TWITTER SENTIMENT ANALYSIS

AARON HANNA, NIKHIL PATHAK
CHRIS LIGHT, SONYA MOHAMMED

Home Users Tools About Tweets

Twitter User Analysis

Username

Analyze

ECE-143 | Team 4
Nikhil Pathak | Chris Light | Sonya Mohammed | Aaron Hanna
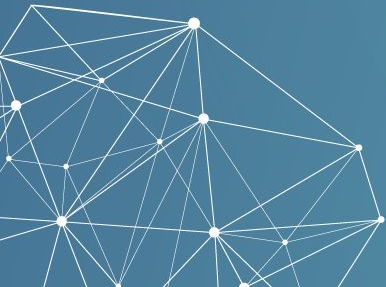
PROJECT MOTIVATION

Rise of Social Media Impact

Influential/Celebrity Accounts

Personal Tweet Improvement

01

Sentiment Model

Sentiment140 Dataset

	target	ids	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
4	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....
...	...	...	...	...	...	...
1599995	4	2193601966	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	AmandaMarie1028	Just woke up. Having no school is the best fee...
1599996	4	2193601969	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	TheWDBoards	TheWDB.com - Very cool to hear old Walt interv...
1599997	4	2193601991	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	bpbabe	Are you ready for your MoJo Makeover? Ask me f...
1599998	4	2193602064	Tue Jun 16 08:40:49 PDT 2009	NO_QUERY	tinydiamondz	Happy 38th Birthday to my boo of alll time!!! ...
1599999	4	2193602129	Tue Jun 16 08:40:50 PDT 2009	NO_QUERY	RyanTrevMorris	happy #charitytuesday @theNSPCC @SparksCharity...

1600000 rows × 6 columns

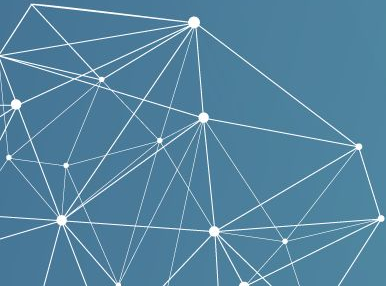1.6 Million Tweets - Labeled Positive or Negative Sentiment

Data Pre-Processing

Extract Only Tweet and Sentiment

Sample to Reduce Dataset Size

Apply NLTK TweetTokenization

Tweet and Token Vectorization

Bag-of-Words

Counts of N-Grams

Tf-Idf*

Importance of N-Grams

Word2Vec

Meaning of Words

Not used for Tweet Classification

*Tf-Idf: Term Frequency-Inverse Document Frequency
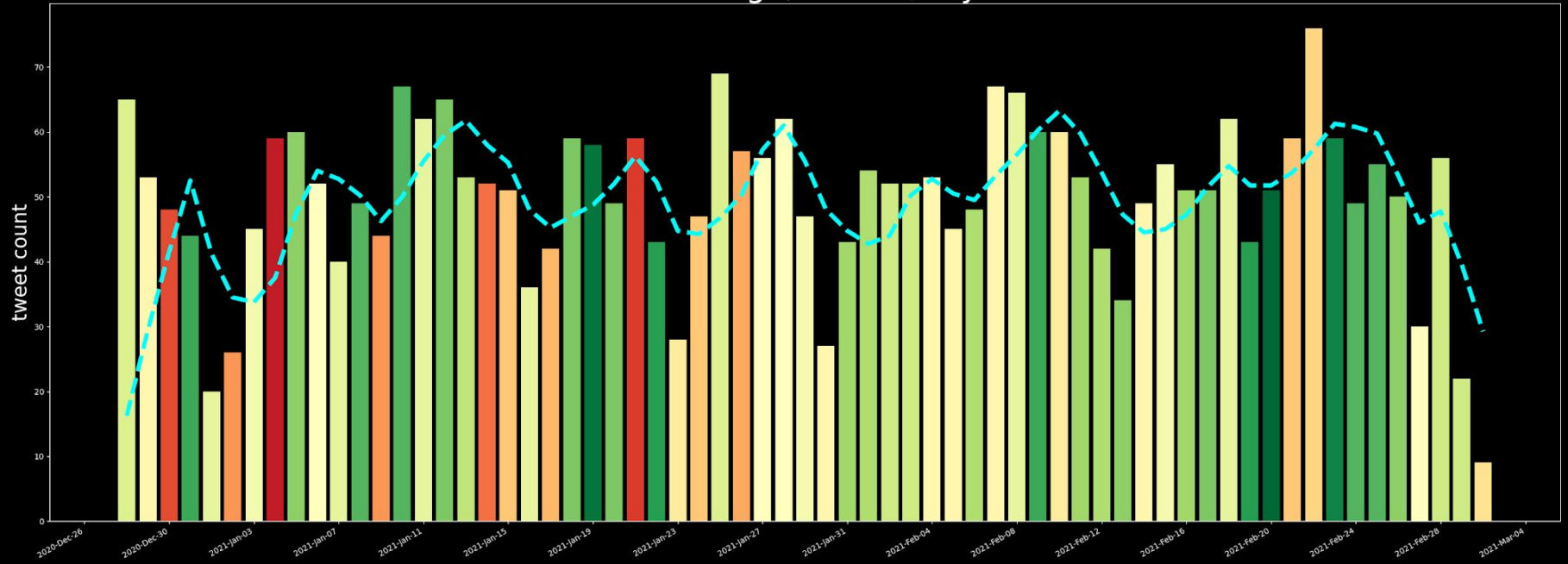
Logistic Regression Classifier Performance

	Training Accuracy	Testing Accuracy
BOW Model	82.1%	78.9%
Tf-Idf Model ✅	84.1%	79.1%

02

Account Case Studies

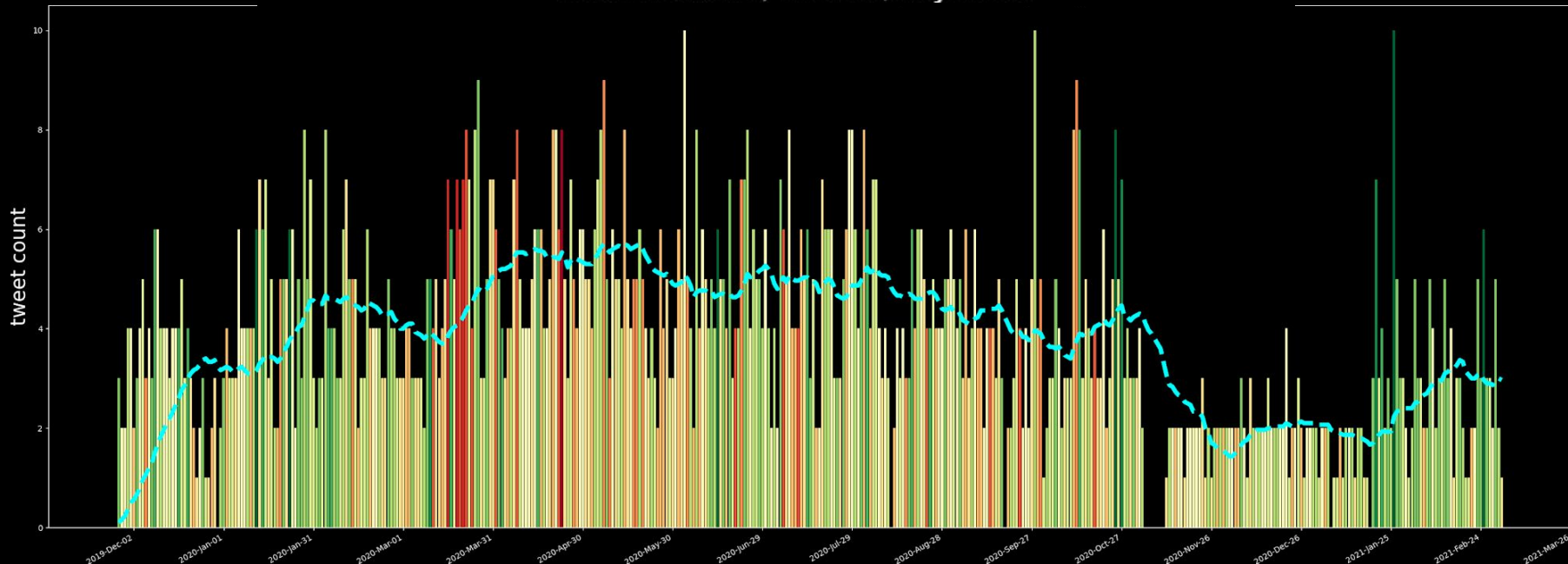FOX 5 San Diego, tweets/day: 50.00

Chronological Tweet Analysis

IRSnews, tweets/day: 23.38
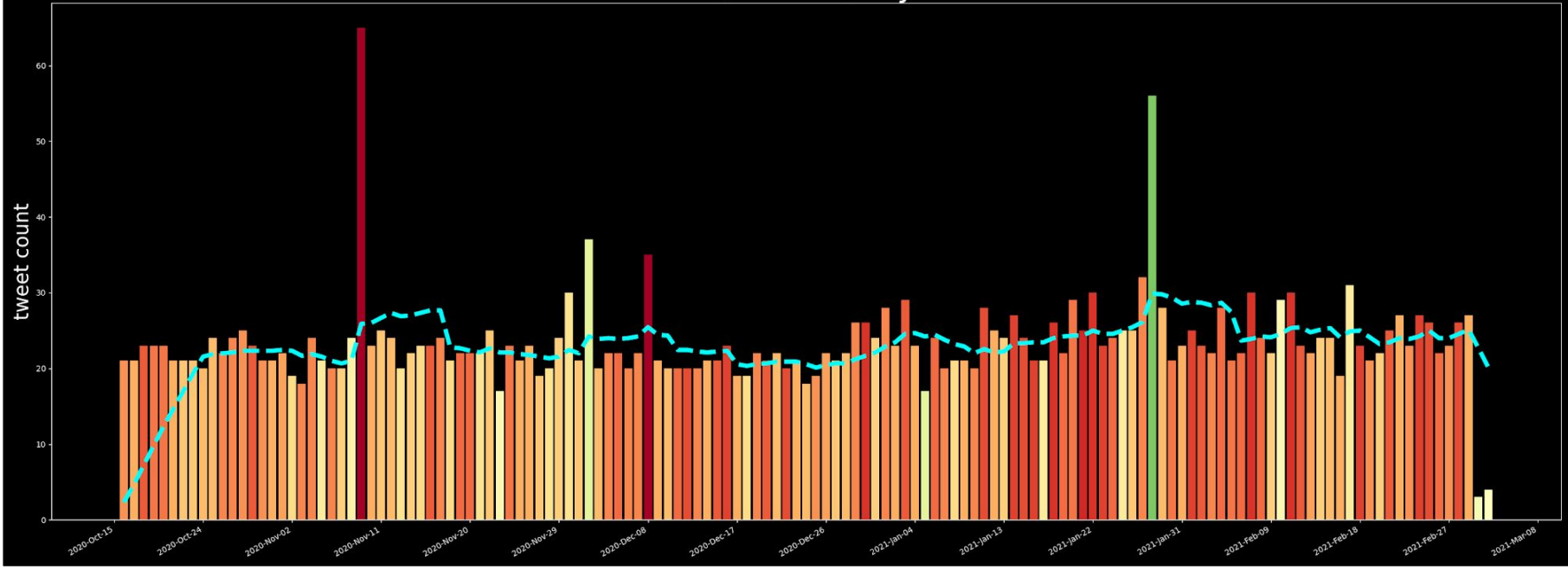
Example Negative Tweet:
"*#IRS doesn't call taxpayers demanding personal info, payments or threats of lawsuits or arrests...*"
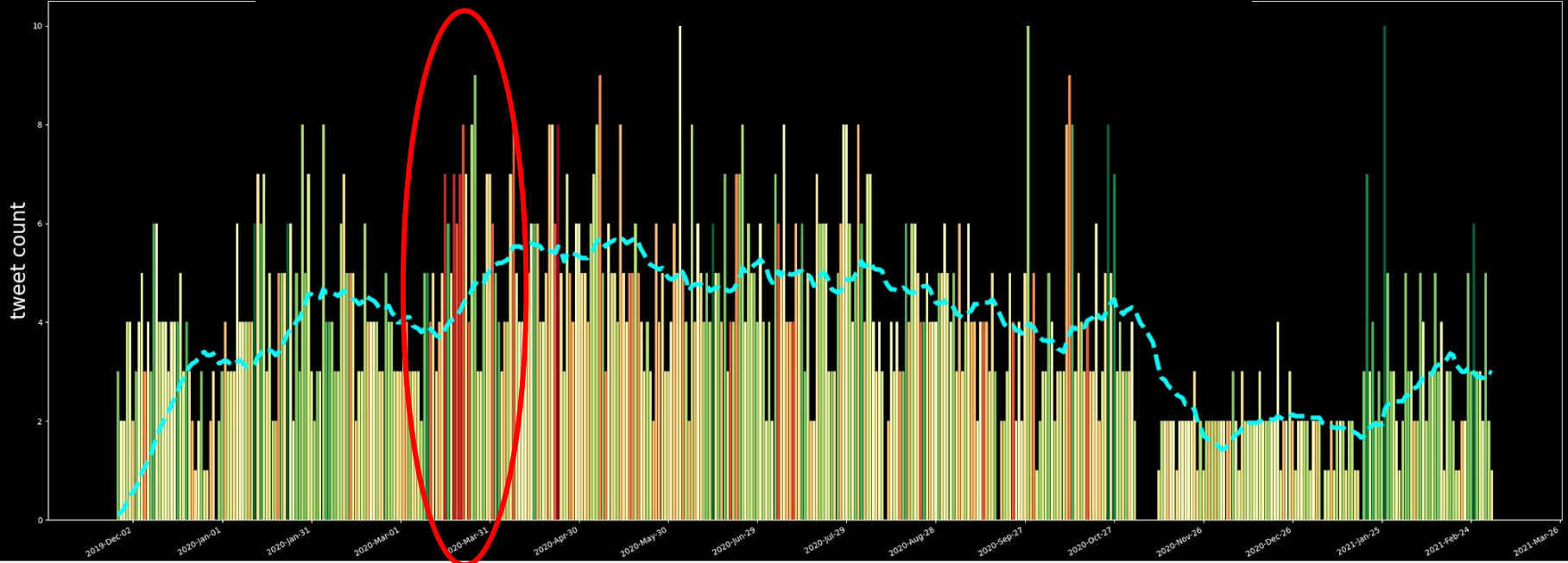
(Joint Federal Account) Vice President, tweets/day: 3.89
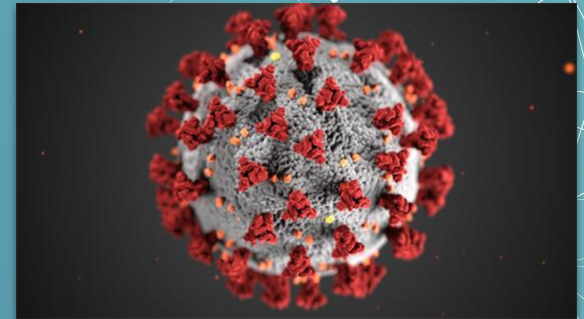
Mike Pence
(2016–2020)

Kamala Harris
(2020–2024)

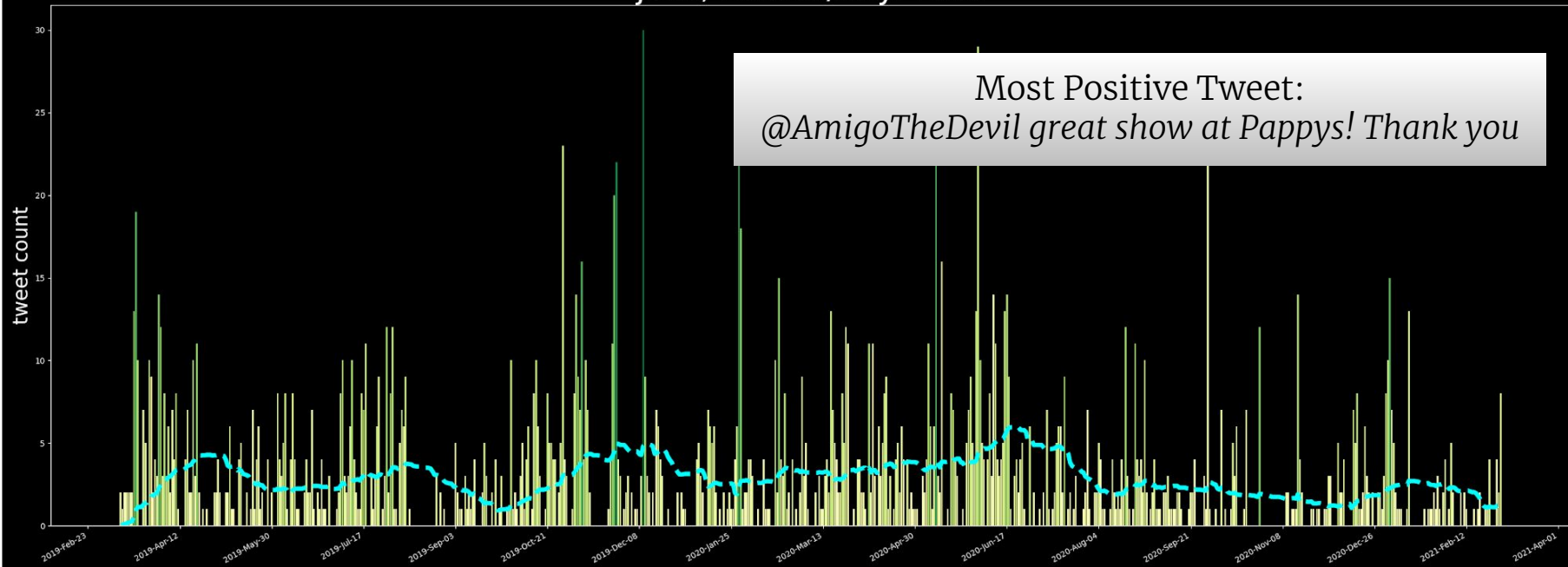(Joint Federal Account) Vice President, tweets/day: 3.89

"*Trump...issued two national emergency declarations...on March 13.*"

(National Conference of State Legislatures, 2020)

jack, tweets/day: 2.89

Most Positive Tweet:
@AmigoTheDevil great show at Pappys! Thank you

Jack Dorsey, CEO of Twitter Inc.

03

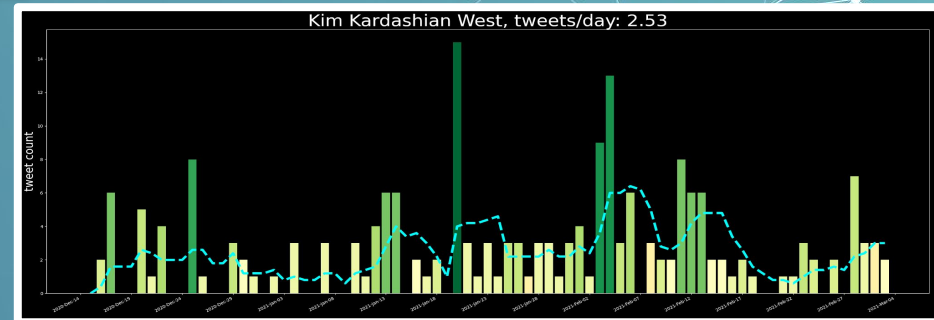# Word Cloud Analysis

Word Cloud of Most Used Words of Kim Kardashian West's Twitter

Kim Kardashian, Public Figure

Word Cloud

- Plot of most used words and their sentiment
- Positive sentiment: shades of red
- Negative sentiment: shades of green

Kamala Harris, U.S. Vice President — Word Cloud of Most Used Words of Kamala Harris's Twitter

Joe Biden, U.S. President — Word Cloud of Most Used Words of Joe Biden's Twitter

04

A Brief Demonstration

THANKS

Questions?