IMDB Movie Analysis

Statistics Final Project-1

Project Description

In this project, we will perform statistical analysis in order to understand IMBD.

Approach

Using MS Excel we will do statistical analysis of data records in IMBD Database

Tech-Stack Used: Microsoft Excel 2016

- MS Excel 2016 is a spreadsheet program where one can record data in the form of tables.
- It is easy to analyse numerical data in an Excel spreadsheet.
- It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications (VBA).

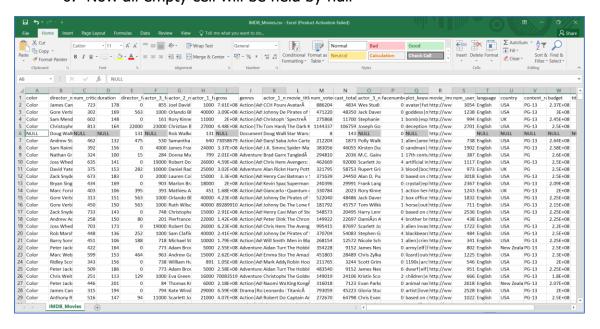
Insights:

A. **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.)

Soln:

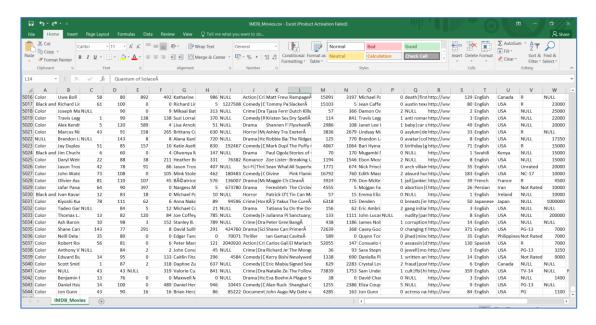
Select and treat all blank cells

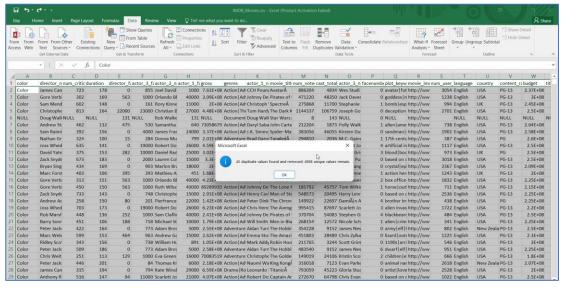
- 1. Selecting all data
- 2. go to special
- 3. click blanks and click ok (all the blank cells in data cell will be selected at the same time)
- 4. Write null in active cell
- 5. Hit ctrl+ enter
- 6. Now all empty cell will be field by null



Removing duplicate rows

- 1. Select all
- 2. Go to data
- 3. Remove duplicates
- 4. Select my data has headers
- 5. Click ok



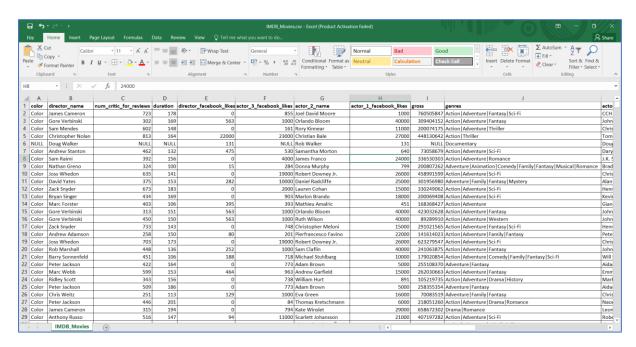


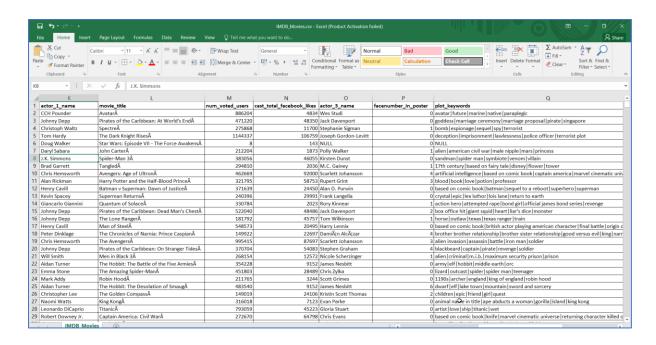
Highlighting errors:

- 1. Select all
- 2. Go to special
- 3. Select formulas
- 4. Select only errors from formulas
- 5. Click ok

Adding borders:

- 1. Select all
- 2. Add border
- 3. Proper spacing to view database properly
- 4. Making header text bold





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

Soln:

Movie name: AvatarÂ
 Budget: 237000000
 Gross: 760505847
 Profit: 523505847

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

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

Soln:

TOP 250 any language:

num voted users: filter >25000

Imbd_score: sort Largest to smallest, filter top 250

Rank: ROW() - 1

rank	movie_title v	imdb_score 🚭	num_voted_users 🕶
1	The Shawshank RedemptionÂ	9.3	1689764
2	The GodfatherÂ	9.2	1155770
3	The Dark KnightÂ	9	1676169
4	The Godfather: Part IIÂ	9	790926
5	FargoÂ	9	170055
6	The Lord of the Rings: The Return of the KingÂ	8.9	1215718
7	Pulp FictionÂ	8.9	1324680
8	Schindler's ListÂ	8.9	865020
9	The Good, the Bad and the UglyÂ	8.9	503509
10	12 Angry MenÂ	8.9	447785
11	Forrest GumpÂ	8.8	1251222
12	Star Wars: Episode V - The Empire Strikes Back	8.8	837759
13	The Lord of the Rings: The Fellowship of the Ri	8.8	1238746
14	InceptionÂ	8.8	1468200

DaredevilÂ	8.8	213483
It's Always Sunny in PhiladelphiaÂ	8.8	133415
Fight ClubÂ	8.8	1347461
Star Wars: Episode IV - A New HopeÂ	8.7	911097
The Lord of the Rings: The Two TowersÂ	8.7	1100446
The MatrixÂ	8.7	1217752
One Flew Over the Cuckoo's NestÂ	8.7	680041
Goodfellas Â	8.7	728685
City of GodÂ	8.7	533200
Friday Night LightsÂ	8.7	42746
Seven SamuraiÂ	8.7	229012
Saving Private RyanÂ	8.6	881236
The Silence of the LambsÂ	8.6	887467
Se7enÂ	8.6	1023511
InterstellarÂ	8.6	928227
The Usual SuspectsÂ	8.6	740918
HannibalÂ	8.6	159910
LutherÂ	8.6	70568
Spartacus: War of the DamnedÂ	8.6	173172
Once Upon a Time in the WestÂ	8.6	216005
It's a Wonderful LifeÂ	8.6	275720
CasablancaÂ	8.6	387508
American History XÂ	8.6	782437
Modern TimesÂ	8.6	143086
Spirited AwayÂ	8.6	417971
The Lion KingÂ	8.5	644348
Raiders of the Lost ArkÂ	8.5	661017
The Dark Knight RisesÂ	8.5	1144337
Back to the FutureÂ	8.5	732212
Terminator 2: Judgment DayÂ	8.5	744891
GladiatorÂ	8.5	982637
The Green MileÂ	8.5	782610
AlienÂ	8.5	563827
Django UnchainedÂ	8.5	955174
Apocalypse NowÂ	8.5	450676
The DepartedÂ	8.5	873649
PsychoÂ	8.5	422432
MementoÂ	8.5	845580
The PrestigeÂ	8.5	844052
WhiplashÂ	8.5	399138
The Lives of OthersÂ	8.5	259379
Children of HeavenÂ	8.5	27882
OutlanderÂ	8.5	50391
	It's Always Sunny in PhiladelphiaÂ Fight ClubÂ Star Wars: Episode IV - A New HopeÂ The Lord of the Rings: The Two TowersÂ The MatrixÂ One Flew Over the Cuckoo's NestÂ GoodfellasÂ City of GodÂ Friday Night LightsÂ Seven SamuraiÂ Saving Private RyanÂ The Silence of the LambsÂ Se7enÂ InterstellarÂ The Usual SuspectsÂ HannibalÂ LutherÂ Spartacus: War of the DamnedÂ Once Upon a Time in the WestÂ It's a Wonderful LifeÂ CasablancaÂ American History XÂ Modern TimesÂ Spirited AwayÂ The Lion KingÂ Raiders of the Lost ArkÂ The Dark Knight RisesÂ Back to the FutureÂ Terminator 2: Judgment DayÂ GladiatorÂ The Green MileÂ AlienÂ Django UnchainedÂ Apocalypse NowÂ The DepartedÂ PsychoÂ MementoÂ The PrestigeÂ WhiplashÂ The Lives of OthersÂ Children of HeavenÂ	It's Always Sunny in PhiladelphiaÂ 8.8 Fight ClubÂ 8.7 The Lord of the Rings: The Two TowersÂ 8.7 The MatrixÂ 8.7 One Flew Over the Cuckoo's NestÂ 8.7 GoodfellasÂ 8.7 City of GodÂ 8.7 Friday Night LightsÂ 8.7 Seven SamuraiÂ 8.7 Saving Private RyanÂ 8.6 The Silence of the LambsÂ 8.6 Se7enÂ 8.6 InterstellarÂ 8.6 Spartacus: War of the DamnedÂ 8.6 Once Upon a Time in the WestÂ 1t's a Wonderful LifeÂ 8.6 CasablancaÂ 8.6 Modern TimesÂ 8.6 Modern TimesÂ 8.6 Spirited AwayÂ 8.6 The Lost ArkÂ 8.6 The Dark Knight RisesÂ 8.5 Back to the FutureÂ 8.5 Terminator 2: Judgment DayÂ 6ladiatorÂ 8.5 The Orea MileÂ 8.5 Django UnchainedÂ 8.5 The Openation A.5 PsychoÂ 8.5 MementoÂ 8.5 Melidren of HeavenÂ 8.5

57	OutlanderÂ	8.5	50391
	EntourageÂ	8.5	135643
	AirliftÂ	8.5	30977
	Dr. Strangelove or: How I Learned to Stop Wo	8.5	342585
	The PianistÂ	8.5	497946
	Star Wars: Episode VI - Return of the JediÂ	8.4	681857
	American BeautyÂ	8.4	822500
	AliensÂ	8.4	488537
	WALL·EÂ	8.4	71883
	A SeparationÂ	8.4	151812
	BraveheartÂ	8.4	736638
68	Reservoir DogsÂ	8.4	664719
	Stargate SG-1Â	8.4	6398
	The ShiningÂ	8.4	61033
	Veronica MarsÂ	8.4	5552
72	The InbetweenersÂ	8.4	5598
73	Rang De BasantiÂ	8.4	7023
74	PsychÂ	8.4	6731
75	M*A*S*HÂ	8.4	3626
76	Batman: The Dark Knight Returns, Part 2Â	8.4	3043
77	To Kill a MockingbirdÂ	8.4	21508
78	OldboyÂ	8.4	35618
79	Requiem for a DreamÂ	8.4	57354
80	Das BootÂ	8.4	16820
81	Lawrence of ArabiaÂ	8.4	19277
82	Baahubali: The BeginningÂ	8.4	6275
83	Once Upon a Time in AmericaÂ	8.4	22100
84	AmélieÂ	8.4	53426

	I		1
85	Princess MononokeÂ	8.4	221552
86	Toy Story 3Â	8.3	544884
87	Inside OutÂ	8.3	345198
88	Toy StoryÂ	8.3	623757
89	The StingÂ	8.3	175607
90	Indiana Jones and the Last CrusadeÂ	8.3	515306
91	Good Will HuntingÂ	8.3	604904
92	UpÂ	8.3	665575
93	UnforgivenÂ	8.3	277505
94	Batman BeginsÂ	8.3	980946
95	Inglourious BasterdsÂ	8.3	885175
96	2001: A Space OdysseyÂ	8.3	427357
97	AmadeusÂ	8.3	270790
98	L.A. ConfidentialÂ	8.3	414219
99	SnatchÂ	8.3	600996
100	Some Like It HotÂ	8.3	175196
101	ScarfaceÂ	8.3	537442
102	Eternal Sunshine of the Spotless MindÂ	8.3	666937
103	RoomÂ	8.3	161288
104	Monty Python and the Holy GrailÂ	8.3	382240

105	LifeÂ	8.3	29450
106	The Great EscapeÂ	8.3	165638
107	The ApartmentÂ	8.3	109335
108	Judgment at NurembergÂ	8.3	44457
109	Singin' in the RainÂ	8.3	150020
110	Inside JobÂ	8.3	55382
111	Taxi DriverÂ	8.3	507063

-	The HuntÂ	8.3	170155
	MetropolisÂ	8.3	111841
114	DownfallÂ	8.3	248354
115	Raging BullÂ	8.3	235133
116	Finding NemoÂ	8.2	692482
117	Gone with the WindÂ	8.2	215340
118	Captain America: Civil WarÂ	8.2	272670
119	Gran TorinoÂ	8.2	561773
120	A Beautiful MindÂ	8.2	610568
121	Die HardÂ	8.2	592582
122	How to Train Your DragonÂ	8.2	485430
123	The Bridge on the River KwaiÂ	8.2	149444
124	Pan's LabyrinthÂ	8.2	467234
125	The Secret in Their EyesÂ	8.2	131831
126	The Wolf of Wall StreetÂ	8.2	780588
127	V for VendettaÂ	8.2	791783
128	TrainspottingÂ	8.2	469561
129	On the WaterfrontÂ	8.2	100890
130	Into the WildÂ	8.2	426359
131	Lock, Stock and Two Smoking BarrelsÂ	8.2	414976
132	The Big LebowskiÂ	8.2	537419
133	IncendiesÂ	8.2	80429
134	The Deer HunterÂ	8.2	232577
135	Buffy the Vampire SlayerÂ	8.2	101902
136	The Elephant ManÂ	8.2	161972
	Lage Raho Munna BhaiÂ	8.2	27569
138	Mr. Smith Goes to WashingtonÂ	8.2	77392

139	RebeccaÂ	8.2	87424
140	It Happened One NightÂ	8.2	64888
141	Blade RunnerÂ	8.2	461609
142	The ThingÂ	8.2	258078
143	CasinoÂ	8.2	333542
144	WarriorÂ	8.2	332276
145	Howl's Moving CastleÂ	8.2	214091
146	The AvengersÂ	8.1	995415
147	DeadpoolÂ	8.1	479047
148	Jurassic ParkÂ	8.1	613473
149	The Sixth SenseÂ	8.1	704766
150	Monsters, Inc.Â	8.1	585659
151	Pirates of the Caribbean: The Curse of the Blac	8.1	809474

	-		
	Pirates of the Caribbean: The Curse of the Blad	8.1	809474
	Guardians of the GalaxyÂ	8.1	682155
	The HelpÂ	8.1	318955
	PlatoonÂ	8.1	291603
-	The MartianÂ	8.1	472488
	The Bourne UltimatumÂ	8.1	491077
	RockyÂ	8.1	375240
158	Gone GirlÂ	8.1	569841
159	Butch Cassidy and the Sundance KidÂ	8.1	152089
160	The Imitation GameÂ	8.1	467613
	Million Dollar BabyÂ	8.1	482064
162	The Truman ShowÂ	8.1	667983
163	Groundhog DayÂ	8.1	437418
164	No Country for Old MenÂ	8.1	612060
165	The RevenantÂ	8.1	406020
166	Shutter IslandÂ	8.1	786092
167	Stand by MeÂ	8.1	271794
168	Kill Bill: Vol. 1Â	8.1	735784
169	12 Years a SlaveÂ	8.1	439176
170	Annie HallÂ	8.1	192940
171	Sin CityÂ	8.1	656640
172	The Grand Budapest HotelÂ	8.1	475518
173	The TerminatorÂ	8.1	600266
174	SpotlightÂ	8.1	195333
175	The Best Years of Our LivesÂ	8.1	40359
176	The Wizard of OzÂ	8.1	291875
177	There Will Be BloodÂ	8.1	372990
178	PrisonersÂ	8.1	383591
179	The Princess BrideÂ	8.1	294163
180	Hotel RwandaÂ	8.1	264533
181	Mad Max: Fury RoadÂ	8.1	552503
182	Amores PerrosÂ	8.1	173551
183	Before SunriseÂ	8.1	183288
184	The CelebrationÂ	8.1	65951
185	SolarisÂ	8.1	54057
186	GandhiÂ	8.1	171726
187	Hachi: A Dog's TaleÂ	8.1	155249
188	Barry LyndonÂ	8.1	101627
189	NetworkÂ	8.1	103493
190	A Christmas StoryÂ	8.1	104908
191	The Man Who Shot Liberty ValanceÂ	8.1	53741
192	Cat on a Hot Tin RoofÂ	8.1	33741

	Touching the VoidÂ	8.1	26926
	High NoonÂ	8.1	80193
	Donnie DarkoÂ	8.1	580999
	Elite SquadÂ	8.1	81644
	The Sea InsideÂ	8.1	64556
198	RushÂ	8.1	312629
199	Tae Guk Gi: The Brotherhood of WarÂ	8.1	31943
200	AkiraÂ	8.1	106160
201	JawsÂ	8	412454
202	The ExorcistÂ	8	284252
203	AladdinÂ	8	260939
204	The IncrediblesÂ	8	479166
205	Dances with WolvesÂ	8	186485
206	The Sound of MusicÂ	8	148172
207	Rain ManÂ	8	383784
208	Slumdog MillionaireÂ	8	641997
209	The King's SpeechÂ	8	503631
210	Catch Me If You CanÂ	8	525801
211	Star TrekÂ	8	504419
212	The Pursuit of HappynessÂ	8	338383
213	Doctor ZhivagoÂ	8	55816
214	Black SwanÂ	8	551363
215	District 9Â	8	531737
216	Young FrankensteinÂ	8	112671
217	Dead Poets SocietyÂ	8	277451
218	Mystic RiverÂ	8	338415
219	RatatouilleÂ	8	473887
		•	
220	Fiddler on the RoofÂ	8	29839
	Kill Bill: Vol. 2Â	8	512749
	X-Men: Days of Future PastÂ	8	514125
	JFKÂ	8	113472
224	The ArtistÂ	8	190030
	Sling BladeÂ	8	72443
	Dallas Buyers ClubÂ	8	326494
	BoyhoodÂ	8	266020
	Bowling for ColumbineÂ	8	123090
_	Casino RoyaleÂ	8	470483
	Casino RoyaleÂ	8	470501
	SickoÂ	8	66610
	Shaun of the DeadÂ	8	395921
	Life of PiÂ	8	440084
	The Perks of Being a WallflowerÂ	8	351274
	A Fistful of DollarsÂ	8	147566
	Before SunsetÂ	8	168398

237	Central StationÂ	8	28951
238	HerÂ	8	355126
239	Waltz with BashirÂ	8	46107
240	The ReturnÂ	8	31589
241	The Diving Bell and the ButterflyÂ	8	89906
242	PattonÂ	8	76398
243	The Wild BunchÂ	8	63192
244	Night of the Living DeadÂ	8	87978
245	Rosemary's BabyÂ	8	140527
246	Days of HeavenÂ	8	37594
247	The HustlerÂ	8	62860
248	A Streetcar Named DesireÂ	8	78454
249	The Man from EarthÂ	8	129799
250	True RomanceÂ	8	163492

Top_Foreign_Lang_Film:

rank	Top_Foreign_Lang_Film	imdb_score 🚭	num_voted_users 🔻	language 🍱
1	The Good, the Bad and the UglyÂ	8.9	503509	Italian
2	City of GodÂ	8.7	533200	Portuguese
3	Seven SamuraiÂ	8.7	229012	Japanese
4	Spirited AwayÂ	8.6	417971	Japanese
5	The Lives of OthersÂ	8.5	259379	German
6	Children of HeavenÂ	8.5	27882	Persian
7	AirliftÂ	8.5	30977	Hindi
8	A SeparationÂ	8.4	151812	Persian
9	Rang De BasantiÂ	8.4	70233	Hindi
10	OldboyÂ	8.4	356181	Korean
11	Das BootÂ	8.4	168203	German
13	Baahubali: The BeginningÂ	8.4	62756	Telugu
14	AmélieÂ	8.4	534262	French
15	Princess MononokeÂ	8.4	221552	Japanese
16	The HuntÂ	8.3	170155	Danish
17	MetropolisÂ	8.3	111841	German
18	DownfallÂ	8.3	248354	German
19	Pan's LabyrinthÂ	8.2	467234	Spanish
20	The Secret in Their EyesÂ	8.2	131831	Spanish
21	IncendiesÂ	8.2	80429	French
22	Lage Raho Munna BhaiÂ	8.2	27569	Hindi
23	Howl's Moving CastleÂ	8.2	214091	Japanese
24	Amores PerrosÂ	8.1	173551	Spanish
25	The CelebrationÂ	8.1	65951	Danish
26	SolarisÂ	8.1	54057	Russian
27	Elite SquadÂ	8.1	81644	Portuguese
28	The Sea InsideÂ	8.1	64556	Spanish
29	Tae Guk Gi: The Brotherhood of WarÂ	8.1	31943	Korean

30	AkiraÂ	8.1	106160	Japanese
31	A Fistful of DollarsÂ	8	147566	Italian
32	Central StationÂ	8	28951	Portuguese
33	Waltz with BashirÂ	8	46107	Hebrew
34	The ReturnÂ	8	31589	Russian
35	The Diving Bell and the ButterflyÂ	8	89906	French

D. **Best Directors:** Group the column using the director_name column.

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

Soln:

Query:

```
use ibdm;
select director_name as top10director, avg(imdb_score) as imdb_score
from ibdm.imdb_movies
group by top10director
order by imdb_score desc
limit 10;
```

Result:

top10director	imdb_score
Christopher Nolan	8.414285714285715
S.S. Rajamouli	8.4
Lee Unkrich	8.3
Pete Docter	8.233333333333334
Hideaki Anno	8.2
Quentin Tarantino	8.16
Denis Villeneuve	8.1
Tim Miller	8.1
Andrei Tarkovsky	8.1
Alejandro G. IñÃirritu	8.1

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

Soln:

Query:

```
select distinct genres, count(genres) as genresCount
from ibdm.imdb_movies
group by genres
order by genresCount desc
limit 10;
```

Result:

genres	genresCount
Action Adventure Sci-Fi	44
Comedy Romance	39
Comedy	35
Action Crime Thriller	31
Comedy Drama Romance	31
Action Adventure Thriller	29
Crime Drama Thriller	27
Action Adventure Sci-Fi Thriller	26
Adventure Animation Comedy Family Fantasy	24
Action Adventure Fantasy	22

F. **Charts:** Create three new columns namely, Meryl_Streep, Leo_Caprio, and Brad_Pitt which contain the movies in which the actors: 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' are the lead actors. Use only the actor_1_name column for extraction. Also, make sure that you use the names 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt' for the said extraction.

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

Group the combined column using the actor_1_name column.

Find the mean of the num_critic_for_reviews and num_users_for_review and identify the actors which have the highest mean.

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

Query 1: Critic favourite

```
select actor_1_name critic_favorite, avg(num_critic_for_reviews) as criticReviewCount
from ibdm.imdb_movies
group by critic_favorite
order by criticReviewCount desc
limit 1;
```

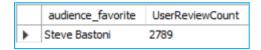
Output:

critic_favorite	criticReviewCount
Albert Finney	750

Query 2: Audience Favourite

```
select actor_1_name audience_favorite, avg(num_user_for_reviews) as UserReviewCount
from ibdm.imdb_movies
group by audience_favorite
order by UserReviewCount desc
limit 1;
```

Output:



Result:

• In this project I have gain practical hands on knowledge to analyse numerical data in an Excel spreadsheet.

•	Learnt various MS Excel functions and formulas that can be used in many companies in day to day analysis data record in their company