Analysis of Customer Support on Twitter

This project dataset is focusing around Customer Support queries and responses from the biggest brands and companies.

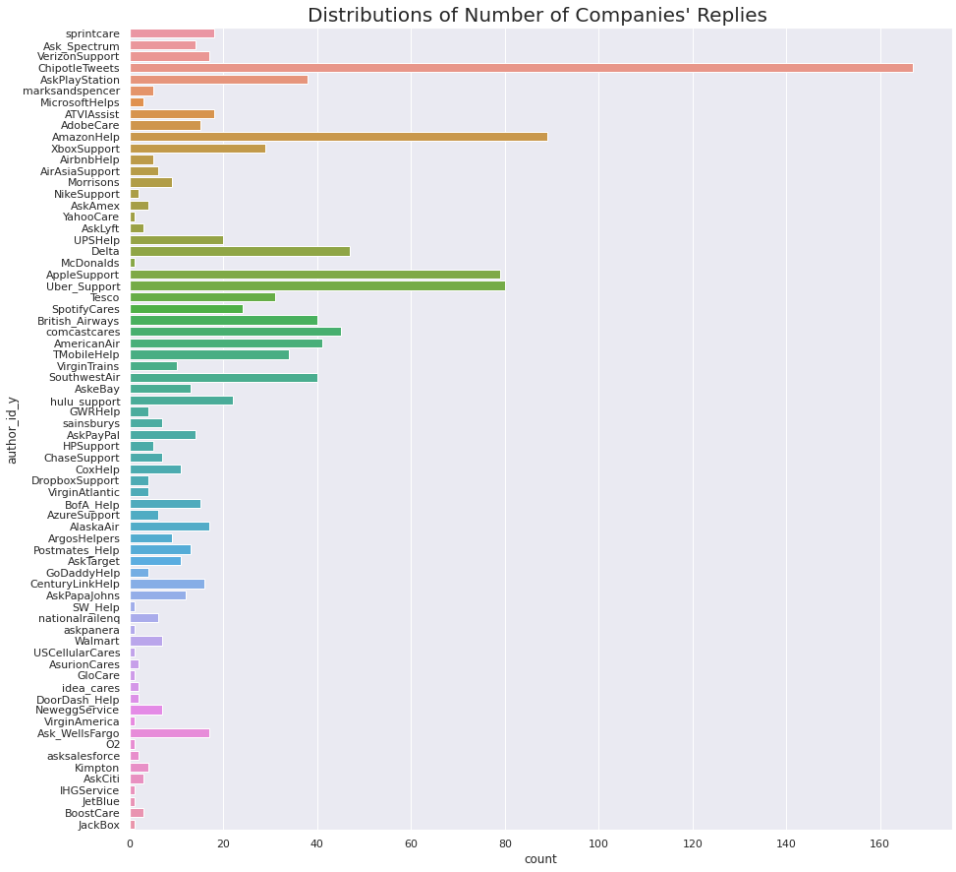
As we learned, in any machine learning task, cleaning or preprocessing the data is very important especially in sentiment analysis use case.

Step 1: To start exploring this dataset, I did Some of the common text preprocessing / cleaning steps are:

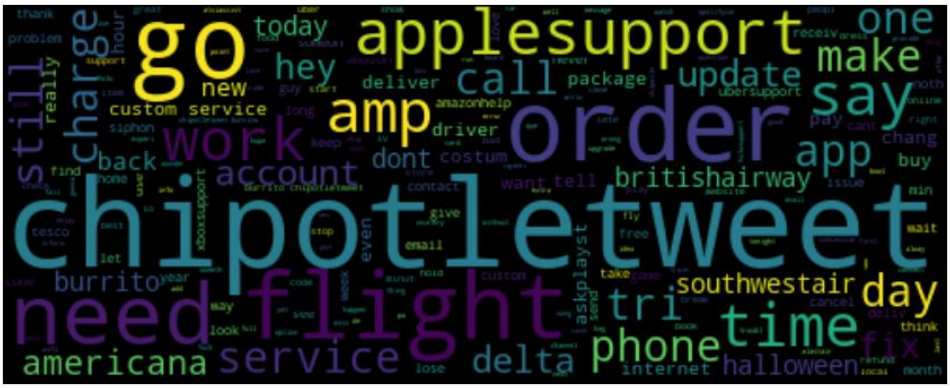
- Lower casing
- Removal of Punctuations
- Removal of Stopwords
- Removal of Frequent words
- Removal of Rare words
- Stemming
- Lemmatization
- Conversion of emoticons to words
- Conversion of emojis to words
- Removal of URLs
- Chat words conversion
- Spelling correction

Step2: I modified the dataframe to make the "inbound" column, which have information on the author of the tweets and whether the tweet was a query or a response. This will get query - response pairs in every row.

In the result of this modification, I get the number of tweets in the dataset for each company and plot the data.



This result suggests that Chipotletweets company has the bigger number of tweets then Amazonhelp followed by Uber and AppleSupport.



And I implement word cloud to visualize the most common words appeared in the tweets of the customers.

Step 3: Using TextBlob to implement Polarity and Subjectivity analysis on the tweets to three parts (Normal, Positive, Negative) Analysis.