Predict Movie Licensing

Watch-It is a media services company whose primary business is its subscription-based streaming service offering access to a library of films and television programs.
Watch-It has a state-of-the-art content team which actively scouts and recommends content that will attract viewers. The more people watch a particular title, the more revenue is generated for Watch-It from advertisements and brand integrations. Management wants to know which titles will generate more revenue so that they can license them.

Data files

- train.csv
- test.csv
- sample_submission.csv

Name	**Description**
title	Title of the movie
country	Countries in which movie was released
genres	Movie Genres (Action ,Adventure, Comedy etc.)
language	Languages in which movie was released
writer_count	Number of writers of the movie
title_adaption	Is movie original screenplay or adapted
censor_rating	Release rating given to the movie (R /PG-13/PG/NR/UR/G)
release_date	Date when movie was released
runtime	Movie runtime
dvd_release_date	Date of release of DVD for sale
users_votes	Number of users who voted for this movie to be included in Watch-It library
comments	Number of comments on movie trailer
likes	Number of likes on movie trailer
overall_views	Number of views on movie trailer
dislikes	Number of dislikes on movie trailer
ratings_imdb	Rating given to movie on IMDB
ratings_tomatoes	Rating given to movie on Rotten tomatoes.
ratings_metacritic	Rating given to movie on Metacritic etc.
special_award	Number of awards nominations/winnings in BAFTA, Oscar or Golden Globe.
awards_win	Awards won by the movie
awards_nomination	Number of awards nominations
revenue_category	Revenue Category (High/Low)

Problem
Perform an analysis of the given data to determine how different features are related to revenue. Build a machine learning model that can predict the revenue_ category, this will help management know what titles would be suitable for licensing.

For each record in the test set (test.csv), predict the value of the revenue_ category variable (Low or High). Submit a CSV file with a header row plus each of the test entries, each on its own line. The file (submissions.cv) should have exactly 2 columns:

- title
- revenue_category(contains Low or High)

Deliverables

- Well commented Jupyter notebook
- submissions.csv

Explore the data, make visualizations, and generate new features if required. Make appropriate plots, annotate the notebook with markdowns and explain necessary inferences. A person should be able to read the notebook and understand the steps you took as well as the reasoning behind them.

Evaluation Metric

a) Accuracy metric will be used for evaluating the model performance.
b) Make use of another metric to evaluate model performance with the reason for using it.