competition and the need for improved methods of distribution to reduce cost and to increase profits. Sales management today is the most important function in a commercial and business enterprise.

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.
- **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.
- **CATEGORY:** A category refers to a grouping or classification system used to categorize products based on shared characteristics, attributes, or functions.
- **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.
- **ACTUAL_PRICE:** The actual price refers to the original or regular price of a product before any discounts or promotional offers are applied.
- **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:

		Product ID	
Main Category	Sub-Category		
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	
Home Audio		16	
Home Theater, TV Video		162	
Mobiles Accessories		161	
Power Accessories		1	
Wearable Technology		76	
Health & Personal Care		HomeMedicalSupplies&Equipment	1
Home & Kitchen	Craft Materials	7	
	Heating, Cooling&AirQuality	116	
	HomeStorage&Organization	16	
	Kitchen Dining	1	
	Kitchen&HomeAppliances	308	
Home Improvement	Electrical	2	
Musical Instruments	Microphones	2	
Office Products	Office Electronics	4	
	OfficePaperProducts	27	
Toys & Games	Arts Crafts	1	

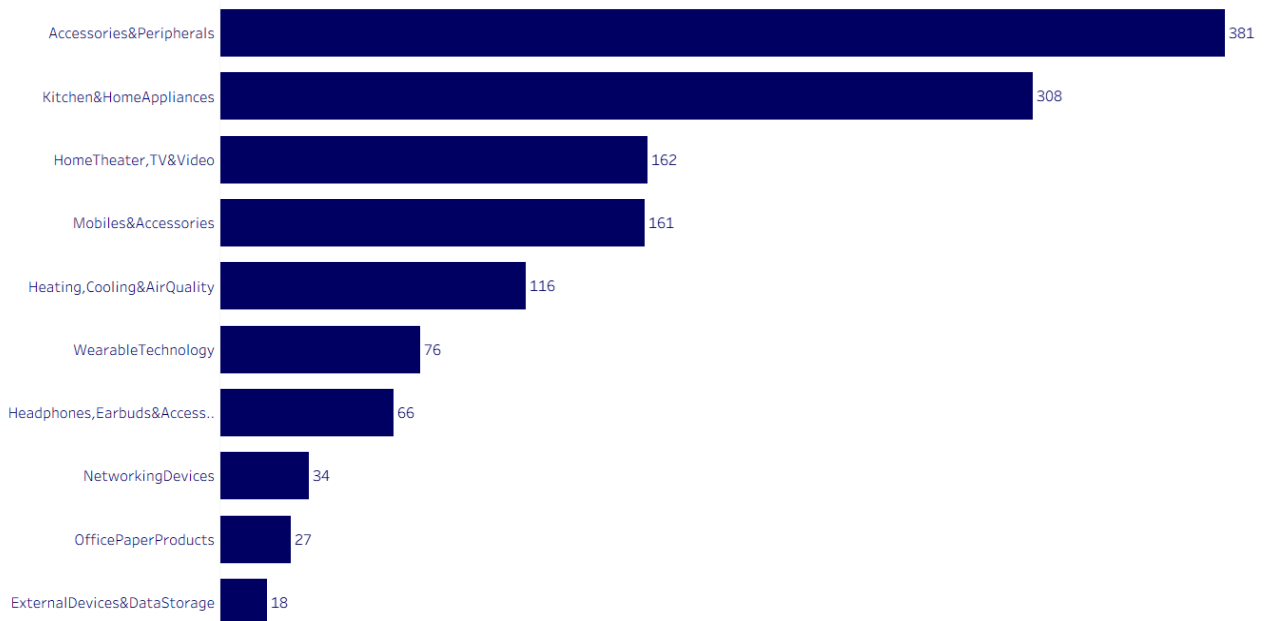
- How many Products belongs to which category and subcategory?

Top 5 Product Category



- We can see that Most of the Product Category is from Electronics, Computers & Accessories then Home & Kitchen.

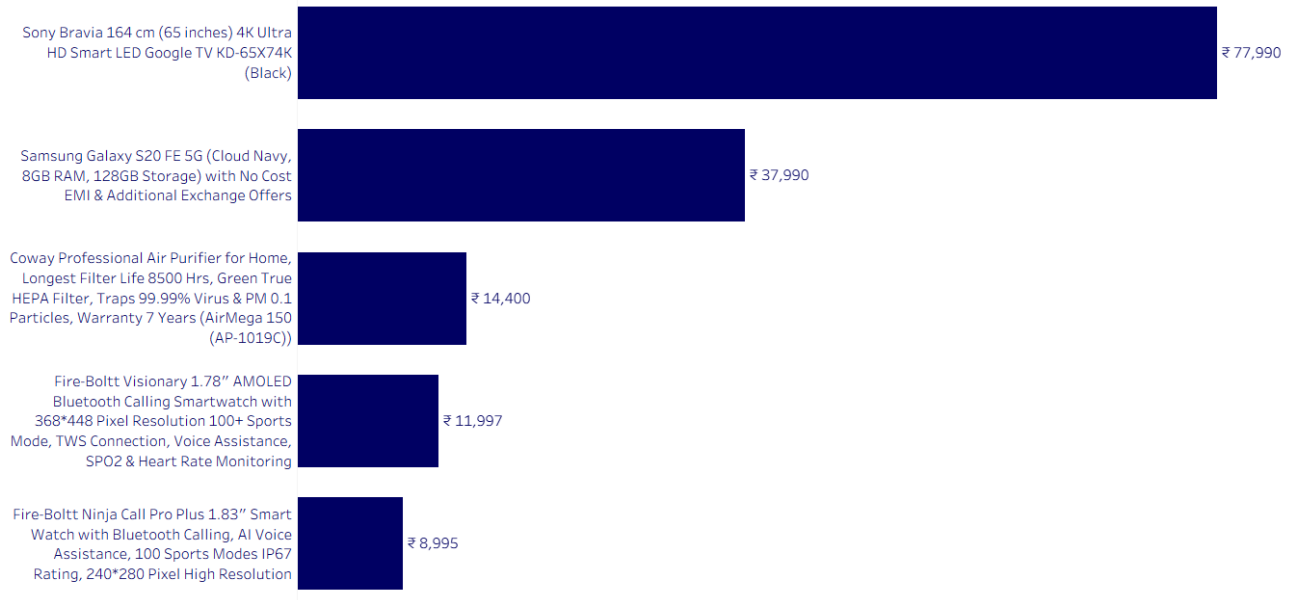
Top 10 Product Sub-Category



- Most of the Sub-Categories from Accessories and Peripherals, Kitchen and Home Appliances, Home theater and TV Video this all sub-Category mostly belongs to Electronics Category.

- What is the Top 5 Most Expensive Products After Discount?

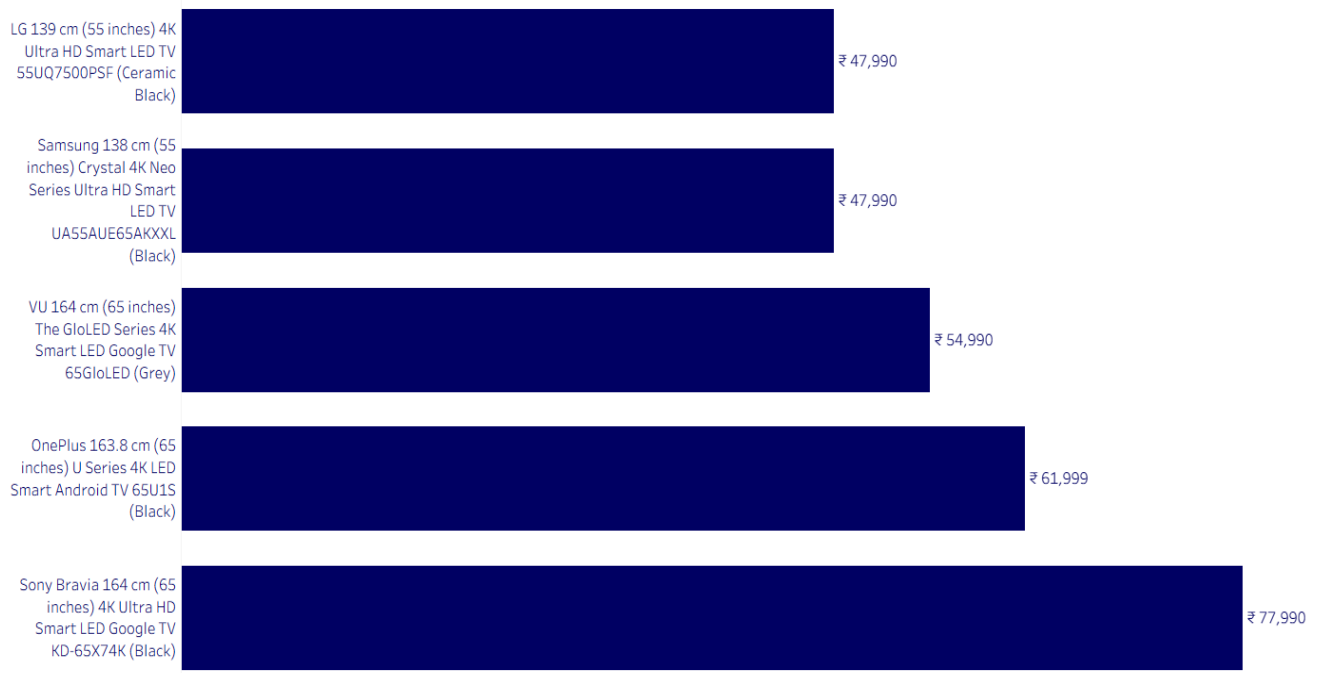
Top 5 Expensive Product After Discount



- Sony Bravia 164 cm 4k Ultra HD Smart LED TV (Black) most expensive product after discount which is from the Category of electronics which is 77990rs.

- What is the Top 5 Cheapest Products after Discount?

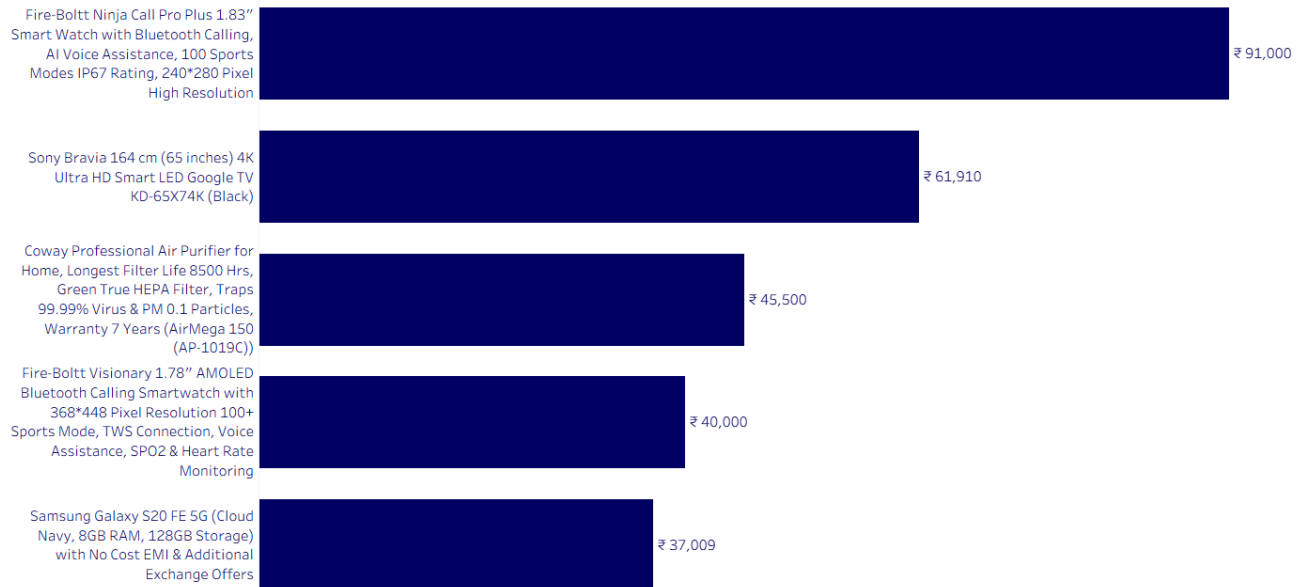
Top 5 Cheapest Product After Discount



- LG 139 cm 4k Ultra HD Smart LED TV is the cheapest product after discount which is 47,990rs.

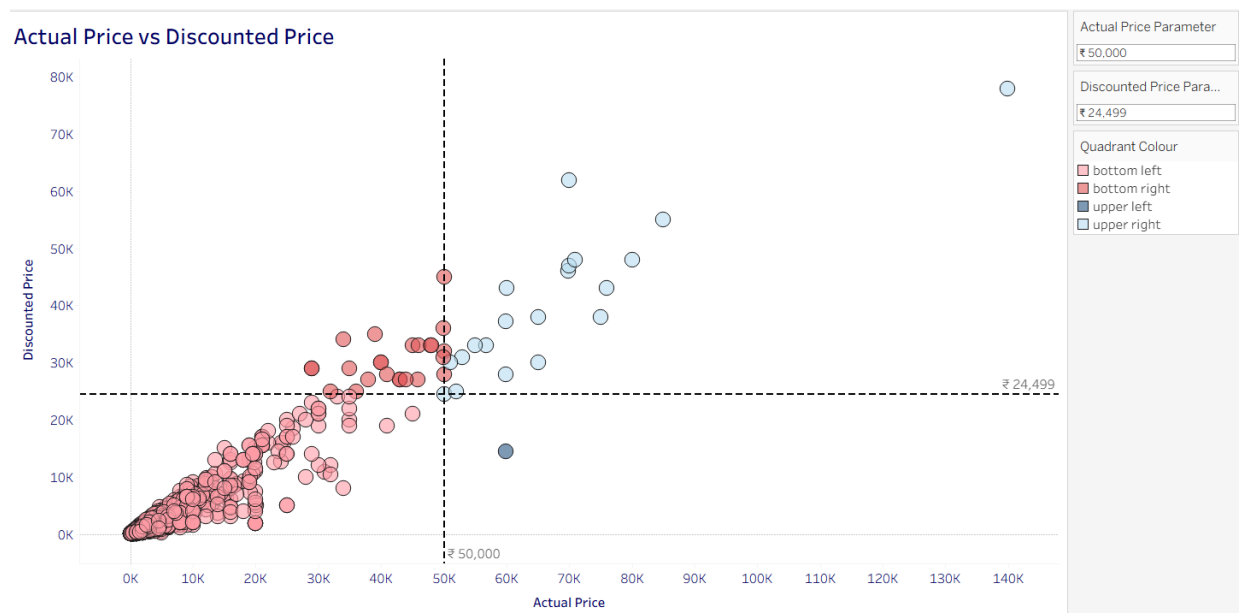
- What is the Price difference due to discount in Product?

Top 5 Largest Price Difference Due to Discount



- Fire Boltt Ninja Call Pro Plus 1.83 Smart Watch product with largest price difference which is 91,000.

- Correlation between features: Actual Price Vs Discounted Price

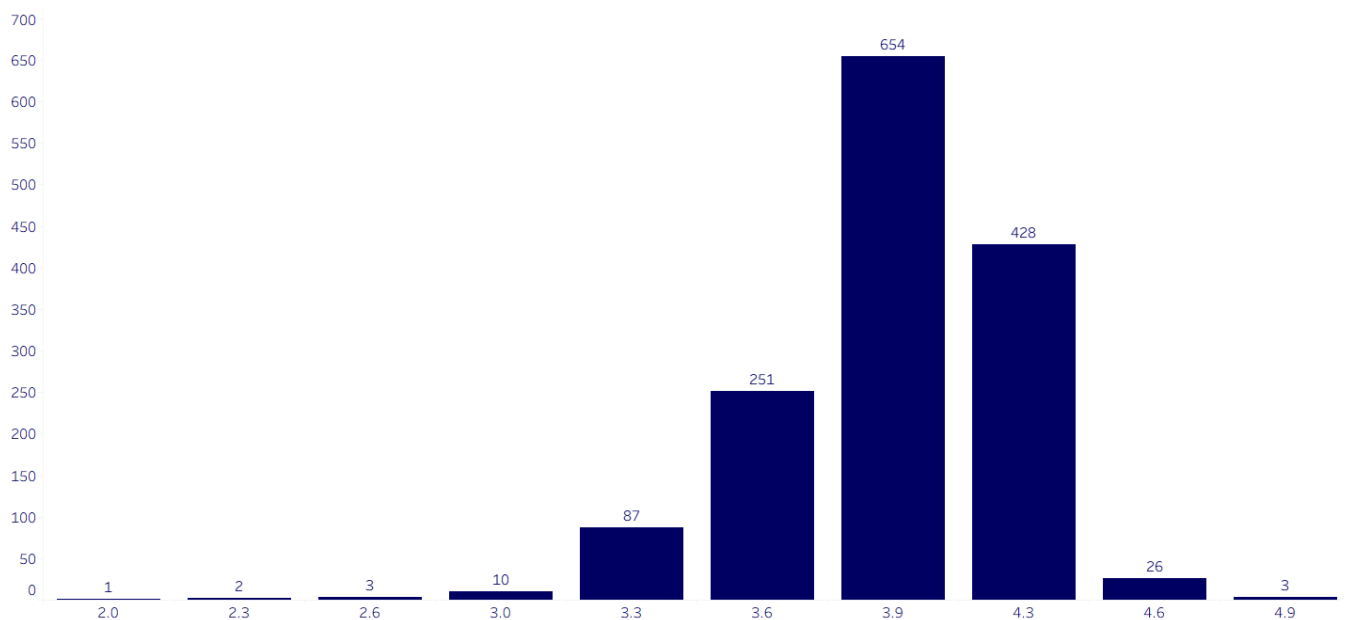


- There is a strong linear relationship the high-ticket items greater than 50k still receive heavy discount.
- We are Sacrificing the large amount of profit by significant margin sacrifice even on premium product.

Product Rating:

- What is the distribution of product rating?

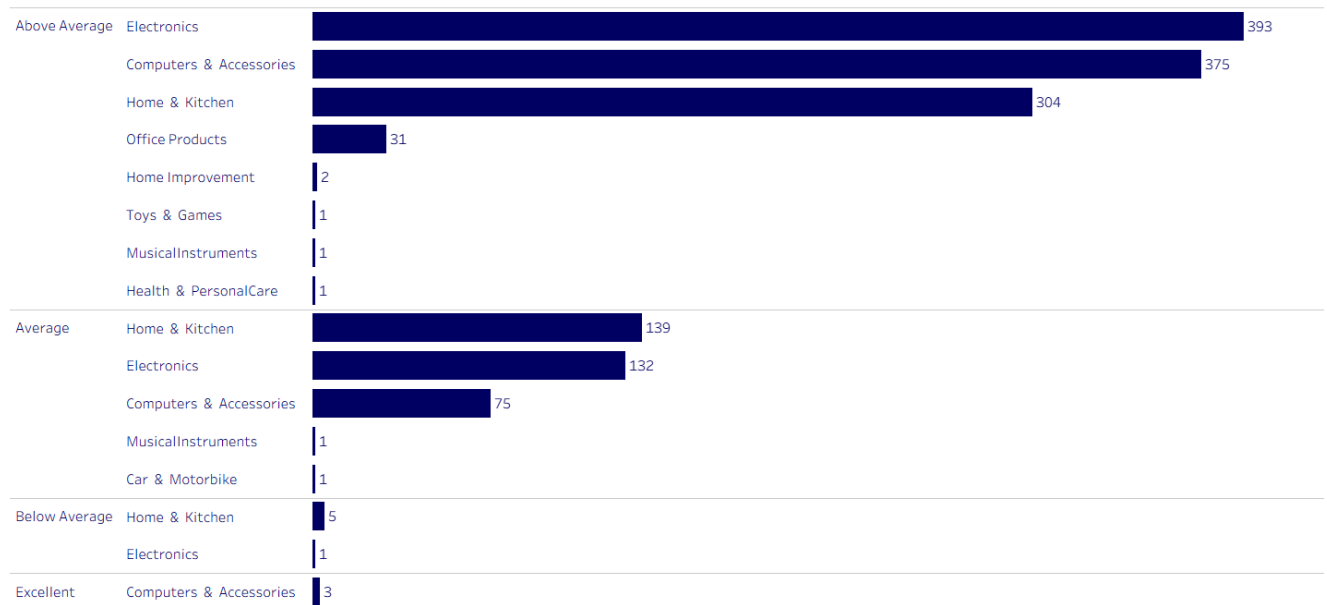
Product Rating Distribution



- Majority of the product lie between 4.0 to 4.5 and very few products are below 3.5 that indicates strong quality control Low rated product are rare but risky.

- What is the Rating score distribution by Product Category?

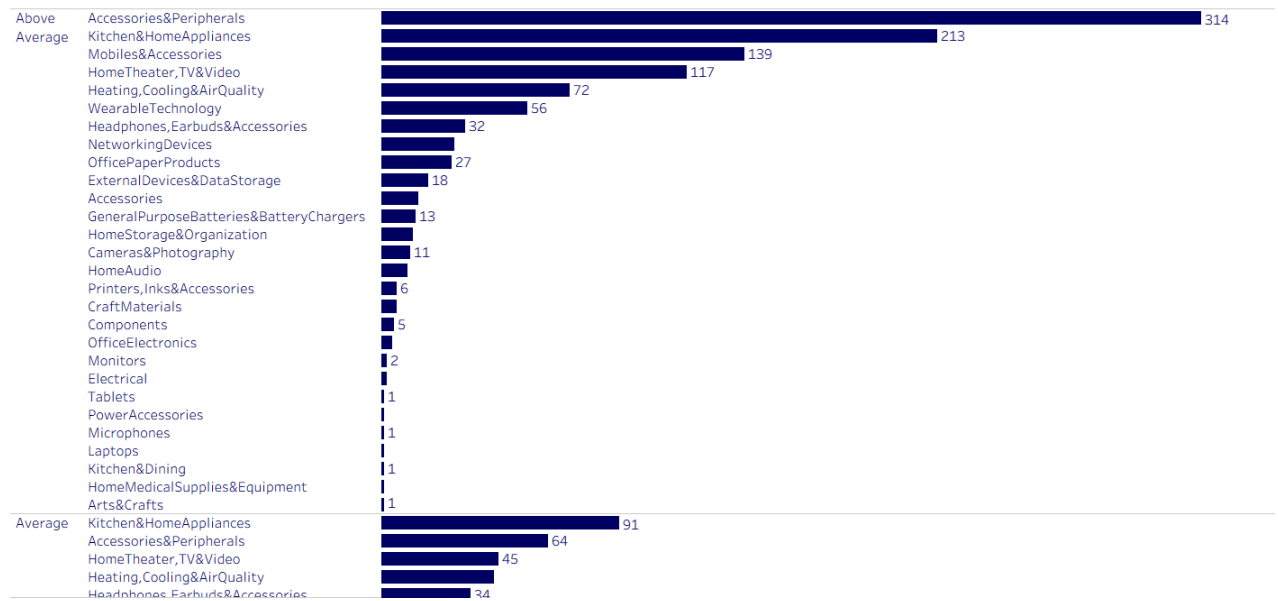
Rating Score Distribution by Category



- The most of the product are above average specially Computers & Accessories, Electronics, Home & kitchen Office Product this are scalable Category.

- What is the Rating of Products Based on Sub-category?

Rating Score Distribution by Sub-Category

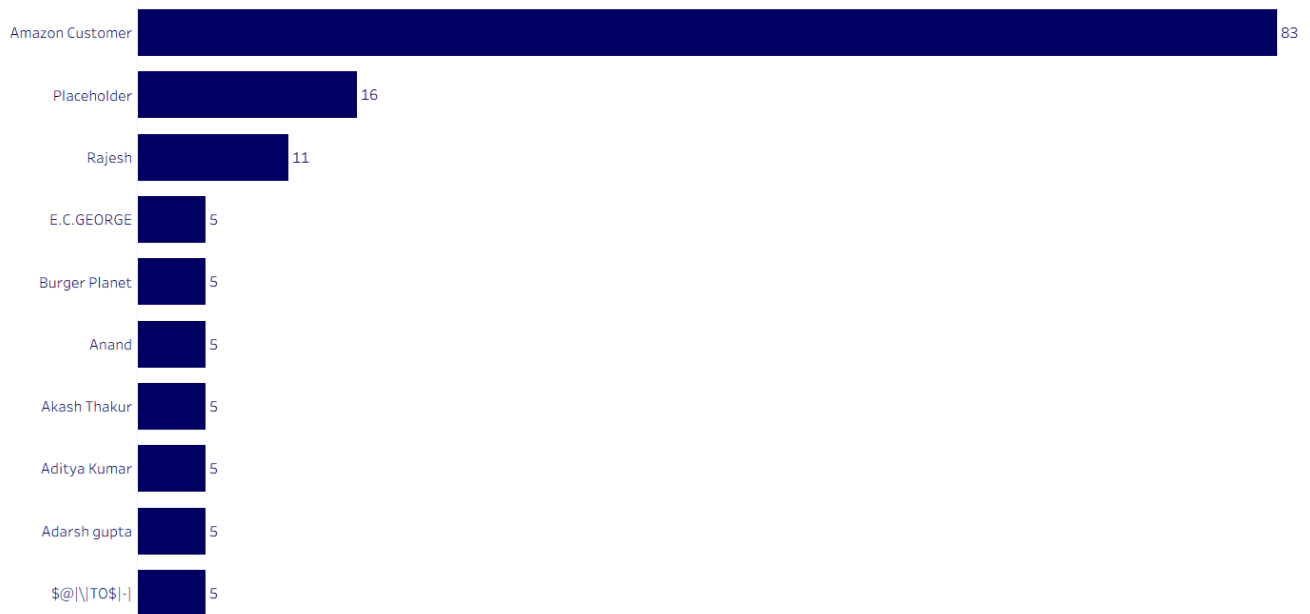


- I can see that most of the subcategory is the category of average and above average Scalable subcategory.

Reviewers:

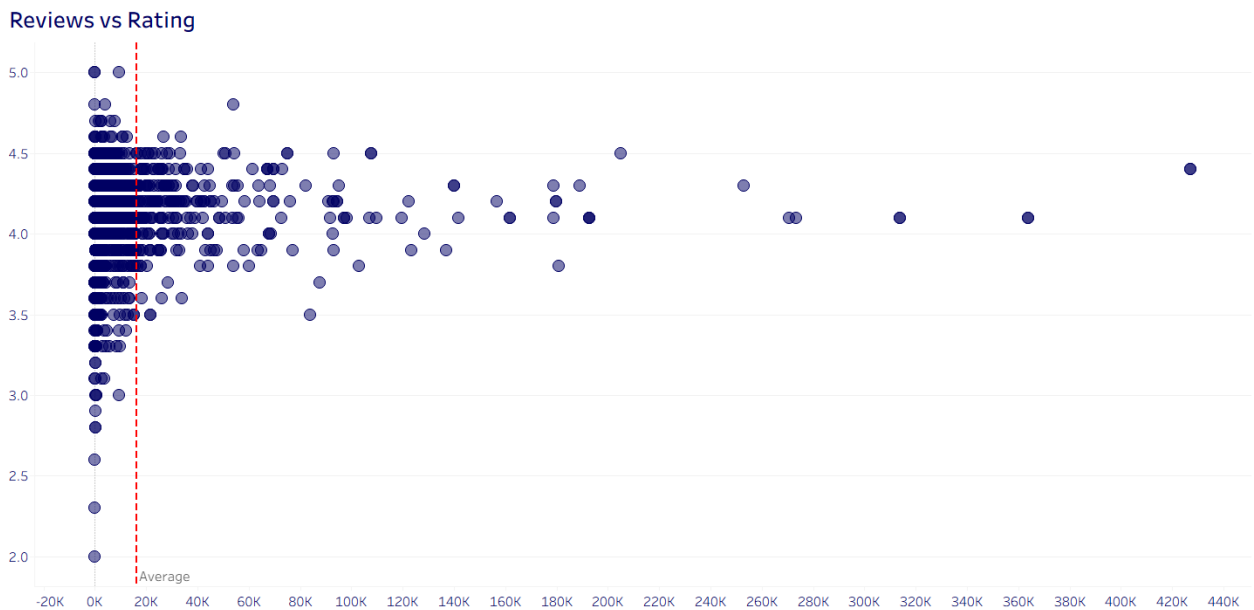
- Who gave rating and reviews for more than one product?

Top 10 Active Reviewers



- More than 500 active reviewers who review the product under Amazon Customer, Placeholder, Kindle Customers. These are active customers who gave more than 10 reviews in the Category.

- What is the compare numbers of review vs rating?

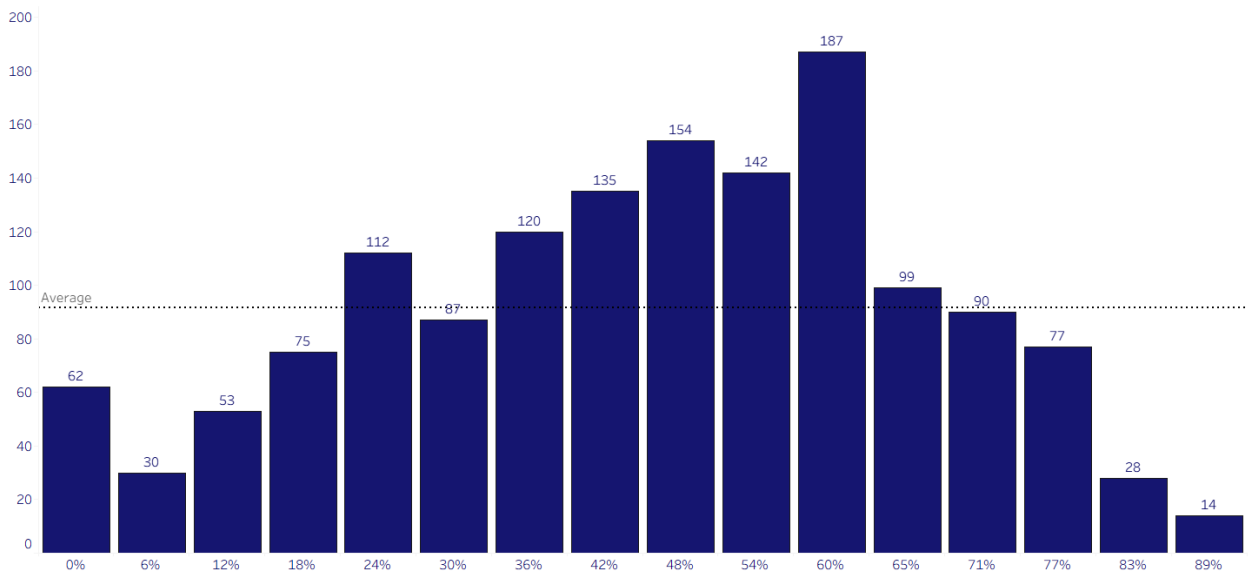


- Product with high rating (100k+) mostly stay around between 4.0 to 4.5 Rating Stability as review increases. Low reviews show high rating volatility.

Product Pricing:

- What is discount percentage distribution amongst the products?

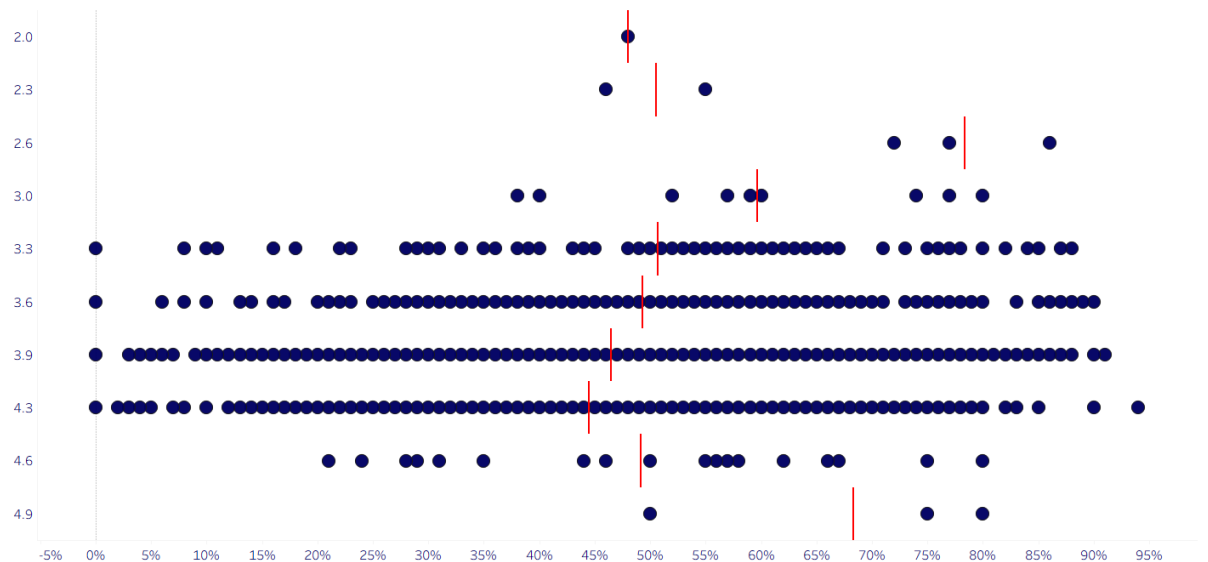
Discount Percentage Distribution



- Most product fall between 30% to 70% discount.
- Extreme discount 80% to 90% exist but are fewer very low discount less the 10% is uncommon.

- What is the compare numbers of discount and Rating?

Discount vs Rating



- No Relationship between Higher discount and higher rating.
- Product between (10%-20%) discount have the rating of 4.5 rating.
- Over discounting does not improve customer satisfaction.

KEY PERFORMANCE INDICATOR(KPI)

- Top 5 Most Expensive Products After Discount
- Top 5 Cheapest Products after Discount
- Rating score Distribution by Category and Sub-Category
- Correlation between features (Reviews Vs Rating)
- Discount vs Rating Distribution
- Actual price and discounted Price distribution
- Discount Percentage distribution
- Top 5 Largest Price Difference due to Discount

CONCLUSION:

- Amazon heavily relies on discount-led selling.
- Despite high discount rating remains strong which indicate customers values perceived deals.
- Large reviews volume indicates high customer engagement and trust.
- Low rated products are rare but risky.
- Majority of product lie between 4.0 to 4.5 very few products below 3.5 indicates strong quality control or review moderation.
- Low rated product should be monitor or delisted.
- Most of the product fall between 30% to 70% discount and extreme discount 80% to 90% exist but are fewer.
- Very Low discount less than 10% are uncommon.
- No strong correlation between higher discount equal to higher rating.
- Product with 0% to 20% discount still achieve 4 + rating.
- Over discounting does not improve customer satisfaction.
- Sellers are losing margin over unnecessarily discount.
- Product with high reviews 100k+ mostly stay around 4.0 and 4.3
- Rating stabilizes as reviews increase.
- Hight ticket items more than 50k still receive heavy discount.
- Early reviews are critical for the product success.
- Significant margin sacrifice even on premium products.
- Electronic dominates the platform.
- Other categories are niche and underrepresented.
- Non -core categories have growth opportunity but low focus.

Strategic Recommendation

- Reduce discounts on **4.3+ rated products**
- Push review acquisition for new products
- Expand **Home, Personal Care & niche categories**
- Focus on **price-value balance**, not just discount %

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