

# PROJECT



# MOVIES

Analyze Ad-Hoc Business Questions

Made By :

Pranav Dharpure



GitHub Link



Linkedin Profile



Email Id

## Table Of Contents

Project Goal .....	2
Language, libraries, tools .....	2
Data .....	2

### **Ad-Hoc Business Questions**

How much did movies cost per minute? .....	3
What is the top-5 movies in terms of budget? .....	3
In terms of release date, how old is every single movie? .....	3
In which years did the producer make a movie? .....	4
Which movies cost less than 50,000 and what were their genres .....	4
Which director did make the most popular movie in English and French .....	4
Make a list of actors (actress) who played after the year 2010 .....	5
How many horror movies did budget exceed the average budget of all movies .....	5
What is the bottom-10 popular movies, and which company made them .....	5
Which companies invested budget between 150K to 200K and for which movie .....	6

# Data Analysis Project: MOVIES

## 1. Project Goals

In this project, I worked on a database in MySQL environment to analyze the data and answer some business questions. For doing this, I work with several tables and run SQL queries and functions to find the best answer for each question.

## 2. Language, libraries, tools:

Language: SQL

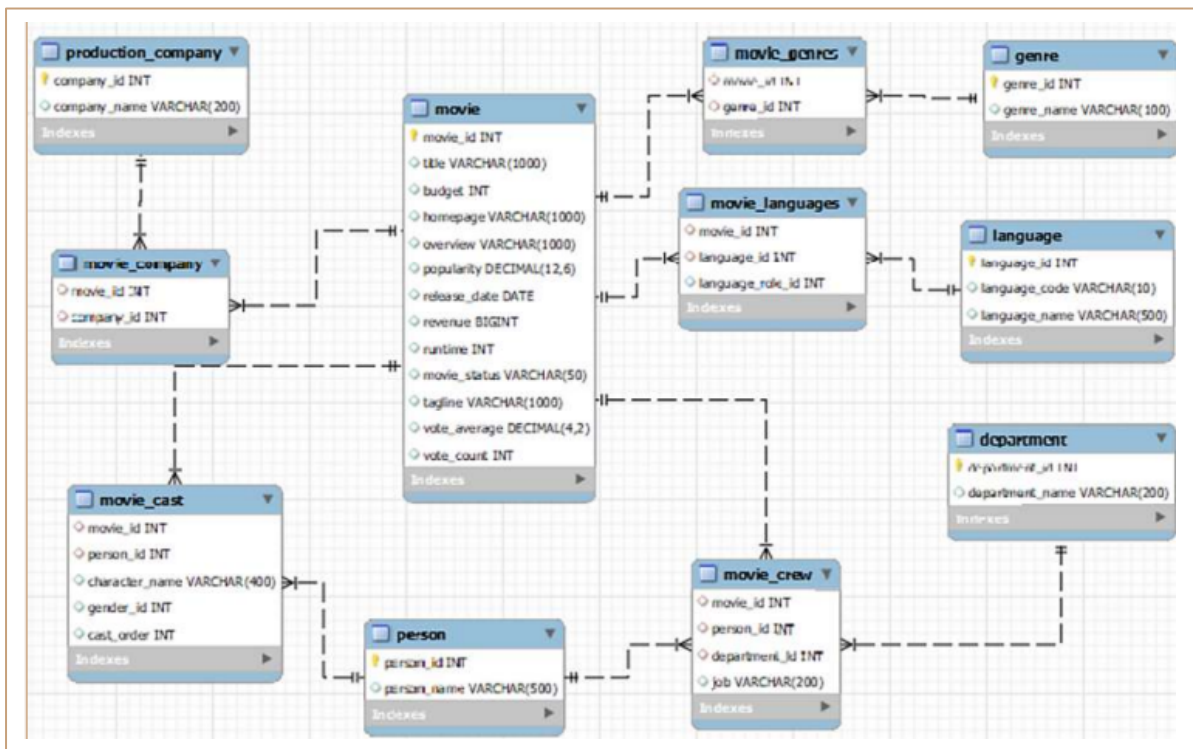
Libraries: -

IDE: MySQL

Application: Microsoft Excel

## 3. Data

There are 17 tables in this database that in the below you can see their attribute type and ER model (entity relationship).





# Ad-Hoc Business Questions

# Data Analysis Project: MOVIES

In this section we want to analyze data in MySQL environment using SQL queries. So, by doing this we can answer some ad-hoc questions that might be asked in daily-basis business.

## 1. How much did movies cost per minute?

```
SELECT title, budget, runtime, round(budget/runtime, 0) as "Budget_Per_Minute"
FROM movie;
```

title	budget	runtime	Budget_Per_Minute
Four Rooms	4000000	98	40816
Star Wars	11000000	121	90909
Finding Nemo	94000000	100	940000
Forrest Gump	55000000	142	387324
American Beauty	15000000	122	122951
Dancer in the Dark	12800000	140	91429
The Fifth Element	90000000	126	714286
Metropolis	92620000	153	605359
My Life Without Me	0	106	0
Pirates of the Caribbean: The Curse of the Blac...	140000000	143	979021

## 2. What is the top-5 movies in terms of budget?

```
select title, budget
from movie
order by budget desc
limit 5;
```

title	budget
Pirates of the Caribbean: On Stranger Tides	380000000
Pirates of the Caribbean: At World's End	300000000
Avengers: Age of Ultron	280000000
Superman Returns	270000000
Tangled	260000000

## 3. In terms of release date, how old is every single movie? Show the top-10 Youngest movies.

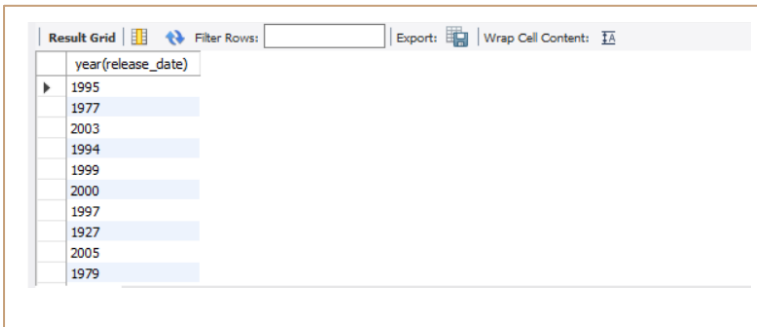
```
select release_date, title, (year(curdate()) - year(release_date)) as age
from movie
order by age asc
limit 10;
```

release_date	title	age
2017-02-03	Growing Up Smith	6
2016-02-04	Pride and Prejudice and Zombies	7
2016-07-14	Ghostbusters	7
2016-02-19	Triple 9	7
2016-06-22	Independence Day: Resurgence	7
2016-08-26	Hands of Stone	7
2016-05-25	Warcraft	7
2016-07-07	Star Trek Beyond	7
2016-05-11	The Angry Birds Movie	7
2016-01-23	Kung Fu Panda 3	7

# Data Analysis Project: MOVIES

4. In which years did the producer make a movie? regardless of the number of movies.

```
select distinct year(release_date)
from movie;
```

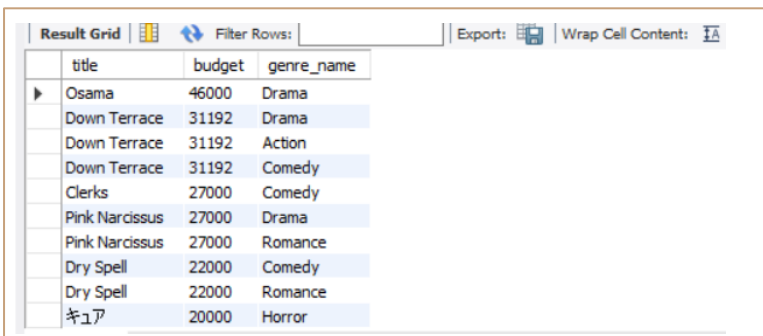


The screenshot shows a database interface with a 'Result Grid' tab. The grid displays the results of a SQL query. The first column is labeled 'year(release\_date)'. The data is as follows:

year(release_date)
1995
1977
2003
1994
1999
2000
1997
1927
2005
1979

5. Which movies cost less than 50,000 and what were their genres? Just top-10

```
select movie.title, movie.budget, genre.genre_name
from movie
join movie_genres on movie.movie_id = movie_genres.movie_id
join genre on genre.genre_id = movie_genres.genre_id
where movie.budget < 50000
order by movie.budget desc
limit 10;
```

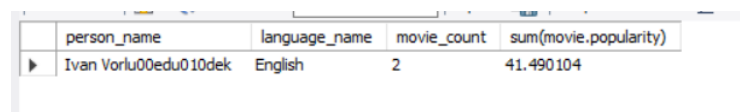


The screenshot shows a database interface with a 'Result Grid' tab. The grid displays the results of a SQL query. The columns are 'title', 'budget', and 'genre\_name'. The data is as follows:

title	budget	genre_name
Osama	46000	Drama
Down Terrace	31192	Drama
Down Terrace	31192	Action
Down Terrace	31192	Comedy
Clerks	27000	Comedy
Pink Narcissus	27000	Drama
Pink Narcissus	27000	Romance
Dry Spell	22000	Comedy
Dry Spell	22000	Romance
キューア	20000	Horror

6. Which director did make the most popular movie in English and French language?

```
select person.person_name, language.language_name, count(*) as movie_count, sum(movie.popularity)
from movie
join movie_crew on movie.movie_id = movie_crew.movie_id
join person on movie_crew.person_id = person.person_id
join movie_languages on movie.movie_id = movie_languages.movie_id
join language on language.language_id = movie_languages.language_id
group by language.language_name, person.person_name
having language.language_name in ('English', 'French')
limit 1;
```



The screenshot shows a database interface with a 'Result Grid' tab. The grid displays the results of a SQL query. The columns are 'person\_name', 'language\_name', 'movie\_count', and 'sum(movie.popularity)'. The data is as follows:

person_name	language_name	movie_count	sum(movie.popularity)
Ivan Vorlu00edu010dek	English	2	41.490104

# Data Analysis Project: MOVIES

## 7. Make a list of actors (actress) who played after the year 2010.

```
select person.person_name, movie.title, movie_crew.job
from movie
join movie_crew on movie.movie_id = movie_crew.movie_id
join person on movie_crew.person_id = person.person_id
where movie_crew.job = 'Characters' and movie.release_date > '2010-01-01'
group by person.person_name, movie.title, movie_crew.job;
```

person_name	title	job
Stan Lee	The Amazing Spider-Man	Characters
Steve Ditko	The Amazing Spider-Man	Characters
Marlon Wayans	Scary Movie 5	Characters
Shawn Wayans	Scary Movie 5	Characters
Buddy Johnson	Scary Movie 5	Characters
Phil Beaman	Scary Movie 5	Characters
Jason Friedberg	Scary Movie 5	Characters
Aaron Seltzer	Scary Movie 5	Characters
Stan Lee	Iron Man 2	Characters
Jack Kirby	Iron Man 2	Characters

## 8. How many horror movies did their budget exceed the average budget of a

```
select movie.title, movie.budget, movie_genres.genre_id, genre.genre_name
from movie
join movie_genres on movie.movie_id = movie_genres.movie_id
join genre on genre.genre_id = movie_genres.genre_id
where genre.genre_name = 'Horror' and movie.budget > (select avg(movie.budget) from movie)
```

title	budget	genre_id	genre_name
Snakes on a Plane	33000000	27	Horror
Aliens vs Predator: Requiem	40000000	27	Horror
Silent Hill	50000000	27	Horror
The Ninth Gate	38000000	27	Horror
Interview with the Vampire	60000000	27	Horror
From Hell	35000000	27	Horror
Gremlins 2: The New Batch	50000000	27	Horror
Ghost Rider	110000000	27	Horror
Resident Evil	33000000	27	Horror
Resident Evil: Apocalypse	45000000	27	Horror

## 9. What is the bottom-10 popular movies, and which company made them?

```
select movie.title, sum(movie.popularity) as popularity, production_company.company_name
from movie
join movie_company on movie.movie_id = movie_company.movie_id
join production_company on production_company.company_id = movie_company.company_id
group by movie.title, production_company.company_name
order by sum(movie.popularity) asc
limit 10;
```

title	popularity	company_name
Love Letters	0.001586	Millennium Films
Love Letters	0.001586	New World Pictures
Arnolds Park	0.001586	Sting Entertainment
Arnolds Park	0.006069	The Picture Factory
Anderson's Cross	0.006323	Illumination Pictures
Le bonheur d'Elza	0.007254	Canal Plus Overseas
Le bonheur d'Elza	0.007254	Aztec Musique
Le bonheur d'Elza	0.007254	France Tu00e9l00e9vision
Four Single Fathers	0.008263	Indiana Production Company
Nothing But a Man	0.008453	DuArt Film and Video

## Data Analysis Project: MOVIES

10. Which movie companies invested the budget between 150K to 200K and for which movie?

```
select movie.title, movie.budget, production_company.company_name
from movie
join movie_company on movie.movie_id = movie_company.movie_id
join production_company on movie_company.company_id = production_company.company_id
where movie.budget between 15000 and 200000
order by budget desc
limit 1;
```

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
title	budget	company_name	
Swingers	200000	Miramax Films	

- END -