

Investigate_a_Dataset

June 24, 2022

1 Project: Investigate a Dataset - [TMDb movies data]

1.1 Table of Contents

Introduction

Data Wrangling

Exploratory Data Analysis

Conclusions

Introduction

1.1.1 Dataset Description

The TMDb movies data is gotten from Kaggle, it has 10866 rows and 21 columns, which will be gotten through a code that will be run in the cell below.

The following are the columns stated in the set ('id', 'imdb_id', 'popularity', 'budget', 'revenue', 'original_title', 'cast', 'homepage', 'tagline', 'overview', 'runtime', 'genres', 'production_companies', 'release_date', 'director', 'vote_count', 'vote_average', 'release_year', 'budget_adj', 'revenue_adj'.

Characteristics of some of the columns

1. The id signify identification of the movies.
2. 'genres', 'cast' and production_companies had a similar feature because it has more than one in each column, so it is seperated with the pipe character '|' for a clear understanding of the column.
3. The release year is gotten from release_date.

1.1.2 Question(s) for Analysis

1. which genres are most popular from year to year
2. what kinds of properties are associated with movies that have high revenue
3. what are the five(5) top most profitable movies(revenue vs budget)
4. write a function for runtime to determine the length of the movie either long or short

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plot
%matplotlib inline
```

Data Wrangling

Read the TMDb csv file and recognize every aspect in the dataset

```
In [2]: # load data from csv files
```

```
df = pd.read_csv('Database_TMDb_movie_data/tmdb-movies.csv')
df.head()
```

```
Out[2]:
```

	id	imdb_id	popularity	budget	revenue	\
0	135397	tt0369610	32.985763	150000000	1513528810	
1	76341	tt1392190	28.419936	150000000	378436354	
2	262500	tt2908446	13.112507	110000000	295238201	
3	140607	tt2488496	11.173104	200000000	2068178225	
4	168259	tt2820852	9.335014	190000000	1506249360	

	original_title	\
0	Jurassic World	
1	Mad Max: Fury Road	
2	Insurgent	
3	Star Wars: The Force Awakens	
4	Furious 7	

	cast	\
0	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	
1	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	
2	Shailene Woodley Theo James Kate Winslet Ansel...	
3	Harrison Ford Mark Hamill Carrie Fisher Adam D...	
4	Vin Diesel Paul Walker Jason Statham Michelle ...	

	homepage	director	\
0	http://www.jurassicworld.com/	Colin Trevorrow	
1	http://www.madmaxmovie.com/	George Miller	
2	http://www.thedivergentseries.movie/#insurgent	Robert Schwentke	
3	http://www.starwars.com/films/star-wars-episod...	J.J. Abrams	
4	http://www.furious7.com/	James Wan	

	tagline	...	\
0	The park is open.	...	
1	What a Lovely Day.	...	
2	One Choice Can Destroy You	...	
3	Every generation has a story.	...	
4	Vengeance Hits Home	...	

	overview	runtime	\
0	Twenty-two years after the events of Jurassic ...	124	
1	An apocalyptic story set in the furthest reach...	120	
2	Beatrice Prior must confront her inner demons ...	119	
3	Thirty years after defeating the Galactic Empi...	136	
4	Deckard Shaw seeks revenge against Dominic Tor...	137	

```

                                genres \
0  Action|Adventure|Science Fiction|Thriller
1  Action|Adventure|Science Fiction|Thriller
2      Adventure|Science Fiction|Thriller
3  Action|Adventure|Science Fiction|Fantasy
4      Action|Crime|Thriller

```

```

                                production_companies release_date vote_count \
0  Universal Studios|Amblin Entertainment|Legenda...      6/9/15      5562
1  Village Roadshow Pictures|Kennedy Miller Produ...      5/13/15      6185
2  Summit Entertainment|Mandeville Films|Red Wago...      3/18/15      2480
3      Lucasfilm|Truenorth Productions|Bad Robot      12/15/15      5292
4  Universal Pictures|Original Film|Media Rights ...      4/1/15      2947

```

```

    vote_average  release_year  budget_adj  revenue_adj
0           6.5         2015  1.379999e+08  1.392446e+09
1           7.1         2015  1.379999e+08  3.481613e+08
2           6.3         2015  1.012000e+08  2.716190e+08
3           7.5         2015  1.839999e+08  1.902723e+09
4           7.3         2015  1.747999e+08  1.385749e+09

```

[5 rows x 21 columns]

```

In [222]: # description of the dataset
df.describe()

```

```

Out[222]:
    count      id  popularity    budget    revenue    runtime \
count  10866.000000  10866.000000  1.086600e+04  1.086600e+04  10866.000000
mean    66064.177434      0.646441  1.462570e+07  3.982332e+07   102.070863
std     92130.136561      1.000185  3.091321e+07  1.170035e+08   31.381405
min         5.000000      0.000065  0.000000e+00  0.000000e+00    0.000000
25%    10596.250000      0.207583  0.000000e+00  0.000000e+00    90.000000
50%    20669.000000      0.383856  0.000000e+00  0.000000e+00    99.000000
75%    75610.000000      0.713817  1.500000e+07  2.400000e+07   111.000000
max    417859.000000     32.985763  4.250000e+08  2.781506e+09   900.000000

    vote_count  vote_average  release_year  budget_adj  revenue_adj
count  10866.000000  10866.000000  10866.000000  1.086600e+04  1.086600e+04
mean     217.389748      5.974922   2001.322658  1.755104e+07  5.136436e+07
std     575.619058      0.935142    12.812941  3.430616e+07  1.446325e+08
min       10.000000      1.500000   1960.000000  0.000000e+00  0.000000e+00
25%       17.000000      5.400000   1995.000000  0.000000e+00  0.000000e+00
50%       38.000000      6.000000   2006.000000  0.000000e+00  0.000000e+00
75%      145.750000      6.600000   2011.000000  2.085325e+07  3.369710e+07
max      9767.000000      9.200000   2015.000000  4.250000e+08  2.827124e+09

```

```

In [223]: # details of the dataset
df.info

```

```

Out[223]: <bound method DataFrame.info of
0      135397  tt0369610  32.985763  150000000  1513528810
1       76341  tt1392190  28.419936  150000000   378436354
2      262500  tt2908446  13.112507  110000000   295238201
3      140607  tt2488496  11.173104  200000000  2068178225
4      168259  tt2820852   9.335014  190000000  1506249360
5      281957  tt1663202   9.110700  135000000   532950503
6       87101  tt1340138   8.654359  155000000   440603537
7      286217  tt3659388   7.667400  108000000   595380321
8      211672  tt2293640   7.404165   74000000  1156730962
9      150540  tt2096673   6.326804  175000000   853708609
10     206647  tt2379713   6.200282  245000000   880674609
11      76757  tt1617661   6.189369  176000003   183987723
12     264660  tt0470752   6.118847   15000000   36869414
13     257344  tt2120120   5.984995   88000000   243637091
14      99861  tt2395427   5.944927  280000000  1405035767
15     273248  tt3460252   5.898400   44000000   155760117
16     260346  tt2446042   5.749758   48000000   325771424
17     102899  tt0478970   5.573184  130000000   518602163
18     150689  tt1661199   5.556818   95000000   542351353
19     131634  tt1951266   5.476958  160000000   650523427
20     158852  tt1964418   5.462138  190000000   209035668
21     307081  tt1798684   5.337064   30000000   91709827
22     254128  tt2126355   4.907832  110000000   470490832
23     216015  tt2322441   4.710402   40000000   569651467
24     318846  tt1596363   4.648046   28000000   133346506
25     177677  tt2381249   4.566713  150000000   682330139
26     214756  tt2637276   4.564549   68000000   215863606
27     207703  tt2802144   4.503789   81000000   403802136
28     314365  tt1895587   4.062293   20000000   88346473
29     294254  tt4046784   3.968891   61000000   311256926
...      ...      ...      ...      ...      ...
10836   38720  tt0061170   0.239435         0         0
10837   19728  tt0060177   0.291704         0         0
10838   22383  tt0060862   0.151845         0         0
10839   13353  tt0060550   0.276133         0         0
10840   34388  tt0060437   0.102530         0         0
10841   42701  tt0062262   0.264925    75000         0
10842   36540  tt0061199   0.253437         0         0
10843   29710  tt0060588   0.252399         0         0
10844   23728  tt0059557   0.236098         0         0
10845    5065  tt0059014   0.230873         0         0
10846   17102  tt0059127   0.212716         0         0
10847   28763  tt0060548   0.034555         0         0
10848    2161  tt0060397   0.207257   5115000   12000000
10849   28270  tt0060445   0.206537         0         0
10850   26268  tt0060490   0.202473         0         0
10851   15347  tt0060182   0.342791         0         0

```

10852	37301	tt0060165	0.227220	0	0
10853	15598	tt0060086	0.163592	0	0
10854	31602	tt0060232	0.146402	0	0
10855	13343	tt0059221	0.141026	700000	0
10856	20277	tt0061135	0.140934	0	0
10857	5921	tt0060748	0.131378	0	0
10858	31918	tt0060921	0.317824	0	0
10859	20620	tt0060955	0.089072	0	0
10860	5060	tt0060214	0.087034	0	0
10861	21	tt0060371	0.080598	0	0
10862	20379	tt0060472	0.065543	0	0
10863	39768	tt0060161	0.065141	0	0
10864	21449	tt0061177	0.064317	0	0
10865	22293	tt0060666	0.035919	19000	0

	original_title \
0	Jurassic World
1	Mad Max: Fury Road
2	Insurgent
3	Star Wars: The Force Awakens
4	Furious 7
5	The Revenant
6	Terminator Genisys
7	The Martian
8	Minions
9	Inside Out
10	Spectre
11	Jupiter Ascending
12	Ex Machina
13	Pixels
14	Avengers: Age of Ultron
15	The Hateful Eight
16	Taken 3
17	Ant-Man
18	Cinderella
19	The Hunger Games: Mockingjay - Part 2
20	Tomorrowland
21	Southpaw
22	San Andreas
23	Fifty Shades of Grey
24	The Big Short
25	Mission: Impossible - Rogue Nation
26	Ted 2
27	Kingsman: The Secret Service
28	Spotlight
29	Maze Runner: The Scorch Trials
...	...
10836	Walk Don't Run

10837	The Blue Max
10838	The Professionals
10839	It's the Great Pumpkin, Charlie Brown
10840	Funeral in Berlin
10841	The Shooting
10842	Winnie the Pooh and the Honey Tree
10843	Khartoum
10844	Our Man Flint
10845	Carry On Cowboy
10846	Dracula: Prince of Darkness
10847	Island of Terror
10848	Fantastic Voyage
10849	Gambit
10850	Harper
10851	Born Free
10852	A Big Hand for the Little Lady
10853	Alfie
10854	The Chase
10855	The Ghost & Mr. Chicken
10856	The Ugly Dachshund
10857	Nevada Smith
10858	The Russians Are Coming, The Russians Are Coming
10859	Seconds
10860	Carry On Screaming!
10861	The Endless Summer
10862	Grand Prix
10863	Beregis Avtomobilya
10864	What's Up, Tiger Lily?
10865	Manos: The Hands of Fate

	cast \
0	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...
1	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...
2	Shailene Woodley Theo James Kate Winslet Ansel...
3	Harrison Ford Mark Hamill Carrie Fisher Adam D...
4	Vin Diesel Paul Walker Jason Statham Michelle ...
5	Leonardo DiCaprio Tom Hardy Will Poulter Domhn...
6	Arnold Schwarzenegger Jason Clarke Emilia Clar...
7	Matt Damon Jessica Chastain Kristen Wiig Jeff ...
8	Sandra Bullock Jon Hamm Michael Keaton Allison...
9	Amy Poehler Phyllis Smith Richard Kind Bill Ha...
10	Daniel Craig Christoph Waltz LÃla Seydoux Ralp...
11	Mila Kunis Channing Tatum Sean Bean Eddie Redm...
12	Domhnall Gleeson Alicia Vikander Oscar Isaac S...
13	Adam Sandler Michelle Monaghan Peter Dinklage ...
14	Robert Downey Jr. Chris Hemsworth Mark Ruffalo...
15	Samuel L. Jackson Kurt Russell Jennifer Jason ...
16	Liam Neeson Forest Whitaker Maggie Grace Famke...

17 Paul Rudd|Michael Douglas|Evangeline Lilly|Cor...
 18 Lily James|Cate Blanchett|Richard Madden|Helen...
 19 Jennifer Lawrence|Josh Hutcherson|Liam Hemswor...
 20 Britt Robertson|George Clooney|Raffey Cassidy|...
 21 Jake Gyllenhaal|Rachel McAdams|Forest Whitaker...
 22 Dwayne Johnson|Alexandra Daddario|Carla Gugino...
 23 Dakota Johnson|Jamie Dornan|Jennifer Ehle|Eloi...
 24 Christian Bale|Steve Carell|Ryan Gosling|Brad ...
 25 Tom Cruise|Jeremy Renner|Simon Pegg|Rebecca Fe...
 26 Mark Wahlberg|Seth MacFarlane|Amanda Seyfried|...
 27 Taron Egerton|Colin Firth|Samuel L. Jackson|Mi...
 28 Mark Ruffalo|Michael Keaton|Rachel McAdams|Lie...
 29 Dylan O'Brien|Kaya Scodelario|Thomas Brodie-Sa...
 ...
 10836 Cary Grant|Samantha Eggar|Jim Hutton|John Stan...
 10837 George Peppard|James Mason|Ursula Andress|Jere...
 10838 Burt Lancaster|Lee Marvin|Robert Ryan|Woody St...
 10839 Christopher Shea|Sally Dryer|Kathy Steinberg|A...
 10840 Michael Caine|Paul Hubschmid|Oskar Homolka|Eva...
 10841 Will Hutchins|Millie Perkins|Jack Nicholson|Wa...
 10842 Sterling Holloway|Junius Matthews|Sebastian Ca...
 10843 Charlton Heston|Laurence Olivier|Richard Johns...
 10844 James Coburn|Lee J. Cobb|Gila Golan|Edward Mul...
 10845 Sid James|Jim Dale|Angela Douglas|Kenneth Will...
 10846 Christopher Lee|Barbara Shelley|Andrew Keir|Fr...
 10847 Peter Cushing|Edward Judd|Carole Gray|Eddie By...
 10848 Stephen Boyd|Raqueel Welch|Edmond O'Brien|Donal...
 10849 Michael Caine|Shirley MacLaine|Herbert Lom|Joh...
 10850 Paul Newman|Lauren Bacall|Julie Harris|Arthur ...
 10851 Virginia McKenna|Bill Travers|Geoffrey Keen|Pe...
 10852 Henry Fonda|Joanne Woodward|Jason Robards|Paul...
 10853 Michael Caine|Shelley Winters|Millicent Martin...
 10854 Marlon Brando|Jane Fonda|Robert Redford|E.G. M...
 10855 Don Knotts|Joan Staley|Liam Redmond|Dick Sarge...
 10856 Dean Jones|Suzanne Pleshette|Charles Ruggles|K...
 10857 Steve McQueen|Karl Malden|Brian Keith|Arthur K...
 10858 Carl Reiner|Eva Marie Saint|Alan Arkin|Brian K...
 10859 Rock Hudson|Salome Jens|John Randolph|Will Gee...
 10860 Kenneth Williams|Jim Dale|Harry H. Corbett|Joa...
 10861 Michael Hynson|Robert August|Lord 'Tally Ho' B...
 10862 James Garner|Eva Marie Saint|Yves Montand|Tosh...
 10863 Innokentiy Smoktunovskiy|Oleg Efremov|Georgi Z...
 10864 Tatsuya Mihashi|Akiko Wakabayashi|Mie Hama|Joh...
 10865 Harold P. Warren|Tom Neyman|John Reynolds|Dian...

homepage \

0 <http://www.jurassicworld.com/>
 1 <http://www.madmaxmovie.com/>

2	http://www.thedivergentseries.movie/#insurgent
3	http://www.starwars.com/films/star-wars-episod...
4	http://www.furious7.com/
5	http://www.foxmovies.com/movies/the-revenant
6	http://www.terminatormovie.com/
7	http://www.foxmovies.com/movies/the-martian
8	http://www.minionsmovie.com/
9	http://movies.disney.com/inside-out
10	http://www.sonypictures.com/movies/spectre/
11	http://www.jupiterascending.com
12	http://exmachina-movie.com/
13	http://www.pixels-movie.com/
14	http://marvel.com/movies/movie/193/avengers_ag...
15	http://thehatefuleight.com/
16	http://www.taken3movie.com/
17	http://marvel.com/movies/movie/180/ant-man
18	NaN
19	http://www.thehungergames.movie/
20	http://movies.disney.com/tomorrowland
21	NaN
22	http://www.sanandreasmovie.com/
23	https://www.facebook.com/fiftyshadesofgreymovie
24	http://www.thebigshortmovie.com/
25	http://www.missionimpossible.com
26	NaN
27	http://www.kingsmanmovie.com/
28	http://www.spotlightthefilm.com
29	http://mazerunnermovies.com
...	...
10836	NaN
10837	NaN
10838	NaN
10839	NaN
10840	NaN
10841	NaN
10842	NaN
10843	NaN
10844	NaN
10845	NaN
10846	NaN
10847	NaN
10848	NaN
10849	NaN
10850	NaN
10851	NaN
10852	NaN
10853	NaN
10854	NaN

10855	NaN
10856	NaN
10857	NaN
10858	NaN
10859	NaN
10860	NaN
10861	NaN
10862	NaN
10863	NaN
10864	NaN
10865	NaN

	director \
0	Colin Trevorrow
1	George Miller
2	Robert Schwentke
3	J.J. Abrams
4	James Wan
5	Alejandro Gonz�lez I����rritu
6	Alan Taylor
7	Ridley Scott
8	Kyle Balda Pierre Coffin
9	Pete Docter
10	Sam Mendes
11	Lana Wachowski Lilly Wachowski
12	Alex Garland
13	Chris Columbus
14	Joss Whedon
15	Quentin Tarantino
16	Olivier Megaton
17	Peyton Reed
18	Kenneth Branagh
19	Francis Lawrence
20	Brad Bird
21	Antoine Fuqua
22	Brad Peyton
23	Sam Taylor-Johnson
24	Adam McKay
25	Christopher McQuarrie
26	Seth MacFarlane
27	Matthew Vaughn
28	Tom McCarthy
29	Wes Ball
...	...
10836	Charles Walters
10837	John Guillermin
10838	Richard Brooks
10839	Bill Melendez

10840	Guy Hamilton
10841	Monte Hellman
10842	Wolfgang Reitherman
10843	Basil Dearden Eliot Elisofon
10844	Daniel Mann
10845	Gerald Thomas
10846	Terence Fisher
10847	Terence Fisher
10848	Richard Fleischer
10849	Ronald Neame
10850	Jack Smight
10851	James Hill
10852	Fielder Cook
10853	Lewis Gilbert
10854	Arthur Penn
10855	Alan Rafkin
10856	Norman Tokar
10857	Henry Hathaway
10858	Norman Jewison
10859	John Frankenheimer
10860	Gerald Thomas
10861	Bruce Brown
10862	John Frankenheimer
10863	Eldar Ryazanov
10864	Woody Allen
10865	Harold P. Warren

	tagline	...	\
0	The park is open.	...	
1	What a Lovely Day.	...	
2	One Choice Can Destroy You	...	
3	Every generation has a story.	...	
4	Vengeance Hits Home	...	
5	(n. One who has returned, as if from the dead.)	...	
6	Reset the future	...	
7	Bring Him Home	...	
8	Before Gru, they had a history of bad bosses	...	
9	Meet the little voices inside your head.	...	
10	A Plan No One Escapes	...	
11	Expand your universe.	...	
12	There is nothing more human than the will to s...	...	
13	Game On.	...	
14	A New Age Has Come.	...	
15	No one comes up here without a damn good reason.	...	
16	It Ends Here	...	
17	Heroes Don't Get Any Bigger	...	
18	Midnight is just the beginning.	...	
19	The fire will burn forever.	...	

20	Imagine a world where nothing is impossible.	...
21	Believe in Hope.	...
22	A rescue pilot survived an earthquake, this is...	...
23	Are you curious?	...
24	This is a true story.	...
25	Desperate Times. Desperate Measures.	...
26	Ted is Coming, Again.	...
27	Manners maketh man.	...
28	Break the story. Break the silence.	...
29	The Maze Was Just the Beginning.	...
...
10836	Run, don't walk to see Walk, Don't Run.	...
10837	There was no quiet on the Western Front!	...
10838	Rough, tough and ready.	...
10839	Every year he rises from the pumpkin patch...	...
10840	NaN	...
10841	Suspenseful desert pursuit in the "High Noon"
10842	NaN	...
10843	Where the Nile divides, the great Cinerama adv...	...
10844	The ORIGINAL man of mystery!	...
10845	How the west was lost!	...
10846	DEAD for Ten Years DRACULA, Prince of Darkness...	...
10847	How could they stop the devouring death...that...	...
10848	A Fantastic and Spectacular Voyage... Through
10849	Shirley MacLaine raises Michael Caine!	...
10850	Harper takes a case - and the payoff is murder.	...
10851	From The Pages Of The Beloved Best Seller... A...	...
10852	All the action you can take...all the adventur...	...
10853	Is any man an Alfie? Ask any girl!	...
10854	The chase is on!	...
10855	G-G-GUARANTEED! YOU'LL BE SCARED UNTIL YOU LAU...	...
10856	A HAPPY HONEYMOON GOES TO THE DOGS!...When a G...	...
10857	Some called him savage- and some called him sa...	...
10858	IT'S A PLOT! ...to make the world die laughing!!	...
10859	NaN	...
10860	Carry On Screaming with the Hilarious CARRY ON...	...
10861	NaN	...
10862	Cinerama sweeps YOU into a drama of speed and
10863	NaN	...
10864	WOODY ALLEN STRIKES BACK!	...
10865	It's Shocking! It's Beyond Your Imagination!	...

	overview runtime \	
0	Twenty-two years after the events of Jurassic ...	124
1	An apocalyptic story set in the furthest reach...	120
2	Beatrice Prior must confront her inner demons ...	119
3	Thirty years after defeating the Galactic Empi...	136
4	Deckard Shaw seeks revenge against Dominic Tor...	137

5	In the 1820s, a frontiersman, Hugh Glass, sets...	156
6	The year is 2029. John Connor, leader of the r...	125
7	During a manned mission to Mars, Astronaut Mar...	141
8	Minions Stuart, Kevin and Bob are recruited by...	91
9	Growing up can be a bumpy road, and it's no ex...	94
10	A cryptic message from Bondâs past sends him...	148
11	In a universe where human genetic material is ...	124
12	Caleb, a 26 year old coder at the world's larg...	108
13	Video game experts are recruited by the milita...	105
14	When Tony Stark tries to jumpstart a dormant p...	141
15	Bounty hunters seek shelter from a raging bliz...	167
16	Ex-government operative Bryan Mills finds his ...	109
17	Armed with the astonishing ability to shrink i...	115
18	When her father unexpectedly passes away, youn...	112
19	With the nation of Panem in a full scale war, ...	136
20	Bound by a shared destiny, a bright, optimisti...	130
21	Billy "The Great" Hope, the reigning junior mi...	123
22	In the aftermath of a massive earthquake in Ca...	114
23	When college senior Anastasia Steele steps in ...	125
24	The men who made millions from a global econom...	130
25	Ethan and team take on their most impossible m...	131
26	Newlywed couple Ted and Tami-Lynn want to have...	115
27	The story of a super-secret spy organization t...	130
28	The true story of how The Boston Globe uncover...	128
29	Thomas and his fellow Gladers face their great...	132
...
10836	British industrialist Sir William Rutland - "B...	114
10837	A young pilot in the German air force of 1918,...	156
10838	The Professionals is a 1966 American Western f...	117
10839	This classic "Peanuts" tale focuses on the thu...	25
10840	Colonel Stok, a Soviet intelligence officer re...	102
10841	A hired gun seeks to enact revenge on a group ...	82
10842	Christopher Robin's bear attempts to raid a be...	25
10843	English General Charles George Gordon, a devou...	134
10844	When scientists use eco-terrorism to impose th...	108
10845	Stodge City is in the grip of the Rumpo Kid an...	93
10846	Whilst vacationing in the Carpathian Mountain,...	90
10847	A small island community is overrun with creep...	89
10848	The science of miniaturization has been unlock...	100
10849	Harry Dean (Michael Caine) has a perfect plan ...	109
10850	Harper is a cynical private eye in the best tr...	121
10851	Born Free (1966) is an Open Road Films Ltd./Co...	95
10852	A naive traveler in Laredo gets involved in a ...	95
10853	The film tells the story of a young man who le...	114
10854	Most everyone in town thinks that Sheriff Cald...	135
10855	Luther Heggs aspires to being a reporter for h...	90
10856	The Garrisons (Dean Jones and Suzanne Pleshett...	93
10857	Nevada Smith is the young son of an Indian mot...	128

10858	Without hostile intent, a Soviet sub runs agro...	126
10859	A secret organisation offers wealthy people a ...	100
10860	The sinister Dr Watt has an evil scheme going...	87
10861	The Endless Summer, by Bruce Brown, is one of ...	95
10862	Grand Prix driver Pete Aron is fired by his te...	176
10863	An insurance agent who moonlights as a carthie...	94
10864	In comic Woody Allen's film debut, he took the...	80
10865	A family gets lost on the road and stumbles up...	74

	genres \
0	Action Adventure Science Fiction Thriller
1	Action Adventure Science Fiction Thriller
2	Adventure Science Fiction Thriller
3	Action Adventure Science Fiction Fantasy
4	Action Crime Thriller
5	Western Drama Adventure Thriller
6	Science Fiction Action Thriller Adventure
7	Drama Adventure Science Fiction
8	Family Animation Adventure Comedy
9	Comedy Animation Family
10	Action Adventure Crime
11	Science Fiction Fantasy Action Adventure
12	Drama Science Fiction
13	Action Comedy Science Fiction
14	Action Adventure Science Fiction
15	Crime Drama Mystery Western
16	Crime Action Thriller
17	Science Fiction Action Adventure
18	Romance Fantasy Family Drama
19	War Adventure Science Fiction
20	Action Family Science Fiction Adventure Mystery
21	Action Drama
22	Action Drama Thriller
23	Drama Romance
24	Comedy Drama
25	Action
26	Comedy
27	Crime Comedy Action Adventure
28	Drama Thriller History
29	Action Science Fiction Thriller
...	...
10836	Comedy Romance
10837	War Action Adventure Drama
10838	Action Adventure Western
10839	Family Animation
10840	Thriller
10841	Western
10842	Animation Family

10843	Adventure Drama War History Action
10844	Adventure Comedy Fantasy Science Fiction
10845	Comedy Western
10846	Horror
10847	Science Fiction Horror
10848	Adventure Science Fiction
10849	Action Comedy Crime
10850	Action Drama Thriller Crime Mystery
10851	Adventure Drama Action Family Foreign
10852	Western
10853	Comedy Drama Romance
10854	Thriller Drama Crime
10855	Comedy Family Mystery Romance
10856	Comedy Drama Family
10857	Action Western
10858	Comedy War
10859	Mystery Science Fiction Thriller Drama
10860	Comedy
10861	Documentary
10862	Action Adventure Drama
10863	Mystery Comedy
10864	Action Comedy
10865	Horror

	production_companies	release_date \
0	Universal Studios Amblin Entertainment Legenda...	6/9/15
1	Village Roadshow Pictures Kennedy Miller Produ...	5/13/15
2	Summit Entertainment Mandeville Films Red Wago...	3/18/15
3	Lucasfilm Truenorth Productions Bad Robot	12/15/15
4	Universal Pictures Original Film Media Rights ...	4/1/15
5	Regency Enterprises Appian Way CatchPlay Anony...	12/25/15
6	Paramount Pictures Skydance Productions	6/23/15
7	Twentieth Century Fox Film Corporation Scott F...	9/30/15
8	Universal Pictures Illumination Entertainment	6/17/15
9	Walt Disney Pictures Pixar Animation Studios W...	6/9/15
10	Columbia Pictures Danjaq B24	10/26/15
11	Village Roadshow Pictures Dune Entertainment A...	2/4/15
12	DNA Films Universal Pictures International (UP...	1/21/15
13	Columbia Pictures Happy Madison Productions	7/16/15
14	Marvel Studios Prime Focus Revolution Sun Studios	4/22/15
15	Double Feature Films The Weinstein Company Fil...	12/25/15
16	Twentieth Century Fox Film Corporation M6 Film...	1/1/15
17	Marvel Studios	7/14/15
18	Walt Disney Pictures Genre Films Beagle Pug Fi...	3/12/15
19	Studio Babelsberg StudioCanal Lionsgate Walt D...	11/18/15
20	Walt Disney Pictures Babieka A113	5/19/15
21	Escape Artists Riche-Ludwig Productions	6/15/15
22	New Line Cinema Village Roadshow Pictures Warn...	5/27/15

23	Focus Features Trigger Street Productions Mich...	2/11/15
24	Paramount Pictures Plan B Entertainment Regenc...	12/11/15
25	Paramount Pictures Skydance Productions China ...	7/23/15
26	Universal Pictures Media Rights Capital Fuzzy ...	6/25/15
27	Twentieth Century Fox Film Corporation Marv Fi...	1/24/15
28	Participant Media Open Road Films Anonymous Co...	11/6/15
29	Gotham Group Temple Hill Entertainment TSG Ent...	9/9/15
...
10836	Columbia Pictures Corporation	1/1/66
10837	Twentieth Century Fox Film Corporation	6/21/66
10838	Columbia Pictures	11/1/66
10839	Warner Bros. Home Video	10/27/66
10840	Lowndes Productions Limited	12/22/66
10841	Proteus Films	10/23/66
10842	NaN	1/1/66
10843	Julian Blaustein Productions Ltd.	6/9/66
10844	20th Century Fox	1/16/66
10845	Peter Rogers Productions	3/1/66
10846	Seven Arts Productions Hammer Film Productions	1/9/66
10847	Planet Film Productions Protelco	6/20/66
10848	Twentieth Century Fox Film Corporation	8/24/66
10849	Universal Pictures	12/16/66
10850	Warner Bros.	2/23/66
10851	High Road	6/22/66
10852	Eden Productions Inc.	5/31/66
10853	NaN	3/29/66
10854	Horizon Pictures Columbia Pictures Corporation	2/17/66
10855	Universal Pictures	1/20/66
10856	Walt Disney Pictures	2/16/66
10857	Paramount Pictures Solar Productions Embassy P...	6/10/66
10858	The Mirisch Corporation	5/25/66
10859	Gibraltar Productions Joel Productions John Fr...	10/5/66
10860	Peter Rogers Productions Anglo-Amalgamated Fil...	5/20/66
10861	Bruce Brown Films	6/15/66
10862	Cherokee Productions Joel Productions Douglas ...	12/21/66
10863	Mosfilm	1/1/66
10864	Benedict Pictures Corp.	11/2/66
10865	Norm-Iris	11/15/66

	vote_count	vote_average	release_year	budget_adj	revenue_adj
0	5562	6.5	2015	1.379999e+08	1.392446e+09
1	6185	7.1	2015	1.379999e+08	3.481613e+08
2	2480	6.3	2015	1.012000e+08	2.716190e+08
3	5292	7.5	2015	1.839999e+08	1.902723e+09
4	2947	7.3	2015	1.747999e+08	1.385749e+09
5	3929	7.2	2015	1.241999e+08	4.903142e+08
6	2598	5.8	2015	1.425999e+08	4.053551e+08
7	4572	7.6	2015	9.935996e+07	5.477497e+08

8	2893	6.5	2015	6.807997e+07	1.064192e+09
9	3935	8.0	2015	1.609999e+08	7.854116e+08
10	3254	6.2	2015	2.253999e+08	8.102203e+08
11	1937	5.2	2015	1.619199e+08	1.692686e+08
12	2854	7.6	2015	1.379999e+07	3.391985e+07
13	1575	5.8	2015	8.095996e+07	2.241460e+08
14	4304	7.4	2015	2.575999e+08	1.292632e+09
15	2389	7.4	2015	4.047998e+07	1.432992e+08
16	1578	6.1	2015	4.415998e+07	2.997096e+08
17	3779	7.0	2015	1.195999e+08	4.771138e+08
18	1495	6.8	2015	8.739996e+07	4.989630e+08
19	2380	6.5	2015	1.471999e+08	5.984813e+08
20	1899	6.2	2015	1.747999e+08	1.923127e+08
21	1386	7.3	2015	2.759999e+07	8.437300e+07
22	2060	6.1	2015	1.012000e+08	4.328514e+08
23	1865	5.3	2015	3.679998e+07	5.240791e+08
24	1545	7.3	2015	2.575999e+07	1.226787e+08
25	2349	7.1	2015	1.379999e+08	6.277435e+08
26	1666	6.3	2015	6.255997e+07	1.985944e+08
27	3833	7.6	2015	7.451997e+07	3.714978e+08
28	1559	7.8	2015	1.839999e+07	8.127872e+07
29	1849	6.4	2015	5.611998e+07	2.863562e+08
...
10836	11	5.8	1966	0.000000e+00	0.000000e+00
10837	12	5.5	1966	0.000000e+00	0.000000e+00
10838	21	6.0	1966	0.000000e+00	0.000000e+00
10839	49	7.2	1966	0.000000e+00	0.000000e+00
10840	13	5.7	1966	0.000000e+00	0.000000e+00
10841	12	5.5	1966	5.038511e+05	0.000000e+00
10842	12	7.9	1966	0.000000e+00	0.000000e+00
10843	12	5.8	1966	0.000000e+00	0.000000e+00
10844	13	5.6	1966	0.000000e+00	0.000000e+00
10845	15	5.9	1966	0.000000e+00	0.000000e+00
10846	16	5.7	1966	0.000000e+00	0.000000e+00
10847	13	5.3	1966	0.000000e+00	0.000000e+00
10848	42	6.7	1966	3.436265e+07	8.061618e+07
10849	14	6.1	1966	0.000000e+00	0.000000e+00
10850	14	6.0	1966	0.000000e+00	0.000000e+00
10851	15	6.6	1966	0.000000e+00	0.000000e+00
10852	11	6.0	1966	0.000000e+00	0.000000e+00
10853	26	6.2	1966	0.000000e+00	0.000000e+00
10854	17	6.0	1966	0.000000e+00	0.000000e+00
10855	14	6.1	1966	4.702610e+06	0.000000e+00
10856	14	5.7	1966	0.000000e+00	0.000000e+00
10857	10	5.9	1966	0.000000e+00	0.000000e+00
10858	11	5.5	1966	0.000000e+00	0.000000e+00
10859	22	6.6	1966	0.000000e+00	0.000000e+00
10860	13	7.0	1966	0.000000e+00	0.000000e+00

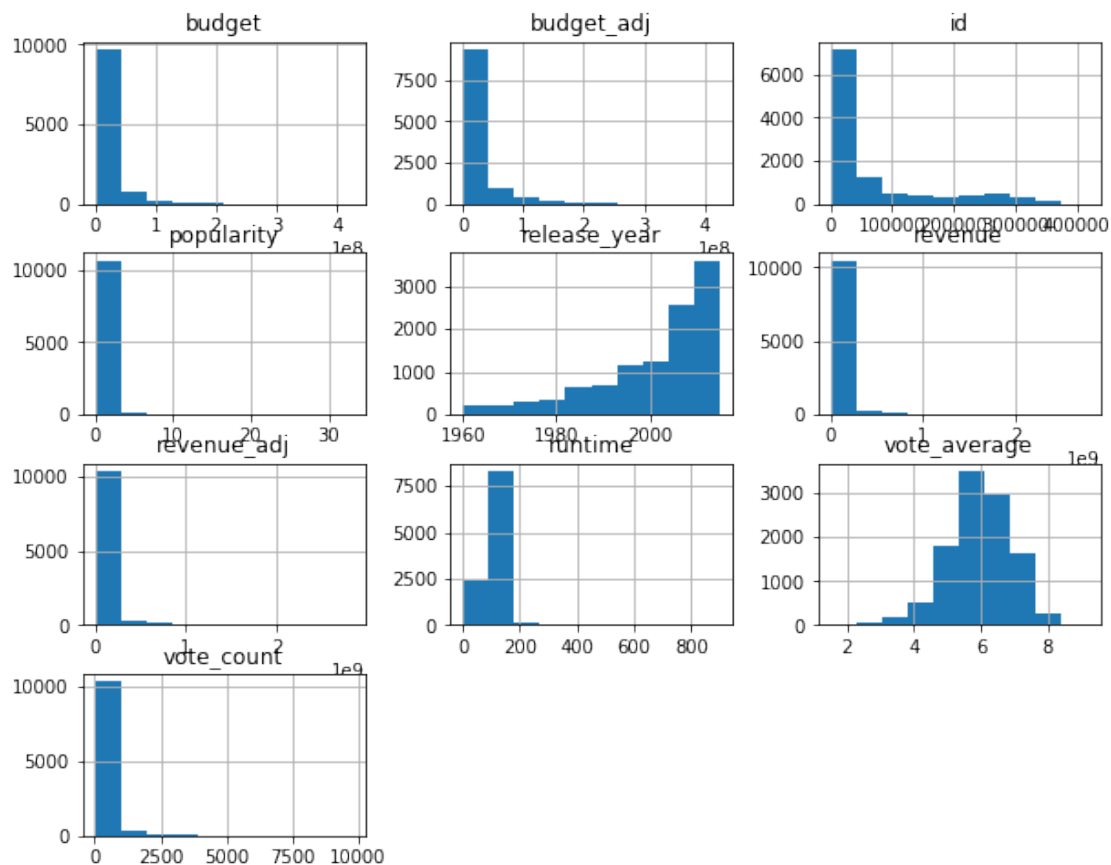
10861	11	7.4	1966	0.000000e+00	0.000000e+00
10862	20	5.7	1966	0.000000e+00	0.000000e+00
10863	11	6.5	1966	0.000000e+00	0.000000e+00
10864	22	5.4	1966	0.000000e+00	0.000000e+00
10865	15	1.5	1966	1.276423e+05	0.000000e+00

[10866 rows x 21 columns]>

In [224]: *# viewing of the total rows and columns*
df.shape

Out[224]: (10866, 21)

In [225]: *# give visuality of dataset*
df.hist(figsize=(10, 8));



1. The chart for the movie dataset is asymmetrical distribution because it all skewed either to right or left.
2. The vote_average chart is a bit similar to a normal distribution but it is not because it didn't create a bell curve.

1.1.3 Data Cleaning

1. The columns that are not needed in analysing will be dropped
2. The missing value will be check before the not needed column will be dropped and after it has being dropped
3. Any missing value that appears after dropping the not needed column, will be drop
4. The dataset for genres will be splitted into individual column

```
In [226]: # find if there is any mising value
df.isna().sum()
```

```
Out[226]: id                0
imdb_id                  10
popularity               0
budget                  0
revenue                 0
original_title          0
cast                   76
homepage              7930
director               44
tagline              2824
keywords             1493
overview               4
runtime                0
genres                 23
production_companies  1030
release_date           0
vote_count             0
vote_average           0
release_year           0
budget_adj             0
revenue_adj            0
dtype: int64
```

```
In [227]: # sum of the total missing value
df.isna().sum().sum()
```

```
Out[227]: 13434
```

```
In [230]: df['revenue_adj'].replace(0,df['revenue_adj'].mean(), inplace=True)
```

```
In [231]: # drop the the set of data we not analysing on
df.drop(columns = ['imdb_id', 'homepage', 'director', 'tagline', 'overview', 'keywords
```

```
In [232]: # show the remaining column
df.head()
```

```

Out[232]:      id  popularity      budget      revenue \
0  135397   32.985763  150000000.0  1.513529e+09
1    76341   28.419936  150000000.0  3.784364e+08
2   262500   13.112507  110000000.0  2.952382e+08
3   140607   11.173104  200000000.0  2.068178e+09
4   168259    9.335014  190000000.0  1.506249e+09

      original_title \
0      Jurassic World
1      Mad Max: Fury Road
2      Insurgent
3  Star Wars: The Force Awakens
4      Furious 7

      cast  runtime \
0  Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...   124.0
1  Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...   120.0
2  Shailene Woodley|Theo James|Kate Winslet|Ansel...   119.0
3  Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...   136.0
4  Vin Diesel|Paul Walker|Jason Statham|Michelle ...   137.0

      genres \
0  Action|Adventure|Science Fiction|Thriller
1  Action|Adventure|Science Fiction|Thriller
2      Adventure|Science Fiction|Thriller
3  Action|Adventure|Science Fiction|Fantasy
4      Action|Crime|Thriller

      production_companies  release_date  vote_count \
0  Universal Studios|Amblin Entertainment|Legenda...    6/9/15         5562
1  Village Roadshow Pictures|Kennedy Miller Produ...    5/13/15         6185
2  Summit Entertainment|Mandeville Films|Red Wago...    3/18/15         2480
3      Lucasfilm|Truenorth Productions|Bad Robot    12/15/15         5292
4  Universal Pictures|Original Film|Media Rights ...    4/1/15         2947

      vote_average  release_year  budget_adj  revenue_adj
0           6.5         2015  1.379999e+08  1.392446e+09
1           7.1         2015  1.379999e+08  3.481613e+08
2           6.3         2015  1.012000e+08  2.716190e+08
3           7.5         2015  1.839999e+08  1.902723e+09
4           7.3         2015  1.747999e+08  1.385749e+09

```

```

In [234]: # check for the missing value after the not needed value is drop
df.isna().sum()

```

```

Out[234]: id          0
popularity          0
budget             0

```

```

revenue                0
original_title         0
cast                   76
runtime                0
genres                 23
production_companies  1030
release_date           0
vote_count             0
vote_average           0
release_year           0
budget_adj             0
revenue_adj            0
dtype: int64

```

```

In [235]: # dropping all missing value
df.dropna(axis = 0, inplace=True)

```

```

In [236]: # to verify that there is no more missing value
df.isna().sum()

```

```

Out[236]: id                0
popularity                0
budget                   0
revenue                  0
original_title           0
cast                     0
runtime                  0
genres                   0
production_companies     0
release_date             0
vote_count               0
vote_average             0
release_year             0
budget_adj               0
revenue_adj              0
dtype: int64

```

```

In [237]: # seperate the dataset of genres into individual column
genres = df['genres'].str.split('|', expand=True)
genres.head()

```

```

Out[237]:
   0      1      2      3      4
0  Action  Adventure  Science Fiction  Thriller  None
1  Action  Adventure  Science Fiction  Thriller  None
2  Adventure  Science Fiction  Thriller  None  None
3  Action  Adventure  Science Fiction  Fantasy  None
4  Action      Crime  Thriller  None  None

```

The missing value was checked twice to know the exact missing value before it is being cleaned and after it was cleaned.

Exploratory Data Analysis

```
In [238]: # set the individual column into one column
joined_genres = genres[0].append(genres[1]).append(genres[2]).append(genres[3]).append(genres[4])
joined_genres.head()
```

```
In [239]: # give a clear visualization to show the most popular genres from year to year
joined_genres.value_counts().plot(kind='bar')
plot.xlabel('Genres')
plot.ylabel('Occurences')
plot.title('Most occuring genres')
```

21

The most popular genres from year to year is Drama, follow by the remaining genres, that is, Comedy, Thriller, Action, etc.

1.1.5 Research Question 2 (what kinds of properties are associated with movies that have high revenue)

```
In [240]: # get revenue greater than zero since we are dealing with high revenue
high_value = df[df['revenue'] > 0].head()
high_value
```

```
Out[240]:
```

	id	popularity	budget	revenue \
0	135397	32.985763	150000000.0	1.513529e+09
1	76341	28.419936	150000000.0	3.784364e+08
2	262500	13.112507	110000000.0	2.952382e+08
3	140607	11.173104	200000000.0	2.068178e+09
4	168259	9.335014	190000000.0	1.506249e+09

	original_title \
0	Jurassic World
1	Mad Max: Fury Road
2	Insurgent
3	Star Wars: The Force Awakens
4	Furious 7

	cast	runtime \
0	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	124.0
1	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	120.0
2	Shailene Woodley Theo James Kate Winslet Ansel...	119.0
3	Harrison Ford Mark Hamill Carrie Fisher Adam D...	136.0
4	Vin Diesel Paul Walker Jason Statham Michelle ...	137.0

	genres \
0	Action Adventure Science Fiction Thriller
1	Action Adventure Science Fiction Thriller
2	Adventure Science Fiction Thriller
3	Action Adventure Science Fiction Fantasy
4	Action Crime Thriller

	production_companies	release_date	vote_count \
0	Universal Studios Amblin Entertainment Legenda...	6/9/15	5562
1	Village Roadshow Pictures Kennedy Miller Produ...	5/13/15	6185
2	Summit Entertainment Mandeville Films Red Wago...	3/18/15	2480
3	Lucasfilm Truenorth Productions Bad Robot	12/15/15	5292
4	Universal Pictures Original Film Media Rights ...	4/1/15	2947

	vote_average	release_year	budget_adj	revenue_adj
0	6.5	2015	1.379999e+08	1.392446e+09
1	7.1	2015	1.379999e+08	3.481613e+08

2	6.3	2015	1.012000e+08	2.716190e+08
3	7.5	2015	1.839999e+08	1.902723e+09
4	7.3	2015	1.747999e+08	1.385749e+09

In [241]: # get the revenue from the highest to the lowest

```
revenue_order = high_value.sort_values(by='revenue', ascending=False).head()
revenue_order
```

Out[241]:

	id	popularity	budget	revenue \
3	140607	11.173104	200000000.0	2.068178e+09
0	135397	32.985763	150000000.0	1.513529e+09
4	168259	9.335014	190000000.0	1.506249e+09
1	76341	28.419936	150000000.0	3.784364e+08
2	262500	13.112507	110000000.0	2.952382e+08

	original_title \
3	Star Wars: The Force Awakens
0	Jurassic World
4	Furious 7
1	Mad Max: Fury Road
2	Insurgent

	cast	runtime \
3	Harrison Ford Mark Hamill Carrie Fisher Adam D...	136.0
0	Chris Pratt Bryce Dallas Howard Irrfan Khan Vi...	124.0
4	Vin Diesel Paul Walker Jason Statham Michelle ...	137.0
1	Tom Hardy Charlize Theron Hugh Keays-Byrne Nic...	120.0
2	Shailene Woodley Theo James Kate Winslet Ansel...	119.0

	genres \
3	Action Adventure Science Fiction Fantasy
0	Action Adventure Science Fiction Thriller
4	Action Crime Thriller
1	Action Adventure Science Fiction Thriller
2	Adventure Science Fiction Thriller

	production_companies	release_date	vote_count \
3	Lucasfilm Truenorth Productions Bad Robot	12/15/15	5292
0	Universal Studios Amblin Entertainment Legenda...	6/9/15	5562
4	Universal Pictures Original Film Media Rights ...	4/1/15	2947
1	Village Roadshow Pictures Kennedy Miller Produ...	5/13/15	6185
2	Summit Entertainment Mandeville Films Red Wago...	3/18/15	2480

	vote_average	release_year	budget_adj	revenue_adj
3	7.5	2015	1.839999e+08	1.902723e+09
0	6.5	2015	1.379999e+08	1.392446e+09
4	7.3	2015	1.747999e+08	1.385749e+09
1	7.1	2015	1.379999e+08	3.481613e+08
2	6.3	2015	1.012000e+08	2.716190e+08

1.1.6 Research Question 3 (what are the five(5) top most profitable movies(revenue vs budget))

```
In [242]: # add revenue/budget to the index
```

```
df['revenue/budget'] = df['revenue_adj']/df['budget_adj']
df_profitable_movies = df.sort_values(by='revenue/budget')
df_profitable_movies.head()
```

```
Out[242]:
```

	id	popularity	budget	revenue	original_title	\
	10307	2259	0.450208	25000000.0	6.0	The House of the Spirits
	5067	13537	0.462609	6000000.0	2.0	Shattered Glass
	8142	2293	0.552091	6000000.0	2.0	Mallrats
	8523	9431	0.578849	15000000.0	5.0	Bordello of Blood
	7158	14844	0.420568	23000000.0	12.0	Chasing Liberty

		cast	runtime	\
	10307	Meryl Streep Glenn Close Jeremy Irons Winona R...	140.0	
	5067	Hayden Christensen Peter Sarsgaard Chloë Sevi...	94.0	
	8142	Jason Lee Jeremy London Shannen Doherty Claire...	94.0	
	8523	Dennis Miller Erika Eleniak Angie Everhart Joh...	87.0	
	7158	Mandy Moore Stark Sands Tony Jayawardena Jerem...	111.0	

		genres	production_companies	\
	10307	Romance Drama	Det Danske Filminstitut Spring Creek Productio...	
	5067	Drama History	Lions Gate Films Cruise/Wagner Productions Bau...	
	8142	Romance Comedy	Gramercy Pictures Alphaville Films View Askew ...	
	8523	Horror Comedy	Tales From The Crypt Holdings Universal City S...	
	7158	Comedy Romance	Alcon Entertainment ETIC Films C.R.G. Internat...	

	release_date	vote_count	vote_average	release_year	budget_adj	\
	10307	10/19/93	52	6.7	1993	3.773675e+07
	5067	11/14/03	46	6.4	2003	7.112116e+06
	8142	10/20/95	201	6.8	1995	8.585801e+06
	8523	8/16/96	27	4.9	1996	2.085325e+07
	7158	1/9/04	86	6.2	2004	2.655224e+07

	revenue_adj	revenue/budget	
	10307	9.056820	2.400000e-07
	5067	2.370705	3.333333e-07
	8142	2.861934	3.333333e-07
	8523	6.951084	3.333333e-07
	7158	13.853345	5.217391e-07

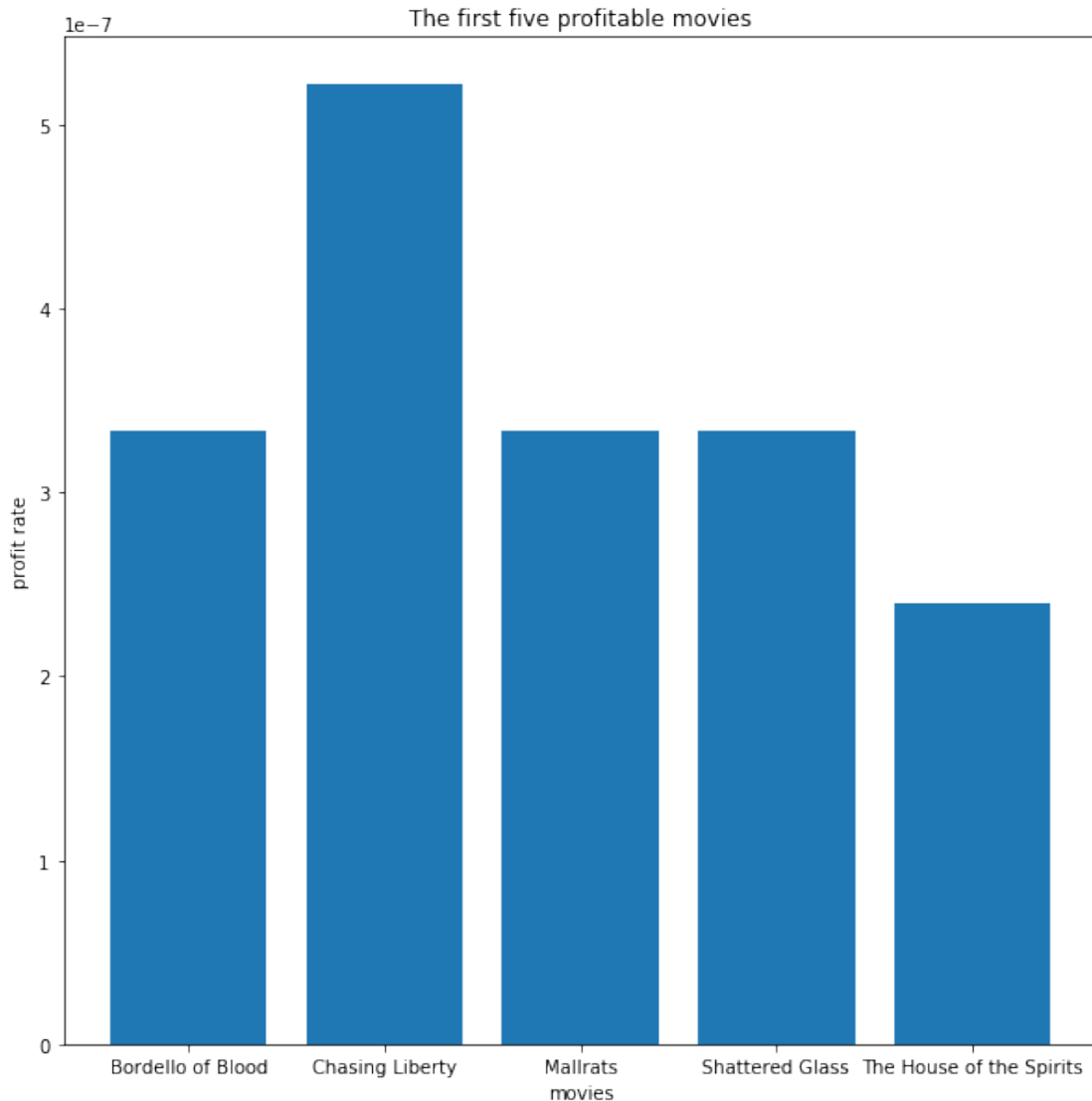
```
In [243]: # showcase the movies that are making much profit
```

```
data = df_profitable_movies[['original_title', 'revenue/budget']].head()
x = data['original_title']
y = data['revenue/budget']
plot.figure(figsize=(10,10))
plot.xlabel('movies')
plot.ylabel('profit rate')
```



```
plot.title('The first five profitable movies')
plot.bar(x,y)
```

Out[243]: <Container object of 5 artists>



1. Chasing Liberty is the movie that is most profitable
2. Bordello of Blood, Mallrats and Shattered Glass is on the same height of making profit

Research Question 4 (write a function for runtime to determine the length of the movie either long or short)

```
In [174]: # give the data in runtime
df['runtime'].head()
```

```
Out[174]: 0    124
          1    120
          2    119
          3    136
          4    137
          Name: runtime, dtype: int64
```

```
In [175]: # the average rate for runtime
          df['runtime'].mean()
```

```
Out[175]: 102.07086324314375
```

```
In [176]: # write to find the length of each movies
          def print_runtime(x):
              if x <= 102.07:
                  return 'short runtime'
              elif x > 102.07:
                  return 'long runtime'
          df['length']=df['runtime'].apply(print_runtime)
```

```
In [177]: # show the column added to the the dataset
          df.head()
```

```
Out[177]:
```

	id	imdb_id	popularity	budget	revenue \
0	135397	tt0369610	32.985763	150000000	1513528810
1	76341	tt1392190	28.419936	150000000	378436354
2	262500	tt2908446	13.112507	110000000	295238201
3	140607	tt2488496	11.173104	200000000	2068178225
4	168259	tt2820852	9.335014	190000000	1506249360

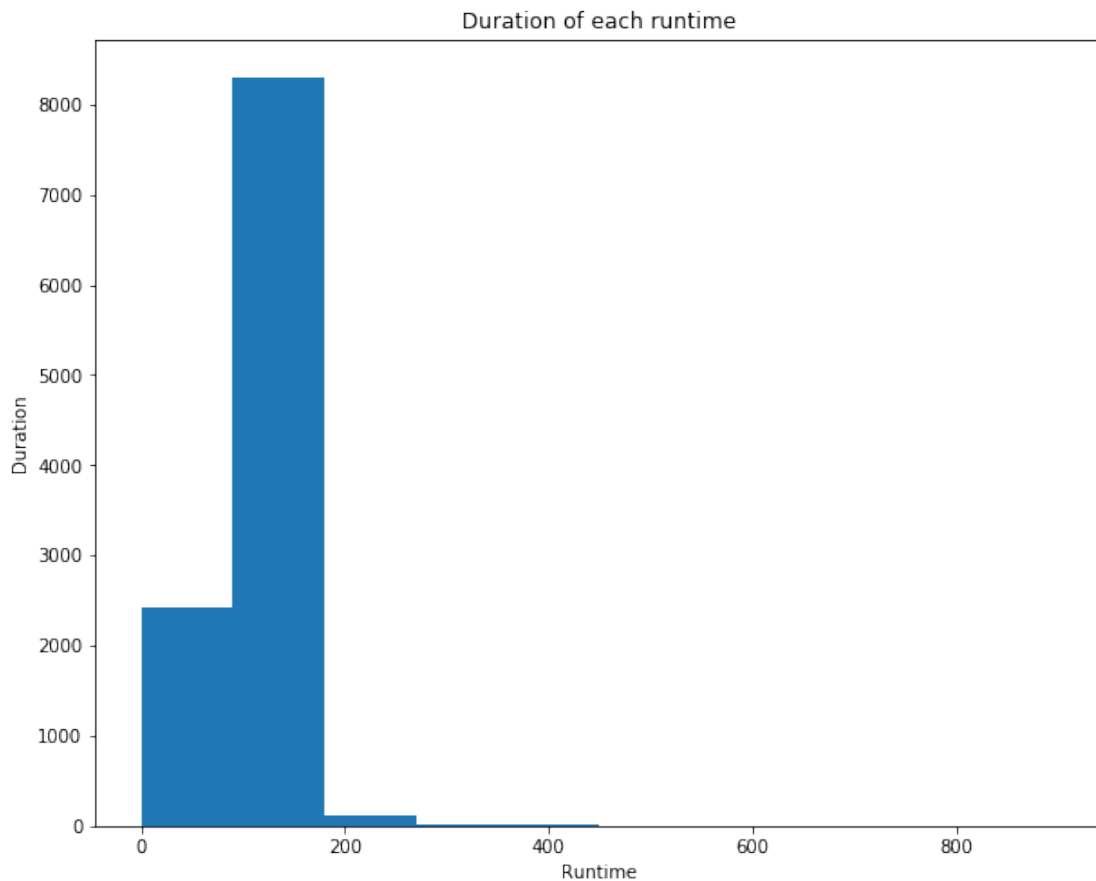
	original_title	runtime \
0	Jurassic World	124
1	Mad Max: Fury Road	120
2	Insurgent	119
3	Star Wars: The Force Awakens	136
4	Furious 7	137

	genres	release_date	release_year \
0	Action Adventure Science Fiction Thriller	6/9/15	2015
1	Action Adventure Science Fiction Thriller	5/13/15	2015
2	Adventure Science Fiction Thriller	3/18/15	2015
3	Action Adventure Science Fiction Fantasy	12/15/15	2015
4	Action Crime Thriller	4/1/15	2015

	budget_adj	revenue_adj	revenue/budget	length
0	1.379999e+08	1.392446e+09	10.090192	long runtime
1	1.379999e+08	3.481613e+08	2.522909	long runtime
2	1.012000e+08	2.716190e+08	2.683984	long runtime
3	1.839999e+08	1.902723e+09	10.340891	long runtime
4	1.747999e+08	1.385749e+09	7.927628	long runtime

```
In [197]: # the highest runtime for each movie
plot.figure(figsize=(10, 8))
plot.hist(df['runtime'])
plot.xlabel('Runtime')
plot.ylabel('Duration')
plot.title('Duration of each runtime')

Out[197]: Text(0.5,1,'Duration of each runtime')
```



The chart show that, runtime that appears most for the movies is lesser than 200
Conclusions

1. The movies with the highest revenue and budget has the highest popularity
2. The dataset has much value that represent zero(0)
3. The release_year is extracted from release_date

1.2 The data limitation

During my work on the project, I found out that there is a lot of missing value, which requires replacement because if it is not replaced, some chat that will be plotted will be vacant.

```
In [27]: from subprocess import call
         call(['python', '-m', 'nbconvert', 'Investigate_a_Dataset.ipynb'])
```

```
Out[27]: 0
```

```
In [ ]:
```