An Analysis of Tweet Button Trends for the New York Times

CSCI 8980: Data Analysis

Kristin Dahl

10/10/2011

I. The Problem

Twitter has exploded in popularity in recent years and has increasingly become a useful means of rapidly sharing news and events. It has become so popular that many websites are choosing to link directly to Twitter via Tweet Buttons, which are widgets that allow people to share content on Twitter without having to leave the webpage. The idea behind the Tweet Button is twofold: promote a specific Twitter account and drive traffic to a specific website.

The problem explored by this analysis is to determine whether patterns exist in Tweet Button usage for articles published by the "New York Times" (NYT). The NYT was selected because it is a popular news publication that integrates the Tweet Button into its articles.

II. The Research Question

The specific question that is discussed in this paper is: when are NYT readers most likely to share an article via the Tweet Button? Does the topic of the article make a difference? The answers to these questions are not immediately clear. On one hand, given the proliferation of devices like smartphones and laptops, internet access is available virtually anytime and anywhere. Under these circumstances, we would expect no discernable pattern in Tweet Button usage because people can, in theory, access any NYT articles and tweet them whenever they want to.

However, many workplaces and schools are blocking social media websites and have implemented strict rules regarding the usage of personal computing devices. And, people typically do not have time to look at the NYT and tweet articles when they're busy with work or class. In this scenario, we would expect Tweet Button usage to peak in the morning, around lunchtime, and in the evening.

We argue that this analysis, while not exhaustive, presents a correlation between articles published by the NYT and the amount of tweets that they receive. Is there a specific time at which NYT should publish an article in order to maximize the number of tweets that are generated? Does the subject of the article have an effect on the optimal publish time?

III. Results

This analysis was done by collecting a random sample of tweets once per minute for one week. All tweets included the text 'nyti.", which indicates that the tweet was created via a Tweet Button in a NYT article. We also included retweets of tweets generated via the Tweet Button. Overall, we collected 2741 unique tweets over the course of one week. For the purposes of this analysis, "unique" means that the tweet collection was filtered to remove any duplicate records (i.e. records in which the same user tweeted the same text).

Each tweet record included the username, the tweet text, and the date and time of the tweet. The collection of tweets was further analyzed to determine the number of tweets submitted per hour, and the top twenty tweeted URLs were analyzed to determine what section of the NYT they were published in. Figure A is the graphical representation of the average number of tweets per hour over the course of a week. The number of tweets peak during 1:00 pm CST and, as would be expected, reaches the lowest point around 5:00 am CST. Figure B is depicts the subject matter of the top twenty most popular tweets over the course of a week.

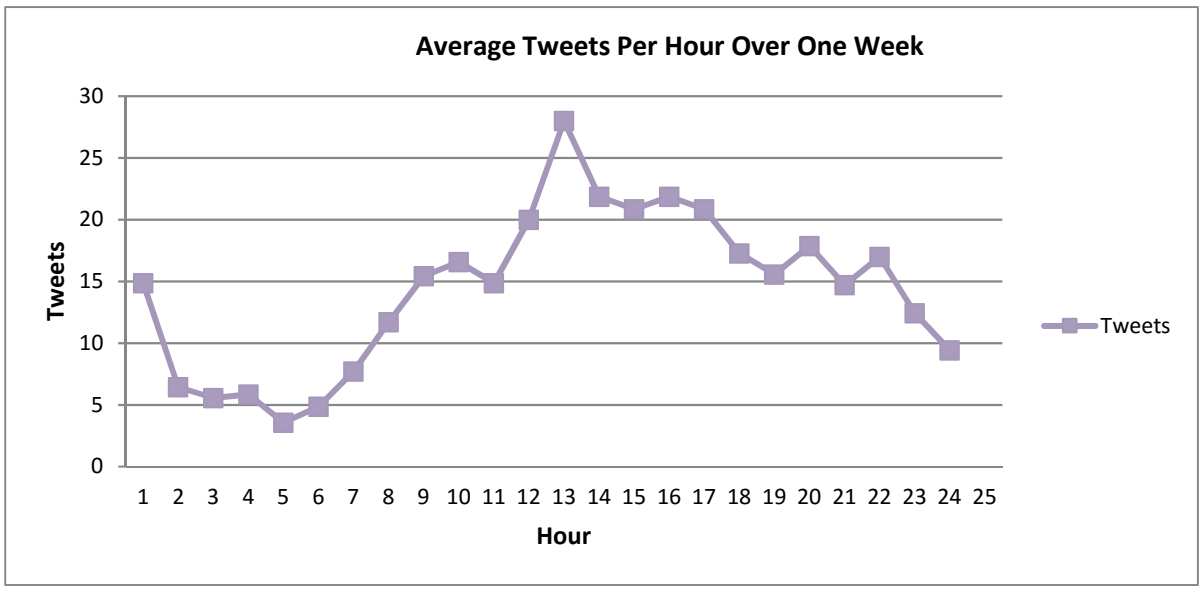

Figure A: Average Tweets Per Hour Over One Week

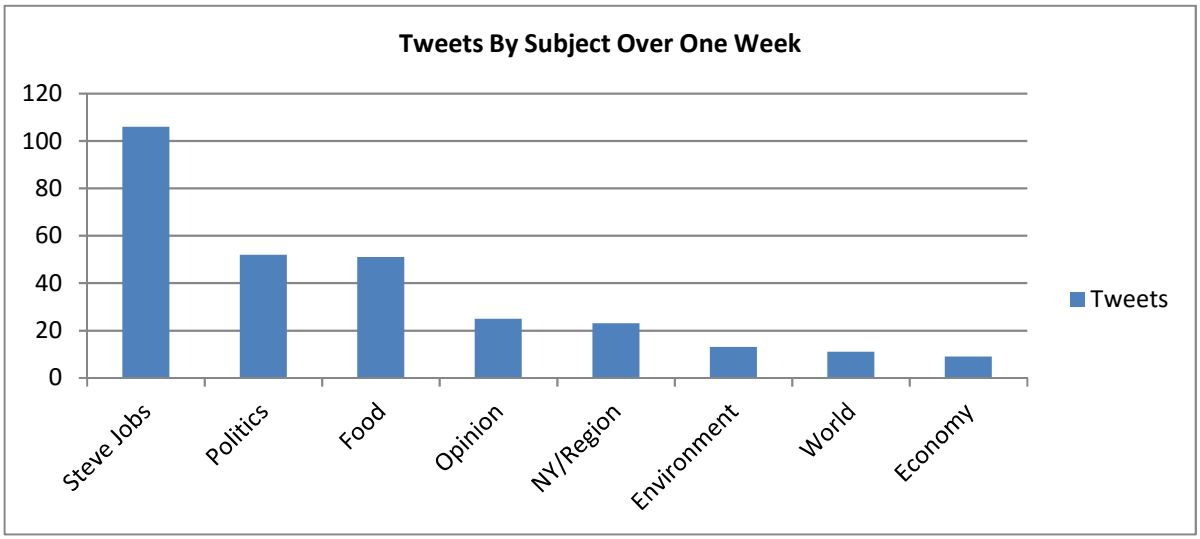

Figure B: Tweets Per Subject Over One Week

IV. Discussion

The results of our analysis show that there is a correlation between the time of day and the number of tweets shared via the Tweet Button. In a 24-hour period, the average number of tweets climbs during the first half of the day, reaches a peak, and then declines. This suggests a definite trend in Tweet Button usage throughout the day. In order to maximize the usage of the Tweet Button for a NYT article, the article should be published early in the day prior to the peak time (as shown in Figure A).

The results also show that there exists a correlation between article subject and the number of tweets generated. Subjects such as "Politics" and "Food" tend to generate more tweets over a one-week period than the subjects of "World" and "Economy." This data reflects the general interests of the NYT and Twitter communities. However, one unanswered question is whether there is a correlation between the number of articles for a particular subject and the number of tweets generated for that subject.

The results also show the effect that key current events can have on the NYT community and their usage of the Tweet Button. Steve Job's death was reported during the week in which we performed the analysis, and the data in Figure B shows that interest in this news has overshadowed general subject matter.