IMDB Movie Analysis

Project Description

The Project involves Analysing a Dataset consisting of various movie Records. However, it has come to notice that certain issues with the data require our attention.

The Primary Objective is to identify and define the problem within the Dataset and proceed with the necessary steps to cleanse the data appropriately. Once the data is cleaned, the focus will shift towards utilizing the data analysis skills to explore the dataset comprehensively, aiming to derive valuable insights.
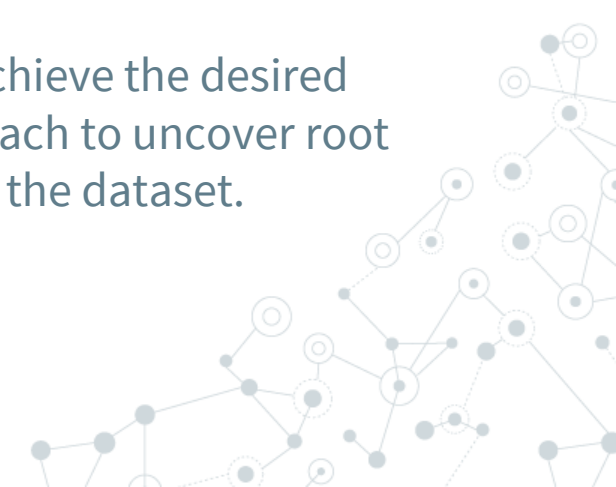
By conducting a thorough analysis of the dataset, I'll be able to obtain the desired output. The output will provide the information needed to effectively analyse the dataset, enabling us to make informed decisions based on the insights gained.

APPROACH

I thoroughly understood the problem statement and gathered the necessary requirements. I devised a strategic plan and executed it using Excel, leveraging its functions and capacities.

Throughout the project, I applied my expertise to achieve the desired results. Additionally, I employed the "5 WHY" Approach to uncover root causes and gain a comprehensive understanding of the dataset.
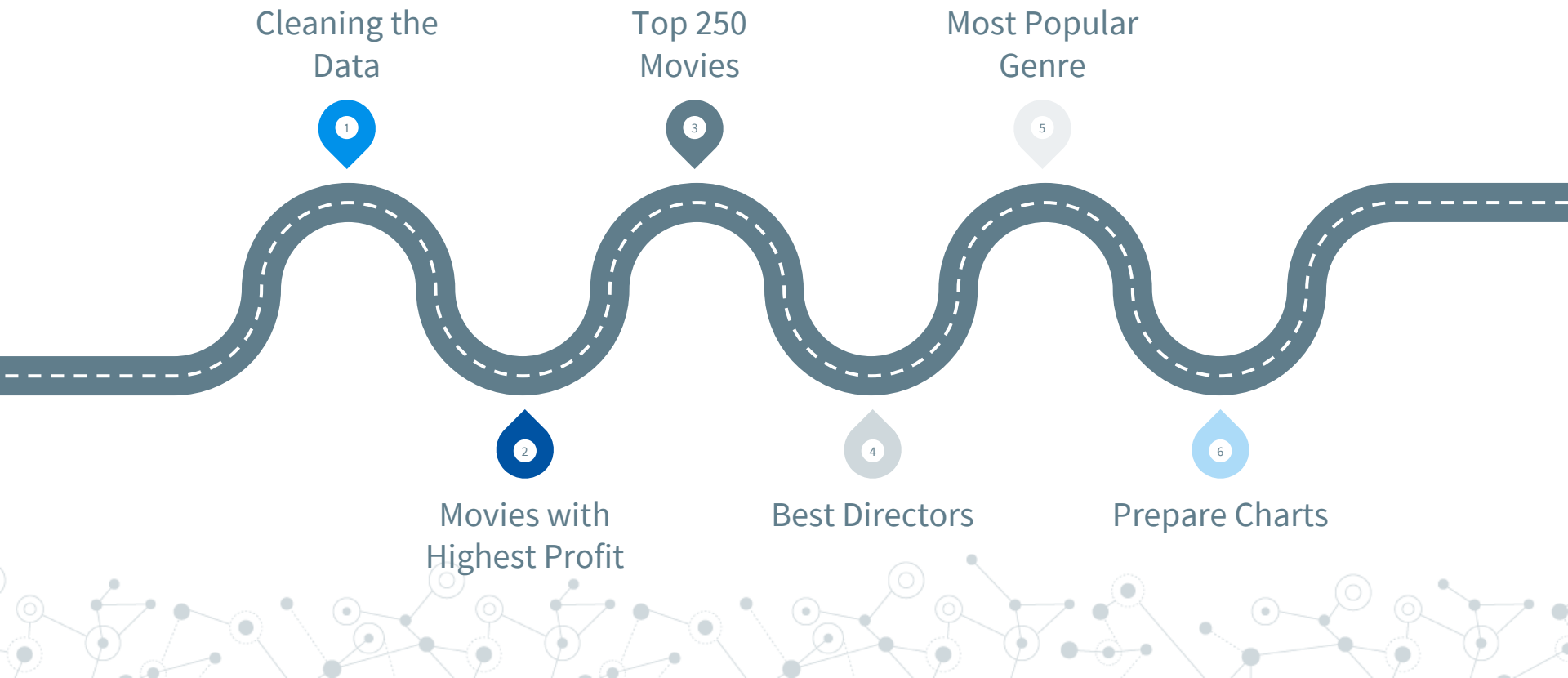
TECH – STACK USED

Microsoft 365
(EXCEL)

Roadmap to Perform Tasks

Cleaning the Data

1

Top 250 Movies

3

Most Popular Genre

5

Movies with Highest Profit

2

Best Directors

4

Prepare Charts

6

Cleaning of Data

In order to accomplish this task, I began by selecting the table and accessing the data function.

Within this function, I utilized the remove duplicate feature to eliminate any duplicate values present in the table.

As a result of this process, I successfully identified and removed a total of 124 duplicate values from the dataset.

Movies with Highest Profit

To Begin with, I initiated the data cleaning process by eliminating rows that contained null values in the "Gross" & "Budget" columns. This step was crucial in ensuring accurate calculations of profit.

Next, I introduced a new column "Profit" which captured the discrepancy between the "Gross" &
"Budget" columns. This calculation allowed me to determine the profitability of each movie accurately.

By utilizing this approach, I successfully compiled a list of movies with the highest profits, providing valuable insights into the dataset.

gross	budget	Profit
162	9000000	-386
703	1500000	651
721	1000000	651
1111	200000	1042
1332	500000	1134
1521	1500000	1373
2436	2000000	2276
2468	300000	2424
2580	1650000	2355
2964	16000000	2486
3330	2100000	3211
3478	1500000	3440
3607	10000000	3431
3830	1200000	3719
4063	4000000	4003
4091	34000000	2953
4584	9000	4406
4756	25000000	4533
4914	150000	4893
4958	2000000	4829
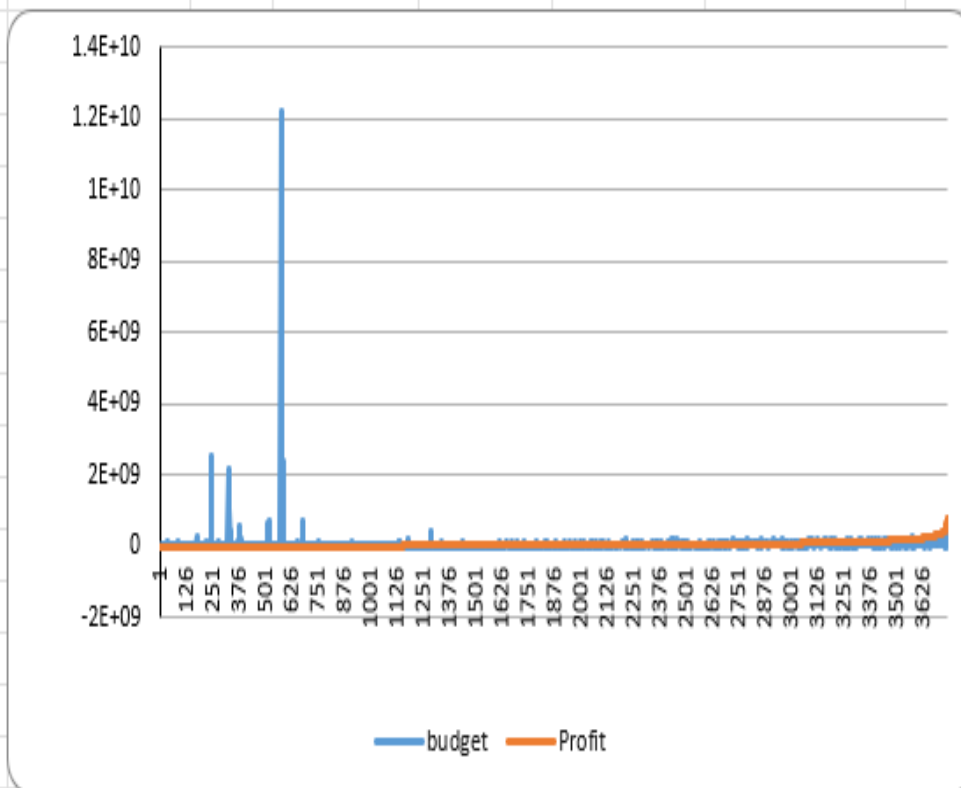
Highest Movie Profit Details

director_name	genres	movie_title	language	country	title_year	imdb_score	gross	budget
James Cameron	Action	AvatarÂ	English	USA	2009	7.9	760505847	237000000

TOP 250 Movies

director_	num_criti	duration	actor_1_n	actor_2_n	actor_3_n	genres	movie_tit	num_vot	num_user_for_review	language	country	title_year	imdb_sco	gross	budget	Profit
Frank Dar	199	142	Morgan F	Jeffrey D	Bob Gunt	Crime|Dr	The Shaw	1689764	4144	English	USA	1994	9.3	54235441	22000000	54231297
Francis Fc	208	175	Al Pacino	Marlon Br	Robert Dt	Crime|Dr	The Godfa	1155770	2238	English	USA	1972	9.2	22202612	2800000	22200374
Francis Fc	149	220	Robert De	Al Pacino	Robert Dt	Crime|Dr	The Godfa	790926	650	English	USA	1974	9	10824921	10000000	10824271
Christoph	645	152	Christian	Heath Lec	Morgan F	Action|Cr	The Dark	1676169	4667	English	USA	2008	9	198332128	175000000	198327461
Quentin 1	215	178	Bruce Wil	Eric Stoltz	Phil LaMa	Crime|Dr	Pulp Ficti	1324680	2195	English	USA	1994	8.9	265107	1500000	262912
Steven Sp	174	185	Liam Nee	Embeth D	Caroline (	Biography	Schindler	865020	1273	English	USA	1993	8.9	28772222	105000000	28770949
Peter Jacl	328	192	Orlando E	Billy Boyc	Bernard H	Action|A	The Lord (	1215718	3189	English	USA	2003	8.9	261437578	92000000	261434389
Irvin Kers	223	127	Harrison I	Kenny Ba	Anthony I	Action|A	Star Wars	837759	900	English	USA	1980	8.8	35033759	19000000	35032859
Robert Ze	149	142	Tom Hank	Siobhan F	Sam Ande	Comedy|	Forrest G	1251222	1398	English	USA	1994	8.8	35707327	55000000	35705929
Peter Jacl	297	171	Christoph	Orlando E	Billy Boyc	Action|A	The Lord (	1238746	5060	English	New Zealand	2001	8.8	35799026	126000000	35793966
Christoph	642	148	Leonardo	Tom Hard	Joseph G	Action|A	Inception	1468200	2803	English	USA	2010	8.8	127490802	150000000	127487999
David Fin	315	151	Brad Pitt	Meat Loaf	Eugenie E	Drama	Fight Clul	1347461	2968	English	USA	1999	8.8	135381507	150000000	135378539
Milos For	149	133	Scatman (	Michael E	Louise Fle	Drama	One Flew	680041	760	English	USA	1975	8.7	424760	7000	424000
George Lu	282	125	Harrison I	Peter Cus	Kenny Ba	Action|A	Star Wars	911097	1470	English	USA	1977	8.7	4131640	7500000	4130170
Peter Jacl	294	172	Christoph	Orlando E	Billy Boyc	Action|A	The Lord (	1100446	2417	English	USA	2002	8.7	11000000	98000000	10997583
Lana Wac	313	136	Keanu Re	Marcus Cl	Gloria Fos	Action|Sc	The Matri	1217752	3646	English	USA	1999	8.7	43532294	65000000	43528648
Martin Sc	192	146	Robert De	Mike Star	Paul Sorv	Biography	Goodfella	728685	989	English	USA	1990	8.7	218051260	207000000	218050271
Bryan Sin	162	106	Kevin Spa	Chazz Pal	Kevin Pol	Crime|Dr	The Usual	740918	1182	English	USA	1995	8.6	194568	1500000	193386
Jonathan	185	138	Anthony I	Scott Gler	Anthony I	Crime|Dr	The Silen	887467	916	English	USA	1991	8.6	531009	10000000	530093
Sergio Lei	164	145	Claudia C	Woody St	Jack Elam	Western	Once Upc	216005	565	English	Italy	1968	8.6	2712293	450000	2711728

Best Directors

director_name	movie_title	imdb_score
Frank Darabont	The Shawshank RedemptionÂ	9.3
Francis Ford Coppola	The GodfatherÂ	9.2
Francis Ford Coppola	The Godfather: Part IIÂ	9
Christopher Nolan	The Dark KnightÂ	9
Quentin Tarantino	Pulp FictionÂ	8.9
Steven Spielberg	Schindler's ListÂ	8.9
Peter Jackson	The Lord of the Rings: The Return of the	8.9
Irvin Kershner	Star Wars: Episode V - The Empire Strik	8.8
Robert Zemeckis	Forrest GumpÂ	8.8

Most Popular Genre

genres
Action
Action
Action
Action
Action
Action
Action
Action
Action
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure
Action|Adventure|Animation|Comedy|Crime|Family|Fantasy
Action|Adventure|Animation|Comedy|Drama|Family|Sci-Fi

GENRE	COUNT
Crime	0
Drama	118
Action	9
Thriller	0
Biography	1
Adventure	2
Fantasy	1
Sci-Fi	1
Comedy	143
Mystery	0
Horror	33
Western	4
War	0
Animation	1
Family	0
Music	143
Romance	0
Sport	0

Comedy & Music are the most popular genre

Prepare Charts

Combined		
movie_title	num_critic_for_reviews	num_user_for_reviews
The HoursÂ	174	660
JuliaÂ	38	44
Out of AfricaÂ	66	200
Florence Foste	87	32
One True Thin	64	112
Julie & JuliaÂ	252	277
The Devil Wea	208	631
A Prairie Hom	211	280
It's Complicate	187	214
The Iron LadyÂ	331	350
Hope SpringsÂ	234	178
The River Wild	42	69
Lions for Lamb	227	298
InceptionÂ	642	2803
Django Uncha	765	1193
The Departed,	352	2054
The Wolf of W	606	1138
Shutter Island	490	964

Actor_1_Name	Mean of Num_critic_for_reviews	Mean of Num_user_for_reviews
Meryl Streep	163.1538462	257.3076923
Leo Caprio	630.3636364	914.4761905
Brad Pitt	462.7777778	1402.222222
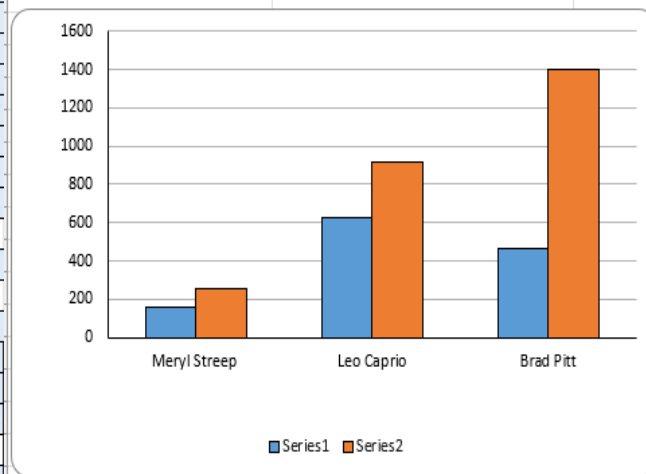
Here, We can clearly see that LEO CAPRIO is Critic Favourite And, BRAD PITT is Audience Favourite

RESULTS

"

The Project successfully analyzed the IMDB Data and provided valuable insights into the Media Industry.

The analysis helped in identifying popular genres, understanding the relationships between ratings and reviews and exploring patterns in movie releases over time.

By leveraging statistical analysis and data visualization, the project enhanced the understanding of IMDB Data and provided meaningful information for further exploration and decision-making in the entertainment industry.

Thank You!

Tushar Agarwal

You can find me at:

LinkedIn

Tushar.agarwal4201@gmail.com