Part 1: Analyzing Best-Seller Trends Across Product Categories

Crosstab Analysis:

I first created a crosstab between product categories and the "isBestSeller" status to understand how prevalent best-sellers are within each category. By calculating the proportion of best-sellers for each category, we observed clear trends.

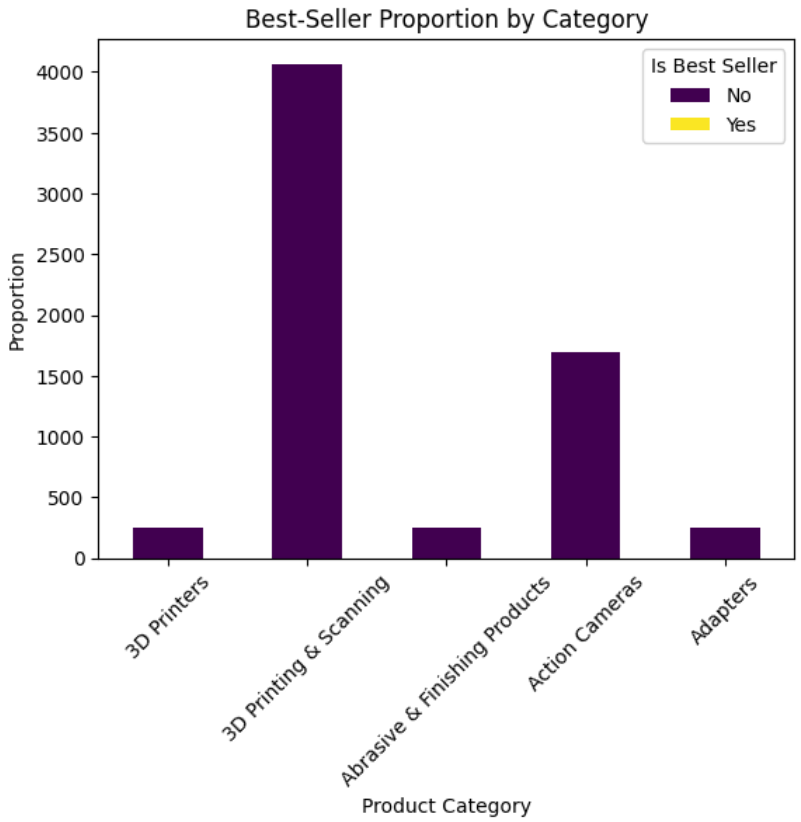
- **Findings:** Categories such as grocery, smart home security & lighting, health & personal care, mobile phone accessories and power & hand tools had notably higher proportions of best-sellers compared to categories like eBook readers & accessories and action cameras.
- **Key Insight:** Categories with larger product ranges tend to have a higher number of best-sellers, likely due to the competitive nature and product volume within these fields.

Statistical Tests:

To confirm the relationship between best-seller status and product category, a **Chi-square test** was conducted. The p-value obtained from the test was higher than 0.05, indicating that there is not enough evidence to conclude a significant **dependency** between product category and best-seller status.

- **Cramér's V:** The calculated Cramér's V value also suggested a **weak strength** of association between best-seller status and product category.

Visualization:



A stacked bar chart was used to visualize the distribution of best-sellers across different categories. The chart revealed that product categories like Arts & Crafts and Baby had a higher proportion of best-sellers, while 3D Printing & Scanning and action cameras categories exhibited lower proportions.

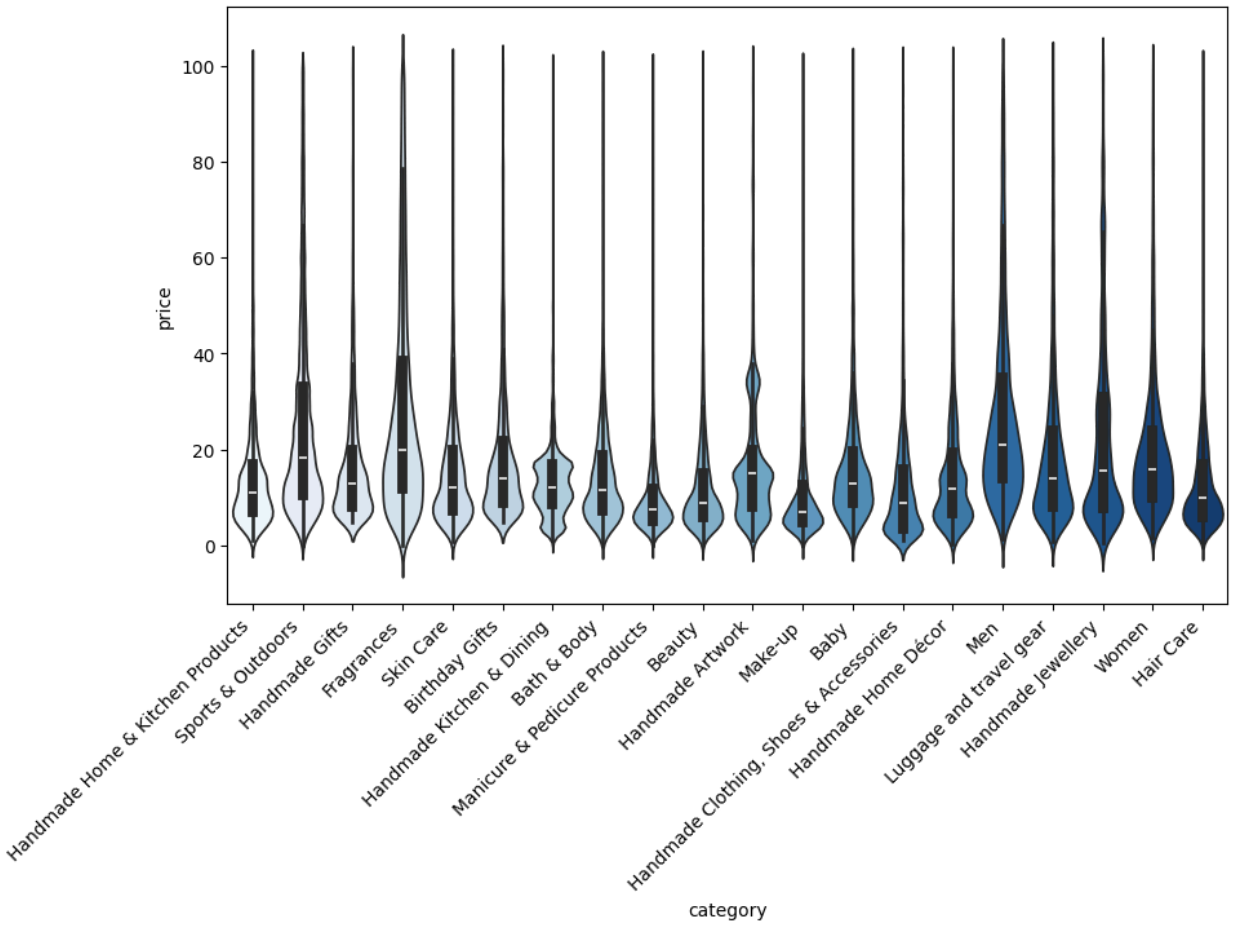
Part 2: Exploring Product Prices and Ratings Across Categories and Brands

Removing Outliers in Product Prices:

Outliers in product prices were removed using the **Interquartile Range (IQR)** method. Products priced below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR were removed from the dataset, ensuring a cleaner analysis.

Violin Plots:

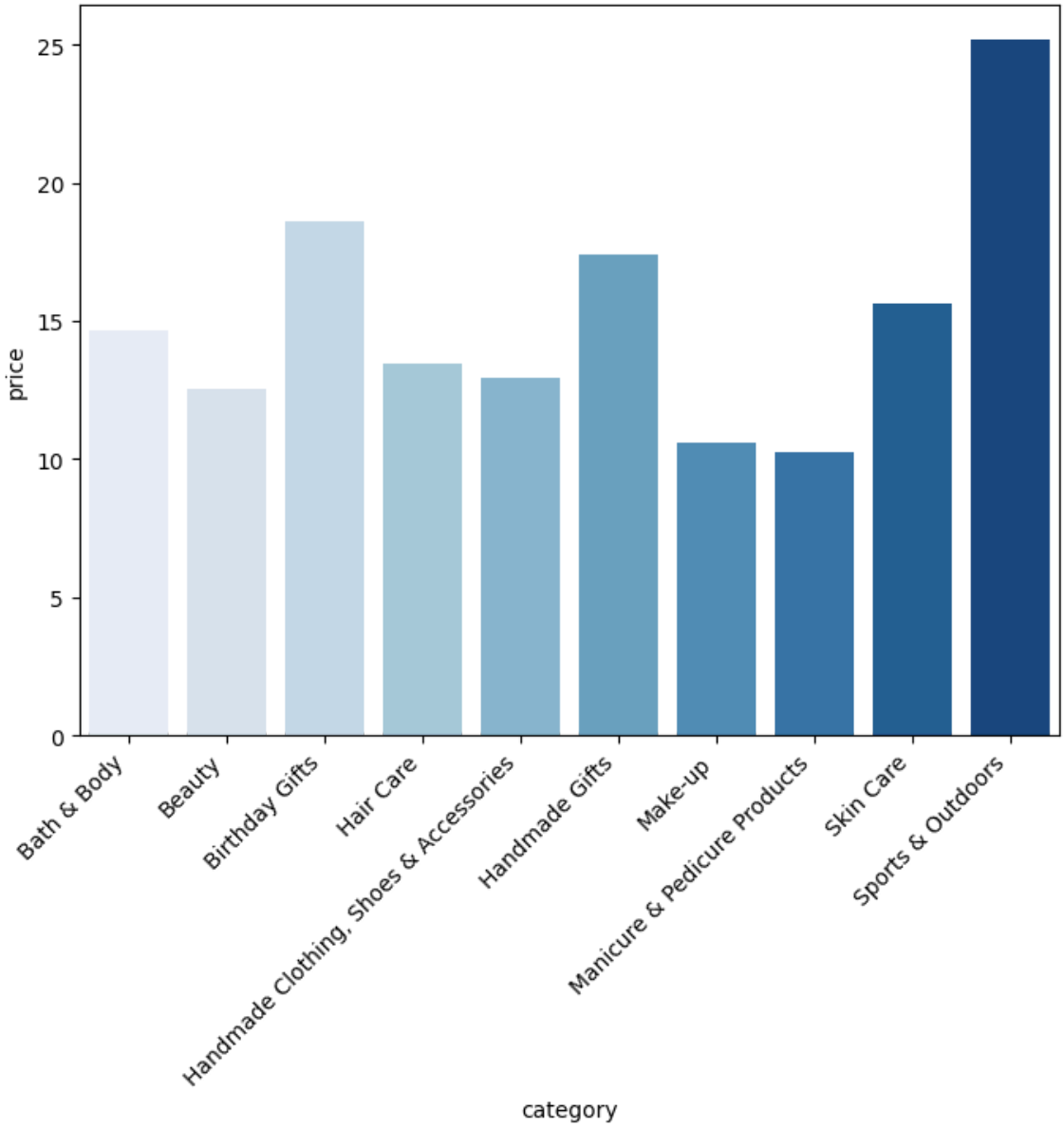
A violin plot was generated to visualize the price distribution across different product categories.



- **Findings:** Some categories, like **Fragrances** and **Man**, showed wide price ranges.
- **Key Insight:** While the price distribution in most categories was skewed, Desktop PCs had the highest **median price** compared to other categories.

Bar Charts:

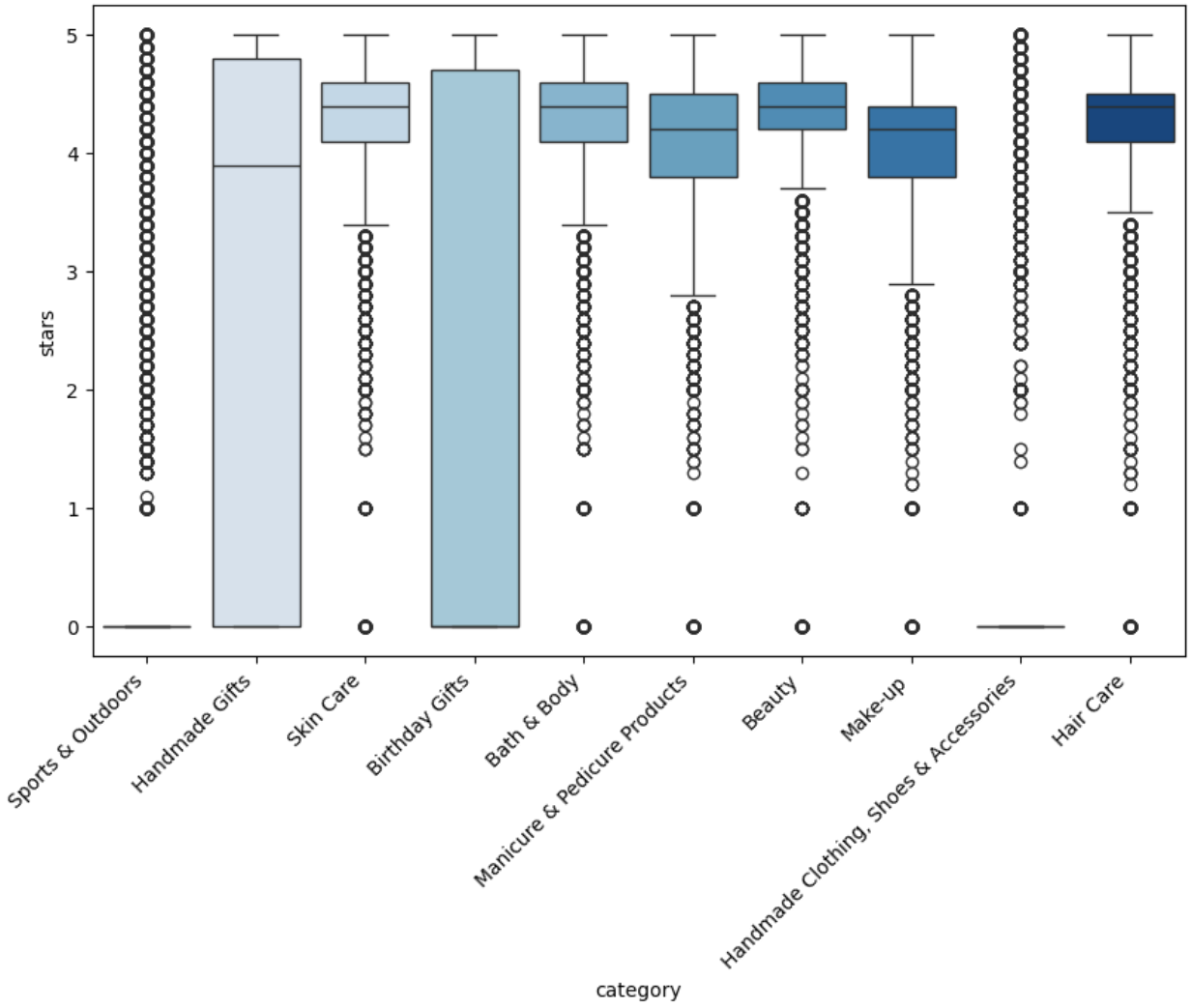
A bar chart was created to compare the average price of products across the top 10 product categories based on product count.



- **Findings:** The **eBooks Readers & Accessories** category had the highest average price.
- **Key Insight: eBooks Readers & Accessories** products have the highest average price, reflecting the high investment and demand for tech products on Amazon UK.

Box Plots:

Side-by-side box plots were used to visualize the distribution of product ratings across different categories.



- **Findings:** Categories like **Skin Care** and **Bath & Body** had a relatively higher median rating compared to others like **Handmade Gifts** and **Sports & Outdoors**.
- **Key Insight: eBooks Readers & Accessories** generally receive the highest median ratings, suggesting that customer satisfaction with this product tends to be higher than other product categories on Amazon UK.

Part 3: Investigating the Interplay Between Product Prices and Ratings

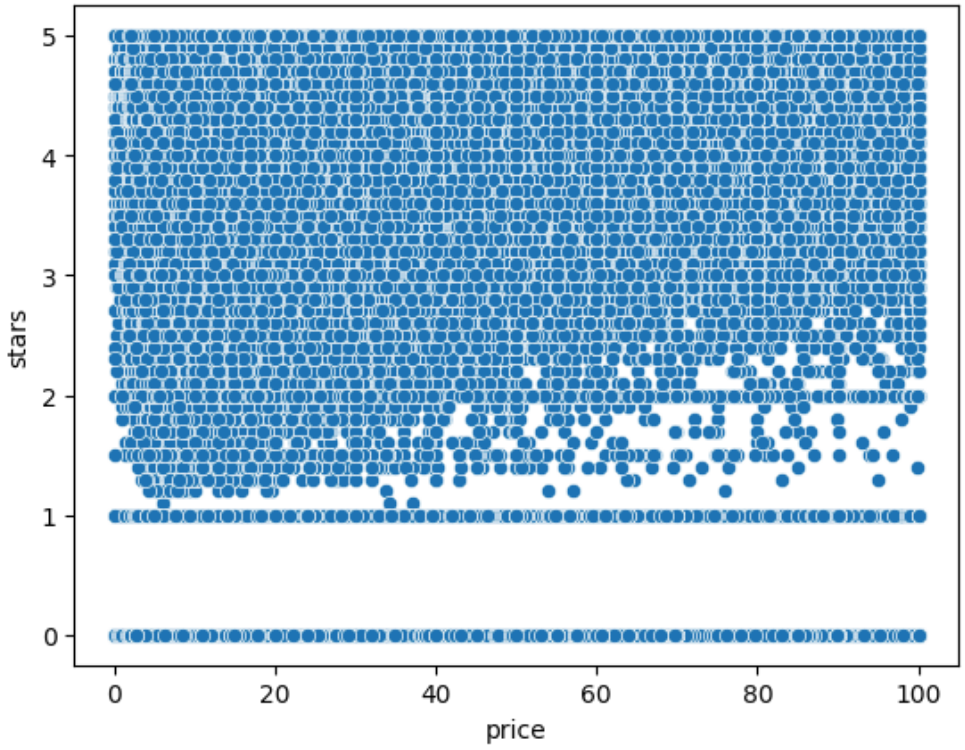
Correlation Coefficients:

The **Pearson correlation coefficient** between product price and rating was calculated, resulting in a very **weak negative correlation**. This indicates that higher-priced products on Amazon UK do not necessarily receive higher ratings.

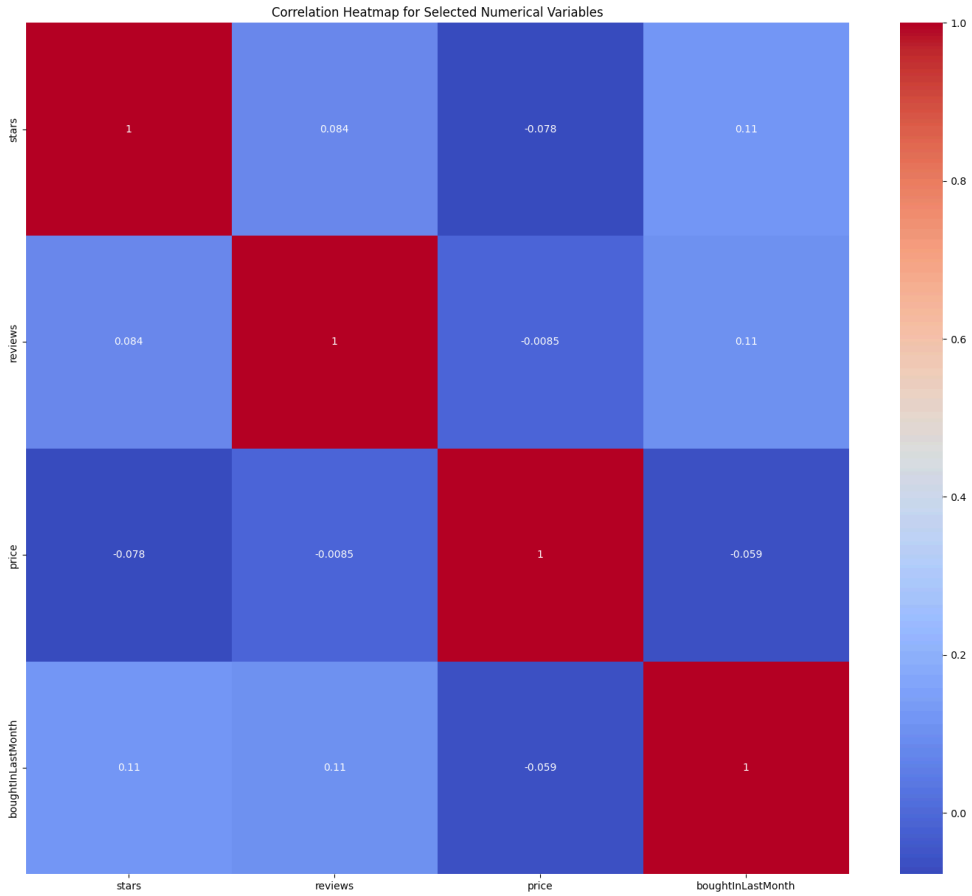
- **Key Insight:** Price has little to no impact on customer ratings. This suggests that factors other than price (such as quality, brand, or customer service) play a more significant role in determining product ratings.

Visualizations:

A scatter plot was used to visualize the relationship between product rating and price. The plot revealed no clear linear relationship, further supporting the weak correlation between these two variables.

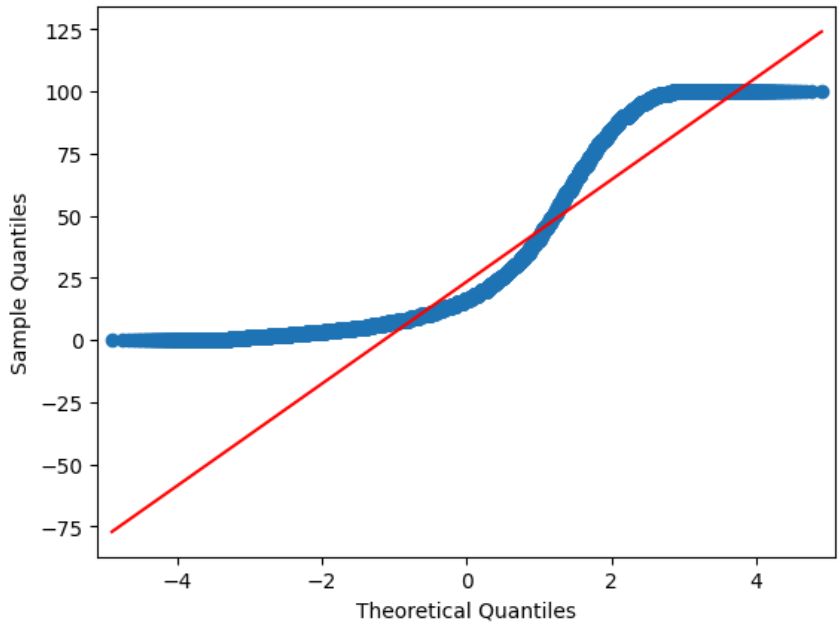


Additionally, a **correlation heatmap** was generated to visualize correlations between all numerical variables. It showed that there were no significant correlations between price and ratings, reinforcing our earlier findings.



A **QQ plot** was also used to assess the normality of the product price distribution, which revealed heavier product tails

.



Bonus: Analysis Without Removing Outliers

Upon repeating the analysis without removing outliers, the results showed that outliers in price could significantly impact the interpretation of price distribution. For instance, categories with extreme price outliers, such as **Sports & Outdoors**, showed a much higher median price when outliers were included.

- **Key Insight:** Excluding outliers provides a more accurate representation of the central tendency and variability in product prices. Including them can distort average price trends and make it harder to interpret patterns.

Summary of Insights:

1. **Best-Seller Status:** Product categories such as Grocery and Health & Personal Care are more likely to have best-sellers. The analysis suggests that competition and product volume are key factors driving best-seller status in these categories.

2. **Price Distribution:** Desktop PCs have the highest median prices, while categories like Manicure and Make-up are typically more affordable. The IQR method for removing price outliers helped better visualize these trends.

3. **Ratings:** Products in categories like eBooks Readers and Accessories tend to receive higher ratings, indicating customer satisfaction might be higher in these categories compared to others.

4. **Price vs Rating Correlation:** The weak correlation between price and ratings suggests that Amazon UK customers' satisfaction with products is not strongly influenced by their price.

5. **Outlier Impact:** Removing outliers leads to a more accurate representation of product price distributions, particularly in categories with extreme price values, like Sports & Outdoors.