**Project Name:**

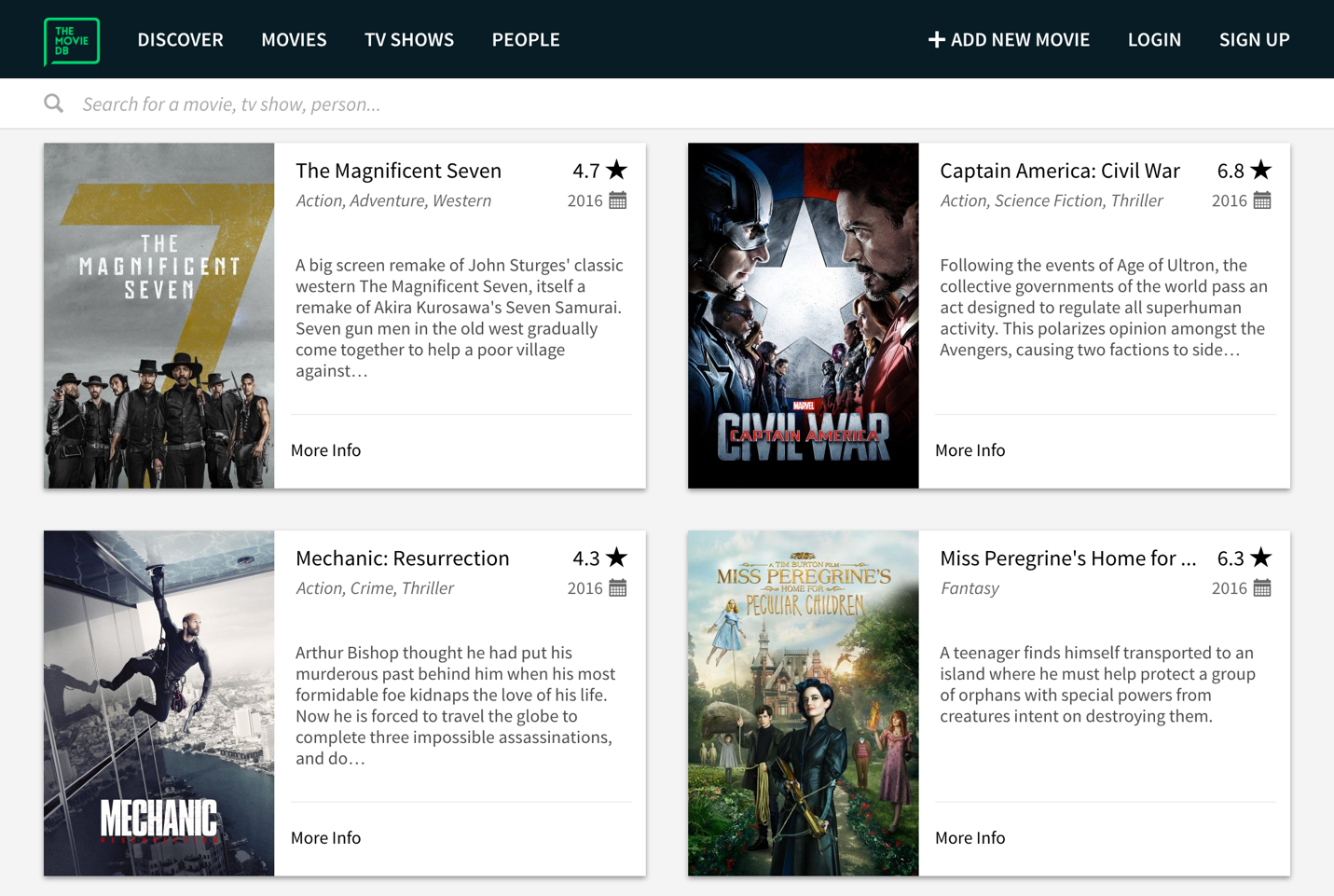
Prediction of Movie Box Office

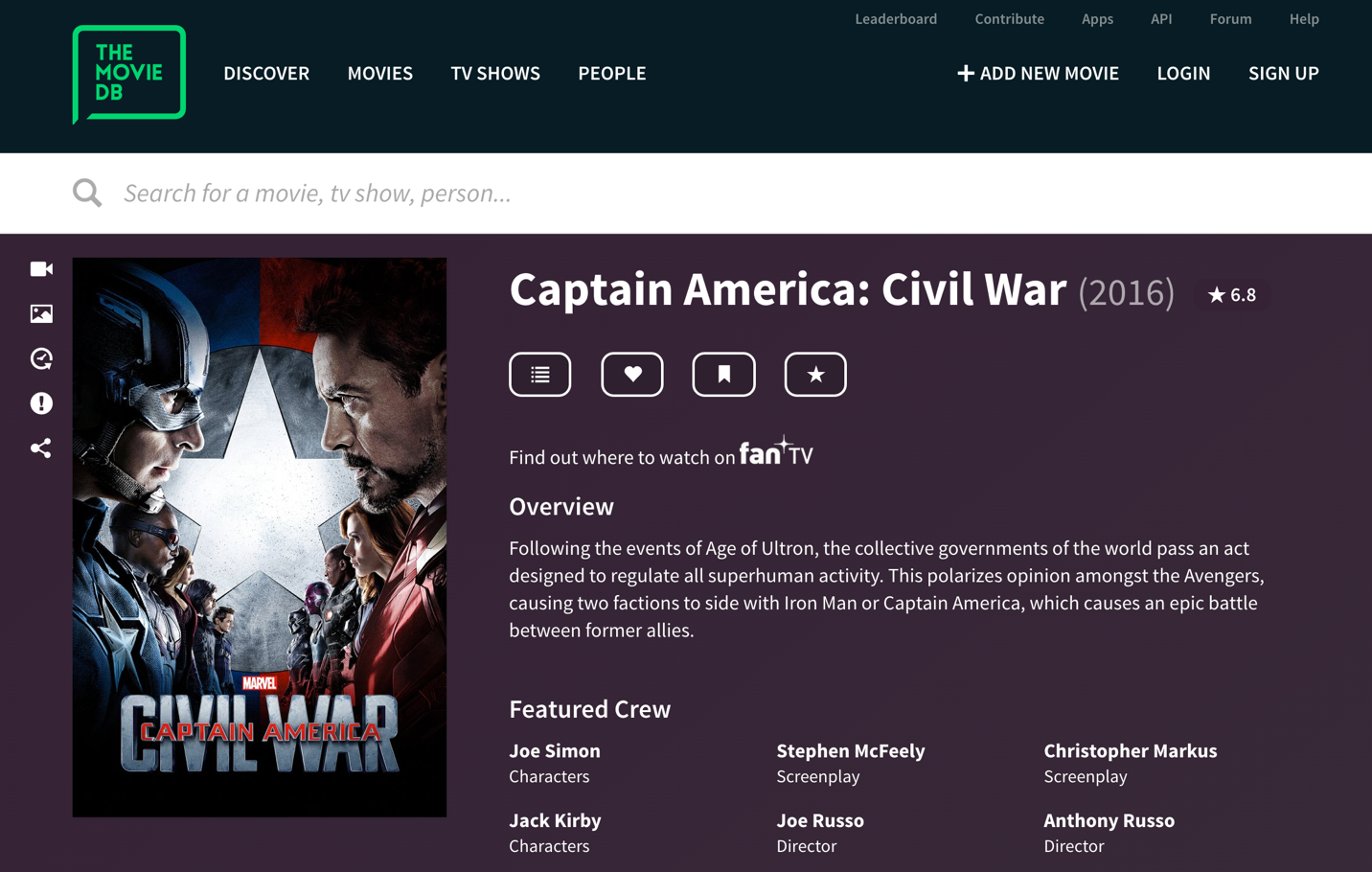
**Objective:**

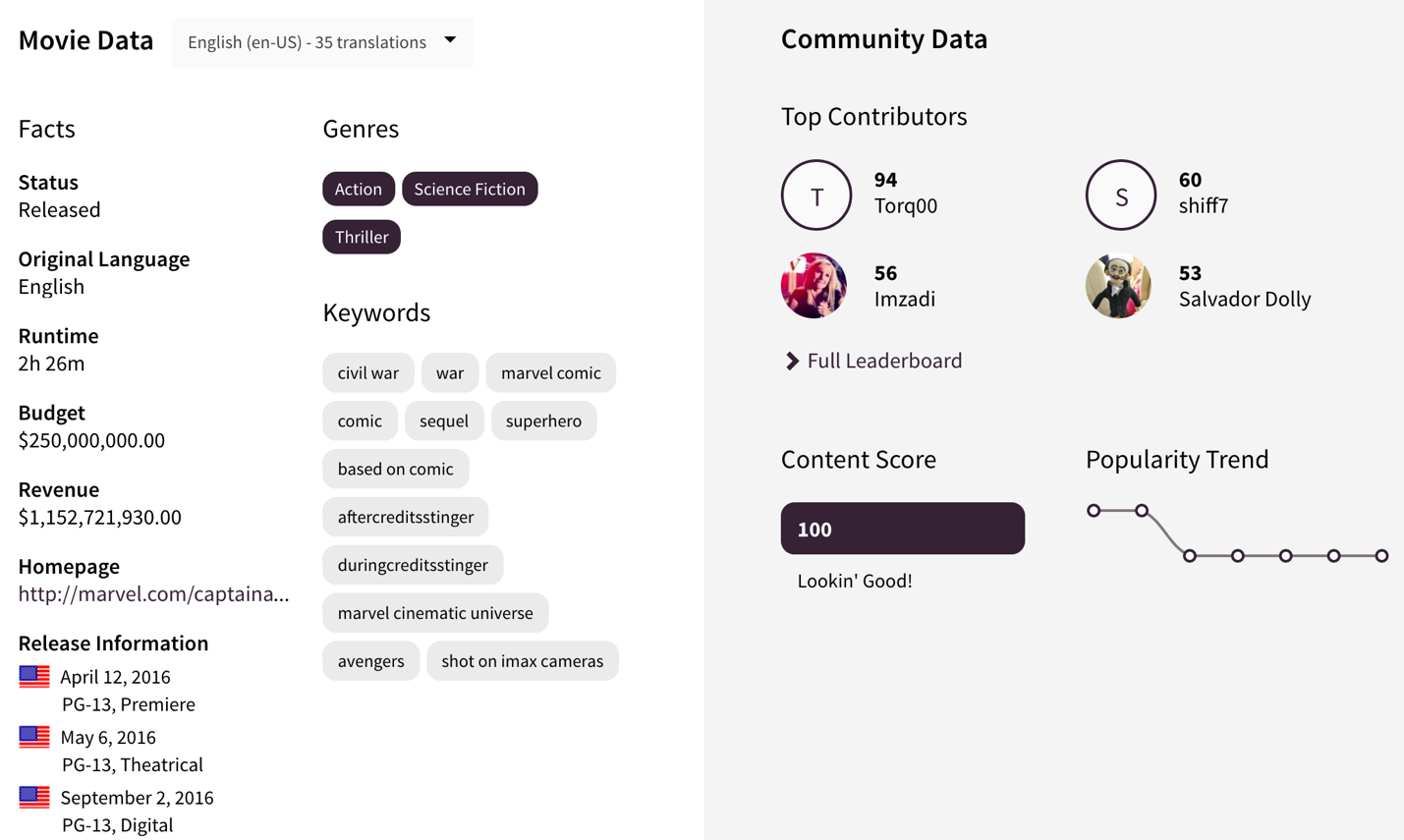
**Data Collection:**

1. Scraping 16677 movies from TMDB , 19453 movies from Boxofficemojo

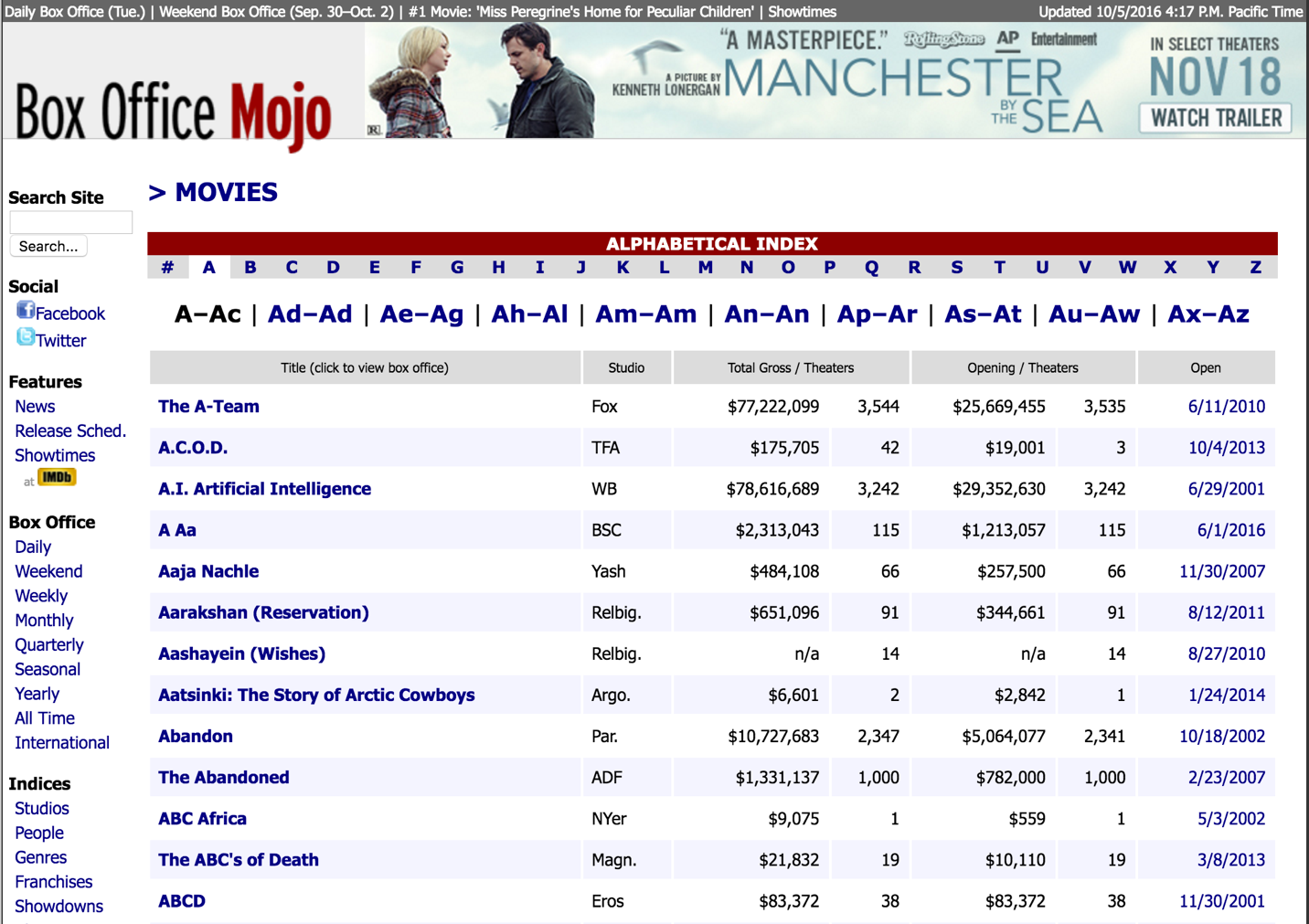
a. https://www.themoviedb.org/movie?page=1&language=en







b. <http://www.boxofficemojo.com/movies/alphabetical.htm?letter=A&p=.htm>



2. Kaggle

<https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset>

**3. EDA**

3.1 Scatter Plot & Box Plot

3.2 EDA of Identifying Important Features

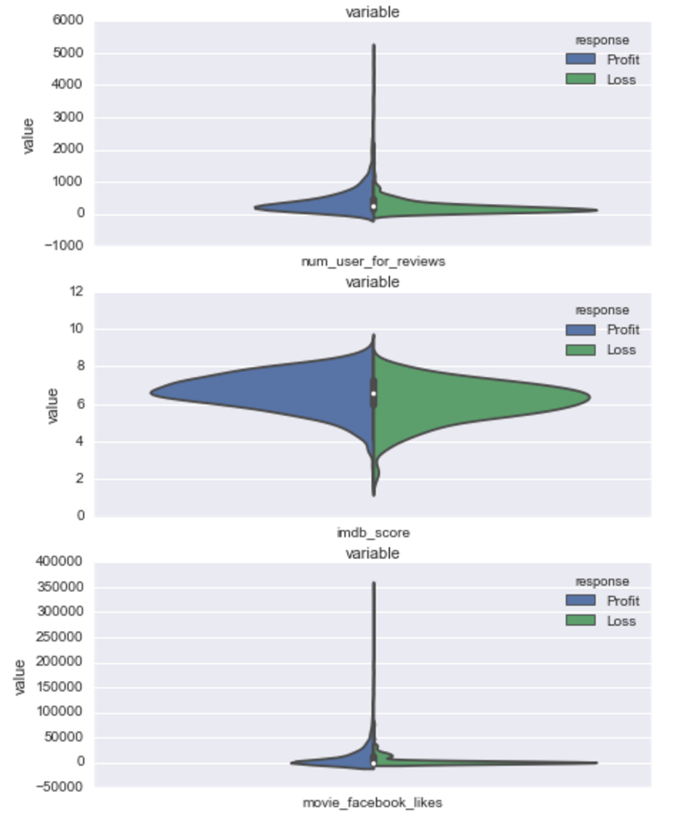
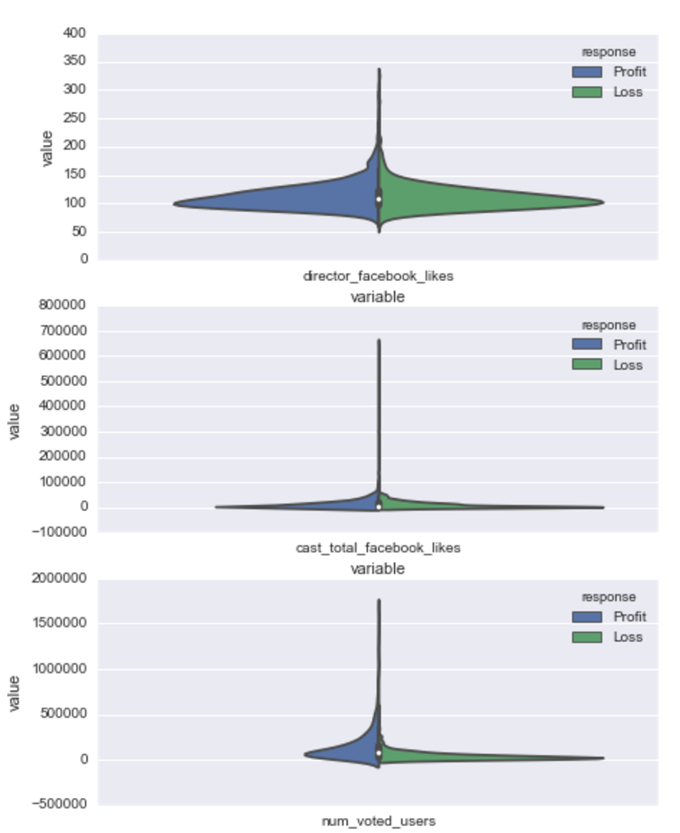
We develop the movie data into two parts, the first part is the movies that are profitable and the second is movies that lose money.

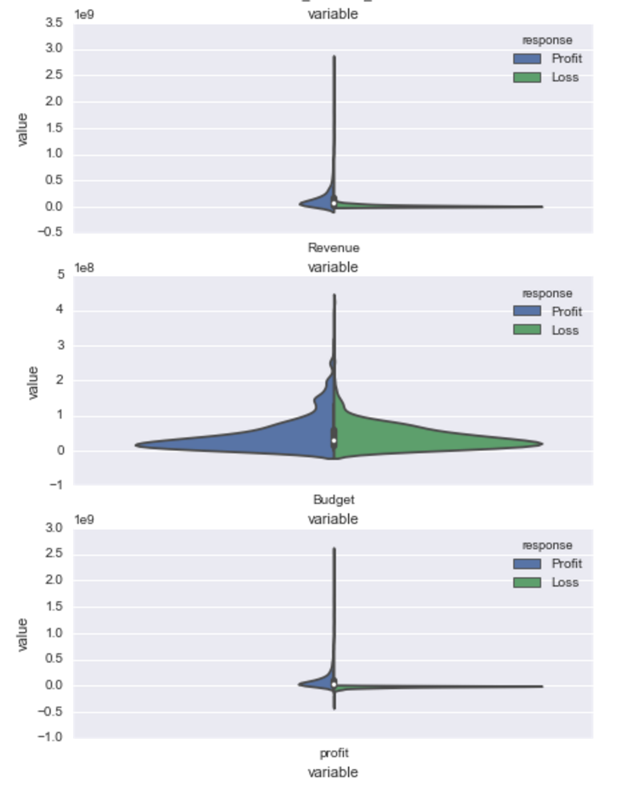
The analysis in this part reveals that the positive and negative samples differ significantly in their correlation structure.

3.2.1 Univariate Characteristics

In order to understand better the predictive power of single features, we compare the univariate distributions of the most important features.

By drawing violin plots, we see that for many of the distribution there is no significant difference between the positive and negative samples.



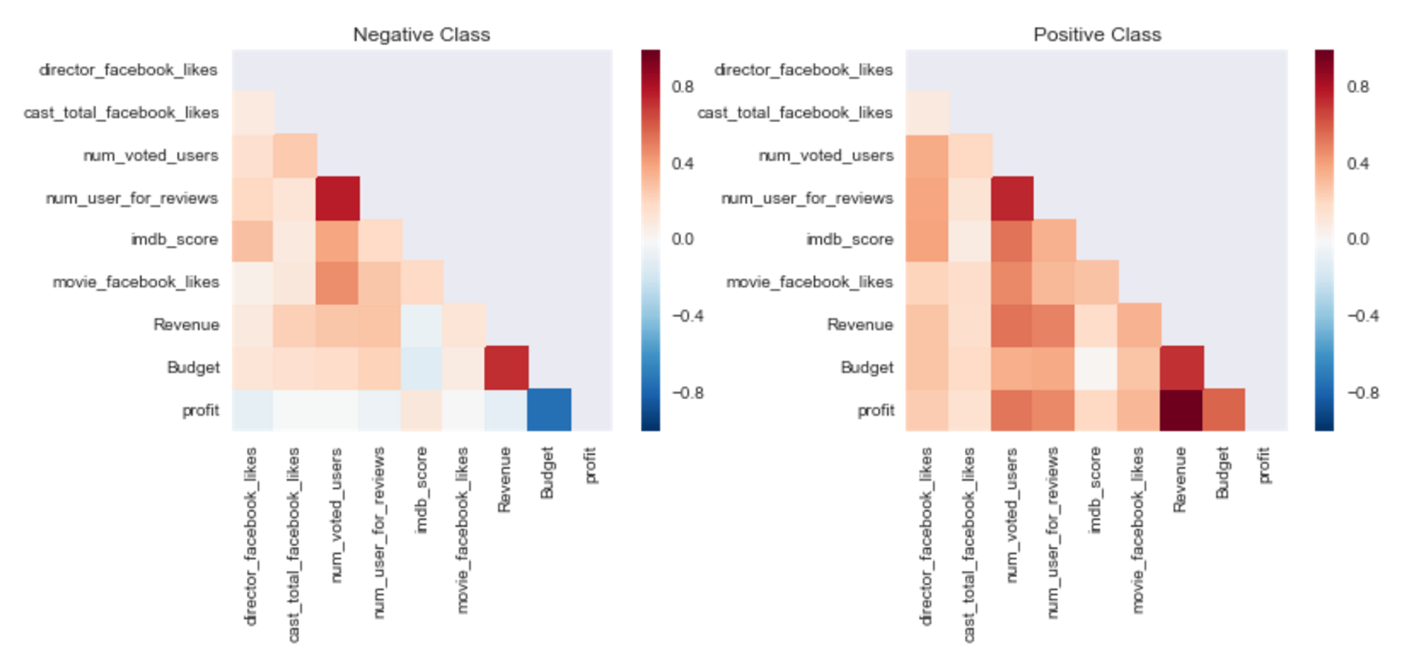


3.2.2 Correlation Analysis

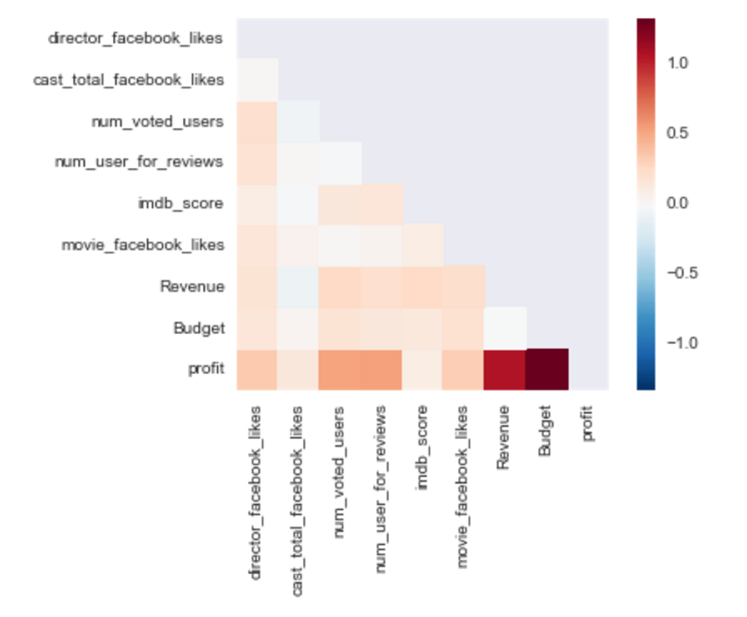
a. Correlation Matrix

b. Heat map

In the previous section we have seen differences between negative and positive samples for univariate characteristics. We go down the rabbit hole a little further and analyze covariances for the negative and positive samples separately.



The difference between the two matrices is sparse except for three specific feature combinations.



**4. Time Series Model**

**5. Cluster Analysis of Actors**

a. Data

b. Method

c. Result(Visu)

**6. Decision Tree Model**

a. Data: Data Merged (Kaggle, TMDB Studio No NAs) (2409 obs.)

b. Method: use studios, year, mean revenue and budget to predict revenue (9 levels)

Training size: 2300; Testing Size: 100

c. Result: 0.65 Accuracy

**7. Random Forest**

a. Data

b. Method

c. Result(Visu)

**8. Neural Network**

a. Data

b. Method

c. Result(Visu)