

Analysis of U.S Movie Audience using Big Data

Iksuk Kim*, Deeksha Lakshmi**, Jongwook Woo***

All authors at California State University Los Angeles, U.S.A

E-mail : ikim@calstatela.edu*, deekshalakshmi@gmail.com**, jwoo5@calstatela.edu***

1. Introduction

Audience demographics play a large role in movie production and marketing. How audiences perceive a film according to genre is clearly one of the most attractive research topics in the area of film marketing. The findings of such research are essential for both academia and the industry in order to better understand the impact of audience demographics on a film's performance. The proposed research focuses on finding the relationship between audience demographics and movie performance. More specifically, it aims to find the relationship between variables that reflect audience demographics and audience evaluations using Big Data.

2. Data

The primary data for this study were obtained from a public data website (MovieLens). A total of one million data points were downloaded in the initial stage. The data set mainly consists of four parts: Movie ID, Audience Rating, Audience Profile, and Movie Genre. In Movie ID, a total 3,882 movies were reviewed. Audience Rating is on a 1 to 5 scale (1 is poor, 5 is excellent). Audience Profile contains the reviewers' age and occupation in categorical format. Age is classified into seven categories (Under 18, 18-24, 25-34, 35-44, 45-49, 50-55, and Over 55). Occupation is classified into 21 categories (Academic/Educator, Artist, Clerical/Admin, College/Graduate Student, Customer Service, Doctor/Health Care, Executive/ Managerial, Farmer, Homemaker, K-12 Student, Lawyer, Programmer, Retired, Sales/Marketing, Scientist, Self-employed, Technician/Engineer, Tradesman/Craftsman, Unemployed, Writer, and Other or Not Specified). Movie Genre is classified into 18 different types (Action, Adventure, Animation, Children, Comedy, Crime, Documentary, Drama, Fantasy, FilmNoir, Horror, Musical, Mystery, Romance, SciFi, Thriller, War, and Western).

3. Analysis

78% of the audience members are between the ages of 18 and 44, 75% of the audience is male, and 13% of the audience members are college or graduate students. The audience ratings range from 1 to 5, but this variable is highly skewed due to higher values. A total of 226,309 cases are given a rating of 5 by audience members, resulting in 23% of total cases. Among the members to give the highest ratings, females were slightly more prone to giving higher ratings than males (female: 24%, male: 22%).

In terms of total count, Comedy and Drama genres ranked first in highest audience ratings (354,788 and 351,793 each). However, in terms of the ratio of total movies in a genre and highest audience ratings, Film-Noir genre ranks first (39%). Although Critic's Rating and Audience Ratings are significantly correlated, Critic's Rating possesses a stronger correlation coefficient (0.114) than Audience Rating (0.018) in predicting movie revenue (all correlations are significant at the 0.01 level). It is obvious that Critic's Rating and Audience Ratings are not symmetric. Among the cases which received the highest score in Audience Rating, only about 40 % of cases received such a score from critics. But of the cases which received a higher rating by critics, more than 80% of cases received a high audience score.

[Table 1] Audience by Age Group

Age	Count	Highest Rating Cases	Ratio
Under 18	27211	6802	25%
18 - 24	183536	40558	22%
25 - 34	395556	85730	22%
35 - 44	199003	44710	22%
45 - 49	83633	19142	23%
50 - 55	72490	18599	26%
Over 55	38780	10768	28%

[Table 2] Audience by Gender

Gender	Count	Highest Rating Cases	Ratio
Female	246440	58546	24%
Male	753765	167763	22%

[Table 3] Audience by Occupation

Occupation	Count	Highest Rating Cases	Ratio
Retired	13754	3839	28%
Clerical/Admin	31623	7825	25%
Doctor/Health Care	37205	9269	25%
K-12 Student	23290	5822	25%
Lawyer	20563	5069	25%
Scientist	22951	5654	25%
Programmer	57214	13670	24%
Artist	50068	11702	23%
College/Grad. Student	131032	30272	23%
Homemaker	11345	2555	23%
Sales/Marketing	49109	11315	23%
Academic/Educator	85351	18603	22%
Executive/Managerial	105425	23044	22%
Self-employed	46021	9902	22%
Technician/Engineer	72816	16209	22%
Other or Not Specified	130499	28178	22%
Customer Service	21850	4655	21%
Unemployed	14904	3179	21%
Writer	60397	12744	21%
Tradesman/Craftsman	12086	2315	19%
Farmer	2706	489	18%

[Table 4] Ratings by Genre

Genre	Count	Highest Rating Cases	Ratio
Film-Noir	18261	7053	39%
War	68527	23423	34%
Documentary	7905	2683	34%
Drama	351793	95507	27%
Crime	79387	20858	26%
Musical	41309	10433	25%
Animation	42951	10763	25%
Mystery	40178	10039	25%
Western	20683	4965	24%
Thriller	188730	41323	22%
Romance	147185	31925	22%
Sci-Fi	156263	32338	21%
Comedy	354788	72689	20%
Action	256376	52074	20%
Adventure	133953	26608	20%
Fantasy	36301	6989	19%
Children's	71950	13567	19%
Horror	75431	12076	16%