IMDb Movie Analysis

Final Project 1



Project Description

- Thousands of movies from all over the world are included in the IMDb movie database.
- It contains all the necessary information about the movies, including the title, year, director, genre, cast, box office revenue, budget, IMDb score, user and critic reviews, and other specifics.
- This information is analyzed to determine the success of the films, the popularity of the actors, the genres, and the directors.
- It can also be used to examine historical trends.

Information I gathered from this database

- Movies with the highest profit.
- Top 250 movies with highest IMDb ratings.
- Top foreign language movies with highest IMDb ratings.
- Best directors with highest mean of IMDb scores.
- Popular genres.
- Best critic and audience favorite actor.
- Trend of number of users voted across decades.

Approach

- First, I examined the data set, including all of the various categories in the sheet like budgets, gross, language, actor, director, IMDb data provided like critic reviews, user reviews, votes etc. to gain a sense of the information accessible to me.
- Then I cleaned the data by removing null values and dropping irrelevant columns.
- Then, I created a profit column and looked at the outliers to determine the lowest and highest in the range.
- Then I examined and comprehended all of the questions asked and summarized my method to solving these questions. Then I applied various methods and formulas in excel sheet and solved the questions.

Tech Stack Used

• I selected Microsoft Excel v2013 to find the answers to the questions and to make all the graphs and charts.



Movies with Highest Profit

• For finding Profit = gross – budget

f_x =12-AE2								
	Н	1	AE	AF	AM	AN	A	
actor_	1_facebook_lik *	gross	budget	budget in millions	profit	profit in millions		
	11000	28341469	25000000	25	3341469	3.341469		
	14000	134821952	6000000	6	128821952	128.821952		
	23000	533316061	185000000	185	348316061	348.316061		
	22000	57300000	13000000	13	44300000	44.3		
	5000	377019252	94000000	94	283019252	283.019252		
	14000	96067179	22000000	22	74067179	74.067179		
	42000	407020000	2222222		00000000	00.00		

• After sorting profit column, got the top 20 highest profit movies.



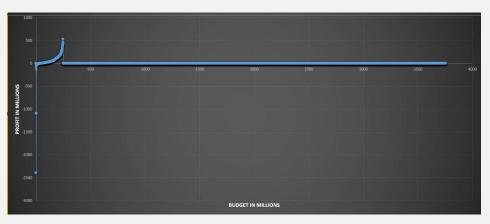
movie_title	budget	budget in millions	imdb_score *	profit 🚚	profit in millions
AvatarÂ	237000000	237	7.9	523505847	523.505847
Star Wars: Episode IV - A New HopeÂ	11000000	11	8.7	449935665	449.935665
E.T. the Extra-TerrestrialÂ	10500000	10.5	7.9	424449459	424.449459
The AvengersÂ	220000000	220	8.1	403279547	403.279547
The AvengersÂ	220000000	220	8.1	403279547	403.279547
The Lion KingÂ	45000000	45	8.5	377783777	377.783777
The Dark KnightÂ	185000000	185	9	348316061	348.316061
DeadpoolÂ	58000000	58	8.1	305024263	305.024263
Jurassic ParkÂ	63000000	63	8.1	293784000	293.784
Finding NemoÂ	94000000	94	8.2	286838870	286.83887
The Lord of the Rings: The Return of the Kir	94000000	94	8.9	283019252	283.019252
Star Wars: Episode VI - Return of the JediÂ	32500000	32.5	8.4	276625409	276.625409
Forrest GumpÂ	55000000	55	8.8	274691196	274.691196
Star Wars: Episode V - The Empire Strikes E	18000000	18	8.8	272158751	272.158751
The Sixth SenseÂ	40000000	40	8.1	253501675	253.501675
JawsÂ	8000000	8	8	252000000	252
The Lord of the Rings: The Two TowersÂ	94000000	94	8.7	246478898	246.478898
Raiders of the Lost ArkÂ	18000000	18	8.5	224374454	224.374454
The Lord of the Rings: The Fellowship of the	93000000	93	8.8	220837577	220.837577
Toy Story 3Â	200000000	200	8.3	214984497	214.984497

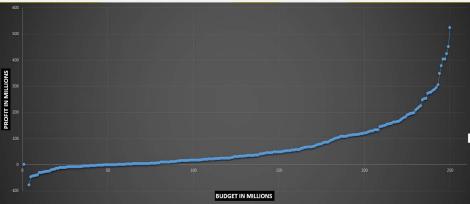
Profit vs Budget

• This graph between profit & budget shows that there are few movies with negative profit which are the outliers for our data.

• After removing outliers the graph looks like this which is a left skewed.

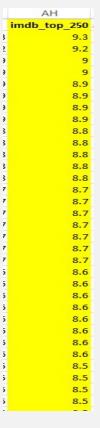
This tells us that as the budget of the movie increased, the profit also increased.





Finding IMDb Top 250

After sorting IMDb score column in descending order



 To find movies where num_voted_users is greater than 25,000 in descending order

 f_x =IF(Q2>25000,Q2,0)

Q	R
num_voted_users	num_voted_user>25k
55597	55597
126893	126893
222099	222099
26413	26413
5557	0
22823	0
439176	439176
279179	279179
122187	122187
47764	47764
209396	209396
42547	42547
112769	112769
146899	146899
4670	0
191912	191912
156898	156898
1622	0
9540	0
13010	0
427357	427357
283418	283418
9010	0
41496	41496
60407	60407
189683	189683

Adding Rank to Top 250 Movies

 To add rank column for top 250 movies on basis of their IMDb scores

 f_x =RANK(AH2,(\$AH\$2:\$AH\$251),0)



N	P	AH	
Rank	Top_250_movies	imdb_top_	250
1	The Shawshank RedemptionÂ		9.3
2	The GodfatherÂ		9.2
3	The Dark KnightÂ		9
3	The Godfather: Part IIÂ		9
5	The Lord of the Rings: The Return of the		8.9
5	Schindler's ListÂ		8.9
5	Pulp FictionÂ		8.9
5	The Good, the Bad and the UglyÂ		8.9
9	InceptionÂ		8.8
9	The Lord of the Rings: The Fellowship o		8.8
9	Fight ClubÂ		8.8
9	Forrest GumpÂ		8.8
9	Star Wars: Episode V - The Empire Strike		8.8
14	The Lord of the Rings: The Two TowersÂ		8.7
14	The MatrixÂ		8.7
14	GoodfellasÂ		8.7
14	Star Wars: Episode IV - A New HopeÂ		8.7
14	One Flew Over the Cuckoo's NestÂ		8.7
14	City of GodÂ		8.7
14	Seven SamuraiÂ		8.7
21	InterstellarÂ		8.6
21	Saving Private RyanÂ		8.6
21	Se7enÂ		8.6
21	The Silence of the LambsÂ		8.6
21	Spirited AwayÂ		8.6
21	American History XÂ		8.6
21	The Usual SuspectsÂ		8.6
21	Modern TimesÂ		8.6
29	The Dark Knight RisesÂ		8.5
29	GladiatorÂ		8.5
29	Terminator 2: Judgment DayÂ		8.5
29	Django UnchainedÂ		8.5
29	The DepartedÂ		8.5

104 Central StationA	8
154 Young FrankensteinÂ	8
154 Before SunsetÂ	8
154 Waltz with BashirÂ	8
154 Sling BladeÂ	8
154 A Fistful of DollarsÂ	8
207 AvatarÂ	7.9
207 The Hobbit: The Desolation of SmaugÂ	7.9
207 Edge of TomorrowÂ	7.9
207 Big Hero 6Â	7.9
207 How to Train Your Dragon 2Â	7.9
207 Iron ManÂ	7.9
207 Toy Story 2Â	7.9
207 Children of MenÂ	7.9
207 The InsiderÂ	7.9
207 ShrekÂ	7.9
207 The Bourne IdentityÂ	7.9
207 Almost FamousÂ	7.9
207 Captain PhillipsÂ	7.9
207 The Hateful EightÂ	7.9
207 HeroÂ	7.9
207 HeroÂ	7.9
207 The NotebookÂ	7.9
207 Straight Outta ComptonÂ	7.9
207 Walk the LineÂ	7.9
207 The Blues BrothersÂ	7.9
207 The Right StuffÂ	7.9
207 TakenÂ	7.9
207 The FighterÂ	7.9

Finding Top Foreign Language Movies

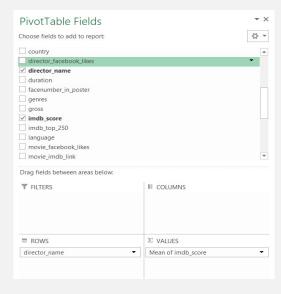
 To find foreign language movies from top 250 movies After removing zeroes we got the desired results



Top Foreign Lang Film The Good, the Bad and the UglyÂ City of GodÂ Seven SamuraiÂ Spirited AwayÂ The Lives of OthersÂ Children of HeavenÂ Princess MononokeÂ Amã@lieÂ Das BootÂ OldboyÂ A SeparationÂ DownfallÂ « Metropolis Â Î The HuntÂ Howl's Moving CastleÂ Pan's LabyrinthÂ IncendiesÂ The Secret in Their EyesÂ AkiraÂ Tae Guk Gi: The Brotherhood of WarÂ The Sea InsideÂ Elite SquadÂ Amores PerrosÂ The CelebrationÂ My Name Is KhanÂ PersepolisÂ Central StationÂ Waltz with BashirÂ A Fistful of DollarsÂ HeroÂ HeroÂ Letters from Iwo JimaÂ Crouching Tiger, Hidden DragonÂ AmourÂ The ChorusÂ

Finding Best Directors

• To find best directors, created a pivot table using director_name column and IMDb mean score

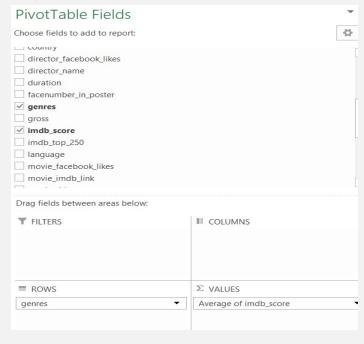


Top Directors	✓ Mean of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Alfred Hitchcock	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Richard Marquand	8.4
Asghar Farhadi	8.4
Lenny Abrahamson	8.3
Lee Unkrich	8.3
Fritz Lang	8.3
Billy Wilder	8.3
Pete Docter	8.233333333
Hayao Miyazaki	8.225
Quentin Tarantino	8.2
Joshua Oppenheimer	8.2
Juan José Campanella	8.2
George Roy Hill	8.2



Finding Popular Genres

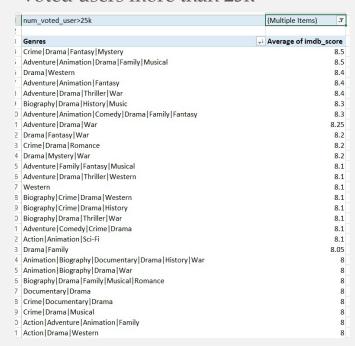
• To find popular genres, first I created a pivot table on basis of genres and average of IMDb score for each of them and sorted the average IMDb score column in descending to get the top scorers.



Genres	Average of imdb_score
Crime Drama Fantasy Mystery	8.5
Adventure Animation Drama Family Musical	8.5
Adventure Animation Fantasy	8.4
Adventure Drama Thriller War	8.4
Documentary Drama Sport	8.3
Biography Drama History Music	8.3
Adventure Animation Comedy Drama Family Fantasy	8.3
Adventure Drama War	8.25
Biography Crime Documentary History	8.2
Drama Fantasy War	8.2
Drama Mystery War	8.2
Documentary History	8.1
Biography Crime Drama History	8.1
Documentary History Music	8.1
Biography Crime Drama Western	8.1
Western	8.1
Action Animation Sci-Fi	8.1
Adventure Comedy Crime Drama	8.1
Adventure Drama Thriller Western	8.1
Biography Drama Family Musical Romance	8
Animation Biography Documentary Drama History War	8
Animation Biography Drama War	8
Crime Documentary Drama	8
Crime Drama Musical	8
Action Adventure Animation Family	8
Drama Family Musical Romance	7.95
Action Adventure Animation Comedy Drama Family Sci-Fi	7.95
Fantasy Romance	7.9
er i let le lure les t	7.0

Top 20 Genres

 Now filter the data by num_voted_user>25k to get the genres with voted users more than 25k



• These are the top 20 genres according to their IMDb average scores with rank.

Rank	Genres	→ I Average of imdb_score
1	Crime Drama Fantasy Mystery	8.5
1	Adventure Animation Drama Family Musical	8.5
3	Drama Western	8.4
3	Adventure Animation Fantasy	8.4
3	Adventure Drama Thriller War	8.4
6	Biography Drama History Music	8.3
6	Adventure Animation Comedy Drama Family Fantasy	8.3
8	Adventure Drama War	8.25
9	Drama Fantasy War	8.2
9	Crime Drama Romance	8.2
9	Drama Mystery War	8.2
12	Adventure Family Fantasy Musical	8.1
12	Adventure Drama Thriller Western	8.1
12	Western	8.1
12	Biography Crime Drama Western	8.1
12	Biography Crime Drama History	8.1
12	Biography Drama Thriller War	8.1
12	Adventure Comedy Crime Drama	8.1
12	Action Animation Sci-Fi	8.1
20	Drama Family	8.05

Critic & User Favorite Actor

Meryl Streep Column

• For meryl_streep column



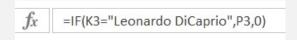
• Filtering out zeroes we get this movie list for Meryl Streep

K		L		P	
actor 1 name		meryl_streep ▼ I		movie_title	
Robert De Niro		0 N		Midnight RunÂ	
Harrison Ford		0 1		Indiana Jones and the Temple of DoomÂ	
Kate Winslet		0		Little ChildrenÂ	
Meryl Streep		The HoursÂ		The HoursÂ	
Jeff Bridges		0		The Fisher KingÂ	
Michael Wincott		0		The CrowÂ	
				^	



Leonardo DiCaprio Column

• For leo_caprio column



 Filtering out zeroes we get this movie list for Leonardo DiCaprio



K	M			
actor_1_name ▼		leo_caprio -T		
Leonardo DiCaprio		InceptionÂ		
Leonardo DiCaprio		Django UnchainedÂ		
Leonardo DiCaprio		The DepartedÂ		
Leonardo DiCaprio		The Wolf of Wall StreetÂ		
Leonardo DiCaprio		The RevenantÂ		
Leonardo DiCaprio		Shutter IslandÂ		
Leonardo DiCaprio		Blood DiamondÂ		
Leonardo DiCaprio		Catch Me If You CanÂ		
Leonardo DiCaprio		TitanicÂ		
Leonardo DiCaprio		The AviatorÂ		
Leonardo DiCaprio		Gangs of New YorkÂ		
Leonardo DiCaprio		The Great GatsbyÂ		
Leonardo DiCaprio		The Great GatsbyÂ		
Leonardo DiCaprio		Revolutionary RoadÂ		
Leonardo DiCaprio		Body of LiesÂ		
Leonardo DiCaprio		Romeo + JulietÂ		
Leonardo DiCaprio		The BeachÂ		
Leonardo DiCaprio		J. EdgarÂ		
Leonardo DiCaprio		The Man in the Iron MaskÂ		
Leonardo DiCaprio		The Quick and the DeadÂ		
Leonardo DiCaprio		Marvin's RoomÂ		

Brad Pitt Column

For brad_pitt column

$$f_{\mathcal{X}}$$
 =IF(K3="Brad Pitt",P3,0)

• Filtering out zeroes we get this movie list for Brad Pitt

Brad Pitt

Brad Pitt

Brad Pitt Brad Pitt

Brad Pitt



ith the Vampire: The Vampire I on of Jesse James by the Cowar

BabelÂ

TroyÂ

Spy GameÂ

Seven Years in TibetÂ

The Tree of LifeÂ

Mr. & Mrs. SmithÂ

Ocean's TwelveÂ

Killing Them SoftlyÂ

By the SeaÂ

ibad: Legend of the Seven Sea 5

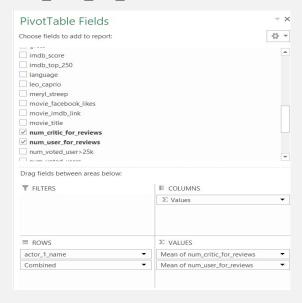
Combined Column

• Combined column for these three columns

Combined The HoursÂ Out of AfricaÂ Julie & JuliaÂ The Devil Wears PradaÂ It's ComplicatedÂ The Iron LadyÂ The River WildÂ Hope SpringsÂ Lions for LambsÂ One True ThingÂ A Prairie Home CompanionÂ InceptionÂ Django UnchainedÂ The DepartedÂ The Wolf of Wall StreetÂ The RevenantÂ Shutter IslandÂ Blood DiamondÂ Catch Me If You CanÂ TitanicÂ The AviatorÂ Gangs of New YorkÂ The Great GatsbyÂ The Great GatsbyÂ Revolutionary RoadÂ Body of LiesÂ Romeo + JulietÂ The BeachÂ J. EdgarÂ The Man in the Iron MaskÂ The Quick and the DeadÂ Marvin's RoomÂ Fight ClubÂ True RomanceÂ The Curious Case of Benjamin ButtonÂ Ocean's ElevenÂ FuryÂ Interview with the Vampire: The Vampire ChroniclesÂ ne Assassination of Jesse James by the Coward Robert Ford.

Mean of Critic & User Reviews

For critic favourite actor & audience favourite actor,
 I created a pivot table using combined column,
 actor_1_name column &
 Mean of num_critic_for_reviews &
 num_user_for_reviews column.



	actor 1 name	■ Mean of num_critic_for_	Mean of num user for re
	■ Leonardo DiCaprio	330.1904762	914.4761905
	Blood DiamondÅ	166	657
	Body of LiesÅ	238	263
	Catch Me If You CanA	194	667
	Django UnchainedÅ	765	1193
	Gangs of New YorkÅ	233	1166
	InceptionÅ	642	2803
	J. EdgarÅ	392	279
	Marvin's RoomÅ	45	71
	Revolutionary RoadÅ	323	414
	Romeo + JulietÅ	106	506
	Shutter Islandå	490	964
	The AviatorÅ	267	799
	The BeachÅ	118	548
	The DepartedÅ	352	2054
	The Great GatsbyÅ	490	753
1	The Man in the Iron MaskÅ	83	244
	The Quick and the DeadÅ	63	216
:	The RevenantÅ	556	1188
1	The Wolf of Wall StreetÅ	606	1138
	TitanicÅ	315	2528
;	⊕ Brad Pitt	245	742.3529412
i	BabelÅ	285	908
	By the SeaÅ	131	61
1	Fight ClubÅ	315	2968
1	FuryÅ	406	701
1	Interview with the Vampire: The Vampire Chronicl	esi 120	406
	Killing Them SoftlvÅ	414	369
1	Mr. & Mrs. SmithÅ	233	798
1	Ocean's ElevenÅ	186	845
	Ocean's TwelveÅ	198	627
;	Seven Years in TibetÅ	76	119
1	Sinbad: Legend of the Seven SeasÅ	98	91
•	Spv GameÅ	142	361
	The Assassination of Jesse James by the Coward	Rc 273	415
1	The Curious Case of Benjamin ButtonA	362	822
1	The Tree of LifeÅ	584	975
	TrouÅ	220	1694
	True RomanceÂ	122	460
-	Meryl Streep	181,4545455	297.1818182
Γ	A Prairie Home CompanionĂ	211	280
ï	Hope SpringsÅ	234	178
i	It's ComplicatedÅ	187	214
1	Julie & JuliaÅ	252	277
1	Lions for LambsÅ	227	298
1	One True ThingÅ	64	112
i.	Out of AfricaA	66	200
	The Devil Wears PradaÅ	208	631
:	The HoursÅ	174	660

Favorite Actor Conclusion

• After sorting on mean columns, we got the below results:

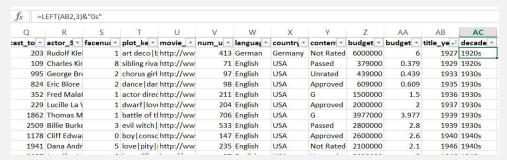
actor_1_name 🚽 Mea	an of num_critic_for_reviews Mean of r	num_user_for_reviews
⊞ Leonardo DiCaprio	330.1904762	914.4761905
⊕ Brad Pitt	245	742.3529412
⊞ Meryl Streep	181.4545455	297.1818182

• This shows that "Leonardo DiCaprio" is both Critic and Audience favourite actor.



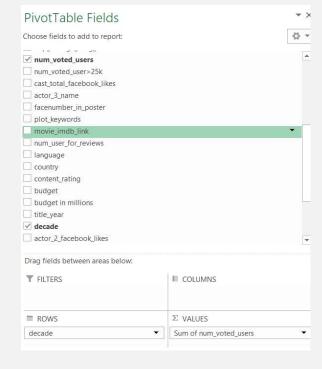
Trend of Voted Users over the Decades

• For decade column:



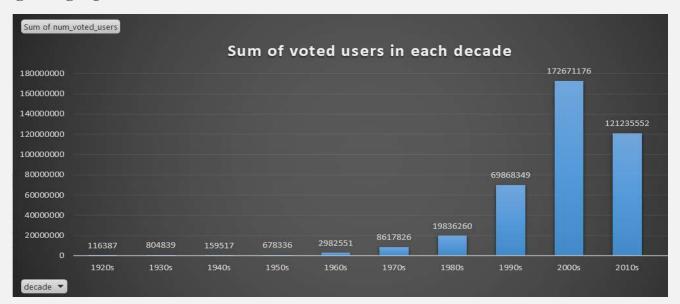
- After sorting decade column, I created a pivot table grouping decade and num_voted_users column.
- The table shows sum of voted users in each decade.

decade	7	Sum of num_voted_users
1920s		116387
1930s		804839
1940s		159517
1950s		678336
1960s		2982551
1970s		8617826
1980s		19836260
1990s		69868349
2000s		172671176
2010s		121235552



Change in Number of Voted Users over Decades

• Plotting bar graph between "decade" & "Sum of num_voted_users" columns.



• This graph shows that there is a significant increase in number of voted users after 1990s & in 2000s.

Results

- After doing the analysis on IMDb database, I was able to find the following results and insights:
- Most profitable movies.
- Higher scorer IMDb movies.
- Foreign language movies with best IMDb ratings.
- Top directors.
- Popular genres.
- Most favorite actor.
- Trend of number of users voted across decades.

THANK YOU