IMDB Movie Analysis

Project Description:

We have dataset of different IMDB movies, we need to clean the data, analysis the data and derive the insights from that dataset by using 5 why analysis and create the bar graph and pivot charts for the insights, answer the questions and should create the report.

Approach:

In order to execute the project, we should download the dataset given and clean the data by removing the duplicate values and null values and by representing the dataset with the bar graphs and pivot tables in order to find the insights.
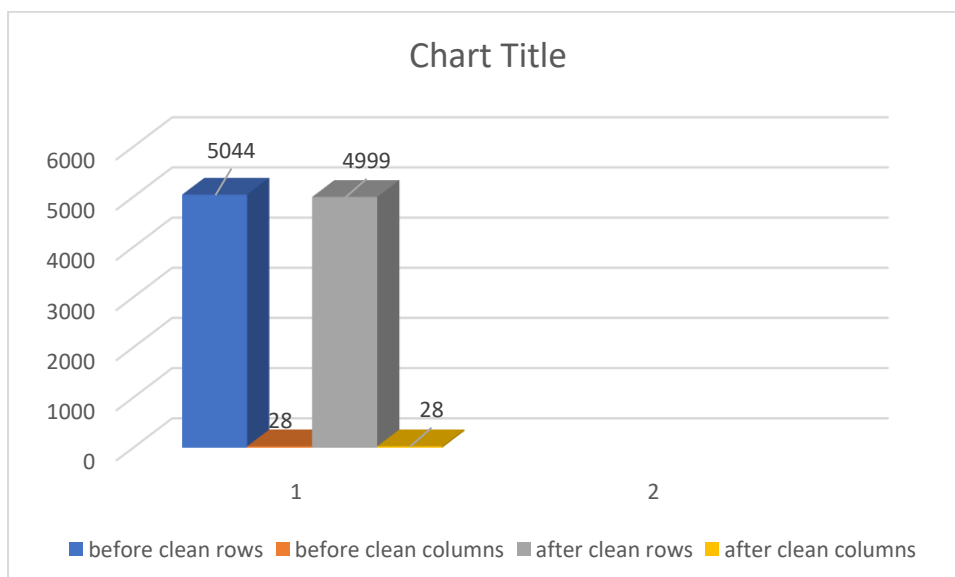
Tech stack used:

To find the insights we have used excel to represent the graphs and download the dataset in excel, Microsoft word to present the report and google drive to submit the project.

Insights:

a. Clean the data:

before clean		after clean	
rows	columns	rows	columns
5044	28	4999	28

Result:

b. Find the movies with the highest profit?

movie_title	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
Star Wars: Episode I - The Phantom MenaceÂ	359544677
The Dark KnightÂ	348316061
The Hunger GamesÂ	329999255
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

Result:
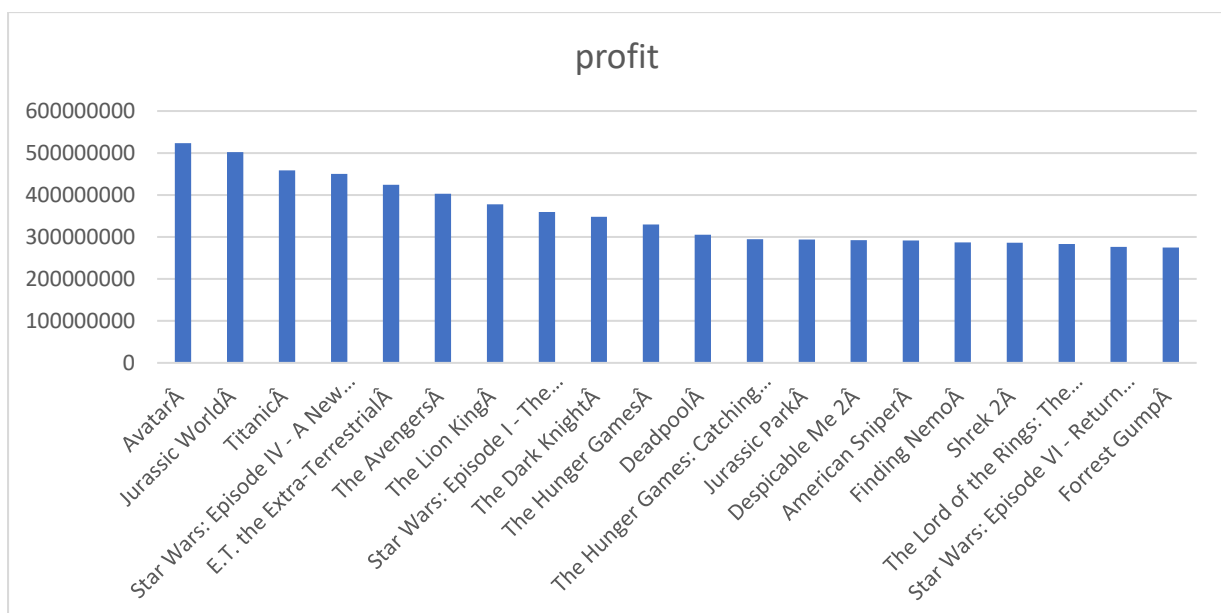
c. Find IMDB Top 250

movie_title	IMDB_Top_250	num_voted_users >25000
The Shawshank RedemptionÂ	9.5	1689764
The Dark KnightÂ	9.3	1676169
InceptionÂ	9.2	1468200
Fight ClubÂ	9.1	1347461
Pulp FictionÂ	9.1	1324680
Forrest GumpÂ	9.1	1251222
The Lord of the Rings: The Fellowship of the RingÂ	9	1238746
The MatrixÂ	9	1217752
The Lord of the Rings: The Return of the KingÂ	9	1215718
The GodfatherÂ	8.9	1155770
The Dark Knight RisesÂ	8.9	1144337
The Lord of the Rings: The Two TowersÂ	8.9	1100446
Se7enÂ	8.9	1023511
The AvengersÂ	8.9	995415
GladiatorÂ	8.8	982637
Batman BeginsÂ	8.8	980946
Django UnchainedÂ	8.8	955174
InterstellarÂ	8.8	928227
Star Wars: Episode IV - A New HopeÂ	8.8	911097
The Silence of the LambsÂ	8.8	887467
AvatarÂ	8.8	886204
Inglourious BasterdsÂ	8.7	885175
Saving Private RyanÂ	8.7	881236
The DepartedÂ	8.7	873649

c.a]Top foreign language film:

Top_foreign_lang_film	Language	imdb_score
Wild GrassÂ	French	9.5
Remember Me, My LoveÂ	Italian	9
Godzilla ResurgenceÂ	Japanese	8.8
Mondays in the SunÂ	Spanish	8.7
AjamiÂ	Arabic	8.7
Journey from the FallÂ	Vietnamese	8.5
The GrandmasterÂ	Mandarin	8.4
The Painted VeilÂ	Mandarin	8.4
Dum Maaro DumÂ	Hindi	8.4
The Barbarian InvasionsÂ	French	8.4
SamsaraÂ	None	8.4

Futuro BeachÂ	Portuguese	8.4
Karachi se LahoreÂ	Urdu	8.4
Lage Raho Munna BhaiÂ	Hindi	8.3
ElzaÂ	French	8.3
3Â	German	8.3
Stories of Our LivesÂ	Swahili	8.3
The Legend of Drunken MasterÂ	Cantonese	8.2
Asterix at the Olympic GamesÂ	French	8.1
The Diving Bell and the ButterflyÂ	French	8.1
The Widow of Saint-PierreÂ	French	8.1
Coming HomeÂ	Mandarin	8.1
FizaÂ	Hindi	8.1
Cries & WhispersÂ	Swedish	8.1
Namastey LondonÂ	Hindi	8
AkiraÂ	Japanese	8

d. Find the best directors:

director_name	Average of imdb_score
John Blanchard	9.5
Cary Bell	8.7
Mitchell Altieri	8.7
Sadyk Sher-Niyaz	8.7
Charles Chaplin	8.6
Mike Mayhall	8.6
Majid Majidi	8.5
Raja Menon	8.5
Ron Fricke	8.5
Damien Chazelle	8.5

Result:
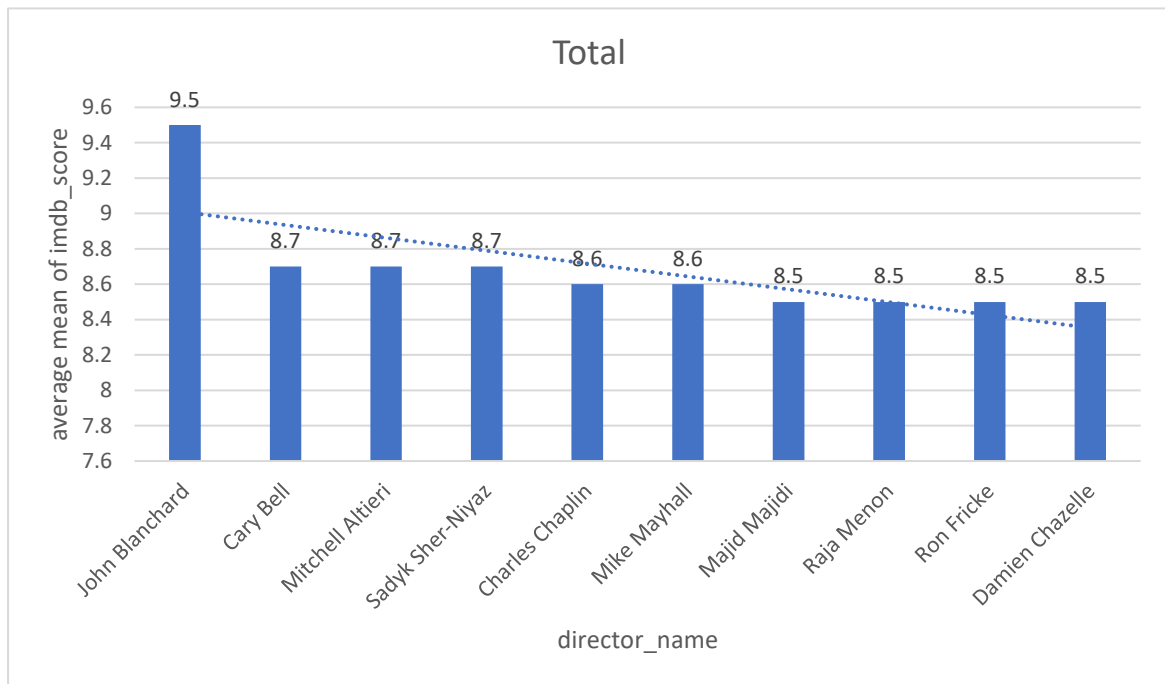
e. Find popular genres

| Genres	Count of imdb_score
Comedy	56
Drama	44
Comedy|Romance	41
Comedy|Drama	38
Horror	38
Comedy|Drama|Romance	36
Crime|Drama	33
Crime|Drama|Thriller	33
Action|Adventure|Sci-Fi	32
Drama|Thriller	31

Result:
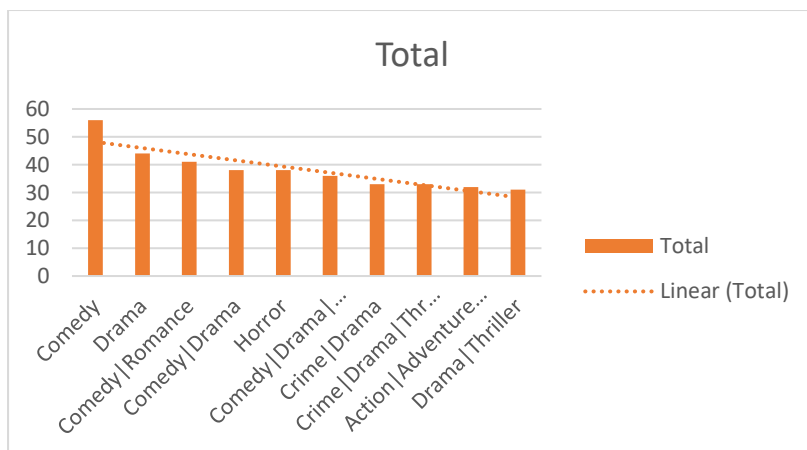
f. Find the critic-favorite and audience-favorite actors

Count of Leonardo DiCaprio	Count of Brad Pitt	Count of Meryl_Streep
21	18	13
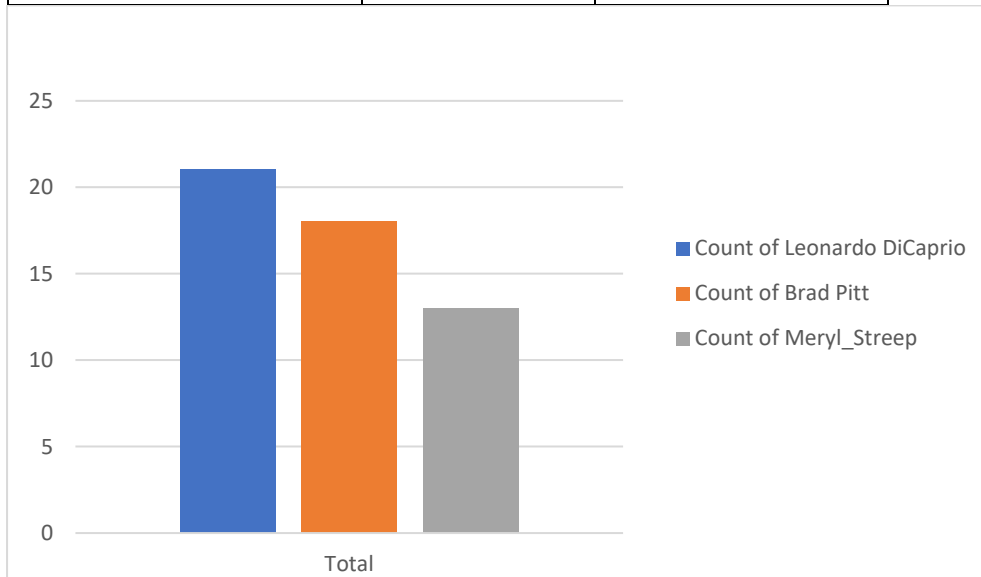
Row Labels	Average of num_critic_for_reviews
Phaldut Sharma	738
Peter Capaldi	654
Craig Stark	596
Bérénice Bejo	576
Suraj Sharma	552
Ellar Coltrane	548
Mike Howard	546
Lou Taylor Pucci	543
Maika Monroe	533
Tim Holmes	525
Albert Finney	510
Elina Alminas	489
Kurt Fuller	487
Iko Uwais	481
Quvenzhané Wallis	478.6666667
Edgar Arreola	478
Sharlto Copley	472
Cory Hardrict	452
Elizabeth McGovern	447
Aidan Turner	447
Wood Harris	432
Anil Kapoor	418

Jessica Barden	417
Danielle Kotch	411
Micah Sloat	409
Christo Jivkov	406
Ivana Baquero	406
Christoph Waltz	405.5
Seamus Davey-Fitzpatrick	405
Giancarlo Giannini	403
Martin Dew	401
John Boyega	399
Jim Meskimen	398
Clea DuVall	396
Jennifer Lawrence	391.8
Chris Hemsworth	390.9285714
Shengyi Huang	377
Daryl Sabara	373.5
Billy West	372
Matt Lucas	371
Ocean James	371
Shailene Woodley	365
Doona Bae	363
Henry Cavill	362.4285714
John Ratzenberger	361.6666667
Rhys Wakefield	361
Benedict Cumberbatch	360.7142857
Heather Donahue	360
James Martin Kelly	360
Grace Phipps	358
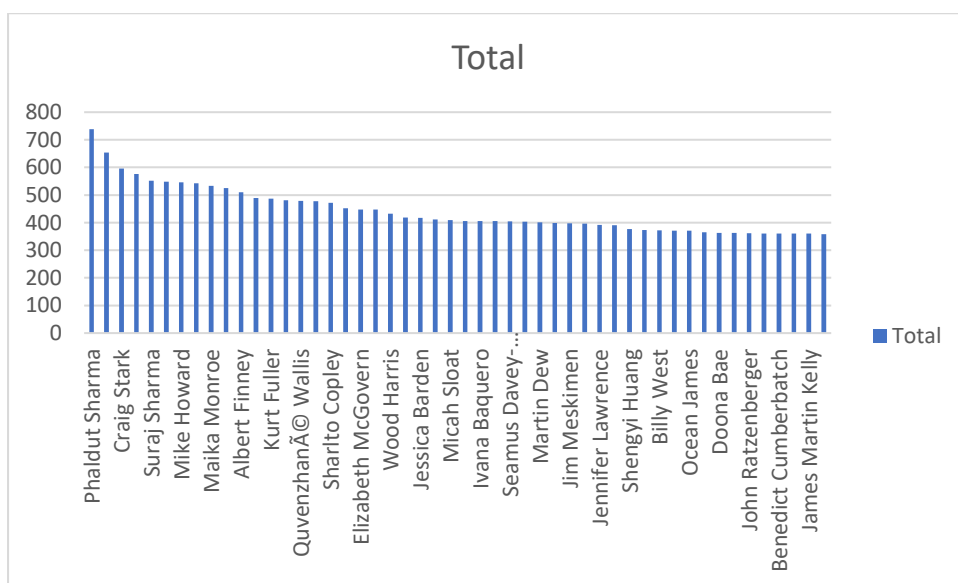
Row Labels	Average of num_user_for_reviews
Heather Donahue	3400
Christo Jivkov	2814
Steve Bastoni	2789
Phaldut Sharma	1885
Keir Dullea	1736
Chen Chang	1641
Nick Stahl	1562
Kevin Rankin	1445
Noah Huntley	1441
Osama bin Laden	1416
Seychelle Gabriel	1382
Mathieu Kassovitz	1314
Eva Green	1290
Essie Davis	1285.5
Sharlto Copley	1262
Giancarlo Giannini	1243
Orlando Bloom	1242.333333
Luenell	1198
Micah Sloat	1189
Fionnula Flanagan	1109
Jim Meskimen	1107
Ivana Baquero	1083
Henry Cavill	1066.857143
Mhairi Calvey	1065
Talulah Riley	1058
Jake Busey	1049
Scatman Crothers	1040
Natalie Portman	1031.9375
Craig Stark	1018
Albert Finney	1017.333333
Anil Kapoor	1017
Peter Capaldi	995
Christopher Lee	990.3333333
CCH Pounder	965.5
Lauren German	962
David Carradine	945.3333333
Kathleen Freeman	945
David Zayas	944

John Doman	935
Glenn Morshower	934.6666667
Sanaa Lathan	932.5
Blanchard Ryan	916
Leonardo DiCaprio	914.4761905
Bunta Sugawara	902
Wayne Knight	895
Aidan Turner	894.25
Jack Warden	888
Rory Culkin	872.6666667
Michael O'Neill	867
Peter Serafinowicz	859

Result:
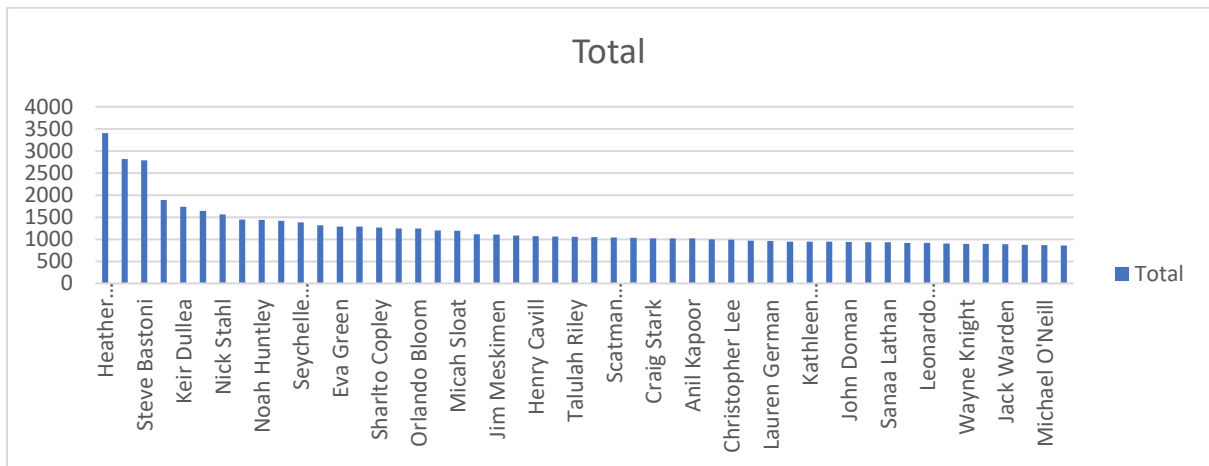
decade	Sum of num_voted_users
1910s	10718
1920s	128672
1930s	984397
1940s	1211888
1950s	1638504
1960s	5065074
1970s	11072372
1980s	23021112
1990s	70278714
2000s	176260317
2010s	124384369

Result:

Total

| | 1910s | 1920s | 1930s | 1940s | 1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s |

y-axis: 0, 20000000, 40000000, 60000000, 80000000, 100000000, 120000000, 140000000, 160000000, 180000000, 200000000