# EE239AS PROJECT 3

## Popularity Prediction on Twitter

Intro here

The average number of tweets per hour, follower of users posting the tweets, and number of retweets are listed below for each of the six hashtags.

#superbowl

Average number of tweets (per hour): 1399

Average number of retweets (per hour): 3341

Average number of followers of users posting tweets: 3341

#nfl

Average number of tweets (per hour): 279

Average number of retweets (per hour): 429

Average number of followers of users posting tweets: 429

#gopatriots

Average number of tweets (per hour): 38

Average number of retweets (per hour): 53

Average number of followers of users posting tweets: 53

The histograms for each of the hashtags are depicted in the figures below. From the plots we see that users were extremely active around the time of the super bowl and posted more tweets then. We see a peak in the number of tweets at one point during the superbowl game. The hashtag #superbowl saw a staggering 272,322 tweets between 5-6 PM on February 2, 2015 which was during the game.

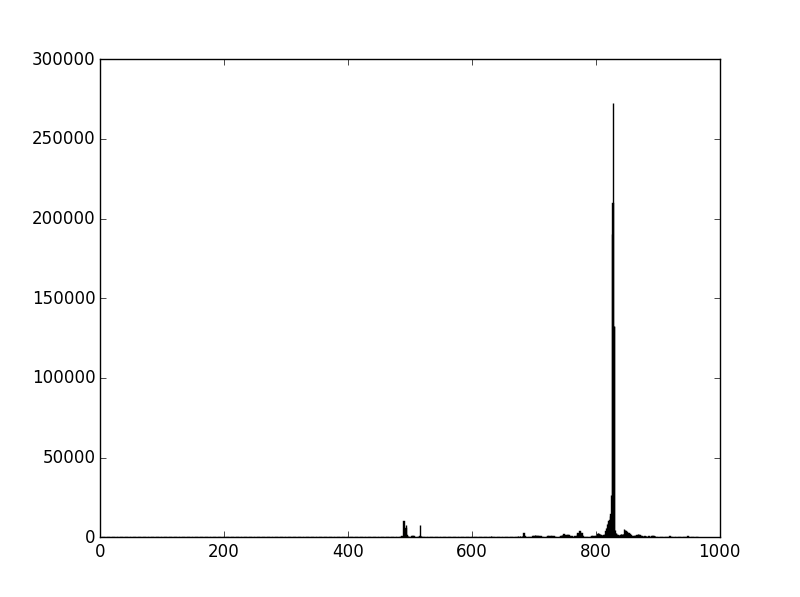


Figure 1: Histogram of number of tweets in hour for #superbowl

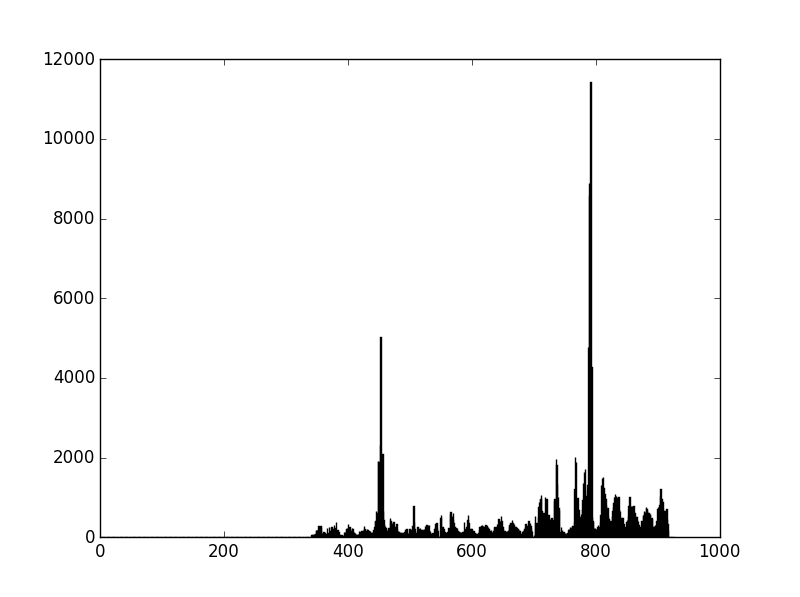


Figure 2: Histogram of number of tweets in hour for #nfl

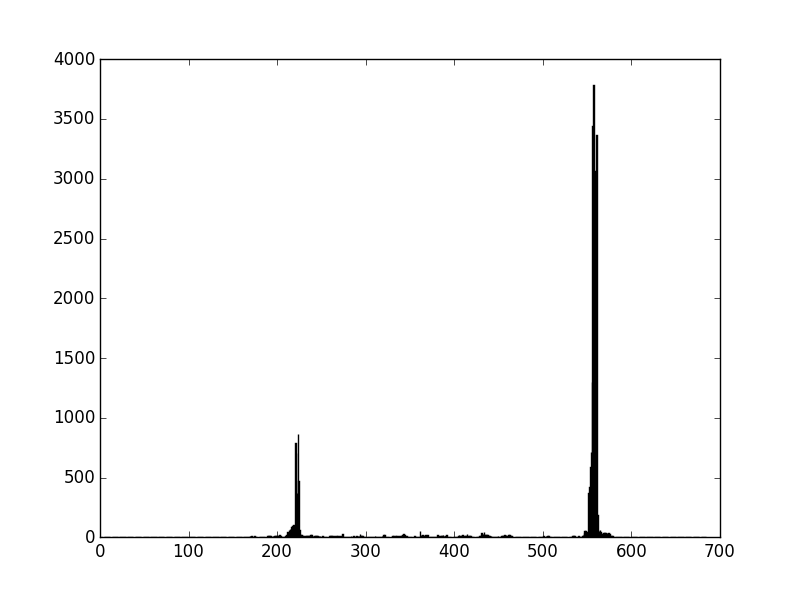


Figure 3: Histogram of number of tweets in hour for #gopatriots

For the regression analysis we used the ordinary least squares model from the statsmodels library. We created a matrix of the five features specified and passed that into the model. For the predicted value, we set yi = num\_tweetsi+1. For the last item in y, we just set it to yn-1. We pass that into the OLS algorithm to generate the model. An example of our regression result is illustrated below.