Untitled

August 9, 2022

[91]:
```
### The project is called the TMDB-Movie Analysis
```

1 Table of Content

1. Introduction
2. Data Wrangling
3. Exploratory Data Analysis
4. Conclusions

2 Introduction

[92]:
```
# This data set contains information about 10,000 movies collected from The␣
↪Movie Database (TMDb), including user ratings and revenue.
```

You can get more information here

3 Questions to answer are

- Which movie title had the highest budget?
- Which movie titles has the highest revenue?
- Which genres are most produced throughout time?

[93]:
```
# Importing important and necessary packages
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

[94]:
```
import matplotlib.pyplot as plt
```

[95]:
```
%matplotlib inline
```

[96]:
```
#Importing the dataset
df_movies= pd.read_csv('Desktop/ALX Project/tmdb-movies.csv')
```

[97]:
```
#Checking for first 10 data
df_movies.head(10)
```

[97]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi…	http://www.jurassicworld.com/
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic…	http://www.madmaxmovie.com/
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley|Theo James|Kate Winslet|Ansel…	http://www.thedivergentseries.movie/#insurgent
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D…	http://www.starwars.com/films/star-wars-episod…
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel|Paul Walker|Jason Statham|Michelle …	http://www.furious7.com/
5	281957	tt1663202	9.110700	135000000	532950503	The Revenant	Leonardo DiCaprio|Tom Hardy|Will Poulter|Domhn…	http://www.foxmovies.com/movies/the-revenant
6	87101	tt1340138	8.654359	155000000	440603537	Terminator Genisys	Arnold Schwarzenegger|Jason Clarke|Emilia Clar…	http://www.terminatormovie.com/
7	286217	tt3659388	7.667400	108000000	595380321	The Martian	Matt Damon|Jessica Chastain|Kristen Wiig|Jeff …	
8	211672	tt2293640	7.404165	74000000	1156730962	Minions	Sandra Bullock|Jon Hamm|Michael Keaton|Allison…	
9	150540	tt2096673	6.326804	175000000	853708609	Inside Out	Amy Poehler|Phyllis Smith|Richard Kind|Bill Ha…	

```
7       http://www.foxmovies.com/movies/the-martian
8                       http://www.minionsmovie.com/
9                http://movies.disney.com/inside-out

                        director  \
0                 Colin Trevorrow
1                   George Miller
2                Robert Schwentke
3                     J.J. Abrams
4                       James Wan
5  Alejandro GonzÃ¡lez IÃ±Ã¡rritu
6                      Alan Taylor
7                     Ridley Scott
8         Kyle Balda|Pierre Coffin
9                      Pete Docter

                                        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.  …

                                           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

                                      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

                                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
5  Regency Enterprises|Appian Way|CatchPlay|Anony…     12/25/15       3929
6            Paramount Pictures|Skydance Productions      6/23/15       2598
7  Twentieth Century Fox Film Corporation|Scott F…      9/30/15       4572
8      Universal Pictures|Illumination Entertainment      6/17/15       2893
9  Walt Disney Pictures|Pixar Animation Studios|W…       6/9/15       3935

   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           7.2          2015  1.241999e+08  4.903142e+08
6           5.8          2015  1.425999e+08  4.053551e+08
7           7.6          2015  9.935996e+07  5.477497e+08
8           6.5          2015  6.807997e+07  1.064192e+09
9           8.0          2015  1.609999e+08  7.854116e+08

[10 rows x 21 columns]
```

[98]:
```
df_movies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10866 entries, 0 to 10865
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    10866 non-null  int64
 1   imdb_id               10856 non-null  object
 2   popularity            10866 non-null  float64
 3   budget                10866 non-null  int64
 4   revenue               10866 non-null  int64
 5   original_title        10866 non-null  object
 6   cast                  10790 non-null  object
 7   homepage              2936 non-null   object
 8   director              10822 non-null  object
 9   tagline               8042 non-null   object
```

```
 10  keywords              9373 non-null   object
 11  overview              10862 non-null  object
 12  runtime               10866 non-null  int64
 13  genres                10843 non-null  object
 14  production_companies  9836 non-null   object
 15  release_date          10866 non-null  object
 16  vote_count            10866 non-null  int64
 17  vote_average          10866 non-null  float64
 18  release_year          10866 non-null  int64
 19  budget_adj            10866 non-null  float64
 20  revenue_adj           10866 non-null  float64
dtypes: float64(4), int64(6), object(11)
memory usage: 1.7+ MB
```

[99]:
```
df_movies.tail(10)
```

[99]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage
10856	20277	tt0061135	0.140934	0	0	The Ugly Dachshund	Dean Jones|Suzanne Pleshette|Charles Ruggles|K…	NaN
10857	5921	tt0060748	0.131378	0	0	Nevada Smith	Steve McQueen|Karl Malden|Brian Keith|Arthur K…	NaN
10858	31918	tt0060921	0.317824	0	0	The Russians Are Coming, The Russians Are Coming	Carl Reiner|Eva Marie Saint|Alan Arkin|Brian K…	NaN
10859	20620	tt0060955	0.089072	0	0	Seconds	Rock Hudson|Salome Jens|John Randolph|Will Gee…	NaN
10860	5060	tt0060214	0.087034	0	0	Carry On Screaming!	Kenneth Williams|Jim Dale|Harry H. Corbett|Joa…	NaN
10861	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson|Robert August|Lord 'Tally Ho' B…	NaN
10862	20379	tt0060472	0.065543	0	0	Grand Prix		
10863	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya		
10864	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?		
10865	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate		

```
10862  James Garner|Eva Marie Saint|Yves Montand|Tosh…      NaN
10863  Innokentiy Smoktunovskiy|Oleg Efremov|Georgi Z…      NaN
10864  Tatsuya Mihashi|Akiko Wakabayashi|Mie Hama|Joh…      NaN
10865  Harold P. Warren|Tom Neyman|John Reynolds|Dian…      NaN

                director                                            tagline  \
10856        Norman Tokar  A HAPPY HONEYMOON GOES TO THE DOGS!…When a G…
10857      Henry Hathaway  Some called him savage- and some called him sa…
10858      Norman Jewison   IT'S A PLOT! …to make the world die laughing!!
10859  John Frankenheimer                                               NaN
10860       Gerald Thomas  Carry On Screaming with the Hilarious CARRY ON…
10861         Bruce Brown                                               NaN
10862  John Frankenheimer  Cinerama sweeps YOU into a drama of speed and …
10863      Eldar Ryazanov                                               NaN
10864         Woody Allen                          WOODY ALLEN STRIKES BACK!
10865    Harold P. Warren       It's Shocking! It's Beyond Your Imagination!

       …                                                overview runtime  \
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  \
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  \
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
10856         14           5.7          1966       0.000000          0.0
10857         10           5.9          1966       0.000000          0.0
10858         11           5.5          1966       0.000000          0.0
10859         22           6.6          1966       0.000000          0.0
10860         13           7.0          1966       0.000000          0.0
10861         11           7.4          1966       0.000000          0.0
10862         20           5.7          1966       0.000000          0.0
10863         11           6.5          1966       0.000000          0.0
10864         22           5.4          1966       0.000000          0.0
10865         15           1.5          1966  127642.279154          0.0

[10 rows x 21 columns]
```

[100]:
```
#Checking for sum of null values
df_movies.isnull().sum()
```

[100]:
```
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
```

[101]:
```
#Checking for Duplicates
df_movies.duplicated().sum()
```

[101]:
```
1
```

df_movies.describe()

[102]:
```
#Checking the number of rows and Columns of the imported data
df_movies.shape
```

[102]:
```
(10866, 21)
```

#Data Cleaning

[103]:
```
# Gathered Information
# 1. Hompage has too much null information and has to be dropped
# 2. Columns like tagline, keywords, production_companies, drectors will have␣
 ↪to be dropped
# 3. The data has one duplicated item and will be dropped.
```

[104]:
```
df_movies.shape
```

[104]:
```
(10866, 21)
```

[105]:
```
df_movies.drop_duplicates(inplace = True)
```

[106]:
```
# Checking if the duplicate has been dropped
df_movies.shape
```

[106]:
```
(10865, 21)
```

[107]:
```
DroppedColumns = ['homepage', 'tagline', 'overview' ,'cast','director',
       'keywords','overview','production_companies','release_date']
```

[108]:
```
df_movies.drop(DroppedColumns, axis=1, inplace=True)
```

[109]:
```
df_movies.head()
```

[109]:

	id	imdb_id	popularity	budget	revenue	original_title	runtime
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	124
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	120
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	119
3	140607	tt2488496	11.173104	200000000	2068178225		
4	168259	tt2820852	9.335014	190000000	1506249360		

```
3  Star Wars: The Force Awakens      136
4                     Furious 7      137

                                      genres  vote_count  vote_average  \
0  Action|Adventure|Science Fiction|Thriller        5562           6.5
1  Action|Adventure|Science Fiction|Thriller        6185           7.1
2         Adventure|Science Fiction|Thriller        2480           6.3
3   Action|Adventure|Science Fiction|Fantasy        5292           7.5
4                     Action|Crime|Thriller        2947           7.3

   release_year    budget_adj   revenue_adj
0          2015  1.379999e+08  1.392446e+09
1          2015  1.379999e+08  3.481613e+08
2          2015  1.012000e+08  2.716190e+08
3          2015  1.839999e+08  1.902723e+09
4          2015  1.747999e+08  1.385749e+09
```

df_movies.info()

[111]:
```
#Checking for Null Value
df_movies.isnull().sum()
```

[111]:
```
id                0
imdb_id          10
popularity        0
budget            0
revenue           0
original_title    0
runtime           0
genres           23
vote_count        0
vote_average      0
release_year      0
budget_adj        0
revenue_adj       0
dtype: int64
```

[112]:
```
# Checking the genre in which it is null
df_movies[df_movies.genres.isnull()]
```

[112]:
```
           id    imdb_id  popularity  budget  revenue  \
424    363869  tt4835298    0.244648       0        0
620    361043  tt5022680    0.129696       0        0
997    287663         NaN    0.330431       0        0
1712    21634  tt1073510    0.302095       0        0
1897    40534  tt1229827    0.020701       0        0
2370   127717  tt1525359    0.081892       0        0
2376   315620  tt1672218    0.068411       0        0
```

```
       id     imdb_id  popularity  budget  revenue
2853    57892  tt0270053    0.130018       0        0
3279    54330  tt1720044    0.145331       0        0
4547   123024  tt2305700    0.520520       0        0
4732   139463  tt2084977    0.235911       0        0
4797   369145        NaN    0.167501       0        0
4890   126909  tt2219564    0.083202       0        0
5830   282848  tt2986512    0.248944       0        0
5934   200204  tt2808968    0.067433       0        0
6043   190940  tt2797242    0.039080       0        0
6530   168891  tt0818519    0.092724       0        0
8234    56804  tt0114844    0.028874       0        0
8614    65595  tt0117880    0.273934       0        0
8878    92208  tt0250593    0.038045       0        0
9307   141859  tt0097446    0.094652       0        0
9799    48847  tt0193716    0.175008       0        0
10659    4255  tt0065904    0.344172    5000        0

                                          original_title  runtime genres  \
424                                        Belli di papÃ       100    NaN
620                                  All Hallows' Eve 2        90    NaN
997                 Star Wars Rebels: Spark of Rebellion        44    NaN
1712                                 Prayers for Bobby        88    NaN
1897             Jonas Brothers: The Concert Experience        76    NaN
2370                                   Freshman Father         0    NaN
2376                     Doctor Who: A Christmas Carol        62    NaN
2853                                         Vizontele       110    NaN
3279                                      ì•„ê¸°ì € ë¸˜        96    NaN
4547   London 2012 Olympic Opening Ceremony: Isles of…       220    NaN
4732                                      The Scapegoat       100    NaN
4797                              Doctor Who: The Snowmen        60    NaN
4890                            Cousin Ben Troop Screening         2    NaN
5830                 Doctor Who: The Time of the Doctor        60    NaN
5934                                       Prada: Candy         3    NaN
6043                                    Bombay Talkies       127    NaN
6530                                       Saw Rebirth         6    NaN
8234                                    Viaggi di nozze       103    NaN
8614                           T2 3-D: Battle Across Time        12    NaN
8878                    Mom's Got a Date With a Vampire         85    NaN
9307                                         Goldeneye       105    NaN
9799                                       The Amputee         5    NaN
10659                     The Party at Kitty and Stud's        71    NaN

       vote_count  vote_average  release_year   budget_adj  revenue_adj
424            21           6.1          2015      0.00000          0.0
620            13           5.0          2015      0.00000          0.0
997            13           6.8          2014      0.00000          0.0
1712           57           7.4          2009      0.00000          0.0
```

1897	11	7.0	2009	0.00000	0.0
2370	12	5.8	2010	0.00000	0.0
2376	11	7.7	2010	0.00000	0.0
2853	12	7.2	2001	0.00000	0.0
3279	11	6.1	2008	0.00000	0.0
4547	12	8.3	2012	0.00000	0.0
4732	12	6.2	2012	0.00000	0.0
4797	10	7.8	2012	0.00000	0.0
4890	14	7.0	2012	0.00000	0.0
5830	26	8.5	2013	0.00000	0.0
5934	27	6.9	2013	0.00000	0.0
6043	12	5.9	2013	0.00000	0.0
6530	24	5.9	2005	0.00000	0.0
8234	44	6.7	1995	0.00000	0.0
8614	14	6.7	1996	0.00000	0.0
8878	16	5.4	2000	0.00000	0.0
9307	10	5.3	1989	0.00000	0.0
9799	11	5.0	1974	0.00000	0.0
10659	10	3.0	1970	28081.84172	0.0

[113]:
```
df_movies.dropna(inplace = True)
```

4 1 Which movie title has the highest budget ?

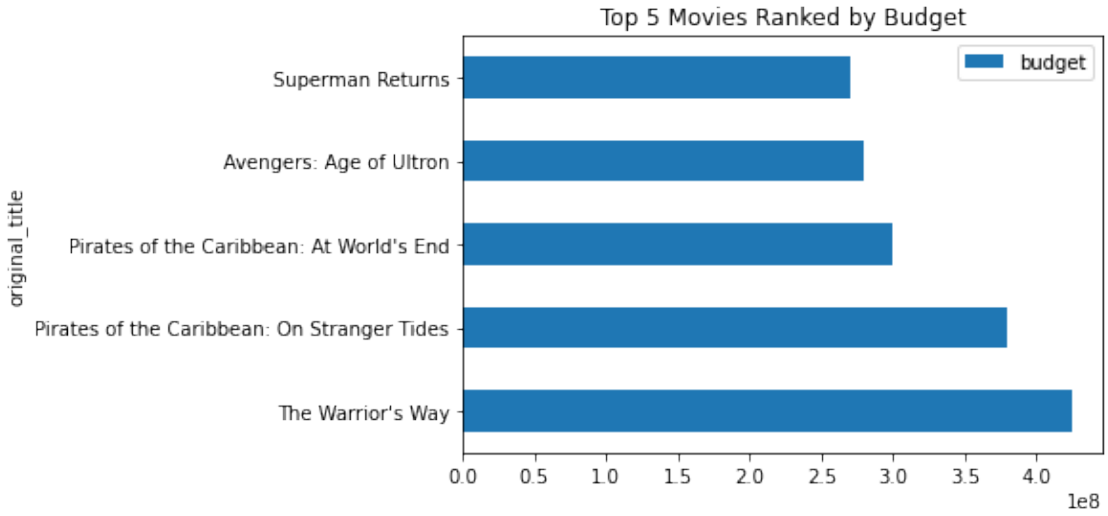
[117]:
```
Top5Movies= df_movies[['original_title', 'budget']].sort_values(by='budget',␣
↪ascending=False).head()
```

[118]:
```
Top5Movies
```

[118]:

	original_title	budget
2244	The Warrior's Way	425000000
3375	Pirates of the Caribbean: On Stranger Tides	380000000
7387	Pirates of the Caribbean: At World's End	300000000
14	Avengers: Age of Ultron	280000000
6570	Superman Returns	270000000

[125]:
```
Top5Movies.set_index('original_title').plot(kind ='barh')
plt.title('Top 5 Movies Ranked by Budget');
```



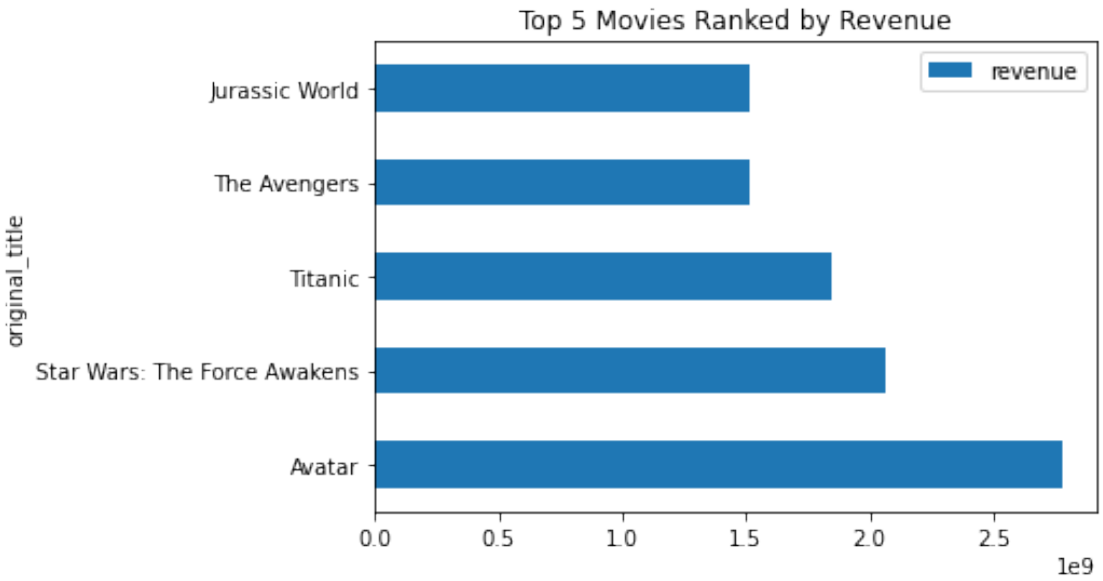
5 2 Which movie titles has the highest revenue?

[138]:
```
Top5MoviesRevenue= df_movies[['original_title', 'revenue']].
↪sort_values(by='revenue', ascending=False).head()
```

[139]:
```
Top5MoviesRevenue
```

[139]:
```
                    original_title     revenue
1386                        Avatar  2781505847
3     Star Wars: The Force Awakens  2068178225
5231                       Titanic  1845034188
4361                   The Avengers  1519557910
0                  Jurassic World  1513528810
```

[140]:
```
Top5MoviesRevenue.set_index('original_title').plot(kind ='barh')
plt.title('Top 5 Movies Ranked by Revenue');
```



6 3 Which genres are most produced throughout time?

[ ]: