

Data Science Research Methods Assignment-3

Sahil Singh

University of Sussex

2021

Introduction

The objective of the report was to analyze a movie dataset from imdb to make recommendations on the type of movie the studio should consider making. Our budget for production was 1.5 million.

Data Analyses

- First step was to control for budget and only consider movies with a budget of less than 1.5 million.
- Next, a new feature, 'profit_percentage' was calculated from the data such that,

$$\text{profit_percentage} = \left(\frac{\text{gross}}{\text{budget}} - 1 \right) \times 100$$

where 'gross' is just the total earning of the movie and 'budget' is the budget of the movie

- We try to analyze the gross and profit of different genres through a boxplot.



- Tried to analyze most prevalent subplots in these movies.

plot_keywords	frequency of occurrence
friend	16
love	9
independent film	8
drugs	7
friendship	7

Hypothesis Generated from Data

- On analysing the subplots, one sees a clear pattern that certain movie genres, for instance, Comedy/Drama/Music have higher earnings.
- With respect to profit_percentage however, one sees that genres like Horror and Horror/Thriller have higher return on investment.
- We try to test if the above mentioned genres have higher returns on average compared to other genres.

Summary

