

TCSS 445: Final Report

Group Name: Tweetrumper

Daniel Looney Samuel Wainright Nazim Zerrouki Adam Shandi

7 MAR 2020

#### Abstract

Tweetrumper App is a joint JavaScript-Python technology that crawls the 2020 election candidates' and the President of United States' Twitter accounts timelines and collects tweets, uses regular expressions to determine keywords, then commits them to a database for searching. The front-end of the applications utilizes serverless commands to organize, aggregate, and display the data from the database on the web browser. The Tweetrumper App is a program that aids user decisions in elections by enabling viewership of tweets related to keywords.

#### Introduction

The Tweetrumper application, at its core, is a program that searches for Twitter data on all Democratic nominee hopefuls' and the current President's latest tweets as they race for the Democratic Convention Nomination and the 2020 Presidential election. The data will highlight key or core issues in society linked with every Tweet. The program will expose those who are concerned with the issues at hand and to what extent they express this for several categories such as: economy, debt, workers, and health care. These are seen as issues that people care for, that people work for, and what people vote for and they have significant influence on the outcome of the election.

### Objectives and Scope

The original goal in the proposed project was to expose or bring to light the manner of President Donald J. Trump's Twitter feed and his views on the topics listed above. However, that viewpoint has expanded as well as the scope of the project linearly. The Tweetrumper application now keeps track of all the candidates involved with the election since late January 2020, before the Democratic National Party's nomination process began in full swing. These candidates include: President Donald J. Trump, Sen. Bernie Sanders, Sen. Elizabeth Warren, Sen. Joe Biden, Sen. Amy Klobuchar, Rep. Tulsi Gabbard, Gov. Pete Buttigieg, Former Mayor Mike Bloomberg, Andrew Yang, and Tom Steyer. This expansion was from notions taught in TCSS 325, Ethics, that the program must be closer to an unbiased form or there will be inconsistencies in its usage. The objective of the program is to enlighten users on viewpoints made by candidates. If the program narrows the scope to one candidate, then viewership of the content will consist of trivial users who only seek a single source of data. The mandate of the program is to broaden each user's view on each topic and to spread knowledge of what each candidate has to bring to the election table.

### **Existing Solutions**

An existing solution is called Trump Twitter Archive which shows all of Trump's tweets based on popular keywords that users might search for as well as a search engine that enables users to find all tweets by use of a search bar. This archive of data does not work in a relational data method because entered keywords do not specifically target the target topics that the Tweetrumper application does target. While their search feature allows for a broader viewset, the data is not presented in a manner that could be organized to the user's interests. However, the Tweetrumper application should define a relationship such that while each keyword can be found in multiple tweets, we want to actually narrow the search down, so

it accounts for a keyword in every message matching topics that show consistently in candidate debates. Essentially, the application should present the information in a faster and more concise way than what Trump Twitter Archive aims to accomplish.

Another difference is that the scope of our project has increased. Trump Twitter Archive only keeps track of the tweets from President Donald J. Trump. While the website still provides information, the Tweetrumper application cannot accept a closed-world view of only having a single political figure as its featureset. Tweetrumper keeps record of all candidates (who are in the best stature for acquiring delegates) and reports their tweets from their timeline. This is an expanded view and truly sets the two applications apart.

### Architecture & Design Overview

The Tweetrumper application utilizes serverless functions to call the database for information to display on the front end. To begin an overview of the design, the numerous technologies that the application uses will be stated. The program uses the following set of languages: HTML, CSS, JavaScript, and Python. The accompanying frameworks and libraries aids the application: Node.js, Next.js, React, and Tweepy (Twitter's REST API). The primary DBMS used is MariaDB (a fork of MySQL).

The Node.js, Next.js, and React half of Tweetrumper comprises the "front-end" of the program. React is a javascript library that allows for easy creation of web applications. Next.js is a framework for React that handles many necessary functions such as file-system routing and server side rendering [1]. Next.js uses Node.js for server-side rendering of the website and database connectivity. Next.js also allows for easy creation of serverless API functions to handle the fetching of data from an external source, such as our SQL server. Although we run these functions on the same server as the web application, it would be possible to export them to run on the cloud, along with our database.

The other half of the application utilizes the Tweepy library in a Python program called "candidate\_reader.py" or the Candidate Reader. The Tweepy library is a REST API created by Twitter, Inc. that allows usage of their read and write functionalities in regards to their data sets on their respectively-named website. The Tweetrumper application uses the read-only functionality of the API, where Twitter handles of candidates (there username followed by the '@' on the website) are the primary key to access data objects it returns known as "Status". Status objects are packaged JSON files that contain bundled user information.

Candidate Reader first makes a connection to the database and sets up a command cursor pointing to the database. This allows Candidate Reader to use prepared SQL queries on the database without producing a .sql extension file. Then, the program employs a query on each candidate's Twitter timeline tweets and retrieves Status objects at a set number (200~1000), starting from the top. Every Status object has an identifier called "full\_text" or the entire text of the tweet. The program scans every text string extracted in this manner and compares it against a list of keywords using regular expressions. The keywords are subjects pertinent to the presidential election and include terms like "economy" and "health care".

Every text string, the keyword, the tweet identifier, and the user identifier are bundled into a list of dictionary items. After collecting these assets for each candidate, the program shifts to building the database if it does not exist, if it doesn't it builds it then builds the relational tables: candidates, handles, tweets, and tweet\_content. These dictionary items are then written line by line to the database using prepared statements, statements that come in the format of "INSERT INTO candidates VALUES (%s, %s, %s, %s)" where "%s" are the values in the dictionary being retrieved and written. After all writes are committed to the database, the program exits. This procedure is required for populating the database with a dataset vital to the program's view.

### Results & Analysis

The Tweetrumper application succinctly does its goal where it enables the user to search for topics like "gun violence" and "economy", then assess the contents of every tweet within the top 200 to 1000 posts of each candidate. Some candidates focus on different topics more than others like Bernie Sanders has talked about the "workers" the most, but in another area, Tom Steyer focuses on the economy. The most surprising of filtering through all these search results is the booming silence the President remains to these topics. One of the queries is able to determine the "strength of concern" in for every keyword input where it shows the count of posts each candidate has made with a given topic and President Trump hardly appears on every keyword. Results where there is reoccurring silence are alarming in this kind of program, because the program truly showcases how polarizing the candidates are or the President.

#### Conclusion

The Tweetrumper application allows users to gain a view on how candidates are effectively using social media today to express their mantra. The program gauges the strength of that mantra and displays it to the users in a fashion they can assess and make sound decisions as they hopefully hit the ballot box in the upcoming 2020 primaries and elections. The purpose of Tweetrumper is that most people follow few candidates and their interests can get pulled into a filter bubble (another topic worthy of another paper) where their worldview is closed even though social media has connected society more than ever before. There is hope that the application can influence people to make clear judgements and know what their voting for and what they're getting as a result in the upcoming elections.

#### Future Work

The Tweetrumper application has room to grow in many areas of design and implementation. Amongst the list of them are the following: transform the application into a progressive web application via the Google guidelines otherwise known as Google Lighthouse[2] and allowing the program to live update/schedule updating tweets. These improvements stand out the most, but there can be others.

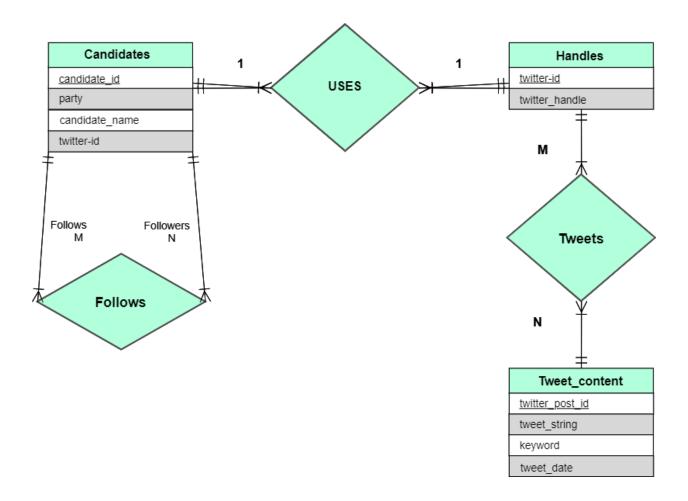
Upgrading to a progressive web application (PWA) is an undertaking that is often unseen. A progressive web app is an installable program that is to look and feel like an app that you get from the mobile ecosystem [3] or installed on the desktop. The program must be reliable and fast, including during and after initial setup. A PWA employs several units (at least one) called service workers [3]. These

service workers ensure that state is not lost if the user pans away from the application. The workers also aid against application crashes by acting as threads for the program, so if one goes down, another service worker comes online. Along with service workers is an integrated cache that self-manages itself [3]. The cache ensures browsed data retention and data remains durable throughout application use, even in offline use. A project like Tweetrumper should accept the mandates of these capabilities because the application is light and should be portable. Once these improvements are in place, it will be more accessible than apps from ecosystems as it's obtained from the browser.

Another improvement would be to allow live-scheduled or scheduled Twitter updates from the candidates, where Tweet updates from candidates are retroactive. An idea that falls synonymous in that regard is having updated candidate statuses when the candidate postpones or drops out of running. Scheduled updates would ensure data integrity and reliability. Currently the database is either in a "static" state of the last snapshot from a Twitter data pull or in a new state from the latest data pull. Neither of these suffice in complete design and breaks the good will of having a database. A secondary live-update condition is to report the status of running candidates in real time. Candidates can drop out of the running or come in at any time, however, it must be limited by a forecast of success.

### References

- [1] "Next.js." Learn Getting Started / Next.js, Feb. 2020, nextjs.org/learn/basics/getting-started.
- [2] "Lighthouse | Tools for Web Developers | Google Developers." Google, Google, 2020, developers.google.com/web/tools/lighthouse.
- [3] Richard, Sam, and Pete LePage. "Progressive Web Apps." Web.dev, Google Developers, 2020, web.dev/progressive-web-apps/.



### Appendix B: Normalization Proof to BCNF

For the candidates relation, the following functional dependencies are outlined:

```
Fd.1 candidate_id → {candidate_name, party, twitter_id}
Fd.2 {candidate_id, candidate_name} → {party, twitter_id}
Fd.3 {candidate_id, candidate_name, party} → twitter_id
Fd.4 {candidate_id, twitter_id} → {candidate_name, party}
```

Since the determinant Fd.1 is a proper subset of the determinants in Fd.2 and Fd.3 and Fd.1 covers all relational attributes, Fd.2 and Fd.3 cannot be considered candidate keys, hence cannot be considered primary keys for BCNF. Fd.1 and Fd.4 are the remaining candidate keys, but since we seek a minimal superkey, Fd.1 has the least number of attributes that satisfy all columns, therefore qualifies for primary key. Nonetheless, the determinant of Fd.1 is also a proper subset of the determinant of Fd.4, hence Fd.1 is the remaining candidate key. Since one candidate key is a determinant of all columns in the candidates relation, candidates is proven BCNF.

For the handles relation, the following function dependencies are provided below:

```
Fd.1 twitter_id → twitter_handle
```

Since the determinant of Fd.1 is a minimal superkey and is the only candidate key for the handles relation, handles is proven BCNF.

For the tweets relation, the following functional dependencies are outlined:

```
Fd.1 twitter_id \rightarrow twitter_post_id
```

Since the determinant of Fd.1 is a minimal superkey and is the only candidate key for the tweets relation, tweets is proven BCNF.

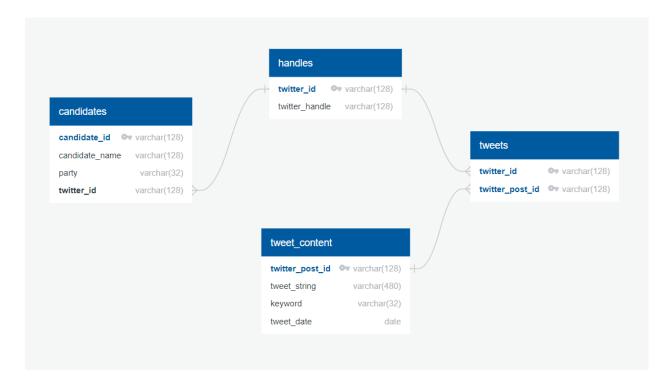
Lastly, for the tweet\_content relation, the following function dependencies are stated:

```
Fd.1 twitter_post_id → {tweet_string, keyword, tweet_date}
Fd.2 {twitter_post_id, tweet_string} → {keyword, tweet_date}
Fd.3 {twitter_post_id, keyword, tweet_string} → tweet_date
Fd.4 {tweet_string, keyword} → tweet_date
Fd.5 keyword → tweet_date
```

Since the determinant of Fd. 1 is a proper subset of the determinants of Fd.2 and Fd.3, and Fd.1 covers all relational attributes uniquely, then the determinants of Fd.2 and Fd.3 cannot be candidate keys. Fd.4 and Fd.5 do not map to both twitter\_post\_id and hence cannot be a super key of the relation, so consequently

cannot be a candidate key for the relationship as well. This leaves the determinant of Fd.1 to be the only remaining candidate key and the only remaining minimal superkey. Since it is the only determinant of that nature for the relation, tweet\_content is proven BCNF.

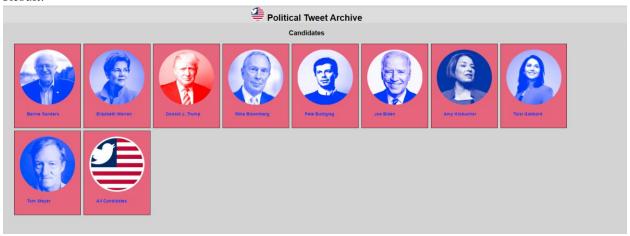
# Appendix C: Relational Schema Diagram



Appendix D: Functional SQL Queries (pictures and description for each)

Query 1: Selects all entries from the table of candidates and orders by the unique candidate ID.

#### Result:



Query 2: Selects the candidate name, keyword, tweet date, and content of tweets made by all candidates that contain a certain keyword in the tweet body, ordered by candidate name.

Result:



### **Tweets from All Candidates**

Search for a Keyword workers Subm

Candidate Name	Keyword	Tweet Date	Post Content	
Amy Klobuchar	workers	2020-02- 16T08:00:00.000Z	Unions help all workers fight for better wages and workplaces. America needs more unions—not fewer. I'm the granddaughter of a miner and the daughter of a union teacher and union newspaperman. I will always defend the right to organize.	
Amy Klobuchar	workers	2020-02- 18T08:00:00.000Z	Nevada: If you're ready to make childcare and long-term care more affordable, If you're ready to take on Big Pharma and stand up for workers' rights, join our campaign. We can do this. https://t.co/4q3egltSgn https://t.co/1zNFi27Oyx	
Amy Klobuchar	workers	2020-02- 19T08:00:00.000Z	I stand in solidarity with the Station Casinos workers fighting for a fair contract in Las Vegas at The Palms. When unions are strong, America is strong. I will continue to fight for workers' rights alongside @culinary226, @teamsters, and everyone united in this effort. https://t.co/0E8SKo2kKW	
Amy Klobuchar	workers	2020-02- 22T08:00:00.000Z	Visited with @Culinary226 workers at Mandalay Bay here in Las Vegas this morning! Culinary has helped over 18,000 immigrants become American citizens. Their work is remarkable and what the American Dream is all about. https://t.co/At25iTzwyB	
Bernie Sanders	workers	2020-02- 24T08:00:00.000Z	I stand with Denver airport workers who organized with @SEIU. Workers stood together, won raises, and voted to join @SEIU105. I hope to see @United honor their choice. https://t.co/5tAvEixrJo	
Bernie Sanders	workers	2020-02- 24T08:00:00.000Z	We do not need a nominee who profits from private equity vultures and the destruction of the planet. We need someone who is willing to stand up for workers and will stop the greed of the fossil fuel industry. Unlike Mr. Bloomberg, as president, that's exactly what I will do. https://t.co/Qsz8gRAxZ8	
Bernie Sanders	workers	2020-02- 29T08:00:00.000Z	This is disgraceful. All workers deserve the right to bargain and strike for better wages and benefits. To Janet Napolitano and @UCSC: stop this outrageous union busting and negotiate in good faith. https://t.co/oaQGTovOdW	
Elizabeth Warren	workers	2020-02- 19T08:00:00.000Z	More than half of hourly workers get their schedules with less than a week's notice & Damp; hours can fluctuate 40-70% each mont That makes it nearly impossible to go to school, get child care, & Damp; pay the bills. That's why @RosaDeLauro & Damp; I have the #SchedulesThatWork Act. https://t.co/cwFiWHzv8Z	
Elizabeth Warren	workers	2020-02- 19T08:00:00.000Z	My #SchedulesThatWork Act would set some ground rules for retail, fast food, & Description of the retail of the re	
Elizabeth Warren	workers	2020-02- 20T08:00:00.000Z	2/3 of part-time retail workers want full-time work, but corporations keep them part-time to skimp on wages & Defits. My Part-Time Worker Bill of Rights would make big companies offer available hrs to their part-time workers before hiring new employees. https://t.co/CAzXuZMST7	

Query 3: Selects all tweets from a certain candidate that contain a certain keyword, ordered by tweet date.

```
const candidates_info = await db.query(escape`

SELECT *

FROM tweet_content

JOIN tweets

ON tweet_content.twitter_post_id = tweets.twitter_post_id

WHERE tweets.twitter_id = ${req.query.id} AND

keyword = ${req.query.keyword}

ORDER BY tweet_date

`);

res.status(200).json({candidates_info});
```

#### Result:



#### Tweets from Elizabeth Warren

Search for a Keyword our violence

Submi

Post ID	Keyword	Tweet Date	Post Content
1222884145191407616	gun violence	2020-01- 30T08:00:00.000Z	My big, bold Gun Violence Prevention and Community Safety Act with @RepHankJohnson combines & Duilds upon Congress's best common-sense gun safety legislation for one comprehensive bill to protect our kids & Durkon with the safer. #EndGunViolenceBill
1222884146646786048	gun violence	2020-01- 30T08:00:00.000Z	I'm grateful that our #EndGunViolenceBill has the support of the @NAACP, @GiffordsCourage, @AMarch4OurLives, @NewtownAction, @stophandguns, & a number of gun violence prevention, civil rights, public health, & community organizations in MA & across the country.
1222932037713133568	gun violence	2020-01- 30T08:00:00.000Z	RT @RepHankJohnson: It's long past time for Congress to treat gun violence in America like the public health crisis that it is. That's why
1222933049979101184	gun violence	2020-01- 30T08:00:00.000Z	RT @MassAGO: Thank you @SenWarren. 100 Americans die from gun violence every day. It's time for Congress to stand up to the gun lobby and
1222933828815204357	gun violence	2020-01- 30T08:00:00.000Z	RT @stophandguns: We are proud of our proven success in gun violence prevention in Massachusetts. We applaud Senator Warren for her support
1222936474213068801	gun violence	2020-01- 30T08:00:00.000Z	RT @NewtownAction: It's long past time for Congress to treat gun violence in America like the public health crisis that it is. That's why w
1222938688046366723	gun violence	2020-01- 30T08:00:00.000Z	RT @SUPGVNetwork: Senator Warren's new Gun Violence Prevention and Community Safety Act is exactly the comprehensive national legislation w
1222982617643069442	gun violence	2020-01- 30T08:00:00.000Z	RT @bradybuzz: Today, @SenWarren & @RepHankJohnson introduced a bill to reduce gun violence. It includes: universal background checks sa
1222984615511625730	gun violence	2020-01- 30T08:00:00.000Z	RT @SenatorMenendez: Joining @SenWarren today to release the Gun Violence Prevention and Community Safety Act —sweeping, commonsense gun leg
1222986009262141456	gun violence	2020-01- 30T08:00:00.000Z	RT @po_murray: Newtown was devastated by gun violence. The incremental approach has failed too many American families. That is why I am thr

### Query 4: Obtains a list of all keywords of issues that candidates tweeted about.

const keywordList = await db.query(escape` SELECT DISTINCT keyword FROM tweet\_content`);

### Result:

# How concerned are candidates about a particular topic?

### Select an issue from the list:

- Russia
- nuclear
- health care
- workers
- gun violence
- weapons
- impeachment
- banks
- economy
- Fed
- Iran
- personal interest
- Bolton
- national security
- corruption
- economic
- sham trial

Query 5: Obtains a count of how many tweets were made by each candidate that contains a particular keyword

```
const keyword = await db.query(escape`
    SELECT C.candidate_name AS "name", TC.Keyword AS "Keyword", COUNT(TC.keyword) AS "strength"
    FROM candidates AS C
    JOIN tweets AS T
    ON C.twitter_id = T.twitter_id
    JOIN tweet_content AS TC
    ON T.twitter_post_id = TC.twitter_post_id
    WHERE TC.keyword = ${req.query.keyword}
    GROUP BY C.candidate_name;
```

#### Result:

Home Issue Tracker About

## Concern Strength for Keyword "health care"

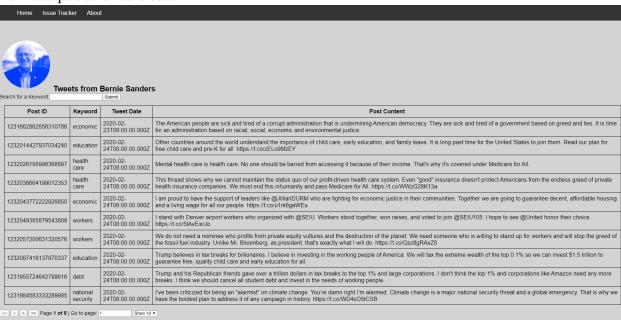
Candidate Name	Tweets with Keyword
Amy Klobuchar	4
Bernie Sanders	10
Elizabeth Warren	2
Joe Biden	9
Mike Bloomberg	9
Pete Buttigieg	5
Tom Steyer	3
Tulsi Gabbard	4

### Appendix E: Web Interface (pictures and description)

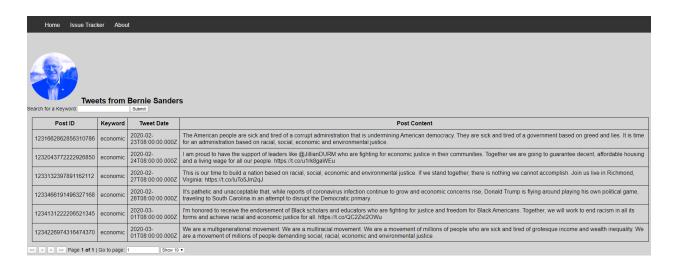
Home Page: Obtains a list of candidates from the database and outputs cards for each candidate that link to their individual page. (Candidate pictures are automatically populated by looking for a file with the name of the candidate. If a new candidate was added to the database it would show up but without a picture)



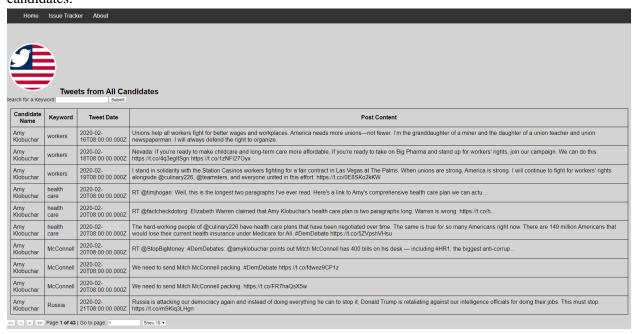
By clicking on an individual candidate, it takes you to a page that is automatically populated by tweets from that particular candidate



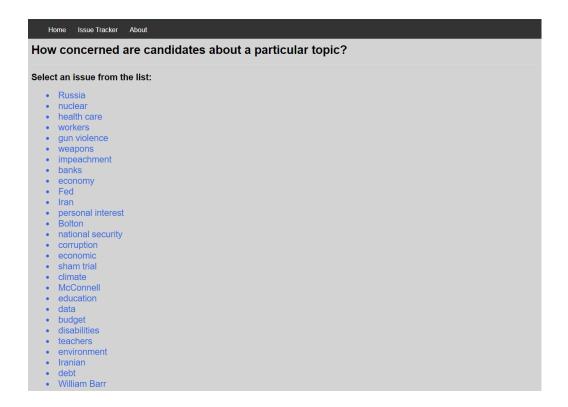
For any candidate, you can filter by only tweets that contain a particular keyword.



By clicking on all candidates, it takes you to a seperate page that displays a list of tweets by all candidates.



By clicking the issue tracker on the navbar, it takes you to a page that lists various issues that candidates are concerned about.



By clicking on a topic from the list, it displays a table that shows how many tweets were made that contain that issue from each candidate

