# IMDB Movie Analysis

• DESCRIPTION:

For your Final Project, we are providing you with dataset having various columns of different IMDB Movies. You are required to Frame the problem. For this task, you will need to define a problem you want to shed some light on.

We can do this by asking 'What?' This is where you frame the problem i.e. What is the problem?

Use these questions to guide your thinking:

What do you see happening?

What is your hypothesis for the cause of the problem? (This will be broadly based on intuition initially)

What is the impact of the problem on stakeholders?

What is the impact of the problem not being solved?

Answering these questions will help you define a problem you are trying to solve and will allow you to find the right data to solve it.

Once you have defined a problem, clean the data as necessary, and use your Data Analysis skills to explore the data set and derive insights.

Make sure to use 5 Whys Analysis in your analysis and use this to create a report which conveys a data story.

Once you have framed the problem and gathered initial insights from the data, you can ask the following questions as you dig deeper into your analysis.

What do you see happening?

What are the specific symptoms of the problem?

What is your hypothesis for the cause of the problem?

• FIVE 'WHYS' APPROACH:

Once you have the problem better defined, you can use 5 Whys technique to determine its root cause by repeatedly asking the question “Why”.

It's also called the Root Cause Analysis, developed by Sakichi Toyoda, founder of Toyota Industries. Here’s an example of how this technique could be used to figure

out the cause of the following problem: A business went over budget on a recent project.

Q: “Why did we go over budget on our project?” A: It took much longer than we expected to complete.

Q: “Why did it take longer than expected to complete?” A: We had to redesign several elements of the product.

Q: “Why did we have to redesign elements of the product?” A: Features of the product were confusing to use.

Q: “Why were the features of the product confusing to use?” A: We made incorrect assumptions about what users wanted.

Q: “Why did we make incorrect assumptions about what users wanted?”

A: Our user experience research team didn’t ask effective questions.

As we see above, what looked like a budgeting problem turned out to be a problem with the user experience team not working effectively. While asking Why is easy, what we're interested in is the answer. Each time we answer why, the next time gets more difficult as we must think deeper behind the reasons for this. As we ask why, we may find that you have multiple answers for the same question.

- APPROACH:

Data Cleaning and Pre-processing: This step involves removing any irrelevant or missing data from the dataset, and formatting the data so that it can be used for analysis.

Feature Engineering: This step involves creating new features or variables from the existing data that can be used to better understand the movies. For example, creating a new variable that represents the budget of the movie, or creating a new variable that represents the genre of the movie.

Exploratory Data Analysis (EDA): This step involves visualizing the data to understand the distribution of different variables and identify any patterns or trends in the data.

Data visualization: This step involves creating visual representations of the data, such as bar charts, line plots, and scatter plots, to help understand the data and draw insights.

Conclusion and Recommendation: This step will be the final step where the data is presented in an understandable way and conclusion is made based on the analysis.

- TECH STACK USED:
  - MS-Excel
  - MySQL
  - Power BI

- INSIGHTS AND SOLUTION:

1) 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 Nulls etc.)

| director_name | num_critic_for_reviews | gross | genres | actor_1_name | movie_title | num_voted_users | num_use |
|---|---|---|---|---|---|---|---|
| James Cameron | 723 | 760505847 | Action\|Adventure\|Fantasy\|Sci-Fi | CCH Pounder | AvatarÂ | 886204 | |
| Gore Verbinski | 302 | 309404152 | Action\|Adventure\|Fantasy | Johnny Depp | Pirates of the Caribbean: At World's EndÂ | 471220 | |
| Sam Mendes | 602 | 200074175 | Action\|Adventure\|Thriller | Christoph Waltz | SpectreÂ | 275868 | |
| Christopher Nolan | 813 | 448130642 | Action\|Thriller | Tom Hardy | The Dark Knight RisesÂ | 1144337 | |
| Andrew Stanton | 462 | 73058679 | Action\|Adventure\|Sci-Fi | Daryl Sabara | John CarterÂ | 212204 | |
| Sam Raimi | 392 | 336530303 | Action\|Adventure\|Romance | J.K. Simmons | Spider-Man 3Â | 383056 | |
| Nathan Greno | 324 | 200807262 | Adventure\|Animation\|Comedy\|Family\|Fantasy\|Musical\|Romance | Brad Garrett | TangledÂ | 294810 | |
| Joss Whedon | 635 | 458991599 | Action\|Adventure\|Sci-Fi | Chris Hemsworth | Avengers: Age of UltronÂ | 462669 | |
| David Yates | 375 | 301956980 | Adventure\|Family\|Fantasy\|Mystery | Alan Rickman | Harry Potter and the Half-Blood PrinceÂ | 321795 | |
| Zack Snyder | 673 | 330249062 | Action\|Adventure\|Sci-Fi | Henry Cavill | Batman v Superman: Dawn of JusticeÂ | 371639 | |
| Bryan Singer | 434 | 200069408 | Action\|Adventure\|Sci-Fi | Kevin Spacey | Superman ReturnsÂ | 240396 | |
| Marc Forster | 403 | 168368427 | Action\|Adventure | Giancarlo Giannini | Quantum of SolaceÂ | 330784 | |
| Gore Verbinski | 313 | 423032628 | Action\|Adventure\|Fantasy | Johnny Depp | Pirates of the Caribbean: Dead Man's ChestÂ | 522040 | |
| Gore Verbinski | 450 | 89289910 | Action\|Adventure\|Western | Johnny Depp | The Lone RangerÂ | 181792 | |
| Zack Snyder | 733 | 291021565 | Action\|Adventure\|Fantasy\|Sci-Fi | Henry Cavill | Man of SteelÂ | 548573 | |
| Andrew Adamson | 258 | 141614023 | Action\|Adventure\|Family\|Fantasy | Peter Dinklage | The Chronicles of Narnia: Prince CaspianÂ | 149922 | |
| Joss Whedon | 703 | 623279547 | Action\|Adventure\|Sci-Fi | Chris Hemsworth | The AvengersÂ | 995415 | |
| Rob Marshall | 448 | 241063875 | Action\|Adventure\|Fantasy | Johnny Depp | Pirates of the Caribbean: On Stranger TidesÂ | 370704 | |
| Barry Sonnenfeld | 451 | 179020854 | Action\|Adventure\|Comedy\|Family\|Fantasy\|Sci-Fi | Will Smith | Men in Black 3Â | 268154 | |
| Peter Jackson | 422 | 255108370 | Adventure\|Fantasy | Aidan Turner | The Hobbit: The Battle of the Five ArmiesÂ | 354228 | |
| Marc Webb | 599 | 262030663 | Action\|Adventure\|Fantasy | Emma Stone | The Amazing Spider-ManÂ | 451803 | |
| Ridley Scott | 343 | 105219735 | Action\|Adventure\|Drama\|History | Mark Addy | Robin HoodÂ | 211765 | |
| Peter Jackson | 509 | 258355354 | Adventure\|Fantasy | Aidan Turner | The Hobbit: The Desolation of SmaugÂ | 483540 | |
| Chris Weitz | 251 | 70083519 | Adventure\|Family\|Fantasy | Christopher Lee | The Golden CompassÂ | 149019 | |
| Peter Jackson | 446 | 218051260 | Action\|Adventure\|Drama\|Romance | Naomi Watts | King KongÂ | 316018 | |
| James Cameron | 315 | 658672302 | Drama\|Romance | Leonardo DiCaprio | TitanicÂ | 793059 | |
| Anthony Russo | 516 | 407197282 | Action\|Adventure\|Sci-Fi | Robert Downey Jr. | Captain America: Civil WarÂ | 272670 | |
| Peter Berg | 377 | 65173160 | Action\|Adventure\|Sci-Fi\|Thriller | Liam Neeson | BattleshipÂ | 202382 | |
| Colin Trevorrow | 644 | 652177271 | Action\|Adventure\|Sci-Fi\|Thriller | Bryce Dallas Howard | Jurassic WorldÂ | 418214 | |
| Sam Mendes | 750 | 304360277 | Action\|Adventure\|Thriller | Albert Finney | SkyfallÂ | 522030 | |
| Sam Raimi | 300 | 373377893 | Action\|Adventure\|Fantasy\|Romance | J.K. Simmons | Spider-Man 2Â | 411164 | |
| Shane Black | 608 | 408992272 | Action\|Adventure\|Sci-Fi | Robert Downey Jr. | Iron Man 3Â | 557489 | |
| Tim Burton | 451 | 334185206 | Adventure\|Family\|Fantasy | Johnny Depp | Alice in WonderlandÂ | 306320 | |
| Brett Ratner | 334 | 234360014 | Action\|Adventure\|Fantasy\|Sci-Fi\|Thriller | Hugh Jackman | X-Men: The Last StandÂ | 383427 | |
| Dan Scanlon | 376 | 268488329 | Adventure\|Animation\|Comedy\|Family\|Fantasy | Steve Buscemi | Monsters UniversityÂ | 235025 | |
| Michael Bay | 366 | 402076689 | Action\|Adventure\|Sci-Fi | Glenn Morshower | Transformers: Revenge of the FallenÂ | 323207 | |
| Michael Bay | 378 | 245428137 | Action\|Adventure\|Sci-Fi | Bingbing Li | Transformers: Age of ExtinctionÂ | 242420 | |

IMDB Movies

To clean the data, I arranged the columns in the correct format and increased the column width to improve readability. I also removed null values and duplicates by selecting them using the "Go to Special" feature option and deleting the entire row. This helped to ensure that the data was accurate and ready for analysis.

For deleting the Null values used the following steps:

- The "Go To Special" feature:
- Select the range of cells that you want to remove blank cells from.
- Press "Ctrl + G" to open the "Go To" dialog box.
- Select "Special" from the options.
- Select "Blanks" and click "OK".
- Press the "Delete" key to remove the blank cells.

For deleting the Duplicate values used the following steps:

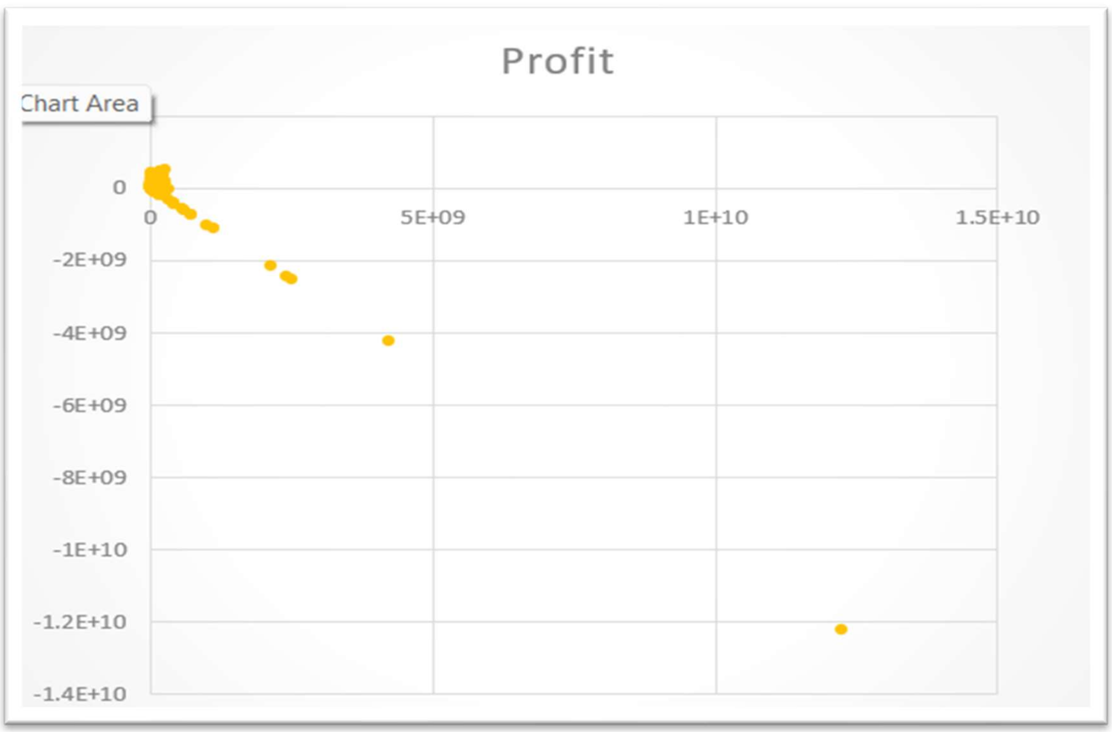
- Select the range of cells that you want to remove blank cells from.
- Go to the "Data" tab in the ribbon.
- Click on "Remove Duplicates" in the "Data Tools" group.
- Make sure that "All" columns are selected, and click "OK".

Before Cleaning: 5044 rows and 28 columns

After Cleaning: 3850 rows and 13 columns

2) 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?



| movie_title | Sum of Profit |
|---|---|
| AvatarÂ | 523505847 |
| Jurassic WorldÂ | 502177271 |
| TitanicÂ | 458672302 |
| Star Wars: Episode IV - A New HopeÂ | 449935665 |
| E.T. the Extra-TerrestrialÂ | 424449459 |
| The AvengersÂ | 403279547 |
| The Lion KingÂ | 377783777 |
| The Jungle BookÂ | 375290282 |
| Star Wars: Episode I - The Phantom MenaceÂ | 359544677 |
| The Dark KnightÂ | 348316061 |
| The Hunger GamesÂ | 329999255 |
| TwilightÂ | 308898950 |
| DeadpoolÂ | 305024263 |
| The Hunger Games: Catching FireÂ | 294645577 |
| Jurassic ParkÂ | 293784000 |
| Despicable Me 2Â | 292049635 |
| American SniperÂ | 291323553 |
| Finding NemoÂ | 286838870 |
| Shrek 2Â | 286471036 |
| The Lord of the Rings: The Return of the KingÂ | 283019252 |
| Star Wars: Episode VI - Return of the JediÂ | 276625409 |
| Forrest GumpÂ | 274691196 |
| Star Wars: Episode V - The Empire Strikes BackÂ | 272158751 |
| JunoÂ | 271985680 |
| Alice in WonderlandÂ | 268370412 |
| Home AloneÂ | 267761243 |
| Star Wars: Episode III - Revenge of the SithÂ | 267262555 |
| Spider-ManÂ | 264706375 |
| MinionsÂ | 262029560 |
| The Sixth SenseÂ | 253501675 |
| JawsÂ | 252000000 |
| FrozenÂ | 250736600 |
| The Secret Life of PetsÂ | 248505540 |
| **Total** | **22049537152** |

In this task we have to first create a new column to store the profit of the movies by taking the difference of the gross and budget

To identify outliers in the data, I plotted a chart and looked for any unusually high or low values. One example of an outlier that I observed was a value of -1.2E+10

I used a tool called Power BI to create this visualization showing that the movie with the highest profit was “Avatar’’

3) 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 are 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.

```
SELECT
 row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
 imdb_score, num_voted_users, movie_title AS
IMDb_Top_250, language
FROM
 imdb.imdb_movies
WHERE
 num_voted_users > 25000
LIMIT 250;
```

I used an SQL query to identify the top 250 movies with the highest IMDB scores and a minimum of 25,000 voted users. Here is the list:

| anking | imdb_scor | num_voted_user | IMDb_Top_250 | languag |
|---|---|---|---|---|
| 1 | 9 | 1676169 | The Dark KnightÃ,Â | English |
| 2 | 8.9 | 1215718 | The Lord of the Rings: The Return of the KingÃ,Â | English |
| 3 | 8.8 | 1468200 | InceptionÃ,Â | English |
| 4 | 8.8 | 1347461 | Fight ClubÃ,Â | English |
| 5 | 8.8 | 1251222 | Forrest GumpÃ,Â | English |
| 6 | 8.8 | 1238746 | The Lord of the Rings: The Fellowship of the RingÃ,Â | English |
| 7 | 8.8 | 213483 | DaredevilÃ,Â | English |
| 8 | 8.7 | 1217752 | The MatrixÃ,Â | English |
| 9 | 8.7 | 1100446 | The Lord of the Rings: The Two TowersÃ,Â | English |
| 10 | 8.6 | 928227 | InterstellarÃ,Â | English |
| 11 | 8.6 | 881236 | Saving Private RyanÃ,Â | English |
| 12 | 8.6 | 159910 | HannibalÃ,Â | English |
| 13 | 8.5 | 1144337 | The Dark Knight RisesÃ,Â | English |
| 14 | 8.5 | 982637 | GladiatorÃ,Â | English |
| 15 | 8.5 | 955174 | Django UnchainedÃ,Â | English |
| 16 | 8.5 | 873649 | The DepartedÃ,Â | English |
| 17 | 8.5 | 844052 | The PrestigeÃ,Â | English |
| 18 | 8.5 | 782610 | The Green MileÃ,Â | English |
| 19 | 8.5 | 744891 | Terminator 2: Judgment DayÃ,Â | English |
| 20 | 8.5 | 644348 | The Lion KingÃ,Â | English |
| 21 | 8.5 | 497946 | The PianistÃ,Â | English |
| 22 | 8.5 | 50391 | OutlanderÃ,Â | English |
| 23 | 8.4 | 736638 | BraveheartÃ,Â | English |
| 24 | 8.4 | 718837 | WALLÃ,Â·EÃ,Â | English |
| 25 | 8.4 | 534262 | AmÃƒÂ©lieÃ,Â | French |
| 26 | 8.4 | 63982 | Stargate SG-1Ã,Â | English |
| 27 | 8.4 | 62756 | Baahubali: The BeginningÃ,Â | Telugu |
| 28 | 8.3 | 980946 | Batman BeginsÃ,Â | English |
| 29 | 8.3 | 885175 | Inglourious BasterdsÃ,Â | English |
| 30 | 8.3 | 665575 | UpÃ,Â | English |
| 31 | 8.3 | 544884 | Toy Story 3Ã,Â | English |
| 32 | 8.3 | 515306 | Indiana Jones and the Last CrusadeÃ,Â | English |
| 33 | 8.3 | 414219 | L.A. ConfidentialÃ,Â | English |
| 34 | 8.3 | 345198 | Inside OutÃ,Â | English |
| 35 | 8.3 | 29450 | LifeÃ,Â | English |
| 36 | 8.2 | 791783 | V for VendettaÃ,Â | English |
| 37 | 8.2 | 780588 | The Wolf of Wall StreetÃ,Â | English |
| 38 | 8.2 | 692482 | Finding NemoÃ,Â | English |
| 39 | 8.2 | 610568 | A Beautiful MindÃ,Â | English |
| 40 | 8.2 | 485430 | How to Train Your DragonÃ,Â | English |
| 41 | 8.2 | 333542 | CasinoÃ,Â | English |
| 42 | 8.2 | 272670 | Captain America: Civil WarÃ,Â | English |
| 43 | 8.2 | 258078 | The ThingÃ,Â | English |
| 44 | 8.1 | 995415 | The AvengersÃ,Â | English |
| 45 | 8.1 | 995415 | The AvengersÃ,Â | English |
| 46 | 8.1 | 809474 | Pirates of the Caribbean: The Curse of the Black Pear | English |
| 47 | 8.1 | 786092 | Shutter IslandÃ,Â | English |
| 48 | 8.1 | 735784 | Kill Bill: Vol. 1Ã,Â | English |
| 49 | 8.1 | 704766 | The Sixth SenseÃ,Â | English |
| 50 | 8.1 | 682155 | Guardians of the GalaxyÃ,Â | English |
| 51 | 8.1 | 667983 | The Truman ShowÃ,Â | English |
| 52 | 8.1 | 656640 | Sin CityÃ,Â | English |

| anking | imdb_scor | num_voted_user | IMDb_Top_250 | languag |
|---|---|---|---|---|
| 53 | 8.1 | 613473 | Jurassic ParkÃ,Â | English |
| 54 | 8.1 | 585659 | Monsters, Inc.Ã,Â | English |
| 55 | 8.1 | 569841 | Gone GirlÃ,Â | English |
| 56 | 8.1 | 552503 | Mad Max: Fury RoadÃ,Â | English |
| 57 | 8.1 | 491077 | The Bourne UltimatumÃ,Â | English |
| 58 | 8.1 | 479047 | DeadpoolÃ,Â | English |
| 59 | 8.1 | 472488 | The MartianÃ,Â | English |
| 60 | 8.1 | 406020 | The RevenantÃ,Â | English |
| 61 | 8.1 | 383591 | PrisonersÃ,Â | English |
| 62 | 8.1 | 312629 | RushÃ,Â | English |
| 63 | 8.1 | 54057 | SolarisÃ,Â | Russian |
| 64 | 8 | 525801 | Catch Me If You CanÃ,Â | English |
| 65 | 8 | 514125 | X-Men: Days of Future PastÃ,Â | English |
| 66 | 8 | 512749 | Kill Bill: Vol. 2Ã,Â | English |
| 67 | 8 | 504419 | Star TrekÃ,Â | English |
| 68 | 8 | 479166 | The IncrediblesÃ,Â | English |
| 69 | 8 | 473887 | RatatouilleÃ,Â | English |
| 70 | 8 | 470483 | Casino RoyaleÃ,Â | English |
| 71 | 8 | 440084 | Life of PiÃ,Â | English |
| 72 | 8 | 400292 | Blood DiamondÃ,Â | English |
| 73 | 8 | 350698 | Big FishÃ,Â | English |
| 74 | 8 | 338383 | The Pursuit of HappynessÃ,Â | English |
| 75 | 8 | 242599 | SerenityÃ,Â | English |
| 76 | 8 | 241030 | MagnoliaÃ,Â | English |
| 77 | 8 | 148238 | Cinderella ManÃ,Â | English |
| 78 | 8 | 128455 | The Iron GiantÃ,Â | English |
| 79 | 8 | 113472 | JFKÃ,Â | English |
| 80 | 7.9 | 886204 | AvatarÃ,Â | English |
| 81 | 7.9 | 696338 | Iron ManÃ,Â | English |
| 82 | 7.9 | 637246 | The Hobbit: An Unexpected JourneyÃ,Â | English |
| 83 | 7.9 | 483540 | The Hobbit: The Desolation of SmaugÃ,Â | English |
| 84 | 7.9 | 467113 | ShrekÃ,Â | English |
| 85 | 7.9 | 431620 | Edge of TomorrowÃ,Â | English |
| 86 | 7.9 | 407601 | The Bourne IdentityÃ,Â | English |
| 87 | 7.9 | 385871 | Toy Story 2Ã,Â | English |
| 88 | 7.9 | 361767 | Children of MenÃ,Â | English |
| 89 | 7.9 | 323353 | Captain PhillipsÃ,Â | English |
| 90 | 7.9 | 279093 | Big Hero 6Ã,Â | English |
| 91 | 7.9 | 272839 | The Hateful EightÃ,Â | English |
| 92 | 7.9 | 221128 | How to Train Your Dragon 2Ã,Â | English |
| 93 | 7.9 | 207287 | Almost FamousÃ,Â | English |
| 94 | 7.9 | 149414 | HeroÃ,Â | Mandari |
| 95 | 7.9 | 133526 | The InsiderÃ,Â | English |
| 96 | 7.8 | 583341 | The HangoverÃ,Â | English |
| 97 | 7.8 | 582917 | GravityÃ,Â | English |
| 98 | 7.8 | 522030 | SkyfallÃ,Â | English |
| 99 | 7.8 | 518537 | X-Men: First ClassÃ,Â | English |
| 100 | 7.8 | 496749 | Captain America: The Winter SoldierÃ,Â | English |
| 101 | 7.8 | 459346 | The Curious Case of Benjamin ButtonÃ,Â | English |
| 102 | 7.8 | 402645 | Ocean's ElevenÃ,Â | English |
| 103 | 7.8 | 395573 | Star Trek Into DarknessÃ,Â | English |
| 104 | 7.8 | 382255 | Harry Potter and the Prisoner of AzkabanÃ,Â | English |

| anking | imdb_scor | num_voted_user | IMDb_Top_250 | languag |
|---|---|---|---|---|
| 105 | 7.8 | 348232 | The Bourne SupremacyÃ,Â | English |
| 106 | 7.8 | 340085 | Back to the Future Part IIÃ,Â | English |
| 107 | 7.8 | 330152 | The Girl with the Dragon TattooÃ,Â | English |
| 108 | 7.8 | 324671 | American GangsterÃ,Â | English |
| 109 | 7.8 | 294810 | TangledÃ,Â | English |
| 110 | 7.8 | 272534 | Wreck-It RalphÃ,Â | English |
| 111 | 7.8 | 261069 | The GameÃ,Â | English |
| 112 | 7.8 | 246698 | The Lego MovieÃ,Â | English |
| 113 | 7.8 | 237872 | 3:10 to YumaÃ,Â | English |
| 114 | 7.8 | 236000 | ApocalyptoÃ,Â | Maya |
| 115 | 7.8 | 225122 | Donnie BrascoÃ,Â | English |
| 116 | 7.8 | 220591 | GattacaÃ,Â | English |
| 117 | 7.8 | 213668 | The FugitiveÃ,Â | English |
| 118 | 7.8 | 199056 | ChangelingÃ,Â | English |
| 119 | 7.8 | 139114 | Fantastic Mr. FoxÃ,Â | English |
| 120 | 7.8 | 113068 | The Last of the MohicansÃ,Â | English |
| 121 | 7.8 | 106072 | The Jungle BookÃ,Â | English |
| 122 | 7.8 | 64989 | The Conjuring 2Ã,Â | English |
| 123 | 7.8 | 38383 | 3rd Rock from the SunÃ,Â | English |
| 124 | 7.8 | 28276 | The Little PrinceÃ,Â | English |
| 125 | 7.7 | 793059 | TitanicÃ,Â | English |
| 126 | 7.7 | 607235 | 300Ã,Â | English |
| 127 | 7.7 | 479453 | The Social NetworkÃ,Â | English |
| 128 | 7.7 | 452465 | ArgoÃ,Â | English |
| 129 | 7.7 | 399651 | Minority ReportÃ,Â | English |
| 130 | 7.7 | 394317 | Cast AwayÃ,Â | English |
| 131 | 7.7 | 392474 | WatchmenÃ,Â | English |
| 132 | 7.7 | 385943 | Despicable MeÃ,Â | English |
| 133 | 7.7 | 343274 | The Fifth ElementÃ,Â | English |
| 134 | 7.7 | 318634 | Love ActuallyÃ,Â | English |
| 135 | 7.7 | 317166 | The Last SamuraiÃ,Â | English |
| 136 | 7.7 | 305929 | Training DayÃ,Â | English |
| 137 | 7.7 | 301279 | ZodiacÃ,Â | English |
| 138 | 7.7 | 292022 | Black Hawk DownÃ,Â | English |
| 139 | 7.7 | 266310 | Man on FireÃ,Â | English |
| 140 | 7.7 | 240962 | True GritÃ,Â | English |
| 141 | 7.7 | 232710 | Seven PoundsÃ,Â | English |
| 142 | 7.7 | 224671 | As Good as It GetsÃ,Â | English |
| 143 | 7.7 | 223127 | The Blind SideÃ,Â | English |
| 144 | 7.7 | 212085 | StardustÃ,Â | English |
| 145 | 7.7 | 200359 | Road to PerditionÃ,Â | English |
| 146 | 7.7 | 189249 | Eastern PromisesÃ,Â | English |

| anking | imdb_scor | num_voted_user | IMDb_Top_250 | languag |
|---|---|---|---|---|
| 157 | 7.6 | 477300 | Sherlock HolmesÃ,Â | English |
| 158 | 7.6 | 421658 | FrozenÃ,Â | English |
| 159 | 7.6 | 403836 | Rise of the Planet of the ApesÃ,Â | English |
| 160 | 7.6 | 385670 | Harry Potter and the Goblet of FireÃ,Â | English |
| 161 | 7.6 | 328159 | Ice AgeÃ,Â | English |
| 162 | 7.6 | 317542 | Dawn of the Planet of the ApesÃ,Â | English |
| 163 | 7.6 | 307029 | Kung Fu PandaÃ,Â | English |
| 164 | 7.6 | 303185 | FuryÃ,Â | English |
| 165 | 7.6 | 299258 | Die Hard with a VengeanceÃ,Â | English |
| 166 | 7.6 | 293662 | CollateralÃ,Â | English |
| 167 | 7.6 | 283563 | MoneyballÃ,Â | English |
| 168 | 7.6 | 280228 | The TownÃ,Â | English |
| 169 | 7.6 | 273108 | Inside ManÃ,Â | English |
| 170 | 7.6 | 269033 | BatmanÃ,Â | English |
| 171 | 7.6 | 267980 | The Godfather: Part IIIÃ,Â | English |
| 172 | 7.6 | 248123 | Lord of WarÃ,Â | English |
| 173 | 7.6 | 243834 | Les MisÃƒÂ©rablesÃ,Â | English |
| 174 | 7.6 | 239752 | Interview with the Vampire: The Vampire Chronicles | English |
| 175 | 7.6 | 224013 | Moulin Rouge!Ã,Â | English |
| 176 | 7.6 | 208817 | Apollo 13Ã,Â | English |
| 177 | 7.6 | 203963 | Lone SurvivorÃ,Â | English |
| 178 | 7.6 | 188887 | Enemy at the GatesÃ,Â | English |
| 179 | 7.6 | 178118 | Bridge of SpiesÃ,Â | English |
| 180 | 7.6 | 176936 | MunichÃ,Â | English |
| 181 | 7.6 | 170684 | TrafficÃ,Â | English |
| 182 | 7.6 | 155496 | GrindhouseÃ,Â | English |
| 183 | 7.6 | 145270 | The ImpossibleÃ,Â | English |
| 184 | 7.6 | 138941 | The Thin Red LineÃ,Â | English |
| 185 | 7.6 | 136580 | The CrowÃ,Â | English |
| 186 | 7.6 | 131217 | The AbyssÃ,Â | English |
| 187 | 7.6 | 97838 | Star Trek: First ContactÃ,Â | English |
| 188 | 7.6 | 76016 | The HurricaneÃ,Â | English |
| 189 | 7.6 | 38690 | The Flowers of WarÃ,Â | Mandari |
| 190 | 7.6 | 25402 | The A-TeamÃ,Â | English |
| 191 | 7.5 | 462669 | Avengers: Age of UltronÃ,Â | English |
| 192 | 7.5 | 444683 | Harry Potter and the Sorcerer's StoneÃ,Â | English |
| 193 | 7.5 | 405973 | X-Men 2Ã,Â | English |
| 194 | 7.5 | 355137 | Harry Potter and the Order of the PhoenixÃ,Â | English |
| 195 | 7.5 | 354228 | The Hobbit: The Battle of the Five ArmiesÃ,Â | English |
| 196 | 7.5 | 338635 | Sherlock Holmes: A Game of ShadowsÃ,Â | English |
| 197 | 7.5 | 321795 | Harry Potter and the Half-Blood PrinceÃ,Â | English |
| 198 | 7.5 | 314033 | Gangs of New YorkÃ,Â | English |

| anking | imdb_score | num_voted_user | IMDb_Top_250 | languag |
|---|---|---|---|---|
| 208 | 7.5 | 171792 | MulanÃ‚Â | English |
| 209 | 7.5 | 154487 | SleepersÃ‚Â | English |
| 210 | 7.5 | 149285 | A Nightmare on Elm StreetÃ‚Â | English |
| 211 | 7.5 | 143835 | Bram Stoker's DraculaÃ‚Â | English |
| 212 | 7.5 | 117719 | Saving Mr. BanksÃ‚Â | English |
| 213 | 7.5 | 90360 | PinocchioÃ‚Â | English |
| 214 | 7.5 | 88270 | The Life of David GaleÃ‚Â | English |
| 215 | 7.5 | 66511 | 42Ã‚Â | English |
| 216 | 7.5 | 53607 | Star Trek BeyondÃ‚Â | English |
| 217 | 7.5 | 49049 | Sleepy HollowÃ‚Â | English |
| 218 | 7.5 | 36919 | ConstantineÃ‚Â | English |
| 219 | 7.4 | 452928 | X-MenÃ‚Â | English |
| 220 | 7.4 | 387616 | Harry Potter and the Chamber of SecretsÃ‚Â | English |
| 221 | 7.4 | 375456 | Crazy, Stupid, Love.Ã‚Â | English |
| 222 | 7.4 | 365104 | Mission: Impossible - Ghost ProtocolÃ‚Â | English |
| 223 | 7.4 | 313866 | Ant-ManÃ‚Â | English |
| 224 | 7.4 | 283480 | Back to the Future Part IIIÃ‚Â | English |
| 225 | 7.4 | 259492 | The RockÃ‚Â | English |
| 226 | 7.4 | 259083 | The Simpsons MovieÃ‚Â | English |
| 227 | 7.4 | 232187 | Mission: Impossible - Rogue NationÃ‚Â | English |
| 228 | 7.4 | 217480 | Law Abiding CitizenÃ‚Â | English |
| 229 | 7.4 | 216032 | Zero Dark ThirtyÃ‚Â | English |
| 230 | 7.4 | 200556 | ContactÃ‚Â | English |
| 231 | 7.4 | 197412 | LincolnÃ‚Â | English |
| 232 | 7.4 | 184795 | The Bucket ListÃ‚Â | English |
| 233 | 7.4 | 177383 | The Adventures of TintinÃ‚Â | English |
| 234 | 7.4 | 168207 | Master and Commander: The Far Side of the WorldÃ‚ | English |
| 235 | 7.4 | 148490 | K-PAXÃ‚Â | English |
| 236 | 7.4 | 142067 | The English PatientÃ‚Â | English |
| 237 | 7.4 | 136973 | The JudgeÃ‚Â | English |
| 238 | 7.4 | 124222 | InvictusÃ‚Â | English |
| 239 | 7.4 | 105446 | PoltergeistÃ‚Â | English |
| 240 | 7.4 | 100743 | Man on the MoonÃ‚Â | English |
| 241 | 7.4 | 99558 | A Time to KillÃ‚Â | English |
| 242 | 7.4 | 96654 | The Phantom of the OperaÃ‚Â | English |
| 243 | 7.4 | 91860 | Across the UniverseÃ‚Â | English |
| 244 | 7.4 | 90932 | Pirate RadioÃ‚Â | English |
| 245 | 7.4 | 77394 | The WalkÃ‚Â | English |
| 246 | 7.4 | 67797 | Mean StreetsÃ‚Â | English |
| 247 | 7.4 | 54643 | Wonder BoysÃ‚Â | English |
| 248 | 7.4 | 47764 | 13 HoursÃ‚Â | English |
| 249 | 7.4 | 36894 | Red CliffÃ‚Â | Mandari |
| 250 | 7.3 | 701607 | The Hunger GamesÃ‚Â | English |

From this I can also extract the films that are not in English language (Because majority of the movies are in English). And they are stored in a new column named Top_foreign_lang_film. Following SQL query used for finding non-English films.

That is, out of the top rated 250 IMDB movies only 31 films are of other languages and they are Italian, German, Hindi, Telugu, Persian, Spanish …etc and the rest of the movies are in English.

```sql
SELECT
 row_number() over(order by imdb_score DESC,
num_voted_users DESC) as ranking,
 imdb_score, num_voted_users, movie_title AS
Top_foreign_lang_film, language
FROM
 imdb.imdb_movies
WHERE
  language != 'English'
LIMIT 250;
```

| anking | imdb_score | num_voted_users | Top_foreign_lang_film | language |
|---|---|---|---|---|
| 1 | 8.4 | 534262 | AmÃƒÂ©lieÃ‚Â | French |
| 2 | 8.4 | 62756 | Baahubali: The BeginningÃ‚Â | Telugu |
| 3 | 8.3 | 15762 | The ReturnedÃ‚Â | French |
| 4 | 8.2 | 374 | Godzilla ResurgenceÃ‚Â | Japanese |
| 5 | 8.2 | 374 | Godzilla ResurgenceÃ‚Â | Japanese |
| 6 | 8.1 | 54057 | SolarisÃ‚Â | Russian |
| 7 | 7.9 | 149414 | HeroÃ‚Â | Mandarin |
| 8 | 7.8 | 236000 | ApocalyptoÃ‚Â | Maya |
| 9 | 7.8 | 7630 | OceansÃ‚Â | French |
| 10 | 7.8 | 5639 | EarthÃ‚Â | Hindi |
| 11 | 7.7 | 62607 | A Very Long EngagementÃ‚Â | French |
| 12 | 7.6 | 38690 | The Flowers of WarÃ‚Â | Mandarin |
| 13 | 7.4 | 36894 | Red CliffÃ‚Â | Mandarin |
| 14 | 7.2 | 24657 | MicmacsÃ‚Â | French |
| 15 | 7.2 | 21912 | Ip Man 3Ã‚Â | Cantonese |
| 16 | 7.2 | 6 | 10,000 B.C.Ã‚Â | |
| 17 | 7.1 | 22897 | The WarlordsÃ‚Â | Mandarin |
| 18 | 7.1 | 8 | Star Wars: Episode VII - The Force AwakensÃ‚Â | |
| 19 | 7 | 36455 | Curse of the Golden FlowerÃ‚Â | Mandarin |
| 20 | 6.8 | 15790 | A Monster in ParisÃ‚Â | French |
| 21 | 6.7 | 18209 | The Great RaidÃ‚Â | Filipino |
| 22 | 6.5 | 24557 | The GrandmasterÃ‚Â | Mandarin |
| 23 | 6.4 | 86152 | The InterpreterÃ‚Â | Aboriginal |
| 24 | 6.4 | 979 | EvolutionÃ‚Â | French |
| 25 | 6.1 | 11584 | Dragon BladeÃ‚Â | Mandarin |
| 26 | 6 | 3322 | Nomad: The WarriorÃ‚Â | Kazakh |
| 27 | 5.9 | 71574 | The Legend of ZorroÃ‚Â | Spanish |
| 28 | 5.3 | 4387 | Obitaemyy ostrovÃ‚Â | Russian |
| 29 | 5.1 | 20567 | Asterix at the Olympic GamesÃ‚Â | French |
| 30 | 4.9 | 590 | Animal Kingdom: Let's go ApeÃ‚Â | French |
| 31 | 4.4 | 230 | Top Cat BeginsÃ‚Â | Spanish |

4) Best Directors: T-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

```
SELECT
 director_name AS top_10_director,
 round(AVG(imdb_score), 2) AS avg_score
FROM
 imdb.imdb_movies
GROUP BY director_name
ORDER BY avg_score DESC , director_name
LIMIT 10
```

| top_10_director | avg_score |
|---|---|
| Christopher Nolan | 8.41 |
| S.S. Rajamouli | 8.4 |
| Lee Unkrich | 8.3 |
| Pete Docter | 8.23 |
| Hideaki Anno | 8.2 |
| Quentin Tarantino | 8.16 |
| Alejandro G. IÃ±Ãirritu | 8.1 |
| Andrei Tarkovsky | 8.1 |
| Denis Villeneuve | 8.1 |
| James Gunn | 8.1 |

Based on the data provided, it appears that Christopher Nolan and S.S. Rajamouli are the best director, with an average IMDB score of 8.4 for his movies.

5) Popular Genres: Perform this step using the knowledge gained while performing previous steps.

Your task: Find popular genres.

```
SELECT genres AS popular_genres,AVG(imdb_score) AS highest_imdb_score
FROM imdb.imdb_movies first
GROUP BY genres
ORDER BY AVG(imdb_score) DESC
limit 10
```

| | | |
|---|---|---|
| 1 | popular_genres | highest_imdb_score |
| 2 | Action\|Adventure\|Crime\|Drama\|Sci-Fi\|Thriller | 8.8 |
| 3 | Crime\|Drama\|Horror\|Mystery\|Thriller | 8.6 |
| 4 | Crime\|Drama\|Fantasy\|Mystery | 8.5 |
| 5 | Action\|Drama\|Romance | 8.5 |
| 6 | Drama\|Western | 8.5 |
| 7 | Adventure\|Animation\|Drama\|Family\|Musical | 8.5 |
| 8 | Drama\|Romance\|Sci-Fi | 8.5 |
| 9 | Action\|Adventure\|Drama\|Fantasy\|War | 8.4 |
| 10 | Adventure\|Animation\|Comedy\|Drama\|Family\|Fantasy | 8.3 |
| 11 | Drama\|Fantasy\|Horror\|Mystery | 8.3 |

Based on the data provided, it appears that the Action|Adventure|Crime|Drama|SciFi|Thriller genre has the highest average IMDB score, indicating that it is a more preferable genre.

6) 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-favourite and audience-favourite actor

```
SELECT actor_1_name,
COUNT(movie_title) AS no_of_movies,
ROUND(AVG(num_user_for_reviews), 2) AS user_reviews,
ROUND(AVG(num_critic_for_reviews), 2) AS critic_reviews
FROM ((SELECT actor_1_name, movie_title,
num_user_for_reviews, num_critic_for_reviews
FROM imdb.imdb_movies
WHERE actor_1_name = ' Meryl Streep')
UNION ALL (SELECT actor_1_name, movie_title,
num_user_for_reviews, num_critic_for_reviews
FROM imdb.imdb_movies
WHERE actor_1_name = ' Leonardo DiCaprio')
UNION ALL (SELECT actor_1_name, movie_title,
num_user_for_reviews, num_critic_for_reviews
FROM imdb.imdb_movies
WHERE actor_1_name = ' Brad Pitt')) c
GROUP BY actor_1_name
ORDER BY user_reviews DESC , critic_reviews DESC
```

| actor_1_name | no_of_movies | user_reviews | critic_reviews |
|---|---|---|---|
| Leonardo DiCaprio | 20 | 922.55 | 322.2 |
| Brad Pitt | 17 | 742.35 | 245 |
| Meryl Streep | 11 | 297.18 | 181.45 |

Based on the data provided, it appears that Leonardo DiCaprio is the audience favourite and critic favourite actor.

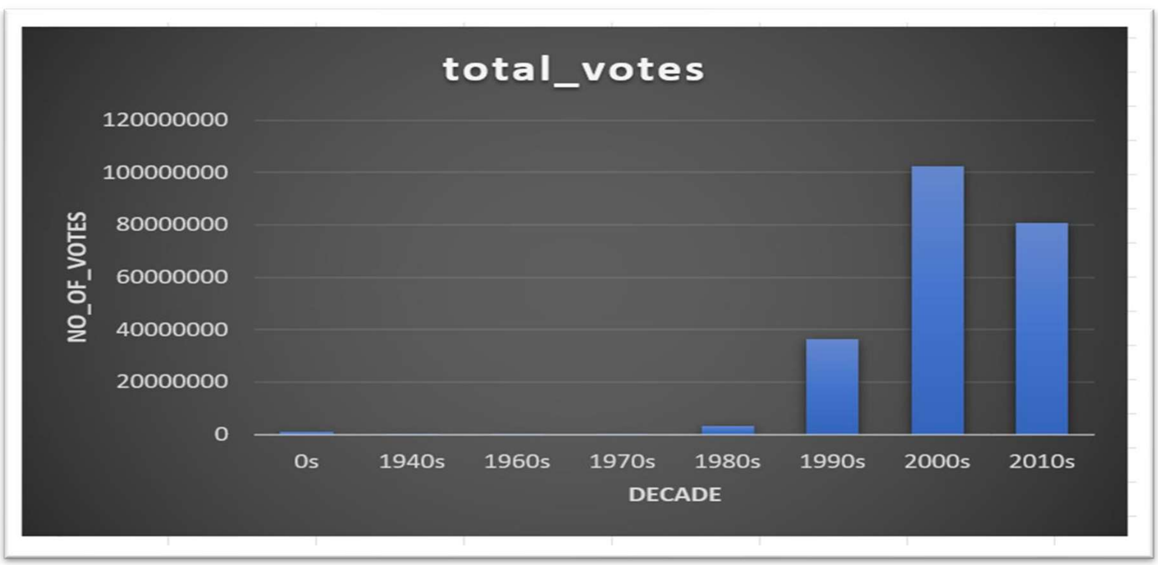
7) Bar Chart: Observe the change in number of voted users over decades using a bar chart. Create a column called decade which represents the decade to which every movie belongs to. For example, the title_year year 1923, 1925 should be stored as 1920s. Sort the column based on the column decade, group it by decade and find the sum of users voted in each decade.

Your task: Find the number of user votes per decade.

```sql
SELECT
 CONCAT(CONVERT( FLOOR(title_year / 10) * 10 , CHAR),
 's') AS decade,
 SUM(num_voted_users) AS total_votes
FROM
 imdb.imdb_movies
GROUP BY decade
ORDER BY decade
```

During the 2000s, there was a high number of users who voted for movies.

| 1 | decade | total_votes |
|---|---|---|
| 2 | 0s | 954256 |
| 3 | 1940s | 90360 |
| 4 | 1960s | 95264 |
| 5 | 1970s | 347092 |
| 6 | 1980s | 3057135 |
| 7 | 1990s | 36486955 |
| 8 | 2000s | 102375694 |
| 9 | 2010s | 80568844 |



- ANALYSIS:

the dataset provides insights into the popularity of Hollywood movies, the distinction between English and foreign language films in the top 250, the preferences of both audiences and critics towards certain actors, and the increase in the number of voters over time. The process of recording and analyzing the findings from the dataset can be complex, but it can lead to valuable insights and trends that can help understand the film industry and its audience.

- CONCLUSION:

this project provides valuable insights into the film industry by analyzing the IMDB ratings of movies, directors, and actors. The results can help filmmakers make informed decisions about future projects, such as which actors and directors to hire or which genres are currently popular. The project also provides practical experience in using Microsoft Excel, including the use of pivot tables and advanced functions. Overall, this project is a useful tool for anyone interested in understanding the trends and patterns in the film industry.