**INFO 6210 Gaurav Suresh Satam 001428363**

**Movie Database**

**Abstract**

Movie database is a SQL database created for the purpose of viewing basic information about movies such as casts, directors, production houses, tweets regarding movies, profile details of users who tweeted about the movie. It is designed as a one - stop destination for the users to access the movies that are released. This database has the tables which store details like movie released date, IMDB rating, etc.

**Introduction**

In the current scenario, a moviegoer ( user ) has to visit more than one website to get following data related to movies:

* Released movies
* IMDB rating
* Year released, cast, director, production house

There are websites like [www.imdb.com](http://www.imdb.com) and [www.rottentomatoes.com](http://www.rottentomatoes.com) with rich amount of the aforementioned data but user has to open at least 3 – 4 websites to view all he data. So, this project was started with the intention of developing a one – stop destination for users to view all the data.

The data from the websites was fetched by calling the APIs and putting it together in the database called “Movie Database”.

**Implementation**

**Technologies Used:**

* Python
* SQL

**Python:**

I have used python to scrape the data from the twitter API and stored the scraped data in excel files.

**SQL:**

Excels obtained as a result of python web scraping are modelled in the form of SQL tables. I have used Microsoft SSMS for this purpose.

**Web Scraping**

I have used twitter API for this project. I scraped the tweets posted by users related to movie and user details like follower counts, retweet count, etc.

**Data Source**

I downloaded a CSV consisting of movie details like movie name, casts, director, production house, IMDB rating from Kaggle.

<https://www.kaggle.com/carolzhangdc/imdb-5000-movie-dataset>

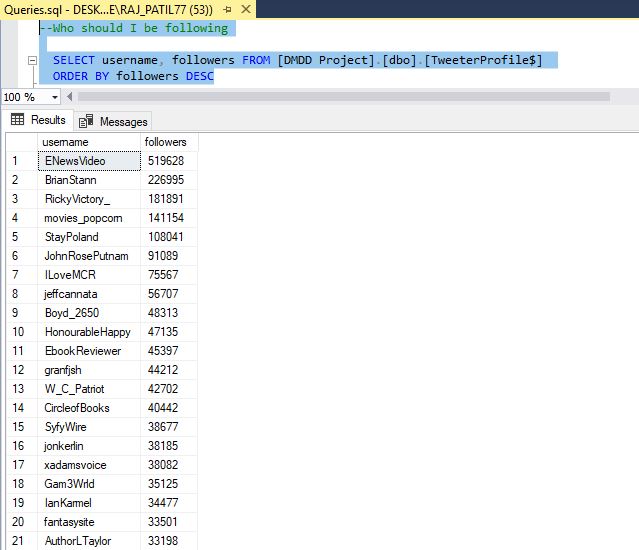
**ER Diagram**



**Use Cases:**

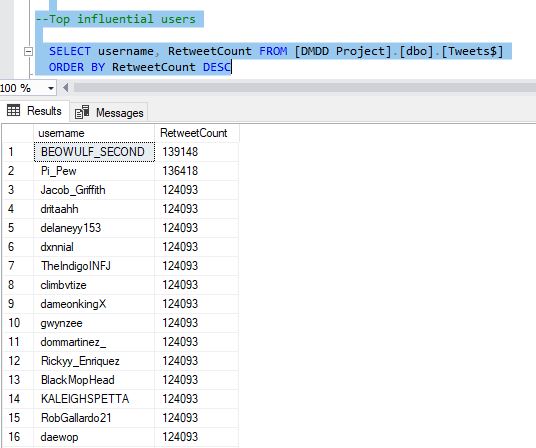
* Who should I be following:

The twitter user who has highest number of followers is the most popular user. And, other users will look forward to following that account.



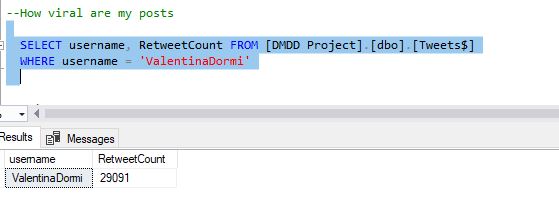
* Who is the most influential user

The user who has highest number of retweet count is the most influential user



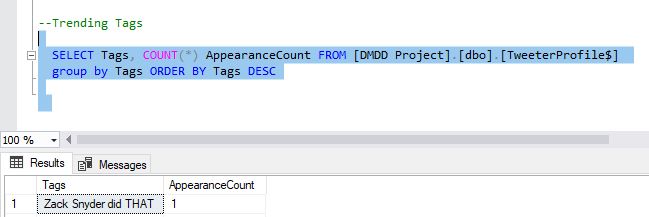
* Checking how viral are posts of a particular user

It can be decided by the number of retweet count

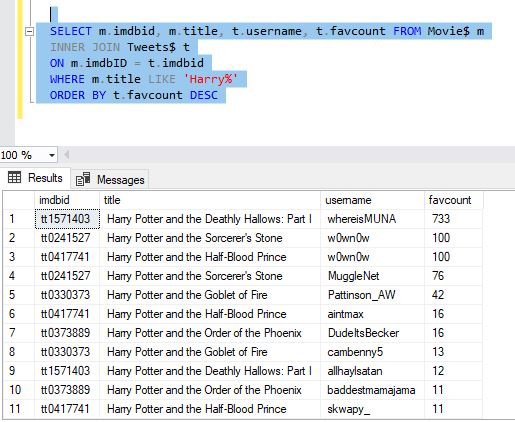


* Finding trending Tags

It can be decided by the number of times a particular tab appears in twitter tweets

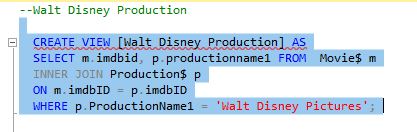


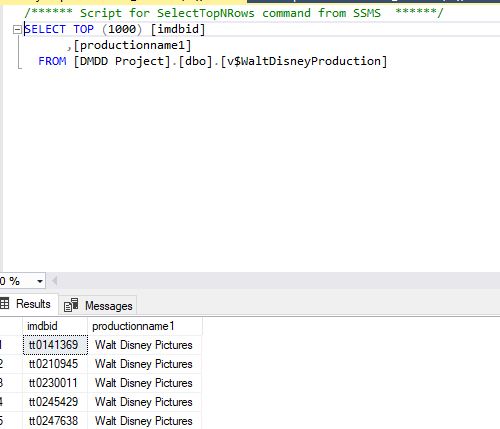
* Query to fetch username and favourite count of tweets associated with all Harry Potter parts



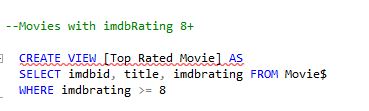
**Views**

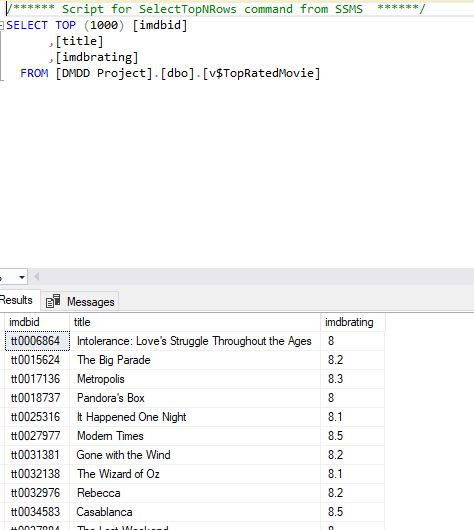
* View to fetch movies under Walt Disney Production



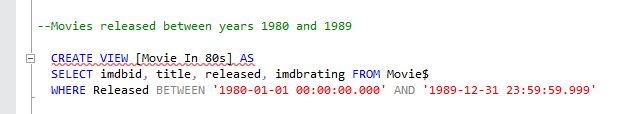
****

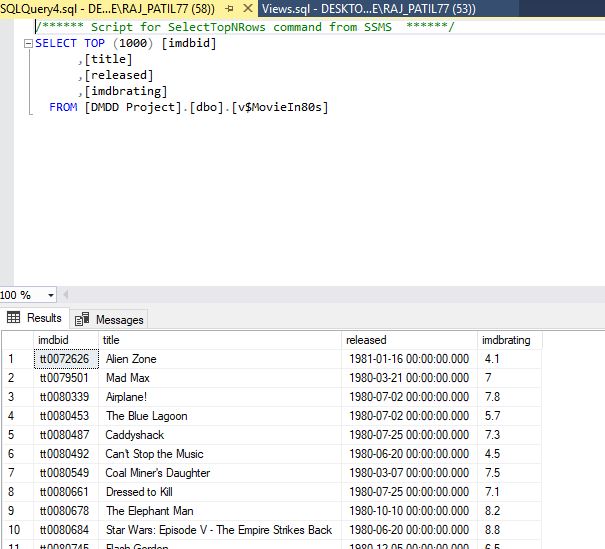
* View to fetch most popular movie ( 8+ rating )



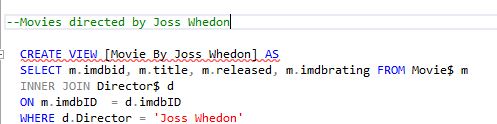


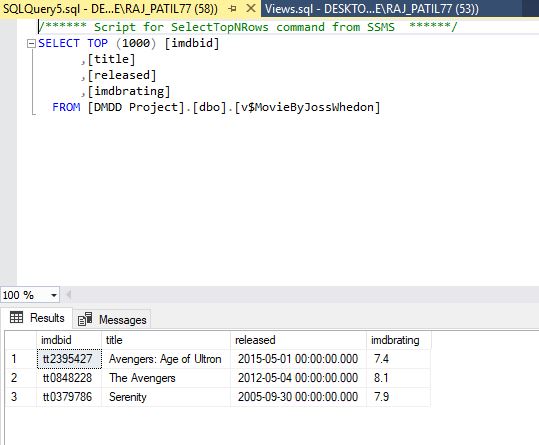
* View to get movies released in 80s



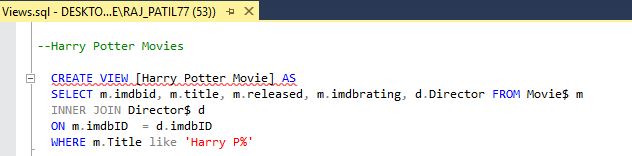


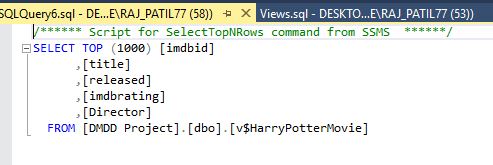
* View to fetch movies directed by particular director





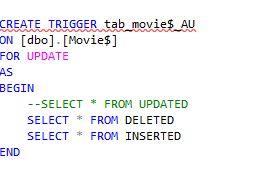
* View to fetch all Harry Potter Movies

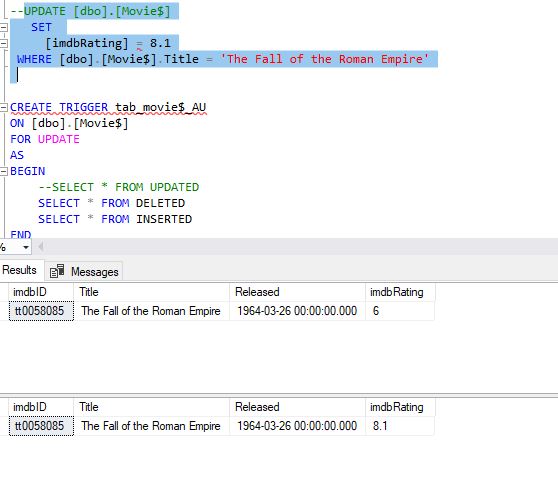




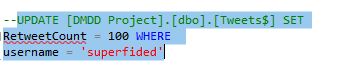
**Triggers**

* After Update (\_AU) Trigger on Movie Table

****

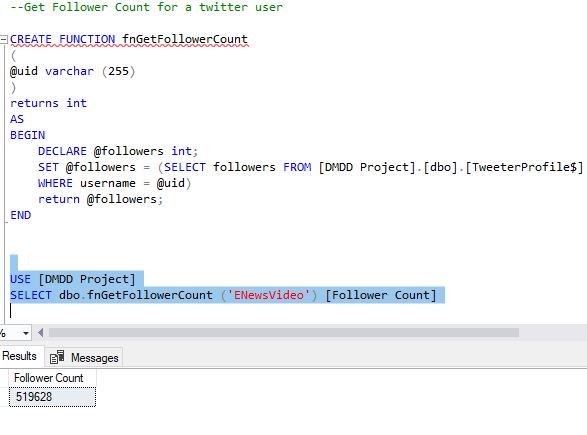
* ****

****

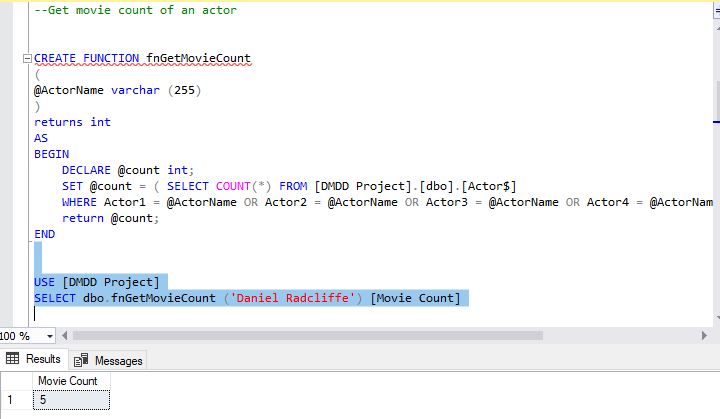
****

**Functions**

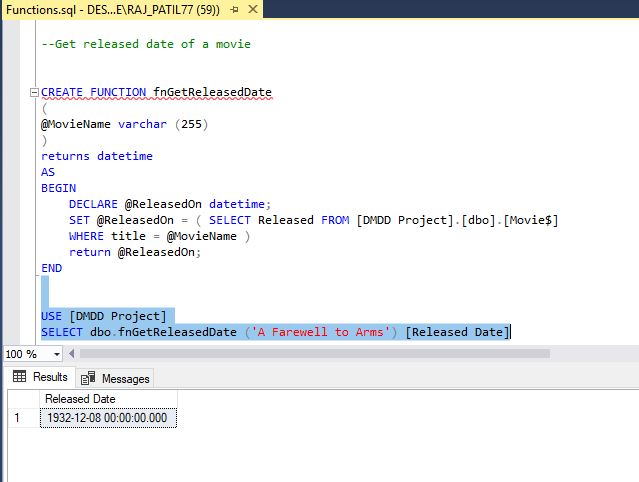
* Get followers count of a twitter user

****

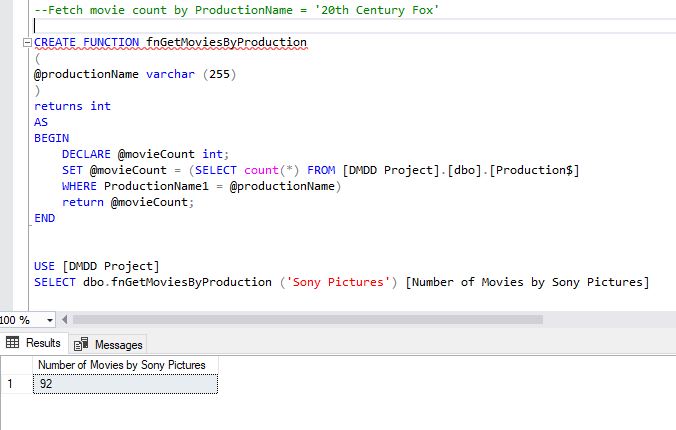
* Get movies count of an actor



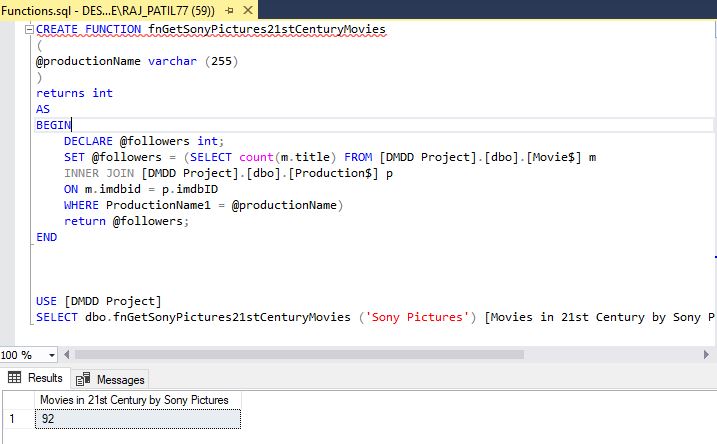
* Get movie released date



* Get movies under particular production



* Get movies by Sony Pictures in 21st Century



**Project Outcome**

This project is the one destination solution for moviegoers. They can obtain all the necessary information on movies like movie released date, casts, production house, director, trending tweets related to movie, etc.