Twitter Users Analysis in the Times of COVID

Capstone Project 2

By Anna Kantur

Springboard

Problem Statement

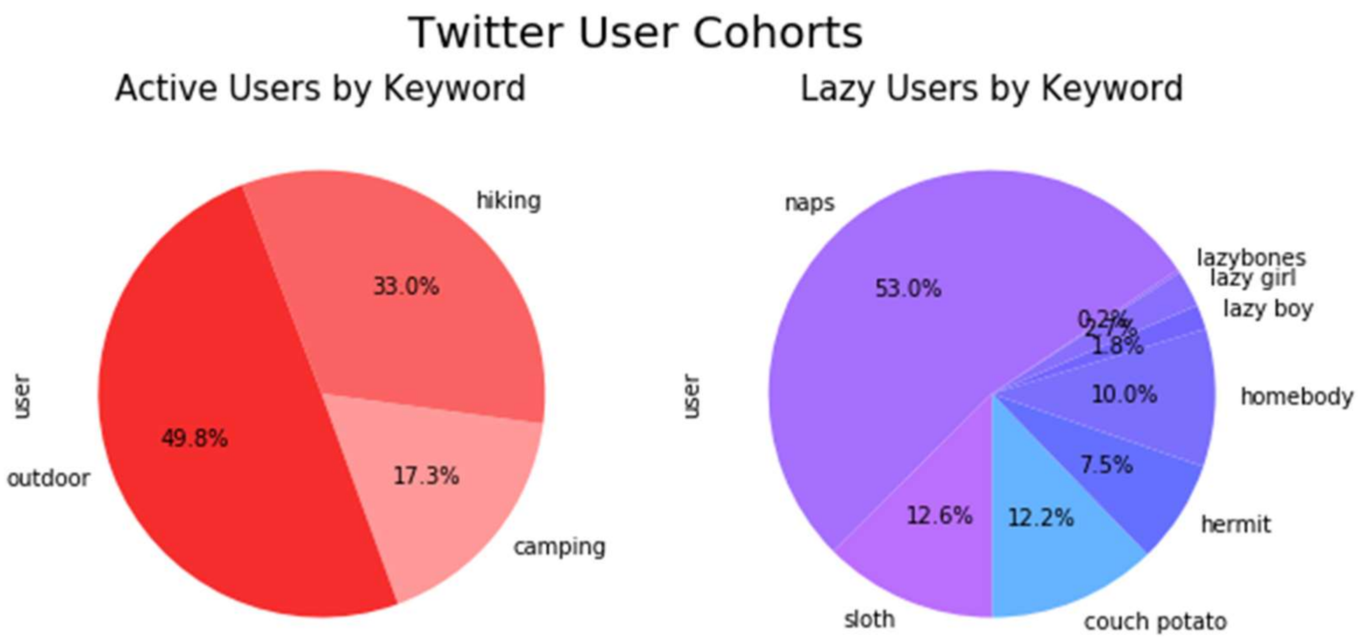
- COVID-19 pandemic lockdown orders made outdoor enthusiasts around the world stay at home
- People turned to social media to express their emotions and share the news about their new routines
- Twitter is a powerful source of information and there are 50K+ users who have “outdoors” in their Twitter bio
- We decided to analyze how “active” and “lazy” Twitter users react to COVID and the necessity to stay at home, what they do while in lockdown, and how the mood changed comparing to the same time period in 2019
- Our analysis is relevant for sociologists who are trying to analyze the social changes that were brought on us, as well as for the outdoor enthusiasts themselves who are trying to find new ways to cope with the lifestyle changes

Data Wrangling

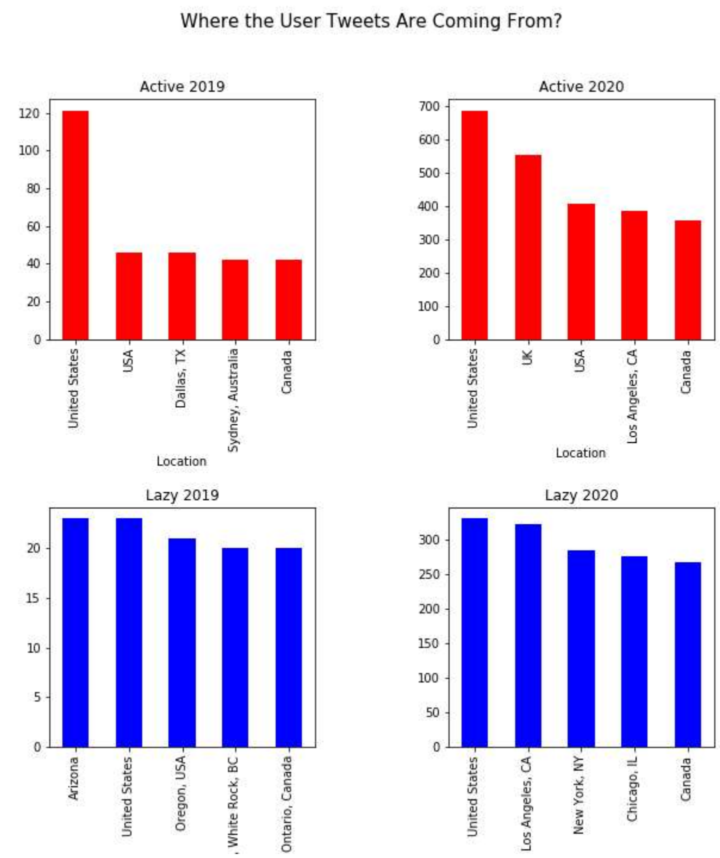
- We choose users using Twitter bios search on followerwonk.com
- Keywords for “active” users: “outdoors”, “hiking”, “camping”
- Keywords for “lazy” users: "homebody", "couch potato", ”hermit”, “lazy boy”, “lazy girl”, “sloth”, “naps”, “lazybones”
- Time period: March 1-April 15, 2019 and March 1-April 15, 2020
- We collected the tweets of ~10K “active” and “lazy” English-speaking Twitter users
- We excluded replies and retweets

Springboard

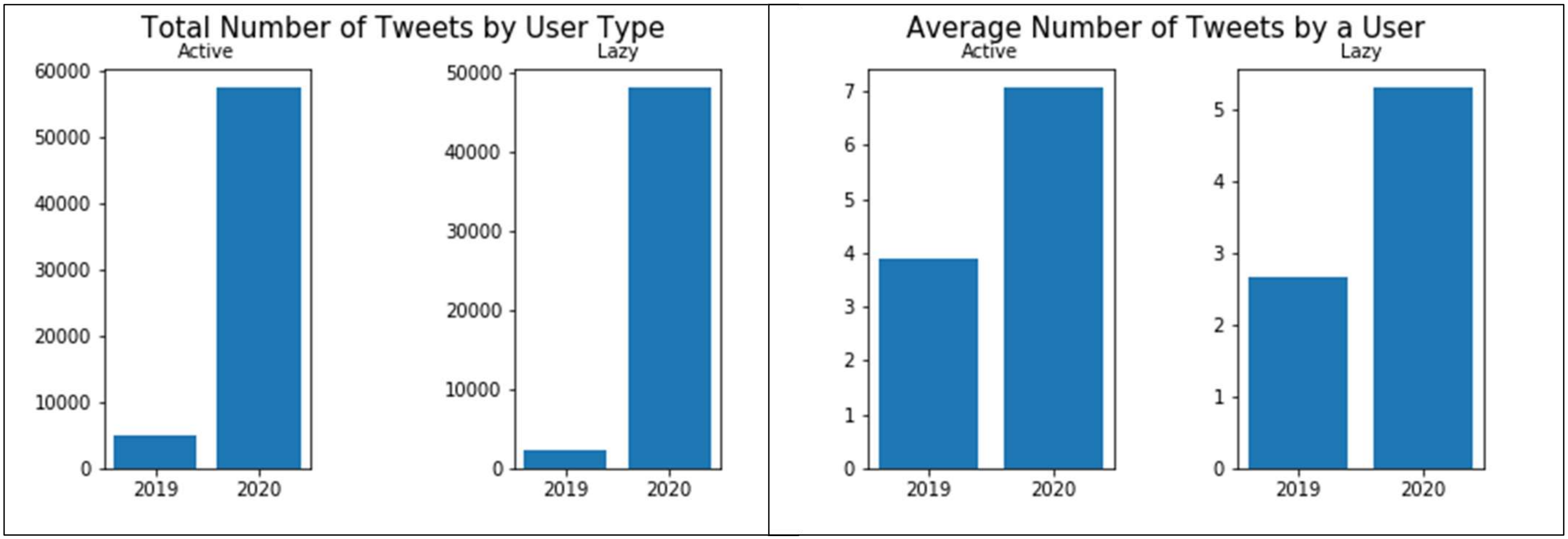
Users by Keyword



Most Tweets Geographies



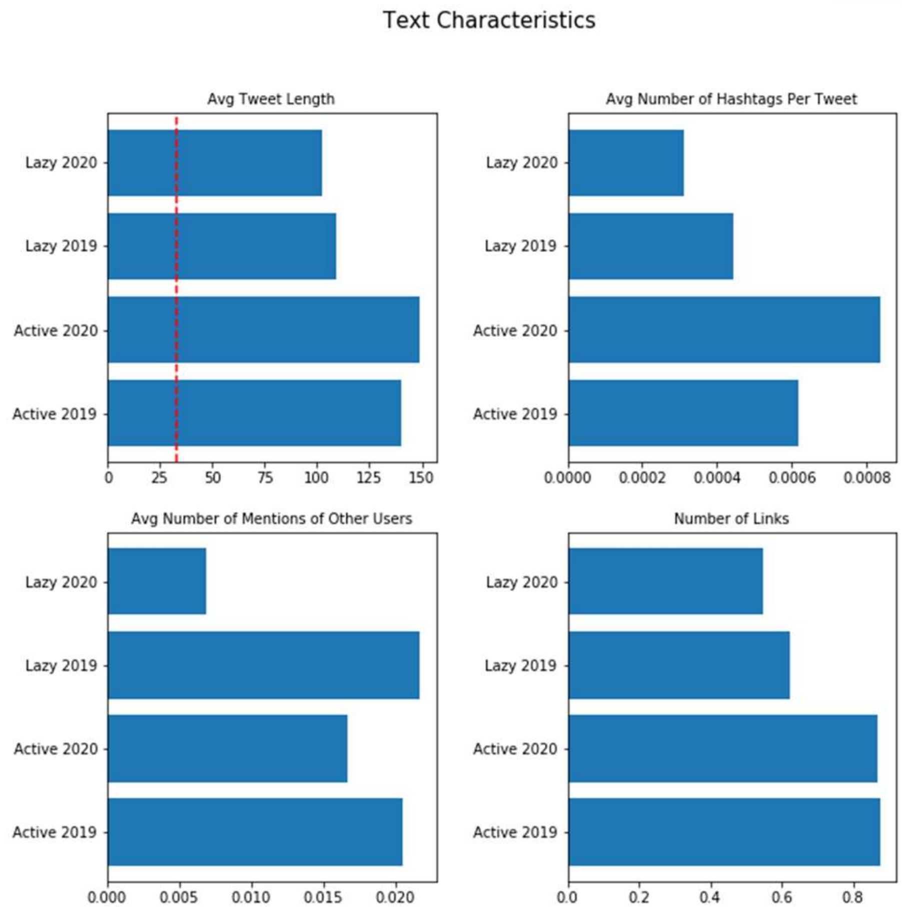
More Tweets in 2020



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.

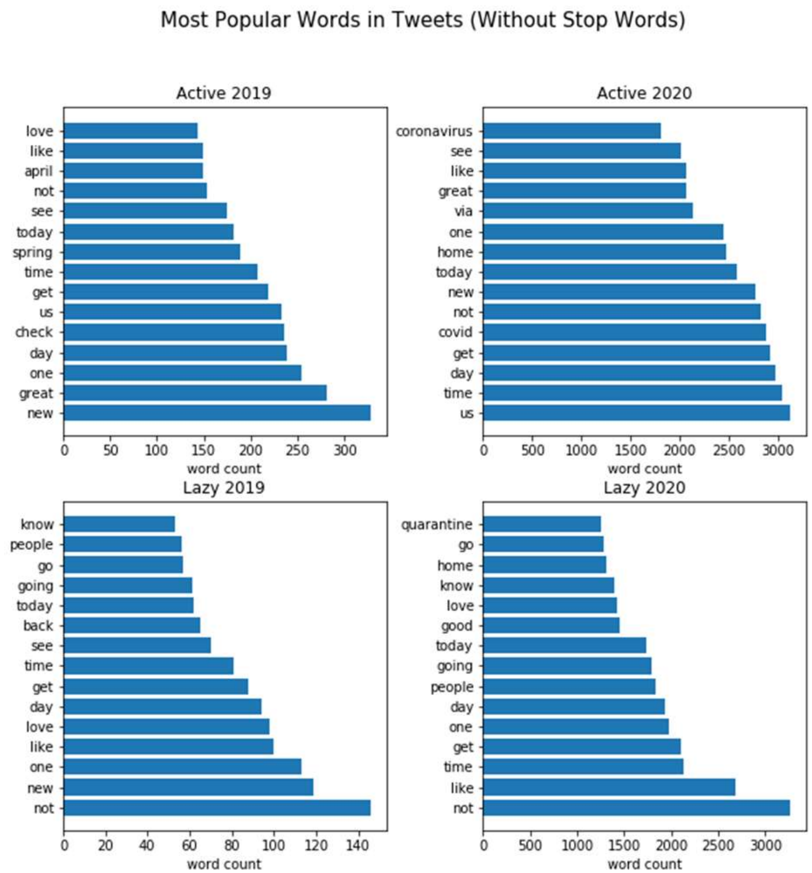
Text Characteristics

- "Active" User Tweets were more than 30 characters longer
- Active users used more mentions and links
- Tweets in both user cohorts are over 100 characters (vs average tweet length of 33 characters)



Springboard

COVID Topic Shows Up in Most Popular Words



Springboard

COVID Theme Shows in Most Popular Emojis

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)]
```

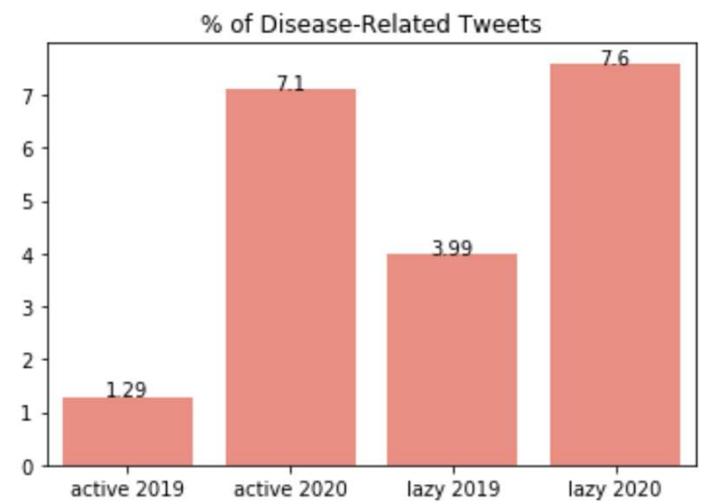
Springboard

Topic Modeling

User Cohort	NMF Model (sklearn)	LDA Model (genism)
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)

COVID Twitter Sentiment Analysis

- We used the concept of semantic similarity and Spacy word vectors to filter disease-related tweets in 2019 and COVID-related tweets in 2020
- Keywords for disease-related tweets in 2019: disease, virus, cough, doctor, nurse, hospital
- Keywords for COVID-related tweets in 2020: covid, coronavirus, virus, pandemic, disease, lockdown, quarantine, cough, doctor, nurse, hospital
- Between 2019 and 2020 there was a 500% YoY increase in disease-related tweets for “active” users and only 100% YoY increase for “lazy” users



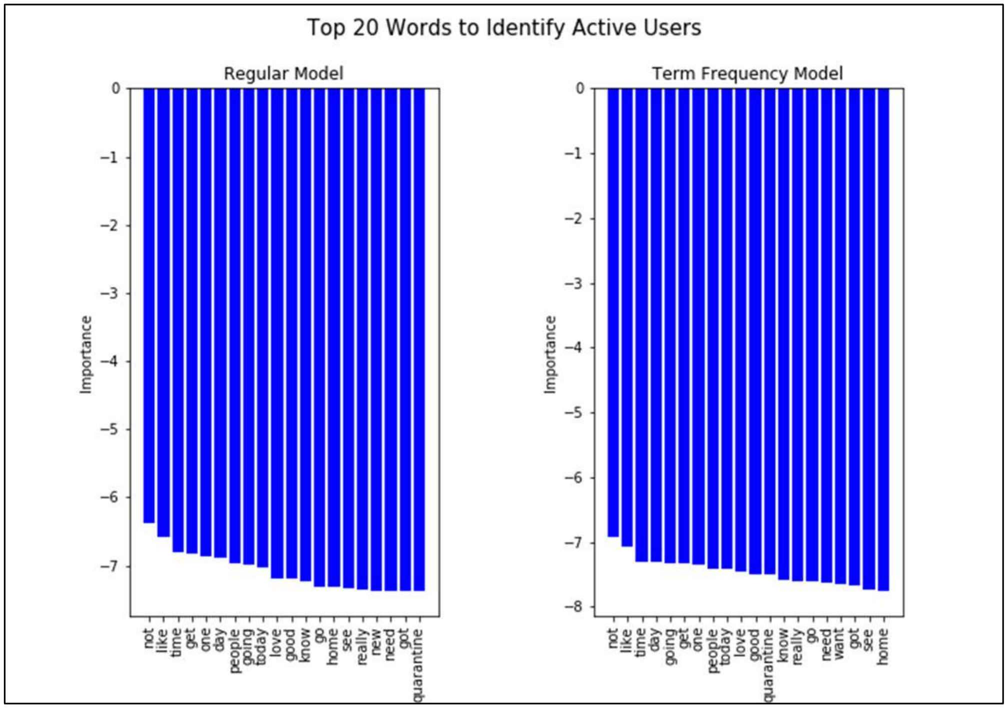
COVID Twitter Sentiment Analysis (Continued)

- We used vaderSentiment out-of-the-box Twitter sentiment model and the recommended thresholds:
  - positive sentiment: compound score >= 0.05
  - neutral sentiment: (compound score > -0.05) and (compound score < 0.05)
  - negative sentiment: compound score <= -0.05
- 2020 to 2019: the positive sentiment increased in 2020 by 1% from 51.5% to 52.5%
- “active” vs “lazy”: “active” users are almost 8% more negative on Twitter about diseases than “lazy” users
- 2019 “active” vs 2020 “active”, 2019 “lazy” vs 2020 “lazy”: COVID pandemic of 2020 affected the user cohorts equally negatively (5.5% more negative tweets for “active” users, 6.5% for “lazy” users)

Springboard

Most Predictive Words for "active" users (Naïve Bayes)

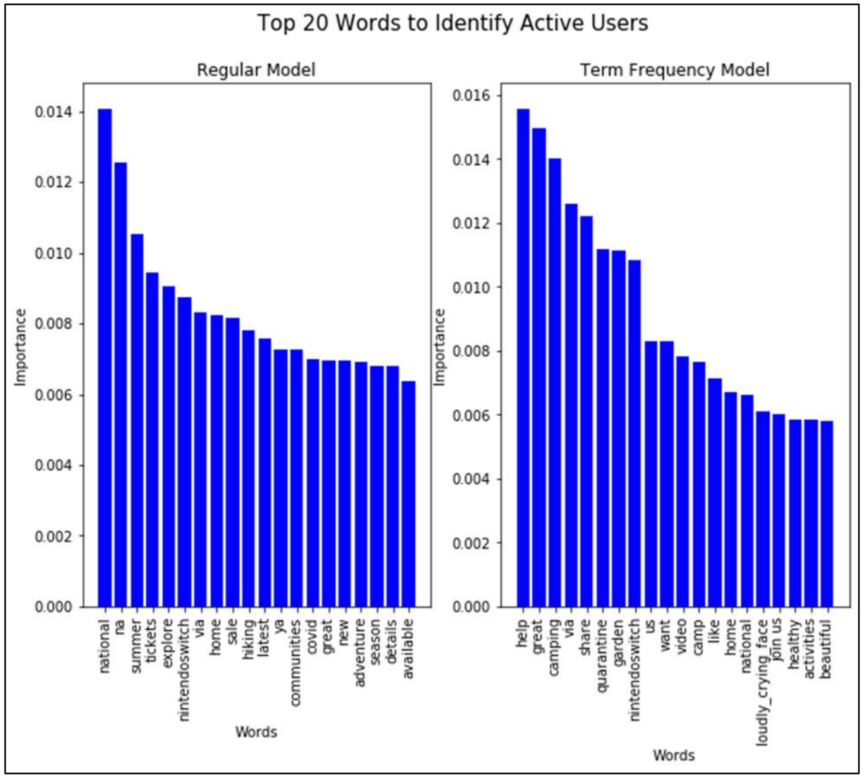
- ROC_AUC best score: 85%
- We used the logarithmic approach to find top words (e.g. feature importance can be negative)
- Even the most highly rated words had negative values, so they were not highly predictive



Springboard

Most Predictive Words for "active" users (Random Forest)

- ROC_AUC best score: 71%
- "active" users turned to Nintendo Switch games during quarantine



Project Findings

- There were 10x more tweets in 2020 in both cohorts, and the top emojis showed confusion and negativity of the COVID uncertain times
- The mood in tweets about diseases stayed as negative in 2020 as it was in 2019, but 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.)
- “active” users changed their hobbies more drastically: while still dreaming about outdoors and planning future trips, they also switched to reading and playing online games
- “lazy” users continued with their home-based hobbies and became less opportunistic by not participating in online contests like in 2019

Springboard

Ideas for Further Research

- Re-do the analysis on larger population (broader definitions for “active” and “lazy”, other languages)
- Consider a wider time period since now the lockdown orders stay until June or longer
- Create an app to predict a Twitter user type by keywords
- Create an app to predict what % chance is that an “active” user converts into a “lazy” user after the extended lockdown at home.

Springboard

Thank You!

Anna Kantur
https://www.linkedin.com/in/annakantur/
https://github.com/a-kantur

Springboard