Data analyze.

1. Strong correlation between revenue and budget (which makes total sense) Also strong correlation between revenue and budget and votes count.

2. Hypothesis: old films are rated higher than new, due to pleasant nostalgic feelings of viewers

3. Next step calculate all possibility average score of movies by year.

4. ROI (it is our gross divide to budget) of genres could be different, and we will calculate it by years, by country and by votes.

5. Calculated cinematography spending by countries log scale.

6. In conclusion, calculated budget and revenue on average every movie by filmmaking companies.

Model creating.

repo: https://github.com/Kasterov/BestMovieAnalyzer.git

1. Successful data analyzing and as a consequence – possibility to create model due to the some features that will be proven to have an impact on evaluation.

2. Cleaning data from information that is corrupted and will not bring any benefit during model training, or even harm it.

3. One-hot-encoding several features so that our model can use all possible data and its numeric form if possible. This process will have to first go through two columns of our dataset.

4. Create three columns from one-hot-encoded rating column according certificate table that was described in business understanding part.

5. Delete features that can not be used for teaching model (nominal data general).

6. Create our model due to the new dataset that is clean and prepared now.

7. Calculate RMSE and figure out is it satisfies our requirements.

8. If requirements are not met, then find way to decrease RMSE (maybe cross validation).

9. Create function for calculation score for supposed movie with different features.