

CAPSTONE PROJECT

(with Germin8)

A

Project Report

on

Identifying Trends through Social Media Analytics
in Retail Space

GROUP (Section A & B)

Sneha Komma (71310051)

Arpit Gupta (71310056)

Garima Jena (71310095)

Vamsi Posemsetty (71310014)

CERTIFICATION IN BUSINESS ANALYTICS

INDIAN BUSINESS SCHOOL

HYDERABAD

ACKNOWLEDGEMENT

We would like to extend our gratitude to many people, who so generously contributed to the work presented in this report.

Special mention goes to our company sponsor, Mr. Jayant Bahel and his team. They have been an amazing experience and we thank them, not only for giving wonderful opportunity to work on an interesting project, but also for supporting us throughout the project tenure and giving us their valuable time and feedbacks.

Similar, profound gratitude goes to our faculty mentor, Prof. Smarajit Bose, who has been a truly dedicated mentor. We thank professor, not only for his tremendous academic support, but also for giving us valuable guidance.

We are also hugely appreciative to Prof. John Trip, especially for teaching us Tableau so well and clearing our doubts several time during the project.

Special mention goes to our Capstone project coordinators, Miss. Reema Gupta and Mr. Amit Ghosh, for their extended support and understanding the difficulty at our part and allowing us project extension.

Finally, but by no means least, thanks go to CBA program and ISB, for building such a wonderful platform to learn and share knowledge in Analytics and understanding the industry requirements.

ABSTRACTS

With the advent of high-speed internet and flourishing social media, today there is enormous textual data available to companies with respect to customer perception about products and services, how are competitors' products doing in the market, what is currently most happening which can be utilized etc. However, sensing all these sentiments from the textual data which comes in variety of forms is an important task.

We were asked by the Germin8 team to analyze the social media data over online retail industry for the span of 3 months starting from April 1st 2014 to July 1st 2014. This dataset contained a total of 9.4 lakh records with 92 attributes for each record. The primary objective of our project was to Identifying meaningful patterns and trends and the extraction of potential knowledge in large volumes of this text data. These insights/trends are presented in the form of interactive dashboards in Tableau that facilitate insights of social and business value.

We undertook this problem from a naïve point of view without using extensive algorithms such as association rules, clustering etc. as it was really difficult to process all the data through these algorithms given our systems' computing power. Therefore, we split the data day-wise and created 92 separate datasets and built 6-gram and 8-gram term-document matrix for each dataset separately. Later, we combined all the day-wise frequencies of 6-gram or 8-gram in a single dataset. This enabled us to identify the 6-gram or 8-grams trending on social media on consecutive days. The visualization of this dataset in Tableau helped us in identifying key phrase trending on social media using dynamic and context filtering.

From this exercise, we were able to identify various trends which were prevalent in social media for some time and then tamed such as "Snapdeal video share contest and win mobile" and "get moto e genuine product"; also we identify trends which were constantly there almost throughout the time frame such as "Flipkart Genuine product day replacement guarantee free"; finally certain trend which started late in June month like "Myntra liveforlikes contest" and similarly certain trend which were there initially in April were "Like share tag and get chance to win flipkart vouchers".

The finding from this project can be extended ahead to analyze the impact of each author on the tweets. There may be the chances that only one or couple of authors are artificially inflating the tweets on some day. Also, the other thing which can be done is to filter the dataset by each domain such as Twitter, Facebook, blogs etc. and then implementing this methodology. This will enable us to understand that this kind of analysis is useful for which domain specifically.

METHODOLOGY

Data Processing

Our first step was to assemble, compile, and clean the required data. This included tasks such as loading the csv data and splitting the data by date. The next step was to clean the text data such as removing punctuations, stop words, numbers, converting the text to lower case. Finally, we constructed n-grams term-document matrix of various lengths to identify key trends. 6-grams and 8-grams identified the trending 6-word and 8-word phrases in social media. The idea is that 6-gram and 8-gram phrases would give more context and insight into the trends relative to the 1-grams.

Visualization & Insights

Once we identified trends in the raw social media, we summarized the results using interactive dashboards in Tableau. This will enable various cross-functional teams within the organization and the concerned consumer/brand have access to the required actionable insights.

IDENTIFIED TRENDS

Figure 1: 6-grams

1. Shows an interactive dashboard that connects a heat map of 6-grams with spark lines of those topics over the 3 month time period.
2. If the user selects a single topic, they can further drill down and highlight the trend over time for that particular topic.
3. It can also be observed that there is a lot of conversation around flipkart deals and vouchers, deals on mcafee, amazon and myntra.com

Figure 2: 8-grams

1. Shows an interactive dashboard that connects a heat map of 8-grams with spark lines of those topics over the 3 month time period
2. It is interesting to note that the trends in 8-grams are differ than the trends found in 6-grams.
3. Trending topics include vouchers on flipkart, amazon, mcafee, deals on extended warranties on mobile phones

Figure 3: Context Filtering with Date

1. This shows the top 10 phrases or tweets for the selected range of dates.
2. If we would like to know the top trends in the end of June then that can be identified from this.
3. Trending topic includes "Flipkart launches schedule delivery service" and "Myntra liveforlike

contest” during end of June.

Figure 1: Interactive dashboard that connects a heat map of 6-grams with a spark lines

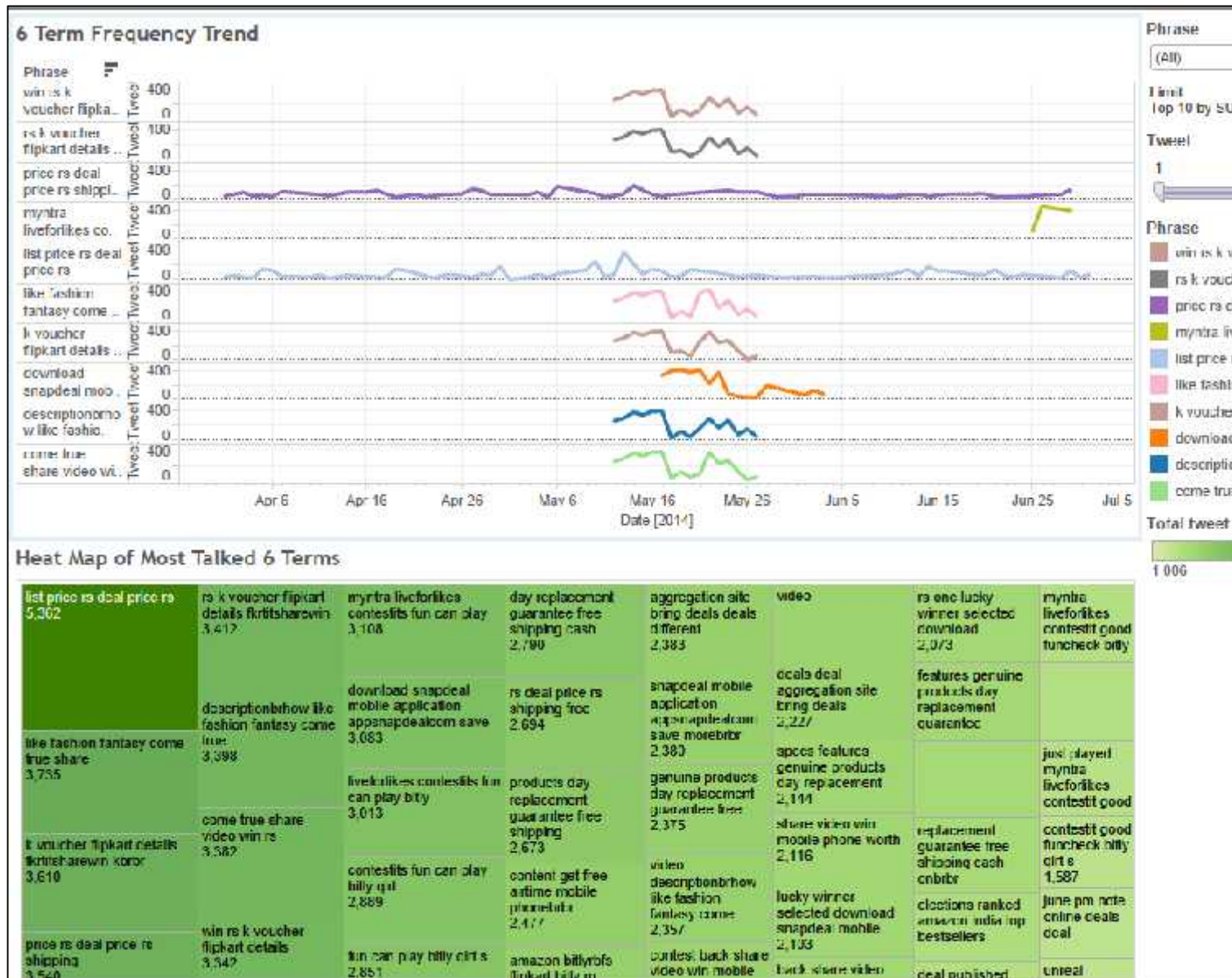


Figure 2: Interactive dashboard that connect a heat map of 8-grams with a spark line

