DATABASE MANAGEMENT AND DATABASE DESIGN
ASSIGNMENT 2

Tv-show/Movie Recommendation System

TEAM MEMBERS-

Shreyas Rai 002769838

Smiti Agrawal 00002781419

Sarthak Srivastava 002756847

GITHUB LINK -

https://github.com/shreyashusky/tv_Shows_Recommendation_System.git

About-
These days, the small screen has some very big things to offer. From sitcoms to dramas to travel and talk shows, all the small screen shows are the best in showing diversity. TV shows are definitely one of the biggest business markets out there. Also, it offers an opportunity to upcoming artists.

Over the years, the number of TV shows has increased exponentially and so has their customer base. A survey conducted by www.deadline.com says that the top TV Shows can have 15-20 million viewership. With the advancement in technology and availability of cheap internet services, the reach of TV Shows is going to increase in the near future.

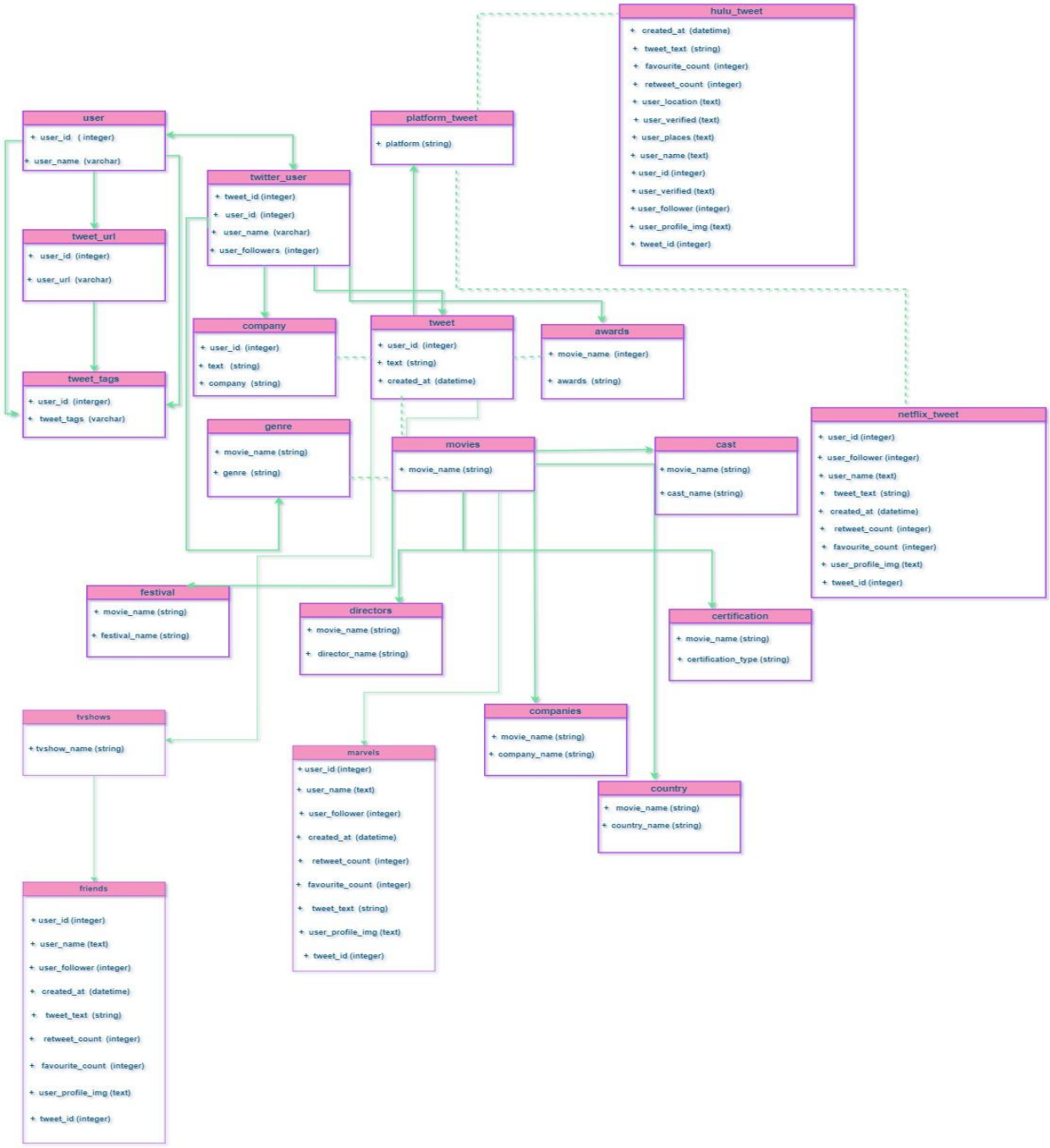
Our domain is Top TV Shows aired till date. We will use social networking APIs to get the data for entities which represent Companies, Genre, Subscription, Ratings, Producers and Consumers for our domain.

Entity Relationship DIAGRAM

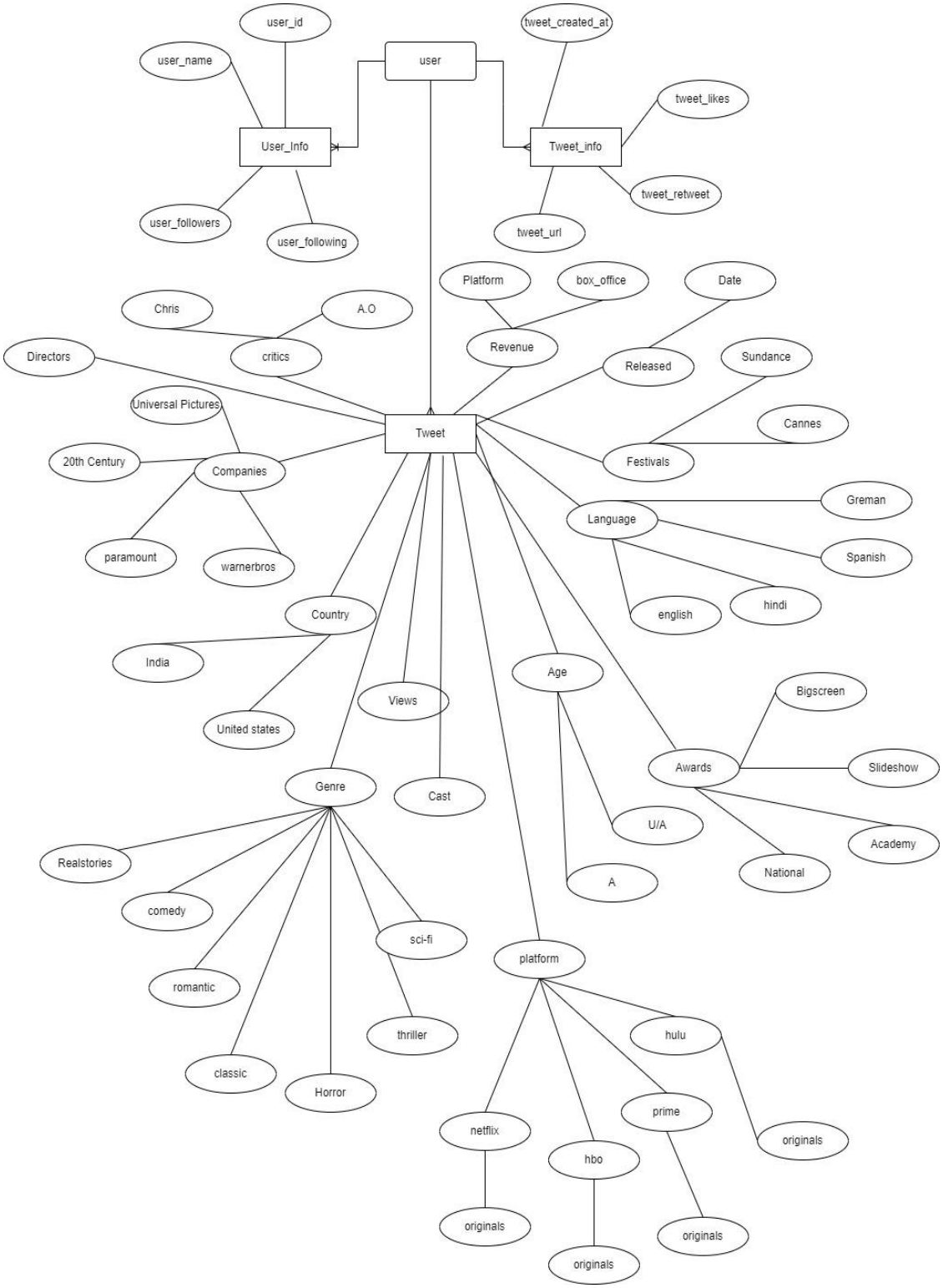
Explanation on some of the design decisions :

- ➔ The Twitter_user is same as user. The Twitter user_id is unique - hence it can also be treated as the primary key of the table.
- ➔ Each user can tweet any number of tweets. The tweet will store the information about the movies and tv show recommend by the people across the world it also store information about user, the text and when the tweet was created.

- ➔ A user can recommend Tvshow / movies on twitter by mentioning or adding a URL. These data will be stored in tweet_url and tweet_tags.
- ➔ 'movies' has the 'tweet' which uniquely distinguish each tweet which is a foreign key.
- ➔ A user can search a movie based on the directors, companies, age, festival, cast and county. Each tweet will be distinguished and have a user_id. Note that each recommendation can have more than one twitter order.



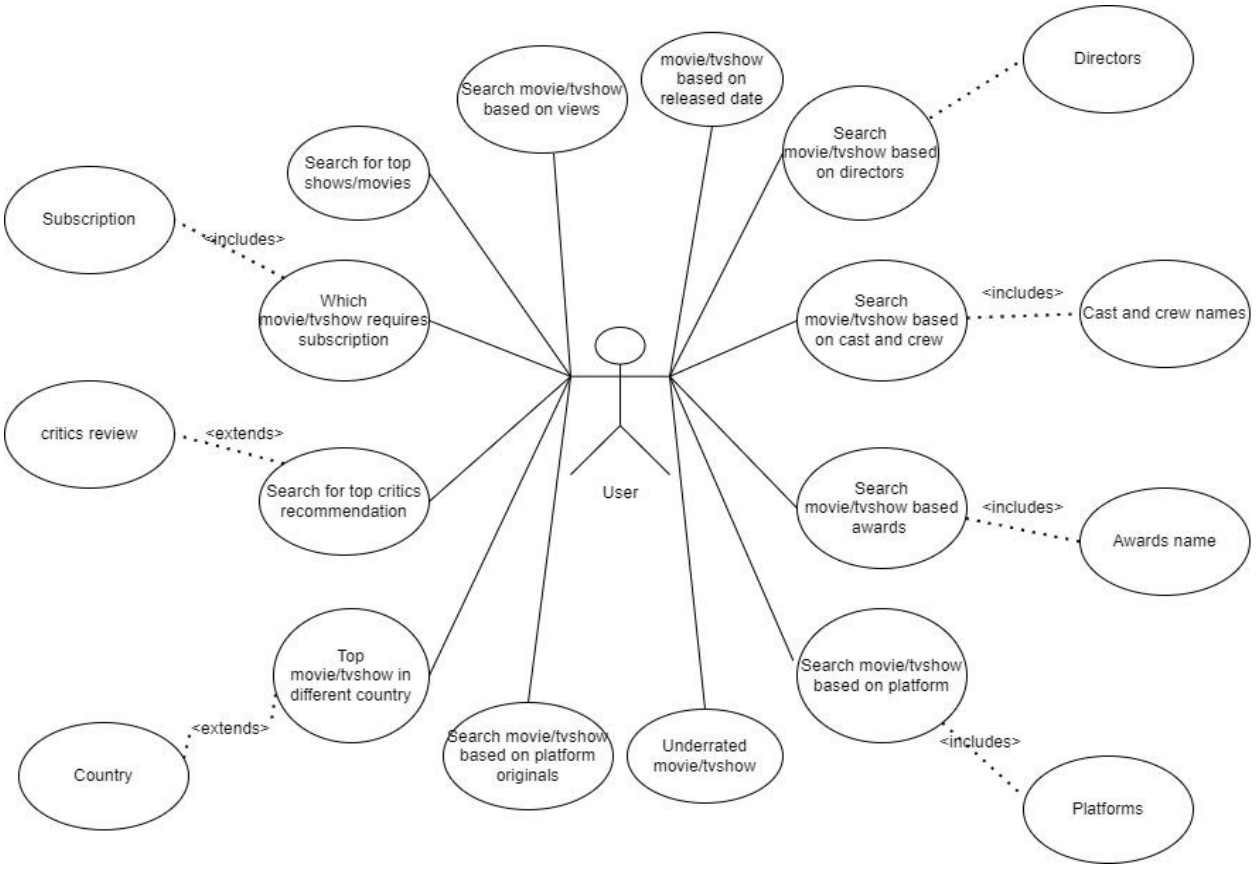
UML DIAGRAM


user_id
user_name
user
tweet_created_at
tweet_likes
User_Info
Tweet_info
tweet_retweet
user_followers
user_following
tweet_url
Platform
box_office
Date
Chris
A.O
Revenue
Directors
critics
Released
Sundance
Universal Pictures
Cannes
Tweet
20th Century
Companies
Festivals
Greman
Language
paramount
warnerbros
Spanish
hindi
english
Country
India
Age
Bigscreen
Views
United states
Awards
Slideshow
Genre
Cast
U/A
Academy
Realstories
National
A
comedy
sci-fi
platform
romantic
hulu
thriller
classic
Horror
prime
netflix
originals
hbo
originals
originals
originals

USE CASE DIAGRAM

Explanation on some of the design decisions :

When a user uses the tv-show/movie recommendations system he can browse for top movies/shows, in fact users can also search movies based on different platforms and which movie/tv-show is available on which platform and it will include the subscription details. Top Critic's recommendation is available which includes critics' reviews. The search can extend to the country and the top movie/tv-shows of different countries. A user can also search with directors, cast, or awards. This will include the recommendations based on cast and crew name and awards name. Users can also search for platform originals available on different platforms. This will help the users to find top recommendations based on his/her choice.



SQL STATEMENT FOR CONCEPTUAL MODEL

TWEETS TABLE

```
CREATE TABLE "tweets" (
        "user_id"       INTEGER,
        "created_at"    TEXT,,
        "tweet_text"    TEXT,
        PRIMARY KEY (‘user_id’)
);
```

TWITTER_USER

```
CREATE TABLE twitter_user(
        "user_id"       INTEGER,
        "user_name"   TEXT,
        "User_followers"    INTEGER,
        PRIMARY KEY (‘user_name’)
);
```

TWEET_URL

```
CREATE TABLE "tweet_url" (
        "user_id"    INTEGER,
        "user_profile_imgl"    TEXT
);
```

NETFLIX_TWEET TABLE

```
CREATE TABLE "netflix" (
        "User_id"   INTEGER,
        "user_name"   TEXT,
        "user_followers"   INTEGER,
        "tweet_id"   REAL,
        "created_at"    TEXT,
        "tweet_text"    TEXT,
        "retweet_count"    INTEGER,
        "favorite_count"    INTEGER,
        "user_profile_img"   TEXT
);
```

HULU_TWEET TABLE

```
CREATE TABLE "hulu" (
        "user_id"  INTEGER,
        "user_name"   TEXT,
        "user_location"    TEXT,
        "user_verified"    TEXT,
        "user_place"  TEXT,
        "user_followers”    INTEGER,
        "tweet_id"   REAL,
        "created_at"    TEXT,
        "tweet_text"    TEXT,
        "retweet_count"   INTEGER,
        "favorite_count"    INTEGER,
        "user_profile_img"      TEXT
);
```

MARVEL TABLE

```
CREATE TABLE "marvels" (
        "user_id"   INTEGER,
        "user_name"   TEXT,
        "user_followers"   INTEGER,
        "tweet_id"       REAL,
        "created_at"    TEXT,
        "tweet_text"    TEXT,
        "retweet_count"   INTEGER,
        "favorite_count"   INTEGER,
        "user_profile_img"   TEXT
);
```

FRIENDS TABLE

```
CREATE TABLE "friends" (
        "field1" INTEGER,
        "user_id"  INTEGER,
        "user_name"   TEXT,
        "user_followers"   INTEGER,
        "tweet_id"   REAL,
        "created_at"    TEXT,
        "tweet_text"    TEXT,
        "retweet_count"   INTEGER,
```

```
        "favorite_count"   INTEGER,
        "user_profile_img"   TEXT
);
```

TWEET_TAGS

```
CREATE TABLE "tweet_tags" (
        "user_id"    INTEGER,
        "tweet_tags"    TEXT
);
```

USER TABLE

```
CREATE TABLE "user" (
        "user_id"    INTEGER,
        "user_name"    TEXT
);
```

COMPANY TABLE

```
CREATE TABLE "company" (
        "user_id"    INTEGER,
        "tweet_text"    TEXT,
        “company”      TEXT
);
```

GENRE TABLE

```
CREATE TABLE "genre" (
        "movie_name"    TEXT,
        "genre"    TEXT
);
```

MOVIES TABLE

```
CREATE TABLE “movies” (
        “movie_name”  TEXT
);
```

DIRECTOR TABLE

```
CREATE TABLE "director" (
```

```
        "movie_name"    TEXT,
        "director_name"    TEXT
);
```

COMPANIES TABLE

```
CREATE TABLE “companies” (
        "movie_name"    TEXT,
        "companies_name"    TEXT
);
```

FESTIVAL TABLE

```
CREATE TABLE “festival” (
        "movie_name"    TEXT,
        "festival_name"    TEXT
);
```

COUNTRY TABLE

```
CREATE TABLE “festival” (
        "movie_name"    TEXT,
        "country_name"    TEXT
);
```

CERTIFICATION TABLE

```
CREATE TABLE “festival” (
        "movie_name"    TEXT,
        "certification_type"    TEXT
);
```

CAST TABLE

```
CREATE TABLE “cast” (
        "movie_name"    TEXT,
        "cast_name"    TEXT
);
```

USE-CASES (By Professor)

1. What user posted this tweet?

   SQL - Select user_name, tweet_text from tweets
   where tweet_id="1590821038711275520";

   Relational Algebra - $\pi_{\text{user\_name, tweet\_text}}\ \sigma_{\text{tweet\_id = "1590821038711275520"}}$ tweets

2. When did the user post this tweet?

   SQL - Select user_name,created_at from tweets where user_name="Harris";

   Relational Algebra - $\pi_{\text{user\_name, created\_at}}\ \sigma_{\text{user\_name = "Harris"}}$ tweets

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

   SQL - select * from tweets where user_name = 'SIMRAN'

   and created_at > '2022-11-10 05:54:19';

   Relational Algebra - $\sigma_{\text{user\_name = "SIMRAN" AND created\_at > "2022-11-10 05:54:19"}}$ tweets

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

   SQL - select count(tweet_id) as Number_of_tweet_created,created_at from tweets where user_name = 'SIMRAN' and created_at > '2022-11-10 05:54:19';

   Relational Algebra -

   $$\pi_{\text{COUNT (tweet\_id)} \rightarrow \text{number\_of\_tweet\_created, created\_at}}$$
   $$\gamma_{\text{COUNT (tweet\_id)}}$$
   $$\sigma_{\text{user\_name = "SIMRAN" AND created\_at > "2022-11-10 05:54:19"}}\ \text{tweets}$$

5. When did this user join Twitter?

   SQL - select distinct(user_name),user_created_at from user where user_name = "HARA TAKASHI";

Relational Algebra - δ

$$\pi_{\text{user\_name, user\_created\_at}}$$
$$\sigma_{\text{user\_name = "HARA TAKASHI"}} \text{user}$$

6. What keywords/ hashtags are popular?

   SQL - select hash_tag, count(hash_tag) as frequency from tweets group by hash_tag order by count(hash_tag) desc;

   Relational Algebra - $\tau_{\text{COUNT (hash\_tag)} \downarrow}$
   $\pi_{\text{hash\_tag, COUNT (hash\_tag)} \rightarrow \text{frequency}}$
   $\gamma_{\text{hash\_tag, COUNT (hash\_tag)}} \text{tweets}$

7. What tweets are popular?

   SQL - Select user_name,tweet_text from tweets where favorite_count>1000;

   Relational Algebra - $\pi_{\text{user\_name, tweet\_text}}$
   $\sigma_{\text{favorite\_count > 1000}} \text{tweets}$

Please find attached screenshots from jupyter notebook-

```
In [26]: run_query('Select user_name, tweet_text from tweets where tweet_id="1590821038711275520"
         #What user posted this tweet?
```

Out[26]:

	user_name	tweet_text
0	Pop Base	Velma actress Linda Cardellini from 'Scooby-Do...

```
In [27]: run_query('Select user_name,created_at from tweets where user_name="Harris"')
         #When did the user post this tweet?
```

Out[27]:

	user_name	created_at
0	Harris	2022-11-12 05:10:42+00:00

```
In [28]: run_query('Select user_name,tweet_text from tweets where favorite_count>1000')
         #What tweets are popular?
```

Out[28]:

	user_name	tweet_text
0	Pop Base	Velma actress Linda Cardellini from 'Scooby-Do...

3 J. Carlos Roldan

In [205]:
```
#what tweet have this user posted in last 24 hours
a="select * from tweets where user_name = 'SIMRAN' and created_at > '2022-11-10 05:54:19'"
run_query(a)
```

Out[205]:

	user_id	user_name	user_followers	tweet_id	created_at	tweet_text	retweet_count	favorite_count	
0	1486041377217839104	SIMRAN	3309	1591297419178037250	2022-11-12 05:10:44+00:00	RT @surelynotparth: I watched three Henry Cavi...	1	0	http://pbs.twimg.com/pr

In [209]:
```
a="select count(tweet_id) as Number_of_tweet_created,created_at from tweets where user_name = 'SIMRAN' and created_at >
#How many tweet have this user posted in 24 hours?
run_query(a)
```

Out[209]:

	Number_of_tweet_created	created_at
0	1	2022-11-12 05:10:44+00:00

In [237]:
```
run_query('select distinct(user_name),user_created_at from user where user_name = "HARA TAKASHI"')
#When did this user join twitter
```

Out[237]:

	user_name	user_created_at
0	HARA TAKASHI	2015-01-14 10:38:51+00:00

USE-CASES (SHREYAS RAI)

1. **Use Case:** How many followers do Marvel have?
   Description: User can find the follower count of marvel.
   Actors: User
   Precondition: User must know the marvel table to execute the query
   Actor Action: User sees the follower count of the marvel
   System Response: Marvel follower count will be displayed
   Post Condition: System displays follower count
   Alternate Path: There is no alternate path

2. **Use Case:** In which country Marvel is most popular?
   Description: Which country has the maximum tweets about Marvels
   Actors: User
   Precondition: Tweets must be present from different countries
   Actor Action: The user sees the maximum tweets from a country
   System Response: System shows the maximum tweets from a country
   Post Condition: System displays the country with maximum number of tweets
   Alternate Path: There are no alternate paths

3. **Use Case:** Is the user tweeting about Hulu having a verified Twitter account?
   Description: To see the tweets are coming from verified users or non verified users
   Actors: User
   Precondition: The twitter account must be verified
   Actor Action: The user sees true or false values for verified accounts
   System Response: System displays the boolean value for verified accounts
   Post Condition: There must be some verified accounts in the dataframe
   Alternate Path: If there are no verified users, the system will return all false values

4. **Use Case:** From where the user is tweeting about Hulu?
   Description: To see from which country maximum tweets are coming for Hulu application
   Actors: user
   Precondition: The tweets must be from different countries for getting proper output
   Actor Action: The user will see the country name with maximum tweet about hulu
   System Response: System will display the country with maximum tweets about hulu
   Post Condition: The tweets must be from different countries
   Alternate Path: There is no alternate path

5. **Use Case:** Count of users that have tweeted using various sources?
   Description: To check whether the user is tweeting from desktop android or IOS
   Actors: user
   Precondition: No precondition
   Actor Action: User will be able to see the device from where
   System Response: System will display the user tweeting from which device
   Post Condition: no post condition
   Alternate Path: There is no alternate path

Please find attached screenshots from jupyter notebook-

```
In [120]: run_query('select sum(user_followers) from marvels')
          #how many followers do Marvel have?
```

Out[120]:

	sum(user_followers)
0	9942453

```
In [124]: run_query('select user_location from hulu where user_followers=(select max(user_followers) from hulu)')
          #In which country marvel is most popular? Roppongi, Minato-ku, Tokyo
```

Out[124]:

	user_location
0	東京都港区六本木

```
In [126]: run_query('select user_location, user_name from hulu where user_name="Tom Costantino"')
          #From where the user is tweeting about hulu?
```

Out[126]:

	user_location	user_name
0	Instagram: tomtheorville	Tom Costantino

```
In [131]: run_query('select user_name from hulu where user_verified is True')
          #Is user tweeting about hulu has a verified Twitter account?
```

Out[131]:

	user_name
0	ORICON NEWS（オリコンニュース）
1	Hulu Japan
2	Tom Costantino

In [132]:
```
run_query('select user_name from hulu where user_verified is False')
#Is user tweeting about hulu does not have a verified Twitter account?
```
Out[132]:

	user_name
0	mi
1	らぴ
2	Kaya
3	さき⛩💎🎟
4	comomo🍑💎
5	nzm
6	緋廼宮 奈由
7	𝔭𝔬𝔯𝔬
8	じゅーん*ժUოe
9	キャラメル
10	きょもST🍅💎@taigaxxxnewera
11	Mimirin 🦋
12	naooonaonao
13	知花華帆
14	えむ
45	🌱🏷うり💚🌱🫧🍓
46	🐸

In [136]:
```
run_query('select user_source, user_name from netflix where user_name="Jeanerie Caling"')
#What is the source of tweet?
```
Out[136]:

	user_source	user_name
0	Twitter for iPhone	Jeanerie Caling

In [139]:
```
run_query('select count(user_name),user_source from netflix group by user_source')
#Count of users that have twitted using various source?
```
Out[139]:

	count(user_name)	user_source
0	1	Hootsuite Inc.
1	1	IFTTT
2	1	M360 Twitter APIs App
3	1	PPE.pl
4	1	TweetDeck
5	11	Twitter Web App
6	19	Twitter for Android
7	1	Twitter for iPad
8	14	Twitter for iPhone

USE-CASES (SMITI AGRAWAL)

6. **Use Case:** Who is the most popular user on twitter, tweeted about Hulu?
   Description: User who has maximum favorite_count will be most popular
   Actors: User
   Precondition: User must have tweeted something containing Hulu
   Steps:
   Actor Action: Get the tweet as many likes as possible
   System Response: If tweet contains word Hulu then it will proceed and use case ends
   Post Condition: Found the most popular user tweeted about Hulu
   Alternate Path: User did not tweeted about Hulu and system throws error
   Error: User is incorrect

7. **Use Case:** What are positive tweets given by users?
   Description: User can find out positive tweets given by another user
   Actors: Users
   Precondition: Tweet text should contain words like “love”,”must watch”

Steps:
Actor Action: Find the tweet text containing positive words
System Response: If the words mentioned by actor is present use case will proceed
Post Condition: Gets the positive tweets about Netflix.
Alternate Path: If no positive words, system throws error
Error: No such words found

8. **Use Case:** Which tweet is most retweeted?
   Description: Gives the tweet that is shared the most
   Actors: Users
   Precondition: A tweet should be present to be retweeted
   Steps:
   Actor Action: Shares the tweet as much as possible
   System Response: If the tweet is shared at least once use case will end
   Post Condition: Gets tweed id and the user who created most retweeted tweet
   Alternate Path: If no retweets, system will throw error
   Error: No retweets

9. **Use Case:** What are the most recent tweets about netflix?
   Description: It will give latest tweet about netflix
   Actors: User
   Precondition: There should be data of netflix tweets
   Steps:
   Actor Action: User will get the latest tweet created for netflix
   System Response: If created_at column is present use case will proceed
   Post Condition: System will display the most recently created tweet
   Alternate Path: If no date mentioned of tweet creation, system will throw error

10. **Use Case:** When were most tweets created?
    Description: It gives date and time when most tweets were created
    Actors: Users
    Precondition: Created time should be there in table data
    Steps:
    Actor Action: User will group the date creating tweet
    System Response: If created_at column is present use case will proceed
    Post Condition: System will give the date when most user tweeted

Please find attached screenshots from jupyter notebook-

8	14	Twitter for iPhone

In [144]:
```
run_query('select created_at,count(created_at) from netflix group by created_at order by count(created_at)desc limit 3')
#On which date and time was most tweet created?
```

Out[144]:

	created_at	count(created_at)
0	2022-11-12 07:33:01+00:00	2
1	2022-11-11 09:01:48+00:00	2
2	2022-11-11 09:00:28+00:00	2

In [151]:
```
run_query('select user_id,user_name,favorite_count from hulu where favorite_count=(select max(favorite_count) from hulu')
#Who is the most popular user?
```

Out[151]:

	user_id	user_name	favorite_count
0	365469866	Hulu Japan	1587

In [175]:
```
a = "select user_name, tweet_text from netflix
WHERE tweet_text like '%comedy%' OR tweet_text like '%#Blockbuster%' OR tweet_text like '%want to watch%'"
run_query(a)
#Users who gave positive tweets?
```

Out[175]:

	user_name	tweet_text
0	Sherri TWDFan4Ever	Watching #Blockbuster on #Netflix.\n\nMakes me...
1	Images	The live-streamed comedy special will air in 2...
2	Perez	This makes me want to watch #TheCrown even mor...

#Who is the most popular user?

Out[151]:

	user_id	user_name	favorite_count
0	365469866	Hulu Japan	1587

In [175]:
```
a = "select user_name, tweet_text from netflix
WHERE tweet_text like '%comedy%' OR tweet_text like '%#Blockbuster%' OR tweet_text like '%want to watch%'"
run_query(a)
#Users who gave positive tweets?
```


Out[175]:

	user_name	tweet_text
0	Sherri TWDFan4Ever	Watching #Blockbuster on #Netflix.\n\nMakes me...
1	Images	The live-streamed comedy special will air in 2...
2	Perez	This makes me want to watch #TheCrown even mor...

In [179]:
```
run_query('select tweet_id,tweet_text,retweet_count from netflix where retweet_count=(select max(retweet_count) from ne
#Which tweet is most retweeted?
```

Out[179]:

	tweet_id	tweet_text	retweet_count
0	1590993186159484928	RT @lawen__27: กรี๊ดดด!! ทุกคน!! นานที #netflix...	16208

In [181]:
```
run_query('select user_name, tweet_text,created_at from netflix group by created_at order by created_at desc limit 1')
#What are the most recent tweets about netflix?
```

Out[181]:

	user_name	tweet_text	created_at
0	野竿 朋子	RT @NetflixJP: #Netflix の今月おすすめの新作をご紹介！ \n\n『Li...	2022-11-12 07:33:59+00:00

USE-CASES (SARTHAK SRIVASTAVA)

11. **Use Case:** Does Tokyo have hulu users?
    Description: Does people use hulu in Tokyo.
    Actors: User
    Precondition: User must have tweeted something about hulu.
    Actor Action: Get the tweets which have location as Tokyo.
    System Response: If the tweet has a location as Tokyo.
    Post Condition: Find all the tweets from Tokyo.
    Alternate Path: Hulu is not available in Tokyo.
    Error: Not available.

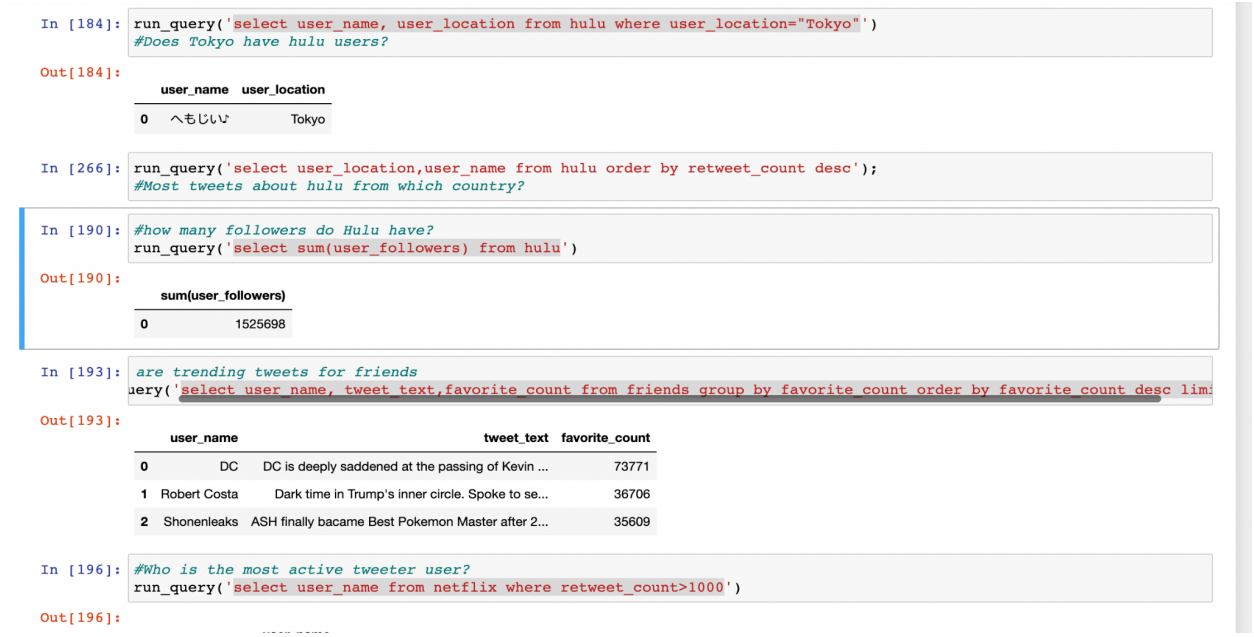
12. **Use Case:** Most tweets about hulu from which country?
    Description: To find which country has most hulu subscriber.
    Actors: User will view the country.
    Precondition: User must have entered the location.
    Actor Action: Get the tweets with the location.
    System Response: Find all the tweets of hulu with their location.
    Post Condition: Find the locations of tweets.
    Alternate Path: Location is not entered
    Error: Location not available

13. **Use Case:** How many followers do Hulu have?
    Description: To find the total no. of hulu subscriber.
    Actors: User
    Precondition: User must have subscribed to Hulu
    Actor Action: Get the total no. of followers
    System Response: Find the total no. of followers of hulu
    Post Condition: Total no. of followers
    Alternate Path: No alternate path

14. **Use Case:** What are trending tweets for friends?
    Description: To find whats trending in friends
    Actors: User get all the new news about friends.
    Precondition: User must have mention friends
    Actor Action: All the tweets related to friends.
    System Response: All tweets with friends mention.
    Post Condition: Various Friends tweet
    Alternate Path: No alternate path

15. **Use Case:** Who is the most active tweeter user?
    Description: The user with most likes
    Actors: Users with likes
    Precondition: User must have tweeted
    Actor Action: All the tweets which were liked
    System Response: Most liked tweet of the user
    Post Condition: Top liked tweet

    Please find attached screenshots from jupyter notebook-



```
In [184]: run_query('select user_name, user_location from hulu where user_location="Tokyo"')
          #Does Tokyo have hulu users?
```

Out[184]:

	user_name	user_location
0	へもじいぃ	Tokyo

```
In [266]: run_query('select user_location,user_name from hulu order by retweet_count desc');
          #Most tweets about hulu from which country?
```

```
In [190]: #how many followers do Hulu have?
          run_query('select sum(user_followers) from hulu')
```

Out[190]:

	sum(user_followers)
0	1525698

```
In [193]: are trending tweets for friends
          uery('select user_name, tweet_text,favorite_count from friends group by favorite_count order by favorite_count desc limi
```

Out[193]:

	user_name	tweet_text	favorite_count
0	DC	DC is deeply saddened at the passing of Kevin ...	73771
1	Robert Costa	Dark time in Trump's inner circle. Spoke to se...	36706
2	Shonenleaks	ASH finally bacame Best Pokemon Master after 2...	35609

```
In [196]: #Who is the most active tweeter user?
          run_query('select user_name from netflix where retweet_count>1000')
```

Out[196]:


```
In [266]: run_query('select user_location,user_name from hulu order by retweet_count desc');
          #Most tweets about hulu from which country?
```

```
In [190]: #how many followers do Hulu have?
          run_query('select sum(user_followers) from hulu')
```

Out[190]:

	sum(user_followers)
0	1525698

```
In [193]: are trending tweets for friends
          uery('select user_name, tweet_text,favorite_count from friends group by favorite_count order by favorite_count desc limi
```

Out[193]:

	user_name	tweet_text	favorite_count
0	DC	DC is deeply saddened at the passing of Kevin ...	73771
1	Robert Costa	Dark time in Trump's inner circle. Spoke to se...	36706
2	Shonenleaks	ASH finally bacame Best Pokemon Master after 2...	35609

```
In [196]: #Who is the most active tweeter user?
          run_query('select user_name from netflix where retweet_count>1000')
```

RELATIONAL ALGEBRA EXPRESSIONS FOR USE CASES(SHREYAS RAI)

1. Usecase1 - How many followers do Marvel have?

   $\pi_{\text{SUM (user\_followers)}}$

   $\gamma_{\text{SUM (user\_followers)}}$ marvels

2. Usecase2 - In which country Hulu is most popular?

   $\pi_{\text{user\_location}}$

   $\sigma_{\text{user\_followers = a}}$ hulu

   $a=\pi_{\text{MAX (user\_followers)}}$

   $\gamma_{\text{MAX (user\_followers)}}$ hulu

3. Usecase3 - Is the user tweeting about Hulu having a verified Twitter account?

   $\pi_{\text{user\_name}}$

   $\sigma_{\text{user\_verified = true}}$ hulu

4. Usecase4 - From where the user is tweeting about Hulu?

   $\pi_{\text{user\_location, user\_name}}$

   $\sigma_{\text{user\_name = "Tom Costantino"}}$ hulu

5. Usecase5 - Count of users that have tweeted using various sources?

   $\gamma_{\text{user\_source, COUNT (user\_name)}}$ netflix

RELATIONAL ALGEBRA EXPRESSIONS FOR USE CASES(SMITI AGRAWAL)

6. Usecase6 - Who is the most popular user on twitter, tweeted about Hulu?

   $a=\pi_{\text{MAX (favorite\_count)}}$

   $\gamma_{\text{MAX (favorite\_count)}}$ hulu

   $\pi_{\text{user\_id, user\_name, favorite\_count}}$

   $\sigma_{\text{favorite\_count = a}}$ hulu

7. Usecase7 - What are positive tweets given by users for netflix?

   $\pi_{\text{user\_name, tweet\_text}}$

   $\sigma_{\text{tweet\_text LIKE "\%comedy\%" OR tweet\_text LIKE "\%\#Blockbuster\%" OR tweet\_text LIKE "\%want to watch\%"}}$ netflix

8. Usecase8 - Which tweet is most retweeted for netflix?

   $a=\pi_{\text{MAX (retweet\_count)}}$

   $\gamma_{\text{MAX (retweet\_count)}}$ netflix

   $\pi_{\text{tweet\_id, tweet\_text, retweet\_count}}$

   $\sigma_{\text{retweet\_count = a}}$ netflix

9. Usecase9 - What are the most recent tweets about netflix?

$\tau_{\text{created\_at}\downarrow}$

$\pi_{\text{user\_name, tweet\_text, created\_at}}$

$\gamma_{\text{created\_at,}}$ netflix

10. Usecase10 - When were most tweets created?

$\tau_{\text{COUNT (created\_at)}\downarrow}$

$\gamma_{\text{created\_at, COUNT (created\_at)}}$ netflix

RELATIONAL ALGEBRA EXPRESSIONS FOR USE CASES(SARTHAK SRIVASTAVA)

11. Usecase11 - Does Tokyo have hulu users?

$\pi_{\text{user\_name, user\_location}}$

$\sigma_{\text{user\_location = "Tokyo"}}$ hulu

12. Usecase12 - Most tweets about hulu from which country?

$\tau_{\text{retweet\_count}\downarrow}$

$\pi_{\text{user\_location, user\_name}}$

$\gamma_{\text{retweet\_count,}}$ hulu

13. Usecase13 - How many followers does Hulu have?

$\pi_{\text{SUM (user\_followers)}}$

$\gamma_{\text{SUM (user\_followers)}}$ hulu

14. Usecase14 - What are trending tweets for friends?

$\tau_{\text{favorite\_count}\downarrow}$

$\pi_{\text{user\_name, tweet\_text, favorite\_count}}$

$\gamma_{\text{favorite\_count,}}$ friends

15. Usecase15 - Who is the most active twitter user?

$\pi_{\text{user\_name}}$

$\sigma_{\text{retweet\_count > 1000}}$ netflix

SQL STATEMENTS(SHREYAS RAI)

1. **Usecase1 - How many followers do Marvel have?**

```
select sum(user_followers) from marvels;
```

2. **Usecase2 - In which country Hulu is most popular?**

```
select user_location from hulu
```

where user_followers=(select max(user_followers) from hulu);

3. **Usecase3 - Is the user tweeting about Hulu having a verified Twitter account?**
   select user_name from hulu where user_verified is True;

4. **Usecase4 - From where the user is tweeting about Hulu?**
   select user_location, user_name from hulu
   where user_name="Tom Costantino";

5. **Usecase5 - Count of users that have tweeted using various sources?**
   select count(user_name),user_source from netflix group by user_source;

SQL STATEMENTS(SMITI AGRAWAL)

6. **Usecase6 - Who is the most popular user on twitter, tweeted about Hulu?**
   select user_id,user_name,favorite_count from hulu where
   favorite_count=(select max(favorite_count) from hulu);

7. **Usecase7 - What are positive tweets given by users?**
   a = "select user_name, tweet_text from netflix
   WHERE tweet_text like '%comedy%' OR tweet_text like '%#Blockbuster%'
   OR tweet_text like '%want to watch%'"
   run_query(a);

8. **Usecase8 - Which tweet is most retweeted?**
   select tweet_id,tweet_text,retweet_count from netflix
   where retweet_count=(select max(retweet_count) from netflix);

9. **Usecase9 - What are the most recent tweets about netflix?**
   select user_name, tweet_text,created_at from netflix
   group by created_at order by created_at desc limit 1;

10. **Usecase10 - When were most tweets created?**
    select created_at,count(created_at) from netflix group by created_at
    order by count(created_at)desc limit 3;

SQL STATEMENTS(SARTHAK SRIVASTAVA)

11. **Usecase11 -** Does Tokyo have hulu users?
    select user_name, user_location from hulu where user_location="Tokyo";

12. **Usecase12 -** Most tweets about hulu from which country?

select user_location,user_name from hulu group by retweet_count order by retweet_count desc;

13. **Usecase13 -** How many followers does Hulu have?
    select sum(user_followers) from hulu;

14. **Usecase14 -** What are trending tweets for friends?
    select user_name, tweet_text,favorite_count from friends group by favorite_count order by favorite_count desc limit 3;

15. **Usecase15 -** Who is the most active twitter user?
    select user_name from netflix where retweet_count>1000

TEAM MEMBERS-

Shreyas Rai 002769838

Smiti Agrawal 00002781419

Sarthak Srivastava 002756847

GITHUB LINK -

https://github.com/shreyashusky/tv_Shows_Recommendation_System.git