E-commerce Data Analysis: Trends and Customer

Sentiment

Documentation

**Name: Logiya Vidhyapathi**

**Date: 16/07/2023**

**E-commerce Sales Data:**

This dataset contains detailed information on over 1,000 Amazon products, including their ratings and reviews. The key features include:

* **Product Details:** ID, name, category, discounted price, actual price, discount percentage, description, image link, and product link.
* **Rating Information:** Rating score and the number of votes.
* **Review Information:** User ID, user name, review ID, review title, and review content.

Overall, the dataset comprises **1,457 rows** and **15 columns**, providing comprehensive insights into product performance and customer feedback on Amazon.

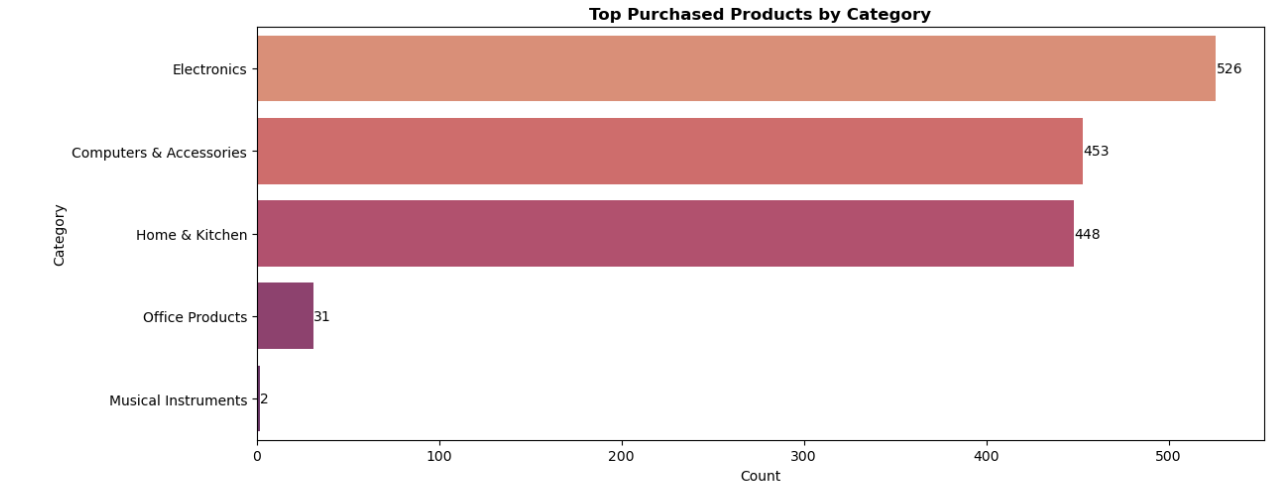
**Business Question:**

How does the discount percentage impact the sales volume of products across different categories on Amazon?

How do customer sentiments and review ratings influence the purchasing decisions and discount strategies for different product categories on Amazon?

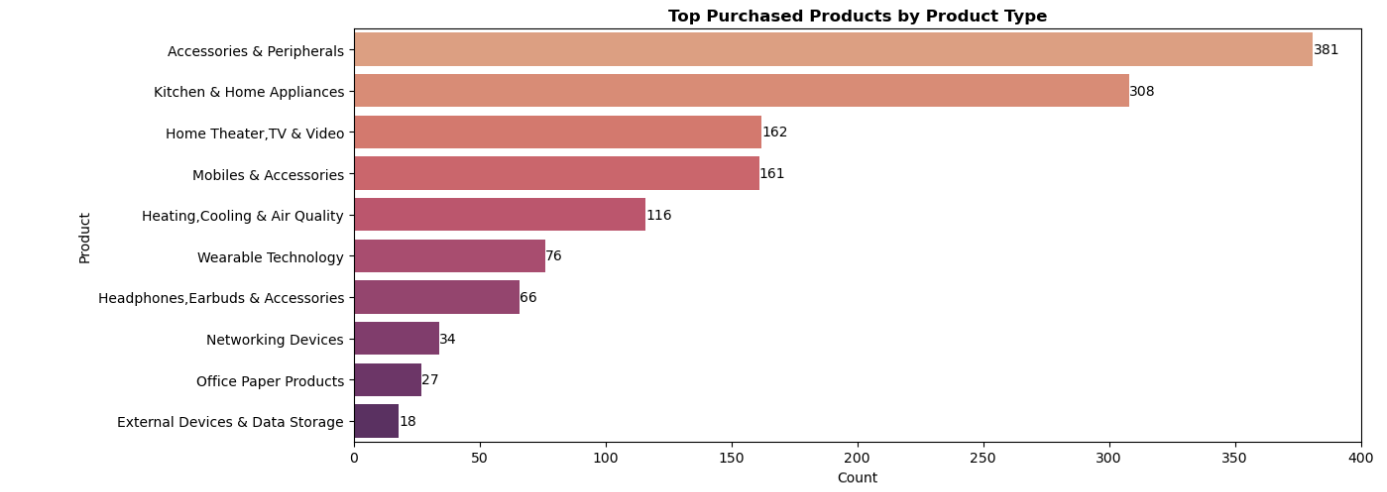
**Bar Plot:**

**Top Purchased Products by Category:**



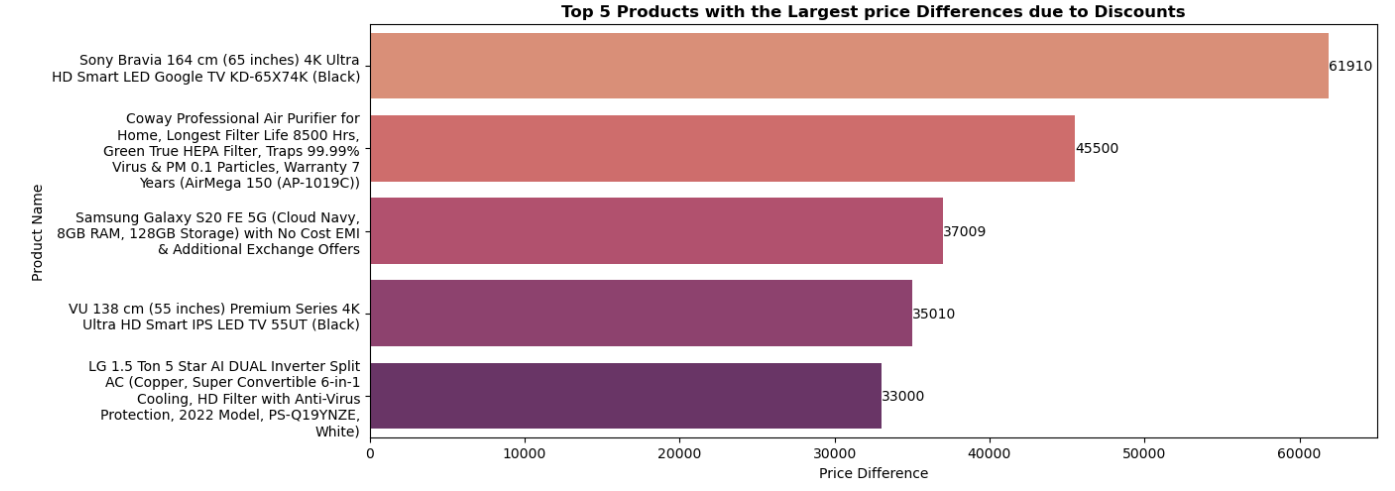
The **bar plot** shows the top purchased products by category. **Electronics** is the **most purchased category**, followed by **Computer & Accessories**. This indicates that these two categories are the most popular among customers compared to other categories.

**Top Purchased Products by Product Type:**



The **bar plot** shows the top purchased products by product. **Accessories & Peripherals** is the **most purchased product**, followed by **Kitchen & Home Appliances**. This suggests that these two product types are the most popular among customers compared to other products.

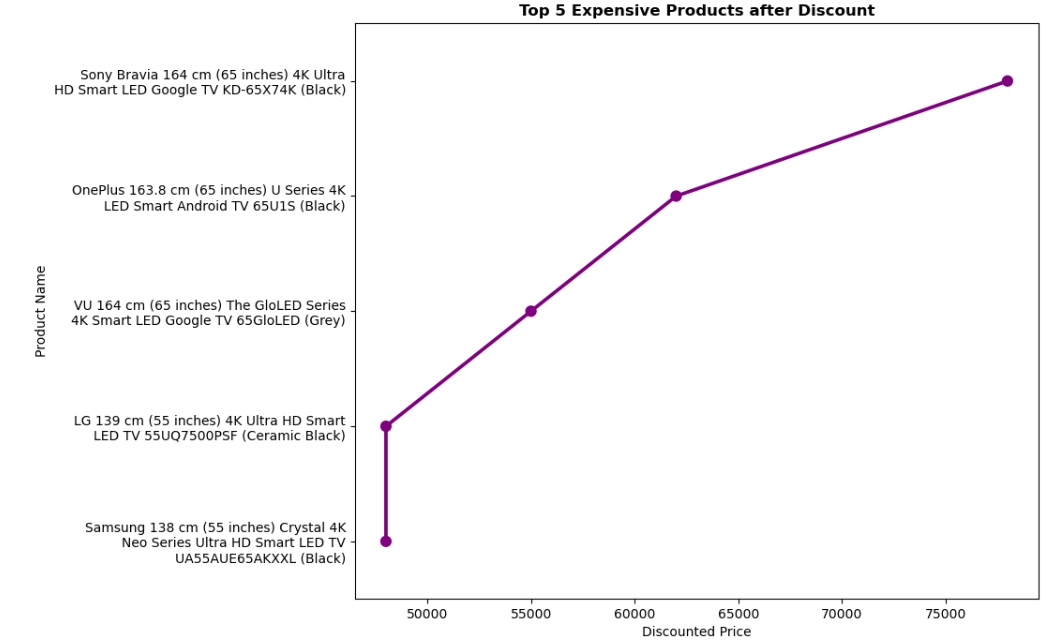
**Top 5 Products with the Largest Price Differences due to Discounts:**



The **bar plot** displays the top 5 products with the largest price differences due to discounts. **Sony Bravia 164 4K Ultra HD Smart LED Google TV** has the **greatest price difference** at **61910** after the discount, followed by **Coway Professional Air Purifier** at **45,500**. This indicates that these two products have the largest price differences due to discounts compared to the other three products.

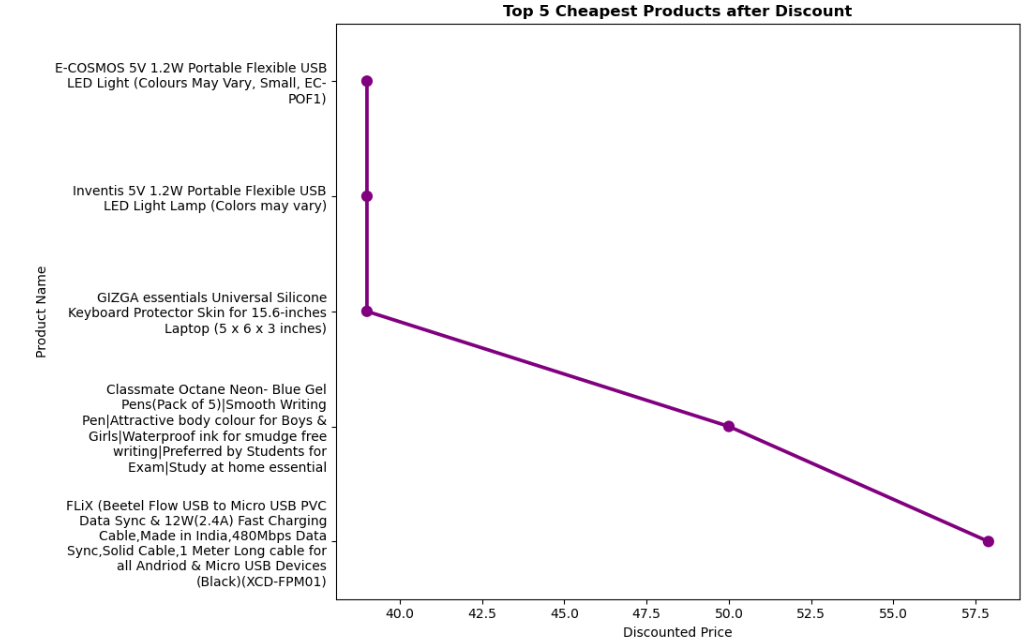
**Point Plot:**

**Top 5 Expensive Products after Discount:**



The **point plot** shows the top 5 expensive products after discounts. **Sony Bravia 164 4K Ultra HD Smart LED Google TV** is the **priciest** followed by **OnePlus U Series 4K LED Smart Android TV**. This indicates that these two products are the most expensive compared to the other three products.

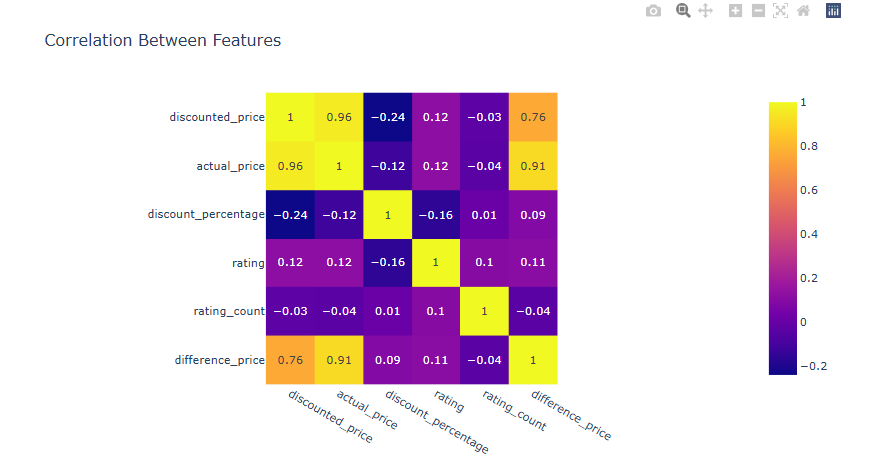
**Top 5 Cheapest Products after Discount:**



The **point plot** displays the top 5 cheapest products after discounts. **E-COSMOS 5V1.2W Portable Flexible USB LED Light** is the **least expensive** followed by the **Inventis 5V 1.2W Portable Flexible USB LED Light Lamp**. This indicates that these two products are the cheapest compared to the other three products.

**Heatmap:**

**Correlation between Features:**

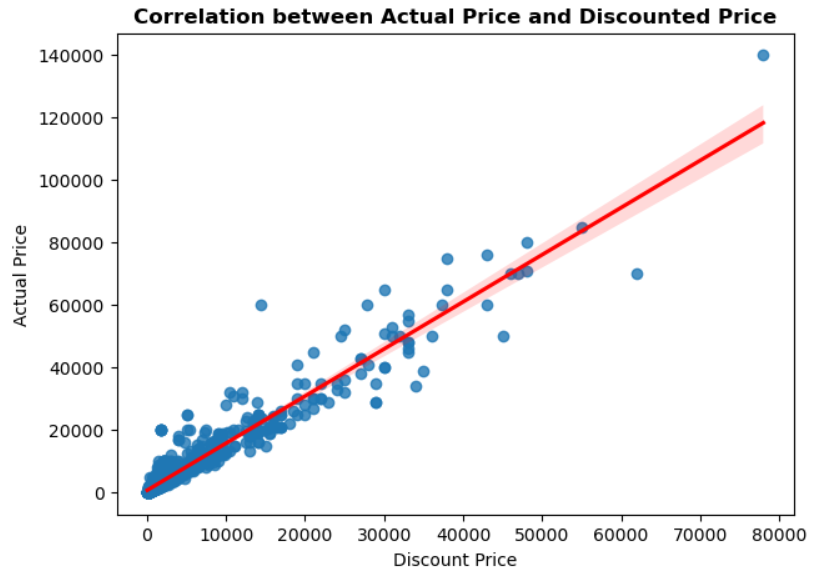


The **heatmap** shows correlations between various factors in the dataset:

* **Actual Price and Discounted Price:** There is a strong positive correlation, indicating that higher original prices often lead to higher discounted prices.
* **Discount Percentage and Actual Price:** This correlation is negative, suggesting that as the original price increases, the percentage discount tends to decrease.
* **Discount Percentage and Discounted Price:** There is a positive correlation, indicating that higher discounts are associated with lower discounted prices.
* **Rating and Rating Count:** There is a positive correlation, implying that products with higher ratings tend to have more reviews.
* **Difference Price and Actual Price:** This correlation is likely positive, indicating that larger differences between original and discounted prices are more common with higher original prices.

**Scatter Plot:**

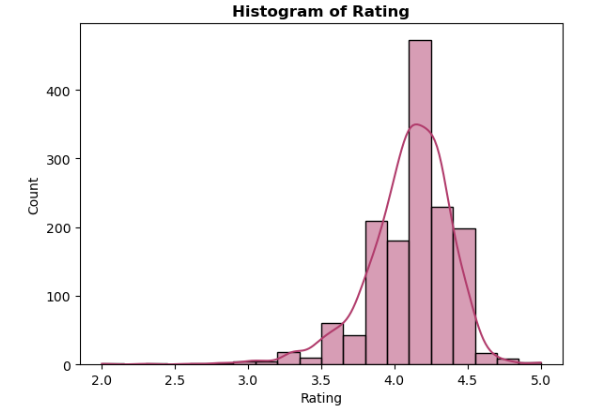
**Correlation between Actual Price and Discounted Price:**



The **scatter plot** illustrates a **clear positive correlation** between the **actual prices** and **discounted prices of products**. As actual prices increase, discounted prices also tend to increase proportionally. This suggests that products with higher initial prices also receive higher discounts, potentially influencing purchasing decisions based on perceived value and affordability.

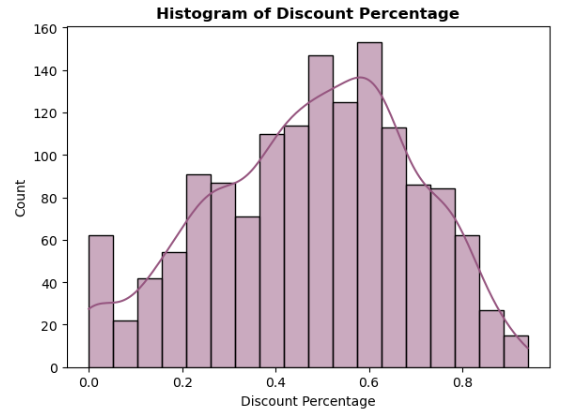
**Histogram:**

**Histogram of Rating:**



The histogram shows that most product ratings fall **within the range of 4.0 to 4.3**, with **no** **ratings below 2.0**. The distribution of ratings is **slightly left-skewed**, indicating a **concentration of higher ratings**.

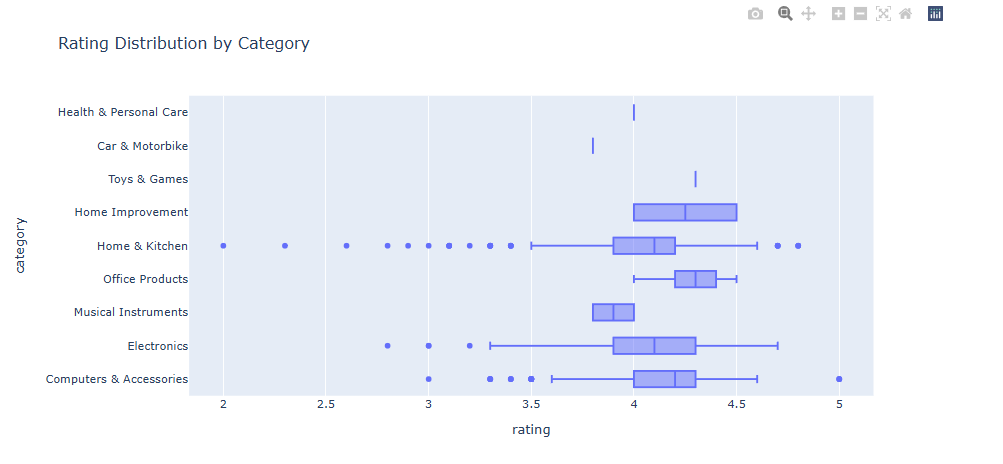
**Histogram of Discount Percentage:**



In the histogram, the **majority of products** in the dataset exhibit discounts ranging between **48% and 63%**. This range indicates that discounts are encountered around this interval, with **fewer products** showing discounts **below 48% or above 63%.**

**Box Plot:**

**Rating Distribution by Category:**



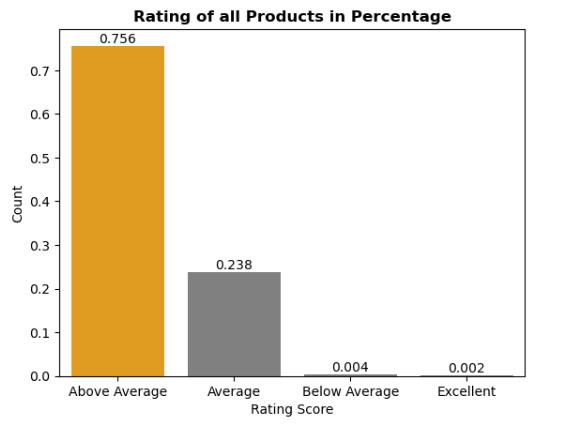
The **Box plot** illustrates the rating distribution by category:

* **Computers & Accessories** and **Electronics** category have ratings mostly **between 3.3 and 4.7**, with some products rated as high as 5.0 and as low as 2.75.
* **Toys & Games**, **Cars & Motorbike** and **Health & Personal Care** category generally have **ratings** **ranging from 3.75 to 4.37**.
* **Home Improvement** and **Office Products** categories consistently have a minimum **rating of 4.0.**
* **Home & Kitchen** category shows a wide range of ratings, **from a high of 4.75 to a low of 2.0**, the lowest rating across all categories. Most products in this category have ratings **between 3.5 and 4.6**.

Overall, **Home & Kitchen** category exhibit the **broadest variation** in ratings, while **Home Improvement** and **Office Products** maintain consistently **higher ratings**.

**Bar Plot:**

**Rating of all Products in Percentage:**



The **bar plot** indicates that the majority of products in the dataset are rated as **Above Average** and **Average**. There are very few products with ratings classified as **Below Average** and **Excellent**, and no products have been rated as **Poor**.

**Sentiment Analysis of Overall Reviews:**



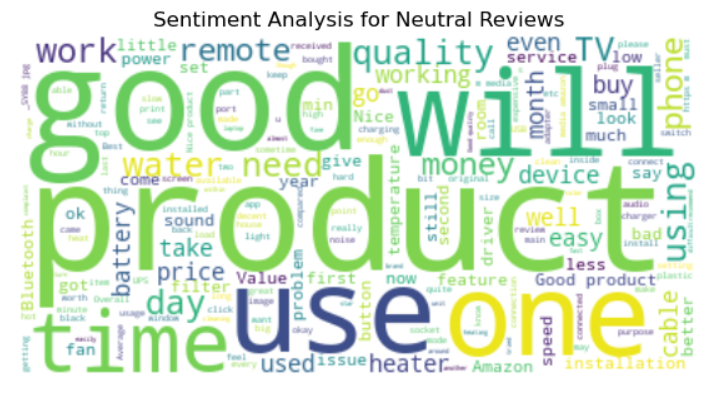
The sentiment analysis for **overall reviews** shows customers express positive sentiments like **‘good’, ‘best’ and ‘look’** indicating widespread **satisfaction with product quality and usability**. However, terms such as **‘poor’, ‘low’, ‘slow’ and ‘bad’** also appear, **suggesting areas for improvement in product reliability and customer service**. This visualization captures diverse customer feedback, highlighting both strengths and areas needing enhancement.

**Sentiment Analysis for Positive Reviews:**



The sentiment analysis for **positive reviews** highlights key terms such as **“Good”, “Best”, “Nice”, Great”** indicating **high customer satisfaction**. Words like **“Quality”, “Value” and “Better”** suggest **appreciation for product functionality and cost-effectiveness**. Terms related to customer service, like **“friendly” and “support”**, emphasize a **positive buying experience**. This visualization succinctly captures the attributes most valued by customers.

**Sentiment Analysis for Neutral Reviews:**



The sentiment analysis for **neutral reviews** shows words like **“good”, “working”, “price”, “small”** indicating a **balanced sentiment among customers**. This visualization offers insights into customer perceptions that are neither overly positive nor negative, pinpointing areas where enhancements could potentially improve satisfaction levels.

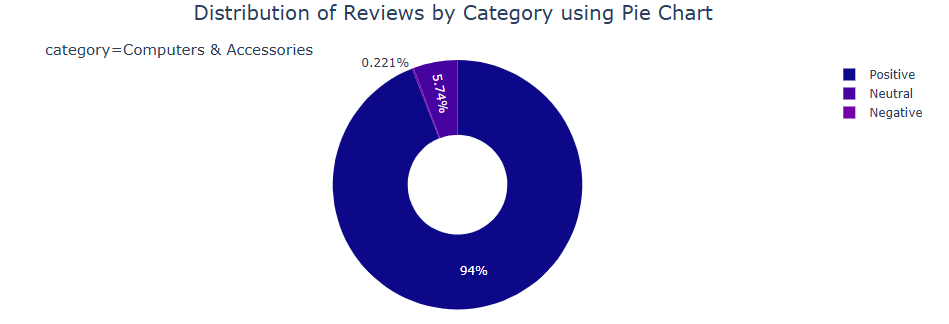
**Sentiment Analysis for Negative Reviews:**



The sentiment analysis for **negative reviews** reveals common complaints like **“bad”, “low”, “slow”, “poor” and “little”** reflecting **dissatisfaction with product quality**. This visualization highlights areas needing improvement to address customer issues effectively.

**Pie Chart:**

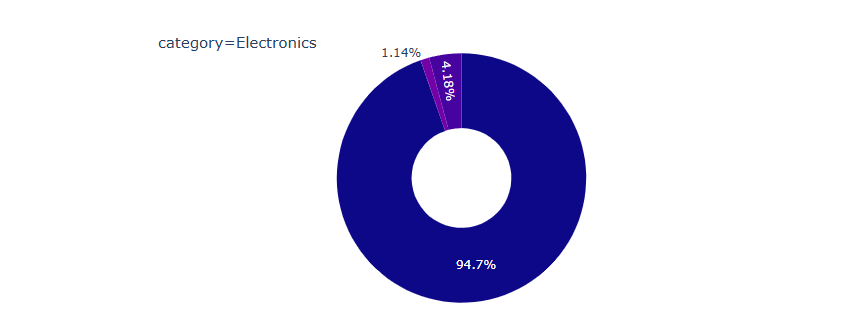
**Distribution of Reviews by Category:**



**Computers & Accessories Category:**

In the Computers & Accessories category:

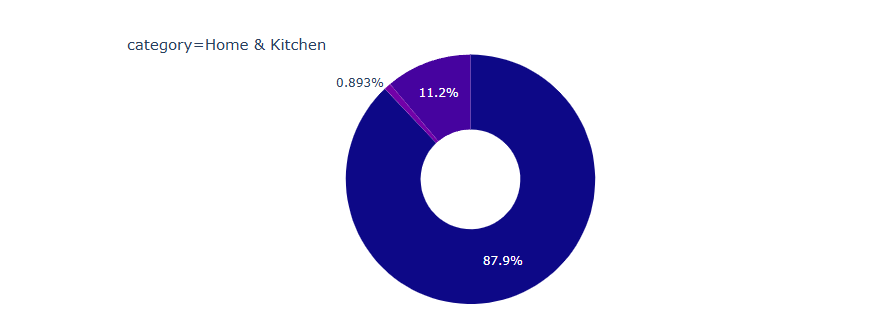
* **Positive Reviews:** Represented the highest proportion of reviews at **94%**, indicating **strong satisfaction levels.**
* **Neutral Reviews:** Accounted for **5.74%** of the reviews, suggesting a **moderate sentiment without strong opinions.**
* **Negative Reviews:** Represented a minimal **0.221%** of the reviews, **indicating few dissatisfaction concerns.**



**Electronics Category:**

In the Electronics category:

* **Positive Reviews:** Represented the highest proportion of reviews at **94.7%**, showing significant **satisfaction among customers.**
* **Neutral Reviews:** Accounted for **4.18%** of the reviews, indicating a **balanced sentiment.**
* **Negative Reviews:** Represented a minimal **1.14%** of the reviews, **suggesting some areas of dissatisfaction.**

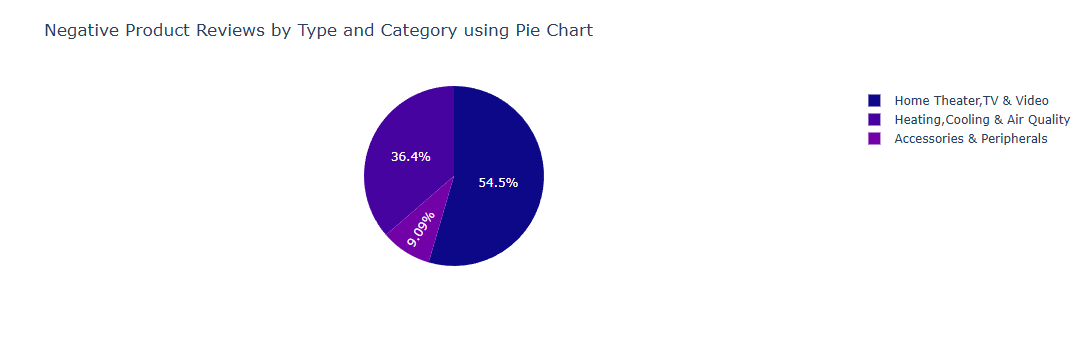


**Home & Kitchen Category:**

In the Home & Kitchen category:

* **Positive Reviews:** Made up **87.9%** of the reviews, indicating overall satisfaction.
* **Neutral Reviews:** Accounted for **11.2%**of the reviews, suggesting a **moderate sentiment.**
* **Negative Reviews:** Represented a minimal **0.893%** of the reviews, **indicating minimal dissatisfaction concerns.**

**Negative Product Reviews by Type and Category:**



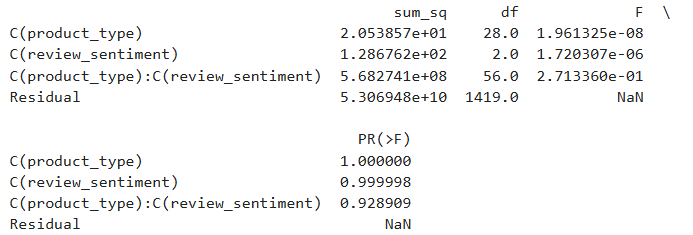
The pie chart visualizes negative reviews across different product types within the specified categories.

* **Home, Theater, TV and Video:** This product type accounted for the highest percentage of negative reviews at **54.5%**, indicating common issues or dissatisfaction with these products.
* **Heating, Cooling & Air Quality:** This product type represented the **36.4%** of negative reviews, suggesting areas where customers experience disappointment or dissatisfaction.
* **Accessories & Peripheral:** This product type represented the remaining **9.09%** of negative reviews, indicating specific areas that may require attention or improvement based on customer feedback.

This visualization helps identify which product types within each category receive the highest proportion of negative feedback, providing insights into potential areas for improvement or customer concerns.

**Statistical Analysis:**

**Two-way ANOVA:**



**Null Hypothesis (Ho)**: There is no significant interaction effect between product type and review sentiment on discounted prices.

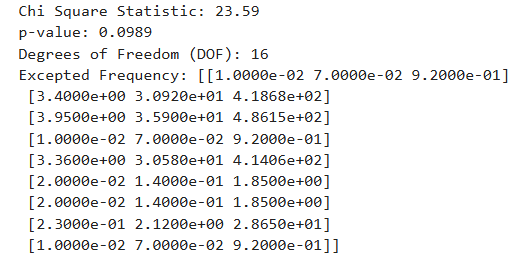
**Alternative Hypothesis (H1):** There is a significant interaction effect between product type and review sentiment on discounted prices.

The **Two-way ANOVA** results indicate that **there are no statistically significant interactions between product types and review sentiments on discounted prices**. Therefore, **fail to reject the null hypothesis.**

* **Product Type**: The variation in discounted prices across different product types is not statistically significant (0.928909 > 0.05).
* **Review Sentiment:** There is no significant difference in discounted prices based on different review sentiments (0.928909 > 0.05).
* **Interaction Effect:** The interaction between product type and review sentiment also shows no significant influence on discounted prices (0.928909 > 0.05).

Overall, neither product type nor review sentiment significantly affects discounted prices individually or in combination.

**Chi Square Test:**

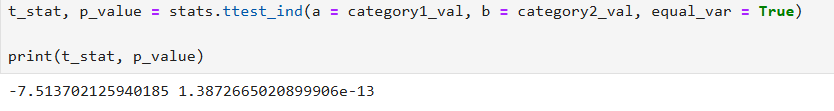
****

**Null Hypothesis (Ho)**: There is no association between the product categories and review sentiments.

**Alternate Hypothesis (H1)**: There is an association between the product categories and review sentiments.

The **Chi-square statistic** calculated was **23.59** with a corresponding **p-value of 0.0989** and **16 degrees of freedom**. This result suggests that **there is no significant association between the categories (product types) and review sentiments (negative, neutral, positive) in the dataset at a 95% confidence level**. Therefore, we **fail to reject the null hypothesis**, indicating that the observed frequencies align closely with the expected frequencies under the assumption of independence between product categories and review sentiments

**Independent t test:**



The **independent t-test** conducted here compares the mean discounted prices between two categories, "Computers & Accessories" and "Home & Kitchen."

**Reject the Null Hypothesis at alpha 0.05.**