

**Topic Modeling Amazon Product Reviews**

University of Colorado Boulder

*Michelle Puglio*

January 31, 2019

*Puma Review Topics using K-means Topic Modeling (Segmented by Rating)*

***Reviews with 1 Star or Less***

|  |  |
| --- | --- |
| **Topic 1:** | quality shoes poor weeks |
| **Topic 2:** | shoe size 34 shoes |
| **Topic 3:** | size shoes ordered wrong |
| **Topic 4:** | watch time water working |
| **Topic 5:** | pair shoes one used |
| **Topic 6:** | small size fit good |
| **Topic 7:** | shoes like running picture |
| **Topic 8:** | uncomfortable shoes hard blisters |
| **Topic 9:** | color wrong ordered different |
| **Topic 10:** | hurt feet el de |
| **Topic 11:** | narrow wide foot feet |
| **Topic 12:** | shirt small bad darker |
| **Topic 13:** | leather synthetic description shoes |
| **Topic 14:** | socks would work shoes |
| **Topic 15:** | 13 size way small |
| **Topic 16:** | apart bag fell ripped |
| **Topic 17:** | made fake poorly knock |
| **Topic 18:** | fit shoes support tight |
| **Topic 19:** | disappointed tight way shoes |
| **Topic 20:** | large lot size somewhat |

***Reviews with 4 Stars or More***

|  |  |
| --- | --- |
| **Topic 1:** | watch great nice good |
| **Topic 2:** | great fit shoes price |
| **Topic 3:** | running shoes great shoe |
| **Topic 4:** | shoe great comfortable good |
| **Topic 5:** | size small half shoes |
| **Topic 6:** | shoes comfortable great like |
| **Topic 7:** | comfortable like great fit |
| **Topic 8:** | bag gym b005lce58a great |
| **Topic 9:** | good quality product shoes |
| **Topic 10:** | love shoes comfortable fit |
| **Topic 11:** | narrow feet shoes wide |
| **Topic 12:** | muy de el la |
| **Topic 13:** | sock great love fit |
| **Topic 14:** | pair shoes another comfortable |
| **Topic 15:** | 34 shoes shoe fit |
| **Topic 16:** | loves son shoes great |
| **Topic 17:** | nice shoes really comfortable |
| **Topic 18:** | stars five four great |
| **Topic 19:** | sneaker great comfortable sneakers |
| **Topic 20:** | cute super shoes daughter |

*Review Topic Descriptions*

***Reviews with 1 Star or Less***

**Topic 1** – shoes tend to fray, wear, or rip, including: Puma Soleil FS and Complete TFX Sprint 3.

**Topic 2** – consumers suggest ordering ½ size larger as theirs were too small (i.e., Adult Suede Classic Shoe).

**Topic 3** – consumers received wrong size or wrong product.

**Topic 4** – complaints that Puma “water-resistant” watches are not truly waterproof (Men’s & Women’s watches).

**Topic 5** – comparisons to Nike shoes which consumers seemed to like better.

**Topic 6** – Puma shoes are too small, consumers continue to request ½ size larger.

**Topic 7** – Puma Cell Voltra Sneaker are not typically the best for running in.

**Topic 8** – consumers felt uncomfortable in Puma Soleil Leather Sneakers and developed blisters.

**Topic 9** – customer received their product in the wrong color than the one ordered.

**Topic 10** – reviews in Spanish.

**Topic 11** – a wide variety of Puma shoe wearers felt they were too narrow.

**Topic 12** – Sizing is too small (t-shirts, shoes, etc.).

**Topic 13** – consumers complain that “leather” shoes are actually synthetic (Puma Soleil Leather Sneakers).

**Topic 14** – Puma shoes tend to give blisters when they are worn without socks.

**Topic 15** – Puma’s children’s shoes are typically too small, including: Voltaic Jr Running Shoe, Whirlwind V Sneaker,

and Attencio II Jr Soccer Cleat.

**Topic 16** – a wide variety of bags made by Puma tend to rip after very few uses.

**Topic 17** – consumers received “fake” Puma products.

**Topic 18** – consumers say Puma Adult Suede Classic Shoe fit too tight.

**Topic 19** – consumers are disappointed with the small-fitting Puma shoes, claiming they are “way too tight”.

**Topic 20** – Amazon customers feel the size chart does not represent the true fit of Puma shoes.

***Reviews with 4 Stars or More***

**Topic 1** – wearers of Puma Men’s PU105P2.0054.004 Futuristic Watch like its style and fit.

**Topic 2**  - the price of Puma’s Adult Suede Classic shoe is deemed reasonable by customers, among other styles.

**Topic 3** – Puma ‘Voltaic’ Running sneakers are the best for walking, hiking, and running.

**Topic 4** – the ‘Adult Suede Classic Shoe’ is found incredibly comfortable by customers.

**Topic 5** – positive reviews about many products, but most suggest ordering shoes at ½ size larger.

**Topic 6** – Puma Men’s Suede Sneakers tend to fit comfortably.

**Topic 7** – female consumers think that Puma’s heels, flats, and wedges are exceptionally comfortable.

**Topic 8** – consumers are very happy with the durability and size of Puma Men’s Teamsport Formation Duffel Bag.

**Topic 9** – male consumers deem Puma Men’s El Rey Cross Perf Leather Slip-On Sneaker as a ‘quality’ product.

**Topic 10** – a wide variety of Puma shoe wearers reported that they were comfortable.

**Topic 11** – Puma shoes are stylish, but way too narrow.

**Topic 12** – reviews in Spanish.

**Topic 13** – a wide variety of different socks tend to fit comfortably.

**Topic 14** – happy Puma consumers intend on buying another pair of their shoes in a different color.

**Topic 15** – positive reviews on a wide variety of shoes, but consumers wished they had purchased a larger size.

**Topic 16** – parent consumers who bought shoes for their son(s) were satisfied with their purchase.

**Topic 17** – Puma’s more casual sneakers are found comfortable by consumers.

**Topic 18** – non-descriptive reviews, (i.e. “four stars”, “five stars”, etc.)

**Topic 19** – the shoes that consumers find comfortable are: Puma Classics, Femmes Mostro Perf Ext Shoes, Soleil

Fashion Sneakers.

**Topic 20** – parent consumers who bought shoes for their daughters felt the quality of Puma Girls’ Shoes were low,

but were overall satisfied with the style.

*Pre-Processing Steps*

First, I chose Puma as the popular clothing and shoe brand to analyze through the Amazon product and review data. Puma is the third-largest sportwear manufacturer in the world and has been around for over seventy years, so it was easy to assume there was extensive information about their products on Amazon. Using Amazon’s meta product data in conjunction with the reviews dataset, I was able to identify the ASINs associated with this brand and leverage the meta data to only contain Puma products. Next, with that list of ASINs, I extracted the 27,149 reviews Amazon users had written about Puma products from the second dataset and processed those into a text file for analyzation. Needless to say, this produced an overwhelming about of data. So, it only made sense to further process and potentially segment this data to get the most insights.

I wanted to find out what consumers liked and did not like about Puma products. When were they buying these products? For who? And, at what price? For this reason, I chose to segment the data by ‘stars’ and focus on the highest-rated and lowest-rated reviews. This was completed easily by separating the data with a for loop and performing topic models on two critical score ranges: low-star ratings (1 or less) and ratings that are positive (4-5). From a marketing standpoint, it is important to look at reviews that are critical. With this information, I felt I could produce the best reflection of what Puma consumers are enjoying about their products (from the 4-5 star category) and what they were not (within the 1 star or less category).

*Models Steps*

I decided to use k-means clustering in Python and performed topic modeling on the two segments to reveal the most popular topics in each cluster. Assuming that the number of 1 star ratings is equal to the number of 2 star ratings and so on, the two segments I had created (1 star or less and 4-5 stars) would each hold about 20% of the total review data, (or 40% total). Since these segments only represented a portion of the entire dataset, I did not feel the need to ask the model for any more than 20 topics each (40 total for both segments) and a parameter of four words per topic. Many different cluster numbers and parameters were explored, but this combination made the most sense in trying to extract meaning from the topic models. Utilizing k-means clustering in Python, I was able to create two topic models (see above) and two output files which contained the organized reviews per cluster. With these, I was able to extract the following actionable marketing and product insights:

*Marketing & Product Insights*

*Attributes that people like about our products:*

Referencing the topic model which encapsulates the reviews with four stars or more, it seemed that the majority of positive reviews were written by men, or about Puma’s men’s products. Topic 1 tells us that consumers were impressed by a specific Puma watch, Men’s PU105P2.004 Futuristic Watch, and its style and fit. The attributes that consumers are most impressed by should be things we choose to accentuate in our ads. For example, quality of fit and its unique style should be attributes we leverage in future campaigns for Puma watches.

Moreover, (and again mostly reported by males), the word “comfortable” appears again and again in the model for a variety of men’s footwear. Topic 6 references the comfortability of Puma Men’s Suede Sneakers; as well as Topic 9 which shows that people love the quality of Puma Men’s El Rey Cross Perf Leather Slip-On. Future advertisements for Puma Men’s footwear should showcase this feature of “exceptional comfortability”.

For the women in this topic model, it seems that their favorite Puma shoes were not so much the sneakers – but rather the Brand’s more fashionable footwear. Female consumers wrote positive reviews about Puma Classics and Soleil Fashion Sneakers (Topic 19). In Topic 7, we see extensive positive feedback about the comfortability of Puma’s heels, flats, and wedges. Since - for women - fashionable shoes usually means sacrificing comfort, future Puma advertisements should highlight the fact that women won’t have to sacrifice comfort for fashion anymore when wearing Puma shoes. (This is huge!)

*Attributes that people dislike about our products:*

A recurring theme across both topic models (1 star or less and 4 stars or more) was that consumers had wished they had ordered ½ size larger, or more. In the first cluster, we see this in Topics 2, 6, 8, 11, 12, 14, 15, 18, 19, and 20. According to this model, 50% of 1 star reviews on Puma products are in reference to the small sizing of their shoes. Moreover, in the positive review model (4 stars or more), we see the words “small” and “narrow” again and again in Topics 5, 11, and 15, regardless of the fact that these reviews were overall positive. We might choose to readjust Puma’s sizing charts to help customers select the right size for them and/or re-adjust the products’ actual sizing down by ½. This way, consumers will receive better-fitting shoes and the company will reduce the amount of returns and complaints received.

Topic 4 of the lowest-ratings topic model suggests that the “water-resistant” feature on Puma watches is not as effective as consumers were led to believe. As such, we’d want to avoid making claims in our ads that say the watch can withstand water.

*Purchase Occasions:*

Topic 16 in the positive-review topic model shows that parents (both moms and dads) often buy Puma shoes for their sons. More specifically, the shoes they are buying are athletic styles, like cleats and tennis shoes, usually for whichever sport their son is playing. An understanding of when and why Puma consumers are buying our products provides guidance for when and how we should advertise them. I’d suggest from these data implications that we should target our advertisements to parents with young children, and boys ages 5-17, since some reviews mentioned the word “teenager”. These advertisements should also show adolescent boys playing a variety of sports in Puma footwear, while highlighting how the shoes can increase their performance, but are also quality, practical, and comfortable.

In Topic 20 of the same cluster, it shows that parents are also buying Puma shoes for their daughters, but are having a different experience with the products. Puma’s Girl’s stylish shoes – like flat’s, slip-ons, fashion sneakers, etc. – are generally leaving consumers satisfied. Generally, we see reviews complimenting these products’ style. However, when parents buy shoes for their daughters intended for playing sports (cleats, skate shoes, etc.), they are often disappointed with the quality and small-size of the shoe overall – regardless of its fashionable look. If Puma does not feel the need to change the production process of their athletic shoes, I suggest focusing marketing efforts towards young girls on Puma’s fashion line. As previously mentioned, it is hard for grown women (and perhaps young women, too) to find fashionable shoes that are comfortable and wearable. The same advertising approach could be taken here, channeling efforts towards showcasing this unique attribute when advertising Puma Girl’s fashion footwear line.

*Product Development/Improvement Ideas:*

In the negative review topic model, Topic 4 shows that consumers want a truly “water-resistant” watch. We might suggest to our product development department that we create dependable water-proof watches, as that’s what customers are really looking for.

In the same cluster, looking into Topic 16, we see that consumers are itching for a durable sports bag from Puma that can hold all of their gear and stays in good condition. This is another suggestion to our product development department: to increase the quality and durability of Puma’s bags – zippers, flaps, straps, and all.

*What’s Not in the Data:*

Personally, I assumed Puma’s main source of revenue and brand power came from their clothing – things like sweatshirts, sweatpants, and accessories. However, I was surprised to see that basically the entirety of Puma reviews from Amazon were about their shoes. Not just athletic sneakers, but a colorful variety of all the different footwear styles Puma offers. Therefore, we can assume one of two things: first, loyal consumers to Puma love their shoes more than their clothing lines, or second, that the consumers on Amazon who are purchasing Puma products are there to buy shoes, and Amazon is their channel of choice for doing do. What we know for sure, is that consumers do want Puma footwear, and the quality of that product is important to them. This should shape our ideas about what products we are selling on Amazon, how our customers are perceiving these products, and how Puma should market those products on Amazon.

*Diving Further*

Although not part of the original datasets, demographic information (specifically gender information) about the reviewers would have been useful in analyzing the Puma data, in my opinion. I felt - while working through the dataset, topic models, and clustered reviews – that the majority of negative reviews were written by women or about Puma’s Women’s footwear. Moreover, the majority of positive reviews seemed to be written by men or about male products. Therefore, I believe demographic data would have been interesting to include, in an attempt to see if there is an existing correlation between gender and satisfaction with Puma products. To either prove or disprove my assumption that Puma is more likely to create a content consumer when he (or she) is buying a product from their Men’s line.