EXECUTIVE SUMMARY

Aim: The aim of this project is to provide insights on the modern customer support practices and different topics of discussion in these conversations and the overall sentiment from the customers about each company via conducting Sentiment Analysis based on the tweets generated by the customer support agents and customers on Twitter. Additionally, we will derive the most frequent Topics in the tweets based on Topic Analysis and Modelling.

The Customer Support on Twitter dataset is a large, modern corpus including tweets and replies to help innovation in natural language understanding and conversational models.

The first three questions address the data exploratory questions which involves calculating the Top 10 brands according to the maximum number of tweets posted (including customers and companies).This is solved by combining the customer + company tweets extracted from the dataset. We come to know that the Top 10 companies are AmazonHelp (12471), AppleSupport(5781) , Uber_Support(3801) ,Tesco(2749), British_Airways(2441), SpotifyCares(2326), VirginTrains(2232), sainsburys(2203), AmericanAir(1135) and GWRHelp(1039).

This is followed by calculating the number of tweets by the weekdays (convert the 'created_at' column using str.split function. Thursday, Wednesday and Friday are the days where highest engagement is seen on Twitter. This could be because these days are halfway to the weekend.

The number of customers that specifically requested for support from AmazonHelp is also calculated by extracting it from the 'customer' text. There are 5547 customers who requested support.

The fourth question revolves around Sentiment Analysis for the top 5 brands and computing the proportion of positive vs. negative tweets between the brands using Lexicon-based sentiment analysis. Firstly, we clean the text by defining a function and then creating a separate dataframe including only the top 5 brands. Following, we import relevant modules from NLTK library : VADER (pre-trained on social media data to give an overall sentiment score -1 to +1. AmazonHelp has the highest count of Positive Sentiment (440), followed by AppleSupport(220) with a negative sentiment of 89 counts. We further plot proportion of reviews with respect to the positive sentiments. We can see that SpotifyCares has the highest proportion of positive reviews, followed by Tesco and Uber_Support. However, AmazonHelp and AppleSupport have the same proportion.

Lastly, we determine the number of topics frequently mentioned by customers of AppleSupport via text-processing and LDA modelling (Latent Dirichlet Allocation).This is carried by doing some text pre-processing. After, data cleaning , I have applied Term Frequency representation of the document for LDA modelling. Furthermore, I have trained

an LDA with 10 topics and used Word Probabilities by Topic. The top 10 topics generated by LDA have been visualised using the word clouds.

The benefits associated with this business problem will in turn help the companies better address the problems that its customers are facing and would likely help improve their customer service in future. It would also aid in healthy competition between the brands and propel them to compete based on the positive and negative sentiments of each company.

PRACTICAL IMPLICATIONS

The above techniques such as Lexicon based sentiment analysis and NLTK methods can be used to analyse huge volumes of Twitter data. Additionally, the algorithms can be applied to any kind of textual data such as news articles or customer service chat data. Due to large sums of corpus, it is becoming more and more imperative to gain insight on how customers view the businesses around them via Sentiment Analysis and calculating the positive and negative sentiment score is beneficial as it helps in :

- Strategically implementing new practices in the business adhering to the sentiments of the customers

- Launching new products and services(for instance : new software updates of the phones purchased) based on customer's preferences.

- Competing with other companies by improving their service quality.

- Targeting more audience for business expansion.