Project description phase2 :

Name :

Menatallh Mohamed

Team member :

Catay senturca

Movies Database : (Project aim)

* The data analysis over 5000 movies from over covering about 66 countries , spotlighting the movie name followed by its genres , no. of votes and criticisms and it’s total IMDB score .
* It mentions some details considering the names of the director and the main actor/actress the no. of their Facebook accounts likes , the year the movie was released and so on either it’s colored or not and the aspect ratio depending on the cinemas through times and the releasing location and language of each .
* The financial sides are as well covered both the gross and budget and so the total profit can be calculated after taking the inflection factors and exchange rate over years into consideration .

the URL of the Google drive file containing the data :

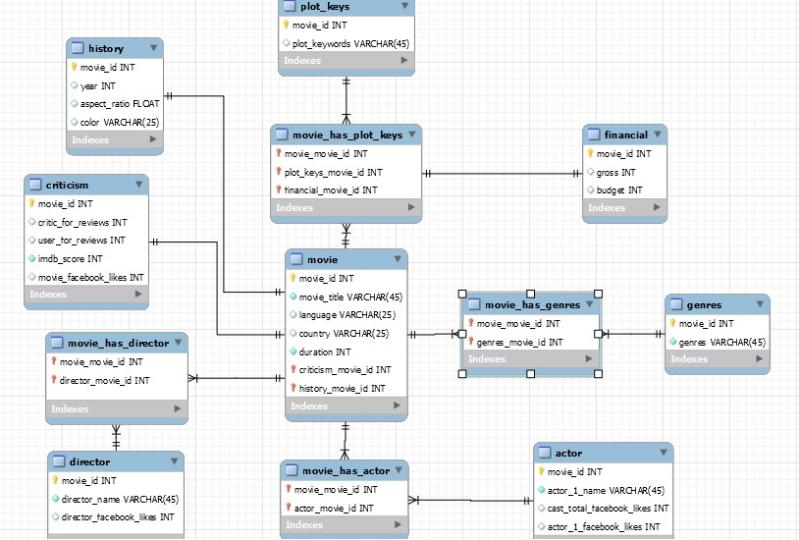
<https://drive.google.com/file/d/0B_QgieOOD8vLZlBueXkxWkRmRHc/view>

List of the ten questions to be answered :

1. List the details of the movie with the highest IMDB score ?
2. Which year got the top no. of released movies(The most frequent appeared year)?
3. What is the average of durations of movies and list those who has a duration over 150 mins and order them ?
4. What is the percentage of movies released outside the U.S in comparison to the total no. of movies ?
5. Which is more the movies released on English or all the other languages ?
6. Who has the higher Facebook likes is it the director or main actor/actress and indicate the movies done by the both (actor/actress,director) with the highest facebook likes ?
7. Which movie is considered the “ most successful ” related to its total Box-office profit , comparing the gross and budget of each is it the same movie with the highest IMDB score ?
8. Comparing the oldest and newest movie of the data how does the aspect ratio changes ?
9. Which is said to be more popular Action or comedy taking the average of the appearance of each at the genres ?
10. Show the num of movies made by Christopher Nolan between 2000 and 2016 and group by imdb\_score ?

Description of the database platform and machine specification:

* MySQL Work branch 6.3 CE through mysql-installer-community-5.6.10.1 taking 172 MG
* Machine Specification :
* O.S name : Microsoft Windows 10 Home Single Language
* System Model : HP 14 Notebook PC
* System Type : x64-based processor , 64-bit based O.S
* Installed Physical Memory(RAM) : 8.00 GB
* The data is to be uploaded via inserting it from excel sheets each includes the entities of each table on its own taken from :
* <https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>
* The primary key of each table is the first value which is for all of an auto incremented ID for each movie , other tables are those which are result of n-n relationship between tables



**MYSQL answers to the 10 previously listed questions:**

**1- List the details of the movie with the highest IMDB score ?**

select movie\_title, year, genres ,duration,actor\_1\_name,imdb\_score

from movie,history,genres,actor,criticism

where imdb\_score = max(imdb\_score);

//OR

SELECT \*

FROM criticism

WHERE imdb\_score = (SELECT MAX(imdb\_score) FROM criticism) as TOP\_IMDB

select movie\_title, year,language ,duration,actor\_1\_name,imdb\_score

from movie,history,actor,criticism

where imdb\_score = TOP\_IMDB;

**2- Which year got the top no. of released movies(The most frequent appeared year)?**

select year

from history

group by year having max(count(\*) desc);

//OR Prints the ranking of each considering equally appearance years as one

select year

from(select year,

count(\*) as year\_count,

rank() over (order by count(\*) desc) as rank

from history

group by year)

where

rank = 1

**3- What is the average of durations of movies and list those who has a duration over 150 mins and order them ?**

select movie\_title, avg (duration),year

from movies  
,history

having avg (duration) > 150

group by year desc;

**4- What is the percentage of movies released out side the U.S in comparison to the total no. of movies ?**

SELECT (count(movie\_title) FROM movie as total\_movies\_count,

count upper(country) FROM movie as total\_usa\_count)

where country = "USA"

SET percentage := (total\_usa\_count/total\_movies\_count) \* 100;

**5-Which is more the movies releaseg on english or all the other languages ?**

SELECT (count(movie\_title) FROM movie as total\_movies\_count,

count upper(language) FROM movie as total\_english\_count)

where language = ("English",

SET difference := (total\_movies\_count-total\_english\_count),

SET most\_common := max(total\_english\_count,difference)

**6- Who has the higher Facebook likes is it the director or main actor/actress and indicate the movies done by the both (actor/actress,director)**

**with the highest facebook likes ?**

select max(avg (director\_facebook\_likes) as acting, avg (actor\_1\_facebook\_likes) as directing)

from actor,director

SELECT all movie\_title

FROM movie ,director,actor

WHERE (SELECT (MAX(actor\_1\_facebook\_likes) FROM actor as TOP\_actor),

(MAX(director\_facebook\_likes) FROM director as TOP\_director)

actor\_1\_name = TOP\_actor or director = TOP\_director

**7- Which movie is considered the “ most successful ” related to it’s total Box-office profit , comparing the gross and budget of each is it**

**the same movie with the highest IMDB score ?**

SELECT movie\_title ,

max(gross-budget) as TOP\_financial ,

MAX((imdb\_score) FROM criticism) as TOP\_IMDB

select case when

(TOP\_IMDB = TOP\_financial)

then 'true'

else 'false'

end as bool

from financial , movie

**8- Comparing the oldest and newest movie of the data how does the aspect ratio changes ?**

SELECT aspect\_artio ,year

FROM history

where year = MAX(year) and year = MIN(year)

**9- Which is said to be more popular Action or comedy taking the average of the appearance of each at the genres ?**

update genres

set genres ="0"

where genres like '%Action%'

set genres ="1"

where genres like '%Comedy%'

SELECT genres, COUNT(genres) as genres\_counter

FROM genres

GROUP BY genres

ORDER BY genres\_counter desc

LIMIT 1;

**10- Show the num of movies made by Christopher Nolan between 2000 and 2016 and group by imdb\_score ?**

select count(director), year, imdb\_score

from history, movie, criticism

where director ='Christopher Nolan'

and year between 2000 and 2016

group by imdb\_score ;