What's in a Tweet?

```
"id"=>12296272736,
"text"=>
"An early look at Annotations:
http://groups.google.com/group/twitter-api-announce/browse_thread/thread/fa5da2608865453"
"created_at"=>"Fri Apr 16 17:55:46 +0000 2010"
"in_reply_to_user_id"=>nil,
"in_reply_to_screen_name"=>nil,
"in_reply_to_status_id"=>nil,
"favorited"=>false,
"truncated"=>false,
"user"=>
 "id"=>6253282,
 "screen_name"=>"twitterapi",
 "name"=>"Twitter API",
 "description"=>
 "The Real Twitter API. I tweet about API changes, service issues and
 happily answer questions about Twitter and our API. Don't get an answer? It's on my website.",
 "url"=>"http://apiwiki.twitter.com"
 "location"=>"San Francisco, CA",
 "profile_background_color"=>"c1dfee",
 "profile_background_image_url"=>
 "http://a3.twimg.com/profile_background_images/59931895/twitterapi-background-new.png",
 "profile_background_tile"=>false,
 "profile_image_url"=>"http://a3.twimg.com/profile_images/689684365/api_normal.png",
 "profile_link_color"=>"0000ff",
 "profile_sidebar_border_color"=>"87bc44",
 "profile_sidebar_fill_color"=>"e0ff92",
 "profile_text_color"=>"000000",
 "created_at"=>"Wed May 23 06:01:13 +0000 2007",
 "contributors_enabled"=>true,
 "favourites_count"=>1,
 "statuses_count"=>1628,
 "friends_count"=>13,
 "time_zone"=>"Pacific Time (US & Canada)",
 "utc_offset"=>-28800,
 "lang"=>"en",
 "protected"=>false,
 "followers_count"=>100581
 "geo_enabled"=>true,
 "notifications"=>false,
 "following"=>true,
 "verified"=>true,
"contributors"=>[3191321],
"geo"=>nil,
"coordinates"=>nil,
"place"=>
 {"id"=>"2b6ff8c22edd9576",
 "url"=>"http://api.twitter.com/1/geo/id/2b6ff8c22edd9576.json"
 "name"=>"SoMa",
 "full_name"=>"SoMa, San Francisco",
 "place_type"=>"neighborhood",
 "country_code"=>"US",
 "country"=>"The United States of America",
 "bounding_box"=>
 {"coordinates"=>
  [[[-122.42284884, 37.76893497],
    [-122.3964, 37.76893497],
    [-122.3964, 37.78752897],
    [-122.42284884, 37.78752897]]],
  "type"=>"Polygon"}},
"source"=>"web"
```

Annotations:

- **DEPRECATED**
- The author's user ID.
- The author of the tweet. This embedded object can get out of sync.
- Number of tweets this user has.
- Whether this user has geo enabled (http://bit.ly/4pFY77).
- The geo tag on this tweet in GeoJSON (http://bit.ly/b8L1Cp).

- The tweet's unique ID. These IDs are roughly sorted & developers should treat them as opaque (http://bit.ly/dCkppc).
- Text of the tweet. Consecutive duplicate tweets are rejected. 140 character max (http://bit.ly/4ud3he).
- Tweet's creation date.
- The ID of an existing tweet that this tweet is in reply to. Won't be set unless the author of the referenced tweet is mentioned.
- The screen name & user ID of replied to tweet author.
- Truncated to 140 characters. Only possible from SMS.
- The author's user name.
- The author's screen name.
- The author's biography.
- The author's URL.
- The author's "location". This is a free-form text field, and there are no guarantees on whether it can be geocoded.
- Rendering information for the author. Colors are encoded in hex values (RGB).
- The creation date for this account.
- Whether this account has contributors enabled (http://bit.ly/50npuu).
- Number of favorites this user has.
- Number of users this user is following.
- The timezone and offset (in seconds) for this user.
- The user's selected language.
- Whether this user is protected or not. If the user is protected, then this tweet is not visible except to "friends".
- Number of followers for this user.
- **DEPRECATED** in this context
- Whether this user has a verified badge.
- The contributors' (if any) user IDs (http://bit.ly/50npuu).
- **DEPRECATED**
- The place ID
- The URL to fetch a detailed polygon for this place
- The printable names of this place
- The type of this place - can be a "neighborhood" or "city"
- The place associated with this Tweet (http://bit.ly/b8L1Cp).
- The country this place is in
- The application that sent this tweet
- The bounding box for this place

Data Features

- Tweet Unique ID
- Tweet Location
- **Tweet Text**
- Twitter Username
- Retweet Count
- Follower Count
- Favorites Count
- Hashtags
- Mentions
- Etc.

900,000K+ Tweets

Objectives
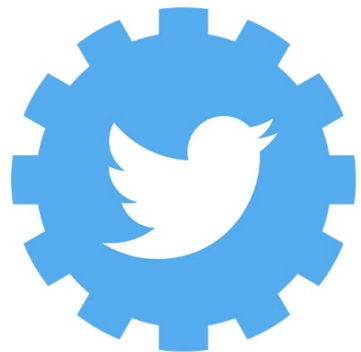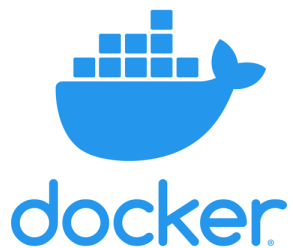
Methodology

Text Processing & Dimensionality Reduction

150,000+ words

4 Topics

543 words

Sentiment Analysis

100 Most Common Words
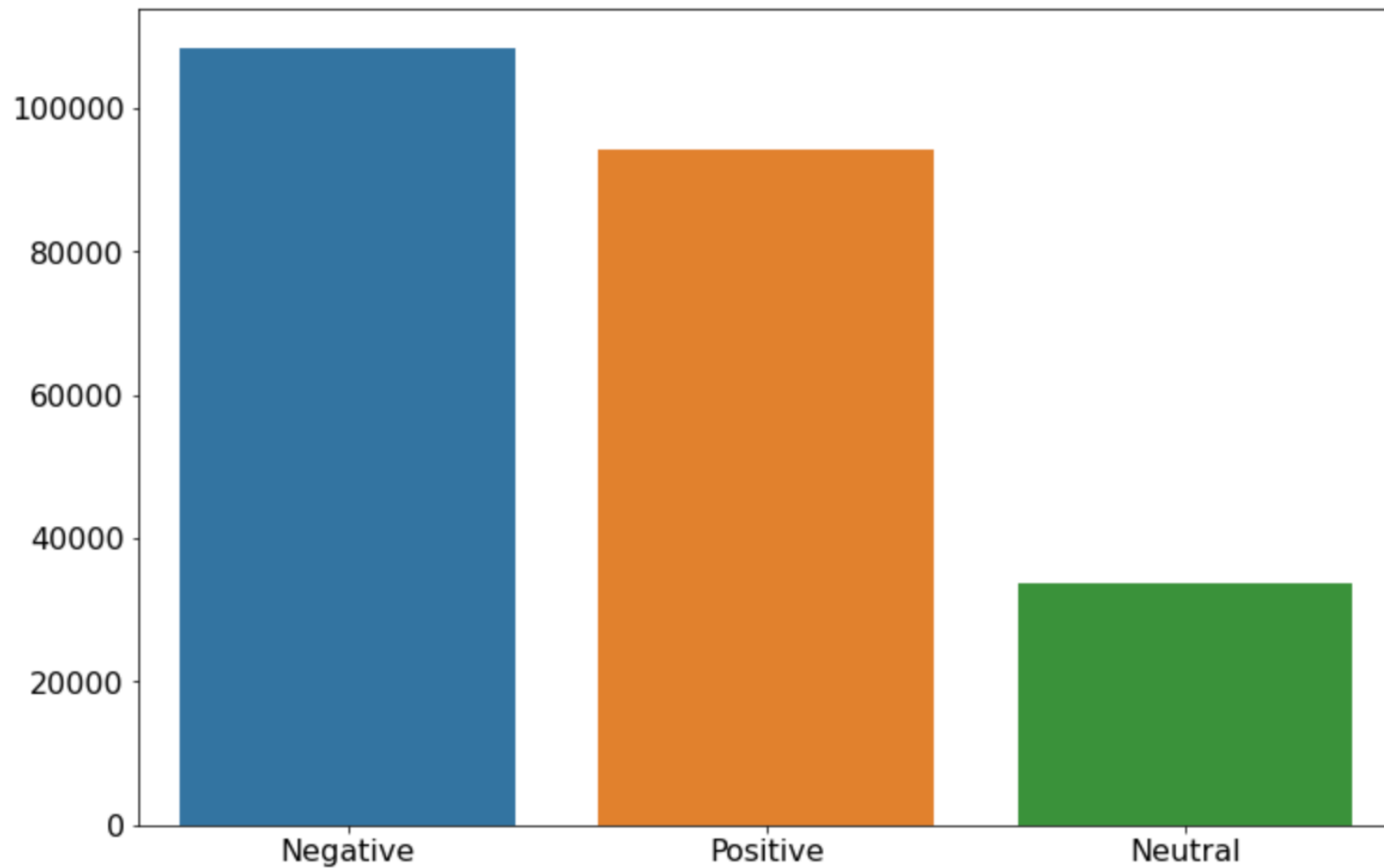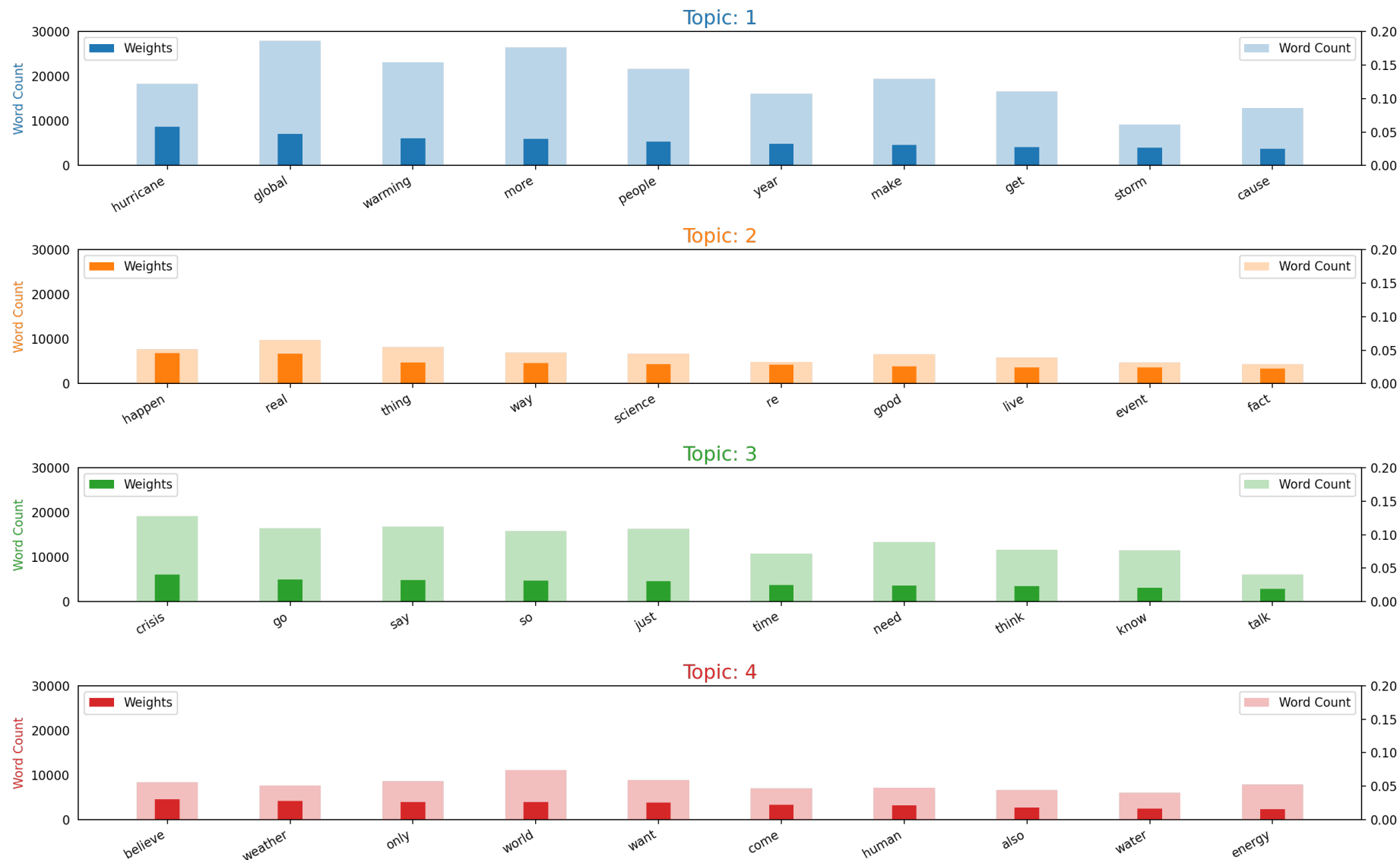
Findings: Topics

Conclusion

Topic 1:
Global warming is caused by people and results in "natural disasters"

Topic 2:
Climate Change is scientifically real and is happening

Topic 3:
Although Climate Change is a crisis, more time and discussion is required:

Topic 4:
Possibly, Climate Change is a believe to be caused by natural world weather:

Key Topic Terms

- Hurricane
- Global Warming
- People
- Cause

- Happen
- Real
- Science
- Fact

- Crisis
- Need
- Time
- Think

- Believe
- Weather
- World
- Human

Potential Future Work

Further Data Engineering:

- AWS
- MongoDB
- Streamlit and/or Heroku

Further Analysis:

- LSA & NMF
- Social Network Analysis
- Gelphi + NetworkX