Twitter Analysis of IPL 2017

Summary

The Indian Premier League (IPL) is a professional Twenty20 cricket league in India contested during April and May of every year by franchise teams representing Indian cities. The league was founded by the **Board of Control for Cricket in India (BCCI)** in 2007 and major contributor to the GDP of the Indian economy. The IPL is the most-attended cricket league in the world and ranks sixth among all sports leagues. Duff & Phelps added that the value of brand IPL has jumped to **$4.16 billion** after the 2016 edition, against **$3.54 billion** in 2015. IPL 2017 started on April 5, 2017 and final is scheduled on May 21, 2017. There are 8 teams participated in IPL tournament.

My goal is to find out most popular Team and batsman/bowler to target IPL sponsors who would be interested in sports marketing -which focuses both on the promotion of sports events and teams as well as the promotion of other products and services through sporting events and sports teams.

My hypothesis is that two popular winning team will play final. Data from Twitter will be collected and analyzed to predict IPL title holder team.
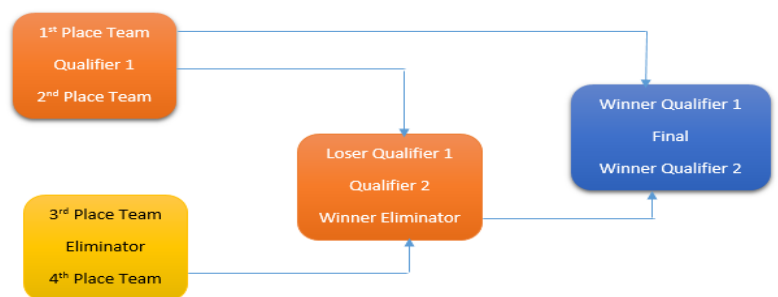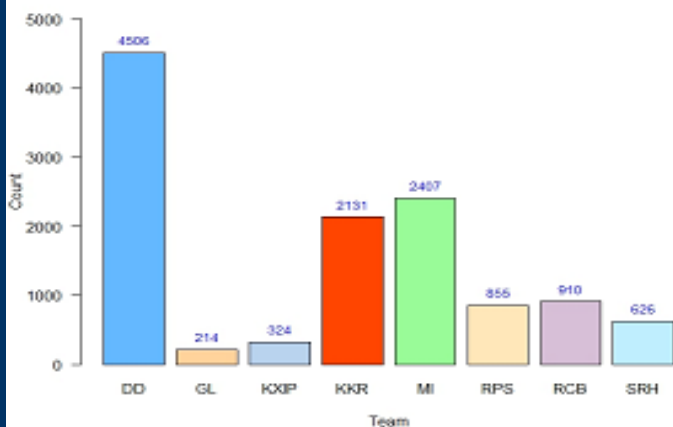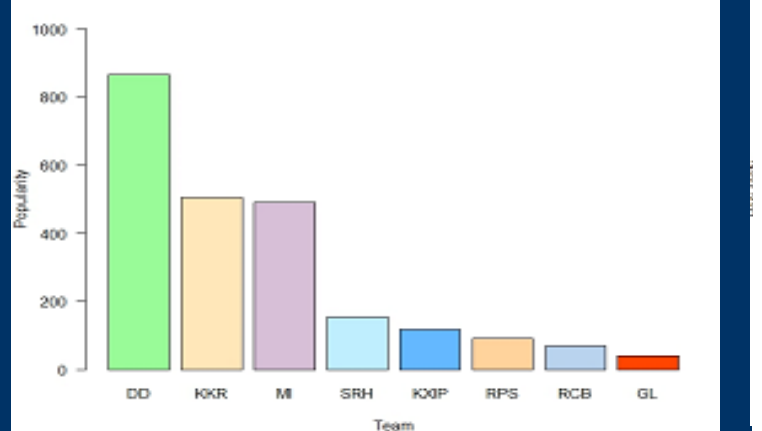

Fig A- IPL Playoff phenomena to qualify final.

Data Collection Method and Analysis

The streaming API method in the tweepy package was used to collect tweets with the keywords, '#RoyalChallengers' and '#GujratLions' on Thursday, April 27. Similarity, tweepy was used to capture and subsequent analysis of tweets to '#DelhiDaredevils', '#KingsXIPunjab', '#KolkataKnightRiders', '#MumbaiIndians', '#RoyalChallengersBangalore', '#SunrisersHyderabad', '#PuneSupergiants','#GujratLions' over the series of the five days (4/22-4/28) of the IPL. No users classification was done while collecting tweets. Tweets were divided into original tweets and retweets and sentiment analysis was performed on each tweet to determine whether it was a positive tweet, a negative tweet, or a neutral tweet. As with most traffic on Twitter, a significant majority of the tweet traffic was retweets However, retweets do not represent fan sentiment as strongly as original tweets. As a result, I used original tweets and retweets only once to determine fan satisfaction. Pandas DataFrame was created to store the tweets and attributes for easy analysis. At the end of the collection period, the daily tweets were stitched together into a "master file". The "master file" was analyzed to find out popular Team and popular batsman/bowler (Which was compared against 30 batsman and 30 bowlers).

Team			Points Table					Batsman				Bowler			
TEAMS	ABBREVIATION		TEAMS	P	W	L	PTS	PLAYER	TEAM	RUNS	AVG	PLAYER	TEAM	BALLS	WKTS
Kolkata Knight Riders	KKR		Kolkata Knight Riders	9	7	2	14	Gautam Gambhir	KKR	376	62.67	Bhuvneshwar Kumar	SRH	168	16
Mumbai Indians	MI		Mumbai Indians	8	6	2	12	Robin Uthappa	KKR	331	41.38	Andrew Tye	GL	120	12
Sunrisers Hyderabad	SRH		Sunrisers Hyderabad	9	5	3	11	Suresh Raina	GL	309	61.8	Chris Morris	DD	150	12
Rising Pune Supergiant	RPS		Rising Pune Supergiant	8	4	4	8	Hashim Amla	KXIP	299	59.8	Mitchell McClenaghan	MI	192	12
Gujarat Lions	GL		Gujarat Lions	8	3	5	6	Sanju Samson	DD	284	40.57	Nathan Coulter-Nile	KKR	86	11
Kings XI Punjab	KXIP		Kings XI Punjab	8	3	5	6								
Royal Challengers Bangalore	RCB		Royal Challengers Bangalore (RCB)	9	2	6	5								
Delhi Daredevils	DD		Delhi Daredevils (DD)	7	2	5	4								

Figure-I: IPL scorecard as of 28-Apr-2017 - Point table of winning teams and stats of top five batsman and bowlers.

No of Tweets (22 Apr 2017 - 28 Apr 2017)

Figure-IA: Distribution of Twitter tweets across Teams. Delhi Dare Devils seems to be most tweeted and Gujrat Lions least tweeted.

Team Popularity (Sentimental Score)

Figure-IB: Sentiment analysis was performed on each tweet to determine popularity based on no. of match being played (**Refer Fig I**) Team popularity found to be matching winning teams points.

Batsman Popularity
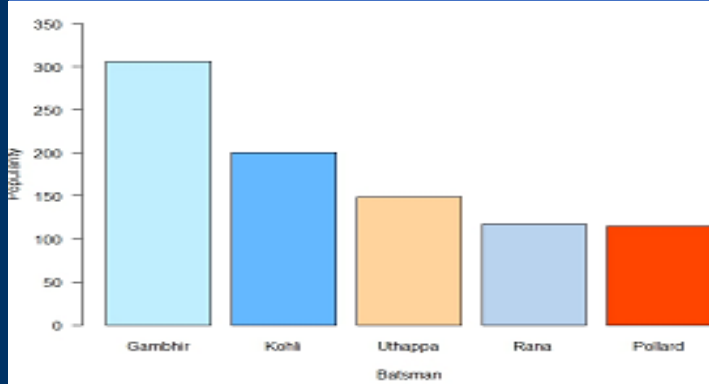
Figure-IC: Twitter text was analyzed to determine top 5 popular batsman out of 30 IPL batsman. Batsman popularity found to be 60% matching with top order batsman (**Refer to Fig I**).

Bowler Popularity
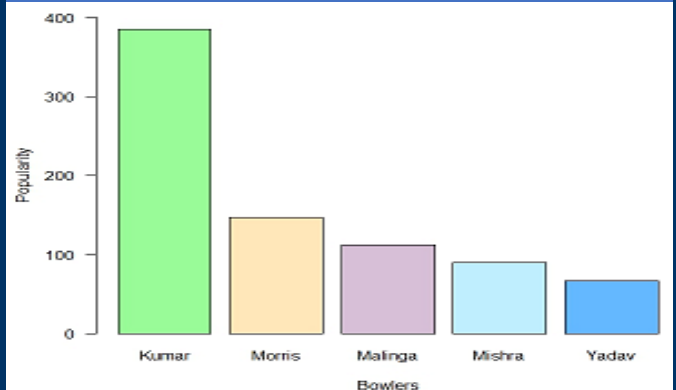
Figure-ID: Twitter text was analyzed to determine top 5 popular bowlers out of 30 IPL bowlers. Bowlers popularity found to be 40% matching with highest wicket taking bowlers (**Refer to Fig I**).

Conclusion

Even though DD has the highest popularity, DD match winning rate is comparatively low granting DD wins next two match. Based on IPL Playoff phenomena (**Fig A**), KKR and MI are potential teams to play final. The t-test was used to analyze KKR vs. MI supporters' feelings.

At α = 0.05 significance level (p-value = 0.7917 which is greater than α), there is not sufficient evidence to claim that the proportion of KKR supporters feelings is larger than the proportion of MI supporters feelings. Thus, it is likely that KKR and MI supporters have the same amount of feelings about winning IPL title.

```
        Welch Two Sample t-test

data:  kkp_tweet and mi_tweet
t = 0.2831, df = 3.842, p-value = 0.7917
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -741.3334  906.6667
sample estimates:
mean of x mean of y
 518.3333  435.6667
```

Figure-1E: Results from statistical tests performed in the study.

*Note: This analysis was performed based on data Collected Until 28 Apr 2017. Remaining 22 qualifying matches are Planned until 14 May 2017.