



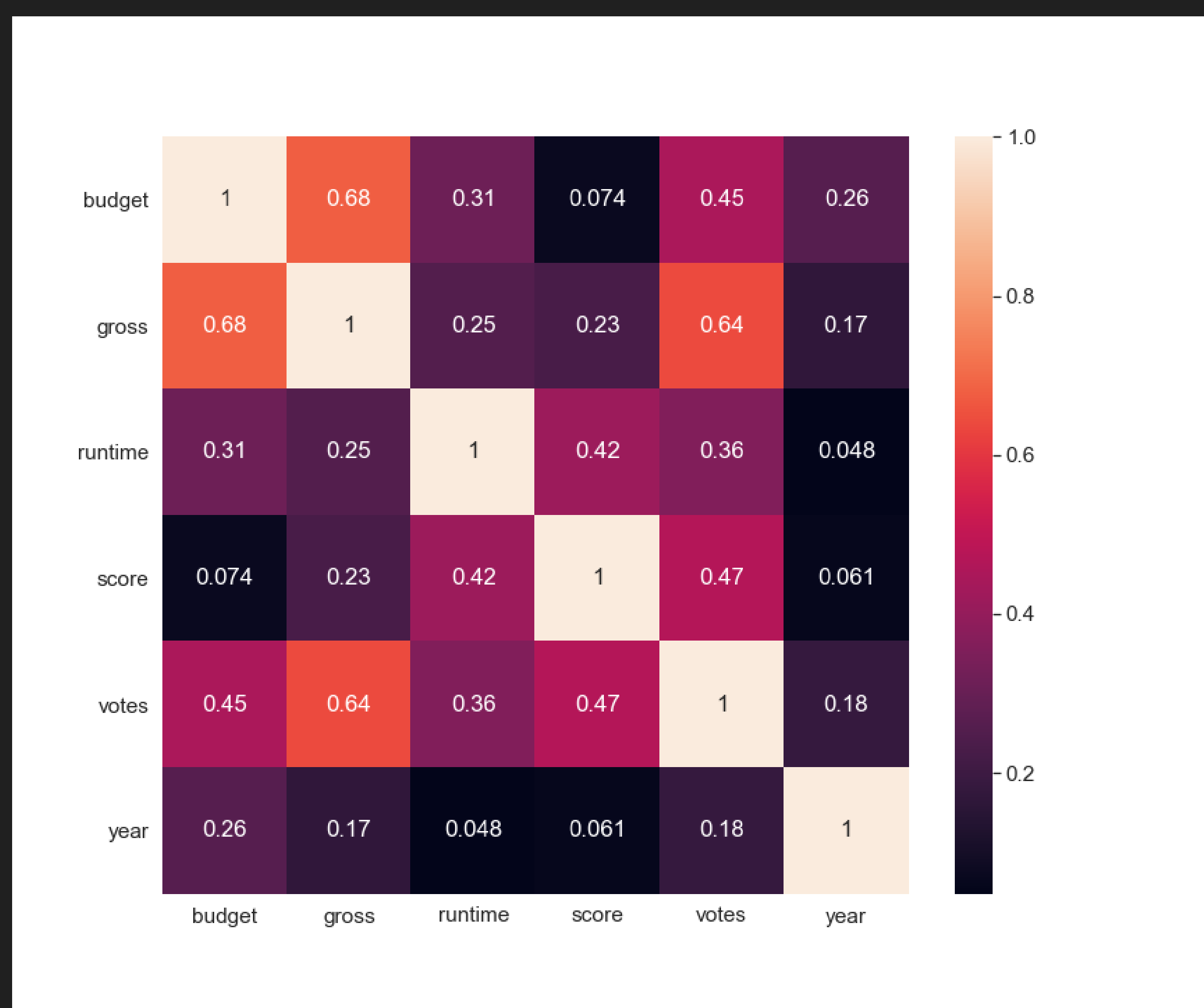
UNIVERSITY OF TARTU

PREDICTING A MOVIE'S BOX OFFICE AND ANALYZING IT'S KEY FEATURES

Silver Spitsõn, Karen Saksakulm

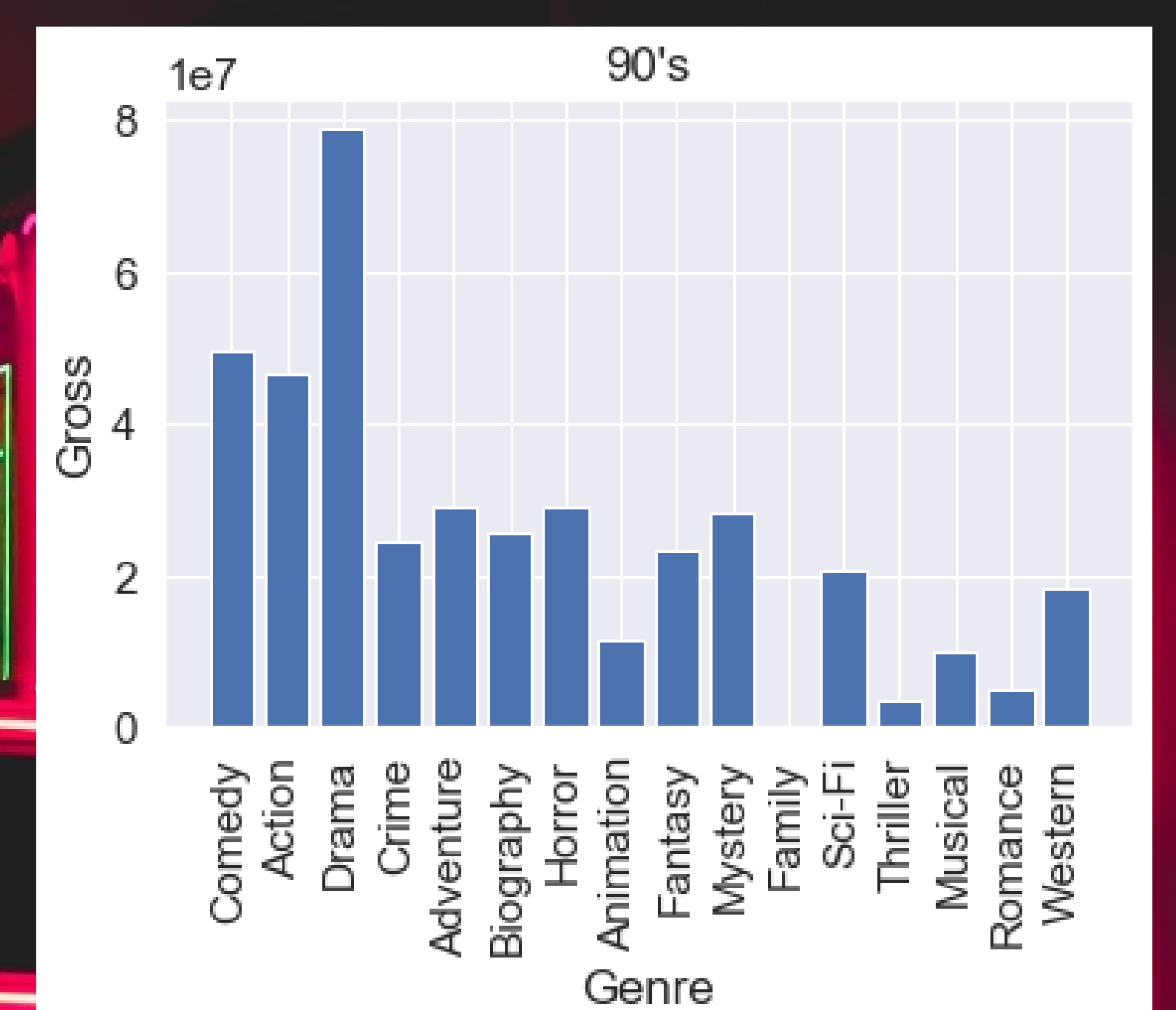
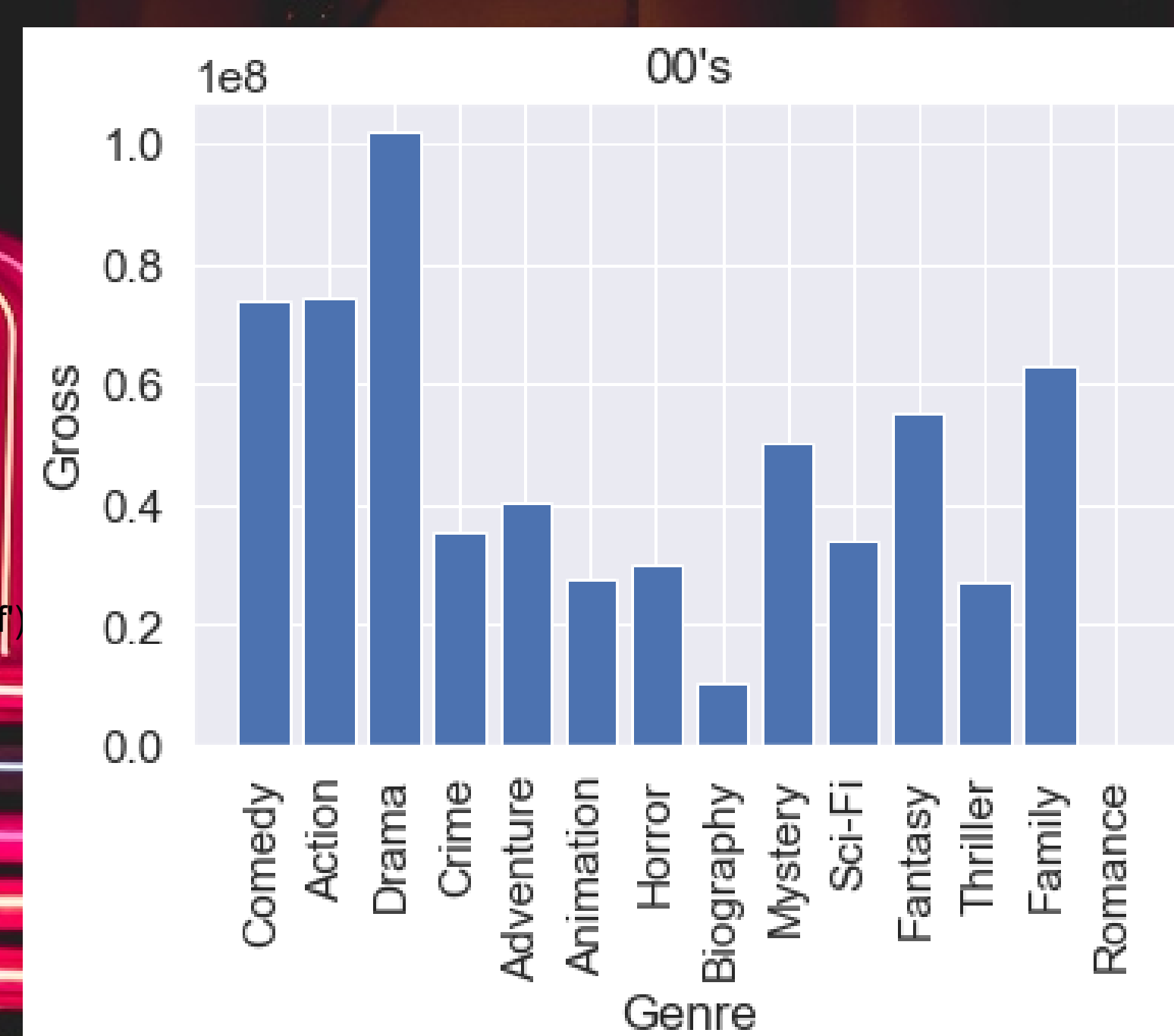
INTRODUCTION

Movies are interesting forms of media and the performance of each of them varies wildly. The aim of this project was to try and predict a movie's box office and find the key features affecting the box office the most and analyzing the results.



DATA

We used data from Kaggle which had 6820 movies from 1986-2016 with each year having 220 movies. The data had been collected from IMDb. Some of the movies were missing the budget value so we filled in about 600 missing budgets by using the Wikipedia API. We ended up dropping the movies with no budget.



PREDICTING AND DATA

ANALYSIS

For predicting the box office we used the Random Forrest Classifier. We used the actual gross of the predicted movies to get the difference between each movie's predicted and actual value. We also compared different features with each other to try and find correlations.

RESULTS

We found that the key features of movie's box office are budget and the number of votes. We also found that the bigger the movie's gross is, the bigger the error between real and predicted gross was. Overall, the prediction didn't seem to produce any reliable results. In hindsight this seems quite obvious, considering the nature of movies and the complexity of their production.