Nareg Kedjejian

Predicting Opening Week Member Hours for New Netflix Titles

**Business Task:**

The goal of this project is to predict opening week viewership for new Netflix Titles that will be released in 2025 based on data from 2019 to 2024. A regularized linear regression machine learning model will be developed and evaluated. A successful model will beat simple heuristics based on average opening week viewership by genre. Feature importance will be assessed to determine what factors impact opening week viewership most.

**Dataset:**
- Data was collected on 2525 titles between 2019 and 2025
- The dataset consisted of 31 features before preprocessing and 66 features after
- The target is opening week viewership in millions of views

**Data Cleaning:**
- Removed duplicate observations
- Fixed incorrect data types and made string and number formats consistent
- Used various techniques to fill missing values (iterative regression-based imputation, distribution-based imputation, random sample imputation, creating new categories)

**Feature Engineering:**
- Performed one-hot, ordinal, and cyclical encoding for categorical variables
- Performed log transformations on log-normal features
- Performed a Box-Cox transformation on the target
- Assessed multicollinearity and skew of features

**Results and Key Insights:**
- Baseline heuristic results: {'MAE': 7.889, 'RMSE': 9.7481, 'MAPE': 0.2698}
- After cross validation and hyperparameter tuning, the best model was determined to be Ridge with an alpha of 412.245 and no polynomial features or interactions
- Ridge regression results: {'MAE': 6.3293, 'RMSE': 8.1248, 'MAPE': 0.2203}
- The model does not overfit or underfit, and the model generalizes well
- The five most important features that impact opening week viewership are marketing spend, whether or not the title is based on a known IP, the number of competition titles released within two weeks, the homepage hero share of a title, and whether or not the primary genre was drama.
- The number of competition titles and whether or not the primary genre was drama have a negative correlation with opening week viewership

**Recommendations:**
- Promote new titles by spending more on marketing and increasing homepage hero time
- Reduce the number of competing titles for new releases
- Audience like titles based on a known IP and does not like dramas