# Steelers Pre-game Tweets

By Mike Scheibel

**Data set source:**

This is a collection of tweets given the search terms #HereWeGo or #Steelers. This was pulled before the game on 11/9/20 when the weekend started. I wanted to analyze things like location of Steelers followers, and what players get mentioned most often.

**Data set description:**

The data was scraped from Twitter using Tweepy and TextBlob. This data is the tweets using the Steelers most used hashtags. I then wrote the data to a csv file: ‘20201122\_070537\_steelers\_tweets.csv’. I collected the following data from those who tweeted with those hashtags : screen\_name, description, location, friends\_count, followers\_count, statuses\_count, created\_at, retweet\_count and entities[‘hashtags’].

**Descriptive statistics:**

Leah Remini:

Tokens: 831459

Unique Tokens: 4412

Average Token Length: 5.643

Lexical Diversity: 0.005

Top 20 words:

Game, Steelers, Officially, Back, Jacksonville, Follow, Fan, Away, Play, Sunday, Win, Love, Jags, Re-tweet, Give, Making, JuJu, Catch, Help, Awesome

**Interesting questions:**

How many of the users who tweeted are from Pittsburgh?

Are there any celebrity followers?

Are there any other common hashtags used?

Who is the most popular player on the Steelers?

What is the sentiment going into this game?

**Approach for a single question:**

What is the sentiment going into this game?

My approach would be to pull out the text column from the csv file. I would then tokenize and deal with the HTML markup, hashtags, and emojis. I would then run a polarity sentiment analysis on the text, and calculate the percentage of positive, neutral, and negative tweets.