MSC BIG DATA ANALYTICS FOR BUSINESS

DOCUMENTATION

SOCIAL MEDIA ANALYSIS
- CLASS 2020 -

TEAM MEMBERS:
CARLOS MONTENEGRO
KRIPA FERNANDEZ
LUCAS OLIVIER

LILLE, FRANCE

DOCUMENTATION

The purpose of this project is to provide a comprehensive analysis of all the twitter buzz about the Citibank. Our main sources of information were the following:

- Tweets on the profile of the username "@Citibank"

- Tweets that mention "Citibank"

- The list of memberships of the followers of the username "@Citibank"

With the purpose of analyzing this data we have used the following techniques seen in class:

- Preprocessing of tweet comments.

- General statistics and geolocation of the tweets.

- Text mining including sentiment, emotion and topic analysis.

The business applications of text mining range from crisis management to evaluation of the ROI in a marketing campaign. In this group project it has been decided to focus on the first following three applications:

- Crisis Management: to avoid reputational damage.

- Lead generation: to adjust the marketing campaign.

- Improve customer service: help manage complaints.

- Improve product quality: in order to meet customer's+needs.

- Evaluate ROI on marketing campaign: helps increase the number of positive discussions.

Some of the limitation that affected our work are the following:

- Getting a great number of tweets from specific periods of time.

- Having an equal sample distribution of the daily tweets.

- Getting multiple lists of membership for the tweeters that have tweet about Citibank.

- Having more variables that could identify different customer segments and a specific customer persona.

With the purpose and restrictions state in the previous section we created the following outputs:

MARKDOWN	SHINY APP
Citibank influencers	Word Cloud
Target audience for Citibank advertisements	Topic Analysis
Top retweeted tweets on Citibank	Emotion Analysis
Citibank vs Competitors in social media presence	Negative & Positive Analysis
Citibank vs. competitors of Citibank tweeters	Negative & Positive vs Time Analysis
Citibank vs competitor's social media presence	
Citibank's response rate on tweets	
Network analysis	

For our outputs, we created three different documents containing our code:

1. The first one which contains the lines of code to download data using Twitter APIs, to clean it and create new datasets summarizing important information for the previous outputs. The main libraries that we used in this step are the following:

> *Rtweet, lubridate, SnowballC, slam, tm, Matrix, tidytext, dplyr, hunspell and purrr.*

2. The second one which contains the lines of code to call the dataset created in the previous step and to create the report in markdown. The main libraries that we used in this step are the following:

> *Readbitmap, rtweet, rjson, dplyr, httr, reshape2, ggplot2, maps, tidyverse, stringr, mapproj, maps, viridis, rworldmap, igraph and networkD3*

3. The third one which contains the lines of code to call the datasets created in the first step and to create the shiny app. The main libraries that we used in this step are the following:

> *Shiny and Shinydashboard*