Identifying popular movie genres & recommending movies for students within IIITD&M-K

By

S.Srikrishnan – ME21B2008 | G.L.Ram Krishna – ME21B2010

Table of Contents

S.No	Title	Page Number	
1	Collected Data		
2	Data Cleaning Code		
3	Cleaned Data		
4	Data Visualization		
5	Analysis/Inference		
6	Data Analytics		

Collected Data

- A google form was created & circulated within the IIITDM-K student fraternity on Sunday the 5th of November '23.
- So far, we've received 76 responses during a span of 5 days.
- https://docs.google.com/spreadsheets/d/1anRMyCv8fkoKuwGIYL7iJHdDsVwrGOZafgwVVKRE6YQ/edit?usp=sharing

Data Cleaning Code

☐ The mentioned code has been implemented in order to clean the data.

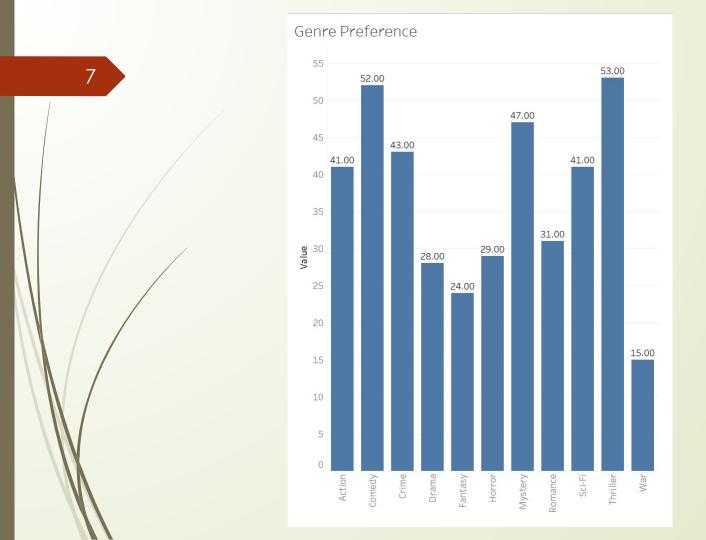
Cleaned Data

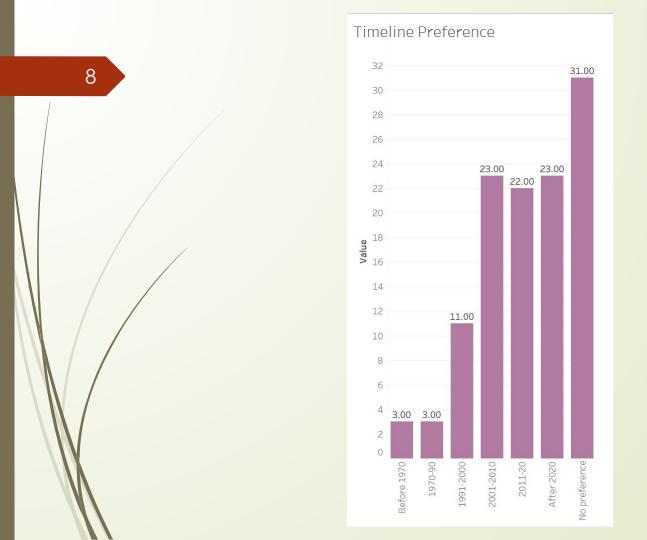
Below presented in the cleaned data.

S.No	How old a	Comedy	Fantasy	Romance	Drama	Action
1	20	1	1	0	1	1
2	30	0	0	0	0	0
3	20	1	0	1	0	1
4	20	1	0	0	0	1
5	19	0	1	0	0	1
6	20	0	1	0	0	0
7	20	1	1	1	1	1
8	20	1	1	1	0	0
9	20	1	0	0	1	0

Data Visualization

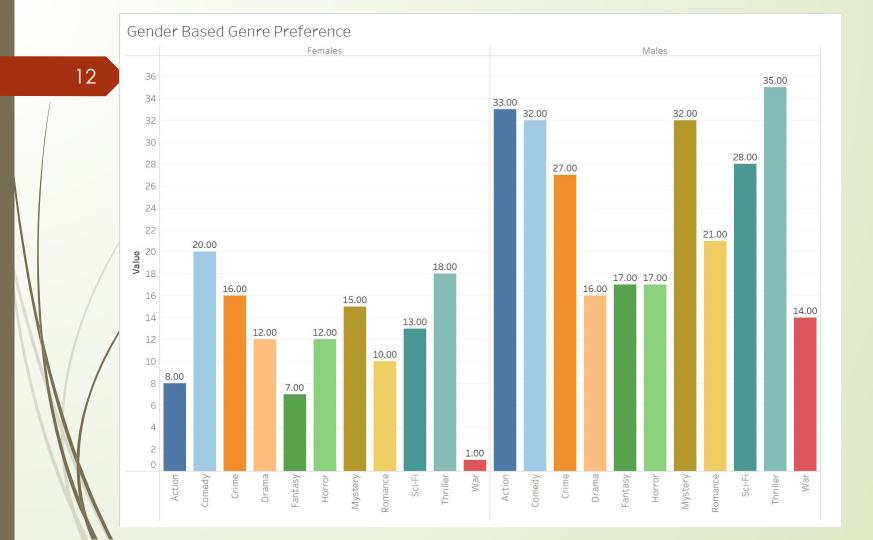
- The cleaned data was fed into Tableau for data visualization.
- Below presented are the results.

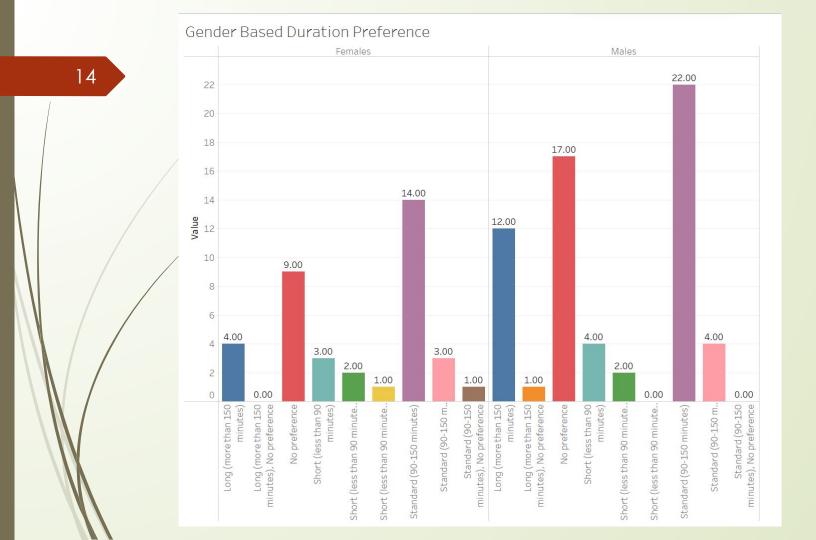


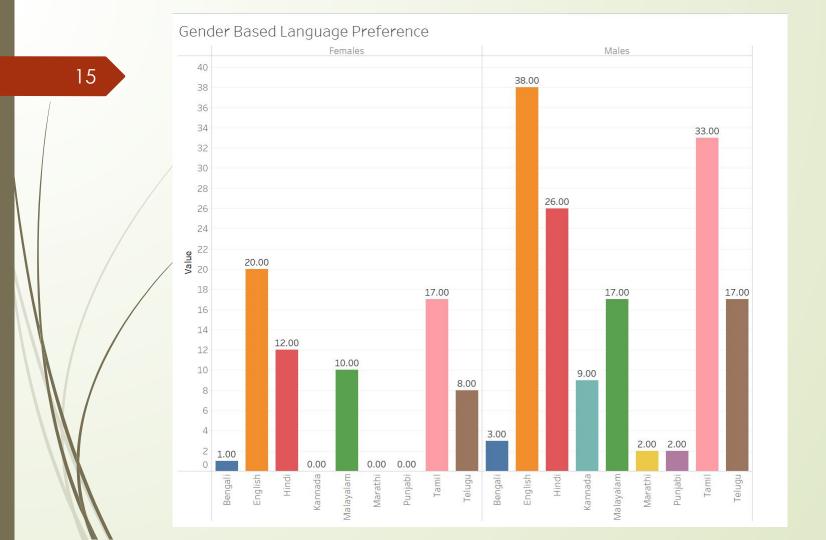












Age Based Genre Preference 18 19 20 21 23 24 25 26 19.00 19.00 20 16.00 16.00 16 11.00 12 Value 10 10.00 9.00 9.00 How old are you?: 20 8.00 Thriller: 19.00 7.00 5.00 5.00 4.00 4.00 4.00 4.00 4.00 3.00 3.00 3.00 3.00 3.00 2.00 2.00 2.00 2.00 2.00 2.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 Comedy Crime Drama Fantasy (Horror Mystery Romance Sci-Fi Comedy Crime Drama Fantasy Horror Mystery Romance Sci-Fi Thriller Comedy Comedy Drama Horror Romance Thriller Comedy Drama Horror Romance Thriller Comedy Drama Horror Romance Thriller Cornedy Drama Horror Romance Thriller Comedy Drama Horror Thriller Drama Horror Romance Thriller Comedy Drama Horror Romance Thriller Comedy Drama Horror Thriller

Romance

Romance

Age Based Timeline Preference 20 22 18 19 21 23 24 25 26 10 9.000 9.000 9 8.000 8 7.000 6 5.000 Value 5 4.000 13.000 13.000 3.000 3.000 3.000 3.000 3.000 3.000 3.000 3 2.000 2.000 2.000 2.000 2.000 2.000 2.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 1970-90 1970-90 Before 1970 1970-90 1970-90 Before 1970 2001-2010 2011-2020 Before 1970 1970-90 2001-2010 Before 1970 1970-90 Before 1970 Before 1970 2011-2020 Before 1970 2001-2010 2011-2020 Before 1970 1970-90 2001-2010 2011-2020 1970-90 2001-2010 2011-2020 Before 1970 1970-90 2001-2010 2011-2020 1970-90 2001-2010 2011-2020 2011-2020 2001-2010 2011-2020 2001-2010 2011-2020 2001-2010 Before 1970

Age Based Language Preference 19 20 21 22 23 24 25 26 21.00 22 20 18 16.00 16 13.00 14 Value 11.00 10.00 10 8.00 7.00 8 6.00 6 4.00 4.00 3.00 3.00 2.00 2.00 2.00 2.00 1.00 2 1.00 1.00 1.00 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 Kannada 0.00 Tamil 0.00 Marathi 0.00 0 English Tamil Marathi Marathi Marathi English English Tamil English Kannada English Kannada Tamil Kannada Marathi Kannada Tamil English Tamil English Kannada English Marathi English Kannada Marathi Marathi Tamil Kannada Marathi Marathi Tamil English Tamil Kannada Tamil Kannada

Age Based Duration Preference 20 22 23 26 19 21 24 25 14 13 12 11 10 9 8.00 Value 6.00 6 5.00 5 4 3.00 3.00 3.00 3.00 3 2.00 2.00 2.00 2.00 2.00 2.00 2.00 2 0.00 0.00 0.00 1.00 1.00 1.00 1.00 1.00 0.00 0.00 0.00 00.0 0.00 0.00 0.00 Standard (90-150 minutes.. Short (less than 90 minute. Long (more than 150 minu. Short (less than 90 minute. Standard (90-150 minutes. Short (less than 90 minute. Short (less than 90 minute. Standard (90-150 minutes. Short (less than 90 minute. Short (less than 90 minute. Standard (90-150 minutes. Short (less than 90 minute. Short (less than 90 minute. Standard (90-150 minutes. Short (less than 90 minute. Short (less than 90 minute. Long (more than 150 minu. Short (less than 90 minute. Short (less than 90 minute. Standard (90-150 minutes. Long (more than 150 minu. Short (less than 90 minute. Short (less than 90 minute. Long (more than 150 minu. Short (less than 90 minute. Short (less than 90 minute. Short (less than 90 minute. Long (more than 150 minu. Long (more than 150 minu.

Analysis/Inference

- From the above displayed data visualization, the following inferences have been made:
- Thriller>Comedy>Mystery>Crime>Sci-Fi ==Action>Romance>Horror> Drama>Fantasy>War is the genre preferred by the IIITD&M-K student fraternity.
- ❖ It can be inferred that students don't consider the period a movie belongs to. They carve for just quality content.
- Also, other preferred timelines are 2001-2010==After 2020>2011-2020>1991-2000>1970-1990==Before 1970.
- ♦ 2001-2010 being equal to After 2020 and being preferred more than movies released during 2011-2020 is quite surprising.
- English>Tamil>Hindi>Malayalam>Telugu>Kannada>Bengali>Marathi is the language preference of the IIITD&M-K student fraternity.
- * Malayalam being preferred more than Telugu is an anomaly considering the significant Telugu Population on the campus.

- This can be attributed to the new generation (Malayalam film movement).
- ❖ We received a total of 48 male & 22 female responses. Considering the minimal population of females on the campus, this is indeed a good number of female participants and male: female ratio turns out to be 2.18.
- ❖ People generally prefer Standard>Long>Short. Long being preferred over short shows that people are ready to invest time as long as the content is engaging. People are craving to be relaxed and get entertained.
- A significant population also are open to duration and doesn't bother over duration.
- Amongst Females Comedy>Thriller>Crime>Mystery>Sci Fi>Horror==Drama>Romance>Action> Fantasy>War is the preferred genre order.
- Amongst Males, Thriller>Action>Comedy==Mystery>Sci-Fi>Crime>Romance>Fantasy==Horror>Drama>War is the preferred genre order.
- This shows that war is least preferred among both males & females. Drama is more preferred among females than males and Action is preferred more among males than females.
- ♦ Both males & females don't bother about the period of the film. The individual data shows that females, only prefer After 2020> 2011-2020==2001-2010>1991-2000.
- The individual data for males shows that 2001-2010>After 2020== 2011-2020>1991-2000>1970-1990>before 1990. This shows that men like to explore different types of cinemas.

- Amongst Females English>Tamil>Hindi>Malayalam>Telugu>Bengali. Whereas Males prefer English>Tamil>Hindi>Malayalam==Telugu> Kannada>Bengali>Marathi==Punjabi. This shows that men want to explore content from different languages than females.
- The age statistics show us that we've received the highest number of responses from the age group of 19 & 20.

Data Analytics

- ☐ The movie recommendation system is a content based recommendation system which use K- nearest neighbors machine learning algorithm to recommend a movie.
- The recommendation system finds the similarity between movies through genres with the help of cosine function.
- This data is then fed to the KNN algorithm with this as a metric. Simultaneously, the TF-IDF (frequency vs inverse frequency matrix) is fit into it allowing the it to learn how to find similar movies based on profiles of the movies (supervised learning).
- ☐ The KNN algorithm finds K nearest neighbours near the input data.
- ☐ The TF-IDF vectorization is NLP used to transform data into numerical features.

