

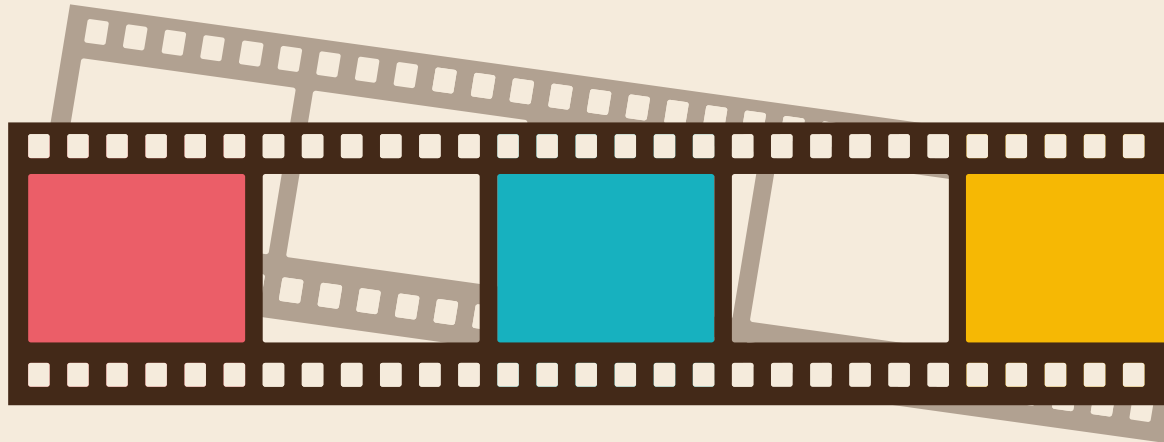
IMDB MOVIE PREDICTIONS

NOLAN ARENDT



“To make a film is easy; to make a good film is war. To make a very good film is a miracle.”

—ALEJANDRO GONZALEZ INARRITU





01. PROJECT OVERVIEW
02. DATA CLEANING
03. MODELING PROCESS
04. EXPLORATORY DATA ANALYSIS
05. FUTURE ANALYSIS

WORKFLOW

01. PROJECT OVERVIEW

Predict a movies score and revenue.

Incorporate actress, actor, and director scores into predictions.

Identify highest scoring genres and highest grossing genres.

Identify how correlated movie score and revenue are.



02. DATA CLEANING

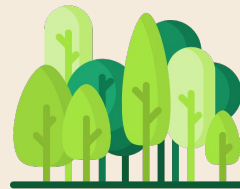
- 1. Acquire data.**
- 2. Merge datasets.**
- 3. Handle null values.**
- 4. Dummify genre and role.**
- 5. Separate actresses, actors, and directors.**



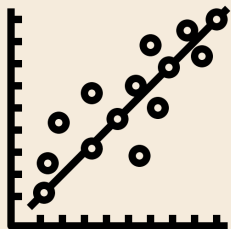
03. MODELING PROCESS



Extra Trees



**Random
Forests**



**Linear
Regression**



XGBoost

MODEL SELECTION IMDB SCORE

IMDB SCORE: LINEAR REGRESSION

TRAINING SCORE: 0.760

BASELINE: 0.9867

TESTING SCORE: 0.767

RMSE: 0.490

COPY_X : TRUE

FIT_INTERCEPT : TRUE

NORMALIZE : TRUE

MODEL SELECTION REVENUE

REVENUE : EXTRA TREES

TRAINING SCORE: 0.829

BASELINE: \$187,909,923

TESTING SCORE: 0.766

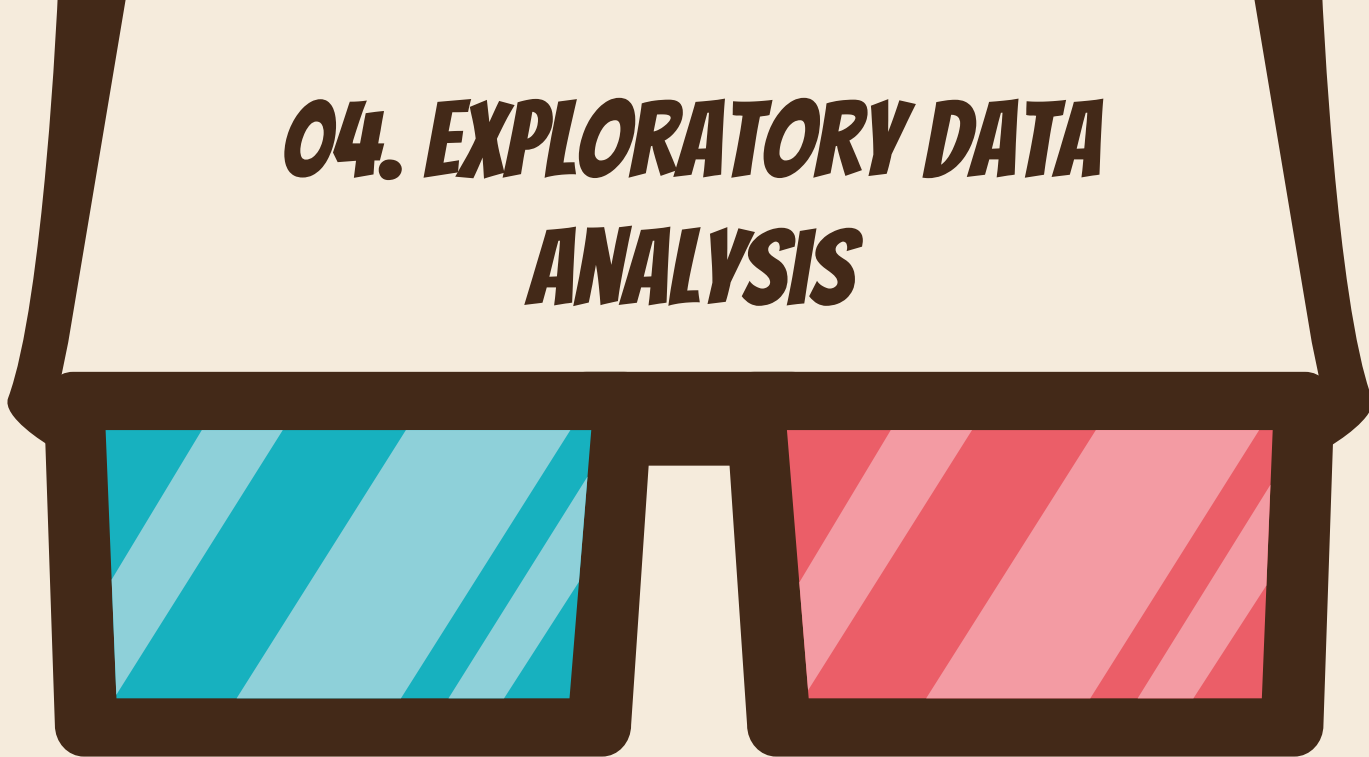
RMSE: \$71,109,157

MAX_DEPTH : 8

MIN_SAMPLES_SPLIT : 3

N_ESTIMATORS : 100

04. EXPLORATORY DATA ANALYSIS



TOP FIVE IMDB SCORES

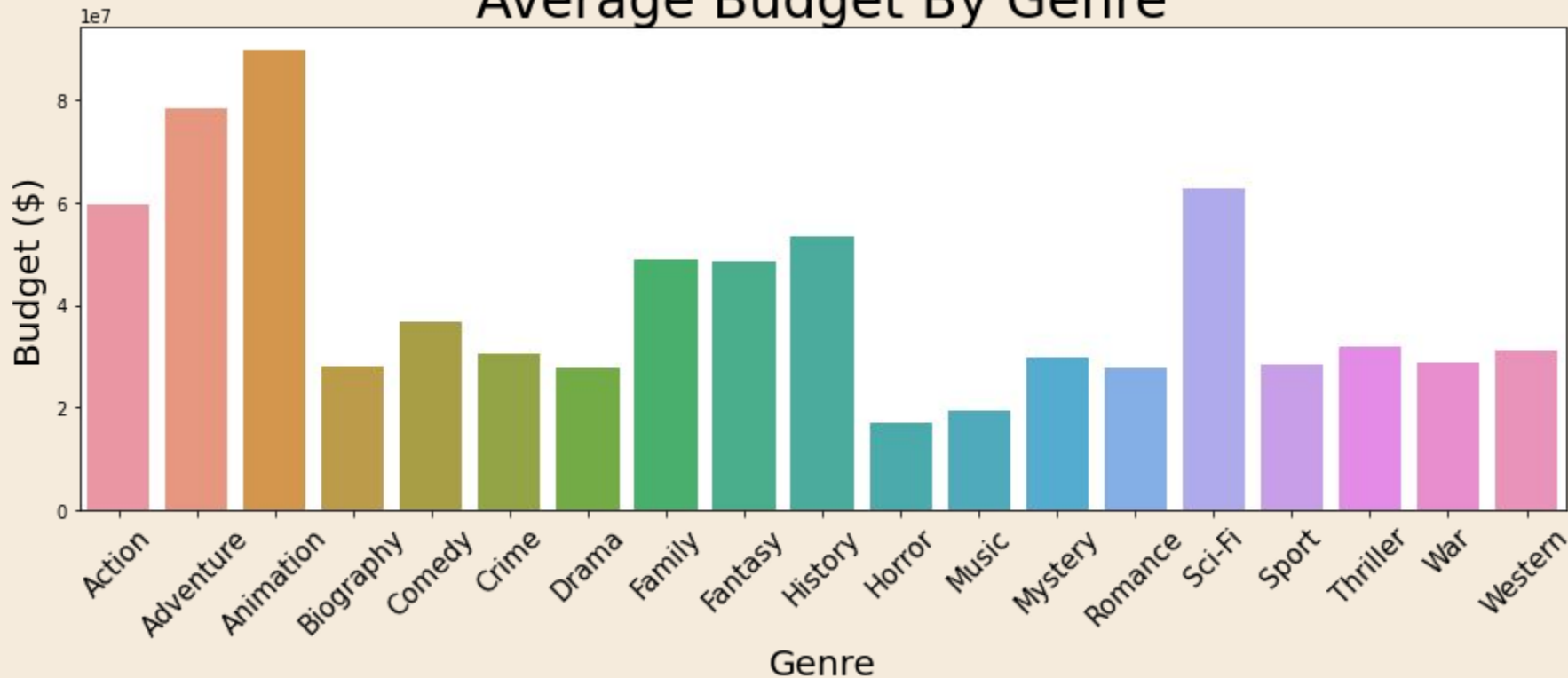
<u>Movie Title</u>	<u>Actual Score</u>	<u>Predicted Score</u>	<u>Difference</u>
Shawshank Redemption	9.30	10.00	↑ 0.70
The Godfather	9.20	9.084	↓ 0.116
The Godfather Part II	9.00	8.245	↓ 0.755
Schindler's List	8.90	8.331	↓ 0.569
Forrest Gump	8.80	8.825	↑ 0.025

TOP FIVE MOVIE REVENUES

<u>Movie Title</u>	<u>Actual Revenue</u>	<u>Predicted Revenue</u>	<u>Difference</u>
Avatar	\$2,781,505,847	\$2,127,454,351	\$654,051,496
The Avengers	\$1,519,557,910	\$1,084,326,906	\$435,231,004
Jurassic World	\$1,513,528,810	\$1,230,437,950	\$283,090,860
Avengers: Age of Ultron	\$1,405,035,767	\$1,163,597,611	\$241,438,156
Minions	\$1,156,730,962	\$909,385,604	\$247,345,358

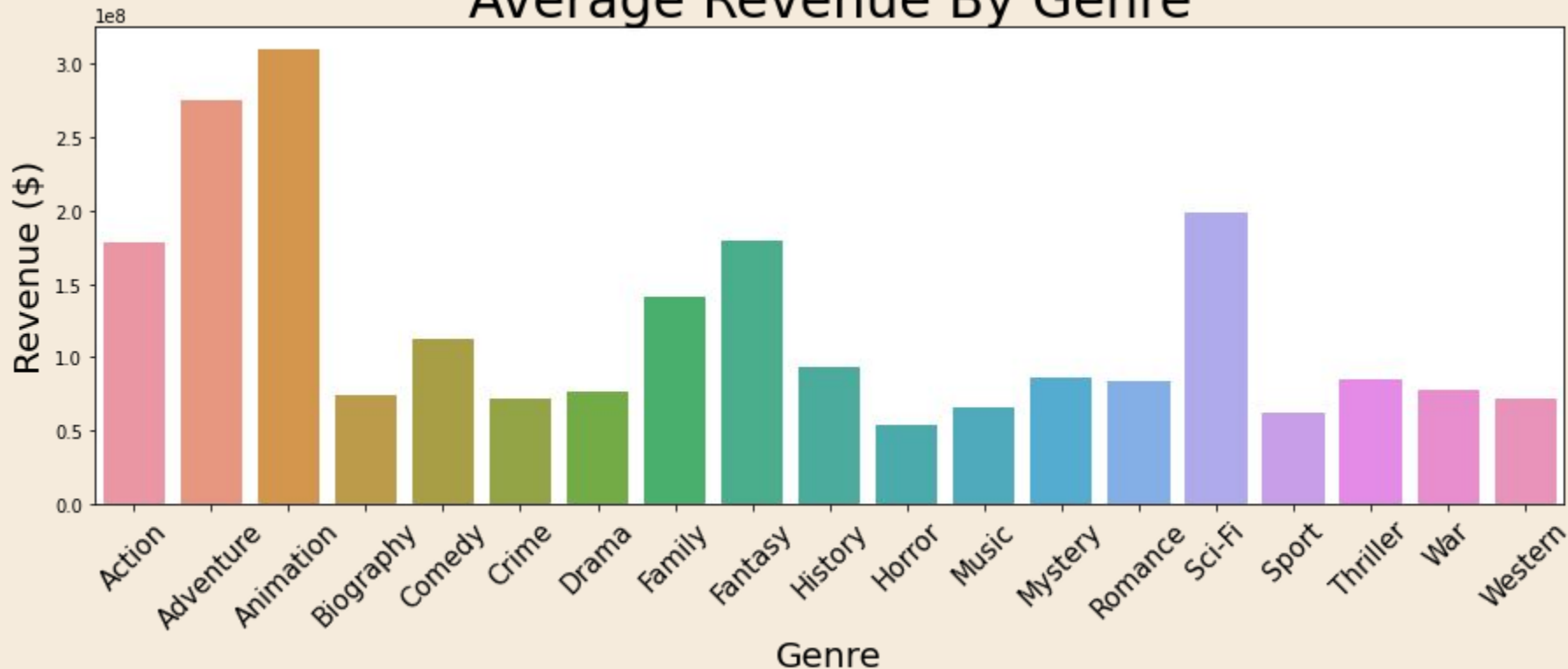
AVERAGE BUDGET

Average Budget By Genre

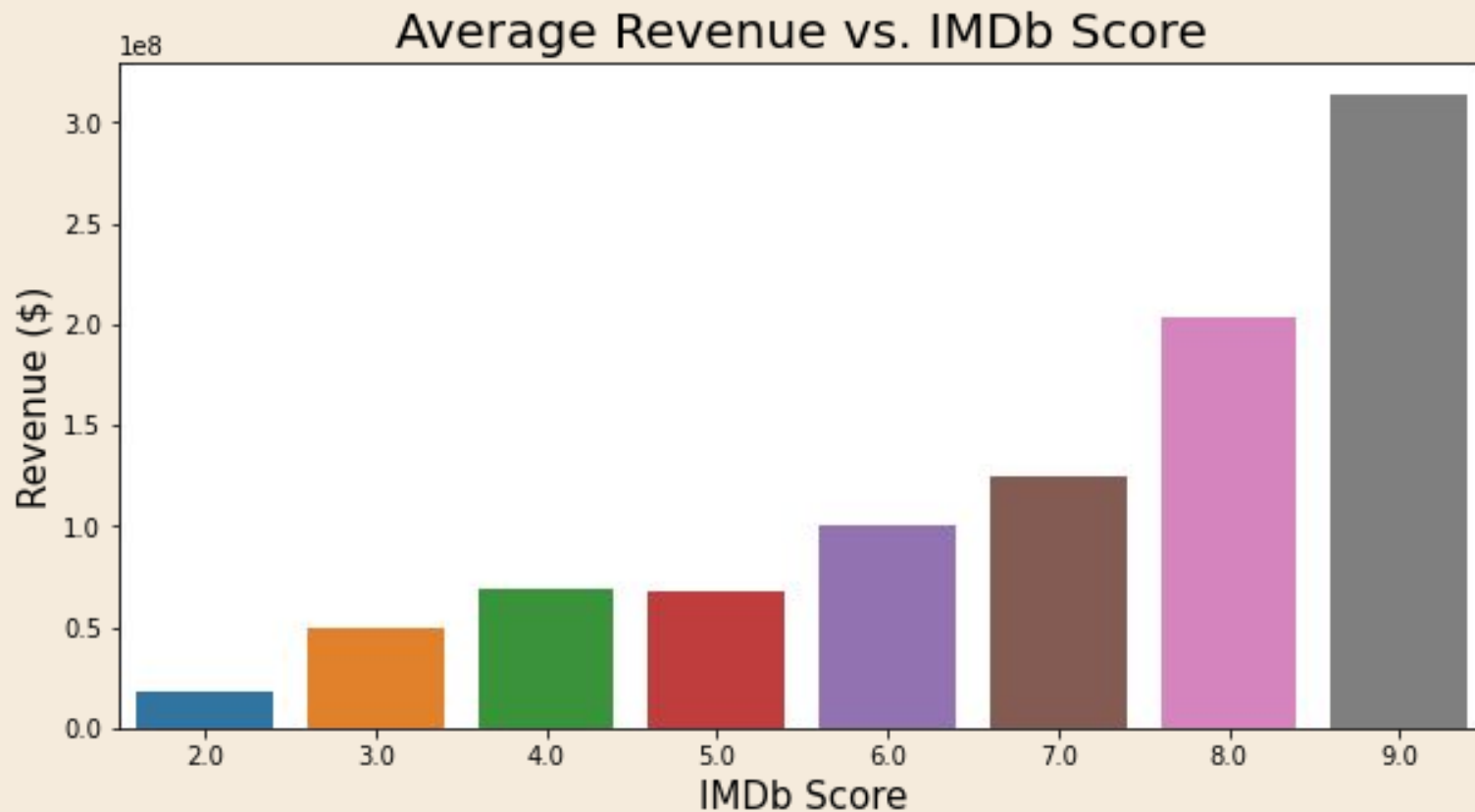


AVERAGE REVENUE

Average Revenue By Genre

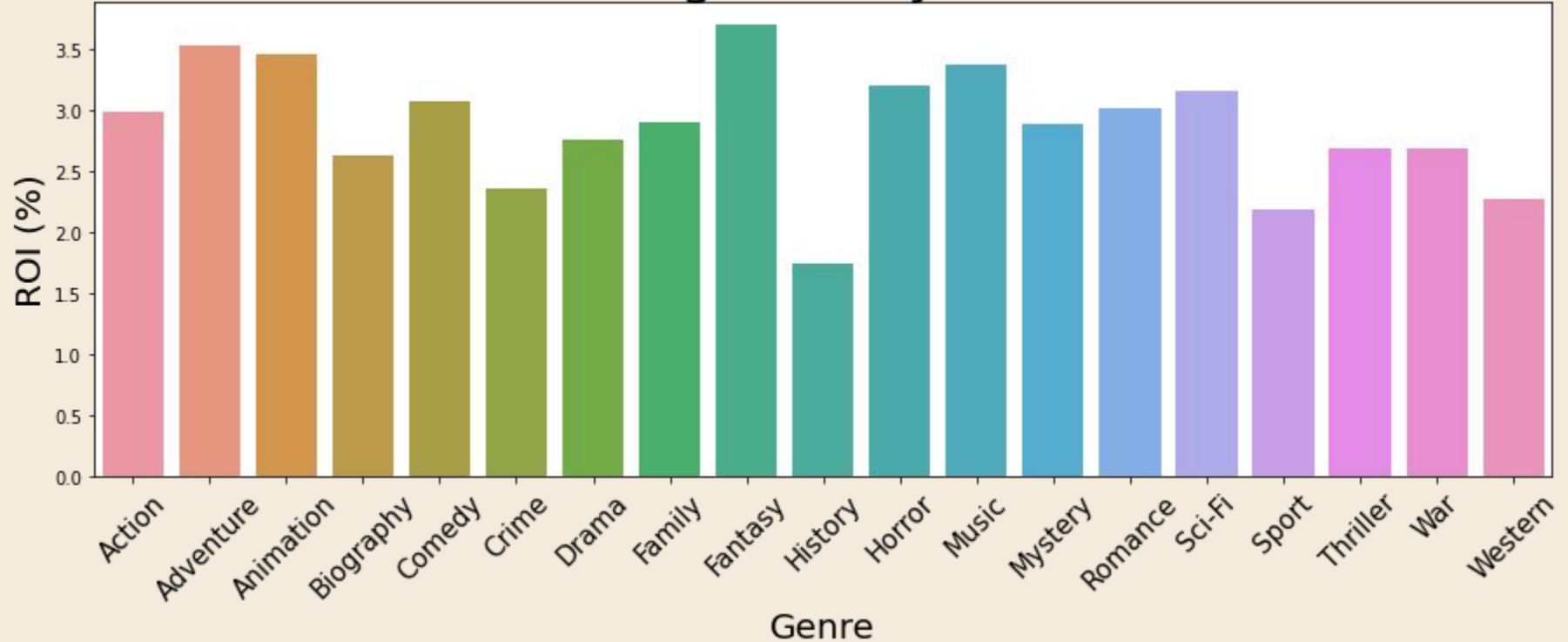


AVERAGE REVENUE



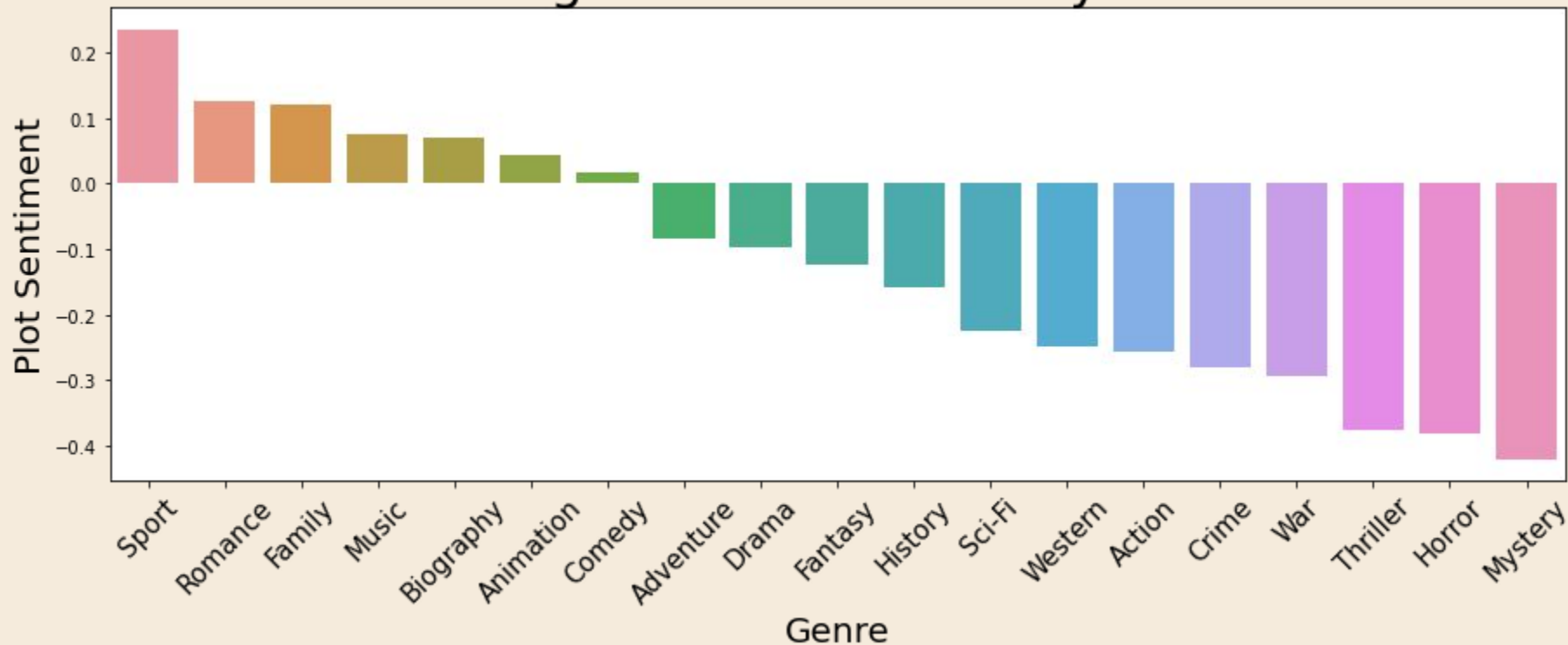
AVERAGE ROI

Average ROI By Genre

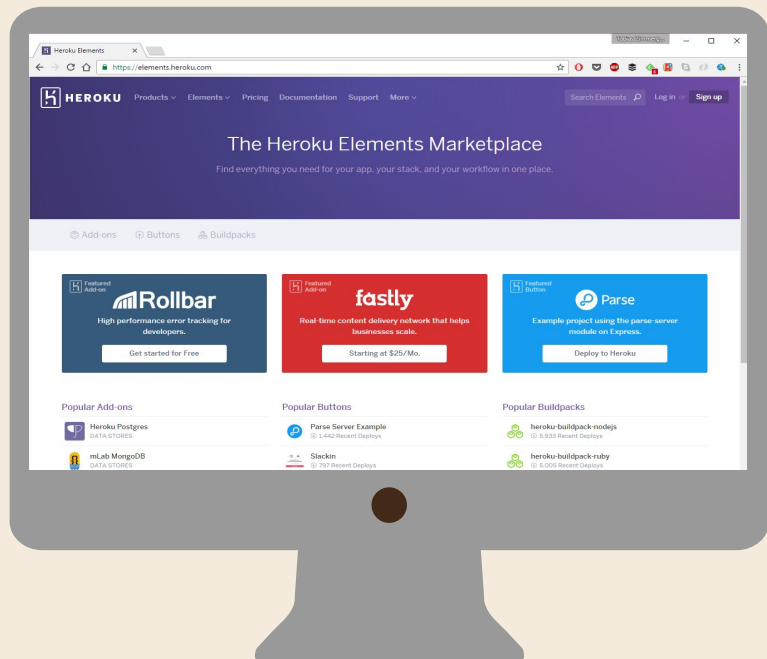


PLOT SENTIMENT

Average Plot Sentiment By Genre



05. FUTURE ANALYSIS



- ★ Heroku Application
- ★ Predict future movie scores
- ★ Feature engineer more roles (writer, producer, etc.)



THANK YOU!

nolanarendt10@gmail.com

[linkedin.com/in/nolanarendt/](https://www.linkedin.com/in/nolanarendt/)
github.com/nolanarendt

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

Please keep this slide for attribution.

