Carina Vallefuoco

ST 472 01

I. INTRODUCTION:

Courtney Datin and I worked as partners to analyze movie theme data. The motivation for this analysis was to create a model that allows us to group people based on what types of movies they like, their age, and their gender.

II. STATEMENT OF THE PROBLEM/QUESTION:

The expected result of the study is to find the connection between movies themes. In other words, people who like a certain movie theme tend to like another movie theme. For example, people who like scary movies tend to like suspenseful movies and people who like romantic movies tend to like relationship movies.

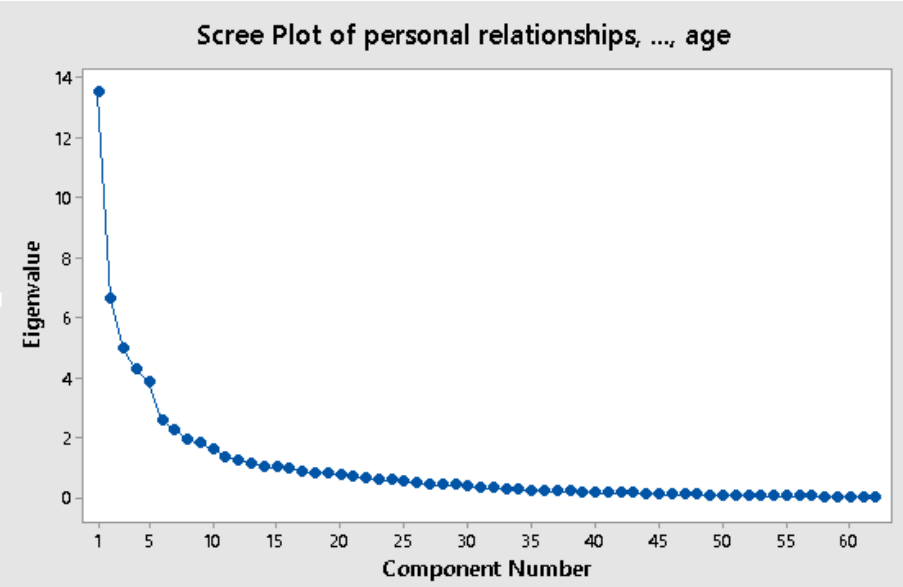
III. METHOD: The data was collected via survey. The survey prompted subjects to rate movie themes from 0 to 10 (0 being the least liked and 10 being their favorite) and also to include their age and gender.

SUBJECTS: The study of 82 people (33 males and 49 females) consists of Loyola students, their friends, and their families.

VARIABLES: The study measured 62 different movie themes, age, and gender.

IV. ANALYSIS PREP:

The first step in our analysis is to figure out how many principal components/factors we need to consider. We ran a principal component analysis to obtain a scree graph and to collect the top 10 eigenvalues.



We can see the elbow of the scree occurs around principal component 7. So initially we chose k, the number of factors, to be 7. Principal component 7 explains 61.3% of the cumulative variance. Ideally, the PC should explain around 75% but 61.3% is sufficient.

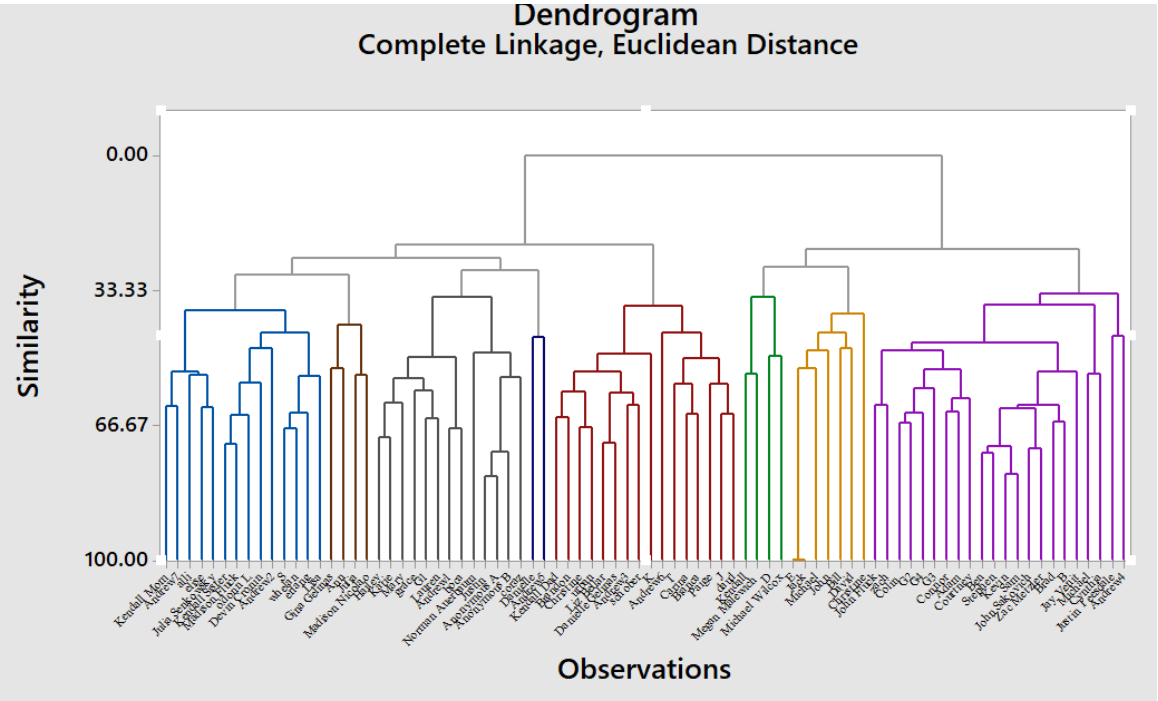
Next, we ran three rotated factor analysis using varimax for k, k-1, and k+1. This allowed us to see which variables where explained by each factor and how heavily they weighed on those factor. We

stayed away from having factors that only explained one variable because we knew it was not going to explain much and wouldn't be helpful in the cluster analysis.

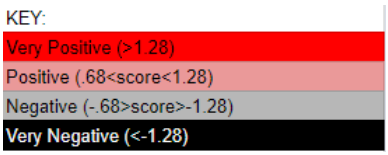
When k=7, love and divorce were explained by factor 1 which doesn't make much sense because they seem to be opposites. When k=6, there is no factor that accounts for humor so we ruled out factor 6. When k=8, factor 8 accounts for humor. We also can see when k=8 that factor 1 is explicitly explaining love related movie themes (and not divorce.) We also gain a factor that can be interpreted as kid friendly. We chose to consider eight factors since the factors are more distinct and we gain a factor when k=8.

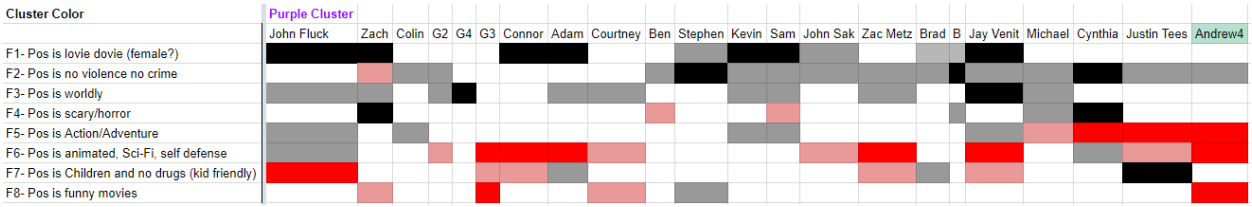
V. ANALYSIS RESULTS:

After the PCA and FA we categorized the factors as the following: Factor 1 'Lovie Dovie', Factor 2 'Anti-violence/anti-crime', Factor 3 'Worldly', Factor 4 'Scary', Factor 5 'Action', Factor 6 'Animation/sci-fi', Factor 7 'kid friendly', and Factor 8 'Humor'. We then ran a cluster analysis and generated a dendrogram with a similarity level of 30. We were able to see eight clusters which helped us interpret each grouping of subjects. We did not include the cluster of only two people in our analysis because we felt it was insignificant.

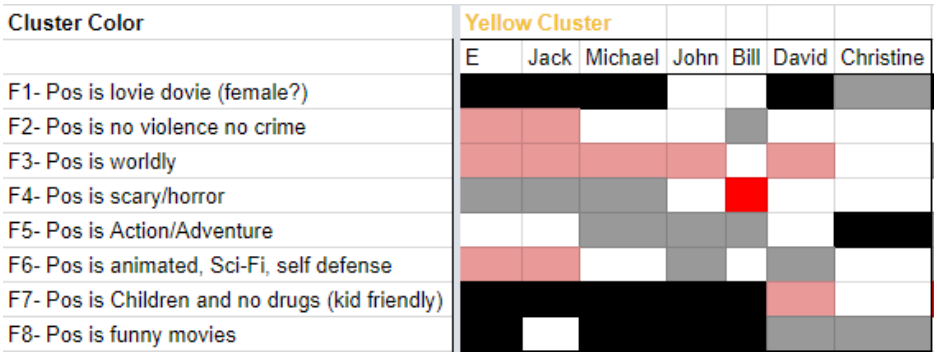


The purple cluster contains more subjects than any other cluster. People in this cluster generally have negative F1 scores, negative F2 scores, negative F3 scores, and positive F6 scores. This means that subjects in this cluster do not like lovie dovie movies, like violent and crime movies, they do not like worldly movies, and they do like animated sci-fi movies.



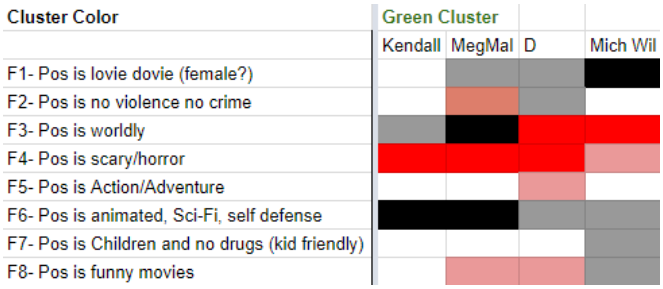


The subjects in the yellow cluster generally have negative F1 scores and negative F8 scores. This means that subjects in this factor do not like funny/comedy movies and also do not like love movies. Despite Christine, it makes sense that there are negative F1 scores in this cluster because the majority of subjects are male.

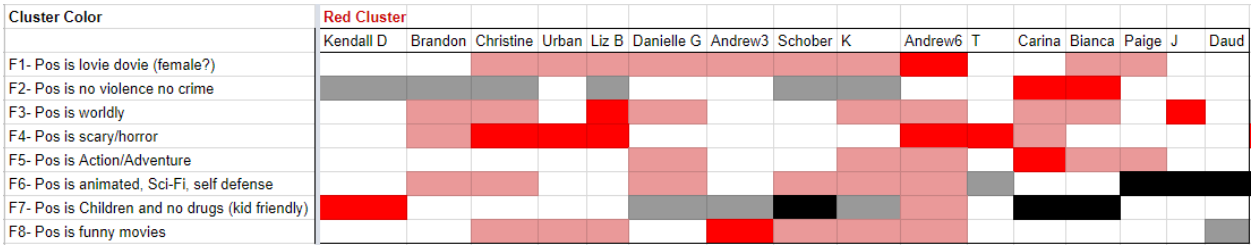


The subjects in the green cluster all have negative F4 scores and positive F6 scores.

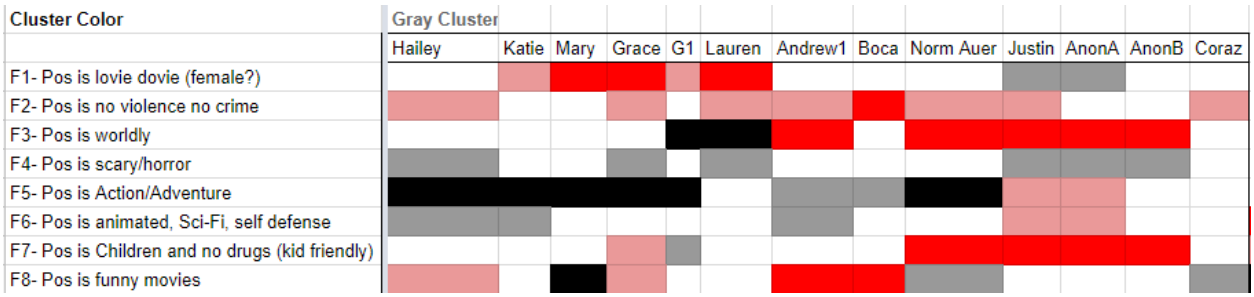
Subjects in this factor really like scary/horror movies and do not like animated/Sci-Fi/self-defense movies.



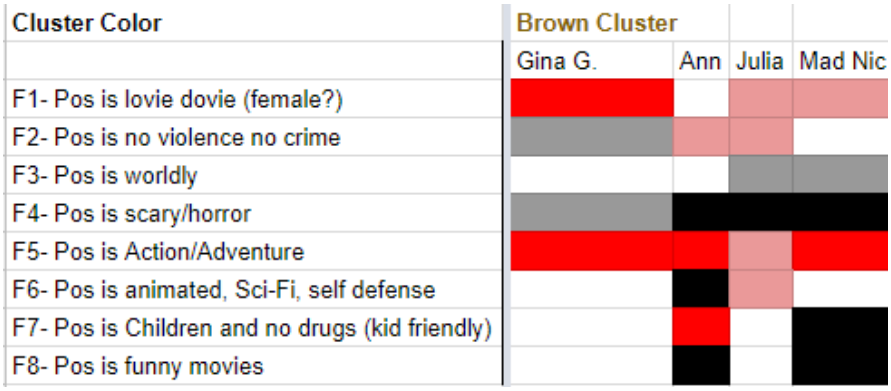
The subjects in the red cluster generally have positive F1 scores, positive F3 scores, and positive F4 scores. This means that the subjects in this cluster like lovie dovie movies, like worldly movies, and they also like scary movies



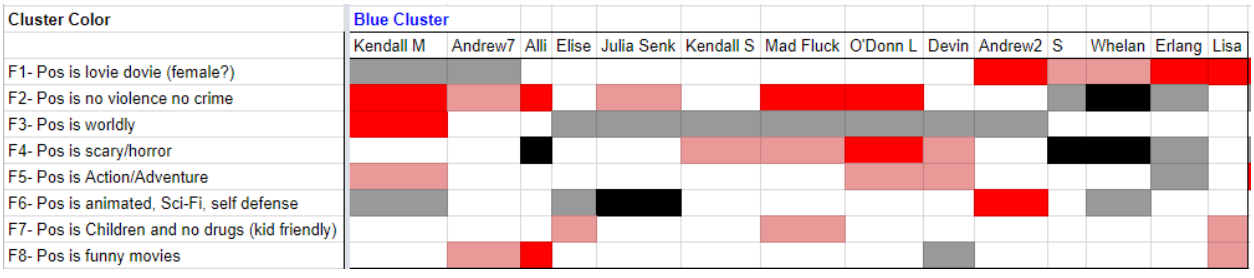
The subjects in the gray cluster all have positive F2 scores. This means the subjects in this cluster do not like movies with violence or crime.



The subjects in the brown cluster all have negative F4 scores and positive F5 scores. This means that subjects in the brown cluster dont like scary movies but do like action and adventure movies.



The subjects in the blue cluster are difficult to characterize. The people in this cluster are most likely the people who did not fit into any other cluster.



VI. ADDITIONAL ANALYSIS OR GRAPHS OR TABLES:

We saw that in the left half of the dendrogram the blue, brown, gray, navy and red clusters generally had positive F1 scores while the green, yellow and purple clusters on the right half had negative F1 scores. In the rotated factor analysis gender loads on F1 where a positive F1 score suggests female and a negative F1 score suggests male. 42 out of 49 of the subjects on the left half are female and 26 out of 33 of the subjects on the right half are male. This confirms our expectations.

V. POTENTIAL IMPACT:

The overall upshot of our analysis was that females really like love related movie themes including found love, romantic, lost love, cute, new love, and personal relationships. Going forward, I would extend the analysis to include the subject's music genre preferences. The variables would still include age, gender, and an even amount of movie themes and music genres. I think it would be interesting to see the connection between movie themes and music genres. I assume subjects who like love movie tend to like country movies and subjects who like scary movies tend to like heavy metal music.