Project 2
Twitter Emoji Analysis
Team 7

Sean Tidd
Timothy Mickle
Zachary Minnich

Objectives ✅

- Filter out emojis from historical and streamed tweets
- Answer analytical questions involving emojis with Spark
- Discover possible trends between emojis and tweets
- Properly graph and label results
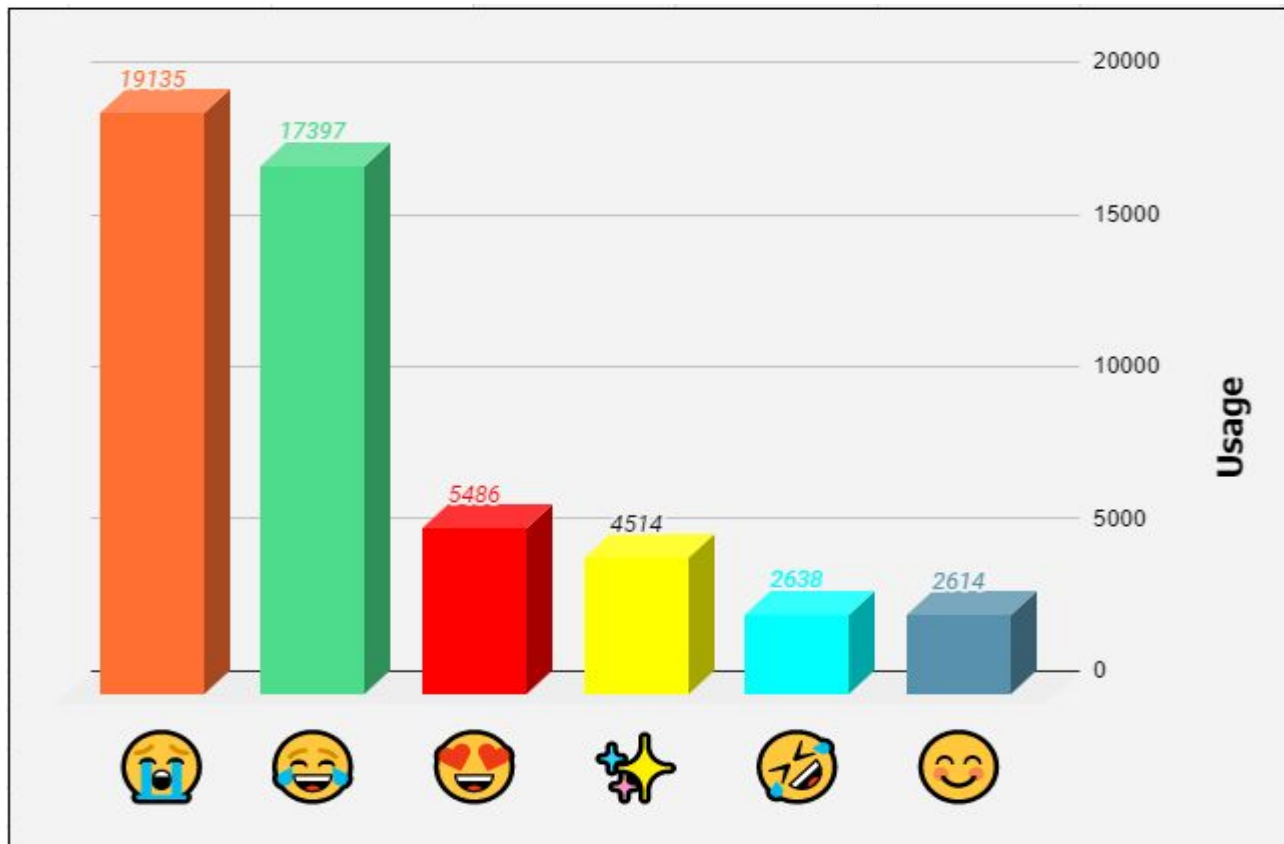
Emoji Structure 🛠️

———

- Represented in unicode with variable lengths and unicode ranges
- Each individual unicode is a separate char
- 😀 -> \uD83D\uDE00, 🦁 -> \uD83E\uDD81,

  🧛 -> \uD83E\uDDDB\u200D\u2642\uFE0F, etc.

- Skin tone variations are represented as base emoji unicode with the skin tone unicode appended at the end
- ✋ -> \u270B, ✋ -> \u270B\uD83C\uDFFB, ✋ -> \u270B\uD83C\uDFFF

How Emojis Are Separated From Tweets ✂️

———

1) Break the string(Tweet) up into chars
2) Filter out all chars that are words, spaces, or punctuation
3) Rebuild emojis char by char by checking if it validates one of the 4 emoji regex
4) If it isn't valid, remove the currently appended char and add the emoji to the list
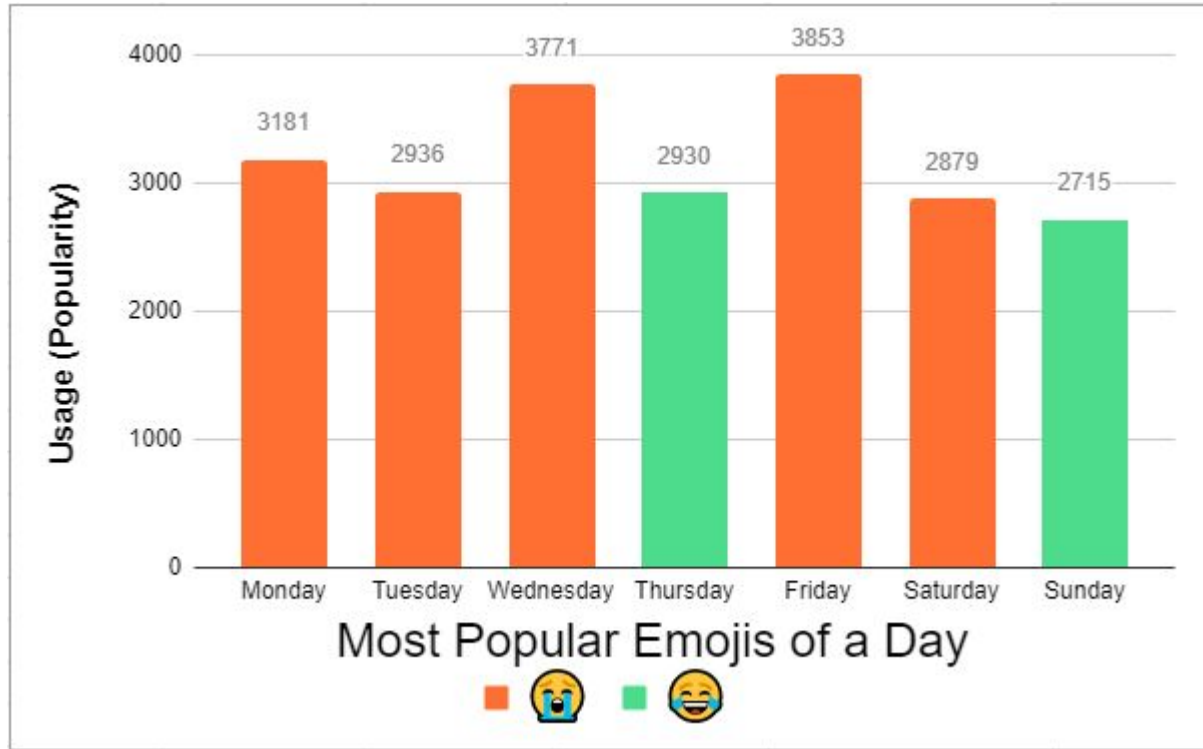5) Restart with the removed char and repeat

Assumptions: Emojis with a gender unicode variation (🧗) attached are not accepted in Java and are just assumed to be two separate emojis: base emoji followed by the gender symbol emoji (🧗♂)
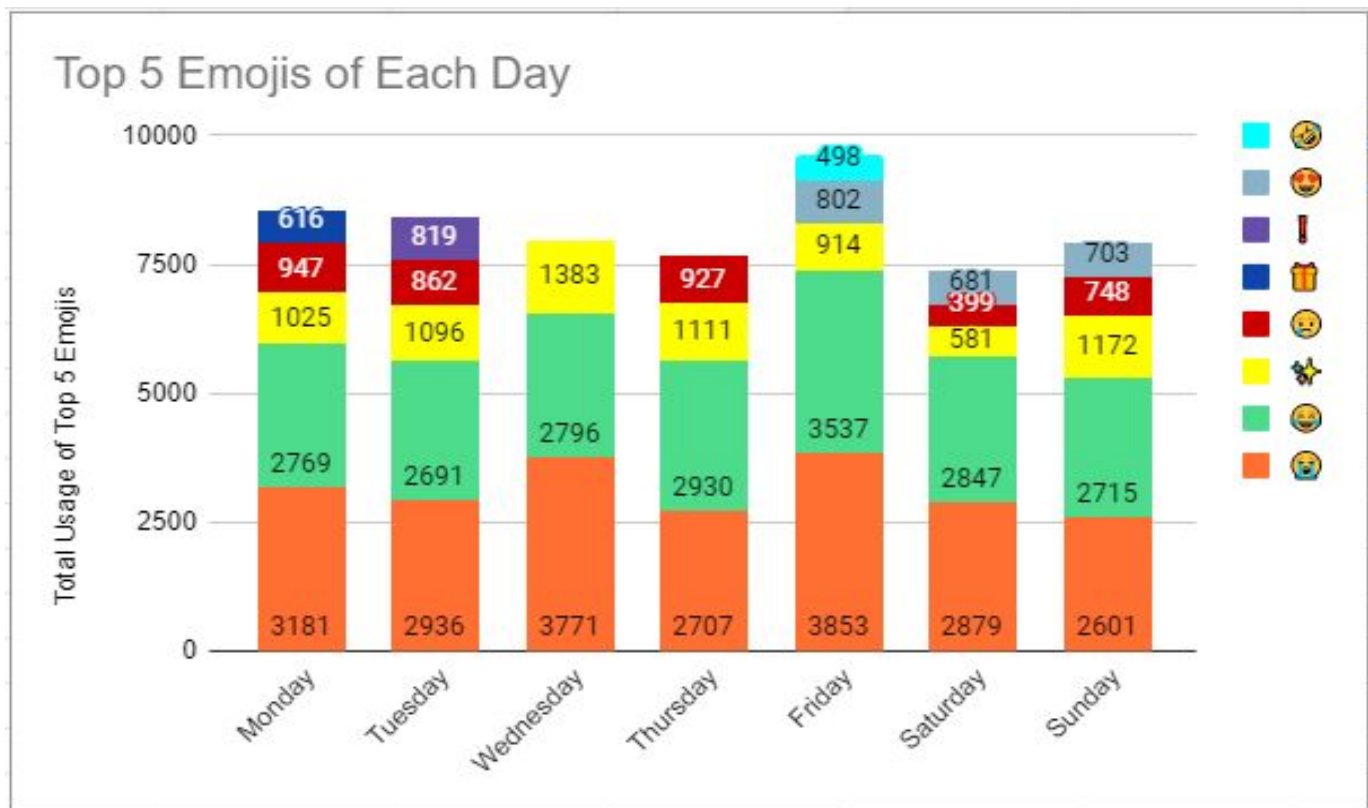
Most Popular Emoji During a Week (33,600 Tweets) 📅
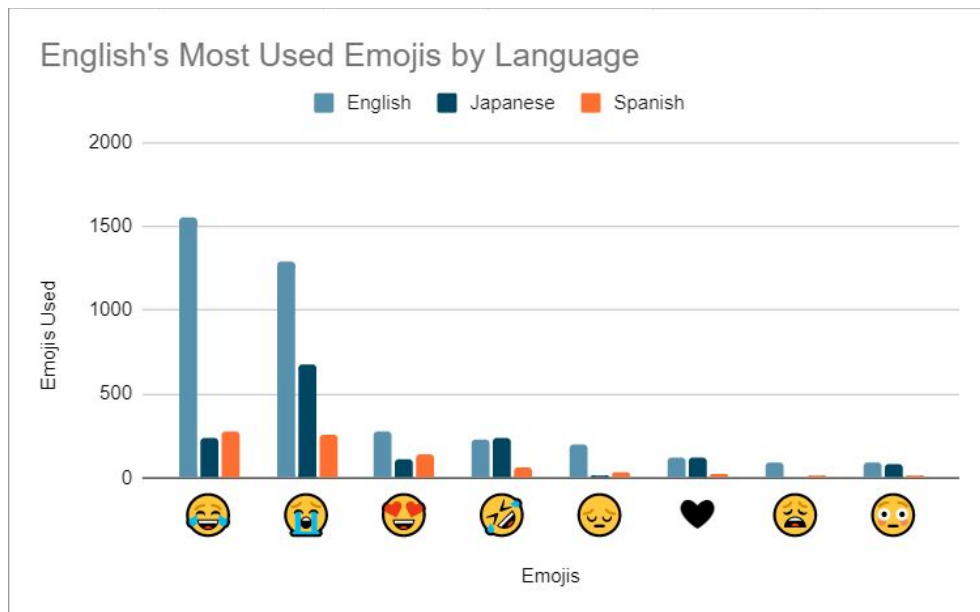
Most Popular Emoji During a Day (4,800 Tweets daily) ☀️
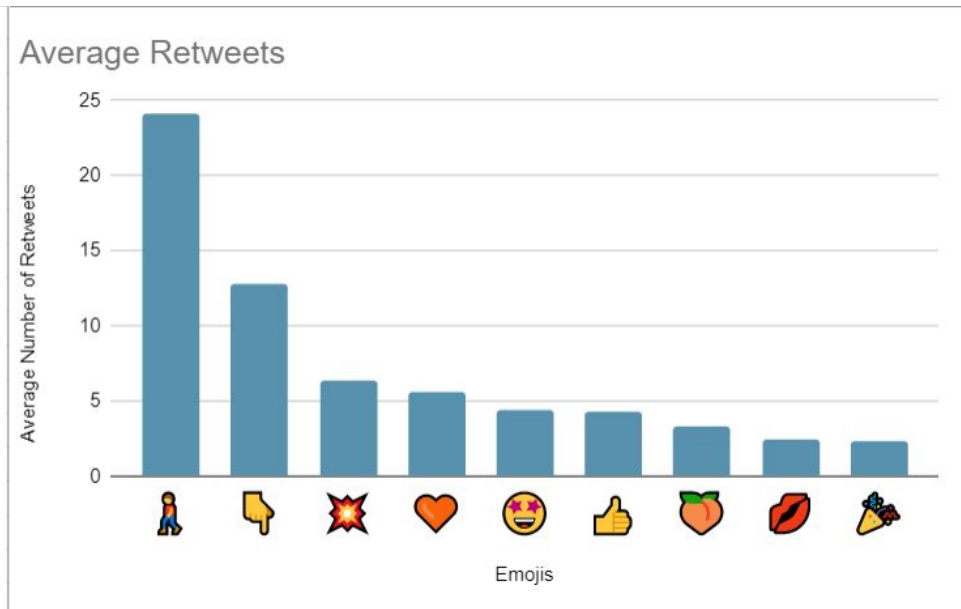
– – –

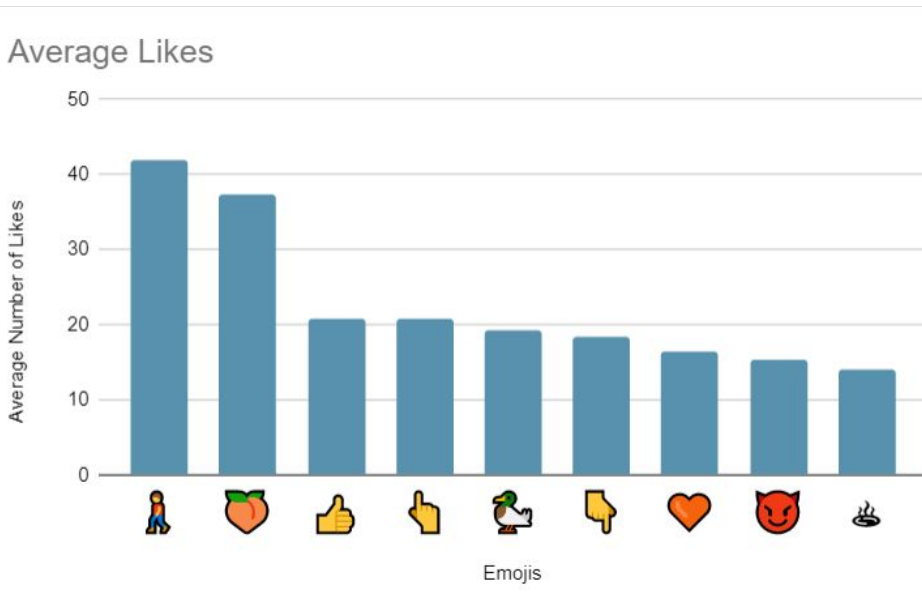# Top 5 Emojis of Each Day (4,800 Tweets daily) 🕐

Most Popular Emoji in Different Languages 🇯🇵

—— ——

Assumptions: This sample of data over a week is representative of all weeks
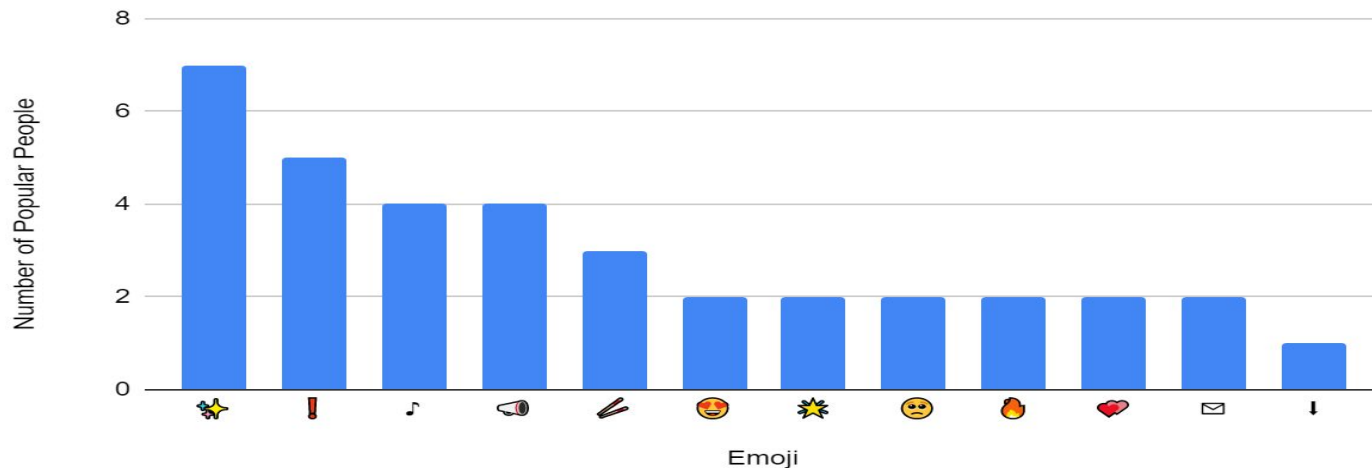
Most Liked and Retweeted Emojis 👍

Assumptions: It is only counted if it is used more than 10 times in a week.
Filters out retweets to only get original posts (as retweets carry over into
subsequent retweets)
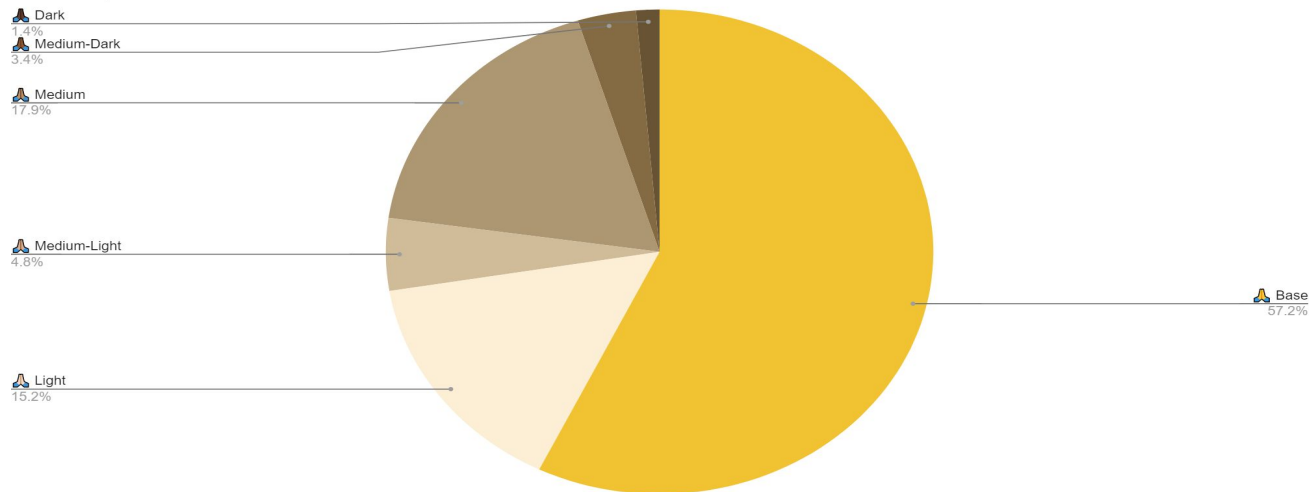
Most Used Emojis by Popular People 🥰

Assumptions: Assumed that you had to have at least 10,000 followers to be considered "popular/famous". Higher values result in less data because there are more accounts with less followers than accounts with more than 10,000 that are tweeting regularly

Most Popular Skin Tone for "Praying Hands" Emoji 🙏

———

Percentage of 🙏Variation Use

Stream of 25,000 live Tweets

🙏 Dark
1.4%

🙏 Medium-Dark
3.4%

🙏 Medium
17.9%

🙏 Medium-Light
4.8%

🙏 Base
57.2%

🙏 Light
15.2%

Assumptions: Everyone chose the variation according to how they identify otherwise they use base

Questions?

Sources

———

GitHub:

- https://github.com/srtidd23/Team7_Project2.git

Emoji Unicode:

- https://www.iemoji.com/#?category=animals-nature&version=36&theme=appl&skintone=default

Twitter:

- www.twitter.com