“BEST IN SHOW” ACADEMY AWARD PREDICTIONS

Presented By:

Cayla Valentine, Joanna Santana, Kelly McDonald, Olivia Ramsey

OBJECTIVE

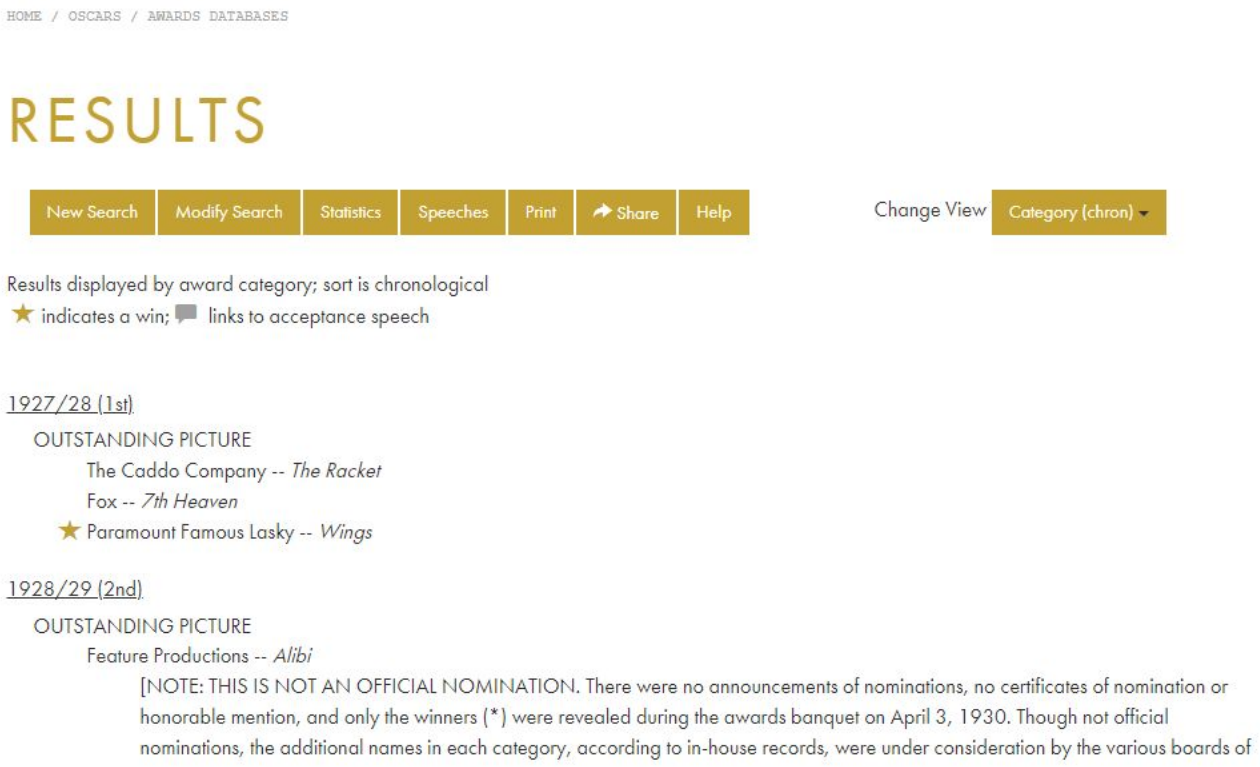
To create a model that would be able to predict movies that have the greatest chance of winning Best Picture at the Academy Awards

DATA SOURCES

The Oscar Awards have their own database with data from 1927-2023.

To optimize our time…we opted for a CSV file of these results from Kaggle.

To get more data, we ran each Best Picture nominee + winner through the OMDB API service.



HOME / OSCARS / AWARDS DATABASES

RESULTS

New Search | Modify Search | Statistics | Speeches | Print | Share | Help

Change View: Category (chron)

Results displayed by award category; sort is chronological
★ indicates a win; 💬 links to acceptance speech

1927/28 (1st)

OUTSTANDING PICTURE

The Caddo Company -- *The Racket*

Fox -- *7th Heaven*

★ Paramount Famous Lasky -- *Wings*

1928/29 (2nd)

OUTSTANDING PICTURE

Feature Productions -- *Alibi*

[NOTE: THIS IS NOT AN OFFICIAL NOMINATION. There were no announcements of nominations, no certificates of nomination or honorable mention, and only the winners (*) were revealed during the awards banquet on April 3, 1930. Though not official nominations, the additional names in each category, according to in-house records, were under consideration by the various boards of

DATA SOURCES

OMDB Search Result example:

Title: barbie Year: 2023 Plot: Short Response: JSON Search Reset

Request:

http://www.omdbapi.com/?t=barbie&y=2023

Response:

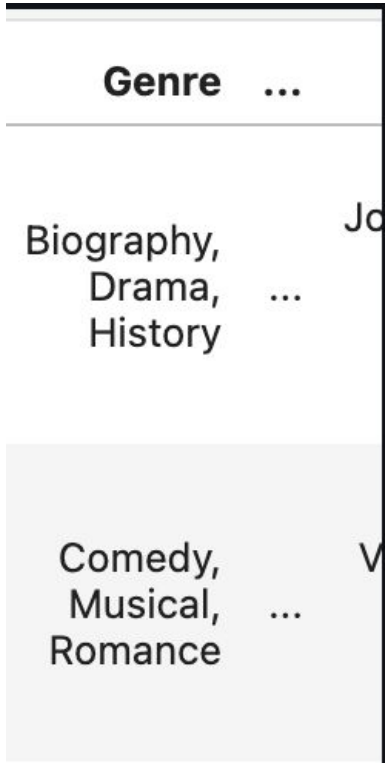
{"Title":"Barbie","Year":"2023","Rated":"PG-13","Released":"21 Jul 2023","Runtime":"114 min","Genre":"Adventure, Comedy, Fantasy","Director":"Greta Gerwig","Writer":"Greta Gerwig, Noah Baumbach","Actors":"Margot Robbie, Ryan Gosling, Issa Rae","Plot":"Barbie suffers a crisis that leads her to question her world and her existence.","Language":"English, Spanish","Country":"United States, United Kingdom","Awards":"2 wins & 1 nomination","Poster":"https://m.media-amazon.com/images/M/MV5BOWIwZGY0OTYtZjUzYy00NzRmLTg5YzgtYWMzNWQ0MmZiY2MwXkEyXkFqcGdeQXVyMTUzMTg2ODkz._V1_SX300.jpg","Ratings":[{"Source":"Internet Movie Database","Value":"7.4/10"},{"Source":"Rotten Tomatoes","Value":"88%"},{"Source":"Metacritic","Value":"80/100"}],"Metascore":"80","imdbRating":"7.4","imdbVotes":"238,040","imdbID":"tt1517268","Type":"movie","DVD":"N/A","BoxOffice":"$541,907,382","Production":"N/A","Website":"N/A","Response":"True"}

FIRST CHALLENGE

- The title of “winner” at the Academy Awards has gone by different names over the years:
  - * “Best Picture”, “Outstanding Motion Picture”, “Outstanding Picture”, “Outstanding Production”, “Best Motion Picture”
- Successfully able to filter the results to “Best Picture” and all previous aliases
- Compared our API Key with data frame

DATA CLEANING

GENRE COLUMN HAD MULTIPLE DATA PIECES



Genre	...
Biography, Drama, History	...
Comedy, Musical, Romance	...

USED THE SPLIT FUNCTION TO SEPARATE GENRES BUT...MORE CHALLENGES!

Romance	Western	...	Comedy	NaN	Fantasy	Sci-Fi	Action	Drama	Short	Film-Noir	Animation	Crime
0	0	...	0	0	0	0	0	1	0	0	0	0
1	0	...	1	0	0	0	0	0	0	0	0	0
0	0	...	0	0	0	0	0	1	0	0	0	0
1	0	...	0	0	0	0	0	1	0	0	0	0
0	0	...	0	0	0	0	0	1	0	0	0	0

DATA CLEANING CONT'D

- New challenge: NaN value was being assigned to a column
- Updated columns, successfully dropped NaN

```
1  df_1.columns
```

```
Index(['year_ceremony',       'winner',        'Rated',       'Runtime',
       'Metascore',     'imdbRating',     'imdbVotes',     'BoxOffice',
       'Thriller',         'Family',         'Short',       'Romance',
       'Fantasy',        'Mystery',        'Comedy',        'Horror',
       'Animation',          'Music',       'Western',        'Action',
             nan,        'History',       'Musical',         'Drama',
       'Biography',      'Film-Noir',         'Crime',           'War',
       'Sci-Fi',          'Sport',     'Adventure'],
      dtype='object')
```

MORE DATA CLEANING

- One aspect of data we needed to account for was some movies were duplicates because of remakes (i.e. “A Star Is Born”)

	year_film	film	Title/Year
71	1937	A Star Is Born	A Star Is Born-1937
544	2018	A Star Is Born	A Star Is Born-2018

MORE DATA CLEANING

- Pieces of data that needed to be changed into integers include Runtime, IMDB Votes, and Box Office
- The dollar sign in Box Office needed to be removed
- We also needed to change much of our data into integers

```
# Remove "," and "$" from BoxOffice column so that we can change it to an integer
def BoxOffice_cleaning(dataframe):
    dataframe["BoxOffice"] = dataframe["BoxOffice"].str.replace(",","")
    dataframe["BoxOffice"] = dataframe["BoxOffice"].str.replace("$","")
    print(dataframe["BoxOffice"].unique())
BoxOffice_cleaning(combined_df)
```

Runtime	imdbVotes	BoxOffice
151 min	98,933	$17,348,945
158 min	80,669	$6,773,650
130 min	627,105	$718,732,821
147 min	150,623	$4,608,096
104 min	33,492	$5,456,531

FIRST MODEL ATTEMPT

Model Type: Logistic regression

Data Included: Runtime, Metascore, imdbRating, imdbVotes, BoxOffice, Genre, Rating, Year of Ceremony

Confusion Matrix

	Predicted loser	Predicted winner
Actual loser	111	0
Actual winner	21	0

Accuracy Score : 0.8409090909090909
Balanced Accuracy Score : 0.5

Classification Report

	precision	recall	f1-score	support
loser	0.84	1.00	0.91	111
winner	0.00	0.00	0.00	21
accuracy			0.84	132
macro avg	0.42	0.50	0.46	132
weighted avg	0.71	0.84	0.77	132

MODEL OPTIMIZATION

- Model types: Logistic Regression + Random Forest
- Data added: Golden Globes results, Director, Producer, Country
- Data transformed: Standard Scaling on numeric columns
- Data removed: Box Office
- Data limited: Years of Ceremony (past 50 years only, 1944 - 2020 only)

BEST MODEL ATTEMPT

Method: logistic regression

Data Transformation: Standard Scaling on numeric columns.

Data added: Golden Globes results

Data limited: Years 1944-2020

Confusion Matrix

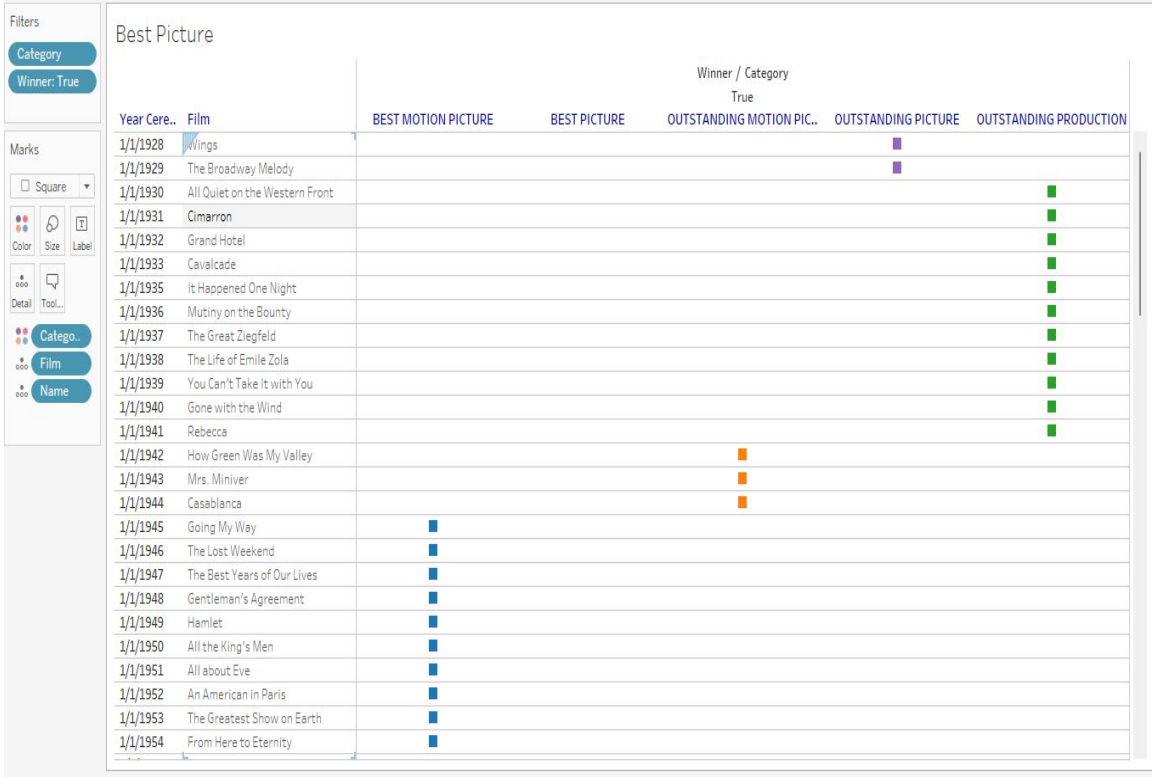
	Predicted loser	Predicted winner
Actual loser	38	3
Actual winner	9	5

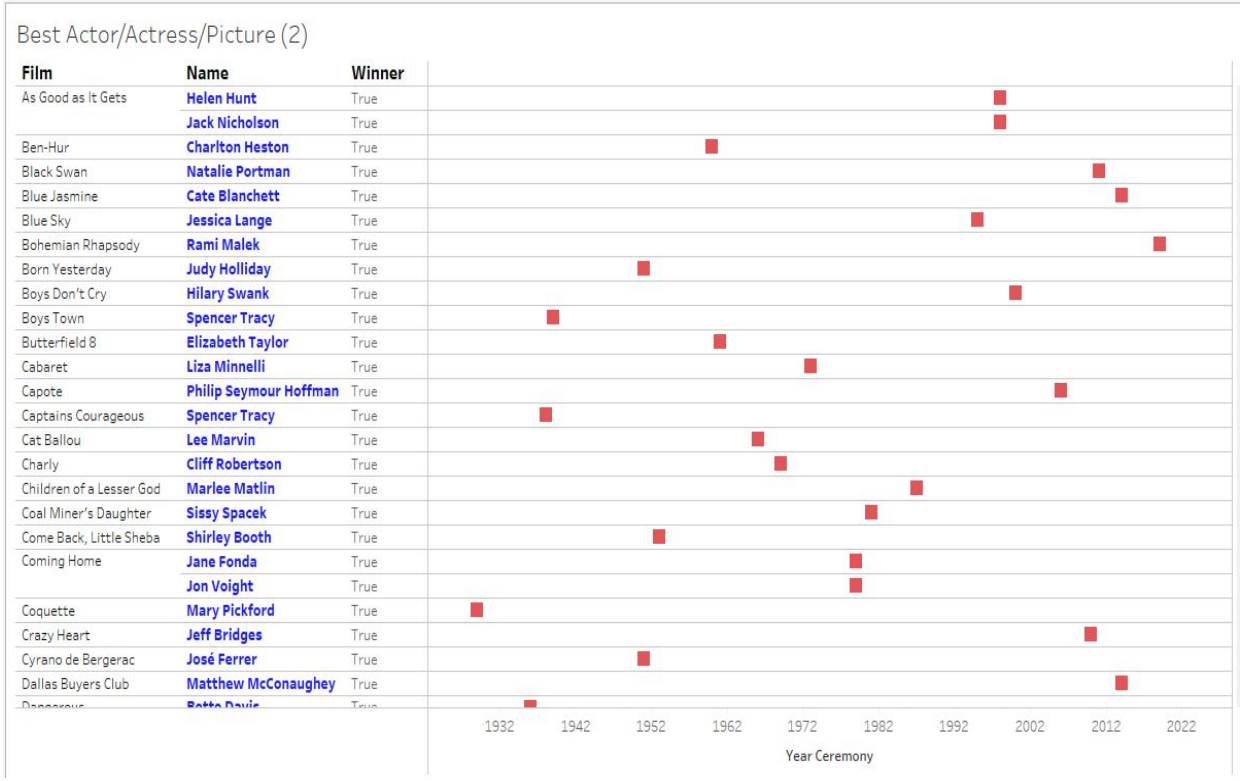
Accuracy Score : 0.7818181818181819
Balanced Accuracy Score : 0.64198606271777

Classification Report

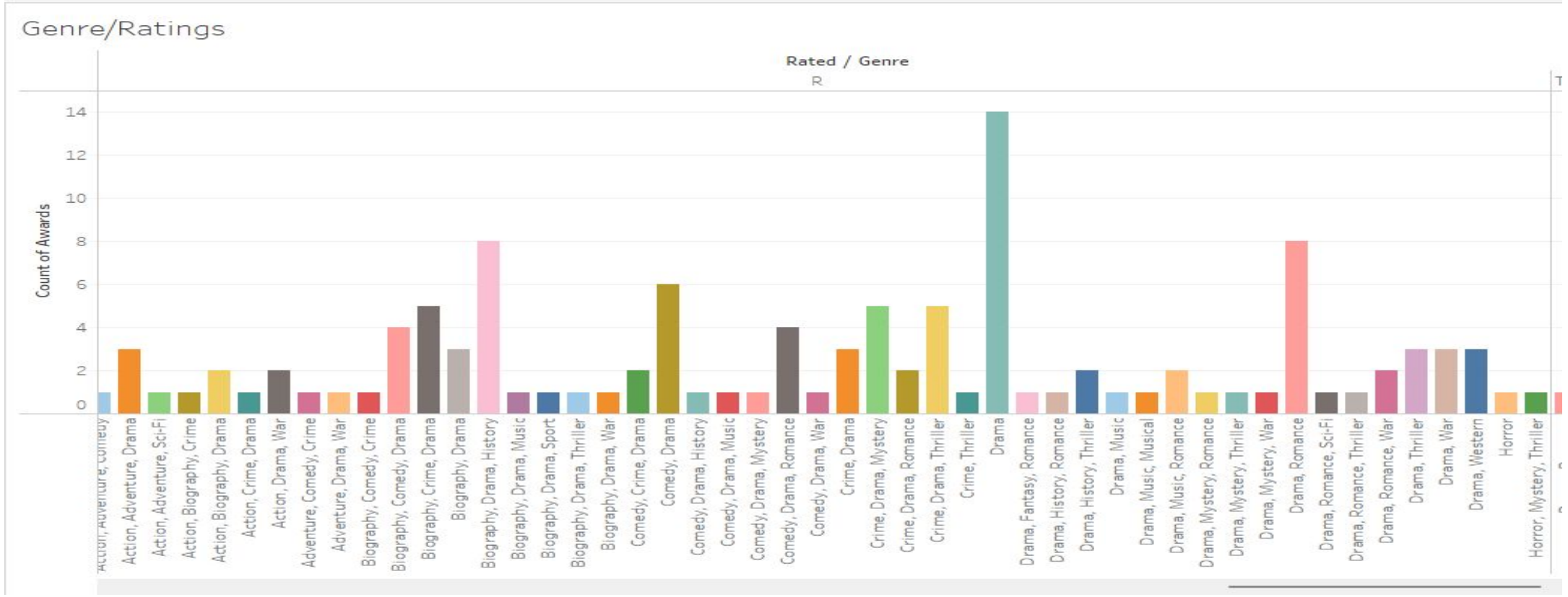
	precision	recall	f1-score	support
loser	0.81	0.93	0.86	41
winner	0.62	0.36	0.45	14
accuracy			0.78	55
macro avg	0.72	0.64	0.66	55
weighted avg	0.76	0.78	0.76	55

DATA INVESTIGATION WITH TABLEAU

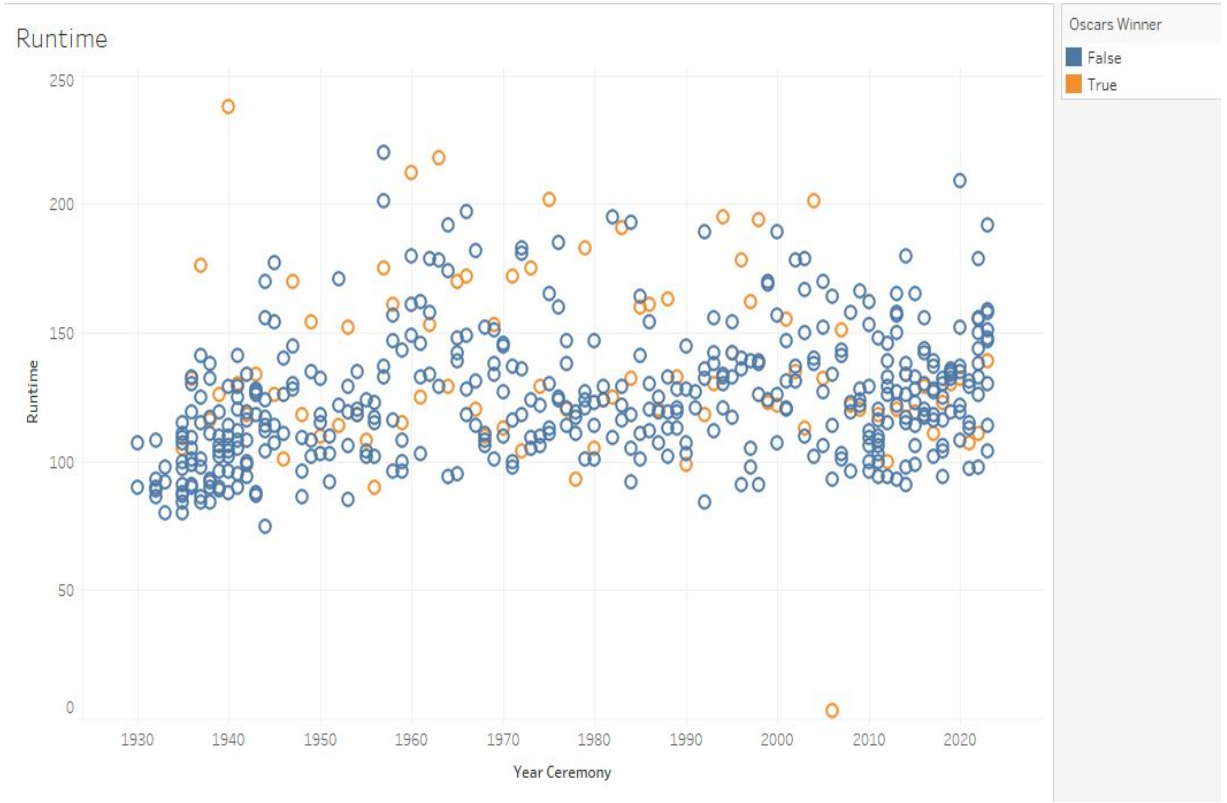


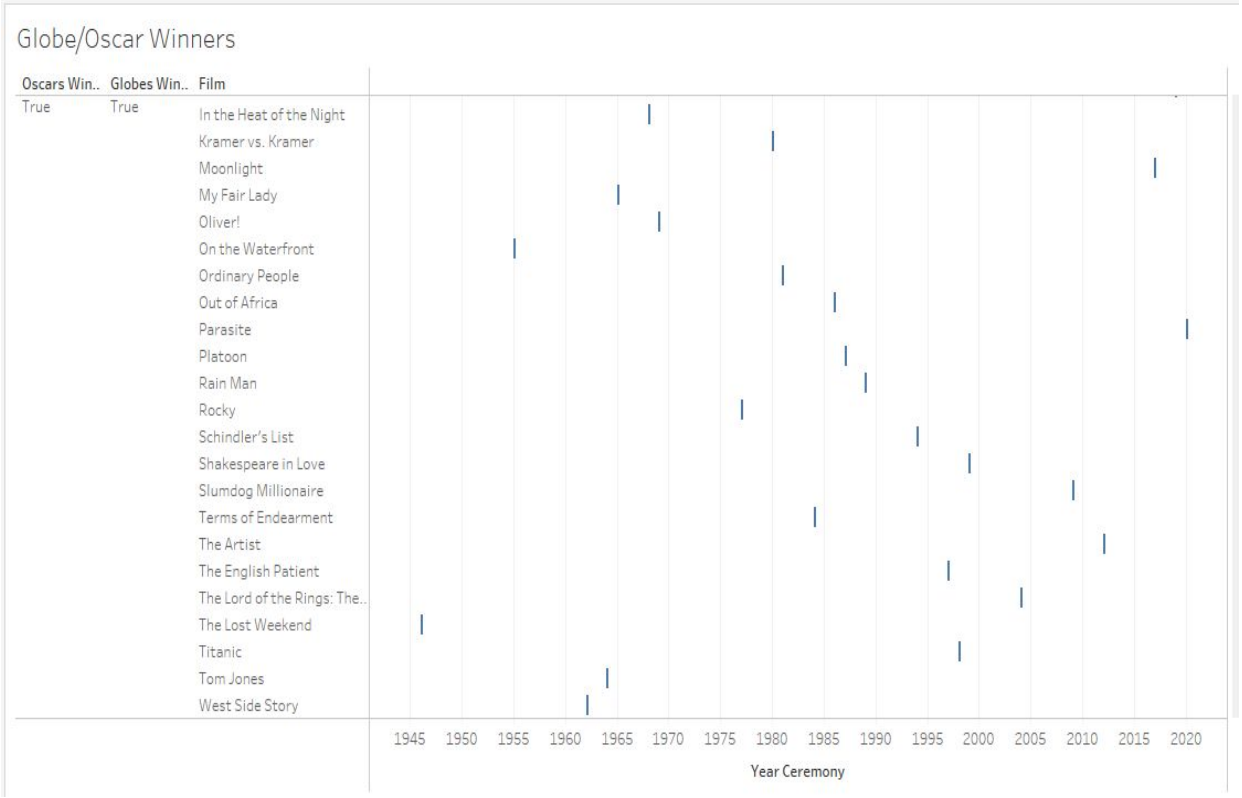


DATA INVESTIGATION WITH TABLEAU



DATA VISUALIZATION WITH TABLEAU





COMPARING TO ANOTHER MODEL

- Difficult data to predict - consider first challenge (names of categories changing)
- Importance of categories for predictions(our comparison model used 150+ categories)
- Sentiment Score

- Logistic Regression: Accuracy: 0.88

Actual\Predicted	0	1
0	209	5
1	24	2

- Random Forest: Accuracy: 0.86

Actual\Predicted	0	1
0	207	7
1	26	0

- Neural Network: Accuracy: 0.88

Actual\Predicted	0	1
0	210	4
1	25	1

- Fusion: Accuracy: 0.89

Actual\Predicted	0	1
0	211	3
1	24	2


Let’s use our model to make some predictions!

2024 Oscar Awards...what will win Best Picture?

No hope for any of these?!

Title	Prediction
Barbie	It's an honor to be nominated
Oppenheimer	It's an honor to be nominated
Past Lives	It's an honor to be nominated
Spider-Man: Across the Spider-Verse	It's an honor to be nominated
Air	It's an honor to be nominated

But wait! Our model includes Golden Globes results…and the 2024 Golden Globes haven't happened yet either. Instead of telling our model that these movies didn't win at the Globes, what if we told the model that they did win?!

After all, it is possible for 3 of these 5 movies to ALL win at the Globes since the Globes splits Best Picture into several categories including Best Drama, Best Comedy/Musical, and Best Animated Film.

And the winner is...

Title	Prediction
Barbie	It's an honor to be nominated
Oppenheimer	WINNER
Past Lives	It's an honor to be nominated
Spider-Man: Across the Spider-Verse	It's an honor to be nominated
Air	It's an honor to be nominated

Oppenheimer!

But the Oscars don’t always get it right either…

So we’re going to run the model on some movies that we think should have been nominated. (Based on their representation of being classic “snubs”)

And the winner is...

Title	Prediction
Hocus Pocus	It's an honor to be nominated
Mean Girls	It's an honor to be nominated
Dazed and Confused	It's an honor to be nominated
Die Hard	It's an honor to be nominated
The Rocky Horror Picture Show	It's an honor to be nominated
The Princess Bride	It's an honor to be nominated
Singin' in the Rain	WINNER

Singin’ in the Rain!


The End

The Academy Awards