Oscar Nominations

Team Members/ Responsibilities

- Jake Littman - Machine Learning model
- Oneil Anderson - Technologies/Presentation
- Breionna Turner- Database
- Elena Rivera - Presentation/ Dashboard
- Christine Ibrahim Puri - Github Setup/Dashboard

Overview

Our project will analyze movies throughout history. Using the datasets at our disposal, we will identify the variables that relate to oscar performance of actors/actresses/directors and writers. In our project, we will use the tools that we have learned throughout the bootcamp to create a full analysis on the data provided from the Oscar nominations. We will also address the questions with this analysis.

Questions to Address

- Why do certain genres outperform others?
- What factors influence movie nominations?
- What genres and production companies have the highest probability of being nominated for an Oscar at the next academy award?

Technologies

Data Cleaning and Analysis

We will use Python to Extract, Transform and Load and perform meaning analysis over the data. We will be using the ImbalancedLearn, Pandas, Numpy, and SciKitLearn packages.

Database Storage

We will use PostgreSQL as our relational database storage.

Machine Learning

Our machine learning models will leverage SciKitLearn and ImbalancedLearn Python packages. Due to the oscar imbalance, we will use random over/undersampling, SMOTE, Cluster Centroid Undersampling, and/or SMOTEEN.

Dashboard for Presentation

We will leverage D3 Javascript/Tableau to create our dashboard to present our findings.

Database

Software used: Jupyter Notebook [Pandas], PostgreSQL

Cleansing the datasets:
The database has come together from bringing in the *movie_metadata.csv* and *oscars_movie*.csv datasets.

The data in both were cleaned via regex processing and taking out null values. In addition, *release_year* and *nominated* columns were added to the *movie_metadata*.

Joining the datasets:
From there, the data was exported into a Postgres database and then the datasets were combined into the combination_table.csv. The datasets were combined on the Oscar's '*film*' column with the Movie's '*title*' column.

The columns in the combined dataset are as follows: *budget, release_year, revenue, runtime, title, genres, nominated, winner, category.*

Database

Machine Learning

- Prelim data preprocessing:
- Feature Engineering and Selection:
- Training/Test data sets:
- Models: Logistic REgression

Machine Learning

Dashboard Integration

Software Used: Tableau

Budget, Revenue, Nominations

- We will use three bar charts side-by-side to display the budgets, revenues, and nominations for each genre in one view.

Average Budget | Oscar Nominations

- A scatter plot could be used to display the relationship between the average budget of films and the number of nominations films received to view if there is any significant positive or negative relationship between the two variables

- A bar chart could also be used along with Tableau Calculation field to form groups of "High", "Medium", and "Low" with regard to the average spending. For example, a budget of 300k - 500k would be Low, 500k - 1M would be medium, and >= 1M would be High.

Genre | Category

- Amongst the categories that receive Oscar awards (Actor, Actress, Art direction, Directing etc), we will display the genres that receive the most nominations in each category.

- The columns of the bar chart would be divided by the categories, and the bars would be color-coded according to genres, with the number of nominations on the y-axis

Interactivity

- Creating sets that group a range of years, for example 1990- 2000, 2001 - 2011, etc would create additional dimensions that would allow a user filter by years, and observe changes in relationships between genre and category over the years.

Dashboard Integration

Example of side-by-side bar charts showing the Budget, Revenue, and Nominations for each genre

Bar: Budget, Revenue, Nominations

Genres	Budget	Revenue	Nominations
Action	62,674,765,555	216,016,045,249	493
Adventure	45,122,499,597	176,435,832,576	418
Animation	13,630,712,588	58,955,330,178	93
Comedy	28,821,895,957	110,209,347,974	442
Crime	7,682,251,765	22,863,823,943	194
Documentary	188,569,415	797,417,515	13
Drama	62,537,811,325	258,662,199,531	1,886
Family	5,347,743,149	30,265,867,890	30
Fantasy	10,132,673,984	31,935,968,169	77
Foreign	8,397,361	25,592,951	0
History	2,676,147,083	10,293,234,169	71
Horror	5,087,307,863	20,624,730,349	58
Music	963,893,015	2,158,563,154	28
Mystery	1,796,644,076	8,286,495,933	38
Romance	4,086,403,630	19,410,098,332	84
Science	7,068,943,003	28,986,318,409	59
Thriller	8,105,783,706	24,739,688,921	94
TV	5,000,000	42,000,000	0
War	1,784,716,356	5,436,844,054	37
Western	3,218,456,783	9,069,030,189	51