

IST736 – Text Mining

Prof: Dr. Gates

Final Project: Sentiment Public – Kobe Bryant

**Name: Nicole Jade Villanueva, Jamyl Jaquez, James Yongjik Eakins, and Yehuda Perry**

**Course: IST736 – Text Mining**

**Professor: Dr. Gates**

**Assignment: Final Project**

**Date: 03/20/2020**

**Table of Contents**

[1. Introduction 3](#_Toc35598345)

[1.1 General 3](#_Toc35598346)

[1.2 Sentiment Analysis 6](#_Toc35598347)

[1.3 Topic Modeling and LDA 6](#_Toc35598348)

[1.4 What is Naive Bayes? 6](#_Toc35598349)

[1.5 Naïve Bayes Classification 7](#_Toc35598350)

[1.6 Multinomial Naïve Bayes 8](#_Toc35598351)

[1.7 Bernoulli Model 8](#_Toc35598352)

[2. Dataset Analysis 9](#_Toc35598353)

[2.1 About the Data 9](#_Toc35598354)

[2.2 Data Collection 10](#_Toc35598355)

[3. Analysis and Models 16](#_Toc35598356)

[3.1 Loading the data to Google ‘Colab’ 16](#_Toc35598357)

[3.2 Review the data 16](#_Toc35598358)

[3.3 Cleans the data 17](#_Toc35598359)

[3.4 Visualization 19](#_Toc35598360)

[3.5 Word Cloud 22](#_Toc35598361)

[3.6 Tweets Sentiment Analysis 23](#_Toc35598362)

[3.6.1 Positive and Negative Tweets 23](#_Toc35598363)

[3.6.2 Time Series 24](#_Toc35598364)

[3.6.3 SVM Model 24](#_Toc35598365)

[3.6.4 Positive and Negative Words 25](#_Toc35598366)

[3.6.5 Bernoulli 25](#_Toc35598367)

[3.6.6 Topic Modeling 26](#_Toc35598368)

[3.6.7 Tweets Location 28](#_Toc35598369)

[4. Conclusions 29](#_Toc35598370)

[5. References 31](#_Toc35598371)

# Introduction

## General

Kobe Bryant spent his early years in Italy and joined the NBA straight out of high school. A dominant scorer, Bryant won five NBA championships and the 2008 MVP Award with the Los Angeles Lakers. Although later seasons were marred by injuries, he surpassed Michael Jordan for third place on the NBA all-time scoring list in December 2014 and retired in 2016 after scoring 60 points in his final game. In 2018, Bryant earned an Academy Award for Best Animated Short Film for *Dear Basketball*. On January 26, 2020, he was in a helicopter crash that killed Bryant, his 13-year-old daughter Gigi and seven others.

Kobe Bean Bryant was born on August 23, 1978, in Philadelphia, Pennsylvania. Named after a city in Japan, Bryant is the son of former NBA player Joe "Jellybean" Bryant.

In 1984, after ending his NBA career, the elder Bryant took the family to Italy, where he played in the Italian League. Growing up in Italy alongside two athletic older sisters, Shaya and Sharia, Bryant was an avid player of both basketball and soccer. When the family returned to Philadelphia in 1991, Bryant joined the Lower Merion High School basketball team, leading it to the state championships four years in a row. With an eye on the NBA, he also started working out with the 76ers.

Though Kobe Bryant boasted good grades and high SAT scores, Bryant decided to go straight to the NBA from high school. He was selected by the Charlotte Hornets with the 13th overall pick of the 1996 NBA draft and was subsequently traded to the Los Angeles Lakers.

In his second season with the Lakers, Bryant was voted a starter for the 1998 All-Star Game, becoming the youngest All-Star in NBA history at 19. The shooting guard then teamed up with superstar center Shaquille O'Neal to win three consecutive NBA championships and was voted first-team all-NBA from 2002-2004. He also inked multi-year endorsement deals with Adidas, Sprite and other top sponsors.

Although the Lakers struggled after O'Neal left in 2004, Bryant performed brilliantly. He scored 81 points against the Toronto Raptors in January 2006, the second-highest single-game mark in NBA history, and led the league in scoring that year and the next.

In 2008, Bryant was named Most Valuable Player and carried his team to the NBA Finals, where they lost to the Boston Celtics. In the 2009 NBA Finals, the Lakers beat the Orlando Magic to win the championship. Shortly afterward, Bryant was part of the memorial service to honor friend and music superstar Michael Jackson. The following year, the Lakers won their second straight title by defeating the Celtics.

Bryant played on both the 2008 and 2012 U.S. Olympic teams, winning consecutive gold medals with teammates Kevin Durant, LeBron James and Carmelo Anthony, among several other top players.

After suffering a torn Achilles tendon in April 2013, Bryant worked hard to return to the court before fracturing his knee just six games into the 2013-2014 season. The veteran All-Star surpassed Michael Jordan for third all-time on the NBA scoring list in December 2014, but his season ended due to injury for the third straight year when he sustained a torn rotator cuff in January 2015.

Although Bryant returned in time for the start of the 2015-2016 NBA season, he personally struggled alongside his young Lakers teammates. In November 2015, he announced that he would retire at the end of the season. "This season is all I have left to give," he wrote on The Players' Tribune website. "My heart can take the pounding. My mind can handle the grind but my body knows it's time to say goodbye."

The announcement drew a strong reaction, particularly from NBA Commissioner Adam Silver. "With 17 NBA All-Star selections, an NBA MVP, five NBA championships with the Lakers, two Olympic gold medals and a relentless work ethic, Kobe Bryant is one of the greatest players in the history of our game," Silver said in a statement. "Whether competing in the finals or hoisting jump shots after midnight in an empty gym, Kobe has an unconditional love for the game."

On April 13, 2016, Bryant dazzled a sold-out crowd at the Staples Center and fans everywhere in the last game of his career, scoring 60 points and leading the Lakers to a win against the Utah Jazz. It was Bryant’s sixth 60-point game of his career.

After the game, Bryant spoke to the crowd. "I can't believe how fast 20 years went by," he said. "This is absolutely crazy ... and to be standing at center court with you guys, my teammates behind me, appreciating the journey that we've been on — we've been through our ups, been through our downs. I think the most important part is we all stayed together throughout."

An all-star lineup of Laker icons also paid tribute to Bryant, including O’Neal, Phil Jackson, Pau Gasol, Derek Fisher, Lamar Odom and Magic Johnson. "We are here to celebrate greatness for 20 years," Johnson said. "Excellence for 20 years. Kobe Bryant has never cheated the game, never cheated us as the fans. He has played through injury, he has played hurt. And we have five championship banners to show for it."

In November 2015, Bryant announced his upcoming retirement from the Lakers with a poem on The Players' Tribune website, titled "Dear Basketball." The athletic great soon sought the best in other fields to turn his poem into a short film, including Disney animator Glen Keane and composer John Williams.

The result was a beautifully rendered five-minute, 20-second film, which debuted at the 2017 Tribeca Film Festival. Oscar voters took note, leading to the unexpected sight of Bryant accepting an Academy Award for Best Animated Short Film at the 2018 ceremony.

The Academy of Motion Picture Arts and Science's short films and animation branch also extended an invitation for Bryant to become a member of the organization. However, in June 2018 it was revealed that the Academy's governors committee had rescinded the invitation, saying the retired basketball great needed to show more efforts in the field before being considered for membership.

In July 2003, Bryant was charged with one count of sexual assault on a 19-year-old female hotel worker in Colorado. Bryant said he was guilty of adultery but innocent of the rape charge. The case against Bryant was dismissed in 2004, and he settled the civil lawsuit filed by the hotel worker against him out of court.

Among his philanthropic endeavors, the basketball great partnered with the non-profit After-School All-Stars as part of the Kobe & Vanessa Bryant Family Foundation. He also ran an annual summer camp called the Kobe Basketball Academy.

Bryant married 19-year-old Vanessa Laine in April 2001. The couple became parents to four daughters: Natalia Diamante (b. 2003), Gianna Maria-Onore (b. 2006, d. 2020), Bianka (b. 2016) and Capri (b. 2019).

On January 26, 2020, Bryant was onboarding a Sikorsky S-76 helicopter that crashed in the Los Angeles suburb of Calabasas. Nine people, including Bryant and his 13-year-old daughter Gianna "Gigi," died. The helicopter was on its way from Orange County to Thousand Oaks, California where Bryant was scheduled to coach a tournament game at Mamba Sports Academy.

Bryant and his daughter were laid to rest in a private funeral on February 7, 2020. On February 24, they were honored in a memorial service at the Staples Center, with Beyoncé and Alicia Keys delivering musical tributes and Jordan, Shaq and Vanessa Bryant among those sharing emotional recollections of the basketball great and family man.[[1]](#footnote-1)

## Sentiment Analysis

Sentiment analysis is the automated process that uses AI to identify positive, negative, and neutral opinions within text. Sentiment analysis models detect polarity within a text (e.g. a positive, negative or neutral opinions), whether it’s a whole document, paragraph, sentence, or clause. Sentiment analysis assumes various forms, from models that focus on polarity (positive, negative, neutral) to those that detect feelings and emotions (angry, happy, sad, etc.), or even models that identify intentions (e.g. interested v. not interested).

## Topic Modeling and LDA

Topic modeling is one of the most powerful techniques in text mining for data mining, latent data discovery, and finding relationships among data and text documents. Researchers have published many articles in the field of topic modeling and applied in various fields such as software engineering, political science, medical and linguistic science, etc. There are various methods for topic modelling; Latent Dirichlet Allocation (LDA) is one of the most popular in this field. Researchers have proposed various models based on the LDA in topic modeling. According to previous work, this paper will be very useful and valuable for introducing LDA approaches in topic modeling. In this paper, we investigated highly scholarly articles (between 2003 to 2016) related to topic modeling based on LDA to discover the research development, current trends and intellectual structure of topic modeling. In addition, we summarize challenges and introduce famous tools and datasets in topic modeling based on LDA.[[2]](#footnote-2)

## What is Naive Bayes?

In machine learning, Naive Bayes classifiers are the part of simple machine learning. Naive Bayes is popular algorithm which is used to find the accuracy of the news whether its real or fake using multinomial NB and pipelining concepts. There are number of algorithms that focus on common principle, so it is not the only algorithm for training such classifiers. For example, to check if the news is fake or real Naive Bayes can be used.

Naive Bayes is a kind of algorithm used in text classification. The use of token is correlated with the news that may be fake or not fake in Naive Bayes classifier and then the accuracy of the news is calculated by using Bayes theorem.

In addition, Naive Bayes is a probabilistic classifier that makes classifications using the maximum A Posteriori decision rule in a Bayesian setting. It can also be represented using a very simple Bayesian network. Naive Bayes classifiers have been especially popular for text classification, and are a traditional solution for problems such as spam detection.[[3]](#footnote-3)

The goal of any probabilistic classifier is, with features x\_0 through x\_n and classes c\_0 through c\_k, to determine the probability of the features occurring in each class, and to return the most likely class. Therefore, for each class, we want to be able to calculate P(c\_i | x\_0, …, x\_n).

there is very little explicit training in Naive Bayes compared to other common classification methods. The only work that must be done before prediction is finding the parameters for the features’ individual probability distributions, which can typically be done quickly and deterministically. This means that Naive Bayes classifiers can perform well even with high-dimensional data points and/or a large number of data points.

## Naïve Bayes Classification

Naive Bayes is a type of classifier considered to be a supervised learning algorithm, which belongs to the Machine Language class and works by predicting “membership probabilities” for each individual class, for instance, the likelihood that the given evidence, or record, belongs to a certain class. The class with the greatest, or highest probability, shall be determined the “most likely class,” which is also known as Maximum A Posteriori (MAP)

Another way of thinking about Naive Bayes classifier is that this method uses the “naive” notion that all features are unrelated. In most cases, this assumption of independence is outrageously false. Suppose Naïve Bayes classifier is scanning an article and comes across “Barack,” in many cases the same article will also have “Obama” contained in it. Even though these two features are clearly dependent, the method will still calculate the probabilities “as if they were independent,” which does end up overestimating “the probability that an article belongs to a certain class”.

Since Naïve Bayes classifier overestimates the probabilities of dependencies, it gives the impression that it would not work well for text classification. On the contrary, Naïve Bayes classifier still has a high-performance rate even with “strong feature dependencies,” since the dependencies will actually end up cancelling out each other for the most part.

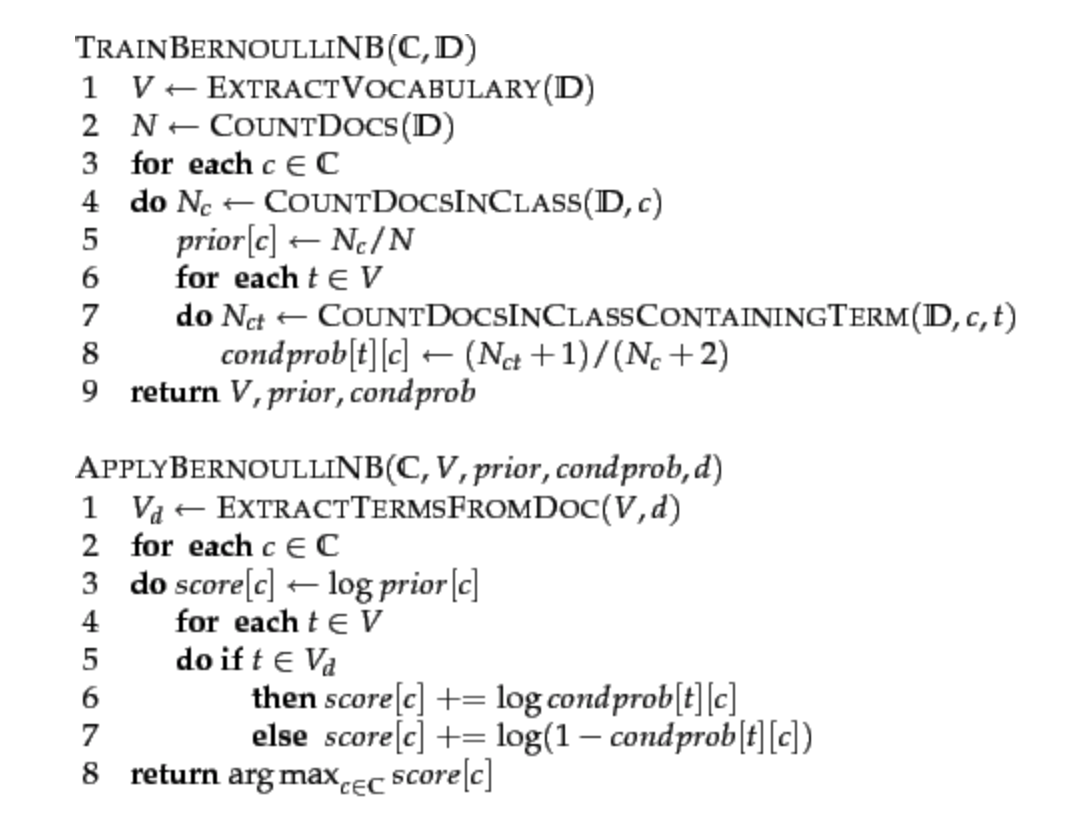
In addition, what makes Naïve Bayes classifier desirable is that it’s relatively fast and a highly accessible technique. It can be used for binary or multiclass classifications, making it an excellent choice for “Text Classification problems” as mentioned earlier. Also, Naïve Bayes classifier is a straightforward algorithm that only really relies on performing many counts. Thus, it can be “easily trained on a small dataset”.[[4]](#footnote-4)

## Multinomial Naïve Bayes

Naive Bayes has three flaws when applied to document classification. First, a word’s non-appearance counts just as much its appearance, whereas surely a document’s class is determined by the words that are in it rather than those that aren’t? Second, Naive Bayes doesn’t take account of the number of appearances of a word, whereas surely frequently occurring words should have a greater influence on the class than ones that only appear once? Third, it treats all words the same, whereas surely unusual words like “weka” and “breakfast” should count more than common ones like “and” and “the”? Multinomial Naive Bayes is a classification method that solves these problems and is generally better and faster than plain Naive Bayes.

## Bernoulli Model

An alternative to the multinomial model is the multivariate Bernoulli model or Bernoulli model. It is equivalent to the binary independence model, which generates an indicator for each term of the vocabulary, either indicating presence of the term in the document or indicating absence. Figure 1, below, presents training and testing algorithms for the Bernoulli model. The Bernoulli model has the same time complexity as the multinomial model.[[5]](#footnote-5)



**Figure 1: NB algorithm (Bernoulli model). Training and testing**

# Dataset Analysis

## About the Data

**Dataset Name**:

kobe\_tweet\_data\_01\_28.csv, kobe\_tweet\_data\_01\_29.csv, kobe\_tweet\_data\_01\_30.csv, kobe\_tweet\_data\_01\_31.csv, kobe\_tweet\_data\_02\_01.csv, kobe\_tweet\_data\_02\_02.csv, kobe\_tweet\_data\_02\_03.csv, kobe\_tweet\_data\_02\_04.csv

**Dataset format**: The data has been collected in comma-separated-value format.

**Number of columns and rows**: 17 columns 35857 rows (combined)

**Columns Names**: id, created\_at, source, original\_text, clean\_text, sentiment, polarity,

subjectivity, lang, favorite\_count, retweet\_count, original\_author, possibly\_sensitive, hashtags, user\_mentions, place, place\_coord\_boundaries

**Data Dictionary**:

Id = Unique id of the tweet

Created\_at = Tweet date and time

Source = Source of the Tweet

Original\_text = The original tweet that posted on the Twitter

Clean\_text = Tweepy preprocessing called for basic preprocessing

Sentiment = Pre-sentiment analysis of the tweet

Polarity = Separated the polarity from Sentiment

Subjectivity = Separated the subjectivity from Sentiment

Lang = Language of the tweet

Favorite\_count = the number of favorites per tweet

Retweet\_count = the number of retweets

Original\_author = Profile username of the tweet’s author

Possibly\_sensitive = the tweet data is for NSFW content on Twitter.

Hashtags = The hashtags used by the original author

User\_mentions = Any other profile mentions in the tweets

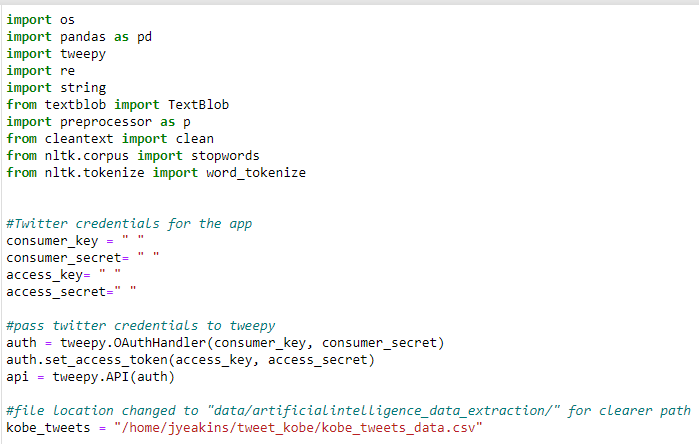
Place = The location of the original author

Place\_coord\_boundaries = The coordinates of the original author’s place

## Data Collection

Collecting historical data from Twitter is not an easy process. The user must understand the limits and guidelines to properly collect the historical data. Twitter limits the accessibility of its users to collect the data. A user can only access data from 7-days prior to the date the user starts the data collection. Also, each account can only collect a limited number of tweets per session before Twitter locks the account from the calling API.

The code contains 17 steps to create one csv file. It starts from importing libraries that are required to do all 17 steps. Once the libraries are imported, the twitter credentials are entered to call API by using ‘Tweepy’. Then the columns of the csv file will be defined. All columns of the csv file are listed above with the data dictionary.



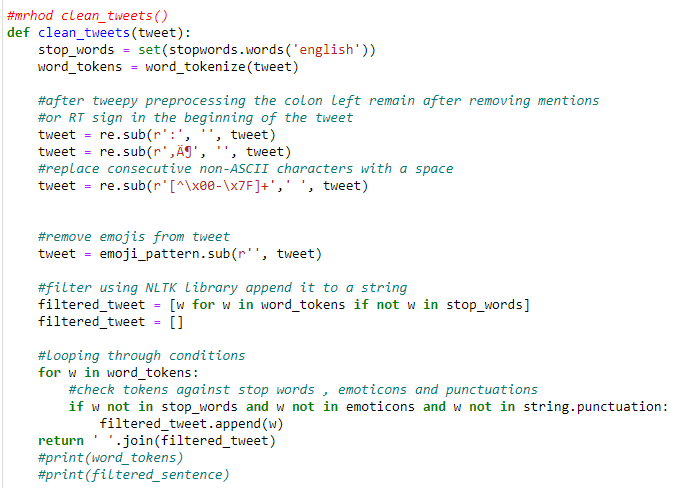
**Figure 2: Python code**

The code will contain a part where it sets the columns of final csv file, the start date of tweet collection and how it handles the emoticons and emojis by setting up the variables for emoticons and emoji and a block of code to recognize them.



**Figure 3: Cont. Python code**

The next portion will be cleaning the data. The tweets will be collected and stored in a raw format, but there will be another column for the clean texts. This won’t be used for the future analysis, but having a column to store cleaned texts will be useful to compare different text cleaning tools and see which gives better results. For the later analysis, this cleaned texts won’t be used.

****

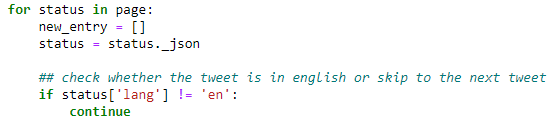
**Figure 4: Cont. Python code**

From this block of code, a method called ‘clean\_twets’ and this will celan some of the twitter data and double check emoticons and emojis because some older version of mobile’s emoticons is not cleansed earlier. The stopwords and tokenizing methods used to break down the words from the tweets.



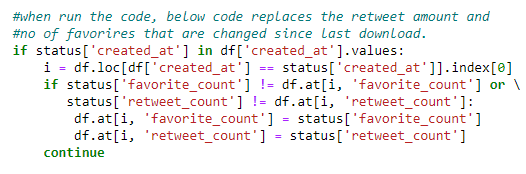
**Figure 5: Cont. Python code**

To connect to Twitter’s API, this code will be using a Python library called ‘Tweepy’, which is a supported tool for accessing the Twitter API. It supports Python 2.6 and 3.6 or above. In the code block below, there are two parameters; one for the file name to take care of as well as the keywords for this file to be filled with which explains the second parameter.



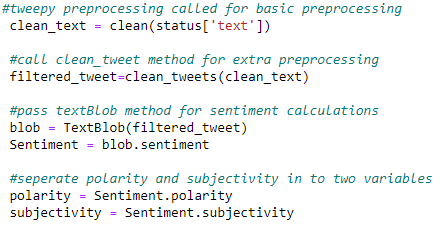
**Figure 6: Cont. Python code**

The tweets that received from the Twitter API is in a JSON format and has quite an amount of information attached. By creating an array of string named ‘new\_entry=[]’ to store all the JSON parsed data on each iteration, and it continues to retrieve data only if the language is English.



**Figure 7: Cont. Python code**

The block of code above is to replace the retweet amount and number of favorites that have changed since the last download.



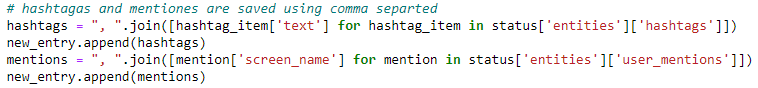
**Figure 8: Cont. Python code**

By using the ‘preprocessing’ and the block of code that was defined earlier, it will clean the tweet texts and save them as ‘filtered\_tweets’. One the filtered\_tweets are created, the sentiment property returns a named tuple of the form sentiment (polarity, subjectivity). The polarity score is a float within the range [-1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective. Pass the filtered\_tweet to TextBlob for sentiment calculation and it separately stored sentiment, polarity and in three different variables.



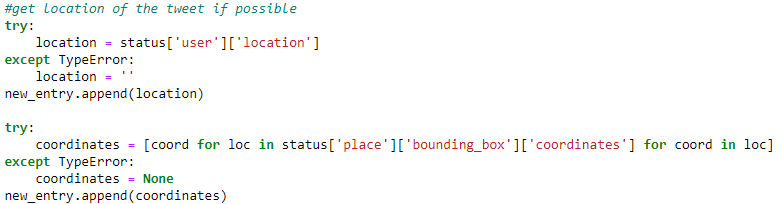
**Figure 9: Cont. Python code**

This block of code is to append the JSON parsed data to the string array that it created. Those appended data is the data that was extracted from twitter by using Tweepy. But there are much more data fields. In order to follow the sequence it gathers the original author’s user name (Twitter profile name of the Tweet). Possibly\_sensitive column of the tweet data is for NSFW content on Twitter.



**Figure 10: Cont. Python code**

The code block above is the most important part since it gathers the hashtags of the tweets. This will store the hashtags that are collected by using the keywords that will be defined later in the code. Also, the mentions are the screen name of other Twitter users in the tweets.



**Figure 11: Cont. Python code**

Next block of code is to track down the location of the tweet but practically this is a bit hard and nearly impossible because most of the users don’t allow Twitter to access their locations. But the codes are written to get the user’s profile location instead of tweet’s location since it basically gives an idea of the region and country the user is located.



**Figure 12: Cont. Python code**

Once everything is done, it is time to wrap up everything. The two lines above is to take a single tweet data and append it to a data frame called ‘df’.



**Figure 13: Cont. Python code**

After the data is stored in ‘df’, ‘to\_csv’ is used to write all the data it gathered into the particular CSV file.



**Figure 14: Cont. Python code**

The keywords (hashtags) are defined here. All the possible hashtags that related to Kobe Bryant and his death will be written here to collect the tweets by using those keywords.

# Analysis and Models

## Loading the data to Google ‘Colab’

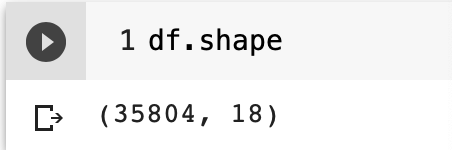
Loading the data to Google ‘Colab’:



**Figure 15: Loading the data to Google Colab**

## Review the data

35,804 observations.



**Figure 16: 35,504 observation**

## Cleans the data

Cleansing the data #1



**Figure 17: Cleansing the data**

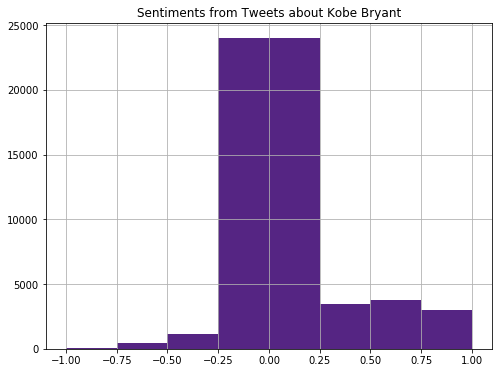
Cleanse the data #2



**Figure 18: Cleansing the data #2**

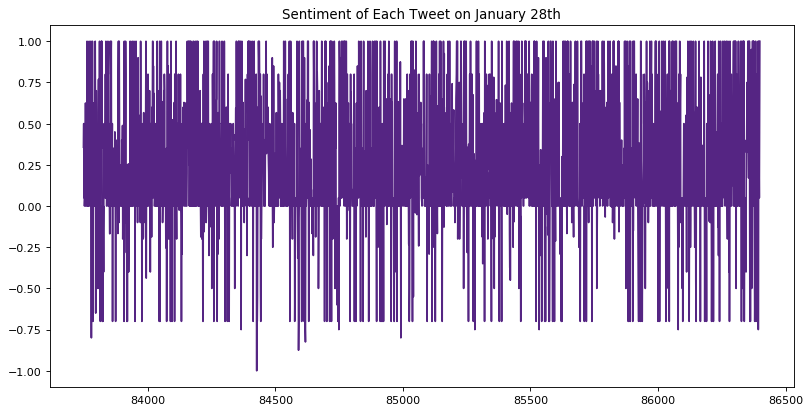
## Visualization

Initial explorations of the data focused on statistical analysis of hashtag and word frequencies.



**Figure 19: Exploratory Analysis #1**

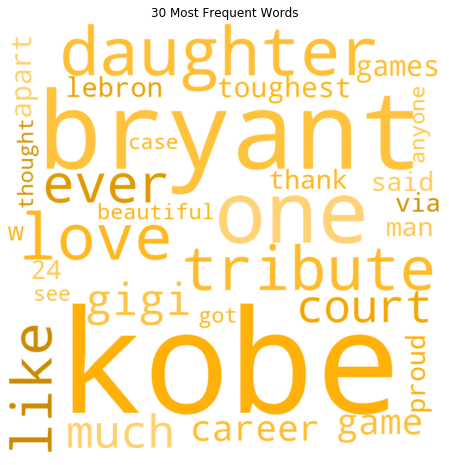
Over all tweets, the sentiments were largely neutral though the trend leaned toward positive more than negative.

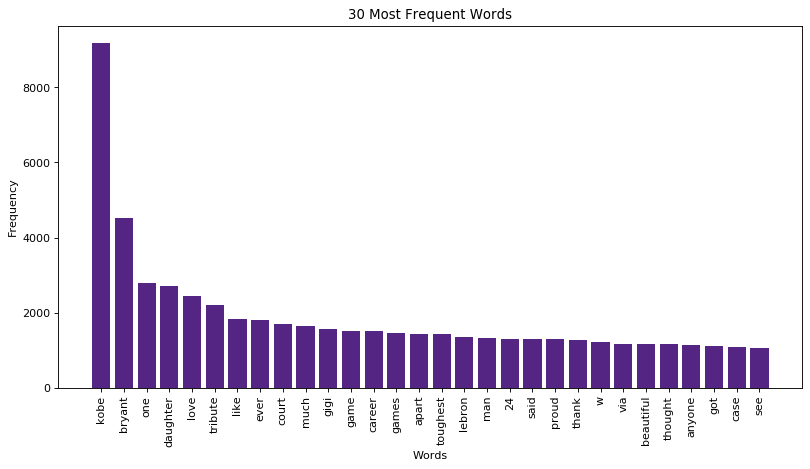


**Figure 20: Exploratory Analysis #2**

A single day was isolated to see if there was any pattern in sentiment of tweets over the course of a day.

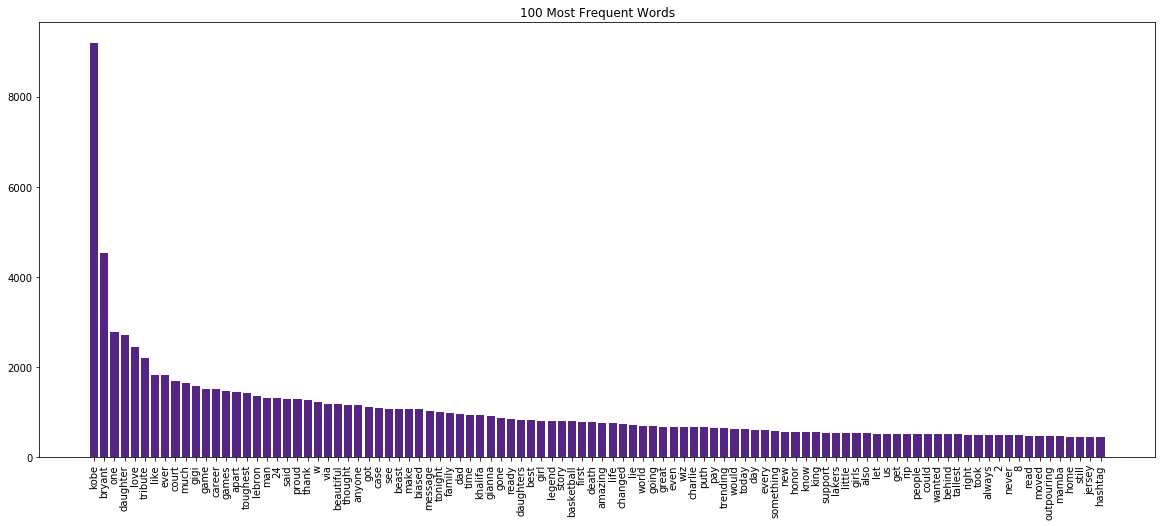
On January 28th, tweets pulled at the time of larger negative spikes proved to be backlash against news that featured false reports or wild speculation.

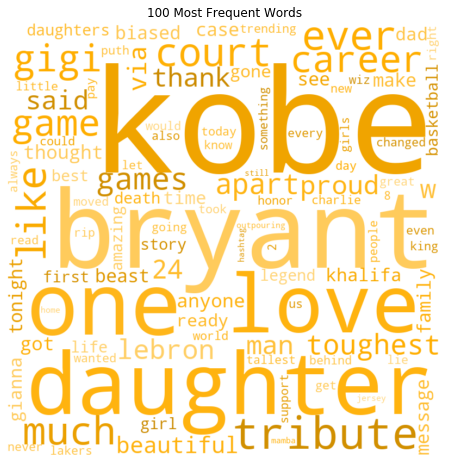
Exploring Word Frequencies and Word Clouds



**Figure 21: Exploratory Analysis 3 and 4**

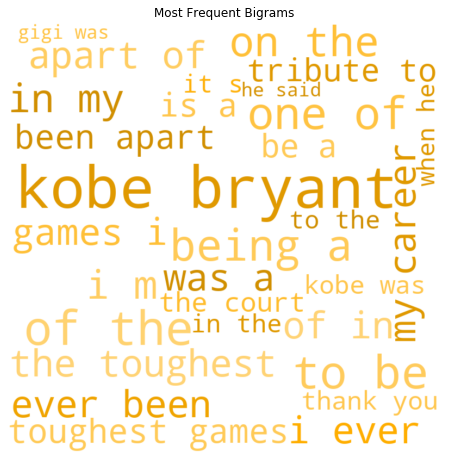
The partially cleaned tweets were all put into a single corpus, tokenized, and had stop words removed. It was then passed through a count vectorizer and the most frequent words were plotted. The bar plot and WordCloud above illustrates the 30 most frequently occurring words. It is evident from the word cloud that the discussion on twitter involved many references to his daughter and to his basketball career.





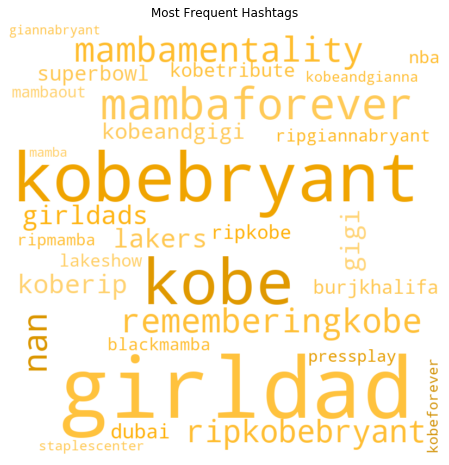
**Figure 22: Exploratory Analysis #5 and #6**

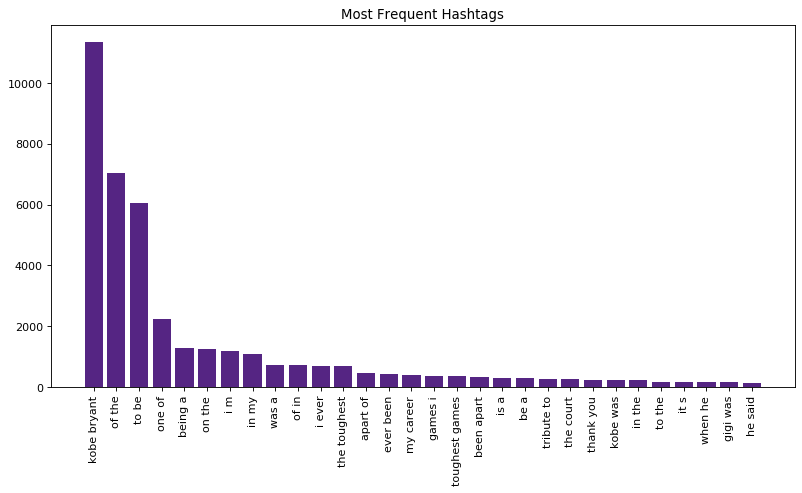
The word cloud involving the 100 most common words revealed additional mentions relevant to his work ethic, his perspective on being a father, and many positive emotional words. This is one of the first indications that the analysis may indicate a marketing campaign could expand beyond Kobe’s reputation as a professional athlete.



**Figure 22: Exploratory Analysis #5 and #6**

Additional consideration was placed on bigram frequencies without stop words removed. This confirmed the theme surrounding positive emotional language as well as references to his career as a professional basketball player.

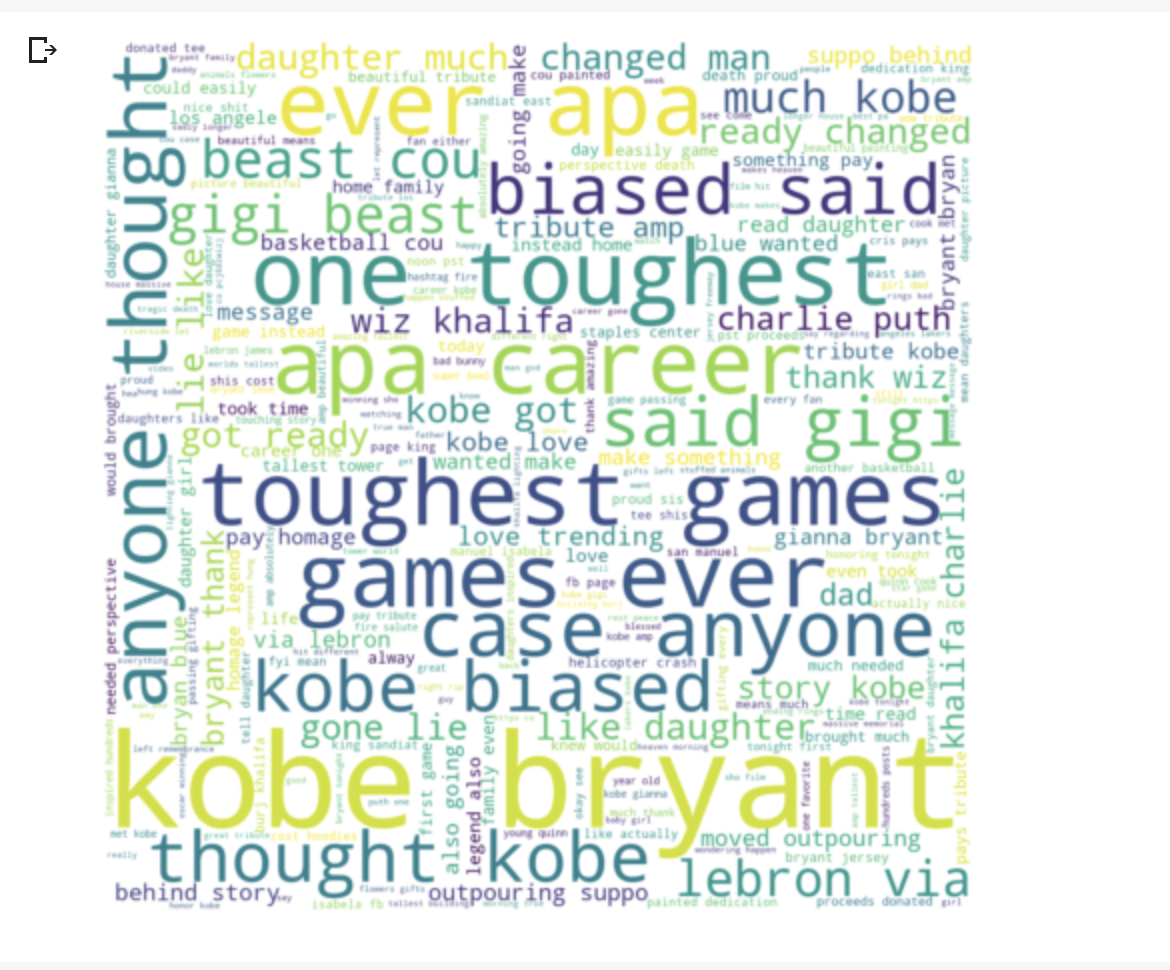




**Figure 23: Exploratory Analysis #7 and #8**

The last frequency bar plot and word cloud focused on hashtags. A key discovery was that ‘#girldad’ was more viral than any other hashtag.

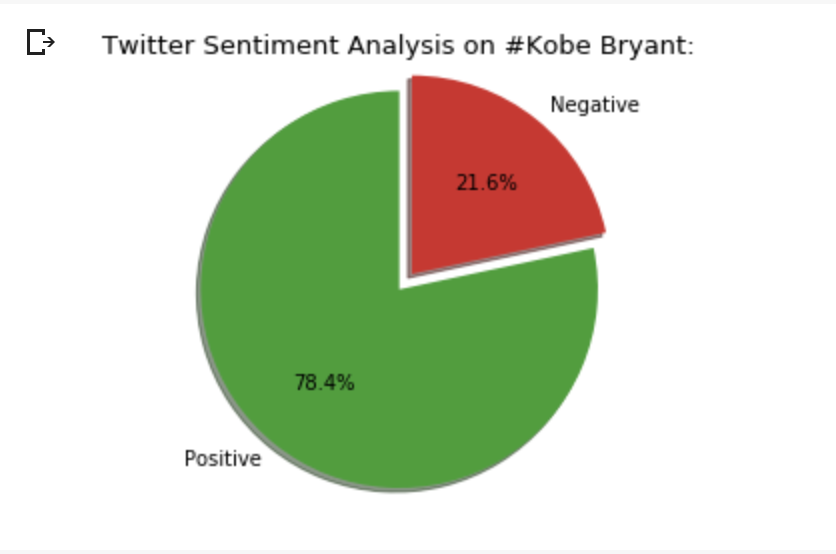
## Word Cloud



**Figure 24: Word Cloud**

## Tweets Sentiment Analysis

At this stage, we are performing sentiment analysis of all tweets:



**Figure 25: Tweets Sentiment Analysis**

## Positive and Negative Tweets

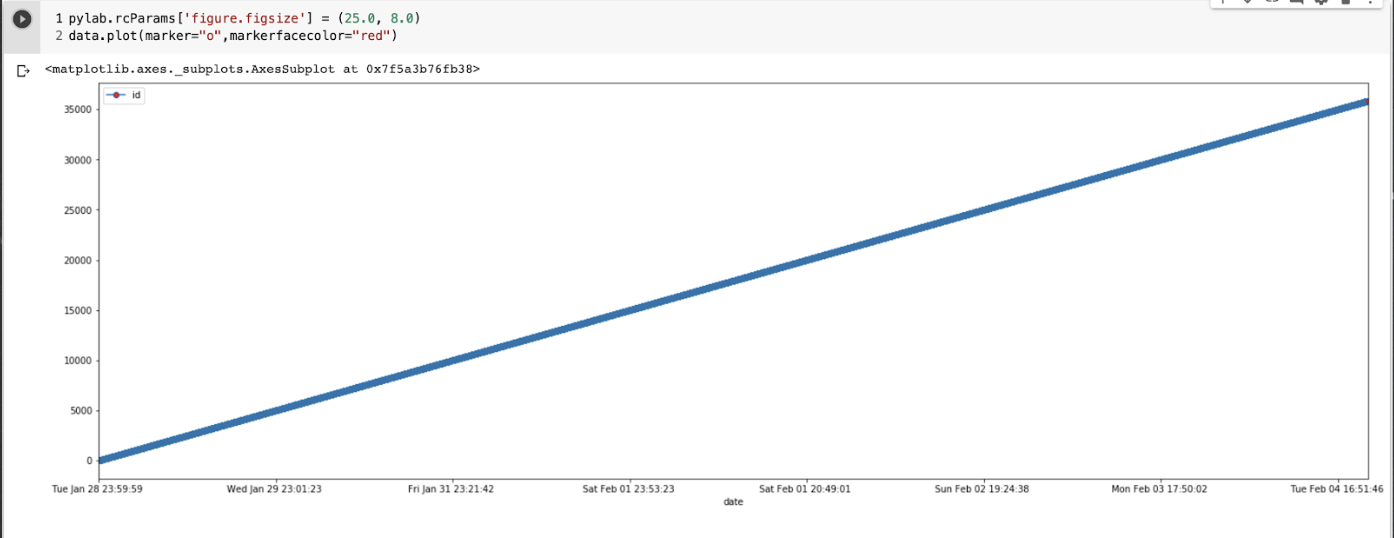
Number of positive and negative tweets:



**Figure 26: Positive and Negative Tweets**

## Time Series

Now, we are building the time series to understand better tweets, users and number of tweets:

****

**Figure 27: Time Series**

## SVM Model

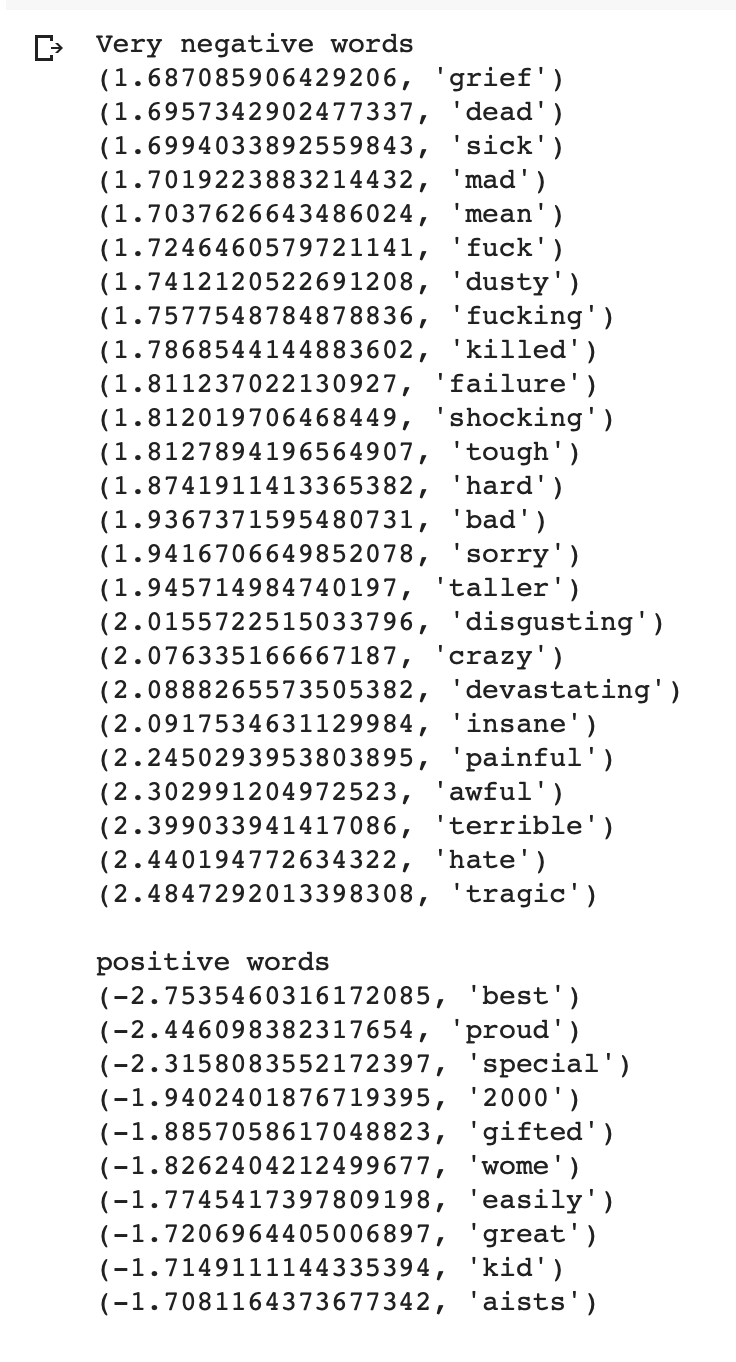
Support Vector Machine came to be the most successful algorithm in this overall process. This algorithm was used to predict based on training data, whether or not a tweet had a negative or positive or neutral sentiment. It also categorized and sorted the words that had an extremely negative sentiment as well as very positive sentiment.

****

**Figure 28: SVM**

## Positive and Negative Words

After carefully going through each algorithm, the best one that gave us the highest accuracy results was the Support Vector Machine that gave an accuracy of 96%. From the results gathered, this was the actual outcome:

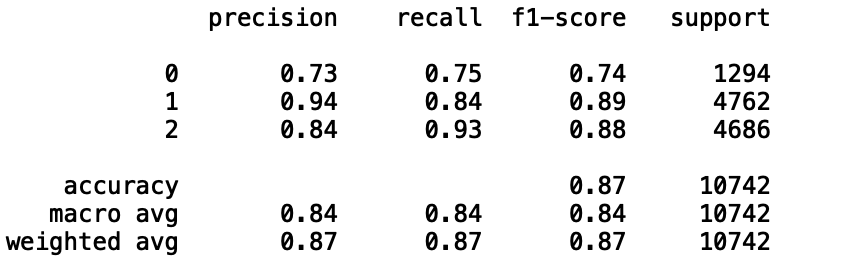
****

**Figure 29: SVM**

## Bernoulli

The Bernoulli algorithm required a vectorization with boolean outputs:

With this line of code, Bernoulli gave the best performance and accuracy:

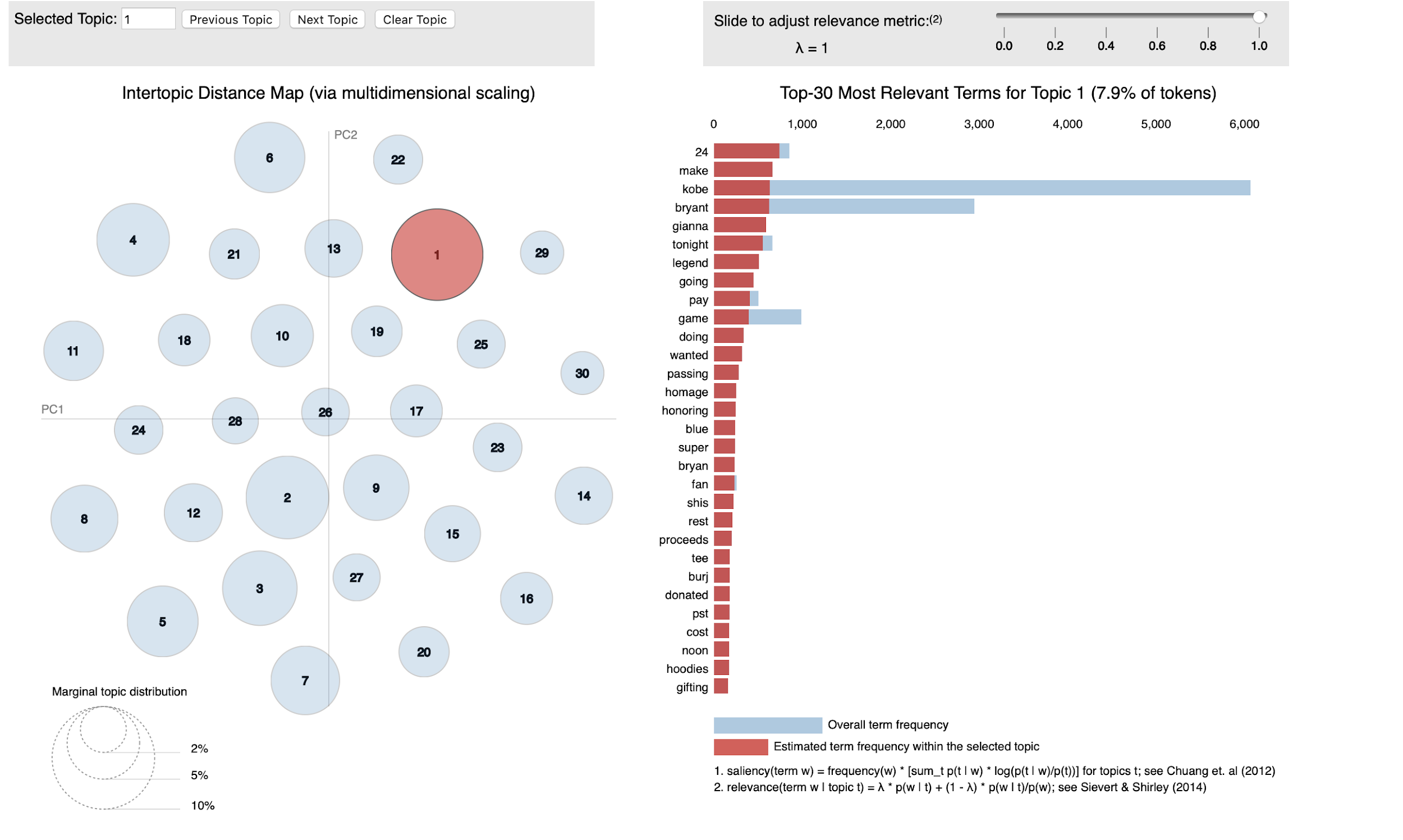


**Figure 30: Bernoulli**

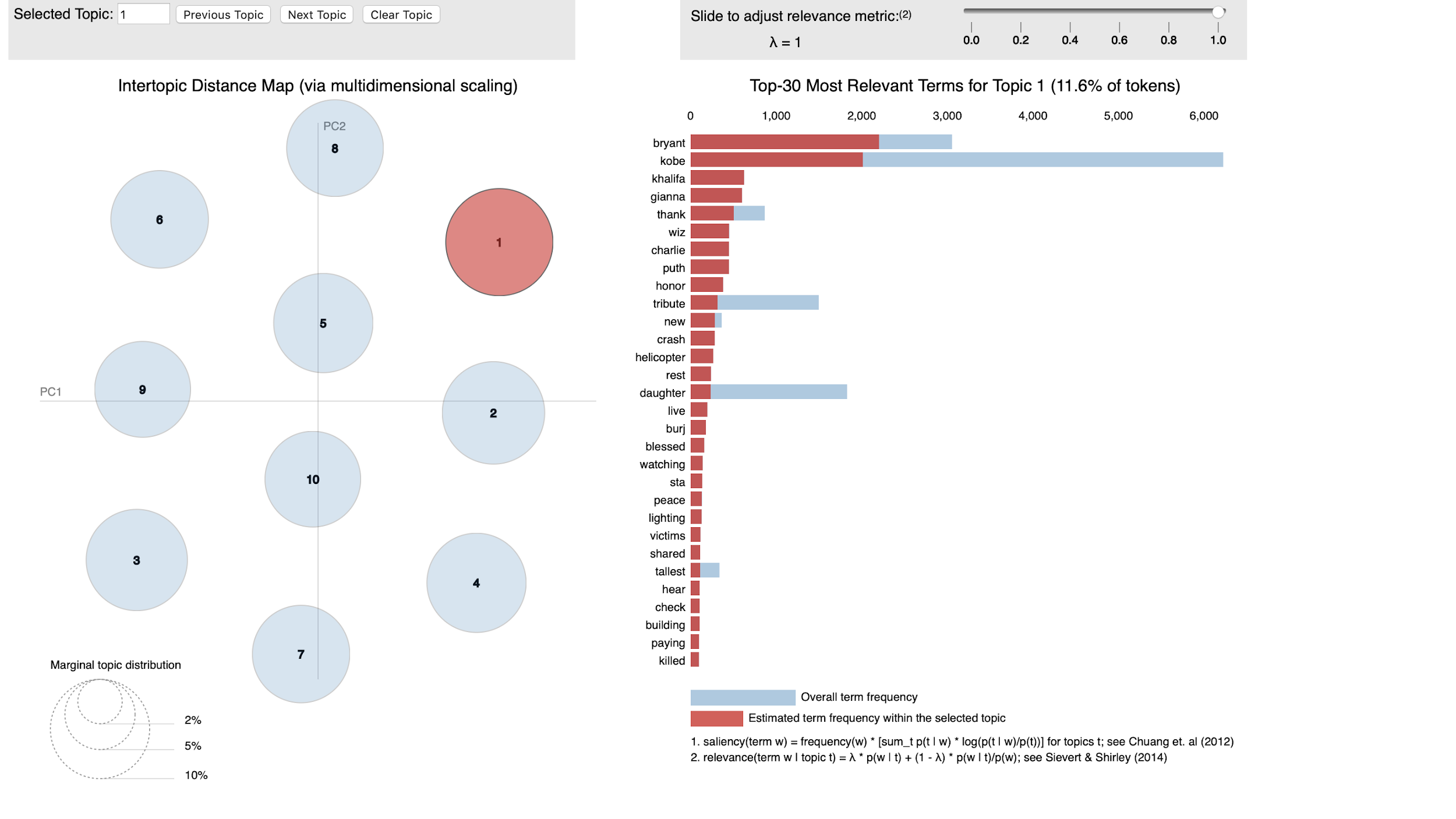
Another factor that helped improve the accuracy of Bernoulli was adding more training data. At one point, adding more training didn't really improve the efficiency so it came down to finalizing it once no improvement was made.

## Topic Modeling

In order to dive more into the tweets that were posted about Kobe and find out any tweets related to the sexual assault accusations. There were two Topic Models developed. The first topic modelling had 30 topics and the second one had 8. The results gathered from the 30-topic modelling didn't really give any sort of topic related to Kobe being involved in any accusations. In the second Topic Modelling with only 8 topics, didn’t really show any results that could’ve given us any information about him involved in any sexual assault.

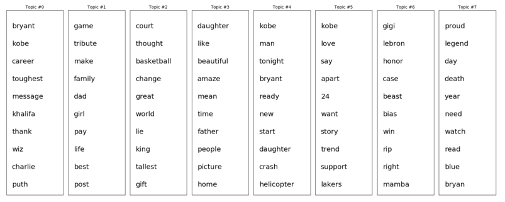


**Figure 31: Top modeling of 30**



**Figure 32: Top modeling of 8**

Another experimentation with topic modeling yielded several potential topic clusters.







**Figure 33: Top modeling Outcome**

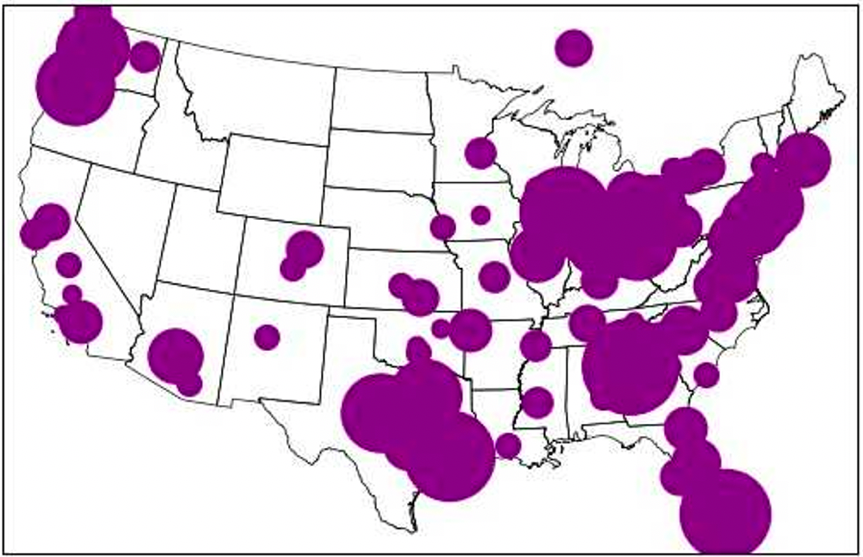
Topic models were run for 8, 6, 4, and 3 topics. The topic clusters confirmed the themes that the initial exploratory visualizations highlighted. Beyond his professional accomplishments, tweets about Kobe Bryant frequently discussed his role as a father and his ability to carry his work ethic into every aspect of his life.

A search for tweets containing the key words identified by topic modeling also highlighted many anecdotes about personal interactions with Kobe that illustrated his compassion, focus, and drive. Should a marketing campaign be centered on Kobe Bryant, a company would find that an emphasis on his family life and work ethic to resonate strongly with a diverse audience.

**Figure 31: Top modeling Outcome**

## Tweets Location

In order to identified the possible location for campaigns around the country, we have identified the tweets location with positive sentiments. As you can see the East Coast as well as the West Coast are the majority locations.



**Figure 34: Positive sentiment location for possible Campaigns**

# Conclusions

Kobe Bryant is one of the greatest basketball players of all time, but his influence extended far beyond the court. Kobe was as passionate about philanthropy as he was about pursuing baskets or championship rings. Kobe is leaving a legacy and he's teaching people how to be more than just an athlete.

The Kobe and Vanessa Bryant Family Foundation was one of the basketball star's most impactful philanthropic works. Bryant and his wife started the foundation to improve the lives of young people and their families. KVBF provides scholarships and operates Mamba FC, a youth soccer club that teaches young athletes how to become leaders and independent thinkers through sports.

Since 2011, the foundation has partnered with the United Way of Greater Los Angeles, Step Up on Second and homeless youth drop-in center My Friend's Place to provide comprehensive services to homeless kids in Los Angeles. Vanessa Bryant regularly visits the young parents at My Friend's Place and hosts an annual holiday celebration for them and their children.

Bryant was an honorary ambassador of After-School All-Stars, a non-profit organization that partners with schools nationwide to extend school hours for low-income children. As a champion for after-school programming for all, Bryant gave frequent motivational speeches and raised money for the organization's Los Angeles chapter, which serves 14,000 students. 98% of those students are youth of color.

Over the last 20 years, Bryant granted more than 250 wishes of children battling life-threatening illnesses through the Make A Wish Foundation. The foundation's Facebook page called Bryant "an amazing wish granter who has brought countless smiles to our wish kids and their families." Bryant was also a champion for people battling cancer. Outside of sending motivational messages to those fighting the disease, he raised money for research and medical expenses. In 2012, he was part of an all-star telecast that raised more than $81 million for research through Stand up to Cancer. Bryant and his wife donated at least $1 million to the National Museum of African American History and Culture. The couple is one of the museum's founding donors. When the museum opened in 2016, Bryant tweeted out this message: "Go. See this. Museum. There is no greater testament to this country than the stories in this building."[[6]](#footnote-6)

Analyzing the sentiments one week after Kobe’s accident was not an easy thing to do and involved with a lot of feelings and touch each one of us a group that need to conduct a scientific research. During the analysis process we have analyzed sentiments, cleansed the data, visualized the data and conduct predictions that will might use commercial companies for possible campaigns even thou Kobe Bryant is not with us anymore. The majority of the sentiments were positive and even the negative tweets didn’t really indicate any kind of negativity toward Kobe’s. Our conclusion that companies decide to conduct commercial campaign will be successful and specially in the East and West Coast.

# References

Twitter.com

<https://developer.twitter.com/en/docs/tutorials/choosing-historical-api>

Tweepy.org

<https://www.tweepy.org/>  
  
Earthscience.org - collect tweet eg.

<https://www.earthdatascience.org/courses/earth-analytics-python/using-apis-natural-language-processing-twitter/get-and-use-twitter-data-in-python/>

The Biography.com

<https://www.cnn.com/2020/01/27/us/honoor-kobe-bryant-charity-iyw-trnd/index.html>

1. The Biography.com [↑](#footnote-ref-1)
2. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey [↑](#footnote-ref-2)
3. <https://towardsdatascience.com/introduction-to-naive-bayes-classification-4cffabb1ae54> [↑](#footnote-ref-3)
4. <https://www.csustan.edu/sites/default/files/groups/University%20Honors%20Program/Journals/02_stahl.pdf> [↑](#footnote-ref-4)
5. <https://nlp.stanford.edu/IR-book/html/htmledition/the-bernoulli-model-1.html> [↑](#footnote-ref-5)
6. <https://www.cnn.com/2020/01/27/us/honoor-kobe-bryant-charity-iyw-trnd/index.html> [↑](#footnote-ref-6)