Text mining is a very powerful tool if used correctly. It was listed in the Gartner Hype Cycle in 2012. Unfortunately it did not meet its expectation. Some seem to say that text mining is a powerful tool. Yet it did not meet its full potential due to the lack of knowledge that we have on it. I, myself believe the exact opposite. Text mining can fulfill it’s potential simply if it’s used correctly. This project is all about that. With the right hands, text mining can allow us to build a business from scratch. With a combination of text mining and social big data, We will show how we can build a customer service team from scratch. In fact, Khalil and I were consulted by MEA because they wanted to put in place an airline customer service, but had no idea how to do it. As a result, we went to explore some Airline customer service examples on twitter. We took American Airlines and Delta Airlines as references and we decided to retrieve tweets from their customer service team. In addition to that, in order to fully solve the problem given to us by MEA, we broke it down into several questions:

* What is the average tweet length of an employee at a customer service?
* What is the frequency of the tweets of an employee at a customer service?
* When does the team most frequently tweet?
* Which method are used by the team in order to assist the customer?
* What words and phrases are most commonly used?
* What is the employee’s tone? Is it positive, negative or neutral?

We are going to answer these questions in order to help MEA with some key decisions regarding their customer team

Plz note that this word document is just here to explain the ppt into more details. So it comes as a support.

1. What is the average tweet length of an employee at a customer service?

First thing first, we thought that it would be interesting to note what is the average tweet length of a customer service team. We realize that in this histogram, the average tweet length is between 100 and 120 depending on the Airline.

An very straightforward conclusion would be that the customer service team uses an almost full length tweet in order to assist customers. (Doc 1)

1. What is the frequency of the tweets of an employee at a customer service?

After that, we decided to check what is the frequency of the tweets of an employee at a customer service. A simple calculation based on the time of the tweets suggest the following. Presented by these histogram, we can notice that on average, each 220 seconds (in both airlines) we can expect 1 tweet.(Doc2 )

1. When Does the Team most Frequently Tweet?

Next up, we went to check when do the customer service teams most often tweet through out the day. We Found some interesting and similar results for both companies. If we take a look at these histograms we can notice that the customer service team tweet frequently between 10 am and 5 am, with tweets averaging between 150 and 250. This is not the case for the hours between 5 and 9, where we can notice that the customer service team barely tweet (75 tweets almost).

This is quite interesting, because a huge insight can be taken from this simple but effect analysis. We can tell MEA, that they can smartly manage their customer service team. A large part of their customer team can be reserved between 10 am and 5 am. The other small part will take charge for the rest of the day, when the work rate seems to dip down.

1. Which method are used by the team in order to assist the customer?

Next up, we wanted to see what methods are implemented by the customer service team in order to help customers. We tried to search the tweets that contains the word “DM”, “http” (symbolizing a tweet), and a number. We calculated the percentages of the tweets containing each word. We found out the following.

In the American Airlines case, we can see through the histogram that x% of tweets contain the word “DM”, y% contain the word “http”, z% contain numbers. On the other hand, in the Delta Airline case, x% contain the word “DM”, y% contain the word “http”, and z% contain numbers.

We can witness two different approaches. American Airlines like to focus on DM. Delta Airline focuses more on their website. It is up to MEA to decide. We just presented the different methods.

Another think to look is if these methods are implemented together. These histograms show exactly that. We can see that American Airline use these method together. This is not the case with Delta airlines where we can see that 15% of the tweet contain both links and DMs

1. What words and phrases are most commonly used?

Next up, we wanted to see what words are mostly used and so we did a wordcloud for both companies. Wordcloud allows us to get a brief overview of the most words used. The bigger the word, the more common its appears. We can see a certain similarity. In both companies, the wordcloud tells us that the words “please”, “our”,”sorry”,”thanks” … are mostly used by both customer services. Very interesting…

A more precise study with histograms will offer us number. In the American Airline company, the top 5 most used words are

* Our with 654 words
* Please with 537 words
* Will with 407 words
* Sorry with 350 words
* Can with 335

In the Delta Airline case, the top 5 most used words are

* Please with 746 words
* Thanks with 459 words
* Sorry with 430 words
* This with 352 words
* Can with 310 words

We then checked word association between the top words. For further explanation, this test will show us with each words, what are the most used terms. For instance, in this example, the word team is most often used “our”, “kind” etc. We’ve done this for multiple words for each company. You can check our PPT. With these word associations, we can witness and take an idea about the structure of sentences used by customer services. Some examples would be:

* “Please send details”
* “Thanks for your patience”
* “Sorry to hear”
* “Happy to assist you”
* “Confirmation Dates”

From here, it is easy to figure out how the team behave. They are very respectful toward the customer. But also they are very reassuring. This is essential because this will give to the customer a sense of security that anything will be fine and that his problem will be fixed.

1. What is the employee’s tone? Is it positive, negative or neutral?

Finally we have done a sentiment analysis to check what is the employee’s tone. The sentiment analysis test that we used is called the polarity test. The polarity is score from -inf to +inf. If a polarity test is positive, it means that the sentiment is positive. If the polarity is negative, it means that the polarity is negative. Please keep in mind that the function that we us try and plot the polarity of each tweet in normal distribution. So we are focusing on the mean of these graphs. We can see that the mean is 0 on both cases. This can only mean one thing. The tone of the customer service team is neutral. This is to be expected since the customer service team is just trying to help people without expressing any sort of feeling

1. Are they using different approaches?

One last thing to mention is the following. Are these two companies using different approaches? Well this pyramid histogram, at first glance, can seem to plot a lot of differences. But that’s not the case. We can see that some key words are used commonly together such as please, will, help. In that case we can conclude that they are using different approaches. This commonality which only put the same words used in both airlines on a cloud confirms our idea.

Conclusion

That’s that. If you torture a data set large enough, it will confess to anything. That is exactly what we have done. By just analyzing tweets, we were able to almost decrypt how a customer service team behave and work. We were also able to see how many tweets a customer service team create on average. We also managed to find what are the methods that these companies use in order to help people with their problem. Finally, we were able to analyze the tone of the customer service which seems to be neutral.