

Investigate_a_Dataset

August 5, 2021

1 Project: Investigate a Movie Dataset

1.1 Table of Contents

Introduction

Data Wrangling

Exploratory Data Analysis

Conclusions

Movie Dataset Analysis

1.2 Introduction

In this project i use TMDb Movie dataset which contains a large number of row and coulumnns, each coulumnns provide detailes information about the movie like relase year and revenues and so on..

My Questions that i explore it from the dataset is:

1-What is the high of top10 movies that get the high revenue?

2- Do high budget generate high revenue?

3- What is the most repetaed type of movies?

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

Data Wrangling

1.2.1 General Properties

```
In [2]: # Load your data and print out a few lines. Perform operations to inspect data
# types and look for instances of missing or possibly errant data.
movie_df = pd.read_csv('tmdb-movies.csv')
movie_df
```

```

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
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a 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...	6/9/2015
1	Village Roadshow Pictures Kennedy Miller Produ...	5/13/2015
2	Summit Entertainment Mandeville Films Red Wago...	3/18/2015
3	Lucasfilm Truenorth Productions Bad Robot	12/15/2015
4	Universal Pictures Original Film Media Rights ...	4/1/2015
5	Regency Enterprises Appian Way CatchPlay Anony...	12/25/2015
6	Paramount Pictures Skydance Productions	6/23/2015
7	Twentieth Century Fox Film Corporation Scott F...	9/30/2015
8	Universal Pictures Illumination Entertainment	6/17/2015
9	Walt Disney Pictures Pixar Animation Studios W...	6/9/2015
10	Columbia Pictures Danjaq B24	10/26/2015
11	Village Roadshow Pictures Dune Entertainment A...	2/4/2015
12	DNA Films Universal Pictures International (UP...	1/21/2015
13	Columbia Pictures Happy Madison Productions	7/16/2015
14	Marvel Studios Prime Focus Revolution Sun Studios	4/22/2015
15	Double Feature Films The Weinstein Company Fil...	12/25/2015
16	Twentieth Century Fox Film Corporation M6 Film...	1/1/2015
17	Marvel Studios	7/14/2015
18	Walt Disney Pictures Genre Films Beagle Pug Fi...	3/12/2015
19	Studio Babelsberg StudioCanal Lionsgate Walt D...	11/18/2015
20	Walt Disney Pictures Babieka A113	5/19/2015
21	Escape Artists Riche-Ludwig Productions	6/15/2015
22	New Line Cinema Village Roadshow Pictures Warn...	5/27/2015

23	Focus Features Trigger Street Productions Mich...	2/11/2015
24	Paramount Pictures Plan B Entertainment Regenc...	12/11/2015
25	Paramount Pictures Skydance Productions China ...	7/23/2015
26	Universal Pictures Media Rights Capital Fuzzy ...	6/25/2015
27	Twentieth Century Fox Film Corporation Marv Fi...	1/24/2015
28	Participant Media Open Road Films Anonymous Co...	11/6/2015
29	Gotham Group Temple Hill Entertainment TSG Ent...	9/9/2015
...
10836	Columbia Pictures Corporation	1/1/1966
10837	Twentieth Century Fox Film Corporation	6/21/1966
10838	Columbia Pictures	11/1/1966
10839	Warner Bros. Home Video	10/27/1966
10840	Lowndes Productions Limited	12/22/1966
10841	Proteus Films	10/23/1966
10842	NaN	1/1/1966
10843	Julian Blaustein Productions Ltd.	6/9/1966
10844	20th Century Fox	1/16/1966
10845	Peter Rogers Productions	3/1/1966
10846	Seven Arts Productions Hammer Film Productions	1/9/1966
10847	Planet Film Productions Protelco	6/20/1966
10848	Twentieth Century Fox Film Corporation	8/24/1966
10849	Universal Pictures	12/16/1966
10850	Warner Bros.	2/23/1966
10851	High Road	6/22/1966
10852	Eden Productions Inc.	5/31/1966
10853	NaN	3/29/1966
10854	Horizon Pictures Columbia Pictures Corporation	2/17/1966
10855	Universal Pictures	1/20/1966
10856	Walt Disney Pictures	2/16/1966
10857	Paramount Pictures Solar Productions Embassy P...	6/10/1966
10858	The Mirisch Corporation	5/25/1966
10859	Gibraltar Productions Joel Productions John Fr...	10/5/1966
10860	Peter Rogers Productions Anglo-Amalgamated Fil...	5/20/1966
10861	Bruce Brown Films	6/15/1966
10862	Cherokee Productions Joel Productions Douglas ...	12/21/1966
10863	Mosfilm	1/1/1966
10864	Benedict Pictures Corp.	11/2/1966
10865	Norm-Iris	11/15/1966

	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 [3]: movie_df.shape

Out[3]: (10866, 21)

In [4]: movie_df.describe()

Out[4]:

	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 [5]: movie_df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10866 entries, 0 to 10865
Data columns (total 21 columns):
id                10866 non-null int64
imdb_id           10856 non-null object
popularity        10866 non-null float64
budget            10866 non-null int64
revenue          10866 non-null int64
original_title    10866 non-null object
cast              10790 non-null object
homepage          2936 non-null object
director          10822 non-null object
tagline           8042 non-null object
```

```

keywords          9373 non-null object
overview          10862 non-null object
runtime           10866 non-null int64
genres            10843 non-null object
production_companies 9836 non-null object
release_date      10866 non-null object
vote_count        10866 non-null int64
vote_average      10866 non-null float64
release_year      10866 non-null int64
budget_adj        10866 non-null float64
revenue_adj       10866 non-null float64
dtypes: float64(4), int64(6), object(11)
memory usage: 1.7+ MB

```

```

In [24]: # This function to find if there is a empty columns in the dataset
         # i choose 2 columns from my dataset

```

```

is_empty = pd.DataFrame(movie_df, columns= ['budget', 'production_companies'])
is_empty= is_empty.empty

```

```

print(is_empty)
# when you get "False" in the output this is mean the column in not empty.
# when the output is "True" this is mean the column in empty.

```

False

1.2.2 Data Cleaning

In this section i cleaned my data to be more clear to use and to get the right information! So first i dropped the coulumn that i will not be use then i see if there is a missing data of my dataset and then i dropped the missing data, Finally i see if there is any duplicate in data then i dropped it also.

That what makes my dataset more clearer to me!

```

In [7]: ## Drop the coulums that i will not be use
        movie_df.drop(['imdb_id', 'homepage', 'cast', 'production_companies', 'keywords', 'tagli

```

```

In [8]: movie_df.head()

```

```

Out[8]:
   popularity  budget  revenue  original_title \
0    32.985763  150000000  1513528810  Jurassic World
1    28.419936  150000000   378436354  Mad Max: Fury Road
2    13.112507  110000000   295238201  Insurgent
3    11.173104  200000000  2068178225  Star Wars: The Force Awakens
4     9.335014  190000000  1506249360  Furious 7

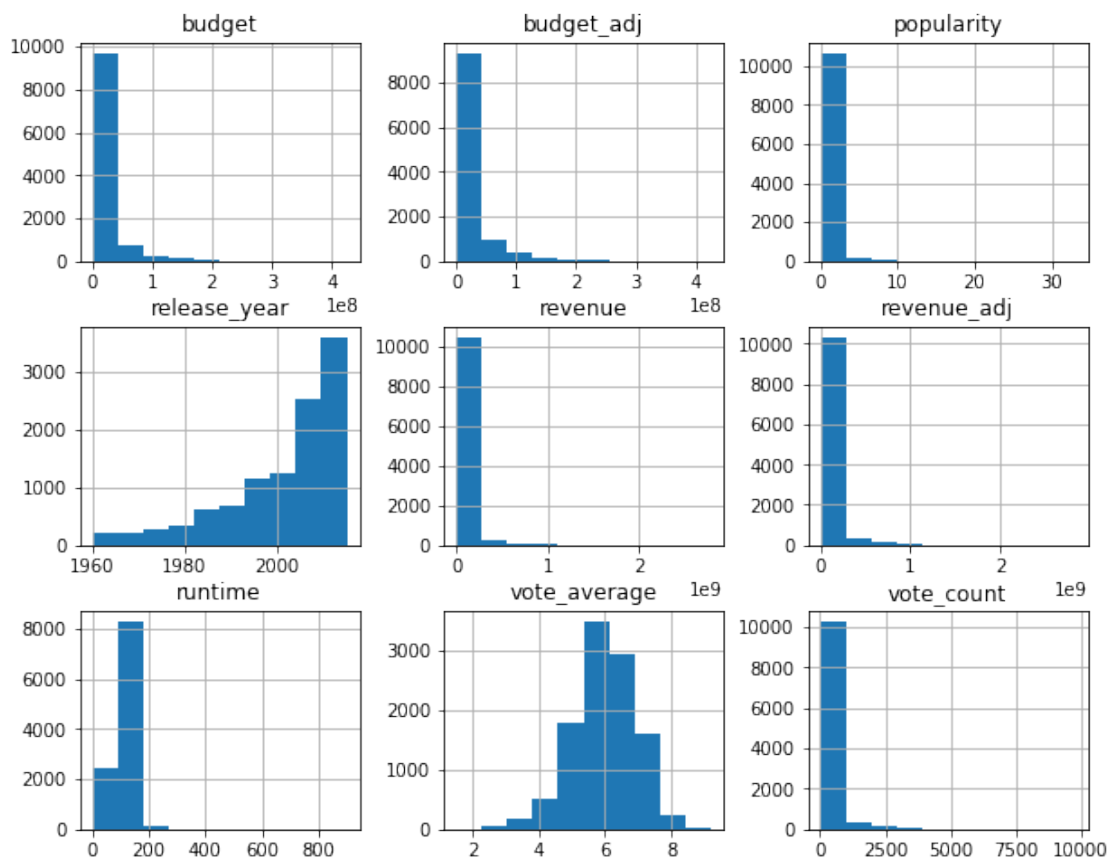
```


	director	runtime	genres \
0	Colin Trevorrow	124	Action Adventure Science Fiction Thriller
1	George Miller	120	Action Adventure Science Fiction Thriller
2	Robert Schwentke	119	Adventure Science Fiction Thriller
3	J.J. Abrams	136	Action Adventure Science Fiction Fantasy
4	James Wan	137	Action Crime Thriller

	release_date	vote_count	vote_average	release_year	budget_adj \
0	6/9/2015	5562	6.5	2015	137999939.3
1	5/13/2015	6185	7.1	2015	137999939.3
2	3/18/2015	2480	6.3	2015	101199955.5
3	12/15/2015	5292	7.5	2015	183999919.0
4	4/1/2015	2947	7.3	2015	174799923.1

	revenue_adj
0	1.392446e+09
1	3.481613e+08
2	2.716190e+08
3	1.902723e+09
4	1.385749e+09

```
In [9]: movie_df.hist(figsize=(10,8));
```



```
In [10]: movie_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10866 entries, 0 to 10865
Data columns (total 13 columns):
popularity          10866 non-null float64
budget              10866 non-null int64
revenue             10866 non-null int64
original_title      10866 non-null object
director            10822 non-null object
runtime             10866 non-null int64
genres              10843 non-null object
release_date        10866 non-null object
vote_count          10866 non-null int64
vote_average        10866 non-null float64
release_year        10866 non-null int64
budget_adj          10866 non-null float64
revenue_adj         10866 non-null float64
dtypes: float64(4), int64(5), object(4)
memory usage: 1.1+ MB
```

```
In [11]: movie_df.dropna(inplace= True)
         movie_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10801 entries, 0 to 10865
Data columns (total 13 columns):
popularity          10801 non-null float64
budget              10801 non-null int64
revenue             10801 non-null int64
original_title      10801 non-null object
director            10801 non-null object
runtime             10801 non-null int64
genres              10801 non-null object
release_date        10801 non-null object
vote_count          10801 non-null int64
vote_average        10801 non-null float64
release_year        10801 non-null int64
budget_adj          10801 non-null float64
revenue_adj         10801 non-null float64
dtypes: float64(4), int64(5), object(4)
memory usage: 1.2+ MB
```

```
In [12]: movie_df.head()
```

```

Out[12]:      popularity      budget      revenue      original_title \
0    32.985763  150000000  1513528810      Jurassic World
1    28.419936  150000000  378436354      Mad Max: Fury Road
2    13.112507  110000000  295238201      Insurgent
3    11.173104  200000000  2068178225  Star Wars: The Force Awakens
4     9.335014  190000000  1506249360      Furious 7

      director  runtime      genres \
0  Colin Trevorrow    124  Action|Adventure|Science Fiction|Thriller
1   George Miller    120  Action|Adventure|Science Fiction|Thriller
2 Robert Schwentke    119      Adventure|Science Fiction|Thriller
3   J.J. Abrams     136  Action|Adventure|Science Fiction|Fantasy
4   James Wan       137      Action|Crime|Thriller

      release_date  vote_count  vote_average  release_year  budget_adj \
0      6/9/2015         5562           6.5         2015  137999939.3
1      5/13/2015         6185           7.1         2015  137999939.3
2      3/18/2015         2480           6.3         2015  101199955.5
3     12/15/2015         5292           7.5         2015  183999919.0
4      4/1/2015         2947           7.3         2015  174799923.1

      revenue_adj
0  1.392446e+09
1  3.481613e+08
2  2.716190e+08
3  1.902723e+09
4  1.385749e+09

```

```
In [13]: sum(movie_df.duplicated())
```

```
Out[13]: 1
```

```
In [14]: movie_df.drop_duplicates(inplace=True)
```

Exploratory Data Analysis

1.2.3 Research Question 1 (What is the high of top10 movies that get the high revenue?)

```
In [15]: movie_sorted= movie_df.sort_values(by='revenue', ascending=False)
```

```
In [16]: movie_sorted
```

```

Out[16]:      popularity      budget      revenue \
1386     9.432768  237000000  2781505847
3      11.173104  200000000  2068178225
5231     4.355219  200000000  1845034188
4361     7.637767  220000000  1519557910
0      32.985763  150000000  1513528810
4      9.335014  190000000  1506249360

```

14	5.944927	280000000	1405035767
3374	5.711315	125000000	1327817822
5422	6.112766	150000000	1274219009
5425	4.946136	200000000	1215439994
8	7.404165	74000000	1156730962
3522	0.760503	195000000	1123746996
4949	7.122455	94000000	1118888979
4365	5.603587	200000000	1108561013
8094	1.136610	22000000	1106279658
4363	6.591277	250000000	1081041287
6555	4.205992	200000000	1065659812
1930	2.711136	200000000	1063171911
1921	5.572950	200000000	1025467110
3375	4.955130	380000000	1021683000
4367	4.218933	250000000	1017003568
2875	8.466668	185000000	1001921825
2634	8.021423	125000000	976475550
5434	3.928789	76000000	970761885
7387	4.965391	300000000	961000000
5431	4.310786	250000000	958400000
634	10.174599	250000000	955119788
1923	4.840588	250000000	954305868
7388	4.878162	150000000	938212738
1389	5.076472	250000000	933959197
...
4608	0.403624	0	0
4609	0.402947	0	0
4629	0.373557	0	0
4642	0.351001	0	0
4641	0.351494	0	0
4640	0.352564	12000000	0
4639	0.312624	0	0
4638	0.356594	0	0
4637	0.356988	0	0
4636	0.358343	0	0
4634	0.361254	3000000	0
4633	0.363004	0	0
4632	0.365206	0	0
4631	0.365381	30000000	0
4630	0.372274	0	0
4627	0.381610	3000000	0
4610	0.402405	0	0
4625	0.383809	0	0
4623	0.385121	0	0
4622	0.385202	0	0
4621	0.386979	3000000	0
4620	0.387190	1000000	0
4619	0.388099	0	0

4618	0.389170	3000000	0
4617	0.392485	12500000	0
4616	0.393660	5000000	0
4615	0.524197	0	0
4613	0.400215	0	0
4612	0.401012	2000000	0
10865	0.035919	19000	0

	original_title \
1386	Avatar
3	Star Wars: The Force Awakens
5231	Titanic
4361	The Avengers
0	Jurassic World
4	Furious 7
14	Avengers: Age of Ultron
3374	Harry Potter and the Deathly Hallows: Part 2
5422	Frozen
5425	Iron Man 3
8	Minions
3522	Transformers: Dark of the Moon
4949	The Lord of the Rings: The Return of the King
4365	Skyfall
8094	The Net
4363	The Dark Knight Rises
6555	Pirates of the Caribbean: Dead Man's Chest
1930	Toy Story 3
1921	Alice in Wonderland
3375	Pirates of the Caribbean: On Stranger Tides
4367	The Hobbit: An Unexpected Journey
2875	The Dark Knight
2634	Harry Potter and the Philosopher's Stone
5434	Despicable Me 2
7387	Pirates of the Caribbean: At World's End
5431	The Hobbit: The Desolation of Smaug
634	The Hobbit: The Battle of the Five Armies
1923	Harry Potter and the Deathly Hallows: Part 1
7388	Harry Potter and the Order of the Phoenix
1389	Harry Potter and the Half-Blood Prince
...	...
4608	I Am Bruce Lee
4609	Eden
4629	The Motel Life
4642	Hijacked
4641	Cockneys vs Zombies
4640	Dungeons & Dragons: The Book of Vile Darkness
4639	3, 2, 1... Frankie Go Boom
4638	Airborne

4637	Marley
4636	Extracted
4634	Dragon Eyes
4633	Battlestar Galactica: Blood & Chrome
4632	The Battery
4631	Bait
4630	Frenemies
4627	El Gringo
4610	What Richard Did
4625	The Forger
4623	Bad Ass
4622	The Babymakers
4621	Outpost: Black Sun
4620	Lovely Molly
4619	Not Suitable For Children
4618	The Thompsons
4617	Halo 4: Forward Unto Dawn
4616	Werewolf: The Beast Among Us
4615	Eating Out: The Open Weekend
4613	Promociãşn Fantasma
4612	Small Apartments
10865	Manos: The Hands of Fate

	director	runtime \
1386	James Cameron	162
3	J.J. Abrams	136
5231	James Cameron	194
4361	Joss Whedon	143
0	Colin Trevorrow	124
4	James Wan	137
14	Joss Whedon	141
3374	David Yates	130
5422	Chris Buck Jennifer Lee	102
5425	Shane Black	130
8	Kyle Balda Pierre Coffin	91
3522	Michael Bay	154
4949	Peter Jackson	201
4365	Sam Mendes	143
8094	Irwin Winkler	114
4363	Christopher Nolan	165
6555	Gore Verbinski	151
1930	Lee Unkrich	103
1921	Tim Burton	108
3375	Rob Marshall	136
4367	Peter Jackson	169
2875	Christopher Nolan	152
2634	Chris Columbus	152
5434	Pierre Coffin Chris Renaud	98

7387	Gore Verbinski	169
5431	Peter Jackson	161
634	Peter Jackson	144
1923	David Yates	146
7388	David Yates	138
1389	David Yates	153
...
4608	Pete McCormack	90
4609	Megan Griffiths	98
4629	Alan Polsky Gabe Polsky	95
4642	Brandon Nutt	90
4641	Matthias Hoene	88
4640	Gerry Lively	90
4639	Jordan Roberts	89
4638	Dominic Burns	78
4637	Kevin Macdonald	144
4636	Nir Paniry	85
4634	John Hyams	88
4633	Jonas Pate	91
4632	Jeremy Gardner	101
4631	Kimble Rendall	93
4630	Daisy von Scherler Mayer	86
4627	Eduardo Rodríguez	98
4610	Lenny Abrahamson	87
4625	Lawrence Roeck	96
4623	Craig Moss	90
4622	Jay Chandrasekhar	98
4621	Steve Barker	101
4620	Eduardo Sánchez	100
4619	Peter Templeman	97
4618	The Butcher Brothers	84
4617	Stewart Hendler	92
4616	Louis Morneau	93
4615	Q. Allan Brocka	80
4613	Javier Ruiz Caldera	88
4612	Jonas Åkerlund	96
10865	Harold P. Warren	74

	genres	release_date	vote_count \
1386	Action Adventure Fantasy Science Fiction	12/10/2009	8458
3	Action Adventure Science Fiction Fantasy	12/15/2015	5292
5231	Drama Romance Thriller	11/18/1997	4654
4361	Science Fiction Action Adventure	4/25/2012	8903
0	Action Adventure Science Fiction Thriller	6/9/2015	5562
4	Action Crime Thriller	4/1/2015	2947
14	Action Adventure Science Fiction	4/22/2015	4304
3374	Adventure Family Fantasy	7/7/2011	3750
5422	Animation Adventure Family	11/27/2013	3369

5425	Action Adventure Science Fiction	4/18/2013	6882
8	Family Animation Adventure Comedy	6/17/2015	2893
3522	Action Science Fiction Adventure	6/28/2011	2456
4949	Adventure Fantasy Action	12/1/2003	5636
4365	Action Adventure Thriller	10/25/2012	6137
8094	Crime Drama Mystery Thriller Action	7/28/1995	201
4363	Action Crime Drama Thriller	7/16/2012	6723
6555	Adventure Fantasy Action	6/20/2006	3181
1930	Animation Family Comedy	6/16/2010	2924
1921	Family Fantasy Adventure	3/3/2010	2853
3375	Adventure Action Fantasy	5/11/2011	3180
4367	Adventure Fantasy Action	11/26/2012	6417
2875	Drama Action Crime Thriller	7/16/2008	8432
2634	Adventure Fantasy Family	11/16/2001	4265
5434	Animation Comedy Family	6/25/2013	3436
7387	Adventure Fantasy Action	5/19/2007	2626
5431	Adventure Fantasy	12/11/2013	2952
634	Adventure Fantasy	12/10/2014	3110
1923	Adventure Fantasy Family	10/17/2010	3467
7388	Adventure Fantasy Family Mystery	6/28/2007	3382
1389	Adventure Fantasy Family	7/7/2009	3220
...
4608	Documentary	2/9/2012	14
4609	Drama	3/11/2012	32
4629	Drama Mystery Thriller	11/16/2012	12
4642	Action Thriller	7/31/2012	11
4641	Comedy Horror	8/31/2012	65
4640	Adventure Fantasy	8/8/2012	10
4639	Comedy	10/12/2012	13
4638	Horror Thriller	3/15/2012	13
4637	Documentary	4/20/2012	88
4636	Science Fiction Thriller	3/10/2012	40
4634	Drama Action Thriller Crime	5/10/2012	18
4633	Drama Science Fiction	11/9/2012	52
4632	Drama Horror	10/13/2012	43
4631	Horror Action Thriller	9/5/2012	99
4630	Drama Family TV Movie	1/13/2012	48
4627	Adventure Drama Action Thriller	5/10/2012	27
4610	Drama	9/9/2012	17
4625	Drama	7/3/2012	25
4623	Action Drama	4/12/2012	87
4622	Comedy	8/3/2012	39
4621	Action Horror	1/1/2012	12
4620	Horror Thriller	5/18/2012	38
4619	Comedy Romance	7/12/2012	35
4618	Horror Thriller	8/21/2012	13
4617	Science Fiction	10/5/2012	199
4616	Drama Horror	10/9/2012	32

4615		Comedy Romance	7/1/2012	13
4613		Comedy	2/3/2012	26
4612		Comedy Crime	3/10/2012	24
10865		Horror	11/15/1966	15

	vote_average	release_year	budget_adj	revenue_adj
1386	7.1	2009	2.408869e+08	2.827124e+09
3	7.5	2015	1.839999e+08	1.902723e+09
5231	7.3	1997	2.716921e+08	2.506406e+09
4361	7.3	2012	2.089437e+08	1.443191e+09
0	6.5	2015	1.379999e+08	1.392446e+09
4	7.3	2015	1.747999e+08	1.385749e+09
14	7.4	2015	2.575999e+08	1.292632e+09
3374	7.7	2011	1.211748e+08	1.287184e+09
5422	7.5	2013	1.404050e+08	1.192711e+09
5425	6.9	2013	1.872067e+08	1.137692e+09
8	6.5	2015	6.807997e+07	1.064192e+09
3522	6.1	2011	1.890326e+08	1.089358e+09
4949	7.9	2003	1.114231e+08	1.326278e+09
4365	6.8	2012	1.899489e+08	1.052849e+09
8094	5.6	1995	3.148127e+07	1.583050e+09
4363	7.5	2012	2.374361e+08	1.026713e+09
6555	6.8	2006	2.163338e+08	1.152691e+09
1930	7.5	2010	2.000000e+08	1.063172e+09
1921	6.3	2010	2.000000e+08	1.025467e+09
3375	6.3	2011	3.683713e+08	9.904175e+08
4367	6.9	2012	2.374361e+08	9.658933e+08
2875	8.1	2008	1.873655e+08	1.014733e+09
2634	7.2	2001	1.539360e+08	1.202518e+09
5434	7.0	2013	7.113853e+07	9.086655e+08
7387	6.8	2007	3.155006e+08	1.010654e+09
5431	7.6	2013	2.340083e+08	8.970944e+08
634	7.1	2014	2.302728e+08	8.797523e+08
1923	7.4	2010	2.500000e+08	9.543059e+08
7388	7.2	2007	1.577503e+08	9.866889e+08
1389	7.3	2009	2.541001e+08	9.492765e+08
...
4608	7.0	2012	0.000000e+00	0.000000e+00
4609	6.5	2012	0.000000e+00	0.000000e+00
4629	4.9	2012	0.000000e+00	0.000000e+00
4642	2.9	2012	0.000000e+00	0.000000e+00
4641	5.6	2012	0.000000e+00	0.000000e+00
4640	4.1	2012	1.139693e+07	0.000000e+00
4639	5.2	2012	0.000000e+00	0.000000e+00
4638	3.3	2012	0.000000e+00	0.000000e+00
4637	7.4	2012	0.000000e+00	0.000000e+00
4636	5.5	2012	0.000000e+00	0.000000e+00
4634	4.1	2012	2.849233e+06	0.000000e+00

4633	6.6	2012	0.000000e+00	0.000000e+00
4632	5.7	2012	0.000000e+00	0.000000e+00
4631	5.2	2012	2.849233e+07	0.000000e+00
4630	5.1	2012	0.000000e+00	0.000000e+00
4627	4.8	2012	2.849233e+06	0.000000e+00
4610	4.9	2012	0.000000e+00	0.000000e+00
4625	6.8	2012	0.000000e+00	0.000000e+00
4623	5.4	2012	0.000000e+00	0.000000e+00
4622	4.8	2012	0.000000e+00	0.000000e+00
4621	3.7	2012	2.849233e+06	0.000000e+00
4620	4.6	2012	9.497443e+05	0.000000e+00
4619	5.5	2012	0.000000e+00	0.000000e+00
4618	4.0	2012	2.849233e+06	0.000000e+00
4617	6.5	2012	1.187180e+07	0.000000e+00
4616	4.7	2012	4.748721e+06	0.000000e+00
4615	6.5	2012	0.000000e+00	0.000000e+00
4613	5.5	2012	0.000000e+00	0.000000e+00
4612	5.1	2012	1.899489e+06	0.000000e+00
10865	1.5	1966	1.276423e+05	0.000000e+00

[10800 rows x 13 columns]

```
In [17]: high_movie= movie_sorted.head(10)
         high_movie
```

```
Out[17]:
```

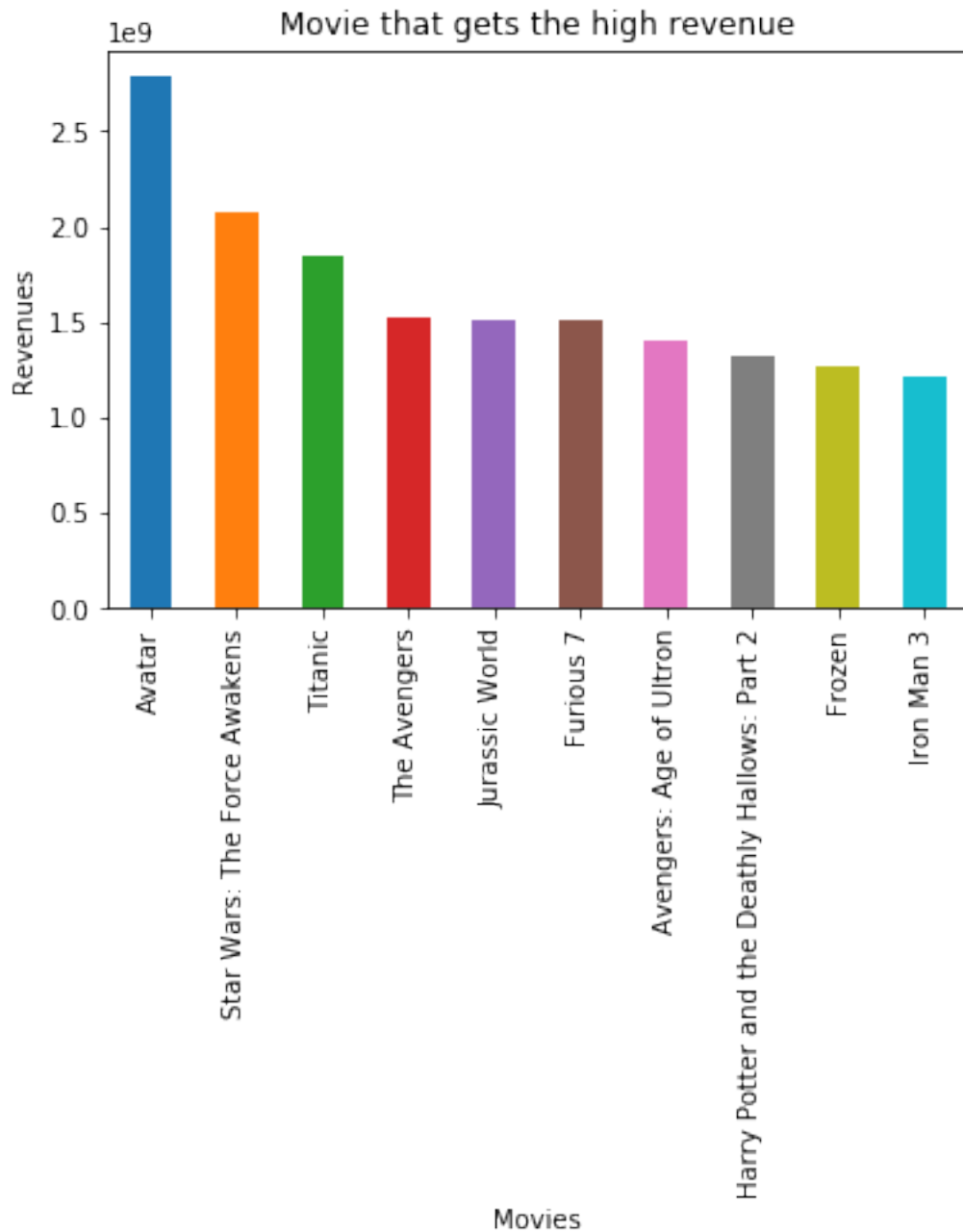
	popularity	budget	revenue \
1386	9.432768	237000000	2781505847
3	11.173104	200000000	2068178225
5231	4.355219	200000000	1845034188
4361	7.637767	220000000	1519557910
0	32.985763	150000000	1513528810
4	9.335014	190000000	1506249360
14	5.944927	280000000	1405035767
3374	5.711315	125000000	1327817822
5422	6.112766	150000000	1274219009
5425	4.946136	200000000	1215439994

	original_title	director \
1386	Avatar	James Cameron
3	Star Wars: The Force Awakens	J.J. Abrams
5231	Titanic	James Cameron
4361	The Avengers	Joss Whedon
0	Jurassic World	Colin Trevorrow
4	Furious 7	James Wan
14	Avengers: Age of Ultron	Joss Whedon
3374	Harry Potter and the Deathly Hallows: Part 2	David Yates
5422	Frozen	Chris Buck Jennifer Lee
5425	Iron Man 3	Shane Black

	runtime	genres	release_date	\
1386	162	Action Adventure Fantasy Science Fiction	12/10/2009	
3	136	Action Adventure Science Fiction Fantasy	12/15/2015	
5231	194	Drama Romance Thriller	11/18/1997	
4361	143	Science Fiction Action Adventure	4/25/2012	
0	124	Action Adventure Science Fiction Thriller	6/9/2015	
4	137	Action Crime Thriller	4/1/2015	
14	141	Action Adventure Science Fiction	4/22/2015	
3374	130	Adventure Family Fantasy	7/7/2011	
5422	102	Animation Adventure Family	11/27/2013	
5425	130	Action Adventure Science Fiction	4/18/2013	

	vote_count	vote_average	release_year	budget_adj	revenue_adj
1386	8458	7.1	2009	240886902.9	2.827124e+09
3	5292	7.5	2015	183999919.0	1.902723e+09
5231	4654	7.3	1997	271692064.2	2.506406e+09
4361	8903	7.3	2012	208943741.9	1.443191e+09
0	5562	6.5	2015	137999939.3	1.392446e+09
4	2947	7.3	2015	174799923.1	1.385749e+09
14	4304	7.4	2015	257599886.7	1.292632e+09
3374	3750	7.7	2011	121174755.3	1.287184e+09
5422	3369	7.5	2013	140405002.9	1.192711e+09
5425	6882	6.9	2013	187206670.5	1.137692e+09

```
In [18]: high_movie.plot.bar(x="original_title", y='revenue',legend=False)
plt.title('Movie that gets the high revenue')
plt.xlabel("Movies")
plt.ylabel("Revenues")
plt.show()
```



1.2.4 Result of First Question:

I was solved this question by sorted my revenue to get the highest revenue first then i got the top 10 of highest revenue, then i did a relation of movie title and the revenues of the top10 movies, the bar chart describe that the Avatar movie gets the highest revenue in all of the movies in this dataset.

1.2.5 Research Question 2 (Do high budget generate high revenue?)

```
In [19]: # Continue to explore the data to address your additional research
# questions. Add more headers as needed if you have more questions to
# investigate.
```

```
budget_sorted= movie_df.sort_values(by='budget', ascending=False)
budget_sorted
```

```
Out[19]:
```

	popularity	budget	revenue \
2244	0.250540	425000000	11087569
3375	4.955130	380000000	1021683000
7387	4.965391	300000000	961000000
14	5.944927	280000000	1405035767
6570	1.957331	270000000	391081192
1929	2.865684	260000000	591794936
4411	1.588457	260000000	284139100
7394	2.520912	258000000	890871626
5508	1.214510	255000000	89289910
634	10.174599	250000000	955119788
5431	4.310786	250000000	958400000
1923	4.840588	250000000	954305868
1389	5.076472	250000000	933959197
643	6.052479	250000000	746000000
4367	4.218933	250000000	1017003568
4363	6.591277	250000000	1081041287
10	6.200282	245000000	880674609
1386	9.432768	237000000	2781505847
5432	3.972460	225000000	662845518
2902	1.628345	225000000	419651413
4381	2.592896	225000000	624026776
4361	7.637767	220000000	1519557910
4370	3.702647	215000000	752215857
6883	0.157586	210000000	234360014
648	4.983782	210000000	245500000
4405	1.630455	209000000	303025485
6215	1.508329	207000000	550000000
5444	2.643496	200000000	743559607
5520	1.094958	200000000	491868548
1930	2.711136	200000000	1063171911
...
3292	0.088877	0	0
7040	0.893922	0	0
3293	0.130306	0	0
3294	0.129800	0	0
3297	0.133564	0	0
3298	0.152615	0	0
3299	0.120405	0	0
3277	0.258333	0	0
7056	0.763058	0	0

3254	0.173639	0	0
7067	0.681008	0	0
3255	0.172834	0	0
3258	0.168750	0	0
3260	0.161881	0	0
3261	0.161594	0	0
3262	0.160986	0	0
3263	0.160950	0	0
7069	0.679386	0	0
3264	0.160382	0	11295
3265	0.158418	0	850575
3274	0.152386	0	0
3267	0.201483	0	0
7064	0.712887	0	0
3269	0.199865	0	0
7062	0.716760	0	0
7061	0.722796	0	0
3270	0.156088	0	0
3271	0.155760	0	0
3273	0.154248	0	0
140	1.144808	0	0

	original_title \
2244	The Warrior's Way
3375	Pirates of the Caribbean: On Stranger Tides
7387	Pirates of the Caribbean: At World's End
14	Avengers: Age of Ultron
6570	Superman Returns
1929	Tangled
4411	John Carter
7394	Spider-Man 3
5508	The Lone Ranger
634	The Hobbit: The Battle of the Five Armies
5431	The Hobbit: The Desolation of Smaug
1923	Harry Potter and the Deathly Hallows: Part 1
1389	Harry Potter and the Half-Blood Prince
643	X-Men: Days of Future Past
4367	The Hobbit: An Unexpected Journey
4363	The Dark Knight Rises
10	Spectre
1386	Avatar
5432	Man of Steel
2902	The Chronicles of Narnia: Prince Caspian
4381	Men in Black 3
4361	The Avengers
4370	The Amazing Spider-Man
6883	X-Men: The Last Stand
648	Transformers: Age of Extinction

4405	Battleship
6215	King Kong
5444	Monsters University
5520	Oz: The Great and Powerful
1930	Toy Story 3
...	...
3292	Signs
7040	Sniper 3
3293	One-Eyed Monster
3294	Food Matters
3297	Witless Protection
3298	Mamãe
3299	Mulligans
3277	La Journée de la jupe
7056	Mickey, Donald, Goofy: The Three Musketeers
3254	Front of the Class
7067	Saved!
3255	Hero Wanted
3258	Bouquet final
3260	Keith
3261	Hamlet 2
3262	Get Smart's Bruce and Lloyd Out of Control
3263	Seventh Moon
7069	Five Children and It
3264	Otto; or, Up with Dead People
3265	Faubourg 36
3274	Red Mist
3267	Donkey Punch
7064	(T)Raumschiff Surprise - Periode 1
3269	Gardens of the Night
7062	New York Minute
7061	Mickey's Twice Upon a Christmas
3270	Global Metal
3271	Bad Biology
3273	Where in the World Is Osama Bin Laden?
140	Dragonheart 3: The Sorcerer's Curse

	director	runtime	\
2244	Sngmoo Lee	100	
3375	Rob Marshall	136	
7387	Gore Verbinski	169	
14	Joss Whedon	141	
6570	Bryan Singer	154	
1929	Nathan Greno Byron Howard	100	
4411	Andrew Stanton	132	
7394	Sam Raimi	139	
5508	Gore Verbinski	149	
634	Peter Jackson	144	

5431	Peter Jackson	161
1923	David Yates	146
1389	David Yates	153
643	Bryan Singer	131
4367	Peter Jackson	169
4363	Christopher Nolan	165
10	Sam Mendes	148
1386	James Cameron	162
5432	Zack Snyder	143
2902	Andrew Adamson	150
4381	Barry Sonnenfeld	106
4361	Joss Whedon	143
4370	Marc Webb	136
6883	Brett Ratner	104
648	Michael Bay	165
4405	Peter Berg	131
6215	Peter Jackson	187
5444	Dan Scanlon	104
5520	Sam Raimi	130
1930	Lee Unkrich	103
...
3292	Patrick Hughes	12
7040	P.J. Pesce	97
3293	Adam Fields	84
3294	James Colquhoun	80
3297	Charles Robert Carner	97
3298	Andy Muschietti	3
3299	Chip Hale	92
3277	Jean-Paul Lilienfeld	88
7056	Donovan Cook	68
3254	Peter Werner	95
7067	Brian Dannelly	92
3255	Brian Smrz	95
3258	Michel Delgado	101
3260	Todd Kessler	93
3261	Andrew Fleming	92
3262	Gil Junger	71
3263	Eduardo S��nchez	87
7069	John Stephenson	88
3264	Bruce LaBruce	94
3265	Christophe Barratier	120
3274	Paddy Breathnach	98
3267	Olly Blackburn	99
7064	Michael Herbig	87
3269	Damian Harris	110
7062	Dennie Gordon	91
7061	Carole Holliday Matthew O'Callaghan Theresa Cu...	68
3270	Sam Dunn	93

3271	Frank Henenlotter	84
3273	Morgan Spurlock	93
140	Colin Teague	97

	genres	release_date	vote_count \
2244	Adventure Fantasy Action Western Thriller	12/2/2010	74
3375	Adventure Action Fantasy	5/11/2011	3180
7387	Adventure Fantasy Action	5/19/2007	2626
14	Action Adventure Science Fiction	4/22/2015	4304
6570	Adventure Fantasy Action Science Fiction	6/28/2006	834
1929	Animation Family	11/24/2010	2019
4411	Action Adventure Fantasy Science Fiction	3/7/2012	1479
7394	Fantasy Action Adventure	5/1/2007	2083
5508	Action Adventure Western	7/3/2013	1607
634	Adventure Fantasy	12/10/2014	3110
5431	Adventure Fantasy	12/11/2013	2952
1923	Adventure Fantasy Family	10/17/2010	3467
1389	Adventure Fantasy Family	7/7/2009	3220
643	Action Adventure Fantasy Science Fiction	5/15/2014	4006
4367	Adventure Fantasy Action	11/26/2012	6417
4363	Action Crime Drama Thriller	7/16/2012	6723
10	Action Adventure Crime	10/26/2015	3254
1386	Action Adventure Fantasy Science Fiction	12/10/2009	8458
5432	Action Adventure Fantasy Science Fiction	6/12/2013	4729
2902	Adventure Family Fantasy	5/15/2008	891
4381	Action Comedy Science Fiction	5/23/2012	3142
4361	Science Fiction Action Adventure	4/25/2012	8903
4370	Action Adventure Fantasy	6/27/2012	4837
6883	Adventure Action Science Fiction Thriller	5/24/2006	2381
648	Science Fiction Action Adventure	6/25/2014	2161
4405	Thriller Action Adventure Science Fiction	4/11/2012	1523
6215	Adventure Drama Action	12/14/2005	1289
5444	Animation Family	6/20/2013	2251
5520	Fantasy Adventure Family	3/7/2013	2855
1930	Animation Family Comedy	6/16/2010	2924
...
3292	Romance	3/10/2008	12
7040	Drama Action Thriller	10/18/2004	15
3293	Comedy Horror	11/11/2008	11
3294	Documentary	1/1/2008	16
3297	Action Adventure Comedy Crime	2/22/2008	12
3298	Horror	10/7/2008	11
3299	Drama	1/1/2008	18
3277	Drama	11/9/2008	25
7056	Adventure Animation Comedy Family	8/3/2004	98
3254	Drama	1/1/2008	22
7067	Comedy Drama Family	1/21/2004	71
3255	Action Adventure Crime Drama Thriller	3/28/2008	15

3258		Comedy	11/5/2008	12
3260	Drama Family Romance		9/13/2008	90
3261	Comedy Music		1/21/2008	24
3262	Action Comedy		7/1/2008	17
3263	Horror		9/20/2008	10
7069	Adventure Fantasy Drama Science Fiction Family		10/15/2004	35
3264	Drama Horror Comedy Thriller Foreign		1/19/2008	10
3265	Drama		9/24/2008	14
3274	Foreign Horror Thriller		8/22/2008	15
3267	Horror Drama Thriller Crime		7/18/2008	19
7064	Comedy Science Fiction		7/14/2004	80
3269	Crime Drama		11/21/2008	12
7062	Comedy		5/1/2004	97
7061	Animation Family		1/1/2004	36
3270	Documentary Music		6/20/2008	13
3271	Fantasy Horror		2/8/2008	20
3273	Documentary		1/21/2008	10
140	Action Adventure Fantasy		2/24/2015	59

	vote_average	release_year	budget_adj	revenue_adj
2244	6.4	2010	425000000.0	1.108757e+07
3375	6.3	2011	368371256.2	9.904175e+08
7387	6.8	2007	315500574.8	1.010654e+09
14	7.4	2015	257599886.7	1.292632e+09
6570	5.3	2006	292050672.7	4.230205e+08
1929	7.3	2010	260000000.0	5.917949e+08
4411	6.0	2012	246933513.2	2.698595e+08
7394	5.8	2007	271330494.3	9.369017e+08
5508	6.0	2013	238688504.9	8.357833e+07
634	7.1	2014	230272762.7	8.797523e+08
5431	7.6	2013	234008338.2	8.970944e+08
1923	7.4	2010	250000000.0	9.543059e+08
1389	7.3	2009	254100108.5	9.492765e+08
643	7.6	2014	230272762.7	6.871339e+08
4367	6.9	2012	237436070.3	9.658933e+08
4363	7.5	2012	237436070.3	1.026713e+09
10	6.2	2015	225399900.8	8.102203e+08
1386	7.1	2009	240886902.9	2.827124e+09
5432	6.5	2013	210607504.4	6.204455e+08
2902	6.1	2008	227876992.6	4.250173e+08
4381	6.1	2012	213692463.3	5.926659e+08
4361	7.3	2012	208943741.9	1.443191e+09
4370	6.4	2012	204195020.5	7.144127e+08
6883	6.2	2006	227150523.2	2.535000e+08
648	5.9	2014	193429120.7	2.261279e+08
4405	5.5	2012	198496554.8	2.877967e+08
6215	6.4	2005	231128601.0	6.141098e+08
5444	7.0	2013	187206670.5	6.959966e+08

5520	5.6	2013	187206670.5	4.604054e+08
1930	7.5	2010	200000000.0	1.063172e+09
...
3292	7.5	2008	0.0	0.000000e+00
7040	5.3	2004	0.0	0.000000e+00
3293	5.1	2008	0.0	0.000000e+00
3294	7.0	2008	0.0	0.000000e+00
3297	4.4	2008	0.0	0.000000e+00
3298	6.0	2008	0.0	0.000000e+00
3299	5.6	2008	0.0	0.000000e+00
3277	7.2	2008	0.0	0.000000e+00
7056	6.3	2004	0.0	0.000000e+00
3254	7.8	2008	0.0	0.000000e+00
7067	6.0	2004	0.0	0.000000e+00
3255	5.2	2008	0.0	0.000000e+00
3258	5.4	2008	0.0	0.000000e+00
3260	7.1	2008	0.0	0.000000e+00
3261	5.8	2008	0.0	0.000000e+00
3262	5.4	2008	0.0	0.000000e+00
3263	4.1	2008	0.0	0.000000e+00
7069	5.4	2004	0.0	0.000000e+00
3264	5.0	2008	0.0	1.143943e+04
3265	5.3	2008	0.0	8.614510e+05
3274	4.8	2008	0.0	0.000000e+00
3267	4.0	2008	0.0	0.000000e+00
7064	6.3	2004	0.0	0.000000e+00
3269	6.4	2008	0.0	0.000000e+00
7062	5.4	2004	0.0	0.000000e+00
7061	7.2	2004	0.0	0.000000e+00
3270	7.1	2008	0.0	0.000000e+00
3271	4.9	2008	0.0	0.000000e+00
3273	5.9	2008	0.0	0.000000e+00
140	4.5	2015	0.0	0.000000e+00

[10800 rows x 13 columns]

```
In [20]: high_budget= budget_sorted.head(10)
         high_budget
```

```
Out[20]:
```

	popularity	budget	revenue	\
2244	0.250540	425000000	11087569	
3375	4.955130	380000000	1021683000	
7387	4.965391	300000000	961000000	
14	5.944927	280000000	1405035767	
6570	1.957331	270000000	391081192	
1929	2.865684	260000000	591794936	
4411	1.588457	260000000	284139100	
7394	2.520912	258000000	890871626	

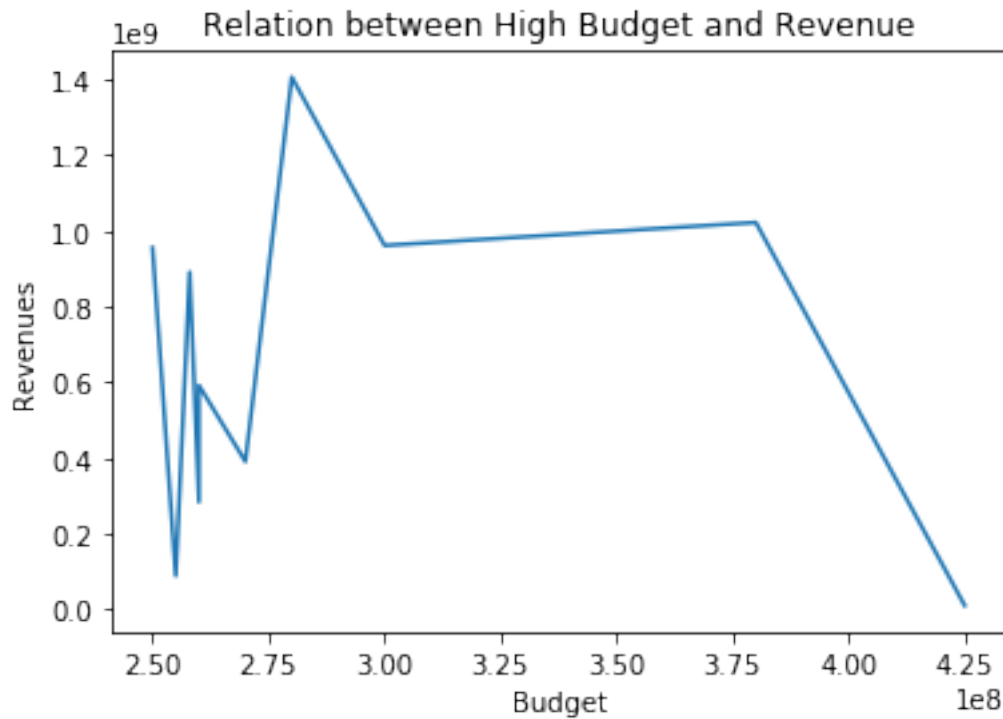
5508	1.214510	255000000	89289910
634	10.174599	250000000	955119788

	original_title	director \
2244	The Warrior's Way	Sngmoo Lee
3375	Pirates of the Caribbean: On Stranger Tides	Rob Marshall
7387	Pirates of the Caribbean: At World's End	Gore Verbinski
14	Avengers: Age of Ultron	Joss Whedon
6570	Superman Returns	Bryan Singer
1929	Tangled	Nathan Greno Byron Howard
4411	John Carter	Andrew Stanton
7394	Spider-Man 3	Sam Raimi
5508	The Lone Ranger	Gore Verbinski
634	The Hobbit: The Battle of the Five Armies	Peter Jackson

	runtime	genres	release_date \
2244	100	Adventure Fantasy Action Western Thriller	12/2/2010
3375	136	Adventure Action Fantasy	5/11/2011
7387	169	Adventure Fantasy Action	5/19/2007
14	141	Action Adventure Science Fiction	4/22/2015
6570	154	Adventure Fantasy Action Science Fiction	6/28/2006
1929	100	Animation Family	11/24/2010
4411	132	Action Adventure Fantasy Science Fiction	3/7/2012
7394	139	Fantasy Action Adventure	5/1/2007
5508	149	Action Adventure Western	7/3/2013
634	144	Adventure Fantasy	12/10/2014

	vote_count	vote_average	release_year	budget_adj	revenue_adj
2244	74	6.4	2010	425000000.0	1.108757e+07
3375	3180	6.3	2011	368371256.2	9.904175e+08
7387	2626	6.8	2007	315500574.8	1.010654e+09
14	4304	7.4	2015	257599886.7	1.292632e+09
6570	834	5.3	2006	292050672.7	4.230205e+08
1929	2019	7.3	2010	260000000.0	5.917949e+08
4411	1479	6.0	2012	246933513.2	2.698595e+08
7394	2083	5.8	2007	271330494.3	9.369017e+08
5508	1607	6.0	2013	238688504.9	8.357833e+07
634	3110	7.1	2014	230272762.7	8.797523e+08

```
In [21]: high_budget.plot.line(x="budget", y='revenue', legend=False)
plt.title('Relation between High Budget and Revenue')
plt.xlabel("Budget")
plt.ylabel("Revenues")
plt.show()
```

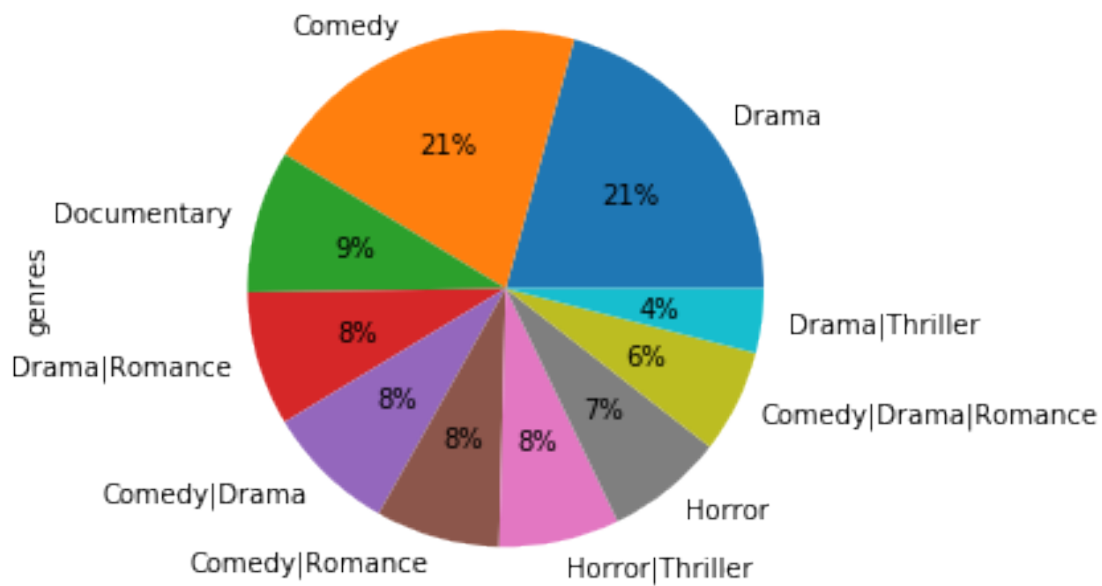


1.3 Result of Second question:

I was solved this question also by sorted my budget to get the highest budget first then i got the top 10 of highest budget, then i did a relation of revenues and the budget of the top10 movies, the line chart describe that there is no relation of that when i have high budget that i will got high revenue! as we can see here in chart the highest budget is 4.25 but this budget didn't get the high revenue. So from this chart we learn there is no relation between them!

1.3.1 Research Question3 (What is the most repeated type of movies?)

```
In [22]: movie_df["genres"].value_counts().head(10).plot(kind="pie", autopct='%1.0f%%');
plt.axis("equal");
```



1.4 Result of Thrid question:

As we know there is a difference between the types of movie such as when we say Drama type and Drama,Comedy type there is a big difference between them!

So from this pie chart we can see the percentage of every type in our dataset, and we notice the Comedy and Drama get the same and high percentage which mean they are the most in our dataset.

Conclusions

Finally, this was a way to clean my dataset and answer my questions!

We learned from this project that there is a difference in revenue and it's impossible to get all the movies the same revenues, Also the most important thing we learned that there is no relation between the high budget and high revenue of the movies as we see in our line chart above, and we learned in each dataset there is perentage of what is the most type that in our dataset as we see in our project that the drama and comedy is the highest types of movies in our dataset. ## Limitations

The limitations is how to make the questions that you can make sure that you will get the answer in your dataset, delete the rows with the missing data this would impact the accuracy of the analysis, Also when you drop the coulumn that you think you will not use it but when you start analysing the data you notice that you shold have these coulumn and you must rearrange the data.

```
In [23]: from subprocess import call
         call(['python', '-m', 'nbconvert', 'Investigate_a_Dataset.ipynb'])
```

```
Out[23]: 0
```

```
In [ ]:
```