Islanders Team Presentation

Objective: Help the NY Islanders understand their Twitter fanbase & improve engagement

Opportunity

@NYIslanders has **535K Twitter followers** and a very limited understanding of those fans. Increasing fan engagement = increasing long-term revenue

Resources

Our primary resource is **Twitter data** about fan accounts, tweet content, and tracked Twitter interactions.

We are using **BigQuery** to manage the data, **Co-lab** notebooks to write code, and various **clustering algorithms** to analyze the data

Our project has two phases: understanding fans, and recommending an engagement strategy

1) Understand Islanders Fans

1A Calculate summary statistics and input variables for clustering

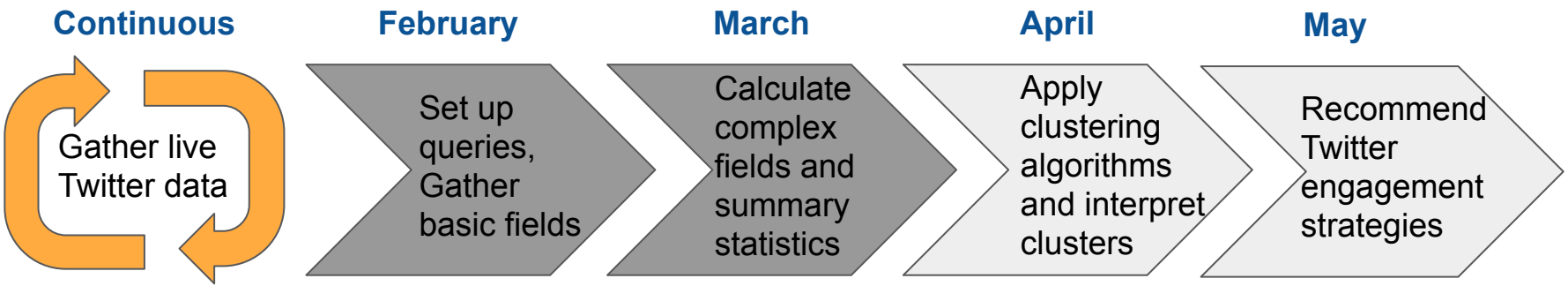
1B Apply clustering algorithms and interpret clusters

2) Recommend engagement strategy

2A Evaluate fan interactions across each cluster

2B Extract insights and recommend content strategy to engage fan clusters

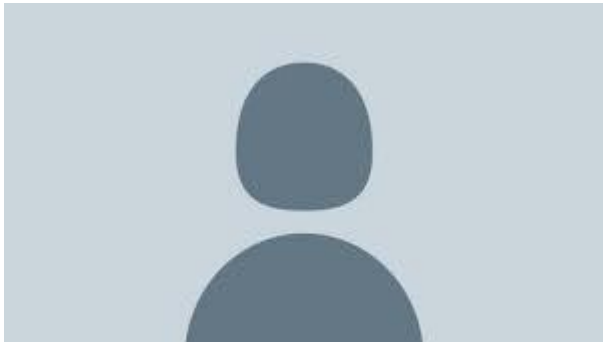
Project Timeline



Initial Findings

Summary Statistics about fanbase on Twitter

Of the Islanders 535,000 followers …



~ 60% have no bio

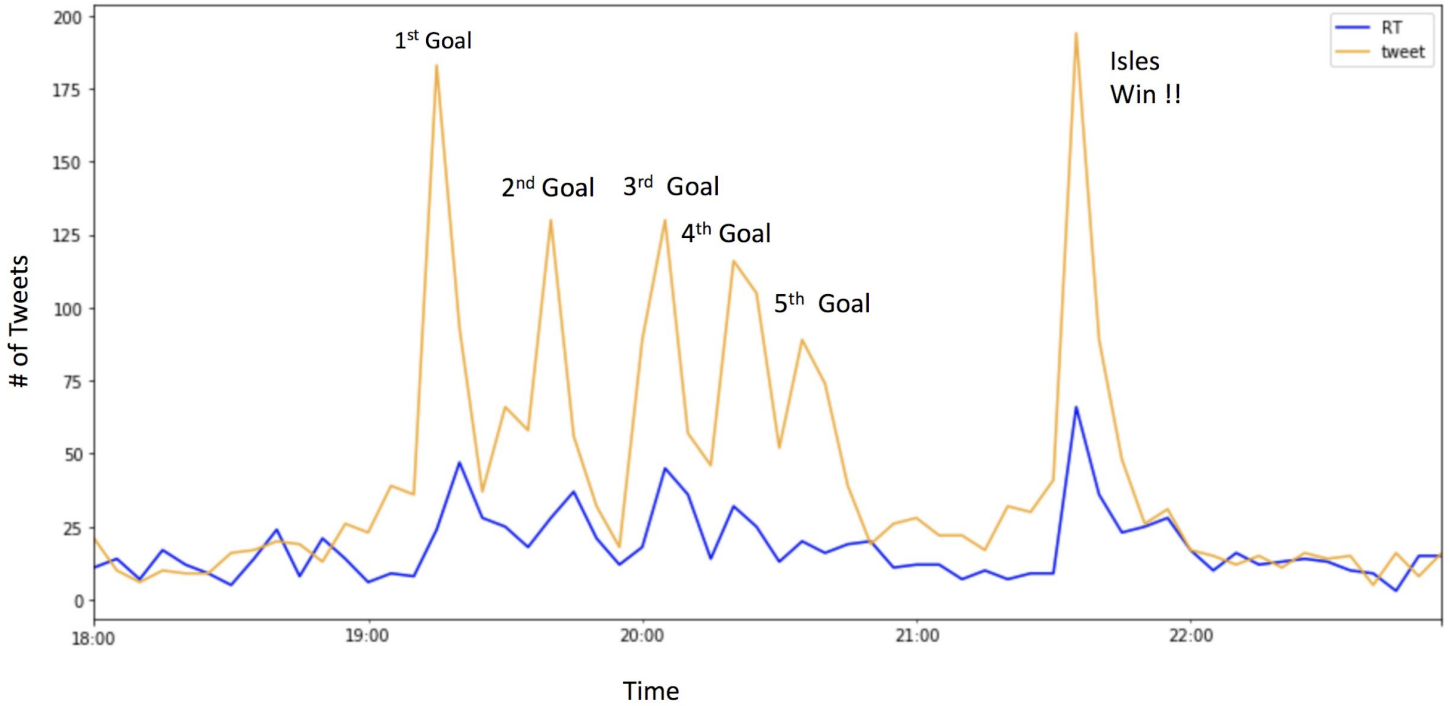
~ 74% have no location

~1,800 or 0.3% have tweeted more than five times about the Islanders, making them avid fans

~ A large number of fans on twitter are anonymous and don't interact with the Islanders account and have very little activity

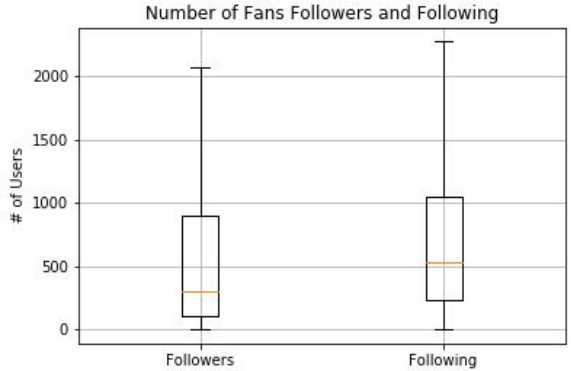
Mid-Game Twitter Activity

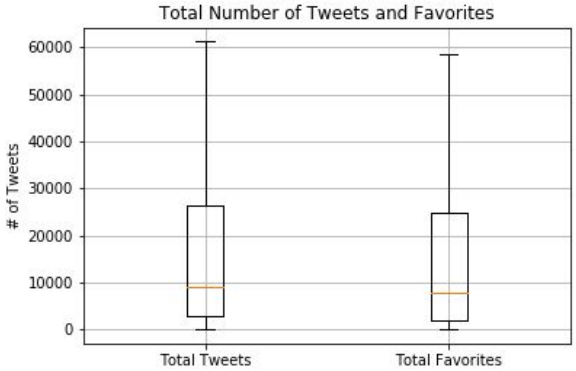
A look at how users interact during the game and what that interaction looks like

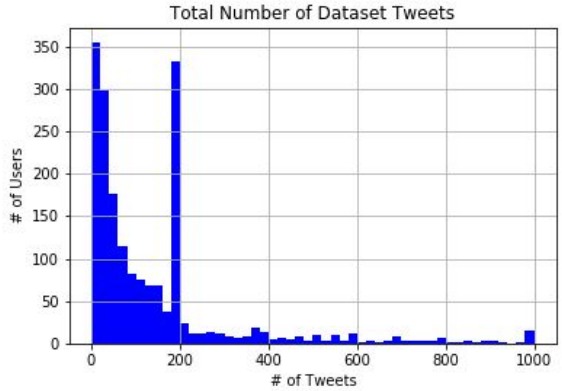


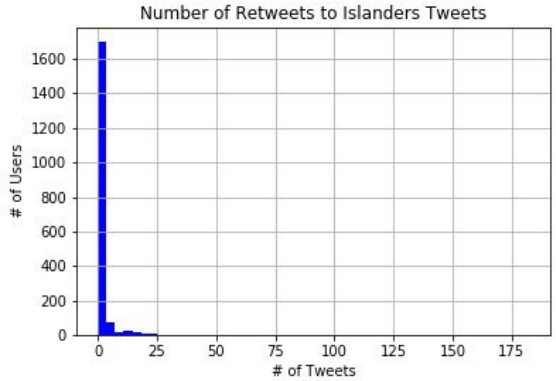
Initial Findings

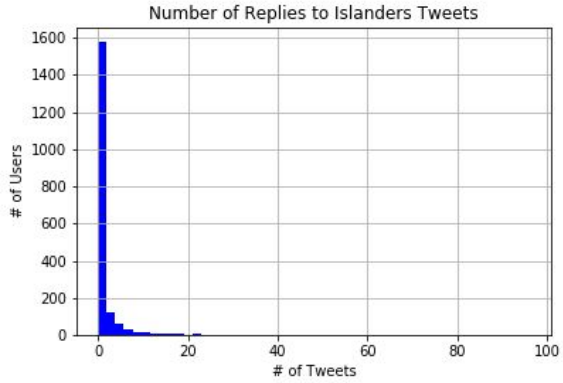
SUMMARY STATISTICS:

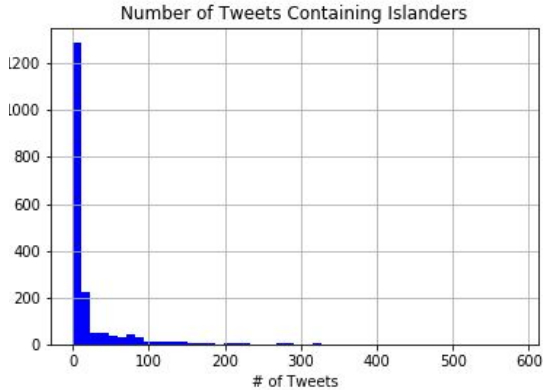






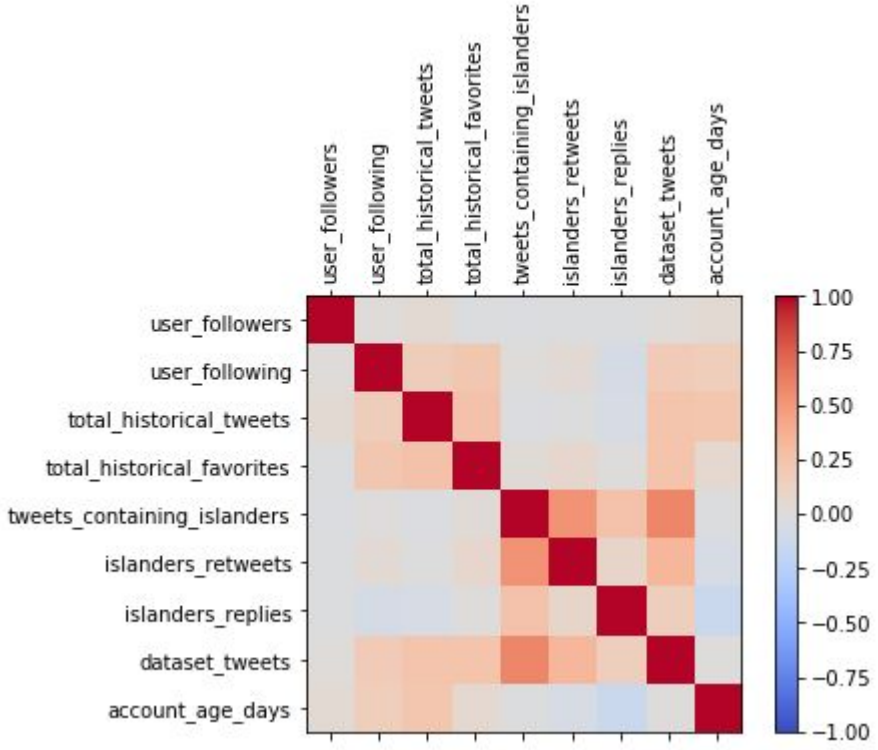






Initial Findings

SUMMARY STATISTICS:



Introduction to K-Means Clustering

- The algorithm begins by randomly placing K centroids (which we must tell the algorithm) and labeling data points based on distance to those centroids.
- Each centroid then moves to the middle of their respective cluster. Following this, the algorithm repeats itself by re-labeling data points based on distance to the new centroids.

- This iterative process completes only once all data points no longer change clusters.

Data for K-Means Clustering

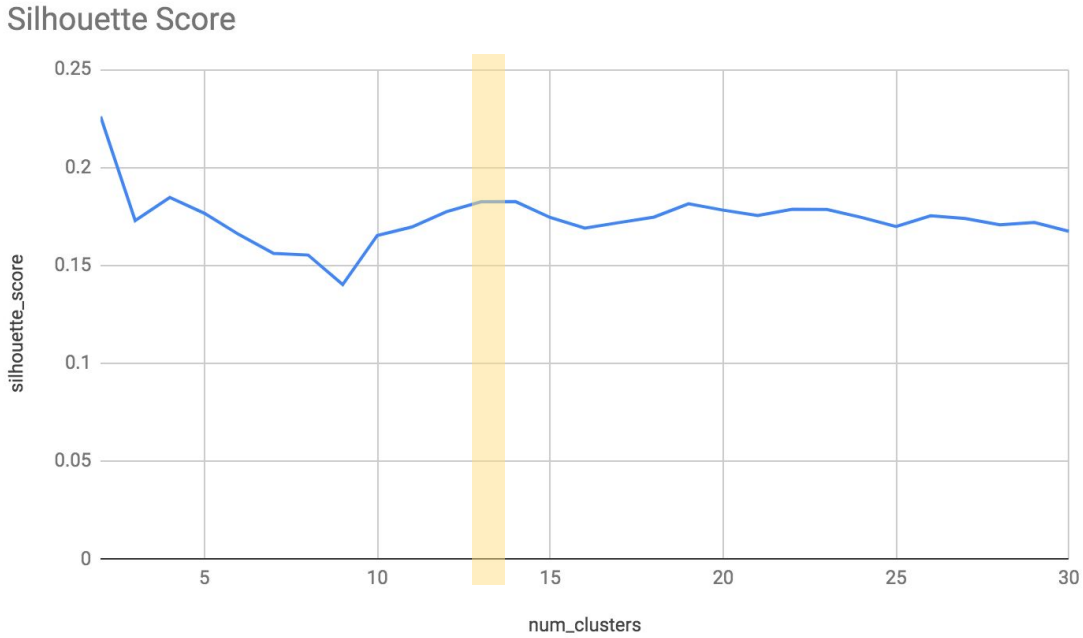
Here is glimpse of the features we have aggregated for the 1,887 fans classified as avid:

handle	user_followers	user_following	total_historical_t	total_historical_f	tweets_containin	islanders_retwee	islanders_replies	dataset_tweets	has_bio	is_private	is_verified	bio_num_charac	has_location	has_profile_pic	account_age_days
hockeyGirl2027	849	295	66905	104423	3	2	0	38	1	0	0	0	0	0	
SmithhBob	343	216	8327	9601	1	0	1	5	1	0	0	0	0	1	3028
369Brain	126	397	569	646	68	5	14	397	1	0	0	0	0	0	
mannyhannies69	66	186	929	5272	29	0	3	327	1	0	0	66	0	1	242
IslanderTed	314	1619	5979	2422	52	0	0	115	1	0	0	140	1	1	2421
IslesGeorgia	246	175	6104	3104	119	3	0	226	1	0	0	0	0	0	
13hayxs	888	685	7652	4057	0	0	0	28	1	0	0	0	0	0	
TParbus	1	54	18	49	9	0	3	10	1	0	0	0	0	0	
bbaeser	14	148	1171	1221	0	0	0	16	1	0	0	51	1	1	451
smn1976	130	1005	7118	2068	8	0	1	11	1	0	0	0	0	0	

We have identified 15 different features for each of the users to use as part of the K-means clustering algorithm

Some of these are binary variables (has_bio, is_private), while others are numerical features we found in the scraped twitter data(followers, total tweets)

We choose the optimal K clusters that maximizes the silhouette score.



The optimal number of clusters is 13, though the K-Means algorithm doesn't appear to do a great job of clustering users (low silhouette scores). We will investigate other clustering techniques (e.g. GMM, hierarchical, etc.) to see if we can build more discernible clusters.

Next steps: discuss initial findings and progress with clustering algorithms

1. **Discuss summary statistics and initial clusters with Islanders sponsor**
    a. Iterate our focus as needed, based on this conversation
2. **Refine clustering models until satisfactory**
    a. Start with k-means clustering, but evolve as needed
3. **Continually interpret clusters**
    a. What intuitive type of fan would each cluster represent?
    b. How are fans in each cluster engaging with Islanders content on Twitter?
    c. What concrete actions could Islanders take to better engage this cluster?