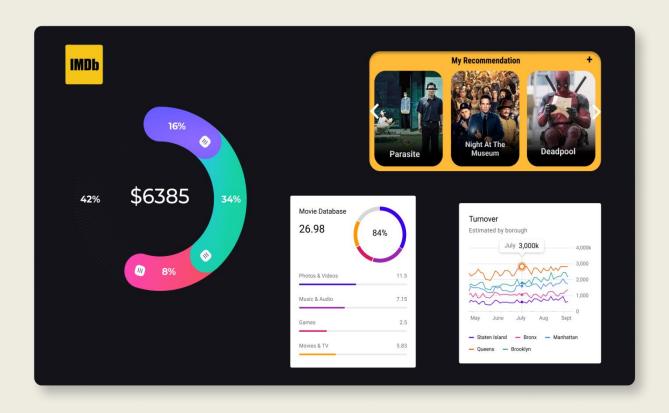
IMDB Movie Analysis

Final Project-1 (Project – 5)



Sowmiya R | Data Analytics | 20 July 2023

PROJECT DESCRIPTION

For this Final Project, they were providing me with dataset having various columns of different IMDB Movies. For this task, I will need to define a problem I want to shed some light on.

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?

Once defined a problem, clean the data as necessary, and using our Data Analysis skills to explore the data set and derive insights.

- 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

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.

I have been solving these questions by using the 5 why's approach analysis and making insights of using the excel data functions and mysql queries to get the desired results as well. Finding out the case studies as there are as follows....

CASE STUDIES

✓ Clean the data

(Dropping columns, removing null values, etc)

✓ Find the movies with the highest profit?

(observing the outliers using the appropriate chart type)

✓ Find IMDB Top 250

(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.)

✓ Find the best directors

(Top 10 director by highest mean of IMDB score)

✓ Find popular genres

(Perform this by step while performing previous steps)

✓ Find the critic-favorite and audience-favorite actors

(Finding the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean) and with decade of new data frame called df_by_decade.

APPROACH

First of all, I have downloaded the provided dataset and then load the dataset in the MS Excel, performing technique of data cleaning strategy Open dataset in the form of tables and perform few process in it, with that I have been used with a MySQL workbench to query certain case studies.

TECH-STACK USED

MySQL workbench and Microsoft Excel is used to answer the required questions with the analysis and insights.





INSIGHTS AND RESULTS OF ANALYSIS

1. Cleaning the data: This is one of the most important step 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

- ❖ First of all I have analyzed that with the provided dataset, there are 5043 rows and 28 columns and 1 header row are present there.
- ❖ After then, I looked a project description that related and required columns can only be used for the upcoming analysis so then I have been used a data cleaning strategy using MySQL workbench and excel functions to solve the data cleaning task.
- So I need only the columns and they are:
- director name
- num_critic_for_reviews
- gross
- genres
- actor_1_name
- movie title
- num_voted_users
- num_user_for_reviews

- language
- budget
- title_year
- imdb_score
- movie facebook likes
- ❖ After these are the only included then I left with a 13 columns. Then I go with a null values and null rows, this was get rid by using the excel function (press F5 − go to special − select blanks − press ok), it will show the blank rows and then (data − remove duplicates) it will remove the duplicates and save this sheet. Then load this currently saved dataset in the MySQL workbench, queried for removing the extra null values precisely.
- ❖ And save the removed null values dataset and again load in the excel to see how many entries has been left .
- ❖ By performing these functions in the excel I found that the total number of rows is then 3789.
- ❖ Below this, I have attached with cleaned dataset as a picture of removing duplicates and with the blanks, required columns, and set the blank(language) column been default as 'English'.
- This is my attached excel file.
 https://docs.google.com/spreadsheets/d/1Ha--2
 2RrVp2mce6NS6hj5dGjqWq4jOpiw/edit?usp=drive_link

director_name num_critic_for_rev	niews gross genres	actor_1_name	movie_title num	voted_users num_user_for_r	eviews language budget title_year imdb_score movie_facebook_likes
James Cameron Gore Verbinski	723 760505847 Action Adventure Fantasy Sci-Fi 302 309404152 Action Adventure Fantasy	CCH Pounder Johnny Depp	AvatarĂ Pirases of the Caribbean: At World's EndĀ Cross-si	886204 471220	3054 English 237000000 2009 7.9 33000 1238 English 300000000 2007 7.1 0
Sam Mendes Christopher Nolan Andrew Stanton	602 200074175 Action Adventure Thriller 813 448130642 Action Thriller 462 73058679 Action Adventure Sci-Fi	Christoph Waltz Tom Hardy Daryl Sabara	SpectreÄ The Dark Kright RisesÄ John CarterÄ	275868 1144337 212204	994 English 245000000 2015 6.8 85000 2701 English 250000000 2012 8.5 164000 738 English 263700000 2012 6.6 24000
Sam Raimi Nathan Greno	462 73058679 Action Adventure Sci-Fi 392 336530303 Action Adventure Romance 324 200807262 Adventure Animation Comedy Family Fantasy Musical Romance	J.K. Simmons	tim Cale Strigger State And School Carteria Spider-Man 3A Tangkedi A	383056	
Joss Whedon David Yates Zack Snyder	224 200007.02 Adventure Animaton Connecy Family Family Musical Romance 635 458991599 Action Adventure Sci-Fi 375 301956980 Adventure Family Family Mystery 673 330249062 Action Adventure Sci-Fi	Chris Hemsworth Alan Rickman Henry Cavill	langesure. Averigen: Age of Ultron À Harry Potter and the Half-Blood Prince Batman v Superman: Duwn of Justice Ä	462669 321795 371639	387 English 26000000 2010 7.8 29000 1117 English 250000000 2015 7.5 118000 973 English 250000000 2009 7.5 10000 3018 English 250000000 2016 6.9 197000
Zack Snyder Bryan Singer Marc Forster	6/3 330249062 Action Adventure Sci-Fi 434 200059408 Action Adventure Sci-Fi 403 168368427 Action Adventure	Kevin Spacey Giancarlo Giannini	Battman v Superminer: Daven of Justice/A Superman ReturnsiA Quantum of SolaceA	371639 240396 330784	3018 English 250000000 2016 6.9 197000 2367 English 209000000 2006 6.1 0 1243 English 20000000 2008 6.7 0
Gore Verbinski Gore Verbinski	313 423032628 Action Adventure Fantasy 450 89289910 Action Adventure Western	Johnny Depp Johnny Depp	Pirates of the Caribbean: Dead Man's ChestĂ Thu Lone RangerÄ Man of StestÄ	522040 181792	1832 English 22500000 2006 7.3 5000 711 English 21500000 2013 6.5 48000
Zack Snyder Andrew Adamson	733 291021565 Action Adventure Fantasy Sci-Fi 258 141614023 Action Adventure Family Fantasy 738 573279647 Action Adventure Sci-Fi	Henry Cavill Peter Dinklage		548573 149922	438 English 225000000 2008 6.6 0
Joss Whedon Rob Marshall Barry Sonnenfeld	703 623279547 Action Adventure Sci-Fi 448 241053875 Action Adventure Fantasy 451 179020854 Action Adventure Comedy Family Fantasy Sci-Fi	Chris Hemsworth Johnny Depp Will Smith	The Avengersk Pirates of the Caribbean: On Stranger Tidesk Men in Black JR	995415 370704 268154	1722 English 220000000 2012 8.1 123000 494 English 250000000 2011 6.7 58000 494 English 225000000 2012 6.8 40000
Peter Jackson Marc Webb	4-51 17902053-402001/400401016 [Cambol/Farmsy Farmsy St3-F1 422 2555108370 Adventure Farmsy 599 262030663 Action Adventure Fartasy 343 105219735 Action Adventure Drama History		nours in bucks 34. The Hobbit: The Battle of the Five ArmiesÂ The Annazing Spider-ManÂ		802 Enalish 250000000 2014 7.5 65000
Ridley Scott Peter Jackson		Emma Stone Mark Addy Aidan Turner	Robin HoodA The Hobbit: The Desolation of SmaueÂ	451803 211765 483540	546 English 200000000 2010 6.7 17000 951 Faulish 235000000 2013 7.9 83000
Chris Weitz Peter Jackson	251 70083519 Adventure Family Fantasy 446 218051260 Action Adventure Drama Romance	Christopher Lee Naomi Watts	The Golden CompassÅ King KongÅ	149019 316018	666 English 18000000 2007 6.1 0 2618 English 20700000 2005 7.2 0
James Cameron Anthony Russo Pater Berg	315 658672302 Drama Romance 516 407197282 Action Adventure Sci-Fi 377 65173160 Action Adventure Sci-Fi Thriller	Leonardo DiCaprio Robert Downey Jr. Liam Neeson	TitanicĂ Captain America: Civil WarĀ BattleshipĀ	793059 272670 202382	2528 English 200000000 1997 7.7 26000 1002 English 250000000 2016 8.2 72000 751 English 26000000 2012 5.9 44000
Colin Trevorrow Sam Mendes	644 652177271 Action Adventure Sci-Fi Thriller	Bryce Dallas Howard	Jurassic WorldÅ	418214 522030	1290 English 15000000 2015 7 150000 1498 English 20000000 2012 7.8 80000
Sam Raimi Shane Black	300 373377893 Action Adventure Fantasy Romance 608 408992272 Action Adventure Sci-Fi	Robert Downey Jr.	Spider-Man 2Å Iron Man 3Å	411164 557489	1498 English 200000000 2012 7.8 80000 1903 English 200000000 2004 7.3 0 1187 English 200000000 2013 7.2 95000
Tim Burton Brett Ratner Dan Scanlon	451 334185206 Adventure Family Fantasy 334 234360014 Action Adventure Family Sci-Fi Thriller 376 268488329 Adventure Animation Cornedy Family Fantasy	Johnny Depp Hugh Jackman	Alice in Wonderlandå X-Men: The Last Standå	306320 383427 235025	736 English 200000000 2010 6.5 24000 1912 English 2100000000 2006 6.8 0 265 English 200000000 2013 7.3 44000
Michael Bay Michael Bay	306 402076689 Action Adventure Sci-Fi 378 245428137 Action Adventure Sci-Fi	Steve Buscemi Glenn Morshower Bingbing Li	Monsters Universityå Transformers: Revenge of the FallenÅ Transformers: Age of ExtinctionÅ	323207 242420	1439 English 20000000 2009 6 0 918 English 21000000 2014 5.7 56000
Sam Raimi Marc Webb	525 234903076 Adventure Family Fantasy 495 202853933 Action Adventure Fantasy Sci-Fi 499 172051787 Action Adventure Sci-Fi	Tim Holmes Emma Stone	Oz the Great and PowerfulĂ The Amazing Spider-Man 2Ă	175409 321227	511 English 215000000 2013 6.4 60000 1067 English 20000000 2014 6.7 41000
Joseph Kosinski John Lasseter	304 191450875 Adventure Animation Comedy Family Sport	Jeff Bridges Joe Mantegna	TRON: LegacyĀ Cars 2Ā	264183 101178	665 English 17000000 2010 6.8 30000 283 English 20000000 2011 6.3 10000
Martin Campbell Lee Unkrich McG	436 116593191 Action Adventure Sci-Fi 433 414984497 Adventure Animation Comedy Family Fantasy 422 125320003 Action Edwards Sci-Fi	Ryan Reynolds Tom Hanks Christian Bale	Green LanternÅ Toy Story 3Å Terminator SalvationÅ	223393 544884 286095	S50 English 200000000 2011 5.6 24000 733 English 200000000 2010 8.3 30000 974 English 200000000 2009 6.6 0 657 English 190000000 2015 7.2 94000
James Wan Marc Forster	422 125320003 Action Adventure Sci-Fi 424 350034110 Action Crime Thriller 654 202351611 Action Adventure Horror Sci-Fi Thriller	Jason Statham	Terminator SalvationĀ Furious 7Ā World War 2Ā	286095 278232 465019	934 English 200000000 2009 6.6 0 657 English 190000000 2015 7.2 94000 995 English 190000000 2013 7 129000
Bryan Singer J.J. Abrams	539 233914986 Action Adventure Fantasy Sci-Fi Thriller 590 228756232 Action Adventure Sci-Fi	Jennifer Lawrence Benedict Cumberbatch	X-Miner Darknesså Star Trek Into Darknesså	514125 395573	995 English 190000000 2013 7 129000 752 English 200000000 2014 8 82000 1272 English 190000000 2013 7.8 92000
Bryan Singer Baz Luhrmann	338 65171860 Adventure Fantasy 490 144812796 Drama Romance	Eddie Marsan Leonardo DiCaprio	Jack the Giant SlayerA The Great GatsbyA	106416 362912	205 English 195000000 2013 6.5 22000 753 English 105000000 2013 7.3 115000
Mike Newell Guillermo del Toro Michael Bay	306 90755643 Action Adventure Fantasy Romance 575 101785482 Action Adventure Sci-Fi 428 35238779 Action Adventure Sci-Fi	Jake Gyllenhaal Charlie Hunnam Glenn Morshower	Prince of Persia: The Sands of TimeÄ Pacific RimÄ Transformers: Dark of the MoonÄ	222403 381148 326180	453 English 20000000 2010 6.6 23000 1306 English 190000000 2013 7 83000 899 English 15000000 2011 6.3 46000
Steven Spielberg Mark Andrews	470 317011114 Action Adventure Fantasy 488 237282182 Adventure Animation Comedy Family Fantasy	Harrison Ford Kelly Macdonald	Indiana Jones and the Kingdom of the Crystal SkullÄ BraveÄ	333847 273556	899 English 19500000 2011 6.3 46000 2054 English 18500000 2008 6.2 5000 418 English 18500000 2012 7.2 39000
Justin Lin Andrew Stanton		Sofia Boutella John Ratzenberser	Star Trek BeyondÅ WALLÄ EÅ	53607 718837 121084	432 English 18500000 2016 7.5 30000 1043 English 18000000 2018 8.4 16000
Brett Ratner Roland Emmerich	222 150-000020 ACCION/ADMINITURE (251-1) Imment 421 22380-6889 Adventure Animation Family Sci-Fi 162 140080850 Action Comady Crime Thriller 367 166112167 Action Adventure Sci-Fi	Tzi Ma Oliver Platt	Rush Hour 3A 2012Â	283418	221 English 14000000 2007 6.2 0 1055 English 20000000 2009 5.8 13000
Robert Zemeckis Lana Wachowski David Yates	240 137850096 Animation Drama Family Fantasy 384 47375327 Action Adventure Sci-Fi 248 124051759 Action Adventure Drama Romance	Robin Wright Channing Tatum Christoph Waltz	A Christmas CarolĂ Jupiter AscendingĂ The Legend of TarzanĀ	72809 139593 42372	249 English 20000000 2009 6.8 0 720 English 17600000 2015 5.4 44000 239 English 8000000 2016 6.6 29000
David Yates Andrew Adamson Bryan Singer	284 291709845 Adventure Family Fantasy 396 154985087 Action Adventure Sci-Fi	Christoph Waltz Jim Broadbent Jennifer Lawrence	The Legend of TarzanA The Chronicles of Namia: The Lion, the Witch and the WardrobeÂ X-Marr. ApocalypseA	42372 286506 148379	239 English 180000000 2016 6.6 29000 1463 English 180000000 2005 6.9 0 622 English 178000000 2016 7.3 54000
Christopher Nolan Pete Docter		Christian Bale John Ratzenberger	A-nontri Apocarypsala The Dark KnightÄ UpÄ	1676169 665575	4667 English 18500000 2008 9 37000 704 English 17500000 2009 8.3 27000
Rob Letterman Jon Favreau	408 202979556 Adventure Animation Comedy Family 219 198332128 Action Adventure Animation Comedy Family Sci-Fi 486 318298180 Action Adventure Sci-Fi	Amy Poehler Robert Downey Jr.	Monsters vs. Alienså Iron Manå	114553 696338	187 English 17500000 2009 6.5 0 1055 English 14000000 2008 7.9 10000
Martin Scorsese Barry Sonnenfeld	682 73820094 Adventure Drama Family Mystery 85 113745408 Action Comedy Sci-Fi Western 264 102176165 Action Adventure Fantasy Horror Thriller	ChloÃe Grace Moretz Will Smith		245333 129601 117927	678 English 17000000 2011 7.5 42000 648 English 17000000 1999 4.8 0
Rob Cohen David Ayer Tom Shadyac		Jet Li Will Smith Limmy Rennett	region Wild Wild Westå The Mummy: Tomb of the Dragon Emperorå Sicrido Squadå Fune Aminishrida		501 English 145000000 2008 5.2 0 971 English 175000000 2016 6.9 80000
Doug Liman Kevin Reynolds	186 100289690 Comedy [Family [Family] 185 100289691 Action Adventure Sci-Fi 191 88346220 Action Adventure Sci-Fi Thriller	Jimmy Bennett Tom Cruise Jeanne Tripplehorn	Evan Almightyå Edge of Tomorrowå Waterworldå	115099 431620 144337	257 English 175000000 2007 5.4 2000 741 English 178000000 2014 7.9 77000 309 English 178000000 1995 6.1 0
Stephen Sommers Pete Docter	91 abavaczu Action (Adventure) Sci-Fi Thrillier 536 356454367 Adventure Adventure Sci-Fi Thrillier 330 366454367 Adventure Janimation Commody Drama Family Fantasy 370 362645141 Adventure Drama Family Fantasy	Joseph Gordon-Levitt Amy Poehler	G.I. Joe: The Rise of CobraÂ Inside OutÂ	174578 345198	534 English 175000000 2009 5.8 0 773 English 175000000 2015 8.3 118000
Jon Favreau		Scarlett Johansson Robert Downey Jr.	The Jungle BookÂ	106072	
Rupert Sanders Robert Stromberg	416 15511815 Action Adventure Frama Fantasy 401 241407328 Action Adventure Family Fantasy Romance	Chris Hemsworth Angelina Jolie Pitt	rom Mari A. Snow White and the Huntsman A. Maleficent A.	228554 252257	723 English 200000000 2010 7 18000 710 English 170000000 2012 6.1 53000 634 English 180000000 2014 7 89000
Matt Reeves Carl Rinsch	521 208543795 Action Adventure Drama Sci-Fi 218 38297305 Action Adventure Drama Fantasy 576 259746958 Action Adventure Sci-Fi	Gary Oldman Keanu Reeves Scarlett Johansson	Dawn of the Planet of the ApesiÄ 47 RoninÄ Captain America: The Winter SoldierÄ Shrek Forever AfterÄ	317542 116994 496749	620 English 170000000 2014 7.6 45000 324 English 175000000 2013 6.3 35000 742 English 170000000 2014 7.8 55000
Anthony Russo Mike Mitchell Brad Bird		Jon Hamm Judy Greer			
Don Hall Rich Moore	443 93417865 Action Adventure Family Mystery Sci-Fi 384 222487711 Action Adventure Animation Comedy Drama Family Sci-Fi 377 189412677 Adventure Animation Comedy Family Sci-Fi	Damon Wayans Ir. Jack McBrayer	TomorrowlandÅ Big Hero 6Å Wreck-it RalphÅ	128306 279093 272534	433 English 16500000 2014 7.9 41000 345 English 16500000 2012 7.8 40000
Robert Zemeckis Roland Emmerich	188 665426 Adventure Animation Family Fantasy 286 102315545 Action Adventure Sci-Fi	Tom Hanks Vivica A. Fox	The Polar ExpressÂ Independence Day: ResurgenceÂ	120798 58137	444 English 165000000 2004 6.6 10000 520 English 165000000 2016 5.5 67000
Dean DeBlois Jonathan Mostow	288 217387997 Adventure Animation Family Fantasy	Gerard Butler Nick Stahl	How to Train Your Dragon A	485430 305340	
James Gunn Christopher Nolan Christopher Nolan	103 33330096 Action Adventure Sci-Fi 712 187991439 Adventure Drama Sci-Fi 642 202588651 Action Adventure Sci-Fi Thriller	Bradley Cooper Matthew McConaughey Leonardo DiCaorio	Guardians of the GalaxyÁ InterstellarÁ IncestinÁ	682155 928227 1468200	1676 English 20000000 2003 6.4 0 1097 English 170000000 2014 8.1 96000 2715 English 165000000 2014 8.6 349000 2803 English 165000000 2010 8.8 175000
Christopher Notan Peter Jackson Rob Cohen	642 29236853 Action Adventure Sci +1 Thrillier 643 303001229 Adventure Famasy 187 144512310 Action Crime Thriller	Aidan Turner Paul Walker	inceptionA The Hobbit: An Unexpected JourneyÂ The Fast and the FuriousÂ	1468200 637246 272223	2803 English 160000000 2010 8.8 175000 1367 English 180000000 2012 7.9 166000 988 English 30000000 2001 6.7 14000
David Fincher Matthew Vaughn		Brad Pitt	The Curious Case of Berjamin ButtonA X-Men: First ClassA	459346 518537	
Francis Lawrence Jon Turteltaub	500 146405371 Action Adventure Sci-Fi 389 281666058 Adventure Sci-Fi 235 63143812 Action Adventure Family Fantasy	Jennifer Lawrence Jennifer Lawrence Nicolas Cage	The Hunger Games: Mockingjay - Part 2A The Sorcerer's ApprenticeA	166137 124185	383 English 16000000 2015 6.6 38000 238 English 15000000 2010 6.1 11000
Wolfgang Petersen James Bobin	231 60655503 Action Adventure Drama Thriller 218 76846624 Adventure Family Fantasy	Jimmy Bennett Johnny Depp	PoseidonÂ Alice Through the Looking GlassÂ	82380 21352	629 English 16000000 2006 5.6 0 131 English 17000000 2016 6.4 30000
Chris Miller Duncan Jones	227 320706665 Adventure Animation Comedy Family Fantasy 275 46978995 Action Adventure Fantasy	Justin Timberlake Dominic Cooper J.K. Simmons	Shrek the ThirdÂ WarcraftÂ	211971 111609	326 English 16000000 2007 6.1 0 781 English 16000000 2016 7.3 89000 887 English 15500000 2015 6.6 82000
Alan Taylor Michael Apted Michael Bay	174 89732035 Action Adventure Sci-Fi 228 104383624 Adventure Family Famasy 191 198339855 Action Drama History Romance War	J.K. Simmons Bruce Spence Jennifer Garner	Terminator Genisysk The Chronicles of Namia: The Voyage of the Dawn Treaderk Pearl Habror k	188457 106446 254111	227 English 155000000 2010 6.3 11000
Michael Bay Oliver Stone	191 190539853 Action (orange instory) (morantice) (war 396 318759914 Action (Adventure) (Sci-Fi 248 34293771 Action (Adventure) (Biography (Drama (History Romance) War	Zack Ward Anthony Hopkins	rvan naroorik Transformersk AlexanderÅ	513158 138863	1999 English 140000000 2001 6.1 0 1782 English 150000000 2007 7.1 8000 1390 English 155000000 2004 5.5 0
David Yates Mike Newell	329 292000866 Adventure Family Fantasy Mystery 295 289994397 Adventure Family Fantasy Mystery	Robert Pattinson Robert Pattinson	Harry Potter and the Order of the PhoenixA Harry Potter and the Goblet of FireA	355137 385670	1108 English 15000000 2007 7.5 0 1896 English 15000000 2005 7.6 0
Peter Berg Francis Lawrence	318 227946274 Action Drama 323 256386216 Drama Horror Sci-Fi	Will Smith Will Smith	HancockÂ I Am LegendÂ	343648 530870	590 English 15000000 2008 6.4 0 1413 English 15000000 2007 7.2 11000
Tim Burton Brad Bird	276 206456431 Adventure Cornedy Family Fantasy 318 206435493 Animation Cornedy Family Fantasy	Johnny Depp Janeane Garofalo	Charlie and the Chocolate FactoryA RatatouilleA	320284 473887	1361 English 15000000 2005 6.7 0 626 English 15000000 2007 8 10000
Christopher Nolan Eric Darnell Shawn Levy	478 205343774 Action Adventure 167 179982968 Action Adventure Animation Comedy Family 18 17234731 Adventure Compdy Employ Compdy Family	Christian Bale Bernie Mac Robin Williams	Batman Begins.Å Madagascar: Escape 2 AfricaÄ Night at the Museum: Battis of the SmithsonianÄ	980946 146019 130272	2665 English 150000000 2005 8.3 15000 10 English 150000000 2008 6.7 0 200 English 150000000 2009 5.9 2000 641 English 150000000 2009 6.7 0
Gavin Hood Lana Wachowski	19983-00 Accomposition printing printin	Hugh Jackman Essie Davis	X-Men Origins: WolvenneA The Matrix RevolutionsÂ	361924 364948	
Chris Buck Lana Wachowski	406 400736600 Adventure Animation Comedy Family Fantasy Musical 275 281492479 Action Sci-Fi	Josh Gad Steve Bastoni	FrozenÂ The Matrix ReloadedÂ	421658 421818	904 English 150000000 2013 7.6 58000 2789 English 150000000 2003 7.2 0
Alan Taylor George Miller	486 206360018 Action Adventure Fantasy 739 153629485 Action Adventure Sci-Fi Thriller	Chris Hemsworth Tom Hardy	Thor: The Dark WorldA Mad Mao: Fury RoadÂ	414070 552503	532 English 17000000 2013 7.1 63000 1588 English 15000000 2015 8.1 191000
Ron Howard Kenneth Branagh Buron Howard	298 133375846 Mystery Thriller 516 181015141 Action Adventure Fantasy 275 114078379 adventure Jerimation Connecty Drama Family	Tom Hanks Chris Hemsworth ChloÃn Grace Moretz	Angels & DemonsĂ ThorĂ BoltĂ	207839 536314 146766	435 English 150000000 2009 6.7 0 738 English 150000000 2011 7 63000 178 English 150000000 2008 6.9 0
Hoyt Yeatman Jonathan Liebesman	245 114035579 Adventure Animaton Comedy Cramary Family F	Kelli Garner Liam Neeson	G-Force A Wrath of the Titans A	33042 152826	90 English 15000000 2009 5.1 0 253 English 15000000 2012 5.8 19000
Tim Burton Christopher McQuarrie	526 79711678 Comedy Fantasy Horror 465 195000874 Action Adventure Thriller 357 61937495 Drama Fantasy Horror Thriller	Johnny Depp Tom Cruise	Dark Shadowsk Mission: impossible - Rogue NationÅ The WolfmanÄ	199039 232187	479 English 10000000 2012 6.2 82000 440 English 15000000 2015 7.4 47000
Joe Johnston Steve Hickner		Anthony Hopkins Matthew Broderick		89442 105902	452 English 150000000 2010 5.8 0 206 English 15000000 2007 6.2 0
Jennifer Yuh Nelson M. Night Shyamalan J.J. Abrams	284 165230261 Action Adventure Animation Comedy Family 280 131564731 Action Adventure Family Fantasy 310 133382309 Action Adventure Timiller	Angelina Jolie Pitt Seychelle Gabriel Philip Seymour Hoffman	Kung Fu Panda 2A The Last AirbanderA Mission: Impossible IIIA	182718 118951 256695	209 English 15000000 2011 7.3 20000 1382 English 15000000 2010 4.2 18000 871 English 15000000 2006 6.9 0
J.J. Abrams Roland Emmerich Simon Wells	310 133382309 Action Adventure Thriller 339 73103784 Action Drama Thriller 132 21379315 Action Adventure Animation Comedy Family Sci-Fi	Philip Seymour Hoffman Channing Tatum Elisabeth Harnois	Mission: Imposible IIIA White House Downä Mars Needs Mornsä	256695 164238 17590	434 English 150000000 2013 6.4 33000
Simon were David Bowers Joe Wright Rob Minkoff	135 64459316 Adventure Animation Comedy Family	Hugh Jackman	Flushed Awayâ Panâ	85086 39956	122 English 149000000 2006 6.7 0
Wolfgang Petersen	196 111505642 Adventure Animation Comedy Family Sci-Fi 220 133228348 Adventure	Ty Burrell Brad Pitt	Mr. Peabody & ShermanÄ TroyÄ	47900 381672	
Eric Darnell Lee Tamahori Paul Feig	211 216366733 Adventure Animation Comedy Family 264 160201106 Action Adventure Thriller 464 118039659 Action Comedy Fantasy Sci-Fi	Jada Pinkett Smith Toby Stephens Ed Begley Jr.	iroyu Madagascar 3: Europe's Most Wantedå Die Another Daylå Ghostbusterså	119213 169914 69757	154 English 145000000 2012 6.9 17000 1185 English 142000000 2002 6.1 0 1211 English 144000000 2016 5.5 62000
Michael Bay		Bruce Willis Will Smith	GhostauseriA ArmageddonÂ Men in Black BÂ	322395	1171 English 140000000 1998 6.6 11000
Barry Sonnenfeld Robert Zemeckis Alessandro Carloni	208 190418803 Action Adventure Comedy Family Fantasy Mystery Sci-Fi 287 82161969 Action Adventure Animation Fantasy 210 143523463 Action Adventure Animation Comedy Family	Robin Wright J.K. Simmons	Becwulf A	270207 142440 64322	505 English 15000000 2007 6.3 3000 145 English 14500000 2016 7.2 24000
Brad Bird Peter Ramsey	432 209364921 Action Adventure Thriller 256 103400592 Adventure Animation Family Fantasy	Tom Cruise Hugh Jackman	Mission: Impossible - Ghost ProtocolĀ Rise of the GuardiansĀ	365104 123553	512 English 145000000 2011 7.4 35000 174 English 145000000 2012 7.3 25000
Dean Parisot Edward Zwick Ridley Scott	135 110332737 Comedy Crime 190 111110575 Action Drama History War 314 65007045 Action Adventure Drama	John Michael Higgins Tom Cruise Christian Bale	Fun with Dick and JaneA The Lact Samurai	110788 317166	258 English 10000000 2005 6.1 2000 928 English 14000000 2003 7.7 0
J.J. Abrams Sam Raimi	314 65007045 Action Adventure Drama 518 257704099 Action Adventure Sci-Fi 291 403706375 Action Adventure Fantasy Romance	Christian Bale Chris Hemsworth J.K. Simmons	Exodus: Gods and KingsÅ Star TreikÅ Spider-ManÅ	128682 504419 544665	1559 English 150000000 2009 8 19000 2012 English 139000000 2002 7.3 5000
Dean DeBlois Alex Proyas	292 176997107 Action Adventure Animation Comedy Family Fantasy 184 31141074 Action Adventure Fantasy	Gerard Butler Gerard Butler	How to Train Your Dragon 2Å Gods of EgyptÅ	221128 51892	343 English 14500000 2014 7.9 46000 273 English 14000000 2016 5.5 24000
Rob Cohen Zack Snyder	145 31704416 Action Adventure Sci-Fi Thriller 451 107503316 Action Drama Mystery Sci-Fi	Sam Shepard Matt Frewer	Stealth Watchmin A Lethal Weapon 4A	45455 392474 127497	388 English 135000000 2005 5 0 1279 English 130000000 2009 7.7 18000
Richard Donner Ang Lee	267 132122995 Action Sci-Fi	Jet Li Kevin Bankin	Lethal Weapon 4Å Hulik G I Ing-Betalistion B	212106	
Jon M. Chu Breck Eisner Hironobu Sakaguchi	351 122512052 Action Adventure Sci-Fi Thriller 163 68642452 Action Adventure Comedy Thriller 166 32131830 Action Adventure Animation Fantacy Romance Sci-Fi	Channing Tatum Matthew McConaughey Steve Buscemi	G. L. Iore: Retalisation A Sahara A Final Fantasy: The Spirits Within A	146352 77673 72259	288 English 13000000 2013 5.8 42000 463 English 130000000 2005 6 0 788 English 137000000 2001 6.4 0
Joe Johnston Michael Apted	E10 176636916 Action Advanture Sci Ei	Chris Evans Colin Salmon	Contain America: The First Assessed	508818 157519	679 English 140000000 2011 6.9 46000
Peter Weir Bill Condon	197 1263030560 Action Adventure Thriller 244 93926336 Action Adventure Drama History War 322 292298923 Adventure Drama Fantasy Romance	James D'Arcy Robert Pottiones	The World Is Not Enough! An Asset See See See See See See See See See S	168207	
George Miller Louis Leterrier	156 63992328 Animation Comedy Family Musical 354 134518390 Action Adventure Sci-Fi	Robin Williams Ty Burrell	The Twilight Saga: Breaking Dawn - Part 2Å Happy Feet 2Å The Incredible HulkÅ	32399 326286	79 English 135000000 2011 5.9 0 643 English 15000000 2008 6.8 0
Steven Spielberg Alejandro G. IĀzĀļrritu	252 52792307 Adventure Family Fantasy 556 183635922 Adventure Drama Thriller Western	Mark Rylance Leonardo DiCaprio	The Broan	12572 406020	106 English 14000000 2016 6.8 27000 1188 English 13500000 2015 8.1 190000
David Soren Gore Verbinski Eric Darnell	166 83024900 Adventure Animation Comedy Family Sport 362 123207124 Adventure Animation Comedy Family Western 53 8346920 Adventure Animation Comedy Family	Ryan Reynolds Johnny Depp Benedict Cumberbatch	TurboÅ RangoÅ Peneuron of Madazascar Å	62424 183208 60230	90 English 135000000 2013 6.5 13000 337 English 135000000 2011 7.2 26000 1138 English 132000000 2014 6.7 11000
Paul Greengrass Mark Osborne	329 227137090 Action Mystery Thriller 266 215395021 Action Adventure Animation Comedy Family	Matt Damon Angelina Jolie Pitt	Penguins of Madagascar Ä The Bourne Ultimatum Ä Kung für Panda Ä	491077 307029	820 English 11000000 2007 8.1 0 360 English 13000000 2008 7.6 6000
Peyton Reed Francis Lawrence	15393634 Action Adventure Comedy Sci-Fi	Judy Greer Jennifer Lawrence Jim Parsons	Ant-Manā The Hunger Games: Catching Fireā	313866 498397 70121	549 English 130000000 2015 7.4 61000 206 English 130000000 2013 7.6 82000
Tim Johnson			HomeA War of the Worldså		214 English 13500000 2015 6.7 26000 2741 English 13200000 2005 6.5 0
Michael Bay Chris Miler	4 133306524 Action [Comedy Crime] Thriller 46 149234747 Action [Adventure [Animation Comedy Family Fantasy 149214747 Action [Adventure [Animation Comedy Family Fantasy 149214747 Action [Adventure [Animation Comedy Family Fantasy 149214747 Action [Adventure [Animation Comedy Family Fantasy 1492147474 Action Adventure [Animation Comedy Family Fantasy 1492147474 Action Adventure [Animation Comedy Family Fantasy 14921474 Action Adventure [Animation Comedy Family Fantasy 14921474 Action Comedy Family Fantasy 14921474 Action Adventure [Animation Comedy Family Fantasy 14921474 Action Action Comedy Family Fantasy 14921474 Action Action Comedy Family Fantasy 14921474 Action Comedy Family Family Fantasy 14921474 Action Action Comedy Family Family Family Fantasy 14921474 Action Action Comedy Family	Will Smith Salma Hayek	Bad Boys IIĀ Pussi in BootsĀ Sulsā	178126 114287	2741 English 132000000 2005 6.5 0 511 English 130000000 2003 6.6 0 137 English 130000000 2011 6.7 16000
Phillip Noyce Darren Aronofsky Steven Snielbere	330 118311368 Action Crime Mystery Thriller 434 101160529 Action Adventure Drama 440 77564037 Action Adventure Eramily Mystery	Angelina Jolie Pitt Anthony Hopkins Toby Jones	Saità Noahà The Adventures of Tintinà	245621 200022 177383	514 English 110000000 2010 6.4 23000 1240 English 125000000 2014 5.8 71000 447 English 13500000 2011 7.4 44000
Steven Spielberg Alfonso CuarĀ ³ n Baz Luhrmann	274 249358727 Adventure Family Fantasy Mystery 245 49551662 Adventure Drama Romance War	Daniel Radcliffe Essie Davis	Harry Potter and the Prisoner of AzkabanĀ AustraliaĀ	382255 102338	1504 English 130000000 2004 7.8 11000
M. Night Shyamalan Eric Leishton	349 60522097 Action Adventure Sci-Fi 145 137748063 Adventure Animation Family Thriller	Will Smith Alfre Woodard	After EarthÂ DinosaurÂ	158720 38438	744 English 13000000 2013 4.9 37000 241 English 127500000 2000 6.5
Shawn Levy Tom McGrath	154 113733726 Adventure Comedy Family Fantasy 233 148337537 Action Animation Comedy Family Sci-Fi 258 317557891 Adventure Family Fantasy	Robin Williams J.K. Simmons Daniel Radciffe	Night at the Museum: Secret of the TombÅ MegamindÄ Harry Potter and the Sorcerer's StoneÅ	67223 172754 444683	126 English 127000000 2014 6.2 11000 187 English 130000000 2010 7.3 13000
Chris Columbus Robert Schwentike		Ryan Reynolds		91640	15/1 English 125000000 2001 7.5 16000 210 English 130000000 2013 5.6 20000
Gore Verbinski Francis Lawrence Ron Howard	271 305388685 Action Adventure Fantasy 403 337103873 Adventure Sci-Fi[Thriller 294 217536138 Mystery Thriller	Johnny Depp Jennifer Lawrence Tom Hanks	Pirates of the Caribbean: The Curse of the Black PearlÄ The Hunger Games: Mockingjay - Part 1Ä The Da Viric GodeÄ	809474 305008 314253	591 English 125000000 2014 6.7 52000 1966 English 125000000 2006 6.6 0
Carlos Saldanha Bryan Singer	159 131536019 Adventure Animation Comedy Family Musical 289 214948780 Action Adventure Fantasy Sci-Fi Thriller	Miguel Ferrer Hugh Jackman	Rio 2Ā X-Men 2Ā	58498 405973	99 English 10300000 2014 6.4 0 1055 English 11000000 2003 7.5 0
Justin Lin Guy Ritchie	342 209805005 Action Crime Thriller	Paul Walker Robert Downey Jr.	Fast FiveÂ Sherlock Holmes: A Game of ShadowsÂ	284792 338635	366 English 125000000 2011 7.3 54000 412 English 125000000 2011 7.5 39000
Louis Leterrier Paul Verhoeven	344 163192114 Action Adventure Fantasy 196 119412921 Action Sci-Fi	Liam Neeson Ronny Cox	Clash of the TitansÄ Total RecallÄ	229679 240241	391 English 65000000 1990 7.5 0
John McTiernan Tony Giroy	85 32694788 Action Adventure History 436 113165635 Action Adventure Thriller	Tony Curran Jeremy Renner	The 13th WarriorĀ The Bourne LegacyĀ	101411 229823	546 English 8500000 1999 6.6 0 504 English 12500000 2012 6.7 31000

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?

- ❖ Now as the values in gross and budget columns are very large so I have used a excel to get rid of these large values is not readable that (go to format cells – select custom – type #,,"million" – then select ok) it will give us the readable values , i.e 274M.
- ❖ Next , we create a column called 'profit' as gross-budget difference in the MySQL workbench to get the desired results. The query can be as follows:

SELECT director_name,
 movie_title,
 ROUND(budget, 0) AS budget,
 ROUND(gross - budget, 0) AS profit
FROM
 moviesp.cleanedm
ORDER BY gross - budget DESC
LIMIT 10;

With this query, I have found out the result of movie_title are as follows:

AvatarÂ

Jurassic WorldÂ

TitanicÂ

Star Wars: Episode IV - A New HopeÂ

E.T. the Extra-TerrestrialÂ

The AvengersÂ
The Lion KingÂ

Star Wars: Episode I - The Phantom MenaceÂ

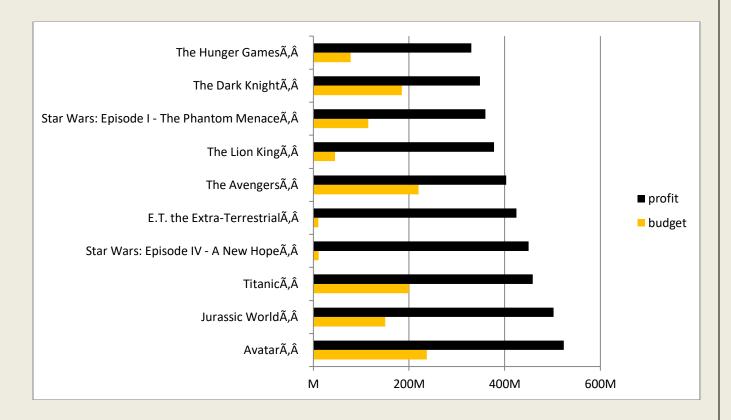
The Dark KnightÂ
The Hunger GamesÂ

director_name	movie_title	budget	profit
James Cameron	AvatarÃ,Â	237M	524M
Colin Trevorrow	Jurassic WorldÃ,Â	150M	502M
James Cameron	TitanicÃ,Â	200M	459M
George Lucas	Star Wars: Episode IV - A New HopeÃ,Â	11M	450M
Steven Spielberg	E.T. the Extra-TerrestrialÃ,Â	11M	424M
Joss Whedon	The AvengersÃ,Â	220M	403M
Roger Allers	The Lion KingÃ,Â	45M	378M
George Lucas	Star Wars: Episode I - The Phantom MenaceÃ,Â	115M	360M
Christopher Nolan	The Dark KnightÃ,Â	185M	348M
Gary Ross	The Hunger GamesÃ,Â	78M	330M

This the result of above query, that we have found that Movies with Highest profit. And the insights of this table are:

This is my attached excel file:

https://docs.google.com/spreadsheets/d/1vU6T3pqE0nJI688PbSpdxeA8LBAjKr0R/edit?usp=drive_link



From the above bar chart represents that the highest profit of respected movies have been found to be analyzed and gave insights.

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 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

a) For the first part we should look for all the movies which has higher imdb scores user votes. So that I had solved this question by query with the user votes of higher than 30000 are as follows:

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 moviesp.cleanedm
WHERE num_voted_users > 30000
LIMIT 250;

IMDB Top 250 Movies					
ranking	imdb_score	IMDb_Top_250	language		
1	9.3	The Shawshank RedemptionÃ,Â	English		
2	9	The Dark KnightÃ,Â	English		
3	9	The Godfather: Part IIÃ,Â	English		
4	8.9	The Lord of the Rings: The Return of the KingÃ,Â	English		
5	8.9	Schindler's ListÃ,Â	English		
6	8.8	InceptionÃ,Â	English		
7	8.8	Fight ClubÃ,Â	English		
8	8.8	Forrest GumpÃ,Â	English		
9	8.8	The Lord of the Rings: The Fellowship of the RingÃ,Â	English		

10	8.8	Star Wars: Episode V - The Empire Strikes BackÃ,Â	English
11	8.7	The MatrixÃ,Â	English
12	8.7	The Lord of the Rings: The Two TowersÃ,Â	English
13	8.7	Star Wars: Episode IV - A New HopeÃ,Â	English
14	8.7	GoodfellasÃ,Â	English
15	8.6	Se7enÃ,Â	English
16	8.6	InterstellarÃ,Â	English
17	8.6	The Silence of the LambsÃ,Â	English
18	8.6	Saving Private RyanÃ,Â	English
19	8.6	American History XÃ,Â	English
20	8.6	Spirited AwayÃ,Â	Japanese
21	8.5	The Dark Knight RisesÃ,Â	English
22	8.5	GladiatorÃ,Â	English
23	8.5	Django UnchainedÃ,Â	English
24	8.5	The DepartedÃ,Â	English
25	8.5	The PrestigeÃ,Â	English
26	8.5	The Green MileÃ,Â	English
27	8.5	Terminator 2: Judgment DayÃ,Â	English
28	8.5	Back to the FutureÃ,Â	English
29	8.5	Raiders of the Lost ArkÃ,Â	English
30	8.5	The Lion KingÃ,Â	English
31	8.5	The PianistÃ,Â	English
32	8.5	Apocalypse NowÃ,Â	English
33	8.5	PsychoÃ,Â	English
34	8.4	American BeautyÃ,Â	English
35	8.4	BraveheartÃ,Â	English
36	8.4	WALLÃ,·EÃ,Â	English
37	8.4	Star Wars: Episode VI - Return of the JediÃ,Â	English
38	8.4	AmÃf©lieÃ,Â	French
39	8.4	AliensÃ,Â	English
40	8.4	Once Upon a Time in AmericaÃ,Â	English
41	8.4	Lawrence of ArabiaÃ,Â	English
42	8.4	Das BootÃ,Â	German
43	8.4	Baahubali: The BeginningÃ,Â	Telugu
44	8.3	Batman BeginsÃ,Â	English
45	8.3	Inglourious BasterdsÃ,Â	English
46	8.3	Eternal Sunshine of the Spotless MindÃ,Â	English
47	8.3	UpÃ,Â	English
48	8.3	Toy StoryÃ,Â	English
49	8.3	Good Will HuntingÃ,Â	English
50	8.3	SnatchÃ,Â	English
51	8.3	Toy Story 3Ã,Â	English

52	8.3	ScarfaceÃ,Â	English
53	8.3	Indiana Jones and the Last CrusadeÃ,Â	English
54	8.3	2001: A Space OdysseyÃ,Â	English
55	8.3	L.A. ConfidentialÃ,Â	English
56	8.3	Inside OutÃ,Â	English
57	8.3	UnforgivenÃ,Â	English
58	8.3	AmadeusÃ,Â	English
59	8.3	DownfallÃ,Â	German
60	8.3	Raging BullÃ,Â	English
61	8.3	RoomÃ,Â	English
62	8.3	MetropolisÃ,Â	German
63	8.2	V for VendettaÃ,Â	English
64	8.2	The Wolf of Wall StreetÃ,Â	English
65	8.2	Finding NemoÃ,Â	English
66	8.2	A Beautiful MindÃ,Â	English
67	8.2	Die HardÃ,Â	English
68	8.2	Gran TorinoÃ,Â	English
69	8.2	The Big LebowskiÃ,Â	English
70	8.2	How to Train Your DragonÃ,Â	English
71	8.2	Pan's LabyrinthÃ,Â	Spanish
72	8.2	Blade RunnerÃ,Â	English
73	8.2	Into the WildÃ,Â	English
74	8.2	CasinoÃ,Â	English
75	8.2	WarriorÃ,Â	English
76	8.2	Captain America: Civil WarÃ,Â	English
77	8.2	The ThingÃ,Â	English
78	8.2	Howl's Moving CastleÃ,Â	Japanese
79	8.1	The AvengersÃ,Â	English
		Pirates of the Caribbean: The Curse of the Black	
80	8.1	PearlÃ,Â	English
81	8.1	Shutter IslandÃ,Â	English
82	8.1	Kill Bill: Vol. 1Ã,Â	English
83	8.1	The Sixth SenseÃ,Â	English
84	8.1	Guardians of the GalaxyÃ,Â	English
85	8.1	The Truman ShowÃ,Â	English
86	8.1	Sin CityÃ,Â	English
87	8.1	Jurassic ParkÃ,Â	English
88	8.1	No Country for Old MenÃ,Â	English
89	8.1	Monsters, Inc.Ã,Â	English
90	8.1	Gone GirlÃ,Â	English
91	8.1	Mad Max: Fury RoadÃ,Â	English
92	8.1	The Bourne UltimatumÃ,Â	English

93	8.1	Million Dollar BabyÃ,Â	English
94	8.1	DeadpoolÃ,Â	English
95	8.1	The Grand Budapest HotelÃ,Â	English
96	8.1	The MartianÃ,Â	English
97	8.1	The Imitation GameÃ,Â	English
98	8.1	12 Years a SlaveÃ,Â	English
99	8.1	Groundhog DayÃ,Â	English
100		The RevenantÃ,Â	English
101	8.1	PrisonersÃ,Â	English
102	8.1	There Will Be BloodÃ,Â	English
103	8.1	The HelpÃ,Â	English
104	8.1	RushÃ,Â	English
105	8.1	The Princess BrideÃ,Â	English
106	8.1	Hotel RwandaÃ,Â	English
107	8.1	SpotlightÃ,Â	English
108	8.1	The Sea InsideÃ,Â	Spanish
109	8.1	Tae Guk Gi: The Brotherhood of WarÃ,Â	Korean
110	8	Slumdog MillionaireÃ,Â	English
111	8	Black SwanÃ,Â	English
112	8	District 9Ã,Â	English
113	8	Catch Me If You CanÃ,Â	English
114	8	X-Men: Days of Future PastÃ,Â	English
115	8	Kill Bill: Vol. 2Ã,Â	English
116	8	Star TrekÃ,Â	English
117	8	The King's SpeechÃ,Â	English
118	8	The IncrediblesÃ,Â	English
119	8	RatatouilleÃ,Â	English
120	8	Casino RoyaleÃ,Â	English
121	8	Casino RoyaleÃ,Â	English
122	8	Life of PiÃ,Â	English
123	8	JawsÃ,Â	English
124	8	Blood DiamondÃ,Â	English
125	8	Rain ManÃ,Â	English
126	8	HerÃ,Â	English
127	8	The Perks of Being a WallflowerÃ,Â	English
128	8	Big FishÃ,Â	English
129	8	Mystic RiverÃ,Â	English
130	8	The Pursuit of HappynessÃ,Â	English
131	8	In BrugesÃ,Â	English
132	8	The ExorcistÃ,Â	English
133	8	Dead Poets SocietyÃ,Â	English
134	8	AladdinÃ,Â	English

135	8	SerenityÃ,Â	English
136	8	MagnoliaÃ,Â	English
137	8	Mulholland DriveÃ,Â	English
138	8	The ArtistÃ,Â	English
139	8	Dances with WolvesÃ,Â	English
140	8	True RomanceÃ,Â	English
141	8	BrazilÃ,Â	English
142	8	Cinderella ManÃ,Â	English
143	8	The Iron GiantÃ,Â	English
144	8	JFKÃ,Â	English
145	8	Dancer in the DarkÃ,Â	English
146	8	Doctor ZhivagoÃ,Â	English
147	7.9	AvatarÃ,Â	English
148	7.9	Iron ManÃ,Â	English
149	7.9	The Hobbit: An Unexpected JourneyÃ,Â	English
150	7.9	TakenÃ,Â	English
151	7.9	The Hobbit: The Desolation of SmaugÃ,Â	English
152	7.9	ShrekÃ,Â	English
153	7.9	Edge of TomorrowÃ,Â	English
154	7.9	The Bourne IdentityÃ,Â	English
155	7.9	The NotebookÃ,Â	English
156	7.9	Toy Story 2Ã,Â	English
157	7.9	Children of MenÃ,Â	English
158	7.9	Edward ScissorhandsÃ,Â	English
159	7.9	Hot FuzzÃ,Â	English
160	7.9	Captain PhillipsÃ,Â	English
161	7.9	E.T. the Extra-TerrestrialÃ,Â	English
162	7.9	Big Hero 6Ã,Â	English
163	7.9	The FighterÃ,Â	English
164	7.9	The Hateful EightÃ,Â	English
165	7.9	How to Train Your Dragon 2Ã,Â	English
166	7.9	The UntouchablesÃ,Â	English
167	7.9	Crouching Tiger, Hidden DragonÃ,Â	Mandarin
168	7.9	Almost FamousÃ,Â	English
169	7.9	Boogie NightsÃ,Â	English
170	7.9	Walk the LineÃ,Â	English
171	7.9	HalloweenÃ,Â	English
172	7.9	HeroÃ,Â	Mandarin
173	7.9	The Blues BrothersÃ,Â	English
174	7.9	Ed WoodÃ,Â	English
175	7.9	The InsiderÃ,Â	English
176	7.9	Letters from Iwo JimaÃ,Â	Japanese

177	7.9	Straight Outta ComptonÃ,Â	English
178	7.9	GloryÃ,Â	English
179	7.9	GloryÃ,Â	English
180	7.9	My Fair LadyÃ,Â	English
181	7.9	The Remains of the DayÃ,Â	English
182	7.9	The Right StuffÃ,Â	English
183	7.9	The World's Fastest IndianÃ,Â	English
184	7.8	The HangoverÃ,Â	English
185	7.8	GravityÃ,Â	English
186	7.8	Silver Linings PlaybookÃ,Â	English
187	7.8	SkyfallÃ,Â	English
188	7.8	X-Men: First ClassÃ,Â	English
189	7.8	Captain America: The Winter SoldierÃ,Â	English
190	7.8	The Curious Case of Benjamin ButtonÃ,Â	English
191	7.8	DriveÃ,Â	English
192	7.8	Ocean's ElevenÃ,Â	English
193	7.8	Star Trek Into DarknessÃ,Â	English
194	7.8	Birdman or (The Unexpected Virtue of Ignorance)Ã,Â	English
195	7.8	Harry Potter and the Prisoner of AzkabanÃ,Â	English
196	7.8	The Bourne SupremacyÃ,Â	English
197	7.8	Back to the Future Part IIÃ,Â	English
198	7.8	The Girl with the Dragon TattooÃ,Â	English
199	7.8	American GangsterÃ,Â	English
200	7.8	TangledÃ,Â	English
201	7.8	PredatorÃ,Â	English
202	7.8	Wreck-It RalphÃ,Â	English
203	7.8	Lucky Number SlevinÃ,Â	English
204	7.8	The GameÃ,Â	English
205	7.8	Being John MalkovichÃ,Â	English
206	7.8	The Fault in Our StarsÃ,Â	English
207	7.8	The Lego MovieÃ,Â	English
208	7.8	3:10 to YumaÃ,Â	English
209	7.8	Moonrise KingdomÃ,Â	English
210	7.8	ApocalyptoÃ,Â	Maya
211	7.8	Donnie BrascoÃ,Â	English
212	7.8	O Brother, Where Art Thou?Ã,Â	English
213	7.8	GattacaÃ,Â	English
214	7.8	The FugitiveÃ,Â	English
215	7.8	About TimeÃ,Â	English
216	7.8	Office SpaceÃ,Â	English
217	7.8	ChangelingÃ,Â	English
218	7.8	Pride & PrejudiceÃ,Â	English

219	7.8	AtonementÃ,Â	English
220	7.8	The Big ShortÃ,Â	English
221	7.8	Finding NeverlandÃ,Â	English
222	7.8	What's Eating Gilbert GrapeÃ,Â	English
223	7.8	South Park: Bigger Longer & UncutÃ,Â	English
224	7.8	Remember the TitansÃ,Â	English
225	7.8	Fantastic Mr. FoxÃ,Â	English
226	7.8	The Boy in the Striped PajamasÃ,Â	English
227	7.8	The Last of the MohicansÃ,Â	English
228	7.8	The Jungle BookÃ,Â	English
229	7.8	The Jungle BookÃ,Â	English
230	7.8	Kung Fu HustleÃ,Â	Cantonese
231	7.8	TombstoneÃ,Â	English
232	7.8	NebraskaÃ,Â	English
233	7.8	Glengarry Glen RossÃ,Â	English
234	7.8	The Last EmperorÃ,Â	English
235	7.8	The Best OfferÃ,Â	English
236	7.8	The Conjuring 2Ã,Â	English
237	7.8	The Color PurpleÃ,Â	English
238	7.8	Black BookÃ,Â	Dutch
239	7.8	The White RibbonÃ,Â	German
240	7.8	HamletÃ,Â	English
241	7.8	The Little PrinceÃ,Â	English
242	7.8	The VerdictÃ,Â	English
243	7.7	TitanicÃ,Â	English
244	7.7	300Ã,Â	English
245	7.7	The Social NetworkÃ,Â	English
246	7.7	ArgoÃ,Â	English
247	7.7	Kick-AssÃ,Â	English
248	7.7	Minority ReportÃ,Â	English
249	7.7	Cast AwayÃ,Â	English
250	7.7	WatchmenÃ,Â	English

From the above query , I got this **IMDb_Top_250** , I had attached as proof of this.

This is my attached excel file:

https://docs.google.com/spreadsheets/d/16NnCBQPx46wDCiuNa tK76Pn8Ra47QAzb/edit?usp=drive_link b)For the next part as , out of these top250 movies with a highest user votes , we search for all the Top_Foreign_Lang_Film using the below query :

```
SELECT row_number() over(order by imdb_score DESC,

num_voted_users DESC) as ranking,

imdb_score,

num_voted_users,

movie_title AS Top_non_english,

language

FROM

moviesp.cleanedm

WHERE

num_voted_users > 30000 and language <> 'English'

LIMIT 36;
```

	imdb_scor	num_voted_user		
ranking	е	S	Top_non_english	language
1	8.6	417971	Spirited AwayÃ,Â	Japanese
2	8.4	534262	AmÃf©lieÃ,Â	French
3	8.4	168203	Das BootÃ,Â	German
4	8.4	62756	Baahubali: The BeginningÃ,Â	Telugu
5	8.3	248354	DownfallÃ,Â	German
6	8.3	111841	MetropolisÃ,Â	German
7	8.2	467234	Pan's LabyrinthÃ,Â	Spanish
8	8.2	214091	Howl's Moving CastleÃ,Â	Japanese
9	8.1	64556	The Sea InsideÃ,Â	Spanish

			Tae Guk Gi: The Brotherhood of	
10	8.1	31943	WarÃ,Â	Korean
11	7.9	217740	Crouching Tiger, Hidden DragonÃ,Â	Mandarin
12	7.9	149414	HeroÃ,Â	Mandarin
13	7.9	132149	Letters from Iwo JimaÃ,Â	Japanese
14	7.8	236000	ApocalyptoÃ,Â	Maya
15	7.8	99353	Kung Fu HustleÃ,Â	Cantonese
16	7.8	59507	Black BookÃ,Â	Dutch
17	7.8	52958	The White RibbonÃ,Â	German
18	7.7	85589	PonyoÃ,Â	Japanese
19	7.7	62607	A Very Long EngagementÃ,Â	French
20	7.7	55516	The Great BeautyÃ,Â	Italian
21	7.6	92295	House of Flying DaggersÃ,Â	Mandarin
22	7.6	68119	The Kite RunnerÃ,Â	Dari
23	7.6	38690	The Flowers of WarÃ,Â	Mandarin
24	7.5	77656	The Painted VeilÃ,Â	Mandarin
25	7.5	41496	2046Ã,Â	Cantonese
26	7.4	36894	Red CliffÃ,Â	Mandarin
27	7.3	63084	Paris, je t'aimeÃ,Â	French
			Mongol: The Rise of Genghis	Mongolia
28	7.3	37635	KhanÃ,Â	n
29	7.2	55928	District B13Ã,Â	French
30	7.1	179235	The Passion of the ChristÃ,Â	Aramaic
31	7	36455	Curse of the Golden FlowerÃ,Â	Mandarin
32	6.7	32003	Coco Before ChanelÃ,Â	French
33	6.4	86152	The InterpreterÃ,Â	Aboriginal
34	5.9	71574	The Legend of ZorroÃ,Â	Spanish
35	4.3	31414	In the Land of Blood and HoneyÃ,Â	Bosnian

- ❖ From the above query , I have found 35 movies list with respect to the num_voted_users for non English movies with the top 250 lists.
- ❖ Then with this results, have been queried with respect to the number of voters for the most favorable movies with ratings of imdb_score, that is non_english that is foreign language movies that have been listed using the query. So in this question Top

250 and Top_Foreign_Lang_Film(35 movies), MySQL workbench had been used to get the desired results and with the results, I had transferred the results to excel to make table to be readable.

This is my file attached here:
https://drive.google.com/file/d/1sDk5gBzVoLVdENnGPDrv
77P4AksNeHbW/view?usp=drive link

4. **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

For this question ,want to find out the top 10 directors whom with the imdb_score of highest one . This will be found out by using the below query:

SELECT

director_name AS Top_10_Director,

ROUND(AVG(imdb_score), 2) AS avg_rating

FROM

moviesp.cleanedm

GROUP BY director_name

ORDER BY avg_rating DESC, director_name

LIMIT 10;

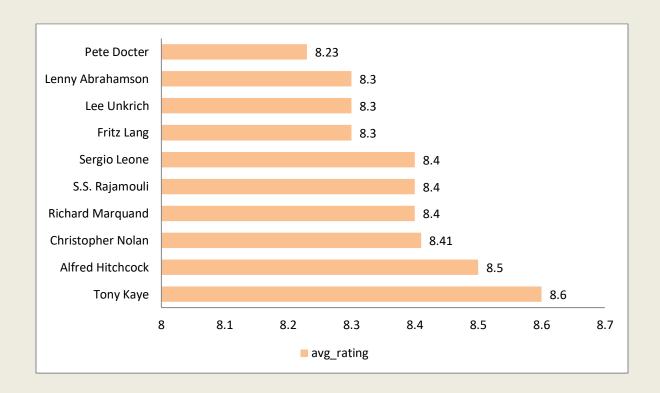
With this query, I have got the results.....

Top 10 Directors					
Top_10_Director	avg_rating				
Tony Kaye	8.6				
Alfred Hitchcock	8.5				
Christopher Nolan	8.41				
Richard Marquand	8.4				
S.S. Rajamouli	8.4				
Sergio Leone	8.4				
Fritz Lang	8.3				
Lee Unkrich	8.3				
Lenny Abrahamson	8.3				
Pete Docter	8.23				

From the above results, I have inferred that the top 10 directors have been listed by using the sql query results with the below insights.

This is my excel file attached:

https://docs.google.com/spreadsheets/d/1kDmzbNlehAevghQAK 8UPAuYi1rA3jlXj/edit?usp=drive_link



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

Your task: Find popular genres

❖ For this question, we first split the genres of each movie into individual genres into 8 separate levels of genres. Then, we make a list of all genres by uniting all levels of genres and then count the popularity of each genre.

❖ I have been using sql query to solve this description of separating 8 levels of genres with list of all genres the with mean of gross and the number of movies.

```
With d8 as
(SELECT *, (CASE WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 8), '|',
-1) = genre_1 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_2
THEN NULL
WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 8), '|', -1) = genre 3
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_4
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_5
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_6
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 8), '|', -1) = genre_7
THEN NULL
ELSE SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 8), '|', -1) END) AS
genre 8
FROM
(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|',
-1) = genre 1 THEN NULL
WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 7), '|', -1) = genre 2
THEN NULL
WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 7), '|', -1) = genre 3
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_4
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_5
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 7), '|', -1) = genre_6
THEN NULL
 ELSE SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 7), '|', -1) END) AS
genre 7
 FROM
(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|',
-1) = genre_1 THEN NULL
```

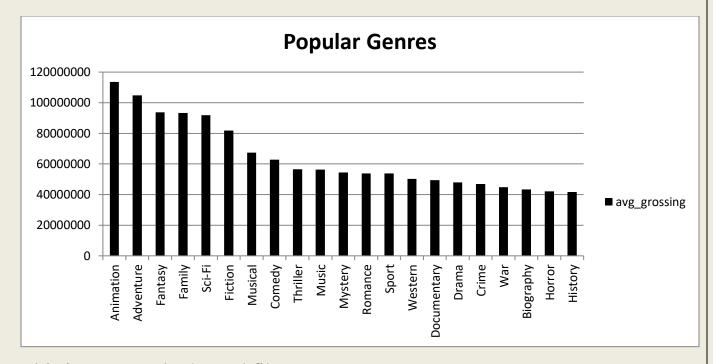
```
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_2
THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_3
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_4
THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 6), '|', -1) = genre_5
THEN NULL
 ELSE SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 6), '|', -1) END) AS
genre 6
 FROM
 (SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|',
-1) = genre 1 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_2
THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_3
THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 5), '|', -1) = genre_4
THEN NULL
 ELSE SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 5), '|', -1) END) AS
genre 5
 FROM(SELECT *, (CASE WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|',
4), '|', -1) = genre 1 THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_2
THEN NULL
 WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) = genre_3
THEN NULL
 ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 4), '|', -1) END) AS
genre 4
 FROM
 (SELECT *, (CASE WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 3), '|',
-1) = genre_1 THEN NULL
WHEN SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 3), '|', -1) = genre_2
THEN NULL
 ELSE SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 3), '|', -1) END) AS
genre 3
 FROM
 (SELECT *, (CASE WHEN SUBSTRING INDEX(SUBSTRING INDEX(genres, '|', 2), '|',
-1) = genre 1 THEN NULL
```

```
ELSE SUBSTRING_INDEX(SUBSTRING_INDEX(genres, '|', 2), '|', -1) END) AS
genre 2
 FROM
 (SELECT
SUBSTRING(genres, 2) as genres,
SUBSTRING INDEX(SUBSTRING INDEX(SUBSTRING(genres, 2), '|', 1), '|', - 1) AS
genre 1
 FROM moviesp.cleanedm) d1) d2) d3) d4) d5) d6) d7),
 list as
 (SELECT DISTINCT genre 1 AS genre list
 FROM
 (SELECT genre 1 FROM d8) UNION ALL
 (SELECT genre 2 FROM d8) UNION ALL
 (SELECT genre 3 FROM d8) UNION ALL
 (SELECT genre 4 FROM d8) UNION ALL
 (SELECT genre 5 FROM d8) UNION ALL
 (SELECT genre 6 FROM d8) UNION ALL
 (SELECT genre 7 FROM d8) UNION ALL
 (SELECT genre 8 FROM d8)) g12
 WHERE genre 1 IS NOT NULL ORDER BY genre 1)
 SELECT genre_list,round(AVG(DISTINCT CASE WHEN INSTR(genres, genre_list)
THEN gross ELSE NULL END), 2) AS avg grossing,
 COUNT(DISTINCT CASE WHEN INSTR(genres, genre list) THEN genres ELSE
NULL END) AS no of movies
 FROM cleanedm JOIN list
 GROUP BY genre list
 ORDER BY avg grossing DESC;
 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 moviesp.cleanedm
 WHERE actor 1 name = 'CCH Pounder')
 UNION ALL
 (SELECT actor 1 name, movie title, num user for reviews, num critic for reviews
 FROM moviesp.cleanedm
```

WHERE actor_1_name = ' J.K. Simmons')
UNION ALL
(SELECT actor_1_name, movie_title, num_user_for_reviews,
num_critic_for_reviews
FROM moviesp.cleanedm
WHERE actor_1_name = ' Jennifer Lawrence')) c
GROUP BY actor_1_name
ORDER BY user_reviews DESC, critic_reviews DESC;

Popular Genres				
genre_list	avg_grossing	no_of_movies		
Animation	113547233.2	84		
Adventure	104744835.7	238		
Fantasy	93680359.4	170		
Family	93247050.7	178		
Sci-Fi	91891509.05	134		
Fiction	81796457.68	235		
Musical	67450846.54	53		
Comedy	62876377.93	263		
Thriller	56544554.87	179		
Music	56385988.68	101		
Mystery	54357927.28	97		
Romance	53797743.97	171		
Sport	53731204.96	39		
Western	50326576.6	33		
Documentary	49478940.43	4		
Drama	47978352.15	310		
Crime	46975720.56	121		
War	44756370.57	60		
Biography	43329712.5	63		
Horror	42187374.12	70		
History	41782106.2	51		

From the above query results, There are 23 Genres got to inferred that the splitted genres with the mean of gross and number of movies match lists. And got the analysis of insights of column chart are as follows:



This is my attached excel file:

https://docs.google.com/spreadsheets/d/13gHz7H3oDULFNXcZJ IRAB2g6VOb-UgVZ/edit?usp=drive_link

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-favorite and audience-favorite actors

- a) For this critic-favourite, the assigned favourite movies has been extracted which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors.
- ❖ So I have queried for extracting the provided favourite actors with the number of movies , user reviews (audience-favourite) and critic-reviews. The query for finding this as follows:

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 moviesp.cleanedm
 WHERE actor_1_name = 'Meryl Streep')
 UNION ALL
 (SELECT actor 1 name, movie title, num user for reviews,
num critic for reviews
FROM moviesp.cleanedm
WHERE actor_1_name = 'Leonardo DiCaprio')
 UNION ALL
 (SELECT actor 1 name, movie title, num user for reviews,
num critic for reviews
FROM moviesp.cleanedm
 WHERE actor_1_name = 'Brad Pitt')) c
 GROUP BY actor_1_name
 ORDER BY user_reviews DESC, critic_reviews DESC;
```

Leonardo DiCaprio	20	922.55	322.2
Brad Pitt	17	742.35	245
Meryl Streep	11	297.18	181.45

- ❖ From the above query, I got this results which inferred of extracting from the actor name 1 with he help of user reviews and critic reviews.
- ❖ And this query helped me to find out that Leonardo DiCaprio is the most favorite actor for both the audience and the critics.
- ❖ This is the file attached here:
 https://docs.google.com/spreadsheets/d/1046zg F9vOST467
 Ewu6HOAwdWET-C-Bn/edit?usp=drive_link
- b) 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**.
 - ❖ For this decades per number user votes, we will bw finding out extracting the years from MySQL query workbench are as follows:

SELECT

CONCAT(CONVERT(FLOOR(title_year / 10) * 10 , CHAR),

's') AS decade,

SUM(num_voted_users) AS total_votes

FROM

moviesp.cleanedm

GROUP BY decade

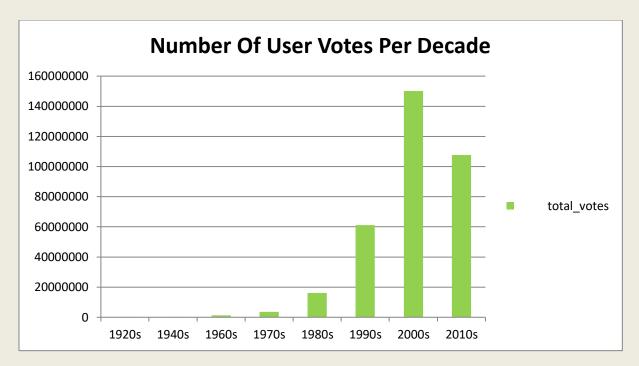
ORDER BY decade;

This is excel attached here:

 $\underline{https://docs.google.com/spreadsheets/d/1B3LzOFwL59h8d2jBRb6YJ73Y3tmHTAiR/edit?usp=drive_link}$

Number Of User Votes Per				
Decade				
decade	total_votes			
1920s	111841			
1940s	90360			
1960s	1218680			
1970s	3679621			
1980s	16136813			
1990s	61213572			
2000s	150036196			
2010s	107714560			

From the above query, I have inferred that the results of decades per number of user votes with the voting that are extracted using the query using aggregation function.



From the above column chart, I have analyzed a insights of decades over the number of user votes.

RESULT

- ➤ So in this IMDB movie analysis project , I have analyzed with the useful insights that are making good impression with a anyone can readable format. In this project , MySQL workbench and MS Excel helped me throughout the one , which helps me getting and extracting the more and more useful strategy.
- ➤ Then I have learned how to remove outliers, data cleaning strategy, Advanced SQL query, exporting the results of sql query to the excel to make visual insights and so on.

	whole excel files)	
https://drive.google.com/drive/folders/1L1Lmf10XD2d6N2DZx08		
<u>34sa0G2RrdUwi?u</u>	sp=drive link	