Input files:

movies\_oscar.xlsx [1]

movies.xlsx [2]

wordcloud.csv [3]

all\_user\_reviews.csv [4]

Full\_critic\_reviews.csv [5]

Critics\_review.csv [6]

User\_reviews.csv [7]

rotten\_tomatoes\_reviews.csv [8]

oscar\_movie\_details.csv [9]

no\_oscar.csv [10]

final\_model.csv [11]

**PYTHON NOTEBOOK:**

1. Scraping Oscar Nominated Movies’ Details from Rotten Tomatoes

Input file: movies\_oscar.xlsx

This part of the code collects data about Oscar nominated movies from 2015-2017

Information collected:

Movie name

Genre

Audience votes

Critics votes

Rating

Run time

Studio

User score

Output file: Oscar\_details.csv

1. Delving Deeper into the Genres of Oscar Nominated Movies  
     
   Input file: movies\_oscar.xlsx  
     
   This part of the code collects the synopsis ( first paragraph) of each Oscar Nominated movie from 2015-2017 from Wikipedia  
     
   Output file: wordcloud.csv

Reading in the Wordcloud.csv file

Input file: wordcloud.csv

Generates a wordcloud that shows insights about the themes of Oscar nominated movies

1. Scraping Crtics’ Reviews for all Movies:

Input file: movies.xlsx

This part of the code collects critic reviews of all the Oscar nominated and Non-Oscar nominated movies

Output file: Full\_critic\_reviews.csv

1. Are There Any Insights In The Critics’ Reviews?

Input file: Critics\_reviews.csv

This part of the code analyzes the critics’ sentiment for all Oscar nominated movies

1. Scraping Users’ Reviews:

Input file: movies.xlsx

This part of the code scrapes user reviews of both the Oscar nominated and Non-Oscar nominated movies

Output file: all\_user\_reviews.csv

1. Analyzing User Reviews’ Positive Sentiment for Oscar Nominated Movies:

Input file: User\_reviews.csv

This part of the code analyzes audience sentiment across all Oscar nominated movies

1. NLP on Critics’ Reviews:

Input files: rotten\_tomatoes\_reviews.csv, Full\_critic\_reviews.csv, oscar\_movie\_details.csv

-This part of the code trains a Naïve Bayesian Classifier on the training data in the rotten\_tomatoes\_reviews.csv file  
  
-Next, the model gets the sentiment score for all both the Oscar and Non-Oscar nominated movies critic reviews from full\_critic\_reviews.csv file  
  
- The sentiment score is then appended to the oscar\_movie\_details.csv file to get the final model for the logistic regression model

Output file: final\_model.csv

**R Studio:**

Loading the final\_model.csv file

Removing the movie name variable

First Logistic Regression Model

Second Logistic Regression Model

Third Regression Model

**File Information:**

1.movies\_oscar.xlsx

-> manually entered these names into an Excel

-> file has names of Oscar nominated movies

-> Used these names to access the Rotten Tomatoes page to collect details of each Oscar nominated movie

-> Used this file to collect data from Wikipedia pages to generate wordcloud

2.movies.xlsx

-> has names of both Oscar nominated and Non-Oscar nominated movies

-> used to collect critics' reviews of both kinds of movies from Rotten Tomatoes

-> used to collect user reviews of both kinds of movies from Rotten Tomatoes

3.wordcloud.csv

-> Obtained this by executing 'Delving Deeper Into The Genres of Oscar Nominated Movies' part of the code

-> Collected this information from the Wikipedia page of each Oscar nominated movie

4.all\_user\_reviews.csv

-> has user reviews of both Oscar Nominated and Non Oscar nominated movies

5.full\_critic\_reviews.csv

-> has critic reviews of both Oscar Nominated and Non Oscar nominated movies

6.Critics\_review.csv

-> obtained this file by subsetting on the full\_critic\_reviews.csv file

-> this file has critic reviews of ONLY Oscar nominated movies

7.User\_reviews.csv

-> Obtained this file by subsetting on the all\_user\_reviews.csv file

-> this file has user reviews of only Oscar nominated movies

8.rotten\_tomatoes\_reviews.csv

-> has 94797 reviews from Rotten Tomatoes ( got this dataset from Kaggle)

-> used to train the NLP Naive Bayesian classifier to get sentiment score of all 237 movies

Source: <https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critics-datasets/version/1#rotten_tomatoes_movies.csv>

9.oscar\_movie\_details.csv

-> has consolidated data of both Oscar nominated and Non-Oscar nominated movies

-> obtained this data by consolidating static data (no\_oscar.csv) with scraped data (Oscar\_Details.csv)

10.no\_oscar.csv

-> obtained this sample from the rotten\_tomatoes\_movies.csv file (sourced from Kaggle)

-> has movies from 2015-2017

Source: <https://www.kaggle.com/stefanoleone992/rotten-tomatoes-movies-and-critics-datasets/version/1#rotten_tomatoes_movies.csv>

11.final\_model.csv

-> obtained this file after integrating sentiment score and performing cleaning and manipulation

-> used to build the logistic regression model