

# Correlation Analysis between Ethnic Diversity and Success Rate on a Massive Repository of Movies Data Set and the Board of Directors of Fortune 500 in Terms of Net Sales and Gross Profit

**Sarah Bamatraf**

Strategic Research Section UAE Space Agency Abu Dhabi, Abu Dhabi, UAE

Email: S.bamatraf@space.gov.ae

**How to cite this paper:** Bamatraf, S. (2023)

Correlation Analysis between Ethnic Diversity and Success Rate on a Massive Repository of Movies Data Set and the Board of Directors of Fortune 500 in Terms of Net Sales and Gross Profit. *Journal of Data Analysis and Information Processing*, 11, 463-479.

<https://doi.org/10.4236/jdaip.2023.114024>

**Received:** February 14, 2023

**Accepted:** November 13, 2023

**Published:** November 16, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In contemporary workplace, organizations are emphasizing on individual's diversity and inclusion initiatives in order to reinforce managerial adaptability, increase competitive advantage and decrease legal risks. Nonetheless, in recent times, there has arisen a debate on whether diversity is a variable that has an immediate effect on success or not. This study focused on determining if diversity in terms of ethnicity, gender, age, etc., has effects on success, by investigating two different data sets; the first one is a massive repository of movies data set and actors to determine if there is a correlation between multiple movie related variables and box office earnings. While the second data focused on Fortunes top 500 companies in the United States (US) vs. 500 less profitable companies in the US. Moreover, the study explores how diversity among Board of Directors (BOD) of fortune 500 companies affects the net sales and gross profits. The movie data set was collected from two main websites; Internet Movie Database (IMDB) and Rotten Tomatoes (RT), the imdb data set contained 107,645 records, while as the rotten tomatoes contained 13,904 records. In addition, information about Fortunes 500 companies was obtained from various websites manually, as immediate data sets were hard to find since it's the first study that focuses on diversity and success of fortune companies. The data set contained data of fortunes top 500 companies with information of all of its BOD about 5358 records, and less profitable companies of 4434 records. The reason in which these data sets were chosen was to study the ethnic diversity factor and its impact on success rate, and also due to the fact that IMDB and Rotten Tomatoes are the most recognized websites

---

that provide access to a massive repository of movie data sets. While the fortune company's data set was chosen to demonstrate diversity in the chosen dataset where one was for movies and the other was enterprise based. Furthermore, the data was analyzed in python to establish the relationship between the various variables. In all of the correlation analysis, the Pearson's coefficient was less than 0.1. Therefore, it was concluded that ethnic diversity has an insignificant effect on the success of movies and the Fortune 500 companies.

## **Keywords**

Diversity, Ethnicity, Repository

---

## **1. Introduction**

### **1.1. Problem Definition**

Recently, organizations are focusing more on corporates diversity to reinforce organizational adaptability and encourage competitive advantage. However, a debate has arisen that tackles whether diversity is a variable that has an immediate effect on success or not. The advancement in technology and the easiness of travel in the 21st century have caused a fundamental change in workplace dynamics. Cross-border trade and investment barriers have been eradicated by the advancement in telecommunication and transportation.

The breaking of the transportation and telecommunication barriers has consequently diversified the workforce in various organizations. According to [1], workforce diversity refers to similarities and differences that occur among a company's employees or a group of people working together to achieve a common goal. The similarities and difference are in terms of race, cultural background, age, physical abilities, physical disabilities, gender, sexual orientation, and religion. Diversity is known to make a working environment heterogeneous. Various researches have been done to establish the effect of diversity on the success of a company or a group of people working together towards a common goal.

### **1.2. Motivation**

One of the ways that organizations seek to improve their performance is through diversification of workforce. In recent years, many organizations and companies have implemented diversity initiatives that are beyond the traditional monolithic structure, compliance standards and affirmative action. Employers seek to employ competitive workforce regardless of age, gender, race, ethnicity, religion, language, perception, and attitude. As the global economy continues to expand, business leaders have learned the value of a multi-cultural workforce with regard to different aspects of their businesses. There is a greater understanding of diversity as a competitive edge when leveraged. Thus, the need for it to be ac-

counted for in the workforce equation for continued development. Advantages of diversity in work place [2]:

- Improves approachability to new and diverse customer marketplaces.
- Increases innovation and productivity.
- Increases revenue.
- Leads to the development of new products and services.
- Allows greater flexibility and adaptability in a more globalized environment.
- And improves social cohesion.

### 1.3. Objectives

1) To investigate impact of ethnic diversity on the success of a movie in terms of movie ratings and Box Office earnings.

2) To investigate the impact of ethnic diversity of the Board of Directors of Fortune 500 companies on the success of the company in terms of net sales and gross profit.

### 1.4. Hypothesis

Since the research is carried out on two independent data set (movie dataset and Fortune 500 company's dataset), there are to null hypothesis and two alternative hypotheses for each dataset.

- Null hypotheses

1) H01 Ethnic diversity of movie writers and directors has no impact on the success of the movie.

2) H02 Ethnic diversity of Board of Directors of Fortune 500 Company has no impact on the net sales and gross profit.

- Alternative Hypotheses

1) Ethnic diversity directly affects the success of a movie.

2) Ethnic diversity of the board of directors has an impact on the net sales and gross profit.

### 1.5. Research Contribution

The study focused on the effect of ethnic diversity to the success of a movie in terms of movie ratings and box office earnings and, the effect of ethnic diversity on the BOD of a company and the success of the company in terms of net sales and net profits. In the research, movie data from The Internet Movie Database (IMDB) and Rotten Tomatoes database were used. Also, data of Fortune 500 companies were used. The Internet Movie Database is a movie repository that store comprehensive data about movies, scriptwriters, movie directors, movie release dates among other relevant details. The IMDB repositories stores more than 900,000 movie titles. Furthermore, the site data contained more than 2.3 million individuals. Also, data contains details of scriptwriters, movie directors, movie producers, movie reviewers among others. The in-depth storage of data in the Internet Movie Database makes it a rich source of information for various

analyses. **Figure 1** displays the most common genres provided by IMDB Website and the number of movies for each genre within the dataset.

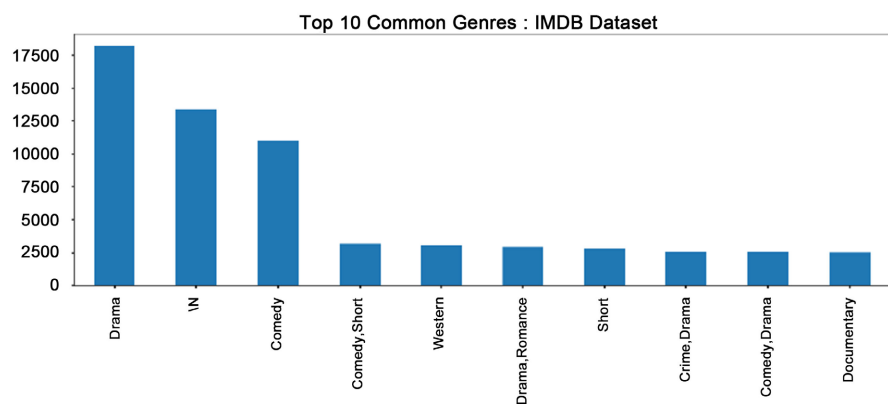
Rotten Tomatoes is a repository for movie data as well as movie review aggregation. The Rotten Tomatoes provides a platform whereby movies can be revised by professional movie reviews, movie critics, and the audience. The website uses a special algorithm to combine the reviews into a single aggregation referred to as the Tomatometer. The Tomatometer is trusted by millions of people around the world in determining whether to watch a movie not based on the review.

The research further utilizes data from Fortune 500 companies, Fortune 500 companies refer to a list that is compiled annually by the Fortune magazine. The list ranks the top 500 companies in the United States of America based on the total revenue per given fiscal year. The data provided by the Fortune magazine includes details about the Chief Executive officers of the company, the list of the Members of Board of Directors, Net profits, and Net sales among other important information. The dataset from the repository mentioned above is collected and analyzed in python using important python libraries.

## 2. Related Work

Social media is considered a massive source for sharing contents, thus, giving the liberty for millions of users to comment on all type of subjects on a daily basis. Furthermore, it is evident that businesses consider these massive repositories as a rich source for valuable data as they have a strong interest in tapping in that world in order to gather information that improves their decision-making process. As an example, using social media for creating predictive models that helps filmmakers make more profitable decisions [3].

The Internet Movie Database (IMDb) is an online comprehensive database that contains information related to movies, actors, television shows, production, etc. Furthermore, imdb features 963,309 movie titles, around 2,297,335 actor's data [4]. And has a separate web page dedicated for each of the actor's history and information. In addition, it offers ratings for each movie by aggregating the results of the overall rating given by the audience [5].



**Figure 1.** Top 10 common genres provided by IMDB.

Rotten Tomatoes is a website that is considered a rich source for movies data. It computes rating by a measurement known as the Tomatometer; which is basically a measurement of quality entertainment, representing the percentage of positive expert reviews for films and TV shows to help users with their entertainment viewing decisions as it displays a comprehensive guide to what to watch for the audience [6]. Oghina, Breuss, Tsagkias, and de Rijke [7] examined Information Retrieval (IR) system; a system that uses various signals from various sources for ranking objects efficiently. Moreover, [7] have focused on predicting movie ratings from numerous social media signals. In order to efficiently improve racial and ethnic diversity in a workplace environment, organizations are required understand some of the principle terms and definitions that includes the following [8]:

- Racial Discrimination: racial discrimination in a workplace is defined as any act of exclusion, restriction or preference that is based on race, color, descent or national or ethnic origin, which prevents an employee's ability to exercise their rights to be equally treated in a workplace [8].
- Ethnic Group: defined as a group of individuals whom members are identified through factors such as common heritage, culture, ancestry, language, dialect, history, identity and geographic origin [8].
- Ethnic Minority: is defined as any ethnic group that is not dominant socially, economically or politically [8].
- Implicit Bias: negative associations that people unknowingly hold. They are articulated automatically, without conscious awareness [8].
- Inclusion: authentically incorporating traditionally excluded individuals and/or groups into processes, activities and decision/policy making in a way that shares power [8].

## 2.1. Variables That Measure Workplace Diversity in Organizations

There are several variables that can be used to measure workplace diversity. The most important ones include the following:

### 1) Age Diversity

Age is a generational difference between employees of an organization. Growing age diversity is increasingly becoming part of quite a number of business organizations. According to Kunze [9], the social identity and categorization theory may be used to understand this relationship. According to this theory, it is suggested that individuals tend to classify themselves based on dimensions that seem relevant to them. As a result, individuals tend to favor employees of their own group, and discriminate employees from other age groups. Thus, the employees generational belonging is an important criterion for distinction that may stir emotional conflict at workplace.

### 2) Gender Diversity

Ali defines gender diversity as psychological and experienced disparities that are culturally or socially attached to being of a male or a female.

### 3) Ethnic Diversity

An ethnic group refers to a group of people with a sense of common origin, and most often, a sense of common destiny. Pitts [10] argue that as firms impress ethnic diversification, there is need to pay more attention on how different ethnic groups interact with each other at workplace. Hoogendoorn Van Praag [11] defines ethnic diversity as the heterogeneity in races, languages, and cultures among employees of an organization.

## 2.2. The Impact of Workplace Diversity on Employees and Organizational Performance

Workforce diversity has an influence on both employee and organizational performance, and consequently, on the organizational performance. This means that a positive effect of workplace diversity at employee level will have a positive effect at organizational level, and vice versa.

## 2.3. Conceptual Framework

### 1) Age Diversity

Most often, organizations avoid utilizing the expertise of old employees because of stereotypes and false assumptions that they are slow in adopting to changes and new technologies, prone to health problems, poor performance, and expensive compared to the older generation [12]. According to a study done by Hamilton Nickerson [13], on simple production technology, it was found that work-forces with age diversity were less productive. A similar finding is reported by Leonard Levine [14], where they indicated that retail store businesses that had age diversity among employees were slow in making profits. In another study by Ilmarinen [15], it was reported that there was no relationship between employees age and work performance. Many studies have shown that older employees perform work-related tasks as effective and efficient as younger employees. According to Williams O'Reilly [16], having a heterogenous age employee team is more productive than having an employee team with a homogenous age.

### 2) Gender Diversity

Mixed gender team of employees performs better compared to a team of employees of the same gender [17]. In studies carried out by [18], it was evident that there was a positive relationship between organizational performance and gender diversity, based on the organization resources. Many other studies have found a negative effect of gender diversity on team performance in male dominated samples, and insignificant effects in female dominated samples [19]. According to Samuel [20], the organizations competitive advantage increases when gender diversity is at a moderate level, while a greater level of gender diversity reduces organizational performance. The study results obtained by [21], showed an inverted U-shaped connection between employees' gender and organizational performance. In a similar study conducted [21], it was found that there was an inverted U-shaped relationship between gender mix and employee productivity.

The research also found that moderately heterogenous teams demonstrated better performance compared to gender homogenous teams. Gender diversity contributed positively to the services industry, while in the manufacturing industry, it had negative effects. Therefore, gender diversity in service industry workforce might bring a positive impact compared to companies in the manufacturing industry.

### 3) Ethnic Diversity

Jackson, *et al.* [14] studied the effect of gender diversity on performance. In his studies, he found that ethnically diverse teams of employees exhibited poor performance compared to homogenous teams. A close study by Jones [22], also demonstrated that ethnic groups were less cohesive compared to teams. Thus, ethnic diversity is likely to have a less positive impact on group performance when compared to team performance. An ethnically diverse team of employees possess high creativity and innovation that comes with learning opportunities. According to Sander Hoogendoorn [12], ethnic diversity at moderate level has no impact on organizational performance. Samuel [20], also reported a positive effect of ethnic diversity on innovation, productivity, market share, and sales. Jones [22] investigated the effect of ethnic diversity in the Oil Gas Industry. The researchers reported a positive relation between ethnic diversity and team performance. In another study conducted by Jones [22], it was found that there was no relationship between ethnic diversity and sales productivity, revenue, and customer satisfaction.

## 3. Data Analysis

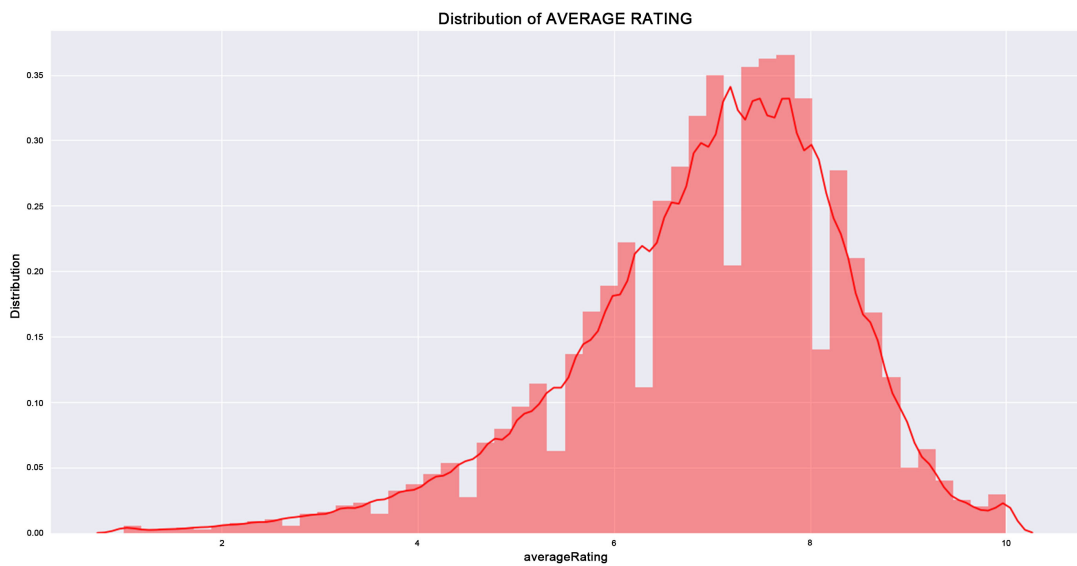
### 3.1. Movies Data Analysis; an Analysis of IMDB Movies Dataset and Rotten Tomatoes Dataset

This study was conducted by choosing two major film websites (Rotten Tomatoes), the reasons in which these websites were chosen due to the fact that these are the essential websites that provides up to date movies ratings and have massive amount of data. Other websites provide datasets that are not up to date or include missing data.

Two main movie websites were chosen as a source for data gathering; which were imdb and rotten tomatoes websites. The objective of the study was to analyze the rotten tomatoes data to determine the relationship between diversity in terms of movie writers, movie directors and the success of a movie in terms of earnings. To carry out the analysis, 14,235 movie sets were scrapped from Rotten Tomatoes website. The data contained 29 variables each describing a varied aspect of the movie. The analysis was done using Python programming language that required use of various libraries to ease the analysis process. The list of libraries used in the analysis includes Numpy, pandas, Scikit-learn, OS, and SYS. The analysis was carried out in Jupyter Notebook. **Table 1** demonstrates variables of the imdb dataset and description for each of the variable studied while **Figure 2** shows distribution of average rating.

**Table 1.** Demonstrates IMDB dataset.

Variable	Description of the variable
Movie ID	Unique ID identifying the movie
Year	The year the movie was release
New Distribution	Data about distribution of the movie
Votes	Number of votes on the movie
Ranks	Raking of the movie
Title Comp	Title of the movie
Budget	Amount of money used in movie production
Box Office	Box office income
Currency Budget	Currency of the budget
Currency Box Office	Currency of the box office income
rtcriticRatings	Movie ratings given by movie critics
rtAllCriticsNumReviews	Number of reviews by all the critics
rtAllCriticsNumFresh	Number of fresh movie critics
rtAllCriticsNumRotten	Number of Rotten Tomatoes Critics
rtAllCriticsScore	Scores by the critics
rtTopCriticsRating	Rating provided by top critics
rtTopCriticsNumReviews	Number of reviews by top rotten tomatoes top critics
rtAudienceRating	Ratings provided by the audience
rtAudienceNumRatings	Number of ratings by the audience
rtAudienceScore	Scores given by the audience
BudgetUSD	Budget of the movie in United States Dollars
BoxOfficeUSD	Box office income in United States Dollars



**Figure 2.** Shows distribution of average rating.



### 3.2. Methodology

Before starting the analysis, it was paramount that the data were cleaned for the analysis. The preparation of the data entailed carrying out the following process;

- 1) Import the required libraries (Pandas, Numpy, OS, Scikit-learn and OS).
- 2) Loading all the files using the imported libraries.
- 3) Merging the data in the files into a single dataset.

Implementation.

After setting up the environment, firstly, the dataset was loaded, the Rotten Tomatoes movies dataset and IMDB movies dataset in the Jupyter environment.

The next task was merging the datasets to find the number of movies which were common in both datasets, an inner join on Title column was used to merge the two datasets. It was found that the number of movies which were common in both datasets were 8553. It was also concluded that in IMDB dataset, Title and Original Title column have same number of unique observations. There were 107,645 distinct movies in IMDB Dataset and 13,904 movies in Rotten Tomatoes Dataset.

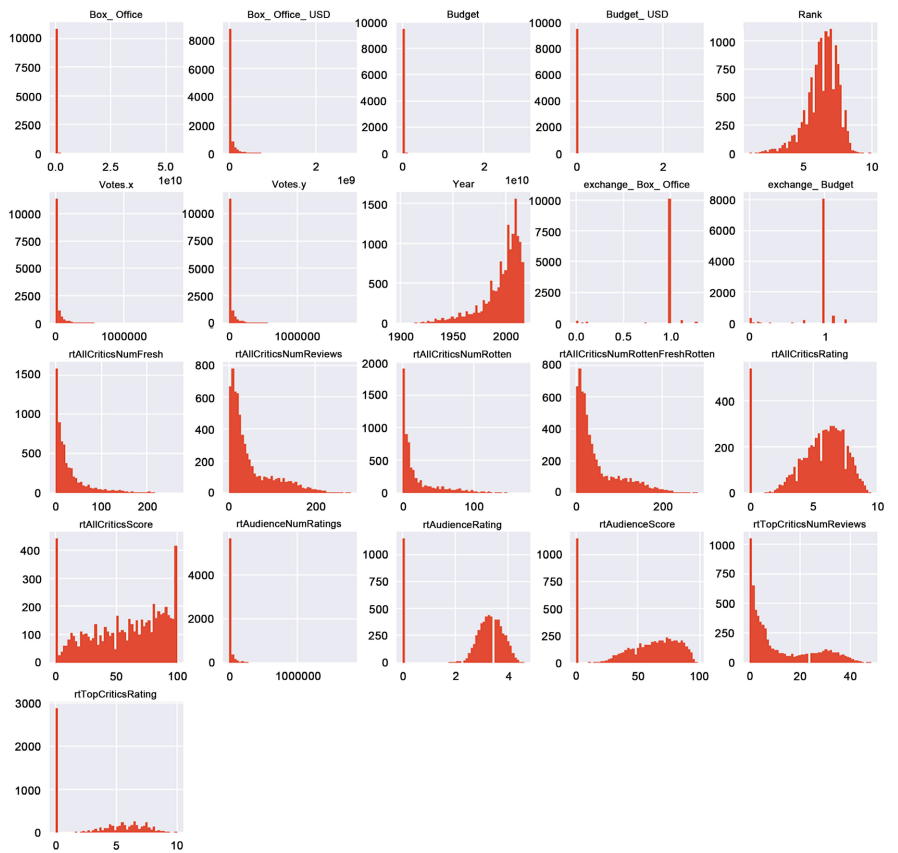
After finding the number of movies which are similar in both datasets, in terms of critics score and audience score and revenues. The Tomatometer rating which was discussed above is based on the published opinions of hundreds of film and television critics—is a trusted measurement of movie and TV programming quality for millions of moviegoers and it can also be observed that Tomatometer rating is greater than the audience ratings which shows that critics rated the movie in more positive manner as compared to the audiences. It can be observed that most of the movies are drama and comedy. Also, for 13,367 movies the genre is not recognized. Next, the number of movies or programmes of various title types present in the dataset was analyzed. And finally, for the IMDB Dataset, the distribution was also analyzed in terms of runtimes in minutes and found that most of the movies were of 90 minutes in runtime.

### 3.3. Rank Versus Ratings (Critics/Audience)

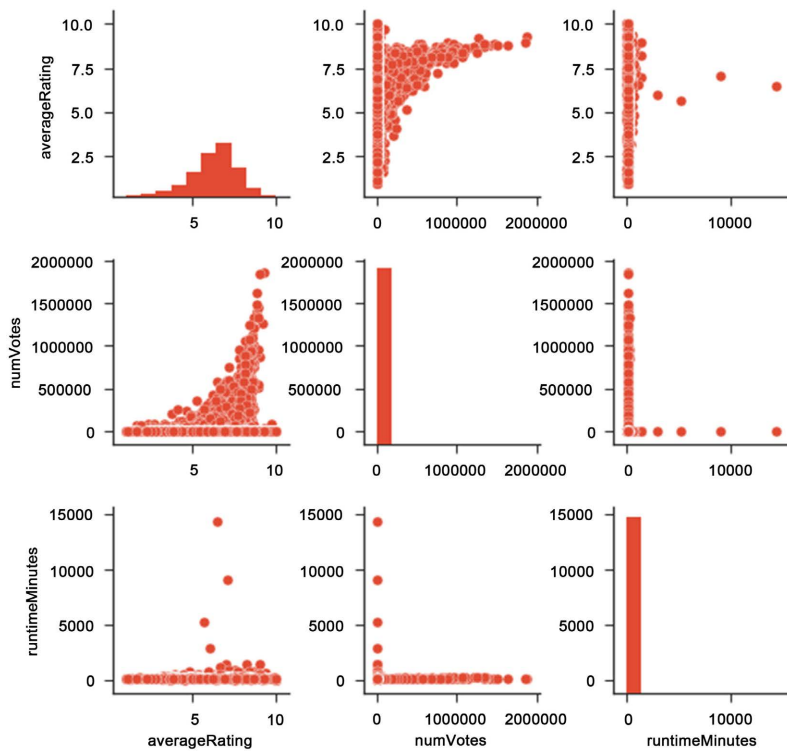
In a similar manner, what was also analyzed was the scatter plots among the audience/critics score and the overall rank of the movie in the two scatterplots. But, here in this case, we can observe that Audience Score is highly correlated to the Rank of the movie as compared to the Critic Score. **Figure 3** demonstrates the various numeric variables. And **Figure 4** exhibits a correlation matrix heat map.

### 3.4. Pearson Method of Correlation

Statistical tests were also performed to understand how these variables are dependent on each other and how significantly they differ from each other. The results shown below are the Pearson's coefficient of Correlation which has been already discussed above in the scatter plots. One of the measures used to establish if there is a relationship between two variables is the Pearson product-moment correlation coefficient normally referred to as Pearson's coefficient.

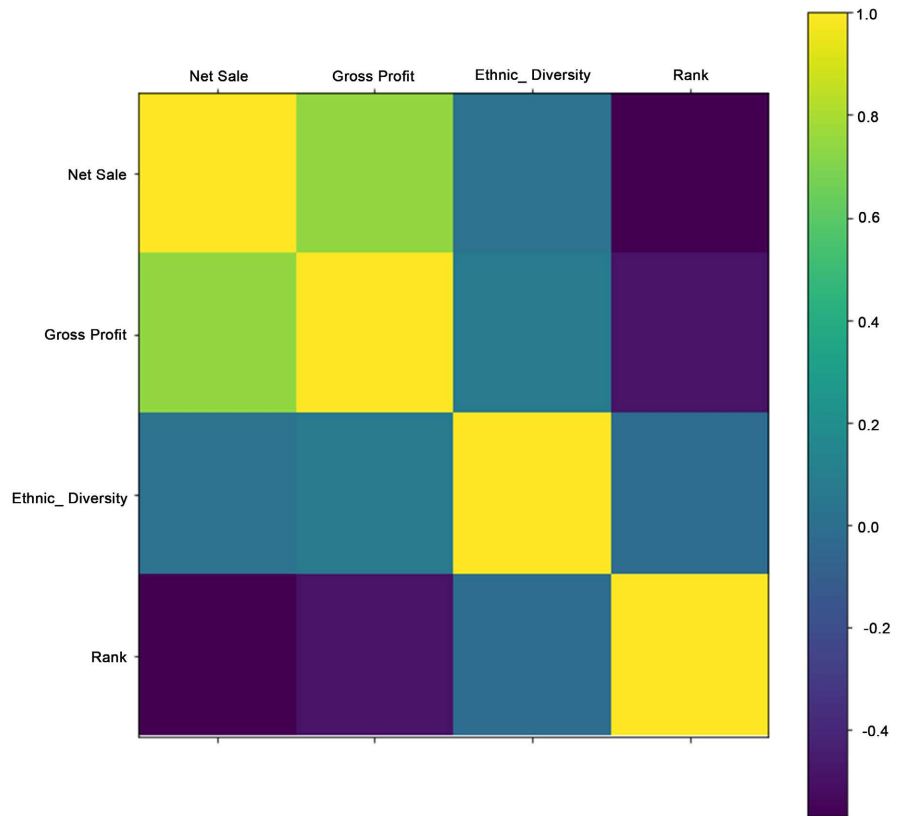


(a)



(b)

Figure 3. Distribution of various numeric variables in the dataset.



**Figure 4.** Correlation matrix heat map.

The Pearson's coefficient measures the strength of the linear relationship between two variables. The formula for calculating the Pearson's coefficient is as shown below:

$$1 - \lambda = 1 - \sum_{i=1}^R \rho_i^2 = 1 - \frac{1}{2D}$$

The value of Pearson product-moment correlation coefficient ranges between  $-1$  and  $1$ . A strong positive correlation tends to lean towards positive one while a strong negative correlation tends to lean towards negative one. If the value of Pearson product-moment correlation coefficient is zero, there is an indication that there is no correlation between the variables under study. There are various ways in which the value of the Pearson product-moment correlation coefficient can be calculated. One of the most common method entails plotting a scatter plot and then establishing a line of best fit through the plotted data points. Pearson product-moment correlation coefficient is then used to indicate how far the plotted points are from the best line of fit. In this project, the Pearson product-moment correlation coefficient was determined using the Numpy library using the `correlf()` function.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

### 3.5. Fortune's Companies Data Analysis

Data about the Board of Directors of Fortune 500 companies was collected. The data was contained in two CSV files. One file contained data about the details of the Board of directors, while the other file contained information about net sales and gross profit. To have an overview of the data, the head () function was used. Furthermore, the data about the company was grouped as either *top gainers* or *top losers*. The following table shows a snippet of the Board of directors' top gainers data and top loser data. **Table 2** shows Fortune's Less Profitable Companies Diversity Score.

The data of the top losers and the top gainers company were merged with the data of the net sales and gross sales of the respective companies. Using the data of the board of directors of each of the fortune 500 companies, the ethnic diversity score was calculated. After the calculation, a histogram was plotted to visualize how diverse the board of directors are in the company basing on the earlier calculated ethnic diversity score. **Figure 5** shows a plot for Ratings versus Diversity for Writers.

## 4. Results

### 4.1. Distribution of Ethnic Diversity among the Writers

**Figure 6** shows Box Office Versus Diversity Score for Writers. The movie writers have varied diversity, consisting of Hispanics, Blacks, and whites, among others. It is thus important to establish how the diverse the movie writers are according to the Rotten Tomatoes data, it can be seen that the diversity among the writers is normally distributed save the maximum number of observations. Furthermore, the chart shows that most of the movies have writer ethnic diversity score that is greater than 0.6 and less than 0.8.

**Table 2.** Fortune's less profitable companies diversity score.

Company Name	Net Sale	Gross Profit	African	American	Asian	British	Indian	total ethnicities	diversity
Big Lots Inc.	5.2 B	2.1B	0	9	0	0	0	9	0
Momentive Performance Materials Inc	544 M	146 m	0	10	1	0	0	11	0.16
Markel Corporation	5.61 B	2.06 B	0	14	0	0	0	14	0
Noble Energy, Inc.	3.49 B	2.41 B	0	11	0	0	0	11	0
Leidos Holding, Inc.	7.04 B	852 M	0	12	1	0	0	13	0.14
Rockwell Collins, Inc.	6.82 B	1.95 B	0	10	0	0	0	10	0
Sprague	2.39 B	210.9 M	0	8	0	0	0	8	0
YRC World Wide Inc.	4.7 B	3.34 B	0	10	0	0	0	10	0
The Hanover Insurance Group, Inc.	4.95 B	1.98 B	0	11	0	0	0	11	0
Fiserv, Inc.	5.51 B	2.55 B	0	9	0	0	0	9	0

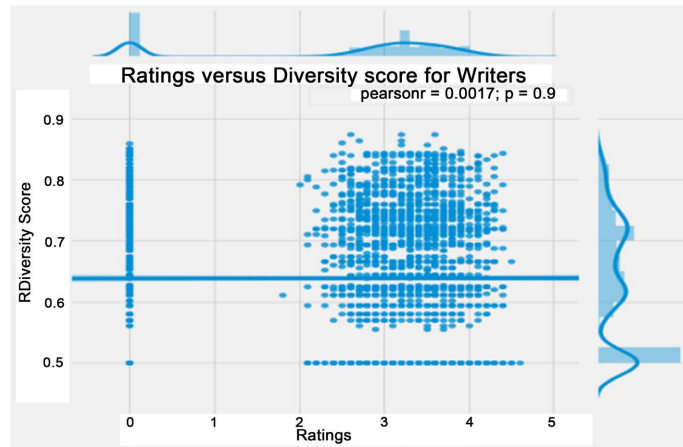
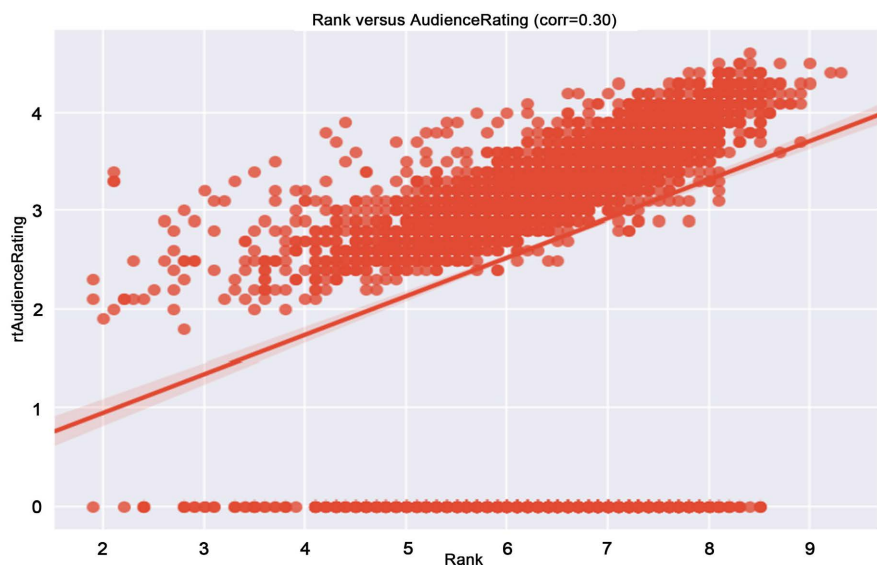


Figure 5. Ratings versus diversity for writers.



(a)

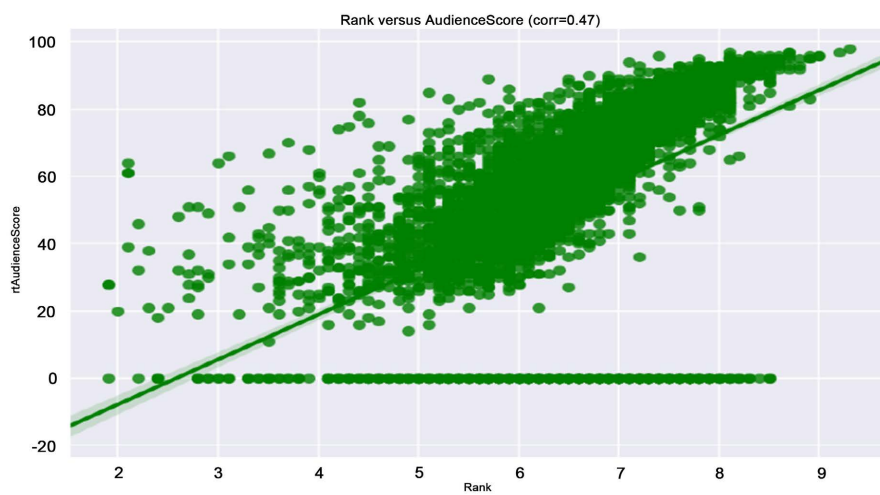


Figure 13 Rank versus Score (Audience)

(b)

Figure 6. Rank versus audience rating and audience score.

## 4.2. Distribution of Ethnic Diversity among the Directors

It can be seen that the diversity score with the highest distribution is between 0.68 and 0.70. Thus, it can be concluded that the distribution of diversity among the movie directors is relatively normal.

## 4.3. Correlation between Writers Ethnic Diversity and Movie Ratings

**Figure 7** shows the Rank Versus Audience Rating and Audience Score. To visualize how diversity among the writers relates the diversity movie ratings, a regression joint plot was drawn. From the regression chart, it can be seen that the Pearson product moment correlation coefficient is 0.00017 (the Pearson correlation product moment coefficient is approximately equal to zero). Since the value of  $r$  is almost close to zero, thus be concluded that there is no relationship between the diversity among the movie writers and he movie rating.

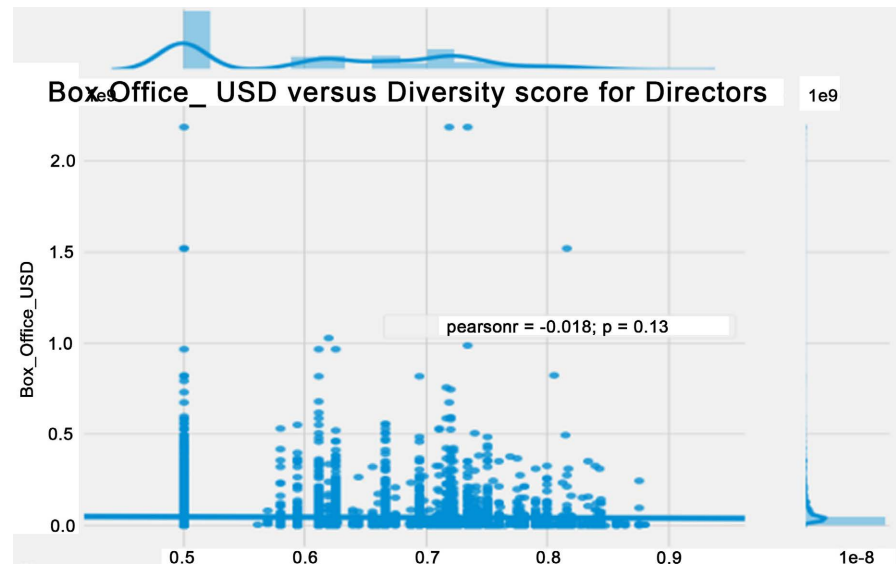
## 4.4. Correlation between Box Office Earnings and Diversity among Writers

Analysis was also done to establish if there is a relationship between writer's diversity score and the movies box office earnings. From the regression chart above, it can be seen that the Pearson's correlation coefficient,  $r$ , is equal to 0.11. The value of  $r$  is significant and thus, it can be concluded that there is a weak positive relationship between movie diversity and box office earnings. The relationship is such that as the diversity score increases, the box office earnings increase.

## 5. Discussion

The analysis in the previous section has presented how ethnic diversity score of movie writers and directors effects on the success of a movie. The section has also analyzed how the ethnic diversity score of Board of Directors of fortune 500 companies affects the net sales and net profits. The general conclusion from the analysis is that ethnic diversity has an impact on success of movies and the fortune 500 companies. In this context, the success of a movie is measured in terms of ratings and box office earnings while the success of fortune 500 companies is measured in terms of net sales and the profits. Despite the established correlation, the impact of ethnic diversity is very small and may be considered to be insignificant. The impact of ethnic diversity is attributed to various factors. Jackson [14] studied the effect of ethnic diversity on performance. In his studies, he found that ethnically diverse teams of employees exhibited poor

Performance compared to homogenous teams. A close study by Jones [22], also demonstrated that ethnic groups were less cohesive compared to teams. Thus, ethnic diversity is likely to have a less positive impact on group performance when compared to team performance. Despite the fact that an ethnically diverse team of employees possess high creativity and innovation that comes



**Figure 7.** Box office versus diversity score for writers.

with learning opportunities and complementary, Hoogendoorn [12], established that ethnic diversity at moderate level has no impact on organizational performance. Samuel [20], also reported a very weak positive effect of ethnic diversity on innovation, productivity, market share, and sales in a company. Also, Jones [22] investigated the effect of ethnic diversity in the Oil Gas Industry. The research reported a weak positive relation between ethnic diversity and team performance. In another study conducted by Jones [22], it was found that there was no relationship between ethnic diversity and sales productivity, revenue, and customer satisfaction. It is evident that the outcomes of this research are in tandem with what other researches had established earlier. All researches agree that ethnic diversity has a weak impact on success of a company in terms of net profits and sales. The same trend extends to the effect of ethnic diversity in terms of writers and movies on the success of a movie in terms of movie rating and box office earnings.

## 6. Future Work

Future work can focus on obtaining the ethnicities of actors from Wikipedia as the ethnicities that were obtained for this study was by the use of a name classifier. Website such as Wikipedia and ethnicelebs were examined for ethnicities but it does not contain all the ethnicity of the actors. Also, Wikipedia was scraped for data regarding ethnicity, nonetheless, it only had information regarding actor's nationality not ethnicity.

## 7. Conclusions

According to the analysis done, the Pearson's coefficient has been found to be less than 0.1 in all the regression analysis. This implies a weak relationship between ethnic diversity and the other variables. Based on this, the null hypothesis

is accepted. Therefore, it can be concluded that ethnic diversity of movie writers and directors has no significant impact on the success of a movie and Ethnic diversity of Board of Directors of Fortune 500 Company has no impact on the net sales and gross profit.

Moreover, the analysis in has presented how ethnic diversity score of movie writers and directors affects the success of a movie. The paper has also analyzed how the ethnic diversity score of board of directors of fortune 500 companies is associated with net sales and net profits.

The general conclusion from the analysis is that ethnic diversity has an impact on success of movies and the Fortune 500 companies. In this context, the success of a movie is measured in terms of ratings and box office earnings while the success of Fortune 500 companies is measured in terms of net sales and the profits. Despite the established correlation, the association of ethnic diversity is very small and may be considered to be insignificant.

It is evident that the outcomes of this research are in tandem with what other researches had established earlier. All the researchers agree with this study findings, which showed that ethnic diversity has a weak impact on success of a company in terms of net profits and sales.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- [1] Saxena, A. (2014) Workforce Diversity: A Key to Improve Productivity. *Procedia Economics and Finance*, **11**, 76-85. [https://doi.org/10.1016/S2212-5671\(14\)00178-6](https://doi.org/10.1016/S2212-5671(14)00178-6)
- [2] Siciliano, J.I. (1996) The Relationship of Board Member Diversity to Organizational Performance. *Journal of Business Ethics*, **15**, 1313-1320. <https://doi.org/10.1007/BF00411816>
- [3] Apala, K.R., Jose, M., Motnam, S., Chan, C.C., Liszka, K.J. and de Gregorio, F. (2013) Prediction of Movies Box Office Performance Using Social Media. *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Niagara*, 25-29 August 2013, 1209-1214.
- [4] Augustine, A. and Pathak, M. (2008) User Rating Prediction for Movies. Technical report, Citeseer.
- [5] Amazon.com, Inc. (1990) Internet Movies Database. <https://www.imdb.com/>
- [6] Fandango (2016) Rotten Tomatoes. <https://www.rottentomatoes.com/about/>
- [7] Oghina, A., Breuss, M., Tsagkias, M. and de Rijke, M. (2012) Predicting IMDB Movie Ratings Using Social Media. In: *European Conference on Information Retrieval*, Springer, Berlin, 503-507. [https://doi.org/10.1007/978-3-642-28997-2\\_51](https://doi.org/10.1007/978-3-642-28997-2_51)
- [8] 2017 PeopleScout, a TrueBlue Company—All Rights Reserved (2017) Improving Racial and Ethnic Diversity in the Workplace. <http://www.peoplescout.com/racial-and-ethnic-diversity-in-the-workplace/>
- [9] Kunze, F., Boehm, S. and Bruch, H. (2009) Age Diversity, Age Discrimination, and Performance Consequences—A Cross Organizational Study. *Academy of Manage-*



- ment Proceedings, 2009, 1-6. <https://doi.org/10.5465/ambpp.2009.44249873>
- [10] Ali, M., Kulik, C.T. and Metz, I. (2011) The Gender Diversity—Performance Relationship in Services and Manufacturing Organizations. *The International Journal of Human Resource Management*, **22**, 1464-1485. <https://doi.org/10.1080/09585192.2011.561961>
- [11] Pitts, D.W. and Wise, L.R. (2010) Workforce Diversity in the New Millennium: Prospects for Research. *Review of Public Personnel Administration*, **30**, 44-69. <https://doi.org/10.1177/0734371X09351823>
- [12] Hoogendoorn, S. and Van Praag, M. (2012) Ethnic Diversity and Team Performance: A Field Experiment. IZA Discussion Paper No. 6731. <https://doi.org/10.2139/ssrn.2114911>
- [13] Selvaraj, P.C. and Darwin, J.R. (2015) The Effects of Work Force Diversity on Employee Performance in Singapore Organisations. *International Journal of Business Administration*, **6**, 17-29. <https://doi.org/10.5430/ijba.v6n2p17>
- [14] Hamilton, B.H., Nickerson, J.A. and Owan, H. (2004) Diversity and Productivity in Production Teams. <https://doi.org/10.2139/ssrn.547963>
- [15] Leonard, J. and Levine, D.I. (2003) Diversity, Discrimination, and Performance. <https://doi.org/10.2139/ssrn.420564>
- [16] Ilmarinen, J. (2005) Towards a Longer Worklife: Ageing and the Quality of Worklife in the European Union. Finnish Institute of Occupational Health, Ministry of Social Affairs and Health.
- [17] Williams, K.Y. and O'Reilly, C.A. (1998) Demography and Diversity in Organizations: A Review of 40 Years of Research. *Research in Organizational Behavior*, **20**, 77-140.
- [18] Wood, R. and Bandura, A. (1989) Impact of Conceptions of Ability on Self-Regulatory Mechanisms and Complex Decision Making. *Journal of Personality and Social Psychology*, **56**, 407-415. <https://doi.org/10.1037/0022-3514.56.3.407>
- [19] Frink, D.D., Robinson, R.K., Reithel, B., Arthur, M.M., Ammeter, A.P., Ferris, G.R., Kaplan, D.M. and Morrisette, H.S. (2003) Gender Demography and Organization Performance: A Two-Study Investigation with Convergence. *Group & Organization Management*, **28**, 127-147. <https://doi.org/10.1177/1059601102250025>
- [20] Idowu, S.O., Capaldi, N., Zu, L.R. and Gupta, A.D. (2013) Encyclopedia of Corporate Social Responsibility, Springer, New York. <https://doi.org/10.1007/978-3-642-28036-8>
- [21] Van der Vegt, G.S. and Bunderson, J.S. (2005) Learning and Performance in Multidisciplinary Teams: The Importance of Collective Team Identification. *Academy of Management Journal*, **48**, 532-547. <https://doi.org/10.5465/amj.2005.17407918>
- [22] Jones, M., Qazi, M. and Young, K.D. (2005) Ethnic Differences in Parent Preference to Be Present for Painful Medical Procedures. *Pediatrics*, **116**, e191-e197.