**IMDB Movie Analysis**

**Problem Statement:** The factors that make a movie more successful than others. What kind of movies are more successful and get higher IMDB score?

**Project Description:** In this project, first we will understand and clean the data as necessary. Then, perform the analysis and derive useful insights.

**Approach:**

* Download the dataset
* Understanding the data
* Find the null and duplicate values
* Data processing as per the requirement
* Solve the problems
* Create charts for representation

**Analysis:**

1. **Cleaning the data:**

* The raw dataset has 5044 columns and 28 rows (including title)
* To get the count of null values in each column and row, COUNTBLANK() function is used
* The columns are dropped which are not required for analysis like color, duration, director\_facebook\_likes, actor\_3\_facebook\_likes, actor\_1\_facebook\_likes, facenumber\_in\_poster, country, content\_rating, actor\_2\_facebook\_likes, aspect\_ratio, movie\_imdb\_link, plot\_keywords, actor\_3\_name, cast\_total\_facebook\_likes, actor\_2\_name
* After cleaning, we are left with 3783 R and 12 C

1. **Movies with highest profit:**

* A new column ‘profit’ is created that have the difference between gross and budget
* The data is sorted in descending order using profit column
* Top 10 movies with highest profit are extracted

|  |  |
| --- | --- |
| **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 |

1. **Top 250:**

* The data is sorted in descending order using imdb\_score
* Then, it is filtered by num\_voted\_users to more than 25,000
* Rank column is added indicating the ranks of the corresponding films
* Extracted top 250 movies sorted in descending order and stored as IMDB\_Top\_250
* All non-English movies were extracted from IMDB\_Top\_250 and stored as Top\_Foreign\_Lang\_Film

|  |  |  |
| --- | --- | --- |
| **movie\_title** | **imdb\_score** | **num\_voted\_users** |
| The Shawshank RedemptionÂ | 9.3 | 1689764 |
| The GodfatherÂ | 9.2 | 1155770 |
| The Dark KnightÂ | 9 | 1676169 |
| The Godfather: Part IIÂ | 9 | 790926 |
| The Lord of the Rings: The Return of the KingÂ | 8.9 | 1215718 |
| Pulp FictionÂ | 8.9 | 1324680 |
| Schindler's ListÂ | 8.9 | 865020 |
| The Good, the Bad and the UglyÂ | 8.9 | 503509 |
| Forrest GumpÂ | 8.8 | 1251222 |
| Star Wars: Episode V - The Empire Strikes BackÂ | 8.8 | 837759 |
| The Lord of the Rings: The Fellowship of the RingÂ | 8.8 | 1238746 |
| InceptionÂ | 8.8 | 1468200 |
| 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 |
| GoodfellasÂ | 8.7 | 728685 |
| City of GodÂ | 8.7 | 533200 |
| 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 |
| 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 |
| 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 |
| WALLÂ·EÂ | 8.4 | 718837 |
| A SeparationÂ | 8.4 | 151812 |
| BraveheartÂ | 8.4 | 736638 |
| Reservoir DogsÂ | 8.4 | 664719 |
| OldboyÂ | 8.4 | 356181 |
| Requiem for a DreamÂ | 8.4 | 573541 |
| Das BootÂ | 8.4 | 168203 |
| Lawrence of ArabiaÂ | 8.4 | 192775 |
| Baahubali: The BeginningÂ | 8.4 | 62756 |
| Once Upon a Time in AmericaÂ | 8.4 | 221000 |
| AmÃ©lieÂ | 8.4 | 534262 |
| Princess MononokeÂ | 8.4 | 221552 |
| Toy Story 3Â | 8.3 | 544884 |
| Inside OutÂ | 8.3 | 345198 |
| Toy StoryÂ | 8.3 | 623757 |
| The StingÂ | 8.3 | 175607 |
| Indiana Jones and the Last CrusadeÂ | 8.3 | 515306 |
| Good Will HuntingÂ | 8.3 | 604904 |
| UpÂ | 8.3 | 665575 |
| UnforgivenÂ | 8.3 | 277505 |
| Batman BeginsÂ | 8.3 | 980946 |
| Inglourious BasterdsÂ | 8.3 | 885175 |
| 2001: A Space OdysseyÂ | 8.3 | 427357 |
| AmadeusÂ | 8.3 | 270790 |
| L.A. ConfidentialÂ | 8.3 | 414219 |
| SnatchÂ | 8.3 | 600996 |
| Some Like It HotÂ | 8.3 | 175196 |
| ScarfaceÂ | 8.3 | 537442 |
| Eternal Sunshine of the Spotless MindÂ | 8.3 | 666937 |
| RoomÂ | 8.3 | 161288 |
| Monty Python and the Holy GrailÂ | 8.3 | 382240 |
| The HuntÂ | 8.3 | 170155 |
| MetropolisÂ | 8.3 | 111841 |
| DownfallÂ | 8.3 | 248354 |
| Raging BullÂ | 8.3 | 235133 |
| Finding NemoÂ | 8.2 | 692482 |
| Gone with the WindÂ | 8.2 | 215340 |
| Captain America: Civil WarÂ | 8.2 | 272670 |
| Gran TorinoÂ | 8.2 | 561773 |
| A Beautiful MindÂ | 8.2 | 610568 |
| Die HardÂ | 8.2 | 592582 |
| How to Train Your DragonÂ | 8.2 | 485430 |
| The Bridge on the River KwaiÂ | 8.2 | 149444 |
| Pan's LabyrinthÂ | 8.2 | 467234 |
| The Secret in Their EyesÂ | 8.2 | 131831 |
| The Wolf of Wall StreetÂ | 8.2 | 780588 |
| V for VendettaÂ | 8.2 | 791783 |
| TrainspottingÂ | 8.2 | 469561 |
| On the WaterfrontÂ | 8.2 | 100890 |
| Into the WildÂ | 8.2 | 426359 |
| Lock, Stock and Two Smoking BarrelsÂ | 8.2 | 414976 |
| The Big LebowskiÂ | 8.2 | 537419 |
| IncendiesÂ | 8.2 | 80429 |
| Blade RunnerÂ | 8.2 | 461609 |
| The ThingÂ | 8.2 | 258078 |
| CasinoÂ | 8.2 | 333542 |
| WarriorÂ | 8.2 | 332276 |
| Howl's Moving CastleÂ | 8.2 | 214091 |
| The AvengersÂ | 8.1 | 995415 |
| DeadpoolÂ | 8.1 | 479047 |
| Jurassic ParkÂ | 8.1 | 613473 |
| The Sixth SenseÂ | 8.1 | 704766 |
| Monsters, Inc.Â | 8.1 | 585659 |
| Pirates of the Caribbean: The Curse of the Black PearlÂ | 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 |
| Gone GirlÂ | 8.1 | 569841 |
| Butch Cassidy and the Sundance KidÂ | 8.1 | 152089 |
| The Imitation GameÂ | 8.1 | 467613 |
| Million Dollar BabyÂ | 8.1 | 482064 |
| The Truman ShowÂ | 8.1 | 667983 |
| Groundhog DayÂ | 8.1 | 437418 |
| No Country for Old MenÂ | 8.1 | 612060 |
| The RevenantÂ | 8.1 | 406020 |
| Shutter IslandÂ | 8.1 | 786092 |
| Stand by MeÂ | 8.1 | 271794 |
| Kill Bill: Vol. 1Â | 8.1 | 735784 |
| 12 Years a SlaveÂ | 8.1 | 439176 |
| Annie HallÂ | 8.1 | 192940 |
| Sin CityÂ | 8.1 | 656640 |
| The Grand Budapest HotelÂ | 8.1 | 475518 |
| The TerminatorÂ | 8.1 | 600266 |
| SpotlightÂ | 8.1 | 195333 |
| The Best Years of Our LivesÂ | 8.1 | 40359 |
| The Wizard of OzÂ | 8.1 | 291875 |
| There Will Be BloodÂ | 8.1 | 372990 |
| PrisonersÂ | 8.1 | 383591 |
| The Princess BrideÂ | 8.1 | 294163 |
| Hotel RwandaÂ | 8.1 | 264533 |
| Mad Max: Fury RoadÂ | 8.1 | 552503 |
| Amores PerrosÂ | 8.1 | 173551 |
| Before SunriseÂ | 8.1 | 183288 |
| The CelebrationÂ | 8.1 | 65951 |
| Donnie DarkoÂ | 8.1 | 580999 |
| Elite SquadÂ | 8.1 | 81644 |
| The Sea InsideÂ | 8.1 | 64556 |
| RushÂ | 8.1 | 312629 |
| Tae Guk Gi: The Brotherhood of WarÂ | 8.1 | 31943 |
| AkiraÂ | 8.1 | 106160 |
| JawsÂ | 8 | 412454 |
| The ExorcistÂ | 8 | 284252 |
| AladdinÂ | 8 | 260939 |
| The IncrediblesÂ | 8 | 479166 |
| Dances with WolvesÂ | 8 | 186485 |
| The Sound of MusicÂ | 8 | 148172 |
| Rain ManÂ | 8 | 383784 |
| Slumdog MillionaireÂ | 8 | 641997 |
| The King's SpeechÂ | 8 | 503631 |
| Catch Me If You CanÂ | 8 | 525801 |
| Star TrekÂ | 8 | 504419 |
| The Pursuit of HappynessÂ | 8 | 338383 |
| Doctor ZhivagoÂ | 8 | 55816 |
| Black SwanÂ | 8 | 551363 |
| District 9Â | 8 | 531737 |
| Young FrankensteinÂ | 8 | 112671 |
| Dead Poets SocietyÂ | 8 | 277451 |
| Mystic RiverÂ | 8 | 338415 |
| RatatouilleÂ | 8 | 473887 |
| Fiddler on the RoofÂ | 8 | 29839 |
| Kill Bill: Vol. 2Â | 8 | 512749 |
| X-Men: Days of Future PastÂ | 8 | 514125 |
| JFKÂ | 8 | 113472 |
| 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 |
| 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 |
| Central StationÂ | 8 | 28951 |
| HerÂ | 8 | 355126 |
| Waltz with BashirÂ | 8 | 46107 |
| True RomanceÂ | 8 | 163492 |
| PersepolisÂ | 8 | 70194 |
| Big FishÂ | 8 | 350698 |
| The Straight StoryÂ | 8 | 63733 |
| BrazilÂ | 8 | 152306 |
| In BrugesÂ | 8 | 307639 |
| Mulholland DriveÂ | 8 | 235992 |
| My Name Is KhanÂ | 8 | 69759 |
| Dancer in the DarkÂ | 8 | 79330 |
| MagnoliaÂ | 8 | 241030 |
| SerenityÂ | 8 | 242599 |
| Cinderella ManÂ | 8 | 148238 |
| Blood DiamondÂ | 8 | 400292 |
| The Iron GiantÂ | 8 | 128455 |
| AvatarÂ | 7.9 | 886204 |
| E.T. the Extra-TerrestrialÂ | 7.9 | 281842 |
| ShrekÂ | 7.9 | 467113 |
| Iron ManÂ | 7.9 | 696338 |
| Toy Story 2Â | 7.9 | 385871 |
| Straight Outta ComptonÂ | 7.9 | 119928 |
| The Hobbit: An Unexpected JourneyÂ | 7.9 | 637246 |
| TakenÂ | 7.9 | 483756 |
| Crouching Tiger, Hidden DragonÂ | 7.9 | 217740 |
| Walk the LineÂ | 7.9 | 188637 |
| The FighterÂ | 7.9 | 275869 |
| The Bourne IdentityÂ | 7.9 | 407601 |
| Big Hero 6Â | 7.9 | 279093 |
| My Fair LadyÂ | 7.9 | 66959 |
| Captain PhillipsÂ | 7.9 | 323353 |
| Little Miss SunshineÂ | 7.9 | 355810 |
| The UntouchablesÂ | 7.9 | 219008 |
| CrashÂ | 7.9 | 361169 |
| HalloweenÂ | 7.9 | 157857 |
| Edward ScissorhandsÂ | 7.9 | 357581 |
| The Hobbit: The Desolation of SmaugÂ | 7.9 | 483540 |
| How to Train Your Dragon 2Â | 7.9 | 221128 |
| The Blues BrothersÂ | 7.9 | 142448 |
| NightcrawlerÂ | 7.9 | 293304 |
| Do the Right ThingÂ | 7.9 | 59524 |
| The WrestlerÂ | 7.9 | 251349 |
| Hot FuzzÂ | 7.9 | 352695 |
| The Remains of the DayÂ | 7.9 | 45703 |
| Boogie NightsÂ | 7.9 | 189032 |
| The Hateful EightÂ | 7.9 | 272839 |
| OnceÂ | 7.9 | 90827 |
| GloryÂ | 7.9 | 101888 |
| Before MidnightÂ | 7.9 | 95362 |
| 4 Months, 3 Weeks and 2 DaysÂ | 7.9 | 44763 |
| MoonÂ | 7.9 | 260607 |
| Nine QueensÂ | 7.9 | 38215 |
| The ChorusÂ | 7.9 | 44151 |
| Veer-ZaaraÂ | 7.9 | 34449 |
| Letters from Iwo JimaÂ | 7.9 | 132149 |
| The Right StuffÂ | 7.9 | 45271 |
| AmourÂ | 7.9 | 70382 |
| Ed WoodÂ | 7.9 | 142416 |
| The World's Fastest IndianÂ | 7.9 | 44198 |
| Almost FamousÂ | 7.9 | 207287 |
| The NotebookÂ | 7.9 | 396396 |
| HeroÂ | 7.9 | 149414 |

1. **Best Directors:**

* The data is grouped by director\_name and mean of imdb\_score
* First, director\_name is sorted in ascending order then, the score is sorted in descending order
* Finally, top 10 directors with mean imdb score are extracted

|  |  |
| --- | --- |
| **Director name** | **Average of imdb\_score** |
| Charles Chaplin | 8.6 |
| Tony Kaye | 8.6 |
| Alfred Hitchcock | 8.5 |
| Damien Chazelle | 8.5 |
| Majid Majidi | 8.5 |
| Ron Fricke | 8.5 |
| Sergio Leone | 8.433333333 |
| Christopher Nolan | 8.425 |
| Asghar Farhadi | 8.4 |
| Marius A. Markevicius | 8.4 |

1. **Popular Genres:**

* Group the data by genres with mean gross value
* Then, sort the gross value in descending order
* Most popular genres are Family and Sci-Fi

|  |  |
| --- | --- |
| **Genres** | **Average of gross** |
| Family|Sci-Fi | 434949459 |
| Adventure|Animation|Drama|Family|Musical | 422783777 |
| Adventure|Animation|Comedy|Drama|Family|Fantasy | 356454367 |
| Action|Biography|Drama|History|Thriller|War | 350123553 |
| Action|Adventure|Fantasy|Sci-Fi | 296684758.3 |
| Adventure|Drama|Fantasy|Romance | 296481890 |
| Action|Adventure|Fantasy|Romance | 289279970.3 |
| Adventure|Sci-Fi | 281666058 |
| Adventure|Family|Fantasy|Mystery | 279056317 |
| Action|Adventure|Animation|Family | 261437578 |

1. **Critic-favourite and Audience-favourite actors:**

* Create a new sheet and name it as Combined which have all the details of actors - 'Meryl Streep', 'Leonardo DiCaprio', and 'Brad Pitt
* Group the data using actor name column
* Find the mean of num\_critic\_for\_reviews and num\_users\_for\_review for each actor
* Leonardo DiCaprio has the highest mean

|  |  |  |
| --- | --- | --- |
| **Actor name** | **Average of num\_critic\_for\_reviews** | **Average of num\_user\_for\_reviews** |
| Brad Pitt | 245.00 | 742.35 |
| Leonardo DiCaprio | 322.20 | 922.55 |
| Meryl Streep | 181.45 | 297.18 |

**Sum of users voted in each decade:**

|  |  |
| --- | --- |
| **Decade** | **number of voted users** |
| 1920 | 116387 |
| 1930 | 804839 |
| 1940 | 230838 |
| 1950 | 678336 |
| 1960 | 2983442 |
| 1970 | 8269828 |
| 1980 | 19344369 |
| 1990 | 69635863 |
| 2000 | 166005315 |
| 2010 | 116239604 |

**Tech-Stack Used:** Dataset provided and MS Excel

**Insights:**

* The movie with highest profit of 523505847 is AvatarÂ
* The movie with highest number of voted users is ‘The Shawshank RedemptionÂ’ with imdb score of 9.3
* The best directors are Charles Chaplin and Tony Kaye with average imdb score of 8.6
* The most popular genre are Family and Sci-Fi with average gross amount of 434949459
* The Critic and Audience favourite actor is Leonardo DiCaprio with maximum number of critic reviews and user reviews
* The maximum number of users voted in 2000 (decade), which is 166005315

**Results:** This project has helped in understanding the problem statement, perform analysis to gather insights from the data and getting hands-on experience on real life data. EDA helped me to make data error free and removing null and duplicate values to make data a good fit for statistical and analytical treatment.