

I have analyzed the TMDb movies database. This data set contains information about 10,000 movies collected from The Movie Database (TMDb), including user ratings and revenue.

Questions:-

Q1) Show the names top 10 highest budget movies and plot the bar graph original title vs budget?

Q2) Show the name of top 10 popular movies, plot the bar graph between original title and their popularity?

Q3) Which director has done most movies. Show the top 10 director in terms of number of movies on bar graph?

Q4) Show the relation between budget vs revenue. Analyse the graph and give a conclusion on high budget movies vs high revenue?

Q5) Show the popularity of movie according to budget. Analyze the graph and give conclusion on the relation between high budget movie and their popularity?

Q6) Group the revenue by month and show that which months have high revenues.? Which will help producer to decide the release date of movies. Plot the graph which will help producer easily to choose his release month

Description about investigation

- I have collected the data from kaggle and I have tried to analyze some of the open question.
- I have investigated the relation between revenue and budget. I have also analyzed which is the best release month for high revenues.
- I have analyzed the popularity of movies and number of movies done by directors.
- I have tried to analyze which is good release month in terms of revenue

Data Wrangling

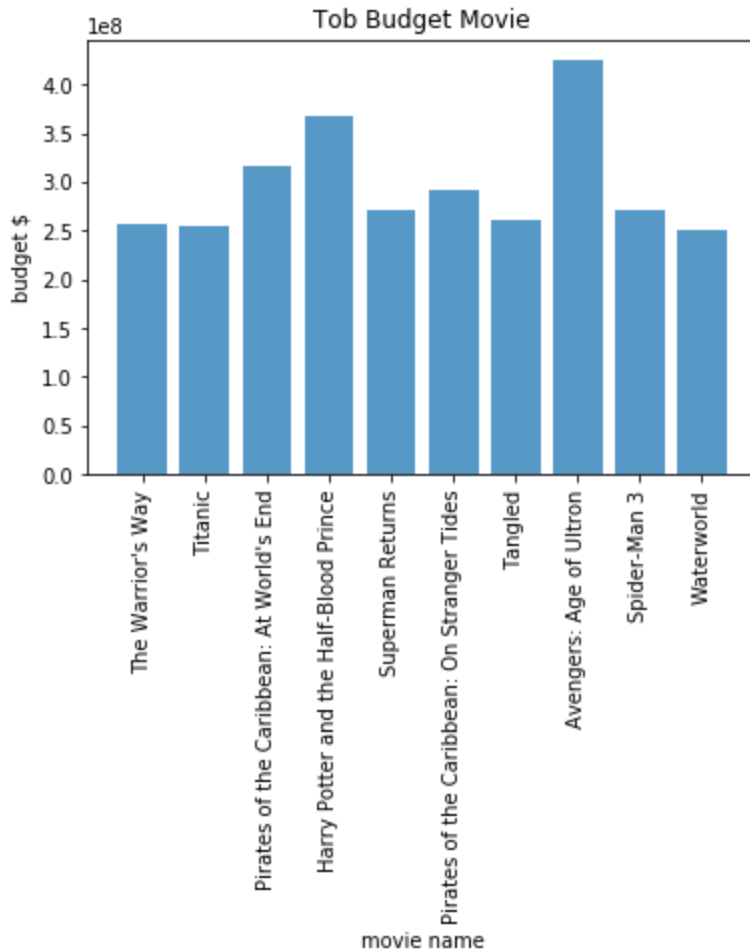
- In analyzing the release month vs revenue. I have first group the revenue by month. Then I have plotted the graph.
- In analyzing the top 10 high budget or revenue. First I have fetched the top movies based on budget or revenue respectively and then plotted the graph.

Summary and plot communicating your final result

Top Budget Movie

Outcome:-

- Avenger-Age of Ultron is very high budget movie and second is Harry potter and half deadly prince.



- Most popular movie is Captain America: The winter soldier and second most popular is guardian of galaxy.

Observation:-

- We can see that Captain America, The winter soldier has very popular actors so the success of movie is also very high. So it has budget.

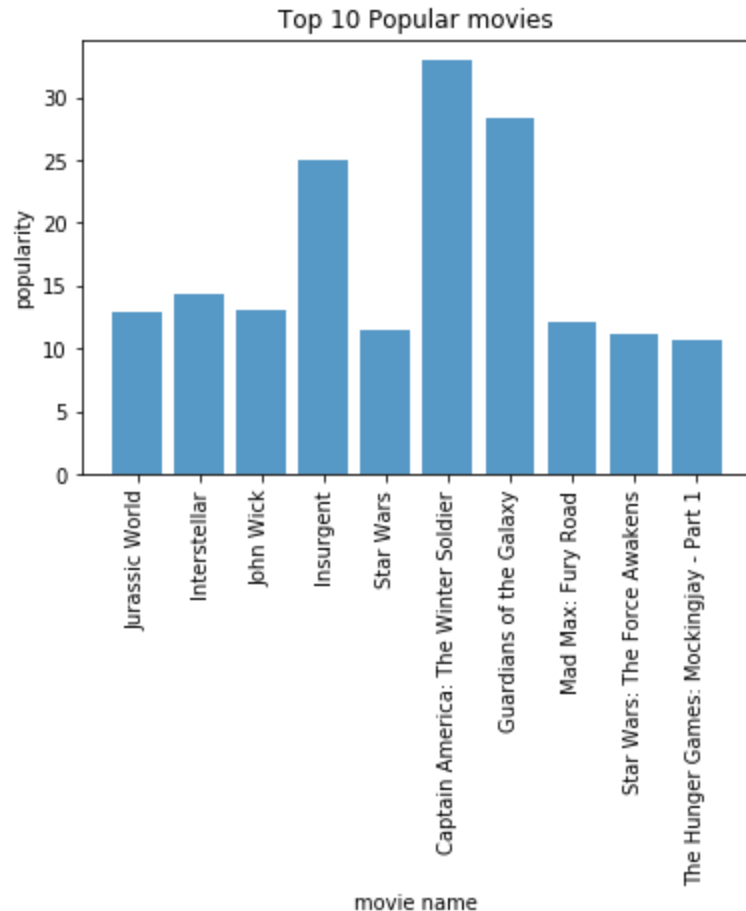
Inferences:-

- We can predict that the other series of Captain America or guardian of galaxy have high budget.

Limitation:-

- There is no direct correlation between movie name and budget.
- Budget on the movie depends on various factors like director, script, actor/actress.
- We can try to find relation between budget of same series of movie like spider man 1, spider man 2, spider man 3.

Top Popular movies



Outcome:-

- We can see that Captain America :The Winter Soldier is most popular and 2nd most popular is Guardian of galaxy

Observation:-

- We can say that Captain America ,The winter soldier has very popular actors/as the success of movie is also very high. So the popularity is very high.

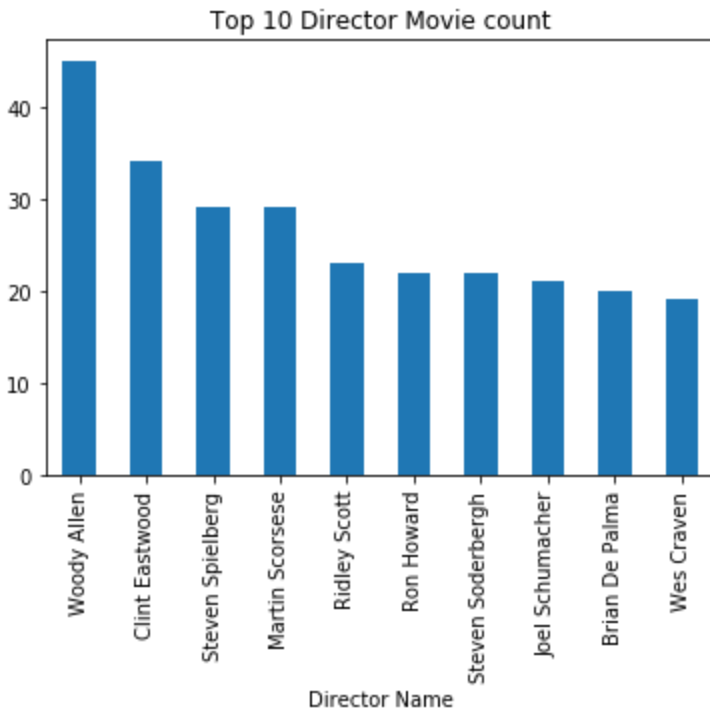
Inferences:-

- We can Predict that the other series of Captain America or guardian of galaxy will also be popular
- Actor/Actress of these movies will also be popular in other movies.

Limitation:-

- There is no direct correlation between movie popularity and movie name .
- popularity of the movie depends on various factor like director,script,actor/actress
- We can try find relation between popularity of same series of movie like spider man 1,spider man 2,spider man3

Top Director movie count



Outcome:-

- We can see from the graph highest number of movie is done by Woody Allen that is more than 40 movies
- Second highest is Clint Eastwood that is more than 30 movies and so on

Observation:-

- Woody Allen has done highest number of movies. It may be because of short movies, which will take less time to produce.

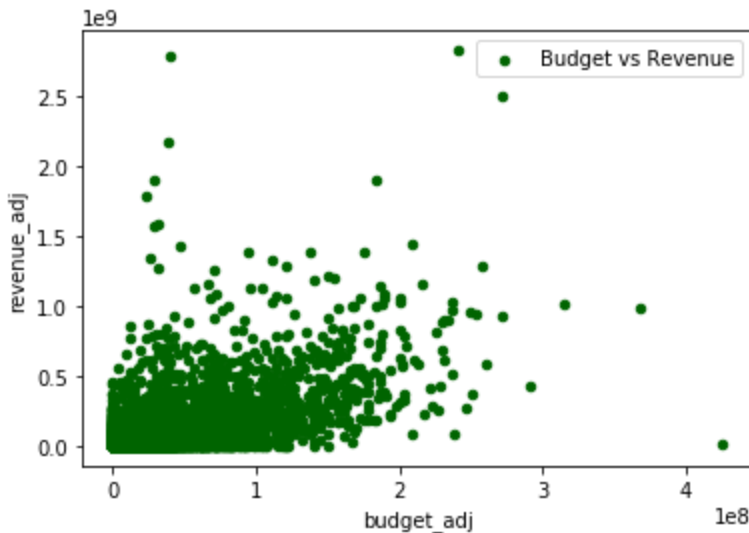
Inference:-

- In future Woody Allen and Clint Eastwood will produce more movies with respect to other directors

Limitation:-

- There is no direct correlation between number of movies and their revenue
- We can't predict the success of a director by more number of movies he has done.
- We should include one more variable like revenue in plotting. Through that we can see that who is the most successful director

Budget vs Revenue



Outcome:-

- Some of the average budget movies has very high popularity. Very few of the very high budget has less popularity.

Observation:-

- We can see some low budget movies has very high revenue .
- Some of very high budget movies have low revenue

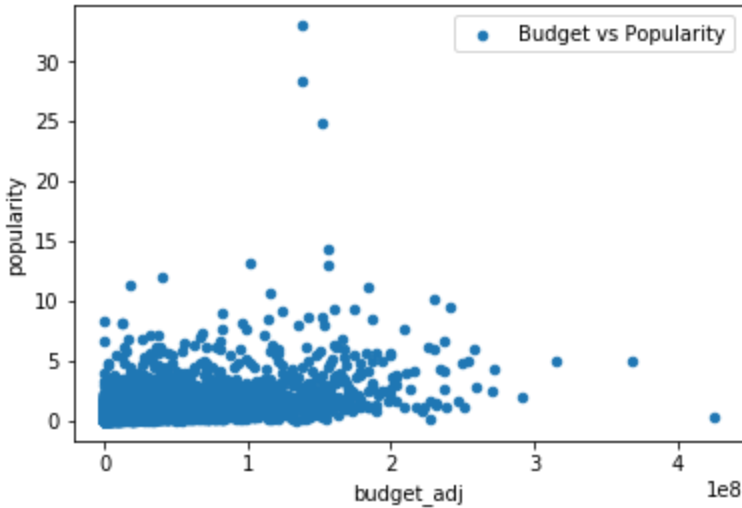
Inference:-

- We can see that mostly average budget movie have average revenue .
- So we can say that average budget movie will have average revenue.

Limitation:-

- We can see the budget and revenue are not directly proportional.
- Some of the high budget movies have low revenue
- We should plot the three variable budget ,revenue and popularity then we will be able to predict more revenue

Budget Vs Popularity of movie



Outcome:-

- December and June has very high revenue. December is holiday time. June is summer vacation time. I think most of the high budget movie will try to release during this June or December.

Observation:-

- We can see some average budget movies have very high popularity.
- Some of the very high budget movies have low popularity.

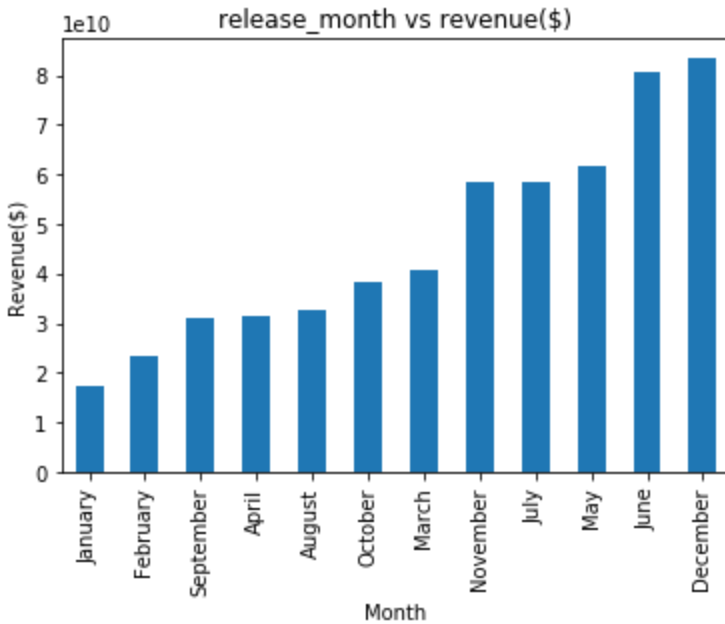
Inference:-

- We can see that mostly average budget movies have average popularity.
- So we can say that average budget movies will have average popularity.

Limitation:-

- We can see that budget and popularity are not correlated exactly.
- Some of the high budget movies have low popularity.
- We should include more variables like actor in the movie, since due to a new actor or not very successful actor, popularity may be low.
- Some of the medium budget movies have high popularity. We can't predict the exact reason. It could be due to various reasons like famous actor/actress or famous director or may be holiday season or could be autobiographical movie or real event movies or any historical movies. It could be due to a very good script. There could be many factors for the popularity of movies.

Release Month vs Revenue



Outcome:-

- We can clearly see that December has very high revenue and June is the second highest revenue month

Observation:-

- December and June are holiday month, so the movie will be able to collect more revenue
- It is due to Christmas holiday in December and summer holiday of children in June

Inference:-

- We can see that every December and June movie will have more revenue.
- So most of the producer will try to launch big budget movie in these month

Limitation:-

- It is not always true that every movie will have high revenue in December and in June
- Revenue of the movie depends on Actor/Actress, Director, script or budget etc
- We can include one more variable like Actor or Director and plot the graph. We will be able to see if this actor or director has very high revenue in December or in June