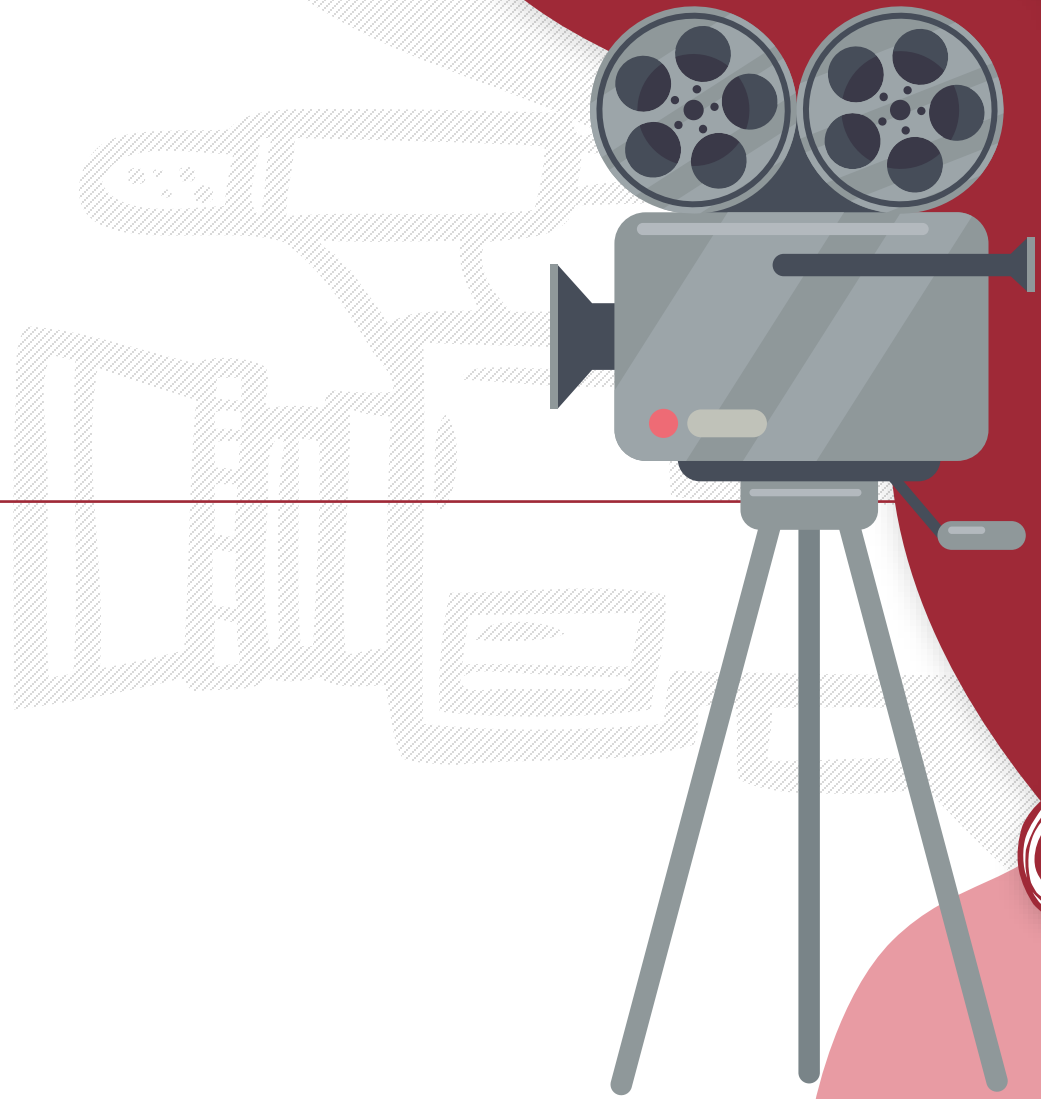


IMDB MOVIE ANALYSIS

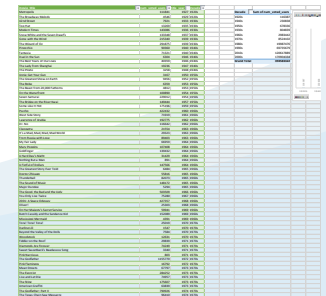


PROJECT DESCRIPTION

- PROJECT is **ABOUT analysis of IMDB Movie dataset.**
- **Purpose to carried out Analysis : (Assuming)**
 - **Event** : Award function is going to happen in New York
 - **Category for award:**
 - Who is best actor ?
 - Who is best director ?
 - Which movie was popular?
 - Which type of movies people's are interested ?
- To give details to this event management team, I am working as data analytics.

PROJECT DESCRIPTION

- The dataset and few questions listed below have helped me to reach the result :
 - Which **movie has highest profit** ?
 - Which are **top 250 movies** with the highest IMDB rating ?
 - Among **top 250 movies** are **except English** Language?
 - Who are the **best director**?
 - Which **genres** are popular?
 - Which **actor's** movie is most popular and leading ?
 - Which actor have the **highest mean of citric for reviews and users for review**?
 - How number of voted **users changes over decade**?
 - This is the [excel sheet](#) which I have worked out to carry out data analysis. (**Double Click on image**)



Movie Title	Year	Genre	Rating	Profit	Director	Actor	Mean Citric	Users for Review
Avatar	2009	Action	7.8	208353589	James Cameron	Sam Worthington	7.8	1000000
Avengers: Endgame	2019	Action	8.4	678811243	Anthony Russo	Robert Downey Jr.	8.4	1000000
Star Wars: The Force Awakens	2015	Adventure	7.8	94222424	J.J. Abrams	Daisy Ridley	7.8	1000000
Avengers: Infinity War	2018	Action	8.4	678811243	Anthony Russo	Robert Downey Jr.	8.4	1000000
Star Wars: The Last Jedi	2017	Adventure	6.9	69541524	Rian Johnson	Daisy Ridley	6.9	1000000
Star Wars: The Rise of Skywalker	2019	Adventure	6.4	66963524	J.J. Abrams	Daisy Ridley	6.4	1000000
Star Wars: The Force Awakens	2015	Adventure	7.8	94222424	J.J. Abrams	Daisy Ridley	7.8	1000000
Star Wars: The Last Jedi	2017	Adventure	6.9	69541524	Rian Johnson	Daisy Ridley	6.9	1000000
Star Wars: The Rise of Skywalker	2019	Adventure	6.4	66963524	J.J. Abrams	Daisy Ridley	6.4	1000000
Star Wars: The Force Awakens	2015	Adventure	7.8	94222424	J.J. Abrams	Daisy Ridley	7.8	1000000
Star Wars: The Last Jedi	2017	Adventure	6.9	69541524	Rian Johnson	Daisy Ridley	6.9	1000000
Star Wars: The Rise of Skywalker	2019	Adventure	6.4	66963524	J.J. Abrams	Daisy Ridley	6.4	1000000

APPROACH



- **First step:**
 - Observing data
 - Deleting non-required columns from the original provided dataset based on task given.
- **Second step: (cleaning data)**
 - Removing duplicates from the dataset.
 - Checking for missing values
 - Imputing missing values wherever it is required based on task
- **Third step:**
 - Arranging data in proper manner to complete the task given.
 - i.e., Ascending or descending order, correcting names, etc., creating columns.
- **Fourth step:**
 - Using required software, project is completed.



TECH-STACK USED



TASK



A. Cleaning the data



B. Movies with highest profit



C. Top 250 & Top 250 by language



D. Best top 10 director



E. Popular Genres



F. Charts



Task:1 :Cleaning Data

Before cleaning

Columns: 28
Rows: 5043

After cleaning

Columns: 12
Rows: 3849

Duplicates

47 Rows

Dataset


It is ready for
performing
analysis.

Task – 2 : Movies with highest profit

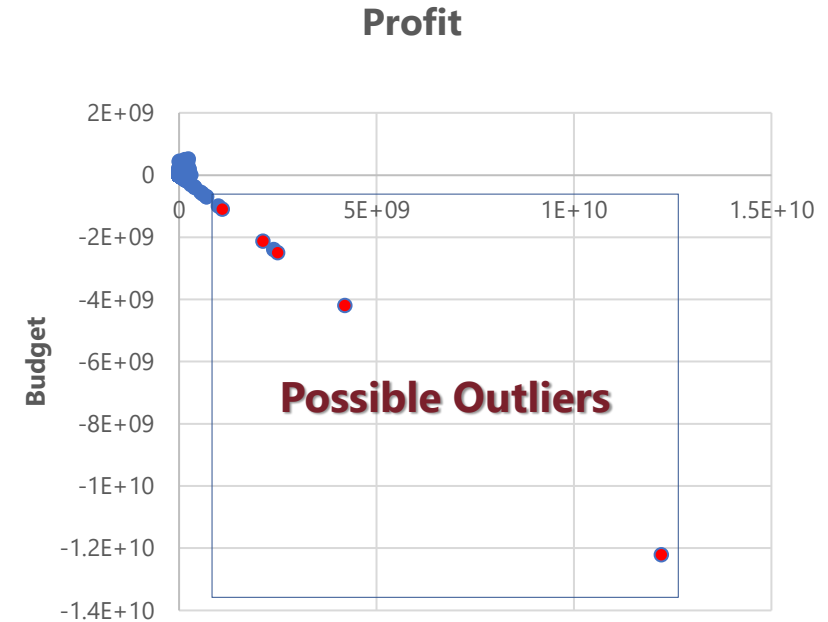
Steps followed:

- Created new column as profit
- Ascending by highest to lowest

Profit
=[@gross]-[@budget]



Movie	Profit
Avatar	523505847
Jurassic World	502177271
Titanic	458672302
Star Wars: Episode IV - A New Hope	449935665
E.T. the Extra-Terrestrial	424449459



Task – 3(A) : IMDB Top 250



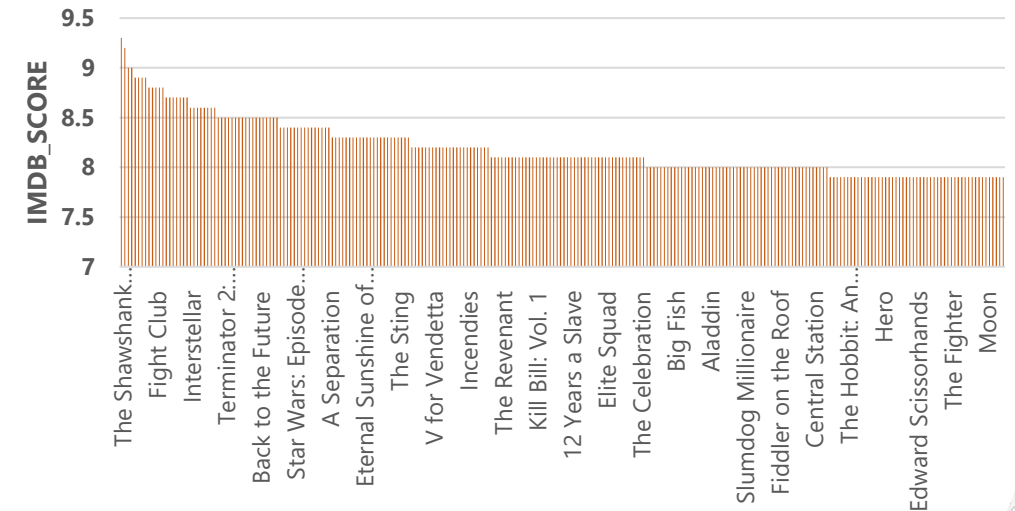
IMDB_TOP_250	num_voted_users	language	imdb_score	Rank
The Shawshank Redemption	1689764	English	9.3	1
The Godfather	1155770	English	9.2	2
The Dark Knight	1676169	English	9	3
The Godfather: Part II	790926	English	9	3
The Lord of the Rings: The Return of the King	1215718	English	8.9	5
Schindler's List	865020	English	8.9	5
Pulp Fiction	1324680	English	8.9	5
The Good, the Bad and the Ugly	503509	Italian	8.9	5
Inception	1468200	English	8.8	9
The Lord of the Rings: The Fellowship of the Ring	1238746	English	8.8	9

Steps followed:

- Movies are sorted by descending order from highest to lowest imdb_score.
- Dataset are filtered based on num_voted_users > 25000.
- Rank is given using the equation RANK.

D	E	F	G	H
imdb_score	Rank			
9.3	=RANK([@[imdb_score]], \$D\$2:\$D\$262, 0)			

Top_250_Movies

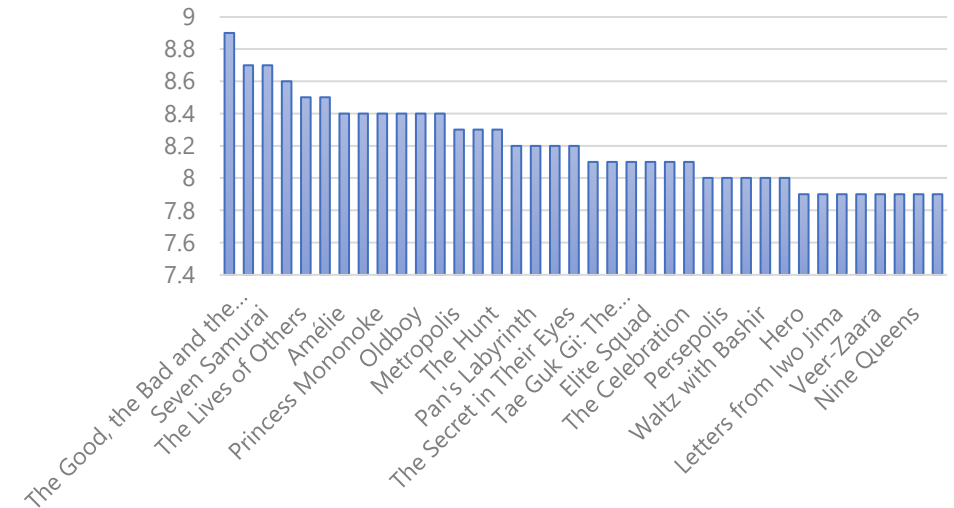


Task – 3(B) : IMDB Top 250 except English



Movie_title	language	imdb_score	Rank	Languages	Count of imdb_score
The Good, the Bad and the Ugly	Italian	8.9	5	Aboriginal	2
City of God	Portuguese	8.7	14	Arabic	1
Seven Samurai	Japanese	8.7	14	Aramaic	1
Spirited Away	Japanese	8.6	21	Bosnian	1
The Lives of Others	German	8.5	29	Cantonese	8
Children of Heaven	Persian	8.5	29	Czech	1
Amélie	French	8.4	47	Danish	3
Baahubali: The Beginning	Telugu	8.4	47	Dari	2
Princess Mononoke	Japanese	8.4	47	Dutch	3
Das Boot	German	8.4	47	Dzongkha	1
Oldboy	Korean	8.4	47	Filipino	1
				French	37
				German	13
				Hebrew	3
				Hindi	10
				Hungarian	1
				Icelandic	1
				Indonesian	2
				Italian	7
				Japanese	12
				Kazakh	1
				Korean	5
				Mandarin	14
				Maya	1
				Mongolian	1
				Norwegian	4
				Persian	3
				Portuguese	5
				Romanian	1
				Russian	1
				Spanish	26
				Swedish	1
				Telugu	1
				Thai	3
				Vietnamese	1
				Zulu	1

Top_250_Non_English Movie



Task – 4 : Best top 10 Director



Top_10_director	Mean_imdb_score
Charles Chaplin	8.6
Tony Kaye	8.6
Alfred Hitchcock	8.5
Damien Chazelle	8.5
Majid Majidi	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Asghar Farhadi	8.4
Marius A. Markevicius	8.4
Richard Marquand	8.4
S.S. Rajamouli	8.4

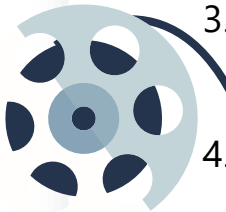
Task – 5 : Popular Genres



Steps followed:

By considering number of voted users, and imdb_score of movie ,the genres are written with | , in which first one is mostly considered the genres of the movie.

1. Created column to separate the genres
2. Made pivot table
3. Counted highest imdb score, considered highest number of voted users
4. Popular genres are noted



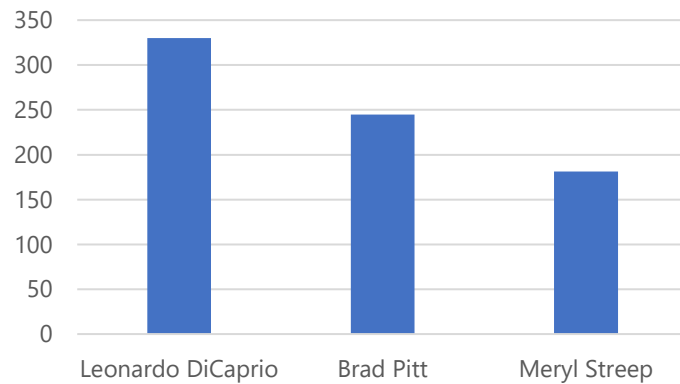
Genres	Sum of num voted users
Action	138489713
Comedy	65900669
Drama	59364726
Adventure	49565555
Crime	34578971
Biography	20470453
Horror	11828176
Mystery	4339683
Animation	3877725
Fantasy	3230512
Sci-Fi	1183437
Western	543103
Documentary	526542
Family	305644
Romance	201486
Musical	174573
Thriller	2354
Grand Total	394583322

genres	num_voted_users	imdb_score
Crime	1689764	9.3
Action	1676169	9
Action	1468200	8.8
Drama	1347461	8.8
Crime	1324680	8.9
Comedy	1251222	8.8
Action	1238746	8.8
Action	1217752	8.7
Action	1215718	8.9
Crime	1155770	9.2
Action	1144337	8.5
Action	1100446	8.7
Crime	1023511	8.6
Action	995415	8.1
Action	982637	8.5
Action	980946	8.3
Drama	955174	8.5
Adventure	928227	8.6
Action	911097	8.7
Crime	887467	8.6
Action	886204	7.9
Adventure	885175	8.3
Action	881236	8.6
Crime	873649	8.5
Biography	865020	8.9
Mystery	845580	8.5
Drama	844052	8.5

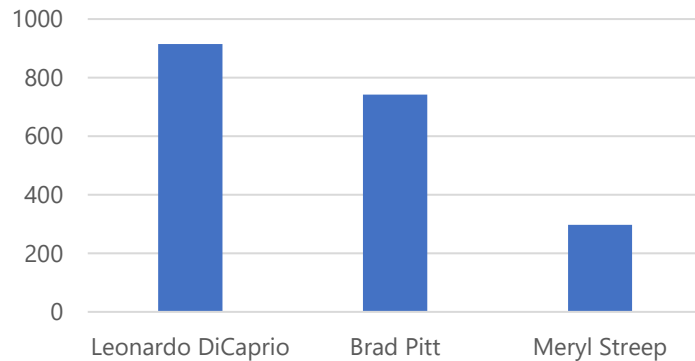
Task – 6 : Critic_Fav and Audience_Fav Actors



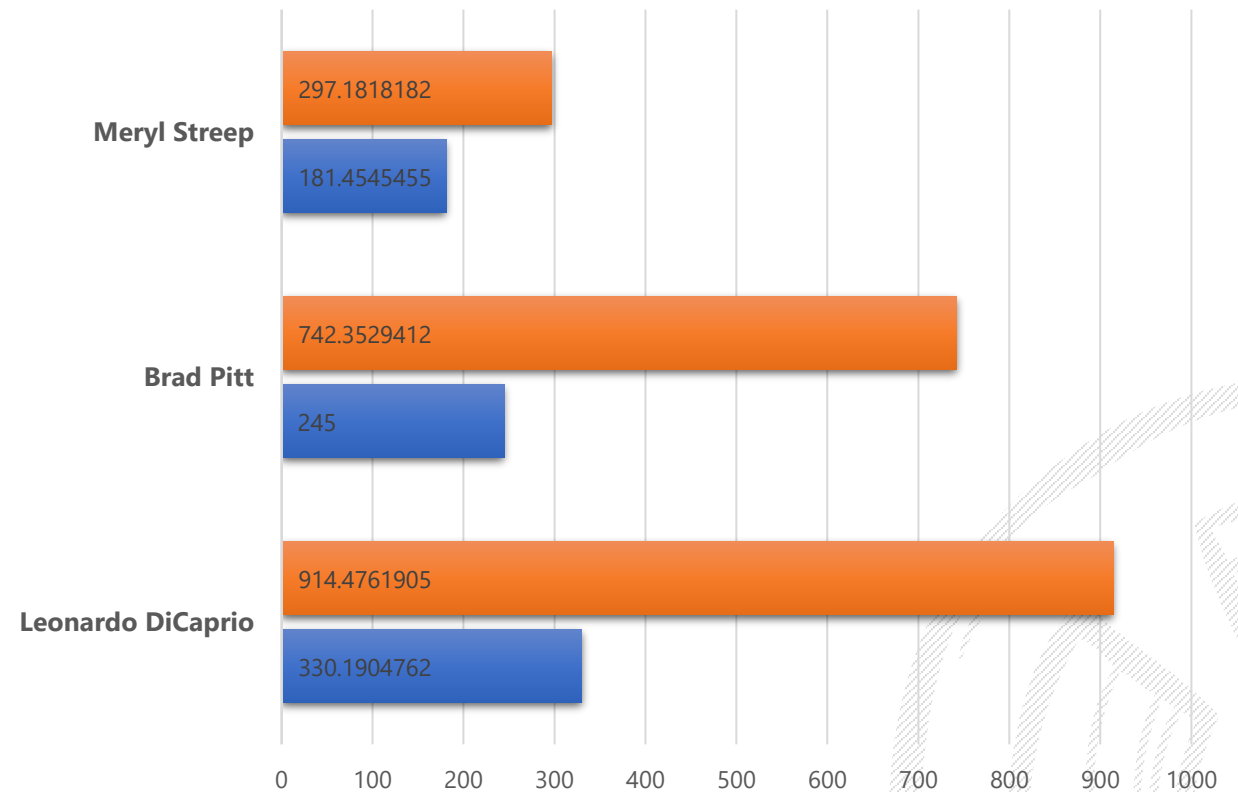
Average of
num_critic_for_reviews



Average of
num_user_for_reviews



Critic_fav & audience_fav actor



ICONIC ACTOR

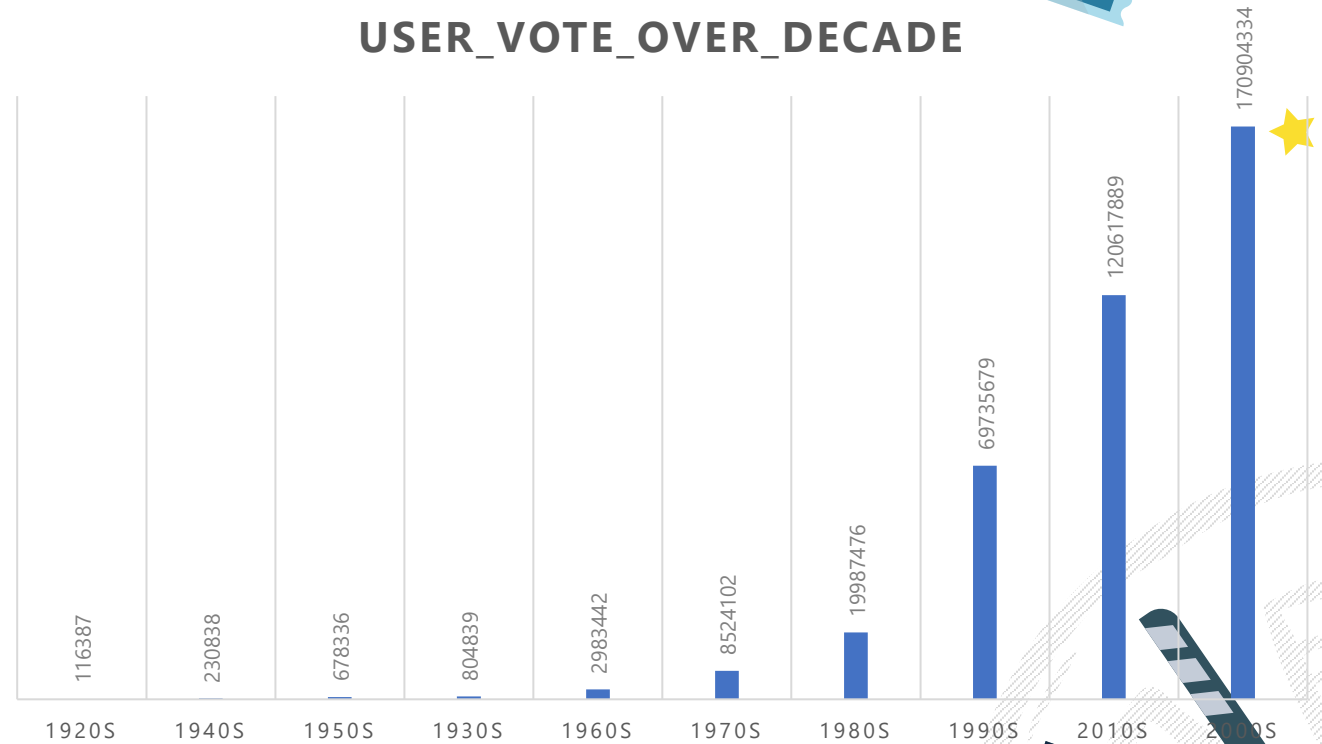
Leonardo DiCaprio

- *Average number of critic for reviews: 330*
- *Average number of users for reviews: 915*
- *Inception, Django Unchained, The departed are most popular with 8.8 & 8.5 imdb_scored movie of this actor.*

Growth of User Votes Over DECADE ★

num_voted_users	title_year	Decade
111841	1927	1920s
4546	1929	1920s
7921	1933	1930s
13269	1935	1930s
143086	1936	1930s
133348	1937	1930s
215340	1939	1930s
291875	1939	1930s
90360	1940	1940s

USER_VOTE_OVER_DECADE



- **Steps followed:**

1. Created column Decade.
2. Made pivot table.
3. Counted User voted as per decade.
4. Ordered by decade from low to high
5. Graph is plotted



Result

Skills developed while performing project

**Advanced
Excel
Skills**

**Pivot
Charts &
Table**

**Equation
formatting**

Graphs

THANK YOU

