IPL 2018 Fan Engagement analysis from twitter & FB

Balaji Venktesh (11910041) | ISB CBA Practicum 1 | 24-Jun-2018

# Table of contents:

|  |  |  |
| --- | --- | --- |
| **Topic #** | **Topic** | **Page #** |
| 1 | Executive summary | 3 |
| 2 | Introduction | 3 |
| 3 | Motivation for the study | 3 |
| 4 | Data collection | 3 |
| 5 | Tools used | 4 |
| 6 | Data cleaning and preparation | 4 |
| 7 | Analysis | 4 |
| 8 | Conclusion and business recommendations | 8 |
| 9 | Appendix | 9 |
| 10 | Reference | 9 |

# Executive Summary:

Indian Premier League (IPL) is one of the most viewed and followed sports events in India. Fans express their views in social media like Twitter & Facebook. In this project, the tweets that were recorded during the IPL 2018 and the user posts from fans in Facebook official pages are extracted and analyzed for sentiments and cluster fans into different segments. The objective of this study is to see how the fans reacted during this IPL season.

# Introduction:

During the peak cricket season in India from April till end of May, IPL is consistently proving to be a major factor of entrainment for many people in India. During this season, there will be more posts and tweets from the fans across the country expressing their feelings for their team. The purpose behind this study is to analyze the sentiments for different teams/players from tweets/posts that were tweeted under the hashtags.

# Motivation for this study:

As it is always for popularity of players and teams, each fan will support for a team and oppose other teams. There are endless discussions and threads on social media arguing that IPL is scripted especially against CSK. The motivation for this study is to hence capture the fan sentiment of how tweets were posted during the IPL finals match and to compare the sentiment level for different players and tweets from different cities.

# Questions & Hypothesis:

The questions that I would like to test using the data collected are:

1. What is the sentiment level for players from both the team? Which player had the highest sentiment score and which player had the lowest sentiment score?
2. Can we cluster, or group fans based on their expressions?
3. Commonly used bi-grams, trending tweets and their expressions

# Data collection:

1. Tweets collected from twitter using different hashtags like IPL, CSK, RCB etc... through out the season and post the IPL finals.
2. Facebook Posts from fans from each team’s official pages were scrapped.

# Tools used:

1. Python:
   1. Tweepy for gathering tweets
   2. Facebook REST API for scrapping Facebook posts
   3. NLTK and related libraries for sentiment scores calculation
   4. Matplotlib in python for visualization

# Data cleaning and preparation:

1. Tweepy twitter API was used to gather the twitter streaming trends for the live tweets. The tweets were dumped in a JSON file.
2. Tweets were collected through the season for different teams under their team name hashtags and saved separately as JSON file.
3. Similarly, Posts from Facebook fan pages were collected using REST API and saved as JSON file throughout the season and post season.
4. Ensured columns/information’s from Twitter & FB matches
5. Now all JSON files were combined into single file using Python
6. The JSON file is loaded back into the python environment and the fields such as: Username, Tweet, Likes, Retweets, DateTime, Language, Geography and Source are converted into a dataframe.
7. Using NLTK library, Tweets were Tokenized first. Removed Stop words from the Tweets.
8. Then special Characters were also removed to get final clean data for further analysis.

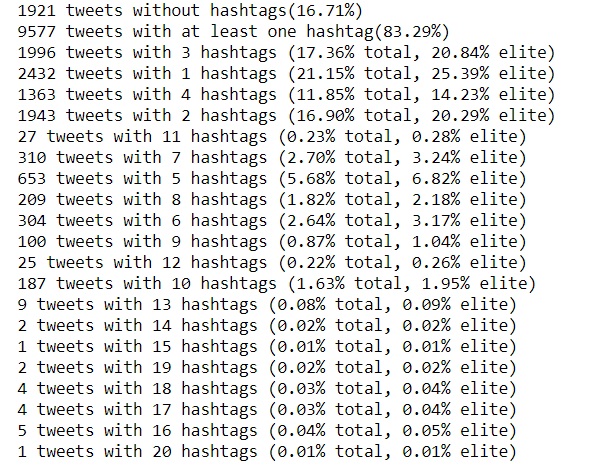
# Analysis:

From the data which I have collected, below are list of summary of the analysis.

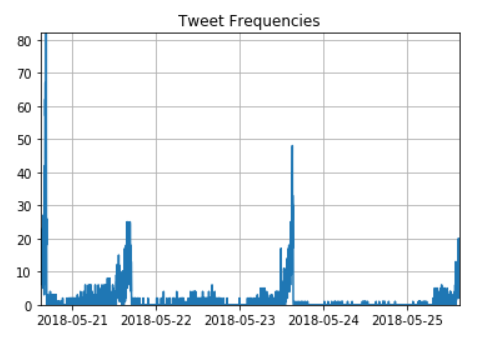
1) Top 20 common Hashtags from Fans



2) Descriptive statistics how Fans are interested in Hashtags



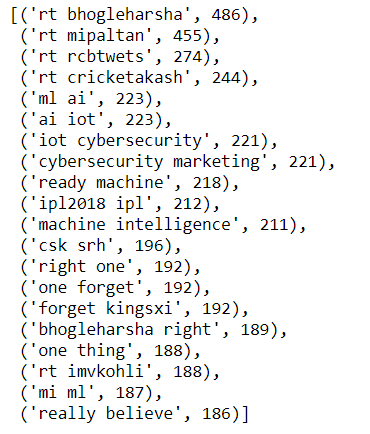
3) Time Series Analysis, During Match Vs Off Match timings



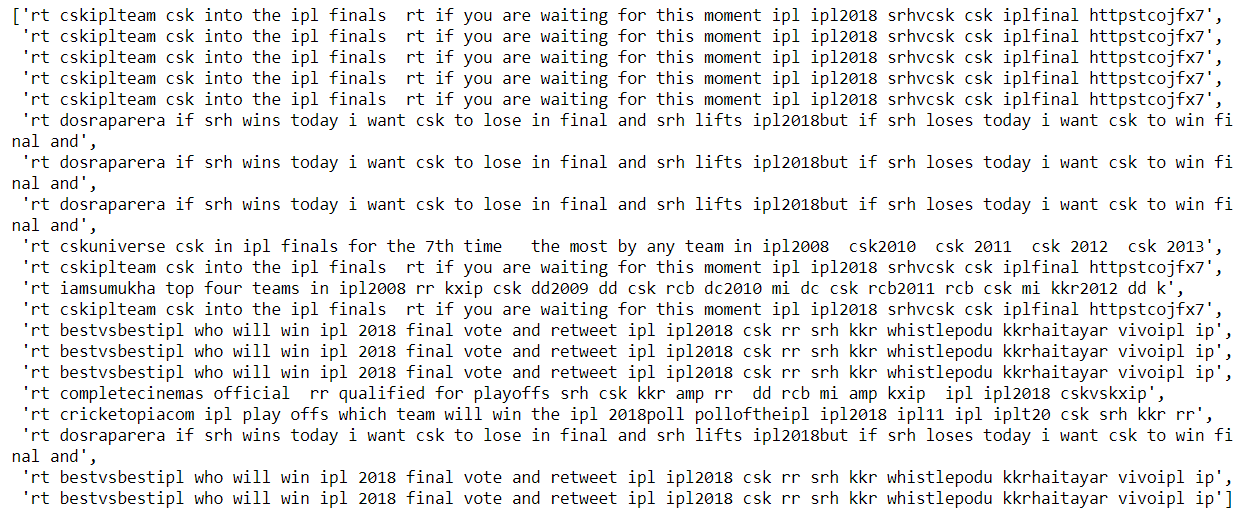
4) Bag of Words from Data



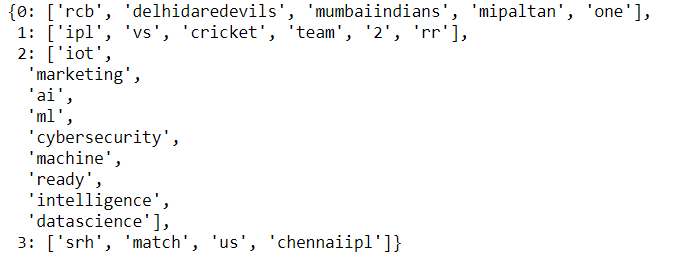
5) Top 20 frequently used Bi-Grams



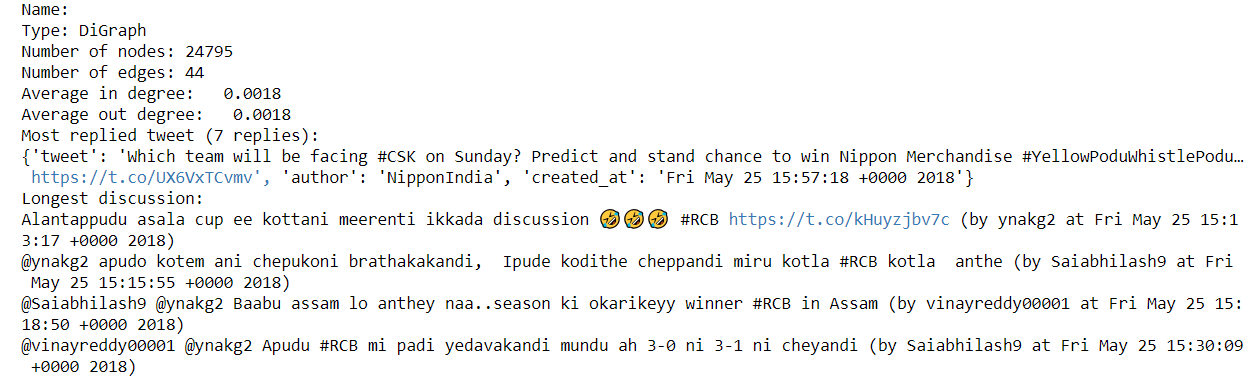
6) Top 20 tweets using rank algorithm



7) Clustering Fans into 4 Categories



8) Network Analysis



# Conclusion and Business Recommendations:

Based on the analysis above, Fans are more sensitive to their team about winning matches. Fan supports particular team based on their geographic location particularly. Fan Engagement is more when particular team quality for Playoff’s and into finals. Fans engaged with team winning more than players.

***Recommendations:***

When analyzing CSK tweets, “#**WhistlePodu**” tag was more equally tagged like #CSK tag. This emotionally connected more fans with team CSK, which is missing with other team. Hence each team must come with different theme which can connect with fans more easily.

2 Cluster of Fan’s identified 1) Team Centric 2) IPL centric i.e. Neutral Audience

# Appendix and other notable related works:

# References and Credits:

1. Tweepy documentation: <http://docs.tweepy.org/en/v3.5.0/>
2. Mastering Social Media Mining with Python - **Marco Bonzanini**
3. Text Analytics with Python - **Dipanjan Sarkar**