**Assignment Overview:**

1) Analysis Based on the product reviews:

a) Promotion Focus: Identifying the product that should be promoted more based on the reviews.

b) Distress Identification: Determining which category is causing the most distress among consumers.

c) Pricing and Review Impact: Examining if there is a correlation between pricing and product reviews and sharing insights.

2) Buyer's Perspective:

Imagine you are the buyer for these categories. Recommend products to buy, remove, or manage based on the data analysis.

3) Marketing Insights:

Identify additional insights from the dataset valuable from a marketing perspective.

**1) Analysis Based on the product reviews:**

**a) Promotion Focus: Identifying the product that should be promoted more based on the reviews.**

To identify products that would require marketing/promotions, I've developed two logics that generally hold true in most cases

Logic 1: Scoring Products for Promotion:

Here, I've devised a scoring system based on two significant factors:

        1- High Percentage of Positive Ratings:

        A product with a high '% of Good Ratings' [(Count of 5-star + Count of 4-star) / (Num Of Reviews)] indicates that the product is of high quality and holds substantial potential.

        2- Low Number of Reviews:

        Lower counts of 'Num Of Reviews' suggest that the product hasn't gained much popularity, possibly due to an absence of an effective marketing campaign.

High Percentage of Positive Ratings: Calculates the '% of Good Ratings' using the formula: ((Count of 5-star + Count of 4-star) / Num Of Reviews) \* 100 Assigns a score based on the percentage of positive ratings:

Rules:

* If % of Good Ratings ≥ 90%, assign a score of 5.
* If 70% ≤ % of Good Ratings < 90%, assign a score of 4.
* If 50% ≤ % of Good Ratings < 70%, assign a score of 3.
* If 30% ≤ % of Good Ratings < 50%, assign a score of 2.
* If 10% ≤ % of Good Ratings < 30%, assign a score of 1.
* Otherwise, assign a score of 0.

Low Number of Reviews: Based on quantiles of the 'Num Of Reviews':

* If the number of reviews is in the top quartile, assign a score of 2.
* If the number of reviews is in the second quartile, assign a score of 3.
* If the number of reviews is in the third quartile, assign a score of 4.
* Otherwise, assign a score of 5.

Logic 2: Identifying Potential Marketing Issues

Calculates the percentage of low ratings (1-star and 2-star ratings) out of the total ratings for each product. Flags products as potential marketing issues if the percentage of low ratings exceeds 50%.

**b) Distress Identification: Pinpointing Buyer Discontent**

Formula to Identify Buyer Discontent:

Calculates the percentage of 1-star and 2-star ratings compared to the total ratings for each product. Identifies product categories causing buyer discontent if the percentage of low ratings is higher than 80%.

**c) Pricing and Review Impact: Examining if there is a correlation between pricing and product reviews and sharing insights.**

Correlation Analysis: Evaluates the percentage of missing data in the 'Price' column to understand its impact on the analysis. Generates a correlation matrix using Pearson's correlation coefficients between 'Price', 'Num Of Reviews', 'Average Rating', and individual star ratings. Visualizes the correlation matrix using a heatmap for interpretation.

**2) Buyer's Perspective: Criteria for Product Selection**

As a buyer, I have certain unsaid guidelines while choosing the right product:

1- I will focus on products which have high average rating and have sufficient amount of reviews.

2- I will definitely stay away from the products having less ratings and less number of reviews.

3- I will make sure the product I am buying stays within my budget. I don't want to overspend.

4- I may consider products with less number of reviews if it has a very high rating.

I have decided to create a scoring mechanism to arrive at the products that I would like to buy.

I have 3 major factors which will decide whether I should buy a specific product or not.

1- Average Rating : As I said in point 1 of the above block that products with a higher average rating will be more suitable to me.

2- Number of Reviews : As I said in point 2 of the above block, if a product has higher number of reviews. I will be more comfortable in buying such a product.

3- Price (let's say I have a budget of 200 dollars. So I wont be comfortable in buying any product which has a price of more than 200 dollars)

The 'Score' column, derived from these factors, helps identify products aligned with the buyer's preferences. The resulting data showcases products meeting these criteria while adhering to the specified budget.

**3)Marketing Insights: Analyzing Product Ratings and Reviews for Marketing Strategies**

Here I did data cleaning, preprocessing and arrived at dataframe C (check the ‘wonderla.py’ or Wonderla Final\_Cut.ipynb file for details)

And created 2 plots which would help us to get some marketing insights:

* 1. Analyzing Rating Distribution:

Plotting a histogram showcasing the distribution of average ratings, excluding products with a 0 rating.

* 1. Review Distribution by Rating:

Plotting a bar chart depicting the total number of reviews corresponding to each star rating (One Star to Five Star).