Sentiment Analysis of Beer Hashtags

Introduction

Billions of dollars[1] are spent every year on TV commercials for the NFL alone. Beer marketing teams in particular have to compete in this heavy market for their product which is often consumed during games. I was personally curious how effective the catch phrase "Dilly Dilly!" from Bud Light commercials was, and if I could compare it to other slogans.To do this I would look at specific hashtags: #budlight, #millerlite, #dillydilly, and #millertime. I figured this would be a good way to compare slogan tweets ,#dillydilly and #millertime, to their baseline counterparts. These common beers of Anheuser-Busch and MillerCoors make a good example for beer companies everywhere. I predict the slogan hashtags will both have higher positive sentiment ratings than their counterparts indicating that these slogans are working and have a positive impact.

Data Collection and Analysis

To study this question, a REST API was utilized to mine tweets off of Twitter with specific hashtags. 1006 tweets were gathered between the four hashtags after retweets were removed. A sentiment was assigned to the tweet itself using the TextBlog package in Python, either 'positive' 'negative' or neutral. The initial data collected can be viewed below in the first two tables with number of instances and their total proportions indicated respectively.

To analyze this data, I used the a chi-square goodness of fit test with the following null and alternative hypotheses:

H_0: There is no difference in proportions of sentiment of tweets between the four hashtags

H_A: There is at least one difference in proportions of sentiment of tweets between the four hashtags.

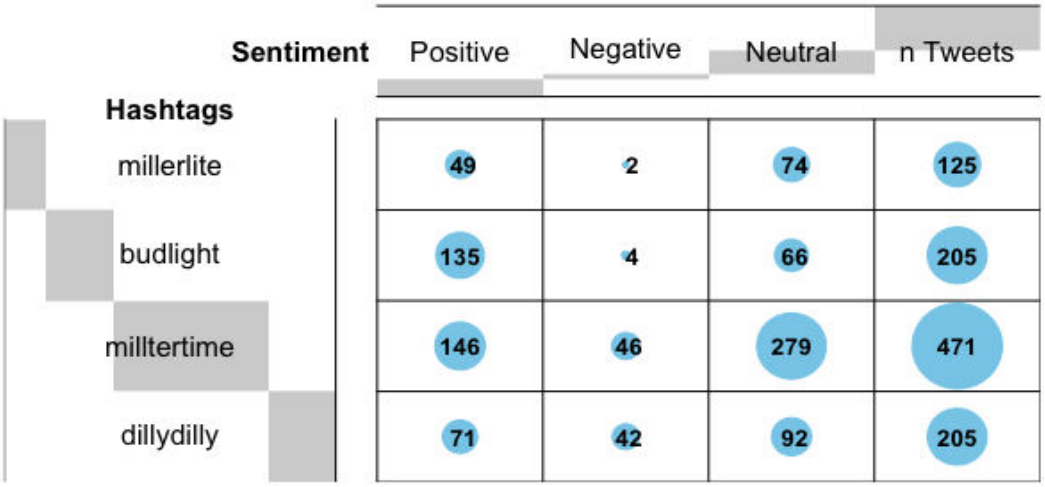
My results initially came up with a X-squared = 116.52, $p < .001$. This indicates that there is actually a difference in proportions of sentiment between these four hashtags. To determine which groups were different a post-hoc test was completed. Individual results are appended below, however all groups were found to be different from one another at the .05 level.

Conclusion

Looking at the proportions chart, the ranking of positivity in tweets for the hashtags are ranked #BudLight, #MillerLite, #Millertime, #dillydilly. Apparently twitter users are tweeting using #budlight most positively of the four. This may indicate the slogans aren't as effective amongst twitter users, especially for Bud Light. Further analysis of the tweets could be considered. Sarcasm for instance does not come through sentiment analysis. Tweets could also be individually looked at to see if the negativity/positivity is directed at the beer itself. Furthermore, tweets could be gathered using a Streaming API to see if sentiment changes over time or in respect to a recent commercial.

[1] https://nypost.com/2018/01/25/advertisers-spent-less-money-on-nfl-games-in-2017/

Hashtag Sentiment Data



Hashtag	Positive Tweet Proportion	Negative Tweet Proportion	Neutral Tweet Proportion
millerlite	39.20%	1.60%	59.20%
budlight	65.85%	1.95%	32.20%
millertime	30.99%	9.77%	59.24%
dillydilly	34.64%	20.49%	44.87%

comparison <fctr>	raw.p <dbl>	adj.p <dbl>
millerlite vs. budlight	0.0000	0.0000
millerlite vs. milltertime	0.0071	0.0071
millerlite vs. dillydilly	0.0000	0.0000
budlight vs. milltertime	0.0000	0.0000
budlight vs. dillydilly	0.0000	0.0000
milltertime vs. dillydilly	0.0004	0.0005