

## **Dataset:**

### **IMDb Movies Extensive Dataset**

Extensive collection of IMDb data

<https://www.kaggle.com/stefanoleone992/imdb-extensive-dataset>

Variables of Interest:

original\_title  
*Title of film*  
year  
*Release year*  
genre  
*Film genres*  
duration  
*Duration in minutes*  
country  
*Production countries*  
language  
*Languages*  
director  
*Directors*  
writer  
*Writers*  
actors  
*Actors*  
avg\_vote  
*Average user score, 1-10\*

### **The Academy Awards, 1927-2020**

Collection of Oscar winners and nominees since 1927

<https://www.kaggle.com/unanimad/the-oscar-award>

Variables of Interest:

category  
*Academy Awards category*  
winner  
*TRUE if entity is winner, FALSE if nominee*  
name  
*Name of individual*

### **IMDb Top 100 Stars of 2020**

Top 100 most popular celebrities of 2020 per IMDb STARmeter

<https://www.imdb.com/list/ls082599715/>

Variables of Interest:

Name  
*Name of individual*

## Our Dataset:

Combines IMDb data with Oscar winners and nominees and IMDb STARmeter top 100  
More than 63,000 observations

### Independent Variables (x):

genre\_1  
*Primary genre*  
genre\_2  
*Secondary genre*  
duration  
*Duration in minutes*  
country  
*Primary production country*  
language  
*Primary language*  
oscar\_actor  
*Indicates whether actors are winners or nominees*  
oscar\_director  
*Indicates whether director is winner or nominee*  
oscar\_writer  
*Indicates whether writer is winner or nominee*  
top\_100  
*TRUE if actors are in IMDb STARmeter top 100*  
continent  
*Primary production continent*

### Response Variables (y):

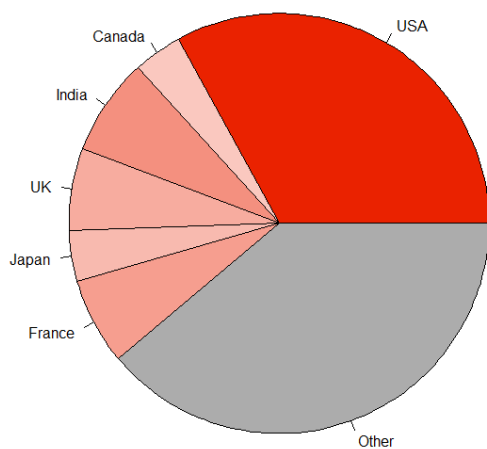
avg\_vote  
*Average user score, 1-10*  
greenlight  
*TRUE if avg\_vote > mean(avg\_vote)*

## EDA:

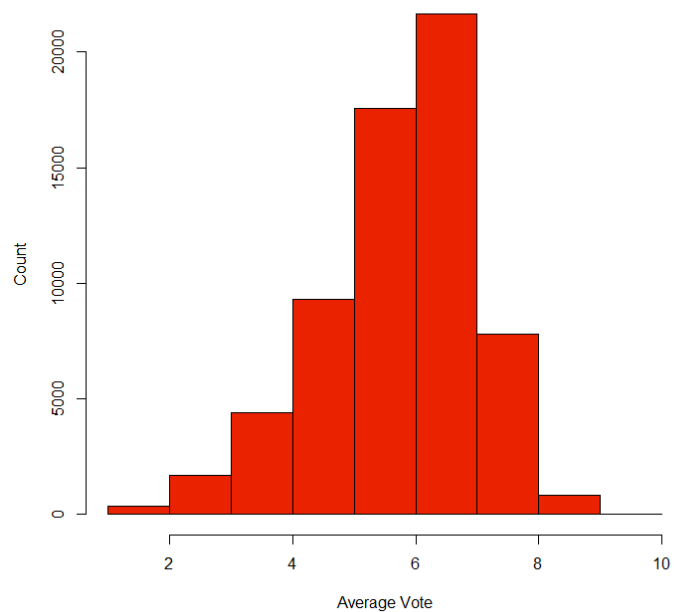
### Summary Statistics:

genre_1	duration	language	oscar_director	top_100	greenlight
Drama 18717	Min 41.0	English 30944	Winner 3507	True 702	True 27386
Comedy 17921	1st Qu 90.0	French 4431	Nominee 39	False 62922	False 36238
Action 9821	Median 97.0	Spanish 2927	None 60078		
Horror 4286	Mean 101.7	Italian 2413			
Crime 3616	3rd Qu 109.0	Japanese 2411			
Other 9263	Max 239.0	Other 20498			
genre_2	country	oscar_actor	oscar_writer	avg_vote	continent
Drama 22787	USA 20979	Winner 3507	Winner 3507	Min 1.000	Africa 522
Comedy 8653	India 4793	Nominee 4060	Nominee 166	1st Qu 5.100	Asia 12557
Romance 5399	France 4245	None 56057	None 59951	Median 6.000	Europe 23559
Thriller 4827	UK 4019			Mean 5.782	N.Amer 23981
Horror 4548	Italy 2778			3rd Qu 6.700	Oceania 1079
Other 17410	Other 26810			Max 9.800	S.Amer 1762

**Country Distribution**



**Average Vote Distribution**



**Average Vote by Duration**

