
#mood

— The Emoji Context Project —

Caroline Locke, Bryce Bible, Ryan Flint, Chris Muncey, Manny Bhidya

Project Overview and Objective

- Emojis are ubiquitous
- There is a certain sentiment implied by them
 - Sentiment can (and frequently does) change based on context
 - Sometimes there is no context
 - Sometimes the context lies outside of the scope of the post
- Goal of this project was to develop a method to determine underlying emoji meaning and the mood/connotation associated with that emoji

Data Collection

- Developed a scraper using python
 - Tweepy API
 - Filtered by the english language and selected emojis
- Ran two bots for 8 hours each
 - Ran them at separate times to avoid overlap in collected tweets
 - First bot looked for 25 most common emojis
 - Second bot looked for next 25 most common
- Collected around 2.5M tweets total
 - Around 592 MB total in case anyone was curious

Data Processing

- **Standardizing Data Access**

- Defining Functions for processing raw scrapped output
- Functions for loading a sample of tweets from processed files
- Generators for delivering tweets.

- **ETL (Extract - Transform - Load)**

- Detecting and fixing encoding issues resulting from various scraping methods
- Calculating Sentiment
- Emoji Aggregate Metadata
- Emoji one-hot-list.

- **Emoji visualizations**

- Updated matplotlib backend to support loading macbook emojis
- optimizing data processing workflows for better visualizations

Data Processing

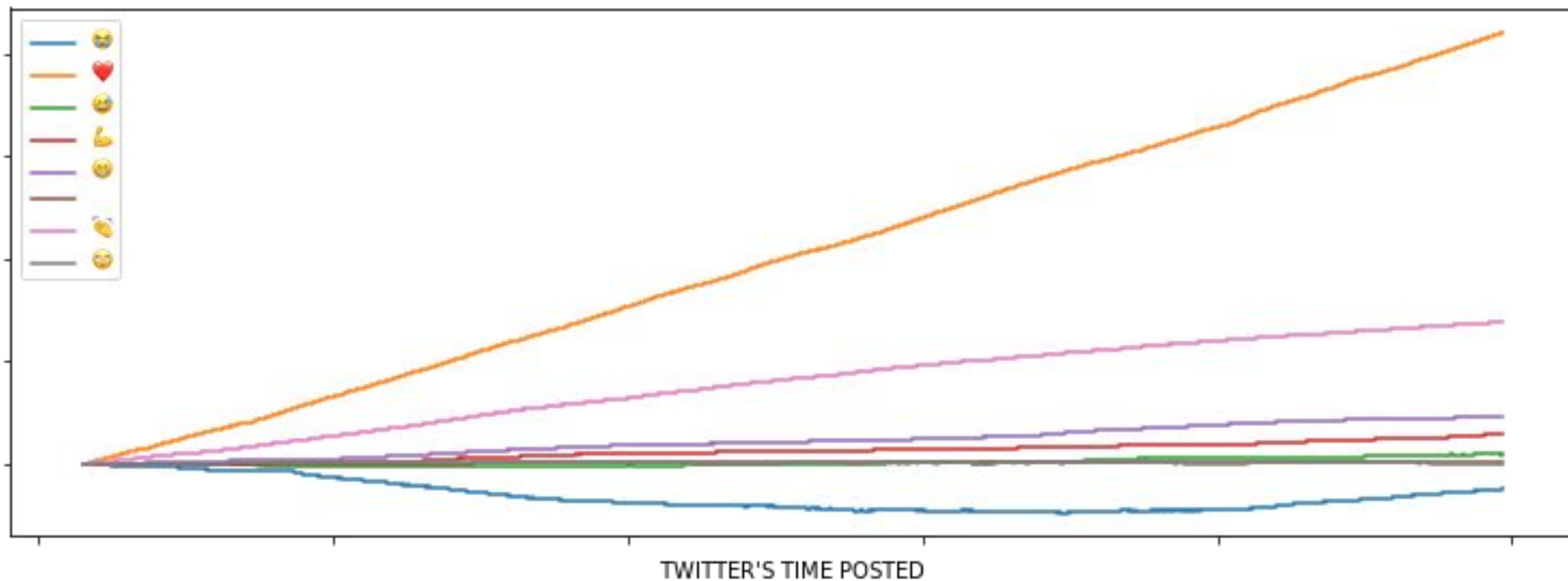
- **Sentiment Analysis**

- Collecting Sentiment Dataset (**Positive** vs **Negative**) **Opinion Lexicon**
- Creating Sentiment Algorithm (**TF-IDF** and **Cosine similarity**)
- Building Emoji features **count mean std min -25% 50% 75% max**
- Building **chronological** and **cumulative** measures for **emojis**
- **Clustering Emojis** based on **features** based

- **Visualizations**

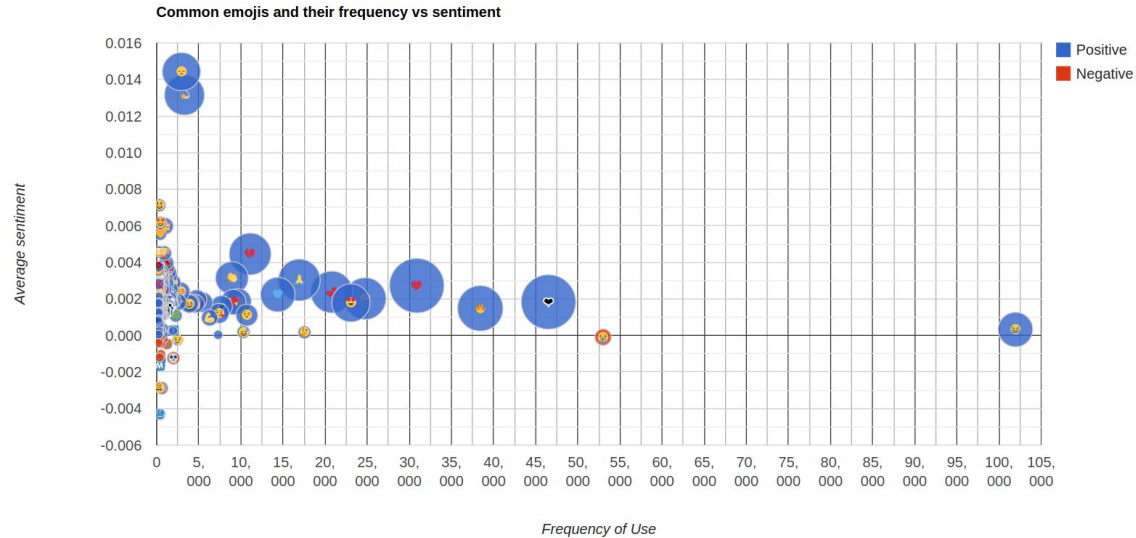
- Visualising **chronological** and **cumulative** measures for emojis
- Visualising **Emoji features** in **bar charts**
- Visualising the **top 25 Emoji features** in **bar charts**
- Visualising the **bottom 25 Emoji features** in **bar charts**
- Visualising the **top Emoji features** in **pie charts**
- Visualising **Emoji Clusters** based on **features** processed by **K Nearest Neighbors** and **Principal Component Analysis**

Sentiment Analysis

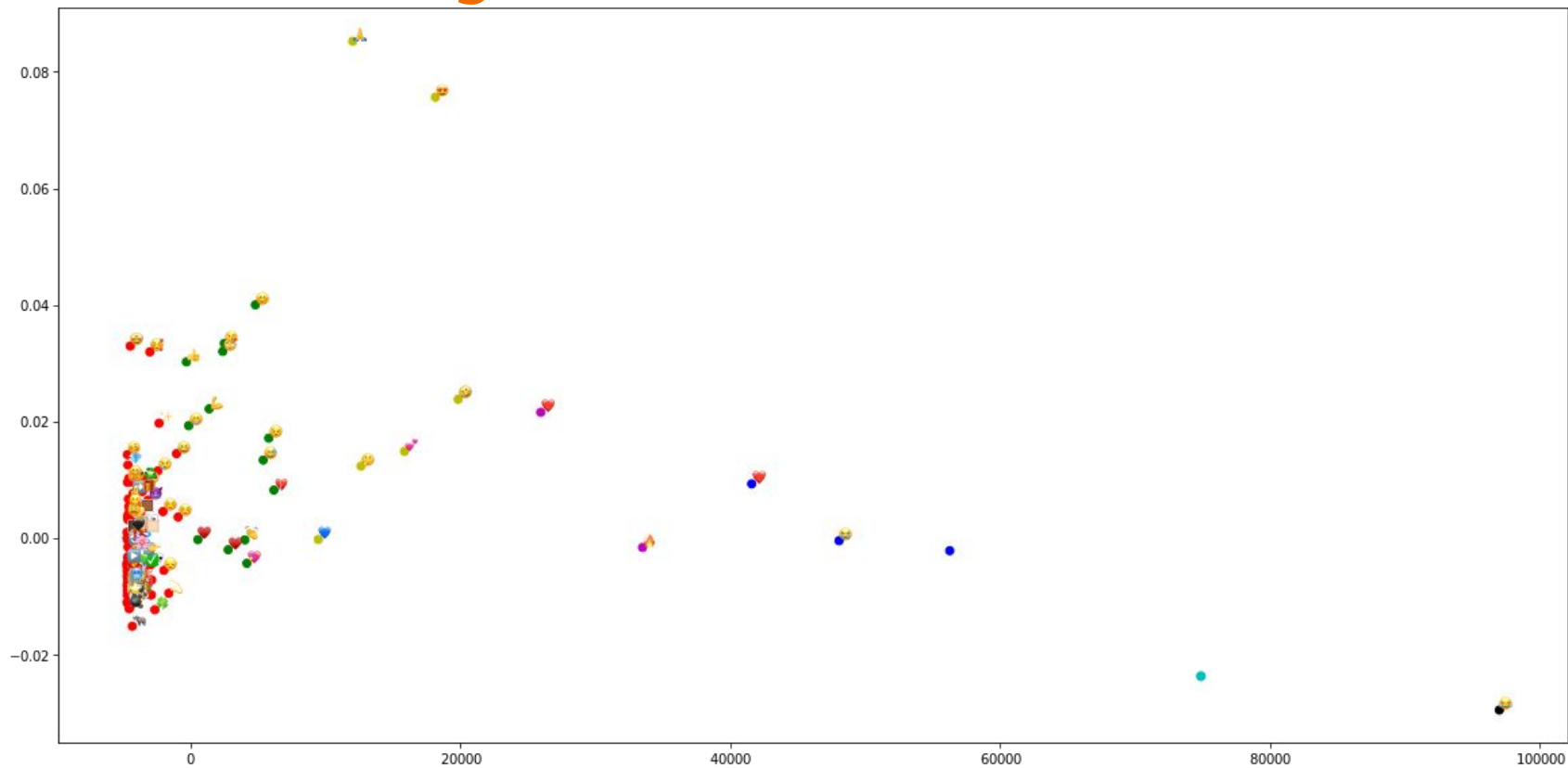


Sentiment Bubble Chart

Interactive page:

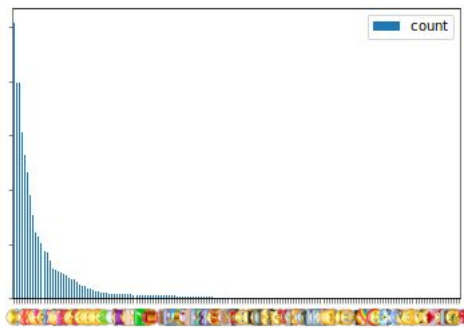


Data Clustering

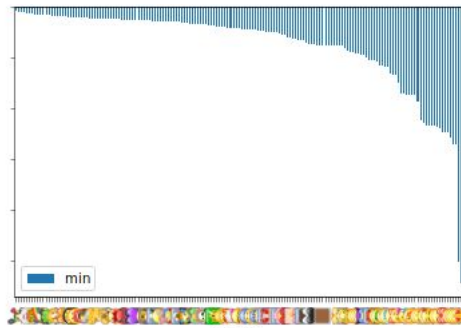


Data Clustering

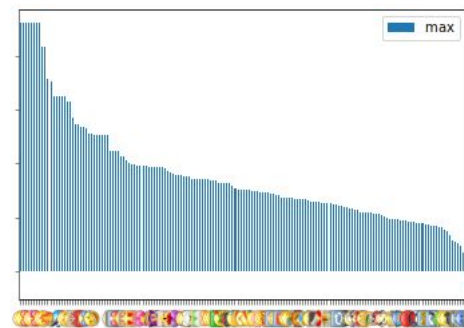
```
{ 'b': [ '😭', '❤️', '' ],  
  'c': [ '' ],  
  'g': [ '😄', '💪', '😁', '👋', '😞', '👍', '😘', '😊', '😊', '❤️',  
        '❤️', '❤️', '💔' ],  
  'k': [ '😄' ],  
  'm': [ '❤️', '🔥' ],  
  'r': [ '😭', '🔴', '😊', '🎉', '😁', '🔄', '🍑', '💯', '😄', '👁️', '😄', '😊',  
        'U', 'A', '😊', 'P', '🟪', '🟪', '🌈', '📺', '🍆',  
        '💧', '😬', 'S', 'white heart', '😬', '🎥', '😞',  
        '❤️', '💀', '❤️', '➡️', '🦋', '🟠', '📞', '🚩', '\\', 'u', '2',  
        '0', 'd', '♀️', '🚒', '🌸', '😞', '😭', '🚀', '👑', '👉',  
        '🟪', '🐱', '✨', '🦋', '💋', '😄', '🟪', '👑', '👉',  
        '✅', '♻️', 'G', '👹', '18', '🔥', '🔄', '❤️', '🏆', '😞',  
        '🎁', '👉', '😞', '😄', 'N', '😞', '📷', '🐰', '👉', '😞',  
        '🎧', '💎', '🕒', '🐱', '😞', '😄', '☀️', '🔗', '💰', '❤️', '📌',  
        'B', 'R', '🔄', '😞', '😞', '😞', '🍀', '👉', '💰', '❤️', '🕶️',  
        '🐱', '🎵', '🔄', '😞', '😞', '👉', '📺', 'M', '!!!',  
        'E', '⚠️', '❤️', '♂️', '😞', '🌸', '✅', '🌹', '🍒', '📺', '❌',  
        '🏀', '😄', '❤️', '👻', '😞', '😞', '💣', '🌟', '🍁', '👻', '🌟', '❌',  
        '🔔', '⭐', '📢', '▶️', 'D', '❤️', '💔', '❤️' ] }  
  'y': [ '🙏', '😄', '🙏', '😞', '❤️', '💔' ] }
```



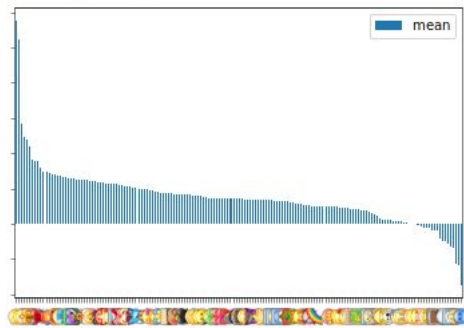
Emojis



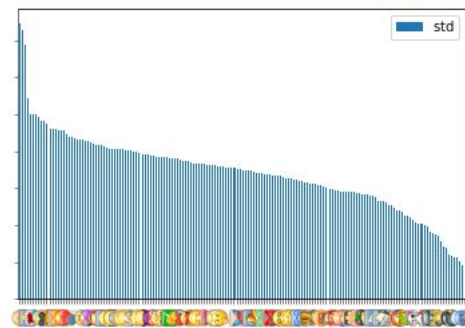
Emojis



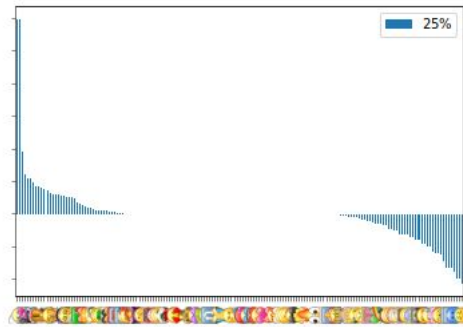
Emojis



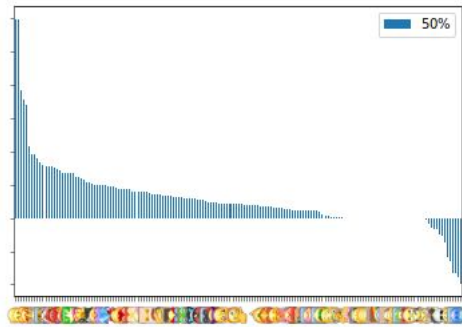
Emojis



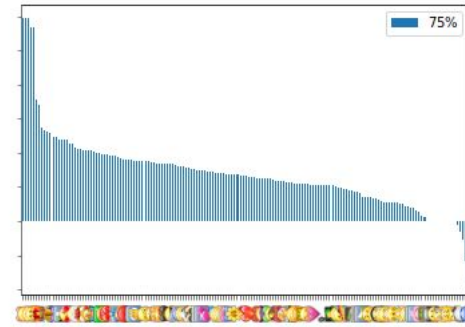
Emojis



Emojis

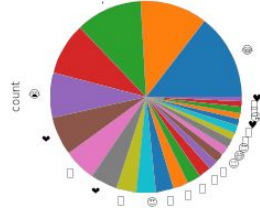
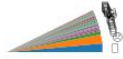


Emojis

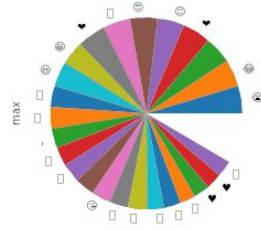
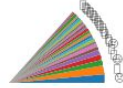


Emojis

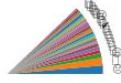
25%



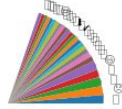
50%



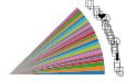
mean

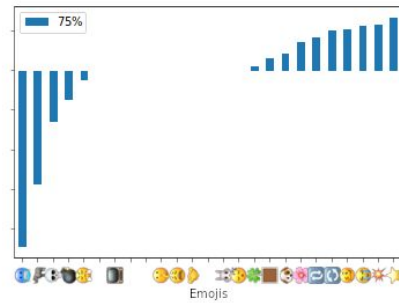
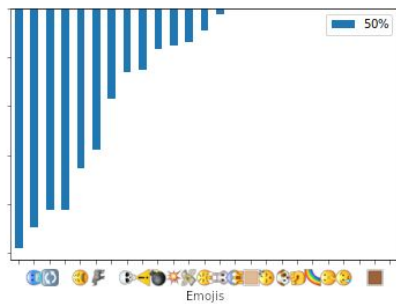
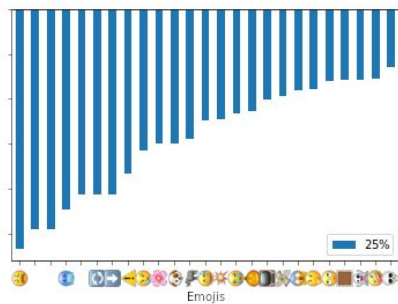
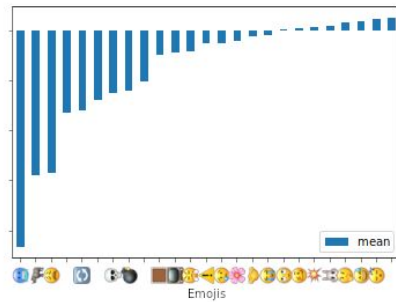
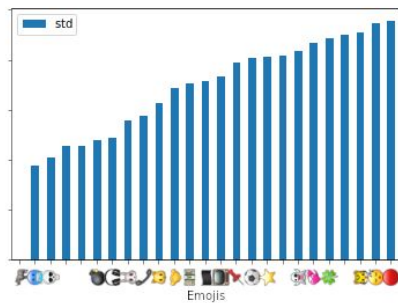
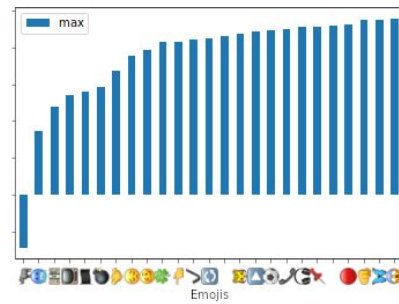
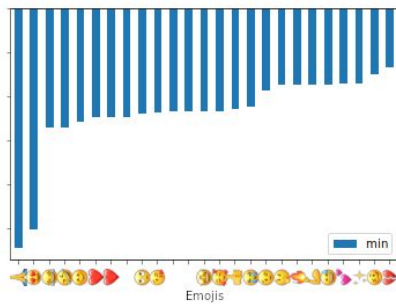
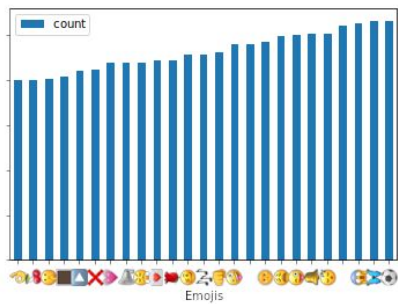


75%



std





Future Work

- Word clouds and word lists associated with each emoji currently available
- Generate text strings from emojis and vice versa -- possibly web display
- Compare results in other language to see how connotation and usage changes
- Possible overlap with Twitter Happiness Index ([Hedonometer](#))