

# IMDB Movie Analysis

## Project Description:

We are provided with dataset having various columns of different IMDB movies. Our goal is to clean and analyse the data to provide meaningful insights by answering set of questions such as finding the movies with highest profit, movies with highest IMDB ratings, finding best directors, popular actors and genres.

## Approach:

- We explore the dataset, to get a better understanding of the dataset.
- Once we are familiar with the dataset, we clean it in a way that is suitable for further analysis.
- Then we continue with our analysis by answering a set of questions which helps us identify trends in the film industry.

## Tech-Stack Used:

- Data Cleaning - MS Excel
- Analysis - MySQL workbench
- Charts and Visualization - Power BI

## Insights:

- A. **Cleaning the data:** This is one of the most important steps to perform before moving forward with the analysis. Use your knowledge learned till now to do this. (Dropping columns, removing null values, etc.)

**Your task:** Clean the data

## Steps taken to clean data:

- Remove duplicate values.
- Check for null values and remove the rows with null values.
- Identify the columns required for analysis and remove the unnecessary columns.
- Arrange the columns in a way which will be easy for us to continue with the analysis.
- **Before Cleaning:** 5043 rows and 28 columns
- **After Cleaning:** 3722 rows and 13 columns

| num_voted_users |                   |                       |          |                         |                        |                      |                        |           |   |
|-----------------|-------------------|-----------------------|----------|-------------------------|------------------------|----------------------|------------------------|-----------|---|
| color           | director_name     | num_critic_for_review | duration | director_facebook_likes | actor_3_facebook_likes | actor_2_name         | actor_1_facebook_likes | gross     | genres                                    |
| Color           | James Cameron     | 723                   | 178      | 0                       | 855                    | Joel David Moore     | 1000                   | 760505847 | Action Adventure Fantasy Sci-Fi           |
| Color           | Gore Verbinski    | 302                   | 169      | 563                     | 1000                   | Orlando Bloom        | 40000                  | 309404152 | Action Adventure Fantasy                  |
| Color           | Sam Mendes        | 602                   | 148      | 0                       | 161                    | Rory Kinnear         | 11000                  | 200074175 | Action Adventure Thriller                 |
| Color           | Christopher Nolan | 813                   | 164      | 22000                   | 23000                  | Christian Bale       | 27000                  | 448130642 | Action Thriller                           |
|                 | Doug Walker       |                       |          | 131                     |                        | Rob Walker           | 131                    |           | Documentary                               |
| Color           | Andrew Stanton    | 462                   | 132      | 475                     | 530                    | Samantha Morton      | 640                    | 73058679  | Action Adventure Sci-Fi                   |
| Color           | Sam Raimi         | 392                   | 156      | 0                       | 4000                   | James Franco         | 24000                  | 336530303 | Action Adventure Romance                  |
| Color           | Nathan Greno      | 324                   | 100      | 15                      | 284                    | Donna Murphy         | 799                    | 200807262 | Adventure Animation Comedy Family Fantasy |
| Color           | Joss Whedon       | 635                   | 141      | 0                       | 19000                  | Robert Downey Jr.    | 26000                  | 458991599 | Action Adventure Sci-Fi                   |
| Color           | David Yates       | 375                   | 153      | 282                     | 10000                  | Daniel Radcliffe     | 25000                  | 301956980 | Adventure Family Fantasy Mystery          |
| Color           | Zack Snyder       | 673                   | 183      | 0                       | 2000                   | Lauren Cohan         | 15000                  | 330249052 | Action Adventure Sci-Fi                   |
| Color           | Bryan Singer      | 434                   | 169      | 0                       | 903                    | Marlon Brando        | 18000                  | 200069408 | Action Adventure Sci-Fi                   |
| Color           | Marc Forster      | 403                   | 106      | 395                     | 393                    | Mathieu Amalric      | 451                    | 168368427 | Action Adventure                          |
| Color           | Gore Verbinski    | 313                   | 151      | 563                     | 1000                   | Orlando Bloom        | 40000                  | 423032628 | Action Adventure Fantasy                  |
| Color           | Gore Verbinski    | 450                   | 150      | 563                     | 1000                   | Ruth Wilson          | 40000                  | 89289910  | Action Adventure Western                  |
| Color           | Zack Snyder       | 733                   | 143      | 0                       | 748                    | Christopher Meloni   | 15000                  | 291021565 | Action Adventure Sci-Fi                   |
| Color           | Andrew Adamson    | 258                   | 150      | 80                      | 201                    | Pierfrancesco Favino | 22000                  | 141614023 | Action Adventure Family Fantasy           |
| Color           | Joss Whedon       | 703                   | 173      | 0                       | 19000                  | Robert Downey Jr.    | 26000                  | 623279547 | Action Adventure Sci-Fi                   |
| Color           | Rob Marshall      | 448                   | 136      | 252                     | 1000                   | Sam Claflin          | 40000                  | 241063875 | Action Adventure Fantasy                  |
| Color           | Barry Sonnenfeld  | 451                   | 106      | 188                     | 718                    | Michael Stuhlbarg    | 10000                  | 179020854 | Action Adventure Comedy Family Fantasy    |
| Color           | Peter Jackson     | 422                   | 164      | 0                       | 773                    | Adam Brown           | 5000                   | 255108370 | Adventure Fantasy                         |
| Color           | Marc Webb         | 599                   | 153      | 464                     | 963                    | Andrew Garfield      | 15000                  | 262030663 | Action Adventure Fantasy                  |
| Color           | Ridley Scott      | 343                   | 156      | 0                       | 738                    | William Hurt         | 891                    | 105219735 | Action Adventure Drama History            |
| Color           | Peter Jackson     | 509                   | 186      | 0                       | 773                    | Adam Brown           | 5000                   | 258353554 | Adventure Fantasy                         |
| Color           | Chris Weitz       | 251                   | 113      | 129                     | 1000                   | Eva Green            | 16000                  | 70083519  | Adventure Family Fantasy                  |

| movie_title                                 | genres  | director_name     | actor_1_name    | language | num_critic_for_reviews | num_voted_users |
|---|---|-------------------|-----------------|----------|------------------------|-----------------|
| Avatar                                      | Action Adventure Fantasy Sci-Fi               | James Cameron     | CCH Pounder     | English  | 723                    | 886204          |
| Pirates of the Caribbean: At World's End    | Action Adventure Fantasy                      | Gore Verbinski    | Johnny Depp     | English  | 302                    | 471220          |
| Spectre                                     | Action Adventure Thriller                     | Sam Mendes        | Christoph Waltz | English  | 602                    | 275868          |
| The Dark Knight Rises                       | Action Thriller                               | Christopher Nolan | Tom Hardy       | English  | 813                    | 1144337         |
| Star Wars: Episode VII - The Force Awakens  | Documentary                                   | Doug Walker       | Doug Walker     |          |                        | 8               |
| John Carter                                 | Action Adventure Sci-Fi                       | Andrew Stanton    | Daryl Sabara    | English  | 462                    | 212204          |
| Spider-Man 3                                | Action Adventure                              |                   |                 | English  | 392                    | 383056          |
| Tangled                                     | Adventure Fantasy                             |                   |                 | English  | 324                    | 294810          |
| Avengers: Age of Ultron                     | Action Adventure                              |                   |                 | English  | 635                    | 462669          |
| Harry Potter and the Half-Blood Prince      | Adventure Fantasy                             |                   |                 | English  | 375                    | 321795          |
| Batman v Superman: Dawn of Justice          | Action Adventure                              |                   |                 | English  | 673                    | 371639          |
| Superman Returns                            | Action Adventure                              |                   |                 | English  | 434                    | 240396          |
| Quantum of Solace                           | Action Adventure                              |                   |                 | English  | 403                    | 330784          |
| Pirates of the Caribbean: Dead Man's Chest  | Action Adventure Fantasy                      |                   |                 | English  | 313                    | 522040          |
| The Lone Ranger                             | Action Adventure Western                      |                   |                 | English  | 450                    | 181792          |
| Man of Steel                                | Action Adventure Fantasy Sci-Fi               |                   |                 | English  | 733                    | 548573          |
| The Chronicles of Narnia: Prince Caspian    | Action Adventure Family Fantasy               |                   |                 | English  | 258                    | 149922          |
| The Avengers                                | Action Adventure Sci-Fi                       |                   |                 | English  | 703                    | 995415          |
| Pirates of the Caribbean: On Stranger Tides | Action Adventure Fantasy                      |                   |                 | English  | 448                    | 370704          |
| Men in Black 3                              | Action Adventure Comedy Family Fantasy Sci-Fi |                   |                 | English  | 451                    | 268154          |
| The Hobbit: The Battle of the Five Armies   | Adventure Fantasy                             |                   |                 | English  | 422                    | 354228          |
| The Amazing Spider-Man                      | Action Adventure Fantasy                      |                   |                 | English  | 599                    | 451803          |
| Robin Hood                                  | Action Adventure Drama History                |                   |                 | English  | 343                    | 211765          |
| The Hobbit: The Desolation of Smaug         | Adventure Fantasy                             |                   |                 | English  | 509                    | 483540          |
| The Golden Compass                          | Adventure Family Fantasy                      |                   |                 | English  | 251                    | 149019          |
| King Kong                                   | Action Adventure Drama Romance                |                   |                 | English  | 446                    | 316018          |
| Titanic                                     | Drama Romance                                 |                   |                 | English  | 315                    | 793059          |

AutoSave OFF

imdb

Search

Fathima Rizwana

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

Undo

CutCopyFormat Painter

Clipboard

Font

Alignment

Wrap TextMerge & Center

Number

Conditional Formatting

Format as Table

Cell Styles

InsertDeleteFormat

FillClear

Sort & Filter

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

Don't show againSave As...

A1

movie\_title

|    | A   | B   | C                 | D                  | E        | F      | G        | H     | I        | J              | K     | L         | M         |
|----|---|---|-------------------|--------------------|----------|--------|----------|-------|----------|----------------|-------|-----------|-----------|
|    | movie_title                                 | genres  | director_name     | actor_1_name       | language | num_ci | num_vote | num_u | movie_fa | imdb_title_yes | gross | budget    |           |
| 1  | Avatar                                      | Action Adventure Fantasy Sci-Fi                           | James Cameron     | CCH Pounder        | English  | 723    | 886204   | 3054  | 33000    | 7.9            | 2009  | 760505847 | 237000000 |
| 2  | Pirates of the Caribbean: At World's End    | Action Adventure Fantasy                                  | Gore Verbinski    | Johnny Depp        | English  | 302    | 471220   | 1238  | 0        | 7.1            | 2007  | 309404152 | 300000000 |
| 3  | Spectre                                     | Action Adventure Thriller                                 | Sam Mendes        | Christoph Waltz    | English  | 602    | 275868   | 994   | 85000    | 6.8            | 2015  | 200074175 | 245000000 |
| 4  | The Dark Knight Rises                       | Action Thriller   | Christopher Nolan | Tom Hardy          | English  | 813    | 1144337  | 2701  | 164000   | 8.5            | 2012  | 448130642 | 250000000 |
| 5  | John Carter                                 | Action Adventure Sci-Fi                                   | Andrew Stanton    | Daryl Sabara       | English  | 462    | 212204   | 738   | 24000    | 6.6            | 2012  | 73058679  | 263700000 |
| 6  | Spider-Man 3                                | Action Adventure Romance                                  | Sam Raimi         | J.K. Simmons       | English  | 392    | 383056   | 1902  | 0        | 6.2            | 2007  | 336530303 | 258000000 |
| 7  | Tangled                                     | Adventure Animation Comedy Family Fantasy Musical Romance | Brad Garrett      | Nathan Greno       | English  | 324    | 294810   | 387   | 29000    | 7.8            | 2010  | 200807262 | 260000000 |
| 8  | Avengers: Age of Ultron                     | Action Adventure Sci-Fi                                   | Joss Whedon       | Chris Hemsworth    | English  | 635    | 462669   | 1117  | 118000   | 7.5            | 2015  | 458991599 | 250000000 |
| 9  | Harry Potter and the Half-Blood Prince      | Adventure Family Fantasy Mystery                          | David Yates       | Alan Rickman       | English  | 375    | 321795   | 973   | 10000    | 7.5            | 2009  | 301956980 | 250000000 |
| 10 | Batman v Superman: Dawn of Justice          | Action Adventure Sci-Fi                                   | Zack Snyder       | Henry Cavill       | English  | 673    | 371639   | 3018  | 197000   | 6.9            | 2016  | 330249062 | 250000000 |
| 11 | Superman Returns                            | Action Adventure Sci-Fi                                   | Bryan Singer      | Kevin Spacey       | English  | 434    | 240396   | 2367  | 0        | 6.1            | 2006  | 200069408 | 209000000 |
| 12 | Quantum of Solace                           | Action Adventure  | Marc Forster      | Giancarlo Giannini | English  | 403    | 330784   | 1243  | 0        | 6.7            | 2008  | 168368427 | 200000000 |
| 13 | Pirates of the Caribbean: Dead Man's Chest  | Action Adventure Fantasy                                  | Gore Verbinski    | Johnny Depp        | English  | 313    | 522040   | 1832  | 5000     | 7.3            | 2006  | 423032628 | 225000000 |
| 14 | The Lone Ranger                             | Action Adventure Western                                  | Gore Verbinski    | Johnny Depp        | English  | 450    | 181792   | 711   | 48000    | 6.5            | 2013  | 89289910  | 215000000 |
| 15 | Man of Steel                                | Action Adventure Fantasy Sci-Fi                           | Zack Snyder       | Henry Cavill       | English  | 733    | 548573   | 2536  | 118000   | 7.2            | 2013  | 291021565 | 225000000 |
| 16 | The Chronicles of Narnia: Prince Caspian    | Action Adventure Family Fantasy                           | Andrew Adamson    | Peter Dinklage     | English  | 258    | 149922   | 438   | 0        | 6.6            | 2008  | 141614023 | 225000000 |
| 17 | The Avengers                                | Action Adventure Sci-Fi                                   | Joss Whedon       | Chris Hemsworth    | English  | 703    | 995415   | 1722  | 123000   | 8.1            | 2012  | 623279547 | 220000000 |
| 18 | Pirates of the Caribbean: On Stranger Tides | Action Adventure Fantasy                                  | Rob Marshall      | Johnny Depp        | English  | 448    | 370704   | 484   | 58000    | 6.7            | 2011  | 241063875 | 250000000 |
| 19 | Men in Black 3                              | Action Adventure Comedy Family Fantasy Sci-Fi             | Barry Sonnenfeld  | Will Smith         | English  | 451    | 268154   | 341   | 40000    | 6.8            | 2012  | 179020854 | 225000000 |
| 20 | The Hobbit: The Battle of the Five Armies   | Adventure Fantasy   | Peter Jackson     | Aidan Turner       | English  | 422    | 354228   | 802   | 65000    | 7.5            | 2014  | 255108370 | 250000000 |
| 21 | The Amazing Spider-Man                      | Action Adventure Fantasy                                  | Marc Webb         | Emma Stone         | English  | 599    | 451803   | 1225  | 56000    | 7              | 2012  | 262030663 | 230000000 |
| 22 | Robin Hood                                  | Action Adventure Drama History                            | Ridley Scott      | Mark Addy          | English  | 343    | 211765   | 546   | 17000    | 6.7            | 2010  | 105219735 | 200000000 |
| 23 | The Hobbit: The Desolation of Smaug         | Adventure Fantasy   | Peter Jackson     | Aidan Turner       | English  | 509    | 483540   | 951   | 83000    | 7.9            | 2013  | 258355354 | 225000000 |
| 24 | The Golden Compass                          | Adventure Family Fantasy                                  | Chris Weitz       | Christopher Lee    | English  | 251    | 149019   | 666   | 0        | 6.1            | 2007  | 70083519  | 180000000 |
| 25 | King Kong                                   | Action Adventure Drama Romance                            | Peter Jackson     | Naomi Watts        | English  | 446    | 316018   | 2618  | 0        | 7.2            | 2005  | 218051260 | 207000000 |

imdb

Ready

Accessibility: Unavailable

- B. **Movies with highest profit:** Create a new column called profit which contains the difference of the two columns: gross and budget. Sort the column using the profit column as reference. Plot profit (y-axis) vs budget (x-axis) and observe the outliers using the appropriate chart type.
- Your task:** Find the movies with the highest profit?

We find the profit by calculating the difference between gross and budget, and find the movie with the highest profit by sorting the data with respect to the profit column in descending order.

The movie with the highest profit is Avatar.

AutoSaveOFFimdb

Search

Fathima Rizwana

CommentsShare

FileHomeInsertPage LayoutFormulasDataReviewViewHelp

UndoPasteCopyFormat PainterClipboard

Calibri11Font

BBIUAlign Center

Wrap TextMerge & Center

GeneralConditional FormattingTable as ListCell StylesInsertDeleteFormatCellsEditing

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

Don't show againSave As...

A2

AvatarA

|    | A  | B   | C                 | D                   | E        | F      | G        | H     | I        | J              | K     | L         | M         | N         |
|----|--|---|-------------------|---------------------|----------|--------|----------|-------|----------|----------------|-------|-----------|-----------|-----------|
|    | movie_title                                    | genres  | director_name     | actor_1_name        | language | num_ci | num_vote | num_u | movie_fs | imdb_title_ave | gross | budget    | profit    |           |
| 1  | Avatar   | Action Adventure Fantasy Sci-Fi                   | James Cameron     | CCH Pounder         | English  | 723    | 886204   | 3054  | 33000    | 7.9            | 2009  | 760505847 | 237000000 | 523505847 |
| 2  | Jurassic World                                 | Action Adventure Sci-Fi Thriller                  | Colin Trevorrow   | Bryce Dallas Howari | English  | 644    | 418214   | 1290  | 150000   | 7              | 2015  | 652177271 | 150000000 | 502177271 |
| 3  | Titanic  | Drama Romance                                     | James Cameron     | Leonardo DiCaprio   | English  | 315    | 793059   | 2528  | 26000    | 7.7            | 1997  | 658672302 | 200000000 | 458672302 |
| 4  | Star Wars: Episode IV - A New Hope             | Action Adventure Fantasy Sci-Fi                   | George Lucas      | Harrison Ford       | English  | 282    | 911097   | 1470  | 33000    | 8.7            | 1977  | 460935665 | 110000000 | 449935665 |
| 5  | E.T. the Extra-Terrestrial                     | Family Sci-Fi                                     | Steven Spielberg  | Henry Thomas        | English  | 215    | 281842   | 515   | 34000    | 7.9            | 1982  | 434949459 | 105000000 | 424449459 |
| 6  | The Avengers                                   | Action Adventure Sci-Fi                           | Joss Whedon       | Chris Hemsworth     | English  | 703    | 995415   | 1722  | 123000   | 8.1            | 2012  | 623279547 | 220000000 | 403279547 |
| 7  | The Lion King                                  | Adventure Animation Drama Family Musical          | Roger Allers      | Matthew Broderick   | English  | 186    | 644348   | 656   | 17000    | 8.5            | 1994  | 422783777 | 450000000 | 377783777 |
| 8  | Star Wars: Episode I - The Phantom Menace      | Action Adventure Fantasy Sci-Fi                   | George Lucas      | Natalie Portman     | English  | 320    | 534658   | 3597  | 13000    | 6.5            | 1999  | 474544677 | 115000000 | 359544677 |
| 9  | The Dark Knight                                | Action Crime Drama Thriller                       | Christopher Nolan | Christian Bale      | English  | 645    | 1676169  | 4667  | 37000    | 9              | 2008  | 533316061 | 185000000 | 348316061 |
| 10 | The Hunger Games                               | Adventure Drama Sci-Fi Thriller                   | Gary Ross         | Jennifer Lawrence   | English  | 673    | 701607   | 1959  | 140000   | 7.3            | 2012  | 407999255 | 780000000 | 329999255 |
| 11 | Deadpool                                       | Action Adventure Comedy Romance Sci-Fi            | Tim Miller        | Ryan Reynolds       | English  | 579    | 479047   | 1058  | 117000   | 8.1            | 2016  | 363024263 | 580000000 | 305024263 |
| 12 | The Hunger Games: Catching Fire                | Adventure Sci-Fi Thriller                         | Francis Lawrence  | Jennifer Lawrence   | English  | 502    | 498397   | 706   | 82000    | 7.6            | 2013  | 424645577 | 130000000 | 294645577 |
| 13 | Jurassic Park                                  | Adventure Sci-Fi Thriller                         | Steven Spielberg  | Wayne Knight        | English  | 308    | 613473   | 895   | 19000    | 8.1            | 1993  | 356784000 | 630000000 | 293784000 |
| 14 | Despicable Me 2                                | Animation Comedy Family Sci-Fi                    | Pierre Coffin     | Steve Carell        | English  | 306    | 286877   | 284   | 56000    | 7.5            | 2013  | 368049635 | 760000000 | 292049635 |
| 15 | American Sniper                                | Action Biography Drama History Thriller War       | Clint Eastwood    | Bradley Cooper      | English  | 490    | 325264   | 916   | 112000   | 7.3            | 2014  | 350123553 | 588000000 | 291323553 |
| 16 | Finding Nemo                                   | Adventure Animation Comedy Family                 | Andrew Stanton    | Alexander Gould     | English  | 301    | 692482   | 866   | 11000    | 8.2            | 2003  | 380838870 | 940000000 | 286838870 |
| 17 | Shrek 2  | Adventure Animation Comedy Family Fantasy Romance | Andrew Adamson    | Rupert Everett      | English  | 205    | 314630   | 483   | 0        | 7.2            | 2004  | 436471036 | 150000000 | 286471036 |
| 18 | The Lord of the Rings: The Return of the King  | Action Adventure Drama Fantasy                    | Peter Jackson     | Orlando Bloom       | English  | 328    | 1215718  | 3189  | 16000    | 8.9            | 2003  | 377019252 | 940000000 | 283019252 |
| 19 | Star Wars: Episode VI - Return of the Jedi     | Action Adventure Fantasy Sci-Fi                   | Richard Marquand  | Harrison Ford       | English  | 197    | 681857   | 647   | 14000    | 8.4            | 1983  | 309125409 | 325000000 | 276625409 |
| 20 | Forrest Gump                                   | Comedy Drama                                      | Robert Zemeckis   | Tom Hanks           | English  | 149    | 1251222  | 1398  | 59000    | 8.8            | 1994  | 329691196 | 550000000 | 274691196 |
| 21 | Star Wars: Episode V - The Empire Strikes Back | Action Adventure Fantasy Sci-Fi                   | Irvin Kershner    | Harrison Ford       | English  | 223    | 837759   | 900   | 17000    | 8.8            | 1980  | 290158751 | 180000000 | 272158751 |
| 22 | Home Alone                                     | Comedy Family                                     | Chris Columbus    | Macaulay Culkin     | English  | 102    | 311075   | 305   | 20000    | 7.5            | 1990  | 285761243 | 180000000 | 267761243 |
| 23 | Star Wars: Episode III - Revenge of the Sith   | Action Adventure Fantasy Sci-Fi                   | George Lucas      | Natalie Portman     | English  | 359    | 520104   | 3286  | 10000    | 7.6            | 2005  | 380262555 | 113000000 | 267262555 |
| 24 | Spider-Man                                     | Action Adventure Fantasy Romance                  | Sam Raimi         | J.K. Simmons        | English  | 291    | 544665   | 2012  | 5000     | 7.3            | 2002  | 403706375 | 139000000 | 264706375 |
| 25 | Minions  | Action Animation Comedy Family Sci-Fi             | Kyle Balda        | Steve Carell        | English  | 308    | 142403   | 275   | 70000    | 6.4            | 2015  | 336029560 | 740000000 | 262029560 |

imdb

100% Accessibility: Unavailable

Average: 180104076.9 | Count: 14 | Sum: 1571036692

Query:

select movie\_title from imdb where profit=(select max(profit) from imdb);

Output:

|   |             |
|---|-------------|
|   | movie_title |
| ▶ | Avatar      |

MySQL80 x

Query Database Server Tools Scripting Help

Query 1 SQL File 3\* SQL File 4\* SQL File 5\* SQL File 7\* SQL File 7\* x

Limit to 1000 rows

```

1 • select count(*) from imdb;
2
3 • select * from imdb;
4
5 • delete from imdb;
6
7 • select movie_title from imdb where profit=(select max(profit) from imdb);
8
9 • select rank() over(order by imdb_score desc) as movie_rank, movie_title as IMdb_Top_250, imdb_score

```

Result Grid

| movie_title |
|-------------|
| Avatar      |

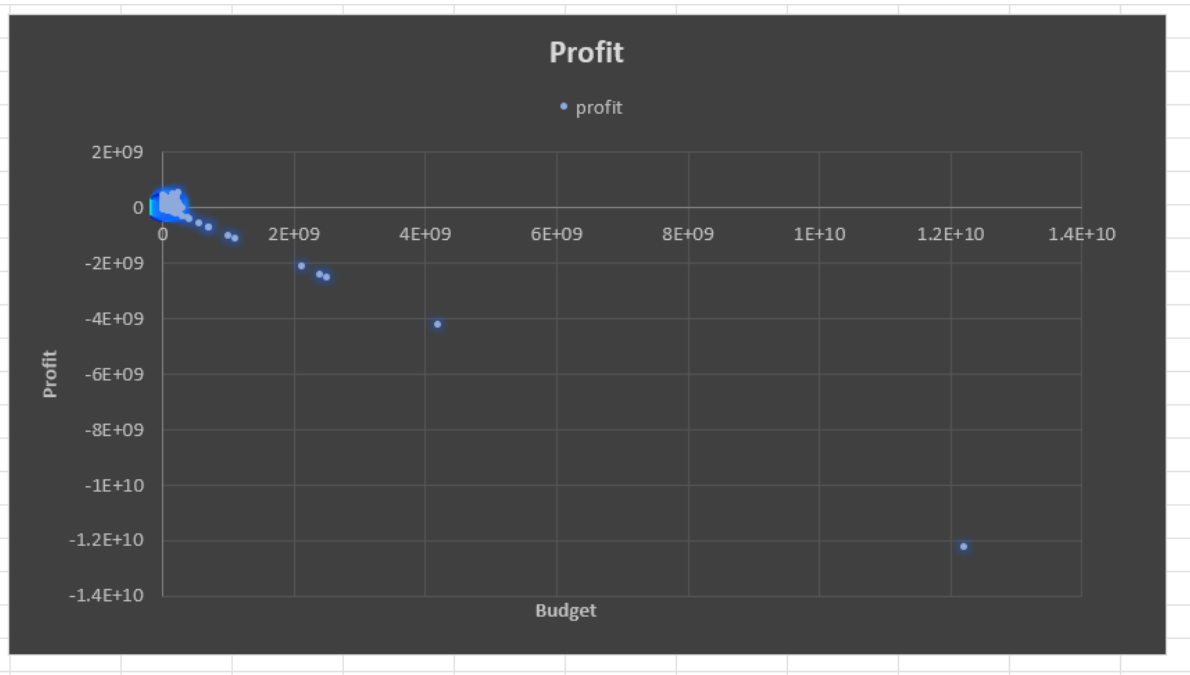
imdb 14 x

Output

Action Output

| #  | Time     | Action  | Message   |
|----|----------|---|---|
| 30 | 20:21:45 | select genres as popular_genres, avg(imdb_score) as highest_imdb_score from imdb_data group by genres or...   | 0 row(s) returned                                     |
| 31 | 20:22:01 | select genres as popular_genres, avg(imdb_score) as highest_imdb_score from imdb group by genres order by ... | 10 row(s) returned                                    |
| 32 | 20:22:59 | select actor_1_name, movie_title, num_critic_for_reviews, num_user_for_reviews from imdb where actor_1_n...   | Error Code: 1064. You have an error in your SQL synta |
| 33 | 20:23:51 | with cte as( select actor_1_name, movie_title, num_critic_for_reviews, num_user_for_reviews from imdb wher... | 3 row(s) returned                                     |
| 34 | 20:24:41 | select floor(title_year/10)*10 as decade, sum(num_voted_users) as df_by_decade from imdb group by decade ...  | 10 row(s) returned                                    |

We find the outliers by using a scatter plot.



- C. **Top 250:** Create a new column `IMDb_Top_250` and store the top 250 movies with the highest IMDb Rating (corresponding to the column: `imdb_score`). Also make sure that for all of these movies, the `num_voted_users` is greater than 25,000. Also add a `Rank` column containing the values 1 to 250 indicating the ranks of the corresponding films.

Extract all the movies in the `IMDb_Top_250` column which are not in the English language and store them in a new column named `Top_Foreign_Lang_Film`. You can use your own imagination also!

**Your task:** Find IMDB Top 250

**Query:**

```
select rank() over(order by imdb_score desc) as movie_rank, movie_title as
IMDb_Top_250, imdb_score
```

```
from imdb where num_voted_users>25000 limit 250;
```

**Output:**

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

Schemas

Filter objects

case\_study\_2  
ig\_clone  
ops  
imdb  
imdb\_data  
Columns  
Indexes  
Foreign Keys  
Triggers  
weather  
Views  
Stored Procedures  
Functions  
Administration  
Schemas

Information

Table: **imdb**

Columns:

movie\_title  
genres  
director\_name  
actor\_1\_name  
language  
num\_critics\_for\_reviews  
num\_voted\_users  
num\_user\_for\_reviews  
movie\_facebook\_likes  
imdb\_score  
title\_year  
gross  
budget  
profit

text  
text  
text  
text  
text  
double  
double  
double  
double  
double  
double  
double  
double

Query 1

SQL File 3\* SQL File 4\* SQL File 5\* SQL File 7\* SQL File 7\*

Limit to 1000 rows

SQL Additions

Automatic context  
the toolbar to man  
current caret po  
autom

```

4
5 delete from imdb;
6
7 select movie_title from imdb where profit=(select max(profit) from imdb);
8
9 select rank() over(order by imdb_score desc) as movie_rank, movie_title as IMDB_Top_250, imdb_score
10 from imdb where num_voted_users>25000 limit 250;
11
12 with cte as(

```

Result Grid

| movie_rank | IMDb_Top_250                                      | imdb_score |
|------------|---|------------|
| 1          | The Shawshank Redemption                          | 9.3        |
| 2          | The Godfather                                     | 9.2        |
| 3          | The Dark Knight                                   | 9          |
| 3          | The Godfather: Part II                            | 9          |
| 5          | The Lord of the Rings: The Return of the King     | 8.9        |
| 5          | Schindler's List                                  | 8.9        |
| 5          | Pulp Fiction                                      | 8.9        |
| 5          | The Good, the Bad and the Ugly                    | 8.9        |
| 9          | Inception   | 8.8        |
| 9          | The Lord of the Rings: The Fellowship of the Ring | 8.8        |

Output

Action Output

| #  | Time     | Action   | Message  |
|----|----------|--|--|
| 31 | 20:22:01 | select genres as popular_genres, avg(imdb_score) as highest_imdb_score from imdb group by genres order by ...  | 10 row(s) returned   |
| 32 | 20:22:59 | select actor_1_name, movie_title, num_critics_for_reviews, num_user_for_reviews from imdb where actor_1_n...   | Error Code: 1064. You have an error in your SQL syntax; check the manual that corresponds to your M... |
| 33 | 20:23:51 | with cte as( select actor_1_name, movie_title, num_critics_for_reviews, num_user_for_reviews from imdb wher... | 3 row(s) returned  |
| 34 | 20:24:41 | select floor(title_year/10)*10 as decade, sum(num_voted_users) as df_by_decade from imdb group by decade ...   | 10 row(s) returned   |
| 35 | 21:22:27 | select movie_title from imdb where profit=(select max(profit) from imdb) LIMIT 0, 1000                         | 1 row(s) returned  |
| 36 | 21:36:00 | select rank() over(order by imdb_score desc) as movie_rank, movie_title as IMDB_Top_250, imdb_score from i...  | 250 row(s) returned  |

Object Info Session

| movie_rank | IMDB_Top_250                                      | imdb_score |
|------------|---|------------|
| 1          | The Shawshank Redemption                          | 9.3        |
| 2          | The Godfather                                     | 9.2        |
| 3          | The Dark Knight                                   | 9          |
| 3          | The Godfather: Part II                            | 9          |
| 5          | The Lord of the Rings: The Return of the King     | 8.9        |
| 5          | Schindler's List                                  | 8.9        |
| 5          | Pulp Fiction                                      | 8.9        |
| 5          | The Good, the Bad and the Ugly                    | 8.9        |
| 9          | Inception   | 8.8        |
| 9          | The Lord of the Rings: The Fellowship of the Ring | 8.8        |
| 9          | Fight Club  | 8.8        |
| 9          | Forrest Gump                                      | 8.8        |
| 9          | Star Wars: Episode V - The Empire Strikes Back    | 8.8        |
| 14         | The Lord of the Rings: The Two Towers             | 8.7        |
| 14         | The Matrix  | 8.7        |
| 14         | Goodfellas  | 8.7        |
| 14         | Star Wars: Episode IV - A New Hope                | 8.7        |
| 14         | One Flew Over the Cuckoo's Nest                   | 8.7        |
| 14         | City of God                                       | 8.7        |
| 14         | Seven Samurai                                     | 8.7        |
| 21         | Interstellar                                      | 8.6        |
| 21         | Saving Private Ryan                               | 8.6        |
| 21         | Se7en   | 8.6        |
| 21         | The Silence of the Lambs                          | 8.6        |
| 21         | Spirited Away                                     | 8.6        |

|    |  |     |
|----|--|-----|
| 21 | American History X                         | 8.6 |
| 21 | The Usual Suspects                         | 8.6 |
| 21 | Modern Times                               | 8.6 |
| 29 | The Dark Knight Rises                      | 8.5 |
| 29 | Gladiator                                  | 8.5 |
| 29 | Terminator 2: Judgment Day                 | 8.5 |
| 29 | Django Unchained                           | 8.5 |
| 29 | The Departed                               | 8.5 |
| 29 | The Lion King                              | 8.5 |
| 29 | The Green Mile                             | 8.5 |
| 29 | The Prestige                               | 8.5 |
| 29 | The Pianist                                | 8.5 |
| 29 | Apocalypse Now                             | 8.5 |
| 29 | Raiders of the Lost Ark                    | 8.5 |
| 29 | Psycho                                     | 8.5 |
| 29 | Back to the Future                         | 8.5 |
| 29 | Alien                                      | 8.5 |
| 29 | Memento                                    | 8.5 |
| 29 | Whiplash                                   | 8.5 |
| 29 | The Lives of Others                        | 8.5 |
| 29 | Children of Heaven                         | 8.5 |
| 47 | WALL·E                                     | 8.4 |
| 47 | Braveheart                                 | 8.4 |
| 47 | Amélie                                     | 8.4 |
| 47 | Star Wars: Episode VI - Return of the Jedi | 8.4 |
| 47 | Once Upon a Time in America                | 8.4 |
| 47 | Princess Mononoke                          | 8.4 |
| 47 | Aliens                                     | 8.4 |
| 47 | American Beauty                            | 8.4 |
| 47 | Lawrence of Arabia                         | 8.4 |
| 47 | Das Boot                                   | 8.4 |
| 47 | Requiem for a Dream                        | 8.4 |
| 47 | Oldboy                                     | 8.4 |
| 47 | Reservoir Dogs                             | 8.4 |
| 47 | A Separation                               | 8.4 |
| 61 | Toy Story 3                                | 8.3 |
| 61 | Up   | 8.3 |
| 61 | Inside Out                                 | 8.3 |
| 61 | Batman Begins                              | 8.3 |
| 61 | Inglourious Basterds                       | 8.3 |
| 61 | Indiana Jones and the Last Crusade         | 8.3 |
| 61 | L.A. Confidential                          | 8.3 |
| 61 | Toy Story                                  | 8.3 |
| 61 | Scarface                                   | 8.3 |
| 61 | Eternal Sunshine of the Spotless Mind      | 8.3 |
| 61 | Amadeus                                    | 8.3 |
| 61 | Raging Bull                                | 8.3 |
| 61 | Metropolis                                 | 8.3 |



|     |  |     |
|-----|--|-----|
| 61  | Unforgiven   | 8.3 |
| 61  | Downfall   | 8.3 |
| 61  | Room   | 8.3 |
| 61  | Snatch   | 8.3 |
| 61  | 2001: A Space Odyssey                                  | 8.3 |
| 61  | Good Will Hunting                                      | 8.3 |
| 61  | The Sting  | 8.3 |
| 61  | The Hunt   | 8.3 |
| 61  | Some Like It Hot                                       | 8.3 |
| 61  | Monty Python and the Holy Grail                        | 8.3 |
| 84  | Captain America: Civil War                             | 8.2 |
| 84  | How to Train Your Dragon                               | 8.2 |
| 84  | The Wolf of Wall Street                                | 8.2 |
| 84  | Finding Nemo   | 8.2 |
| 84  | A Beautiful Mind                                       | 8.2 |
| 84  | Casino   | 8.2 |
| 84  | V for Vendetta   | 8.2 |
| 84  | The Thing  | 8.2 |
| 84  | Die Hard   | 8.2 |
| 84  | Blade Runner   | 8.2 |
| 84  | Gran Torino  | 8.2 |
| 84  | Warrior  | 8.2 |
| 84  | Howl's Moving Castle                                   | 8.2 |
| 84  | Into the Wild  | 8.2 |
| 84  | Pan's Labyrinth  | 8.2 |
| 84  | The Big Lebowski                                       | 8.2 |
| 84  | Incendies  | 8.2 |
| 84  | Gone with the Wind                                     | 8.2 |
| 84  | The Secret in Their Eyes                               | 8.2 |
| 84  | Trainspotting  | 8.2 |
| 84  | The Bridge on the River Kwai                           | 8.2 |
| 84  | Lock, Stock and Two Smoking Barrels                    | 8.2 |
| 84  | On the Waterfront                                      | 8.2 |
| 107 | The Avengers   | 8.1 |
| 107 | Guardians of the Galaxy                                | 8.1 |
| 107 | Mad Max: Fury Road                                     | 8.1 |
| 107 | The Revenant   | 8.1 |
| 107 | The Bourne Ultimatum                                   | 8.1 |
| 107 | Pirates of the Caribbean: The Curse of the Black Pearl | 8.1 |
| 107 | Monsters, Inc.   | 8.1 |
| 107 | The Martian  | 8.1 |
| 107 | Shutter Island   | 8.1 |
| 107 | Jurassic Park  | 8.1 |
| 107 | Gone Girl  | 8.1 |
| 107 | The Truman Show  | 8.1 |
| 107 | Deadpool   | 8.1 |
| 107 | Kill Bill: Vol. 1                                      | 8.1 |

|     |                                    |     |
|-----|------------------------------------|-----|
| 107 | Prisoners                          | 8.1 |
| 107 | The Sixth Sense                    | 8.1 |
| 107 | Sin City                           | 8.1 |
| 107 | Rush                               | 8.1 |
| 107 | The Grand Budapest Hotel           | 8.1 |
| 107 | Million Dollar Baby                | 8.1 |
| 107 | The Help                           | 8.1 |
| 107 | No Country for Old Men             | 8.1 |
| 107 | There Will Be Blood                | 8.1 |
| 107 | 12 Years a Slave                   | 8.1 |
| 107 | Spotlight                          | 8.1 |
| 107 | Hotel Rwanda                       | 8.1 |
| 107 | The Imitation Game                 | 8.1 |
| 107 | The Princess Bride                 | 8.1 |
| 107 | Groundhog Day                      | 8.1 |
| 107 | The Sea Inside                     | 8.1 |
| 107 | Tae Guk Gi: The Brotherhood of War | 8.1 |
| 107 | Stand by Me                        | 8.1 |
| 107 | Akira                              | 8.1 |
| 107 | Elite Squad                        | 8.1 |
| 107 | The Terminator                     | 8.1 |
| 107 | Platoon                            | 8.1 |
| 107 | Butch Cassidy and the Sundance Kid | 8.1 |
| 107 | Donnie Darko                       | 8.1 |
| 107 | Annie Hall                         | 8.1 |
| 107 | The Wizard of Oz                   | 8.1 |
| 107 | Before Sunrise                     | 8.1 |
| 107 | The Best Years of Our Lives        | 8.1 |
| 107 | Amores Perros                      | 8.1 |
| 107 | The Celebration                    | 8.1 |
| 107 | Rocky                              | 8.1 |
| 152 | X-Men: Days of Future Past         | 8   |
| 152 | Ratatouille                        | 8   |
| 152 | Star Trek                          | 8   |
| 152 | Life of Pi                         | 8   |
| 152 | Casino Royale                      | 8   |
| 152 | Blood Diamond                      | 8   |
| 152 | The Incredibles                    | 8   |
| 152 | Cinderella Man                     | 8   |
| 152 | Big Fish                           | 8   |
| 152 | The Pursuit of Happyness           | 8   |
| 152 | Kill Bill: Vol. 2                  | 8   |
| 152 | Catch Me If You Can                | 8   |
| 152 | The Iron Giant                     | 8   |
| 152 | JFK                                | 8   |
| 152 | Serenity                           | 8   |
| 152 | Magnolia                           | 8   |
| 152 | District 9                         | 8   |

|     |                                     |     |
|-----|-------------------------------------|-----|
| 152 | Mystic River                        | 8   |
| 152 | Aladdin                             | 8   |
| 152 | Rain Man                            | 8   |
| 152 | Her                                 | 8   |
| 152 | Dances with Wolves                  | 8   |
| 152 | Dead Poets Society                  | 8   |
| 152 | The Artist                          | 8   |
| 152 | The King's Speech                   | 8   |
| 152 | Brazil                              | 8   |
| 152 | In Bruges                           | 8   |
| 152 | Mulholland Drive                    | 8   |
| 152 | Slumdog Millionaire                 | 8   |
| 152 | Black Swan                          | 8   |
| 152 | The Perks of Being a Wallflower     | 8   |
| 152 | True Romance                        | 8   |
| 152 | Dancer in the Dark                  | 8   |
| 152 | The Exorcist                        | 8   |
| 152 | Jaws                                | 8   |
| 152 | Casino Royale                       | 8   |
| 152 | Doctor Zhivago                      | 8   |
| 152 | The Straight Story                  | 8   |
| 152 | Fiddler on the Roof                 | 8   |
| 152 | Sicko                               | 8   |
| 152 | My Name Is Khan                     | 8   |
| 152 | The Sound of Music                  | 8   |
| 152 | Persepolis                          | 8   |
| 152 | Dallas Buyers Club                  | 8   |
| 152 | Shaun of the Dead                   | 8   |
| 152 | Sling Blade                         | 8   |
| 152 | Boyhood                             | 8   |
| 152 | Bowling for Columbine               | 8   |
| 152 | Central Station                     | 8   |
| 152 | Young Frankenstein                  | 8   |
| 152 | Before Sunset                       | 8   |
| 152 | Waltz with Bashir                   | 8   |
| 152 | A Fistful of Dollars                | 8   |
| 205 | Avatar                              | 7.9 |
| 205 | The Hobbit: The Desolation of Smaug | 7.9 |
| 205 | Iron Man                            | 7.9 |
| 205 | Edge of Tomorrow                    | 7.9 |
| 205 | Big Hero 6                          | 7.9 |
| 205 | How to Train Your Dragon 2          | 7.9 |
| 205 | Toy Story 2                         | 7.9 |
| 205 | Children of Men                     | 7.9 |
| 205 | The Insider                         | 7.9 |
| 205 | The Hateful Eight                   | 7.9 |
| 205 | The Bourne Identity                 | 7.9 |
| 205 | Almost Famous                       | 7.9 |

|     |                                |     |
|-----|--------------------------------|-----|
| 205 | Captain Phillips               | 7.9 |
| 205 | Shrek                          | 7.9 |
| 205 | Hero                           | 7.9 |
| 205 | The Notebook                   | 7.9 |
| 205 | Glory                          | 7.9 |
| 205 | Walk the Line                  | 7.9 |
| 205 | Straight Outta Compton         | 7.9 |
| 205 | The Blues Brothers             | 7.9 |
| 205 | The Right Stuff                | 7.9 |
| 205 | Taken                          | 7.9 |
| 205 | The Untouchables               | 7.9 |
| 205 | The World's Fastest Indian     | 7.9 |
| 205 | Edward Scissorhands            | 7.9 |
| 205 | Glory                          | 7.9 |
| 205 | Ed Wood                        | 7.9 |
| 205 | My Fair Lady                   | 7.9 |
| 205 | Halloween                      | 7.9 |
| 205 | Hot Fuzz                       | 7.9 |
| 205 | Crouching Tiger, Hidden Dragon | 7.9 |
| 205 | The Remains of the Day         | 7.9 |
| 205 | Boogie Nights                  | 7.9 |
| 205 | Letters from Iwo Jima          | 7.9 |
| 205 | The Fighter                    | 7.9 |
| 205 | E.T. the Extra-Terrestrial     | 7.9 |
| 205 | Crash                          | 7.9 |
| 205 | Amour                          | 7.9 |
| 205 | Nightcrawler                   | 7.9 |
| 205 | Little Miss Sunshine           | 7.9 |
| 205 | The Wrestler                   | 7.9 |
| 205 | The Chorus                     | 7.9 |
| 205 | Do the Right Thing             | 7.9 |
| 205 | Moon                           | 7.9 |
| 205 | Before Midnight                | 7.9 |
| 205 | Nine Queens                    | 7.9 |

## Top Foreign Language Film

Query:

with cte as(

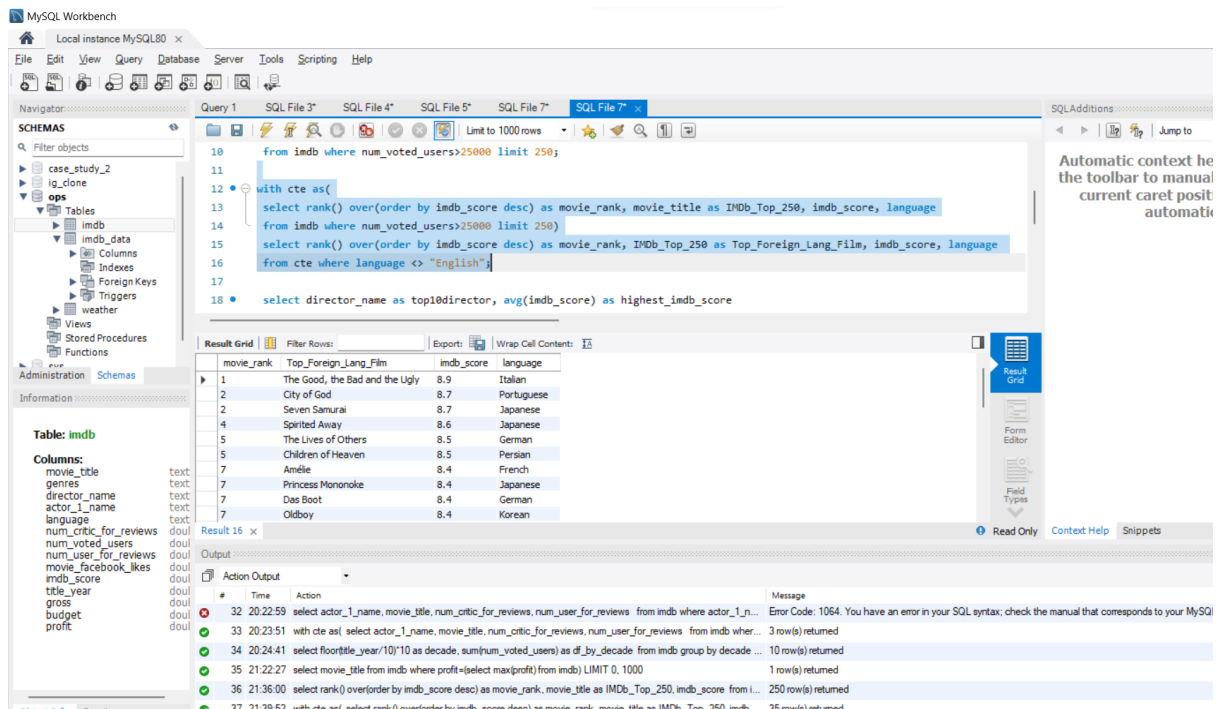
```
select rank() over(order by imdb_score desc) as movie_rank, movie_title as
IMDb_Top_250, imdb_score, language
```

```
from imdb where num_voted_users>25000 limit 250)
```

```
select rank() over(order by imdb_score desc) as movie_rank, IMDb_Top_250 as
Top_Foreign_Lang_Film, imdb_score, language
```

```
from cte where language <> "English";
```

## Output:



The screenshot shows the MySQL Workbench interface. The SQL editor contains the following query:

```

10 from imdb where num_voted_users>25000 limit 250;
11
12 with cte as(
13   select rank() over(order by imdb_score desc) as movie_rank, movie_title as IMDb_Top_250, imdb_score, language
14   from imdb where num_voted_users>25000 limit 250)
15   select rank() over(order by imdb_score desc) as movie_rank, IMDb_Top_250 as Top_Foreign_Lang_Film, imdb_score, language
16   from cte where language <> "English";
17
18 select director_name as top10director, avg(imdb_score) as highest_imdb_score

```

The Results Grid shows the following data:

| movie_rank | Top_Foreign_Lang_Film          | imdb_score | language   |
|------------|--------------------------------|------------|------------|
| 1          | The Good, the Bad and the Ugly | 8.9        | Italian    |
| 2          | City of God                    | 8.7        | Portuguese |
| 2          | Seven Samurai                  | 8.7        | Japanese   |
| 4          | Spirited Away                  | 8.6        | Japanese   |
| 5          | The Lives of Others            | 8.5        | German     |
| 5          | Children of Heaven             | 8.5        | Persian    |
| 7          | Amélie                         | 8.4        | French     |
| 7          | Princess Mononoke              | 8.4        | Japanese   |
| 7          | Das Boot                       | 8.4        | German     |
| 7          | Oldboy                         | 8.4        | Korean     |

The Action Output pane shows the execution progress of the query, including the CTE creation and the final result set retrieval.

| movie_rank | Top_Foreign_Lang_Film          | imdb_score | language   |
|------------|--------------------------------|------------|------------|
| 1          | The Good, the Bad and the Ugly | 8.9        | Italian    |
| 2          | City of God                    | 8.7        | Portuguese |
| 2          | Seven Samurai                  | 8.7        | Japanese   |
| 4          | Spirited Away                  | 8.6        | Japanese   |
| 5          | The Lives of Others            | 8.5        | German     |
| 5          | Children of Heaven             | 8.5        | Persian    |
| 7          | Amélie                         | 8.4        | French     |
| 7          | Princess Mononoke              | 8.4        | Japanese   |
| 7          | Das Boot                       | 8.4        | German     |
| 7          | Oldboy                         | 8.4        | Korean     |
| 7          | A Separation                   | 8.4        | Persian    |
| 12         | Metropolis                     | 8.3        | German     |
| 12         | Downfall                       | 8.3        | German     |
| 12         | The Hunt                       | 8.3        | Danish     |

|    |                                    |     |            |
|----|------------------------------------|-----|------------|
| 15 | Howl's Moving Castle               | 8.2 | Japanese   |
| 15 | Pan's Labyrinth                    | 8.2 | Spanish    |
| 15 | Incendies                          | 8.2 | French     |
| 15 | The Secret in Their Eyes           | 8.2 | Spanish    |
| 19 | The Sea Inside                     | 8.1 | Spanish    |
| 19 | Tae Guk Gi: The Brotherhood of War | 8.1 | Korean     |
| 19 | Akira                              | 8.1 | Japanese   |
| 19 | Elite Squad                        | 8.1 | Portuguese |
| 19 | Amores Perros                      | 8.1 | Spanish    |
| 19 | The Celebration                    | 8.1 | Danish     |
| 25 | My Name Is Khan                    | 8   | Hindi      |
| 25 | Persepolis                         | 8   | French     |
| 25 | Central Station                    | 8   | Portuguese |
| 25 | Waltz with Bashir                  | 8   | Hebrew     |
| 25 | A Fistful of Dollars               | 8   | Italian    |
| 30 | Hero                               | 7.9 | Mandarin   |
| 30 | Crouching Tiger, Hidden Dragon     | 7.9 | Mandarin   |
| 30 | Letters from Iwo Jima              | 7.9 | Japanese   |
| 30 | Amour                              | 7.9 | French     |
| 30 | The Chorus                         | 7.9 | French     |
| 30 | Nine Queens                        | 7.9 | Spanish    |

D. **Best Directors:** Group the column using the director\_name column.

Find out the top 10 directors for whom the mean of imdb\_score is the highest and store them in a new column top10director. In case of a tie in IMDb score between two directors, sort them alphabetically.

**Your task:** Find the best directors

**Query:**

```
select director_name as top10director, avg(imdb_score) as highest_imdb_score
from imdb group by director_name order by highest_imdb_score desc, director_name
limit 10;
```

**Output:**

|   | top10director     | highest_imdb_score |
|---|-------------------|--------------------|
| ▶ | Akira Kurosawa    | 8.7                |
|   | Charles Chaplin   | 8.6                |
|   | Tony Kaye         | 8.6                |
|   | Alfred Hitchcock  | 8.5                |
|   | Damien Chazelle   | 8.5                |
|   | Majid Majidi      | 8.5                |
|   | Ron Fricke        | 8.5                |
|   | Sergio Leone      | 8.433333333333334  |
|   | Christopher Nolan | 8.425              |
|   | Asghar Farhadi    | 8.4                |

MySQL Workbench

Local instance MySQL80 x

File Edit View Query Database Server Tools Scripting Help

Navigator

Filter objects

SCHMAS

case\_study\_2

ig\_clone

ops

Tables

imdb

imdb\_data

Columns

Indexes

Foreign Keys

weather

Views

Stored Procedures

Functions

Administration Schemas

Information

Table: imdb

Columns:

movie\_title text

genres text

director\_name text

actor\_1\_name text

language text

num\_crit\_for\_reviews doubl

num\_voted\_users doubl

num\_user\_for\_reviews doubl

movie\_facebook\_likes doubl

imdb\_score doubl

title\_year doubl

gross doubl

budget doubl

profit doubl

Query 1 SQL File 3\* SQL File 4\* SQL File 5\* SQL File 7\* SQL File 7\*

Limit to 1000 rows

SQL Additions

Automatic context help is disabled the toolbar to manually get help current caret position or to toggle automatic help.

16 from cte where language <> "English";

17

18 • select director\_name as top10director, avg(imdb\_score) as highest\_imdb\_score

19 from imdb group by director\_name order by highest\_imdb\_score desc, director\_name limit 10;

20

21

Result Grid

Filter Rows:

Exports

Wrap Cell Content: 15

Fetch rows:

Read Only

Context Help

Snippets

Output

Action Output

# Time Action Message Duration / #

37 21:39:52 with cte as( select rank() over(order by imdb\_score desc) as movie\_rank, movie\_title as IMDb\_Top\_250, imdb... 35 row(s) returned 0.000 sec /

38 21:43:57 select director\_name as top10director, avg(imdb\_score) as highest\_imdb\_score from imdb group by director\_n... 10 row(s) returned 0.032 sec /

39 21:44:36 select director\_name as top10director, avg(imdb\_score) as highest\_imdb\_score from imdb group by director\_n... 10 row(s) returned 0.016 sec /

E. **Popular Genres:** Perform this step using the knowledge gained while performing previous steps.

**Your task:** Find popular genres

**Query:**

select genres as popular\_genres, avg(imdb\_score) as highest\_imdb\_score

from imdb group by genres order by highest\_imdb\_score desc, genres limit 10;

## Output:

The most popular genres are Adventure|Animation|Drama|Family|Musical and Crime|Drama|Fantasy|Mystery, with the highest IMDb score of 8.5

| popular_genres                                | highest_imdb_score |
|---|--------------------|
| Adventure Animation Drama Family Musical      | 8.5                |
| Crime Drama Fantasy Mystery                   | 8.5                |
| Adventure Animation Fantasy                   | 8.4                |
| Adventure Drama Thriller War                  | 8.4                |
| Adventure Animation Comedy Drama Family Fa... | 8.3                |
| Biography Drama History Music                 | 8.3                |
| Documentary Drama Sport                       | 8.3                |
| Adventure Drama War                           | 8.25               |
| Biography Crime Documentary History           | 8.2                |
| Drama Fantasy War                             | 8.2                |

The screenshot shows the MySQL Workbench interface. The 'Schemas' pane on the left shows the 'imdb' database selected. The 'Query' editor in the center contains the following SQL query:

```
select genres as popular_genres, avg(imdb_score) as highest_imdb_score
from imdb group by genres order by highest_imdb_score desc, genres limit 10;
```

The 'Result Grid' on the right displays the results of the query, which are identical to the table shown in the previous block. The 'Action Output' pane at the bottom shows the execution log with three entries:

- 38 21:43:57 select director\_name as top10director, avg(imdb\_score) as highest\_imdb\_score from imdb group by director\_n... 10 row(s) returned 0.032 sec / 0.00
- 39 21:44:36 select director\_name as top10director, avg(imdb\_score) as highest\_imdb\_score from imdb group by director\_n... 10 row(s) returned 0.016 sec / 0.00
- 40 21:46:20 select genres as popular\_genres, avg(imdb\_score) as highest\_imdb\_score from imdb group by genres order by ... 10 row(s) returned 0.015 sec / 0.00

- F. **Charts:** Create three new columns namely, Meryl\_Streep, Leo\_Caprio, and Brad\_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor\_1\_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo



DiCaprio', and 'Brad Pitt' for the said extraction.

Append the rows of all these columns and store them in a new column named Combined.

Group the combined column using the actor\_1\_name column.

Find the mean of the num\_critic\_for\_reviews and num\_users\_for\_review and identify the actors which have the highest mean.

Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title\_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade. Store this in a new data frame called df\_by\_decade.

**Your task:** Find the critic-favorite and audience-favorite actors

**Query:**

```
with cte as(
```

```
select actor_1_name, movie_title, num_critic_for_reviews, num_user_for_reviews
```

```
from imdb where actor_1_name in ("Meryl Streep","Leonardo DiCaprio","Brad Pitt")
```

```
order by actor_1_name)
```

```
select actor_1_name, round(avg(num_critic_for_reviews),3) as  
mean_num_critic_for_reviews,
```

```
round(avg(num_user_for_reviews),3) as mean_num_user_for_reviews
```

```
from cte group by actor_1_name
```

```
order by mean_num_critic_for_reviews desc , mean_num_user_for_reviews desc;
```

**Output:**

```
Leonardo DiCaprio is the most popular actor among both critics and users
```

|   | actor_1_name      | mean_num_critc_for_reviews | mean_num_user_for_reviews |
|---|-------------------|----------------------------|---------------------------|
| ▶ | Leonardo DiCaprio | 330.19                     | 914.476                   |
|   | Brad Pitt         | 245                        | 742.353                   |
|   | Meryl Streep      | 181.455                    | 297.182                   |

MySQL Workbench

Local instance MySQL80 x

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

case\_study\_2

lg\_done

ops

Tables

imdb

imdb\_data

Columns

Indexes

Foreign Keys

Triggers

Views

Stored Procedures

Functions

Administration Schemas

Information

Table: imdb

Columns:

movie\_title text

genres text

director\_name text

actor\_1\_name text

language text

num\_critc\_for\_reviews text

num\_voted\_users doubl

num\_user\_for\_reviews doubl

movie\_facebook\_likes doubl

imdb\_score doubl

title\_year doubl

gross doubl

budget doubl

profit doubl

Query 1 SQL File 3\* SQL File 4\* SQL File 5\* SQL File 7\* SQL File 7\*

Limit to 1000 rows

SQLAdditions

Automatic context help is available in the toolbar to manually get the current caret position or automatic help

```

24
25
26 with cte as(
27   select actor_1_name, movie_title, num_critc_for_reviews, num_user_for_reviews
28   from imdb where actor_1_name in ("Meryl Streep","Leonardo DiCaprio","Brad Pitt")
29   order by actor_1_name)
30   select actor_1_name, round(avg(num_critc_for_reviews),3) as mean_num_critc_for_reviews,
31   round(avg(num_user_for_reviews),3) as mean_num_user_for_reviews
32   from cte group by actor_1_name
33   order by mean_num_critc_for_reviews desc , mean_num_user_for_reviews desc;
34

```

Result Grid

|   | actor_1_name      | mean_num_critc_for_reviews | mean_num_user_for_reviews |
|---|-------------------|----------------------------|---------------------------|
| ▶ | Leonardo DiCaprio | 330.19                     | 914.476                   |
|   | Brad Pitt         | 245                        | 742.353                   |
|   | Meryl Streep      | 181.455                    | 297.182                   |

Result 20 x

Output

Action Output

| #  | Time     | Action  | Message            |
|----|----------|---|--------------------|
| 39 | 21:44:36 | select director_name as top10director, avg(imdb_score) as highest_imdb_score from imdb group by director_n... | 10 row(s) returned |
| 40 | 21:46:20 | select genres as popular_genres, avg(imdb_score) as highest_imdb_score from imdb group by genres order by ... | 10 row(s) returned |
| 41 | 21:47:29 | with cte and select actor_1_name movie_title num_critc_for_reviews num_user_for_reviews from imdb where       | 3 row(s) returned  |

Read Only Context Help Snippets

**Change in number of voted users over decades:**

**Query:**

```

select floor(title_year/10)*10 as decade, sum(num_voted_users) as df_by_decade
from imdb group by decade order by decade;

```

**Output:**

|   | decade | df_by_decade |
|---|--------|--------------|
| ▶ | 1920   | 116387       |
|   | 1930   | 804839       |
|   | 1940   | 159517       |
|   | 1950   | 678336       |
|   | 1960   | 2982551      |
|   | 1970   | 8523299      |
|   | 1980   | 19987476     |
|   | 1990   | 69581866     |
|   | 2000   | 170681343    |
|   | 2010   | 119432961    |

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator: Filter objects

SCHEMAS

- case\_study\_2
- lg\_clone
- ops
  - Tables
    - imdb
      - imdb\_data
        - Columns
        - Indexes
        - Foreign Keys
        - Triggers
      - weather
    - Views
    - Stored Procedures
    - Functions
  - Administration
  - Schemas

Information

Table: **imdb**

Columns:

- movie\_title: text
- genres: text
- director\_name: text
- actor\_1\_name: text
- language: text
- num\_critc\_for\_reviews: doubl
- num\_voted\_users: doubl
- num\_user\_for\_reviews: doubl
- movie\_facebook\_likes: doubl
- imdb\_score: doubl
- title\_year: doubl
- gross: doubl
- budget: doubl
- profit: doubl

Query 1 SQL File 3\* SQL File 4\* SQL File 5\* SQL File 7\* SQL File 7\*

```

33 order by mean_num_critc_for_reviews desc , mean_num_user_for_reviews desc;
34
35
36 select floor(title_year/10) as decade, sum(num_voted_users) as df_by_decade
37 from imdb group by decade order by decade;
38
39

```

Result Grid

|   | decade | df_by_decade |
|---|--------|--------------|
| ▶ | 1920   | 116387       |
|   | 1930   | 804839       |
|   | 1940   | 159517       |
|   | 1950   | 678336       |
|   | 1960   | 2982551      |
|   | 1970   | 8523299      |
|   | 1980   | 19987476     |
|   | 1990   | 69581866     |
|   | 2000   | 170681343    |
|   | 2010   | 119432961    |

SQLAdditions

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

Read Only Context Help Snippets

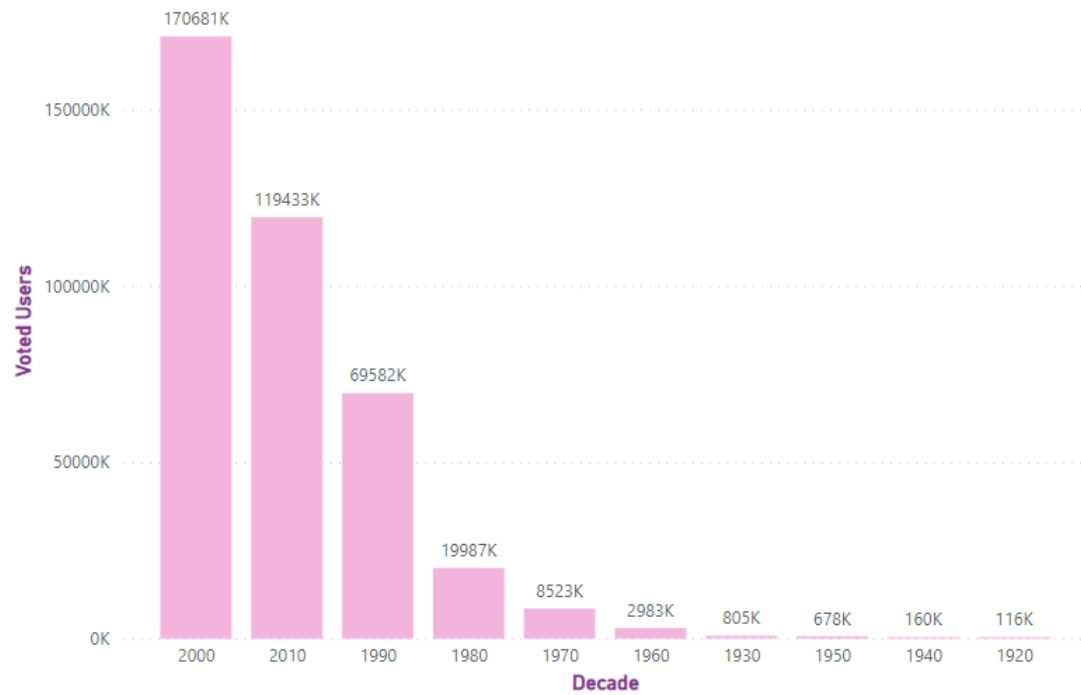
Result 21 x

Output

| #  | Time     | Action  | Message            | Duration / Fetch      |
|----|----------|---|--------------------|-----------------------|
| 40 | 21:46:20 | select genres as popular_genres, avg(imdb_score) as highest_imdb_score from imdb group by genres order by ... | 10 row(s) returned | 0.015 sec / 0.000 sec |
| 41 | 21:47:29 | with cte as( select actor_1_name, movie_title, num_critc_for_reviews, num_user_for_reviews from imdb wher ... | 3 row(s) returned  | 0.016 sec / 0.000 sec |
| 42 | 21:48:22 | select floortitle_year/10 as decade, sum(num_voted_users) as df_by_decade from imdb group by decade ...       | 10 row(s) returned | 0.000 sec / 0.000 sec |

Object Info Session

**Number of voted users over Decade**



**Result:**

By working on this project, I gained hands-on experience in cleaning the data in MS Excel, improve my skills in SQL, learnt how to create charts using Power BI and experienced how to analyse data as a data analyst.