DS 501 Report

Outline:

Background

Procedure

Outcome

Business Applications

Background:

The main purpose of this experiment is to determine whether free product research can be done online by mining social network sites. The secondary purpose of the experiment is determining whether influential users can be identified and then influenced in order to positively sway public opinion on the product. The product being investigated is the Nexus 5 phone, and the social media platform being used is Twitter. We are investigating users on this platform who have tweeted about this product with the identifier '#Nexus5'.

The value from this experiment is that if successful a company will be able to collect massive amounts of product and customer feedback, all of which the company does not have to directly pay for. Typically gathering such customer feedback is costly and time-consuming for the company. This feedback can then be given an in-depth product analysis to determine customer sentiment of the product, and what are the most liked/disliked features. However the most value to the company is in the secondary purpose of this experiment, which is identifying and influencing users who have a strong community influence.

Strong community influence can be determined by the number of 'followers' they have as the primary metric, and a secondary metric of how many users they 'follow' themselves (It is important to note that an influential user should have a followers to following ratio above 1).

The reason to target 'following' as a secondary metric is because it is expected that companies and celebrities will have the most followers, and that is not the group we are trying to pursue. We are looking to target your 'average' non-commercial user, whom should have a network of friends they follow as well. We are specifically targeting this user, so that we can leverage their network and their credibility within to promote the product/brand. The purposed way to influence these users would be a marketing directive specially targeted to them, something along the lines of a free promotional giveaway (the specifics of which are better suited for a marketing department and are outside the scope of this study).

When targeting these influential users there are three expected categories of outcome, with six total outcomes. The expected outcome categories are as follows: user is swayed positively, user is swayed negatively, and no change. The optimal outcome of course would be to positively sway the user in favor of the brand and/or product.

For a user being swayed positively there are four potential outcomes the first of which and best case scenario being the user becomes a brand evangelists (an extremely loyal customer and promoter of the company). The second and more probably outcome is the user becomes a buzz marketer for the product within their network, which is the target outcome of the study. The third expected outcome is that a user who was voicing negative opinions on a product will no longer be critical of the brand or products after the persuasion attempt. It is important to note for the third outcome that the user does not voice any positives about the product or brand; they have only stopped their negative influence. The fourth and final outcome is that a user who was voicing negative opinions now voices them at a lower frequency.

For the user being swayed negatively category there is one possible outcome. This outcome is that upon persuasion the user continues to voice negatives at a higher intensity.

Finally the last category of no change has one outcome in that the persuasion had no discernable effect on the user, and has not affected their behavior in anyway. Neither of these outcomes are desirable however the no change is more desirable as the cost for persuasion is very minimal.

Buzz marketing has been shown to be one of the most effective forms of marketing. For those unfamiliar with the concept it is getting 'non-corporate' entities to write product reviews/blogs and discuss with their friends a product in a positive light. Since it is coming from a 'non-corporate' source people perceive it as word of mouth instead of as an advertisement, and are therefore less apprehensive about listening. If we can realize the goal of being able to identify key users and then are able to successfully convert them into buzz marketers, this yields incredible value for a company. As this equates to a new highly-effective member of your sales force, who's only cost was whatever promotional material was given.

Procedure:

For this project, we used the Twitter API to collect tweets. Twitter provides developers with multiple ways to gather data from them through their API's. We began by making a call to the API to return tweets that contained a certain term (in this case, #Nexus5). Twitter allows one developer (or more specifically, one set of access keys) to make this type of call 180 times in every 15 minute interval. It can return up to a maximum of 100 tweets, so the max one person could retrieve in 15 minutes is 18,000 tweets. After we were satisfied with our code, we set it to collect this maximum amount of tweets. However, the way we wrote it, the program will stop calling for tweets when it detects that there are no more to return. Through multiple times of running this, the most tweets we ever collected at once was 2,396. We suspect that there is something going on in our code that is unintentionally stopping the call, because we know from real-world experience that there are more than 2,400 tweets that mention #Nexus5. This is

something left to be investigated more, so for the time being our analysis was one this limited batch of tweets.

The first thing we looked at was popular words. The most popular words in our data are printed below, based on word count. This eliminates words that contain "Nexus", because when we did include that, most of our words were some variation of "Nexus5" and were hashtags. We looked at hashtags separately, so any words that because with "#" were eliminated. We also removed stop words (words like "the", "in", etc.) and instances of just punctuation.

Word	Count
sleeve	1026
leather	874
iphone	413
vs	403
5:	398
updated	397
however,	397
5c	397
inclusion	397
longer	330
production,	328

bring	328
6:	327
sleeve/case	191
mid	185
smar...	151
choice	142
multiple	142
google	121
quote	120
concrete	116
case	103
smartphones	103
sleeve/cover	102
papaya	102
store:	102
deerskin	99
diary	96
rum	96
bluelake	93

Interesting observations: iPhone is near the top, as is "vs". This probably indicates that some tweets were comparing the Nexus phone to the new iPhone. That is the iPhone 6, and "6:" was also one of our top words. "5c" was also one of the top words, which is a type of iPhone,

although an older version. It is possible that our phone was compared to older versions of the iPhone. It would be helpful for us to explore specifically those tweets that mention both the Nexus 5 and iPhone and see what words are included in them, and if they are generally positive or negative.

An approach to explore as we investigate more could be to look at various words n-grams. That is, popular instances of two words together, 3 words together, etc. This could help us understand the comparisons we expect are in our data.

Another observation is the multiple instances of words related to a phone case. "Sleeve", "leather", "sleeve/case", "case", "sleeve/cover", and "deerskin" are probably all related to this. This could interest Google's marketing division and help them to consider how to effectively market cases for the Nexus 5 based on what people are saying about them.

RETWEETS

Next we looked at the most popular retweets in our data. Below are the results. It appears that the most popular retweets were essentially ads. Google may want to look at these users closer to understand who they are, why they are mentioning the Nexus 5, and how to best leverage their popularity.

```
+-------+--------------+--------------------------------------------------+
| Count | Screen Name  | Text                                             |
+-------+--------------+--------------------------------------------------+
```

269	Android	RT @Android: Forgot where you've parked? With Now
		cards in the #Googleapp on your #Nexus5, you'll be
		able to find your car easily. http://t…

269	Android	RT @Android: Forgot where you've parked? With Now
		cards in the #Googleapp on your #Nexus5, you'll be
		able to find your car easily. http://t…

269	Android	RT @Android: Forgot where you've parked? With Now
		cards in the #Googleapp on your #Nexus5, you'll be
		able to find your car easily. http://t…

269	Android	RT @Android: Forgot where you've parked? With Now
		cards in the #Googleapp on your #Nexus5, you'll be
		able to find your car easily. http://t…

44	msebastianco	RT @msebastianco: Testing @TARGIT #mobility on the
		new #Nexus5 super high resolution screen. All
		good so far #Targit2014 #Mobility2014 htt…

15	desi_dime	RT @desi_dime: Find Stunning #nexus5 32 GB at
		lowest price of just - http://t.co/TB1aLafrAr
		#LikeLike #deals #offers #Coupons http://t.co/3…

15	desi_dime	RT @desi_dime: Find Stunning #nexus5 32 GB at
		lowest price of just - http://t.co/TB1aLafrAr
		#LikeLike #deals #offers #Coupons http://t.co/3…

| 15 | desi_dime | RT @desi_dime: Find Stunning #nexus5 32 GB at |
| | | lowest price of just - http://t.co/TB1aLafrAr |

| | | | #LikeLike #deals #offers #Coupons http://t.co/3… |

| 15 | desi_dime | RT @desi_dime: Find Stunning #nexus5 32 GB at |

| | | | lowest price of just - http://t.co/TB1aLafrAr |

| | | | #LikeLike #deals #offers #Coupons http://t.co/3… |

| 15 | desi_dime | RT @desi_dime: Find Stunning #nexus5 32 GB at |

| | | | lowest price of just - http://t.co/TB1aLafrAr |

| | | | #LikeLike #deals #offers #Coupons http://t.co/3… |

+-------+--------------+--+

Users:

Next we looked at the most popular user mentions, based on count. Below are the results.

+-------------+-------+

| Screen Name | Count |

+-------------+-------+

| desi_dime | 15 |

| googlenexus | 9 |

| google | 8 |

| Android | 7 |

| GooglePlay | 6 |

| nxtab | 6 |

| _binkybear | 4 |

| swept | 4 |

| dbrandSkins | 4 |

| cetges | 4 |

+-------------+-------+

Not surprisingly, google, Googleplay, Android, and googlenexus are on the list. However it also includes non-Google entities, and one of those is even in the top spot. This absolutely would be important to Google's analysis of this data. They need to look at each of the most popular users to see why they are talking about the phone, what they are saying, and what users are saying about them. The top user was also in the top Retweets list, so that offers more comparison points to analyze.

HASHTAGS:

+--------------------+-------+

| Hashtag (Non-Nexus) | Count |

+--------------------+-------+

| tech | 572 |

| iphone5c | 397 |

| google | 330 |

| apple | 312 |

| technology | 265 |

| android | 178 |

| iphone | 171 |

| techno | 142 |

| blog | 93 |
| post | 93 |

+--------------------+-------+

+-----------------+-------+

| Hashtag (Nexus) | Count |

+-----------------+-------+

nexus5	2524
nexus5sleeve	1344
nexus5cover	1273
nexus5portel	1273
nexus5pouch	1248
nexus5leather	1242
nexus6	333
nexus4	78
nexus	75
nexus5case	71

+-----------------+-------+

Above are the results for most popular hashtags. We split the tags into two groups, based on their inclusion of "Nexus" or not. We did this because we wanted to see important hashtags that were not just a version of our search term. The non-Nexus tag group includes tags we would probably immediately associate with the phone and technology in general. Two important tags do stand out: iphone and blog. Again we see the iphone included in tweets about our phone, which

probably indicates that people are making comparisons at the time of the new iPhone's release. Seeing the hashtag blog indicates that maybe people are writing longer posts about our phone, or phone technology in general, and are linking to them. Google should explore these posts, and find a way to understand the general sentiment of what they are saying.

Business Applications:

The results collected from the exploratory research proves companies can successfully mine the social internet for both brand and product information. For a business this means they have an unlimited virtually free database of feed, public perception, etc. of either their brand or products. There are three core areas a company could use this information for, real-time consumer impressions, gauge brand sentiment, and determine influential users for a buzz marketing campaign.

The first application is being able to track effectiveness of marketing campaigns. By determining the average or 'normal' frequency the product/brand is mentioned over time (assuming no outstanding events are going on), we now have a consistent baseline of our consumer impact. When a new commercial or product launch is about to take place, the twitter live stream can be monitored in unison to give us the real-time customer touches. The baseline can then be used to compare against the results after a new marketing campaign or product launch. If a significant impact is observed than the event can be considered a success, or conversely a failure if there is no significant increase.

The second application is gauging sentiment of the user population. As illustrated through the study, tweets can be archived and then monitored for their content. If an advanced word search algorithm or even a somewhat simple n-gram solution was used, researched could determine the context in which their products are being discussed. By determining the words or

series of words following or leading to the brand identifier, the user sentiment could be captured. For example if words with negative connotations are surrounding the product identifier it would be relatively safe to assume that a consumer had a negative experience. This type of data collection and analysis provides the company with feedback, for instance Google could review tweets on the Nexus5 to determine that people enjoy the longer display screen.

The final business application is determining influential users, and utilizing them to promote the brand. As illustrated we have the ability to determine who has the most followers among people tweeting about the Nexus5 or whose tweet was the most retweeted. Combining this information with full tweet text analysis provides us the ability to see how they are influencing others (either positively or negatively). The business can reach out to the influential users and as described early could target them using enhanced user support or free promotions to help sway their opinion.