THE AGE OF OTT PLATFORM

"Education for Knowledge, Science and Culture"

-Shikhanmaharshi Dr. Bapuji Salunkhe

(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE, KOLHAPUR (Autonomous)

DEPARTMENT OF STATISTICS

A PROJECT REPORT
On

THE AGE OF OTT PLATFORM

Submitted by

Mr. Shinde Pranav Tanaji
Mr. Kanire Darshan Sharad
Mr. Jadhav Pankaj Parasharam
Mr. Mane Manoj Jagannath

in partial fulfillment for the award of

the degree of

BACHELOR OF SCIENCE

in

STATISTICS

2022-23

(स्वायत्त) कोल्हापूर

VIVEKANAND COLLEGE, KOLHAPUR(Autonomous)
DEAPRTMENT OF STATISTICS

Certificate

This is to Certify that,

Sr. No.	Name	Roll No.
1	Mr. Shinde Pranav Tanaji	8355
2	Mr. Kanire Darshan Sharad	8342
3	Mr. Jadhav Pankaj Parasharam	8337
4	Mr. Mane Manoj Jagannath	8345

Have satisfactorily completed the project work on **"The age of OTT Platform"** as a part of skill enhancement course for **B. Sc. III,** prescribed by the Department of Statistics, *Vivekanand College, Kolhapur (Autonomous)* in the academic year **2022-23**.

This project has been completed under our guidance and supervision. To the best of our knowledge and belief, the matter presented in this project report is original and has not been submitted elsewhere for any other purpose.

Project Guide

(Ms. Patil P. C.)

Examiner

Head

(Ms. Pawar V. V.)
HEAD
DEPARTMENT OF STATISTICS
VIVEKANAND COLLEGE, KOLHAPUR
(AUTONOMOUS)

<u>ACKNOWLEDGEMENT</u>

<u>INDEX</u>

INTRODUCTION

The days of the television as we know it are almost gone. These days you don't even need to have a TV set to watch your favourite shows – everything can be found on the Internet. Dozens of live streaming platforms across the globe provide different types of content: movies, TV series, news, sports events, etc.

Streaming services started as an add-on to DVD and digital download offerings with a trickle of second-run movies and TV shows. They were supplements to the programs you watched on their first (and second) runs on cable TV. But speedier internet connections and an abundance of video streaming devices have accelerated the decline of traditional cable. More and more viewers are cutting the cord entirely in favour of dedicated streaming alternatives.

Entertainment and tech giants are not blind to the threat, however, and the media landscape is rapidly changing. Consolidation and curation (that is, owning the most media properties and serving the best content) seems to be the overarching goals of the players involved.

Video Streaming is a technology that has completely changed the entertainment industry as well as consumption models among audience members. A lot has changed since that very first Real Player transmission in 1995. Since then, technology has been constantly improving, making content delivery and access easier no matter the platform trying to access it.

Video content used to be delivered to consumers through cable providers with a box connected to a television. As the world and technology, we use evolved, the internet became a big part of our lives and with it, mobile connectivity. This resulted in content providers such as Hulu and even more popular, Netflix, launching and having a different type of television model. With their business model, users could stream TV shows through the internet on any device connected to the internet. This includes smartphones, tablets, laptops, and connected TVs.

Although different payment models exist (subscription or those that include advertising), the new system goes over the top of the traditional closed TV infrastructure. This is where the term OTT or over the top comes into play. Netflix,

Inc. is one of the best examples regarding commercial applications for Video Streaming. With millions of subscribing to the service all over the world, the company has found a way to capitalize its services using its title stock and outsourced infrastructure.

We have seen some repercussions appear regarding Video Streaming. Film experts fear audiences may be turning "platform agnostic" consuming content regardless of the size of the screen or the image quality. Nonetheless, audiences have proven they are willing to return to cinemas if the movie is worth the price of admission. However, content abundance has made audiences "socially autistic". Always connected to a device and somehow always isolated from others. Apparently, audiences are willing to sacrifices social experiences "offline" in the sake of personalized content.

OBJECTIVES

1. To identify whether people are aware of streaming services or not based on socio-demographic factors.
2. To know reasons or factors due to which people are switching from Cable/DTH to Online Video Streaming.
3. To predict customer's satisfaction level who are using Cable/DTH services.
4. To analyze customer's preference is selecting the best streaming platform.
5. To compare and classify the viewers based on time spent on watching/viewing.

METHODOLOGY

Keeping our objective in mind and the techniques to be used, we have designed our questionnaire accordingly. In our survey, we have collected our data through online applications such as WhatsApp, Facebook and mail as well as from offline such as housing societies, colleges, vicinities etc. We have used convenient sampling in our study. We have conducted a pilot study on 50 samples and after a slight modification in our questionnaire; we conducted the actual survey and have collected 106 sample data.

TECHNIQUES USED

- Graphical Representation
- Pareto Analysis
- Chi-Square Test
- Decision Tree

GRAPHICAL REPRESENTATION

The main purpose of graphical representation is to readily give some idea about the entire data and draw instant conclusions.

➢ **Graph of User Percentage for Free Streaming Services:**

USER PERCENTAGE FOR FREE STREAMING SERVICES

Service	Percentage
YouTube	87.50%
Hotstar	41.30%
Zee5	23.10%
Voot	13.50%
MX Player	36.50%
Vidmate	20.20%
None	5.80%
Other	8.70%
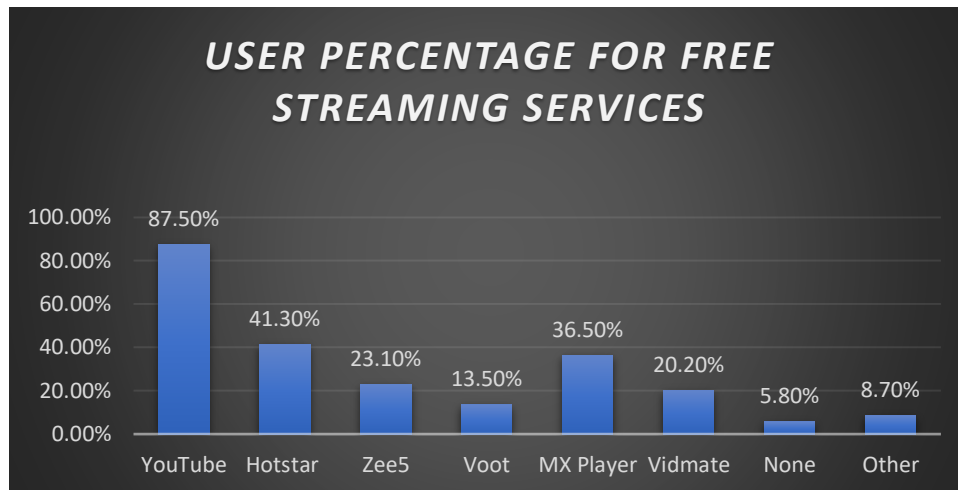
Interpretation:

Here, the maximum user prefer YouTube as free online streaming platform.

➢ **Graph of User Percentage for Paid Streaming Services:**

USER PERCENTAGE FOR PAID STREAMING SERVICES

Service	Percentage
Netflix	24.30%
Amazon Prime	14.60%
YouTube Premium	17.50%
Hotstar Premium	16.50%
Zee5 Premium	7.80%
None	57.30%
Other	5.80%
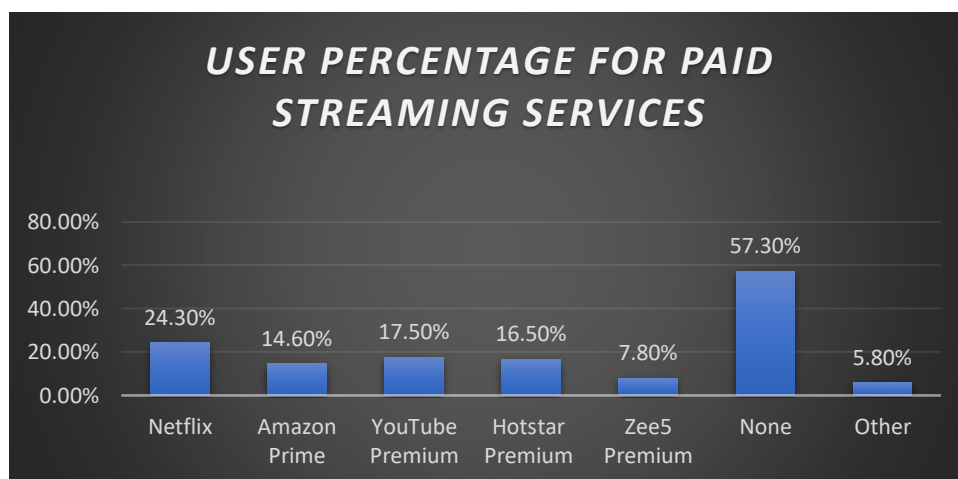
Interpretation:

Here, 57.3% users don't use any of the paid services while 24.3% of users use Netflix as one of the paid services.

➤ **Graph of Types of Programs on Cable/DTH vs Online Streaming:**

Types of Programs on Cable/DTH vs Online Streaming

(Percentages by category)

Category	Percentage Online Streaming	Percentage Cable/DTH
Daily Soaps	30.20%	67.80%
Reality Shows	33.00%	36.30%
Movies	64.20%	45.70%
Web Series	50.00%	0.00%
Sport	44.30%	47.90%
News	28.30%	37.50%
Kids	22.60%	27.40%
Documentary	19.80%	14.60%
Travel & Holiday	17.90%	17.40%
Others	12.30%	14.80%

■ Percentage Online Streaming   ■ Percentage Cable/DTH
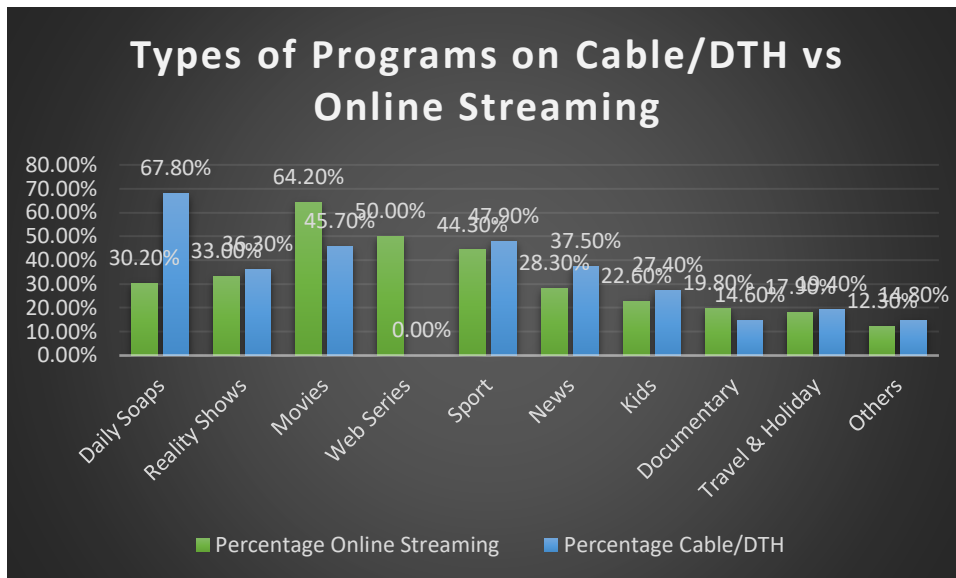
Interpretation:

Here, People mostly viewed Daily Soaps and Sports on Cable/DTH while Movies and Web Series on Online Streaming.

➤ **Graph of Types of Content on Online Streaming Services:**

TYPES OF CONTENT ON ONLINE STREAMING

Category	Percentage
Original	43.30%
Entertaining	81.70%
Educaational	58.70%
Emotinal	16.30%
Political Drama	30.80%
Religious	29.80%
Insightful	18.30%
Nudity	4.80%
Other	5.80%
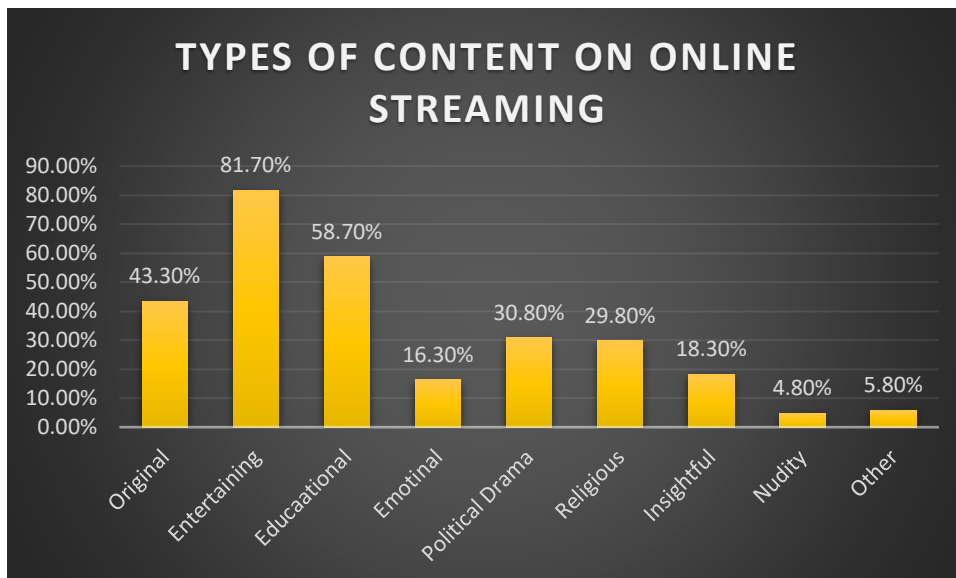
Interpretation:

Here, users like Entertaining, Educational and Original content than other types of content.

➤ **Graph of Language Preferred on Cable/DTH vs Online Streaming:**
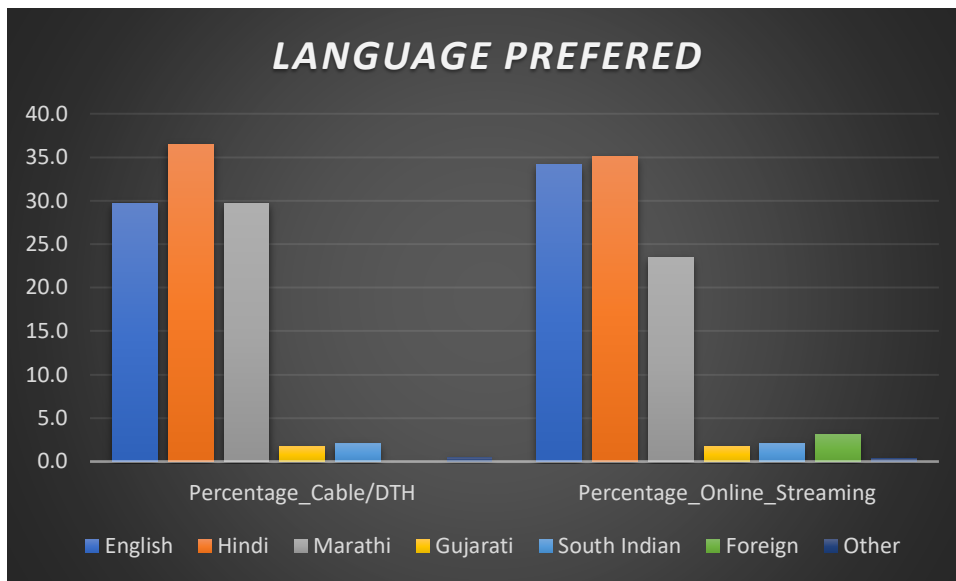
LANGUAGE PREFERED

Interpretation:

Here, Hindi is the most preferred language used by users while watching followed by English.

➤ **Graph of Devices Used for Online Streaming Services:**

DEVICES USED FOR ONLINE STREAMING
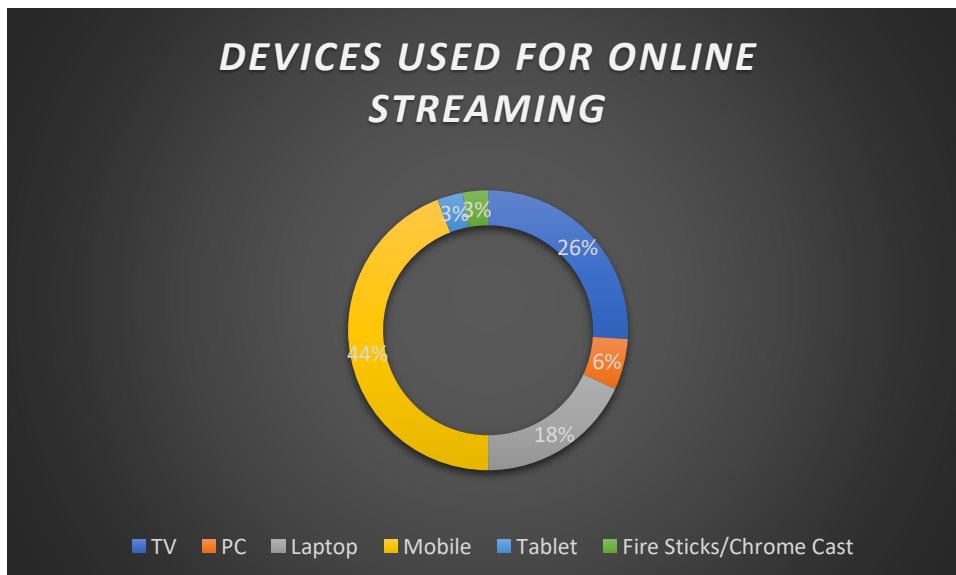
Interpretation:

Here, for above diagram we can conclude that 44% of users use mobile phones for online streaming.

➢ **Graph of Reasons for Using Streaming Services:**

REASONS FOR USING ONLINE STREAMING

Categories (x-axis): Addiction, Stress Buster, Peer Pressure, Lack Of Boredom, Affordable Prices, Well Curated Stuff, Personal Interest, Can Watch Entire…, Other

Interpretation:

Here, the maximum users use streaming services due to Personal Interest and Stress Buster.

➢ **Graph of User When they Prefer to Watch Streaming Services:**

WHEN DO YOU PREFER TO WATCH

Category	Percentage
During Leisure Time	74.50%
While Travelling	38.70%
While Work	8.50%
While Eating	21.70%
On Holiday	34.90%
On Weekends	39.60%
Late Night	32.10%
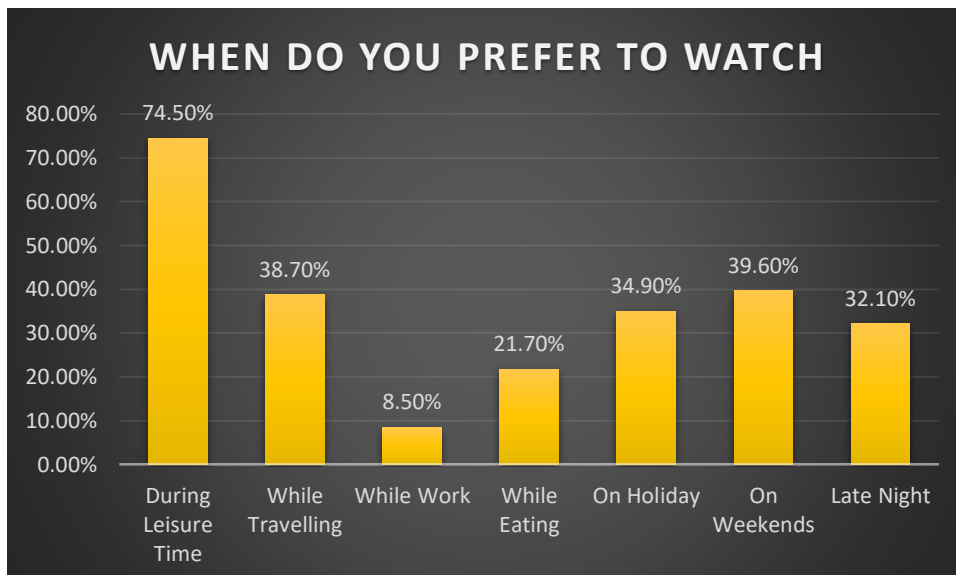
Interpretation:

Here, maximum 74.5% users watch streaming during leisure hours and 39.6% users also watch streaming on Weekends.

TEST OF INDEPENDENCE

1. Chi-Square:

The Chi-Square test of independence is used to determine if there is a significant relationship between two nominal (categorical) variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable. The data can be displayed in a contingency table where each row represents a category for one variable and each column represents a category for the other variable.

H_0: The two variables X and Y are independent.

H_1: The two variables X and Y relate to each other.

The test statistic is given by,
$$\chi^2 = \sum \left((O_i - E_i)^2 / E_i \right)$$

Where, O_i: Observed Frequency

E_i: Expected Frequency

Decision Criteria: Reject H_0 if p-value < 0.05

In our survey, we used Chi-square test to check whether people are aware of streaming services.

ANALYSIS:

CASE 1: To check the association between Gender and Awareness of Online streaming.

H_{01}: There is no association between Awareness of Online Streaming and Gender.

V/s

H_{11}: There is an association between Awareness of Online Streaming and Gender.

Awareness	Female	Male
Yes	9	52
No	5	40

Chi squared value = 4.6725, df = 1, p-value = 0.002998

Here, p-value=0.002998 < 0.05.

CONCLUSION: We reject H_{01} at 5% level of significance as p-value < 0.05 and conclude that there is no association between awareness of online streaming and gender.

2. Fisher's Exact Test:

The Chi-squared test is not suitable when the expected values in any of the cells of a contingency table are below 5. In such a case, Fisher's exact test is used.

ANALYSIS:

CASE 2: To check the association between Age and Awareness of Online Streaming

H_{02}: There is no association between Awareness of Online Streaming and Age.

V/s

H_{12}: There is an association between Awareness of Online Streaming and Age.

Awareness	Below 18	18-31	32-59
Yes	1	9	1
No	7	70	19

Chi squared value =9.348, df = 3, p-value = 0.6267
Here, p-value=0.6267 > 0.05.

CONCLUSION: We accept H_{02} at 5% level of significance as p-value > 0.05 and conclude that there is an association between awareness of online streaming and age.

CASE 3: To check the association between Education and Awareness

H_{03}: There is no association between Awareness of Online Streaming and Education.

V/s

H_{13}: There is an association between Awareness of Online Streaming and Education.

Awareness	SSC	HSC	Graduation	Post-graduation	Other
No	0	1	8	2	0
Yes	7	10	65	12	2

Fisher's Exact Test for count data with simulated p-value (based on 2000 replicates)
p-value = 0.905
Here, p-value=0.905 > 0.05.

CONCLUSION: We accept H_{03} at 5% level of significance as p-value > 0.05 and conclude that there is an association between awareness of online streaming and education.

PARETO ANALYSIS

Pareto Analysis is a Statistical Technique in decision making that is used for the selection of a limited number of tasks that produce a significant overall effect. It uses the Pareto Principle. It is also known as 80/20 rule. The idea is that by doing 20% of the work you can generate 80% of the benefit of doing the whole job. This is also known as 'vital few' and the 'trivial many' effects.

The Pareto Principle has many applications in quality control. It is the basis for the Pareto Diagram, one of the key tools used in total quality control and Six Sigma. A Pareto chart is used to graphically summarize and display the relative importance of the differences between groups of data. Pareto chart organizes and displays information to show the relative importance of the differences between groups of data. Pareto chart organizes and displays information to show the relative importance of various problems or causes of problems. It is essentially a special form of vertical bar chart that puts items in order (highest to lowest) relative to some measurable effect of interest such as frequency, cost or time.

The chart is based on the Pareto Principle, which states that when several factors affect a situation, a few factors will account for most of the impact. The Pareto describes a phenomenon in which 80 percent of variation observed can be explained by a mere 20 percent of the causes of that variation. The Pareto curve makes it clear as to where effort must be concentrated to give maximum effect.

The Pareto Chart is a very simple but effective tool for prioritizing problem causes, which is why it is widely used for problem-solving in the manufacturing industry.

The Pareto Chart is a descending bar graph that shows the frequencies of occurrences or relative sizes of the various problems or causes of a particular problem.

The problem categories or causes are shown on the x-axis of the bar graph. Aside from its main bar graph, the Pareto chart may also include a line graph that indicates the cumulative percentage of occurrences at each bar of the bar graph.

This line graph, referred to as the 'cumulative percentage line', is used to determine which of the bars belong to the 'vital few' and which ones are relegated to the 'trivial many'.

Reasons For Not Using Online Streaming
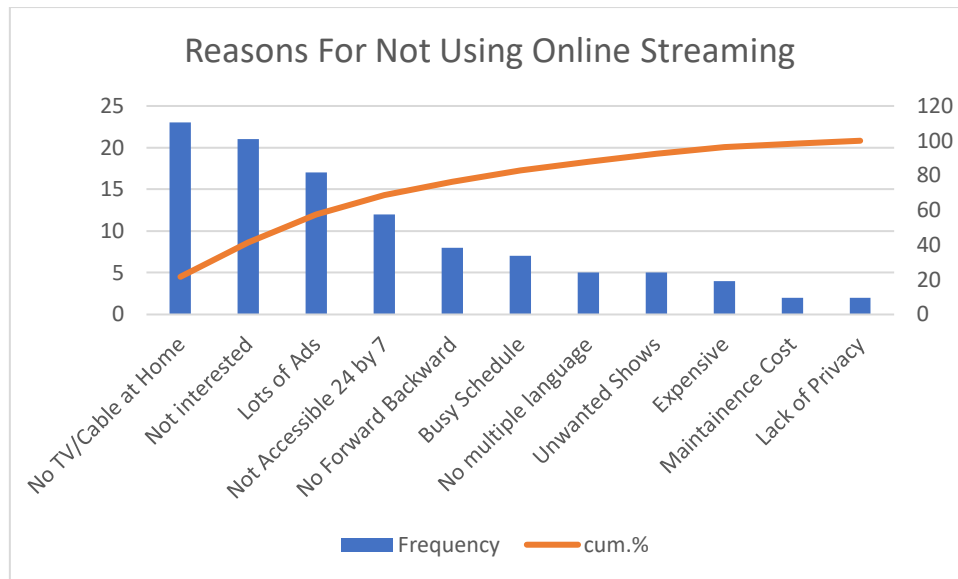
CONCLUSION:

From Pareto Analysis, we conclude the main reasons for people switching from Cable/DTH to Online Streaming are 'No TV/Cable at home', 'Not Interested', 'Lots of Ads', 'Not Accessible 24 x 7', 'No Froward-Backward Option', 'Busy Schedule', , 'No Multi-Language Option'.

WORD CLOUD

Word Cloud is a data visualization technique used for representing text data in which the size of each word indicates its frequency or importance. Significant textual data points can be highlighted using a word cloud. Word clouds are widely used for analysing data from social network websites.

Word clouds are also known as text clouds or tag clouds work simply: the more a specific word appears in a source of textual data (such as a speech, blog post, or database), the bigger and bolder it appears in the word cloud.

For Cable/DTH:

The above word cloud highlights *LESS COST, CONTENT, QUALITY SHOWS,* etc. the most as compared to other words.

This shows that these larger sized words have been expressed by most people in their responses. Some sort of improvement is needed by them in this area of Cable/DTH.

For Online Streaming Services:

The above word cloud highlights *LESS COST, DEBIT CARD, PAYMENT,* etc. the most as compared to other words.

This shows that these larger sized words have been expressed by most people in their responses. Some sort of improvement is needed by them in this area of Online Streaming Services.

<u>**CONCLUSION**</u>

From the study of our project, we conclude that,

> ➢ Most of peoples aware about the online streaming.
> ➢ The main reason for switching to DTH to online streaming is No cable/ TV at home.
> ➢ Most of the peoples prefer YouTube as free online streaming platform.
> ➢ Most of the peoples avoid the paid online streaming platform.

SUGGESTION

For Cable/DTH:

1. Reduce the cost of the monthly packs.
2. Provide better customer service.
3. Improve customer care communication system.
4. Make some quality shows for customers.
5. Provide additional perks.
6. Mostly during monsoon, the Cable/DTH faces fickle issues which they must handle with care.
7. Reduce advertisements as it disturbs the viewers.

For Online Streaming:

1. Provide affordable subscription plans.
2. Increase the mode of payments (e.g., UPI payments).
3. Allow payment by different types of cards (e.g., RuPay Debit Card).
4. Focus more on originality of shows.
5. Recommend new and featured content using personalized banners.
6. Increase the safety and security of personal data.
7. Speed up content discovery by creating stories.
8. Collect user's feedback when they delete account to deliver a better streaming experience to other customers.

<u>SCOPE</u>

The data sample used for our study was from Kolhapur. We can extend this survey to different regions and give more suggestions to improve Cable/DTH and Online Streaming. Major problems can be identified and worked upon by finding good solutions.

There are so many streaming services like music streaming, video streaming, live video streaming, etc. Our topic is limited to Video Streaming platforms only. So, this study can be explored further to other streaming services.

Large mobile penetration in developing countries and cheap data prices is one of the reasons for streaming being more popular these days. We can study the relationship between data consumption and use of streaming services.

Future research should examine the impact of how consumers watch entertainment and explore the different ways that Cable/DTH or streaming services are used.

BIBLIOGRAPHY

Online References

- https://in.pcmag.com/internet/95943/the-best-video-streaming-services-for-2020
- https://www.grabon.in/indulge/featured/most-popular-ott-platforms-in-india/
- https://www.consumerreports.org/tv-service/cable-vs-streaming-live-tv-services/
- https://www.monetizemore.com/blog/ott-vs-ctv/
- https://www.inventiva.co.in/uncategorized/surbhi/the-streaming-war-is-revolutionizing-entertainment-in-india/
- https://www.surveymonkey.com/curiosity/how-do-consumers-feel-about-streaming-services-vs-cable/

Sites Used

- YouTube
- Analytics Vidhya
- Statistics Solutions
- Statistics How To
- IDRE Stats – Statistical Consulting Web Resources
- Medium

Book References

- Programmed Statistics by B. L. Agarwal
- Fundamental of Applied Statistics
- Fundamental of Mathematical Statistics by S. C. Gupta & V. K. Kapoor
- Google

Statistical Software Used

R-Software

MS-Excel

R-Code

```
> data=matrix(c(0,10,1,0,7,70,19,0),c(2,4),byrow=T)
> data
     [,1] [,2] [,3] [,4]
[1,]   0   10   1    0
[2,]   7   70  19    0
> fisher.test(data)

     Fisher's Exact Test for Count Data


data:  data
p-value = 0.6267
alternative hypothesis: two.sided


> data=matrix(c(0,1,8,2,0,7,10,65,12,2),c(2,5),byrow=T)
> fisher.test(data)


     Fisher's Exact Test for Count Data
data:  data
p-value = 0.905
alternative hypothesis: two.sided
```

 _____ THANK YOU _____