Kalyani Government Engineering College
(Affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal)

Kalyani - 741235, Nadia, WB

Project Report on

MOVIE RECOMMENDATION AND PERSONALITY ANALYSIS BASED ON MOVIE/SHOW PREFERENCES

(A dissertation submitted in partial fulfilment of the requirements of Bachelor of Technology in Computer Science and Engineering of Maulana Abul Kalam Azad University of Technology, West Bengal)

Submitted by

ARKAJYOTI BHATTACHARYAY
University Roll Number: 10200119050
Registration Number: 037801 of 2019-20
Session 2019-2023
6th Semester

Under the guidance of
Prof. Swapan Kumar Mondal

Professor,
Department of Computer Science and Engineering

2019-2023

কল্যাণী - ৭৪১ ২৩৫
নদীয়া, পশ্চিমবঙ্গ

Kalyani 741 235
Nadia, West Bengal,India

পত্রাঙ্ক /Ref. No.:
তারিখ / Date :

কল্যাণী গভঃ ইঞ্জিনিয়ারিং কলেজ
Kalyani Government Engineering College
(Govt. of West Bengal)

Certificate of Approval

This is to certify that the project report on "Movie Recommendation and Personality Analysis Based on Movie Show Preferences" is a record of bonafide work, carried out by Arkajyoti Bhattacharyay under my guidance and supervision.

In my opinion, the report in its present form is in conformity as specified by Kalyani Government Engineering College and as per regulations of the Maulana Abul Kalam Azad University of Technology. To the best of my knowledge, the results presented here are original in nature and worthy of incorporation in the project report for the B.Tech. program in Computer Science and Engineering.

_____

Signature of Supervisor

Name and affiliation: Professor Swapan Kumar Mondal, Dept. of CSE

_____

Signature of Head, Dept. of CSE

Declaration by the student

I Arkajyoti Bhattacharyay, a student of B.Tech, CSE 3nd year declare that I have submitted this report in partial fulfilment of the requirements of Bachelor of Technology in Computer Science and Engineering of Maulana Abul Kalam Azad University of Technology, West Bengal.

I solemnly declare that the project report on "Movie Recommendation and Personality Analysis Based on Movie Show Preferences" is based on my own work carried out during the course of our study under the supervision of Professor Swapan Kumar Mondal sir.

I assert the statements made and conclusions drawn are an outcome of my research work. I further certify that

I. The work contained in the report is original and has been done by me under the general supervision of my supervisor.

II. The work has not been submitted to any other Institution for any other degree/diploma/certificate in this university or any other University of India or abroad.

III. We have followed the guidelines provided by the university in writing the report.

IV. Whenever we have used materials (data, theoretical analysis, and text) from other sources, we have given due credit to them in the text of the report and giving their details in the references.

_____

Signature

Name of the student: Arkajyoti Bhattacharyay

University Roll Number: 10200119050

Date: 03.06.2022

ACKNOWLEDGEMENT

I would like to express my profound gratitude and deep regards to my project guide Professor Swapan Kumar Mondal (Professor, Department of Computer Science and Engineering, Kalyani Government Engineering College, West Bengal) for his encouragement and inspiration in this project. This project would never have been successful without his great guidance.

I would also like to specially thanks to Dr. Kousik Dasgupta (Assistant Professor, Department of Computer Science and Engineering, Kalyani Government Engineering College, West Bengal) for his advice and valuable suggestions.

_____

Signature

Name of the student: Arkajyoti Bhattacharyay

Roll No: 10200119050

ABSTRACT

This project work is about Movie Recommendation and Personality Analysis Based on Movie Show Preferences. The main goal of our project is to conduct a survey to collect sample datasets about popular interests. To make this project more advanced, we will remove the outliers and make a more generalized model to analyse their preferences and implement those to a recommendation system to produce desired recommendation of movies. This project can be used further for developing a personality profile to make these recommendations even better.

CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Motivation

Everyone loves movies irrespective of age, gender, race, colour, or geographical location. We all in a way are connected to each other via this amazing medium. Yet what most interesting is the fact that how unique our choices and combinations are in terms of movie preferences. Some people like genre-specific movies be it a thriller, romance, or sci-fi, while others focus on lead actors and directors. When we take all that into account, it's astoundingly difficult to generalize a movie and say that everyone would like it. But with all that said, it is still seen that similar movies are liked by a specific part of the society.

1.2 Background

The basics of recommender systems were founded by researches into cognition science and information retrieval, and its first manifestation was the Usenet communication system created by Duke University in the second half of the 1970s, where users were able to share textual content with each other. These were categorized into newsgroups and subgroups for easier search, but it was not directly built on or targeted the preferences of users.

The first known such solution was the computer librarian Grundy, which first interviewed users about their preferences and then recommended books to them considering this information. Based on the information collected the system allocated the user into a stereotype group using a rather primitive method, thus recommending the same books to all persons in the same group. For more information about the results of Grundy's solution and its popularity among users, see Rich's 1979 article. This approach may seem a little outdated today, but at the time it was a paradigm shift in automated services, since it was personalised. It is important to note that this milestone has not been reached by all web shops, even now.

Basically, two very different directions of recommender systems have evolved over time: collaborative filtering and content-based filtering. The former attempts to map (profile) the taste of users and offers content to them that users with similar preferences liked. The content-based filtering is about knowing the dimensions of the entity to be recommended (for example, a musical content recommendation system can consider the following dimensions: style, artist, era, orchestration, etc.) and the user's preferences for these dimensions or characteristics. Thus, every time a user likes another song, this new information is added to their profile.

The first example of collaborative filtering and also the origin of the term was the Tapestry system developed by Xerox PARC, which allowed its users to take notes and comment on the documents they were reading (initially in binary form: liking or disliking it). Therefore, users could not only use the content of the documents to manually narrow their search, but on the basis of notes and reviews from other users, which once reaching an appropriate number of users, was able to rank the thematic documents rather well on the basis of their relevance and usefulness.

The first solution to combine collaborative and content-based filtering solutions was Fab, developed by Stanford students, presented in 1994. They point out that their objective with the hybrid system is to eliminate the disadvantages of the two procedures which became known by that time. Their model consists of two basic processes: first they collect content for specific topics (such as websites or articles about financial topics), then for each individual user they select those items collected from specific subjects which highly likely will interest them specifically and finally these contents will reach the user.

1.3 Summary of present work

Combining the two approaches can be conceived in several ways, one procedure can be embedded in another, as Fab's example shows, or it is possible to give a joint recommendation as a result of the two procedures, as Netflix does. Netflix's algorithm, CineMatch was the most successful recommender system for online movie sales in the early 2000s. It was a serious catalyst for such research and the scientific field – which only started its independent existence in the 90s – started

rapidly to develop. The 2006 Netflix Award's challenge was to create a recommender algorithm based on the 100 million film reviews made available by them, which makes recommendations at least 10% better than the results of CineMatch. The 1-million-dollar prize in 2009 was awarded for a solution that included 107 different algorithms and mixed their recommendations depending on the circumstances. We can't omit the biggest example of online referral systems today, amazon.com, which recommends products to the user based on a cooperative filtering technique, taking into account previously browsed and purchased products and what they are currently viewing.

1.4 Organization of the thesis

The rest of the thesis is described in this section. The second chapter describes the recommendation system and its importance

The third chapter describes the process of the survey and analysis of the responses

The fourth chapter is the main part where the model of movie recommendation and personality analysis report is explained.

The fifth chapter describes the project result and discussion.

And in the sixth chapter, the project ends with the conclusion and the scope for future work.

1.5 Hardware/Software used

Here is a list of components used in the project.

Hardware Requirements:

- Processor: Intel® Core™ i5
- Memory (RAM): Minimum 8.00 GB

Software Requirements:

- Operating System: Windows 10
- System Type: 64-bit Operating System
- Programming Language: Python 3.6.7

Software used:

- Google forms
- Visual Studio Code
- Anaconda 3
- Jupyter Notebook

CHAPTER 2 OVERVIEW OF THE PROJECT

2.1 Purpose

The purpose of a recommendation system basically is to search for content that would be interesting to an individual. Moreover, it involves a number of factors to create personalised lists of useful and interesting content specific to each user/individual. Recommendation systems are Artificial Intelligence based algorithms that skim through all possible options and create a customized list of items that are interesting and relevant to an individual. These results are based on their profile, search/browsing history, what other people with similar traits/demographics are watching, and how likely are you to watch those movies. This is achieved through predictive modelling and heuristics with the data available.

2.2 Procedure

For the survey, a ideal set of question should be prepared which should mainly consist multiple choice questions or distinct one word answer type question because that will ensure maximum participation in this project.

The responses of the questions can be used for individual recommendation as well as to make a general pattern for a certain criterion by using some statistical model and ML model.

These can also be further analysed for making a decent profile of their taste or their personality up to an extent by taking help from various pre-existing resources.

2.3 Project design

Project design is an early phase of the project lifecycle where ideas, processes, resources, and deliverables are planned out. As this is a survey based project, a survey will be done via google forms to make the collection of the data hassle-free. Google forms also make the pictorial representation of the statistics of the given responses in a very befitting way. It may be used directly in some cases but

here we are going to use it in a later ML model as an input to make the recommendation system even better.

Project design is an important piece of executing a successful project. From gathering the necessary information and resources, it hepls to bring the details to life. With the right project design, one can tackle anything that comes in the way.

2.4 Limitation

The limitations of the project are those characteristics of design or methodology that impacted or influenced the application or interpretation of the results of your study. They are the constraints on generalizability and utility of findings that are the result of the ways in which you chose to design the study and/or the method used to establish internal and external validity.

Simply it can be said that, the limitations of a study are its flaws or shortcomings. Study limitations can exist due to constraints on research design, methodology, materials, etc., and these factors may impact the findings of your study. However, researchers are often reluctant to discuss the limitations of their study in their papers, feeling that bringing up limitations may undermine its research value in the eyes of readers and reviewers.

The following are some major potential methodological limitations that I have faced and can impact the conclusions of the research,

1. Issues with research samples and selection

Sampling errors occur when a probability sampling method is used to select a sample, but that sample does not reflect the general population or appropriate population concerned.

For example, here I conducted a survey to obtain my research results, samples (participants) were asked to respond to the survey questions. However, I saw that I might have had limited ability to gain access to the appropriate type or geographic scope of participants. In this case, the people who responded to my survey questions may not truly be a random sample.

2. Insufficient sample size for statistical measurements

When conducting a study, it is important to have a sufficient sample size in order to draw valid conclusions. The larger the sample, the more precise your results will be. If your sample size is too small, it will be difficult to identify significant relationships in the data.

Here in my case the sample size is very small which makes the identification of the relations even harder.

3. Limited access to data

The research involved surveying certain people or organizations, so I have faced the problem of having limited access to these respondents. Due to this limited access, I have to redesign or restructure your research in a different way. In this case, I have to explain the reasons for limited access and be sure that the finding is still reliable and valid despite this limitation.

4. Time constraints

The time available to study a research problem and to measure change over time might be constrained by such practical issues. The time constraints negatively impacted my study in a way, I acknowledge this and may need for a future study to answer this research problem.

5. Conflicts arising from cultural bias and other personal issues

Researchers might hold biased views due to their cultural backgrounds or perspectives of certain phenomena, and this can affect a study's legitimacy. Also, it is possible that researchers will have biases toward data and results that only support their hypotheses or arguments. In order to avoid these problems, I examined the results without any bias and the data-gathering process was carried out appropriately.

CHAPTER 3 SURVEY CONUCTION

3.1 Overview

Survey research is sometimes regarded as an easy research approach. However, as with any other research approach and method, it is easy to conduct a survey of poor quality rather than one of high quality and real value.

Survey research is common in studies of health and health services, although its roots lie in the social surveys conducted in Victorian Britain by social reformers to collect information on poverty and working-class life, and indeed survey research remains most used in applied social research. The term ‘survey’ is used in a variety of ways, but generally refers to the selection of a relatively large sample of people from a pre-determined population (the ‘population of interest’; this is the wider group of people in whom the researcher is interested in a particular study), followed by the collection of a relatively small amount of data from those individuals. The researcher therefore uses information from a sample of individuals to make some inference about the wider population.

Data are collected in a standardized form. This is usually, but not necessarily, done by means of a questionnaire or interview. Surveys are designed to provide a ‘snapshot of how things are at a specific time’. There is no attempt to control conditions or manipulate variables; surveys do not allocate participants into groups or vary the treatment they receive. Surveys are well suited to descriptive studies, but can also be used to explore aspects of a situation, or to seek explanation and provide data for testing hypotheses. It is important to recognize that ‘the survey approach is a research strategy, not a research method’. As with any research approach, a choice of methods is available and the one most appropriate to the individual project should be used. This paper will discuss the most popular methods employed in survey research, with an emphasis upon difficulties commonly encountered when using these methods.

Advantages and disadvantages of survey research are briefly discussed in the following section,

Advantages:

1. The research produces data based on real-world observations (empirical data).
2. The breadth of coverage of many people or events means that it is more likely than some other approaches to obtain data based on a representative sample, and can therefore be generalizable to a population.
3. Surveys can produce a large amount of data in a short time for a fairly low cost. Researchers can therefore set a finite time-span for a project, which can assist in planning and delivering end results.

Disadvantages:

1. The significance of the data can become neglected if the researcher focuses too much on the range of coverage to the exclusion of an adequate account of the implications of those data for relevant issues, problems, or theories.
2. The data that are produced are likely to lack details or depth on the topic being investigated.
3. Securing a high response rate to a survey can be hard to control, particularly when it is carried out by post, but is also difficult when the survey is carried out face-to-face or over the telephone.

3.2 Questionnaire Layout

Questionnaires used in survey research should be clear and well presented. The use of capital (upper case) letters only should be avoided, as this format is hard to read. Questions should be numbered and clearly grouped by subject. Clear instructions should be given and headings included to make the questionnaire easier to follow.

The researcher must think about the form of the questions, avoiding 'double-barrelled' questions (two or more questions in one, e.g. 'How satisfied were you with your personal nurse and the nurses in general?'), questions containing double negatives, and leading or ambiguous questions. Questions may be open (where the respondent composes the reply) or closed (where pre-coded response options are available, e.g., multiple-choice questions). Closed questions with pre-coded response options are most suitable for topics where the possible responses are

known. Closed questions are quick to administer and can be easily coded and analysed. Open questions should be used where possible replies are unknown or too numerous to pre-code. Open questions are more demanding for respondents but if well answered can provide useful insight into a topic. Open questions, however, can be time consuming to administer and difficult to analyse. Whether using open or closed questions, researchers should plan clearly how answers will be analysed.

For this survey, I have prepared the questionnaire keeping those points in mind. A combination of open and closed questions has been used to ensure the proper parameters are included for a recommendation system besides ensuring a quick analysis. The closed questions are necessary for both individual and general recommendation whereas the open questions will provide the extra parameters needed for a personalized recommendation of an individual.

Layout:

Only 'Full Name', 'Phone number/ Email' is taken as respondents' information.

The closed questions (where pre-coded response options are available) are given below,

'What is your gender?',

'What is your age?',

'What is/are your favourite genre of film?',

'What makes you choose a film?',

'What is important to you in a film?',

''What don't you want to see in a movie?'',

'Do you prefer a male or female lead?',

'How important is the cast in a film?',

'How important is the director to you in a film?',

'How important is the soundtrack to you in a film?',

'How important is the production company/studio to you in a film?',

'How often do you watch a film?',

'Where do you prefer to watch films?',

'Which is/are your favourite animation style?',

'Do you like the following? [Series, Documentaries, TV shows, short films, Anime, Cartoons]',

'What is your favourite genre of Series?',

'What is your favourite genre of Anime?',

'What is/are your favourite Series?',

'What is/are your favourite Documentary?',

'What is/are your favourite TV show?',

'What is/are your favourite Anime?'

The open questions (where the respondent composes the reply) are given below,

'What is your overall favourite film and why? (P.S. You may choose more than one film. For "why" part you may use only keywords or may completely ignore that part.)',

'What is/are your favourite English film?',

'What is/are your favourite film in your national and native language?',

'What is/are your favourite animated film and anime film?'

3.3 Structure of the form

Survey for "Movie Recommendation and Personality Analysis Based on Movie/Show Preferences"

Everyone loves movies irrespective of age, gender, race, color, or geographical location. We all in a way are connected to each other via this amazing medium. Yet what most interesting is the fact that how unique our choices and combinations are in terms of movie preferences. Some people like genre-specific movies be it a thriller, romance, or sci-fi, while others focus on lead actors and directors. When we take all that into account, it's astoundingly difficult to generalize a movie and say that everyone would like it. But with all that said, it is still seen that similar movies are liked by a specific part of the society. The main focus of recommendation system is to filter and predict only those movies which a user would prefer given some data about the user him or herself.

This survey focuses on basic parameters which can help to determine a feasible movie recommendation system as well as a simple personality analysis.

Thanks in advance for taking part in this survey.

P.S. Name and Contact info is added to reduce fake submissions.

*Required

1. Full Name *

2. Phone number/ Email *

3. What is your gender? *

Mark only one oval.

- ◯ Male
- ◯ Female
- ◯ Other: _____

4. What is your age? *

5. What is/are your favourite genre of film? *

Tick all that apply.

- ☐ Roman
- ☐ Film-
- ☐ Ho
- ☐ Thri
- ☐ Com
- ☐ Musi
- ☐ Adventu
- ☐ Fant
- ☐ Cr
- ☐ Dr
- ☐ Documentar
- ☐ Science-
- ☐ Act
- ☐ West
- ☐ Animati

Ot ☐ _____

6. What makes you choose a film? *

Tick all that apply.

- [ ] Trailer
- [ ] Advertisements (Both online and offline)
- [ ] Reviews
- [ ] Actors/directors involved
- [ ] Social media
- [ ] Online Talk
- [ ] Word of Mouth

7. What is important to you in a film? *

Tick all that apply.

- [ ] Entertainmen
- [ ] Cinematograph
- [ ] Storyli
- [ ] Characte
- [ ] Act
- [ ] Special Effects/ Animation quality
- [ ] Cinemati
- [ ] Soundtra
- [ ] Ot _____

8. What don't you want to see in a movie? *

Tick all that apply.

- [ ] Promotion of a political agenda
- [ ] Illogical
- [ ] Reboots, Remakes and Sequels
- [ ] Too many superheroes
- [ ] Bad animation/CGI
- [ ] Unrealistic lifestyles
- [ ] Excessive violence
- [ ] Dancing and singing (not related to the plot)
- [ ] Ot _____

9. Do you prefer a male or female lead? * *Mark only one oval.*

- Male
- Female
- Don't mind

10. How important is the cast in a film? *

Mark only one oval.

- Important
- Unimportant
- Don't mind

11. How important is the director to you in a film? *

Mark only one oval.

- Important
- Unimportant
- Don't mind

12. How important is the soundtrack to you in a film? *

Mark only one oval.

- Important
- Unimportant
- Don't mind

13. How important is the production company/studio to you in a film? *

Mark only one oval.

- Important
- Unimportant
- Don't mind

14. How often do you watch a film? *

Mark only one oval.

- Daily
- Weekly
- Monthly
- Once a year

15. Where do you prefer to watch films? *

Mark only one oval.

- Cinema
- DVD
- TV
- Streaming platforms
- Pirated
- copy

Other: _____

16. What is your overall favourite film and why? (P.S. You may choose more than one film. For "why" part you may use only keywords or may completely ignore that part.) *

17. What is/are your favourite English film? *

18. What is/are your favourite film in your national and native language? *

19. What is/are your favourite animated film and anime film? *

20. Which is/are your favourite animation style? *

Tick all that apply.

- [ ] Traditional animation (e.g., The Lion King 1994)
- [ ] Anime cel animation (e.g., Your Name)
- [ ] Stop motion animation (e.g., Shaun The Sheep Movie)
- [ ] Motion capture (e.g., The Adventures of Tintin)
- [ ] CGI (Both Traditional (e.g., Disney or Pixar Movies) and Advanced (e.g., Live Action Movies))
- [ ] Performance-capture animation (Motion capture fused with CGI) (e.g., Avatar)

Other: [ ] _____

21. Do you like the following? *

Mark only one oval per row.

	Y	N	Ma
Se			
Documentarie			
TV			
Short			
An			
Cartoo			

Series, Documentary, TV shows, Short films, Anime etc.

These parameters are also helpful for recommendation system. But this is a optional part, so you may skip & submit!

22. What is your favourite genre of Series? *Mark only one oval.*

- Drama
- Crime
- Comedy
- Science fiction
- Fantasy
- Historical
- Mystery
- Thriller
- Superhero

23. What is your favourite genre of Anime?

Mark only one oval.

- [ ] Action/Adventure
- [ ] Comedy/Slice-of-Life
- [ ] Drama/Tragedy
- [ ] Psychological
- [ ] Historical
- [ ] Mecha/Military
- [ ] Supernatural/Magic
- [ ] Romance

24. What is/are your favourite Series?

_____

25. What is/are your favourite Documentary?

_____

26. What is/are your favourite TV show?

_____

27. What is/are your favourite Anime?

_____

Thank you for your time. Your response will surely help us to create a dataset.

3.4 Feedback Analysis

3.4.1 Sample size

What sample size is required for a survey? There is no definitive answer to this question: large samples with rigorous selection are more powerful as they will yield more accurate results, but data collection and analysis will be proportionately more time consuming and expensive. Essentially, the target sample size for a survey depends on three main factors: the resources available, the aim of the study, and the statistical quality needed for the survey. For 'qualitative' surveys using focus groups or interviews, the sample size needed will be smaller than if quantitative data is collected by questionnaire.

Larger samples give a better estimate of the population but it can be difficult to obtain an adequate number of responses. It is rare that everyone asked to participate in the survey will reply. To ensure a sufficient number of responses, include an estimated non-response rate in the sample size calculations. Response rates are a potential source of bias. The results from a survey with a large non-response rate could be misleading and only representative of those who replied.

In this survey, the minimum sample size will be around 100 responses.

3.4.2 Data collection

I have approached data collection in a rigorous and ethical manner. The following information is clearly recorded for convenience:

The respondents were contacted via social media for the survey.

The form was circulated to about 120 people and about 80 of those agreed to participate.

The survey questionnaire was checked by our supervisor Professor Swapan Kumar Mondal sir and administered by myself.

The response rate was better than my expectation. Some records are given below;

	Timestamp	Full Name	Phone number/ Email	What is your gender?	What is your age?	What is/are your favourite genre of film?	What makes you choose a film?	What is important to you in a film?	What don't you want to see in a movie?	Do you prefer a male or female lead?	...	Do you like the following? [TV shows]
0	20/04/2022 17:01:44	Arkajyoti Bhattacharyay	9330436509	Male	21	Romance, Horror, Thriller, Comedy, Adventure, ...	Trailer, Reviews, Actors/directors involved	Entertainment, Cinematography, Action, Soundtrack	Promotion of a political agenda, Bad animation...	Don't mind	...	Yes
1	20/04/2022 18:15:36	Harshita Agarwal	8653540167	Female	27	Comedy, Fantasy	Trailer, Reviews, Actors/directors involved	Entertainment	Bad animation/CGI	Don't mind	...	Yes
2	20/04/2022 18:20:29	Sayan Das	das.sayan128@gmail.com	Male	26	Thriller, Comedy, Musical, Adventure, Fantasy,...	Trailer, Reviews, Actors/directors involved	Entertainment, Cinematography, Storyline, Soun...	Illogical plots, Reboots, Remakes and Sequels,...	Don't mind	...	Maybe
3	20/04/2022 21:43:37	Poushali Bhattacharyay	7044335397	Female	25	Horror, Thriller, Comedy, Adventure, Crime, Sc...	Trailer, Actors/directors involved	Entertainment, Cinematography, Storyline, Char...	Promotion of a political agenda, Illogical plo...	Don't mind	...	Maybe
4	21/04/2022 17:21:30	Pushpita dhar	8697740249	Female	30	Romance, Horror, Thriller, Comedy	Reviews, Social media	Entertainment, Cinematography	Promotion of a political agenda	Female	...	Maybe
5	25/04/2022 19:57:04	Swarupa Chakraborty	6291413640	Female	30	Comedy, Musical, Fantasy, Drama	Advertisements (Both online and offline), Acto...	Entertainment, Cinematography, Storyline	Promotion of a political agenda, Excessive vio...	Don't mind	...	Yes
6	25/04/2022 22:13:25	Sourav Naskar	naskar.sourav818@gmail.com	Male	21	Adventure, Sports	Reviews	Storyline, Characters	Promotion of a political agenda, Illogical plo...	Male	...	Yes
7	25/04/2022 22:45:16	Debjit Pal	debjit.pal.cse2019@kgec.ac.in	Male	21	Romance, Thriller, Comedy, Musical, Adventure,...	Trailer, Advertisements (Both online and offli...	Entertainment, Cinematography, Storyline, Char...	Promotion of a political agenda, Illogical plo...	Male		
8	26/04/2022 01:31:46	Sarthak Chatterjee	sarthakchat2000@gmail.com	Male	21	Film-Noir, Horror, Thriller, Comedy, Adventure...	Trailer, Advertisements (Both online and offli...	Cinematography, Storyline, Characters, Special...	Illogical plots, Bad animation/CGI, Unrealisti...	Don't mind		
9	26/04/2022 09:29:43	Arka Dutta	arkadutta.cg@gmail.com	Male	21	Thriller, Comedy, Crime, Science-Fiction, Anim...	Reviews, Actors/directors involved	Cinematography, Storyline, Characters	Illogical plots, Reboots, Remakes and Sequels,...	Don't mind		
10	26/04/2022 09:38:51	Sayan Kundu	sayankundupiku.2001@gmail.com	Male	21+	Comedy, Western	Social media, Word of Mouth	Entertainment, Characters	Promotion of a political agenda	Don't mind		
11	26/04/2022 10:09:39	Sahasree	sahasreechatterjee@gmail.com	Female	21	Romance, Horror, Thriller, Comedy, Adventure, ...	Trailer, Social media, Online Talk	Entertainment, Cinematography, Storyline, Char...	Promotion of a political agenda, Reboots, Rema...	Female		
12	26/04/2022 10:51:17	Soumyadeep Chowdhury	soumyadeep.rkm@gmail.com	Male	21	Romance, Thriller, Crime	Trailer, Reviews, Social media	Entertainment, Storyline	Promotion of a political agenda, Illogical plo...	Male		
13	26/04/2022 12:03:28	Sayan Dey	sayandey15102001@gmail.com	Male	20	Romance, Thriller, Comedy, Adventure, Crime, S...	Trailer, Reviews, Actors/directors involved	Entertainment, Cinematography, Storyline	Promotion of a political agenda, Illogical plo...	Don't mind		
14	26/04/2022 12:24:38	Shraya Mutsuddi	9062005495	Female	20	Romance, Horror, Thriller, Comedy, Documentary...	Trailer, Reviews, Online Talk	Entertainment, Storyline	Promotion of a political agenda, Illogical plo...	Male		

Fig. response sheet of some of the respondents in python notebook

	Timestamp	Full Name	Phone number/ Email	What is your gender?	What is your age?	What is/are your favourite genre of film?	What makes you choose a film?	What is important to you in a film?	What don't you want to see in a movie?	Do you prefer a male or female lead?	...	Do you like the following? [TV shows]
17	26/04/2022 16:02:43	Rohan Ghosh	9123342264	Male	21	Thriller, Fantasy, Crime, Action	Trailer, Reviews	Entertainment, Cinematography, Storyline, Char...	Illogical plots, Unrealistic lifestyles, Danci...	Don't mind	...	No
18	27/04/2022 14:12:11	Arnab Pal	9064211483	Male	20	Thriller	Trailer	Storyline	Promotion of a political agenda, Unrealistic l...	Don't mind	...	No
19	27/04/2022 15:17:39	Ghanashyam Nandi	gmnandi00@gmail.com	Male	21	Romance, Horror, Thriller, Comedy, Adventure, ...	Trailer, Advertisements (Both online and offli...	Entertainment, Storyline, Characters, Action, ...	Reboots, Remakes and Sequels, Excessive violence	Female	...	Maybe
20	27/04/2022 18:49:24	Protyusha Chaudhuri	protyushachaudhuri274@gmail.com	Female	21	Film-Noir, Thriller, Adventure, Fantasy, Crime...	Actors/directors involved, Word of Mouth	Cinematography, Storyline	Illogical plots, Reboots, Remakes and Sequels,...	Don't mind	...	Yes
21	27/04/2022 18:55:48	Sampurna Biswas	bsampurna2001@gmail.com	Female	21	Musical, Adventure, Crime, Documentary, Animation	Trailer, Reviews, Actors/directors involved, S...	Entertainment, Storyline, Characters, Special ...	Promotion of a political agenda, Illogical plo...	Don't mind	...	Maybe
22	28/04/2022 08:30:43	Pratyay Roy	pratyayroy9051@gmail.com	Male	21	Horror, Thriller, Adventure, Crime, Documentar...	Trailer, Actors/directors involved, Social media	Storyline, Characters, Action, Special Effects...	Promotion of a political agenda, Reboots, Rema...	Male	...	Maybe
23	28/04/2022 19:59:04	Arnab Das	arnabadd@gmail.com	Male	21	Romance, Thriller, Comedy, Adventure, Crime, D...	Reviews, Actors/directors involved, Word of Mouth	Cinematography, Storyline, Cinematics	Illogical plots, Unrealistic lifestyles, Exces...	Don't mind	...	Yes
24	29/04/2022 22:35:15	Wasim Aktar	7679761423	Male	22	Romance, Film-Noir, Horror, Thriller, Comedy, ...	Trailer, Advertisements (Both online and offli...	Entertainment, Cinematography, Storyline, Char...	Promotion of a political agenda, Illogical plo...	Don't mind	...	Yes
25	30/04/2022 15:11:53	Animesh Bhakat	bhakat.animesh1024@gmail.com	Male	21	Science-Fiction	Trailer, Reviews, Actors/directors involved	Cinematography, Storyline	Excessive violence, Dancing and singing (not r...	Don't mind	...	No
26	30/04/2022 19:38:46	Aarati Shah	aaratishah10@gmail.com	Female	21	Romance, Thriller, Comedy, Musical, Adventure,...	Trailer, Reviews, Actors/directors involved	Entertainment, Cinematography, Storyline, Acti...	Promotion of a political agenda, Too many supe...	Don't mind	...	Yes
27	30/04/2022 20:35:13	Abhishek Baral	8250546861	Male	21	Thriller, Comedy, Crime, Documentary, Science-...	Advertisements (Both online and offline), Revi...	Cinematography, Storyline, Characters, Cinematics	Too many superheroes, Bad animation/CGI	Male	...	Yes
28	30/04/2022 22:17:22	Shuvranshu Khanra	khanra.shuvranshu@gmail.com	Male	21	Thriller, Comedy, Adventure, Science-Fiction, ...	Trailer	Entertainment, Storyline, Action	Reboots, Remakes and Sequels, Bad animation/CGI	Male	...	No

Fig. response sheet of some of the respondents in python notebook

3.4.3 Data analysis

The purpose of all analyses is to summarize data so that it is easily understood and provides the answers to our original questions. In order to do this, one must carefully examine their data. Researchers spend substantial time on the data analysis phase of a survey. When analysis is rushed, often important aspects of the data are missed and sometimes the wrong analyses are conducted, leading to both inaccurate results and

misleading conclusions. The method of data analysis will depend on the design of the survey and should have been carefully considered in the planning stages of the survey.

Data collected from the mandatory questions of the response sheet is analysed below,

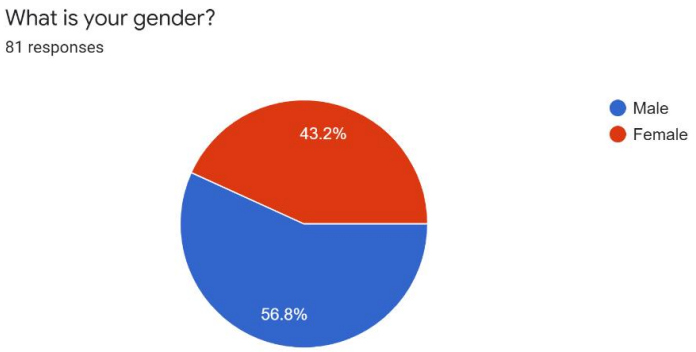


Fig. Gender

Analysis: The gender equality is almost maintained in the survey. So, we can expect a generalised distribution which is not biased on gender.

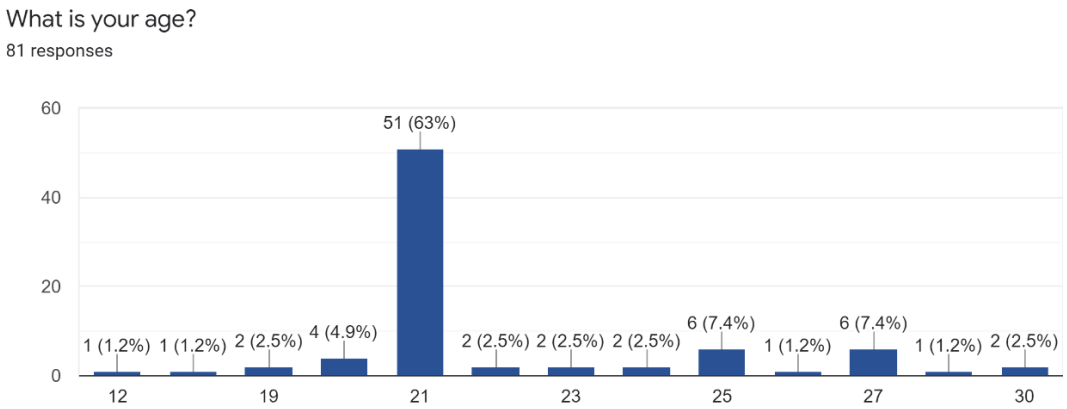


Fig. Age

Analysis: The distribution is focused on the 21 years old respondents, so the generalised result will be best suited for this particular age segment.

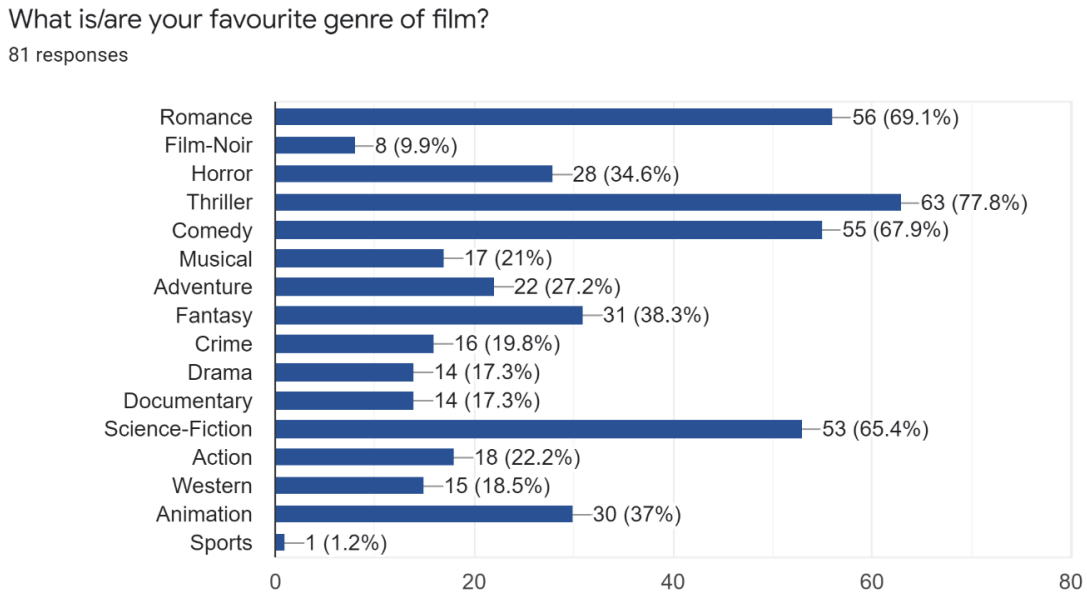


Fig. Genre

Analysis: The respondents' generalised common choice of genres are Thriller, Romance, Comedy and Science-Fiction.

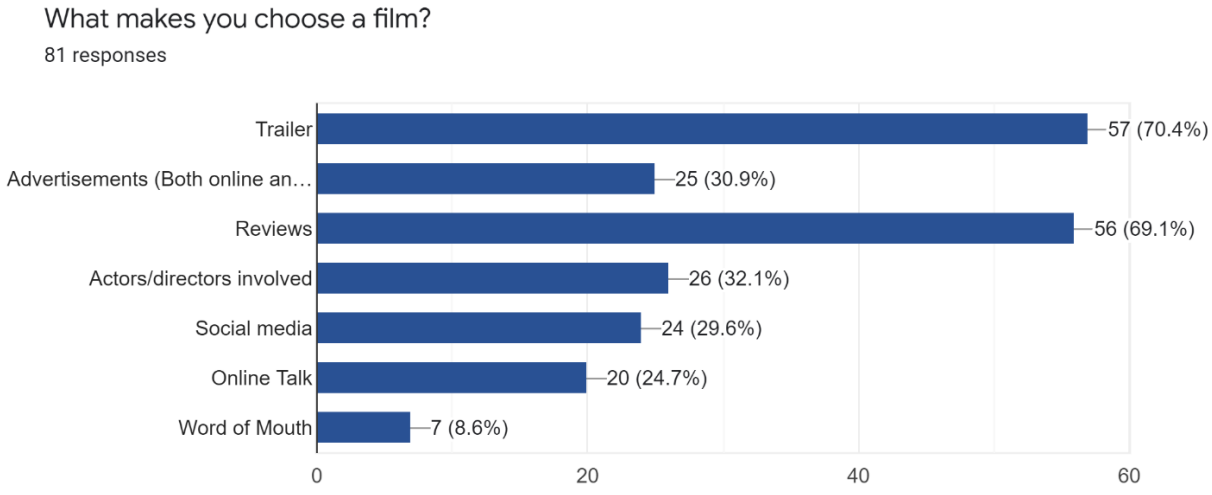


Fig. Source of Information

Analysis: The respondents' generalised source of film selection are Trailers and Reviews of the respective films.

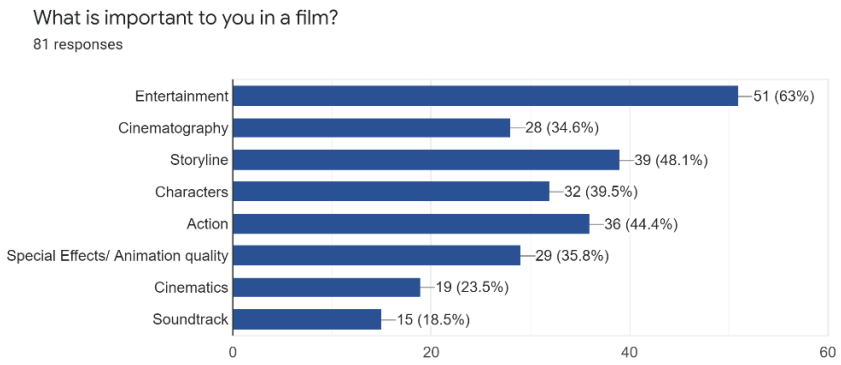


Fig. Important Features

Analysis: The respondents have mixed priorities but Entertainment is common factor.

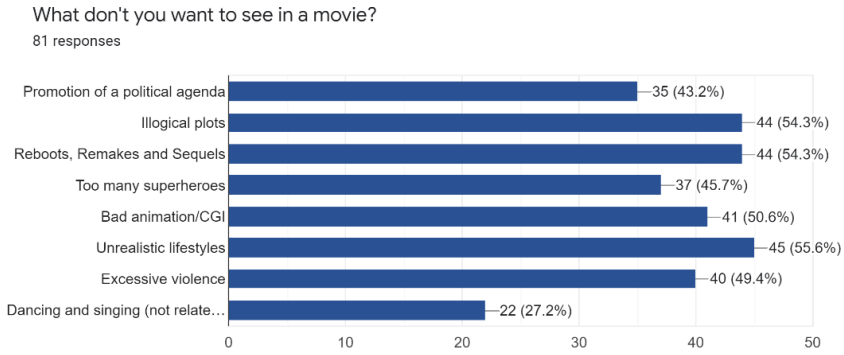


Fig. Dislikes

Analysis: The respondents disliked all the mentioned parameters equally.

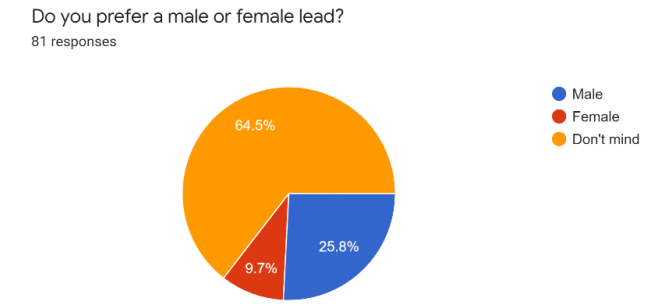


Fig. Lead Actor

Analysis: The respondents don’t mind the gender of the lead actor.

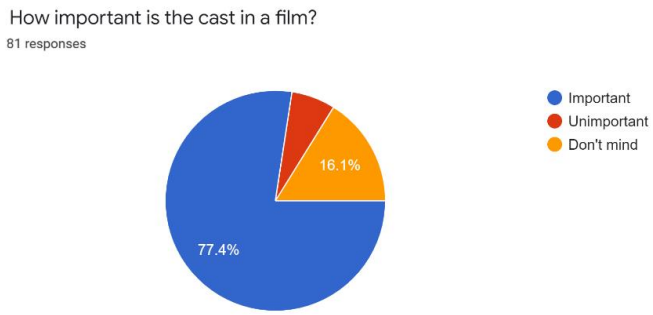


Fig. Cast

Analysis: The respondents think the cast is very important in any film.

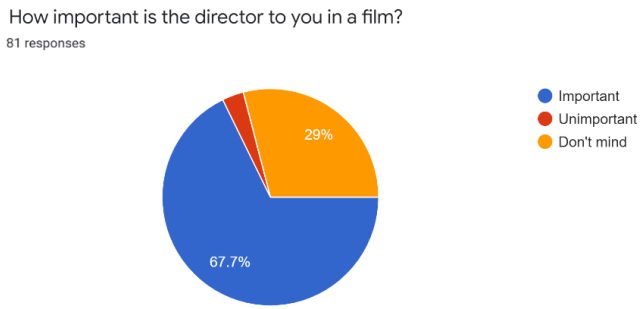


Fig. Director

Analysis: The respondents think the director is very important in any film.

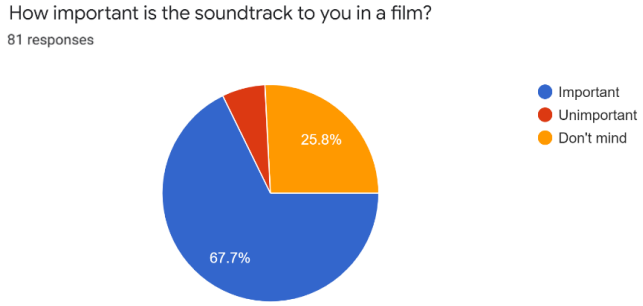


Fig. Soundtrack

Analysis: The respondents think the soundtrack making is very important in any film.

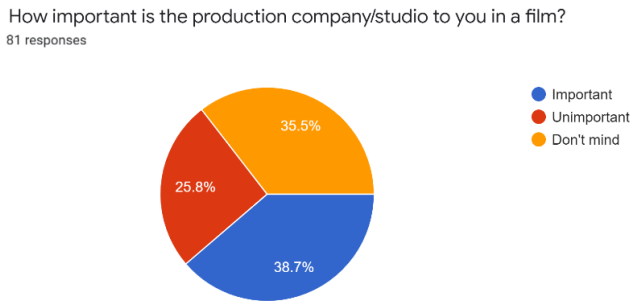


Fig. Studio

Analysis: The respondents seemed to have interest in the production companies/studios as big studios produce superior quality of production.

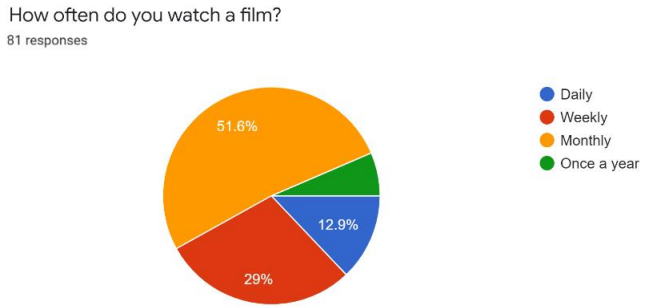


Fig. Frequency

Analysis: The respondents' lion share watch movies weekly or monthly.

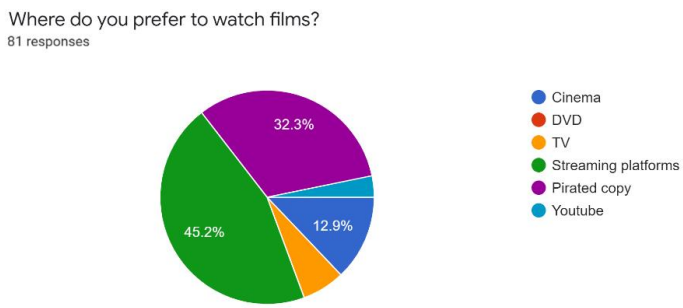


Fig. Platform

Analysis: The respondents' first choice is streaming platforms which provide a vast range of movies.

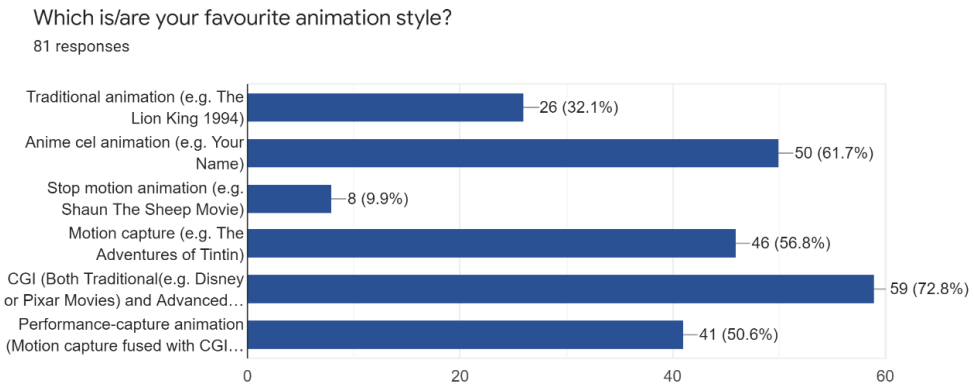


Fig. Animation Style

Analysis: The respondents prefer CGI as it is robust, flexible and used in any kind of digital production or movies. It can also be mixed with other animation technologies. Also, Japanese Anime animation is gaining popularity globally for its uniqueness and variation of styles.

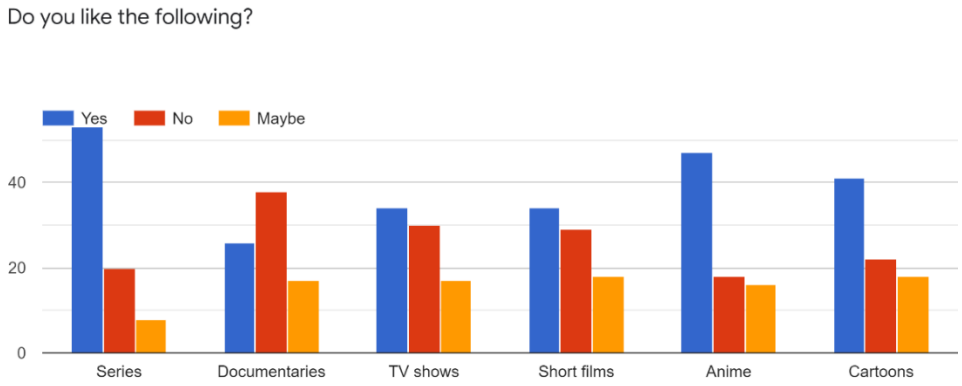
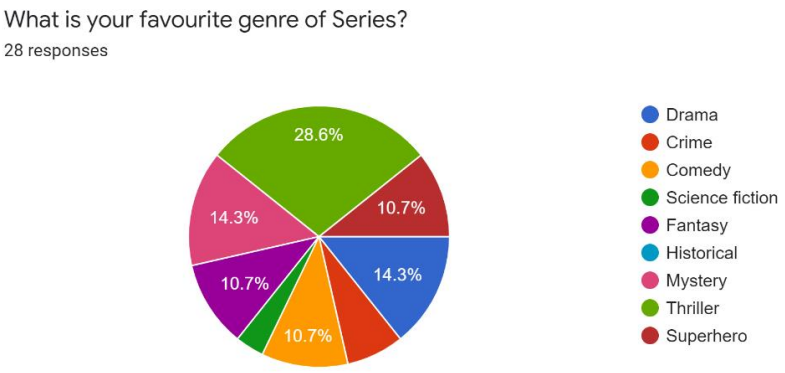


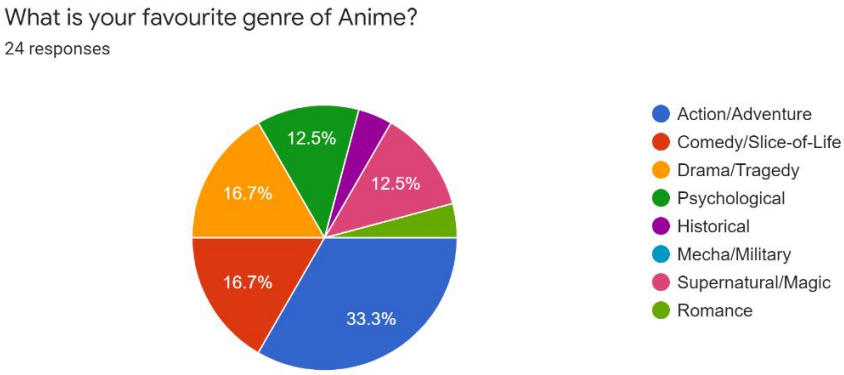
Fig. Other Digital Contents

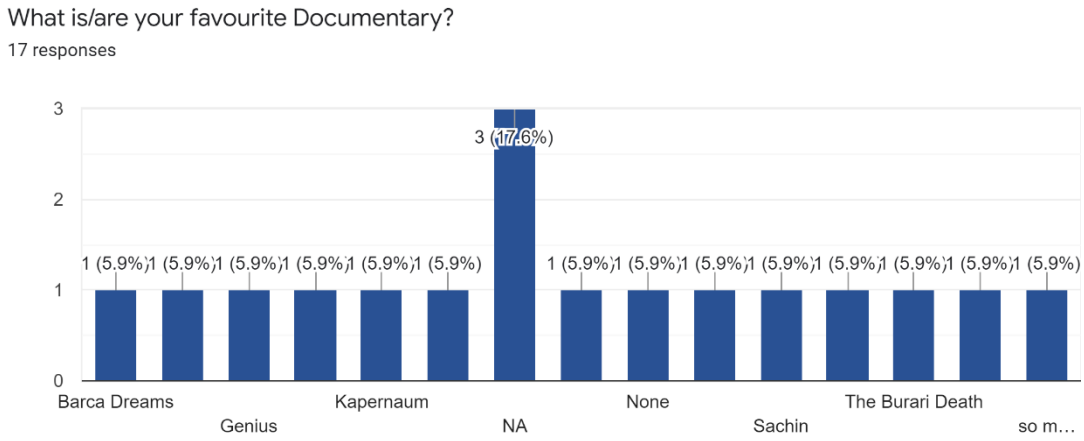
Analysis: The respondents like Series, Anime and Cartoons without any doubt. But TV shows and short films are not preferred choice. Documentaries are not liked by the majority.

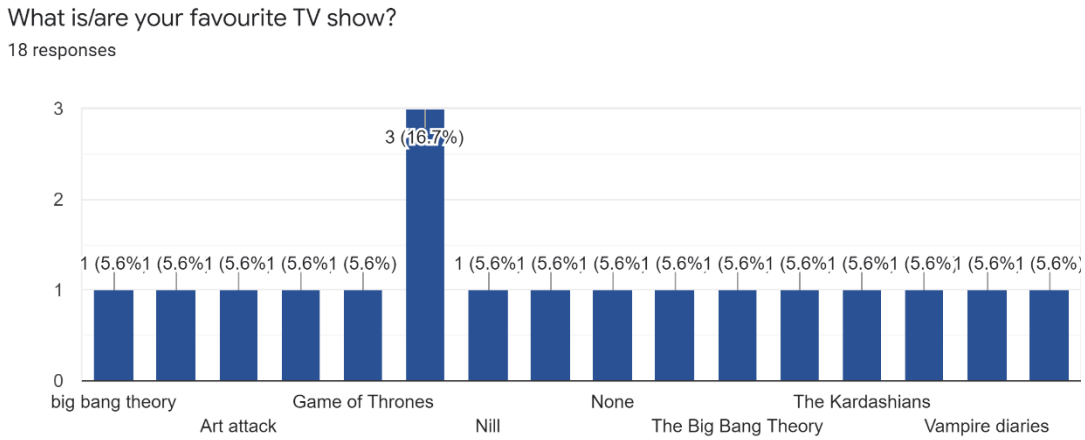
Data collected from the optional questions of the response sheet is analysed below,

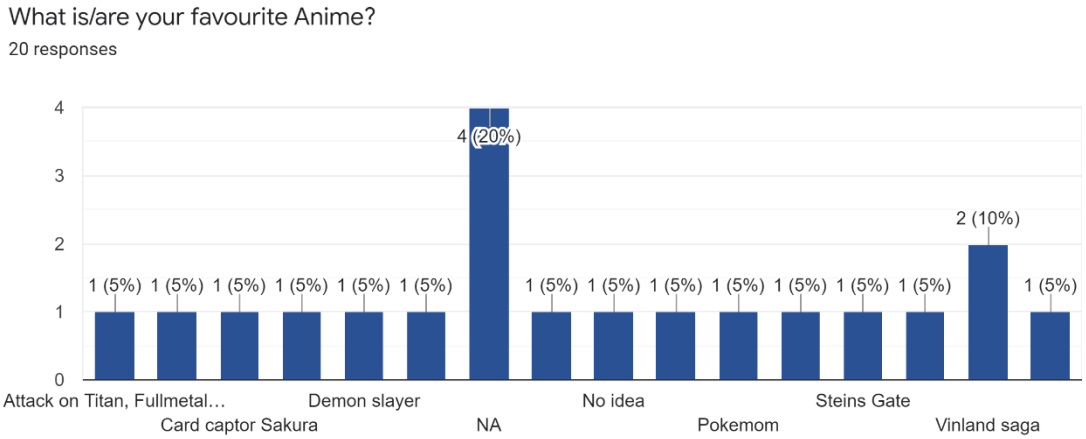
These responses are in very less frequency so it can't determine any general result but it is very helpful for personalised recommendation.











3.4.4 Reporting

The following reporting is clearly acceptable from the analysis of the data gathered from the response sheet:

The majority of the respondents are 21 years old so this analysis can generally be used for the likes and dislikes of that particular age group. This report may drastically change for other age groups or for same age group of another demographic region. It can be said that it is only applicable for this specific set of respondents only.

3.4.5 Ethics

Anyone involved in collecting data from patients has an ethical duty to respect each individual participant's autonomy. Any survey should be conducted in an ethical manner and one that accords with best research practice. Two important ethical issues to adhere to when conducting a survey are confidentiality and informed consent.

The respondent's right to confidentiality should always be respected and any legal requirements on data protection adhered to. In this survey the collective data is also used anonymously.

3.5 Conclusion

Survey research demands the same standards in research practice as any other research approach, and journal editors and the broader research community will judge a report of survey research with the same level of rigour as any other research report. This is not to say that survey research need be particularly difficult or complex; the point to emphasize is that researchers should be aware of the steps required in survey research, and should be systematic and thoughtful in the planning, execution, and reporting of the project. Above all, survey research should not be seen as an easy, 'quick and dirty' option; such work may adequately fulfil local needs (e.g., a quick survey of hospital staff satisfaction), but will not stand up to academic scrutiny and will not be regarded as having much value as a contribution to knowledge.

This survey has partially fulfilled its goal as it mainly represents only a very small portion of the total population. It may not be useful for the general recommendation of the overall population of greater scale.

CHAPTER 4

MAKING MODEL OF RECOMMENDATION SYSTEMS

4.1 Overview of Movie Recommendation Model

Recommender System is a system that seeks to predict or filter preferences according to the user's choices. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general.

4.2 Movie Recommendation Model Analysis

In this project, for the personalised recommendations I will be using two models, I will also use **TMDB 5000 Movie Dataset** for personalised recommendation of movies.
Credits: https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata

1. Simple Recommender

The Simple Recommender offers generalized recommendation to every user based on movie popularity and (sometimes) genre. The basic idea behind this recommender is that movies that are more popular and more critically acclaimed will have a higher probability of being liked by the average audience. This model does not give personalized recommendations based on the user.

The implementation of this model is extremely trivial. All we have to do is sort our movies based on ratings and popularity and display the top movies of our list. As an added step, we can pass in a genre argument to get the top movies of a particular genre.

Simple Demograpic Recommender

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.metrics.pairwise import linear_kernel
from sklearn.metrics.pairwise import cosine_similarity
from ast import literal_eval
```

In [2]:

```
path = "c:/users/arkaj/desktop" #Folder path
credits_df = pd.read_csv(path + "/tmdb_5000_credits.csv")
movies_df = pd.read_csv(path + "/tmdb_5000_movies.csv")
```

In [3]:

```
movies_df.head()
```

Out[3]:

	budget	genres	homepage	id	keywords	origi
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	

In [4]:

```
credits_df.head()
```

Out[4]:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...

In [5]:

```
credits_df.columns = ['id','tittle','cast','crew']
movies_df = movies_df.merge(credits_df, on="id")
```

In [6]:

```
movies_df.head()
```

Out[6]:

	budget	genres	homepage	id	keywords	origi
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	

In [7]:

```
C = movies_df["vote_average"].mean()
m = movies_df["vote_count"].quantile(0.9)

print("C: ", C)
print("m: ", m)

new_movies_df = movies_df.copy().loc[movies_df["vote_count"] >= m]
print(new_movies_df.shape)
```

```
C:  6.092171559442011
m:  1838.4000000000015
(481, 23)
```

In [8]:

```
def weighted_rating(x, C=C, m=m):
    v = x["vote_count"]
    R = x["vote_average"]

    return (v/(v + m) * R) + (m/(v + m) * C)
```

In [9]:

```
new_movies_df["score"] = new_movies_df.apply(weighted_rating, axis=1)
new_movies_df = new_movies_df.sort_values('score', ascending=False)

new_movies_df[["title", "vote_count", "vote_average", "score"]].head(10)
```

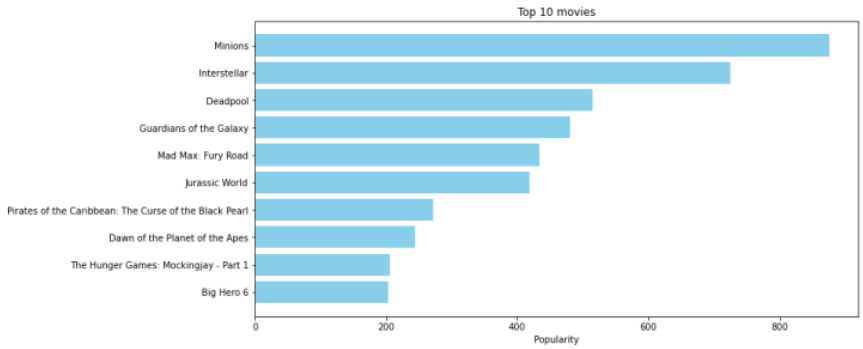
Out[9]:

	title	vote_count	vote_average	score
1881	The Shawshank Redemption	8205	8.5	8.059258
662	Fight Club	9413	8.3	7.939256
65	The Dark Knight	12002	8.2	7.920020
3232	Pulp Fiction	8428	8.3	7.904645

In [10]:

```
# Plot top 10 movies
def plot():
    popularity = movies_df.sort_values("popularity", ascending=False)
    plt.figure(figsize=(12, 6))
    plt.barh(popularity["title"].head(10), popularity["popularity"].head(10), align="ce
nter", color="skyblue")
    plt.gca().invert_yaxis()
    plt.title("Top 10 movies")
    plt.xlabel("Popularity")
    plt.show()


plot()
```



2. Content Based Recommender

The recommender in the previous section suffers some severe limitations. For one, it gives the same recommendation to everyone, regardless of the user's personal taste. If a person who loves romantic movies (and hates action) were to look at our Top 15 Chart, s/he wouldn't probably like most of the movies. If s/he were to go one step further and look at our charts by genre, s/he wouldn't still be getting the best recommendations.

For instance, consider a person who loves Dilwale Dulhania Le Jayenge, My Name is Khan and Kabhi Khushi Kabhi Gham. One inference we can obtain is that the person loves the actor Shahrukh Khan and the director Karan Johar. Even if s/he were to access the romance chart, s/he wouldn't find these as the top recommendations.

To personalise our recommendations more, I am going to build a model that computes similarity between movies based on certain metrics and suggests movies that are most similar to a particular movie that a user liked. Since we will be using movie metadata (or content) to build this engine, this also known as Content Based Filtering.

Content based Recommender, based on only overview/title

In [13]:

```
# Compute similarity
cosine_sim = linear_kernel(tfidf_matrix, tfidf_matrix)
print(cosine_sim.shape)

indices = pd.Series(movies_df.index, index=movies_df["title"]).drop_duplicates()
print(indices.head())
```

```
(4803, 4803)
title
Avatar                                      0
Pirates of the Caribbean: At World's End    1
Spectre                                     2
The Dark Knight Rises                       3
John Carter                                 4
dtype: int64
```

In [11]:

```
print(movies_df["overview"].head(5))
```

```
0    In the 22nd century, a paraplegic Marine is di...
1    Captain Barbossa, long believed to be dead, ha...
2    A cryptic message from Bond’s past sends him o...
3    Following the death of District Attorney Harve...
4    John Carter is a war-weary, former military ca...
Name: overview, dtype: object
```

In [12]:

```
tfidf = TfidfVectorizer(stop_words="english")
movies_df["overview"] = movies_df["overview"].fillna("")

tfidf_matrix = tfidf.fit_transform(movies_df["overview"])
print(tfidf_matrix.shape)
```

```
(4803, 20978)
```

In [14]:

```
def get_recommendations(title, cosine_sim=cosine_sim):
    """
    in this function,
        we take the cosine score of given movie
        sort them based on cosine score (movie_id, cosine_score)
        take the next 10 values because the first entry is itself
        get those movie indices
        map those indices to titles
        return title list
    """
    idx = indices[title]
    sim_scores = list(enumerate(cosine_sim[idx]))
    sim_scores = sorted(sim_scores, key=lambda x: x[1], reverse=True)
    sim_scores = sim_scores[1:11]
    # (a, b) where a is id of movie, b is sim_score

    movies_indices = [ind[0] for ind in sim_scores]
    movies = movies_df["title"].iloc[movies_indices]
    return movies
```

In [15]:

```
print("################ Content Based Filtering - plot##############")
print()
print("Recommendations for The Dark Knight Rises")
print(get_recommendations("The Dark Knight Rises"))
print()
print("Recommendations for Avengers")
print(get_recommendations("The Avengers"))
```

```
################ Content Based Filtering - plot##############

Recommendations for The Dark Knight Rises
65                              The Dark Knight
299                              Batman Forever
428                              Batman Returns
1359                                    Batman
3854    Batman: The Dark Knight Returns, Part 2
119                               Batman Begins
2507                                  Slow Burn
9           Batman v Superman: Dawn of Justice
1181                                       JFK
210                              Batman & Robin
Name: title, dtype: object

Recommendations for Avengers
7               Avengers: Age of Ultron
3144                            Plastic
1715                            Timecop
4124                This Thing of Ours
3311             Thank You for Smoking
3033                     The Corruptor
588     Wall Street: Money Never Sleeps
2136         Team America: World Police
1468                       The Fountain
1286                        Snowpiercer
Name: title, dtype: object
```

4.3 Limitation of the model

1. TMDB 5000 Movie Dataset (https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata) is a small dataset and do not contain even all English movies.
2. This model do not have any regional movies or non-English movies so the recommendation will be incomplete.
3. The Content based Recommender is only based on overview/title, other metrics are not even included so the recommendations may cause some outliers.

This model can also be applied for recommendations in

- Films in national and native languages
- Animated films and Anime films
- Series
- Documentary
- TV show
- Anime
- Cartoons

But the size of this project will be drastically increased, so I have to skip these parts.

4.4 Personality profile Analysis

We watch movies in cinemas. We watch movies online. We watch movies on our televisions. Our consumption is increasing based on ease of access to such content. While it is clear that we consume much media content on a daily basis, what is less evident is the characteristics of people who have a preference toward a specific genre. Much of the research done on this topic has focused on gender and personality characteristics independently as they relate to media preferences. This is a framework that many contemporary psychologists advocate and that many researchers use to determine personality types based on movie preferences.

Comedy

Individuals who chose the comedy genre were more open (more creative and adventurous) and slightly less conscientious (less attention to detail and disorganized). And females who showed a preference towards this genre (when both sexes did) were more open than males.

Horror

Individuals who gravitated towards horror movies were less agreeable (less altruistic), less extroverted (more reserved), and more neurotic (more nervous and tense). According to the study, the lower agreeableness can be explained by the fact that people who dislike horror films are more agreeable and prefer a move that displays images of kindness and warmth (not brutality), that is in line with their personality traits.

With regard to lower levels of extroversion, this finding is perhaps a little puzzling as it has been suggested that extroverts tend to enjoy horror films. Finn provides a possible explanation by mentioning that extroverts avoid a lot of media consumption and gravitate toward social interaction.

Action

People who like action movies are more conscientious (hard working), less neurotic (less emotionally stable), and more open (creative and adventurous). And females who showed a preference towards this genre (when both sexes did) were more open than males, as with the comedy genre.

The levels of conscientiousness can be explained by the fact that such individuals often have a preference for familiarity. This is compatible with the predictable and familiar plot that is often associated with action movies.

Being less neurotic is supported by Conway and Rubin (1991) who state that people who are more neurotic will gravitate towards movies that are lighter (such as comedy) that free them from their neuroticism.

Romance

More conscientious (hard working) and more neurotic (more emotionally unstable) people seem to favour romance movies. And males who showed a preference towards this genre (when both sexes did) were more open than females. Romantic movies have predictable plots and similar characters; hence compatibility with conscientious viewers. They also provide happy endings, which provides comfort to the neurotic who may seek to break free from the tension and anxiety in his own life.

Fantasy

Liking fantasy films seems to reveal greater openness (creative and adventurous) and lower levels of extroversion (more reserved). Greater openness can be explained by the originality often associated with these movies. The plots are often also very creative and appeal to the intellectual. A plausible explanation for the second trait is that imagination and fantasy films go hand in hand. And imagination it seems is something introverts develop more than extroverts.

It seems then that movie preferences may reveal more about one than one perhaps initially thought. The researchers acknowledge that there are some limitations with regard to the research (as is the case with any research) such as the sample and the data source.

Thus, if we map these traits with the respondents' priority of genres then their basic personality can be easily determined, though there will be some anomalies in the result.

CHAPTER 5 RESULT AND DISCUSSION

5.1 Result

This survey has partially fulfilled its goal as it mainly represents only a very small portion of the total population. It may not be useful for the general recommendation of the overall population of greater scale.

The majority of the respondents are 21 years old so this analysis can generally be used for the likes and dislikes of that particular age group. This report may drastically change for other age groups or for same age group of another demographic region. It can be said that it is only applicable for this specific set of respondents only.

The respondents' generalised common choice of genres are Thriller, Romance, Comedy and Science-Fiction. Their source of film selection are Trailers and Reviews of the respective films. The respondents don't mind the gender of the lead actor but They think the cast, the director and the soundtrack is very important in any film. The respondents' lion share watch movies weekly or monthly in streaming platforms.

The respondents like Series, Anime and Cartoons without any doubt. But TV shows and short films are not preferred choice. Documentaries are not liked by the majority.

5.2 Discussion

What I wanted in my survey project has been fully implemented. I have to face many minor problems to make it happen, since we are new to this survey project, it took a lot of time to configure and find a solution.

CHAPTER 6 CONCLUSION AND FUTURE WORK

6.1 Conclusion

I have completed my proposed work on time and I have included a description of it here. A lot of work is being done on this project. And our idea is that next time it will cover more type of digital contents.

This project report is based on “Movie Recommendation and Personality Analysis Based on Movie Show Preferences”. Here the data is analysed for general recommendation, a model has been made to create generalised recommendations.

This is how we decided to finish our project. However, this is not the end because I am very keen to work on this project, so I am willing to do some more work on this project if I get any next opportunity in the future.

6.2 Future work

Following is what I want to do in the future with this project:

- The Content based Recommender will be based on every possible parameter with a huge external dataset.

- This model can be modified to recommend English and non-English films, Films in national and native languages, Animated films and Anime films, Series, Documentary, TV show, Anime, Cartoons etc. all in a single model.

References

1. Richard Williams (2010), "The Animator's Survival Kit", ISBN-13: 978-0571202287, Faber & Faber, 2nd Edition, pp.1-500.
2. S. Ascher & E. Pincus (2013), "The Filmmakers' Handbook", Penguin Putnam Inc., ISBN: 9780452297289, 9780452297289, pp. 1-400.
3. Toby Segaran (2016), "Programming Collective Intelligence: Building Smart Web 2.0 Applications", O'Reilly Media, pp.51-129
4. Tom M. Mitchell (2018), "Machine Learning", McGraw Hill Education, pp.131
5. Dawn Griffiths (2016), "Head First Statistics – A Brain-Friendly Guide", O'Reilly Media, pp.21, 34-36.
6. Martha Nochimson, Lester D. Friedman, Sarah Kozloff, David Desser, Stephen Prince (2013), "An Introduction to Film Genres", W. W. Norton & Company, pp.1-624