General observations about swiss products based on amazon reviews

Bojan Petrovski, Luca Rizzello, Lucas Monnin

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Introduction

Products origins are deduced from manufacturer as it is the only information about location amazon provides.

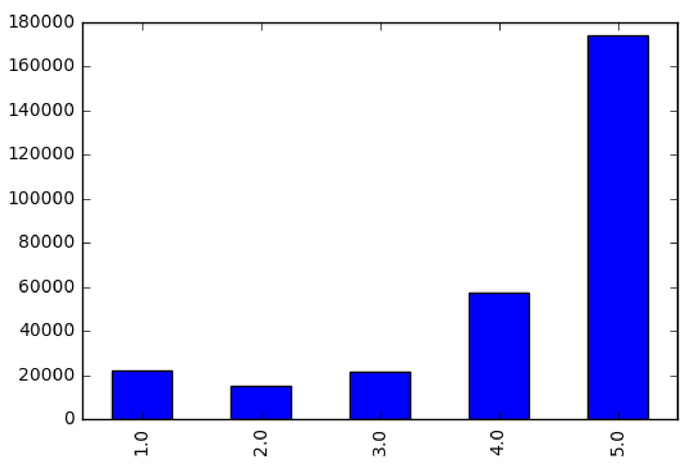
Number of reviews 289'878

Mean 4.19

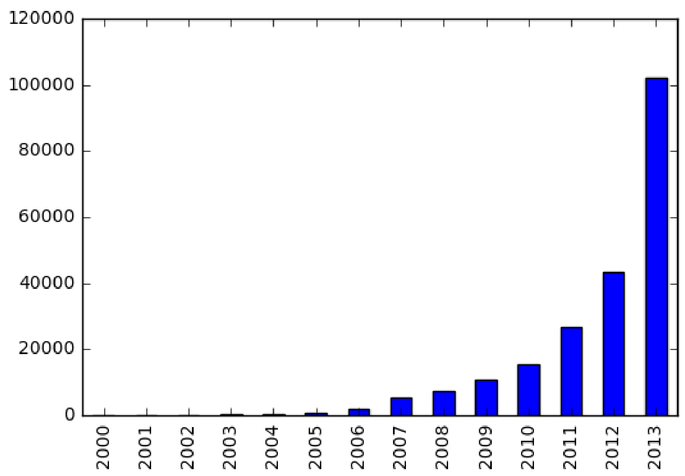
Std 1.23

Reviews

The reviews does not follow a normal distribution since 75% of the reviews are 5 stars reviews.

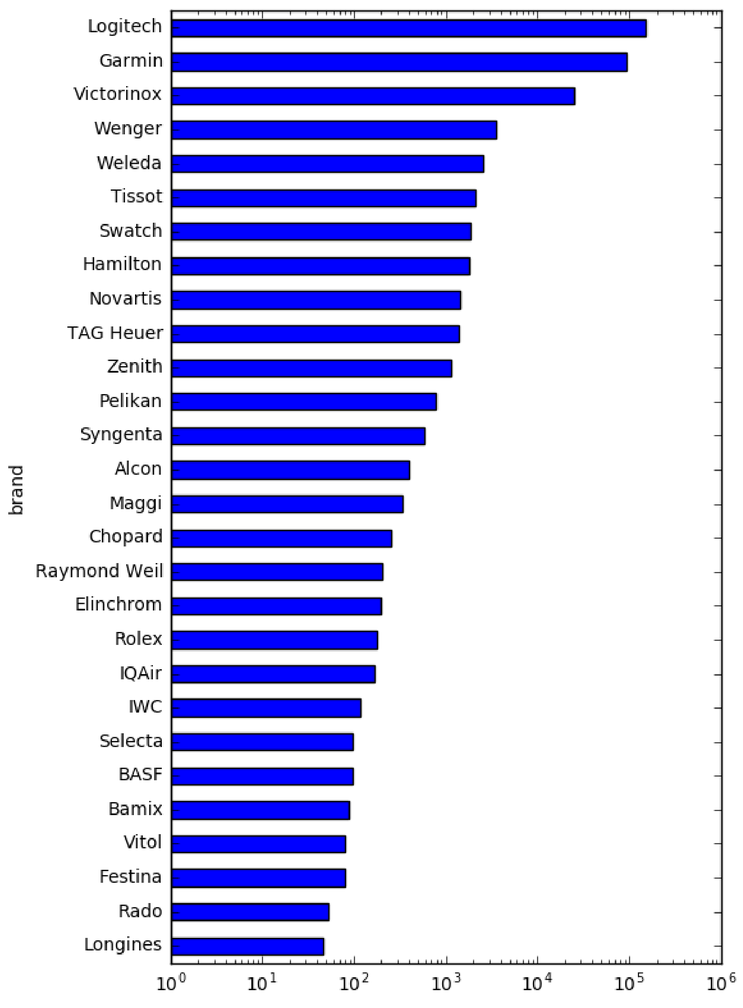


Here we can observe the exponential growth of the number of reviews on swiss products (2014 was removed as data was collected in july 2014).



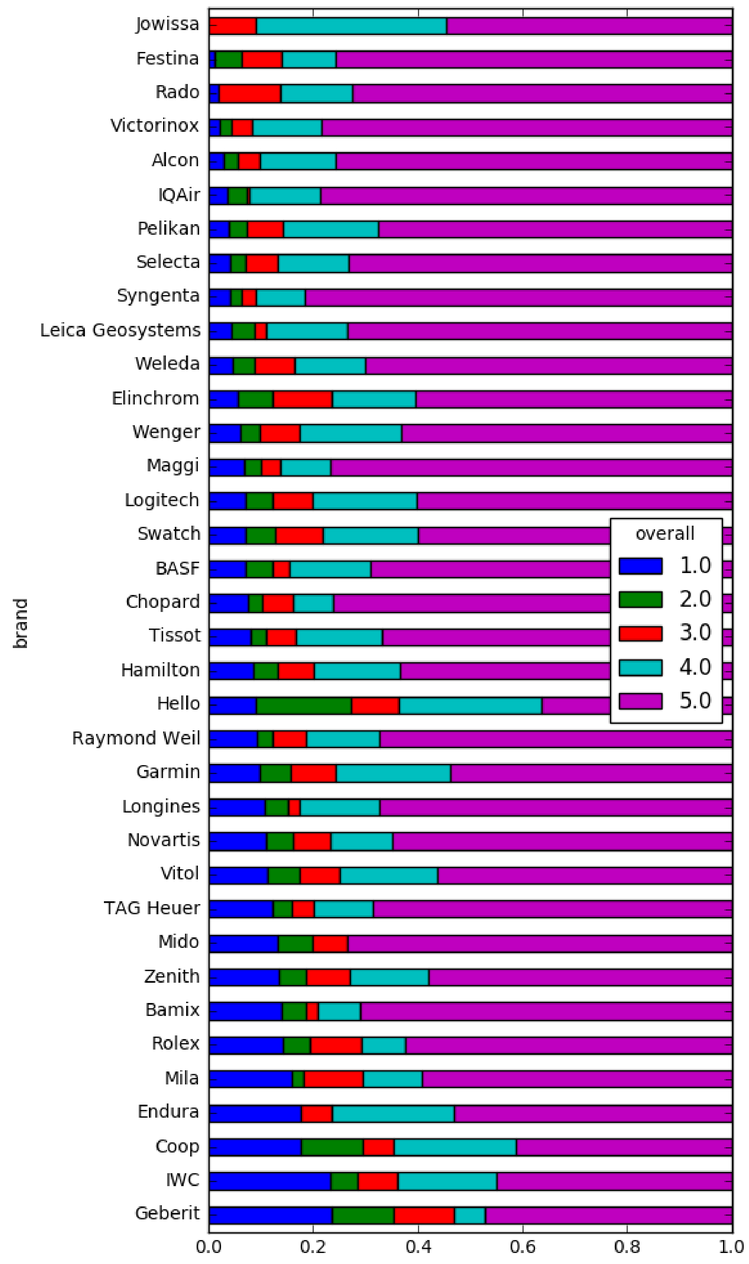
Top 20 reviewed brands

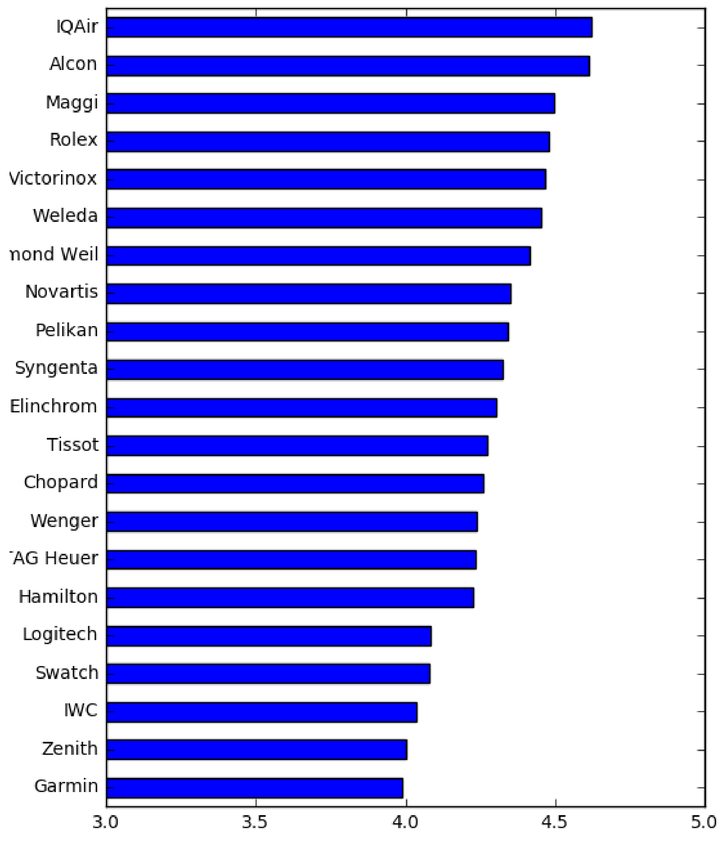
This graph shows the number of reviews for each brand.



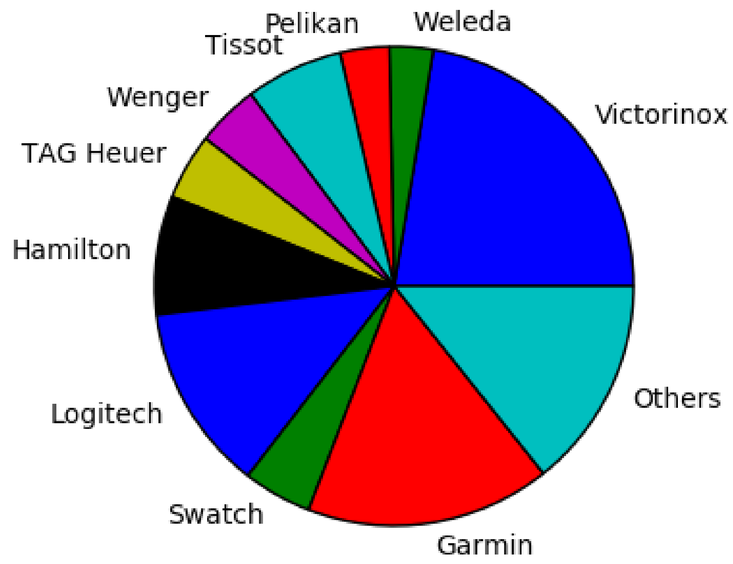
Brands comparisons

On the left, each brand and its reviews (1, 2... to 5), sorted by the ratio of bad reviews (1 star). Brands with less than 10 reviews have been removed from the list.
And on the right, the best rated brands *with more than 100 reviews*.





This pie chart shows the Swiss companies depending on the number of products they sell. The category *Others* contains 38 companies.



Comments

Unsurprisingly, the most exported goods are watches, knives and electronics (note that Wenger does not exists anymore, bought by Victorinox). This latter category of products is mostly shared between Logitech and Garmin (US company based in Switzerland...), but interestingly they seem to have a slightly lower rating mean compared to watches companies.
The difference is not significant but one may conclude that the rating is correlated to the manufacturing location.

Another observation is that electronics products tend to be more reviewed than the others (fancy watches, but one may assume that people who buy these kind of watches are more likely not to buy them online).

And finally some numbers: in these 289'878 reviews, there are 4718 different products for 48 Swiss companies, 3 of them having 51% of the swiss products: Victorinox with 1056 products, Garmin with 780 and Logitech with 595.

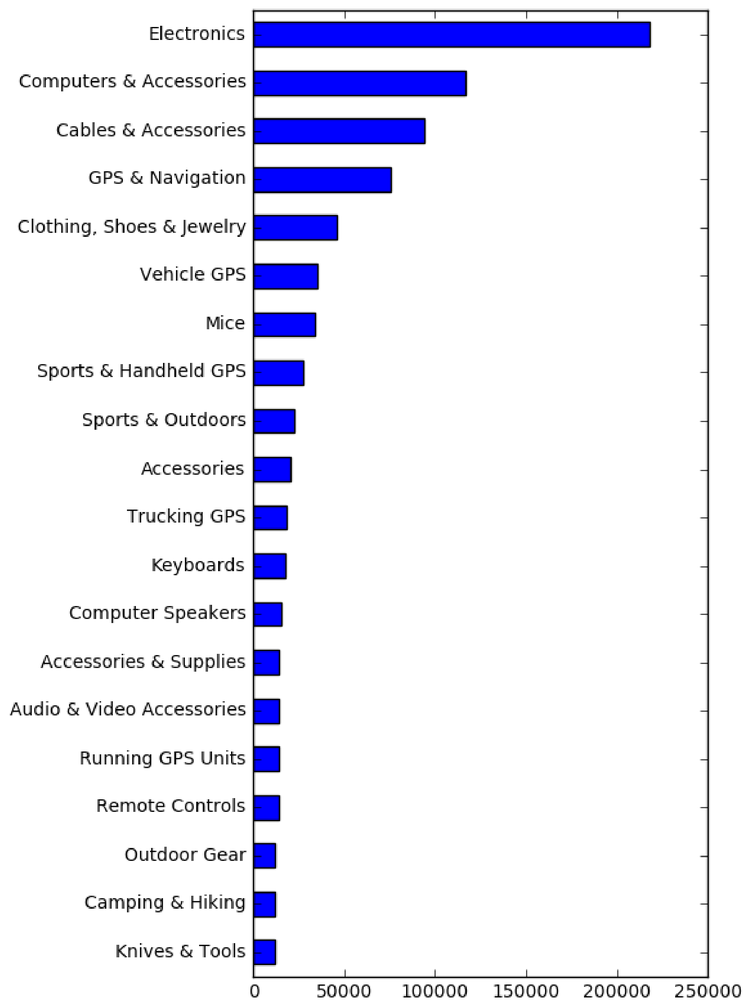
General observations about swiss products based on amazon reviews

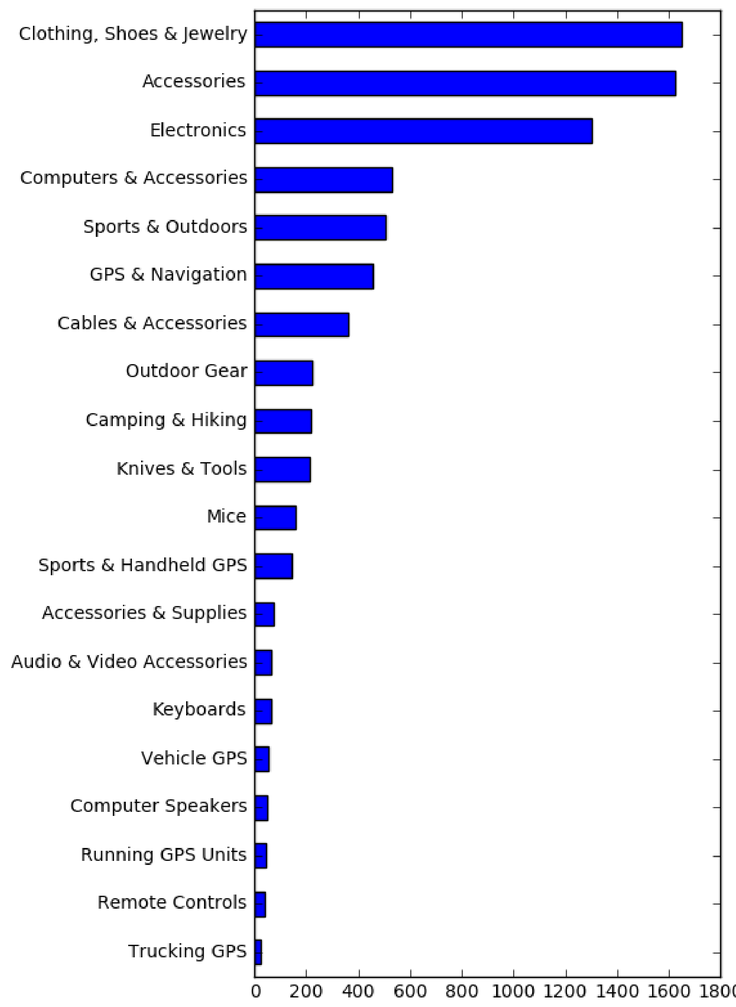
Bojan Petrovski, Luca Rizzello, Lucas Monnin

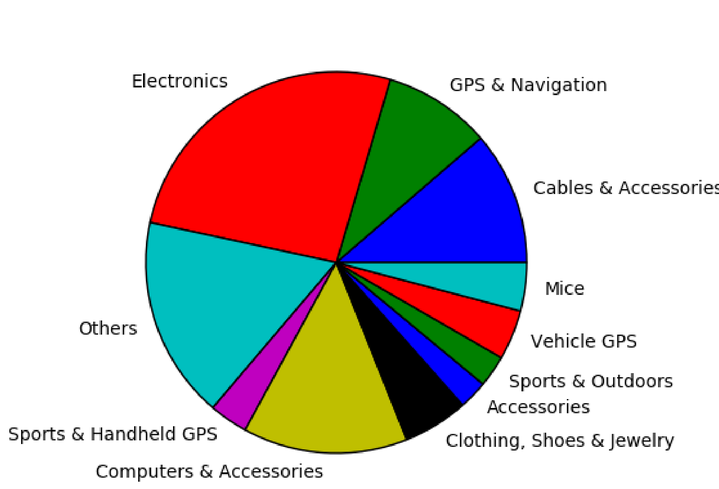
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

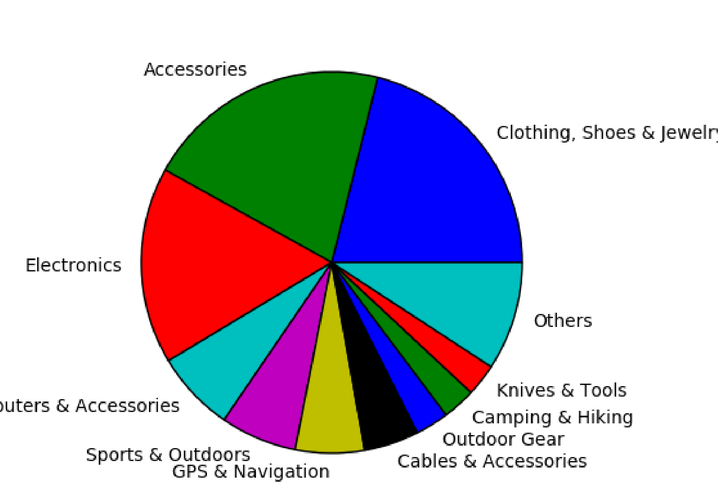
Categories distribution

On the left, the most reviewed categories, and on the right the categories with the most products. At the bottom left, the pie charts shows the top 10 reviewed categories, and at the bottom right, the distribution among the top 10 categories with the most products (*Others* contains 10 more categories for both).



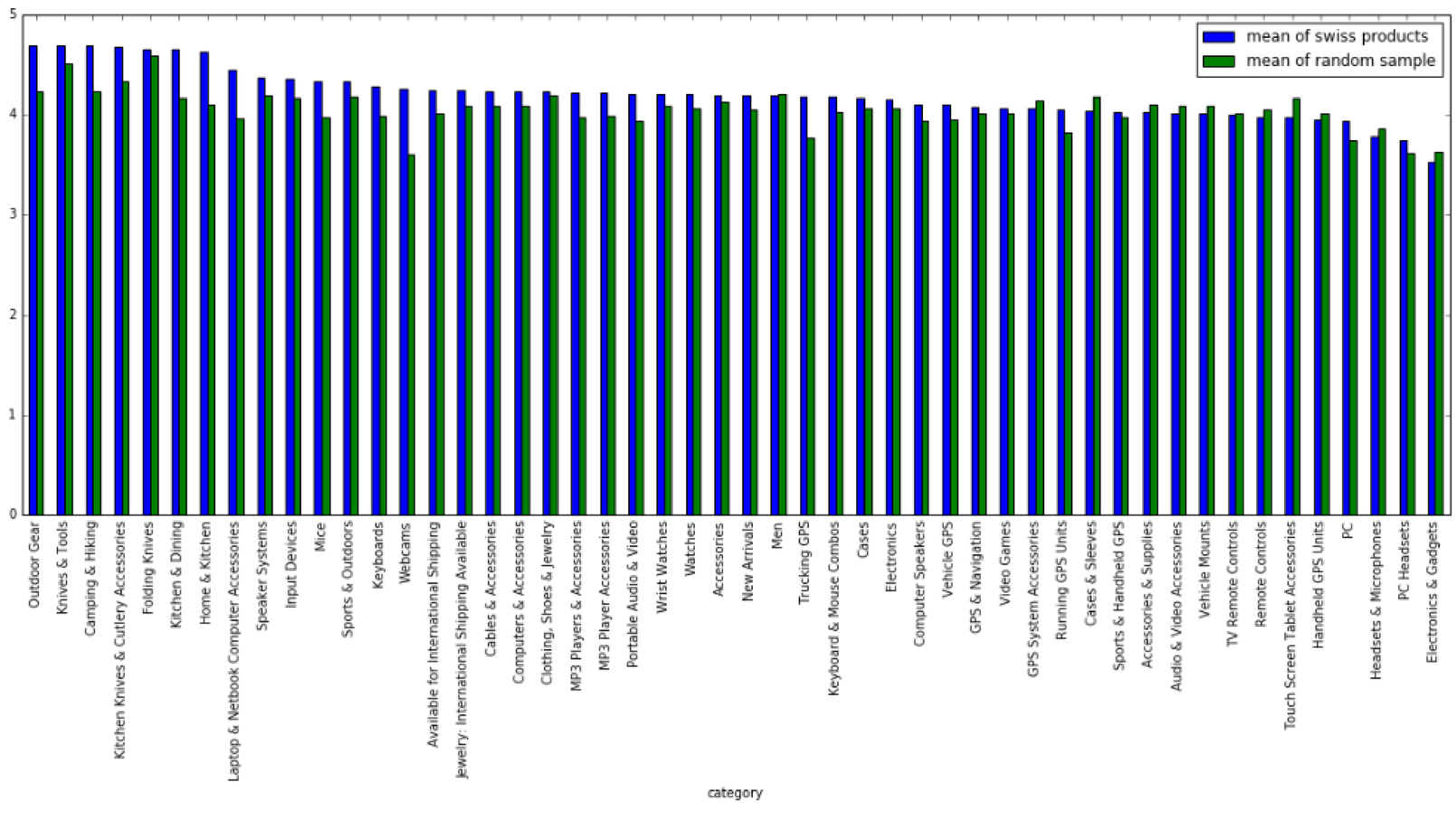






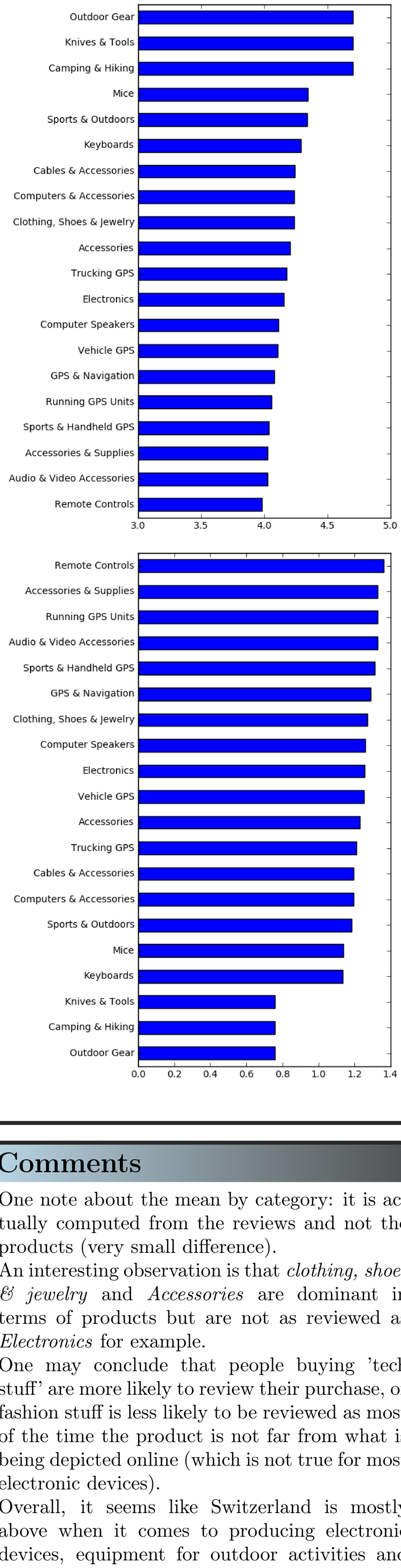
Categories comparisons

In the following graph we have compared the mean of reviews of the most popular categories of Switzerland to those of the rest of the world (n = 300'000 for both samples)



Categories ratings

The chart shows the mean rating of each category, followed by the std of each category.



Comments

One note about the mean by category: it is actually computed from the reviews and not the products (very small difference).

An interesting observation is that *clothing, shoes & jewelry* and *Accessories* are dominant in terms of products but are not as reviewed as *Electronics* for example.

One may conclude that people buying 'tech stuff' are more likely to review their purchase, or fashion stuff is less likely to be reviewed as most of the time the product is not far from what is being depicted online (which is not true for most electronic devices).

Overall, it seems like Switzerland is mostly above when it comes to producing electronic devices, equipment for outdoor activities and watches.

Statistical Comparison of Reviews

Bojan Petrovski, Luca Rizzello, Lucas Monnin

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Introduction

In the first comparison we directly compare the set of Swiss reviews with 4 other sets: a random control set, French set, German set and US set. All data sets are generated using the same spark script and Using the Wikipedia listings for the respective country brand. The random control set is generated using the spark Bernoulli sampling. We use the Mann-Whitney U test instead of the more frequently used t-test since we can not say with certainty that the samples adhere to a normal distribution. Furthermore the Mann-Whitney U test for large samples asymptotically approaches the t-test if the distribution is normal so it is a better choice in our case.

One of the possible problems with using the raw reviews is that some very popular products with lot's of great reviews can skew the results so it is very difficult to generalize results while using individual results. That is why we aggregated the reviews by product. After aggregation we now have a sample of some continuous distributions so the KS test can also be used to check if the distributions are different.

In the third experiment we try to eliminate some effects of fake reviews, or reviews that do not address the product, so we use the helpfulness as a weight when we do the average of each product.

Descriptive Statistics

Comparison of Reviews

Country	Mean	Std
Swiss	4.19	1.23
German	4.15	1.29
French	4.12	1.37
US	4.09	1.33

Comparison after grouping by product

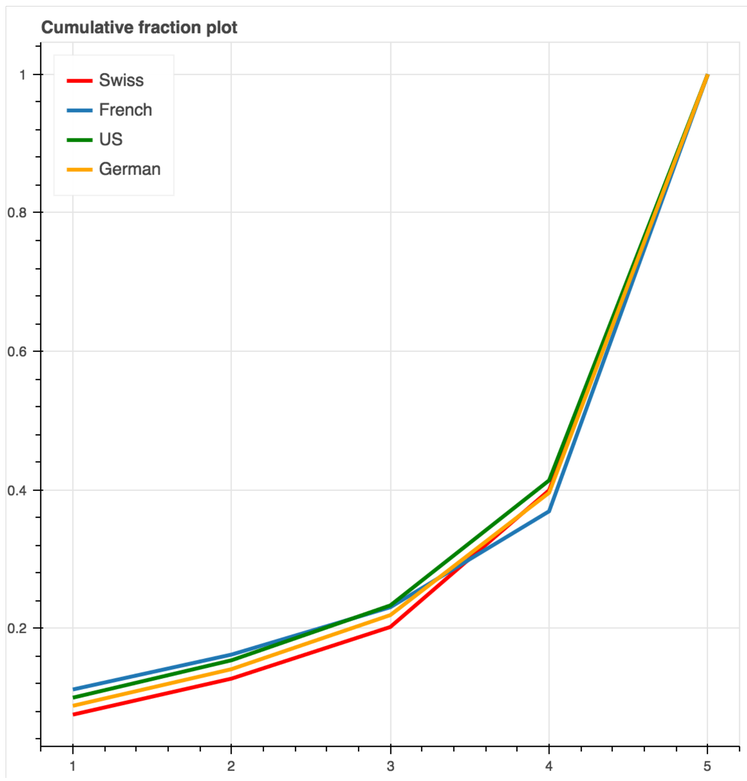
Country	Mean	Std
Swiss	4.24	0.81
German	4.21	0.92
French	4.26	0.94
US	4.14	0.96

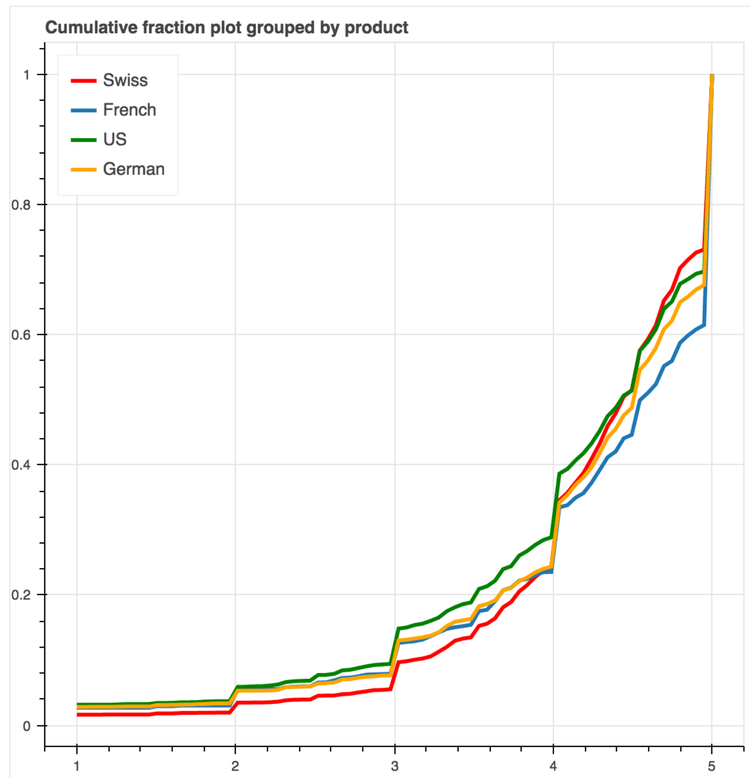
Comparison after grouping by product using weight

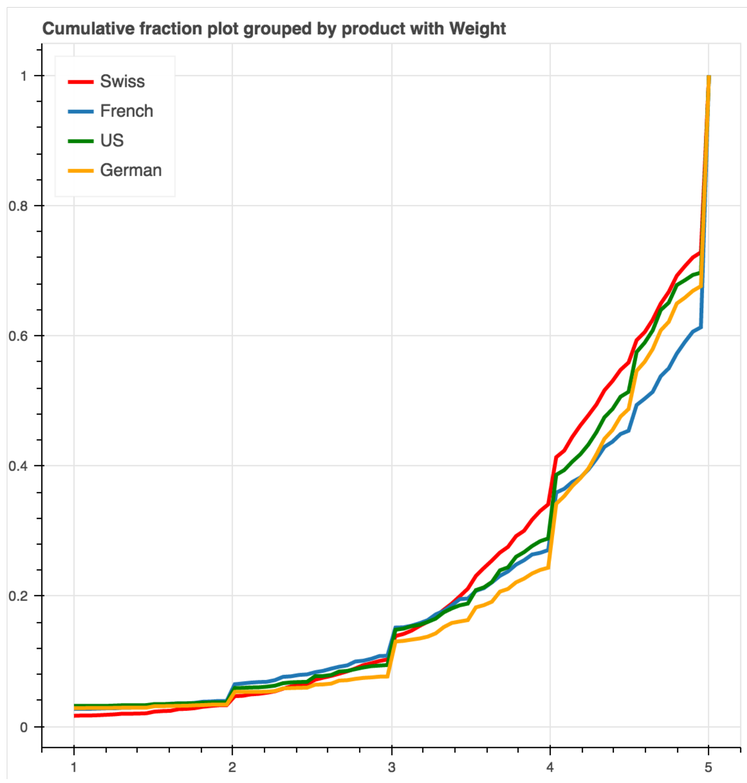
Country	Mean	Std
Swiss	4.11	0.92
German	4.13	0.99
French	4.20	1.00
US	4.05	1.03

Cumulative fraction plots

The first plot is represents the raw data form the reviews and it is a discrete distribution. The second plot on the right is the distribution after grouping by products. The Last plot at the bottom represents the distribution after taking into account the helpfulness of each review.







Comments

Comparison of Reviews

We can not find a statistically significant difference between the Swiss products and the control, but we can reject the null hypothesis in the other tests. For German and US products it implies that there is a significant chance that a Swiss product chosen at random will have a higher rating than a German or US product chosen at random. For French and Swiss products the reverse is true.

Comparison after grouping by product

Using the KS test we discovered that in all cases the distributions are different. After that we preformed the Mann-Whitney U test and this time we could not find any statistical differences between the Swiss, German and US products, the only test that showed a statistically significant result was between Swiss and French products, the result being that a Swiss product chosen at random will have a smaller rating than a French product chosen at random.

Comparison after grouping by product using weight

We again perform our tests and this time we find two significant results, both German and French appear better than Swiss, while there is no difference between US and Swiss Products. The K-S tests again show different distributions.

Sentiment Statistical Comparison of Reviews

Bojan Petrovski, Luca Rizzello, Lucas Monnin

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Introduction

One additional metric that we can use is Sentiment Analysis on the Text of the reviews. We Analyzed all reviews and even found out that there is some correlation between the Sentiment score of the review and the number of Starts awarded. While this is not a hard metric, it may give us an indication of the attitude of Consumers towards the different products.

Descriptive Statistics

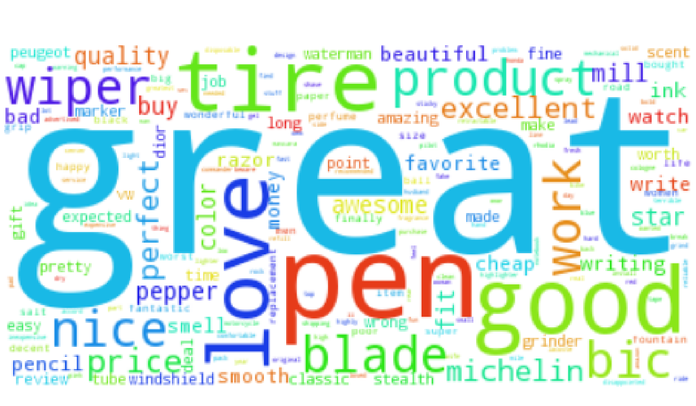
Comparison of Reviews

Country	Mean	Std
Swiss	0.64	0.49
German	0.61	0.51
French	0.61	0.50
US	0.59	0.52

Comparison after grouping by product

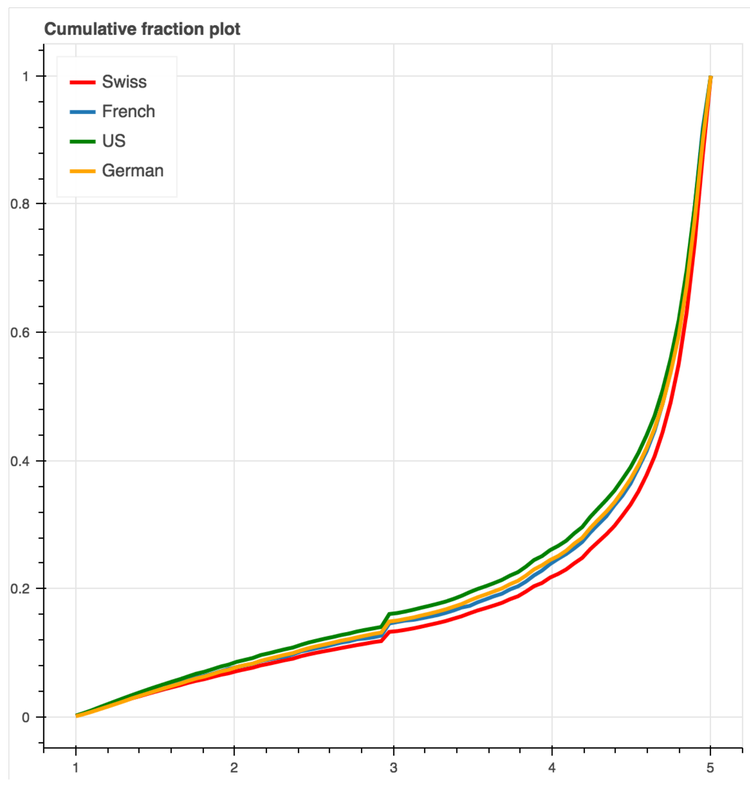
Country	Mean	Std
Swiss	0.65	0.31
German	0.62	0.34
French	0.66	0.33
US	0.60	0.36

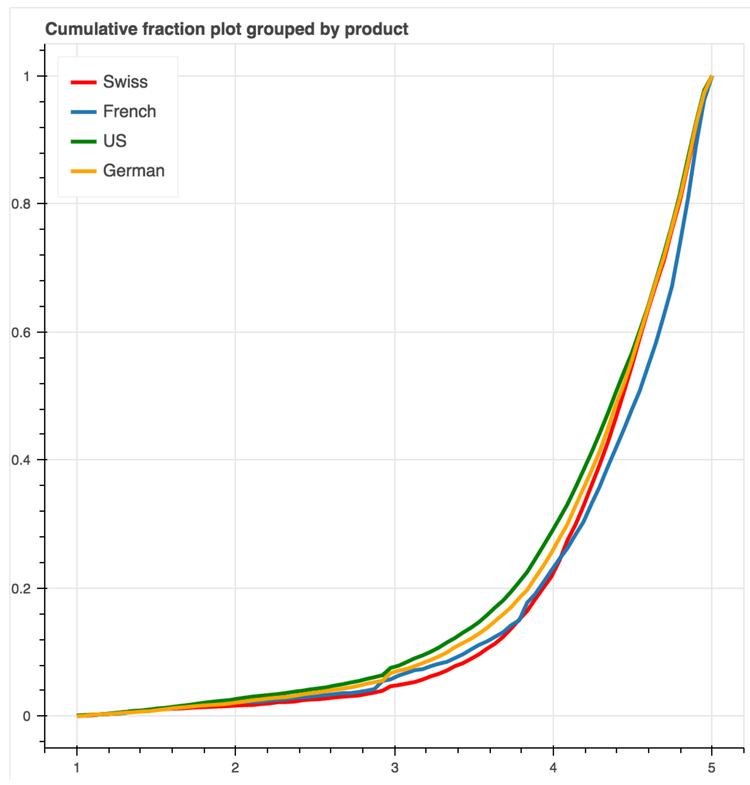
Word Clouds

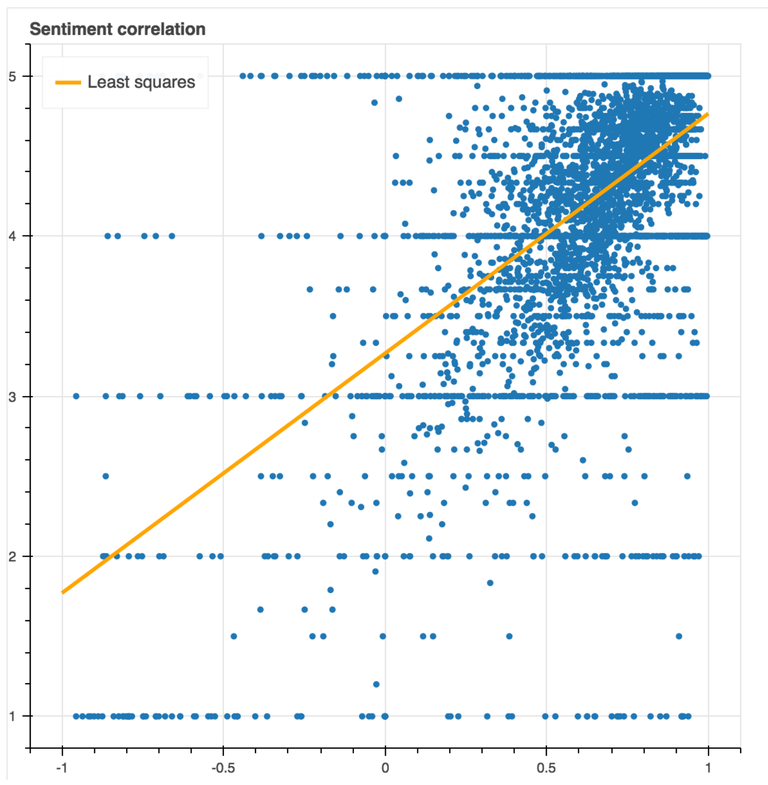


Cumulative fraction plots

The first plot is represents the raw data form the sentiment analysis of the reviews. The second plot on the right is the distribution after grouping by products. The Last plot at the bottom represents the scatter plot of the number of stars vs the sentiment of the review. The yellow line represents a least squares approximation.







Comments

Comparison of Sentiment Data

After analyzing the data we found out the the sentiment score of Swiss Reviews is greater that the The Scores of US, French and German Reviews. The Mann-Whitney U test showed statistical significance in all three tests. The mean of the Swiss reviews was 64.7 while the US, French and German Reviews had means of 59.1, 61.8 and 61.3 respectively. Additionally we grouped the data by product and in that case we can only show that Swiss products have a greater score than German and US products but a worse score then French Product.

Conclusion

Or goal was to try multiple different methodologies and to see if there is some general pattern that would be present in all the experiments. While we did find some statistical significant results there was no pattern and in fact sometimes the results flip deepening on how you group the data. Furthermore even when there is a statistically significant difference the difference is generally small and only 1-3%, and can easily be due to our sampling.