Movie Search Engine

Bekhem Horne
Vanessa Martinez
Alexandra Flores
Alyssa Difurio

Table of contents

Introduction

The purpose of this project was to create a machine learning search engine that will allow users to create their own movie from a selection of menu items to see if the movie would be a success or not. We used a dataset from Kaggle called “The Movies Dataset” to create our model. Our inspiration came from the fact that as most people, we all collectively love to escape reality by going to see a movie. Our expectation of our search engine is to provide the user with a fun model to experience building their own movie, as well as see several visualizations of how the data works and compares.

Introduction Cont.



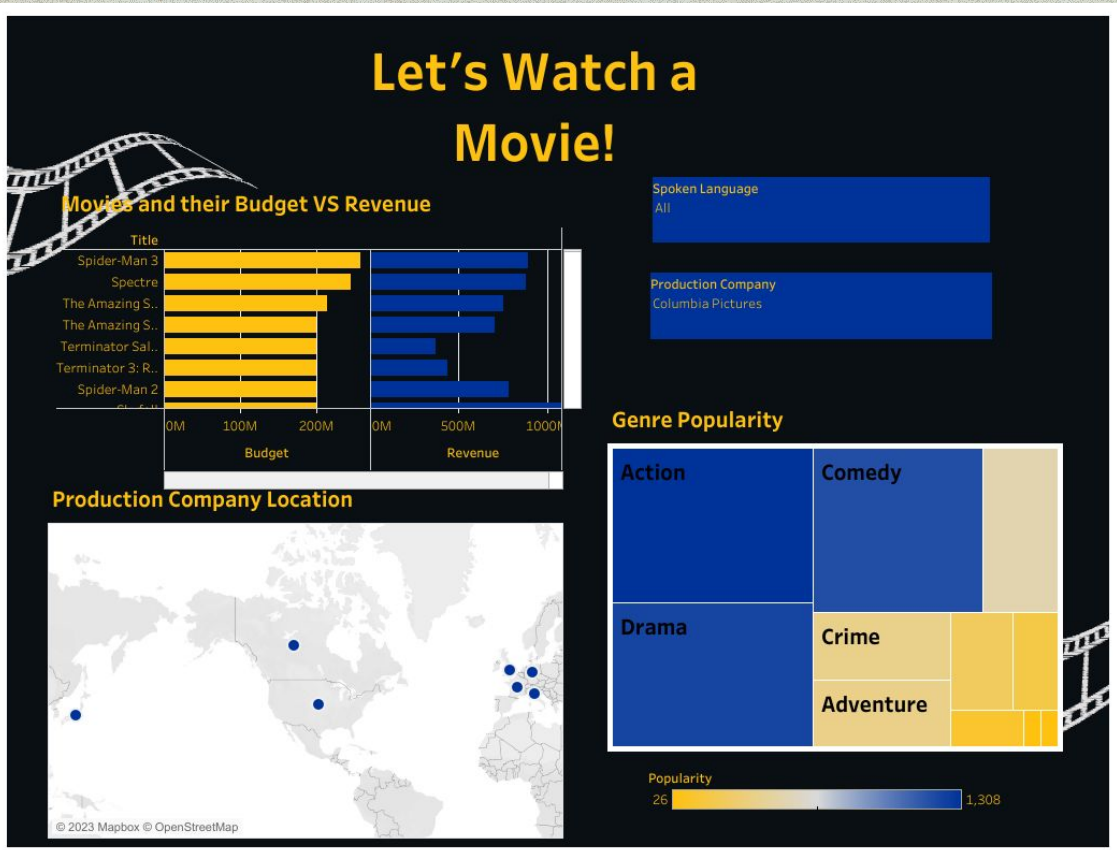



Types of Genres:
Drama
Comedy
Horror
Action
Crime
Animation
Adventure
Documentary
Fantasy
Romance
Other

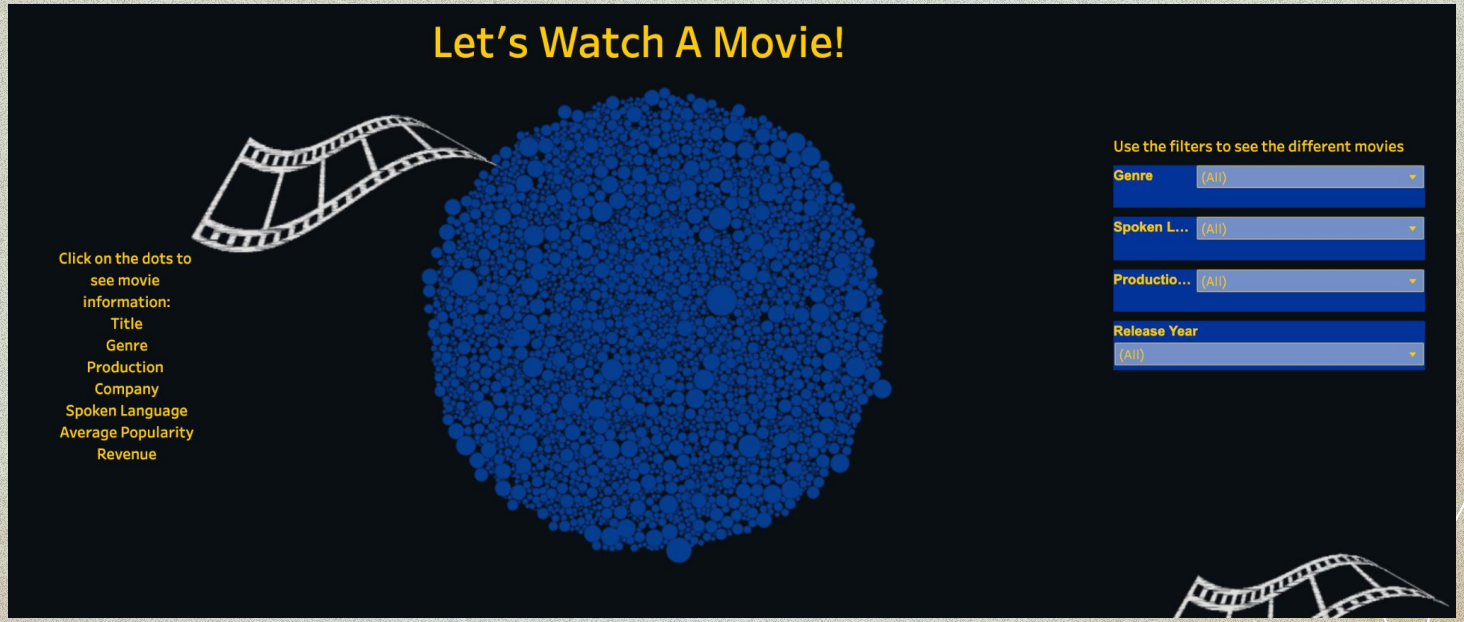
Production Companies:
Paramount Pictures
Metro-Goldwyn-Mayer (MGM)
Twentieth Century Fox Film Corp
Universal Pictures
Walt Disney Pictures
Warner Bros.
Columbia Pictures
RKO Radio Pictures
United Artists

01. Tableau

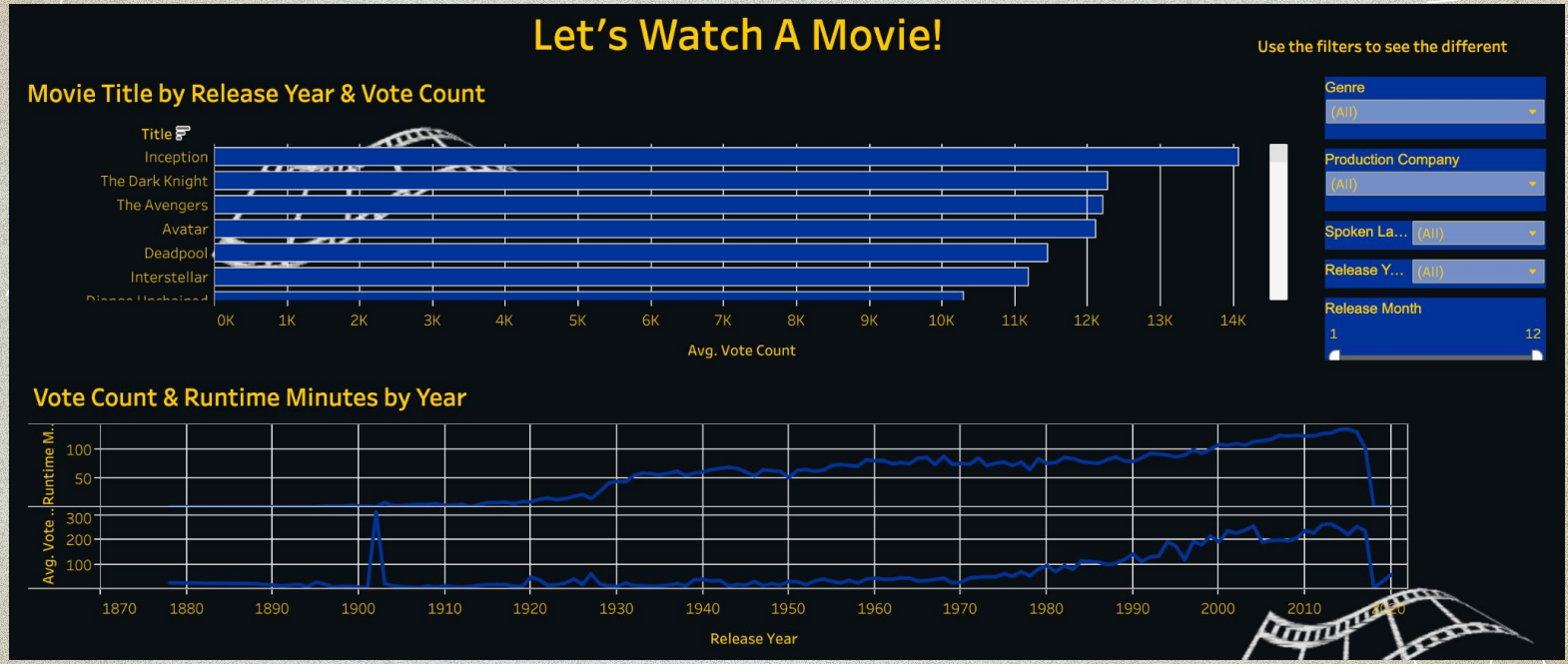
Dashboard 1


Let's Watch a Movie!
Movies and their Budget VS Revenue
Title
Spider-Man 3
Spectre
The Amazing S..
The Amazing S..
Terminator Sal..
Terminator 3: R..
Spider-Man 2
0M
100M
200M
0M
500M
Budget
Revenue
Spoken Language
All
Production Company
Columbia Pictures
Genre Popularity
Action
Comedy
Drama
Crime
Adventure
Popularity
26
1,308
Production Company Location
© 2023 Mapbox © OpenStreetMap

Dashboard 2


Let’s Watch A Movie!
Click on the dots to see movie information:
Title
Genre
Production Company
Spoken Language
Average Popularity
Revenue
Use the filters to see the different movies
Genre
(All)
Spoken L...
(All)
Productio...
(All)
Release Year
(All)

Dashboard 3


Let's Watch A Movie!
Use the filters to see the different
Movie Title by Release Year & Vote Count
Title
Inception
The Dark Knight
The Avengers
Avatar
Deadpool
Interstellar
0K
1K
2K
3K
4K
5K
6K
7K
8K
9K
10K
11K
12K
13K
14K
Avg. Vote Count
Genre
(All)
Production Company
(All)
Spoken La...
(All)
Release Y...
(All)
Release Month
1
12
Vote Count & Runtime Minutes by Year
Runtime M...
100
50
Avg. Vote ...
300
200
100
1870
1880
1890
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
Release Year

Machine Learning

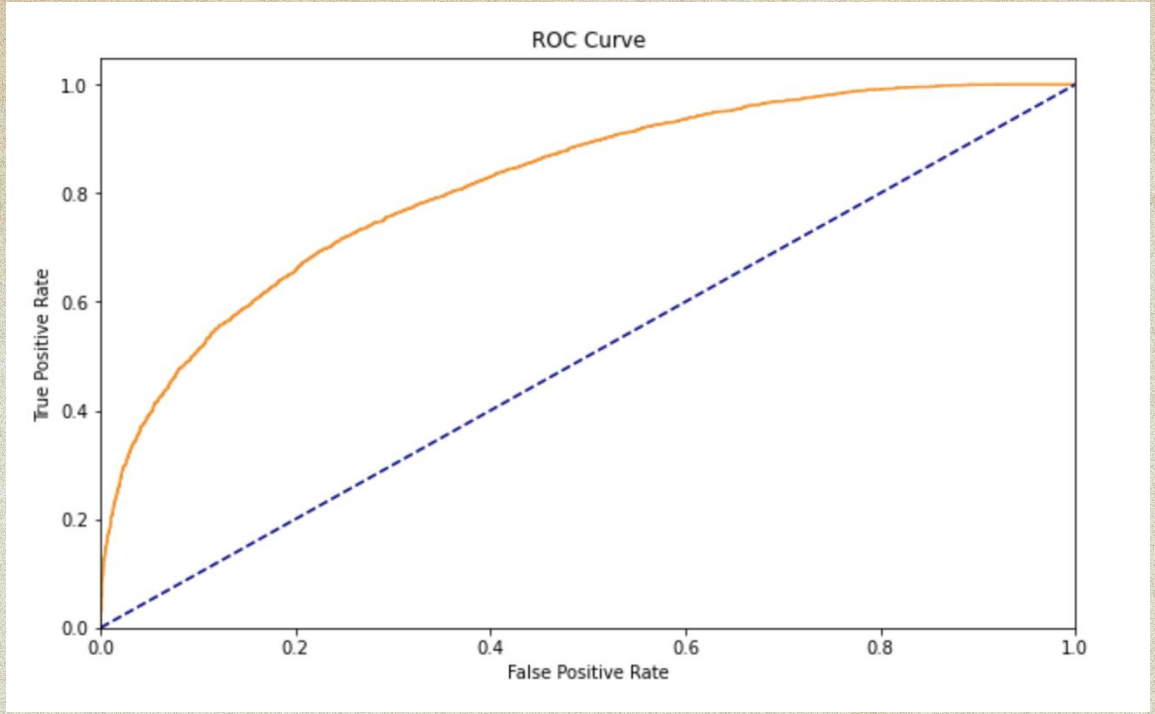
02.

LightGBM Model

```
#Lgb Model
lgb = LGBMClassifier(random_state=42)

evaluateModel(lgb, X_train, X_test, y_train, y_test)
```

```
    METRICS FOR THE TESTING SET:
    ----------------------------
    [[2883 1011]
 [1126 2985]]

                  precision    recall  f1-score   support

       False       0.72      0.74      0.73      3894
        True       0.75      0.73      0.74      4111

    accuracy                           0.73      8005
   macro avg       0.73      0.73      0.73      8005
weighted avg       0.73      0.73      0.73      8005


AUC for the Model Test Set: 0.8177237726534982
```

The LightGBM Model was our most accurate model. There are no signs of overfitting, with an F1 score of 73% and an AUC of .8177 we chose this model to use.

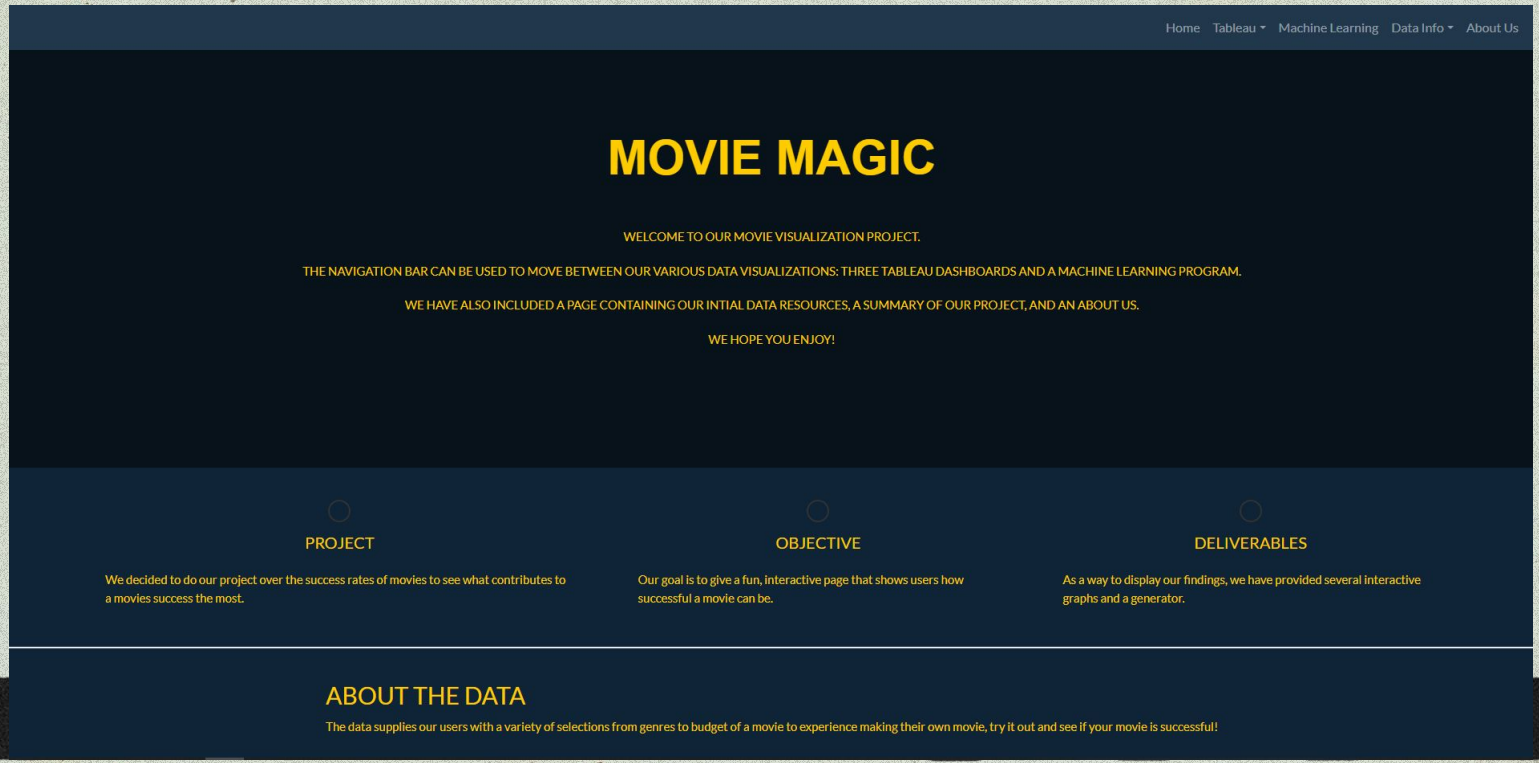


LightGBM Model

03. WebApp

Homepage

Live Demo

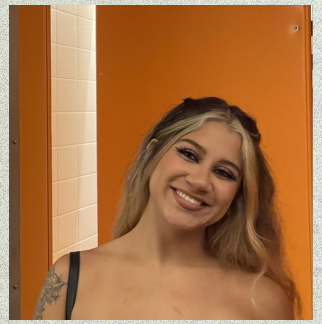


Conclusions | 04.

Future Work & Limitations

Limitations: Data cleaning in the beginning was difficult due to there being many different ID's tied to the data.
Future Work: We would like to transform what we made into a movie recommender.

Our team



Alex Flores

Alyssa Difurio

Bekhem Horne

Vanessa Martinez