



# MOVIE STUDIO CREATION.

Exploration of various Data for most actionable insights for a new movie studio.

By : Brandon K , Mercy J, Faith W, David W , Emmanuel C .



## What it entails:

1. Reading into databases relevant to our analysis that have been collected over time to understand the parameters used to analyse movies.
2. Joining relevant tables together.
3. Performing data cleaning using relevant methods in SQL (Standard Query Language)
4. Visualizing Data to give better comparisons.
5. Perform different data analysis techniques to draw relevant assumptions that suit the Business question.
6. From the results of analysis come up with relevant conclusions on the type of movie the company should focus on producing.



## 1.1).Database Connecting, Reading and Cleaning.

To carry out this analysis various datasets were acquired from Box office movies company movie basics and movie ratings. The database used in this case are [im.db](#) and the [bom.movie\\_gross.csv.gz](#).

Import necessary libraries present in the python working environment and other SQL modules.

- Make a connection to sql using sqlite3
- Read the database using Pandas.

Using SQL, query the databases separately to get the parameters of the tables used within each db table, hence proceed to JOIN tables. To join the tables we look for a parameter that is common for both tables known as a unique identifier, hence tables are joined ON that.

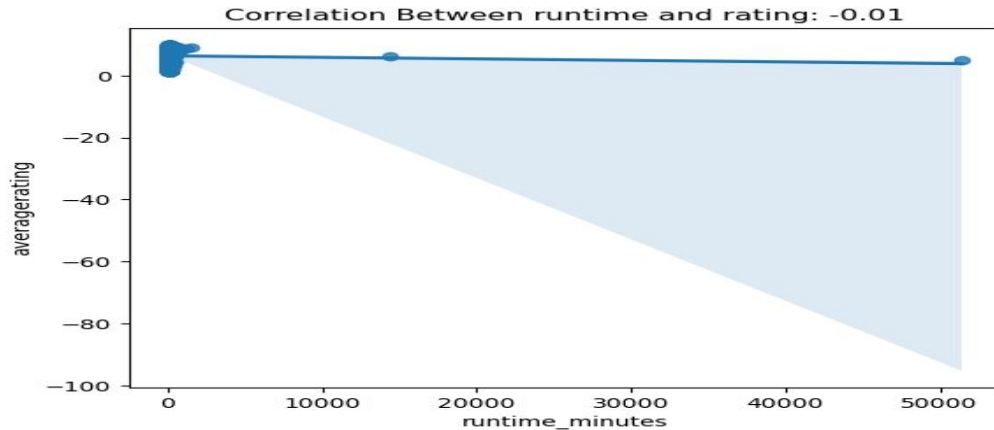
Tables once joined are displayed and cleaning is done, cleaning in this sense is done thru:

- Checking for Null values and handling them in the joint tables
- Converting some values to Floats and Integers.
- Cleaning CSV files and dropping rows with missing values
- Changing formats of some values in the database tables in order for the values to be interpretable by the machine.

## 1.2). Data Analysis, Exploration, Visualization

For this data, the method that was found most in analysing was exploratory analysis.

1. Correlation between most important variables is done. This being correlation between movie runtime and movie rating. The correlation gives a negative result indicating an increase in movie runtime slightly decreases the rating of the movie but by a very negligible difference of -0.01.
  - Conclusion made is : **movie runtime has no effect on movie rating**



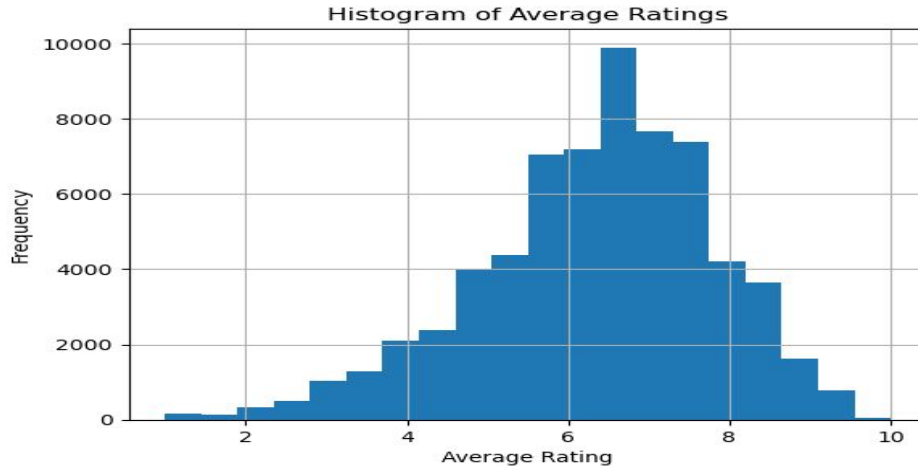


## 2. Exploration of the data thru visualizing the Average ratings of movies.

Average ratings of movies in the data set will give us an area where most data points fall.

From the Histogram we can see that average moving rating is mostly spread near the centre . This implies that the distribution of the ratings is Normal.

Conclusion: **Most ratings are within the average between 5-7/ 10**

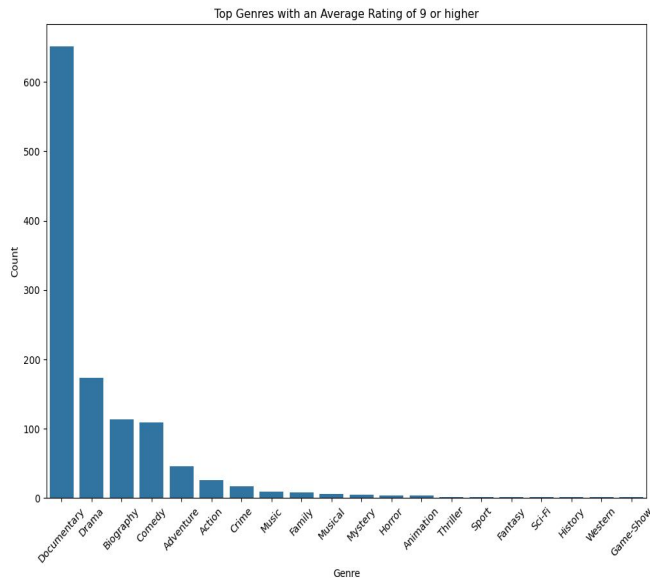


## 1.3). Rating based on genre to choose a particular genre for the company.



When the Genre averages is visualized we can see that the movie genre with the highest customer base is Documentary movies, Drama movies, Biography movies, Comedy movies and Adventure movies making the top 5 movie genres.

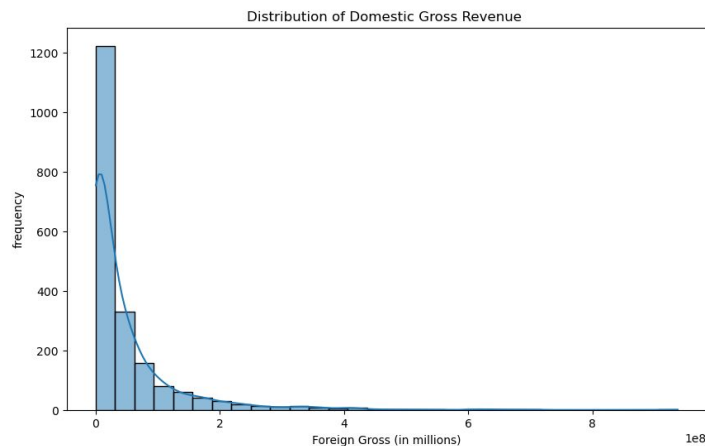
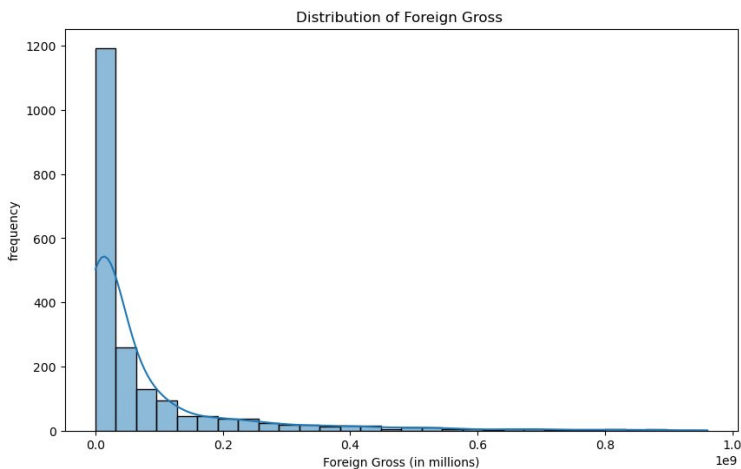
Conclusion: **Documentary movies have the highest rating and are most watched**





The Vizualizations below relate gross revenues to Frequencies.

Conclusions : **With increase in gross revenues the number of movies earning a particular amount decreases**





## 1.4). Recommendations.

From the Analysis and exploration done we can come up with the following incites that will be relevant for the new Movie studio.

1. The studio should major on the top 5 list of movies with highest rating, giving more attention to documentary movies.
2. Movies with Balanced aspects will do better in the market.
3. Spend moderately on the movie budget to avoid losses.

\*\*\* FIN\*\*\*