SOCIAL MEDIA ANALYSIS OF AFLAC

ROHITHA REDDY, ARIMANDA

TABLE OF CONTENTS

INTRODUCTION

Companies around the world are overwhelmed with the huge influx of data, mostly unstructured through multiple sources on daily basis. Insurance companies are no exception. For example, insurance companies collect huge amounts of text data in the form of claims and complaints, survey results, expert and health reports, consumer feedbacks via social networks etc. This superabundant data brings new possibilities but many challenges as well for the Insurance industry in this information age. So, the purpose of this project is to provide insights to Aflac by analyzing its consumer and employee conversations over social media networks like Twitter, Glassdoor and Reddit with help of text analytics and natural language processing.

Questions need to be Analyzed

- What is this distribution of Aflac social media sentiment (positive/negative/neutral) categorized by brand, product, service, or other?
- What is this distribution of Aflac social media sentiment (positive/negative/neutral) compared to other insurance companies?
- What topics are being discussed and what keywords are associated with Aflac?

To answer the above questions, I mainly focused on sentiment analysis, topic modeling and keyword extraction.

DATA DESCRIPTION

The text data related to Aflac and its competitors like MetLife, Cigna, Allstate, and Colonial Life was collected programmatically using Python from three different leading social media networks i.e. Twitter, Glassdoor and Reddit.

Twitter

The data from Twitter was scraped, cleaned and finally stored into a CSV file having columns:

1. User_ID: Unique identifier for the user
2. Tweet_ID: Unique identifier for the tweet
3. Timestamp: Time at which a message was tweeted
4. Tweet: Message tweeted by a user
5. Hashtags: Hashtags used by the user
6. Replies: Replies received by a user
7. Retweets: Number of times a tweet was forwarded by other users
8. Likes: Number of likes received by a user
9. Location: Geographical location of the user

Additional columns were added from preprocessing and analysis. These columns were State(state of the user), City(city of the user), Sentiment(type of the sentiment), Sentiment_score(confidence score), Emotion (type of emotion), Emotion_score(likelihood score), Company(name of the company), Category(name of product or service). This dataset was primarily used to study the user sentiments on different products and services.

This dataset has 61,362 rows and 17 columns.

The shape for each company's data files after preprocessing is as follows:
Aflac: 26948 rows and 17 columns
All State: 4797 rows and 17 columns
Cigna: 10629 rows and 17 columns
Met Life: 18756 rows and 17 columns
Colonial Life: 1306 rows and 17 columns

Glassdoor

The data from Glassdoor was collected by building a web scraper in Python and stored in CSV file with following columns.

1. Date – Date of when the review was posted
2. Review No – Unique number given to every review
3. Employee Type – Current or Former Employee
4. Designation – Job title
5. Position – Role described by an employee
6. Summary – Review title given by an employee
7. Pro - Upside of the company
8. Con – Downside of the company
9. Advice – Suggestion given by an employee
10. Overall Star - Overall Rating of the company between 1.0 and 5.0
11. Work Life Star - Rating given for Work Life
12. Culture Star - Rating given for company's culture
13. Career Opp Star - Rating given for career opportunities
14. Com Benefit Star - Rating given for company's Benefits
15. Sr Management Star - Rating given to Senior management of a company
16. Recommend – Sentiment related to recommendation
17. Outlook - overall sentiment about the company
18. CEO - Sentiment related to top management
19. Company – Name of Company
20. Review Link – Link of review

This Glassdoor dataset has 12,020 rows and 21 columns.

The shape for each company's data files is as follows:
Aflac: 2336 rows and 21 columns
All State: 4253 rows and 17 columns
Cigna: 1954 rows and 17 columns
Met Life: 3068 rows and 17 columns
Colonial Life: 408 rows and 17 columns

Reddit

Reddit data was collected only for Aflac with help of Reddit API named PRAW which was further cleaned and finally stored into a CSV file having columns:

1. ID: Unique identifier for a comment
2. Comment: Comment made by users on a post

This dataset has 700 comments and 2 columns.

DATA COLLECTION & PREPROCESSING

Twitter

Twitter is a popular social network where users can share short messages called tweet. Users share blogs, opinions, links, and pictures on Twitter, companies promote products and engage with customers. There are many ways to collect data from Twitter using Python, the most popular is Twitter Streaming API. The Python packages here used to scrape data are Beautiful Soup and Selenium. The main purpose of using Beautiful Soup is to scrape old data which is not possible to do with Twitter Streaming API.

Beautiful Soup is a Python library that allows getting data out of HTML, XML and markup languages. Beautiful Soup helps in pulling selected content from a web page, remove the HTML markup and save the information. Selenium is library usually used to automate the process of web browser interaction from python. Selenium needs a driver to interface with the browser, for example, ChromeDriver is required for Google Chrome and must be installed before implementing the code. So, with the help of Beautiful Soup and Selenium, I made a Twitter scroller and collected the data of Aflac, MetLife, Allstate, Cigna, and Colonial on basis of different hashtags and search queries for every company. Then the collected data was organized and stored in a CSV file for text mining and analysis.

The information extracted from Twitter was in raw form. In order to use them and get insights, I cleaned and transformed it into a more usable structured database. Generally, a tweet consists of special characters, emotions, punctuations, URLS, digits and actual message. The Python library named Regular Expression Operations has been used to remove the unnecessary data and get the clean message which could be used to do sentiment analysis.

Also, there were so many redundant tweets from the agents. In order to get rid of these tweets, I simply cleaned them and identified the duplicate ones and finally removed them. I also noticed that most of agents had words like ‘agent, the company name, ‘insurance’, ‘agency’, ‘insure’ etc. I used this pattern to identify the remaining agents and filtered them out. This removed thousands of unwanted information from the data.

For Topic Modelling, the message column has been further transformed by using the NLTK library in Python. First, tweets were converted to lowercase in order to bring the consistent form. Second, I removed all the stop words from the tweets with help of NLTK library. Third, I normalized our data with the help of Lemmatization technique by using the WordNetLemmatizer function of NLTK library. In the end, I tokenized the data by splitting it and making a large corpus of words.

The problem of scraping tweets by using Beautiful Soup is that you cannot get the location details of the user. So, to get the user's location I used a Tweepy and GeoText library in Python. I used our Twitter user credentials to access Twitter API. With help of API and username, I got every user's information which also includes the location details. Then I used GeoText to further identify the location into state and city. This method of extracting user's location could only be successful if the users have shared their location in privacy setting.

Glassdoor

Glassdoor is a popular job search engine and review website. Employees and former employees anonymously post a review about the company and their management. Glassdoor's data helps in learning about the organization's employer brand, company rating across key categories, employee reviews pros and cons and much more.

The process to collect data from Glassdoor is no different than Twitter. The scraper has been built with Beautiful Soup which scraped the data of Aflac and its competitors. The Request function of Urllib library has been used for opening and reading URL. Once the data is collected, it is arranged and stored in CSV file for text analysis.

Reddit

Reddit data was collected only for Aflac with help of Reddit API named PRAW. First, the posts from the subreddits **r/Frugal**, **r/personalfinance** and **r/jobs** mentioning 'Aflac' were identified and then they were shortlisted based on the number of comments received in order to take only popular posts into account. Finally, these comments were scraped with the help of Python and API and stored into CSV file for text analytics. In this way, 30 posts having around 700 comments in total were scraped.

The Natural Language Toolkit, or more commonly NLTK which is a suite of Python libraries and programs for symbolic and statistical natural language processing for English was used to perform part-of-speech tagging to extract nouns, adjectives, verbs and adverbs from the comments available. This processed data was then used for the purpose of topic modeling.

METHODS FOR DATA COLLECTION AND ANALYSIS

METHODS

Sentiment Analysis

Sentiment analysis is an automated process of understanding an opinion about a given subject from written or spoken language. It is a key tool for making sense of the data which has allowed companies to get key insights and automate all kind of processes. Here, I have performed sentiment analysis and made visualizations to compare social media sentiment of Aflac to that of its competitors like MetLife, Allstate, Cigna etc. across various products and services like health insurance, disability insurance, cancer insurance, accidental insurance etc.

Topic Modeling and Keyword Extraction

Topic modeling provides us with methods to organize, understand and summarize large collections of textual information. It helps in discovering hidden topical patterns that are present across the collection, annotating documents according to these topics and using these annotations to organize, search and summarize texts.

Here I will use topic modeling to infer the cause of the negative and positive sentiments for Aflac. I will also try to filter out important keywords with the help of word cloud and word correlations.

TOOLS USED

Amazon Comprehend API

Amazon Comprehend is a Natural Language Processing (NLP) service that uses machine learning to find insights and relationships in text. Amazon Comprehend understands natural language text from documents, social network posts, articles and any other textual data stored in AWS. It can identify text entities, the language of the text, the sentiment expressed in text, key phrases with concepts and adjectives and topic modeling that helps in finding out most common topics from a corpus of documents. Amazon Comprehend has been trained on a number of different datasets including customer reviews from Amazon.com, to build best-in-class language models that extract valuable insights from text data. The Amazon Comprehend get trained continuously on a large body of text so there is no need for the user to provide training data.

Each tweet has been passed through the Amazon API which gives sentiment and the sentiment score associated with that tweet. The tweet is categorized as Positive, Negative, Neutral and Mixed.

Indico

Indico is a python wrapper which allows you to do text and image analysis. The API supports to do tasks like sentiment analysis, facial image recognition, language detection, text topic tagging, and facial feature extraction and image feature extraction. I have used Indico API to do tone analysis which helps us in detecting emotions like anger, joy, fear, sadness, and surprise. It also gives the emotion score for each tweet. Indico uses pre-built machine learning algorithms, thus no training data is required.

Gensim

Gensim is a package for processing text, working with word vector models like Word2Vec and FastText and for building topic models. Topic modeling is a technique to extract the hidden topics from large volumes of text. LDA is a popular algorithm for topic modeling which generates topics based on word frequency from a set of documents. Each topic generated is combination of keywords and each keyword contributes a certain weight to the topic.

ANALYSIS & RESULTS

TOPIC MODELLING RESULTS

As mentioned earlier, the topic modeling technique helps us to make sense of data by revealing the latent topics present in the text corpus. Here I will analyze the results of topic modeling only on the negative comments and tweets so that I can somehow understand the source of negativity or concerns and issues mentioned by the customers and employees on the social media. I will also have a look at word clouds to get additional insight.

Twitter

For Twitter, I will analyze the negative tweets associated with Aflac, Cigna, Allstate and finally Colonial Life.

Aflac

Word Cloud

I can see from the above image that words ‘health insurance’, ‘coverage’ and some abusive words are highly visible. This indicates people are showing negative sentiments towards Aflac’s services like health insurance, policies, and coverage.

Topics

```
Topic 1: coverage insurance better cancer employee accident tasteless thought real racist
Topic 2: commercial fuck also hurt fire_voice lmao life_insurance glad time walk
Topic 3: tweet take hope today service looking aint lose fucking nascar
Topic 4: work much stupid back years sound dude major_medical many bullshit
Topic 5: fire really know need shit hate getting life funny give
Topic 6: claim never damn tsunami ever say time going lost right
Topic 7: make money still annoying company week business dumb go try
Topic 8: bitch suck look build_empire unspeakable_nightmare twitter trouble_economics month good help
Topic 9: health_insurance joke even pay call cover policy offer guess scam
Topic 10: voice want think people making tell well hear do lame
```

Here I have top 10 words for each of the 10 topics. I have to infer about the topic by looking at the words. But as I can observe that words under each topic are not so coherent. This might be because of the low lexical diversity of tweets.

But still, I can see that in topics 1,3,7 and 9 people are related to either bad customer service or concerns about health insurance. And topics 2,5 and 10 are representing sarcasm or insults regarding duck or maybe Gilbert Gottfried.

Cigna

Word Cloud

This word cloud above is showing most frequent words in the negative tweets related to Cigna. It suggests that people are discussing Medicare, dental insurance, health insurance, some issues related to claims, coverage, bills, and plans. There is also a mention of horrible customer service

Topics

```
Topic 1: health_insurance company good want refuse customer_service commercial work fail prescription
Topic 2: medicare complaint terrible le choose rate order provider health nothing
Topic 3: year medicare_medicaid program think already rather much nightmare life frustrate
Topic 4: medicare_advantage anthem claim well humana look people blame making service
Topic 5: aetna fraud health_insurance call premium customer different system phone sent
Topic 6: fuck cancel expensive hospital home profit_slump medicare_suspension ding_outlook actually disgust
Topic 7: medicare health_insurance plan pay insurance health suck coverage care cover
Topic 8: hate really never wish better med enroll money medicaid hope
Topic 9: worst first health_insurance scam horrible benefit fucking medical come years
Topic 10: insurance_company drop month thanks insurance drug work money give worse
```

From the topics above I can infer that Medicaid, Medicare and health insurance are the main topics being discussed in a negative context. This is indicated by topics 2,34,6,7,8. Also, topics 1,5,9,10 are related to health insurance.

Allstate

Word Cloud

Here it is evident that people are discussing customer support or customer service, calls, claims, agents, insurance and something about Laura Ingraham.

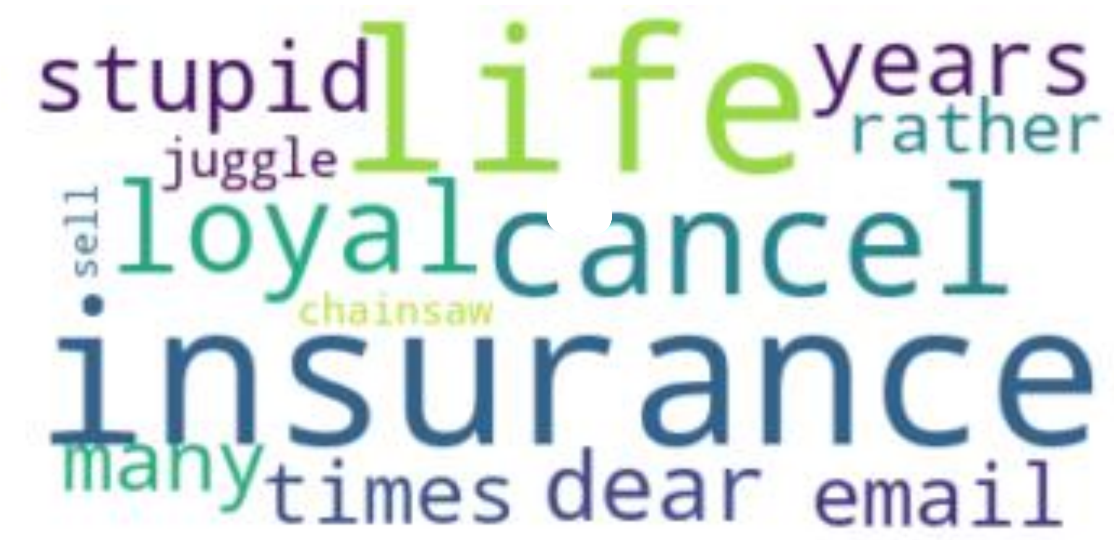
Topics

```
Topic 1: policy agent tell account contact give state sent morning thank
Topic 2: money even commercial horrible joke customer_service people sick stop phone
Topic 3: insurance good_hands insurance_company make customer business company rental geico auto
Topic 4: look work take free person accident hours think phone thing
Topic 5: try someone time insurance much month something worse friend actually
Topic 6: service tell week email cancel insult already home good suck
Topic 7: really years time take tweet send anymore find little message
Topic 8: care going company show customer back health_insurance probably dear hospital
Topic 9: claim health_insurance pay terrible adjuster customer wrong mess change worst
Topic 10: keep cover bill health_insurance provide client fraud looking least damage
```

Topic 1 is about agent and policy, topic 2 has something to do with customer service, topic 3 is about auto insurance and Geico company and topics 8,9 and 10 are related to health insurance.

Colonial Life

Word Cloud



Since there were not many useful tweets about Colonial Life, therefore the word cloud is not much informative. But still, I can see that words life and insurance are dominating the cloud.

Topics

```
Topic 1: times life juggle email fuck dear chainsaw insurance sell years
Topic 2: life insurance stupid cancel years loyal dear juggle chainsaw sell
Topic 3: insurance life years stupid juggle loyal fuck cancel email sell
Topic 4: insurance life loyal cancel stupid years sell chainsaw dear fuck
Topic 5: life insurance years stupid cancel times juggle loyal email dear
```

Here I are facing the problem of topics not being distinct and coherent again because of less amount of data available. But from whatever data I have, one thing I can say confidently that customers are expressing their negative views about their life insurance product.

Met Life

Word Cloud

As I can see from the world cloud that people are using lots of abusive words. There is also a mention of words like policy, claim, insurance and website. So, it seems people are not happy about these products and services.

Topics

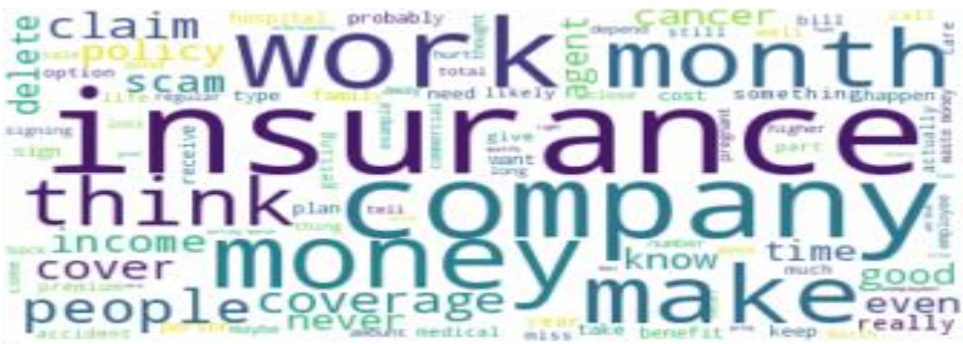
```
Topic 1: cancel email tweet fraud address never cancel_policy team guess left
Topic 2: customer kind anyone enough reply total seriously chat company live
Topic 3: right today claim people service days work provide insurance piss
Topic 4: help answer issue literally years fucking life taylor hold bill
Topic 5: twitter company talk make bitch first representative manager nothing end
Topic 6: play name home instead bring hour friend rep care change
Topic 7: game miss week website sorry thing anything jet second adjuster
Topic 8: night site worst stadium better getting jersey sure experience cause
Topic 9: policy call fuck someone customer_service shit fail terrible damn hope
Topic 10: going support time business month death everyone refund full stay
```

Topic 1 is about some email and fraud. Topics 2, 3 and 9 are about customer service. Topic 5 is about some customer representative or manager. And, topic 10 is about some refund.

Reddit

Just like Twitter, I also did topic modeling of negative comments regarding Aflac on Reddit. Reddit comments are more elaborate and have better lexical diversity than tweets but sometimes in these comments, people tend to diverge from the main topic which can create a problem in text mining and topic modeling.

Word Cloud



Topics

```
Topic 1: employee make business commission market care full next sell glassdoor
Topic 2: right situation definitely good kind break sound happen change move
Topic 3: anything high comment agent paycheck money duck department call buy
Topic 4: insurance company work claim offer little many illegal much get_pay
Topic 5: chart buying_pressure edit back fairly stocks help piss pullback relatively
Topic 6: hospital month something pretty pay thing happen friend great benefit
Topic 7: thought bottom base_commission feel delete kid place reason working thinking
Topic 8: money week time back plan actually surgery accidental price injure
Topic 9: dental policy short_term cancer disability income similar insurance sign save
Topic 10: work look case want time injury policy interview healthcare include
```

It seems that topics 1 and 7 are related to the commission of Aflac employees.Topic 3 suggests about paycheck or payroll deduction and topic 5 tells us about Aflac stocks. Topic 9 is related to dental, cancer and short-term disability insurance.

Glassdoor

Summary

Word Cloud

Here word cloud suggests some good things about Aflac like good company, great company, great place etc.

Topics

```
Topic 1: worker days professional change well_know nothing sales_school solid room_growth much_little
Topic 2: flexible_hours love best client sell wonderful friendly industry appointment quickly
Topic 3: company support employee business benefit insurance pay bonus team offer
Topic 4: sales provide freedom office learn agent everyone selling incentive manager
Topic 5: schedule commission experience flexible income unlimited_income difference large organization given
Topic 6: years reward year trip little salary awesome effort performance still
Topic 7: time working much compensation business high earn amaze life others
Topic 8: hard place know successful always plan getting residual_income free flexibility_schedule
Topic 9: people help need excellent nice flexibility culture think ability decent
Topic 10: great good product training hours opportunity money potential environment management
```

Topic 2 is about discipline in the company, topic 3 is about customer service, topic 4 is about office support, topic 5 mentions of internship and high turnover, topic 7 mentions of cold calling, topic 8 is about insurance agents.

Pros

Word Cloud

From the above word cloud, I can observe that positive words like ‘good’, ‘great’ and ‘friendly ’are dominant. Also, employees are mentioning of training, team management, support, flexibility and commission in the pros section of Glassdoor reviews.

Topics

```
Topic 1: worker days professional change well_know nothing sales_school solid room_growth much_little
Topic 2: flexible_hours love best client sell wonderful friendly industry appointment quickly
Topic 3: company support employee business benefit insurance pay bonus team offer
Topic 4: sales provide freedom office learn agent everyone selling incentive manager
Topic 5: schedule commission experience flexible income unlimited_income difference large organization given
Topic 6: years reward year trip little salary awesome effort performance still
Topic 7: time working much compensation business high earn amaze life others
Topic 8: hard place know successful always plan getting residual_income free flexibility_schedule
Topic 9: people help need excellent nice flexibility culture think ability decent
Topic 10: great good product training hours opportunity money potential environment management
```

Topics 2,5,8 and 9 are about flexible hours, flexible income and friendly environment. Topics 3 and 6 are about pay bonus, salary rewards and offers. Topic 7 is about awesome compensation. Topic 10 is about good training and company's environment.

Cons

Word Cloud

Topics

```
Topic 1: call experience associate product however owner saturate do connection_business straight_commission
Topic 2: people company account manager week hire going place sell office
Topic 3: work business commission money sales month hard benefit start pay
Topic 4: cold_calling everyone market already times felt expense territory claim contact
Topic 5: product marketing challenge often walk definitely like sales_school internship waste_time
Topic 6: employee support expect hours salary commission_base making provide career still
Topic 7: training management time much cold_call area policy care long promise
Topic 8: lack change con else days learn potential process putting woman
Topic 9: know recruit business_owner starting independent_contractor check treat sign understand network
Topic 10: better difficult others rep beginning corporate bunch looking pretty suck
```

Topics 4,5 and 7 are about challenges faced in cold calling and internship. Topic 6 is about concerns related to salary, career and commission base. Topic 8 is related to women employees.

Advice

Word Cloud

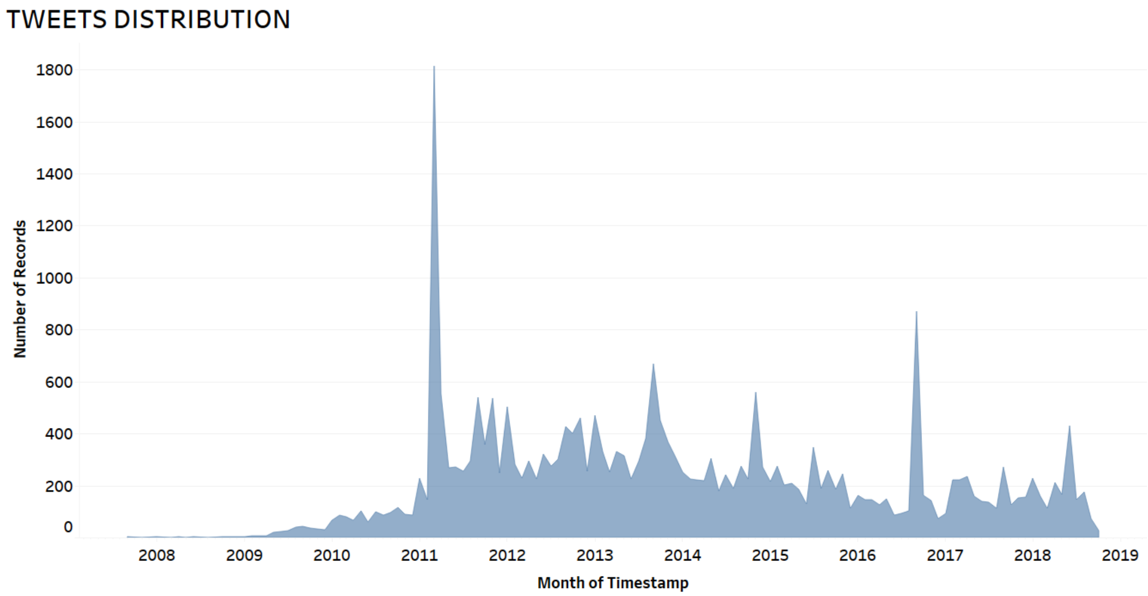
Topics

```
Topic 1: better provide account come years salary email getting hard sales_force
Topic 2: recruit nothing pay team beginning environment awesome coordinator become company
Topic 3: associate help lead position need change business state cold_calling understand
Topic 4: base_salary find create policy sell improve constantly quit long rather
Topic 5: business time rep look something meeting system client experience able
Topic 6: need management manager person good support sales intern level reduce
Topic 7: great company commission sales structure office offer business honest skill
Topic 8: stop maybe little turnover stick career motivate development quality several
Topic 9: people training hire employee money train field process sales please
Topic 10: work keep keep_good base promote month bunch back mentor implement
```

Topics 1 and 4 is about advice related to salary. Topic 3 seems to be related to cold calling. Topic 6 is about advice related to the management. Topic 8 is about employee motivation and turnover. Topic 9 is related to the employee training.

TABLEAU VISUALIZATIONS

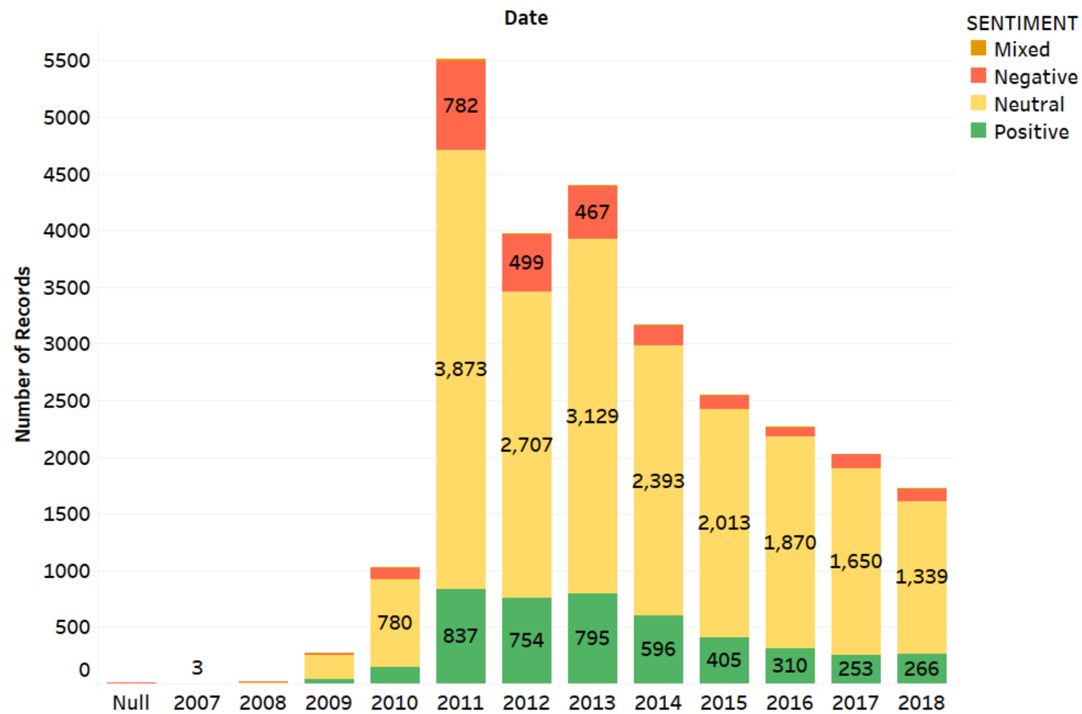
Aflac Twitter Analysis



The plot of sum of Number of Records for Timestamp Month.

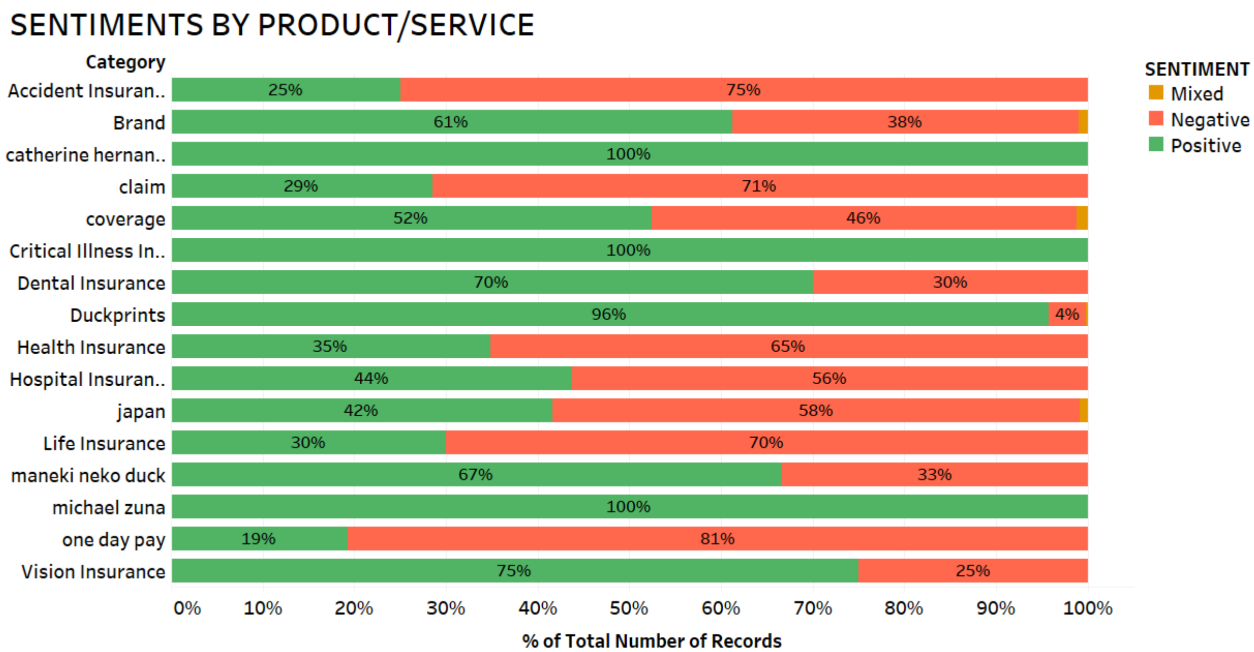
The above plot shows the distribution of Aflac's tweets through the years 2008 to 2018. The plot shows a **peak in the year 2011**. That was when comedian Gilbert Gottfried, voice of the Aflac duck was fired by the insurance company for jokes he tweeted about Japan and the disaster there.

DISTRIBUTION OF TWEETS BY SENTIMENT



Sum of Number of Records for each Date Year. Color shows details about Sentiment.

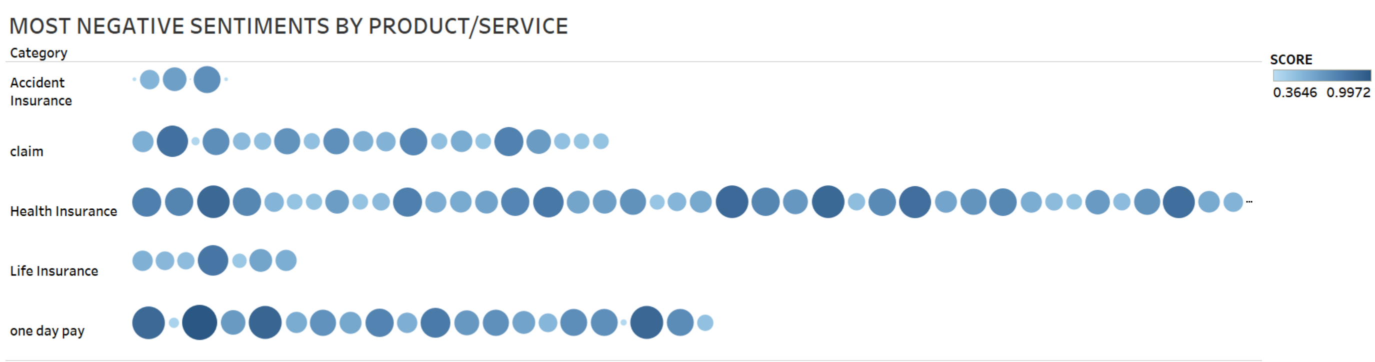
The tweets have been drilled down further and a bar chart has been plotted showing the distribution of tweets by sentiment i.e. Positive, Negative, Neutral and Mixed from the year 2008 to 2018. The number of tweets has been gradually decreasing from 2013. All the years are dominated by neutral tweets compared to Negative and Positive. The number of positive and negative tweets in the year 2011 is similar. Some of the negative tweets in the years 2012 and 2013 are about the Aflac Duck commercial which was released.



% of Total Number of Records for each Category. Color shows details about Sentiment. The marks are labeled by % of Total Number of Records. The data is filtered on Timestamp Year, which keeps 13 of 13 members. The view is filtered on Sentiment, which keeps Mixed, Negative and Positive.

The above stacked bar chart shows the distribution of Positive, Negative and Mixed tweets of the different products and services offered by Aflac. The Neutral sentiments are filtered out since many of them seemed to be not relevant for this analysis. The overall sentiment is majorly negative (considered above 65%) for the following products ***Accident Insurance***, ***Claim***, ***Health Insurance***, ***Life Insurance*** and ***One Pay Day***.

The tweets have been drilled down further to study the negative reviews about the above products.



Score (color) and Score (copy) (size) broken down by Category. The data is filtered on Sentiment, Year of Timestamp and Score. The Sentiment filter keeps Negative. The Year of Timestamp filter keeps 13 of 13 members. The Score filter ranges from 0.276059568 to 0.999721825 and keeps Null values. The view is filtered on Category, which keeps Accident Insurance, claim, Health Insurance, Life Insurance and one day pay.

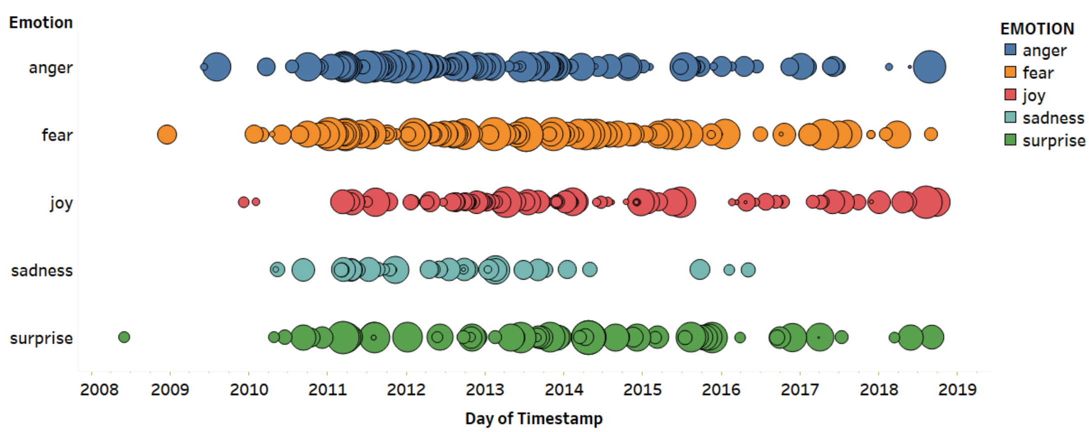
If I hover over the circle, the tooltip gives the details about the tweets, each circle representing a tweet. The smaller circle represents tweet with the least sentiment score and the bigger circle representing the highest sentiment score.

SENTIMENT ANALYSIS BY PRODUCT/SERVICE

Day of Timestamp for each Category. Color shows details about Sentiment. Size shows Score. The data is filtered on Score and Year of Timestamp. The Score filter ranges from 0.539 to 0.999721825 and keeps Null values. The Year of Timestamp filter keeps 13 of 13 members. The view is filtered on Sentiment, which keeps Negative and Positive.

The above plot shows the analysis of each negative and positive tweet by all the products and services. Majority of the tweets are for the Brand. And there are equal proportion of positive and negative tweets. The tweets can be filtered by year, sentiment (Positive/Negative/Neutral/Mixed) and the confidence score as well. The above plot shows all the tweets with confidence score from 0.5 to 0.9 with sentiment being positive and negative.

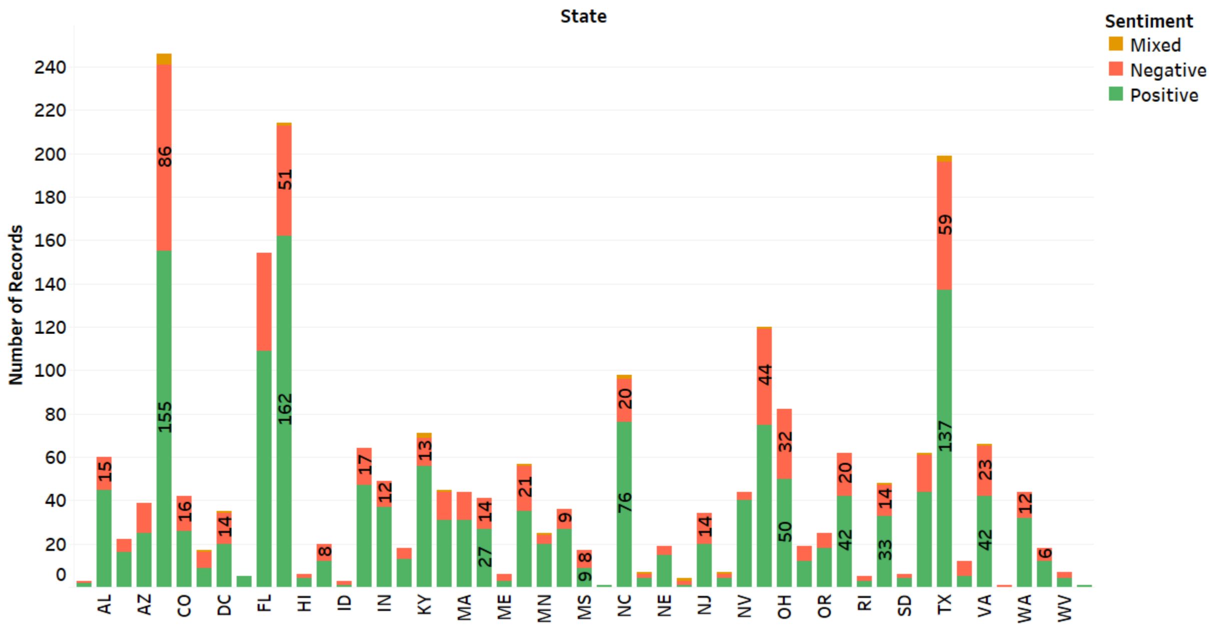
EMOTION ANALYSIS OF AFLAC



Day of Timestamp for each Emotion. Color shows details about Emotion. Size shows Emotion score (copy). The data is filtered on Category and Emotion score (copy). The Category filter keeps 17 of 17 members. The Emotion score (copy) filter ranges from 0.837 to 0.985658884 and keeps Null values.

The above plot shows the different emotions (anger, fear, joy, sadness, surprise) for each tweet through the years 2008 to 2018. The tweets can be filtered by product/service and confidence score of emotion.

DISTRIBUTION BY STATE



Sum of Number of Records for each State. Color shows details about Sentiment. The view is filtered on State and Sentiment. The State filter excludes Null and Unknown. The Sentiment filter keeps Mixed, Negative and Positive.

The above bar chart shows the distribution of positive and negative tweets by State. Most of the tweets are from the following states ***California, Florida, Georgia and Texas***.

Aflac Twitter Analysis with Competitors

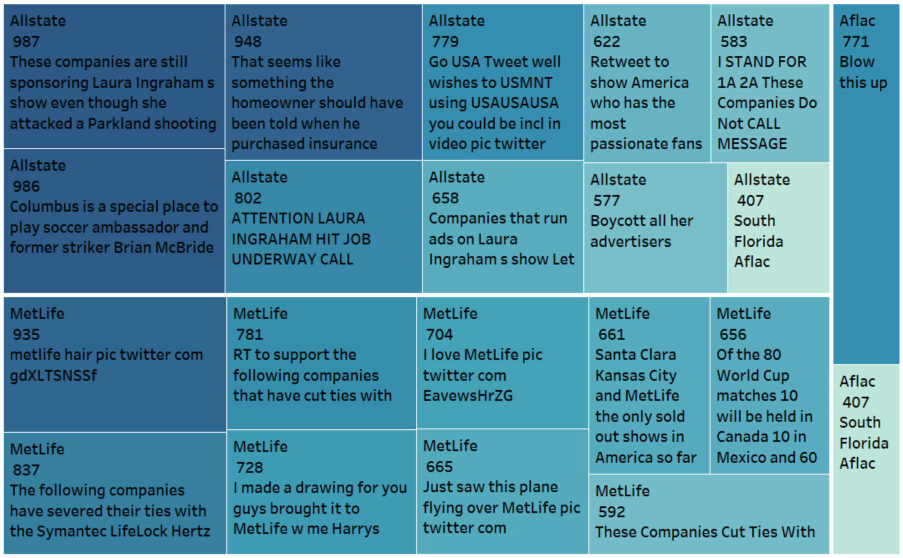
TOP 20 RETWEETS/LIKES/REPLIES

Company, Message and sum of Metric. Color shows sum of Metric. Size shows sum of Metric. The marks are labeled by Company, Message and sum of Metric. The context is filtered on Company, which keeps Aflac. The view is filtered on Message, which has multiple members selected.

The above tree map shows the top 20 most retweeted tweets. The sum of Retweets, sum of Likes and sum of Replies are the metrics. Any of the three metrics can be selected for analysis. The metric Retweets has been selected for the above visualization. The tree map can also be filtered by company. In the above visualization the filter has been applied on Aflac. All the companies can be selected as well.

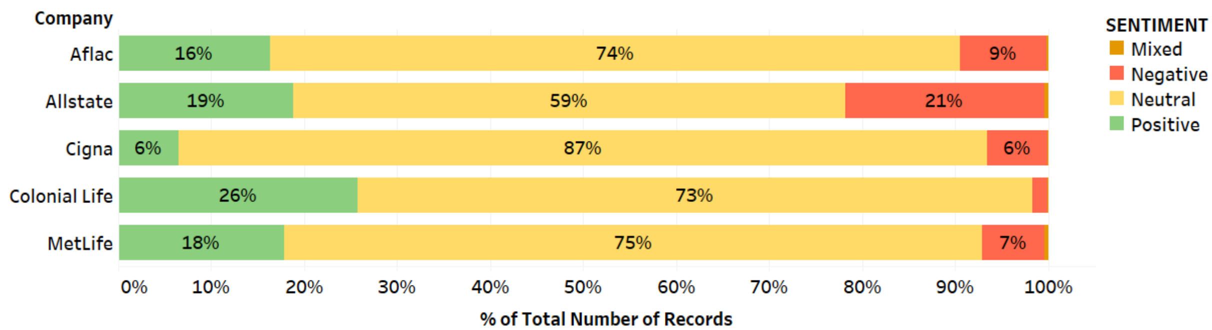
The below tree map shows the top 20 retweets of all the companies (All State, Aflac, MetLife, Cigna and Colonial Life)

TOP 20 RETWEETS/LIKES/REPLIES



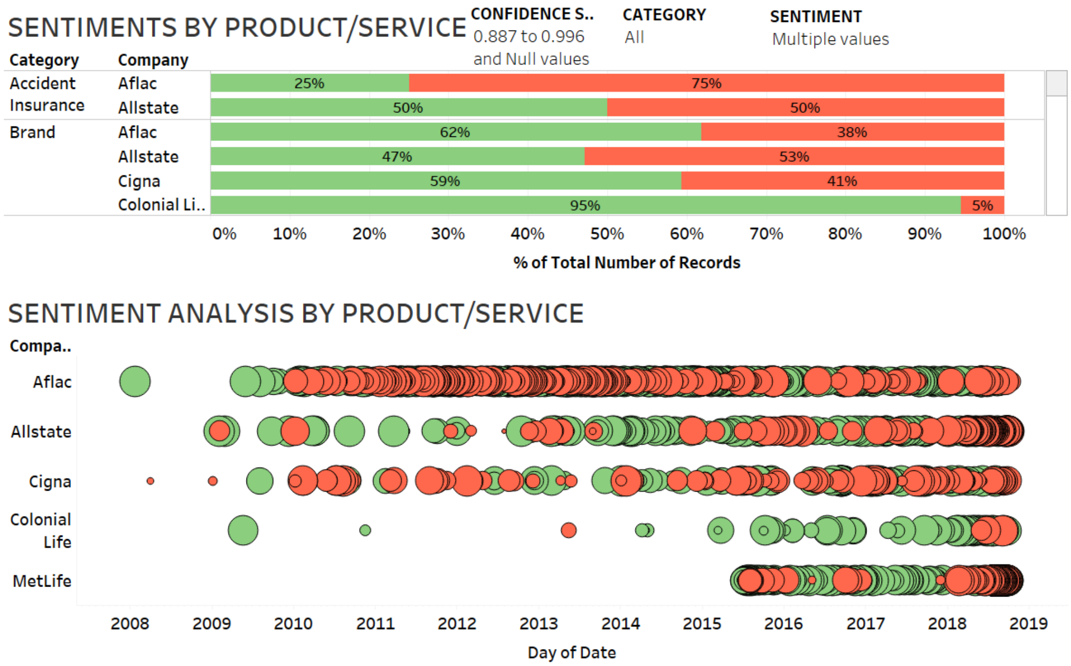
Company, Message and sum of Metric. Color shows sum of Metric. Size shows sum of Metric. The marks are labeled by Company, Message and sum of Metric. The context is filtered on Company, which keeps Aflac, Allstate, Cigna, Colonial Life and MetLife. The view is filtered on Message, which has multiple members selected.

OVERALL SENTIMENTS



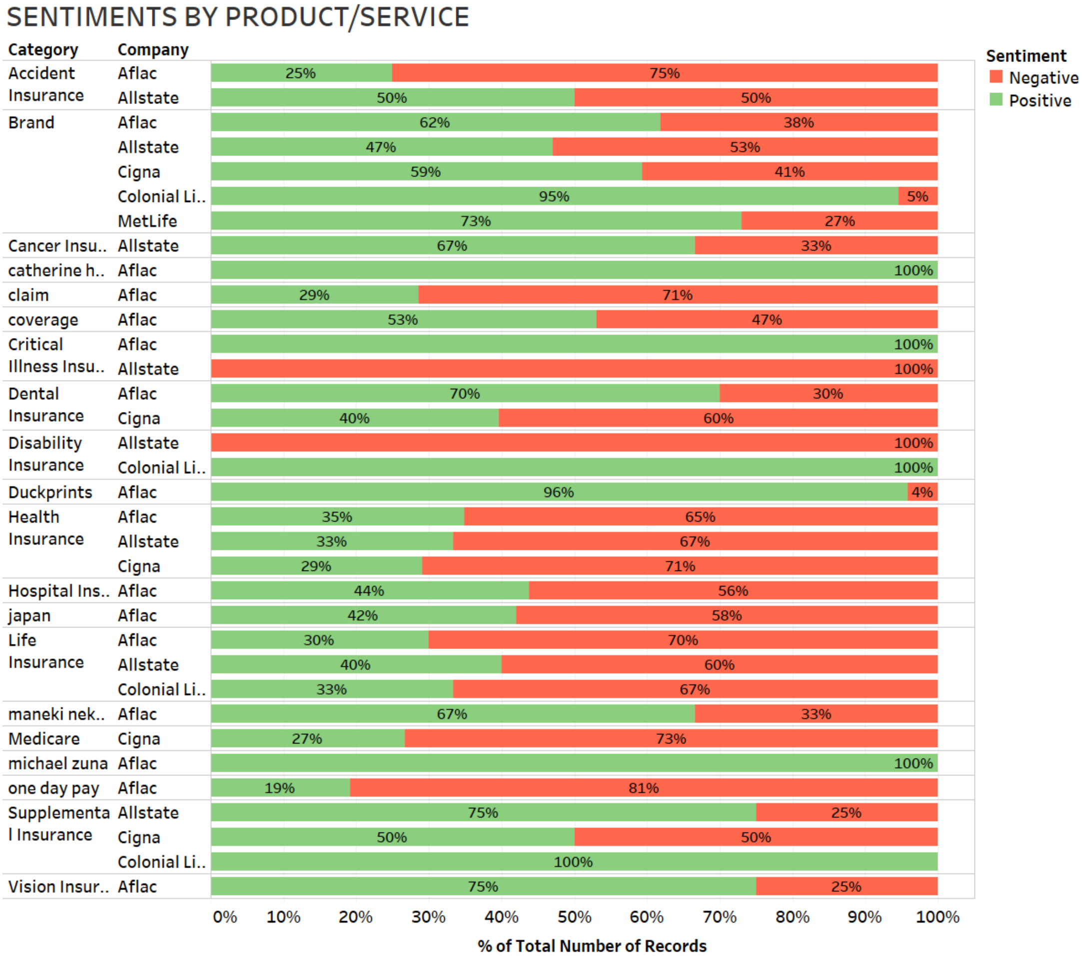
% of Total Number of Records for each Company. Color shows details about Sentiment. The marks are labeled by % of Total Number of Records. The view is filtered on Sentiment, which keeps Mixed, Negative, Neutral and Positive.

The above stacked bar chart shows the distribution of sentiments for all the companies. Aflac stands fourth in terms of positive tweets. But Aflac has the highest number total of tweets compared to other companies.

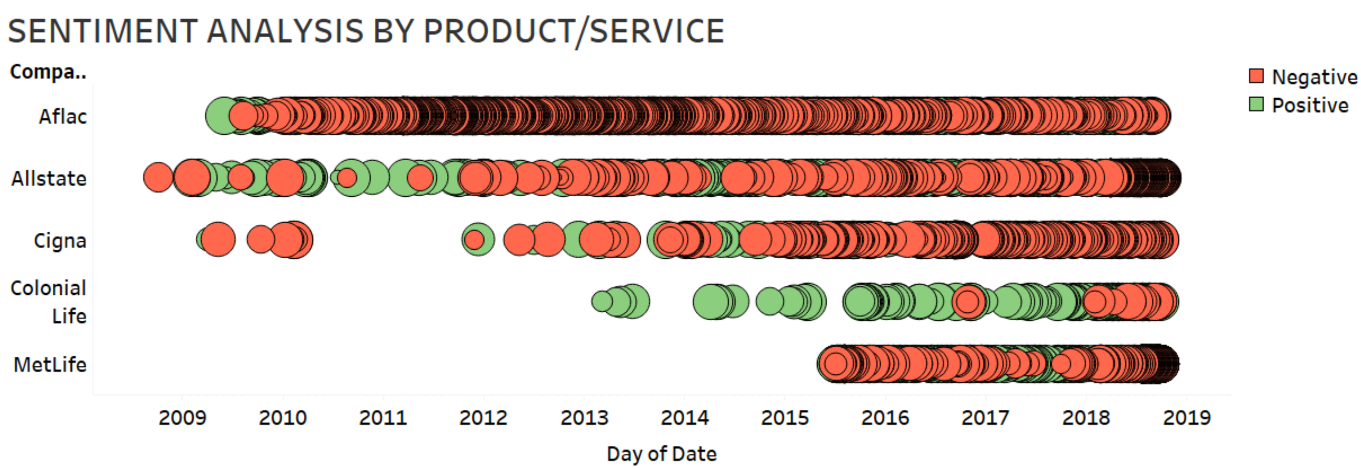


The above dashboard shows the sentiment analysis of all 5 companies by its product and services. I can filter out the category (product/service) as per our requirement and see the percentage distribution of Aflac's sentiments with its competitors and analyze the tweets further in the below visualization.

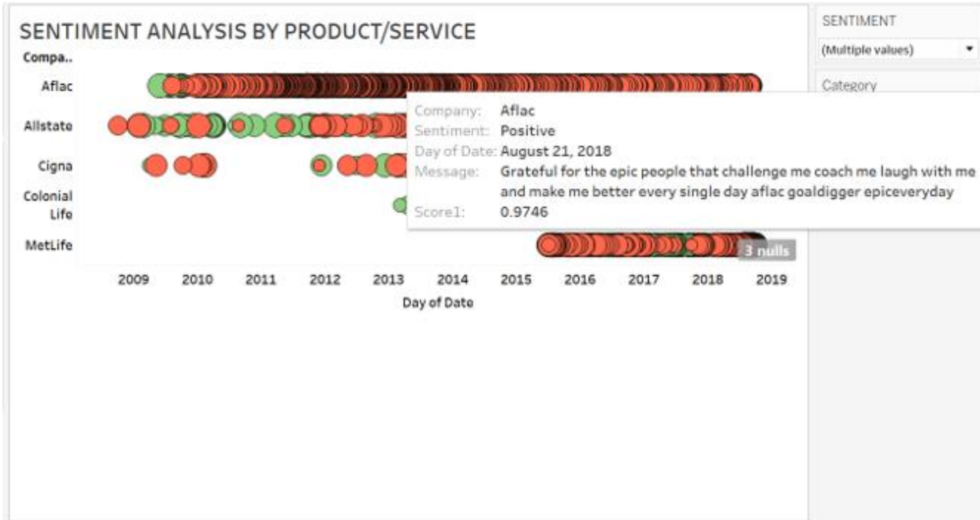
The below two charts show the individual worksheets of the same.



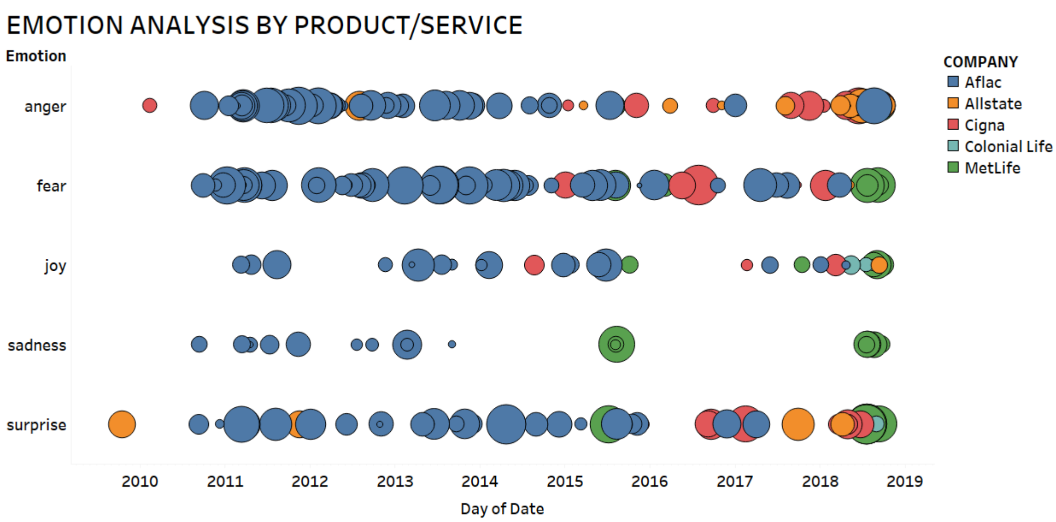
% of Total Number of Records for each Company broken down by Category. Color shows details about Sentiment. The marks are labeled by % of Total Number of Records. The view is filtered on Category and Sentiment. The Category filter keeps 21 of 21 members. The Sentiment filter keeps Negative and Positive.



Day of Date for each Company. Color shows details about Sentiment. Size shows Score1. The data is filtered on Category and Score1. The Category filter keeps Brand. The Score1 filter ranges from 0.282 to 0.996 and keeps Null values. The view is filtered on Sentiment, which keeps Negative and Positive.



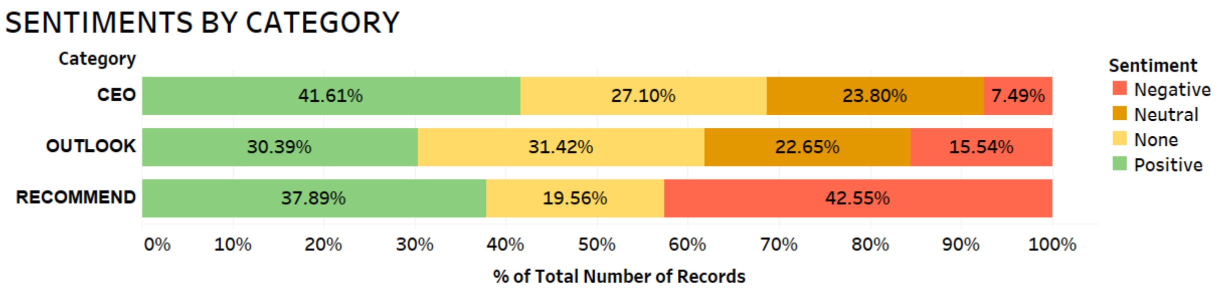
The above picture shows one of the Aflac's positive tweet and its confidence score.



Day of Date for each Emotion. Color shows details about Company. Size shows Emotion score. Details are shown for Category. The context is filtered on Company, which keeps Aflac, Allstate, Cigna, Colonial Life and MetLife. The data is filtered on Emotion score, which ranges from 0.892 to 0.991736054 and keeps Null values. The view is filtered on Category, which keeps Brand.

The visualization above shows the emotion analysis of Aflac and other companies. The category can be filtered as per the requirement. This one is filtered by Brand and the emotion score ranges from 0.8 to 0.9.

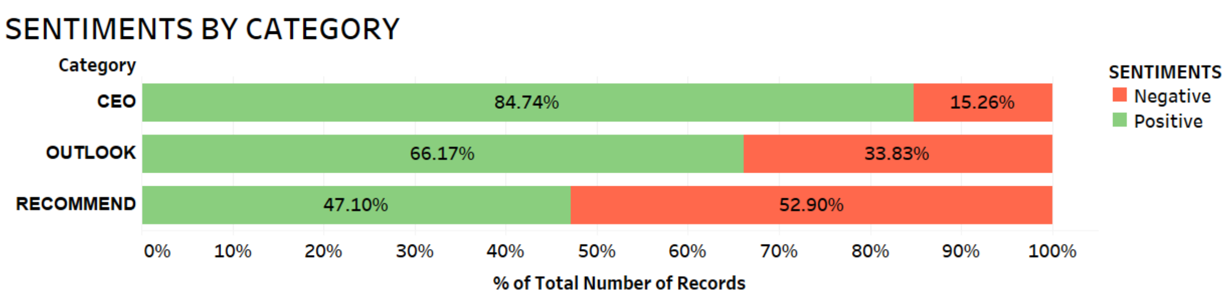
Glassdoor Analysis



% of Total Number of Records for each Category. Color shows details about Sentiment.

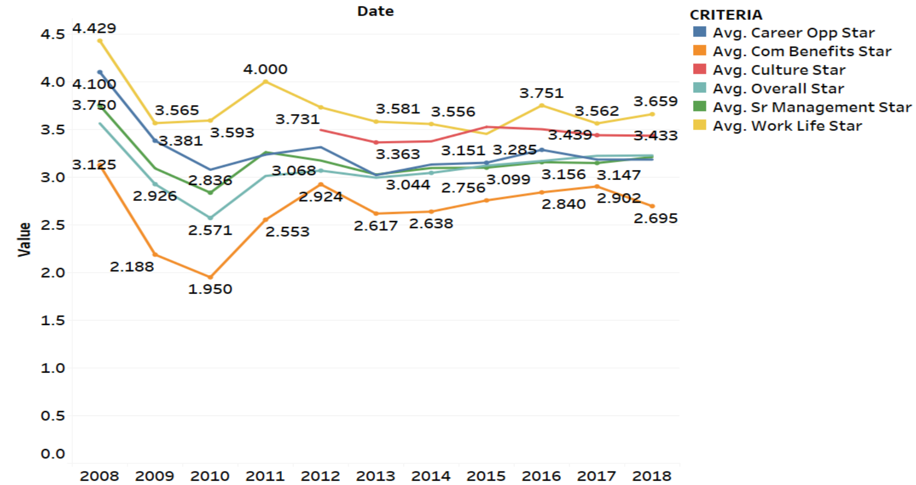
The above stacked bar chart shows the distribution of Aflac sentiments given by the candidates for the three categories CEO, the overall OUTLOOK about Aflac and whether they RECOMMEND Aflac or not. Recommend category has most negative sentiments compared to CEO and OUTLOOK.

Sentiments has been drilled down further to show only the negative and positive sentiments. And it is obvious from the below visualization that RECOMMEND has the highest number of negative sentiments.



% of Total Number of Records for each Category. Color shows details about Pos_Neg_Sentiments. The marks are labeled by % of Total Number of Records. The view is filtered on Pos_Neg_Sentiments, which keeps 2 members.

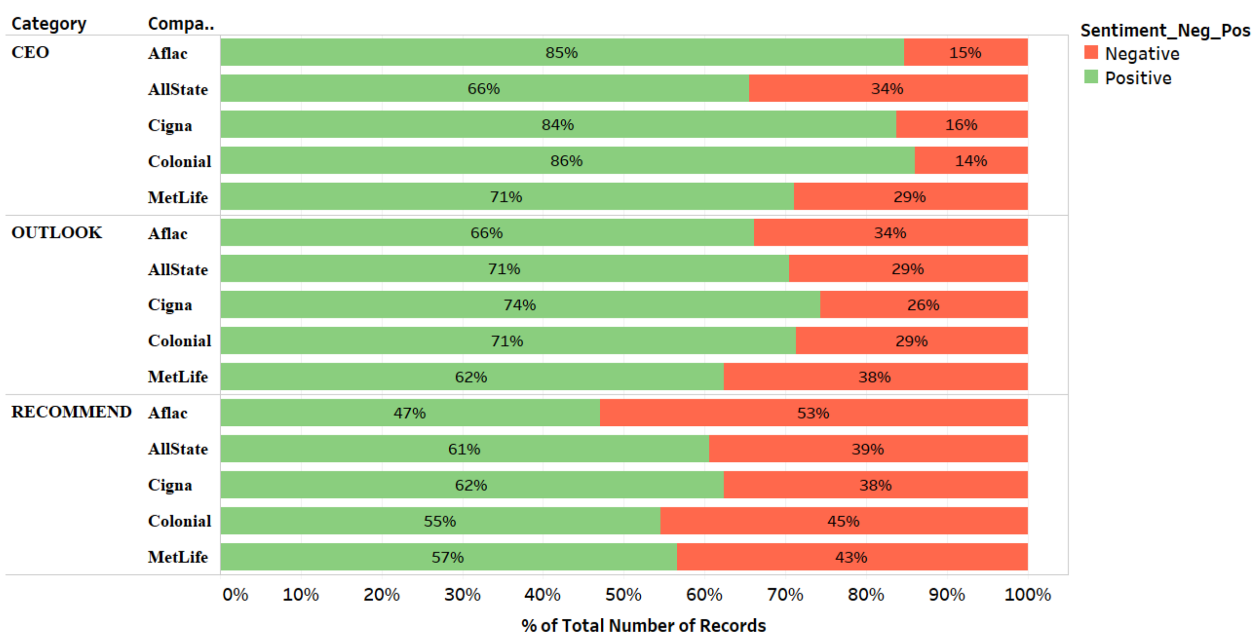
RATING OVER THE YEARS



The trends of Avg. Career Opp Star, Avg. Com Benefits Star, Avg. Culture Star, Avg. Overall Star, Avg. Sr Management Star and Avg. Work Life Star for Date Year. Color shows details about Avg. Career Opp Star, Avg. Com Benefits Star, Avg. Culture Star, Avg. Overall Star, Avg. Sr Management Star and Avg. Work Life Star. The marks are labeled by Avg. Career Opp Star, Avg. Com Benefits Star, Avg. Culture Star, Avg. Overall Star, Avg. Sr Management Star and Avg. Work Life Star. The view is filtered on Date Year, which keeps 11 of 11 members.

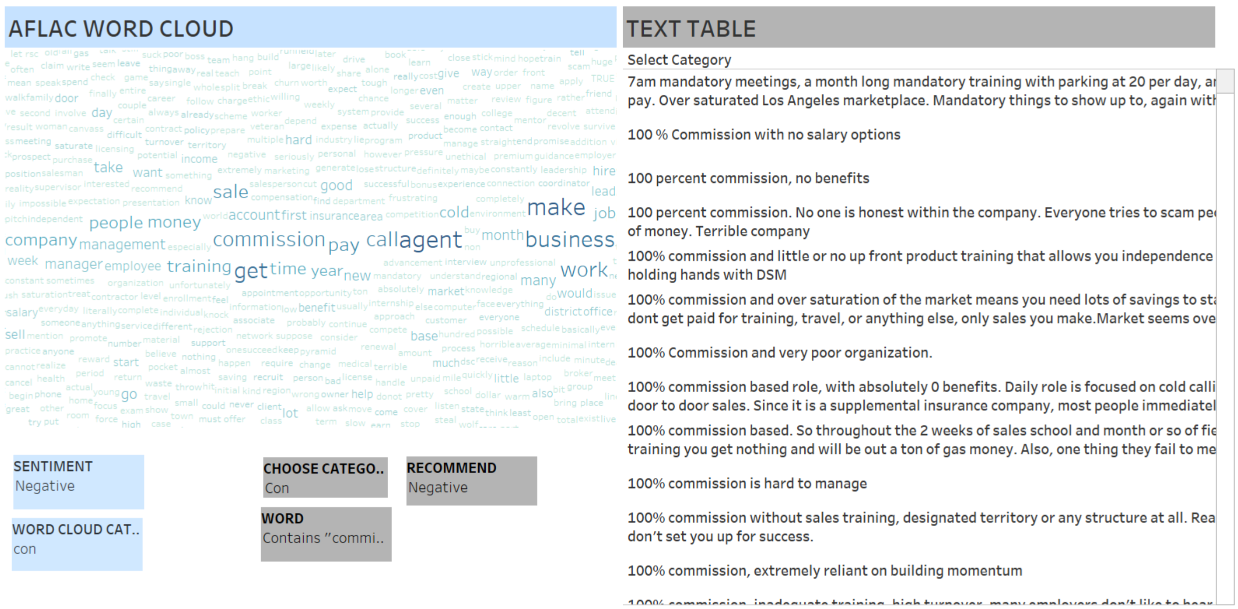
The above line chart shows the average rating of Aflac across different criteria through the years from 2008 to 2018. Average Work Life Star has the highest rating and Company Benefits Star has the least rating. The ratings have not shown much significant changes from 2013.

SENTIMENTS BY CATEGORY



% of Total Number of Records for each Company1 broken down by Category. Color shows details about Sentiment_Neg_Pos. The marks are labeled by % of Total Number of Records. The view is filtered on Sentiment_Neg_Pos, which keeps 2 members.

The above visualization shows the distributions of positive and negative sentiments of Aflac with its competitors. Again, the RECOMMEND has the highest number of negative sentiments and OUTLOOK being the next. In terms of CEO Aflac has the highest number of positive sentiments.
The RECOMMED sentiments has been drilled down for further analysis.



The Aflac Word Cloud shows the most frequent words that are being used by the employees and have been filtered by sentiment and category (Con, Pro, Advice, Summary). Only the RECOMMEND sentiments have been considered for the Word Cloud as RECOMMED has the least number of positive reviews compared to CEO and OUTLOOK. The right side of the dashboard shows the Text Table. This can be filtered by the most frequent word from the word cloud and the specific reviews consisting that word could be analyzed in the Text Table.

Example: Word Cloud has been filtered by the negative sentiments of RECOMMED and Category CON. ***"Commission"*** is one of the most frequent word which can be seen from the world cloud. The same filter for Category (CON), RECOMMEND sentiment (Negative) and reviews that contain the word commission has been applied on the Text Table. The text table thus displays all the results containing the word commission.

RATING COMPARISON

YEAR
All

Compa..	Avg. Career Opp Star	Avg. Com Benefits Star	Avg. Culture Star	Avg. Overall Star	Avg. Sr Management Star	Avg. Work Life Star
Aflac	3.1934	2.7573	3.4576	3.1199	3.1361	3.6288
AllState	3.1524	3.2276	3.4652	3.3602	2.9506	3.5550
Cigna	3.2385	3.3882	3.3734	3.3567	3.0628	3.4974
Colonial	3.3718	2.9771	3.5294	3.2451	3.1754	3.5341
MetLife	2.9735	3.4070	3.1892	3.2441	2.8249	3.4467

PROS

Year of..	Pro	
2018	You're given a lot of responsibilities if you're up for it	AllState
	Your hours depended of the office owner. They worked with your schedule	AllState
	Your pay is based on how hard you work. Great potential. Team environment. The owner helps you to succeed! Recommend to anyone willing to put in the time.	AllState
	Your work for yourself. This is a lot of potential for earnings.	Aflac

CONS

Year of..	Con	
2018	young brains could improvement	Cigna
	Your day can lag making the same calls all day but other then that it's great.	AllState
	Your only as good as your leadership. Network and make friends with other owners. The more allstate friends you make the less likely to get screwed.	AllState
	Zero benefits, you'll be told what to do but you'll be 1099	Aflac
	Zero benefits.This is the most hypocritical company I have ever come across. During training, they tell you how employee benefits are so vital to the success of..	Aflac

The above dashboard compares the average rating of companies across different aspects and shows the Pros and Cons of the companies. The results can also be filtered by year. The above visualization shows all the years from 2008 to 2018.

Aflac has the highest rating for Work Life Star and least rating for Company Benefits Star when compared to its competitors.

CONCLUSION

- Aflac could introduce a twitter handle to address customer grievances like All State Cares. The customers can talk about the issues or any concerns they are facing.

- Aflac has better social media presence as compared to its competitors. Majority of tweets collected were, for Aflac than the other 4 companies. And, many of the tweets were talking about its products and services. Where as majority of MetLife tweets were about their events happening, top celebrity and agents trying to sell the event tickets.

- Aflac’s Company Benefits rating is less when compared to its competitors. The average rating of other aspects of the company is above 3 except the Company Benefits Star which has been less than 3 from 2008.

- Customers are extremely positive about Duckprints and Aflac’s contribution towards cancer. Customers are really showing positive response about Aflac contributing $2 for using #duckprints on social media for fighting childhood cancer.

- There is a rise in the number of tweets being tweeted whenever a new Aflac commercial has been released. Aflac Duck is widely recognized because of these commercials. People show positivity and sarcasm towards Aflac Duck in their tweets.

- On Glassdoor and Reddit, the major concerns were about salary, commissions, payroll deduction and challenges faced in cold-calling.

REFERENCES

1. https://www.meaningcloud.com/blog/text_mining_insurance_industry
2. https://monkeylearn.com/sentiment-analysis/
3. https://www.kdnuggets.com/2016/07/text-mining-101-topic-modeling.html
4. https://aws.amazon.com/comprehend/
5. https://indico.io/
6. https://scikit-learn.org/stable/
7. https://radimrehurek.com/gensim/
8. https://www.analyticsvidhya.com/blog/2016/08/beginners-guide-to-topic-modeling-in-python/
9. https://www.crummy.com/software/BeautifulSoup/
10. https://pandas.pydata.org/