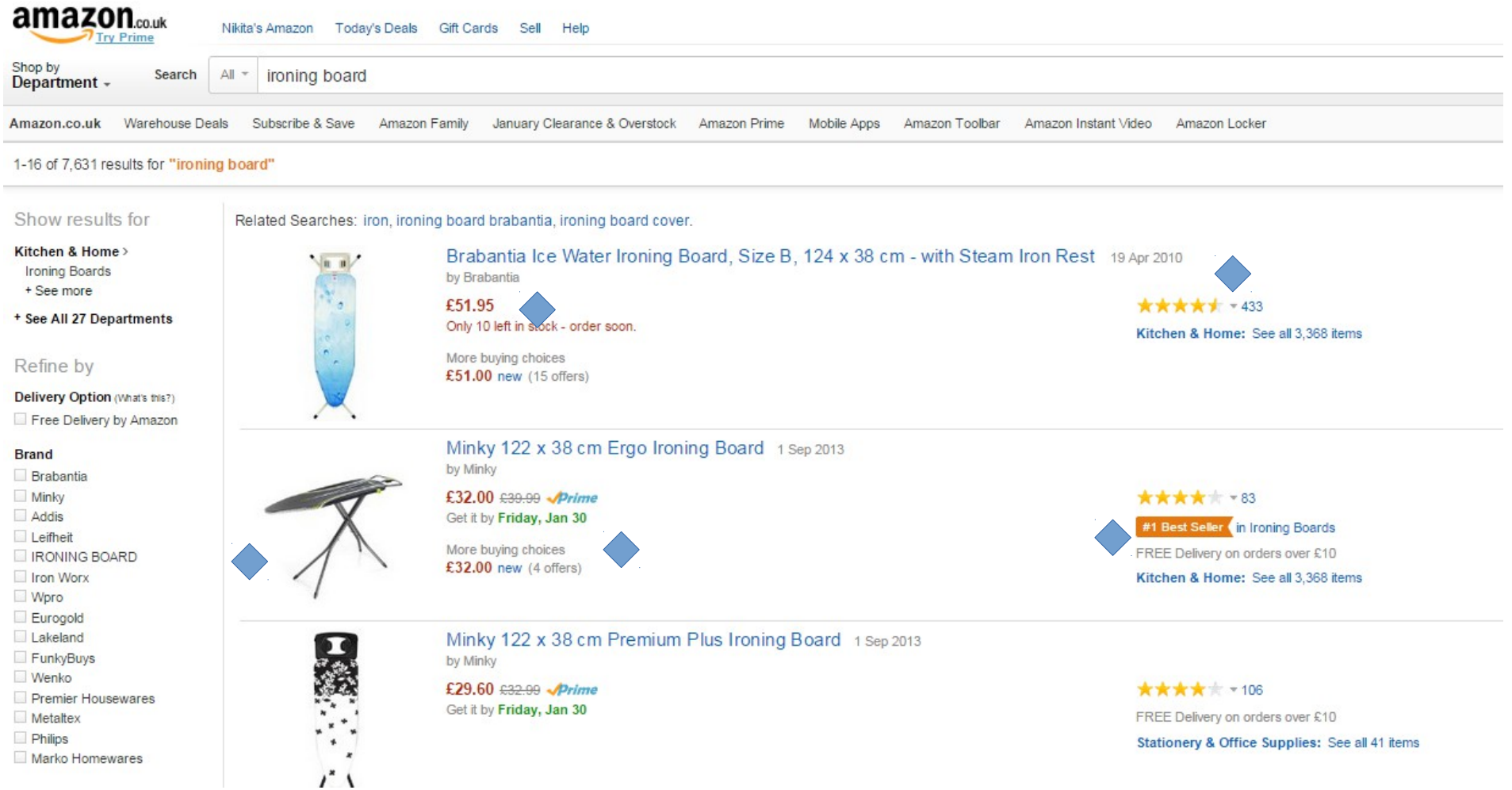
Amazon category mining

Ironing boards example

Lot's of useful information, but hard to analyze 'from screen'



Our features:

- Regular daily downloading (parsing) the Amazon range
- Intelligent disparsing text data from item description (get all descriptions with parametrized sizes, details, usage properties)
- Storing images and price history
- MDS classification of items by parameters: size, color, use purpose
- Proposals for new range based over analysis of current range
- Theoretical sales volume calculation by Rating parsing

Rating parsing

You could be more effective if you manage your Amazon sales looking to competitors ratings. Ating in Amazon are quite complex, and here's short description:

Amazon does indeed know what they're doing. They are skewing the ratings. They add all the stars and divide it by the number of reviewers.

Let's say 25 people reviewed a product, let's say, a toaster, and each star got 5 peoples votes. With 1 being the lowest rating, and 5 being the highest. (1-5 stars being the only choices)
5 x 5 stars= 25
5 x 4 stars= 20
5 x 3 stars= 15
5 x 2 stars= 10
5 x 1 stars= 5
Total stars is 75 divided by 25 reviewers = an average 3 star rating. Not a bad rating. Amazon will probably sell a lot of these toasters.

Somebody suggested they should start with a zero star rating. Let's try that, and still do it Amazon's way of adding up the stars and dividing it by the number of reviewers.
Now, lets say the same 25 people reviewed the toaster, and each star got 5 peoples votes. This time with zero being the lowest rating, and 4 being the highest. (you still have 5 choices)
5 x 4 stars= 20
5 x 3 stars= 15
5 x 2 stars= 10
5 x 1 stars= 5
5 x 0 stars= 0
Total stars is 50 divided by 25 reviewers = an average 2 star rating. Uh oh. The toaster doesn't look so good now does it?

Somebody suggested a 6 star rating, 0-5. Lets try that. Only this time we'll have 30 reviewers since we have 6 choices. We are still neutral because each star gets the same number of votes, 5.
5 x 5 stars= 25
5 x 4 stars= 20
5 x 3 stars= 15
5 x 2 stars= 10
5 x 1 stars= 5
5 x 0 stars= 0
Total stars is 75 divided by 30 reviewers = an average 2.5 star rating. The toaster still doesn't look so good, but this is more accurate, and it should be because we gave people the opportunity to vote zero.

Obviously, since Amazon forces any reviewer to give a product at least a 1 star rating, it skews the rating in favor of the product, especially when other people vote 4 and 5 stars. Why? Well, if only 5 people voted 1 star, it would have a 1 star rating which means you DO NOT recommend the product. The problem here is, the 5 star rating is built on the 1 star ratings, and in this case that first star equals recommendation.

If you want a more accurate star rating, simply take Amazon's average rating and divide it by 1.25. In the above example where 5 people each voted for one of the star categories for a 3 star rating, it would actually equal a 2.4. Not too good. The bottom line: Amazon's ratings are ALWAYS high, but lets not forget, Amazon is selling things here. Trust the reviewers that have a lot of reviews, both positive and negative.

Classification

PCA / FAC classification would give you latent structure of sales in Amazon in order to quick enroll to this market

Principal components (eigenvectors)

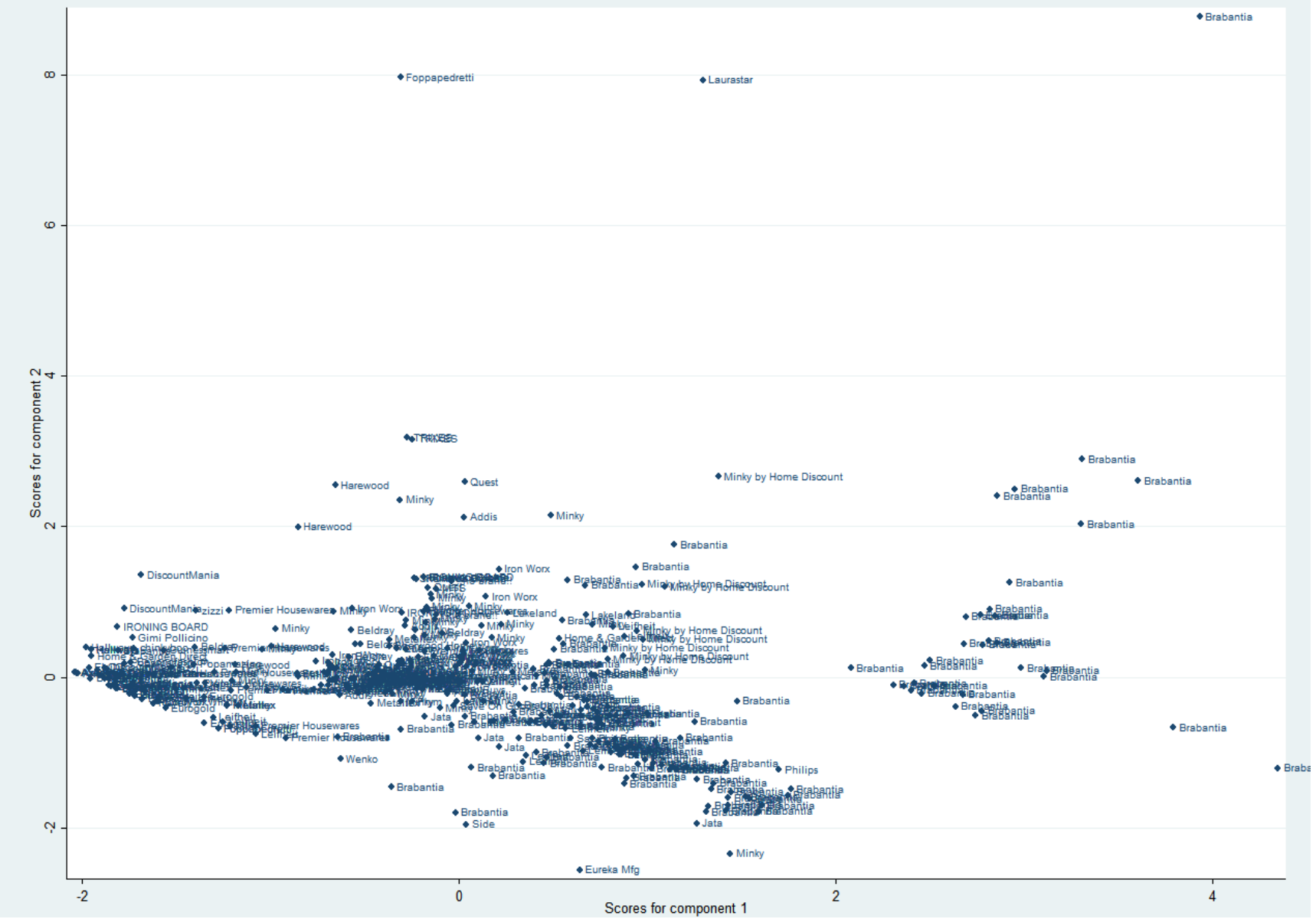
Variable	Comp1	Comp2	Comp3	Comp4	Unexplained
rating	0.6333	-0.0319	-0.0307	-0.7726	0
reviewcount	0.5763	0.2399	-0.6110	0.4868	0
price	0.4956	-0.4850	0.5977	0.4025	0
sellerrevi~t	0.1455	0.8403	0.5182	0.0640	0

Example of Ironing board Principal Component Analysis:

1 group (37% of sales): High rating, medium review count, medium price, small Seller review count [sales leaders in here]
2 group (26%): Small rating, small review count, small price, high Seller review count
3 group (18%): Small rating, negative reviewcount, high price, high seller review count

Classification

How does the overall picture looks for ~500 sku in Ironing board by classifying:



No time to analyze this?

Trust it to us! mail@getrealprice.com