Amazon Ratings insights resume

csv of amazon from kaggle

ABOUT THE CSV

The csv contains info of buys, categories, ratings, discount and price on some amazon products, it has more than 1,000 rows and 16 columns.

cleaning and formatting

A lot of the rows are text values that are not really useful. It is also necessary to transform and clean some of the useful columns like categories, that have too many subcategories, and price that we clean the euros symbol and make it a number category.
After all the processing of data I end up getting fewer columns but these are the ones of which we can make insights.

| Discount percentage	Rating	Index	Category	Discounted price	Actual price
0.5	4.2	9	Computers&Accessories	499	999
0.7	4.3	30	Computers&Accessories	299	999
0.8	3.9	35	Computers&Accessories	199	999
0.8	4	38	Computers&Accessories	199	999
0.67	3.3	46	Computers&Accessories	333	999
0.6	3.6	49	Electronics	399	999
0.86	4	63	Computers&Accessories	139	999
0.6	4.1	75	Computers&Accessories	399	999
0.8	4.5	89	Computers&Accessories	199	999
0.8	4.1	94	Computers&Accessories	199	999
0.65	3.7	98	Computers&Accessories	345	999
0.65	4.2	105	Computers&Accessories	349	999
0.6	4.3	106	Computers&Accessories	399	999
0.7	4.3	108	Computers&Accessories	299	999
0.73	4.3	112	Computers&Accessories	273.1	999
0.6	4.1	114	Computers&Accessories	399	999
0.65	3.5	118	Computers&Accessories	347	999
0.85	4	119	Computers&Accessories	149	999
0.6	4.3	132	Computers&Accessories	399	999
0.66	4.3	162	Computers&Accessories	339	999
0.67	4.3	174	Computers&Accessories	325	999
0.15	4.1	183	Computers&Accessories	849	999
0.65	4	204	Electronics	349	999

final insights

the conclusions and info that we can gather from this csv are not very useful but can serve to get some knowledge on the ratings of products, i think the 2 most important are:
1.- the rating of the 2 best categories are not link to the discount (its not recommended to increase discount if you want to bust your ratings)

Average of rating by category.
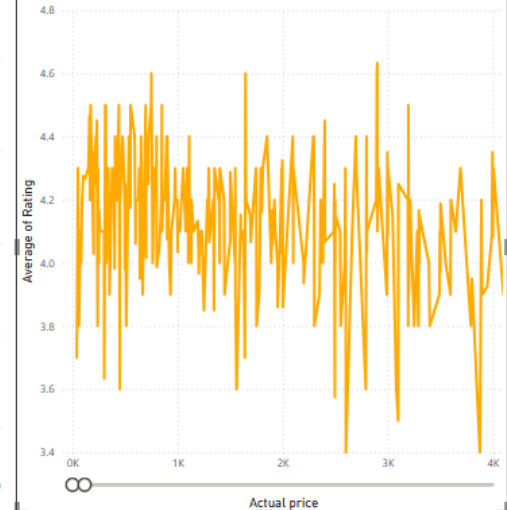
Average of discount_percentage by category
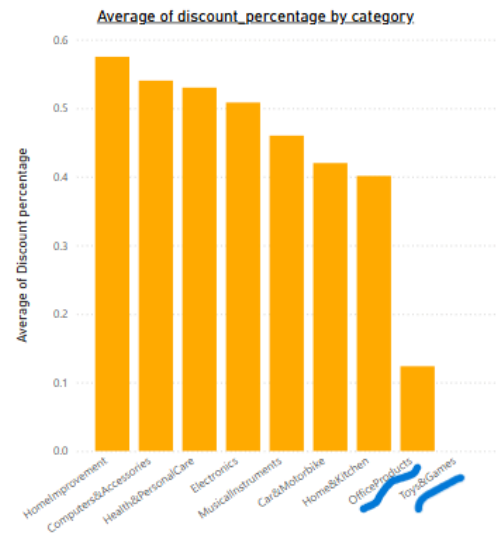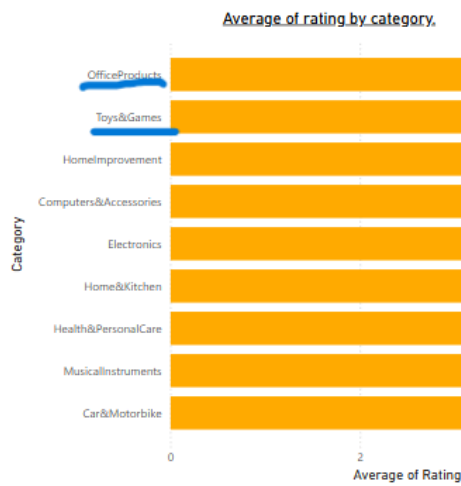
2.- the ratings are not dependent on the discount or price of the products, some other factors should be more important on that regard.

Average of Discount percentage by Rating
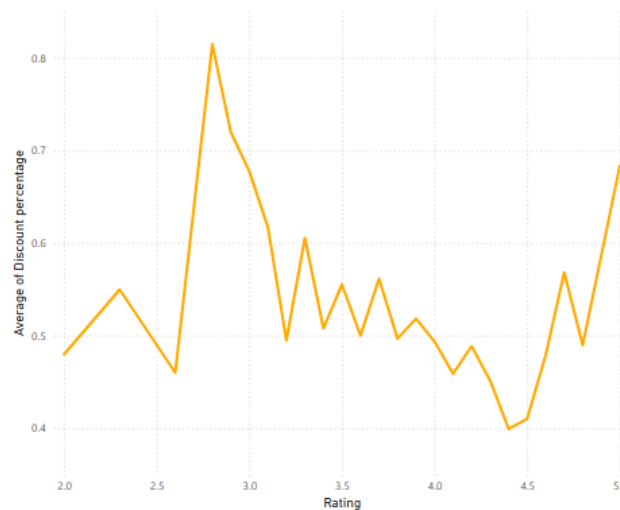
Average of Rating by Actual price

at the end the most relevant information we get from this data is the we need more numerical data in order to make better inferences on this topic.