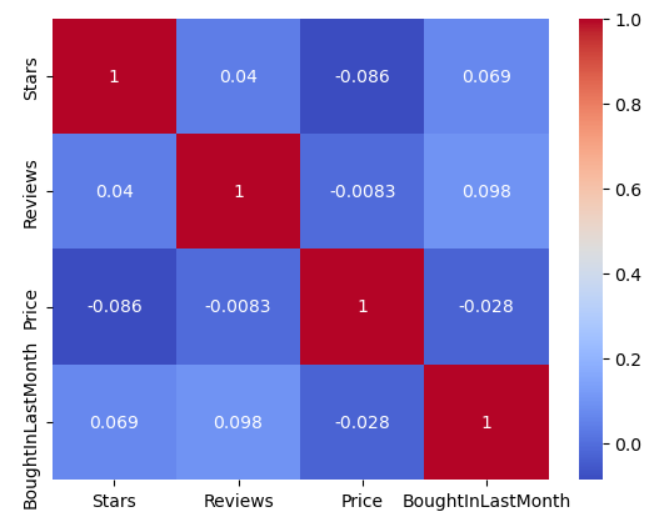
**Statistical Analysis and Recommendations**

**Correlation Analysis:**



**Insights:**

**Stars and Reviews (0.04):** Almost no correlation, suggesting that higher ratings do not necessarily correlate with a higher number of reviews.

**Stars and Price (-0.086):** A very weak negative correlation, hinting that higher-priced products may not always have higher ratings.

**Reviews and Price (-0.0083):** Negligible correlation, indicating that the number of reviews does not increase with higher prices.

**BoughtInLastMonth with Other Variables (0.069 with Stars, 0.098 with Reviews, -0.028 with Price):** Weak correlations show that recent purchases are slightly more influenced by the number of reviews than by stars or price.

**Recommendations:**

**Product Strategy:**

Reviews seem slightly more correlated with purchases in the last month than stars or price, suggesting that increasing the number of reviews could potentially boost sales more effectively than adjusting pricing strategies.

**Marketing and Sales:**

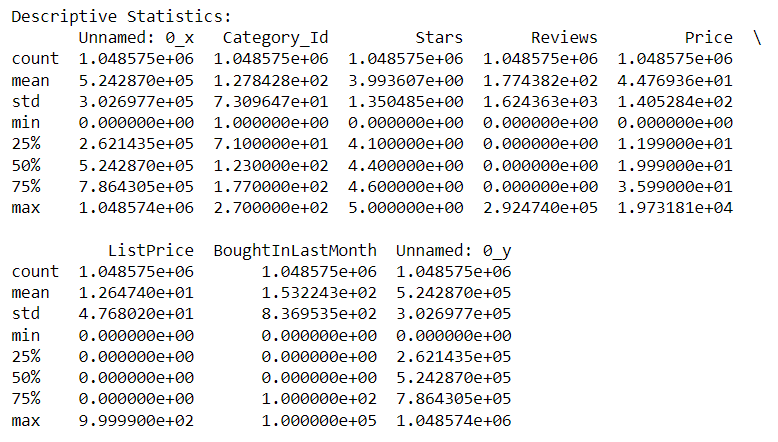
Encourage more customer reviews through post-purchase emails or incentives, as they have a slightly stronger relationship with recent purchases.

Explore other factors that might influence sales, such as marketing campaigns, seasonal trends, or product visibility on the platform, since the traditional metrics like price, stars, and reviews do not strongly predict monthly purchases.

**Machine Learning Algorithms: Decision Tree**

"With an RMSE of 653.16, the Decision Tree model shows promise in forecasting sales for the following month by using things purchased in the previous month as a basis. However, forecasts deviate from actual sales by about 653 units. This suggests that the model is doing a good job of picking up on underlying patterns in the sales data, which makes it a valuable tool for sales forecasting. This model is a good starting point for understanding sales trends and can help with initial inventory management and sales strategy planning, while there is space to improve accuracy and decrease forecast error. It could be helpful to further enhance the model or investigate more complex algorithms in order to maximise performance and produce even more accurate forecasts."

**Statistical Analysis: Descriptive Analysis**



1. **Count:** Each column has 1,048,575 entries, indicating a large dataset with no missing values in the reported columns, ensuring a robust analysis.
2. **Mean:**
   1. Category\_Id: The average category ID is approximately 128, suggesting a wide range of categories.
   2. Stars: The average rating is about 3.99, indicating that items are generally rated favorably.
   3. Reviews: The average number of reviews per item is 177, though this figure can be misleading as we'll see with the median.
   4. Price: The average price is approximately $44.77, showing a moderate price point for items.
   5. ListPrice: The average list price is around $12.65, suggesting that many items may not have a recorded list price or it's set at a base rate.
   6. BoughtInLastMonth: On average, 153 items were bought in the last month per listing.
3. **Standard Deviation:**
   1. Reviews, Price, and BoughtInLastMonth have high standard deviations relative to their means, indicating significant variability and suggesting a wide range of values including potential outliers.
4. **Percentiles (25%, 50% - Median, and 75%):**
   1. Stars: 75% of items have a rating of 4.6 or lower, with 25% rated at least 4.1, showcasing generally high ratings.
   2. Reviews and BoughtInLastMonth: Both show a median of 0, indicating that more than half of the listings did not receive reviews or sales in the last month, which could point to a long tail of rarely purchased or new items.
   3. Price: The interquartile range (IQR) from $11.99 to $35.99 shows that most items are modestly priced.
   4. ListPrice: 75% of items have a list price of 0, which could suggest many items are sold without a discount or marked list price.

These statistics indicate a dataset with a wide range of product interactions, from high to no sales or reviews, and generally favorable ratings. The high variability in reviews and sales suggests that some items are much more popular or visible than others. This skewed distribution is typical in e-commerce settings where a small number of items dominate sales and review activities.