Capstone 2 Project Report

by Anna Kantur

date: 05/04/2020

i. Problem Statement

People come to Twitter to discover what's happening in the world, to share information instantly, and to connect with people and businesses around the globe. With hundreds of millions of users and more than 500 million tweets sent each day[1], there is a steady pulse of conversation where people talk about their lives, their needs, their wants, and what they're doing right now. As an advertiser or a marketer, you only want to reach the right segment of the population - the people who care most about your brand and will be most receptive to your message. Targeting on Twitter lets you reach the right people based on their expressed interests so you will always have the ability to connect with your audience when and where it matters most. In short, marketers are always in search of ways to categorize customers to better target them.

COVID-19 pandemic lockdown orders made outdoor enthusiasts around the world stay at home for over 2 months. People turned to social media to express their emotions and share the news about their new routines.

We analyzed what people who love outdoors talked about during COVID 19 lockdown. Specifically, we identified the main themes of tweets for active and lazy people in the period of March-April 2020 and a year ago for the users of English-speaking Twitter. We also compared the two user cohorts attitudes towards the pandemic.

Our analysis helps to make decisions about customer segmentation and is useful for marketers and advertisers whose products are more suited to users that tend to spend their leisure time indoors or outdoors. Also, the outdoor enthusiasts will be interested in our research as they are trying to decide what to do with their time during the home lockdown.

Ii. Methodology

We collected the users for each cohort by searching Twitter bios on followerwonk.com. Then we used Twitter API to collect their tweets between March 1 and April 15 2019 and 2020.

Most of the machine learning techniques used in our research were from NLP. We identified the main topics of discussion in 2019 and 2020 within each user cohort by using the LDA model from gensim package and the NMF model from sklearn package. Then we used the concepts of semantic similarity and word vectors in spacy to identify COVID-related tweets. We analyzed Twitter sentiment towards Coronavirus for each user cohort by using vaderSentiment. Finally, we used the Multinomial Naive Bayes and Random Forest models from sklearn package to identify the top words that define active users in 2019 and 2020.

The deliverables for this project are the Jupyter notebook with analysis, this report summarizing our approach and findings, and the slides for presentation purposes.

[1] https://marketing.twitter.com/

iii. Data Wrangling

Collecting the data

We first used followerwonk.com to identify active and lazy Twitter users cohorts. We searched Twitter bios by keywords and restricted the search results to users with 10+ tweets. We used the keywords “outdoors”, “hiking” and “camping” for the active cohort, and "homebody", "couch potato", ”hermit”, “lazy boy”, “lazy girl”, “sloth”, “naps”, “lazybones” for the lazy cohort. (We had to use more keywords to identify the lazy cohort, as it was not easy to find these people).

We collected the data from Twitter using Tweepy library for tweets past March 1, 2020 and between March 1, 2019 and April 15, 2019. We wrote the code to take into account the Twitter restrictions for private accounts and 3,200 requests, and Tweepy status code = 401 error for the private users. We also removed retweets and replies. The resulting data frames were of the following size:

- active19 = 4,871 rows
- active20 = 57,424 rows
- lazy19 = 2,257 rows
- lazy20 = 48,111 rows

The increase of tweets in 2020 is the obvious consequence of the pandemic and the stay at home orders when people are turning to social media.

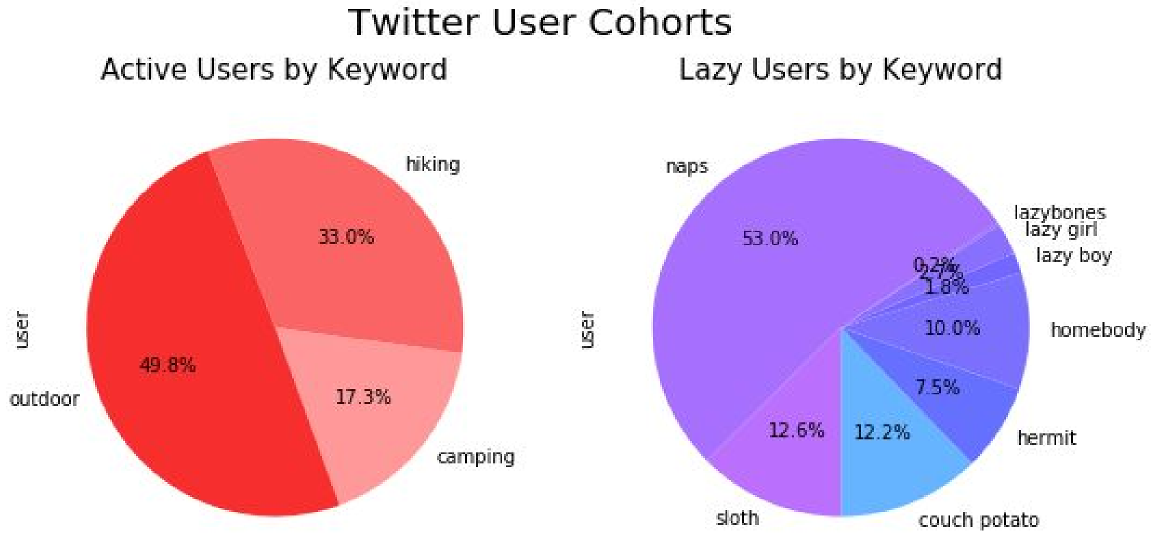
Preprocessing the data

We preprocessed the data by converting the text to lowercase, removing the numbers, removing punctuation, removing white space, removing the stopwords using NLTK package, lemmatizing and converting emojis to text using Emoji package.

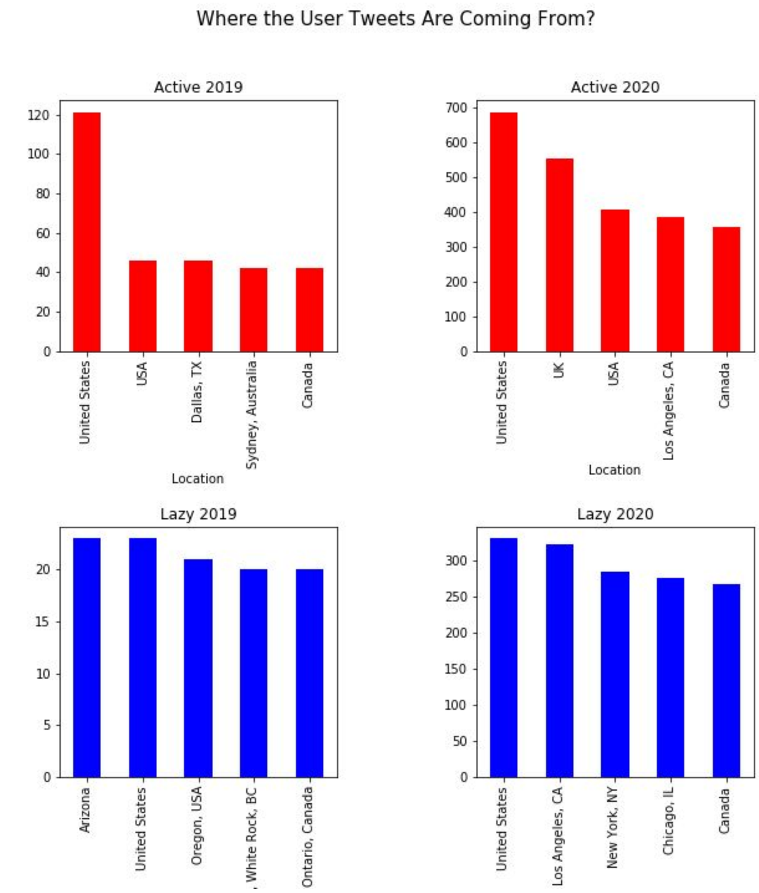
iv. EDA and Data Story

User Cohorts

We collected tweets of 9,995 users in the active cohort and 12,284 users in the lazy cohort:



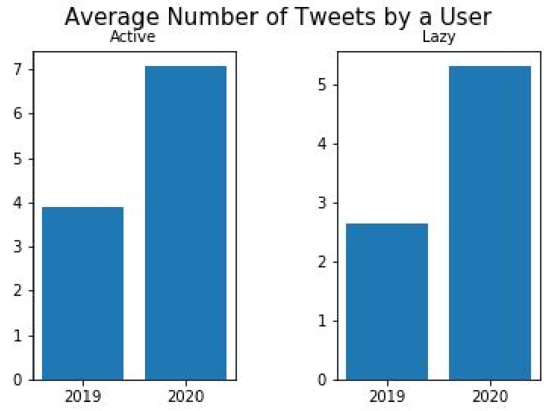
Half of the active users had the word “outdoors” in their Twitter bio, while half of the lazy users had the word “naps” in their Twitter bio.



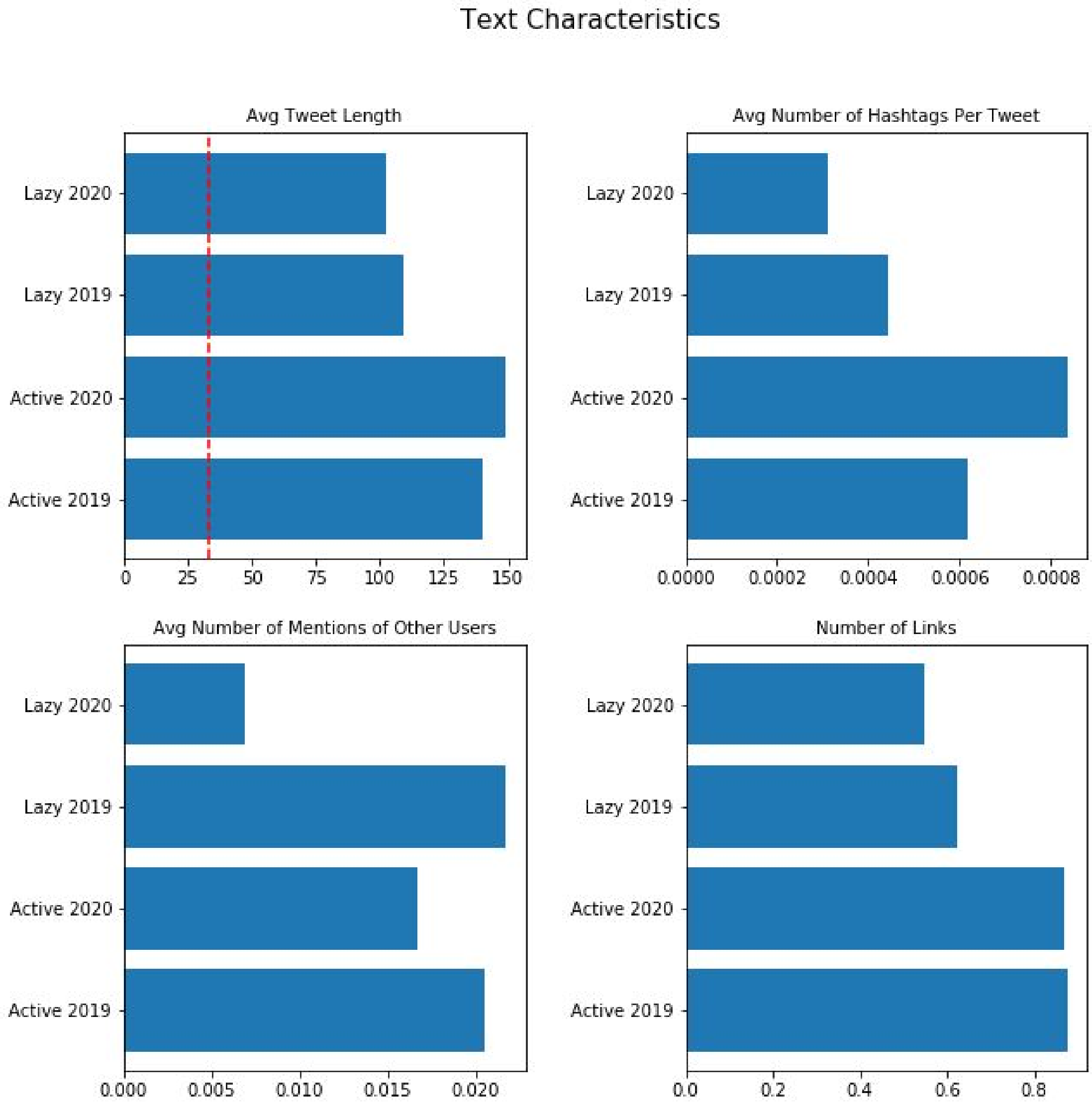
We searched Tweets on the English speaking Twitter. The active user cohort Tweets are from US, Canada, Australia (Sydney) and UK. The lazy user cohort Tweets are from US and Canada. Notably, Los Angeles was a top location in 2020 for both user cohorts.

How Often Do Users Tweet?

The same users tweeted much more in 2020 than in 2019 (excluding replies and retweets) in both user cohorts. The average number of tweets per user increased by 1.5-2 times in 2020. This is possibly a result of COVID stay at home orders and lockdowns.

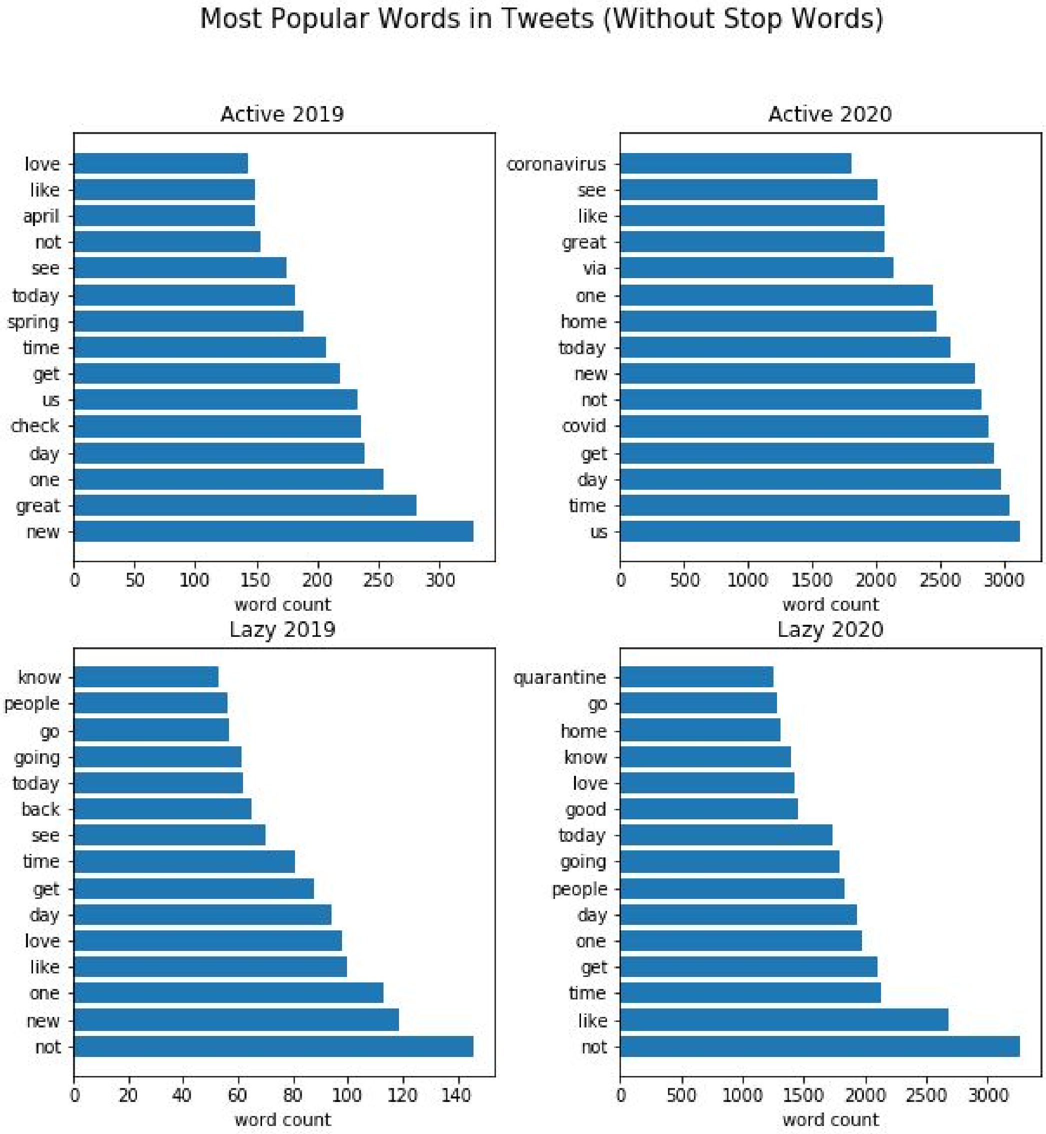


Text Characteristics



User tweets in the active cohort were on average 30 characters longer than those in the lazy cohort. In both 2019 and 2020 the active users used more mentions and links than the lazy users, which increased the length of their tweets. Tweets in both cohorts were over 100 characters, which is quite lengthy considering that the average length of a Tweet in 2018 was 33 characters[2]. The length of tweets didn't increase much in 2020 in any cohort, even though there was still room before the Twitter 280 character limit.

[2] Techcrunch

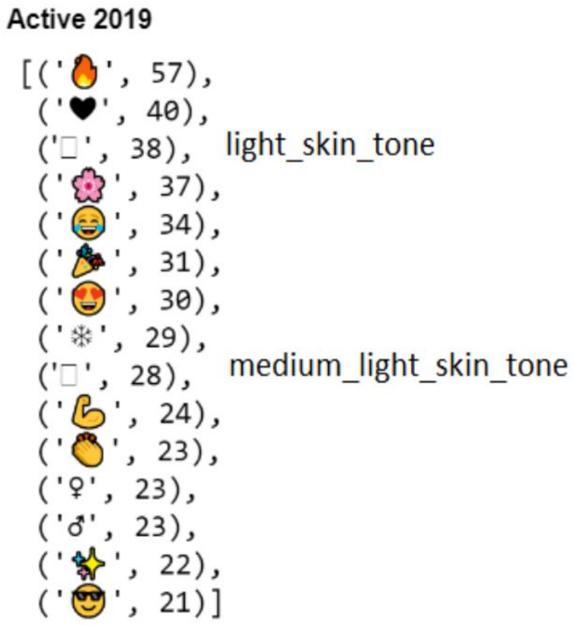


In 2019 users in the active cohort used words like 'spring' and 'april', while in 2020 they talked about 'home', ‘coronavirus’ and 'covid'. The words for the lazy cohort are quite similar in 2019 and 2020, with the main difference being that the top word for 2020 became quarantine.

For the graphical representation of words in Tweets also see word clouds by year by cohort in Appendix A.

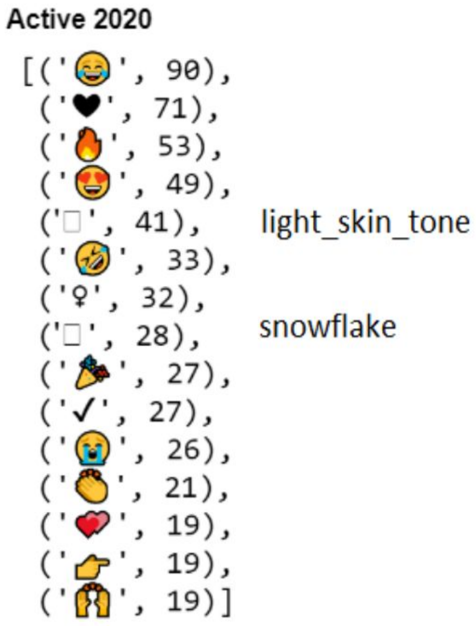
Most Common Emojis

A picture speaks louder than words. That’s why we also analyzed most common emojis in tweets using unicode emoji recognition in emojis package.



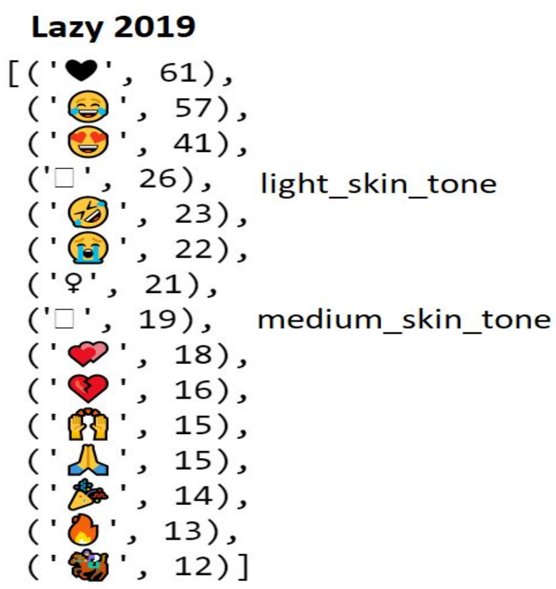
Active 2019

```
[('🔥', 57),
 ('🖤', 40),
 ('□', 38),   light_skin_tone
 ('🌸', 37),
 ('😂', 34),
 ('🎉', 31),
 ('😍', 30),
 ('❄', 29),
 ('□', 28),   medium_light_skin_tone
 ('💪', 24),
 ('👏', 23),
 ('♀', 23),
 ('♂', 23),
 ('✨', 22),
 ('😎', 21)]
```



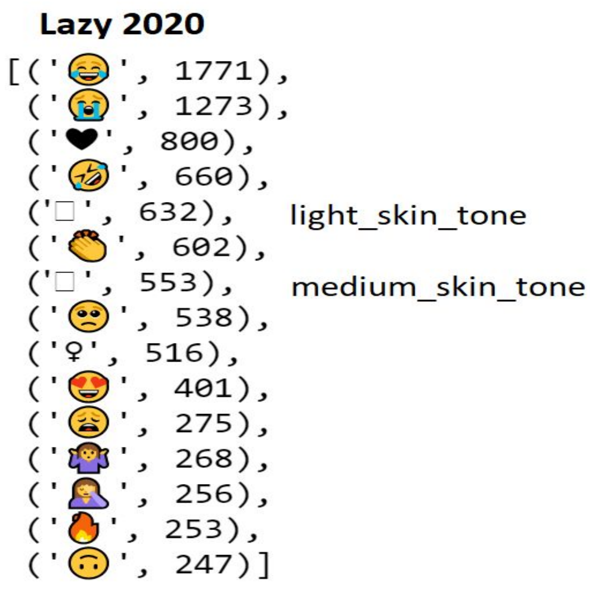
Active 2020

```
[('😂', 90),
 ('🖤', 71),
 ('🔥', 53),
 ('😍', 49),
 ('□', 41),   light_skin_tone
 ('🤣', 33),
 ('♀', 32),
 ('□', 28),   snowflake
 ('🎉', 27),
 ('✓', 27),
 ('😭', 26),
 ('👏', 21),
 ('💕', 19),
 ('👉', 19),
 ('🙌', 19)]
```



Lazy 2019

```
[('🖤', 61),
 ('😂', 57),
 ('😍', 41),
 ('□', 26),   light_skin_tone
 ('🤣', 23),
 ('😭', 22),
 ('♀', 21),
 ('□', 19),   medium_skin_tone
 ('💕', 18),
 ('💔', 16),
 ('🙌', 15),
 ('🙏', 15),
 ('🎉', 14),
 ('🔥', 13),
 ('🏇', 12)]
```



Lazy 2020

```
[('😂', 1771),
 ('😭', 1273),
 ('🖤', 800),
 ('🤣', 660),
 ('□', 632),   light_skin_tone
 ('👏', 602),
 ('□', 553),   medium_skin_tone
 ('🥺', 538),
 ('♀', 516),
 ('😍', 401),
 ('😩', 275),
 ('🤷', 268),
 ('🤦', 256),
 ('🔥', 253),
 ('🙃', 247)]
```

In 2019 active users put emojis with flowers and sparkles, while in 2020 sad crying face and hearts became more popular. Similarly for lazy users, the hearts and flower emojis in 2019 got replaced by crying faces, woman shoulder shrug and women facepalm in 2020, indicating the change of mood.

v. Machine Learning

Topic Modeling

In order to understand what active and lazy users discussed on Twitter in Spring 2019 and 2020, we used topic modeling. Topic modeling is an unsupervised machine learning technique. This means it can infer patterns and cluster similar expressions without needing to define topic tags or train data beforehand. We used this approach to identify main topics of discussion in 2019 and 2020 in both cohorts. Based on the EDA findings, we expect to see Spring season-related topics in active 2019 tweets and covid-related topics in both active and lazy in 2020.

Topic modeling algorithms are statistical methods that analyze the words of documents to discover the themes that pervade a large collection of documents. The basic idea of topic modeling is that a document is a mixture of latent topics and each topic is expressed by a distribution of words.

We used 2 algorithms: Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) to identify top 3 topics in each year in each user cohort. LDA is the most popular topic modeling method in the field of text mining that has been known to be an effective tool for text mining of large datasets. The output of LDA models is in Appendix B. We used LDA from gensim package. Overall, the term frequency of words was almost the same, so topics were not particularly evident. Unlike LDA, NMF relies on linear algebra. It approximates a nonnegative matrix by the product of two low-rank nonnegative matrices. Since it gives semantically meaningful results that are easily interpretable in clustering applications, NMF has been widely used as a clustering method especially for document data, and as a topic modeling method. We used NMF from sklearn package. The output of NMF is in Appendix C.

Top 3 Topics Identified by the Models

User Cohort	NMF	LDA
active users in 2019	1) Links to Facebook videos 2) Great time hiking in the Spring 3) Advertising of camping in India by chhatrasagar Twitter user	1) Happy time hiking 2) Social media (Facebook, videos) about the weekend fun outdoors 3) Social media (Facebook, videos) about camping and the Spring season
active users in 2020	1) Coronavirus (stay at home, wellness, #getbetter, etc.) 2) Thanking people (emojis of clapping_hands,two_hearts, #thankopolis, etc.) 3) Activities during Coronavirus (newsreading, hiking, camping, #dailydoseofgreenspace)	1) Coronavirus staying at home and trying to stay positive (us, best, great) 2) Coronavirus keeping in touch via social media (family, video, social) 3) Missing outdoors during Coronavirus (#stayindoorsdreamoutdoors)

lazy users in 2019	1) Hating college (hate, college, face_with_rolling_eyes emoji, etc.) 2) Good day feeling great (best, day, happy, great, etc.) 3) Twitter contests (bestbuyatplaylistlive, gift card, meetup, etc.)	1) Not enjoying college 2) General Twitter chatter 3) Best Buy Playlist Live event
lazy users in 2020	1) Lockdown and staying at home (#homeorwork, #reclaimingmyhealth, #newyorklockdown, etc.) 2) Playing online games (#achndesign, animalcrossing, #gameitin, etc.) 3) Other activities (goodreads, #homeorwork, #daylong, etc.)	1) Covid support and activities (home, red_heart, party_pooper) 2) Quarantine activities (playing Animal Crossing and Nintendo Switch) 3) Quarantine emotions (face_with_tears_of_joy, miss)

Overall, the two models identified very similar top 3 topics. The evolution of topics also indicates the mood change between 2019 and 2020.

COVID Topic Classification and Twitter Sentiment Analysis

The questions we answered in this section:

- How many more people tweeted about diseases and being sick in 2020 vs 2019? Any difference between active and lazy user cohorts?
- What is the overall change in sentiment on disease-related tweets between 2019 vs 2020 and lazy vs active?

To answer these questions, we used a topic classification technique on the known topic of COVID pandemic in 2020. We tagged the related data in order to train a topic classifier by using the concept of semantic similarity and SpaCy NLP package. We then analyzed the Twitter sentiment on disease-related tweets for 2019 vs 2020 and lazy vs active using vaderSentiment.

Semantic Similarity and Word Vectors in Spacy

We picked all the tweets that are semantically similar to Coronavirus-related words in 2020 and to the disease-related words in 2019 using the algorithm word2vec and spaCy's prebuilt word embeddings.

We created a word vector for Coronavirus in 2020 that included the following words:

'covid coronavirus virus pandemic disease lockdown quarantine cough doctor nurse hospital'

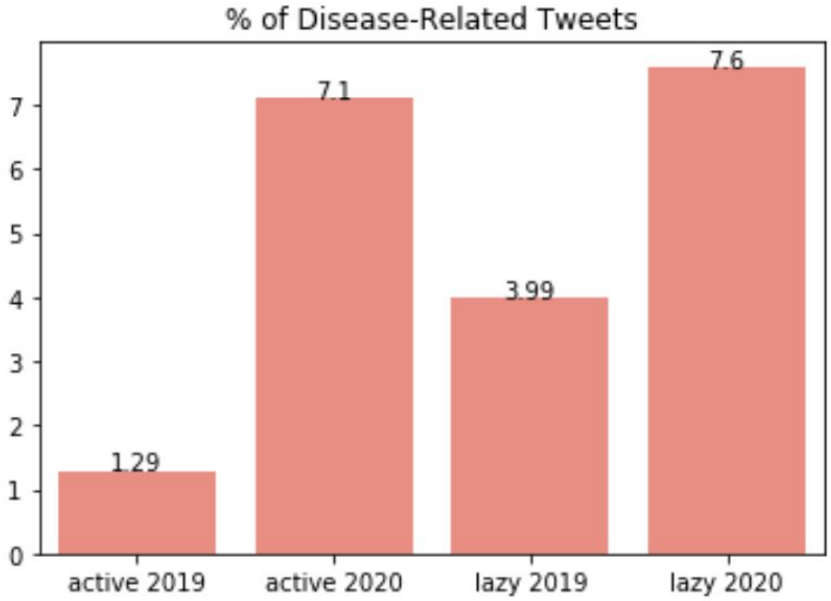
We also created a word vector for diseases in 2019 that included the following words:

'disease virus cough doctor nurse hospital'

We intentionally set the threshold for similarity as high as 55% to ensure that only the relevant tweets got selected. We wanted to avoid tweets like "This day was sick!" (i.e. great day) Also, after some consideration we decided to exclude the words "fever" and "emergency" from the disease-related words, because we wanted to avoid the mix up with the expression "cabin fever" and exclude tweets about natural disasters like fires and hurricanes (even though one can argue that these events also change the predictable course of people’s life similar to COVID or disease).

We choose tweets related to the topic by comparing them to the word vectors described above. We considered a tweet related to the topic if it had semantic similarity of 0.55 or higher.

Between 2019 and 2020 there was a 500% YoY increase in disease-related tweets for active users and a 100% increase for lazy users.



Covid Twitter Sentiment

We used an out-of-the-box Twitter Sentiment model called SentimentIntensityAnalyzer from vaderSentiment package. The model produces negative, neutral, positive and compound scores. We will follow the thresholds for sentiment determined on the vaderSentiment website:

"The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is the most useful metric if you want a single unidimensional measure of sentiment for a given sentence. Calling it a ‘normalized, weighted composite score’ is accurate.

It is also useful for researchers who would like to set standardized thresholds for classifying sentences as either positive, neutral, or negative. Typical threshold values (used in the literature cited on this page) are:

positive sentiment: compound score >= 0.05
neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
negative sentiment: compound score <= -0.05"

2019 vs 2020 Comparison

2020 to 2019 YoY change in sentiment towards disease was quite small: the positive accuracy increased by 1% from 51.5% to 52.5%. Interestingly, there was slightly more positive sentiment in 2020 during the COVID pandemic.

Active vs Lazy Comparison

Combining both 2019 and 2020 years, users in the active cohort are almost 8% more negative on Twitter about diseases than the users in the lazy cohort.

2019 vs 2020 for Each User Cohort Separately

The mood on Twitter in 2020 worsened vs a year ago by 5.5% for active users, and by 6.5% for lazy users, which means that the events of 2020 COVID pandemic affected lazy and active users equally negatively.

Most Predictive Words for Active User Cohort

We built a model that determines the most predictive words for active users in the Springs of 2019 and 2020.

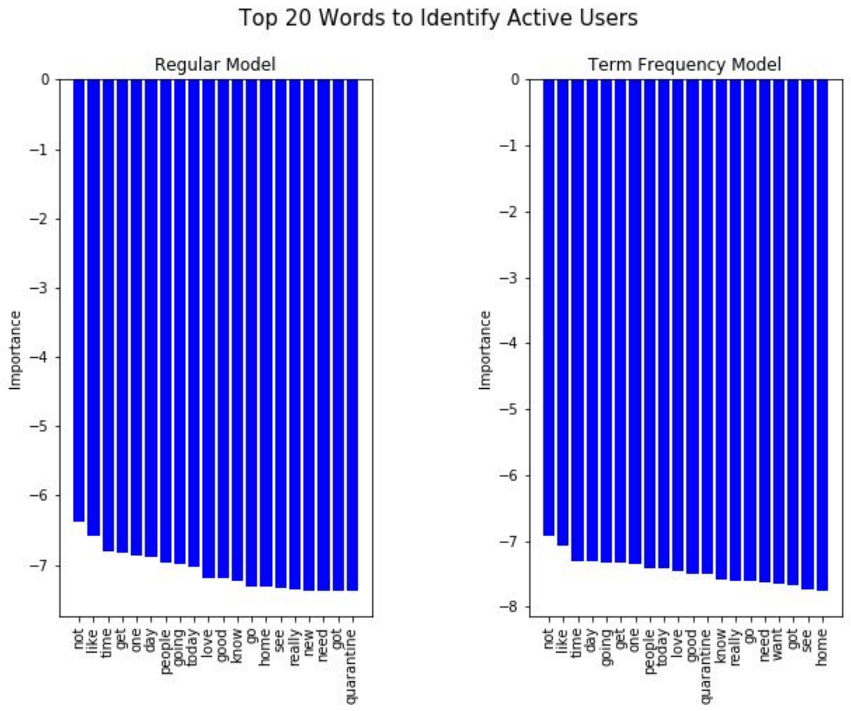
Multinomial Naive Bayes Model

We used a Multinomial Naive Bayes Model because this model was appropriate to predict a category with labelled data and over 100K samples of text data. We used CountVectorizer and TfidfVectorizer from sklearn to build two models: the regular one, and the one that accounts for the term frequency distribution.

After adjusting the hyperparameter alpha through cross-validation on a parameter grid, the best ROC_AUC score for the model was 85%.

We used coef_ attribute of MultinomialNB to get the most important features (e.g. top words for predicting active users). coef_ attribute is a re-parameterization of the Naive Bayes model as a linear classifier model. For binary classification problems this is basically the log of the estimated probability of a feature given the positive class. It means that higher values mean more important features for the positive class.

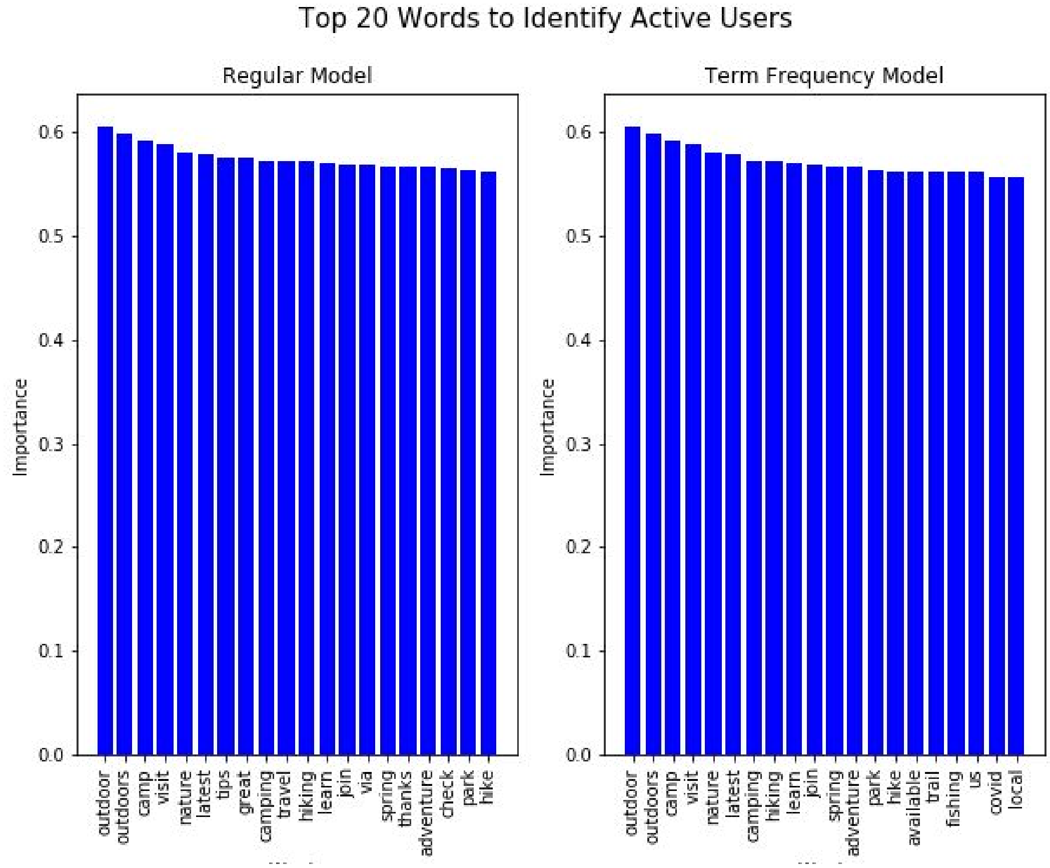
In our case even the most highly rated words (features) had negative values, meaning that they are not highly predictive. The most important words for active users ended up being very generic (e.g. “time”, “day”, “quarantine”, “great”) and not particularly specific to one type of users.



Random Forest Model

We also used a Random Forest Model to double check the results of Naive Bayes. Random Forest is an ensemble model that uses decision trees logic to arrive at the outcome.

After adjusting the hyperparameters for max depth of a tree to 5 and n_estimators (number of trees) to 100, the best ROC_AUC score was 71%, which is 13% less than what we got for the Naive Bayes model.



The Random Forest model shows that covid was one of the most predictive words for active users, indicating that this is a big topic for this user cohort.

vi. Findings

Our analysis showed a few findings about the shift in main topics between 2019 and 2020, the user attitude towards COVID, and top words describing active users in 2019 and 2020.

Overall, there were much more tweets in 2020 in both cohorts, and the top emojis showed confusion and negativity of the COVID uncertain times. (Keep in mind that we analyzed tweets in the period of Mar 1 to April 15, and the COVID pandemic and the lockdowns continued much longer)

While the mood in tweets about diseases stayed as negative in 2020 as it was in 2019, the active users are slightly more negative on Twitter when it comes to diseases. (If you compare active 2019 vs active 2020, the negative change is the same as for lazy 2019 vs lazy 2020, but if you compare all active vs all lazy in both years, active users' mood dropped 8% more than the lazy users.) This was a logical outcome since the active user cohort likes spending time outdoors and now they had to be stuck inside.

From the main topics analysis and defining most predictive words for active users we can infer that the active users are still dreaming about outdoors and planning future trips, but COVID became a big topic for Twitter discussion. At the same time the lazy user cohort continued with their home-based hobbies and became less opportunistic by not participating in online contests like in 2019. A gift card prize stopped being an incentive when people are staying at home and cannot spend the money because the stores are closed.

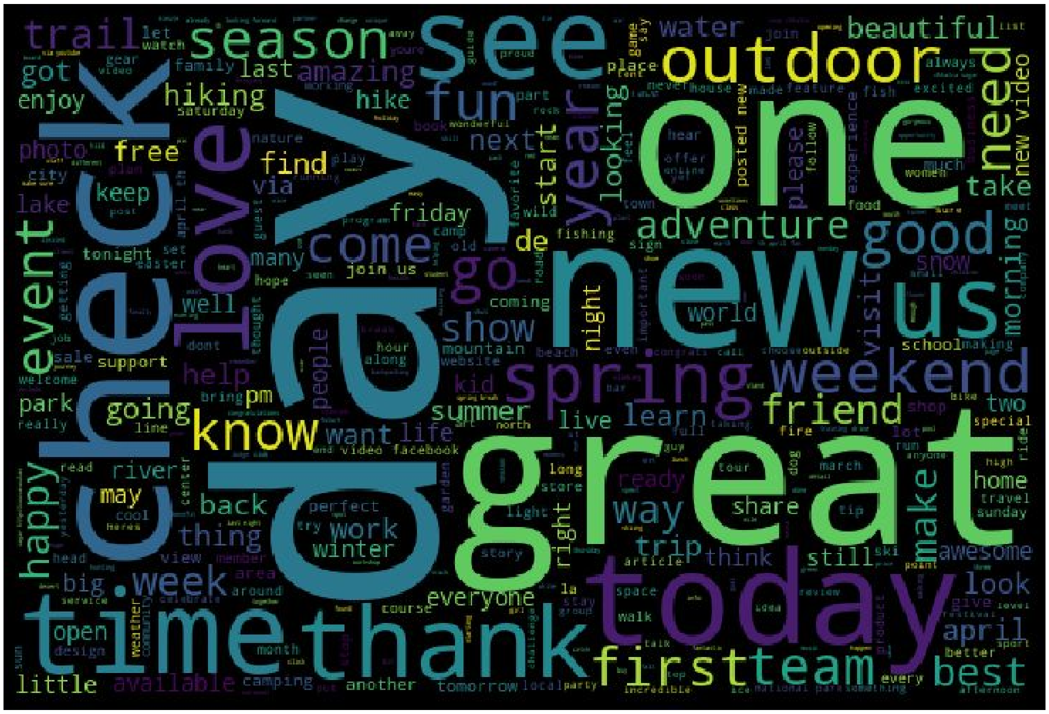
vii. Ideas for Further Research

While our research highlighted important changes in Twitter sentiment, further research can be done on the same topic with a larger user base or a longer tweeting time period. Since the COVID stay at home orders are continuing into June at the time when this report is written, it would be interesting to see what other words will become descriptive of active users in 2020, and how the overall mood changes as more and more time people who love outdoors are asked to stay at home.

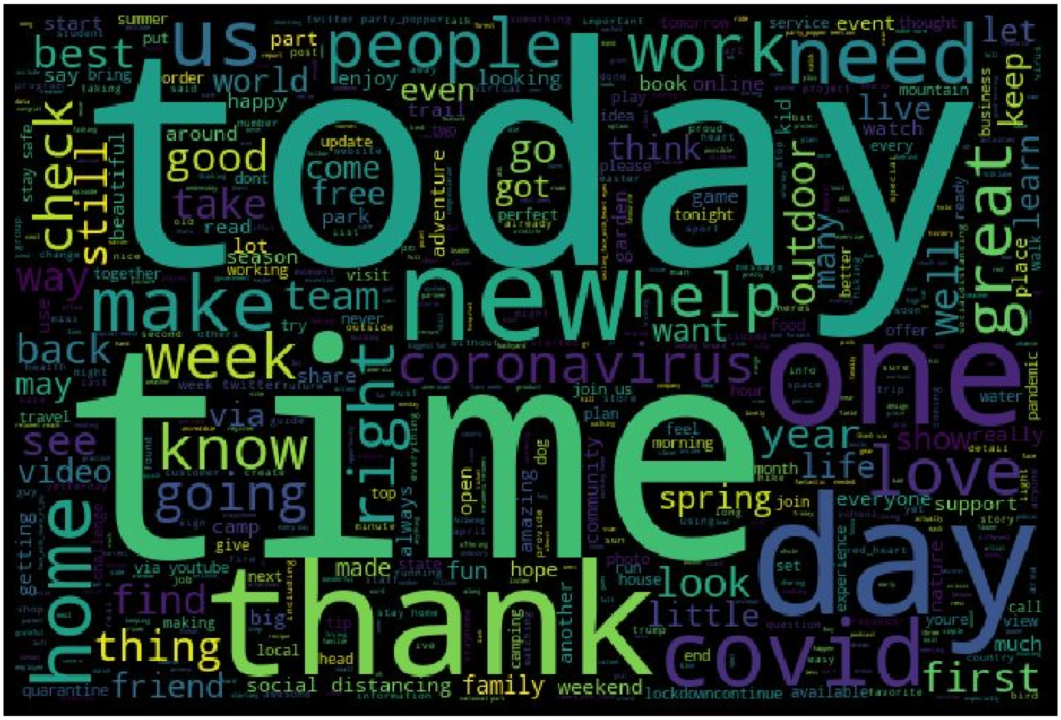
Also, the next steps can be to create an app to predict a Twitter user type by keywords, or an app to predict what % chance is that an active user converts into a lazy user after the extended lockdown at home.

Appendix A: Word Clouds

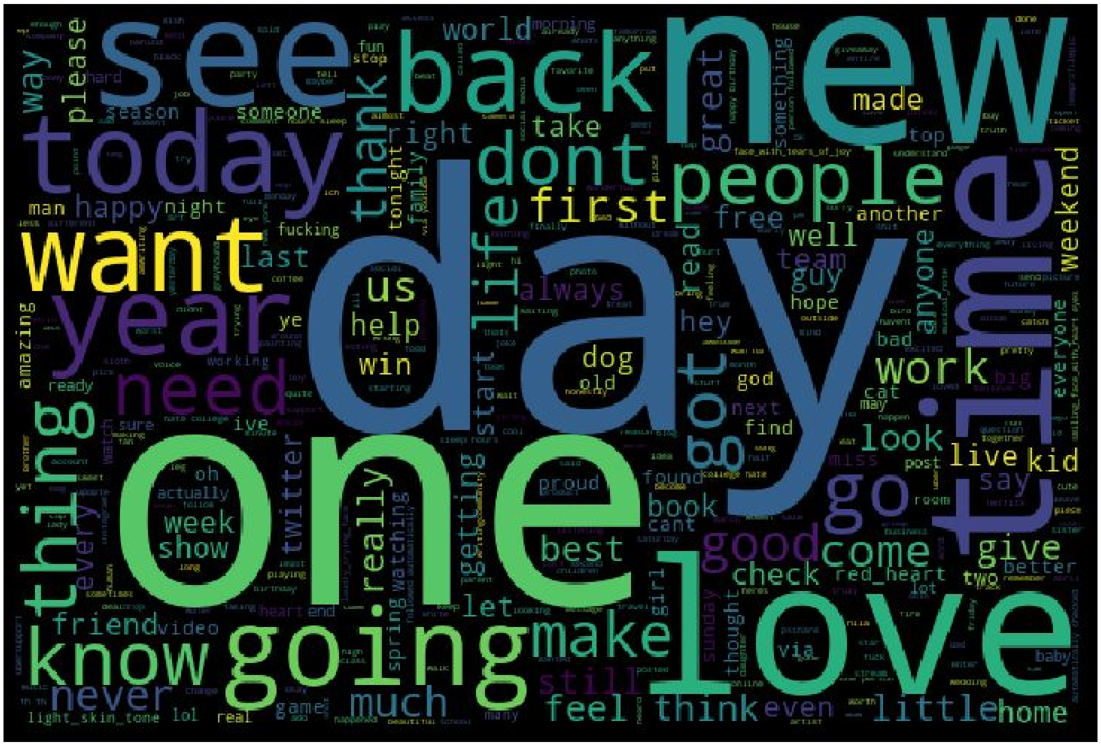
Word Cloud - Active 2019



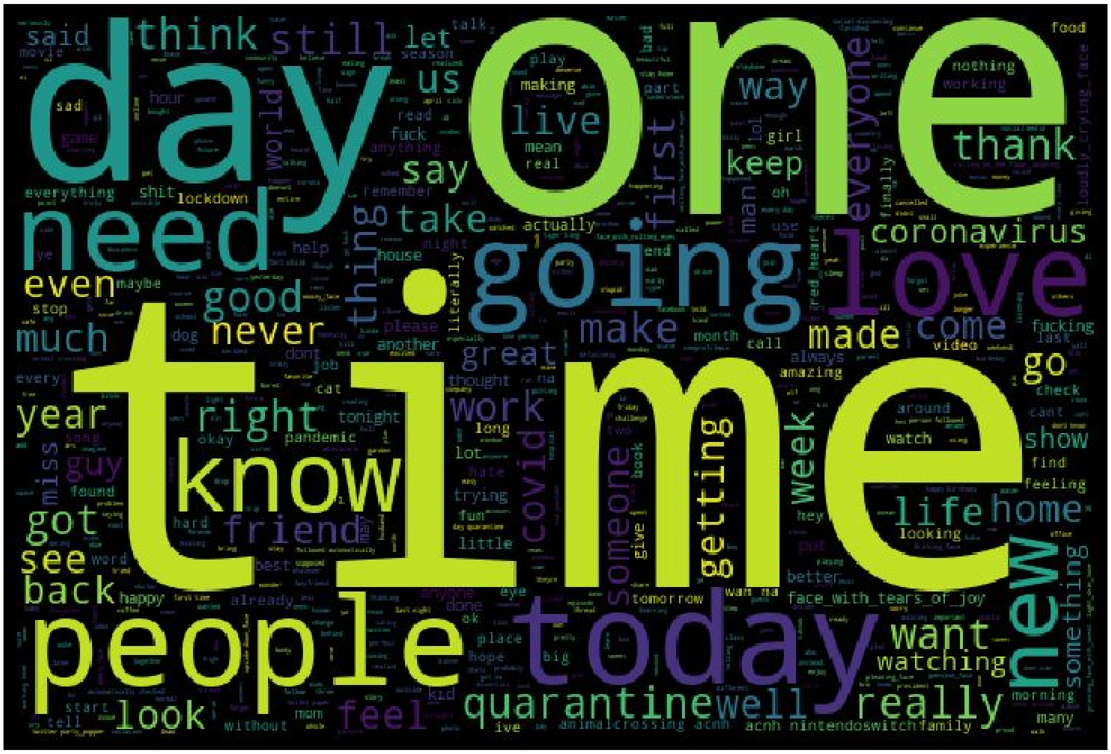
Word Cloud - Active 2020



Word Cloud - Lazy 2019

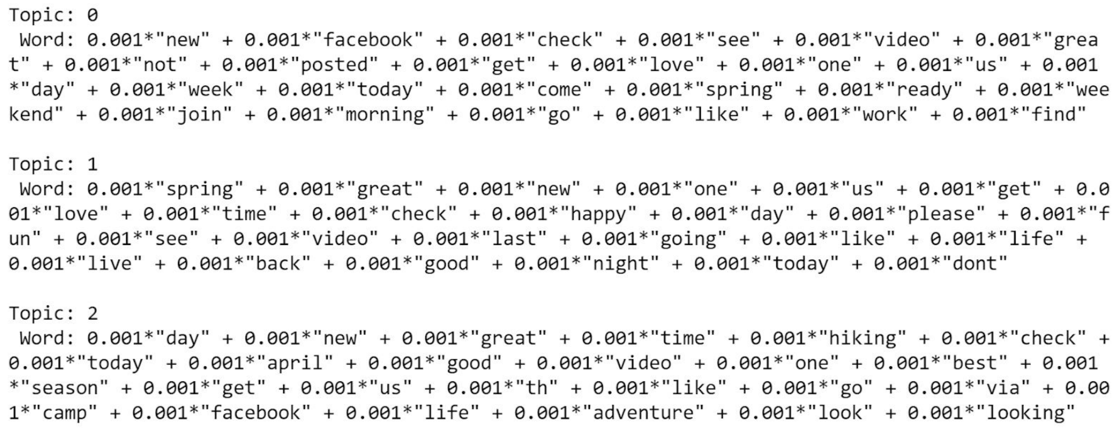


Word Cloud - Lazy 2020



Appendix B: gensim LDA Model Output

2019 Active:

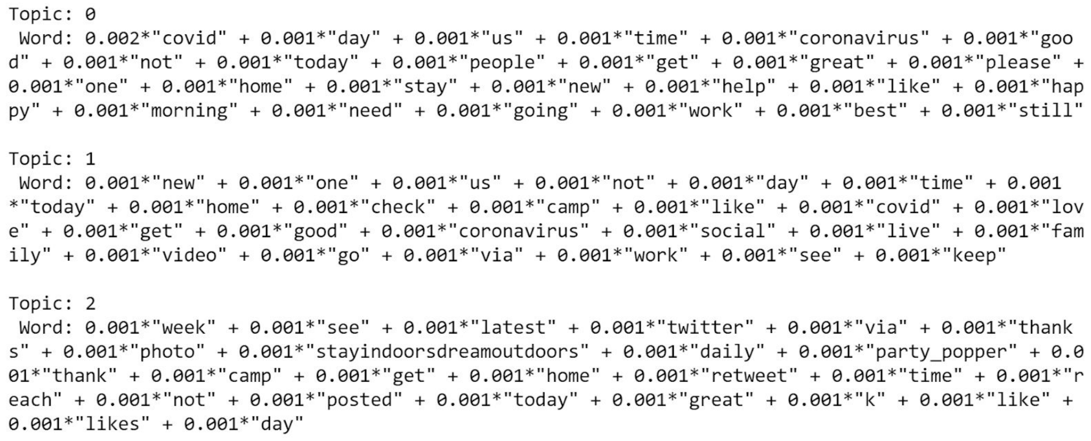


```
Topic: 0
 Word: 0.001*"new" + 0.001*"facebook" + 0.001*"check" + 0.001*"see" + 0.001*"video" + 0.001*"grea
t" + 0.001*"not" + 0.001*"posted" + 0.001*"get" + 0.001*"love" + 0.001*"one" + 0.001*"us" + 0.001
*"day" + 0.001*"week" + 0.001*"today" + 0.001*"come" + 0.001*"spring" + 0.001*"ready" + 0.001*"wee
kend" + 0.001*"join" + 0.001*"morning" + 0.001*"go" + 0.001*"like" + 0.001*"work" + 0.001*"find"

Topic: 1
 Word: 0.001*"spring" + 0.001*"great" + 0.001*"new" + 0.001*"one" + 0.001*"us" + 0.001*"get" + 0.0
01*"love" + 0.001*"time" + 0.001*"check" + 0.001*"happy" + 0.001*"day" + 0.001*"please" + 0.001*"f
un" + 0.001*"see" + 0.001*"video" + 0.001*"last" + 0.001*"going" + 0.001*"like" + 0.001*"life" +
0.001*"live" + 0.001*"back" + 0.001*"good" + 0.001*"night" + 0.001*"today" + 0.001*"dont"

Topic: 2
 Word: 0.001*"day" + 0.001*"new" + 0.001*"great" + 0.001*"time" + 0.001*"hiking" + 0.001*"check" +
0.001*"today" + 0.001*"april" + 0.001*"good" + 0.001*"video" + 0.001*"one" + 0.001*"best" + 0.001
*"season" + 0.001*"get" + 0.001*"us" + 0.001*"th" + 0.001*"like" + 0.001*"go" + 0.001*"via" + 0.00
1*"camp" + 0.001*"facebook" + 0.001*"life" + 0.001*"adventure" + 0.001*"look" + 0.001*"looking"
```

2020 Active:

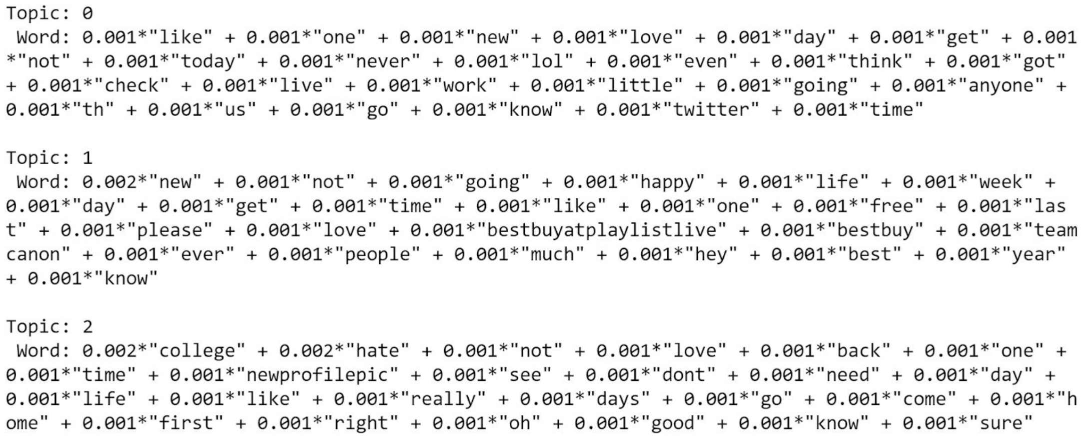


```
Topic: 0
 Word: 0.002*"covid" + 0.001*"day" + 0.001*"us" + 0.001*"time" + 0.001*"coronavirus" + 0.001*"goo
d" + 0.001*"not" + 0.001*"today" + 0.001*"people" + 0.001*"get" + 0.001*"great" + 0.001*"please" +
0.001*"one" + 0.001*"home" + 0.001*"stay" + 0.001*"new" + 0.001*"help" + 0.001*"like" + 0.001*"hap
py" + 0.001*"morning" + 0.001*"need" + 0.001*"going" + 0.001*"work" + 0.001*"best" + 0.001*"still"

Topic: 1
 Word: 0.001*"new" + 0.001*"one" + 0.001*"us" + 0.001*"not" + 0.001*"day" + 0.001*"time" + 0.001
*"today" + 0.001*"home" + 0.001*"check" + 0.001*"camp" + 0.001*"like" + 0.001*"covid" + 0.001*"lov
e" + 0.001*"get" + 0.001*"good" + 0.001*"coronavirus" + 0.001*"social" + 0.001*"live" + 0.001*"fam
ily" + 0.001*"video" + 0.001*"go" + 0.001*"via" + 0.001*"work" + 0.001*"see" + 0.001*"keep"

Topic: 2
 Word: 0.001*"week" + 0.001*"see" + 0.001*"latest" + 0.001*"twitter" + 0.001*"via" + 0.001*"thank
s" + 0.001*"photo" + 0.001*"stayindoorsdreamoutdoors" + 0.001*"daily" + 0.001*"party_popper" + 0.0
01*"thank" + 0.001*"camp" + 0.001*"get" + 0.001*"home" + 0.001*"retweet" + 0.001*"time" + 0.001*"r
each" + 0.001*"not" + 0.001*"posted" + 0.001*"today" + 0.001*"great" + 0.001*"k" + 0.001*"like" +
0.001*"likes" + 0.001*"day"
```

2019 Lazy:

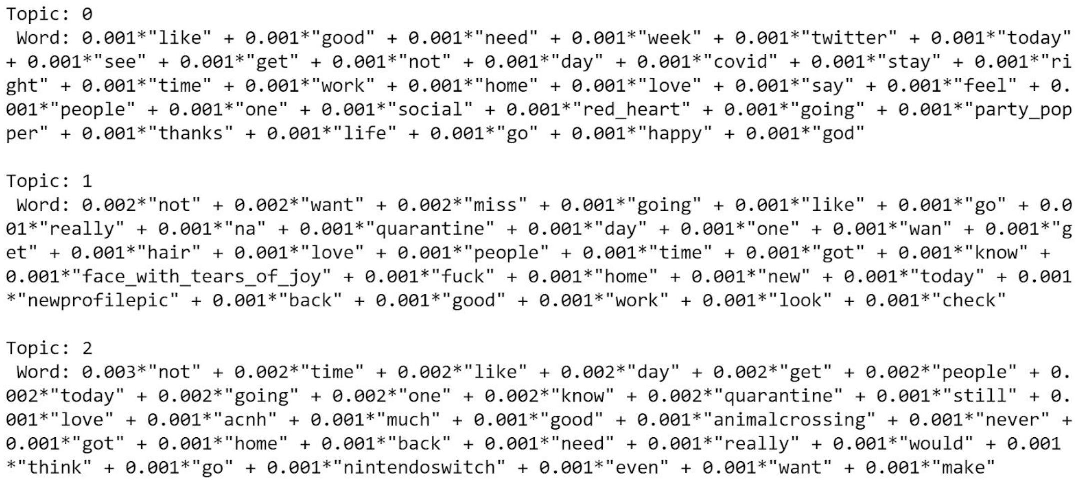


```
Topic: 0
 Word: 0.001*"like" + 0.001*"one" + 0.001*"new" + 0.001*"love" + 0.001*"day" + 0.001*"get" + 0.001
*"not" + 0.001*"today" + 0.001*"never" + 0.001*"lol" + 0.001*"even" + 0.001*"think" + 0.001*"got"
+ 0.001*"check" + 0.001*"live" + 0.001*"work" + 0.001*"little" + 0.001*"going" + 0.001*"anyone" +
0.001*"th" + 0.001*"us" + 0.001*"go" + 0.001*"know" + 0.001*"twitter" + 0.001*"time"

Topic: 1
 Word: 0.002*"new" + 0.001*"not" + 0.001*"going" + 0.001*"happy" + 0.001*"life" + 0.001*"week" +
0.001*"day" + 0.001*"get" + 0.001*"time" + 0.001*"like" + 0.001*"one" + 0.001*"free" + 0.001*"las
t" + 0.001*"please" + 0.001*"love" + 0.001*"bestbuyatplaylistlive" + 0.001*"bestbuy" + 0.001*"team
canon" + 0.001*"ever" + 0.001*"people" + 0.001*"much" + 0.001*"hey" + 0.001*"best" + 0.001*"year"
+ 0.001*"know"

Topic: 2
 Word: 0.002*"college" + 0.002*"hate" + 0.001*"not" + 0.001*"love" + 0.001*"back" + 0.001*"one" +
0.001*"time" + 0.001*"newprofilepic" + 0.001*"see" + 0.001*"dont" + 0.001*"need" + 0.001*"day" +
0.001*"life" + 0.001*"like" + 0.001*"really" + 0.001*"days" + 0.001*"go" + 0.001*"come" + 0.001*"h
ome" + 0.001*"first" + 0.001*"right" + 0.001*"oh" + 0.001*"good" + 0.001*"know" + 0.001*"sure"
```

2020 Lazy:



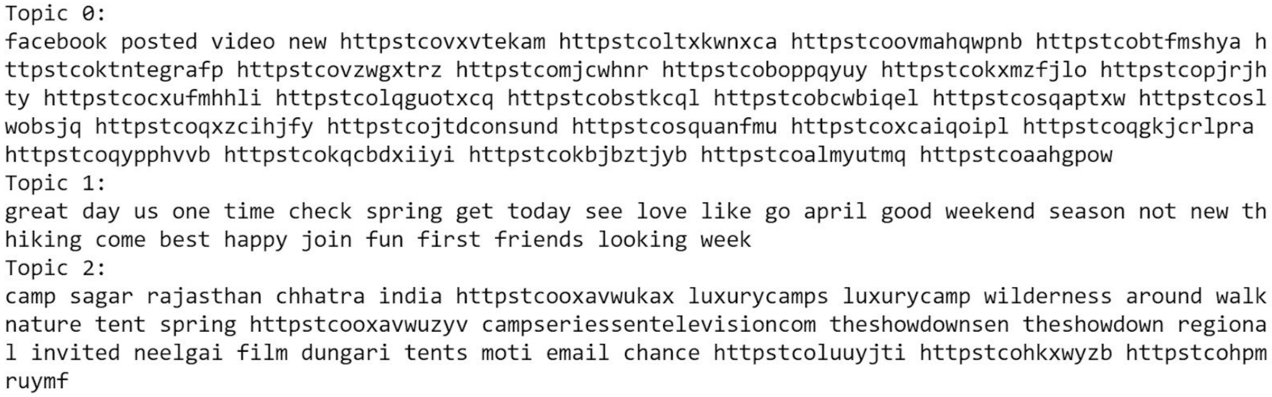
```
Topic: 0
 Word: 0.001*"like" + 0.001*"good" + 0.001*"need" + 0.001*"week" + 0.001*"twitter" + 0.001*"today"
+ 0.001*"see" + 0.001*"get" + 0.001*"not" + 0.001*"day" + 0.001*"covid" + 0.001*"stay" + 0.001*"ri
ght" + 0.001*"time" + 0.001*"work" + 0.001*"home" + 0.001*"love" + 0.001*"say" + 0.001*"feel" + 0.
001*"people" + 0.001*"one" + 0.001*"social" + 0.001*"red_heart" + 0.001*"going" + 0.001*"party_pop
per" + 0.001*"thanks" + 0.001*"life" + 0.001*"go" + 0.001*"happy" + 0.001*"god"

Topic: 1
 Word: 0.002*"not" + 0.002*"want" + 0.002*"miss" + 0.001*"going" + 0.001*"like" + 0.001*"go" + 0.0
01*"really" + 0.001*"na" + 0.001*"quarantine" + 0.001*"day" + 0.001*"one" + 0.001*"wan" + 0.001*"g
et" + 0.001*"hair" + 0.001*"love" + 0.001*"people" + 0.001*"time" + 0.001*"got" + 0.001*"know" +
0.001*"face_with_tears_of_joy" + 0.001*"fuck" + 0.001*"home" + 0.001*"new" + 0.001*"today" + 0.001
*"newprofilepic" + 0.001*"back" + 0.001*"good" + 0.001*"work" + 0.001*"look" + 0.001*"check"

Topic: 2
 Word: 0.003*"not" + 0.002*"time" + 0.002*"like" + 0.002*"day" + 0.002*"get" + 0.002*"people" + 0.
002*"today" + 0.002*"going" + 0.002*"one" + 0.002*"know" + 0.002*"quarantine" + 0.001*"still" + 0.
001*"love" + 0.001*"acnh" + 0.001*"much" + 0.001*"good" + 0.001*"animalcrossing" + 0.001*"never" +
0.001*"got" + 0.001*"home" + 0.001*"back" + 0.001*"need" + 0.001*"really" + 0.001*"would" + 0.001
*"think" + 0.001*"go" + 0.001*"nintendoswitch" + 0.001*"even" + 0.001*"want" + 0.001*"make"
```

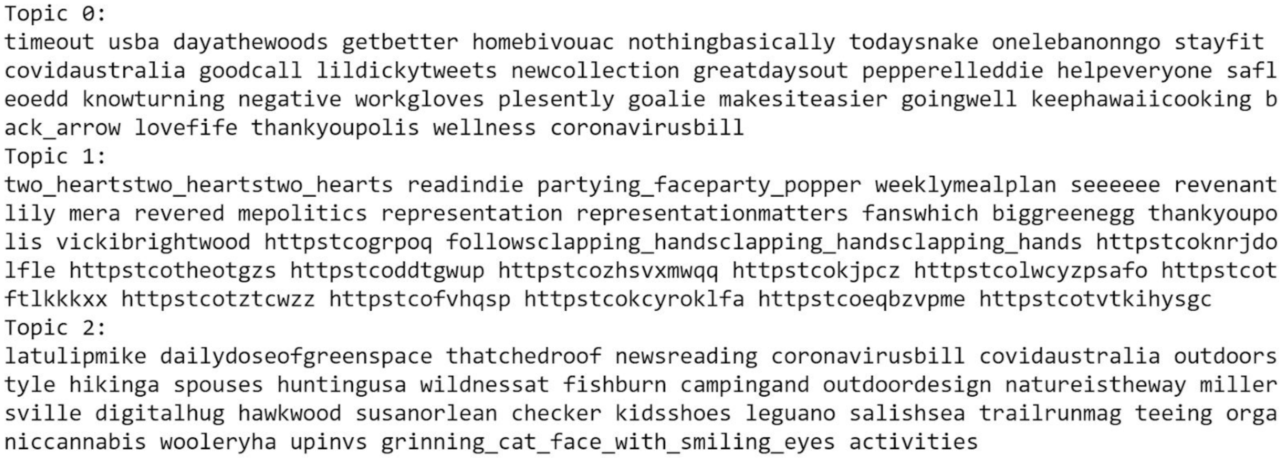
Appendix C: sklearn NMF Model Output

2019 Active:



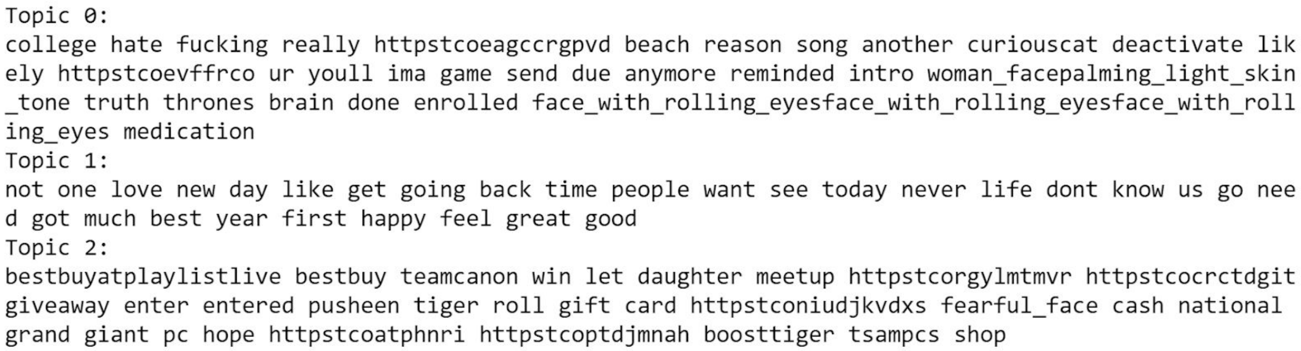
```
Topic 0:
facebook posted video new httpstcovxvtekam httpstcoltxkwnxca httpstcoovmahqwpnb httpstcobtfmshya h
ttpstcoktntegrafp httpstcovzwgxtrz httpstcomjcwhnr httpstcoboppqyuy httpstcokxmzfjlo httpstcopjrjh
ty httpstcocxufmhhli httpstcolqguotxcq httpstcobstkcql httpstcobcwbiqel httpstcosqaptxw httpstcosl
wobsjq httpstcoqxzcihjfy httpstcojtdconsund httpstcosquanfmu httpstcoxcaiqoipl httpstcoqgkjcrlpra
httpstcoqypphvvb httpstcokqcbdxiiyi httpstcokbjbztjyb httpstcoalmyutmq httpstcoaahgpow
Topic 1:
great day us one time check spring get today see love like go april good weekend season not new th
hiking come best happy join fun first friends looking week
Topic 2:
camp sagar rajasthan chhatra india httpstcooxavwukax luxurycamps luxurycamp wilderness around walk
nature tent spring httpstcooxavwuzyv campseriessentelevisioncom theshowdownsen theshowdown regiona
l invited neelgai film dungari tents moti email chance httpstcoluuyjti httpstcohkxwyzb httpstcohpm
ruymf
```

2020 Active:



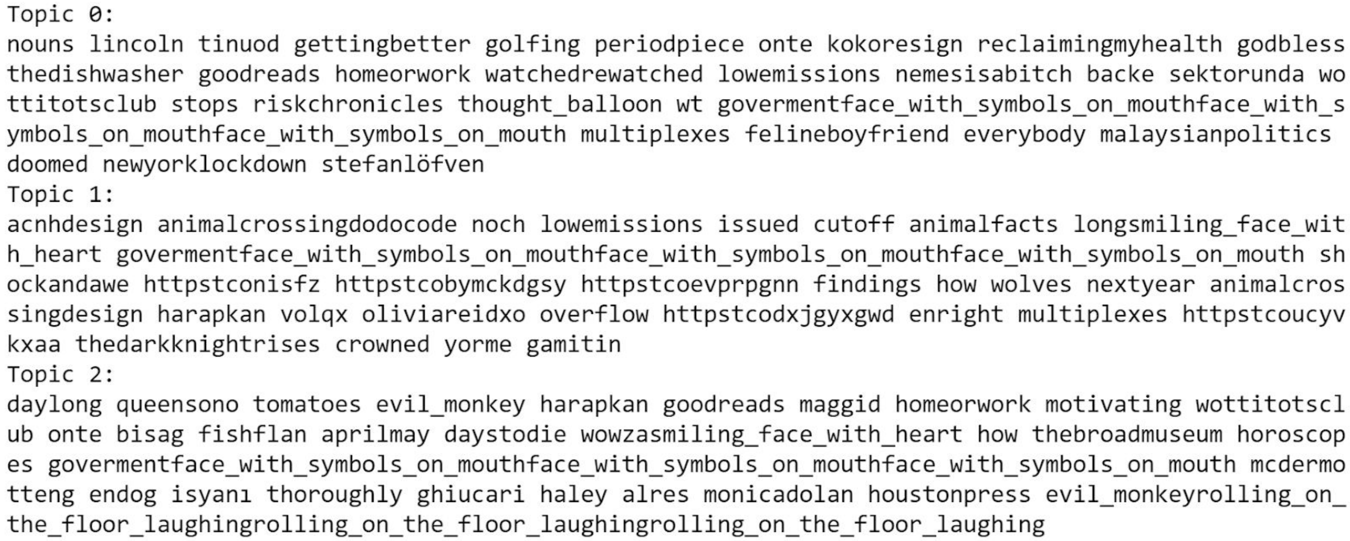
```
Topic 0:
timeout usba dayathewoods getbetter homebivouac nothingbasically todaysnake onelebanonngo stayfit
covidaustralia goodcall lildickytweets newcollection greatdaysout pepperelleddie helpeveryone safl
eoedd knowturning negative workgloves plesently goalie makesiteasier goingwell keephawaiicooking b
ack_arrow lovefife thankyoupolis wellness coronavirusbill
Topic 1:
two_heartstwo_heartstwo_hearts readindie partying_faceparty_popper weeklymealplan seeeeee revenant
lily mera revered mepolitics representation representationmatters fanswhich biggreenegg thankyoupo
lis vickibrightwood httpstcogrpoq followsclapping_handsclapping_handsclapping_hands httpstcoknrjdo
lfle httpstcotheotgzs httpstcoddtgwup httpstcozhsvxmwqq httpstcokjpcz httpstcolwcyzpsafo httpstcot
ftlkkkxx httpstcotztcwzz httpstcofvhqsp httpstcokcyroklfa httpstcoeqbzvpme httpstcotvtkihysgc
Topic 2:
latulipmike dailydoseofgreenspace thatchedroof newsreading coronavirusbill covidaustralia outdoors
tyle hikinga spouses huntingusa wildnessat fishburn campingand outdoordesign natureistheway miller
sville digitalhug hawkwood susanorlean checker kidsshoes leguano salishsea trailrunmag teeing orga
niccannabis wooleryha upinvs grinning_cat_face_with_smiling_eyes activities
```

2019 Lazy:



```
Topic 0:
college hate fucking really httpstcoeagccrgpvd beach reason song another curiouscat deactivate lik
ely httpstcoevffrco ur youll ima game send due anymore reminded intro woman_facepalming_light_skin
_tone truth thrones brain done enrolled face_with_rolling_eyesface_with_rolling_eyesface_with_roll
ing_eyes medication
Topic 1:
not one love new day like get going back time people want see today never life dont know us go nee
d got much best year first happy feel great good
Topic 2:
bestbuyatplaylistlive bestbuy teamcanon win let daughter meetup httpstcorgylmtmvr httpstcocrctdgit
giveaway enter entered pusheen tiger roll gift card httpstconiudjkvdxs fearful_face cash national
grand giant pc hope httpstcoatphnri httpstcoptdjmnah boosttiger tsampcs shop
```

2020 Lazy:



```
Topic 0:
nouns lincoln tinuod gettingbetter golfing periodpiece onte kokoresign reclaimingmyhealth godbless
thedishwasher goodreads homeorwork watchedrewatched lowemissions nemesisabitch backe sektorunda wo
ttitotsclub stops riskchronicles thought_balloon wt govermentface_with_symbols_on_mouthface_with_s
ymbols_on_mouthface_with_symbols_on_mouth multiplexes felineboyfriend everybody malaysianpolitics
doomed newyorklockdown stefanlöfven
Topic 1:
acnhdesign animalcrossingdodocode noch lowemissions issued cutoff animalfacts longsmiling_face_wit
h_heart govermentface_with_symbols_on_mouthface_with_symbols_on_mouthface_with_symbols_on_mouth sh
ockandawe httpstconisfz httpstcobymckdgsy httpstcoevprpgnn findings how wolves nextyear animalcros
singdesign harapkan volqx oliviareidxo overflow httpstcodxjgyxgwd enright multiplexes httpstcoucyv
kxaa thedarkknightrises crowned yorme gamitin
Topic 2:
daylong queensono tomatoes evil_monkey harapkan goodreads maggid homeorwork motivating wottitotscl
ub onte bisag fishflan aprilmay daystodie wowzasmiling_face_with_heart how thebroadmuseum horoscop
es govermentface_with_symbols_on_mouthface_with_symbols_on_mouthface_with_symbols_on_mouth mcdermo
tteng endog isyanı thoroughly ghiucari haley alres monicadolan houstonpress evil_monkeyrolling_on_
the_floor_laughingrolling_on_the_floor_laughingrolling_on_the_floor_laughing
```