

DS 785 Data Science Capstone

Capstone Activities Report

Name: Spencer Swartz

Project Title: Social Media Content Optimization and Data Driven Brand Advocate Identification

Activity Report #3

Agenda/goals:

At the completion of this activity period to remain on track for the project it is expected that I was able to collect and analyze data on brand advocates for The Brewers Association Twitter related Data. In addition I decided to expand the scope of the project to include a predictive model identifying the number of potential retweets based on the text of the potential tweet.

Contacts Made/Method of Contact:

I have contacted Bart Watson (BA) via email with a similar update notifying him of the progress made in the data analysis period on 4/9/18.

Resources and Investigation Methods:

As stated in previous reports the tools user for this portion of the project includes Tableau, the Twitter API and Python. Additional tools used this period includes R and the igraph library for network analysis, as well as two additional Python libraries: gensim and scikitlearn. These additional libraries are used for the predictive modeling.

Progress:

Progress that was possibly lost last reporting period seems to be caught up. Once it was identified that a proof of concept plan of attack was needed for the brand advocate portion of the project. This allowed me to limit the amount of data needed to be gathered, while still providing insights for BA and helping to inform them if gather additional data would be worthwhile.

With this new data collected I was able to run some network analysis within R to identify similar connections between the separated retweeters, this helps to identify users that share BA tweets to fresh eyes. This will help BA to target these individuals in future tweets, which in turn will hopefully optimize the reach of their social media campaigns.

A Tableau Dashboard was developed to help to show the connectedness of the retweeters.

Once I had realized that this third phase of the project was coming to an end a couple of weeks earlier than expected, I decided that I could add an additional phase that I was contemplating and decided against when setting up the project proposal. This phase as explained above will build a predictive model for number of retweets for a given tweet. To do so a couple of standard models will be used. First is called Doc-to-Vec, this models is common in text analysis and is very similar to Word-to-Vec. Both models convert the data from text to vector form. The plan is to use the values of the vector to train a

linear model for predicting retweet numbers. Doc-to-Vec will be implemented with gensim and the linear model will be fit with scikitlearn.

Achievements:

In this reporting period I have completed phase 3 of the project and delivered a dashboard that helps to identify brand advocates. I have also identified specific aspects of a potential model to be implemented in the new phase of the project.

Questions:

I have no current questions for the professor.

Next step:

The next reporting period will be focused on completing the entirety of the analysis part of the project. This will allow for sufficient time to complete the write-up for the remainder of the class