

## **INFO 6210**

### **Data Management and Database Design**

#### **Adding Social Media and Tags Assignment 2**

#### **Assignment 2**

I chose my domain as European soccer. I scraped the dataset from Kaggle. <https://www.kaggle.com/hugomathien/soccer>. In this assignment, I have 3 entities that represent people( Players ), places( Country ) and things( Teams ).

First of all, I decided to make #Lionel Messi, #Cristiano Ronaldo, #Neymar, #Wayne Rooney, #Mario Balotelli, #Andrea Pirlo, #Gareth Bale, #Robert Lewandowski as the tags of soccer players.

Also, by searching social media users' mentions and hashtags, we can compare they similarity.

From these analyses, we can realize that Lionel Messi is a very popular soccer players, because he has a lot of tags and mention times. Moreover, Lionel Messi, LaLiga, Barça Girona are trending in the domain.

These are some testing results of my working.

#### measure popularity of above soccer players

```
In [75]: conn = psycopg2.connect(database='postgres', user='postgres', password='')
cur = conn.cursor()
df1=pd.read_sql('''select tag as tag, count(*) AS times from EuropeanSoccer_tags WHERE team='Lionel Messi' GROUP BY tag ORDER BY times DESC''', con=conn)
df1
```

```
Out[75]:   tag    times
0   Lionel      32
1     Messi      23
2    lionel      19
3    LaLiga      14
4    Girona       9
5    Barca       9
6  BarçaGirona      8
7  Barcelona      8
8  ForcaBarca      8
9 LaLigaSantander      7
10    messi       3
```

#### similarity

```
In [82]: user1 = api.get_user('footy90com')
print(user1.name, user1.screen_name, str(user1.statuses_count), str(user1.friends_count), str(user1.followers_count))

footy90.com footy90com 127875 16451 26163
```

```
In [83]: from datetime import datetime, date, time, timedelta
tweets = user1.statuses_count
account_created_date = user1.created_at
delta = datetime.utcnow() - account_created_date
account_age_days = delta.days
print("Account age (in days): " + str(account_age_days))
if account_age_days > 0:
    print("Average tweets per day: " + "%.2f" % (float(tweets) / float(account_age_days)))
```

```
Account age (in days): 849
Average tweets per day: 150.62
```

```
In [87]: from tweepy import Cursor
hashtags = []
mentions = []
tweet_count = 0
end_date = datetime.utcnow() - timedelta(days=30)
for status in Cursor(api.user_timeline, id='footy90com').items():
    tweet_count += 1
    if hasattr(status, "entities"):
        entities = status.entities
        if "hashtags" in entities:
            for ent in entities["hashtags"]:
                if ent is not None:
                    if "text" in ent:
```

```
In [89]: mentions
```

```
Out[89]: []
```

```
In [91]: from collections import Counter
for user1, count in Counter(hashtags).most_common(10):
    print(user1 + "\t" + str(count))
```

```
Premier_League 429
Süper_Lig 313
Bundesliga 289
Championship 277
LaLiga 252
MLS 233
Serie A 207
```

```
In [98]: hashtags2
Out[98]: ['Trump',
 'Trump',
 'Parkland',
 'Afghanistan',
 'Afghanistan',
 'Turkish',
 'Erdogan',
 'Trump',
 'ISIS',
 'ISIS',
 'Alex',
 'Sanchez',
 'Alex',
 'Sanchez',
 'Brighton',
 'tattooist',
 'India',
 'Arsenal',
 'Arsenal',
 'Gunners',
 ...]
```

```
In [100]: mentions2
Out[100]: ['realDonaldTrump',
 'ForeignAffairs',
 'DailyMirror',
 'bet365',
 'Abdiwaasa_h',
 'AFP',
 'MailSport',
 'MailSport',
 'TheInsiderUG',
 'WashTimes',
 'HevTammvBruce']
```

## Sample Data

### Country

	index	id	name
1	0	1	Belgium
2	1	1729	England
3	2	4769	France
4	3	7809	Germany
5	4	10257	Italy
6	5	13274	Netherlands
7	6	15722	Poland
8	7	17642	Portugal
9	8	19694	Scotland
10	9	21518	Spain
11	10	24558	Switzerland
12	0	1	Belgium
13	1	1729	England
14	2	4769	France
15	3	7809	Germany
16	4	10257	Italy
17	5	13274	Netherlands
18	6	15722	Poland
19	7	17642	Portugal
20	8	19694	Scotland
21	9	21518	Spain
22	10	24558	Switzerland
23	0	1	Belgium
24	1	1729	England
25	2	4769	France
26	3	7809	Germany
27	4	10257	Italy
28	5	13274	Netherlands
29	6	15722	Poland
30	7	17642	Portugal
31	8	19694	Scotland
32	9	21518	Spain
33	10	24558	Switzerland

## Team

	index	id	team_api_id	team_long_name	team_short_name
1	0	1	9987	KRC Genk	GEN
2	1	2	9993	Beerschot AC	BAC
3	2	3	10000	SV Zulte-Waregem	ZUL
4	3	4	9994	Sporting Lokeren	LOK
5	4	5	9984	KSV Cercle Brugge	CEB
6	5	6	8635	RSC Anderlecht	AND
7	6	7	9991	KAA Gent	GEN
8	7	8	9998	RAEC Mons	MON
9	8	9	7947	FCV Dender EH	DEN
10	9	10	9985	Standard de Lige	STL
11	10	11	8203	KV Mechelen	MEC
12	11	12	8342	Club Brugge KV	CLB
13	12	13	9999	KSV Roeselare	ROS
14	13	14	8571	KV Kortrijk	KOR
15	14	15	4049	Tubize	TUB
16	15	16	9996	Royal Excel Mouscron	MOU
17	16	17	10001	KVC Westerlo	WES
18	17	18	9986	Sporting Charleroi	CHA
19	18	614	9997	Sint-Truidense WV	STT
20	19	1034	9989	Lierse SK	LIE
21	20	1042	6351	KAS Eupen	EUP
22	21	1513	1773	Oud-Heverlee Leuven	O-H
23	22	2004	8475	Waasland-Beveren	WAA
24	23	2476	8573	KV Oostende	OOS
25	24	2510	274581	Royal Excel Mouscron	MOP
26	25	3457	10260	Manchester United	MUN
27	26	3458	10261	Newcastle United	NEW
28	27	3459	9825	Arsenal	ARS
29	28	3460	8659	West Bromwich Albion	WBA
30	29	3461	8472	Sunderland	SUN
31	30	3462	8650	Liverpool	LIV
32	31	3463	8654	West Ham United	WHU
33	32	3464	8528	Wigan Athletic	WIG
34	33	3465	10252	Aston Villa	AVL
35	34	3466	8456	Manchester City	MCI
36	35	3467	8668	Everton	EVE
37	36	3468	8655	Blackburn Rovers	BLB
38	37	3469	8549	Middlesbrough	MID
39	38	3470	8586	Tottenham Hotspur	TOT
40	39	3471	8559	Bolton Wanderers	BOL
41	40	3472	10194	Stoke City	STK
42	41	3473	8667	Hull City	HUL
43	42	3474	9879	Fulham	FUL
44	43	3475	8455	Chelsea	CHE
45	44	3476	8462	Portsmouth	POR

## Player

	index	id	player_api_id	player_name	player_fifa_api_id	birthday	height	weight
1	0	1	505942	Aaron Appindangoye	218353	1992-02-29 00:00:00	182.88	187
2	1	2	155782	Aaron Cresswell	189615	1989-12-15 00:00:00	170.18	146
3	2	3	162549	Aaron Doran	186170	1991-05-13 00:00:00	170.18	163
4	3	4	30572	Aaron Galindo	140161	1982-05-08 00:00:00	182.88	198
5	4	5	23780	Aaron Hughes	17725	1979-11-08 00:00:00	182.88	154
6	5	6	27316	Aaron Hunt	158138	1986-09-04 00:00:00	182.88	161
7	6	7	564793	Aaron Kuhl	221280	1996-01-30 00:00:00	172.72	146
8	7	8	30895	Aaron Lennon	152747	1987-04-16 00:00:00	165.1	139
9	8	9	528212	Aaron Lennox	206592	1993-02-19 00:00:00	190.5	181
10	9	10	101042	Aaron Meijers	188621	1987-10-28 00:00:00	175.26	170

## Tags associated with soccer players, location and time.

tag	name	time	location
Chivas	10_soccerplayer	2018-03-01 01:57:40	
Cibao	10_soccerplayer	2018-03-01 01:57:40	
Chivas	10_soccerplayer	2018-03-01 01:57:30	
SCCL2018	10_soccerplayer	2018-03-01 01:05:30	
TimCup	CalciatoriHot	2018-03-01 00:19:42	Italia
leonardobonucci	CalciatoriHot	2018-03-01 00:19:42	Italia
PatrickCutrone	CalciatoriHot	2018-03-01 00:19:42	Italia
soccerplayer	CalciatoriHot	2018-03-01 00:19:42	Italia
shirtless	CalciatoriHot	2018-03-01 00:19:42	Italia
AcMilan	CalciatoriHot	2018-03-01 00:19:42	Italia
TimCup	CalciatoriHot	2018-03-01 00:09:08	Italia
ACMilan	CalciatoriHot	2018-03-01 00:09:08	Italia

Top 100 tags related to each European soccer player.

europansoccer\_tags [postgres@localhost] × Tab-se...d (TSV)

<Filter criteria>

tag	team
lionel	Lionel Messi
lionel	Lionel Messi
messi	Lionel Messi
wallpaper	Lionel Messi
wallpapers	Lionel Messi
high	Lionel Messi
quality	Lionel Messi
lionel	Lionel Messi
lionel	Lionel Messi
10	Lionel Messi
Lionel	Lionel Messi
Messi	Lionel Messi
Harus	Lionel Messi
Dilarang	Lionel Messi
Karena	Lionel Messi
Bukan	Lionel Messi
Manusia	Lionel Messi
MésQueUnClub	Lionel Messi
ViscaBarca	Lionel Messi
FCB	Lionel Messi
Messi	Lionel Messi
Lionel	Lionel Messi
LeoMessi	Lionel Messi
Messi10	Lionel Messi
robert	Lionel Messi
bayern	Lionel Messi
lionel	Lionel Messi
GironaFC	Lionel Messi
LaLiga	Lionel Messi
lionel	Lionel Messi
FCBarcelona	Lionel Messi
LaLiga	Lionel Messi
lionel	Lionel Messi
lionel	Lionel Messi

## Tweets contents

	screen_name	created_at	text
1	C_Ronaldo_Agent	2018-02-28 23:39:13	What Lionel Messi said about Cristiano Ronaldo before th...
2	PowerBroadcast	2018-02-28 13:48:02	Lionel messi wallpaper wallpapers high quality <a href="https://...">https://...</a>
3	C_Ronaldo_Agent	2018-02-28 11:14:15	Football legend Pele settles the Cristiano Ronaldo and L...
4	C_Ronaldo_Agent	2018-02-28 11:14:14	Football legend Pele settles the Cristiano Ronaldo and L...
5	LintasCoid	2018-02-28 03:49:43	#Lionel #Messi #Harus #Dilarang Main Oleh FIFA #Karena D...
6	barcanewsonly	2018-02-27 12:16:17	#MésQueUnClub #ViscaBarca #FCB #Messi #Lionel #LeoMessi ...
7	footy90com	2018-02-27 03:43:12	Football News: <a href="#">#robert #bayern #lione... https://t.co/K...</a>
8	LeoMessi_Agent	2018-02-26 13:35:06	Lionel Messi&#039;s individual highlights vs Girona ...
9	LeoMessi_Agent	2018-02-26 13:35:06	Lionel Messi&#039;s individual highlights vs Girona ...
10	LeoMessi_Agent	2018-02-26 05:23:07	Barcelona news: Lionel Messi makes Cristiano Ronaldo adm...
11	footy90com	2018-02-26 00:12:19	Lionel Messi's Ronaldinho-esque free kick surprised Barc...
12	footy90com	2018-02-26 00:00:25	Lionel Messi enjoys record-breaking day as Barcelona win...
13	C_Ronaldo_Agent	2018-02-25 21:16:14	Barcelona news: Lionel Messi makes Cristiano Ronaldo adm...
14	footy90com	2018-02-25 13:03:30	Lionel Messi goals in the Champions League: The teams Ba...
15	DergisiPlatform	2018-02-25 13:01:27	#Lionel #Messi, #LaLiga da yine #tarih yazdi <a href="https://t...">https://t...</a>
16	barcanewsonly	2018-02-25 12:16:05	#MésQueUnClub #ViscaBarca #FCB #Messi #Lionel #LeoMessi ...
17	footy90com	2018-02-25 11:01:24	Lionel Messi never fails to surprise us, says Barcelona ...
18	HotTuxfr	2018-02-25 08:52:07	La playmate Iryna Ivanova fait grimper la température av...
19	HotTuxfr	2018-02-25 00:52:00	La playmate Iryna Ivanova fait grimper la température av...
20	LtzurboiChris	2018-02-25 00:00:35	#G.O.A.T# Greatest Of All Time #Lionel Messi <a href="https://t.c...">https://t.c...</a>
21	GonzalezNacho	2018-02-24 22:47:05	El mejor el partido #Lionel Messi
22	mexnewz	2018-02-24 22:44:16	#Lionel es extraterrestre #BarcaGirona 6-1 en la #LaLiga...
23	Tufairi	2018-02-24 22:28:38	@MailSport Meanwhile, #Lionel #Messi has been RATED 11/1...
24	Tufairi	2018-02-24 22:10:52	@marca @pablogegea MEANWHILE, #Lionel #Messi was RATED 11...
25	footy90com	2018-02-24 22:05:56	Lionel Messi: Barcelona star breaks two records against ...
26	Tufairi	2018-02-24 21:59:41	#ADONBILIVIT! #Lionel #Messi RATED 11/10. WHY? <a href="#">"This ...</a>
27	Tufairi	2018-02-24 21:59:34	#ADONBILIVIT! #Lionel #Messi RATED 11/10. WHY? <a href="#">"This ...</a>
28	h_jeeja	2018-02-24 21:57:15	Oh Messi #D10S #Lionel <a href="#">https://t.co/8BgFzu0eDY</a>
29	HotTuxfr	2018-02-24 21:52:01	La playmate Iryna Ivanova fait grimper la température av...
30	axadletn	2018-02-24 21:47:04	Latest post:#Lionel Messi Awoowqisa dhanka hooyo oo ku ...
31	JavierBustosOk	2018-02-24 21:26:08	La definición de lo que no es noticia: hizo un gol Messi...
32	ricmagarcia	2018-02-24 21:26:04	Yo viví en la época de Lionel Messi <a href="#">#messi #D10S #Lion...</a>