

Data Management and Database Design

INFO 6210

Project Report

Group Members

HAN WU
JIANYU HU
YU HSUAN CHENG

Abstract

Movie, one of the most important entertainment tools. With the development of the film area, quite a lot of movie studios were built. To introduce the updated movies to the consumers, most companies open their twitter account to do the promotion. However, this phenomenon brings a problem that if someone is a movie fan who cannot read all movies tweets, movie comments or studios tweets, this might be a challenge to find all upcoming movie information as there are enormous accounts to read. Therefore, based on the movies, we want to build a database that contains the comments and tags of the movies, not only the movies but also studios, and collect all the useful information like comments and YouTube watch links for the movie which being or will being to hit the cinema screens.

I. Introduction

What we proposed to build is a database which contains several useful entities and information from different social sites and social medias. They all relate to the movies which have come to the screen or will come to the screen in 2019. The database will be updated to catch the newest post from the movie twitter account for people who are interested in recent news about movies. People can also choose films they are interested based on the tags built for the movies. Also, we can see the popular tags and topics that people talk about and comment on to know the trends regarding the movies from Reddit. Besides, people can select the studios to find the movies from their favorite studios then watch them through the YouTube watching link provided.

II. Data Collected

The project fetches real-time data from Twitter, Reddit and YouTube website corresponding to the 2019 movies from Wiki.

A. Twitter API

1. We harvested the 2019 data from wiki and imported the table to Jupyter to obtain other information and data through Python. The dataset consists of 144 movies and we attached the screen name, searched from Twitter, to the table manually. However, not all movies have official Twitter account. There are 100 screen names while there are 44 movies without screen name.

	Opening date	Title	Studio	Cast and crew	Genre	Country	Screen_name
0	Jan-04	Escape Room	Columbia Pictures	Adam Robitel (director); Bragi F. Schut, Maria...	Drama, Horror, Thriller, Mystery	US	Escape_Room
1	Jan-11	A Dog's Way Home	Columbia Pictures	Charles Martin Smith (director); W. Bruce Came...	Drama, Family	US	ADogsWayHome
2	Jan-11	The Upside	STX Entertainment / Lantern Entertainment	Neil Burger (director); Jon Hartmere (screenpl...	Comedy, Drama	US	TheUpsideFilm
3	Jan-11	Replicas	Entertainment Studios	Jeffrey Nachmanoff (director); Chad St. John (...)	Sci-Fi, Thriller	US	replicas_movie
4	Jan-18	Glass	Universal Pictures / Buena Vista International...	M. Night Shyamalan (director/screenplay); Jame...	Superhero, Horror, Thriller	US	GlassMovie
5	Jan-25	The Kid Who Would Be King	20th Century Fox / Working Title Films	Joe Cornish (director/screenplay); Louis Ashbo...	Fantasy, Adventure, Family	UK, US	KidWouldBeKing
6	Jan-25	Serenity	Aviron Pictures / Global Road Entertainment	Steven Knight (director/screenplay); Matthew M...	Drama, Thriller	US	SerenityFilm

2. Assessing to Twitter API through tweepy package:

```
consumer_key = "LaOJGTHUiYOCkJvKrQvcSn5mE"
consumer_secret = "wTbZQKcDsQWyYUuWlRn0kBX7lSuzAu2tJOqjYYHvQsvIPBaPnI"
access_token = "1151636618-uVTZH53QJ64V1Yw2sPXgWoFx3ilGwlWnHZntbvT"
access_token_secret = "bYOD512TWVXrLzCWz5pceojQWdi6RuHQOopccqf0TOeQ4"

api = twitter.Api(consumer_key=consumer_key,
                  consumer_secret=consumer_secret,
                  access_token_key=access_token,
                  access_token_secret=access_token_secret)

# Pass OAuth details to tweepy's OAuth handler
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
```

```
# Test whether access is successful
```

3. Define a function to convert the Twitter API time format to the format that SQL can read. We will convert all Twitter time data to the format yyyy-mm-dd hh:mm:ss

```
## the function that convets time to the time format that sql readable
def parse_datetime(value):
    time_tuple = parsedate_tz(value)
    timestamp = mktime_tz(time_tuple)

    return datetime.datetime.fromtimestamp(timestamp)
```

3. Retrieved official movie Twitter accounts' post and related attributes like, the count of favourites, Hashtags.etc., for each screen name through "GetUserTimeline" and made the data into a table: 'official_movie_tw_post'. Be careful to skip none-screen-names.

```

for i, movie in enumerate(movies['Screen_name']):



    if movies['Screen_name'].isnull()[i]==True:
        continue
    results = api.GetUserTimeline(screen_name= movie, count=100)

    for each in results:
        status = json.loads(str(each))
        posts_12.append(status['text'])
        post_id_12.append(status['id'])
        favourites_12.append(each.favorite_count)

        retweets_12.append(each.retweet_count)
        time_12.append(parse_datetime(status['created_at']))
        screen_name_12.append(status['user']['screen_name'])

    try:
        hashtags_12.append(status['hashtags'][0]['text'])
    except:
        hashtags_12.append(status['hashtags'])

```

	Screen_name	Post	Post_id	Favourites	Retweets	Hashtags	Posted_time
0	Escape_Room	Get to know the amazing cast and check out a n...	1121524570408624128	12	3	[]	2019-04-25 17:21:09
1	Escape_Room	New phone, who dis? #NationalTelephoneDay  ht...	1121486031671291904	8	3	NationalTelephoneDay	2019-04-25 14:48:00
2	Escape_Room	Centered on six strangers who must overcome li...	1121138475527315456	28	4	[]	2019-04-24 15:46:56
3	Escape_Room	Can I take a message? #AdminProfessionalsDay  ...	1121095688706646017	5	1	AdminProfessionalsDay	2019-04-24 12:56:55
4	Escape_Room	Welcome to the Escape Room. Find the clues	1120849977494786048	6	3	EscapeRoomMovie	2019-04-23 11:01:00

4. Retrieved official movie Twitter accounts' information and related attributes like movie_id, user name, the count of followers, for each screen name through "GetUserTimeline" and made the data into a table: 'movie_tw_info'. Be careful to skip none-screen-names.

```

for i, movie in enumerate(movies['Screen_name']):

    if movies['Screen_name'].isnull()[i]==True:
        continue
    results_1 = api.GetUser(screen_name= movie)

    movie_id_1.append(results_1.id)
    name_1.append(results_1.name)
    screen_name_1.append(results_1.screen_name)

    followers_1.append(results_1.followers_count)
    friends_count_1.append(results_1.friends_count)
    no_tweet_1.append(results_1.statuses_count)
    time_1.append(parse_datetime(results_1.created_at))

```

;

	Movie_id	Screen_name	User_name	Followers	Friends_count	tweets_count	Joined_time
0	1042668670617440256	Escape_Room	Escape Room	1730	5	101	2018-09-20 02:55:57
1	918927445029302272	ADogsWayHome	A Dog's Way Home	2401	8	163	2017-10-13 15:52:30
2	788766161852854273	TheUpsideFilm	The Upside	2434	20	957	2016-10-19 11:38:02
3	1074869646594068481	replicas_movie	ReplicasMovie	1633	303	341	2018-12-17 22:31:08
4	920092226503434241	GlassMovie	Glass	23849	7	135	2017-10-16 21:00:56
5	1001973055747338240	KidWouldBeKing	The Kid Who Would Be King	1466	19	278	2018-05-30 19:46:07

5. Made Cast and Crew into another table:

The “Cast and crew” column corresponding movie titles is the important information but it is not atomic Therefore, we need to split the names into separate rows corresponding to each movie, attached the jobs they did, like actors, directors, screenplay and set the cast and crew id to each row.

Opening date	Title	Studio	Cast and crew	Genre	Country	Screen_name
0 Jan-04	Escape Room	Columbia Pictures	Adam Robitel (director); Bragi F. Schut, Maria...	Drama, Horror, Thriller, Mystery	US	Escape_Room
1 Jan-11	A Dog's Way Home	Columbia Pictures	Charles Martin Smith (director); W. Bruce Came...	Drama, Family	US	ADogsWayHome
2 Jan-11	The Upside	STX Entertainment / Lantern Entertainment	Neil Burger (director); Jon Hartmere (screenpl...	Comedy, Drama	US	TheUpsideFilm
3 Jan-11	Replicas	Entertainment Studios	Jeffrey Nachmanoff (director); Chad St. John (...)	Sci-Fi, Thriller	US	replicas_movie
4 Jan-18	Glass	Universal Pictures / Buena Vista International...	M. Night Shyamalan (director/screenplay); Jame...	Superhero, Horror, Thriller	US	GlassMovie
5 Jan-25	The Kid Who Would Be King	20th Century Fox / Working Title Films	Joe Cornish (director/screenplay); Louis Ashbo...	Fantasy, Adventure, Family	UK, US	KidWouldBeKing
6 Jan-25	Serenity	Aviron Pictures / Global Road Entertainment	Steven Knight (director/screenplay); Matthew M...	Drama, Thriller	US	SerenityFilm

	cast_id	Name	Job	Movie
0	00101	Adam Robitel	director	Escape Room
1	00102	Bragi F. Schut	actor	Escape Room
2	00103	Maria Melnik	screenplay	Escape Room
3	00104	Logan Miller	actor	Escape Room
4	00105	Deborah Ann Woll	actor	Escape Room

6. Made Studios into a new table: M_to_S table:

In the same way as above, the “Studio” column corresponding movie titles is the important information but not atomic. Therefore, we need to split the studios into separate rows corresponding to each movie, and set the cast and crew id to each row.

```

# Make a new company table
studio = []
movie_3 = []
M_to_S_id = []
company = movies['Studio']
# Split studios corresponding to movies into different rows
for i, stu in enumerate(company):

    studio_list = stu.split(" / ") ## studio list for each movie

    for j, s in enumerate(studio_list):
        studio.append(s)
        movie_3.append(movies['Title'][i])

        temp = chr(ord('a')+j) ##creat id
        start_num = i+1
        M_to_S_id.append(str(start_num)+temp)

```

	M_to_S_id	Movie	Studio
0	1a	Escape Room	Columbia Pictures
1	2a	A Dog's Way Home	Columbia Pictures
2	3a	The Upside	STX Entertainment
3	3b	The Upside	Lantern Entertainment
4	4a	Replicas	Entertainment Studios

7. We searched for the screen name of each studio on Twitter. However, not all studios have official Twitter account, especially the studios outside of US. There are totally distinct 103 studios on list while there are 23 studios without screen name. (The table below does not need to import to database, just for attaching screen name manually)

	Studio	Screen_name
0	Columbia Pictures	ColumbiaStudios
1	STX Entertainment	STXEnt
2	Lantern Entertainment	lantern_ent
3	Entertainment Studios	ESGlobalMedia

8. Retrieved studio Twitter accounts' information and related attributes like user name, the count of followers, corresponding to each studio screen name through "GetUserTimeline" and made the data into a table: 'tw_studio_table'. Be careful to skip none-screen-names.

```

for i, company in enumerate(s_studio['Screen_name']):

    if s_studio['Screen_name'].isnull()[i]==True:
        continue
    results_3 = api.GetUser(screen_name= company)

    company_id.append(results_3.id)
    screen_name_3.append(results_3.screen_name)

    followers_3.append(results_3.followers_count)
    friends_count_3.append(results_3.friends_count)
    no_tweet_3.append(results_3.statuses_count)
    time_3.append(parse_datetime(results_3.created_at))
    ## Add studio title correspodng the info to the new table
    company_name_3.append(s_studio['Studio'][i])

```

	Studio	Screen_name	Studio_id	Followers	Friends_count	tweets_count	Joined_time
0	Columbia Pictures	ColumbiaStudios	1465344716	1865	6	12	2013-05-28 14:15:24
1	STX Entertainment	STXEnt	2998868947	40281	206	1084	2015-01-27 20:15:53
2	Lantern Entertainment	lantern_ent	993917042343403520	61	2	7	2018-05-08 14:14:23
3	Entertainment Studios	ESGlobalMedia	34728131	2669	246	2231	2009-04-23 17:01:54
4	Universal Pictures	UniversalPics	21904217	3695593	704	13312	2009-02-25 14:14:14

9. Retrieved studio Twitter accounts' post and related attributes like, the count of favourites, Hashtags.etc., for each screen name through “GetUserTimeline” and made the data into a table: ‘studio_tw_post’. Be careful to skip none-screen-names.

```

for i, stu in enumerate(df_3['Screen_name']):

    results = api.GetUserTimeline(screen_name= stu, count=100)

    for each in results:
        status = json.loads(str(each))
        posts_32.append(status['text'])
        post_id_32.append(status['id'])
        favourites_32.append(each.favorite_count)

        retweets_32.append(each.retweet_count)
        time_32.append(parse_datetime(status['created_at']))
        screen_name_32.append(status['user']['screen_name'])

    try:
        hashtags_32.append(status['hashtags'][0]['text'])
    except:
        hashtags_32.append(status['hashtags'])

```

	Screen_name	Post	Post_id	Favourites	Retweets	Hashtags	Posted_time
0	ColumbiaStudios	Jeri Ryan joins the cast of Helix as Ilaria's ...	434954542367514625	12	12	[]	2014-02-16 02:36:58
1	ColumbiaStudios	With stunning performances that have wowed cri...	426329337600569344	9	6	[]	2014-01-23 07:23:29
2	ColumbiaStudios	She's always hot on the trail of danger. #Gwen...	426329060306722817	8	6	Gwensday	2014-01-23 07:22:23
3	ColumbiaStudios	Captain Phillips just landed on Blu-ray and Di...	426018623895789569	10	4	[]	2014-01-22 10:48:49
4	ColumbiaStudios	Who are your real-life heroes? #MonumentsMen h...	426018153626210304	8	2	MonumentsMen	2014-01-22 10:46:57

10. To see how users comment on movies, we retrieved Twitter users' post and related attributes like, the count of favourites, Hashtags.etc., corresponding to each movie as the query-term through "GetSearch" and made the data into a table:

'tw_user_comment'. However, some movie terms, for example, 'Us' and 'After' are not suitable for searching comments because both are very common words in our daily life. You may obtain something not relating to movies. Therefore, we replace 'Us' and 'After' with 'Us%20movie' and 'After%20movie' as the terms respectively.

```
for i, title in enumerate(movies['Title']):
    if title is 'Replicas':
        results_4 = api.GetSearch(term='Replicas%20movie', count= 100)

    elif title is 'Glass':
        results_4 = api.GetSearch(term='Glass%20movie', count= 100)

    elif title is 'Serenity':
        results_4 = api.GetSearch(term='Serenity%20movie', count= 100)
```

	User	Movie	Followers	Post	Post_id	Hashtags	Posted_time	Favourites
0	Heather Wixson	Escape Room	19294	Hey y'all! #EscapeRoom is out on Blu/DVD today...	1120748112945901568	EscapeRoom	2019-04-23 13:55:47	59
1	CLL at UNC-CH SON	Escape Room	377	#FlashBackFriday to when we went to @bullcitye...	1121840033718263808	FlashBackFriday	2019-04-26 14:14:41	0
2	Colby Deane	Escape Room	176	RT @BaptieGretchen: Today we	1121840033718263808		2019-04-26	0

B. Reddit API

	Opening date	Title	Studio	Cast and crew	Genre	Country	Screen_name
0	Jan-04	Escape Room	Columbia Pictures	Adam Robitel (director); Bragi F. Schut, Maria...	Drama, Horror, Thriller, Mystery	US	Escape_Room
1	Jan-11	A Dog's Way Home	Columbia Pictures	Charles Martin Smith (director); W. Bruce Came...	Drama, Family	US	ADogsWayHome
2	Jan-11	The Upside	STX Entertainment / Lantern Entertainment	Neil Burger (director); Jon Hartmere (screenpl...	Comedy, Drama	US	TheUpsideFilm
3	Jan-11	Replicas	Entertainment Studios	Jeffrey Nachmanoff (director); Chad St. John (...)	Sci-Fi, Thriller	US	replicas_movie
4	Jan-18	Glass	Universal Pictures / Buena Vista International...	M. Night Shyamalan (director/screenplay); Jame...	Superhero, Horror, Thriller	US	GlassMovie
5	Jan-25	The Kid Who Would Be King	20th Century Fox / Working Title Films	Joe Cornish (director/screenplay); Louis Ashbo...	Fantasy, Adventure, Family	UK, US	KidWouldBeKing
6	Jan-25	Serenity	Aviron Pictures / Global Road Entertainment	Steven Knight (director/screenplay); Matthew M...	Drama, Thriller	US	SerenityFilm

1. Based on the movie titles, we retrieved the data from Reddit under the subreddit—movies (under the category: movies.) Then, we made the data into a table.

	id	title	body	movie	score	comms_num	url	created
0	adjst5	Box Office Week: Aquaman is #1 again for third...	["Rank*"]*Title*["Domestic Gross (Weekend)"]*Wo...	Escape Room	1322	695	https://www.reddit.com/r/movies/comments/adjst...	1.546881e+09
1	bae8r1	Movies similar to Escape Room	I watched the movie Escape Room the other day ...	Escape Room	12	25	https://www.reddit.com/r/movies/comments/bae8r...	1.554625e+09
2	ac84qv	Official Discussion: Escape Room [SPOILERS]	# Poll\n\n**If you've seen the film, please ra...	Escape Room	203	949	https://www.reddit.com/r/movies/comments/ac84q...	1.546571e+09
3	bf0bzk	What Did You Think Of: Escape Room?	I got around to watching it last night, and it...	Escape Room	1	13	https://www.reddit.com/r/movies/comments/bf0bz...	1.555687e+09

2. Notice that the “created” column is the unix format of time. Therefore, we need to convert the form of time to that form that is readable.

```
# Fixing the date column("created")
def get_date(created):
    return dt.datetime.fromtimestamp(created)

_timestamp = df_1["created"].apply(get_date)
df_1 = df_1.assign(timestamp = _timestamp)

df_1 = df_1.drop(columns='created')
```

	id	title	body	movie	score	comms_num	url	timestamp
0	adjst5	Box Office Week: Aquaman is #1 again for third...	["Rank*"]*Title*["Domestic Gross (Weekend)"]*Wo...	Escape Room	1322	695	https://www.reddit.com/r/movies/comments/adjst...	2019-01-07 12:16:42
1	bae8r1	Movies similar to Escape Room	I watched the movie Escape Room the other day ...	Escape Room	12	25	https://www.reddit.com/r/movies/comments/bae8r...	2019-04-07 04:19:31
2	ac84qv	Official Discussion: Escape Room [SPOILERS]	# Poll\n\n**If you've seen the film, please ra...	Escape Room	203	949	https://www.reddit.com/r/movies/comments/ac84q...	2019-01-03 22:00:24
3	bf0bzk	What Did You Think Of: Escape Room?	I got around to watching it last night, and it...	Escape Room	1	13	https://www.reddit.com/r/movies/comments/bf0bz...	2019-04-19 11:21:23

C. YouTube API

```
In [1]:
# -*- coding: utf-8 -*-

import os
import google.oauth2.credentials

import google_auth_oauthlib.flow
from googleapiclient.discovery import build
from googleapiclient.errors import HttpError
from google_auth_oauthlib.flow import InstalledAppFlow

# The CLIENT_SECRETS_FILE variable specifies the name of a file that contains
# the OAuth 2.0 information for this application, including its client_id and
# client_secret.
CLIENT_SECRETS_FILE = "client_secret.json"

# This OAuth 2.0 access scope allows for full read/write access to the
# authenticated user's account and requires requests to use an SSL connection.
SCOPES = ['https://www.googleapis.com/auth/youtube.force-ssl']
API_SERVICE_NAME = 'youtube'
API_VERSION = 'v3'
```

1. Use YouTube API V3 to get data through search a specific topic about our domain - movies. My google account and password was stored as json file so that I can use this API with authorized execute.

```

def get_authenticated_service():
    flow = InstalledAppFlow.from_client_secrets_file(CLIENT_SECRETS_FILE, SCOPES)
    credentials = flow.run_console()
    return build(API_SERVICE_NAME, API_VERSION, credentials = credentials)

def print_response(response):
    print(response)

# Build a resource based on a list of properties given as key-value pairs.
# Leave properties with empty values out of the inserted resource.
def build_resource(properties):
    resource = {}
    for p in properties:
        # Given a key like "snippet.title", split into "snippet" and "title", where
        # "snippet" will be an object and "title" will be a property in that object.
        prop_array = p.split('.')
        ref = resource
        for pa in range(0, len(prop_array)):
            is_array = False
            key = prop_array[pa]

            # For properties that have array values, convert a name like
            # "snippet.tags[]" to snippet.tags, and set a flag to handle
            # the value as an array.
            if key[-2:] == '[]':
                key = key[0:len(key)-2:]
                is_array = True

            if pa == (len(prop_array) - 1):
                # Leave properties without values out of inserted resource.
                if properties[p]:
                    if is_array:
                        ref[key] = properties[p].split(',')
                    else:
                        ref[key] = properties[p]
            elif key not in ref:
                # For example, the property is "snippet.title", but the resource does
                # not yet have a "snippet" object. Create the snippet object here.
                # Setting "ref = ref[key]" means that in the next time through the
                # "for pa in range ..." loop, we will be setting a property in the
                # resource's "snippet" object.
                ref[key] = {}
                ref = ref[key]
            else:
                # For example, the property is "snippet.description", and the resource
                # already has a "snippet" object.
                ref = ref[key]
    return resource

# Remove keyword arguments that are not set

```

```

# Remove keyword arguments that are not set
def remove_empty_kwargs(**kwargs):
    good_kwargs = {}
    if kwargs is not None:
        for key, value in kwargs.items():
            if value:
                good_kwargs[key] = value
    return good_kwargs

def videos_list_most_popular(client, **kwargs):
    # See full sample for function
    kwargs = remove_empty_kwargs(**kwargs)

    response = client.videos().list(
        **kwargs
    ).execute()

    return print_response(response)

if __name__ == '__main__':
    # When running locally, disable OAuthlib's HTTPs verification. When
    # running in production *do not* leave this option enabled.
    os.environ['OAUTHLIB_INSECURE_TRANSPORT'] = '1'
    client = get_authenticated_service()

    videos_list_most_popular(client,
        part='snippet,contentDetails,statistics',
        chart='mostPopular',
        regionCode='US',
        videoCategoryId='20')

```

Please visit this URL to authorize this application: https://accounts.google.com/o/oauth2/auth?response_type=code&client_id=784889702052-u7mnuuacm8069moa361pvn2cgrd6u2nid.apps.googleusercontent.com&redirect_uri=https://www.googleapis.com/oauth2/v1/userinfo&scope=https://www.googleapis.com/auth/youtube.force-ssl&state=3QL39uHfT5Fhw15tUuIE0HPw41a1d6pmpmt=content&access_type=offline

Enter the authorization code: 4/FwH0ScXjPuLtylcFAyAvzLQqhl.c0lFqW5S5b0NjGw67-AQJr0b8mtCo0

```

{u'nextPageToken': u'CAUQAA', u'items': [{u'kind': u'youtube#video', u'statistics': {u'commentCount': u'19714', u'viewCount': u'6503190', u'favoriteCount': u'0', u'dislikeCount': u'2601', u'likeCount': u'147260'}, u'contentDetails': {u'definition': u'hd', u'projection': u'rectangular', u'caption': u'false', u'duration': u'PT10M2S', u'licensedContent': True, u'dimension': u'2d', u'snippet': {u'thumbnails': {u'default': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/default.jpg', u'width': 120, u'height': 90}, u'high': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/hqdefault.jpg', u'width': 480, u'height': 360}, u'medium': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/mqdefault.jpg', u'width': 320, u'height': 180}, u'maxres': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/maxresdefault.jpg', u'width': 1280, u'height': 720}, u'standard': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/sddefault.jpg', u'width': 1280, u'height': 720}}, u'videoId': u'b6yDfgly-FA', u'videoCategory': u'20'}}, {u'kind': u'youtube#video', u'statistics': {u'commentCount': u'19714', u'viewCount': u'6503190', u'favoriteCount': u'0', u'dislikeCount': u'2601', u'likeCount': u'147260'}, u'contentDetails': {u'definition': u'hd', u'projection': u'rectangular', u'caption': u'false', u'duration': u'PT10M2S', u'licensedContent': True, u'dimension': u'2d', u'snippet': {u'thumbnails': {u'default': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/default.jpg', u'width': 120, u'height': 90}, u'high': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/hqdefault.jpg', u'width': 480, u'height': 360}, u'medium': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/mqdefault.jpg', u'width': 320, u'height': 180}, u'maxres': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/maxresdefault.jpg', u'width': 1280, u'height': 720}, u'standard': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/sddefault.jpg', u'width': 1280, u'height': 720}}, u'videoId': u'b6yDfgly-FA', u'videoCategory': u'20'}}, {u'kind': u'youtube#video', u'statistics': {u'commentCount': u'19714', u'viewCount': u'6503190', u'favoriteCount': u'0', u'dislikeCount': u'2601', u'likeCount': u'147260'}, u'contentDetails': {u'definition': u'hd', u'projection': u'rectangular', u'caption': u'false', u'duration': u'PT10M2S', u'licensedContent': True, u'dimension': u'2d', u'snippet': {u'thumbnails': {u'default': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/default.jpg', u'width': 120, u'height': 90}, u'high': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/hqdefault.jpg', u'width': 480, u'height': 360}, u'medium': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/mqdefault.jpg', u'width': 320, u'height': 180}, u'maxres': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/maxresdefault.jpg', u'width': 1280, u'height': 720}, u'standard': {u'url': u'https://i.ytimg.com/vi/b6yDfgly-FA/sddefault.jpg', u'width': 1280, u'height': 720}}, u'videoId': u'b6yDfgly-FA', u'videoCategory': u'20'}}]}]}

```

2. After we authorized, python file will get the data which contains detailed information about movies we entered in the search area.

II. Database Design

Shown as the Figure 1:

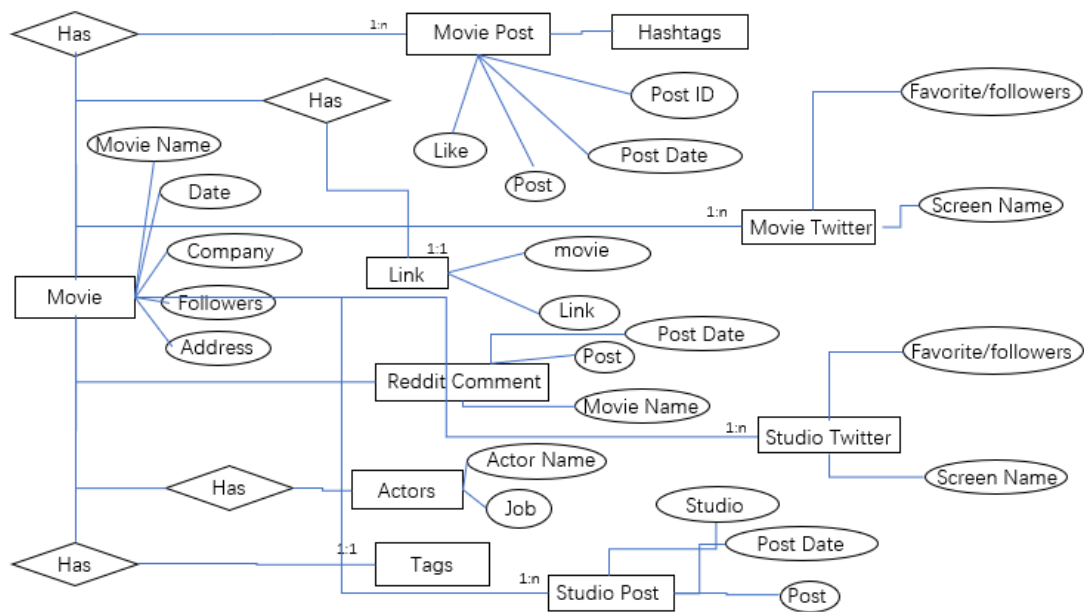


Figure 1: Database Conceptual Model

IV. Database Realization

Shown as the figure 2:

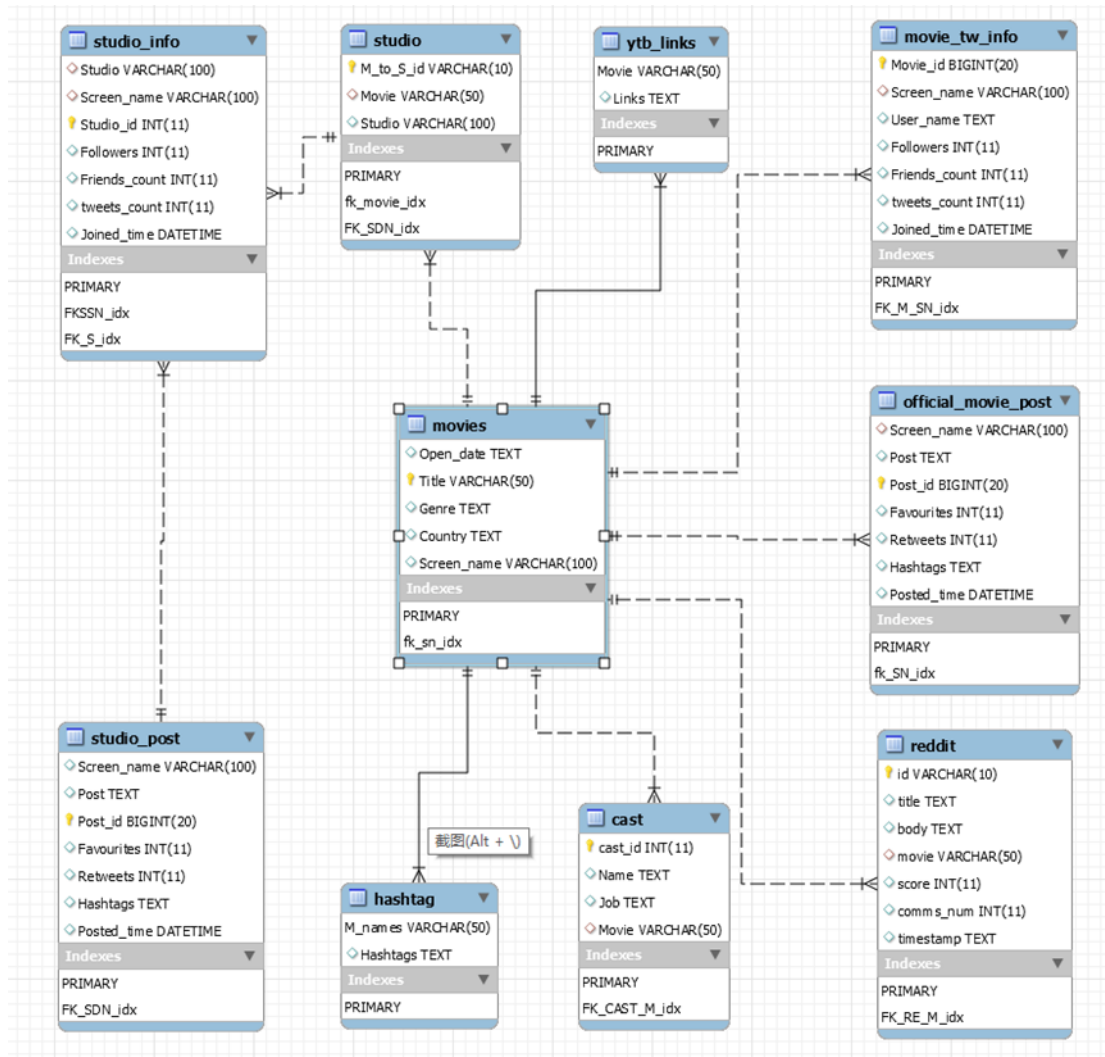


Figure 2: Database EER Diagram

V. Use Cases

Use case 1:

What are peoples saying about the movie? (user comments about the movie)

```

1 • select movie,title,body,timestamp from project.reddit
2   order by movie; /*use case 1 /

```

movie	title	body	timestamp
1917	Restoration of Lost 1917 Feminist Film [Xpost r/silentcinema]		2016-01-08 15:17:34
1917	World War 1 Film '1917 Has Begun Filming - Directed by Sam Mendes ('Skyfall'...		2019-04-01 14:28:07
1917	Looking for info about a silent movie	http://imgur.com/a/yRIqf So I went to an estat...	2016-06-29 13:59:44
1917	Charlie Chaplin's 1917 masterpiece, The Immigrant, gets a detailed commenta...		2015-02-26 06:20:47
1917	Charlie Chaplin Success	So lately I have been watching Chaplin movies: ...	2018-07-07 21:12:19
1917	POOR LITTLE RICH GIRL (1917) -- Mary Pickford		2015-04-10 08:49:09
1917	Yearly Cinema: 1917	Post your favorite movie of the year of current ...	2014-08-25 10:22:28
1917	The unknown women of Joan Fontaine (1917-2013): Fontaine was a remarka...		2014-02-01 09:58:30
1917	History of the Metro-Goldwyn-Mayer Logo (1917-2015)		2016-02-01 21:05:17
1917	I'm gonna have a mega marathon of "historical" movies depicting the entire n...	I haven't seen a huge chunk of these movies w...	2018-01-29 23:04:23
21 Bridges	First Poster for Russo Brothers produced 21 Bridges		2019-04-24 09:05:47
21 Bridges	So why is there so much backlash for The Greatest Showman when movies ha...	The backlash for The Greatest Showman is so o...	2017-12-25 11:57:50
21 Bridges	UNEXPECTED, GREAT MOVIES: THE LIST	Avalon Curious Case of Benjamin Button Mouse ...	2012-08-09 21:18:34
21 Bridges	Alternative Reddit Top 250	There has been quite a bit of criticism of the 20...	2017-06-08 01:26:48
21 Bridges	I've seen a movie everyday since the beginning of the year and will do so for ...	All of these movies I had not seen before, BTW...	2011-05-19 02:18:45
21 Bridges	Introducing r/MovieDirectory: A Place to Explore & Promote Movie Related S...	##Greetings from r/MovieDirectory! I've sta...	2013-09-01 13:06:54
21 Bridges	Learning Excel features + reading film lists = interesting idea.	So I was looking over [AFT's 100 years/100 mov...	2011-05-02 18:07:46
21 Bridges	Box Office Week: Miserable weekend sees bad openings for every new film.	Top Five Of The Week:** Rank Title Do...	2015-10-26 16:08:50
21 Bridges	Top 100 movies of all time!	So I found this list of the top 100 movies ever. A...	2011-01-31 11:45:38

Use case 2:

How viral are the movie posts?

```

1 • select Screen_name,Post,Favourites, Retweets from project.official_movie_post
2   order by Favourites desc; /*use case 2 /

```

result Grid Filter Rows: <input type="text"/> Export: Wrap Cell Content:				
Screen_name	Post	Favourites	Retweets	
disneylionking	July 19, 2019. #TheLionKing https://t.co/CYCR...	580301	259426	
toystory	Summer 2019. #ToyStory4 https://t.co/jown2p...	561716	236832	
SpiderManMovie	Watch the #SpiderManFarFromHome teaser tra...	455084	234388	
jokermovie	Put on a happy face. #JokerMovie - in theaters...	350584	151722	
toystory	Meet Ducky & Bunny. #ToyStory4 https://...	313591	120202	
toystory	On the road of life there are old friends, new fri...	270224	127980	
disneylionking	In 100 days, the king arrives. Watch the brand ...	249362	119252	
toystory	To Infinity and... #ToyStory4 https://t.co/dt06...	214946	87272	
Avengers	#DontSpoilTheEndgame https://t.co/Y1S1VRuvOb	161666	71927	
disneyaladdin	Watch the teaser-trailer for Disney's #Aladdin. ...	143431	70192	
JohnWickMovie	Whatever it takes. https://t.co/ZpLZdqfAkm	130454	52079	
Avengers	"Before we're done, we still have one promise t...	130186	47383	
the_prodigy	It is with deepest shock and sadness that we ca...	129747	47628	
SpiderManMovie	In theaters July 5. #SpiderManFarFromHome ...	126573	47863	
disneyaladdin	A whole new world. Watch the brand-new traile...	124705	49230	
Avengers	Screenshot. Share. #DontSpoilTheEndgame htt...	116330	83919	
Avengers	Don't do it. #DontSpoilTheEndgame https://t.co...	115863	48636	
DisneyFrozen	November 2019. #Frozen2 https://t.co/PhmDd...	113260	63986	
yesterdaymovie	Yesterday, everyone knew The Beatles. Today,...	99926	26347	
Avengers	We salute you, @ChrisEvans! #CaptainAmerica...	97155	19192	
captainmarvel	#CaptainMarvel herself, @brielarson, popped i...	90015	21218	
disneyaladdin	"Your life begins now... Aladdin". Watch this sp...	87596	41922	
DoraMovie	Get ready to unleash your wild side. 🦊 Watch t...	87279	24745	
Avengers	Welcome to the party, @RobertDowneyJr! #Ir...	85918	18013	

fficial movie post 7 ×

Use case 3:

What posts are like the movie I chose? (the movie tags)

```

1 • SELECT a.Hashtags, b.Open_date, b.Title, b.Country, b.Screen_name FROM project.hashtag as a
2   inner join
3   project.movies as b on
4     a.M_names = b.Title
5   ; /*use case 3 /

```

Hashtags	Open_date	Title	Country	Screen_name
War	25-Dec	1917	UK	
Action	12-Jul	21 Bridges	USA	21BridgesMovie
Horror, Thriller	28-Jun	47 Meters Down: Uncaged	US	47MDown
Biography, Drama	22-Nov	A Beautiful Day in the Neighborhood	USA	ABeautifulDay
Fantasy, Drama	17-May	A Dog's Journey	US	a_dogs_journey
Drama, Family	Jan-11	A Dog's Way Home	US	ADogsWayHome
Animation, Sci-Fi, Comedy, Family	18-Oct	A Shaun the Sheep Movie: Farmageddon	UK	shaunthesheep
Animation, Fantasy, Adventure, Comedy	27-Sep	Abominable	US, CN	
Sci-Fi, Drama, Thriller	24-May	Ad Astra	US	
Drama	Apr-19	After	US	aftermovie
Musical, Fantasy, Romance, Adventure	24-May	Aladdin	US	disneyaladdin
Sci-Fi, Action, Adventure, Romance	Feb-14	Alita: Battle Angel	US	AlitaMovie
Action, Thriller	23-Aug	Angel Has Fallen	US	
Horror, Thriller	28-Jun	Annabelle Comes Home	US	annabellemovie
Adventure, Fantasy, Sci-Fi	9-Aug	Artemis Fowl	USA	artemisfowl
Superhero, Epic, Action, Adventure	Apr-26	Avengers: Endgame	US	Avengers
Comedy, Drama	25-Oct	Bad Trip	USA	BadTripTheMovie
Action, Thriller	20-Sep	Black and Blue	USA	
Drama	14-Aug	Blinded by the Light	UK	BBTLmovie
Comedy	24-May	Booksmart	US	Booksmart
Faith, Drama	Apr-17	Breakthrough	US	SeeBreakthrough
Horror, Superhero, Sci-Fi	24-May	Brightburn	US	Brightburn
Live-Action Animation, Adventure, Drama	25-Dec	Call of the Wild	USA	
Superhero, Action, Adventure, Sci-Fi	Mar-08	Captain Marvel	US	captainmarvel

Use case 4:

Who should the new user be following? (the studio which has most followers)

```

1 • SELECT a.Studio, a.Followers, a.Friends_count, a.tweets_count, b.Movie FROM project.studio_info as a
2   inner join project.studio as b
3   on a.Studio = b.Studio
4   order by a.Followers desc
5   ; /*use case 4 /

```

Studio	Followers	Friends_count	tweets_count	Movie
Pixar	11390336	148	7318	Toy Story 4
Marvel Entertainment	7808087	780	50277	Dark Phoenix
Marvel Entertainment	7808087	780	50277	The New Mutants
Netflix	5976921	1203	27977	Isn't It Romantic
Netflix	5976921	1203	27977	Triple Frontier
Netflix	5976921	1203	27977	The Highwaymen
Netflix	5976921	1203	27977	The Dirt
Netflix	5976921	1203	27977	Extremely Wicked, Shockingly Evil, and Vile
Netflix	5976921	1203	27977	Velvet Buzzsaw
Walt Disney Pictures	5289469	278	10047	Maleficent: Mistress of Evil
Walt Disney Pictures	5289469	278	10047	Frozen 2
Walt Disney Pictures	5289469	278	10047	Dumbo
Walt Disney Pictures	5289469	278	10047	Aladdin
Walt Disney Pictures	5289469	278	10047	Toy Story 4
Walt Disney Pictures	5289469	278	10047	The Lion King
Walt Disney Pictures	5289469	278	10047	Artemis Fowl
Warner Bros.	4078659	873	21731	It: Chapter Two
Warner Bros.	4078659	873	21731	The Goldfinch
Warner Bros.	4078659	873	21731	The Lego Movie 2: The Second Part
Warner Bros.	4078659	873	21731	Joker
Warner Bros.	4078659	873	21731	Torrance
Warner Bros.	4078659	873	21731	Motherless Brooklyn
Warner Bros.	4078659	873	21731	Doctor Sleep
Warner Bros.	4078659	873	21731	The Good Liar

Use case 5:

What topics are trending in the domain? / which movie is most popular?


```

7 • SELECT a.Screen_name, a.Followers, a.Friends_count, a.tweets_count, b.Title FROM project.movie_tw_info as a
8   inner join project.movies as b
9   on a.Screen_name = b.Screen_name
10  order by a.Followers desc
11  ; /*use case 5 */

```

Screen_name	Followers	Friends_count	tweets_count	Title
Avengers	4390931	48	46741	Avengers: Endgame
starwars	4011167	418	35779	Star Wars: The Rise of Skywalker
SpiderManMovie	940531	17	4006	Spider-Man: Far From Home
captainmarvel	813896	22	292	Captain Marvel
XMenMovies	775738	200	24492	Dark Phoenix
DowntonAbbey	642854	66	13861	Downton Abbey
UnplannedMovie	351234	4446	921	Unplanned
the_prodigy	252877	703	2521	The Prodigy
JohnWiddMovie	129463	14828	2539	John Wick: Chapter 3 Parabellum
aftermovie	127189	28	2914	After
DetPikachuMovie	101404	16	440	Pok?mon: Detective Pikachu
toystory	100452	10	30	Toy Story 4
TheLEGOMovie	91268	33	2296	The Lego Movie 2: The Second P...
GodzillaMovie	91034	31	553	Godzilla: King of the Monsters
DisneyFrozen	77719	3	6	Frozen 2
UniversalHorror	72498	146	1234	Happy Death Day 2U
ShazamMovie	61891	8	241	Shazam!
disneylionking	55101	11	12	The Lion King
Terminator	54993	547	8	Terminator: Dark Fate
Maleficent	51963	8	282	Maleficent: Mistress of Evil
jokermovie	50447	0	2	Joker
shaunthesheep	32826	194	3998	A Shaun the Sheep Movie: Farm...
PetsMovie	32759	54	12820	The Secret Life of Pets 2

Use case 6:

What keywords/ hashtags should I add to the posts?

```

• SELECT a.title, a.body, a.timestamp, b.Title, b.Genre as Tags FROM project.reddit as a
  inner join project.movies as b
  on a.movie = b.Title
  order by b.Title
  ; /*use case 6 */

```

title	body	timestamp	Title	Tags
POOR LITTLE RICH GIRL (1917) -- Mary Pickford		2015-04-10 08:49:09	1917	War
How macabre cinema has evolved over the last 122 years.	1.The Execution Of Mary, Queen of Scots - 189...	2016-10-09 16:22:24	1917	War
Sam Mendes' WWI Film '1917' Is Reportedly Comprised Of One Long Shot		2019-04-05 11:50:35	1917	War
The time frame of the Best Picture winners: How many years in the past/futu...	With the Oscars coming up this weekend, I star...	2016-02-25 19:09:32	1917	War
Cleopatra (1917) has 136 votes on IMDb. It's a missing film!		2011-06-11 11:27:05	1917	War
History of the Metro-Goldwyn-Mayer Logo (1917-2015)		2016-02-01 21:05:17	1917	War
Charlie Chaplin Success	So lately I have been watching Chaplin movies: ...	2018-07-07 21:12:19	1917	War
Can someone recommend some movies? I've included some of my favorites in...	###Movies I absolutely LOVE ##Discloser: I li...	2015-01-15 17:44:44	21 B...	Action
Your top 25 performances by an actor or actress	You're allowed no more than ONE cheat (an ens...	2013-11-15 17:08:56	21 B...	Action
your picks for best scores in cinema 1970s -2016	1971 Straw Dogs – Jerry Fielding 1972 Sleuth ...	2017-05-06 11:05:59	21 B...	Action
So why is there so much backlash for The Greatest Showman when movies ha...	The backlash for The Greatest Showman is so o...	2017-12-25 11:57:50	21 B...	Action
Every movie i watched throughout the year.	In 2014 recovering from a spine operation with ...	2016-01-03 13:05:02	21 B...	Action
These are all the movies/documentaries I watched in 2014 for the first time	1. La Fille Sur Le Pond (girl on the bridge) 2.La ...	2015-01-05 23:57:00	21 B...	Action
Your Top 25 Films Of All-Time?	I had taken a long break from consistently watc...	2018-07-27 17:48:16	21 B...	Action
Now that the year's wrapped up, what's everyone's Best of the Year List? My...	I did have the great fortune of seeing some of t...	2015-12-31 20:48:52	21 B...	Action
What's everyones top movies of 1977?	Since the year is now 40 years old, what are yo...	2017-01-02 10:15:26	21 B...	Action
Box Office Week: Goosebumps surprises with impressive #1 debut at \$23.5 ...	***Top Five Of The Week:*** Rank Title Do...	2015-10-19 12:46:43	21 B...	Action
A devoted Oscar fan has timed the nominated performances of 2015.	Matthew Stewart has timed every Oscar winnin...	2016-06-14 17:20:43	21 B...	Action
All the movies I watched in 2015	Movies of 2015 I did this in 2014 and no one ca...	2016-01-05 00:25:56	21 B...	Action
UNEXPECTED, GREAT MOVIES: THE LIST	Avalon Curious Case of Benjamin Button Mouse ...	2012-08-09 21:18:34	21 B...	Action
First Poster for Russo Brothers produced 21 Bridges		2019-04-24 09:05:47	21 B...	Action
21 Bridges Official Trailer Chadwick Boseman,		2019-04-25 09:04:03	21 B...	Action
What movie from my favorite movie list is your personal favorite?	1.Life of Brian 2.Goodfellas 3.Inglourious Bast...	2015-08-07 11:31:52	21 B...	Action

Use case 7:

What movies would be shown in May?

```
18
19 • SELECT * FROM project.movies where Open_date like '%May%'
20 ; /*use case 7 */
```

	Open_date	Title	Genre	Country	Screen_name
▶	17-May	A Dog's Journey	Fantasy, Drama	US	a_dogs_journey
	24-May	Ad Astra	Sci-Fi, Drama, Thriller	US	
	24-May	Aladdin	Musical, Fantasy, Romance, Adventure	US	disneyaladdin
	24-May	Booksmart	Comedy	US	Booksmart
	24-May	Brightburn	Horror, Superhero, Sci-Fi	US	Brightburn
	31-May	Domino	Crime, Thriller	BE, DK, FR, ESP	
	May-03	Extremely Wicked, Shockingly Evil, and Vile	Biography, Crime, Thriller, Drama	US	
	31-May	Godzilla: King of the Monsters	Action, Adventure, Sci-Fi	US	GodzillaMovie
	17-May	John Wick: Chapter 3 Parabellum	Action, Thriller	US	JohnWickMovie
	May-03	Long Shot	Comedy	US	LongShotMovie
	31-May	Ma	Horror, Thriller	US	Mamovie
	May-10	Pokémon: Detective Pikachu	Sci-Fi, Fantasy, Action, Mystery, Com...	US	DetPikachuMovie
	May-10	Poms	Comedy	US	PomsMovie
	31-May	Rocketman	Biography, Drama, Musical	US	rocketmanmovie
	May-10	The Hustle	Comedy	US	hustlemovie
	May-03	The Intruder	Thriller	US	IntruderMovie
	10-May	The Professor and the Madman	Drama	US	
	17-May	The Sun Is Also a Star	Drama, Romance	US	sunisalsoastar
	17-May	The Tomorrow Man	Drama	US	TomorrowManFilm
	May-10	Tolkien	Biography, Drama	UK	TolkienMovie
	May-03	UglyDolls	Animation, Comedy	US	uglydolls

Use case 8:

Shown the newest 10 posts from the movie Avengers.

```
22 • SELECT * FROM project.official_movie_post where Screen_name = 'Avengers'
23 order by Posted_time desc
24 limit 10
25 ; /*use case 8 */
```

Screen_name	Post	Post_id	Favourites	Retweets	Hashtags	Posted_time
Avengers	See the film critics call a "once in a generation e...	1121791047443468288	3849	721	AvengersEndgame	2019-04-26 11:00:02
Avengers	RT @FortniteGame: Take on Thanos and the Ch...	1121505464632500224	0	4279		2019-04-25 16:05:13
Avengers	We're in the endgame now. Download the bran...	1121489054372106242	16083	3372	AvengersEndgame	2019-04-25 15:00:01
Avengers	Nothing can prepare you for the endgame. See ...	1121466399866544129	12203	1872	AvengersEndgame	2019-04-25 13:30:00
Avengers	Don't do it. #DontSpoilTheEndgame https://t.co...	1121451301290663938	115863	48636	DontSpoilTheEndgame	2019-04-25 12:30:00
Avengers	Tonight, see the film critics call "epic like you've ...	1121428665114382338	27250	4998		2019-04-25 11:00:03
Avengers	Here's the Marvel Studios' #AvengersEndgame-...	1121187059102699521	26768	4878	AvengersEndgame	2019-04-24 19:00:00
Avengers	Part of the journey is the end. Marvel Studios' ...	1121171960430518272	22558	3353	AvengersEndgame	2019-04-24 18:00:00
Avengers	Paul Rudd gets an exclusive sneak peek at #Av...	1121113188399890432	22821	4176	AvengersEndgame	2019-04-24 14:06:28
Avengers	Be the first to see Marvel Studios' #AvengersEn...	1121096466112577537	15370	2214	AvengersEndgame	2019-04-24 13:00:01

Use case 9:

Shown the movie shown in 2019 acted by Robert Downey Jr..

```
28 • select a.* ,b.*from project.cast as a
29 inner join project.movies as b
30 on a.Movie = b.Title
31 where name = 'Robert Downey Jr.' ;
32 /*use case 9 */
33
34
```

cast_id	Name	Job	Movie	Open_date	Title	Genre	Country	Screen_name
▶ 4603	Robert Downey Jr.	actor	Avengers: Endgame	Apr-26	Avengers: Endgame	Superhero, Epic, Action, Adventure	US	Avengers

Use case 10:

Shown the movies are belong to action movies.

```
--
34 • select a.Title,a.Open_date,a.Country ,b.*from project.movies as a
35 inner join project.hashtag as b
36 on a.Title = b.M_names
37 where b.Hashtags like '%Action%'
38 /*use case 10 */
```

Result Grid				
Filter Rows:		Export:		Wrap Cell Content: IA
Title	Open_date	Country	M_names	Hashtags
21 Bridges	12-Jul	USA	21 Bridges	Action
Alita: Battle Angel	Feb-14	US	Alita: Battle Angel	Sci-Fi, Action, Adventure, Romance
Angel Has Fallen	23-Aug	US	Angel Has Fallen	Action, Thriller
Avengers: Endgame	Apr-26	US	Avengers: Endgame	Superhero, Epic, Action, Adventure
Black and Blue	20-Sep	USA	Black and Blue	Action, Thriller
Call of the Wild	25-Dec	USA	Call of the Wild	Live-Action Animation, Adventure, Drama
Captain Marvel	Mar-08	US	Captain Marvel	Superhero, Action, Adventure, Sci-Fi
Captive State	Mar-15	US	Captive State	Sci-Fi, Thriller, Action
Charlie's Angels	15-Nov	USA	Charlie's Angels	Action, Comedy
Cold Pursuit	Feb-08	US	Cold Pursuit	Action, Thriller
Dark Phoenix	7-Jun	US	Dark Phoenix	Superhero, Action, Adventure
Fast & Furious Pres...	2-Aug	USA	Fast & Furious Pres...	Action, Adventure
Ford v. Ferrari	15-Nov	US	Ford v. Ferrari	Biography, Action, Drama
Gemini Man	11-Oct	TBA	Gemini Man	Sci-Fi, Action, Thriller
Godzilla: King of th...	31-May	US	Godzilla: King of th...	Action, Adventure, Sci-Fi
How to Train Your ...	Feb-22	US	How to Train Your ...	Animation, Fantasy, Action, Adventure
John Wick: Chapter...	17-May	US	John Wick: Chapter...	Action, Thriller

Use case 11:

Show the movie comments on movie Avengers.

```
--
40 • select title,body,score,comms_num from project.reddit
41 where movie like '%Avengers%'
42 order by score desc;
43 /*use case 11 */
```

Result Grid			
Filter Rows:		Export:	
Wrap Cell Content: IA			
title	body	score	comms_num
Marvel Studios' Avengers: Endgame - Official Trailer		73289	12058
AVENGERS: ENDGAME official poster		42642	1731
Avengers: Endgame (2019) Official Poster		33678	3248
Tom Holland to Star Crime-Drama 'Cherry' - Directed by the Russo Brothers (...)		7385	357
Marvel's subversive humour is now predictable and cliché	We all love that moment in The Avengers when ...	3266	1767
So apparently there's this film called Avengers: Endgame coming out this wee...	*This post is going to be locked in order to avoi...	1465	0
Live Thread - Avengers: Endgame [SPOILERS]	*Welcome to the official live thread for Avenger...	887	10663
Why Not Bring Back Intermissions in Movies?	With *Avengers: Endgame* coming out I keep s...	329	472
Tomorrow I will see my 2,000th movie	I recently counted all the movies I'd ever seen ...	282	111
What's your favourite non-MCU superhero movie?	With the climax of the MCU (and probably the e...	59	311
Paul Verhoeven would have been the perfect Director for Ready Player One	Paul Verhoeven is one of my favorite directors. ...	37	31
Avengers Endgame is predicted to make 282 Million at the Box Office Opening...		31	28
The 52 Most Anticipated Movies of 2019		21	34
Who are some of the rising-star actors in The Hollywood that has the best ag...	What comes to mind first for me is Domnhall Gle...	15	38
What movie will be the biggest at the box office this year? What will be the bi...	Worldwide box office. There are a few giants t...	14	67
Is there really any difference in seeing a blockbuster film on opening night vs ...	As we all know avengers endgame, among man...	11	21
Does 4D theatre worth the price tag and experience?	So my country just opened the first ever 4D the...	10	27

Use case 12:

Show the YouTube watching link for the movie Avengers.

```
45 • select movie,links from project.ytb_links
46     where movie like '%Avenger%';
47     /*use case 12 */
48
49
```

movie	links
Avengers: Endgame	https://www.youtube.com/watch?v=ePpJDKfR...
NULL	NULL

Use case 13:

Show the movies belong to the 20th Century Fox.

```
48
49 • select * from project.studio
50     where studio like '%20th%';
51     /*use case 13 */
52
```

M_to_S_id	Movie	Studio
104a	Spies in Disguise	20th Century Fox
111a	The Art of Racing in the Rain	20th Century Fox
113a	The Woman in the Window	20th Century Fox
130a	Ford v. Ferrari	20th Century Fox
143a	Call of the Wild	20th Century Fox
16a	Alita: Battle Angel	20th Century Fox
41a	Breakthrough	20th Century Fox
61a	Ad Astra	20th Century Fox
68a	Dark Phoenix	20th Century Fox
6a	The Kid Who Would Be King	20th Century Fox
82a	Stuber	20th Century Fox
90a	The New Mutants	20th Century Fox
NULL	NULL	NULL

Use case 14:

Show how many movies would be on the screen in May.

```
53 • select count(*) from project.movies
54     where Open_date like '%May%';
55     /*use case 14 */
56
57
```

count(*)
21

Use case 15:

Show the studio which has the most followers/fans.

```

57 • select *,max(Followers) from project.studio_info;
58 /*use case 15 */

```

Studio	Screen_name	Studio_id	Followers	Friends_count	tweets_count	Joined_time	max(Followers)
Fox Searchlight Pictures	foxsearchlight	3291841	722164	1245	8004	2007-04-03 02:03:55	11390336

VI. Functions

Function 1:

DELIMITER //

CREATE FUNCTION Counttag(tags TEXT)

RETURNS INT

DETERMINISTIC

BEGIN

DECLARE count INT;

SELECT COUNT(*)

INTO count

FROM project.hashtag

where Hashtags like concat('%',tags,'%');

RETURN count;

end;//

DELIMITER ;

The screenshot shows a SQL IDE window with a tab titled 'SQL File 4'. The main editor contains the following SQL code:

```
1 DELIMITER //  
2 CREATE FUNCTION Counttag(tags TEXT)  
3 RETURNS INT  
4 DETERMINISTIC  
5 BEGIN  
6 DECLARE count INT;  
7 SELECT COUNT(*)  
8 INTO count  
9 FROM project.hashtag  
10 where Hashtags like concat('%',tags,'%');  
11 RETURN count;  
12 end; //  
13 DELIMITER ;  
14  
15 select project.Counttag('Action');  
16
```

Below the editor, there is a 'Result Grid' section. It includes a 'Filter Rows' input field, an 'Export' button, and a 'Wrap Cell Content' checkbox. The result grid shows the following data:

project.Counttag('Action')
36

Function 2:

```
DELIMITER //  
CREATE FUNCTION Search(movies TEXT)  
RETURNS BOOLEAN  
DETERMINISTIC  
BEGIN  
DECLARE ct_movies TEXT;  
SELECT count(*)  
INTO ct_movies  
FROM project.movies  
where Title = movies;  
IF ct_movies IS NULL THEN RETURN FALSE;  
ELSEIF ct_movies > 0 THEN RETURN TRUE;  
ELSE RETURN FALSE;  
END IF;  
end; //  
DELIMITER ;
```

```

16
17 DELIMITER //
18 • CREATE FUNCTION Search(movies TEXT)
19 RETURNS BOOLEAN
20 DETERMINISTIC
21 BEGIN
22 DECLARE ct_movies TEXT;
23 SELECT count(*)
24 INTO ct_movies
25 FROM project.movies
26 where Title = movies;
27 IF ct_movies IS NULL THEN RETURN FALSE;
28 ELSEIF ct_movies > 0 THEN RETURN TRUE;
29 ELSE RETURN FALSE;
30 END IF;
31 end; //
32 DELIMITER ;
33
34 • select project.Search('Avengers: Endgame');

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	project.Search('Avengers: Endgame')			
▶	1			

Function 3:

```

DELIMITER //
CREATE FUNCTION Matches(g text)
RETURNS INT
DETERMINISTIC
BEGIN
DECLARE ct_g INT;
SELECT COUNT(*)
INTO ct_g
FROM project.reddit
where title like concat('%',g,'%');
RETURN ct_g;
end; //
DELIMITER ;

```

```

35
36 DELIMITER //
37 • CREATE FUNCTION Matches(g text)
38 RETURNS INT
39 DETERMINISTIC
40 BEGIN
41 DECLARE ct_g INT;
42 SELECT COUNT(*)
43 INTO ct_g
44 FROM project.reddit
45 where title like concat('%',g,'%');
46 RETURN ct_g;
47 end;//
48 DELIMITER ;
49
50 • select Matches("Avengers");

```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	Matches("Avengers")			
▶	23			

Function 4:

```

DELIMITER //
CREATE FUNCTION SearchTimeBefore(time text)
RETURNS INT
DETERMINISTIC
BEGIN
DECLARE ct_date INT;
SELECT count(*)
INTO ct_date
FROM project.reddit
where timestamp < time;
RETURN ct_date;
end;//
DELIMITER ;

```



```

51
52 DELIMITER //
53 • CREATE FUNCTION SearchTimeBefore(time text)
54 RETURNS INT
55 DETERMINISTIC
56 BEGIN
57 DECLARE ct_date INT;
58 SELECT count(*)
59 INTO ct_date
60 FROM project.reddid
61 where timestamp < time;
62 RETURN ct_date;
63 end;//
64 DELIMITER ;
65
66 • select SearchTimeBefore('2019-04-01 14:02:29');

```

SearchTimeBefore('2019-04-01 14:02:29')
4689

Function 5:

```

CREATE FUNCTION Search_Cast(Cast TEXT)
RETURNS BOOLEAN
DETERMINISTIC
BEGIN
DECLARE ct_cast TEXT;
SELECT count(*)
INTO ct_cast
FROM project.cast
where Name = cast;
IF ct_cast IS NULL THEN RETURN FALSE;
ELSEIF ct_cast > 0 THEN RETURN TRUE;
ELSE RETURN FALSE;
END IF;
end;//
DELIMITER ;

```

```

67
68 DELIMITER //
69 • CREATE FUNCTION Search_Cast(Cast TEXT)
70 RETURNS BOOLEAN
71 DETERMINISTIC
72 BEGIN
73 DECLARE ct_cast TEXT;
74 SELECT count(*)
75 INTO ct_cast
76 FROM project.cast
77 where Name = cast;
78 IF ct_cast IS NULL THEN RETURN FALSE;
79 ELSEIF ct_cast > 0 THEN RETURN TRUE;
80 ELSE RETURN FALSE;
81 END IF;
82 end;//
83 DELIMITER ;
84
85 • select Search_Cast('Robert Downey Jr.');
```

Result Grid		Filter Rows:		Export:		Wrap Cell Content:	
	Search_Cast('Robert Downey Jr.')						
	1						

Function 6:

```

DELIMITER //
CREATE FUNCTION Count_movies_producedby_studio(m TEXT)
RETURNS INT
DETERMINISTIC
BEGIN
DECLARE count INT;
SELECT COUNT(*)
INTO count
FROM project.studio
where studio like concat('%',m,'%');
RETURN count;
end;//
DELIMITER ;
```

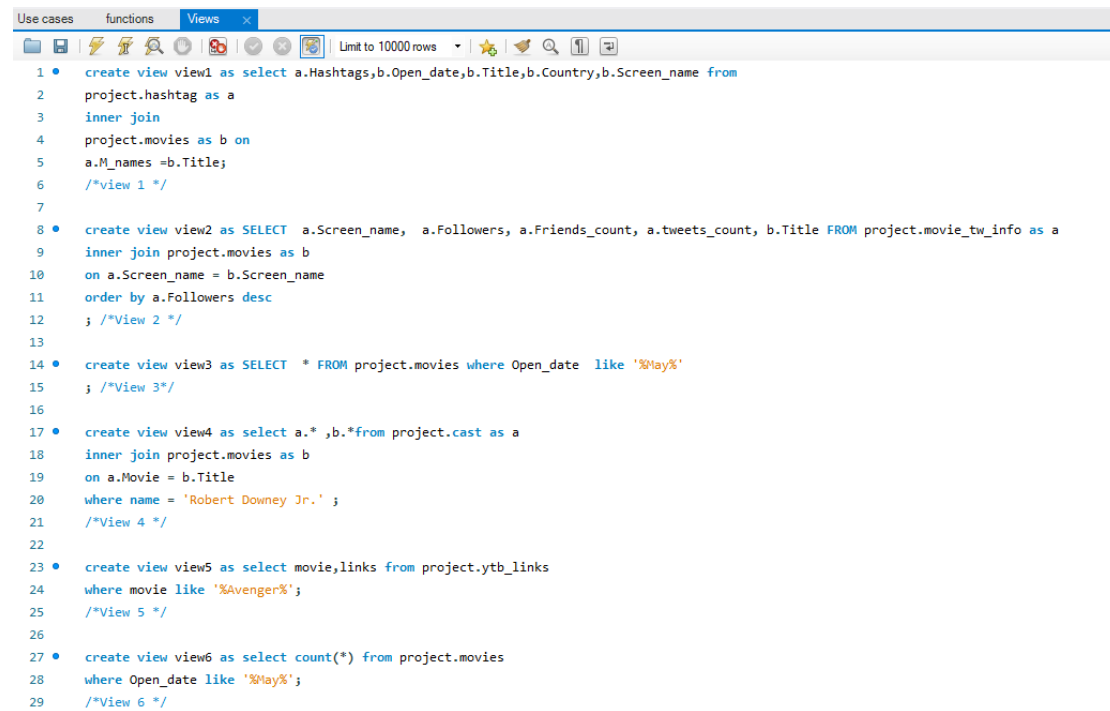
```

86
87     DELIMITER //
88 • CREATE FUNCTION Count_movies_producedby_studio(m TEXT)
89     RETURNS INT
90     DETERMINISTIC
91     BEGIN
92     DECLARE count INT;
93     SELECT COUNT(*)
94     INTO count
95     FROM project.studio
96     where studio like concat('%',m,'%');
97     RETURN count;
98     end;//
99     DELIMITER ;
100
101 • select project.Count_movies_producedby_studio('Warner Bros.');
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	project.Count_movies_producedby_studio('Warner Bros.')			
▶	20			

VII. Views

Views built:



The screenshot shows a SQL editor interface with a toolbar at the top and a list of six SQL queries. The queries are numbered 1 through 6 and are used to create database views. The queries are as follows:

```
1 • create view view1 as select a.Hashtags,b.Open_date,b.Title,b.Country,b.Screen_name from
2 project.hashtag as a
3 inner join
4 project.movies as b on
5 a.M_names =b.Title;
6 /*view 1 */
7
8 • create view view2 as SELECT a.Screen_name, a.Followers, a.Friends_count, a.tweets_count, b.Title FROM project.movie_tw_info as a
9 inner join project.movies as b
10 on a.Screen_name = b.Screen_name
11 order by a.Followers desc
12 ; /*View 2 */
13
14 • create view view3 as SELECT * FROM project.movies where Open_date like '%May%'
15 ; /*View 3*/
16
17 • create view view4 as select a.* ,b.*from project.cast as a
18 inner join project.movies as b
19 on a.Movie = b.Title
20 where name = 'Robert Downey Jr.' ;
21 /*View 4 */
22
23 • create view view5 as select movie,links from project.ytb_links
24 where movie like '%Avenger%';
25 /*View 5 */
26
27 • create view view6 as select count(*) from project.movies
28 where Open_date like '%May%';
29 /*View 6 */
```

View 1:

```
create view view1 as select a.Hashtags,b.Open_date,b.Title,b.Country,b.Screen_name from
project.hashtag as a
inner join
project.movies as b on
a.M_names =b.Title;
/*view 1 */
```


Use cases view1 view2 view3 view4 view5 view6

Limit to 10000 rows

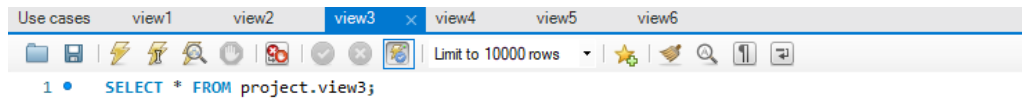
1 • SELECT * FROM project.view2;

Result Grid Filter Rows: Export: Wrap Cell Content: [IA](#)

	Screen_name	Followers	Friends_count	tweets_count	Title
▶	Avengers	4390931	48	46741	Avengers: Endgame
	starwars	4011167	418	35779	Star Wars: The Rise of Skywalker
	SpiderManMovie	940531	17	4006	Spider-Man: Far From Home
	captainmarvel	813896	22	292	Captain Marvel
	XMenMovies	775738	200	24492	Dark Phoenix
	DowntonAbbey	642854	66	13861	Downton Abbey
	UnplannedMovie	351234	4446	921	Unplanned
	the_prodigy	252877	703	2521	The Prodigy
	JohnWickMovie	129463	14828	2539	John Wick: Chapter 3 Parabellum
	aftermovie	127189	28	2914	After
	DetPikachuMovie	101404	16	440	Pok?mon: Detective Pikachu
	toystory	100452	10	30	Toy Story 4
	TheLEGOMovie	91268	33	2296	The Lego Movie 2: The Second P...
	GodzillaMovie	91034	31	553	Godzilla: King of the Monsters
	DisneyFrozen	77719	3	6	Frozen 2
	UniversalHorror	72498	146	1234	Happy Death Day 2U
	ShazamMovie	61891	8	241	Shazam!

View 3:

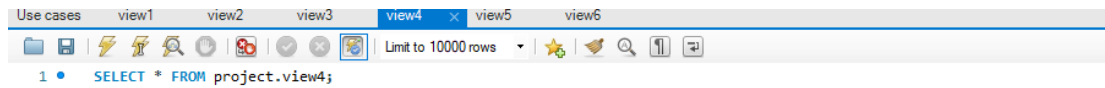
```
create view view3 as SELECT * FROM project.movies where Open_date like '%May%'
; /*View 3*/
```



Open_date	Title	Genre	Country	Screen_name
17-May	A Dog's Journey	Fantasy, Drama	US	a_dogs_journey
24-May	Ad Astra	Sci-Fi, Drama, Thriller	US	
24-May	Aladdin	Musical, Fantasy, Romance, Adventure	US	disneyaladdin
24-May	Booksmart	Comedy	US	Booksmart
24-May	Brightburn	Horror, Superhero, Sci-Fi	US	Brightburn
31-May	Domino	Crime, Thriller	BE, DK, FR, ESP	
May-03	Extremely Wicked, Shockingly Evil, and Vile	Biography, Crime, Thriller, Drama	US	
31-May	Godzilla: King of the Monsters	Action, Adventure, Sci-Fi	US	GodzillaMovie
17-May	John Wick: Chapter 3 Parabellum	Action, Thriller	US	JohnWickMovie
May-03	Long Shot	Comedy	US	LongShotMovie
31-May	Ma	Horror, Thriller	US	Mamovie
May-10	Pokémon: Detective Pikachu	Sci-Fi, Fantasy, Action, Mystery, Com...	US	DetPikachuMovie
May-10	Poms	Comedy	US	PomsMovie
31-May	Rocketman	Biography, Drama, Musical	US	rocketmanmovie
May-10	The Hustle	Comedy	US	hustlemovie
May-03	The Intruder	Thriller	US	IntruderMovie
10-May	The Professor and the Madman	Drama	US	

View 4:

```
create view view4 as select a.* ,b.*from project.cast as a
inner join project.movies as b
on a.Movie = b.Title
where name = 'Robert Downey Jr.' ;
/*View 4 */
```



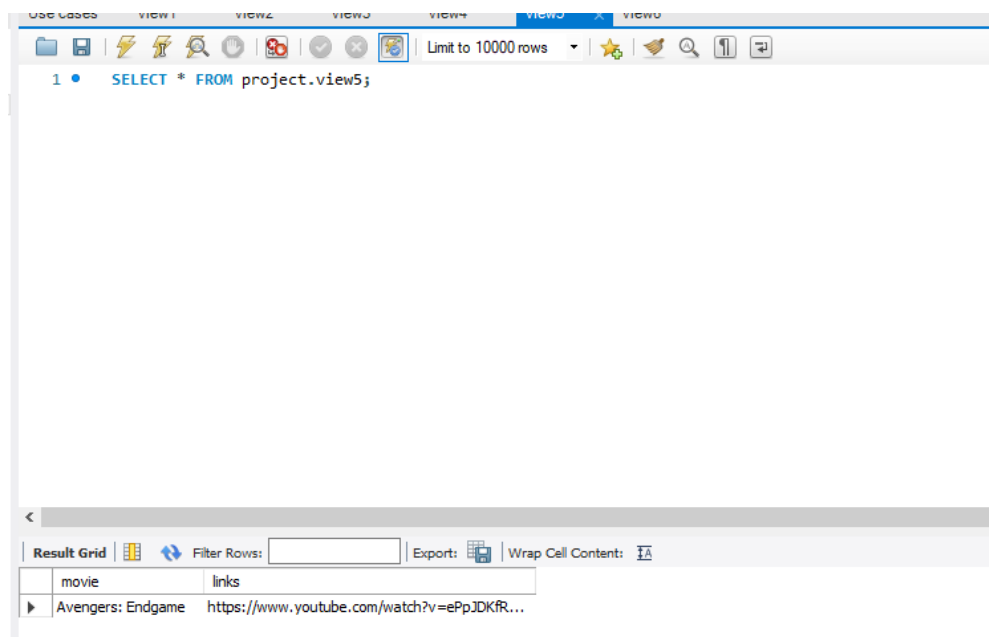
cast_id	Name	Job	Movie	Open_date	Title	Genre	Country	Screen_name
4603	Robert Downey Jr.	actor	Avengers: Endgame	Apr-26	Avengers: Endgame	Superhero, Epic, Action, Adventure	US	Avengers

View 5:

create view view5 as select movie,links from project.ytb_links

where movie like '%Avenger%';

/*View 5 */

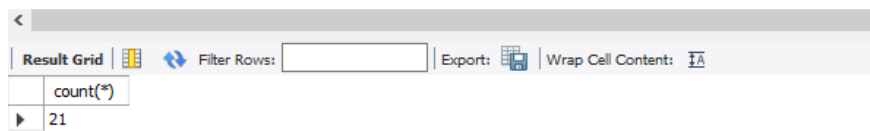
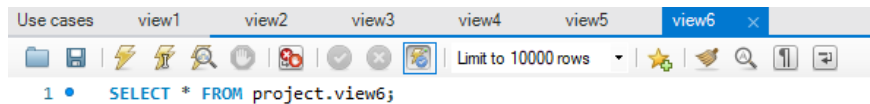


View 6:

create view view6 as select count(*) from project.movies

where Open_date like '%May%';

/*View 6 */



VIII. Summary

In conclusion, we can use this database catch the newest information about the movies shown in 2019 for people who are interested in recent news about movies. People can also choose films they are interested based on the tags built for the movies. Also, we can see the popular tags and topics that people talk about and comment on to know the trends regarding the movies from Reddit. Besides, people can select the studios to find the movies from their favorite studios then watch them through the YouTube watching link provided.

IX. References

1. <https://medium.com/@plog397/webscraping-reddit-python-reddit-api-wrapper-praw-tutorial-for-windows-a9106397d75e>
2. <http://t-redactyl.io/blog/2015/11/analysing-reddit-data-part-2-extracting-the-data.html>
3. <http://t-redactyl.io/blog/2015/11/analysing-reddit-data-part-2-extracting-the-data.html>
4. <https://stackoverflow.com/questions/45529507/python-unicodedecodeerror-utf-8-codec-cant-decode-byte-0x96-in-position-35>
5. <https://stackoverflow.com/questions/7703865/going-from-twitter-date-to-python-datetime-date>
6. <https://stackoverflow.com/questions/7703865/going-from-twitter-date-to-python-datetime-date>
- 7.