A STUDY OF CONSUMER BEHAVIOR ON THE RISE OF OTT PLATFORMS IN INDIA

A PROJECT REPORT

Submitted by

Ms. Asma Shaikh

in partial fulfillment for the award of the degree

of

M.Sc. Statistics Part 2

Academic year 2022 – 2023

Submitted to

DEPARTMENT OF STATISTICS

RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ART'S,

SCIENCE & COMMERCE (AUTONOMOUS),

GHATKOPAR (W)

RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ART'S, SCIENCE & COMMERCE (AUTONOMOUS), GHATKOPAR (W)

(Affiliated to University of Mumbai)

CERTIFICATE

This is to certify that the project entitled **A Study of Consumer Behavior on The Rise of OTT Platforms in India** is the bonafied work of **Ms. Asma Shaikh** bearing seat no**. 908** during the year 2022-2023 in partial fulfillment of the requirements for the award of Degree Master of Science in Statistics.

Signature of Internal guide
(Mr. Atharva Potdar)

Signature of Co-ordinator
(Mr. Jaishankar Singh)

Seal of the college

Signature of Examiner

ABSTRACT

With rapid technological improvements and increasing Internet penetration levels across India, the way people consume media is changing. The emergence of Over-the-Top media services has started to change the media space significantly. OTT Platforms give us the freedom to watch movies and TV series, at our own convenience. While a majority of the population of the county uses Television, we can see a shift towards OTT Platforms like Netflix, Amazon Prime, Disney + Hotstar, Sony Liv, and Zee5. Some of these services may charge a premium cost when compared with Television, they offer various benefits which make them an attractive option. Content streaming on online platforms has much fewer breaks and advertisements compared to Television. They have given Indians better access to International Content with different languages and Subtitles available. This Study aims to understand the behavior of Consumers towards OTT Platforms which includes various indicators like their age, average monthly spending of time and money before and after Covid-19, their preferred OTT Platform, and reasons behind preferring OTT over TV, etc.

To Conduct the Survey, the Sample population has been Chosen by Random Sampling. The primary data has been collected through questionnaires with responses from 202 respondents from Students, working professionals, Unemployed people, Retired people & Housewives from every- age group. Also, we have used secondary data taken from Kaggle which has around 519 respondents. We merge Primary and Secondary data then data has been tabulated & analyzed to understand the perception of individuals towards OTT Platforms.

ACKNOWLEDGEMENT

Success is never achieved single-handedly So, it is our duty to acknowledge all those who have provided a helping hand in making this project successful.

Firstly, I would like to express my deep sense of gratitude towards my project guide, **Mr. Atharva Potdar** for his valuable guidance, co-operation, and approval given in the successful completion of the project work.

I also wish to express my sincere gratitude to **Mr. Jaishankar Singh**, Coordinator of the Department of Statistics for providing us an opportunity to do my project work on **"A Study of Consumer Behavior on the Rise of OTT Platforms in India".**

TABLE OF CONTENTS

INTRODUCTION

In the year 1959, television was introduced. The evolution of entertainment has seen drastic changes over the years from conventional dish or satellite modes, theaters towards on demand mode, over the top platforms. In India, BigFix was the first OTT Platform which was launched by Reliance Entertainment in 2008. The market keeps evolving with the entry of new players and in the year 2016, big revolution took place into the entertainment consumption space with the entry of new media services that are over the top platform with Netflix, amazon prime videos, Disney plus hotstar, alt Balaji, zee 5, Voot and many OTT platforms for regional cinemas. OTT Platforms has large repository of movies and TV Series from across the globe which can be accessed through different devices like Phones, Laptops etc. These Platforms are even producing their original movies/web series & shows, giving their own users a completely different experience. It is basically a media streaming service which is available for the users through the Internet. Increase in the consumption of internet services and smart phones and reduction in its cost are one of the prime reasons for the growth of OTT platforms. These Platforms are becoming more & more customer oriented & have been consistent innovations. As per ministry of information and broadcasting there are around 48 paid broadcasters, 850 plus registered TV channels. In the Indian market most of the users like to have advertisements led free to view models e.g.- YouTube than subscription-based model. The reasons for its growth are - India is a growing mobile market in the world, the launch of 4G services in India, it is tapping mass audiences from various genres from movies, shows, web series to all forms of sports. It has turned out to be a new parallel space to meet the growing consumption needs of audiences as well as for the creators to have such actors who have talents but can't draw box office numbers due to less star power for the Indian mass audiences.

REVIEW OF LITERATURE

According to (purdy,2018) The OTT market is a complex space dominated by three powerful players-Netflix, Amazon Prime, Hulu. It is challenging for small players to find their place in fierce competition where content rules but customers can be dragged if they focus on providing curated content, limiting account sharing, getting the price right and reducing churn through improved customer experience.

Article (singh,2020) OTT Platforms has experienced a surge in consumption and subscriber impacted by the COVID-19. The recent survey by InMobi states that 46% viewers are watching more content online. “As these channels have been unable to shoot due to the lockdown, they are running old content. In such a scenario people will gravitate towards OTT to watch fresh content.” says Paresh Joshi, media consultant & Principal, Provocateur Advisor, on the topic that television channels run out of content.

(Singh D.P.,2019) The advent of the Internet is changing the trend of Indian Cinema very rapidly. Reduced price of internet services lay the foundation for increase in OTT services like Netflix, Hotstar, Zee5, Amazon Prime and Alt Balaji. Many ventures are producing video content solely for these OTT Services. These new ventures are changing the way of Indian Television and Cinemas. The OTT Services analyze the video content watched by the users and keep providing them the content on similar genres.

This article (basu,2020) gives a brief idea about the relationship between TV and OTT in Indian market and the scope for it. When it comes to subscription television, India is the second largest market in the Asia Pacific region. Digital development has led to exponential growth in the number of TV channels. The television penetration in India is 61%.

This article (laghate, How OTT Players are geared up to woo Indian viewers,2017) talks about how people have experienced an increase in the use of OTT especially through handheld devices. People prefer watching movies on OTT platforms rather than waiting for movies and shows to be released on linear platforms. People in India have more OTT platforms than people in any other country with options like Hotstar, Voot, Amazon Prime, Sony Liv, Netflix etc. Although there were few companies that tried to establish themselves in the Indian market, it started whenHotstar launched in 2015 with a huge budget for advertising and marketing.

In a report titled „Digital Media: Rise of on-demand Content“ by Deloitte it is noted that the rise of internet- enabled digital devices capable of supporting digitized content has led to an increase in the use of digital content globally. In India, this trend is observed across diverse platforms such as audio, visual, news, music etc. It mentions that an Indian youth, on

an average spends 14% of their time and nearly 17% of their monthly expenditures on entertainment. An internet content consumer in India consumes an average of 6.2 hours of content on an everyday basis out of which 21% of the time is spent on audio-visual entertainment. A shift in consumer attitude with regard to a favoritism for OTT content and easy access to vast libraries at any time and place over content ownership is notable.

A study titled “UNDERSTANDING ADOPTION FACTORS OFOVER-THE-TOP VIDEO SERVICES AMONGMILLENNIAL CONSUMERS” by Dr. Sabyasachi Dasgupta and Dr. Priya Grover also reiterates that Indian audiences have swayed towards OTT content and are willing to spend for easy and unlimited access to content without a place and time limitation. It again notes the inverse impact of the pricing strategy of OTT on its popularity. Data consumption is another attribute PJAEE, 17(6) (2020) 4215 that makes it a tough choice for Indian viewers and so is habits and preferences for TV as a medium.

Another study by Sidneyeve Matrix on Netflix observes that viewers, especially youth, are becoming active curators of content than couch potatoes taking in „whatever producers feed them.“ Supporting this paradigm shift in consumers is the need to share, stay connected and discuss the content on social media forums. In the process of these social media transactions, the viewers are setting new standards of expectations from producers thereby becoming an active catalyst in the production process too.

A paper by Paramveer Singh finds that Netflix, Hotstar and Jio are most popular among Indian youth. The youth is skewed towards free trials available on these platforms, are nocturnal viewers and prefer web series format over films. The respondents affirm that over-the-top applications are changing media consumption patterns in India. The shift can be attributed to convenience of service, personalized experience and availability of global content etc. The study finds the future of OTT in India to be promising due to increasing smartphone penetration, economic convergence of media companies (takeovers/ mergers) at national or international level, and quality of reception of digital content. Also instrumental is the competitive internet data plans offered by telecom service providers in India.

Ritu Bhavsar in her research paper entitled “The Burgeoning Digital Media Consumption: A Challenge for Traditional Television and Advertising Industries – An Analysis” mentions that digital media has become an indispensable part of everyday lives and is a prominent medium used for gathering and disseminating information, socialization, entertainment and marketing. An ever-increasing consumption of content via digital media effects a change in the consumer preferences and attitudes and this transformation trend can be associated with better internet connectivity, advanced digital devices, competitive data prices in India and the accessible, on-the-go nature of internet media.

AIM & OBJECTIVE

AIM:

To perform Market Research and Analytics on Rise of OTT Platforms: A Study of Consumer Behavior.

OBJECTIVES:

1. To Study the concept of OTT Platforms.
2. To know which age group is more likely to use OTT Platforms.
3. To Know which OTT Platform is most preferred by Indians.
4. To compare the usage of OTT Platforms with TV.
5. To Know the average monthly spending of Indians on OTT Platforms.
6. To Know the average time spent by Indians on OTT Platforms.
7. To Know what Indians, prefer mostly to watch on OTT Platforms.
8. To Know if Covid-19 increased the demand for OTT Platforms.
9. To Know what type of OTT model Indians would like to avail.
10. To Know the satisfaction level of Indians from OTT Platforms.

RESEARCH METHODOLOGY AND DESIGN

For this project, we first decided on the parameters we want to study. Followed by conducting a survey for the same. This process can be described further:

Area Coverage:

This study is conducted in various regions of Mumbai.

Collection of Data:

We collected primary data by conducting a survey. This survey was circulated to various regions of Mumbai. The questionnaire asked for all the appropriate questions related to the research project. The data was collected between the time period of 14th February 2023 to 15th April 2023.

Apart from the primary data we took secondary data of around 527 respondents from Kaggle which was updated on May 22nd, 2021.

Analysis of Data:

The master table was prepared for entering the responses of each respondent and small cross tables were made from the master table for analysis. Master table consists of the entry of all the entries from each respondent for each question of the questionnaire. From this table we took the particular variables for statistical analysis.

Statistical Method:

1. Google Forms – Statistical and Qualitative Data Analysis Software is used to analyses the data collected.
2. Excel - It is a spreadsheet program which is used to save, analyse and visualize few part of our numerical data.
3. Python – Data Analysis Software is used for analyses of collected data.
4. Pewer BI – Data Visualization software is used for visualization.

Research Limitation:

Despite all the efforts and dedication towards this study, the only limitation to this research is that this online questionnaire is conducted and distributed online with channels such as email and text messenger (WhatsApp, Facebook Messenger, etc.). Therefore, it limits the correspondents with internet access or mobile data access.

QUESTIONNAIRE

QUESTIONNAIRE OF PRIMARY DATA :

- Timestamp
- Name
- Gender
- Age
- Occupation
- Where do you live?
- How frequently do you watch TV?
- How frequently do you watch OTT?
- How much time do you approximately spend on OTT platforms daily ?
- How much time do you approximately spend on TV platforms daily ?
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Amazon Prime]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Disney + Hotstar]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Voot]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [YouTube]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Netflix]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Zee5]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [SonyLIV]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [ALTBalaji]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [MX Player]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [JioCinema]
- Which OTT platform do you prefer other than the above specified?

- Which television provider do you prefer the most?
- Have you started watching OTT more due to lockdown?
- Are OTT platforms more convenient than TV?
- How much do you pay for OTT platforms on monthly basis ?
- How much do you pay for TV on monthly basis ?
- What do you prefer mostly to watch on OTT?
- What do you prefer mostly to watch on TV?
- What would you prefer to watch and does it affect your studies/work ?

QUESTIONNAIRE OF SECONDARY DATA :

- Age
- Occupation
- Where do you live?
- How much time do you approximately spend on OTT platforms daily ?
- How much time do you approximately spend on TV platforms daily ?
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Amazon Prime]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Disney + Hotstar]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Voot]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [YouTube]
- Rate the following OTT platforms according to your preference. (1= least preferred,5=Most preferred) [Netflix]
- Which OTT platform do you prefer other than the above specified?
- Which television provider do you prefer the most?
- Have you started watching OTT more due to lockdown?
- Are OTT platforms more convenient than TV?
- How much do you pay for OTT platforms on monthly basis ?
- How much do you pay for TV on monthly basis ?
- What would you prefer to watch and does it affect your studies/work ?

DATA ANALYSIS

Since we have a limitation of a few questions in secondary data which was only present in primary data, we have tried to do the objective-wise analysis where we have considered only primary data for a few objectives and only secondary data for a few objectives and for all other objectives we have used merged data.

- **To know which age group is more likely to use OTT Platforms**

(Based on 202 responses)

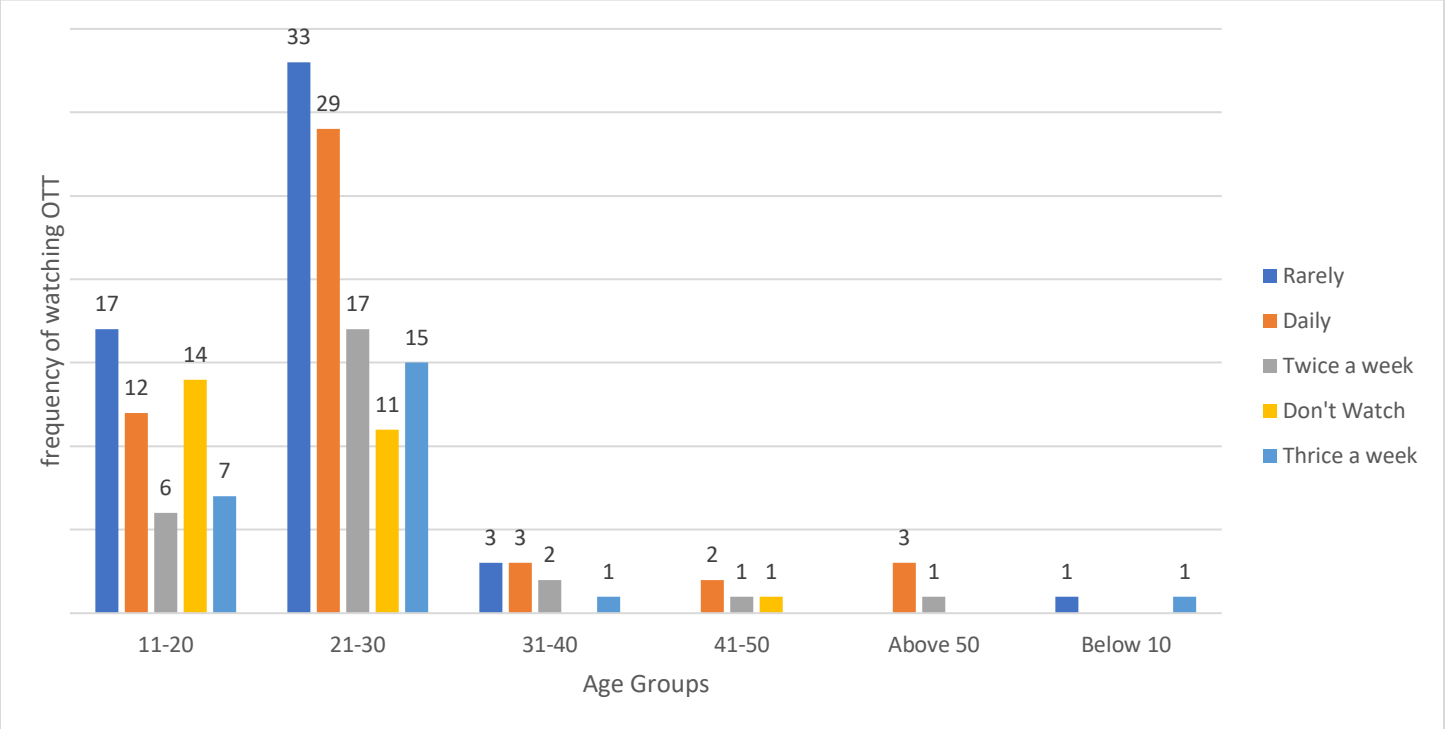

Fig. 1

Fig. 1 shows that most of the respondents in our research study who belong to the age categories of **21-30** and **11-20** have a very high frequency of watching OTT. So, it reveals those consumers belonging to these age categories are more in touch with OTT Platforms and hence those are key consumers of OTT.

- **To Know which OTT Platform is most preferred by Indians.**

Rating of different OTT platforms according to people's preferences.

(1= least preferred, 5= most preferred, 0= Don't Watch)

(Based on 729 responses)

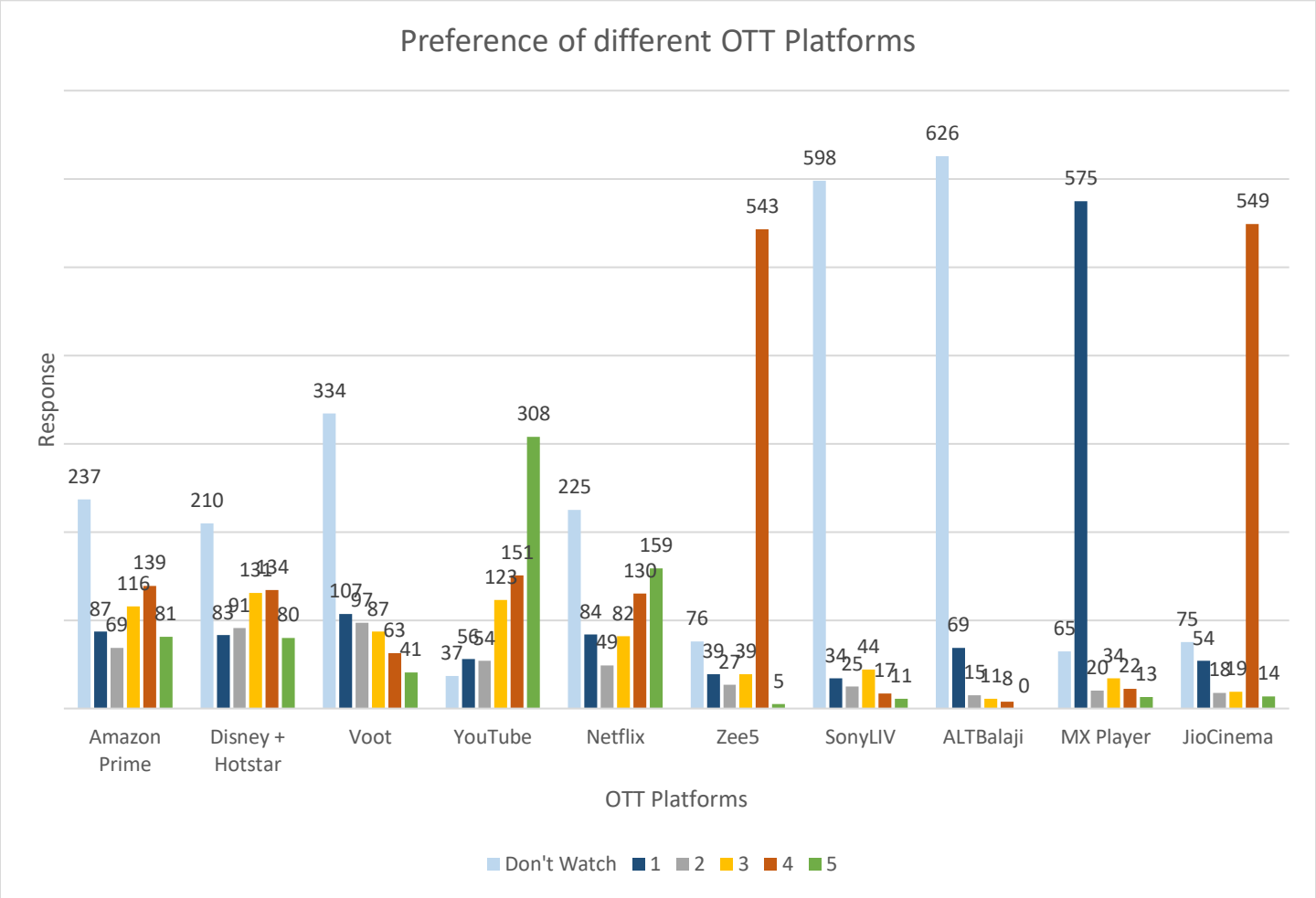

Fig. 2

Fig. 2 shows respondents' preferences among the top 10 OTT Platforms mentioned in our study.

Here, we can clearly observe that Youtube has a maximum of votes for rating 5(i.e., highly preferred) followed by Netflix, Amazon Prime, Disney + Hotstar, etc. Other than that, Jio Cinema and Zee5 have got a maximum of votes for rating 4(i.e. second highest preference).

Chi-Square Test:

Results						
	Netflix	Amazon	Hotstar	Voot	YT	***Row Totals***
Group 1	85 (84.80) [0.00]	88 (84.80) [0.12]	84 (84.80) [0.01]	108 (84.80) [6.35]	59 (84.80) [7.85]	424
Group 2	52 (74.60) [6.85]	72 (74.60) [0.09]	95 (74.60) [5.58]	99 (74.60) [7.98]	55 (74.60) [5.15]	373
Group 3	83 (109.20) [6.29]	116 (109.20) [0.42]	132 (109.20) [4.76]	88 (109.20) [4.12]	127 (109.20) [2.90]	546
Group 4	131 (125.80) [0.21]	141 (125.80) [1.84]	135 (125.80) [0.67]	64 (125.80) [30.36]	158 (125.80) [8.24]	629
Group 5	399 (355.60) [5.30]	333 (355.60) [1.44]	304 (355.60) [7.49]	391 (355.60) [3.52]	351 (355.60) [0.06]	1778
Column Totals	750	750	750	750	750	**3750 (Grand Total)**

The chi-square statistic is 117.589. The *p*-value is < 0.00001.
Therefore, The result is significant at $p < .05$.

- **Which OTT platform do you prefer other than the top 10 mentioned in our study?**

(Based on 750 responses)

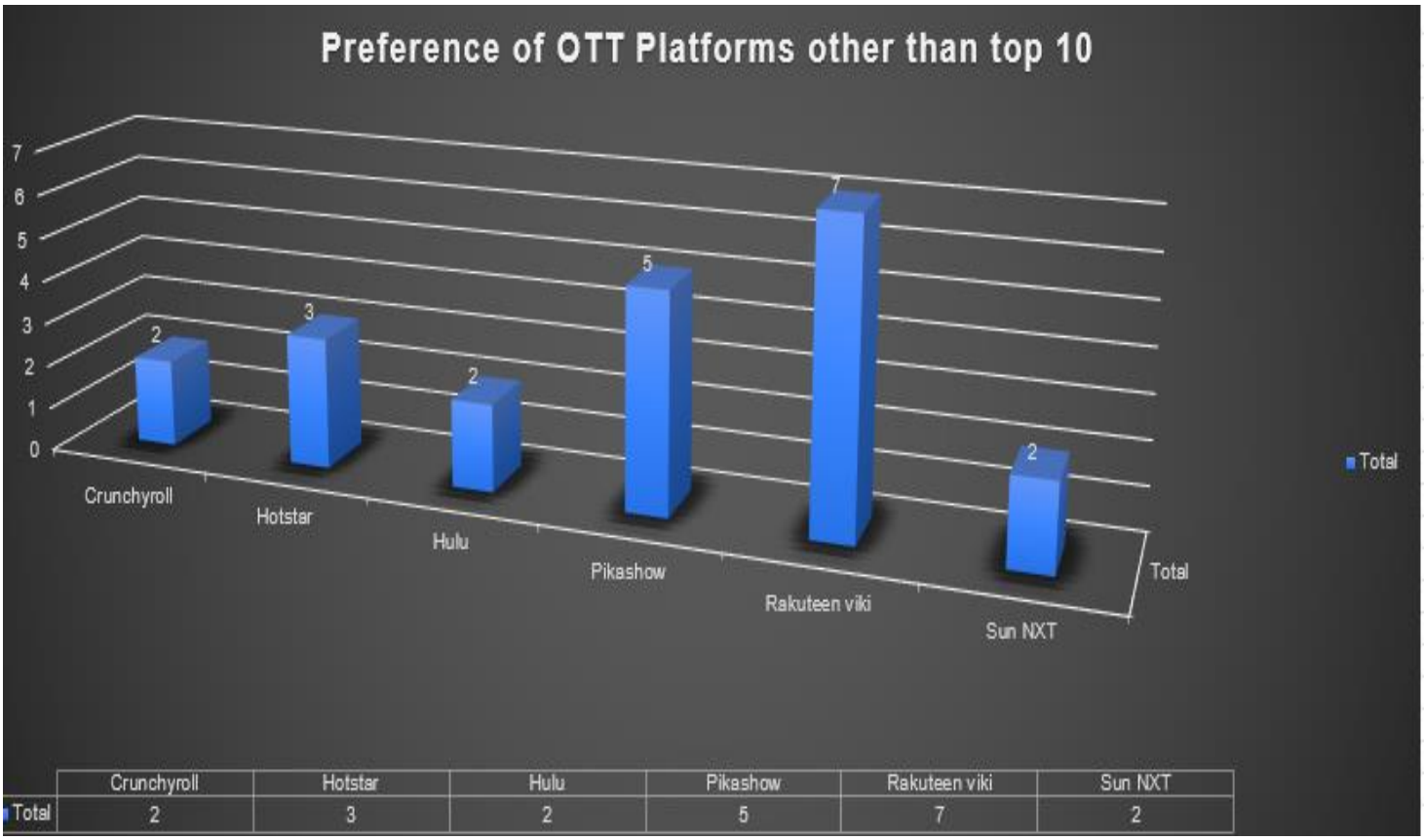

Fig 3

Fig. 3 shows OTT Platforms other than the top 10 OTT platforms mentioned in our study most of our respondents who are a part of our research study have voted for Rakuten Viki followed by Pikashow, Hotstar, Sun NXT, etc.

- **To Know the average monthly spending of Indians on OTT Platforms.**

(Based on 750 responses)

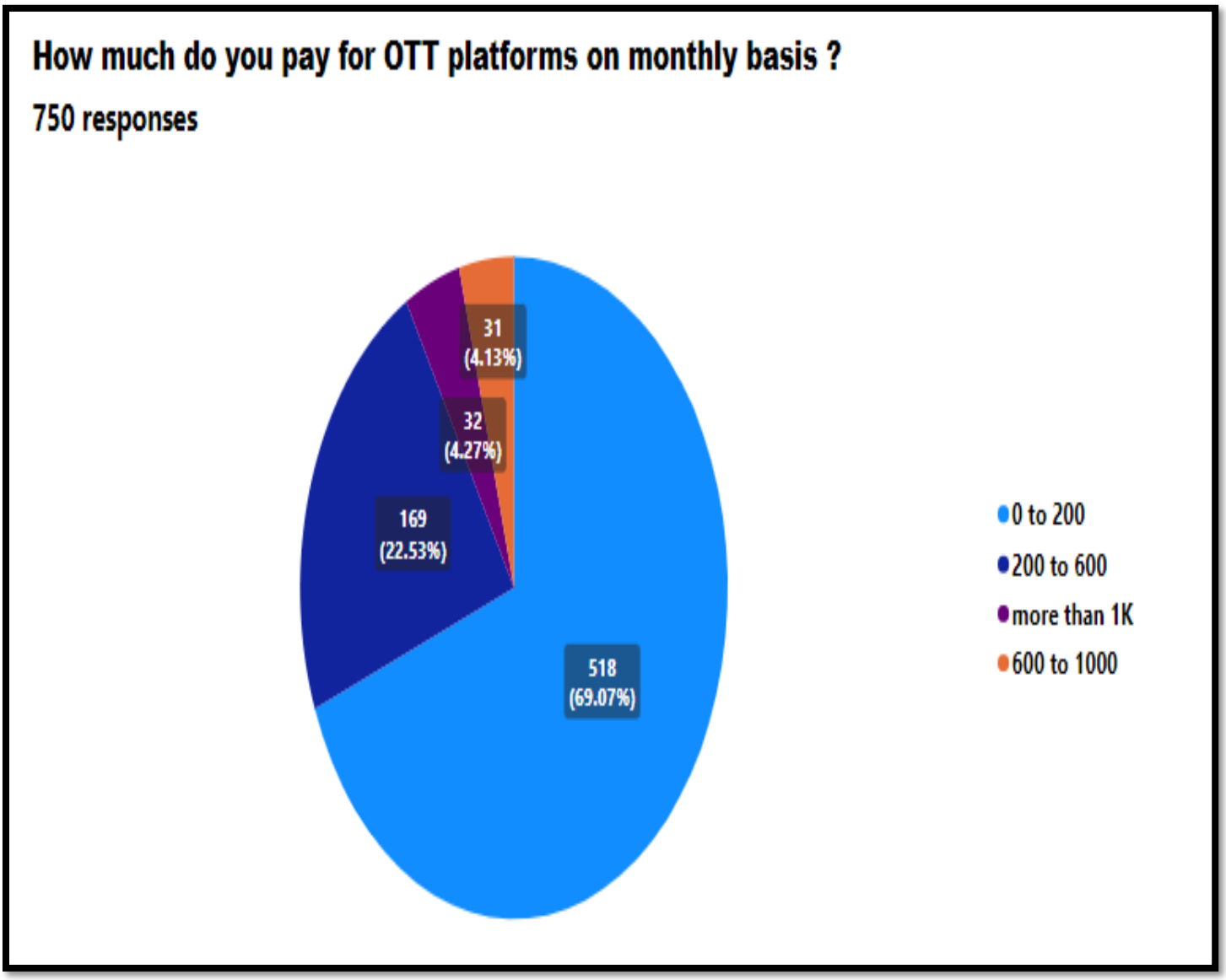

Fig 4

Fig. 4 shows the average monthly spending on subscriptions on OTT Platforms by consumers. So, It reveals that on average consumers are spending 0-200 rupees followed by 69.07% out of the total, and the remaining are spending more than that.

Impact of COVID-19 on OTT

(Based on 750 responses)

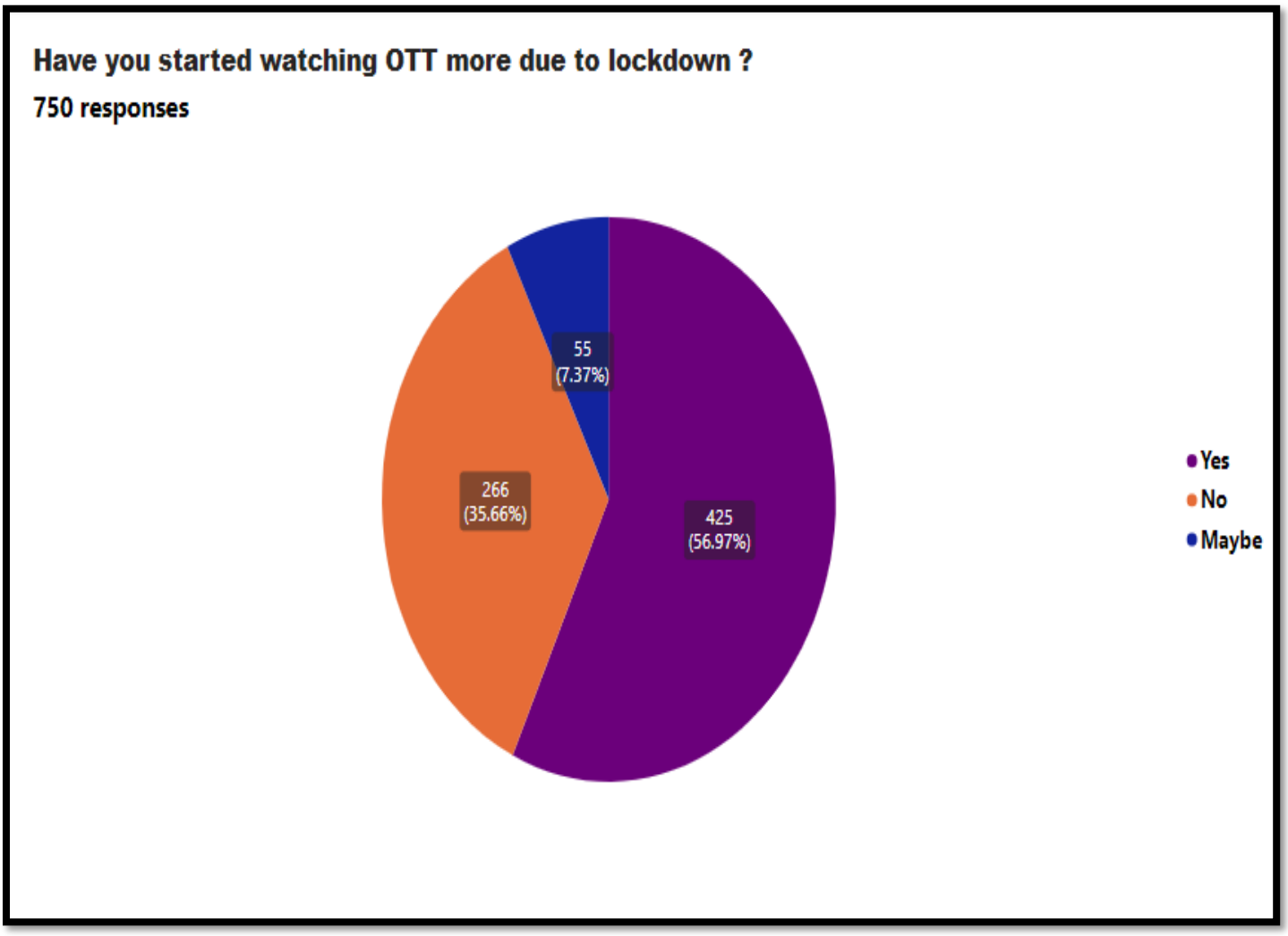

Fig 5

Fig 5 shows that **56.97%** of consumers feels that the lockdown has led to an increase in the consumption pattern of OTT.

Consumer Preferences on Content

(Based on 172 responses)

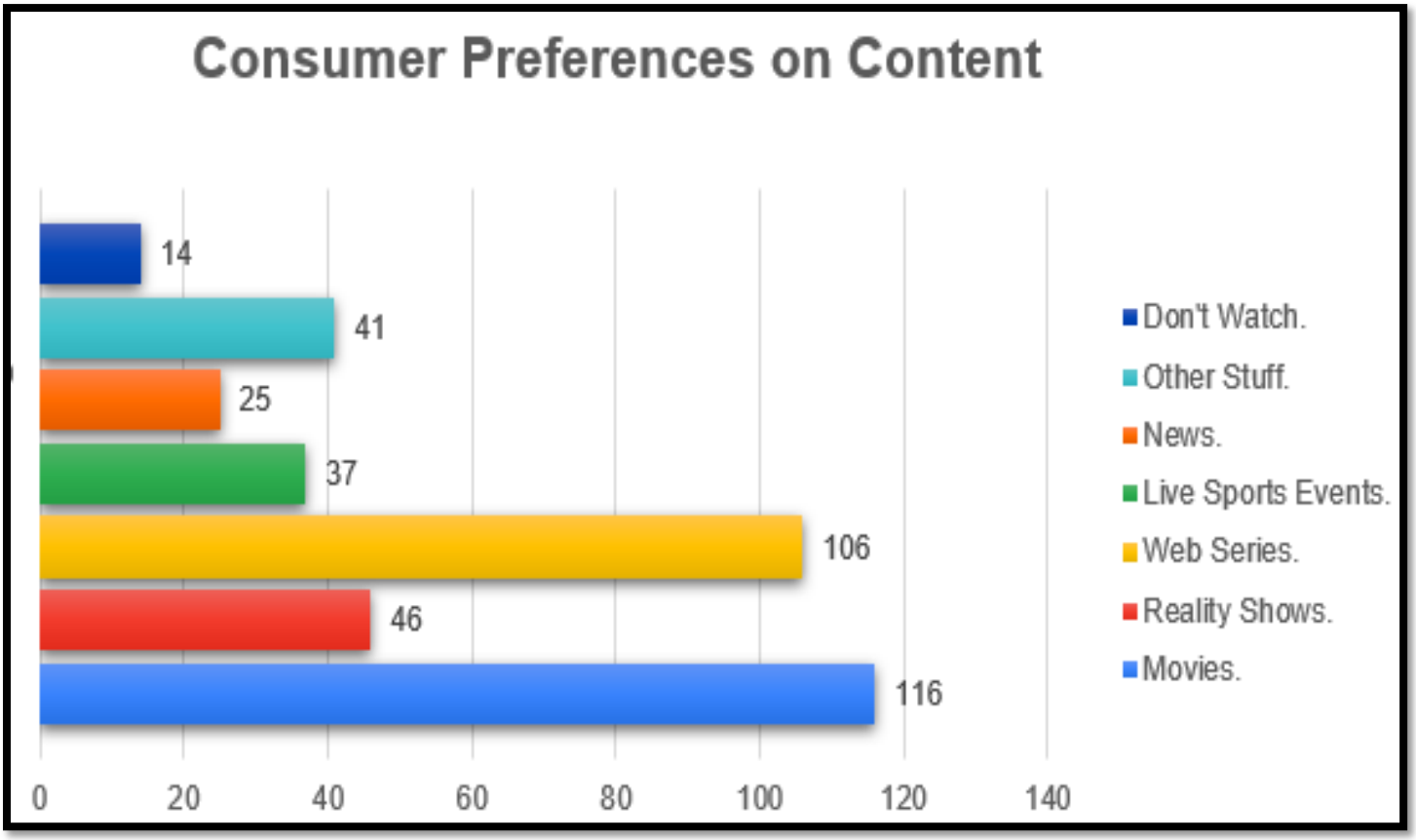

Fig 6

Fig. 6 reveals that dominantly consumers are preferring Movies with 116 votes followed by web series with 106 votes, reality shows with 46 votes, etc.

Time Spending pattern of Consumers on OTT

(Based on 750 responses)

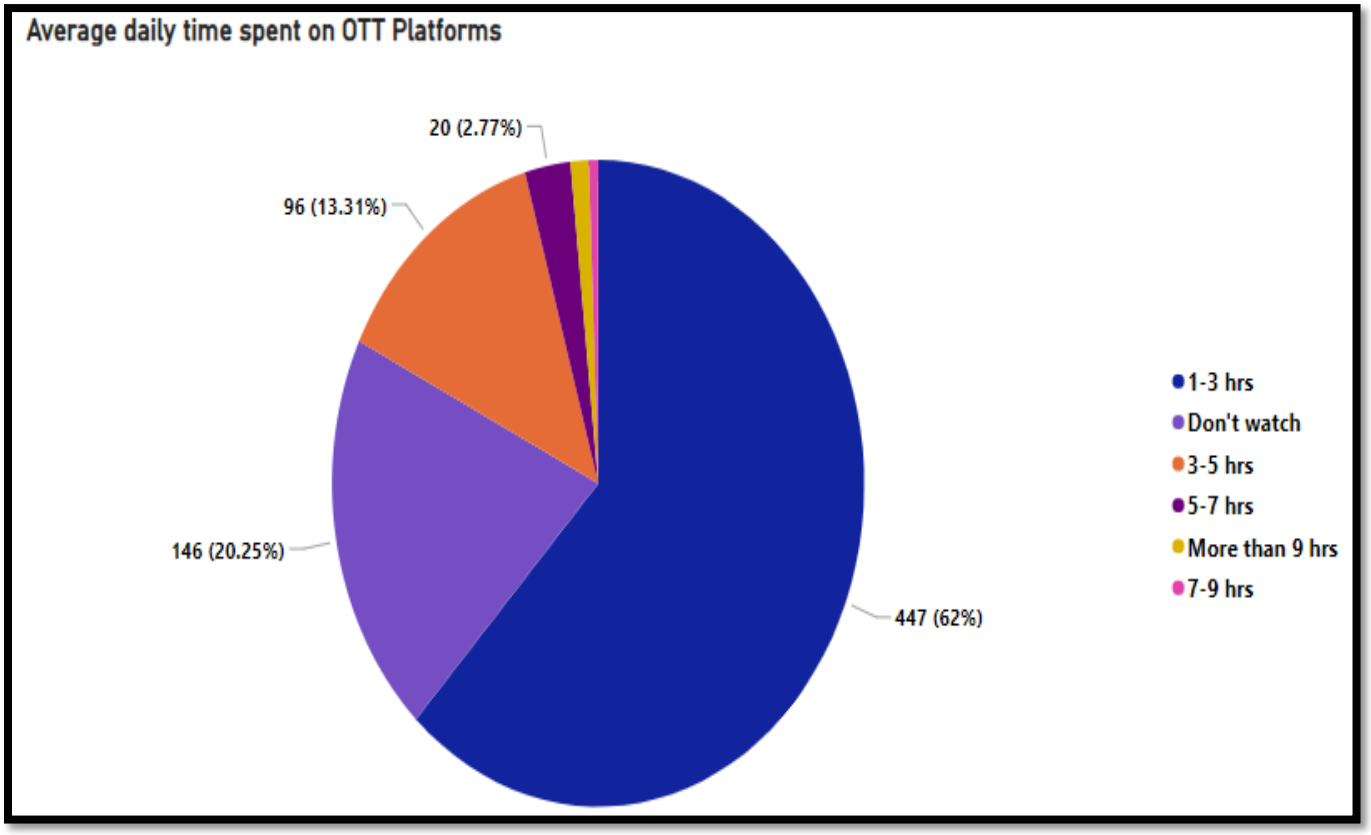

Fig 7

Fig 7 shows that on an average 62% of consumers are spending 1-3 hours on OTT out of total and the remaining spend less or more than that.

To Know the satisfaction level of Indians from OTT Platforms

The below Fig 8 reveals that 69.07% of respondents out of the total have voted for preference on OTT Platforms over TV and claim that consumption of OTT Platforms doesn't affect their studies/work.

This reveals that Indians are very satisfied with the use of OTT Platforms.

(Based on 750 responses)

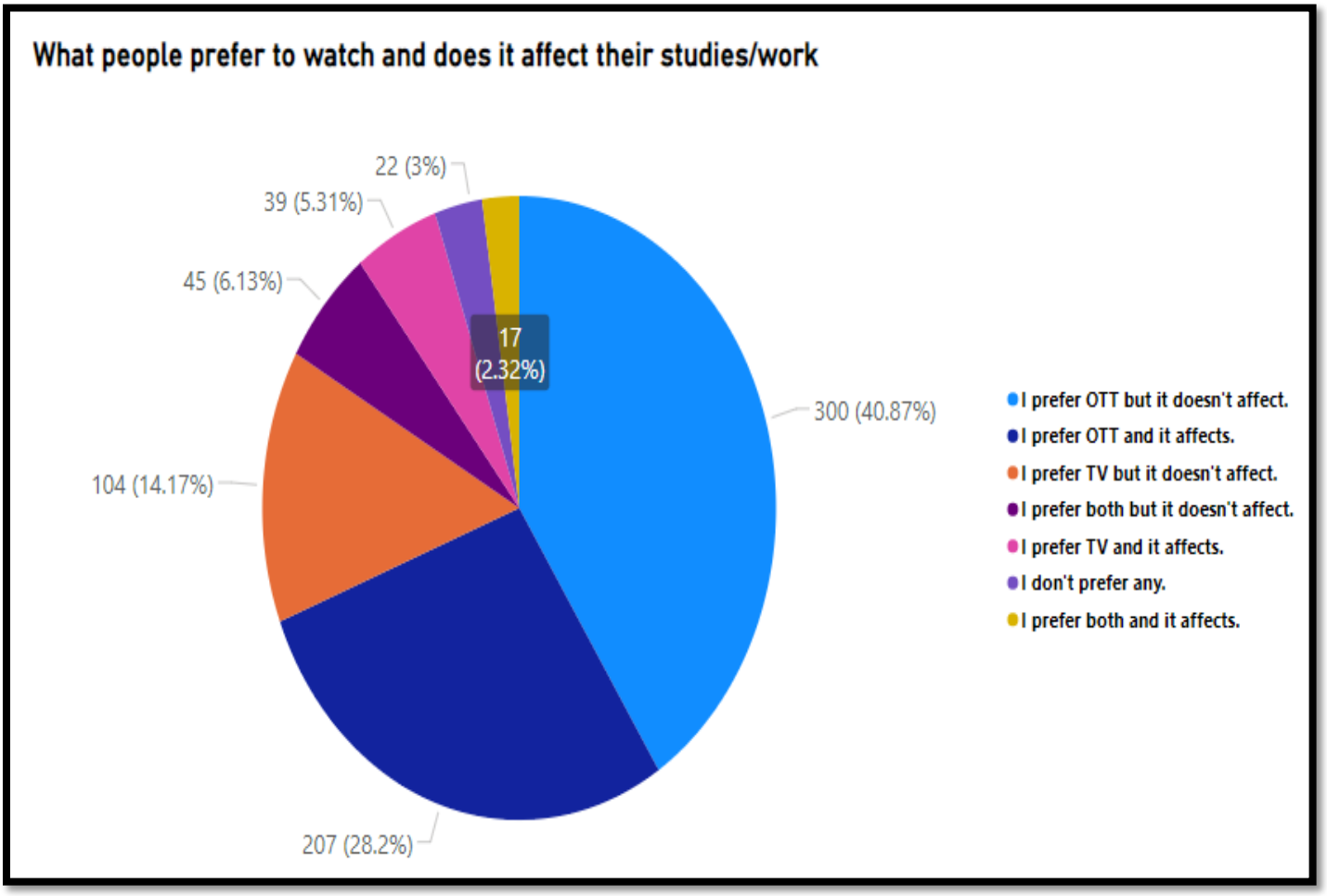

Fig 8

ANALYSIS IN PYTHON:

Comparing the usage of OTT Platforms with TV

	Time Spend on Platform					
Platform	1-3 H	3-5 H	5-7 H	7-9 H	More Than 9	Total
OTT	447	96	20	4	8	575
TV	347	55	16	1	3	422
Total	794	151	36	5	11	997

Chi-Sq Score table:

	Time Spend on Platform					
Platform	1-3 H	3-5 H	5-7 H	7-9 H	More Than 9	
OTT	457.92377	87.08625878	20.76229	2.8837	6.3440321	575
TV	336.07623	63.91374122	15.23771	2.1163	4.6559679	422
Total	794	151	36	5	11	997

The above table shows that in each category/group of OTT, consumers are more in score proving that people are more associated with OTT across groups as compared to TV.

This indicates that in today's world, people spend more time on OTT platforms and feel that OTT is more convenient than TV.

Descriptive Statistics and Chi-Sq test

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

plt.style.use('ggplot')

import nltk
```

```
# Reading the data set
df = pd.read_csv("C:\\Users\\Administrator\\Desktop\\Major Project\\RatingPrp.csv")
```

```
df.head()
```

In [4]:

Out[4]:

	ID	Data Type	Age	Occupation	Where do you live?	Amazon Prime	Disney + Hotstar	Voot	YouTube	Netflix
0	1	Primary	21-30	Student	Central	Don't watch	Don't watch	3	5	Don't watch
1	2	Primary	21-30	Student	Harbour	1	1	1	4	1
2	3	Primary	11-20	Student	Central	Don't watch	1	1	4	Don't watch
3	4	Primary	11-20	Student	Central	Don't watch	Don't watch	Don't watch	5	Don't watch
4	5	Primary	Above 50	Unemployed	Central	Don't watch	Don't watch	Don't watch	5	Don't watch

In [14]:
```
df.ID.value_counts().describe()
```

Out[14]:
```
count    750.0
mean       1.0
std        0.0
min        1.0
25%        1.0
50%        1.0
75%        1.0
max        1.0
Name: ID, dtype: float64
```

In [15]:
```
df.describe()
```

Out[15]:

	ID
count	750.00000
mean	375.50000
std	216.65064
min	1.00000
25%	188.25000
50%	375.50000
75%	562.75000
max	750.00000

In [16]:
```
# Ch- Square Implementation
```

In [17]:
```
import pandas as pd
import scipy.stats as stats
```

In [18]:
```
# create sample data according to survey
data = [['1-3 H','OTT'] for i in range(447)] + \
       [[ '3-5 H','OTT'] for i in range(96)] + \
       [['5-7 H','OTT'] for i in range(20)] + \
       [['7-9 H','OTT'] for i in range(4)] + \
       [['More than 9','OTT'] for i in range(8)] + \
       [['1-3 H','TV'] for i in range(347)] + \
       [['3-5 H','TV'] for i in range(55)] + \
       [['5-7 H','TV'] for i in range(16)] + \
       [['7-9 H','TV'] for i in range(1)] + \
       [['More than 9','TV'] for i in range(3)]
```

In [19]:
```
# defining the table
data1 = [[447,96,20,4,8], [347,55,16,1,3]]
```

In [20]:
```
df = pd.DataFrame(data ,columns = ['OTT','TV'])
```

In [21]:
```
alpha = 0.05
```

In [22]:
```
df
```

Out[22]:

	OTT	TV
0	1-3 H	OTT
1	1-3 H	OTT
2	1-3 H	OTT
3	1-3 H	OTT
4	1-3 H	OTT
...	...	...
992	5-7 H	TV
993	7-9 H	TV
994	More than 9	TV
995	More than 9	TV
996	More than 9	TV

997 rows × 2 columns

In [24]:
```
# create contingency table
data_crosstab1 = pd.crosstab(df['OTT'],
                             df['TV'],
                            margins=True, margins_name="Total")
```

In [25]:
```
#data_crosstab=data_crosstab1.transpose()
```

In [26]:
```
#data_crosstab
```

In [27]:
```
# Calcualtion of Chisquare
chi_square = 0
rows = df['OTT'].unique()
columns = df['TV'].unique()
```

In [28]:
```
for i in columns:
    for j in rows:
        O = data_crosstab1[i][j]
        E = data_crosstab1[i]['Total'] * data_crosstab1['Total'][j] / data_crosstab1['Total']['Total']
        chi_square += (O-E)**2/E
```

In [29]:
```
# The p-value approach
print("Approach 1: The p-value approach to hypothesis testing in the decision rule")
p_value = 1 - stats.chi2.cdf(chi_square, (len(rows)-1)*(len(columns)-1))
conclusion = "Failed to reject the null hypothesis."
if p_value <= alpha:
    conclusion = "Null Hypothesis is rejected."

print("chisquare-score is:", chi_square, " and p value is:", p_value)
print(conclusion)
```

```
Approach 1: The p-value approach to hypothesis testing in the decision rule
chisquare-score is: 4.879556151235902  and p value is: 0.2998805293281992
Failed to reject the null hypothesis.
```

In [30]:
```
# The critical value approach
print("\n-------------------------------------------------------------------------------------")
print("Approach 2: The critical value approach to hypothesis testing in the decision rule")
critical_value = stats.chi2.ppf(1-alpha, (len(rows)-1)*(len(columns)-1))
conclusion = "Failed to reject the null hypothesis."
if chi_square > critical_value:
    conclusion = "Null Hypothesis is rejected."

print("chisquare-score is:", chi_square, " and critical value is:", critical_value)
print(conclusion)
```

```
-------------------------------------------------------------------------------------
Approach 2: The critical value approach to hypothesis testing in the decision rule
chisquare-score is: 4.879556151235902  and critical value is: 9.487729036781154
Failed to reject the null hypothesis.
```

Conclusion:

Based on the chi-square statistical analysis conducted, the comparison between the tabular value and the calculated value reveals that the tabular value is found to be greater. In light of this result, we accept the null hypothesis and refrain from rejecting it. The outcome of this analysis suggests that there is no significant relationship between the watching time and the number of audience members on a specific platform. In other words, the duration of time spent watching does not appear to have a noteworthy impact on the number of individuals viewing content on the given platform.

Predictive Analysis

Library

In [711]:
```
from sklearn.feature_selection import SelectKBest, chi2
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import pandas as pd
from sklearn.preprocessing import LabelEncoder
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier,GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
```

Loading a Data Set

In [712]:
```
data = pd.read_csv("Final.csv")
```

In [713]:
```
data.head(5)
```

Out[713]:

	Age	Occupation	Region	Watch time OTT	Watch Time TV	Amazon Rate	Disney Hotstart Rate	Voot Rate	Netflix Rate	Other OTT	Other TV	Have you started watching OTT more due to lockdown ?	OTT Best than TV	Spend OTT	Spend TV
0	21-30	Student	Central	1-3 hrs	1-3 hrs	Don't watch	Don't watch	3	Don't watch	No	Dish TV(Tata sky, Airtel, etc)	Maybe	Maybe	0 to 200	200 to 600
1	21-30	Student	Harbour	Don't watch	Don't watch	1	1	1	1	No	Local cable	Yes	Yes	0 to 200	200 to 600
2	Nov-20	Student	Central	1-3 hrs	1-3 hrs	Don't watch	1	1	Don't watch	Pikashow	Dish TV(Tata sky, Airtel, etc)	Yes	Maybe	0 to 200	200 to 600
3	Nov-20	Student	Central	3-5 hrs	1-3 hrs	Don't watch	Don't watch	Don't watch	Don't watch	No	Dish TV(Tata sky, Airtel, etc)	No	No	0 to 200	200 to 600
4	Above 50	Unemployed	Central	1-3 hrs	1-3 hrs	Don't watch	Don't watch	Don't watch	Don't watch	No	Dish TV(Tata sky, Airtel, etc)	Maybe	Maybe	0 to 200	200 to 600

In [714]:
```
# Apply LabelEncoder to all columns
le = LabelEncoder()
df = data.apply(le.fit_transform)
df
```

Out[714]:

	Age	Occupation	Region	Watch time OTT	Watch Time TV	Amazon Rate	Disney Hotstart Rate	Voot Rate	Netflix Rate	Other OTT	Other TV	Have you started watching OTT more due to lockdown ?	OTT Best than TV	Spend OTT	Spend TV	What w you pref watch do affect studies/
0	0	4	0	0	0	6	6	2	6	22	1	0	0	0	1	
1	0	4	1	4	4	0	0	0	0	22	4	2	2	0	1	
2	5	4	0	0	0	6	0	0	6	24	1	2	0	0	1	
3	5	4	0	1	0	6	6	6	6	22	1	1	1	0	1	
4	3	5	0	0	0	6	6	6	6	22	1	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
716	0	4	0	0	0	5	0	5	5	22	0	1	1	0	1	
717	0	4	0	0	0	5	5	5	5	22	4	1	2	0	1	
718	0	4	1	0	0	5	5	5	5	22	0	1	1	1	1	
719	3	0	2	4	0	5	0	5	5	22	4	1	1	0	0	
720	2	0	1	2	4	3	3	1	4	22	0	2	2	3	0	

721 rows × 19 columns

Split Data X & Y

```
In [715]: X = df.drop('You Tube Rate', axis=1)
          y = df['You Tube Rate']
```

```
In [716]: X
```

Out[716]:

	Age	Occupation	Region	Watch time OTT	Watch Time TV	Amazon Rate	Disney Hotstart Rate	Voot Rate	Netflix Rate	Other OTT	Other TV	Have you started watching OTT more due to lockdown ?	OTT Best than TV	Spend OTT	Spend TV	What w you pref watch do affect studies/
0	0	4	0	0	0	6	6	2	6	22	1	0	0	0	1	
1	0	4	1	4	4	0	0	0	0	22	4	2	2	0	1	
2	5	4	0	0	0	6	0	0	6	24	1	2	0	0	1	
3	5	4	0	1	0	6	6	6	6	22	1	1	1	0	1	
4	3	5	0	0	0	6	6	6	6	22	1	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
716	0	4	0	0	0	5	0	5	5	22	0	1	1	0	1	
717	0	4	0	0	0	5	5	5	5	22	4	1	2	0	1	
718	0	4	1	0	0	5	5	5	5	22	0	1	1	1	1	
719	3	0	2	4	0	5	0	5	5	22	4	1	1	0	0	
720	2	0	1	2	4	3	3	1	4	22	0	2	2	3	0	

721 rows × 18 columns

```
In [717]: y = data.iloc[:, -1]
```

```
In [718]: y
```

Out[718]:

```
0                5
1                4
2                3
3      Don't watch
4                4
          ...
716              5
717              5
718              5
719              5
720              5
Name: You Tube Rate, Length: 721, dtype: object
```

Apply feature Engineering

In [746]:
```
# Apply SelectKBest
k = 10  # Number of top features to select
selector = SelectKBest(score_func=chi2, k=k)
selector.fit(X, y)
```

Out[746]:
```
SelectKBest(score_func=<function chi2 at 0x000001F4A4932C10>)
```

In [747]:
```
# Get selected feature indices
feature_indices = selector.get_support(indices=True)
```

In [748]:
```
# Get selected feature names
feature_names = X.columns[feature_indices]
```

In [749]:
```
# Print selected features with original column names
print("Selected Features:")
for name in feature_names:
    print(name)
```

```
Selected Features:
Age
Region
Watch time OTT
Watch Time TV
Voot Rate
Netflix Rate
OTT Best than TV
What would you prefer to watch and does it affect your studies/work ?
What do you prefer mostly to watch on OTT?
What do you prefer mostly to watch on TV?
```

In [750]:
```
cols = selector.get_support(indices=True)
```

In [751]:
```
cols = selector.get_support(indices=True)
x = X.iloc[:,cols]
pd.DataFrame(x)
```

Out[751]:

	Age	Region	Watch time OTT	Watch Time TV	Voot Rate	Netflix Rate	OTT Best than TV	What would you prefer to watch and does it affect your studies/work ?	What do you prefer mostly to watch on OTT?	What do you prefer mostly to watch on TV?
0	0	0	0	0	2	6	0	6	44	32
1	0	1	4	4	0	0	2	6	44	32
2	5	0	0	0	0	6	0	5	44	32
3	5	0	1	0	6	6	1	6	44	32
4	3	0	0	0	6	6	0	6	44	32
...	...	...	...	...	...	...	...	...	...	...
716	0	0	0	0	5	5	1	4	44	32
717	0	0	0	0	5	5	2	1	44	32
718	0	1	0	0	5	5	1	3	44	32
719	3	2	4	0	5	5	1	4	44	32
720	2	1	2	4	1	4	2	2	44	32

721 rows × 10 columns

Train Test

In [752]:
```
from sklearn.model_selection import train_test_split

# Assuming you have your features (x) and target variable (y) defined

# Split the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=0)
```

In [753]:
```
x_train.shape
```

Out[753]:
```
(540, 10)
```

In [754]:
```
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
lr
```

Out[754]:
```
LogisticRegression()
```

In [755]:
```
lr.fit(x_train,y_train)
```

```
D:\Anaconda\lib\site-packages\sklearn\linear_model\_logistic.py:814: ConvergenceWarning: lbfgs failed to con
verge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:
    https://scikit-learn.org/stable/modules/preprocessing.html (https://scikit-learn.org/stable/modules/prep
rocessing.html)
Please also refer to the documentation for alternative solver options:
    https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression (https://scikit-learn.org/
stable/modules/linear_model.html#logistic-regression)
  n_iter_i = _check_optimize_result(
```

Out[755]:
```
LogisticRegression()
```

In [756]:
```
y_pred=lr.predict(x_test)
y_pred
```

Out[756]:
```
array(['5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '3', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '1', '5', '5', '5', '5', '5',
       '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '4', '5', '5', '5', '5', '5', '5', '5', '5', '5', '5',
       '5', '5', '5', '5', '5', '5', '5', '5', '3', '5', '5', '5'],
      dtype=object)
```

In [757]:
```
y_test
```

Out[757]:
```
503    5
646    5
245    5
142    4
493    5
      ..
306    5
77     3
665    5
258    5
109    5
Name: You Tube Rate, Length: 181, dtype: object
```

```
In [758]: from sklearn.metrics import confusion_matrix
          confusion_matrix(y_test,y_pred)

Out[758]: array([[  0,   0,   0,   0,   1,   0],
                 [  0,   0,   0,   0,   1,   0],
                 [  0,   0,   2,   0,   3,   0],
                 [  0,   0,   0,   0,   8,   0],
                 [  1,   0,   0,   3, 161,   0],
                 [  0,   0,   0,   0,   1,   0]], dtype=int64)

In [784]: rf = RandomForestClassifier()
          gb = GradientBoostingClassifier()
          svc = SVC()
          lr = LogisticRegression()

In [790]: from sklearn.metrics import accuracy_score
          print("Accuracy of Logistic Regression:", accuracy_score(y_test, y_pred))

          Accuracy of Logistic Regression: 0.8895027624309392

In [799]: gb.fit(x_train,y_train)
          y_pred2 = gb.predict(x_test)
          print("Accuracy of Gradient Boosting:", accuracy_score(y_test, y_pred2))

          Accuracy of Gradient Boosting: 0.861878453038674
```


```
In [784]: rf = RandomForestClassifier()
          gb = GradientBoostingClassifier()
          svc = SVC()
          lr = LogisticRegression()

In [790]: from sklearn.metrics import accuracy_score
          print("Accuracy of Logistic Regression:", accuracy_score(y_test, y_pred))

          Accuracy of Logistic Regression: 0.8895027624309392

In [799]: gb.fit(x_train,y_train)
          y_pred2 = gb.predict(x_test)
          print("Accuracy of Gradient Boosting:", accuracy_score(y_test, y_pred2))

          Accuracy of Gradient Boosting: 0.861878453038674
```


```
In [800]: rf = RandomForestClassifier(max_samples=0.75,random_state=42)
          rf.fit(x_train,y_train)
          y_pred3 = rf.predict(x_test)
          print("Accuracy of Support Vector Classifier:", accuracy_score(y_test, y_pred3))

          Accuracy of Support Vector Classifier: 0.8895027624309392

In [801]: svc.fit(x_train,y_train)
          y_pred4 = svc.predict(x_test)
          print("Accuracy of Support Vector Classifier:", accuracy_score(y_test, y_pred4))

          Accuracy of Support Vector Classifier: 0.9116022099447514
```

Conclusion:

In our project dataset, we encountered the need to handle character values and convert them into a numeric format. To address this, we performed imputation techniques to transform the character values into appropriate numerical representations.

Afterwards, we utilized the chi-square SelectKBest feature selection technique in Python. By employing this approach, we were able to determine the top 10 features that have a significant impact on the YouTube rating. These selected features were considered as the most influential factors affecting the rating.

Following the feature selection step, we proceeded to train and evaluate various classification models using the dataset. The accuracy scores achieved by these models were found to be quite similar, indicating that they performed comparably in terms of predictive accuracy. However, among the different models, the Support Vector Classifier (SVC) stood out with the highest accuracy score of 91%. This outcome suggests that the SVC model is particularly well-suited for our dataset and provides a strong fit for our classification task.

CONCLUSION

Based on the analysis, the research on the topic “Market Research and Analytics on the Rise of OTT Platforms: A Study of Consumer Behaviour” has been successfully completed. It is concluded that people in India are fully aware of the concept of OTT Platforms and even use them in their daily lives for watching entertainment content like web series, movies, and reality shows, etc. We have seen that Over-the-top content streaming platforms in India have led to the emergence of distinct patterns of content consumption in the last few years. In this research, we have observed that mostly young generations are attracted to this medium of entertainment although every age group consumes it. After YouTube, Netflix is found to be the most used platform among all, as it is well known that content produced by Netflix is unique and addictive in terms of entertainment, so do follow by Amazon Prime, as Prime produces originally Indian content mostly.

There are various reasons for choosing OTT over TV as we have discussed before, Entertainment, Mobility, Content on demand, Unique content, Cost Effectiveness, language and ease of use. If we talk about monthly spending it depends on the consumer but mostly consumers are spending 0-200 rupees on an average.

The average time spent on OTT platforms by consumers also varies upon their flexibility and availability of time for entertainment but as research reveals it varies between 1-3 hours. Most consumers are from a young age and nowadays youngsters are preferring Movies as well as Web series maybe because series continue a plot at more realistic pace, permit more character growth, and create a comfortable sense of familiarity for the audience.

Recently because of Covid-19, the majority of content consumption occurred inside, and that too on a very high scale. The lockdown has led to an increase in consumption patterns of the users by 1-3 hours per day on an average.

The data collected from the questionnaire survey was interpreted and analyzed in previous sections. Since we have observed massive consumption of OTT platforms the statement Rise of OTT Platforms has been proved.

KEY FINDINGS

1. 100% of respondents are aware of OTT Platforms.

2. Majority of the respondents using OTT Platforms are from the age of 11-30.

3. Netflix with 41% and Amazon Prime with 26% are Dominant OTT Platforms followed by Disney+ Hotstar with 17.2%.

4. Entertainment, Mobility, Content on demand, Unique content, Cost Effectiveness, language, and ease of use are some of the primary reasons for preferring OTT over TV by consumers.

5. On average consumers are spending 0-200 rupees on OTT monthly subscriptions followed by 69.07% out of total and remaining are spending more than that.

6. On average consumers are spending 1-3 hours on OTT followed by 62% out of total and remaining spend less or more than that.

7. Consumers are preferring web series with 75.6% followed by movies with 59.2%, reality shows with 26.9% etc., to watch entertainment content on OTT.

8. COVID-19 increased the average consumption of OTT among users with 1-3 hours per day followed by 66.2% out of total.

RECOMMENDATIONS

This Study was limited within India. Further Studies can be done by comparing India with other countries for OTT Platforms.

REFERENCES

- Market Research and Analytics on Rise of OTT Platforms A Study of Consumer Behaviour.pdf (ijaem.net)
- Google
- Kaggle