



TMDB Movies project Report

By Ahmed Abdelhamed Elgayar

The Udacity Data Advanced Track Project 1 : this is a report of the main steps for data wrangling and exploratory data analysis for “TMDB Movies”.

A- Data Gathering:

- 1- Download the csv file and upload it in project workspace .
- 2- load the csv file and store it in the variable "df_movie"

B- Data Assessment & Cleaning:

Here i investigated the imported datasets both visually and programmatically for issues in data .

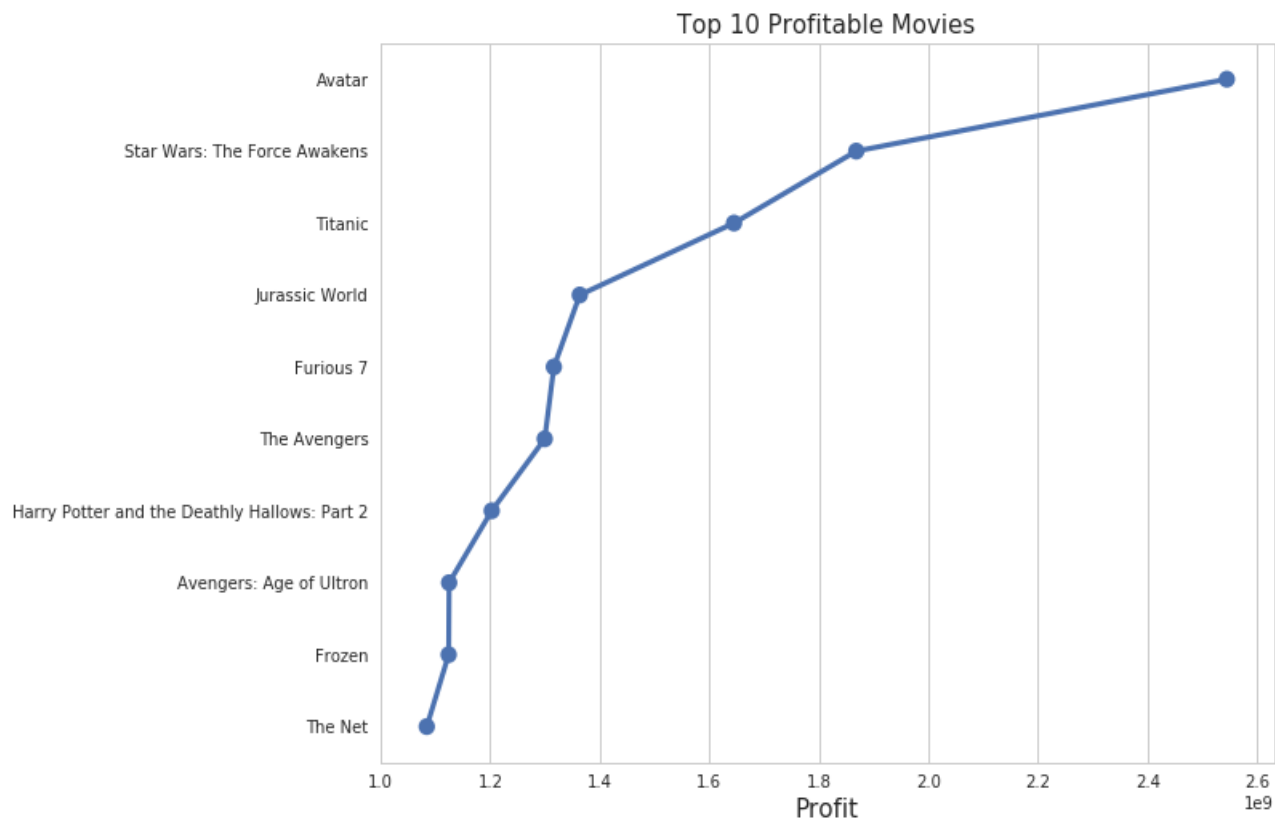
- 1- The visual assessment is done on “tmdb-movies.csv” spreadsheet and then the programmatic assessment in Jupiter notebook for the dataset.
- 2- In the jupiter notebook i mentioned all issues in data.
- 3- Cleaning data : for each issue in the project I defined the issues then the code to solve it then I tested my code.

Issue	Solution
Quality Issues	
There are columns that don't help you in analysis the project	Delete columns that don't help you in analysis the project
relese_date is object type.	Change relese_date to be data time
There are many movies that has illogical budget like (0, 75,115,...) so it will give us alot of outlier and bad analysis for data	Remove illogical budget , to make sense we will make min budget more than 10000.
There are many movies that has unuseful revenue like (0, 3,15,16,) so it will give us alot of outlier and bad analysis for data.	Remove unuseful revenue , to make sense we will make min revenue not less than 10000.
There is a duplicated row that must be removed.	Delete duplicated rows.
There are a null values in (director,production_companies) so we will remove null values.	Remove null values in (director,production_companies).

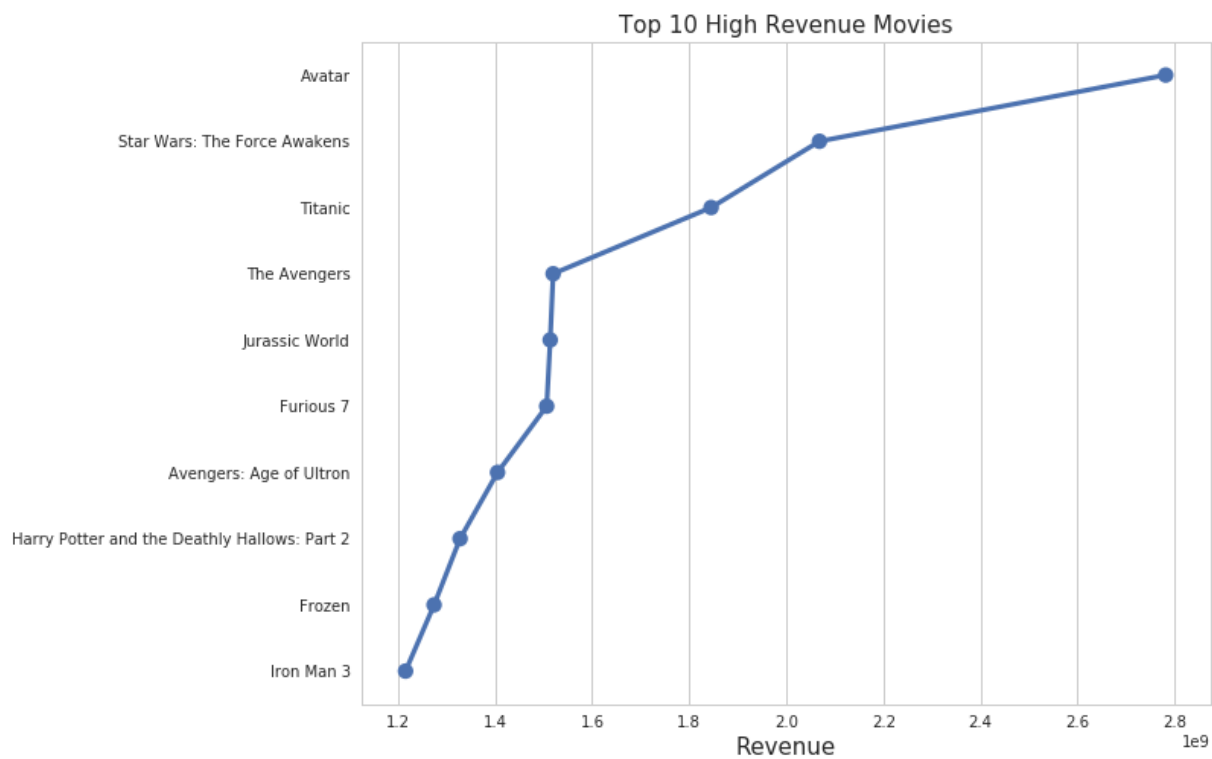
D- Exploratory Data Analysis:

Insights from data analysis and visualization:

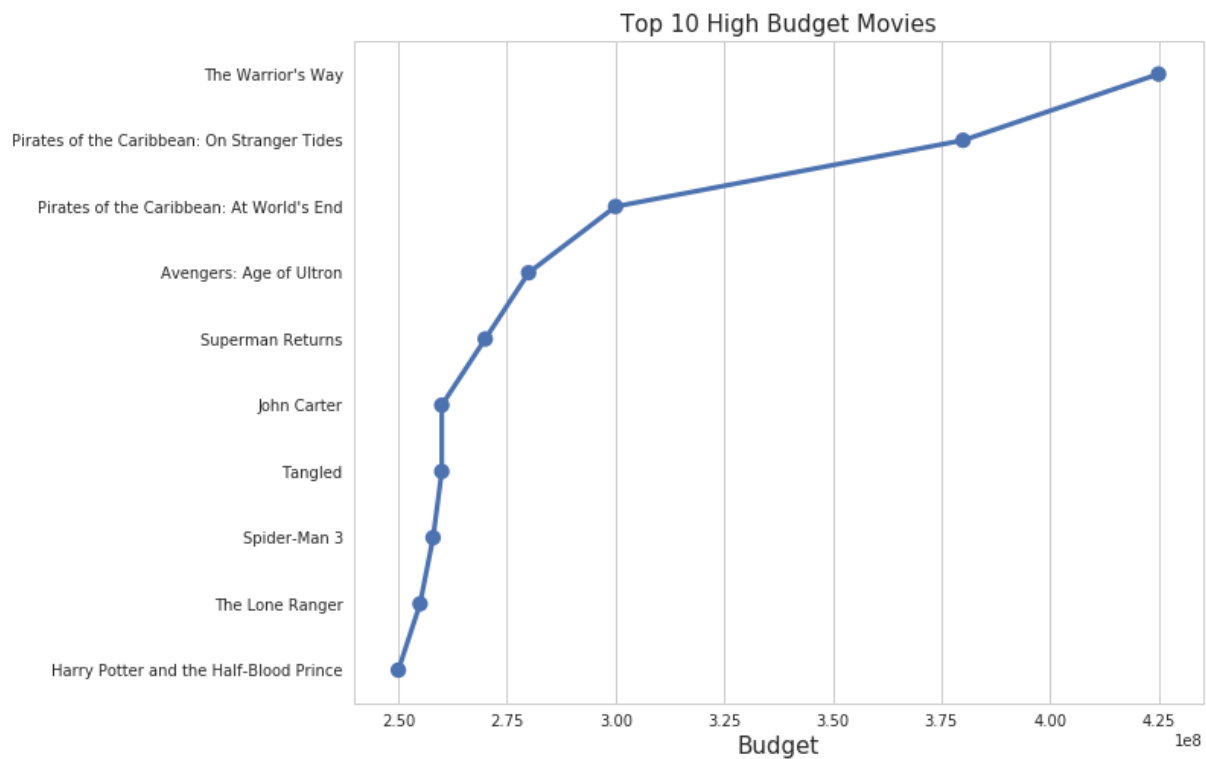
1) The top 10 profitable movies .



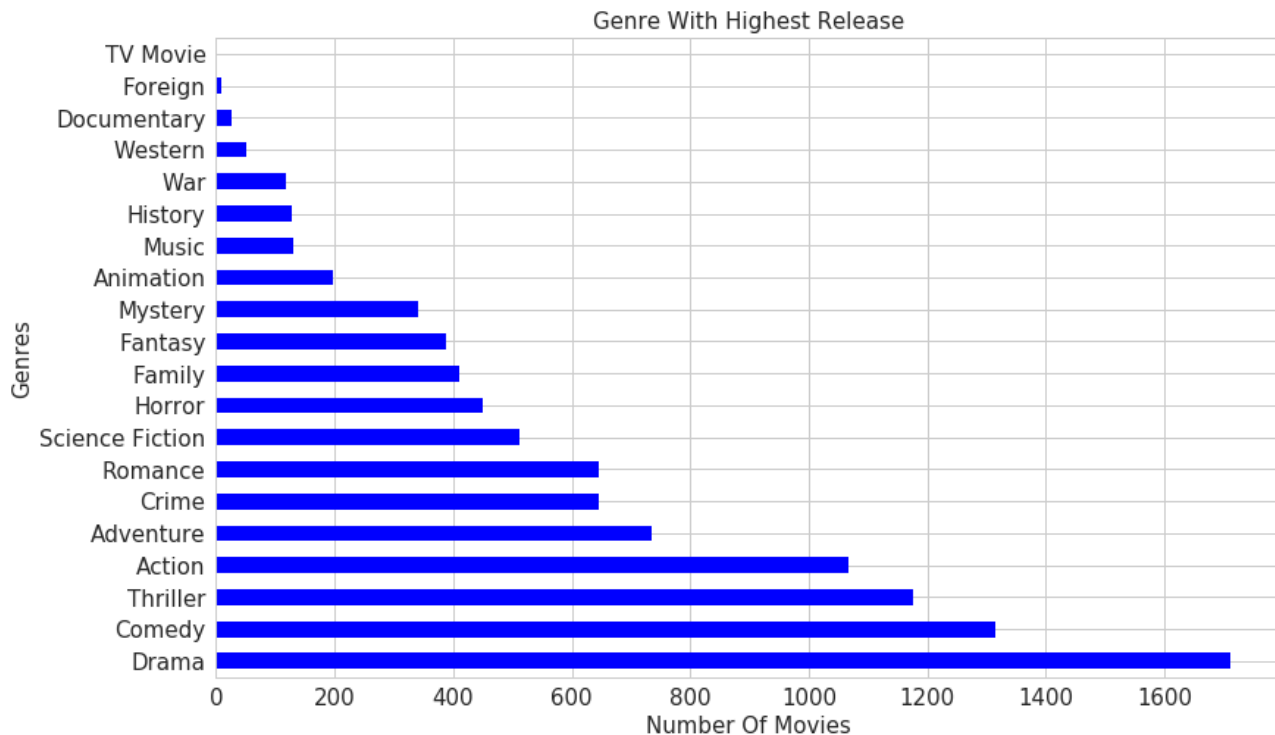
2- The Top 10 High Revenue Movies.



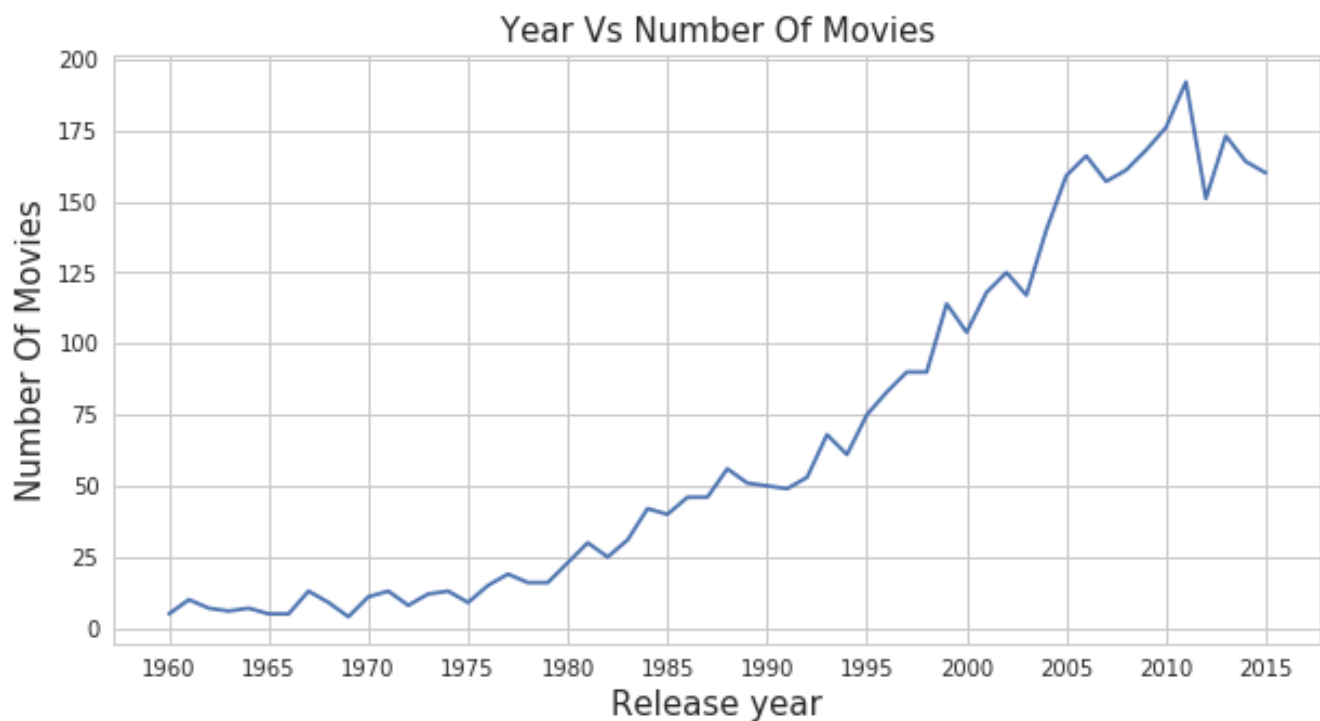
3- The Top 10 High Budget movies



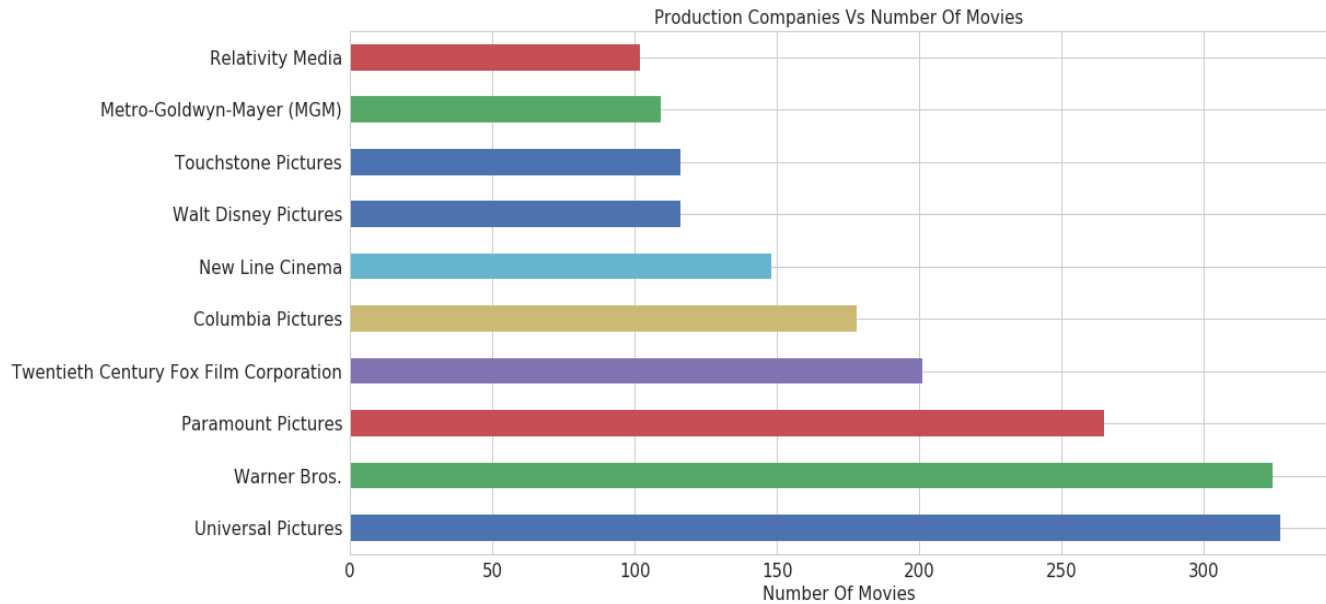
4- Genre With Highest Release



5- The year that has the highest release of movies



6- Top 10 production companies with number of movies



Conclusions from data

- * Avater movie has the highest profit in all movies , making over 2.5B.
- * The Warrior's Way movie is the least in all movies it lost more than 400M.
- * Avater movie has the highest revenue in all movies , making over 2.7B.
- * Chaos movie has the lowest revenue among all movies.
- * The Warrior's Way movie has the highest badget in all movies.
- * Eraserhead movie has the lowest revenue among all movies.
- * Drama genre has the highest release of movies followed by Comedy then Thriller.
- * year 2011 has the highest release of movies followed by year 2013 .
- * the Universal Pictures is on the top of companies releases movies .