INFO 6210
Data Management and Database
Assignment 3

Changrong Chen 001276880

I chose games as my domain. The original data comes from steam and it has created a postgreSQL database successfully. In this assignment I need to convert my postgreSQL database to mongoDB.

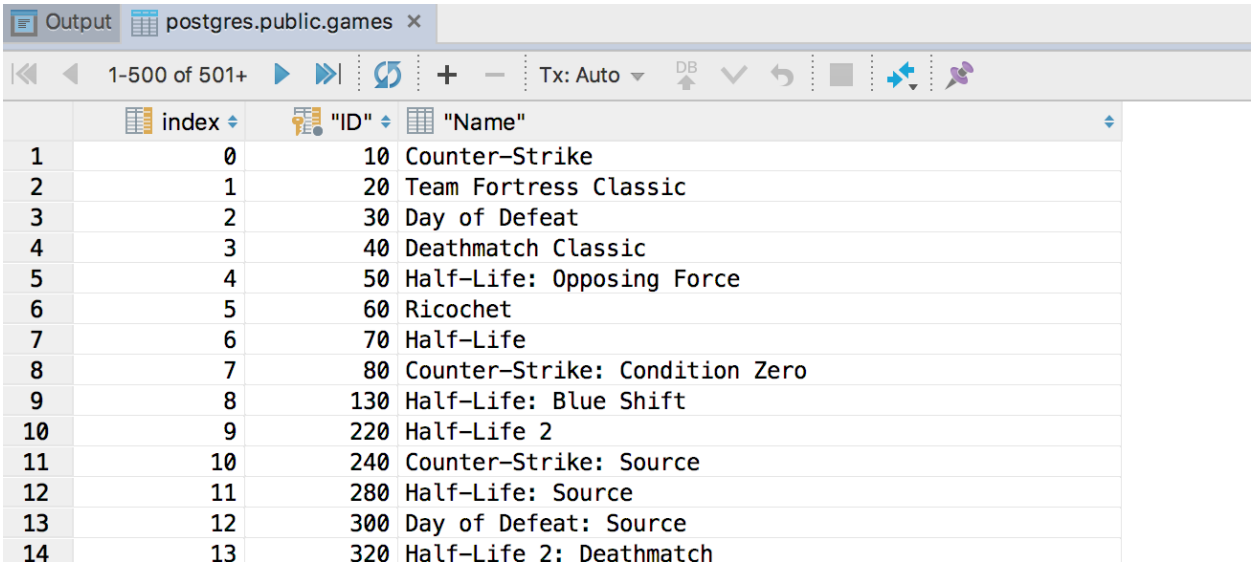
The original database: 2 parts, game & social media data.

Game has 4 tables: games, games_features, games_reviews, games_users.

Social media data has 2 tables: game_tweets, game_hashtags.

The original sample data:

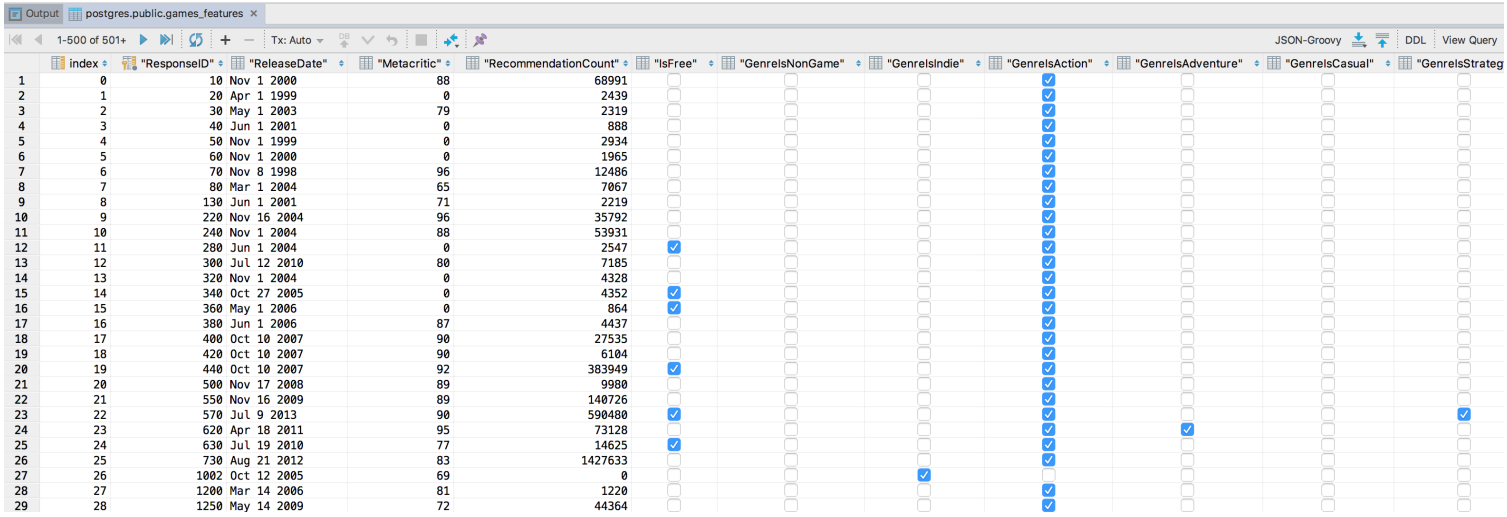
games:


Output postgres.public.games ×

1-500 of 501+ Tx: Auto

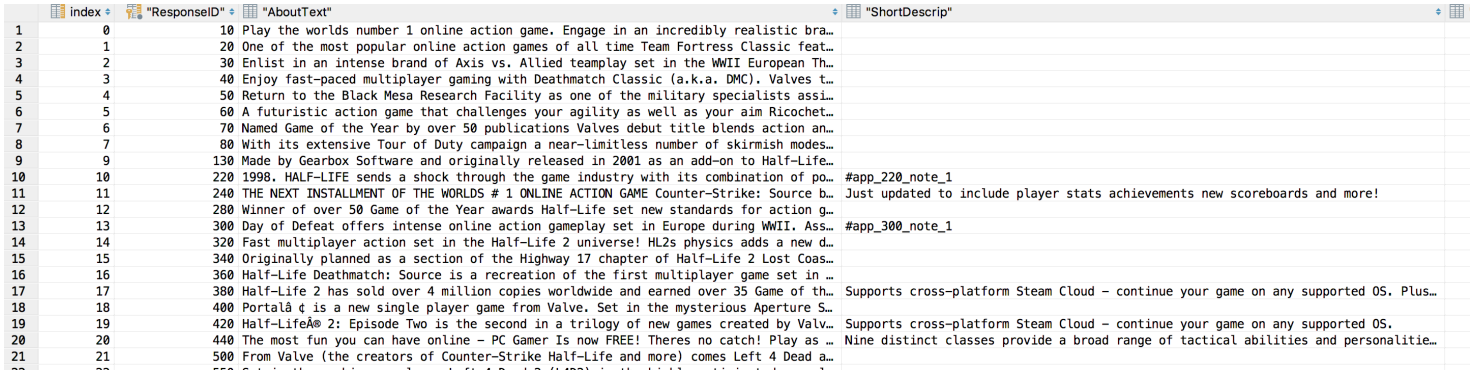
	index	"ID"	"Name"
1	0	10	Counter-Strike
2	1	20	Team Fortress Classic
3	2	30	Day of Defeat
4	3	40	Deathmatch Classic
5	4	50	Half-Life: Opposing Force
6	5	60	Ricochet
7	6	70	Half-Life
8	7	80	Counter-Strike: Condition Zero
9	8	130	Half-Life: Blue Shift
10	9	220	Half-Life 2
11	10	240	Counter-Strike: Source
12	11	280	Half-Life: Source
13	12	300	Day of Defeat: Source
14	13	320	Half-Life 2: Deathmatch

games_features:



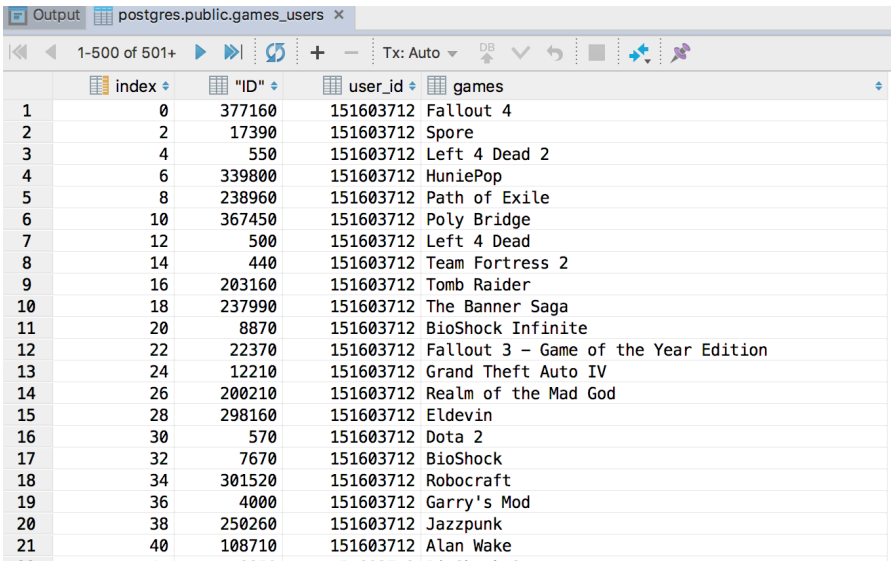
	index	"ResponseID"	"ReleaseDate"	"Metacritic"	"RecommendationCount"	"IsFree"	"GenreIsNonGame"	"GenreIsIndie"	"GenreIsAction"	"GenreIsAdventure"	"GenreIsCasual"	"GenreIsStrateg
1	0	10	Nov 1 2000	88	68991	☐	☐	☐	☑	☐	☐	☐
2	1	20	Apr 1 1999	0	2439	☐	☐	☐	☑	☐	☐	☐
3	2	30	May 1 2003	79	2319	☐	☐	☐	☑	☐	☐	☐
4	3	40	Jun 1 2001	0	888	☐	☐	☐	☑	☐	☐	☐
5	4	50	Nov 1 1999	0	2934	☐	☐	☐	☑	☐	☐	☐
6	5	60	Nov 1 2000	0	1965	☐	☐	☐	☑	☐	☐	☐
7	6	70	Nov 8 1998	96	12486	☐	☐	☐	☑	☐	☐	☐
8	7	80	Mar 1 2004	65	7067	☐	☐	☐	☑	☐	☐	☐
9	8	130	Jun 1 2001	71	2219	☐	☐	☐	☑	☐	☐	☐
10	9	220	Nov 16 2004	96	35792	☐	☐	☐	☑	☐	☐	☐
11	10	240	Nov 1 2004	88	53931	☐	☐	☐	☑	☐	☐	☐
12	11	280	Jun 1 2004	0	2547	☑	☐	☐	☑	☐	☐	☐
13	12	300	Jul 12 2010	80	7185	☐	☐	☐	☑	☐	☐	☐
14	13	320	Nov 1 2004	0	4328	☐	☐	☐	☑	☐	☐	☐
15	14	340	Oct 27 2005	0	4352	☑	☐	☐	☑	☐	☐	☐
16	15	360	May 1 2006	0	864	☑	☐	☐	☑	☐	☐	☐
17	16	380	Jun 1 2006	87	4437	☐	☐	☐	☑	☐	☐	☐
18	17	400	Oct 10 2007	90	27535	☐	☐	☐	☑	☐	☐	☐
19	18	420	Oct 10 2007	90	6104	☐	☐	☐	☑	☐	☐	☐
20	19	440	Oct 10 2007	92	383949	☑	☐	☐	☑	☐	☐	☐
21	20	500	Nov 17 2008	89	9980	☐	☐	☐	☑	☐	☐	☐
22	21	550	Nov 16 2009	89	140726	☐	☐	☐	☑	☐	☐	☐
23	22	570	Jul 9 2013	90	590480	☑	☐	☐	☑	☐	☐	☑
24	23	620	Apr 18 2011	95	73128	☐	☐	☐	☑	☑	☐	☐
25	24	630	Jul 19 2010	77	14625	☑	☐	☐	☑	☐	☐	☐
26	25	730	Aug 21 2012	83	1427633	☐	☐	☐	☑	☐	☐	☐
27	26	1002	Oct 12 2005	69	0	☐	☐	☑	☐	☐	☐	☐
28	27	1200	Mar 14 2006	81	1220	☐	☐	☐	☑	☐	☐	☐
29	28	1250	May 14 2009	72	44364	☐	☐	☐	☑	☐	☐	☐

games_reviews:



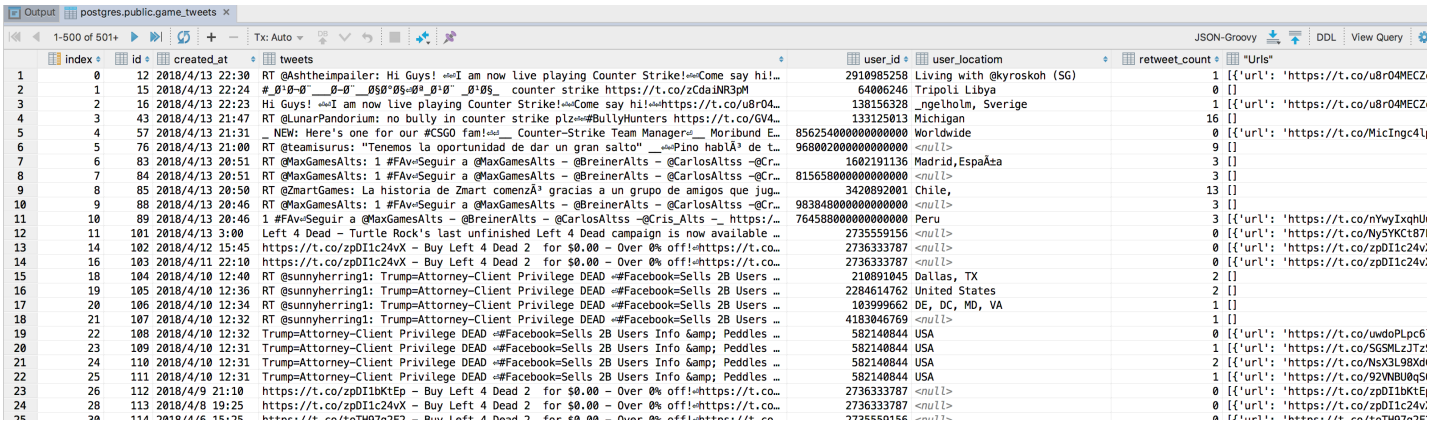
	index	"ResponseID"	"AboutText"	"ShortDescrip"
1	0	10	Play the worlds number 1 online action game. Engage in an incredibly realistic bra…	
2	1	20	One of the most popular online action games of all time Team Fortress Classic feat…	
3	2	30	Enlist in an intense brand of Axis vs. Allied teamplay set in the WWII European Th…	
4	3	40	Enjoy fast-paced multiplayer gaming with Deathmatch Classic (a.k.a. DMC). Valves t…	
5	4	50	Return to the Black Mesa Research Facility as one of the military specialists assi…	
6	5	60	A futuristic action game that challenges your agility as well as your aim Ricochet…	
7	6	70	Named Game of the Year by over 50 publications Valves debut title blends action an…	
8	7	80	With its extensive Tour of Duty campaign a near-limitless number of skirmish modes…	
9	9	130	Made by Gearbox Software and originally released in 2001 as an add-on to Half-Life…	
10	10	220	1998. HALF-LIFE sends a shock through the game industry with its combination of po…	#app_220_note_1
11	11	240	THE NEXT INSTALLMENT OF THE WORLDS # 1 ONLINE ACTION GAME Counter-Strike: Source b…	Just updated to include player stats achievements new scoreboards and more!
12	12	280	Winner of over 50 Game of the Year awards Half-Life set new standards for action g…	
13	13	300	Day of Defeat offers intense online action gameplay set in Europe during WWII. Ass…	#app_300_note_1
14	14	320	Fast multiplayer action set in the Half-Life 2 universe! HL2s physics adds a new d…	
15	15	340	Originally planned as a section of the Highway 17 chapter of Half-Life 2 Lost Coas…	
16	16	360	Half-Life Deathmatch: Source is a recreation of the first multiplayer game set in …	
17	17	380	Half-Life 2 has sold over 4 million copies worldwide and earned over 35 Game of th…	Supports cross-platform Steam Cloud - continue your game on any supported OS. Plus…
18	18	400	Portalâ ¢ is a new single player game from Valve. Set in the mysterious Aperture S…	
19	19	420	Half-LifeÂ® 2: Episode Two is the second in a trilogy of new games created by Valv…	Supports cross-platform Steam Cloud - continue your game on any supported OS.
20	20	440	The most fun you can have online - PC Gamer Is now FREE! Theres no catch! Play as …	Nine distinct classes provide a broad range of tactical abilities and personalitie…
21	21	500	From Valve (the creators of Counter-Strike Half-Life and more) comes Left 4 Dead a…	

games_users:



	index	"ID"	user_id	games
1	0	377160	151603712	Fallout 4
2	2	17390	151603712	Spore
3	4	550	151603712	Left 4 Dead 2
4	6	339800	151603712	HuniePop
5	8	238960	151603712	Path of Exile
6	10	367450	151603712	Poly Bridge
7	12	500	151603712	Left 4 Dead
8	14	440	151603712	Team Fortress 2
9	16	203160	151603712	Tomb Raider
10	18	237990	151603712	The Banner Saga
11	20	8870	151603712	BioShock Infinite
12	22	22370	151603712	Fallout 3 - Game of the Year Edition
13	24	12210	151603712	Grand Theft Auto IV
14	26	200210	151603712	Realm of the Mad God
15	28	298160	151603712	Eldevin
16	30	570	151603712	Dota 2
17	32	7670	151603712	BioShock
18	34	301520	151603712	Robocraft
19	36	4000	151603712	Garry's Mod
20	38	250260	151603712	Jazzpunk
21	40	108710	151603712	Alan Wake

game_tweets:



	index	id	created_at	tweets	user_id	user_location	retweet_count	"Urls"
1	0	12	2018/4/13 22:30	RT @Ashtheimpailer: Hi Guys! ...I am now live playing Counter Strike!...Come say hi!...	2910985258	Living with @kyroskoh (SG)	1	[{'url': 'https://t.co/u8rO4MECZ
2	1	15	2018/4/13 22:24	#_Ø¹Ø¬Ø¨___Ø¬Ø¨__Ø§Ø°Ø§_Ø¹Ø¨ _Ø¹Ø§_ counter strike https://t.co/zCdaiNR3pM	64006246	Tripoli Libya	0	[]
3	2	16	2018/4/13 22:23	Hi Guys! ...I am now live playing Counter Strike!...Come say hi!...https://t.co/u8rO4...	138156328	_ngelholm, Sverige	1	[{'url': 'https://t.co/u8rO4MECZ
4	3	43	2018/4/13 21:47	RT @LunarPandorium: no bully in counter strike plz...#BullyHunters https://t.co/GV4...	133125013	Michigan	16	[]
5	4	57	2018/4/13 21:31	_ NEW: Here's one for our #CSGO fam!... Counter-Strike Team Manager... Moribund E...	856254000000000000	Worldwide	0	[{'url': 'https://t.co/MicIngc4l
6	5	76	2018/4/13 21:00	RT @teamisurus: "Tenemos la oportunidad de dar un gran salto" ...Pino hablÃ³ de t...	968002000000000000	<null>	9	[]
7	6	83	2018/4/13 20:51	RT @MaxGamesAlts: 1 #FAv...Seguir a @MaxGamesAlts - @BreinerAlts - @CarlosAltss -@Cr...	1602191136	Madrid,EspaÃ±a	3	[]
8	7	84	2018/4/13 20:51	RT @MaxGamesAlts: 1 #FAv...Seguir a @MaxGamesAlts - @BreinerAlts - @CarlosAltss -@Cr...	815658000000000000	<null>	3	[]
9	8	85	2018/4/13 20:50	RT @ZmartGames: La historia de Zmart comenzÃ³ gracias a un grupo de amigos que jug...	3420892001	Chile,	13	[]
10	9	88	2018/4/13 20:46	RT @MaxGamesAlts: 1 #FAv...Seguir a @MaxGamesAlts - @BreinerAlts - @CarlosAltss -@Cr...	983848000000000000	<null>	3	[]
11	10	89	2018/4/13 20:46	1 #FAv...Seguir a @MaxGamesAlts - @BreinerAlts - @CarlosAltss -@Cris_Alts -_ https:/...	764588000000000000	Peru	3	[{'url': 'https://t.co/nYwyIxqhU
12	11	101	2018/4/13 3:00	Left 4 Dead - Turtle Rock's last unfinished Left 4 Dead campaign is now available ...	2735559156	<null>	0	[{'url': 'https://t.co/Ny5YKCt87
13	14	102	2018/4/12 15:45	https://t.co/zpDI1c24vX - Buy Left 4 Dead 2 for $0.00 - Over 0% off!...https://t.co...	2736333787	<null>	0	[{'url': 'https://t.co/zpDI1c24v
14	16	103	2018/4/11 22:10	https://t.co/zpDI1c24vX - Buy Left 4 Dead 2 for $0.00 - Over 0% off!...https://t.co...	2736333787	<null>	0	[{'url': 'https://t.co/zpDI1c24v
15	18	104	2018/4/10 12:40	RT @sunnyherring1: Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users ...	210891045	Dallas, TX	2	[]
16	19	105	2018/4/10 12:36	RT @sunnyherring1: Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users ...	2284614762	United States	2	[]
17	20	106	2018/4/10 12:34	RT @sunnyherring1: Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users ...	103999662	DE, DC, MD, VA	1	[]
18	21	107	2018/4/10 12:32	RT @sunnyherring1: Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users ...	4183046769	<null>	1	[]
19	22	108	2018/4/10 12:32	Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users Info & Peddles ...	582140844	USA	0	[{'url': 'https://t.co/uwdoPLpc6
20	23	109	2018/4/10 12:31	Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users Info & Peddles ...	582140844	USA	1	[{'url': 'https://t.co/SGSMLzJTz
21	24	110	2018/4/10 12:31	Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users Info & Peddles ...	582140844	USA	2	[{'url': 'https://t.co/NsX3L98Xd
22	25	111	2018/4/10 12:31	Trump=Attorney-Client Privilege DEAD ...#Facebook=Sells 2B Users Info & Peddles ...	582140844	USA	1	[{'url': 'https://t.co/92VNBU0qS
23	26	112	2018/4/9 21:10	https://t.co/zpDI1bKtEp - Buy Left 4 Dead 2 for $0.00 - Over 0% off!...https://t.co...	2736333787	<null>	0	[{'url': 'https://t.co/zpDI1bKtE
24	28	113	2018/4/8 19:25	https://t.co/zpDI1c24vX - Buy Left 4 Dead 2 for $0.00 - Over 0% off!...https://t.co...	2736333787	<null>	0	[{'url': 'https://t.co/zpDI1c24v

game_hashtags:

0	12	VarityStreaming	Counter-Strike
1	15	_Ø¹Ø¬Ø¨___Ø¬Ø¨__Ø§Ø°Ø§	Counter-Strike
2	16	VarityStreaming	Counter-Strike
3	43	BullyHunters	Counter-Strike
4	57	CSGO	Counter-Strike
5	76	CSGO	Counter-Strike
6	83	FAv	Counter-Strike
7	84	FAv	Counter-Strike
8	85	ThrowbackThursday	Counter-Strike
9	88	FAv	Counter-Strike
10	89	FAv	Counter-Strike

In order to convert SQL database to NoSQL database, we need to denormalization table.

So, games, games_features, games_reviews, games_users will be denormalized to one table; game_tweets, game_hashtags will be denormalized to one table.

In SQL database, use SQL syntax to do social media denormalization:

Sample code:

```
In [6]: import psycopg2
conn = psycopg2.connect(database='postgres',user='postgres', password='')
df_k=pd.read_sql('''select  id,created_at,tweets,user_id,user_locatiom,retweet_count,"Urls","Media",game,array_agg(tag)
from full_tweet_data
group by id,created_at,tweets,user_id,user_locatiom,retweet_count,"Urls","Media",game;''', con=conn)
```

```
In [7]: df_k
```

Out[7]:

	id	created_at	tweets	user_id	user_locatiom	retweet_count	Urls	Media
0	12	2018-04-13 22:30:06	RT @Ashtheimpailer: Hi Guys! \n\nI am now live...	2910985258	Living with @kyroskoh (SG)	1	[{'url': 'https://t.co/u8rO4MECZc', 'expanded_...	None
1	15	2018-04-13 22:24:03	#ÙØ¹Ø¬Ø'ÙÙ_Ø-Ø'Ù_Ø§Ø°Ø§\nØ°ÙØ'Ø¨ ÙØ¹Ø§Ù...	64006246	Tripoli Libya	0	[]	[{'id': 984920156047540230, 'id_str': '9849201...
2	16	2018-04-13 22:23:01	Hi Guys! \n\nI am now live playing Counter Str...	138156328	Ängelholm, Sverige	1	[{'url': 'https://t.co/u8rO4MECZc', 'expanded_...	None
3	43	2018-04-13 21:47:31	RT @LunarPandorium: no bully in counter strike...	133125013	Michigan	16	[]	[{'id': 984478243385368576, 'id_str': '9844782...
4	57	2018-04-13 21:31:01	â¡ NEW: Here's one for our #CSGO fam!\n\nð¼...	856254411173179394	Worldwide	0	[{'url': 'https://t.co/MicIngc4lp', 'expanded_...	None
5	76	2018-04-13	RT @teamisurus: "Tenemos la	968002038318723072	None	9	[]	None

Then convert it to a csv file and then use mongoimport command to import data to mongoDB.

Sample code:

```
In [8]: df_k.to_csv('full_tweets_data.csv', encoding='utf-8', index=False)
```

```
In [11]: twitter_file = pd.read_csv('/Users/chen/Desktop/data/full_tweets_data.csv', encoding='latin1')
```

```
In [12]: twitter_file.duplicated().sum()
```

```
Out[12]: 0
```

```
In [13]: twitter_csv_path='/Users/chen/Desktop/data/full_tweets_data.csv'
```

```
In [14]: db_name='local'
coll_name1='game_twitter'
```

```
In [15]: mongoimport(twitter_csv_path, db_name, coll_name1)
```

```
/Users/chen/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:11: DeprecationWarning: remove is deprecated.
Use delete_one or delete_many instead.
  # This is added back by InteractiveShellApp.init_path()
/Users/chen/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:12: DeprecationWarning: insert is deprecated.
Use insert_one or insert_many instead.
  if sys.path[0] == '':
```

```
Out[15]: 3261
```

In SQL database, use SQL syntax to do game data denormalization:

Sample code:

First, denormalization 4 table to one table.

In [5]:
```
import psycopg2
conn = psycopg2.connect(database='postgres',user='postgres', password='')
df1=pd.read_sql('''
SELECT A."Name", B.*, C."AboutText",C."ShortDescrip", C."Reviews", D.user_id
FROM
  (
    (SELECT "ID","Name" FROM games) A
    FULL OUTER JOIN
    (SELECT * FROM games_features) B on A."ID" = B."ResponseID"
    FULL OUTER JOIN
    (SELECT * FROM games_reviews) C on A."ID" = C."ResponseID"
    FULL OUTER JOIN
    (SELECT * FROM games_users) D on A."ID" = D."ID"
  );''', con=conn)
```

In [10]:
```
df1
```

Out[10]:

	Name	index	ResponseID	ReleaseDate	Metacritic	RecommendationCount	IsFree	GenreIsNonGame	GenreIsIndie	GenreIsAction	...	GenreIsEar
0	Fallout 4	6635.0	377160.0	Nov 9 2015	84.0	72929.0	False	False	False	False	...	
1	Left 4 Dead 2	21.0	550.0	Nov 16 2009	89.0	140726.0	False	False	False	True	...	
2	HuniePop	5043.0	339800.0	Jan 19 2015	0.0	9827.0	False	False	True	False	...	
3	Path of Exile	2147.0	238960.0	Oct 23 2013	86.0	33124.0	True	False	True	True	...	

Then use group command to correct the 1 to n relationship:

In [13]:
```
df2=pd.read_sql('''
select  "Name","ResponseID","ReleaseDate","Metacritic","RecommendationCount","IsFree",
        "GenreIsAction","GenreIsAdventure","GenreIsCasual","GenreIsEarlyAccess","GenreIsRPG",
        "GenreIsStrategy","GenreIsSimulation","GenreIsFreeToPlay","GenreIsSports","GenreIsRacing",
       "GenreIsMassivelyMultiplayer","PriceInitial","AboutText","ShortDescrip","Reviews",
       array_agg(user_id)
from fulldata1
group by "Name","ResponseID","ReleaseDate","Metacritic","RecommendationCount","IsFree",
         "GenreIsAction","GenreIsAdventure","GenreIsCasual","GenreIsEarlyAccess","GenreIsRPG",
         "GenreIsStrategy","GenreIsSimulation","GenreIsFreeToPlay","GenreIsSports","GenreIsRacing",
         "GenreIsMassivelyMultiplayer","PriceInitial","AboutText","ShortDescrip","Reviews"''', con=conn)
```

In [14]:
```
df2
```

Out[14]:

	Name	ResponseID	ReleaseDate	Metacritic	RecommendationCount	IsFree	GenreIsAction	GenreIsAdventure	GenreIs
0	The Count of Monster Disco	277751.0	Oct 24 2014	0.0	0.0	False	False	False	
1	ÃâšÃâ¥_ÃâŠ__ÃâšÃâ¤Ãâ_Ãâ¥ÃâÃâ¤Ãâ©ÃâÂ...	490390.0	Jul 1 2016	0.0	259.0	False	False	True	
2	! That Bastard Is Trying To Steal Our Gold !	449940.0	Mar 3 2016	0.0	0.0	False	True	True	
3	"Glow Ball" - The billiard puzzle game	388390.0	Jul 20 2015	0.0	0.0	False	False	False	
4	"Heroes of Card War"	512560.0	10-2016	0.0	0.0	False	False	False	

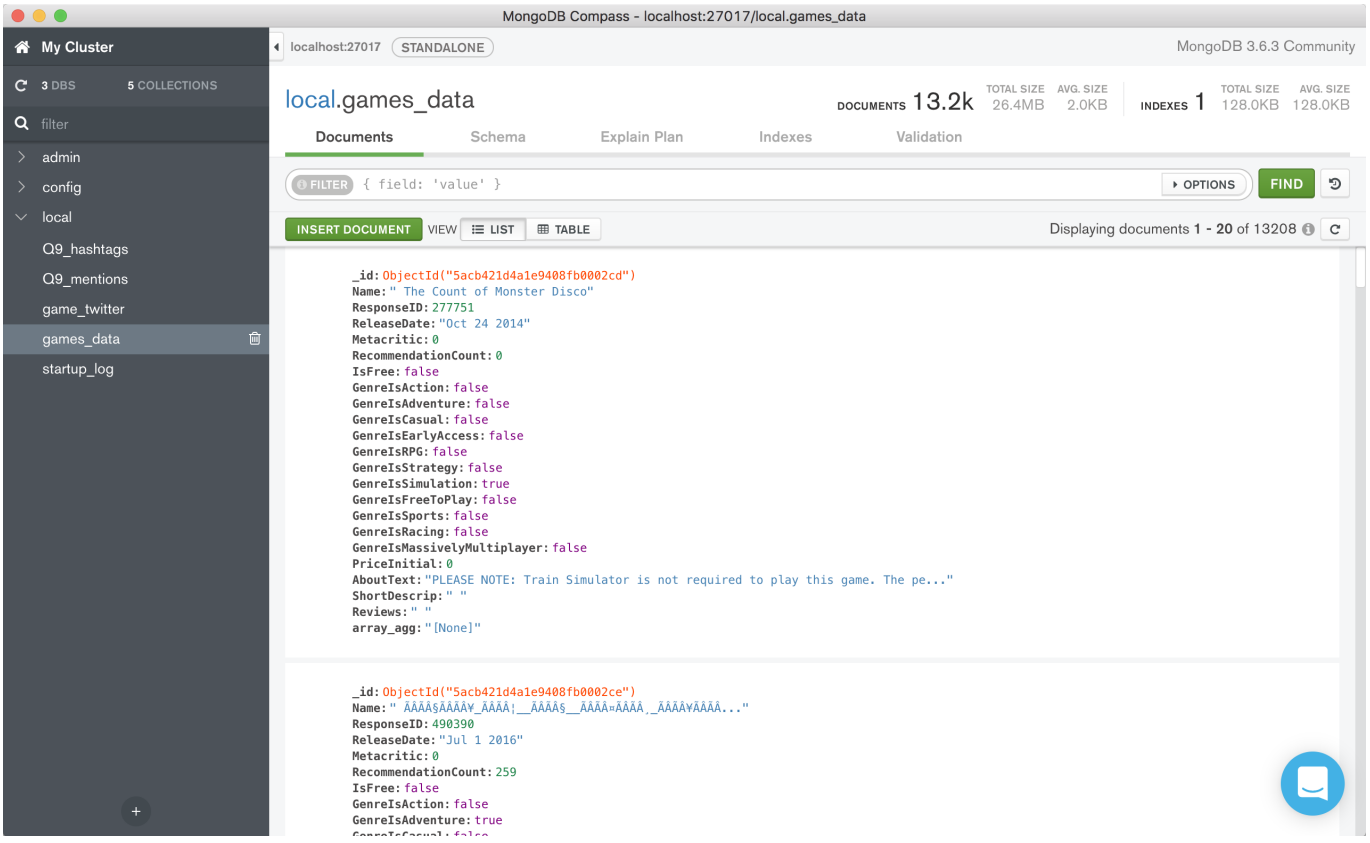
In [15]:
```
df2.to_csv('fulldata2.csv', encoding='utf-8', index=False)
```

Then convert it to a csv file and then use mongoimport command to import data to mongoDB.

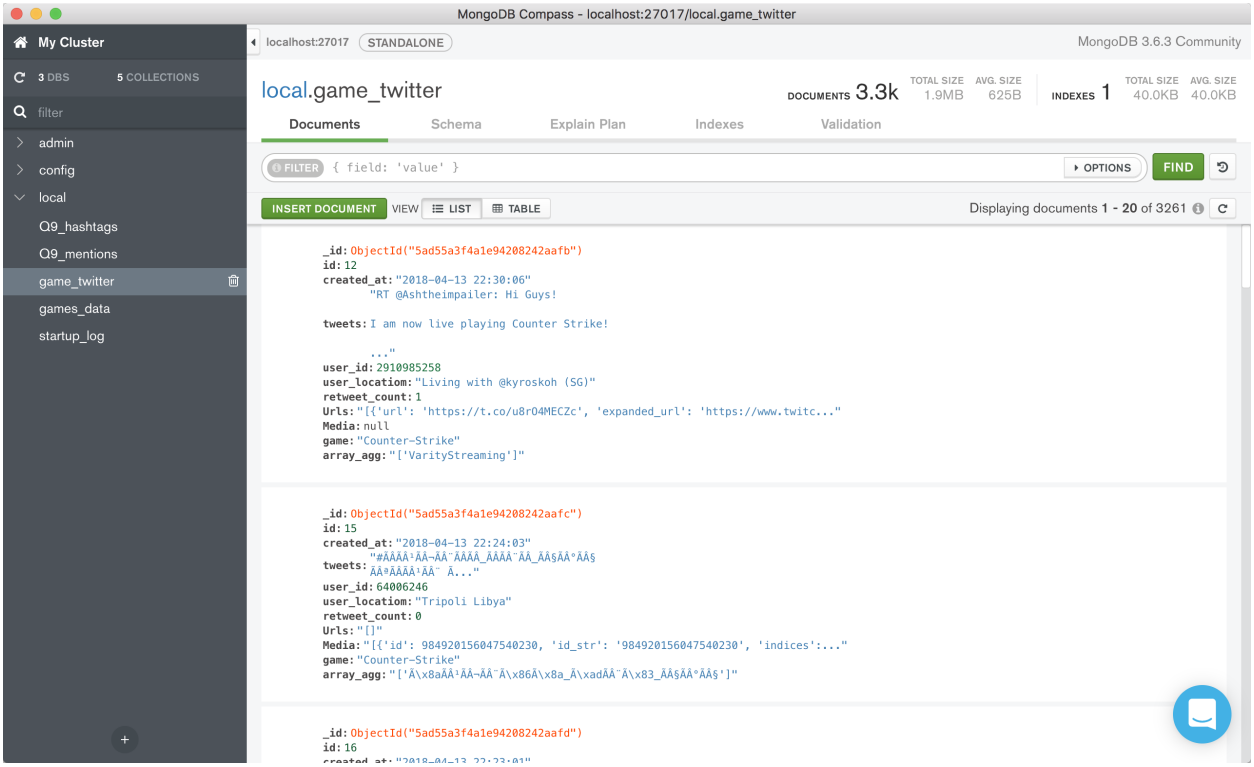
```
In [20]: gamedata_csv_path='/Users/chen/Desktop/steam-data/datafile/data1/fulldata2.csv'
         db_name='local'
         coll_name3='games_data'

In [21]: mongoimport(gamedata_csv_path, db_name, coll_name3)

         /Users/chen/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:11: DeprecationWarning: remove is deprecated.
         Use delete_one or delete_many instead.
           # This is added back by InteractiveShellApp.init_path()
         /Users/chen/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:12: DeprecationWarning: insert is deprecated.
         Use insert_one or insert_many instead.
           if sys.path[0] == '':

Out[21]: 13208
```

Sample data in mongoDB:

games:



Social Media data:



Question Sample Code:

Answer Questions

In [16]:
```
from pymongo import MongoClient
client = MongoClient('localhost', 27017)
```

In [17]:
```
db = client.local
collection=db.game_twitter
```

In [43]:
```
pprint.pprint(collection.find_one({"game": "Dota 2"}))
```

```
{'Media': None,
 'Urls': "[{'url': 'https://t.co/u8rO4MECZc', 'expanded_url': "
         "'https://www.twitch.tv/sagemblack', 'display_url': "
         "'twitch.tv/sagemblack', 'indices': [84, 107]}]",
 '_id': ObjectId('5ad55a3f4a1e94208242add2'),
 'array_agg': "['VarityStreaming']",
 'created_at': '2018-04-13 22:30:06',
 'game': 'Dota 2',
 'id': 961,
 'retweet_count': 1,
 'tweets': 'RT @Ashtheimpailer: Hi Guys! \n'
           '\n'
           'I am now live playing Counter Strike!\n'
           '\n'
           'Come say hi!\n'
           '\n'
           'https://t.co/u8rO4MECZc\n'
           '\n'
           '#VarityStreaming \n'
           '#SmallStreamÃ¢Â\x80Â¦',
 'user_id': 2910985258,
 'user_locatiom': 'Living with @kyroskoh (SG)'}
```

In [21]:
```
#Question1 What are tags are associated with a person, place or thing
list(db.game_twitter.aggregate([
    {"$match":{"game":"Dota 2"}},
    {"$group":{"_id":"$array_agg","count":{"$sum":1}}},
    {"$sort":{"count":-1}}]))
```

```
 {'_id': "['DanaDurnford', 'BanTokyo2020', 'StopTokyo2020']", 'count': 13},
 {'_id': "['quake']", 'count': 12},
 {'_id': "['Divinity']", 'count': 12},
 {'_id': "['Call']", 'count': 9},
 {'_id': "['Fallout']", 'count': 9},
 {'_id': "['Facebook']", 'count': 8},
 {'_id': "['Fallout', 'ThrowbackThursday']", 'count': 7},
 {'_id': "['Divinity', 'PS4']", 'count': 6},
 {'_id': "['DivinityOriginalSin2', 'XboxOne', 'Playstation4']", 'count': 6},
 {'_id': "['Dota']", 'count': 6},
 {'_id': "['Dota', 'Steam']", 'count': 6},
 {'_id': "['Quake', 'Quake', 'DadJokes']", 'count': 6},
 {'_id': "['Streamhype', 'supportsmallstreamers', 'Call']", 'count': 5},
 {'_id': "['Left', 'Dead']", 'count': 5},
 {'_id': "['QuakeChampions']", 'count': 4},
 {'_id': "['Repost', 'BandaiNamco']", 'count': 4},
 {'_id': "['FAv']", 'count': 4},
 {'_id': "['Repost', 'BandaiNamco', 'PS4']", 'count': 4},
 {'_id': "['dota', 'meepo', 'techies', 'SupportSmallerStreamers']",
  'count': 4},
```

In [22]:
```
list(db.game_twitter.aggregate([
    {"$match":{"game":"Left 4 Dead"}},
    {"$group":{"_id":"$array_agg","count":{"$sum":1}}},
    {"$sort":{"count":-1}}]))
```

Out[22]:
```
[{'_id': "['DOTA']", 'count': 56},
 {'_id': "['earthquake', 'quake']", 'count': 20},
 {'_id': "['Divinity', 'XboxOne']", 'count': 13},
 {'_id': "['DanaDurnford', 'BanTokyo2020', 'StopTokyo2020']", 'count': 13},
 {'_id': "['quake']", 'count': 12},
 {'_id': "['Divinity']", 'count': 12},
 {'_id': "['Call']", 'count': 9},
 {'_id': "['Fallout']", 'count': 9},
 {'_id': "['Facebook']", 'count': 8},
 {'_id': "['DivinityOriginalSin2', 'XboxOne', 'Playstation4']", 'count': 7},
 {'_id': "['Fallout', 'ThrowbackThursday']", 'count': 7},
 {'_id': "['Divinity', 'PS4']", 'count': 6},
 {'_id': "['Dota']", 'count': 6},
 {'_id': "['Quake', 'Quake', 'DadJokes']", 'count': 6},
 {'_id': "['Repost', 'BandaiNamco', 'PS4']", 'count': 5},
 {'_id': "['Dota', 'Steam']", 'count': 5},
 {'_id': "['Streamhype', 'supportsmallstreamers', 'Call']", 'count': 5},
```

In [28]:
```
#Question2 What social media users are like other social media users in your domain?
match = {
    'tweets':{"$regex": "Counter Strike"}
}

group = {
    '_id':"$user_id"
}

ret2 = db.game_twitter.aggregate([
    {'$match':match},
    {'$group': group}
])

list(ret2)
```

Out[28]:
```
[{'_id': 138156328}, {'_id': 2910985258}]
```

In [23]:
```
#Question3 What people, places or things are popular in your domain
list(db.game_twitter.aggregate([
    {"$group":{"_id":"$array_agg","count":{"$sum":1}}},
    {"$sort":{"count":-1}}]))
```

Out[23]:
```
[{'_id': "['DOTA']", 'count': 504},
 {'_id': "['earthquake', 'quake']", 'count': 139},
 {'_id': "['quake']", 'count': 108},
 {'_id': "['Divinity']", 'count': 103},
 {'_id': "['Divinity', 'XboxOne']", 'count': 95},
 {'_id': "['Fallout']", 'count': 81},
 {'_id': "['Call']", 'count': 81},
 {'_id': "['DanaDurnford', 'BanTokyo2020', 'StopTokyo2020']", 'count': 79},
 {'_id': "['Facebook']", 'count': 72},
 {'_id': "['Dota']", 'count': 54},
 {'_id': "['Fallout', 'ThrowbackThursday']", 'count': 46},
 {'_id': "['Divinity', 'PS4']", 'count': 46},
 {'_id': "['Dota', 'Steam']", 'count': 43},
 {'_id': "['DivinityOriginalSin2', 'XboxOne', 'Playstation4']", 'count': 41},
 {'_id': "['quake', 'earthquake']", 'count': 41},
 {'_id': "['Quake', 'Quake', 'DadJokes']", 'count': 40},
 {'_id': "['QuakeChampions']", 'count': 36},
 {'_id': "['FAv']", 'count': 36},
```

In [33]:
```
#Question4 What people, places or things are trending in your domain? (A trend is popularity over time.)
list(db.game_twitter.aggregate([
    {"$sort":{"created_at":1}}]))
```

```
 {'_id': ObjectId('5ad55a3f4a1e94208242ad91'),
  'id': 885,
  'created_at': '2018-04-03 12:00:03',
  'tweets': 'Double Cross @larianstudios #Divinity Original Sin 2 https://t.co/t6sWDWNIHn',
  'user_id': 754208650319630337,
  'user_locatiom': 'Japan',
  'retweet_count': 0,
  'Urls': "[{'url': 'https://t.co/t6sWDWNIHn', 'expanded_url': 'http://youtu.be/7P7gbUfBJmw?a', 'display_url': 'youtu
.be/7P7gbUfBJmw?a', 'indices': [53, 76]}]",
  'Media': None,
  'game': 'Left 4 Dead',
  'array_agg': "['Divinity']"},
 {'_id': ObjectId('5ad55a3f4a1e94208242aefd'),
  'id': 1360,
  'created_at': '2018-04-03 12:00:03',
  'tweets': 'Double Cross @larianstudios #Divinity Original Sin 2 https://t.co/t6sWDWNIHn',
  'user_id': 754208650319630337,
  'user_locatiom': 'Japan',
  'retweet_count': 0,
  'Urls': "[{'url': 'https://t.co/t6sWDWNIHn', 'expanded_url': 'http://youtu.be/7P7gbUfBJmw?a', 'display_url': 'youtu
```

In [80]:
```
for doc in collection.find({},{"game":1, "_id":0, "created_at":1}).sort('created_at', pymongo.ASCENDING):
    print(doc)
```

```
{'created_at': '2018-04-03 12:00:03', 'game': 'Counter-Strike'}
{'created_at': '2018-04-03 12:00:03', 'game': 'Left 4 Dead'}
{'created_at': '2018-04-03 12:00:03', 'game': 'Dota 2'}
{'created_at': '2018-04-03 12:00:03', 'game': 'Empire: Total War'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Counter-Strike'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Left 4 Dead'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Dota 2'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Fallout 3'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Empire: Total War'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Call of Duty'}
{'created_at': '2018-04-03 15:44:53', 'game': 'NBA 2K13'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Divinity 2'}
{'created_at': '2018-04-03 15:44:53', 'game': 'Quake'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Counter-Strike'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Left 4 Dead'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Dota 2'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Fallout 3'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Empire: Total War'}
{'created_at': '2018-04-03 17:50:05', 'game': 'Call of Duty'}
```

In [84]:
```
list(db.game_twitter.aggregate([
    {"$match":{"game":"Left 4 Dead"}},
    {"$sort":{"created_at":1}}]))
```

Out[84]:
```
[{'_id': ObjectId('5ad55a3f4a1e94208242ad91'),
  'id': 885,
  'created_at': '2018-04-03 12:00:03',
  'tweets': 'Double Cross @larianstudios #Divinity Original Sin 2 https://t.co/t6sWDWNIHn',
  'user_id': 754208650319630337,
  'user_locatiom': 'Japan',
  'retweet_count': 0,
  'Urls': "[{'url': 'https://t.co/t6sWDWNIHn', 'expanded_url': 'http://youtu.be/7P7gbUfBJmw?a', 'display_url': 'youtu
.be/7P7gbUfBJmw?a', 'indices': [53, 76]}]",
  'Media': None,
  'game': 'Left 4 Dead',
  'array_agg': "['Divinity']"},
 {'_id': ObjectId('5ad55a3f4a1e94208242ad90'),
  'id': 884,
  'created_at': '2018-04-03 15:44:53',
  'tweets': "Je viens de m'acheter #Divinity Original Sin 2. J'ai fait le premier sur PS4 et j'ai adorÃ\x83Â©... j'es
pere que je vaisÃ¢Â\x80Â¦ https://t.co/GEpMQNxzxK",
  'user_id': 2706780463,
  'user_locatiom': 'Nowhere',
```

In [92]:
```
#Question5 use case
for doc in collection.find({},{"game":1, "_id":0, "tweets":1}).sort('game', pymongo.ASCENDING):
    print(doc)
```

...

In [105]:
```
list(collection.find({"game": "Empire: Total War"}).limit(10))
```

Out[105]:
```
[{'_id': ObjectId('5ad55a3f4a1e94208242b0a8'),
  'id': 1910,
  'created_at': '2018-04-13 22:30:06',
  'tweets': 'RT @Ashtheimpailer: Hi Guys! \n\nI am now live playing Counter Strike!\n\nCome say hi!\n\nhttps://t.co/u
8rO4MECZc\n\n#VarityStreaming \n#SmallStreamÃ¢Â\x80Â¦',
  'user_id': 2910985258,
  'user_locatiom': 'Living with @kyroskoh (SG)',
  'retweet_count': 1,
  'Urls': "[{'url': 'https://t.co/u8rO4MECZc', 'expanded_url': 'https://www.twitch.tv/sagemblack', 'display_url': 'tw
itch.tv/sagemblack', 'indices': [84, 107]}]",
  'Media': None,
  'game': 'Empire: Total War',
  'array_agg': "['VarityStreaming']"},
 {'_id': ObjectId('5ad55a3f4a1e94208242b0a9'),
  'id': 1913,
  'created_at': '2018-04-13 22:24:03',
  'tweets': '#Ã\x99Â\x8aÃ\x98Â¹Ã\x98Â¬Ã\x98Â¨Ã\x99Â\x86Ã\x99Â\x8a_Ã\x98Â\xadÃ\x98Â¨Ã\x99Â\x83_Ã\x98Â§Ã\x98Â°Ã\x98Â§\n
```

In [106]:
```
#return this cursor with a limit of 5 and skip of 20 applied
list(collection.find()[20:25])
```

Out[106]:
```
[{'_id': ObjectId('5ad55a3f4a1e94208242ab0f'),
  'id': 110,
  'created_at': '2018-04-10 12:31:23',
  'tweets': 'Trump=Attorney-Client Privilege DEAD \n#Facebook=Sells 2B Users Info &amp;amp; Peddles 2 Doctors\nCA Legislature=Ban FreeSpeeÃ¢Â\x80Â¦ https://t.co/NsX3L98XdC',
  'user_id': 582140844,
  'user_locatiom': 'USA',
  'retweet_count': 2,
  'Urls': "[{'url': 'https://t.co/NsX3L98XdC', 'expanded_url': 'https://twitter.com/i/web/status/983683862965489665', 'display_url': 'twitter.com/i/web/status/9Ã¢Â\x80Â¦', 'indices': [121, 144]}]",
  'Media': None,
  'game': 'Counter-Strike',
  'array_agg': "['Facebook']"},
 {'_id': ObjectId('5ad55a3f4a1e94208242ab10'),
  'id': 111,
  'created_at': '2018-04-10 12:31:02',
  'tweets': 'Trump=Attorney-Client Privilege DEAD \n#Facebook=Sells 2B Users Info &amp;amp; Peddles 2 Doctors\nCA Legislature=Ban FreeSpeeÃ¢Â\x80Â¦ https://t.co/92VNBU0qS0',
  'user_id': 582140844,
```

In [99]:
```
list(collection.find({
    "created_at": {
        "$gte":"2018-04-10 12:00:06",
        "$lte": "2018-04-10 12:30:06"
    }
}))
```

Out[99]:
```
[{'_id': ObjectId('5ad55a3f4a1e94208242ab8b'),
  'id': 239,
  'created_at': '2018-04-10 12:01:26',
  'tweets': 'Fallout 3 in Fallout 4 mod cancelled due to voice acting copyright https://t.co/w3QtPSttd7 #Fallout #Fallout',
  'user_id': 901492458307112962,
  'user_locatiom': None,
  'retweet_count': 0,
  'Urls': "[{'url': 'https://t.co/w3QtPSttd7', 'expanded_url': 'https://capitalwasteland.com/#cancellation-announcement-section', 'display_url': 'capitalwasteland.com/#cancellation-Ã¢Â\x80Â¦', 'indices': [67, 90]}]",
  'Media': None,
  'game': 'Counter-Strike',
  'array_agg': "['Fallout', 'Fallout']"},
 {'_id': ObjectId('5ad55a3f4a1e94208242abd8'),
```