

The Factors Influence the Gross Earning in the Movies Industry

Jhih-Rou Huang

Abstract

This project aims to understand the factors that influence the gross earning in the movie industry. Our customers want to join and invest in the film industry. However, they are new to the field. Moreover, the budget and gross earning are all different for different movies. Hence, this project aims to investigate the relationship of each factor with the gross earning. By doing so, we can help the customer better understand the gross earning of the movie and make a better decision. The result shows that the vote and IMDB have a strong coefficient with the gross earning. Furthermore, the ridge with the best alpha value has the best r squared value in the regression analysis.

Design

The movie is one of the most popular entertainment activities. However, different kinds of movies have various budgets and gross earning. Therefore, the investor will be interested in whether the rating and the number of voting correlate to the profit of movies. This project investigates the correlation between different factors of movies and provides information to investors for their reference.

Data

Multiple pages, five pages of data were scraped from the IMDb website for the data investigation. Two datasets were merged based on the movie's name. In the process of data cleaning, duplicate and invalid data were dropped. A new column called years since the release was calculated based on the years released.

Algorithms

- Delete duplicates and drop rows with null
- Feature analysis and linear regression
- Rescale data
- K-fold/ Train/ Test/ Validation

Tools

BeautifulSoup/ Numpy/ Pandas/ Scikit-learn/ Statsmodels/ Matplotlib/ Seaborn

Communication

