Team 1

Dr. Schubert

DASC 1223

26 February 2024

Amazon E-commerce Analysis: Optimizing Sales & Productivity

***Introduction***

E-commerce is a very important and rapidly growing industry. In the past, e-commerce, like any new and emerging concept, was slow to be viewed as critically important to the daily and long-term operations of any business. As time has passed and business began realizing and acknowledging the importance of technology and the digital era we’ve been transitioning into, e-commerce has shown significant signs of growing and soon enough it will dominate the delivery of products to customers globally. E-commerce is no longer an option for businesses, it’s now a *necessity*. If a business is planning to be profitable and run a reputable/well-known business, they must be active and innovative in the e-commerce sector.

The COVID-19 pandemic really highlighted how critical a role e-commerce could and will play in the economy. The ability of not being able to freely and physically go to retailers or businesses that supplied consumers with products, almost forced us as a society to rely on e-commerce. Even if you could go to the store, why would you? After becoming a safety precaution, it became a convenience matter. Small businesses that survived the pandemic and thrived during the pandemic transitioned to becoming more e-commerce dependent. Personally, I saw this firsthand, I worked for a local sporting goods store in my hometown. We weren’t in complete quarantine in the country I was in, but small businesses still faced financial hardship because the public was personally choosing to stay home. The business reduced the number of employees and started shipping orders through Facebook orders and orders placed through Instagram DM’s. This is just one example of how e-commerce can be beneficial for businesses, no matter how small.

One of the leading e-commerce companies, Amazon, has taken this industry space by storm. They are by far the leading company in this sector. There are many components that make Amazon the leading entity in this industry, its sheer size as a company is intimidating. Amazon is so big and that comes with one advantage; the amount of data they have access to. E-commerce is erupting with data, number of sales, product name, price, star rating, reviews, and the list goes on and on. As I have mentioned before, we are entering a new era of technology and one that is primarily data driven. Business in e-commerce is placing more importance and, most importantly, resources to big data analytics to gain insights to primarily minimize cost and improve profitability.

Our team combed through Amazon’s best sellers and the data associated with those products to develop insights that might be useful to businesses in the e-commerce market or Amazon itself. We chose to analyze the data associated with these products, because they are among the top sellers in the biggest and most important e-commerce platform available. There must be a reason why these products sell the most among the thousands of products on Amazon. We asked ourselves what were those reasons/factors that differentiated these products and looked to quantify those factors and statistically prove them. The goal of this analysis is to understand pricing trends and patterns from this e-commerce data set and provide data-driven pricing recommendations.

As a team we first we decided to possibly identify if any factors directly affected the price of a product. We’re aware that ratings and reviews are very important in Amazon and the community in Amazon is very active, so we decided to test if these two factors have a relationship to the change in price. Then, we analyzed the price distribution and department with the goal of finding insights into whether certain departments sell more products that are relatively more expensive or vice versa. Lastly, we sought to see if brand popularity was a factor in products being the top sellers on Amazon.

This analysis was very extensive and grueling since most e-commerce sales data is not publicly available, but after meticulously web scraping Amazon’s top sellers by department the team was able to derive more than one insight from the data and was able to provide data-driven pricing recommendations.

***Executive Summary***

Web scraping was used to create this data set of 2171 of Amazon’s Top Sellers. The Beautiful Soup library was used to parse through HTML documents, Selenium was used for browser automation, and WebDriver was used to interact with web elements. The dataset has 2171 rows and the main variables (columns) used for this analysis include general data about these products like product name, price, color, brand, and the department it was classified into. More in depth data was gathered including rank within department, star rating, and number of ratings.

Before we started this analysis, we as a group theorized that certain factors affected the price of a product and that was the reasoning behind most of the prices we see among the best-selling products on Amazon. We first hypothesized that star rating and number of ratings had a correlation with price and maybe there was some basis in assuming that there was some form of causation there. We also focused on brand and popularity. Our though process was that brands that were more popular and recognized would sell a lot better than other small name brands. What really helped us was visualizing the price distributions between departments and visualizing what price ranges were more common.

An ordinary least squares (OLS) regression test was conducted to determine if there is a statistical correlation between independent variables of star rating and number of ratings and the dependent variable of the price of the product. In essence, we were testing to determine if star ratings and number of ratings positively or negatively affected price. After the test, the results indicated that these two variables don’t affect the price of the product, at least statistically speaking. So, we concluded that since star rating and number of ratings don’t have a statistical relationship with price, businesses shouldn’t solely rely on star ratings or number of ratings to justify pricing and should instead focus on highlighting the features and quality of the products. This led us to explore other features of products and we began to analyze products by department. We also believe that the brand of the product has a relationship with the price of the product and maybe looking at brand popularity will lead us to interesting insights.

We divided all the products into eight prices ranges and analyzed the distribution of the histogram that visualized the count of products in each price range. Then, we went a step further and analyzed frequency by department in each of the price ranges. We observed that each department showed varying price preferences, so we recommend that when looking to optimize sales, analyze the most current distribution of that department’s sales and price ranges.

We also recommend that businesses should also leverage brand recognition to maximize sales. We formulated this recommendation by analyzing brand popularity by department and tried identifying patterns or trends. It was observed by one of our group members that in each department there was brand that was more prevalent, and it was typically a recognized price. As a business if you highlight the reputation and reliability of established brands you are bound to boost sales.

This analysis was very fruitful and shed light on some of the factors that determine a good price of a product when it comes to e-commerce. Good prices are comparable to other prices in a specific department, competitive with potential substitutes product price, and in some cases, products should be at a reasonably low price, as customers prefer cheaper products in certain departments. There is growth and expansion when it comes to the continuation of this project. The more data that could be extracted, the more data that is available and the more reliable the results are, as well as more tests or factors that can come into play. Another thing to note would be that it would be of interest to redirect our analysis for small to mid-size ecommerce businesses. That part of the industry is different to Amazon and there might be some insights to derive from there.

***Regression Analysis***

To test if there was any strong correlation between the price and star rating, as well as number of ratings an ordinary least squares regression analysis was conducted. By doing this type of statistical analysis, we wanted to gain some insights into the relationship between the dependent variable (price in dollars) and the two independent variables (number of ratings and star ratings). Below are the calculations for the regression analysis of all the products in the dataset.

A screenshot of a computer screen

Description automatically generated

Fig. 1. OLS regression analysis results. Dependent value (price in dollars) and independent values (star ratings & number of ratings).

To summarize the data in [Fig. 1](#Bookmark8), we first need to look at the R-squared value. The r-squared value is 0.037, this means that only 3.7% of the variance in the price can be explained by the number of ratings and star ratings. This very low R-squared value indicates that the amount a product cost can’t really be explained by the star ratings or number of ratings, at least statistically speaking. If we examine this a little deeper, we can see that the other values point towards this same conclusion. If we look specifically at the star rating coefficient, we can see that it is a 60.1175, meaning that (holding all other variables constant) for each additional unit increase in the star rating, the price is expected to increase by approximately $60.12. For number of ratings the coefficient is -0.0887, so if the number of ratings increase by one unit, the price decreases by 0.0887. The relationship between star rating and price seems promising and possibly an insight to how increasing star rating may lead to a higher price, but for both factors the associated p-value is extremely higher than 0.05 (conventional level of significance). This means that both coefficients are not statistically significant and there is no correlation between price or star rating/number of ratings.

***Price Distribution***

The distribution of prices among multiple factors was a big key in forming our insights and examining trends in e-commerce pricing. Since our data contained no time-series variables, it was hard to effectively look at and predict trends in prices. Different price distributions helped to answer the question of how items are priced online currently and to understand potential factors that affect a product's price more heavily.

To begin analyzing price distributions, I first looked at the distribution of all products within $10 increment price bins (until $50, where I then went by $25 increments until $100+). This first distribution ([Fig. 2](#Bookmark1)) provided a broad overview of the products in our data and how they were priced.

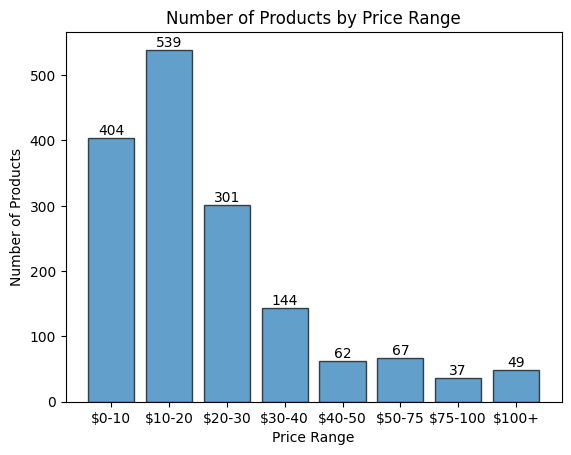


Fig. 2. Distribution of all products by price bins.

As shown in this first figure, most of the products were under $30. This shows that many of the top products bought by Amazon customers tend to be less expensive. This likely is due to many of the departments on Amazon such as *Industrial & Scientific*, *Pet Supplies*, and *Arts, Crafts & Sewing* having many low-cost items that are top sellers. However, since our data had roughly the same number of products per department, this led me to be curious about how the top bins ($0-10, $10-20, & $20-30) were distributed by department. Thus, I created three more distributions for these top bins to see how different departments broke down by price range.

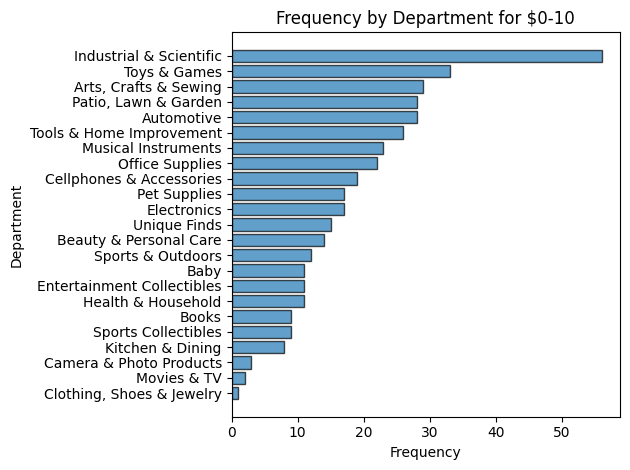


Fig. 3. Distribution of the frequency of products by department in the $0-10 price range.

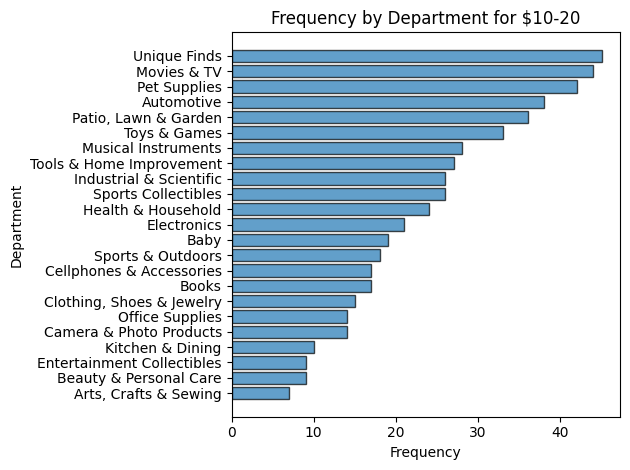


Fig. 4. Distribution of the frequency of products by department in the $10-20 price range.

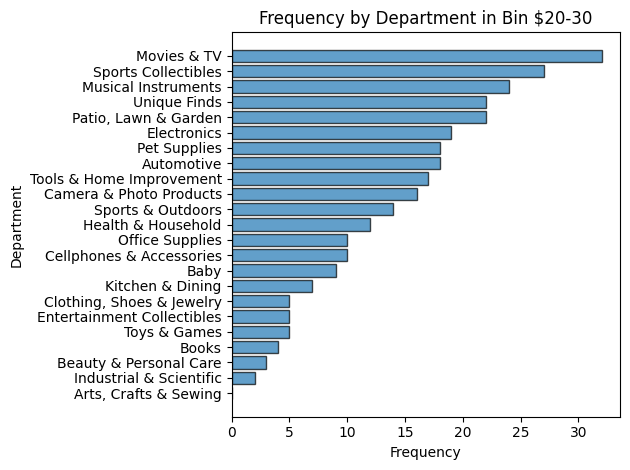


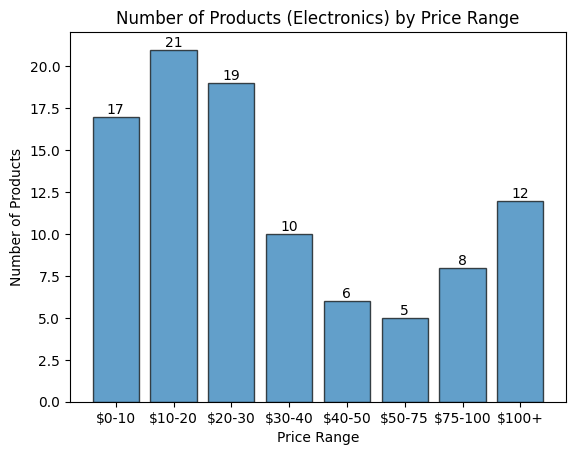
Fig. 5. Distribution of the frequency of products by department in the $20-30 price range.

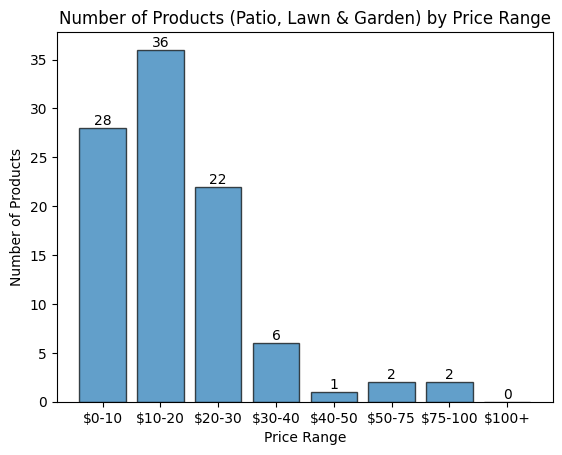
These three figures showed some interesting findings in terms of how the different department’s top products are priced. As expected in the $0-10 range ([Fig. 3](#Bookmark2)), the top departments tended to be ones of cheaper products overall, with *Industrial & Scientific*, *Toys & Game*s, and *Arts, Crafts & Sewing* making the top three. *Electronics* landed in the middle surprisingly, likely due to products like chargers and headphone jacks. *Camera & Photo*, *Movies & TV*, and *Clothing, Shoes & Jewelry* took the bottom three, which made a lot of sense considering their products tend to be more expensive.

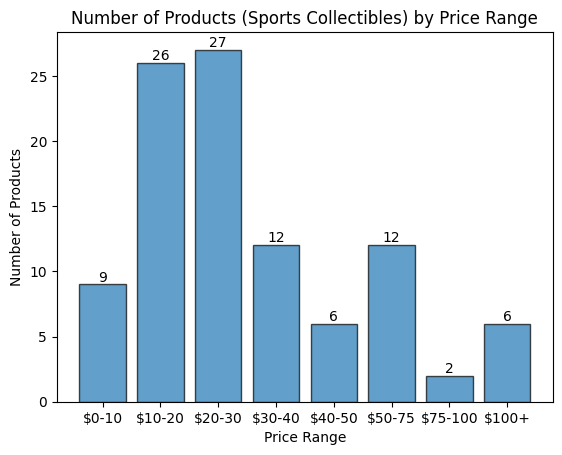
For the $10-20 price range ([Fig. 4](#Bookmark3)), *Arts, Crafts & Sewing* immediately fell to last place, which I found interesting as it was the top department in $0-10. *Unique Finds* had the highest frequency, which I also found interesting considering products in this department tend to be more random and niche, which I thought would lead to many varying prices. However, referring to [Fig. 2](#Bookmark1), the $10-20 range had the highest frequency of products overall, which is shown through a tighter and more linear distribution and no clear dominant department.

Lastly, the $20-30 range ([Fig. 5](#Bookmark4)) showed a more similar distribution to that of the $0-10 range, with clearer top departments compared to the $10-20 range. *Movies & TV*, *Sports Collectibles*, and *Musical Instruments* formed a solid top three, while two of the top products in the $0-10 range (*Industrial & Scientific* and *Arts & Crafts*) barely combined for any products in the distribution. These three distributions clearly show that the pricing within a department contributes to a product's price as it is difficult to be competitive in a market for similar products when competitors offer goods for better prices. This leads to markets plateauing around particular prices for products within the same department as customers aren’t willing to spend more on a product if it is roughly the same. However, do all departments follow this trend or do different departments behave differently? This led me to dig deeper into the price distributions by seeing how specific different departments price their products.

To begin diving into specific departments, I felt that selecting three vastly different departments would effectively answer this question, as looking at different departments with known pricing tendencies likely would not provide any helpful insight. The three departments I chose for this were *Electronics*, *Patio, Lawn & Garden*, and *Sports Collectibles*. I first chose *Electronics* as it is a popular department and had the highest sum of ratings overall. I then selected *Patio, Lawn & Garden* since it was the only department in the top five of each of the $0-10, $10-20, and $20-30 ranges. Lastly, I selected *Sports Collectibles* since it is a more niche department which I expected to have a variety of different price ranges.

  
Fig. 6. Frequency of Products by price range for the Electronics department.

  
Fig. 7. Frequency of Products by price range for the Patio, Lawn & Garden department.

  
Fig. 8. Frequency of Products by price range for the Sports Collectibles department.

*Electronics* (Fig. 6) interestingly had most of its products in the cheaper (less than $30) range*.* However, it then showed a large dip and then increased after $75, accounting for the more expensive products in the department. Patio*, Lawn & Garden* showed that it had hardly any products over $30, explaining why it was a top five department for each of the top, lower price ranges. *Sports Collectibles* behaved very similarly to how I expected, with an erratic up and down distribution, but with a notable peak around $10-30. These distributions display how different departments in fact behave differently in terms of pricing distribution, so the kind of product being sold and the department it is categorized in will influence the price. In a category such as *Sports Collectibles*, sellers have much more room to upcharge for a product they believe is worth more as there isn’t a clear market standard for price. It would be quite difficult on the contrary to sell a top product at a price higher than $30 in the *Patio, Lawn & Garden* department. In *Electronics*, it matters more about the individual product, as products have different market standards, distributing the department prices around those standard prices.

***Brand Analysis***

Brands are an important aspect that can determine what a person buys. Even outside of e-commerce, brand images and popularity can impact sales heavily. Wendy’s recently announced the implementation of surge pricing, and received heavy backlash from the people until it was reversed.

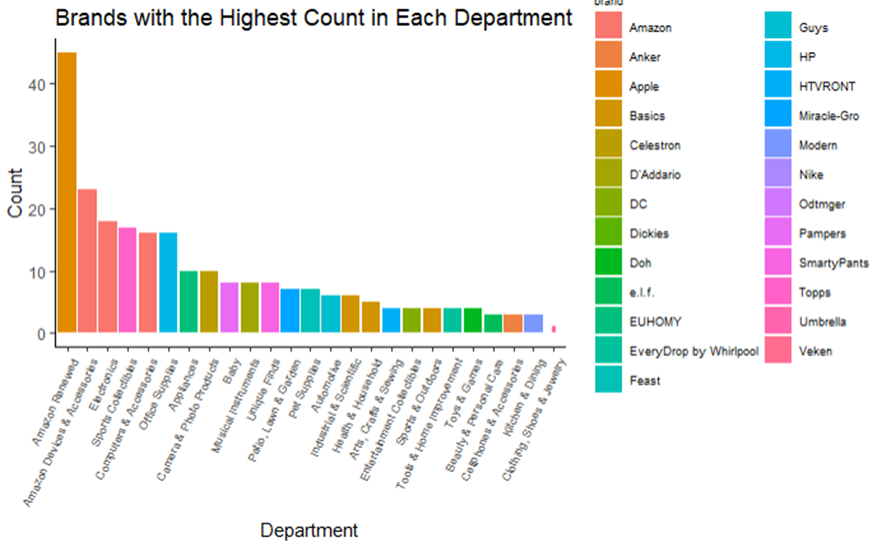


Fig. 9. Top Brand per Department.

This initial graph ([Fig. 9.](#Bookmark9)) was to find which brand has the most products in each department. There are three departments with the top brand of Amazon, and another three with the top brand of Amazon Basics. One of the departments where Amazon products lead, Electronics, is a category that tends to have more expensive items within. Throughout the departments, and noticeable on the legend, are mostly well-known brands. This makes sense, as people trust the brands they know so having more of their products is a good idea. Another thing of note is that there is a high amount of Apple products in the Amazon Renewed department. This was because most of the products were from the various iPhone generations.

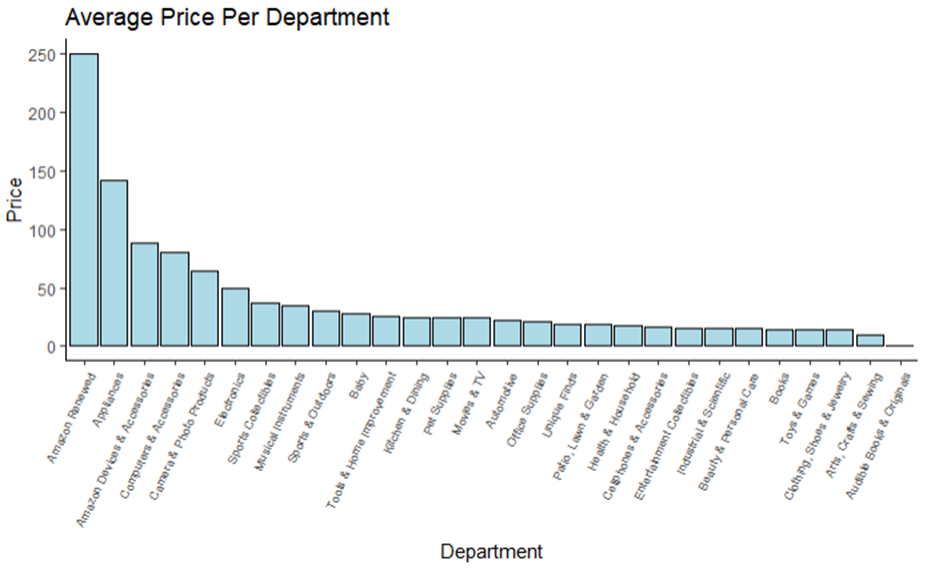


Fig. 10. Average Price Per Department.

The next graph ([Fig. 10.](#Bookmark10)) investigates the average price per item in each department. Unsurprisingly, Amazon Renewed has the highest average price because of the influence of Apple products. The next departments, in order, are Appliances, Amazon Devices & Accessories, Computers & Accessories, and Electronics. These are all products which tend to have the highest prices on Amazon. However, from the graph, it seems that these departments are lower than what someone would think. Computers & Accessories ([Fig. 12](#Bookmark12)), and Electronics ([Fig. 11](#Bookmark11)), are generally expensive items, especially when you look on Amazon.

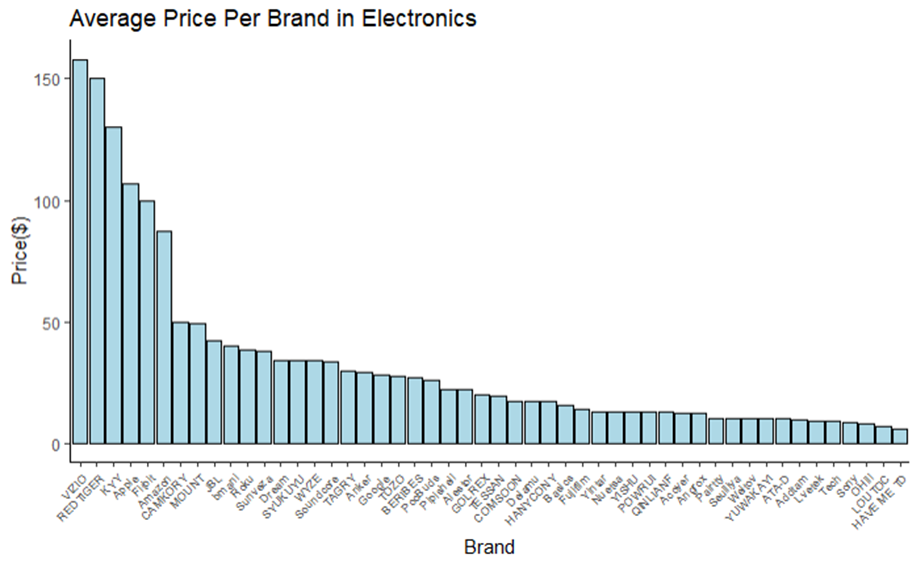


Fig. 11. Average Price by Brand in the Electronics Department.

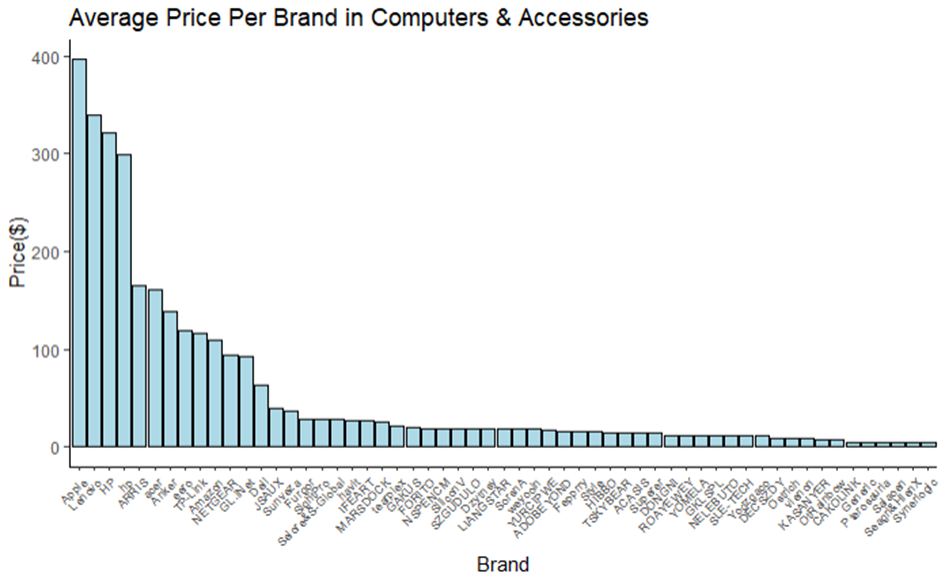


Fig. 12. Average Price by Brand in the Computers & Accessories Department.

Looking specifically at the average prices in brands within Electronics and Amazon Devices & Accessories, we see that the highest is Apple in Computers, at almost $400. Several of the top and well-known brands in both departments also have a high average price. But what brings the average price so low on the department scale are the lesser-known brands with one or two low-cost products. From this, we can see that well-known brands will sell for higher prices. Note that in Computers & Accessories, Amazon, which has the most products, has the 9th highest average price in the department. The same can be said in Electronics, where it is 6th. This means that the number of products you have per department does not necessarily mean you have the highest prices, but that isn't always what people are looking for anyway.

***Insights & Recommendations***

Based on our findings, there were three key insights/recommendations we would make to companies looking to make their product competitive with the other top products on Amazon. First, companies should tailor their prices to the department in which their product is sold in. By analyzing how the products in a particular department are priced, companies can know the prices that customers are already paying for similar products to make their price around the same range. Next, we recommend customers focus on value proposition. Obviously, it is important to have good ratings and stars on products, but past a certain threshold, they don’t make a huge difference as opposed to the value of the product and price that it is offered. These things are much more important and should require more intention and focus. Lastly, companies should leverage brand recognition. Having a brand that is recognizable and well-known helps make a big difference and companies should try to create and maintain a good reputation. This will help boost a product onto the top products page in general and will boost sales overall as more of a following is created for the brand.

Works Cited

“Amazon.Com Best Sellers: The Most Popular Items on Amazon.” *Amazon.Com*, www.amazon.com/gp/bestsellers/?ref\_=nav\_cs\_bestsellers. Accessed 19 Feb. 2024.