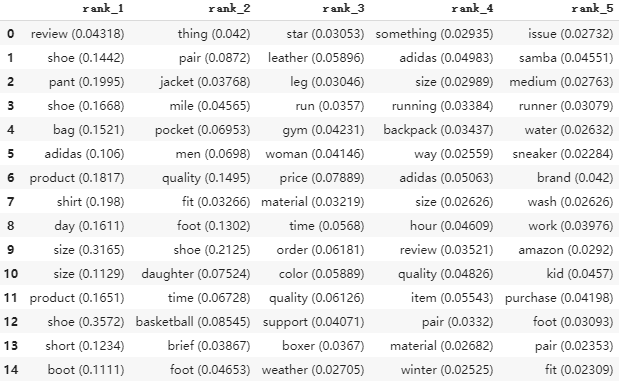
Topic Modeling

Yuxiang Wang

I selected ADIDAS as the brand to analyze. Because adidas has 9457 ASINs and 27511 reviews, I set 50 topics by using Topic Modeling. I also created a table to show the score of each word in each topic.

**Topics:**

|  |  |
| --- | --- |
| 0: review thing star something issue  1: shoe pair leather adidas samba  2: pant jacket leg size medium  3: shoe mile run running runner  4: bag pocket gym backpack water  5: adidas men woman way sneaker  6: product quality price adidas brand  7: shirt fit material size wash  8: day foot time hour work  9: size shoe order review amazon  10: size daughter color quality kid  11: product time quality item purchase  12: shoe basketball support pair foot  13: short brief boxer material pair  14: boot foot weather winter fit  15: shoe foot heel sole toe  16: gift husband son christmas bought  17: shoe golf course round foot  18: shoe support foot arch bit  19: store amazon pair price adidas  20: pair shoe year time adidas  21: lace thing way problem top  22: great fit feel price highly  23: product amazon order shipping return  24: shoe price money deal time  25: muy la que el para | 26: shoe toe foot box size  27: shoe foot pair running support  28: sport comfy perfect work bra  29: shoe lot color compliment people  30: size shoe foot bit width  31: color picture blue pink stripe  32: sock foot bottom way thing  33: adidas shoe brand nike love  34: adidas model shoe pair design  35: cleat soccer ball turf game  36: month time wear week pair  37: shoe tennis time adidas year  38: shoe water bottom hole beach  39: shoe weight gym training running  40: shoe style color fit comfort  41: one quality bit something material  42: shoe son price great quality  43: watch time band button wrist  44: sneaker pair adidas time style  45: son soccer year daughter school  46: sandal foot pair slide house  47: shoe trail running traction grip  48: size order half fit men  49: adidas material quality side stripe  50: shoes great buy good comfortable |



**Topic Description:**

0: It will bring an issue to review something by star.

1: A pair of adidas leather shoes is good for practicing samba.

2: If the leg length is medium, the size of pants and jackets will be the same.

3: Running shoes can be better for the runner to run more miles.

4: When going to the gym, there must be water in your bag or the pocket of backpack.

5: There are different ways between Adidas man and woman sneakers.

6: Adidas, the brand has good product with satisfied quality and reasonable price.

7: The shirt’s material is easy to wash and always fit to size.

8: Feet are still comfortable after a whole day of working time like nine hours.

9: Orders of shoes in Amazon have more problems about size in reviews.

10: The kid shoes are good in size, color and quality to be sent to daughter.

11: Since the item has good quality, people will purchase the product for several times.

12: A pair of basketball shoes can better support your feet when playing basketball.

13: The pairs of shorts and briefs for boxer have good materials.

14: The boots fit the feet under the weather in winter.

15: Shoes with heels are comfortable with feet, especially sole and toes.

16: Adidas would always be bought as Christmas gift for son and husband.

17: Shoes are fit for feet to attend gold courses for several rounds.

18: Some shoes don’t support the arch of feet a little bit.

19: The price of adidas in amazon is paired with the price in store.

20: People always buy a pair of adidas shoes several times a year.

21: Things and ways to do with the lace would be the top problem.

22: The feet feel highly fit and the price is great.

23: Amazon is responsible for shipping order and return product.

24: The price of shoes deals with money problem many times.

25: (This one could be from a Spanish speaker) The price can be very reasonable for the products.

26: The shoe box’s size does not fit the foot toe position.

27: A pair of running shoes can support the feet during running.

28: Sport bra works perfectly and comfily.

29: People have a lot of compliment shoe color to choose.

30: The size of shoes does not fit the width of the feet a little bit sometimes.

31: The blue and pink stripe on the picture are the most popular colors.

32: It’s an important thing to have socks support the bottom of feet in its way.

33: Adidas and Nike are the shoe brands people love most.

34: A pair of Adidas shoes are depended on the model and design.

35: Cleats are important in soccer and other ball game on the turf.

36: A pair of shoes can be worn many times every week or even every month.

37: People purchase tennis shoes several times a year.

38: Though the bottom of shoes does not have holds but still cannot prevent the water so that should not be used to walking in the beach.

39: Running and training shoes can be used to do weights in gym.

40: The shoes with good looking styles and colors are also fit the feet and feel comfortable.

41: The quality is the top one issue which quite a bit of something with material.

42: The shoes for son have great price and quality.

43: The wrist band has to use button to watch time.

44: People can purchase different styles of adidas sneakers many times in a year.

45: Adidas shoes can be gift for son playing soccer and daughter going to school in a new year.

46: Feet with the pair of sandals are easy to slide inside house.

47: Running shoes should have strong traction and grip for the trails.

48: Half of the orders for man have the fitting size problem.

49: The Adidas shoes with side stripes have good quality with good material.

50: The shoes are good to buy since they are great and comfortable.

**Preprocessing Steps:**

The goal of text preprocessing is to make the texts easier to perform eventual analysis method. In order to achieve this goal, I loaded Jason file which contains all the reviews of Adidas. Because I stored the reviews as lists, I converted the reviews texts to dictionary as the first step. Then I used “tmtoolkit” package to process the data. I made the sampling reproducible and added the documents to Corpus. Then I checked the length of Corpus which is 27511 and is the same as the length of reviews. During the whole preprocessing, a corpus of document is tokenized, which means I split the document into individual words and these tokens can be transformed, filtered or annotated. I used TMPPreproc class method to token the documents and change the documents to lowercase. I also removed the special characters and added stop words (‘heep’, ’nt). Meanwhile, I used ‘vocabulary ()’ to find the number of times each token appears in the corpus because the vocabulary is the set of unique tokens in the corpus. The last step I did is to filter the documents by filtering the pos tagging, removing numbers and shortening tokens. After the whole text preprocessing, I found the texts of reviews decrease to 18503, which is working as intended.

**Models Steps:**

I used LDA topic model to get 50 topics of adidas’ reviews. Firstly, I suppressed the “info” messages and warning from LDA, then set the data which I already did in the text preprocessing and fixed the hyperparameter. After that, I printed top5 words with their score (TF-IDF) to check the most popular words. Finally, I combined the words in each topic to finish a sentence.

During the creating model part, I used 3 methods (arun\_2010, cao\_juan\_2009, coherence\_minmne\_2011) to evaluate my model since topic modeling typically requires defining some parameters beforehand. The plot shows all metrics that were calculated during the evaluation. When alpha = 1/k and beta = 0.1, I found the arun\_2010 and cao\_juan\_2009 have the maximizes for values of K greater than 20. As for coherence\_mimno\_2011, it maximizes for values of K between 5-10. It is like that even for a large k, the calculated metric will be very low, because the “density” in the topic-word distribution will still be very high. Then I changed beta to 0.01, I found arun\_2010 and cao\_juan\_2009 didn’t change too much, but coherence\_mimno\_2011 had the maximizes for values of K between 10-20, which might mean that a higher value for beta also leads to a lower information density in the topic-word distribution when using a model with many topics. Finally, I kept the beta=0,1, changed alpha to 1(10k), and I found cao\_juan\_2009 and coherence\_mimno\_2011 have the maximizes of values of K between 0-5. Because the objective cases, I only can say my model has 50% to valid.

**Marketing & Product Insights:**

As for the attributes of Adidas liked by the costumers, quality and price could be considered as the most important aspects. Everyone likes the product with good quality but low price, which is realized in Adidas. For instance, topic 6, 10, 11 and 42 all mentioned about this point. Meanwhile, there are specific different types shoes for different sports, which is an attractive for the costumers as well. For example, topic 3 indicates the good use of running shoes and topic 12 states the basketball shoes. Since each sport needs specific strength function, shoes can be a necessary part to support the athletes. Moreover, the style and color are also attractive the costumers, especially the side stripes. Topic 31 and 49 both suggest this pattern, which could be regarded as the symbol design of Adidas.

As for the attributes people dislike, one of the main problems might be the size. Topic 48 implies that half order of man shoes will be returned since the size problem. Hence, we might need to have a more clear and accurate size guide to help the consumers find the correct and fit size when shopping online. Besides, topic 46 also complains the sandals are easy to slide in the house, which has much potential safety risks and need to be considered seriously. Also, topic 43 mentioned the wrist band had to use button to watch time, which is not convenient during the exercising. The better way might be shown as motion, like showing the time when rising the wrist.

Across the topics, many people will purchase Adidas products several times a year, which could be understood as some essentials in daily life. As for some special occasions, topic 16 pointed out the Christmas and topic 45 mentioned the new year. Therefore, Adidas products could not only be used as daily necessities, but also used as gifts for family. Overall, we’d suggest people will purchase Adidas products as needs since its good quality and reasonable price. They would also purchase Adidas as gifts depending on its style and functions.

As I mentioned above, people always have problem choosing the correct size, so that Adidas need to do more experiments or practices to improve the size guide to make it more precise and accurate. Also, the sandals are supposed to increase the traction, otherwise people might have some accidents. The wrist band could be improved to be more user-friendly, like replacing the button as motion.

By looking at these data, consumers consider the price is fair with its great quality. Especially, the topic in Spanish also indicates the price is reasonable for the products. Therefore, I would suggest the price is fair for most consumers all over the world. Meanwhile, many topics relate the quality with the price, then I’d suggest consumers are satisfied with the product they got with how much money they paid.

Since Adidas has a lot of fashion items, especially in the clothing area, I did not find topics about this. Through the topics, people more focus on the shoes. Thus, they could increase the promotion on clothes, which is one of their advantages. To be honest, Nike has a more solid costumer group on shoes, while Adidas could have more design and style on clothing as the fashion trend. Also, as for the shoes’ design, the consumers only mentioned the side stripes, which is the classical one, so that Adidas could have more innovative design on shoes as well.

Overall, both the products and price of Adidas are close the most people, but they actually have some lines which are more innovative for the young people. They could increase the conscious by using social media, printed ad, create a campaign, or to do more collaboration. The brand could be more youthful and creative, but not only the daily needs for most people.

**Being Creative & Segmenting Data:**

I selected “By Sales Rank” to find the products that are doing well in sales or not. The final result is pretty similar with the previous one. The attributes need to be improved are also about the size, waterproof, wrist band, and sandal and the improvement could be the same as those I mentioned in Marketing and Insights. Though the results are the same, it better suggests that these aspects are really the main problems.

**URL:**

Adidas part 1: <https://colab.research.google.com/drive/1awLBZtCr0LR3SEdZZa5kZ5KgXKX_mgjM>

Adidas part 2:

<https://colab.research.google.com/drive/19xQezWlDK_Y_-LKtNJPUBd3XqFEb7EYW>

Adidas\_SalesRank part 1:

<https://colab.research.google.com/drive/1TYIBv7np9sK6eowYOu_pZAMoy23fBUNP>

Adidas\_SalesRank part 2:

<https://colab.research.google.com/drive/1kYmxVREzpje11t31IrYgXmEECx4XhwQh>