IPhone and Galaxy Sentiment Analysis

Summary of Findings

for Helio Corporation

Alert! Analytics

Celeste Hofer

May 1, 2019

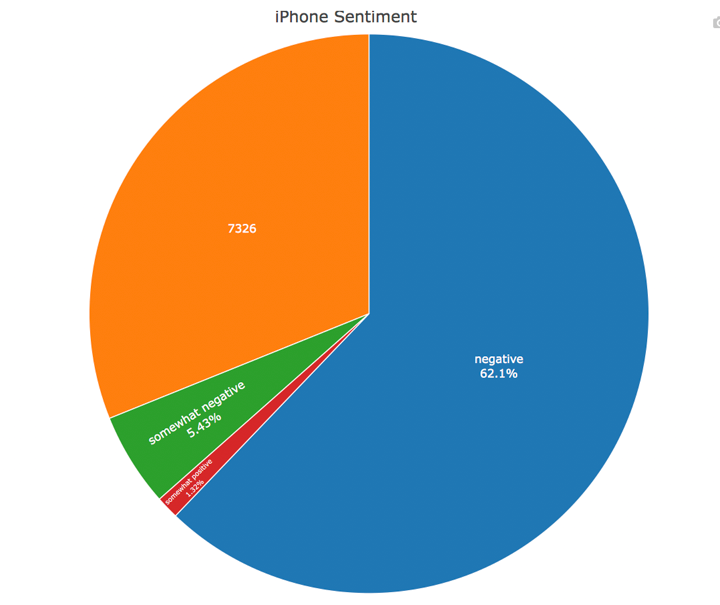
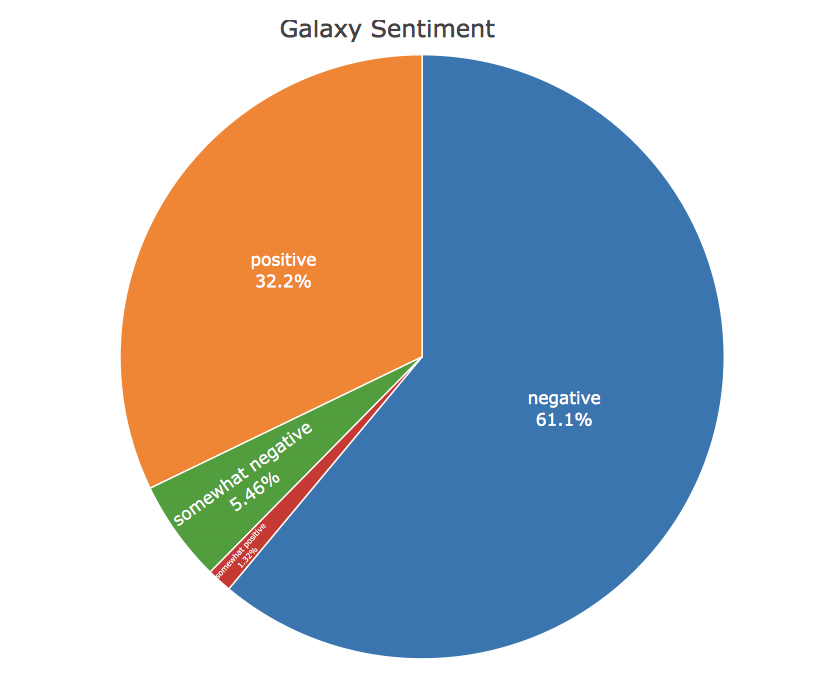
# Project Overview

The goal of this project was to compare both positive and negative sentiment expressed towards both the Apple iPhone and the Samsung Galaxy, in order to determine which smart phone has the better public perception. A summary of positive sentiment, negative sentiment, etc. for each phone was created so that a comparison between the two products could be made. The time period analyzed was March of 2019. The sentiment was gathered by conducting a web crawl using Amazon Web Services, (AWS).

This project is being conducted for Helio Corporation, in order to provide information for them to be used in their discussions on a project with Apple and or Samsung.

# Findings

The sentiment, for each phone, was captured within 4 distinct categories: (1) negative, (2) somewhat negative, (3) somewhat positive, and (4) positive. The category of negative includes both very negative and negative sentiment. Similarly, the category of positive includes both very positive and positive sentiment.



**Corresponding Predicted Sentiment Values**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Negative | Somewhat Negative | Somewhat Positive | Positive |
| iPhone | 14638 | 1279 | 310 | 7326 |
| Galaxy | 14413 | 1290 | 312 | 7538 |

Sentiment towards the Apple iPhone was predominantly negative—62%. Of interest, however, is that the second major category for sentiment was positive—31%. Sentiment expressed as somewhat negative or somewhat positive is expressed much less frequently.

Sentiment towards the Samsung Galaxy was also predominantly negative—61%. 32% of the sentiment expressed was positive, however. The remaining responses were somewhat negative (5.5%) and somewhat positive (1%).

# Confidence in Findings

The machine learning analysis that was conducted for both Apple iPhone and Samsung Galaxy have high confidence outcomes. This confidence assessment is based on the Accuracy and Kappa values received. Additionally, the Positive Predicted Values and Negative Predicted Values reflect high confidence that the predicted values for Sentiment are classified correctly. This machine learning analysis was conducted using R.

The models selected for both the iPhone and Galaxy sentiment analysis both utilized all possible feature inputs (independent variables). In other words, no columns were removed from the data used in the final models used to predict sentiment. The analysis found that the most accurate models were created using all independent variables and re-engineering the dependent or target variable. While separate analyses were conducted creating models that only contained select features, those models were found to not have as high a performance characteristic as the selected models.

## Apple iPhone

The model used for predicting sentiment towards iPhone was created from the Random Forest algorithm with sentiment recoded for the target variable. The sentiment was recoded into 4 categories instead of 6. No feature inputs were removed in this model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Kappa | Predicted Value Type | Negative Sentiment | Somewhat negative Sentiment | Somewhat positive  Sentiment | Positive  Sentiment |
| RF Recoded Sentiment | 0.8452 | 0.6114 | Positive Predicted Value | 0.96834 | 0.941176 | 0.93469 | 0.8236 |
|  |  |  | Negative Predicted Value | 0.90373 | 0.969016 | 0.96516 | 0.9735 |

## Samsung Galaxy

The model used for predicting sentiment towards the Galaxy was created from the C5.0 algorithm with sentiment recoded for the target variable. The sentiment was recoded into 4 categories instead of 6. No feature inputs were removed in this model.

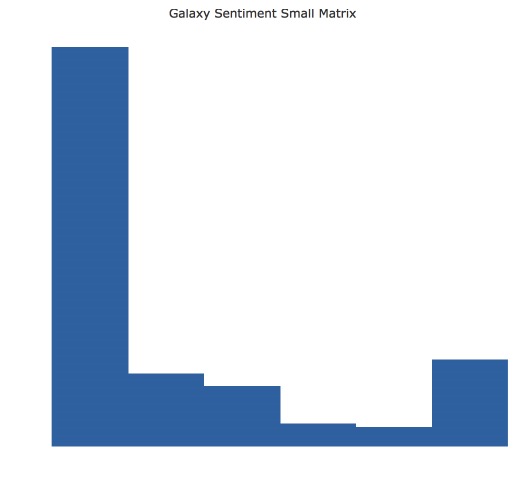
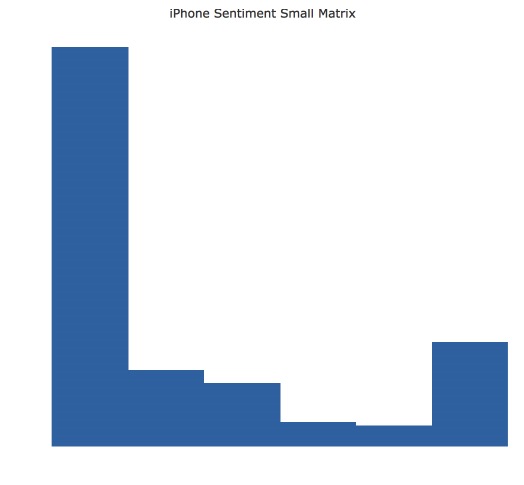
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Kappa | Predicted Value Type | Negative Sentiment | Somewhat negative Sentiment | Somewhat positive  Sentiment | Positive  Sentiment |
| C5 Recoded Sentiment | 0.8479 | 0.6065 | Positive Predicted Value | 0.88941 | 0.785714 | 0.85590 | 0.8421 |
|  |  |  | Negative Predicted Value | 0.92892 | 0.967859 | 0.95718 | 0.9042 |

# Implications

The results of sentiment for both models is predominantly negative. Additionally, the positive sentiment expressed, for both models, was always around 50% of the negative sentiment. The results for each phone are very similar, thus we cannot conclude that one phone has an advantage over the other. More investigation will need to be done.

It is also interesting to note that the two original data sets that the models were trained against (one for iPhone sentiment, the other Galaxy), both had heavily negative assessments of sentiment. One idea is to determine if the sentiment in the data is accurately assessed. Possibly the sentiment analysis is skewed too heavily towards assessment of negative sentiment, and needs more terms denoting positive and neutral sentiment. A linguistic expert that is knowledgeable of text mining could be consulted to assess the corpus.

These are charts representing the data that the models were trained on. Note the skew in sentiment towards Very Negative responses, with Very Positive responses coming in second.



Additionally, sentiment expressed towards both the iPhone and Galaxy product may vary from month to month depending on several factors. Did a new model of the phone just come out? Was there a bad update? Do seasonal factors influence the mood of consumers? It may be beneficial to sample sentiment over a longer time period, or even different quarters of the year.

# Methodology

The methodology used to solve this problem had several steps.

* Amazon Web Services was accessed, and sentiment related to several smart phones was gathered for the month of March 2019.
* The following process was repeated separately for both the iPhone and the Galaxy.
  + Machine learning, using the R language, was used to create models, adhering to the Data Science process.
  + The most accurate model was then used to analyze the sentiment gathered from AWS in. order to classify responses into the categories of negative, somewhat\_negative, somewhat\_positive, and positive.