Conclusion: for **TMDb movie data analysis**

Overview:

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

Limitations:

Assessing & Cleaning Data:

- Dataset contains some missing values & also zero value in revenue & budget columns. All of these values need to be removed before conducting the required analysis.
- Also, the dataset contains one duplicated rows which is also removed to make the dataset cleaner.
- Also, some columns in the dataset are not required in the analysis like:" id", so the list of these columns has been dropped.
- In this analysis, revenue_adj & budget_adj, is used instead of absolute revenue & budget. From my point of view, this gives better way for comparison of revenues & budgets through different years with the same base. This also is useful in avoiding the inflation & currency depreciation effect.
- Also, some columns (like: cast, genres) have the values separated with character "|". This make the analysis more difficult, since you can't deal with one value without separate the values in a single value.

Findings:

Comments on each visualization/result:

Q1: Which genres are most popular?

- The most produced movies are under the '**Drama'** then '**Comedy**" section.
- The 'Family' movies comes in the 10th place according to no of produced movies.

Q2: Is there progress in no of movies produced from year to year?

- There is obvious increase in the no. of movies through the years. The histogram is highly left-skewed which reflects this fact.
- Also from the graph, we can see that the most of movies runtime is between (90 – 140) minutes.
- The average votes form the movies is mostly at "**6**".

Q3: Does the rating has a relation with the revenue?

- We can see the relation between average rating & revenue. This is reflecting the facts that as the audience like the movie as the revenue increase.

Q4: Does the movie with high budget generate higher revenue?

- There is weak relation between budget & revenue. It is not always to have high revenue as the budget increases. We can't say that the movies with higher budget gives higher revenue. Some movies with high budget gives low revenue compared with the most movies & vice versa.
- **Some Limitation:** some values of revenue & budget comes in zero value, so we have to filter to drop these data.

Q5: What are the top directors whose movies has the highest & lowest revenue?

- "**Stephen Spielberg**" is the director who achieve the highest sum of revenue through his movies.
- **"Shinichiro Watanabe"** is the director who achieve the lowest sum of revenue through his movies.

Q6: What are the top cast whose movies has the highest revenue?

- "**Tom Cruise**", then "**Tom Hanks**" is the actor whose movies achieve the highest revenue, while "**Nicolas Cage**" is the actor who achieve the highest average rating for his movies.
- **Some Limitation:** Here we need to have some work to divide the values in the column, since the values was separated by character "|".

Q7: What are the top production companies whose movies has the highest & lowest revenue?

- The most companies achieve the higher mean revenue are: (**Ingenious Media with Twentieth Century & Fox Film Corporation**).
- While the most company achieve the highest sum of revenue is: "**Paramount Pictures**"

Q8: What are the top directors whose movies has the highest & lowest rating?

- The top rated director as proportional rating is: "**Damien Chazelle**", while "**Stephen Spielberg"** is the director who achieve the highest sum of rating.
- The lowest director as proportional & sum rating is: "**Lawrenece Kasanoff**".

Q9: What are the top cast whose movies has the highest rating?

- "**Tom Cruise**", then "**Tom Hanks**" is the actor whose movies achieve the highest revenue, while "**Nicolas Cage**" is the actor who achieve the highest average rating for his movies.
- **Some Limitation:** Here we need to have some work to divide the values in the column, since the values was separated by character "|".

Q10: What are the top movies whose movies has the highest & lowest rating/revenue?

- The most revenue movie is "**Avatar**", then "**Star Wars**", while "**Titanic**" comes in the third place.
- The highest rated movie is "**The Shawshank Redemption**" followed by "**Stop making Sense**", then "**The Godfather**".
- The lowest revenue movie is "**Shattered Glass**", & the lowest rated movie is "**Foodfight!**"

Q11: What are the top companies with the highest no of produced movies?

- The most company produces movies are "**Universal Pictures**", then "**Warner Bros**".

Q12: Which year has the highest no of movies?

- The year (2011) has the highest no of produced movies.

Q13: Which year has the highest budget & highest revenue?

- Year (**2015**) is the highest revenue year, while the (**2010**) is the highest budget year.
- There is weak relation between budget & revenue. It is not always to have high revenue as the budget increases.

Q14: Is there a relation between popularity, revenue?

- There is a weak relation between popularity & revenue, but it is not the rule.

Q15: Is there any relation between runtime & revenue?

- There is some relation, the most revenue movies come in runtime (**90 – 150**) minute.