

Amazon Product Reviews Topic Modeling Project

Brand: Frye

Author: Annie Frenette

Topic Model Topics

Super Negative Sentiment	Negative Sentiment	Positive Sentiment	Super Positive Sentiment
1. Bag Purse Leather Like	1. Boots Pair Buy Wore	1. Get Boots Foot Good	1. Boots Return Get Zipper
2. Boot Foot Boots Get	2. Calf Boot Extended 34	2. Color Amazon Bag Ordered	2. Boots Disappointed Like Quality
3. Boots Pair Disappointed 34	3. Boots Boot Size Like	3. Boots Leather Size Heel	3. Boots Put Feet Dissapointed
4. Size Boots Small Return	4. Beautiful Fine Boots Difficult	4. Boots Big Size Comfortable	4. Shoe Shoes Ballet Flats
5. Shoe Shoes Uncomfortable Comfortable	5. Shoes Shoe Feet Size	5. Calf Boot Extended Boots	5. Small Made Boot Well
			6. Beautiful Soft Leather Boots
			7. Size Shoe Half Fit
			8. Boots Pair Great Leather
			9. Boot Calf Get Extended
			10. 11 size 10 Black

Topic Model Descriptions

The above topics are organized across different sentiments of the text reviews. The first three sentiments, Super Negative, Negative and Positive all have five different topics and the Super Positive Sentiment has ten different topics. These topics were derived from performing a topic model on reviews that gave off a certain sentiment. The process for this analysis is described further below. From a marketing and advertising perspective, it is very important to analyze every single topic that is created from the analysis. In terms of which topics to gain the most insights from is determined by the content of the topic themselves. It is also very important to focus on aspects of your brand that are garnishing a very negative sentiment and a very positive sentiment. I have described a select number of the topics below.

Super Negative Topics

Topic 1- The shoes or purses made by Frye have stiff, hard or unsmooth leather and users have noticed a decline in leather quality over the years.

Topic 3- There is a general sentiment of overall disappointment especially related to sizing issues. There also is concern with Amazon's exchange and return policies regarding these issues.

Topic 5- There is an overall feeling of purchasers not finding the product comfortable.

Negative Topics

Topic 2- There is a general unhappy sentiment with the way Frye determines how extended calf sizing works and many purchasers are unhappy with the fit of Frye shoes around their calves.

Topic 4- This topic describes purchasers' sentiments toward the difficulties they have with walking in the shoes as well as putting them on and taking them off.

Topic 5- This topic describes the overall sentiment of users being upset with the various sizing issues they had when purchasing Frye shoes.

Positive Topics

Topic 2- There are various types of reviews regarding the shoes but there is an overall sentiment towards the process with Amazon. Some users are happy with their boots but disappointed with Amazon whereas some users are happy with their boots but pleased with Amazon's process.

Topic 5- This topic describes users' sentiments toward the calf sizing issues but are overall happy with their purchase.

Super Positive Topics

Topic 6- Purchasers are very happy with their Frye purchase and are impressed by the quality of the soft leather and the aesthetic of the shoe.

Topic 10- Users express various frustrations with the sizing issues especially for size 10 & 11 boots especially for the Men.

Topic 2- Many Frye brand loyalists discuss their love for the brand but are disappointed in the decrease in quality with specific shoes or overall.

Topic 4- This topic relates to the overall positive sentiment users have towards their ballet flats.

Topic 5- This topic has an overall positive sentiment regarding how Frye Boots are made well especially related to the Frye Women's Harness 12R Boot.

Data Processing

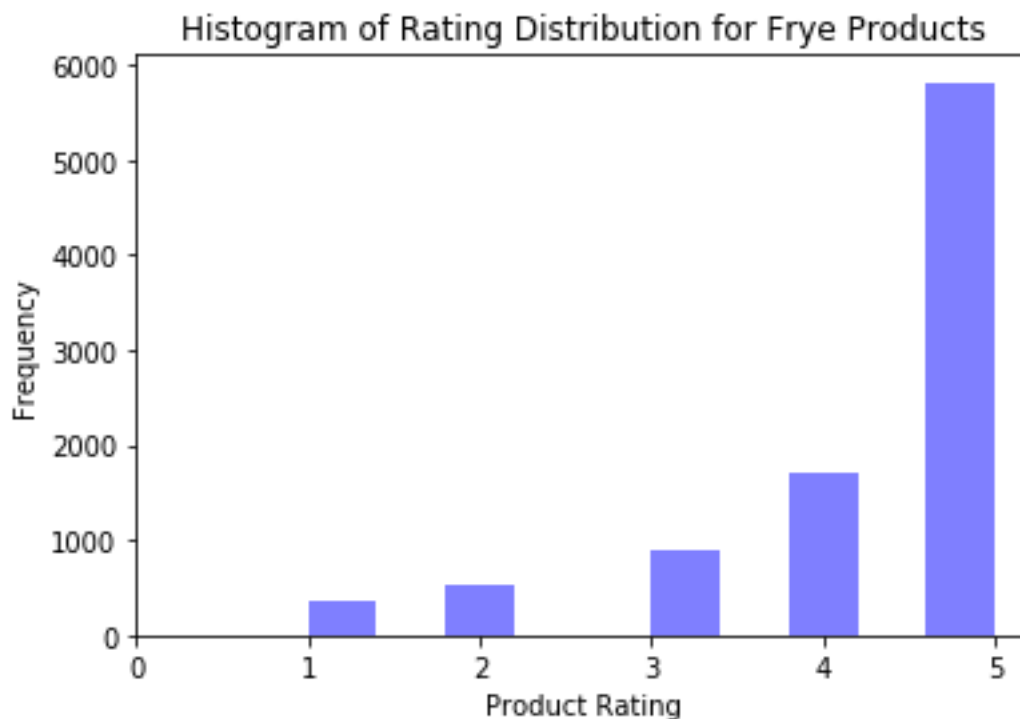
The data for this analysis was provided in two JSON Files. One JSON file contained metadata about the products listed on Amazon in their Clothing, Shoes and Jewelry section. The most important aspect of this data set was the category and the ASIN. The first step was to parse through this JSON and create a python dictionary of all the categories in this dataset. These categories consisted of brand names as well as Amazon specific categories. These categories acted as the key for the dictionary. When creating this list, I established the number of ASINS per category and had this be the value for the key. There were 2773 different categories. I then selected Frye as my specific brand which had 2205 products. The next step was to parse through the dataset and create a set of all the Frye ASINS. These ASINS are the unique Amazon identifier code. After I had a set of the ASINS, I had to merge this with the review data.

The review data consisted of the ASIN, ReviewerID, ReviewerName, helpfulscore, overall rating, review text, a summary as well as the time of the review. I created a dictionary that would contain all the reviews for Frye Products that were at the ReviewerID.ASIN level. I parsed through the review data set and only added the reviews to my dictionary if they matched an ASIN from the Frye ASIN set that was previously created. After this was done, I had a dictionary containing all the reviews for Frye Products.

Modeling Steps

I chose to use a K-Means Clustering model to perform my Topic Model Analysis. This method focuses on putting some essence of structure to a large set of unstructured text reviews. The overall goal of a Topic Model is to form clusters that have similar documents within but are dissimilar from other clusters. The K-Means technique leverages the concept of TF*IDF or Text Frequency interacted with Inverse Document Frequency. This allows for terms or themes that are very common in the specific document or review but not common overall to be emphasized.

The first step I took was to do an overall topic model on all the Frye Reviews. From these results, I decided it would be best to segment the reviews further. My first idea was to segment by the rating of the review which is scored from 1-5 as whole numbers. After examining the distribution (see figure 1. below), I determined that it would be best to choose a different segmentation method. While there is a majority of positive reviews the examination of them told a different story. Very often, a highly rated review would discuss frustrations that they had with the initial sizing of the shoes and the overall buying process. I was hesitant to proceed as I wanted to gain more insights on the product as opposed to the buying process. With this assumption in mind, I decided to use a sentiment analysis on the text of the reviews.



I chose to use the VADER sentiment analysis package. This sentiment analyzer determines the polarity of a string of text. It provides scores on the neutrality, positivity, negativity and a compound score of the text. The compound score normalizes the scores of the words in the text. The package suggests that a compound score of less than -0.05 is a negative sentiment and a score greater than 0.05 is a positive sentiment. The neutral texts lie in between that range. Using this tool, I created four different segments to topic model on. The segments were super negative [**compound score < -0.30**], negative [**compound score \geq -0.30 and \leq -0.05**], positive [**compound score \geq 0.05 and \leq 0.30**] and super positive [**compound score $>$ 0.30**]. I chose not to model on reviews that were neutral. For each segment, I performed K-Means clustering on that segment's review text. For the first three categories, I chose to use 5 different topics. The super positive segment was modeled with ten topics. The reason for this was that of the 9324 reviews, 8190 of them had a positive sentiment. Once the topics were established, I exported the documents for each topic into a classified text file. These text files consisted of only the reviews that matched that specific topic. This allowed for each topic to be

analyzed in depth to gain insights related to the product and how to leverage them to make marketing decisions.

Marketing & Product Insights

As a marketer, it is very important to have a grasp on what type of feedback your product or service is receiving. Very often, this feedback is delivered in an unstructured way and topic modeling provides an easy way to make sense out of the feedback. The first insight from these reviews is to take in what the overall ratings and sentiments of your products are. The good news is that a very large majority of the reviews are receiving high ratings and have a positive sentiment. The best approach to take to leverage these findings is to do a deep analysis on both ends of the spectrum. Analyzing negative feedback can often lead to marketers determining different ways to position their brand as well as translate findings to R&D, Finance or other departments. With positive feedback, marketing and advertising materials can be strengthened for areas that consumers love and appreciate. Positive feedback also helps discover new areas for marketing departments to focus on. It is important to determine if changing negative sentiment has a greater ROI than improving and maintaining positive sentiments. From this topic model, there were a few major themes that could be turned into marketing decisions and potentially higher revenues.

Leather Quality Issues & Positive Feedback

Both ends of the spectrum, had topics that discussed the leather quality of the boots. In *Topic 1*. of the super negative segment, purchasers mentioned how they found the leather stiff and uncomfortable. In *Topic 6*. of the super positive segment, purchasers raved about the quality of the soft leather in these shoes. *Topic 1*. of the super positive segment discussed how brand loyalists were noticing differences in the quality of the leather over time. Many of these reviews mentioned concerns over where the product was being manufactured. There seemed to be an overall distaste for products being manufactured in China as Frye historically has been known as an American brand.

Overall, there is more positive sentiment than negative sentiment and this positive sentiment should be used to drive advertising campaigns. These campaigns can focus on the quality of the leather and potentially use the historical and nostalgia aspect of creating leather shoes to retain customers who are looking for the quality that Frye has always had. However, if there really is an issue with the quality of the leather deteriorating this may be something that should be passed on to other departments.

In relation to the topics discussing the leather quality, *Topic 5*. in the super negative segment and *Topic 4*. in the negative segment relate to comfortability of the product. Very often these reviews mentioned blisters and the inability to pull them on which are commonplace with high quality leather shoes. Frye could produce a series of educational materials to accompany a purchase that discusses the proper way to break in and care for leather shoes to avoid uncomfortable moments. To take this further, if Frye published these materials and leveraged SEO and AdWords these materials may show up when all consumers are searching for ways to break in leather boots. This would make Frye's brand more visible to consumers of leather goods and potentially convert consumers of competing brands.

Sizing Issues

Every single segment contained a topic that related to the sizing of the boots. There didn't seem to be any specific trend of sizing issues with two exceptions. First, *Topic 10*. of the super positive segment was related to the men's shoes and specific size ranges. In this case, it may be best to send this information to another department to combat these issues from the initial creation of the shoe. There may be an underlying problem with the way Frye determines size 10 shoes. The other exception was related to calf sizing. Frye offers extended calf options and the overall sentiment across *Topic 2*. in the negative sentiment and *Topic 5*. in the positive sentiment was that the calf sizing was very often wrong and confusing. The fix here would be to have better descriptions that clearly lay out how the calf sizing works.

Aside from these exceptions, there are two insights from the sizing topics. First, it may be very profitable to incorporate an add-on to their website that deals with the issue. Many shoe brands offer a True Fit calculator that helps consumers select the right size shoe. These calculators ask the consumer questions about the fit of their existing shoes and determines which Frye size will fit them best. If Frye were to implement these, many users' issues of having to return their purchase would be eliminated. The other insight is to make better educational materials on Frye sizing available to consumers. Both of these approaches will help to eliminate costs associated with returns and exchanges.

Products of Interest

One of the first things I noticed when analyzing the reviews with positive sentiments was a very common ASIN. This ASIN was B000AM211U or the Women's Phillip Harness 12R. These are very clearly a well-liked shoe. Reviews consisted of consumers returning to buy their 2nd or 3rd pair as well as first time purchasers. The first way to leverage this information is to consider the excitement these shoes garnish and create marketing and advertising campaigns surrounding these specific shoes. It may be beneficial to analyze the target demographic of purchasers and target look-alike audiences to gain new customers. This approach can easily be replicated across other popular products Frye offers. Additionally, Amazon currently offers 14 different color options. Frye should target all past purchasers of these shoes when they design a new color to retain customers. These campaigns could come in various forms but it is very important to keep the brand relevant to past purchasers over time so they continue to shop at Frye.

There was also an entire topic related to ballet flats, *Topic 4*. in the super positive sentiment. This potentially is a great way for Frye to tap into a new market. Frye is typically associated with western style riding boots. This topic analysis shows that consumers are also a fan of ballet flats. One specific pair that was common in this topic is the Regina Ballet Flat. Frye could target consumers who typically aren't interested in purchasing western style boots and open a whole new way to position their brand. This should run concurrently with their advertising and marketing related to their western rugged appeal to not lose existing customers.

Conclusion

This topic model analysis for Amazon Review Data of Frye Products begins to tap into a whole new area of business intelligence for the brand. Without the ability to efficiently add structure and segment reviews, it is hard for brands to have a good understanding of what their consumers sentiments are. Topic Modeling provides an initial glance into what consumers are thinking outside of structured surveys and purchase history. There are so many avenues to continue the analysis with the results produced by a topic model. While there are many use cases for topic modeling in marketing, many other business functional areas should consider using this approach for their large unstructured datasets.