**Amazon Fashion Products**

Lachandra Ash

November 23, 2022

**Topic**

Amazon's reputation for offering highly reliable product recommendations online has earned the company much praise. Artificial intelligence algorithms and machine learning are used to enhance the shopping experience and boost revenue. When a consumer visits Amazon, they will be presented with a list of suggested items based on Amazon's sophisticated analysis and prediction of the user's buying interests (Krysik, 2021).

**Business Problem**

Amazon needs dependable product recommendations for their consumers. Amazon would like to boost their profits from their products. Amazon wants their reviews, price, and ratings to be analyzed.

**Background History**

Jeff Bezos founded Amazon in 1994 (Krysik, 2021). Since its debut one month ago, the firm has made a weekly net profit of $20,000 on book sales to clients in all 50 states and 45 countries across the globe (Krysik, 2021). In the 25 years since its founding, Amazon has become the dominant eCommerce platform, with 2019 expected to bring in a net profit of up to $ 280.5 billion (Krysik, 2021). Amazon launched their Polish online shop in 2021 (Krysik, 2021). Amazon's worldwide success may be attributed in large part to the company's unique business strategy and well-considered management, but also to the platform's ongoing iteration and the introduction of fresh ideas.

**Data Explanation**

The amazon products dataset was compiled using PromptCloud's proprietary web-crawling service. The dataset includes ten thousand products from Amazon. There are seventeen rows within the dataset. The seventeen features are:

* Product name
* Manufacturer
* Price
* Number available in stock
* Number of reviews
* Number of answered questions
* Average review rating
* Amazon category and sub-category
* Customers who brought this item also brought
* Description
* Product\_information
* Product\_description
* Items customers buy after viewing this item
* Customer questions and answers
* Customer reviews
* Sellers (Data.World, 2022).

**Methods**

I imported the libraries and modules needed for the project. The amazon dataset was uploaded into a pandas dataframe. After I displayed the first five records, I cleaned and processed the data using isna, fillna, dropping columns with seventy percent missing data, checked for number of products available in stock, created new columns, duplicated, and replace methods. I dropped and renamed the columns.

I converted the euro pound sign into usd currency, for the price feature. I converted the integers within the number available in stock column into float. I split the category into three sub-categories and created the sub-category columns. I changed the average review rating column into a float type. I checked the dataframe for non-numeric values. I used the isnull methods to seek missing prices within the dataframe and used dropna method to drop price column and create price new column.

During exploratory data analysis, I viewed the dataframe’s shape, description, info, correlation, covariance, revealed the columns, checked dtypes, sum, and unique features. I examined the manufactures and their prices, revealed the maximum price row details , displayed the minimum price row details, created dataframes with top ten manufacturers, displayed the manufacturers average rating, revealed the various present main categories, chose five popular main categories, created a new dataframe with five most popular main categories, and created a popular sub-category table.

**View Appendix, Table 1: Popular Sub-Category Table**

I defined the sub-values and sub-counts for two sub-categories, created dataframes from each sub-category, reviewed the products average ratings and new average price, checked the new datframe for new and used items, used mean method for average review ratings, reviewed price new feature’s statistics, used quartiles to display outliers, defined outlier prices, checked outlier value, and printed categories and sub-categories with outliers.

**Data Analysis**

1)Category of Popular Manufacturers: The oxford diecast manufacturer is the most popular manufacturer. Gorgi, Hasbro, and Mattel are the least popular manufacturers.

Chart, bar chart

Description automatically generated

2)Main Categories: The characters and brands had the highest number of products. The party supplies, electronic toys, arts & crafts, sports toys & outdoor, cooking & dinning, and games had the least number of products.

**Chart, histogram

Description automatically generated**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

3)Data Price Distribution: The data price distribution plot is right skewed and there are outliers within the plot.

Chart, histogram

Description automatically generated

4)Price Distribution: The boxplot was created to check for outliers within the products’ price. There are 5.95 outliers in quartile 1, 10.37 outliers in quartile 2, and 20.12 outliers in quartile 3.

**Chart, box and whisker chart

Description automatically generated**

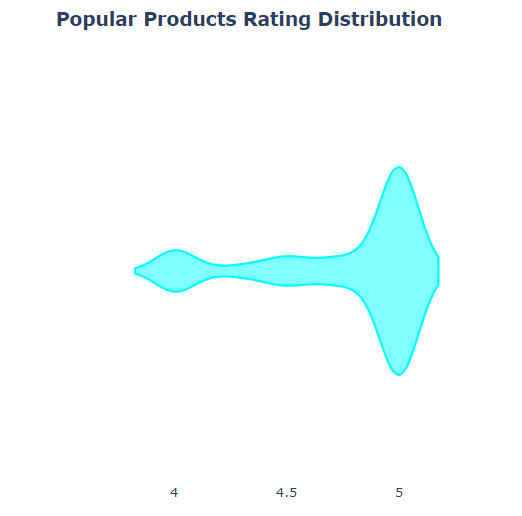
**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

5)Outlier Distribution: The outlier distribution pie chart displays 93.5% of normal outlier price, and 6.55% outliers within the outlier price.

Chart, pie chart

Description automatically generated

6)Popular Products Rating Distribution: The most popular products had 5.2 ratings, and the least popular products had 3.8 ratings.



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

7)Product Reviews Distribution: One product received the greatest number of reviews. Most products received the least number of reviews.

Graphical user interface, application

Description automatically generated

8)Reviews Vs. Products Available in Stock: The highest number of products available in stock has thirty reviews. The highest number of reviews has nineteen products available in stock. The least number of products available in stock has one review.

Chart, scatter chart

Description automatically generated

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

9)Products Available in Stock: Two to three products are mostly in stock.

**Chart

Description automatically generated with low confidence**

**Conclusion**

The products that were priced below fifty dollars were popular. The Disney and Lego manufacturers contained outlier price. The character and brand sub-category have outlier price. The used items were within the outlier price. The 5-star rating range contained the highest number of popular brands. Each product has reviews. There was only one product with the highest number of reviews. The highest number of products had the least number of reviews.

**References**

Data.World (2022). Fashion Products on Amazon.com. Retrieved from [Fashion products on Amazon.com - dataset by promptcloud | data.world](https://data.world/promptcloud/fashion-products-on-amazon-com), on November 23, 2022.

Krysik, A. (2021). eCommerce Giant Work. Retrieved from [Amazon's Product Recommendation System In 2021: How Does The Algorithm Of The eCommerce Giant Work? - Recostream](https://recostream.com/blog/amazon-recommendation-system), on November 23, 2022.

**Appendix:**

**Table 1: Popular Sub-Category Table**

Graphical user interface

Description automatically generated with medium confidence