Michelle Dezihan

Marriott versus Hilton Twitter Sentiment Analysis

**Executive Summary**

With how prevalent opinions are on social media and how quickly perceptions can change, I was curious on how fast a negative announcement can affect sentiment of a company. With Marriott’s recent announcement of a data breach affecting millions of its guests, I gathered over 6,000 tweets from the 7 days that followed the announcement and conducted a sentiment analysis on tweets containing “#marriot” and “#hilton” to compare how the public felt about each hotel brand. As expected, sentiment scores for Marriott dropped immediately and continued to drop at a fairly rapid pace. Hilton showed a gradual increase in positive sentiment, although did not seem to increase much in light of the Marriott data breach.

**Introduction**

Given my career experience in the hospitality industry, I wanted to explore how much the general public complains on social media about their experience at hotels. As a current hotel employee, I am well aware that those who have poor experiences are much more likely to complain publicly than those who have excellent experiences are to share praises. When I first decided to do a Twitter sentiment analysis on Marriott, it was during a period of employee strikes against Marriott in a number of cities. When I actually began to pull tweets from Twitter’s API, I managed to capture the day that Marriott announced its massive data breach, which has been occurring as far back as 2014. This report includes data of the week that followed the announcement of the data breach.

**Data**

To create my dataset, I pulled tweets that contained “#marriott” in the text. While restricted by the 7-day limit for Twitter’s resting API, I was able to pull 6,132 tweets that had “#marriott.”

For Hilton, I was only able to gather 794 tweets that had “hilton” in the text. Regardless of whether my code included a specific number to pull, I could not extract more than 794 tweets.

**Methods**

Through Python, I accessed Twitter’s API and extracted as many tweets containing the hashtags “marriott" or “hilton" as I could. Before storing my tweets into pandas dataframes, I assigned sentiment scores to each tweet using the tidytext document of positive and negative words. From there, I stored all the Marriot tweets into one pandas dataframe and Hilton tweets into another pandas dataframe, and began to clean up the tweets. I did my best to remove all RTs, although I suspect there may be some duplication in the tweets collected, removed hashtags, links, and transformed all the text to lowercase.

To visualize the sentiments of each over the 7-day period, I chose to graph the data in R, which can be found in the Exhibits section of this report. I also included wordclouds for each hotel to show what words the general public frequently mentioned in their tweets.

Finally, I build a simple naïve bayes classifier to further explore whether we could predict whether a specific word in the tweet referred to Marriott or Hilton users.

**Results**

Unsurprisingly, Marriott’s sentiment score tanked in the days following the data breach announcement. The drop in sentiment started at a slower rate than I had expected, but seriously dropped as the number of tweets mentioning “#marriott” continued to greatly increase. As for Hilton, the news of Marriott’s data breach did not seem to cause any significant jump in positivity for Hilton, but instead saw a gradual increase over the week.

The classifier had a pretty successful accuracy of 0.8922, and returned more words related to charity and holidays for Hilton than Marriott. For Marriott, only the odds of 20.7 to 1.0 for the word “million” appeared under the 20 most informative features, which makes sense given the fact that its data breach affected millions.

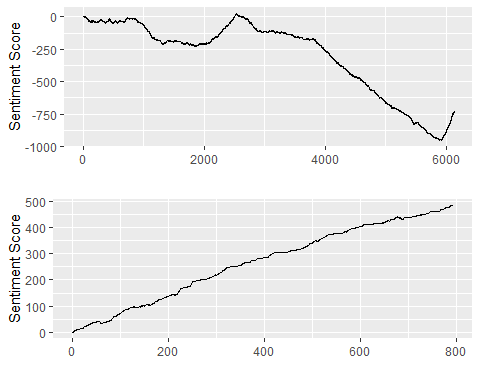
**Future Direction**

To improve this analysis, I would definitely try to gather more tweets over a longer period of time. My original hope was to gather tweets before the strikes began and compare a before/after sentiment analysis, but realized too late I was unable to do so. With the limitations of only being able to pull data from the past 7 days, I would consider running this analysis on tweets extracted from the streaming API or try to pull tweets on an every-other-day basis and build up a database of tweets over a longer period of time in order to get a better sense of how the sentiments of these hotels adjust over time. In addition to pulling tweets based on # content, I would try to capture more tweets using @mentions for Hilton and Marriott as well.

**Exhibits**

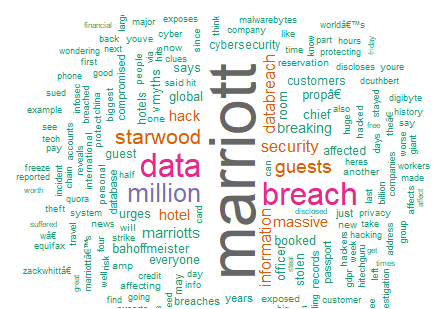
Sentiment Plots

The following graphs display the sentiment of tweets over the past 7 days. The data gathered starts on November 30th, 2018, which was the day Marriott went public with the news of its massive data breach.



Wordclouds

For Marriott, the most frequent words mentioned in tweets are unsurprisingly “data”, “breach”, and “million.” Notable mentions also include “security” and “hack.”



For Hilton, the most frequent words appear to be related to the location of the hotel, events, and charity terms such as “Freecember”, “toy” and “drive.”

