INVESTIGATE A DATASET

INTRODUCTION TO DATASET

I am investigating a TMDb dataset that contains about 10k+ details of movies. This dataset contains full credits for both the cast and the crew, rather than just the first three actors,but some of the entry has no accurate information.

QUESTIONS TO BE ANALYSED

From when Women centered movies are made?  
Which film has made the most money ever?  
Which genre of movies has been relatively preferred over the years?  
who are the most preferred Actors and Directors Before & After 2000?  
What are the top successful movies with respect to runtime before and after 2000?

DATA WRANGLING

Data Acquisition

Data has been loaded using Pandas and operations performed to inspect them.

I deduced from the observations that data contains 10866 rows, 21 columns and currency has no unit.

Data Cleansing

Exploratory of data analysis is easier with cleaned and less data.

REMOVING UNUSED COLUMNS

Columns like ‘id', 'imdb\_id', 'homepage', 'tagline', 'overview', 'budget\_adj', 'revenue\_adj' are not helpful for my data analysis, So I removed them.

REMOVING DUPLICATES

Duplicate data are harder to deal with. So I removed duplicates from the dataset.

DETECT MISSING VALUES

Isnull().sum shows that some of the columns contains NaN values. Dropping the rows contains NaN values is efficient step of data clensing.

DATA EXPLORATION

Column ‘Keywords’ in TMDb reveals about peculiarity of a movie. Its is not a story outline. If a move is women oriented movie , possibly keywords include words like women and feminism.

To understand the paradigm shift in female centric movies, I have matched the words ‘women’ and ‘feminism’ in keywords and found that female centric movies are released when the era of feminism began in early 2000.

Profit is nothing but a difference of revenue and budget. Idxmax() returns index of first occurrence of maximum over profit axis. It shows that the movie Avatar has made great profit out of all these years.

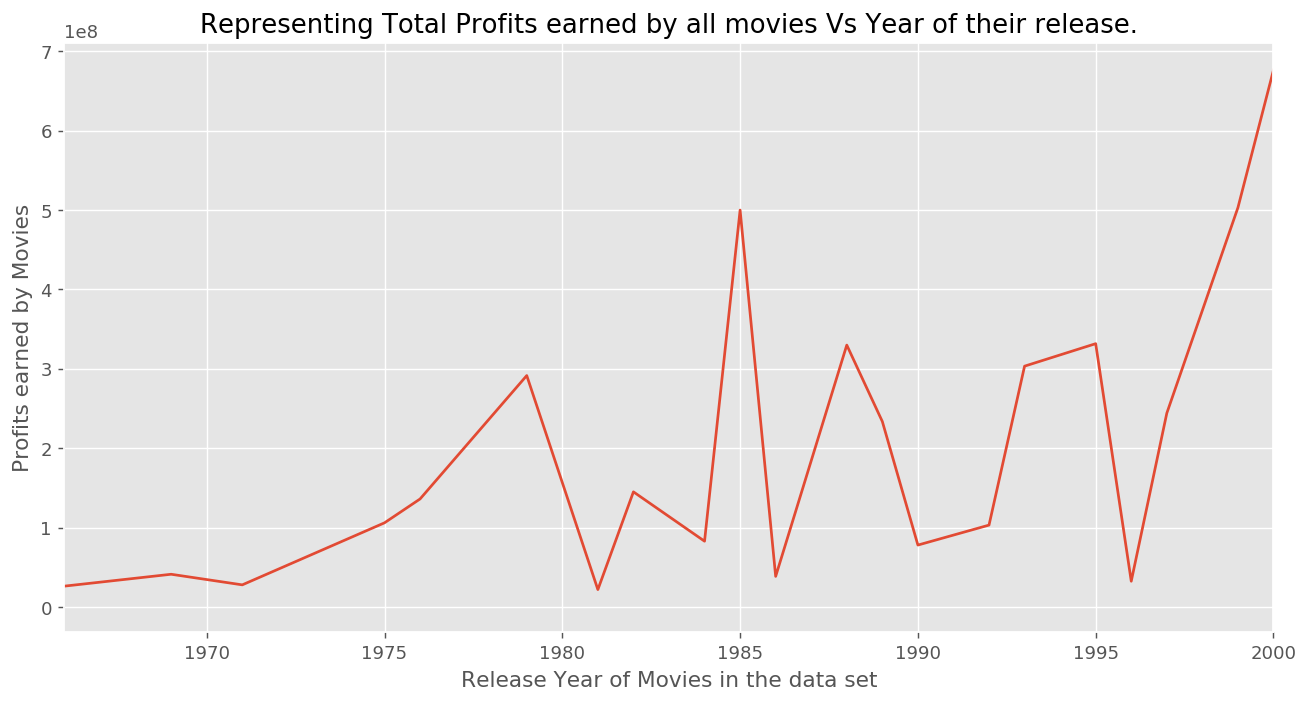
Movies are made based on the interest of viewers. Increased number of movies in particular genre shows which genre Film makers select to make Blockbuster Movies. The count of particular genre of mega hit movies explicitly answers the question which genre of movies has been relatively preferred over the years.

Time is a game changer in every aspect of universe. People’s taste, culture changes with time. Film makers must know to pick the genre, cast and everything else to stand out in a competition.

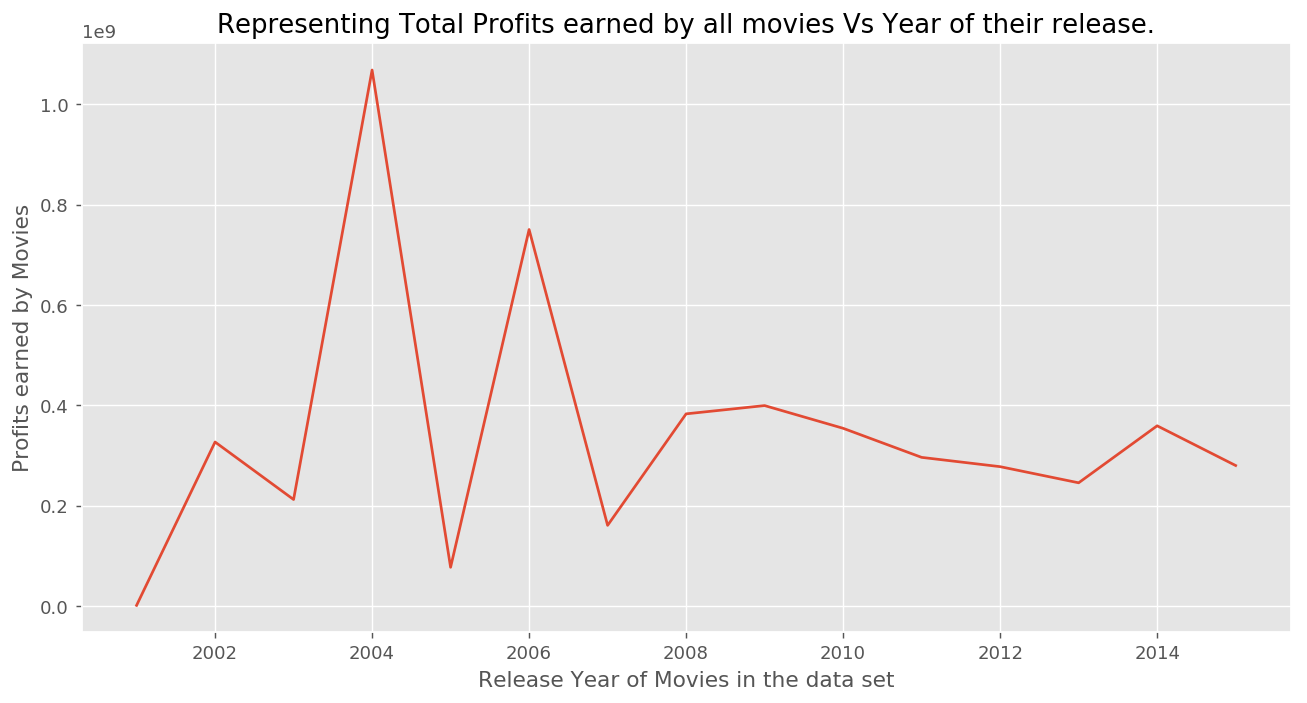
Movie is considered to be Blockbuster if it’s profit is to times the Budget. Based on this condition I filtered a blockbuster movies from the database. Out of the filtered Blockbuster movie data I separated them into two datasets based on their year of release – (Blockbuster movies before 2000 and Blockbuster movies after 2000)

Conclusion

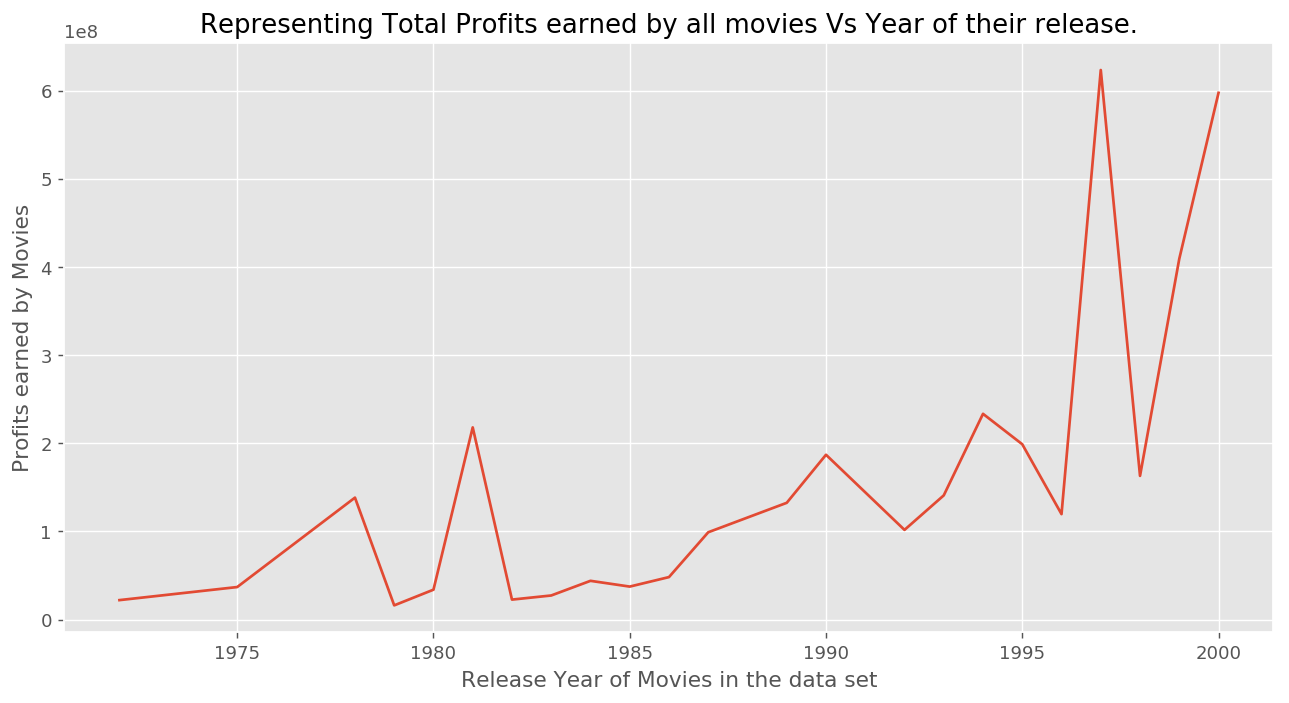
Profitability of Drama movies before 2000



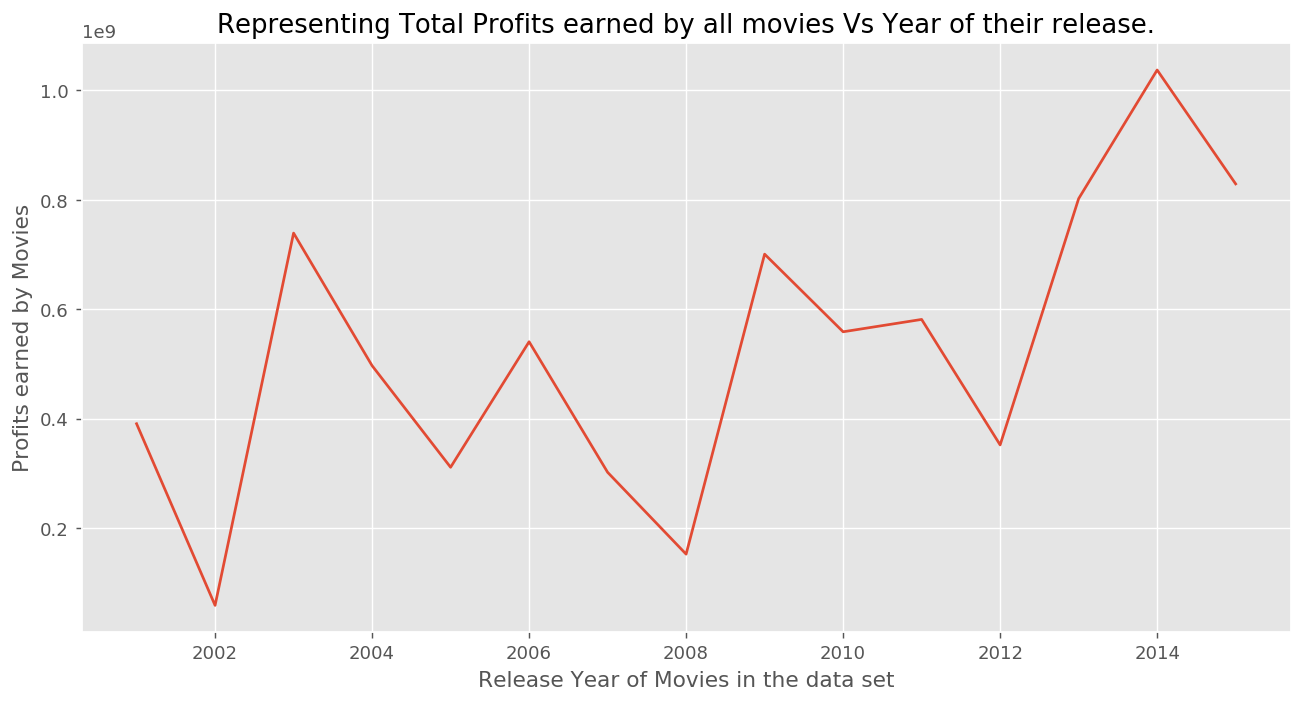
Profitability of Drama movies after 2000



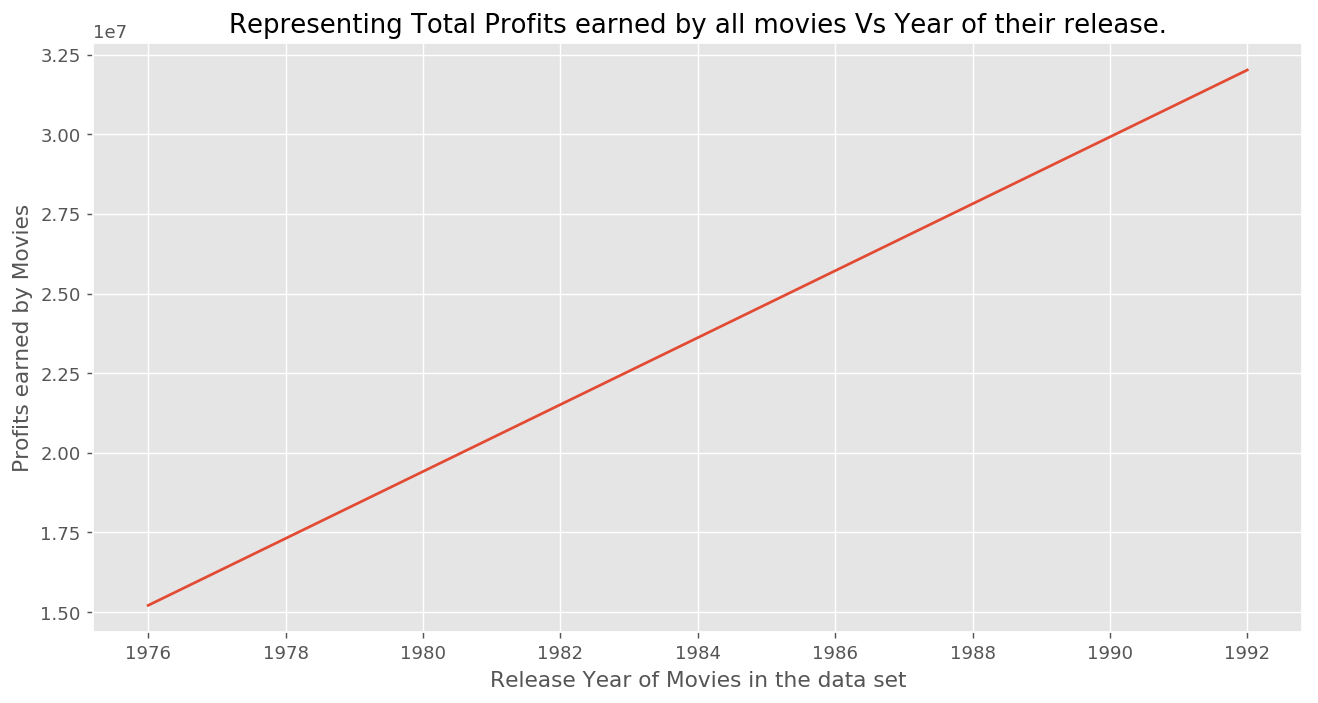
Profitability of comedy movies before 2000



Profitability of comedy movies after 2000



Profitability of Thriller movies before 2000



Profitability of Thriller movies before 2000

