

Investigate a Dataset (TMDb Movie Data)

The primary goal of the project is to go through the dataset and the general data analysis process using Numpy, pandas and Matplotlib. This contain four parts:

Data Set:

TMDb movies data set ,This data set contains information about 10 thousand movies Overview including user ratings and revenue. It Contains 21 columns with 10866 Rows, Dataset contain null values that we will handle in our process.

Questions that we will answer on this project:

We will explore our data by showing:

1. Number of Movies per year
2. Movies which had highest and lowest profit
3. Top 10 movies, which earn highest profit?
4. Most profitable year
5. Highest released Genres

Steps:

Data Cleaning

1. Removing movies in the database have zero budget or zero revenue.
2. Removing the duplicated rows from the dataset.
3. removing unused columns
4. Changing release date column into date format.
5. Changing format of budget and revenue column.
1. Removing movies in the database have zero budget or zero revenue.

Exploratory Data Analysis

Compute statistics and create visualizations with the goal of addressing the research questions.

Conclusions

- Maximum Number of Movies Release In year 2011.
- 'Avatar' then 'Star Wars' then 'Titanic' are the most profitable movies.

- Drama, Comedy, Thriller and Action are four most-made genres.
- Max profit year is 2015