

Data science project

principles of Data Science

Danah Alotaibi-Latifah Allaboun
Reem Almusharraf-Deema Alresheed

Project Poster

May 2023

TMDB Movies

Introduction

TMDB (The Movie Database) is a popular online database that provides information about movies and TV shows around the world. That got us thinking about which language is the most popular for films. We identified the actionable insights and objectives that can be driven from the collected data. By using data science techniques and a classifier model, we came to this conclusion.



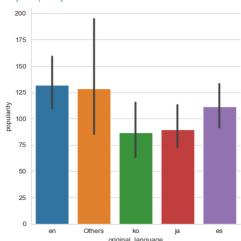
Objectives

- 1-Notice if the popularity of top-rated movies affects the chance to be from a particular language.
- 2-find out the most frequent language from the top-rated movies.
- 3-find out the language that got the highest vote count.
- 4-find out if movies with the highest vote count have the highest popularity.
- 5-find out wich language got the highest rating.

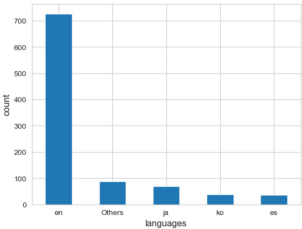
Data Collection

We collect our dataset through TMDB API, we connect to TMDB to retrieve Top rated movies on TMDB website that presents the most movies preferred by users so we can study these movie' features that made them appear in the most popular movies. We retrieved 980 movies data and fill them into a data frame and apply some analysis and visualization techniques.

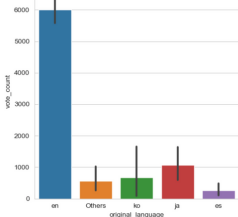
Data analysis



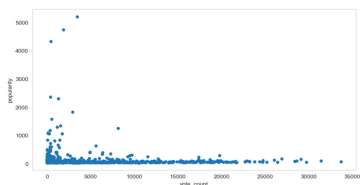
english movies have the highest popularity



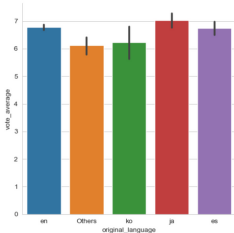
english movies have the highest count



english movies have the highest vote count



their is a negative relationship between the vote_count and the popularity.



Japanese movies have the heights rating average

Findings

- The accuracy is almost 73% from the confusion matrix and classifier model. so the result is accurate and has a good prediction of the original language
- We find out that English movies have the highest popularity and vote count.
- We notice that most of the movies are in English language.
- We notice that there is a negative relationship between the vote_count and the popularity , high vote doesn't imply high popularity.
- We find out that Japanese movies have the highest rating average.

Conclusion

At the end, our dataset contains knowledge about the most popular movies based on the original language. Briefly, our insights about data analysis and visualization that we applied to the dataset can be shown as follows:
The website we studied has good prediction results, and we visualize them using different chart types. We notice that the English language was the most common type that predicted more successfully than the other languages, followed by the Japanese language, so we conclude that if you want to make your film popular and have a high rating, make it in the English language.