Investigate a Dataset

1. Dataset we used

TMDB Dataset was used for the analysis

2. A statement of the question(s) you posed

1. How much has the revenue generation changed over time?
2. Which movie genres has the best ratings in the entire time period of the dataset?
3. Which movie genres have the highest revenue generation?
4. Which director is responsible for the highest revenue generation?

3. A description of what you did to investigate those questions

1. We removed the unnecessary data and extract the relevant data from the main dataset. Now there is some null data in our genre and director column, so we need to remove this using the *dropna* method.
2. We can remove the budget and revenue coloumns, since the columns, 'revenue_adj' and 'budget_adj' are more appropriate since they take into consideration inflation.
3. Also we will remove certain columns like id, imdb_id, cast, etc. as they are currently not needed in our analysis.
4. Delimiting of genre and director columns will be done through the *assign* and *explode* methods.
5. When using the mean method on 'budget_adj' and 'revenue_adj' columns, we need to make sure to remove the cells with the 0 values.

4. Documentation of any data wrangling you did

1. Many unwanted columns present in our dataset that we do not need for the process of analysis.
2. There are some missing values in the 'director' and 'genre' section.
3. Incase of the revenue and budget section, we have many '0' values. We cannot find the proper mean of these columns, if these null values exist.
4. Genre and director column contain numerous values per cell delimited by '|', which will prevent us from forming correct groups of these data.

4. Summary statistics and plots communicating your final results

Note that the 'explode' function used to separate the movie genres from the genre columns and directors from the director column was taken from the following stack overflow question
https://stackoverflow.com/questions/12680754/split-explode-pandas-dataframe-string-entry-to-separate-rows