2.98: Sports Technology: Engineering & Innovation
Individual Deliverable
Kevin Ma

1. You are given a database with tweets that include at least one of the hashtags #Islanders and #Isles, tweeted on Monday's 2/10 matchday. Examine the time relationship between the tweets and the match development and draw conclusions. How many different users used the hashtags? For the 100 users that tweeted the most, draw some stats from their social media profile (followers, tweets, language, account age).
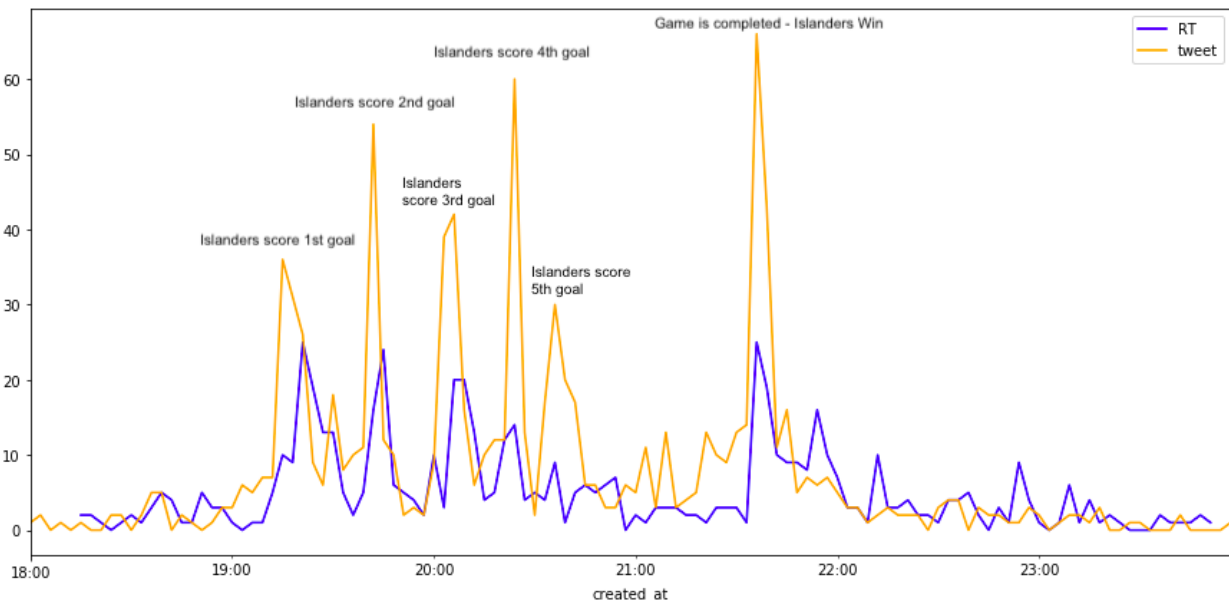

Figure 1. Twitter activity during the Feb 10th Islanders vs. Capitals game (tweets w/ #Islanders or #Isles hashtags)

Figure 1, a time series plot of twitter activity during the February 10th Islanders vs. Capitals game, shows that tweet volume corresponds to events that happen during the game in near real-time. Over the course of the game, there were six major spikes in tweet volume - the first five spikes occurred each time the Islanders scored a goal, while the sixth spike occurred once the game finished and the Islanders won.

If you look at the tweets themselves during the spikes, the language revolves heavily around the player that scored the goal. For instance, Anthony Beauvillier, who scored the Islanders' first two goals, was among the most frequently mentioned names during the first two spikes.

	Minimum	Maximum	Mean	Standard Deviation
Tweets	2	44	6.44	5.97
User Followers	2	671,702	9,862.02	67,029.47
User Following	0	26,650	1,186.07	2,721
User Total Tweets	80	1,374,876	47,100.86	138,935.20

Figure 2. Summary Statistics for top 100 tweeters during Feb 10th Islanders vs. Capitals game

	user_name	user_followers	user_following	user_total_tweets	tweets
0	CordUpTime	3036	1796	125631	44
1	ob1moroney	322	476	85454	27
2	AGrossNewsday	31828	613	74236	23
3	stefen_rosner	181	993	909	20
4	eyesonislesFS	7464	545	28938	19
...	...	...	...	...	...
95	sticktoyourtaco	1578	2487	83159	2
96	OldHockeyDude	55	166	2016	2
97	drv45	19	45	460	2
98	RealBryGuy	213	574	11045	2
99	TronofSeven	3	37	342	2

Figure 3. Top 100 users sending tweets during Feb 10th Islanders vs. Capitals game

In terms of user engagement with the hashtag, there were 849 distinct users that used either the #Islanders or #Isles hashtags throughout February 10th. During the game itself (I estimated the game to start at 7:00pm and end at 9:30pm), 410 different users used these hashtags and sent out a total of 1,013 tweets. Figure 2 shows summary statistics for the top 100 tweeters during the game, while Figure 3 shows each individual user that made up the top 100 tweeters.

2. Brainstorm possible ways of approaching the problem. Since you'll be meeting your sponsor in person on 2/20, use the time to discuss with him your thoughts and what will be your initial approach to the project.

From my perspective, this project has two major goals. The first is to assess fan engagement with the New York Islanders on social media. The second is to build a strategy for the Islanders that will allow the team to improve its engagement with its fans on social media and ultimately build greater awareness using social media. In order to accomplish these two goals, I think it's worth considering the following ideas:

- **Use clustering techniques to gain a better understanding of the Islanders' fan base.** Using clustering techniques on the Islanders' Twitter followers, we can begin to

identify different types of users and build different engagement strategies for each user type depending on their twitter activity. The types of data we'll want to build models on include user account information (age, location, occupation, gender), the types of accounts that they follow, and the types of tweets and retweets that they post.

- **Connect spikes in Twitter activity with real-world events to identify what triggers build the most social media momentum.** We know that in-game events such as goals made by the Islanders spur more Twitter activity, but what other events have similar effects? By identifying the various spikes in Twitter activity over time and using NLP to understand the themes behind those spikes, we can identify sources of content that we should be driving so that Twitter followers and non-followers will want to engage with the Islanders more.

- **Do a comprehensive review of the Islanders' twitter activity relative to its follower growth over time.** Are there any trends in the types of tweets from the Islanders account that is building a larger follower base? If we can track the number of followers over time, we can analyze the Islanders' tweet history and see if there are any patterns that stand out.

- **Analyze what non-followers are saying about the Islanders.** Presumably, the data we currently have is only for Islanders followers. Is there a way we can collect tweets from non-followers as well? We can begin by searching for tweets that mention the Islanders ("Islanders", "Isles") and its players, and use unsupervised machine learning to understand what people are generally saying about the team.