<u>How Movie Reviews Influence Human Decision Making</u>

Gisele Tumutuyimana
Department of Computer Science, Hartwick College
Big Data
Professor Wang
9 November 2021

Table of Contents

Biography:

Gisele Tumutuyimana is a computer science major in her junior year at Hartwick College. Rwanda, East Africa, is where she was born and raised. When she was 12 years old, her father bought their first television ever, and everything appeared fantastic from such a little screen. This piqued her curiosity in learning how computers function. That's why she decided to pursue computer science. she has experience with Python, SQL, and a couple of experiences she gained while working at Hartwick's technology resource center, such as (hardware repair and software install for rental computers for students, Assisting network technicians with network troubleshooting computer labs and printing support, Blackboard transaction support Gmail support, and so on). She is still exploring her career interest as computer science is very huge.

Abstract:

The topic of this project is how human behavior can be influenced by movie reviews. Our theory and the null hypothesis were that movie reviews will POSITIVELY boost a person's intention to watch that specific movie. It is a good thing to note that not all human behavior is the same and therefore there will always be different results depending on how a project is presented. That's why our project is represented by three different movies, varying genres. The first is drama/romance, the second, being action/sci-fi, and the last is horror/thriller. We also kept in mind that there will be outside factors affecting our project as a whole.

Problem Statement:

The reason we wanted to find the impact of movie reviews on movie watchers is that we felt as sometimes movie critics would over-exaggerate the integrity of a movie. This could be due to the movie director of that movie having a past reputation of making bad movies, the reputation of a movie franchise i.e. *Fast & Furious*, the movie actors being infamous for something beforehand, and etc. To add to this, a movie franchise with a loyal and very supportive fanbase may give an upcoming movie a bad review simply because that upcoming movie didn't live up to or exceed their expectations. On the other hand, we considered the professionalism of the website the review is being read from. Common movie review sites can receive good, bad, or mixed reviews; whether that website only accepts professional movie criticism or not. An example of what is being said is movie critic websites such as Rotten Tomatoes, Roger-Ebert, The Guardian, IMDb are considered professional critic sites. In comparison, Netflix, Hulu, Prime Video reviews are meant to represent the population of casual watchers.

Figure 1:

Happy Madison Films: Critical Reception in Review

Year	Film Title	MetaCritic	Rotten Tomatoes	Roger Ebert
1999	*Deuce Bigalow: Male Gigolo*	30	0	1.5
2000	*Little Nicky*	38	22	2.5
2001	*The Animal*	43	30	-
2001	*Joe Dirt*	20	11	1.5
2002	*Mr. Deeds*	24	22	1.5
2002	*The Master of Disguise*	12	2	1
2002	*Eight Crazy Nights*	23	12	2
2002	*The Hot Chick*	29	21	0.5
2003	*Anger Management*	52	43	2
2003	*Dickie Roberts: Former Child Star*	36	23	2
2004	*50 First Dates*	48	44	3
2005	*The Longest Yard*	48	31	3
2005	*Deuce Bigalow: European Gigolo*	23	10	0
2006	*Grandma's Boy*	33	18	-
2006	*The Benchwarmers*	25	12	-
2006	*Click*	45	33	2
2007	*I Now Pronounce You Chuck and Larry*	37	14	-
2007	*Reign Over Me*	61	63	-
2008	*Strange Wilderness*	12	0	-
2008	*The House Bunny*	55	40	-
2008	*You Don't Mess with the Zohan*	54	36	3
2008	*Bedtime Stories*	33	25	2.5
2009	*Paul Blart: Mall Cop*	39	33	3
2009	*The Shortcut*	-	-	-
2009	*Funny People*	60	68	3.5
2010	*Grown Ups*	30	10	2
2011	*Just Go with It*	33	20	1
2011	*Zookeeper*	30	14	3
2011	*Bucky Larson: Born to Be a Star*	9	0	-
2012	*Jack and Jill*	23	3	-
	Average:	**34.7**	**22.7**	**2.025**

Note: MetaCritic and RT scores are out of 100. Ebert scores are out of 4.

As of: 4/6/2012

Figure 1: This is a table of three notable movie critic websites; MetaCritic, Rotten Tomatoes, and Roger Ebert. It is meant to compare the averages of each movie and the total averages together.

Related Work / Literature:

We used an honors thesis from a student from the University of New Hampshire, Jacob R. Pentheny, made not too long ago (2015), as a reference. This was also so we can find out other factors we have not thought about that may impact our project. In his 2015 thesis, he states that the influence of more known stars and genres will affect his research rather than less popular genres (Pentheny, 2015). Our group agreed with this theory because this statement applies to the real world as market performance will be impacted by the type of genre and movie stars are in a movie. Also in his thesis, he mentions that word-of-mouth interactions with say your friends or family may have a greater impact on watching a movie rather than a random review online that you may not deem trustworthy (Pentheny, 2015). Again, this theory was agreed upon by our group, as most people would rather trust movie sources from close friends and family rather than outsiders. In addition, friends and families are likely to have the same taste in entertainment.

Dataset Description:

The way we thought about tackling this problem was first through a survey. Originally, we considered using existing databases from websites such as Kaggle and started data cleansing ahead of time. However, it was apparent that that would be too difficult as not a lot of existing databases go with our envision of the data we needed. So what our team did was release a survey through Google Forms, containing specific questions with our intention to follow a certain theory model, the Elaboration Likelihood Model. This model is based on human behavioral choices and actions.

From there, we released the survey via email links, posters, the spread of word, other media, and etc. We gained 60+ responses over the course of the month and imported the results into an excel spreadsheet once we felt it was good to close the survey out from the public.

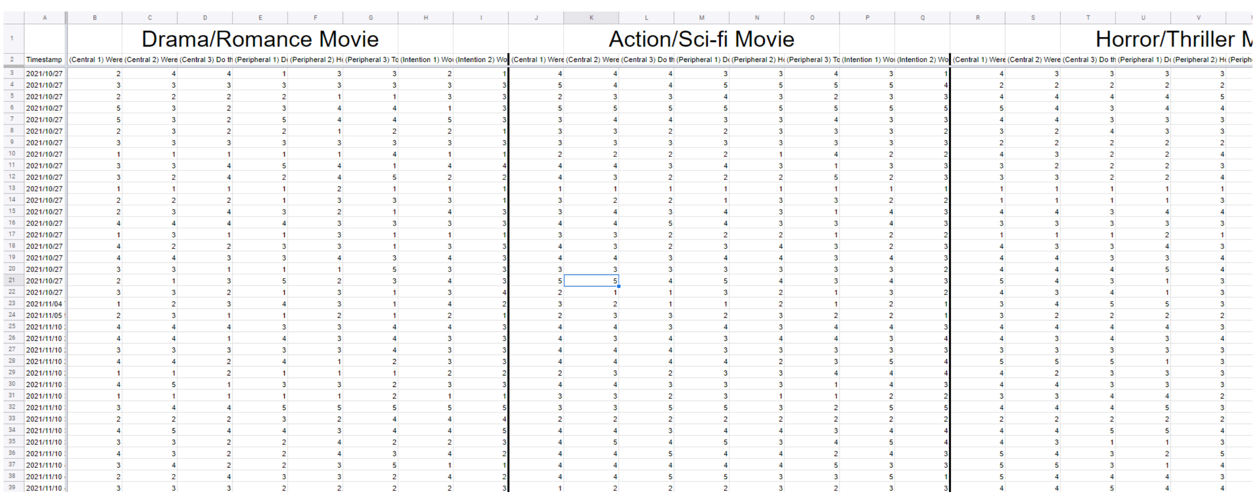


Figure 2: Initial dataset when imported from Google Forms (with a slight organization based on genre)

Elaboration Theory Model:

This model contains three key factors in making a final decision. There is the central, peripheral, and intention route to a final decision. First, a person must have either a logical understanding of something or non-logical. From this, logical knowledge is tied toward a central route of thinking while peripheral thoughts are associated with non-logical thoughts.

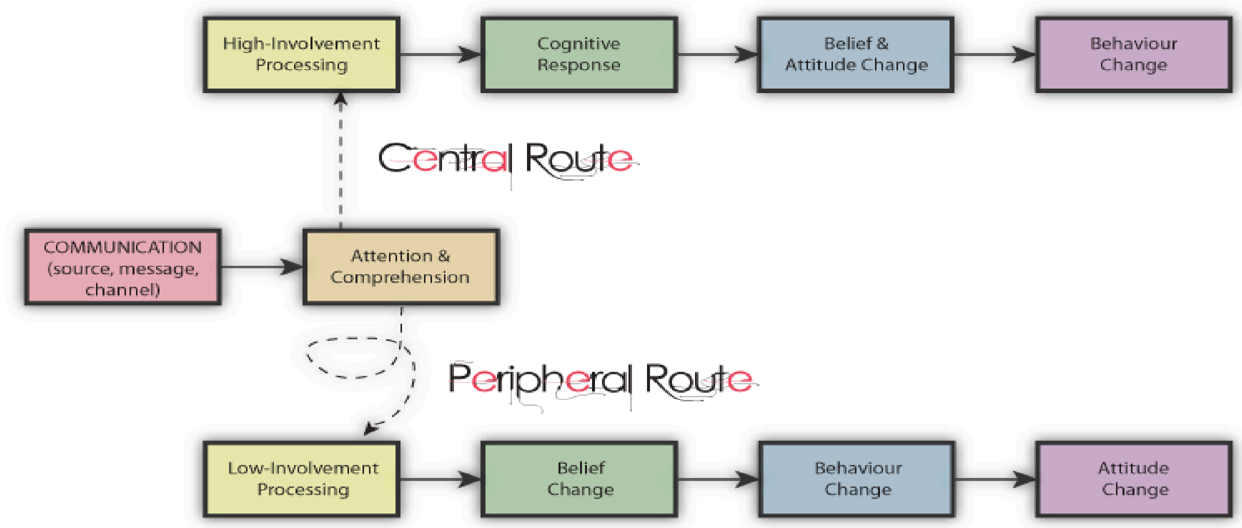


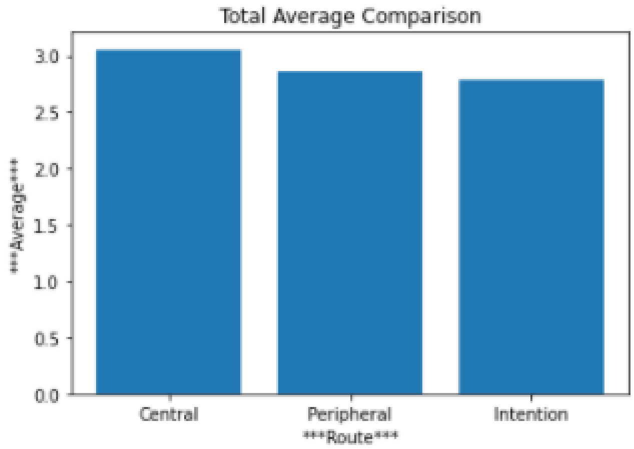
Figure 3: Elaboration Theory Model that describes the Central and Peripheral routes after inputting the message to the user from "Elaboration Likelihood Model (ELM): Article 12/8/2011"

Methodology and Exploratory Data Analysis:

After getting the responses from our survey in the google form and data cleansing in Excel sheets by separating the central, peripheral and intention questions based on the movie genres that were displayed in the survey. It was 3 hidden movies, one was romance and drama, the second one was action and sci-fi and the third one was horror and thriller. Next, we combined the hidden movies reviews central into one, so we can get the total average of the central questions. Same with peripheral and intention questions in the excel sheet. The total average for central was about 35.1%, the peripheral was 32.9% and the intention was 32.0%. Meaning the central was the bigger average compared to the other routes in our project.

Final Data Modeling:

```
barGraph('Total Average Comparison', r_central, r_peripheral, r_intention)
```



Results Discussion:

Based on the results we collected, we can conclude that our theory based on the hidden movie review was correct after getting the results from the survey and from the data that we received from the excel sheet.

Conclusion:

Our conclusion of the project is that Hidden Movie reviews WILL positively be associated with people's intention to watch that certain movie. All thanks for the data was shown in the excel sheets and Jupyter Notebook script, which helped us with the bar graphs with the data we used.

References:

"Elaboration Likelihood Model (ELM): Article 12/8/2011: @Pbworks #Persuasion #Attitudes: Marketing Professor, Persuasion, Elaborate." Pinterest, https://www.pinterest.com/pin/9359111702398478/.

Geddes, John. "Elaboration Likelihood Model Theory – Using Elm to Get inside the User's Mind." The Interaction Design Foundation, https://www.interaction-design.org/literature/article/elaboration-likelihood-model-theory-using-elm-to-get-inside-the-user-s-mind.

Pentheny, Jacob R. "University of New Hampshire University of New Hampshire ..." The Influence of Movie Reviews on Consumers, University of New Hampshire, 2015, https://scholars.unh.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1267&context=honors.

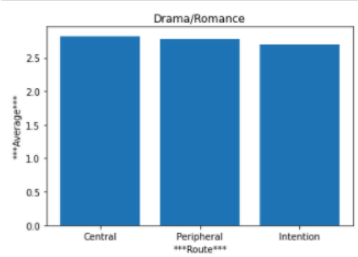
Appendix:

```
In [129]: # Gisele T., Yanira M., Mohamed B.
          # 12/09/2021
          # Final Project

In [130]: import sys
          import pandas as pd
          import matplotlib.pyplot as plt

In [131]: hms = pd.read_excel('pre-Hidden Movie Critic Survey TEST.xlsx')

In [132]: hms.info()
```

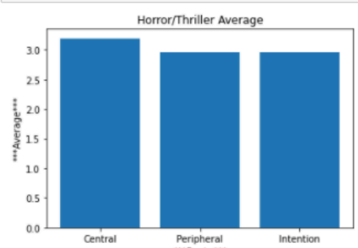
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 52 entries, 0 to 51
Data columns (total 63 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Movie Genre               24 non-null     float64
 1   Question                  51 non-null     object
 2   2021/10/27 10:05:14 AM EST    48 non-null     float64
 3   2021/10/27 10:05:16 AM EST    26 non-null     float64
 4   2021/10/27 10:05:36 AM EST    36 non-null     object
 5   2021/10/27 10:05:38 AM EST    33 non-null     float64
 6   2021/10/27 10:05:39 AM EST    24 non-null     float64
 7   2021/10/27 10:05:53 AM EST    24 non-null     float64
 8   2021/10/27 10:05:55 AM EST    34 non-null     object
 9   2021/10/27 10:06:04 AM EST    34 non-null     object
 10  2021/10/27 10:06:04 AM EST.1  34 non-null     object
 11  2021/10/27 10:06:07 AM EST    24 non-null     float64
 12  2021/10/27 10:06:25 AM EST    24 non-null     float64
 13  2021/10/27 10:06:33 AM EST    24 non-null     float64
```

```
In [133]: '''Drama/Romance Movie Average'''
          c1 = hms.iloc[0:3].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          p1 = hms.iloc[3:6].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          i1 = hms.iloc[6:8].drop(['Movie Genre', 'Question'], axis=1).mean().mean()

In [134]: '''Action/Sci-Fi Movie Averages'''
          c2 = hms.iloc[8:11].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          p2 = hms.iloc[11:14].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          i2 = hms.iloc[14:16].drop(['Movie Genre', 'Question'], axis=1).mean().mean()

In [135]: '''Horror/Thriller Movie Averages'''
          c3 = hms.iloc[16:19].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          p3 = hms.iloc[19:22].drop(['Movie Genre', 'Question'], axis=1).mean().mean()
          i3 = hms.iloc[22:24].drop(['Movie Genre', 'Question'], axis=1).mean().mean()

In [136]: def barGraph(Topic, y1, y2, y3):
              x_values = ["Central", "Peripheral", "Intention"]
              y_values = [y1, y2, y3]
```

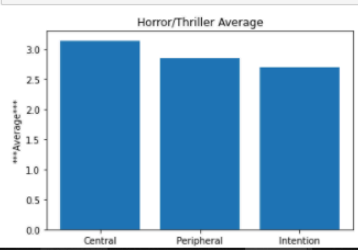
```
barGraph('Drama/Romance', c1, p1, i1)
```




```
barGraph('Action/Sci-fi Average', c2, p2, i2)
```

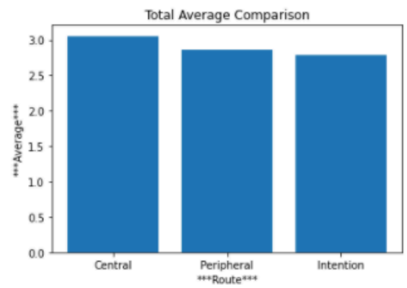



```
barGraph('Horror/Thriller Average', c3, p3, i3)
```



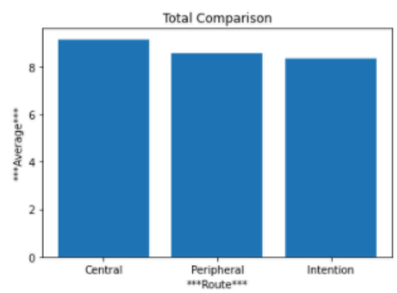
```
'''Total Average Comparison'''
a_central = (c1+c2+c3)/3
a_peripheral = (p1+p2+p3)/3
a_intention = (i1+i2+i3)/3
```

```
barGraph('Total Average Comparison', r_central, r_peripheral, r_intention)
```




```
'''Total Comparison'''
t_central = c1+c2+c3
t_peripheral = p1+p2+p3
t_intention = i1+i2+i3
```

```
barGraph('Total Comparison', t_central, t_peripheral, t_intention)
```



Movie Genre	Question	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/27	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/10/2	2021/11/0
1	C1	2	3	2	5	5	2	3	1	3	3	1	2	2	4	1	4	4	3	2	3	1
1	C2	4	3	2	3	3	3	3	1	3	2	1	2	3	4	3	2	4	3	1	3	2
1	C3	4	3	2	2	2	2	3	1	4	4	1	2	4	4	1	2	3	1	3	2	3
1	P1	1	3	2	3	5	2	3	1	5	2	1	1	3	4	1	3	3	1	5	1	4
1	P2	3	3	1	4	4	1	3	1	4	4	2	3	2	3	3	3	4	1	2	3	3
1	P3	3	3	1	4	4	2	3	4	1	5	1	3	1	3	1	1	3	5	3	1	1
1	I1	2	3	3	1	5	2	3	1	4	2	1	3	4	3	1	3	4	3	4	3	4
1	I2	1	3	3	3	3	1	3	1	4	2	1	1	3	3	1	3	3	3	3	4	2
2	C1	4	5	2	5	3	3	3	2	4	4	1	3	3	4	3	4	4	3	5	2	3
2	C2	4	4	3	5	4	3	3	2	4	3	1	2	4	4	3	3	4	3	5	1	2
2	C3	4	4	3	5	4	2	3	2	3	2	1	2	3	5	2	2	3	3	4	1	1
2	P1	3	5	4	5	3	2	3	2	4	2	1	1	4	4	2	3	4	3	5	3	1
2	P2	3	5	3	5	3	3	3	1	3	2	1	3	3	3	2	4	4	3	4	2	2
2	P3	4	5	2	5	4	3	3	4	1	5	1	3	1	3	1	3	3	3	3	1	1
2	I1	3	5	3	5	3	3	3	2	3	2	1	2	4	4	2	2	4	3	4	3	2
2	I2	1	4	3	5	3	2	3	2	3	3	1	2	3	3	2	3	3	2	3	2	1
3	C1	4	2	4	5	4	3	2	4	3	3	1	1	4	3	1	4	4	4	5	4	3
3	C2	3	2	4	4	4	2	2	3	2	3	1	1	4	3	1	3	4	4	4	3	4
3	C3	3	2	4	3	3	4	2	2	2	2	1	1	3	3	1	3	3	4	3	4	5
3	P1	3	2	4	4	3	3	2	2	2	2	1	1	4	3	2	4	3	5	1	1	5
3	P2	3	2	5	4	3	3	2	4	3	4	1	3	4	3	1	3	4	4	3	3	3
3	P3	3	2	3	4	3	3	2	2	1	2	1	2	1	3	3	2	3	5	1	2	2
3	I1	2	2	4	3	4	1	2	1	2	1	1	2	4	2	1	3	4	2	2	1	4
3	I2	1	2	3	3	4	1	2	2	2	2	1	1	3	2	1	3	3	2	5	1	2

Drama/Romance Average			Drama/Romance Average					
C1	2.836066	2.836	Central	2.825137		C	P	I
C2	2.95082	2.574	Peripheral	2.786885		2.84	2.69	2.84
C3	2.688525		Intention	2.704918		2.95	2.84	2.57
P1	2.688525					2.69	2.84	2.7
P2	2.836066					3.34	3.16	3.18
P3	2.836066					3.26	3.02	2.75
I1	2.836066					2.98	2.7	2.97
I2	2.57377					3.46	2.75	2.77
Action/Sci-fi Average			Action/Sci-fi Average			3.13	3.15	2.62
C1	3.344262		Central	3.196721		2.85	2.64	2.7
C2	3.262295		Peripheral	2.961749				
C3	2.983607		Intention	2.967213				
P1	3.163934							
P2	3.016393							
P3	2.704918							
I1	3.180328							
I2	2.754098							
Horror/Thriller Average			Horror/Thriller Average					
C1	3.459016		Central	3.147541				
C2	3.131148		Peripheral	2.846995				
C3	2.852459		Intention	2.696721				
P1	2.754098							
P2	3.147541							
P3	2.639344							
I1	2.770492							
I2	2.622951							