Social Media Data analysis of NBA Players

Group: Spring

Group Member:

Tong Liu
College of Engineering
Northeastern University
Boston, MA
liu.tong1@husky.neu.edu

Quan Wen
College of Engineering
Northeastern University
Boston, MA
wen.qu@husky.neu.edu

Congxiang Dong
College of Engineering
Northeastern University
Boston, MA
dong.c@husky.neu.edu

Github Link: https://github.com/wenquan1996/Info-6210-Spring-Project-Repository.git

April 25, 2019

ABSTRACT

People nowadays are curious about famous NBA players life behind the screen. However there are too much social media sites that different people talks about different topics of these players, which can may cause misunderstanding of these players. We scraping information of NBA players and their fans from Twitter, Instagram, Reddit. And then, we construct a particular database in SQL and finish normalization of these data. As a result, people may know popular topics of these NBA players and may get a clear view about players' social media data in a more efficient way by using our database

***K*eywords:** Database · Normalization · More

1 Introduction

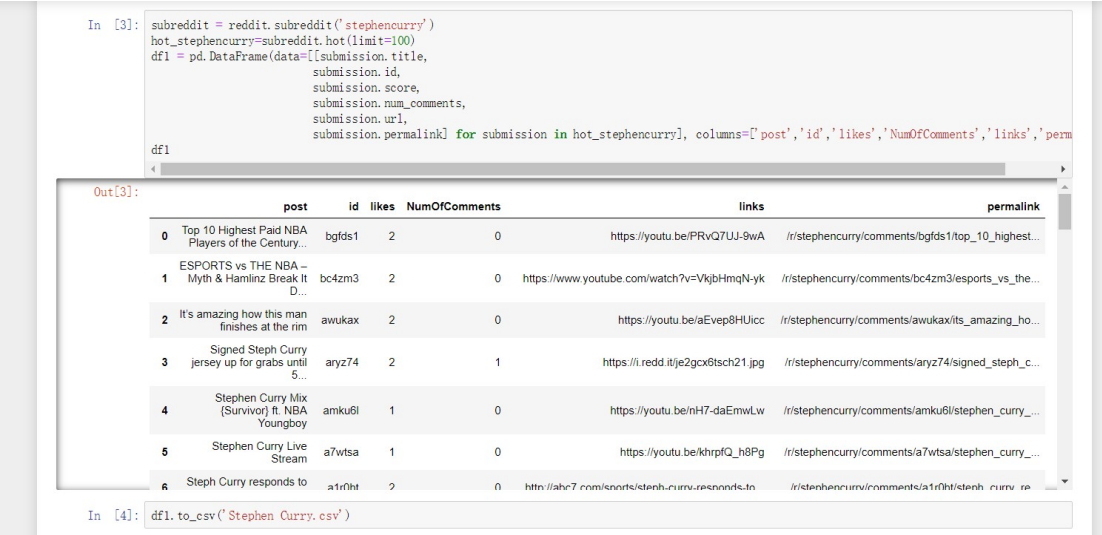
We use API method to get information of posts which are posted by NBA players and their fans from Twitter, Instagram, Reddit, which can indicate social influence of NBA players. We use SQL to normalize these data so that they are in Three Normal Form.

We construct a particular database for NBA players so that people can understand what others' opinion of these NBA players, the popularity of posts of NBA players, what posts are likely to be interesting to them, what users post are similar to each other, what kind of people they should follow so that they can have a better understanding of NBA players, what kind of topics are trending of NBA players, what kind of keywords they should add to their posts, etc.

2 Social Media Scraping

Reddit Scraping:

Reddit is consisted with topics which is regarded as subreddit. We use players' name as subreddit to search for posts. Each player has his own subreddit and we scrape every posts in the subreddit and other attributes of subreddit.



```
In [3]: subreddit = reddit.subreddit('stephencurry')
hot_stephencurry=subreddit.hot(limit=100)
df1 = pd.DataFrame(data=[[submission.title,
                          submission.id,
                          submission.score,
                          submission.num_comments,
                          submission.url,
                          submission.permalink] for submission in hot_stephencurry], columns=['post','id','likes','NumOfComments','links','perm
df1
```

Out[3]:

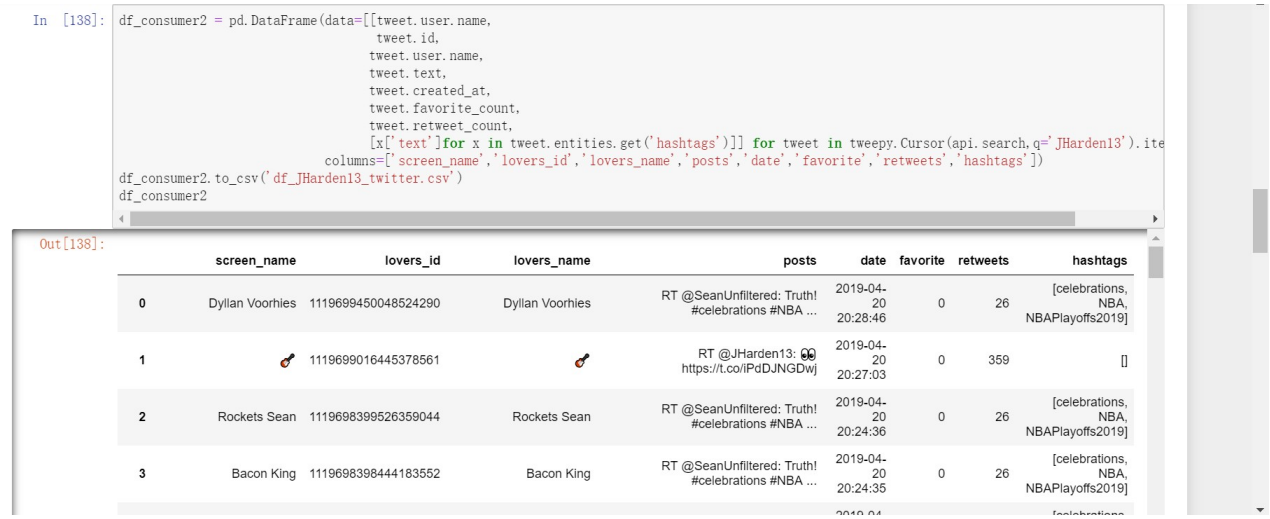
	post	id	likes	NumOfComments	links	permalink
0	Top 10 Highest Paid NBA Players of the Century...	bgfds1	2	0	https://youtu.be/PRvQ7UJ-9wA	/r/stephencurry/comments/bgfds1/top_10_highest...
1	ESPORTS vs THE NBA – Myth & Hamlinz Break It D...	bc4zm3	2	0	https://www.youtube.com/watch?v=VkjbHmqN-yk	/r/stephencurry/comments/bc4zm3/esports_vs_the...
2	It's amazing how this man finishes at the rim	awukax	2	0	https://youtu.be/aEvep8HUicc	/r/stephencurry/comments/awukax/its_amazing_ho...
3	Signed Steph Curry jersey up for grabs until 5...	aryz74	2	1	https://i.redd.it/je2gcx6tsch21.jpg	/r/stephencurry/comments/aryz74/signed_steph_c...
4	Stephen Curry Mix {Survivor} ft. NBA Youngboy	amku6l	1	0	https://youtu.be/nH7-daEmwLw	/r/stephencurry/comments/amku6l/stephen_curry_...
5	Stephen Curry Live Stream	a7wtsa	1	0	https://youtu.be/khrpfQ_h8Pg	/r/stephencurry/comments/a7wtsa/stephen_curry_...
6	Steph Curry responds to	a1r0ht	2	0	http://abc7.com/sports/steph-curry-responds-to	/r/stephencurry/comments/a1r0ht/steph_curry_re

```
In [4]: df1.to_csv('Stephen Curry.csv')
```

Figure 1: Python Code View of Reddit Scraping

Twitter Scraping:

We select NBA players personal account as players' own information and we scrape posts whose hashtags contain these players names. Every posts will be scraped with its own attributes such as data, favorites, retweets, etc.



```
In [138]: df_consumer2 = pd.DataFrame(data=[[tweet.user.name,
                                             tweet.id,
                                            tweet.user.name,
                                            tweet.text,
                                            tweet.created_at,
                                            tweet.favorite_count,
                                            tweet.retweet_count,
                                            [x['text']for x in tweet.entities.get('hashtags')]] for tweet in tweepy.Cursor(api.search,q='JHarden13').ite
                                  columns=['screen_name','lovers_id','lovers_name','posts','date','favorite','retweets','hashtags'])
df_consumer2.to_csv('df_JHarden13_twitter.csv')
df_consumer2
```

Out[138]:

	screen_name	lovers_id	lovers_name	posts	date	favorite	retweets	hashtags
0	Dyllan Voorhies	1119699450048524290	Dyllan Voorhies	RT @SeanUnfiltered: Truth! #celebrations #NBA ...	2019-04-20 20:28:46	0	26	[celebrations, NBA, NBAPlayoffs2019]
1	♂	1119699016445378561	♂	RT @JHarden13: 👀 https://t.co/iPdDJNGDwj	2019-04-20 20:27:03	0	359	[]
2	Rockets Sean	1119698399526359044	Rockets Sean	RT @SeanUnfiltered: Truth! #celebrations #NBA ...	2019-04-20 20:24:36	0	26	[celebrations, NBA, NBAPlayoffs2019]
3	Bacon King	1119698398444183552	Bacon King	RT @SeanUnfiltered: Truth! #celebrations #NBA ...	2019-04-20 20:24:35	0	26	[celebrations, NBA, NBAPlayoffs2019]

Figure 2: Python Code View of Twitter Scraping

Instagram Scraping:

We scrape posts which are posted by NBA players throw players' personal account and posts which contain keywords of players' name.

```
In [23]: df1=df.drop('hashtags',axis=1).join(df["hashtags"].str.split(',',expand=True).stack().reset_index(level=1,drop = True).rename("tags"))
         df1.to_csv("curry_twitter.csv")
         df1
```

Out [23]:

	Unnamed: 0	post_id	screen_name	player_id	posts	date	likes	retweets	tags
0	0	1118658675697016837	StephenCurry30	42562446	I couldn't be more humbled to be a part of bri...	2019-04-17 23:33:06	17133	2103	[]
1	1	1117504662947581952	StephenCurry30	42562446	Greatest comeback story in sports! Congrats @T...	2019-04-14 19:07:28	80501	8758	[]
2	2	1117164541396209664	StephenCurry30	42562446	tune in 5/2 €€ https://t.co/zfMROKcqmh https://...	2019-04-13 20:35:57	8445	804	[]
3	3	1116737618333409281	StephenCurry30	42562446	@INFINITIUSA back at it with another 💧 concept...	2019-04-12 16:19:30	61	11	['QSInspiration'
3	3	1116737618333409281	StephenCurry30	42562446	@INFINITIUSA back at it with another 💧 concept...	2019-04-12 16:19:30	61	11	'ad']
					New @RakutenUS spot drops	2019-04-			

Figure 3: Python Code View of Instagram Scraping

3 ER Diagram

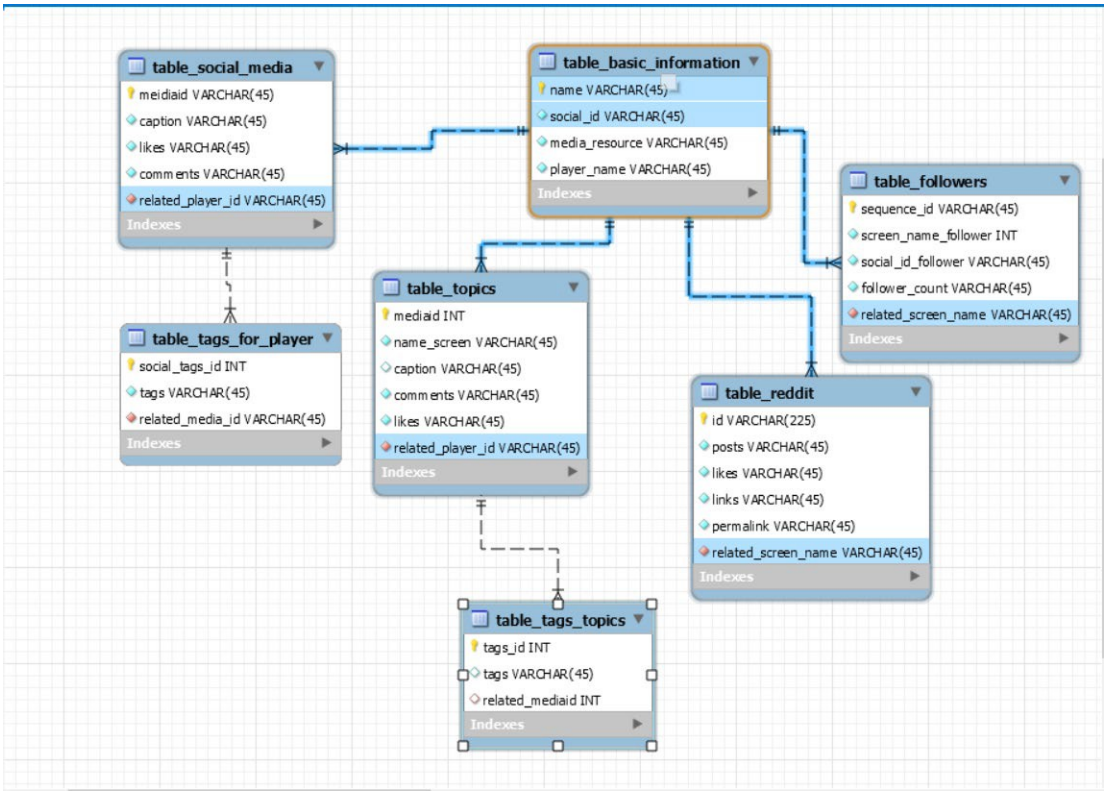


Figure 4: ER Diagram in SQL Workbench

We extract a plenty of data from every social media sites. After exporting csv file of these tables we import these files into SQL workbench and divide these table into several tables which have more specific information. According to ER Diagram, we construct a domain table which is called table_basic_information. Table_basic_information contains NBA players' name, players' social account id, their social media name and the resources name of their social media name.

Table_reddit and table_topics includes information of posts of players' topics from Reddit, Instagram and Twitter, they are connected throw table_basic information. Each post is related to a player's id or name, which are indicated as related_screen_name and related_player_id in table_reddit and table_topics. Id column in table_reddit is unique and mediaid in table_topics is unique in table_topics.

The information of fans is mentioned in table_followers and every name of fans has a related_screen_name of an NBA player. We generate sequence_id in an ascending order so that posts posted by the same follower will be distinct to each other. There are also follower_count which can indicate the number of followers of each follower of a player. If a follower also has a large number of followers, other followers of the player should also follow this follower as there must be something different under this follower's post about this player that others show interests with this follower.

Table_social_media contains information of NBA players social personal posts and posts are grouped by particular player's id which is indicated as related_player_id in the table. Each post has its unique media id as their primary key.

The hashtags are sorted into table_tags_topics and table_tags_for_player. Each hashtag in table_tags_topics and

table_tags_for_player is related to a particular mediaid which is the id of post in the table_topics and table_social_media

4 Normalization

4.1 Check that tables are in First normal form (1NF)

Each table has a primary key which can uniquely identify a record in each tableThe values in each column of the four tables are atomic.There are no repeating groups in each table

4.2 A check that tables are in Second normal form (2NF)

All players data are divided in several tables. Table_social_media contains NBA players personal posts. Table_topics and table_reddit contains topics about NBA Players sent by fans. The information of fans are encapsulated in table_followers. There are no partial dependencies and calculated data in these tables and these tables are already in 1NF.

4.3. A check that your tables are in Third normal form (3NF)

As one player has several post and one post may have several hashtags. We stored player's information in table_basic_information and information of players posts in table_social_media and information of hashtags in table_tags_for_player. As one player has several followers' posts, these posts are stored in table_followers. There are no fields that do not directly depend on the primary key in these tables and these tables are already in 2NF.

5 Output of SQL

table_basic_information: Includes players' social media name, social id, media, media resource and player_name
table_followers: Includes social_id of fans, the screen name of fans and the number of NBA fans and related Player name
table_reddit: Includes posts of fans in reddit, id, likes, links, permalink, link in the post and related player name of the post
table_social_media: Includes media id, caption, likes, comments of posts about players topics and related players id
table_tags_for_player: Includes social_tags_id, post of tags and related player name
table_tags_topics: Includes tags_id and post of tags and related player name
table_topics: Includes media id, caption, comments, likes, poster name(name_screen) of each post and related player id

5.1 Question 1 what are people saying about me?

We select fans' id, post, related_screen_name, player_name from table_reddit, table_basic_information and union posts which represent post from table_topics. As a result, we may see posts that says about players

fans	post	related_screen_name	player_name
1t2hyw	Steph Curry goes HAM	r_stephencurry	Stephen Curry
1t2ktb	Steph Curry has yet to be elected to an All-Star…	r_stephencurry	Stephen Curry
1uy3iq	Just another game-winner for Steph…	r_stephencurry	Stephen Curry
1v66bf	First Post = Highlights	r_giannis_an34	giannis antetokounmpo
1vrg4e	Giannis's best game so far	r_giannis_an34	giannis antetokounmpo
1vsc60	Breaking a screen to the block on Durant, all in …	r_giannis_an34	giannis antetokounmpo
1vuisz	Keep it movin Giannis	r_giannis_an34	giannis antetokounmpo
1vus10	Who Is The Greek Freak? Your Guide To The NB…	r_giannis_an34	giannis antetokounmpo
1vx0gr	His hands are absurdly large	r_giannis_an34	giannis antetokounmpo
1wcspk	Is Kevin Durant the best player in the league? …	r_kd35warriors	Kevin Durant
1wm89u	Giannis put at pick #2 in the mid-season re-draft	r_giannis_an34	giannis antetokounmpo
1x1iws	Reminiscent of Dr. J?	r_giannis_an34	giannis antetokounmpo
1x50ca	The (Unlimited?) Potential of Giannis Antetokou…	r_giannis_an34	giannis antetokounmpo
1xhcan	A Day In The Life Of Giannis Antetokounmpo	r_giannis_an34	giannis antetokounmpo

5.2 Question 2 How viral are my posts?

We select player_name, caption, likes from table_social_media, table_basic_information and the two tables are connected as table_social_media.related_player_id are the same with table_basic_information.social_id and then we order each post by the number of likes so that you can see how viral are the players' posts

```
11      #2 How viral are my posts?
12 •    select player_name, caption, likes
13      from table_social_media, table_basic_information
14      where table_social_media.related_player_id =table_basic_information.social_id
15      order by related_player_id,likes desc;
16
17
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

player_name	caption	likes
James Harden	BRO!!!! Where did you go?? We had some shit …	924137
James Harden	NEEEDS > wants	918194
James Harden	Gotta tip yo □ off to one of the best to do it. A…	728862
James Harden	2009-2019…Time flies. (Ya'll see the yu…	708160
James Harden	omg	701725

Result 2

5.3 Question 3 What posts are likely to be interesting to me?

We select mediaid, name_screen, caption, player_name from table_topics,table_basic_information to find out the fans who @the players' name and the two tables are joined by related_player_id and social id

```
18    #3 What posts are likely to be interesting to me? the fans who @ me
19    select mediaid, name_screen, caption, player_name from table_topics,table_
20    where table_topics.related_player_id = table_basic_information.social_id
21    and caption like '%@stephencurry30'
22    or caption like '%@StephenCurry30'
23    or caption like '%@giannis_an'
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

mediaid	name_screen	caption	player_name
725873	glvder	欸?these two softies that deserves more appre...	giannis antetokounmpo
393753	packers.bucks	| finally posted again偓楹 - - 崇€崇荆 videosta...	giannis antetokounmpo
588022	da_best_fx	#Lebronjamesedit @kingjames	giannis antetokounmpo
55560	Bo	@Chris_Broussard If you could work on your sh...	giannis antetokounmpo
532947	Tay	it's only mid-way through the 1st round and I mi...	giannis antetokounmpo

Result 3

5.4 Question 4 What posts are like mine?

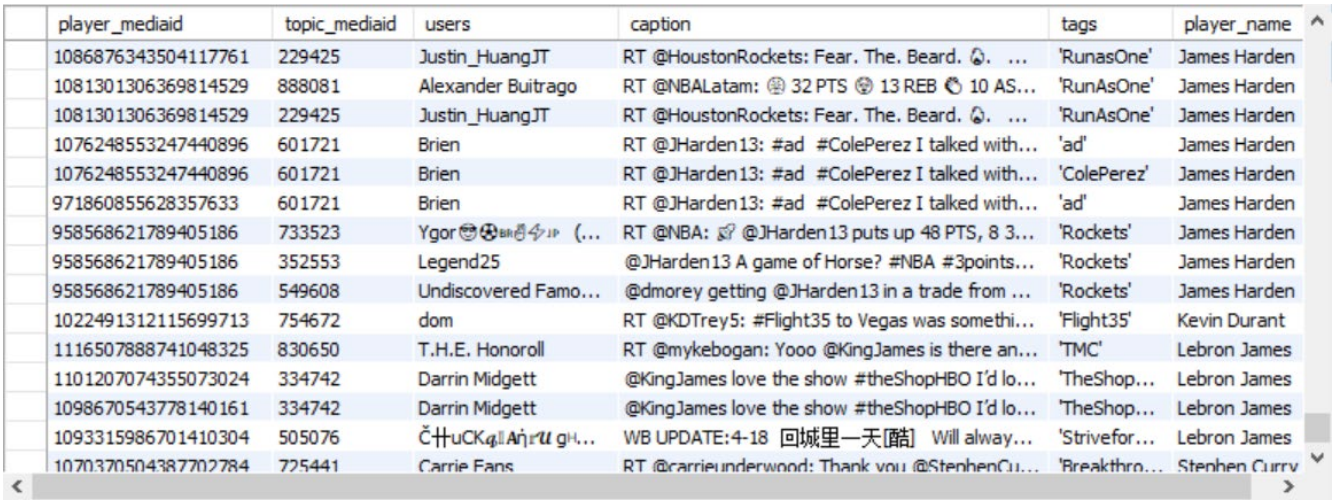
We create two indexes to increase the efficiency of SQL

```
create index itags on table_tags_topics(tags) ;
create index otags on table_tags_for_player(tags) ;
```

First of all, we connect five tables as table_tags_for_player.tags is the same to table_tags_topics.tags and table_tags_for_player.related_media_id is the same to table_social_media.mediaid and table_social_media.related_player_id is the same to table_basic_information.social_id and table_basic_information.social_id is the same to table_topics.related_player_id and table_topics.mediaid is the same to table_tags_topics.related_mediaid
and then we can find out posts which share the same hashtag in five tables. The output of question 4 is listed in question five

5.5 Question 5 What users post like me?

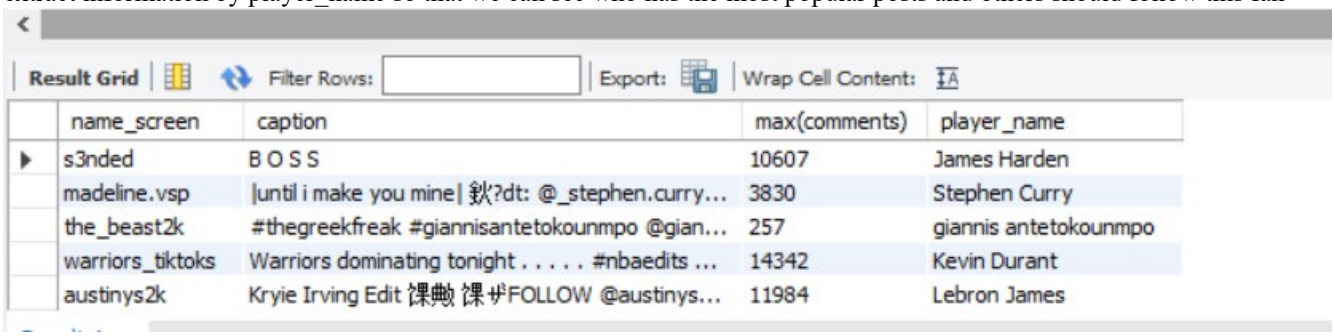
First of all, we connect five tables as table_tags_for_player.tags is the same to table_tags_topics.tags and table_tags_for_player.related_media_id is the same to table_social_media.mediaid and table_social_media.related_player_id is the same to table_basic_information.social_id and table_basic_information.social_id is the same to table_topics.related_player_id and table_topics.mediaid is the same to table_tags_topics.related_mediaid
and then we can find out whose posts which share the same hashtag in five tables



player_mediaid	topic_mediaid	users	caption	tags	player_name
1086876343504117761	229425	Justin_HuangJT	RT @HoustonRockets: Fear. The. Beard. ...	'RunasOne'	James Harden
1081301306369814529	888081	Alexander Buitrago	RT @NBALatam: 32 PTS 13 REB 10 AS...	'RunAsOne'	James Harden
1081301306369814529	229425	Justin_HuangJT	RT @HoustonRockets: Fear. The. Beard. ...	'RunAsOne'	James Harden
1076248553247440896	601721	Brien	RT @JHarden13: #ad #ColePerez I talked with...	'ad'	James Harden
1076248553247440896	601721	Brien	RT @JHarden13: #ad #ColePerez I talked with...	'ColePerez'	James Harden
971860855628357633	601721	Brien	RT @JHarden13: #ad #ColePerez I talked with...	'ad'	James Harden
958568621789405186	733523	Ygor	RT @NBA: @JHarden13 puts up 48 PTS, 8 3...	'Rockets'	James Harden
958568621789405186	352553	Legend25	@JHarden13 A game of Horse? #NBA #3points...	'Rockets'	James Harden
958568621789405186	549608	Undiscovered Famo...	@dmorey getting @JHarden13 in a trade from ...	'Rockets'	James Harden
1022491312115699713	754672	dom	RT @KDTrey5: #Flight35 to Vegas was somethi...	'Flight35'	Kevin Durant
1116507888741048325	830650	T.H.E. Honoroll	RT @mykebogan: Yooo @KingJames is there an...	'TMC'	Lebron James
1101207074355073024	334742	Darrin Midgett	@KingJames love the show #theShopHBO I'd lo...	'TheShop...	Lebron James
1098670543778140161	334742	Darrin Midgett	@KingJames love the show #theShopHBO I'd lo...	'TheShop...	Lebron James
1093315986701410304	505076	ČㅓuCK	WB UPDATE:4-18 回城里一天[酷] Will alway...	'Striveforf...	Lebron James
1070370504387702784	725441	Carrie Fans	RT @carrieunderwood: Thank you @StephenCu...	'Breakthro...	Stephen Curry

5.6 Question 6 Who should I be following?

We select name_screen, caption, the max number of comments, player_name from table_topics, table_basic_information as player_id are the same with social_id in table_basic_information and table_topics.related. And then we group the extract information by player_name so that we can see who has the most popular posts and others should follow this fan

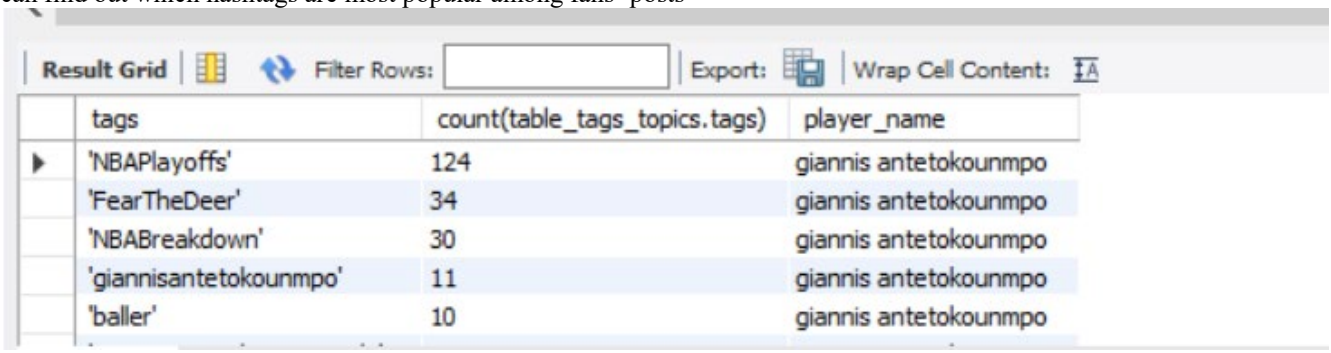


Result Grid | Filter Rows: | Export: | Wrap Cell Content:

name_screen	caption	max(comments)	player_name
s3nded	B O S S	10607	James Harden
madeline.vsp	|until i make you mine| 欸?dt: @_stephen.curry...	3830	Stephen Curry
the_beast2k	#thegreekfreak #giannisantetokounmpo @gian...	257	giannis antetokounmpo
warriors_tiktoks	Warriors dominating tonight . . . . . #nbaedits ...	14342	Kevin Durant
austinys2k	Kryie Irving Edit 傈幽 傈 FOLLOW @austinys...	11984	Lebron James

5.7 Question 7 What topics are trending in my domain?

We select table_tags_topics.tags, the number of table_tags_topics.tags and player_name from table_topics, table_basic_information and table_tags_topics and connect these tables as table_topics.related_player_id is the same with table_basic_information.social_id and table_topics.mediaid is the same with table_tags_topics.related_mediaid so that we can find out which hashtags are most popular among fans' posts



Result Grid | Filter Rows: | Export: | Wrap Cell Content:

tags	count(table_tags_topics.tags)	player_name
'NBAPlayoffs'	124	giannis antetokounmpo
'FearTheDeer'	34	giannis antetokounmpo
'NBABreakdown'	30	giannis antetokounmpo
'giannisantetokounmpo'	11	giannis antetokounmpo
'baller'	10	giannis antetokounmpo

5.8 Question 8 What keywords/ hashtags should I add to my post?

We select tags, the number of tags, table_basic_information.player_name from table_tags_topics, table_topics and table_basic_information. We create a view which is called all_same_tag which contains the same tags in table_tags_for_player and table_topics. We use where not in to select tags that are not the same. The tables are connected as the media id in table_tags_topics are the same with media id in table_topics and related_player_id in table_topics are

the same with social id in table_basic_information so that we may find out the number of mentioned times of hashtags of a player

tags	count	player_name
'NBABreakdown'	30	giannis antetokounmpo
'giannisantetokounmpo'	11	giannis antetokounmpo
'baller'	10	giannis antetokounmpo
'follow4follow'	6	giannis antetokounmpo
'giannisantetokounmpoedit'	6	giannis antetokounmpo
'giannisantetokounmpoedit'	6	giannis antetokounmpo
'dope'	5	giannis antetokounmpo
'milwaukeebucks'	4	giannis antetokounmpo
'omgpageunderrated'	4	giannis antetokounmpo
'dunk'	4	giannis antetokounmpo
'bucksedit'	3	giannis antetokounmpo
'videostarapp'	3	giannis antetokounmpo
'americanfootball'	3	giannis antetokounmpo
'america'	3	giannis antetokounmpo
'packersedit'	2	giannis antetokounmpo
'2k'	2	giannis antetokounmpo

5.9 Question 9 Should I follow somebody back?

We select all information from table_followers and the information has more followers than average and we order them in an descending order so that these followers are worth others to follow back

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

sequence_id	social_id_follower	screen_name_follower	followers_count	related_screen_name
342	161555958	YourFavBlackGuy	75475	t_JHarden13
625	103064205	Antwan Odom	59908	t_Giannis_An34
339	859099743070412800	Lit memes	4419	t_JHarden13
971	439620215	gr1ndasmagazine	4182	t_KDTrey5
373	300508333	Harkinsfamilymattres	2205	t_JHarden13

table_followers7

5.10 Function 1

```
1    #show the players'post which has the highest likes when giving related_player_id
2 •  select info_project.best_caption(324599988);
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

info_project.best_caption(324599988)
No words!! #triplecrown #dubnation

Result 1

5.11 Function 2

```
4    #show the count of topics in reddit when giving related_screen_name
5 •  select info_project.find_reddit('r_stephencurry');
```

info_project.find_reddit('r_stephencurry')
88

5.12 Function 3

```
#show the name of the fan in the topic whose post has the highest likes when giving player's name
select info_project.hot_fans('Lebron James');
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

info_project.hot_fans('Lebron James')
chen.mp4

5.13 Function 4

```
#the player's follower who has the most followers when giving the player's screen name
select info_project.hot_follower('t_StephenCurry30');
```

Result Grid | Filter Rows: | Export:

info_project.hot_follower('t_StephenCurry30')
Tony Bowles

5.14 Function 5

```
#the number of followers about the player in the database when giving player's name
select info_project.num_followers('James Harden');
```

info_project.num_followers('James Harden')
200

5.15 Function 6

```
15
16      #the sum of the followers of followers of the player when giving the player's screen name
17 •    select info_project.`sum_followers'followers`('t_StephenCurry30');
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

info_project.`sum_followers'followers`('t_Stepher
3460

Result Grid

Result 8 × Read Only

5.16 View 1

```
#all same tags between table_tags_topics and table_tags_for_players
SELECT * FROM info_project.all_same_tags;
```

Result Grid | Filter Rows:

tags
'2'
'23'
'30'
'34'
'ad'
'allstar'
'allstars'

5.17 View 2

player_name	caption	likes
James Harden	BRO!!!! Where did you go?? We had some shit ...	924137
James Harden	NEEEDS > wants	918194
James Harden	Gotta tip yo 🎩 off to one of the best to do it. A...	728862
James Harden	2009-2019👀👀...🤔 Time flies. (Ya'll see the yu...	708160
James Harden	😳😳😳😳😳 omg	701725
James Harden	GENERATIONAL Lefty to lefty ..... @zmane2...	690919
James Harden	Nights like this.....	672593

viral posts 8 ×

```
# show the players'posts according to the likes-count
SELECT * FROM info_project.`viral posts`;
```

5.18 View 3

```
#show all posts from fans from three social medias according to players
select * from info_project.`say about`
order by player_name;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

fans	post	related_screen_name	player_name
G!lboe??oe???	RT @NBA: e?|? 25 PPG	14.5 RPG e?|?	t_Giannis_An34	giannis antetokounmpo
Michela Wood	RT @NBA: e?|? 25 PPG	14.5 RPG e?|?	t_Giannis_An34	giannis antetokounmpo
Kiel	RT @NBA: e?|? 25 PPG	14.5 RPG e?|?	t_Giannis_An34	giannis antetokounmpo
1v66bf	First Post = Highlights	r_giannis_an34	giannis antetokounmpo
1vrg4e	Giannis's best game so far	r_giannis_an34	giannis antetokounmpo
1vsc60	Breaking a screen to the block on Durant, all in ...	r_giannis_an34	giannis antetokounmpo
1vuisz	Keep it movin Giannis	r_giannis_an34	giannis antetokounmpo
1vus10	Who Is The Greek Freak? Your Guide To The NB...	r_giannis_an34	giannis antetokounmpo
1vx0gr	His hands are absurdly large	r_giannis_an34	giannis antetokounmpo
OD Engineeri...	RT @NBA: e?|? 25 PPG	14.5 RPG e?|?	t_Giannis_An34	giannis antetokounmpo
1wm89u	Giannis put at pick #2 in the mid-season re-draft	r_giannis_an34	giannis antetokounmpo
1x1iws	Reminiscent of Dr. J?	r_giannis_an34	giannis antetokounmpo

5.19 View 4

```
#the sum of likes and comments of players'posts in different social media
SELECT * FROM info_project.sum_likes;
```

sum(likes)	sum(comments)	player_name	media_resource
276901646	2243411	Stephen Curry	instagram
56421680	304846	giannis antetokounmpo	instagram
84242520	635572	James Harden	instagram
454849370	4880364	Lebron James	instagram
5005945	150957	Kevin Durant	instagram
413686	287085	giannis antetokounmpo	twitter
1853473	389246	James Harden	twitter
466261	338846	Kevin Durant	twitter
6813397	1154785	Lebron James	twitter
3794321	581425	Stephen Curry	twitter

5.20 View 5

```
#the count of tags showed in the topic posts with ranking
SELECT * FROM info_project.tags_topic;
```

Result Grid | Filter Rows:

count	tags
115	'striveforgreatness'
78	'dubnation'
78	'striveforgreatness'
52	'lockedin'
49	'jamesgang'
39	'jamesgang'
37	'ad'
32	'antetokounbros'
23	'thekidfromakron'
23	'thekidfromakron'
18	'rwtw'
18	'antetokounbros'

tags_topic 11 ×

5.21 View 6

```
#the count of tags showed in the players'posts with ranking
SELECT * FROM info_project.`trending topics`;
```

tags	count(table_tags_topics.tags)	player_name
'NBAPlayoffs'	124	giannis antetokounmpo
'FearTheDeer'	34	giannis antetokounmpo
'NBABreakdown'	30	giannis antetokounmpo
'giannisantetokounmpo'	11	giannis antetokounmpo
'baller'	10	giannis antetokounmpo
'giannisantetokounmpoedit'	6	giannis antetokounmpo
'giannisantetokounmpoedit'	6	giannis antetokounmpo
'follow4follow'	6	giannis antetokounmpo
'dope'	5	giannis antetokounmpo
'giannis'	5	giannis antetokounmpo
'dunk'	4	giannis antetokounmpo
'milwaukeebucks'	4	giannis antetokounmpo

6 Conclusion

We have scrape NBA players' social media information and thousands of information of posts from their fans by using Python. We have now mastered design of information systems from a data perspective for engineering and business applications. At the same time, we have improved data modeling, including entity-relationship (E-R) and object approaches. We are able to understand user-centric information requirements and data sharing, fundamental concepts of database management systems (DBMS) and we are able to manipulate their applications and construct alternative data models. We have gained proficiency in SQL, data normalization, data-driven application design for personal computer, server-based, Internet databases and distributed data applications. We are grateful to our teacher and TAs as they help us improve our understanding of database management and database design and they are willing to give us a hand when we encounter difficulties during our final project.

References

[1] Elvis C. Foster, Shripad V. Database Systems A Pragmatic Approach. *2014*.

[2] M. Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems. *2014 Third Edition*

[3] Clare Churcher. Beginning Database Design from Novice to Professional. 2012

[4] Magnus Lie Hetland. Beginning Python from Novice to Professional. 2017

[5] M. Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems. 2011 Third Edition

[6] David Hows, Peter Membrey, Eelco Plugge, Tim Hawkins. The Definitive Guide to MongoDB: A complete guide to dealing with Big Data using MongoDB. 2015

[7] Michael Frampton. Big Data Made Easy A Working Guide to the Complete Hadoop Toolset. 2015

[8] Grant Allen, Mike Owens. The Definitive Guide to SQLite. 2010

MIT License