Name: Mohit Milind Shukla
RUID: 195009793

DAV Assignment: 3

Dataset Name: Movie Industry (Three Decades of movies)

Context: Is the movie industry dying? Is Netflix the new entertainment king? Those were the first questions that lead me to create a dataset and visualize focused on movie revenue and analyze it over the last decades. But, why stop there? There are more factors that intervene in this kind of thing, like actors, genres, user ratings and more. And now, anyone with experience can ask specific questions about the movie industry and get answers.

1) Number of Datasets investigated: 2

a) Movie Industry (Three Decades of movies)
b) Netflix Dataset

I decided to work on the Movie Industry (Three Decades of movies) dataset.
Number of Variables: 15
Number of Rows: 6821

2) Overview of the Variables & Variable Types:

There are 6820 movies in the dataset (220 movies per year, 1986-2016). Each movie has the following attributes:

- Budget: the budget of a movie. Some movies do not have this, so it appears as 0
- Company: the production company
- Country: country of origin
- Director: the director
- Genre: main genre of the movie.
- Gross: revenue of the movie
- Name: name of the movie
- Rating: rating of the movie (R, PG, etc.)
- Released: release date (YYYY-MM-DD)
- Runtime: duration of the movie
- Score: IMDb user rating
- Votes: number of user votes
- Star: main actor/actress
- Writer: writer of the movie
- Year: year of release

The Variables have correct types. Dataset Size is manageable, and the speed of the software is sufficient.

Name: Mohit Milind Shukla
RUID: 195009793

3) Important Variables of Interest:

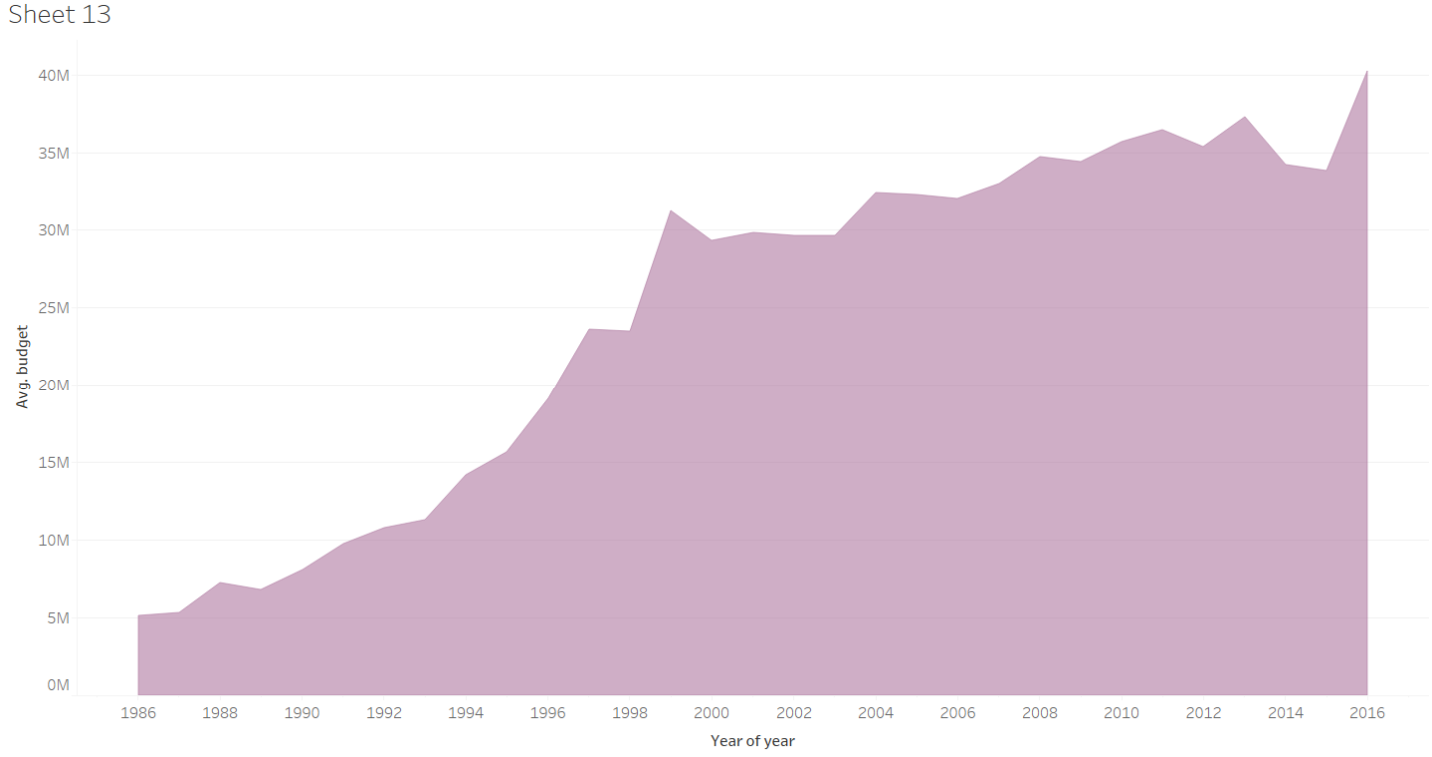
From the 15 variables, let us consider the following variables of interest:

- **Budget**: A films Budget is an integral part for defining success or failure in the box office, financially as well as with respect to popularity.
- **Country**: Which countries make the most movies? What is the average budget each country invests in a movie in dollars? What are the reasons, a country is not able to bring out more movies, is it because of the size, budget, stars?
- **Genre**: Which genre movies are the best. Business wise, popularity wise.
- **Gross**: How did the movie do business wise? Does the genre and rating or any other variables affect the gross directly?
- **Rating**: Are adult movies doing good than R rated movies? Does the rating matter?
- **Runtime**: Do shorter runtime movies do good or longer? Which time range is the best for a movie to do business more and get a higher rating.
- **Votes**: Number of user votes to the movie.

4) Exploratory Data Analysis (EDA):

- The Dataset contains almost 6821 total movies ranging from years 1986 till 2016.
- Every year around top 220 movies from all over the world are considered in our dataset.

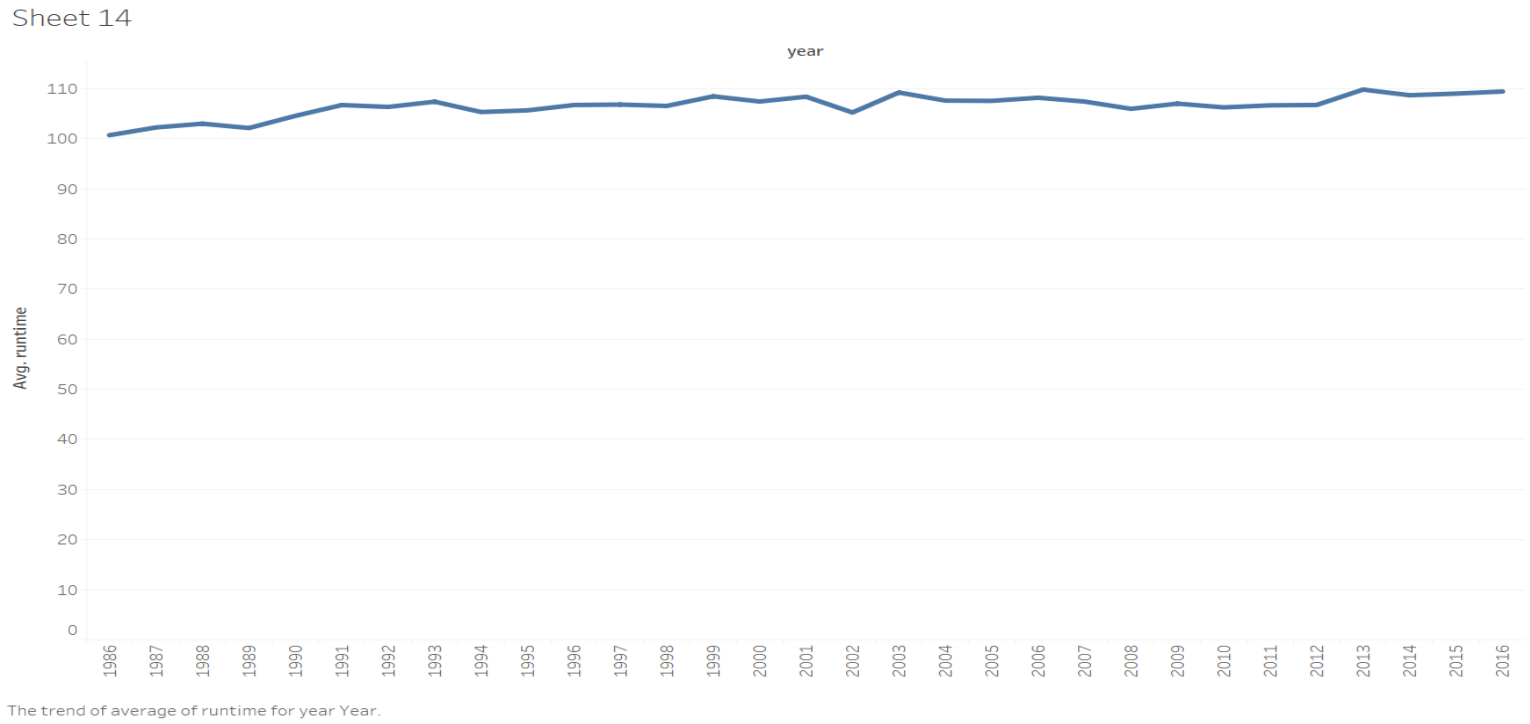
Following are some of the important variables with timelines of the dataset:



Name: Mohit Milind Shukla
RUID: 195009793

As we can observe from the above image that the average budget of the movies have increased from year 1986 to 2016.

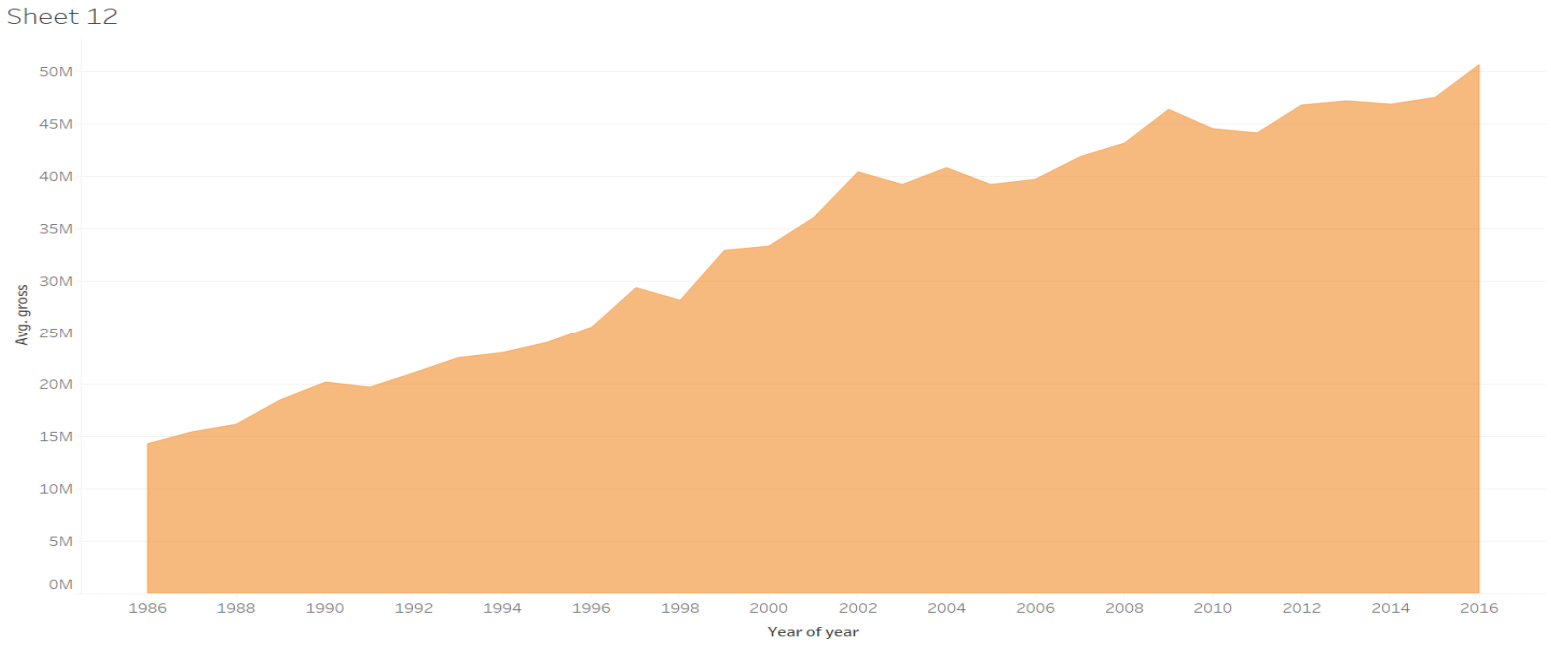
- Average runtime of the movies timeline over the years:-



The trend of average of runtime for year Year.

As we can observe above, the average runtime of the movies in our dataset ranges from 100 to 110 minutes.

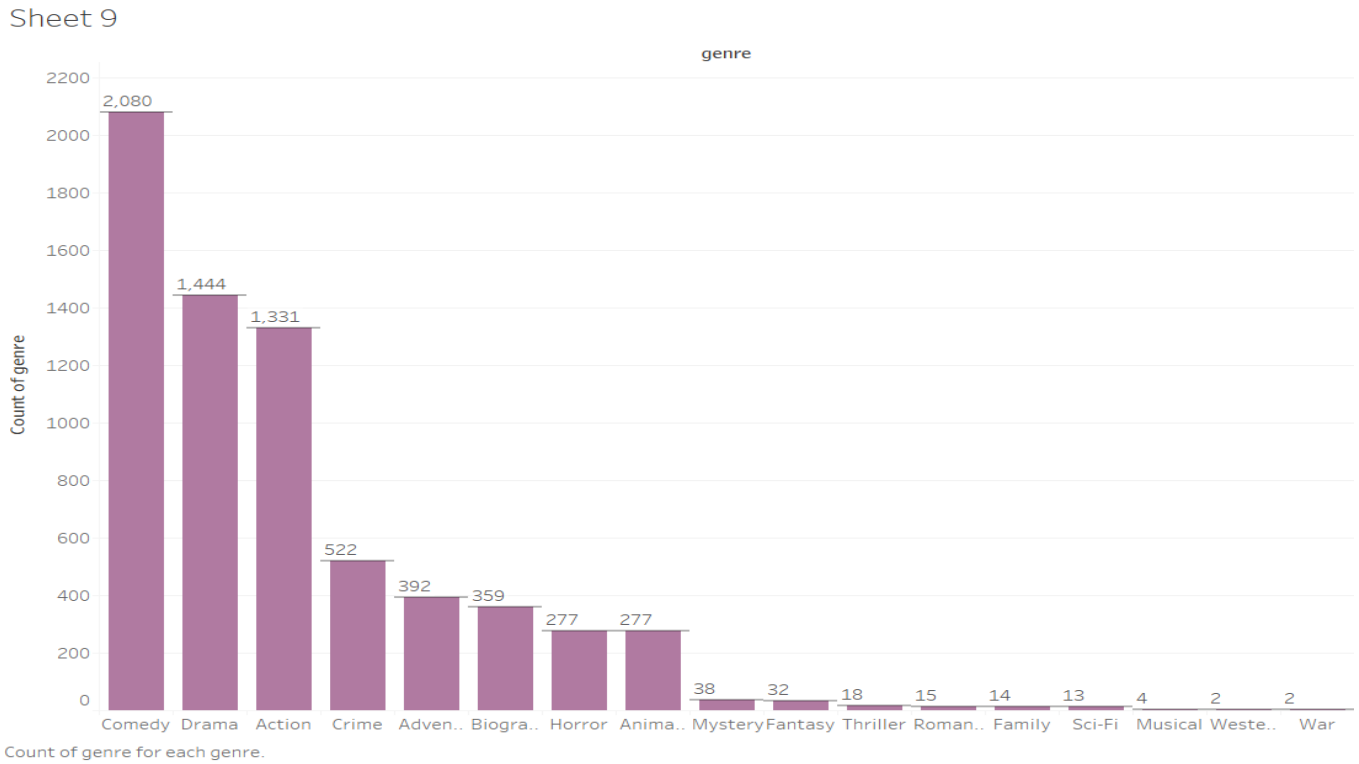
- Gross (Revenue) earned by the movies over the years:-



The plot of average of gross for year Year.

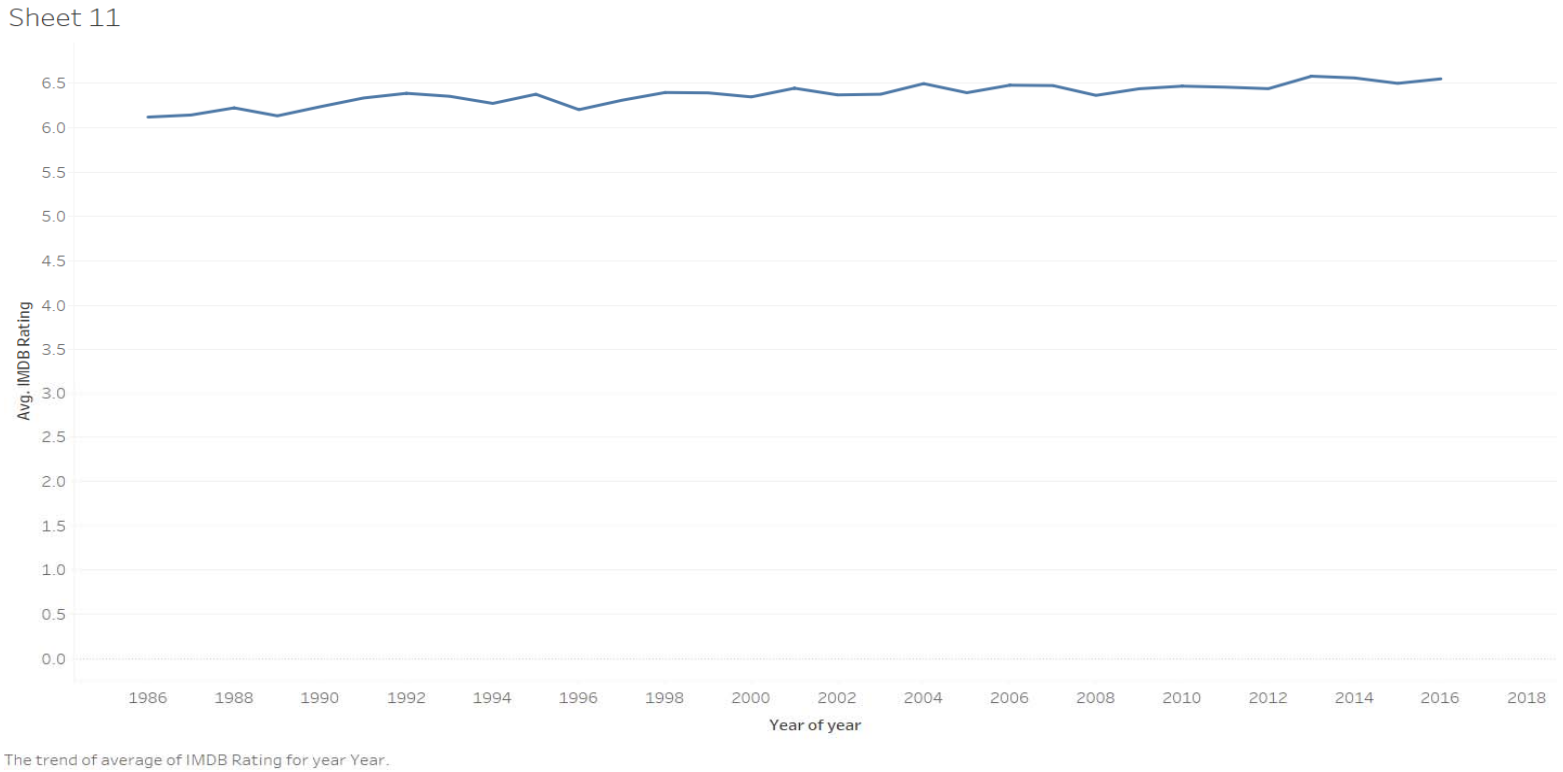
Name: Mohit Milind Shukla
RUID: 195009793

- Total genre count of the movies in our dataset



The above visualization shows us the genre count of our movies we are considering. In our data, we have almost 2080 movies with genre comedy which is the highest and war movies are lowest which are only 2.

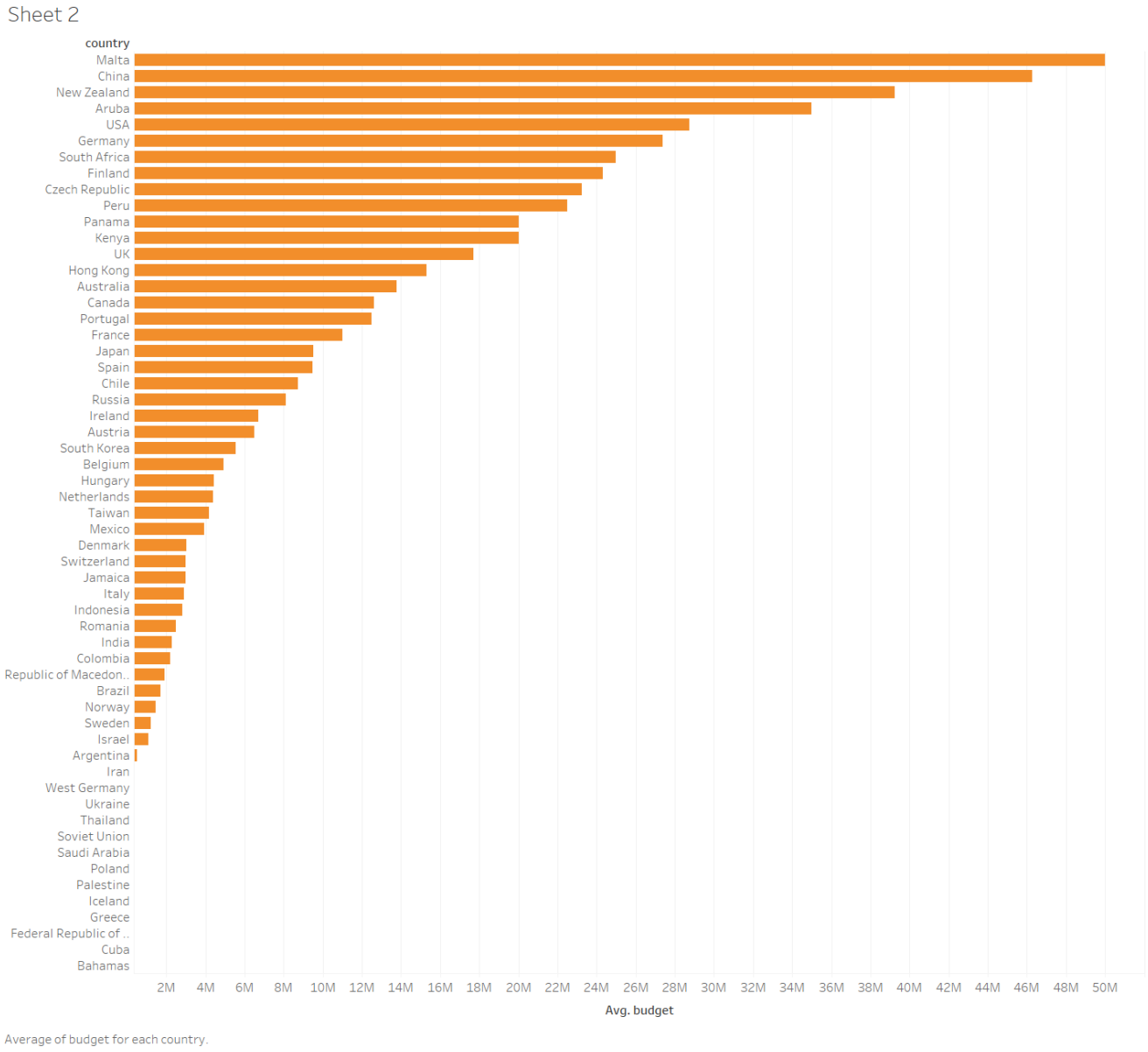
- Imdb rating timeline:- 6 to 6.5 out of 10 Imdb rating.



Name: Mohit Milind Shukla
RUID: 195009793

- **5) Correlation of Important Variables:**

1. Countries with Average budget for the movies:

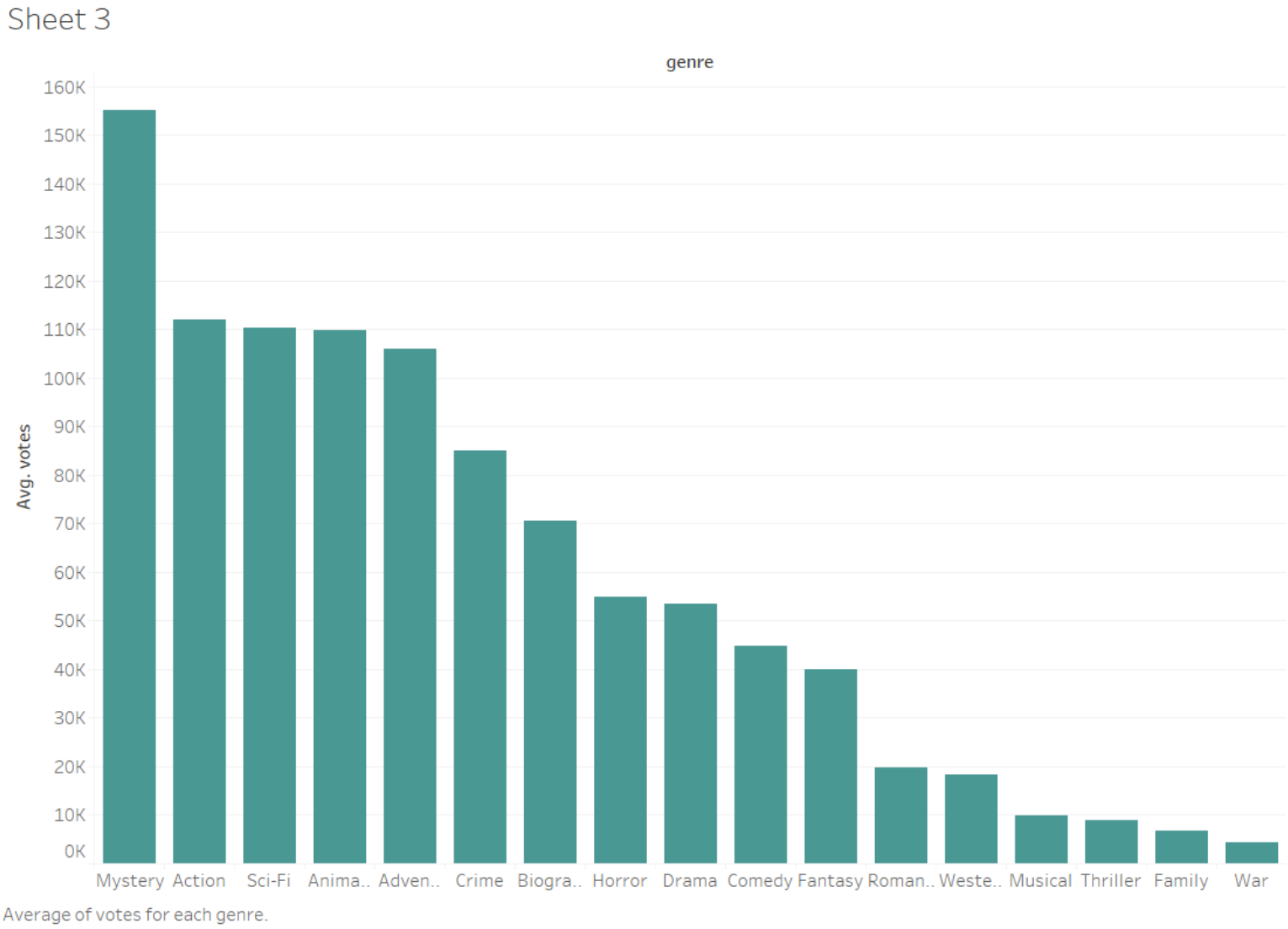


Above visualization shows Malta, China, Aruba, NZ and USA are amongst top 5 countries in the world with highest budgets.

With average budget of around ranging from 35 – 50 Million dollars.

Name: Mohit Milind Shukla
RUID: 195009793

2. Genre of the movie and Average votes:

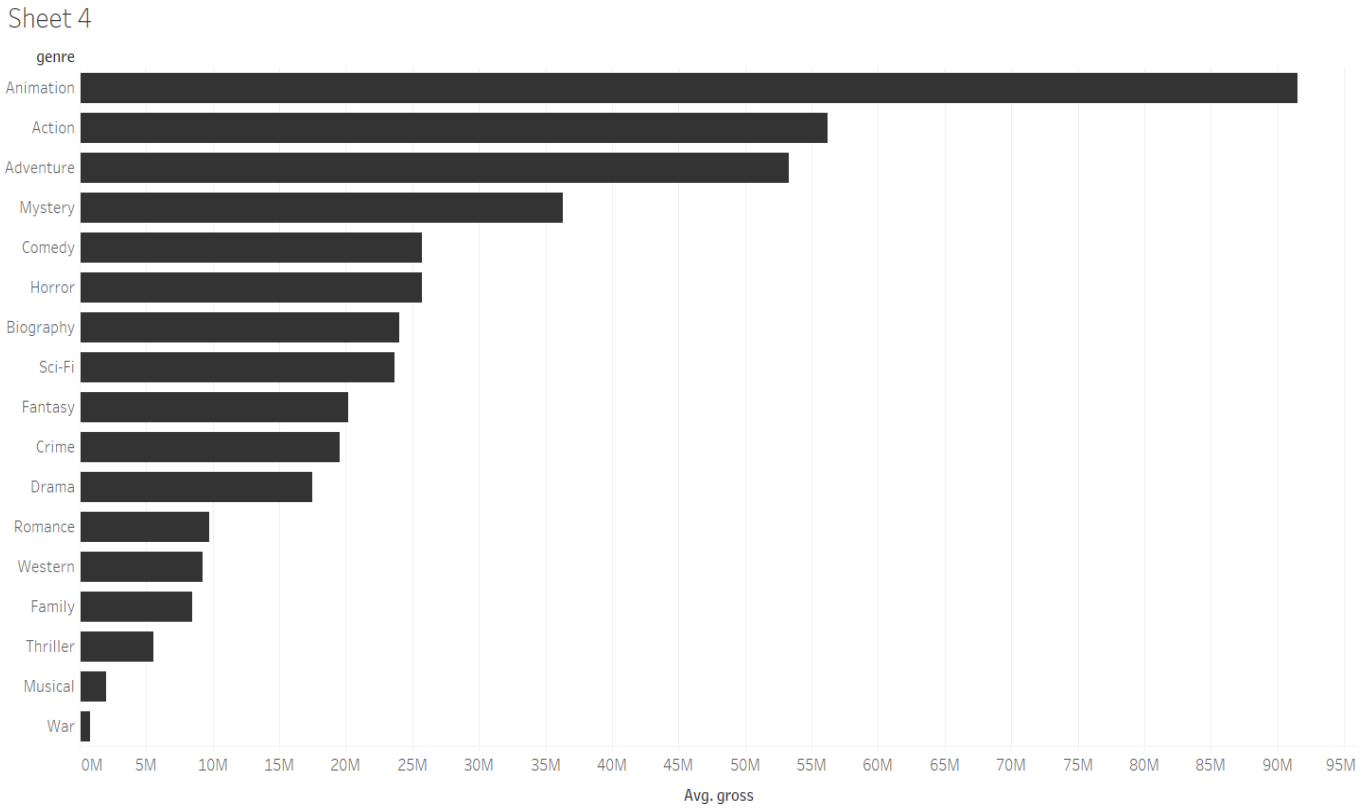


As above we can observe from the visualization that the Mystery genre has received the maximum average votes from the audience since three decades in the world cinema.

And war genre movies have received the least votes from the audience.

Name: Mohit Milind Shukla
RUID: 195009793

3. Genre and Average gross of the movie correlation:



Average of gross for each genre.

Since three decades in the movie Industry, the animation genre has the most highest average gross, that is it has done more business in Box office.
And again, the war genre has the lowest number game in gross earnings too.

We can infer from this that the votes given by the audience affect weather a movie is going to earn at the box office or not.

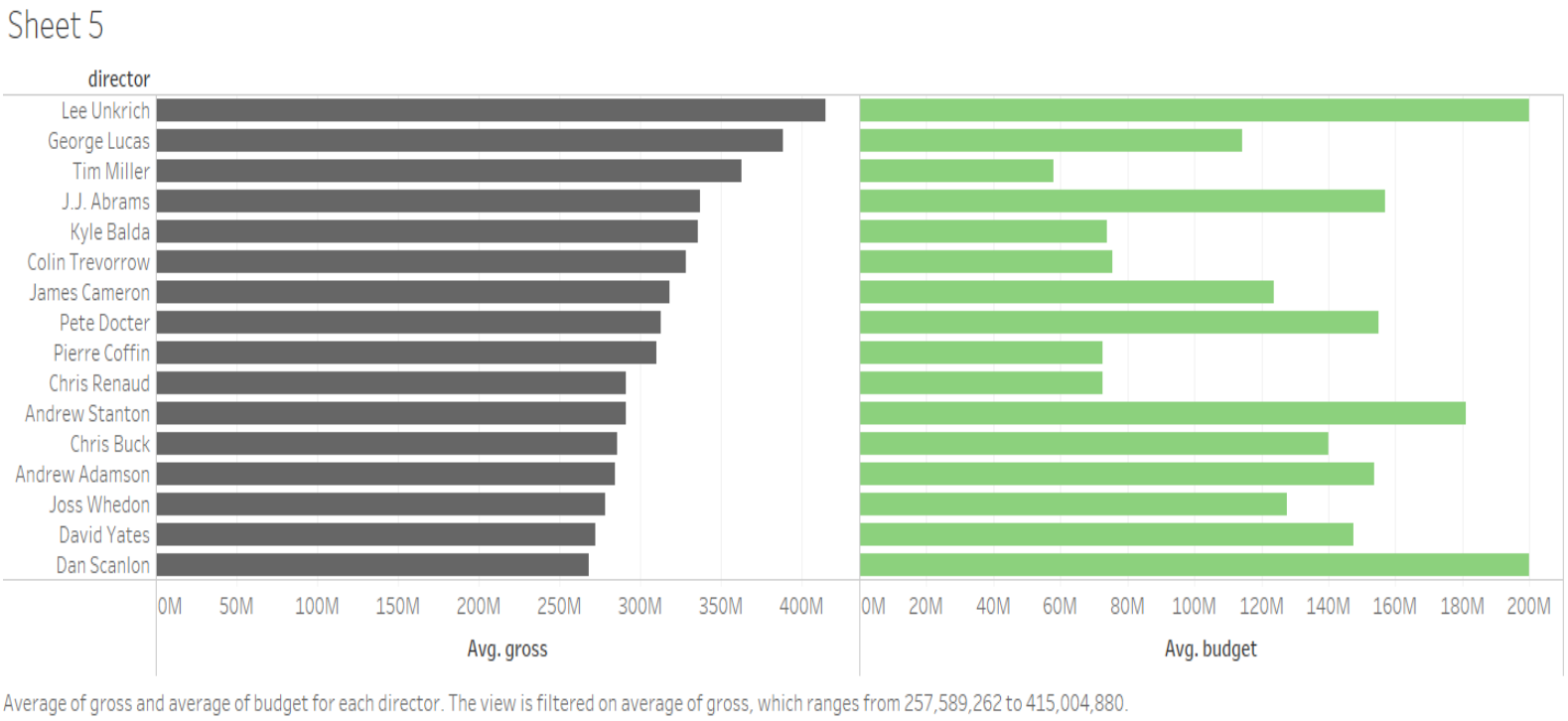
4. Top Directors in the Movie Industry with their Average gross and Average Budget:

Here we are visualizing weather the top grossing directors in the movie industry for the past three decades in the world had higher budget or not.

Answer to the question that will higher budget lead to a movie with higher gross (income in the box office)?

Below is the visualization for the same:

Name: Mohit Milind Shukla
RUID: 195009793



Results: From the above visualization it is evident that the Budget of the movie is not directly proportional to the success of the movie with respect to the gross income.

6) Amount of Data Cleaning:

There were no null values in the dataset and hence there was no need for data cleaning as such. But, the variable year had numerical year in the 1998 format which had to be changed in tableau. We certainly have some 0 values in the budget column, which we will ignore as outliers.

7) Merging two datasets together:

Since the dataset contains all the top grossing movies from 1986 – 2016, the data is sufficient right now to visualize and get proper insights for popular movies.

Therefore there is no further need to merge the dataset with any other data.

8) Does the Dataset have any story to tell?

Well, by visualizing some variables and comparing them with each other there is certainly a story to tell for a movie industry dataset.

Name: Mohit Milind Shukla
RUID: 195009793

1) Consider a movie producer who wants to produce a movie and has various scripts in front of him and wants to choose which one is the best where he can attain maximum box office success.
2) Analyzing the dataset we can provide this information, weather to choose which type of movie according to the country, budget and various other variables.
3) Deciding the budget will be a very important criteria for the producer. How much runtime should be kept will be decided by the editor.
4) Since the budget cannot be a deciding factor for a success of the movie, the investors can look upto genre, rating and votes to decide weather to pursue with the same director or not.
5) There are many aspects more such as company, release month, name length, company which can be visualized and see if we are getting any useful insights.