Elixir twitter bot

Description:
This twitter bot which is a python script, scrapes the data utilising Twitter API. The bot essentially extracts the Tweets related to Diseases, Drugs and Hospitals, this also stores the user information like user name, user handle, twitter joining date, followers count etc. associated with the Tweets that were extracted. Additional feature of this bot is to store the tags used by the users to Tweet in Twitter. One important functionality is to add a sentiment score to all the tweets, by this it would be possible to know the user's opinion on the medicines and hospitals. Extraction is done on the basis of the details stored in the Diseases, Drugs and Hospitals table. All the extracted details are committed in the Database used using a python script. Below are the details that are being scrapped from Twitter.

Tweet Details:

1. Tweet ID
2. Twitter Handle
3. Tweet
4. Tweet creation date
5. Retweet count
6. Likes count

User Details:

1. User ID
2. User Name
3. User Handle
4. User Profile Picture link
5. Followers Count

Tag Details:

1. Tag Name

Sentiment score:
Sentiment score of all the tweets are calculated
score>0 implies positive opinion
score<0 implies negative opinion
score=0 implies neutral opinion

Below are the non trivial questions:

1. What is the most popular and liked drug and its associated disease?

   ```
   SELECT Drugs.DiseaseId, DiseaseName, DrugName from Drugs join Diseases on Diseases.DiseaseId= Drugs.DiseaseId where DrugName in (select DISTINCT query from query_sentiment where sentiment = (SELECT max(sentiment) from query_sentiment where type='medicine'));
   ```

2. What is the best hospital for the given disease?

   ```
   SELECT max(avg_score), query FROM (SELECT avg(sentiment) AS avg_score, query FROM query_sentiment WHERE type='hospital' GROUP BY query HAVING query like '%Cancer');
   ```

3. What are the top 3 dates when the users got the attention of a particular drug/hospital?

SELECT count(tweet_text) as count, date(created_at) as created_date FROM Tweets as t JOIN query_sentiment as qs on qs.tweet_id=t.tweet_id WHERE query ='Herceptin' GROUP BY created_date ORDER BY count DESC LIMIT 3;

4. List of all the users who tweeted regarding a particular hospital/drug/disease.

   SELECT user_name from Twitter_User as tu join Tweets as t on tu.user_handle=t.twitter_handle join query_sentiment as qs on qs. tweet_id=t.tweet_id WHERE query like '%Stroke%';

   π user_name
   σ query = "Stroke"
   (ρ tu twitter_user ⋈ tu . user_handle = t . twitter_handle
   ρ t tweets ⋈ qs . tweet_id = t . tweet_id
   ρ qs query_sentiment)

5. List of all the users who used a particular hashtag?

   select user_name from Twitter_User where user_handle in (select twitter_handle from Tags join Tweets on Tags.tweet_id=Tweets.tweet_id where tags is not null and tags='EAPCI');

   π user_name
   σ user_handle = "π twitter_handle
   σ NOT (tags = NULL)AND tags = "EAPCI" (tags ⋈ tags . tweet_id = tweets . tweet_id tweets)"
   twitter_user

6. Which disease requires attention?

   SELECT DiseaseName from Diseases where DiseaseId in (SELECT DiseaseId from Drugs where DrugName in (select query from query_sentiment where sentiment=(select min(sentiment) from query_sentiment where type='medicine')));

7. What tags are being promoted by a particular user?

   select tags from Tags as tg join Tweets as t on tg.tweet_id=t.tweet_id join Twitter_User as tu on tu. user_handle=t.twitter_handle where user_name='Dr. Thomas Ichim';

   π tags
   σ user_name = "Dr. Thomas Ichim"
   (ρ tg tags ⋈ tg . tweet_id = t . tweet_id
   ρ t tweets ⋈ tu . user_handle = t . twitter_handle
   ρ tu twitter_user)

8. List of all the users and their sentiment score who posts positive tweets regarding medical domain

   select avg(sentiment) as average_score, user_name from Twitter_User as tu join Tweets as t on tu.user_handle=t.twitter_handle join query_sentiment as qs on qs. tweet_id=t.tweet_id group by user_handle having average_score>0 order by average_score desc;

   τ average_score ↓
   π AVG (sentiment) → average_score, user_name
   σ average_score > 0
   γ user_handle, AVG (sentiment)
   (ρ tu twitter_user ⋈ tu . user_handle = t . twitter_handle
   ρ t tweets ⋈ qs . tweet_id = t . tweet_id

ρ qs query_sentiment)

9. Retrieve all the tweets that Doctors have posted about a particular medicine/drug to get accurate information about a drug.

   SELECT tweet_text, query, user_name from Twitter_User as tu join Tweets as t on tu.user_handle=t.twitter_handle join query_sentiment as qs on qs. tweet_id=t.tweet_id where user_name like 'Dr%' and type='medicine' and query='Albuterol';

   π tweet_text, query, user_name
   σ user_name LIKE "Dr%" AND type = "medicine" AND query = "Albuterol"
   (ρ tu twitter_user ⋈ tu . user_handle = t . twitter_handle
   ρ t tweets ⋈ qs . tweet_id = t . tweet_id
   ρ qs query_sentiment)

10. Identify all the negative reviews by the users about a hospital and it's treatment for a disease to improve the performance of the hospital.

    SELECT DISTINCT tweet_text, query FROM Tweets AS t JOIN query_sentiment AS qs ON t.tweet_id=qs.tweet_id WHERE type='hospital' AND query like 'Baystate Medical Center%Cancer' AND sentiment<0;

    δ
    π tweet_text, query
    σ type = "hospital" AND query LIKE "Baystate Medical Center%Cancer" AND sentiment < 0
    (ρ t tweets ⋈ t . tweet_id = qs . tweet_id
    ρ qs query_sentiment)

Required queries:

1. What user posted this tweet?

   SELECT DISTINCT tu.user_name FROM Tweets as t
   JOIN Twitter_User as tu on tu.user_handle=t.twitter_handle
   where tweet_text='Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par… https://t.co/SKTiQZ9Cey';

   δ
   π tu . user_name
   σ tweet_text = "Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par...https://t.co/SKTiQZ9Cey"
   (ρ t tweets ⋈ tu . user_handle = t . twitter_handle
   ρ tu twitter_user)

2. When did the user post this tweet?

   SELECT DISTINCT created_at FROM Tweets
   WHERE tweet_text='Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par… https://t.co/SKTiQZ9Cey';

   δ

π created_at
σ tweet_text = "Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par...https://t.co/SKTiQZ9Cey" tweets

3. What tweets have this user posted in the past 24 hours?

```
SELECT t.tweet_text FROM Tweets as t
join Twitter_User as tu ON t.twitter_handle=tu.user_handle
WHERE   t.created_at  >= datetime('now','-1 day')  AND tu.user_name='Mantis Toboggan M.D';
```

```
π t . tweet_text
 σ t . created_at >= datetime('now','-1 day') AND tu . user_name = "Mantis Toboggan M.D"
  (ρ t tweets ⋈ t . twitter_handle = tu . user_handle
   ρ tu twitter_user)
```

4. How many tweets have this user posted in the past 24 hours?

```
SELECT count(t.tweet_text) FROM Tweets as t
join Twitter_User as tu ON t.twitter_handle=tu.user_handle
WHERE   t.created_at  >= datetime('now','-1 day')  AND tu.user_name='Mantis Toboggan M.D';
```

```
π COUNT (tweet_text)
 γ COUNT (tweet_text)
  σ t . created_at >= datetime('now','-1 day') AND tu . user_name = "Mantis Toboggan M.D"
   (ρ t tweets ⋈ t . twitter_handle = tu . user_handle
    ρ tu twitter_user)
```

5. When did this user join Twitter?

```
SELECT DISTINCT created_at FROM Twitter_User WHERE user_name='Kate Yandell';
```

```
δ
 π created_at
  σ user_name = "Kate Yandell" twitter_user
```

6. What keywords/ hashtags are popular?

```
SELECT DISTINCT tags from Tags AS t
JOIN Tweets AS ts ON ts.tweet_id=t.tweet_id
WHERE ts.likes_count=(SELECT max(likes_count) FROM Tweets AS ts JOIN Tags t ON
ts.tweet_id=t.tweet_id WHERE  t.tags IS NOT NULL);
```

```
δ
 π tags
  σ ts . likes_count = (π MAX (likes_count)
 γ MAX (likes_count)
  σ NOT (t . tags = NULL)
   (ρ ts tweets ⋈ ts . tweet_id = t . tweet_id
    ρ t tags))
   (ρ t tags ⋈ ts . tweet_id = t . tweet_id
    ρ ts tweets)
```

7. What tweets are popular?

   SELECT DISTINCT tweet_text FROM Tweets WHERE likes_count=(SELECT max(likes_count) from Tweets);

   δ
   π tweet_text
   σ likes_count = (π MAX (likes_count)
   γ MAX (likes_count) tweets) tweets

Citations:

Hospitals List: https://health.usnews.com/best-hospitals/area/ma