
Exploratory Data Analysis on OTT Platforms

(Effects of OTT Platforms on the mindset of Indian Youth)

THESIS

SUBMITTED TO

VISHWAKARMA UNIVERSITY, PUNE

FOR THE DEGREE

OF

BACHELOR OF SCIENCE

IN

STATISTICS

BY

PRATIK PATIL

AMISHA SAHU

Under the supervision of

Prof. Dipti Joshi

Assistant Professor

Dept. of Statistics

Prof. Madhulika Mishra

Assistant Professor

Dept. of Statistics

Dr. Nazia Wahid

Head of Department, Statistics

DECLARATION

We,

Pratik Patil (201800665),

Amisha Sahu (201800667)

Hereby declare that the work embodied in this project entitled **“Effects of OTT Platforms on the mindset of Indian Youth”** submitted to the Vishwakarma University, Pune is a record of an original work carried out by us under the guidance of **Mrs. Dipti Joshi** and **Ms. Madhulika Mishra**, Assistant Professor, Department of Statistics, Vishwakarma University, Pune.

This project work is submitted in the partial fulfillment of the requirements for the award of the degree of Batchelor of Science in Statistics (Hons.).

I further declare that the work reported in this project has not been submitted and will not be submitted, either in part or full, for the award of any other degree or diploma in this Institute or any other Institute or University.

Pratik Patil
Amisha Sahu

B.Sc. Statistics
Department of Statistics
Vishwakarma University

Pune

Certificate

This is to certify that the project titled **“Effects of OTT Platforms on the mindset of Indian Youth”** submitted by **Pratik Patil** and **Amisha Sahu** is an original work and has not been previously submitted in part or full for the award of any degree or diploma to this or any other university. This project is submitted to **Vishwakarma University** Pune, in partial fulfillment of the requirement for the award of the degree of **B.Sc.** in the subject **Statistics**.

Date:

Place: Pune

Dr. Nazia Wahid

(H.O.D. of Statistics Dept.)

Prof. Dipti Joshi

(Project Guide/Mentor)

Prof. Madhulika Mishra

(Project Guide/Mentor)

Acknowledgement

Inspiration and Motivation have always played key roles in the success of any venture. It is our good fortune and a matter of pride and privilege to have the esteemed supervision of **Mrs. Dipti Joshi and Ms. Madhulika Mishra**, Assistant professor, Department of Statistics, Vishwakarma University, Pune. It is only their personal influence, expert guidance and boundless support that enabled us to complete the work.

We would also like to extend our gratitude to **Dr. Nazia Wahid**, Head of Department, Statistics for her support.

We express our sincere thanks and deep sense of gratitude towards everyone involved in any way whatsoever for the shaping of this Project.

Date:

Pratik Patil

Amisha Sahu

B.Sc. Statistics
Department of Statistics
Vishwakarma University
Pune

Index

Sr. No	Particulars	Page No.
1	Objective	6
2	Introduction	6
3	Literature Review	7
4	Project Description	9
5	Methodology	9
6	Tools/Software Requirements	10
7	Data Cleaning	11
8	Analysis	13
9	Conclusion	34
10	Future Scope of study	36
11	References	37

Objective

1. To find out if there are significant psychological effects of online web series and video content on youth.
2. To find relation between Behavioural, Life Style or other changes and the nature of content streamed on OTT Platforms.
3. To find impact on academic performance and physical activities in the youth due to binge watching.
4. To explore the perception of youth regarding nature of online video content and web series in India.

Introduction

OTT stands for over-the-top, an over-the-top media service is a media service offered directly to viewers via the Internet. OTT bypasses cable, broadcast, and satellite television platforms, the types of companies which traditionally act as controllers or distributors of such content.

In the past few years, the online entertainment industry in India has marked a major growth. This has all been possible because of the changing lifestyle of the Indian viewers, significant change in the type of content consumed and the introduction of the various online video streaming platforms on which the content is served to them. With the introduction of faster and cheaper internet services in the form of mobile connection and wide availability of broadband services, there has been a major growth in the access to online content, in terms of the number of people visiting these online websites or apps, and the time they spend on watching online videos.

A web series in simpler words is a series of short video episodes, which are released over a certain period of time in order to tell a story to the viewers. Earlier, it was much popular in western countries, but with globalization, the trend of online content streaming became popular in India as well.

The introduction of Netflix, Amazon Prime, ALT Balaji, MX Player, TVF and a number of other web and App based online video streaming platforms, together with faster internet services have become very popular in the Indian entertainment industry.

With growing popularity, the online content streaming has become a profitable opportunity for the content producers, and it further has resulted in more and more content being uploaded on the internet.

The main target population of the web series and other online streaming content is the younger generation, which is mainly comprised of teenagers, and the youth between 18 to 30 years of age.

Working on this model helps these platforms to attract the youth, who is looking for entertainment on their smartphone or laptop devices. As the youth finds today's television soap boring, meaningless, and lack of originality, they tend to watch a lot of drama, comedy, action, love & romance, thrill and more, which comes to them in the form of the web series or episodes and movies.

Literature Review

The Culture of Watching Online content in India.

It is becoming apparent that the attraction for traditional television viewing is declining amongst the Indian youth. The famous OTT platforms like Netflix, Amazon Prime and other video streaming websites provide an instant access to wide categories of video content, which attracts the Indian youth.

In the last decade, India has witnessed a technologically transformative media landscape that has promoted the entry of mass-market of niche channel, content and television technologies. The new platforms of engaging with television media, largely through smart phones and now through Internet TV/ Video-On-Demand (VOD) and OTT platforms gave rise to changing viewing pattern and platform choices.

It must be noted that the most popular amongst OTT platforms in India is Hotstar, launched in 2015 and owned by Star TV (India's biggest private broadcaster), with a total loyal user base of 75 million people in the urban and rural areas, which is followed by Amazon Prime video with 11 million subscribers (Aadeetya, 2018). While Netflix only comes in the third with around 5 million subscribers, it nevertheless offers a more interesting focus of enquiry on account of its many differences from other players and the more particular audiences it targets – niche, English speaking, affluent, urban youth (Joglekar, 2018).

Chun-Mei-Chen (2019), studied the rise of OTT platforms due to the increase in data availability and usage. It has always proved to be a great advantage for people by ensuring them with a personalized digital life due to the collaboration of telecommunications operators and the OTT service providers, like, Amazon, Reliance, and Eros International. Initially, the OTT services had brought about a turning point in phone calls and text messages of the telecommunications operators as it provided all of these for free through the Internet along with media related content. Internet penetration has eliminated the additional assistance provided by these operators with the Internet population rate crossing over 70% in India and other countries, as well.

Moving ahead, OTT platforms took an advantage of a strong audience base that is increasingly growing accustomed to consuming entertainment content through smartphones (Laghate, 2018). The online streaming platforms have now swooped in to take their slice of the market. As the features, contents and contexts of this online video streaming differ sufficiently from traditional TV, it remains a challenge to investigate how far traditional media theories and methodologies can inform research on the changing audience and their viewing pattern. (Livingstone, 2003, 2004, 2007).

Scott Fitzgerald (2019), noted that India is a battleground for Amazon and Netflix and shall emerge to be territory for the expansion of OTT video platforms. YouTube has been an open platform for major Indian broadcasters containing media libraries for all sorts of video content. The journal depicts the birth of a new global Internet TV network across numerous countries with millions of subscribers worldwide. The three types of OTT apps in the Indian market include: communication services, application ecosystems and audio and video content. Several regional content entertainment channels have expanded the OTT sector and media economy in the country including global leaders and major broadcast network-backed platforms, VOOT and OZeeTV. The digital infrastructure with respect to pipes, data, devices and habits have increased the video consumption, and by 2021, the total number of mobile connected devices shall reach to 59% from the current 35% with fee-based subscription model. As a result, we have seen consolidation of television and film via digitalization in the International Journal of Modern Agriculture, Volume 10, No.2, 2021 ISSN: 2305-7246

Sumitra Saha & V.K Satya Prasad (2021), in their research paper “Consumption Pattern Of OTT Platforms in India” determined how the consumption pattern of the growing OTT industry in India is changing rapidly and how new app/ web based platform are being introduced in the market. In their study, they showed hoe YouTube, Netflix, Amazon Prime, MX Player, SonyLIV and other such platforms are gaining new subscribers and expanding their shares in the market.

As per them entertainment is the key driver on digital platforms that draws the attention and generates interest among the widespread urban and rural viewers the online video platforms are

The online video platforms are promising enough in bringing a lot more opportunities for business expansion and growth of the bottom line. Enriched and qualitative touch points amalgamated with technological investments and impacted branding will assist the online video streaming platforms in building a critical mass. Deeper brand connect and stronger affinities have led to successful consumer engagement with the right content

Project Description

This paper is an analysis of various effects of web series and streaming content available on various OTT Platforms on the Indian Youth. We have examined the perception of youth regarding the web series and online streaming content that is available on online platforms like YouTube, Netflix, Amazon Prime, MX Player, and more. This study examines the psychological effects and behavioural changes amongst the youth because of web shows. This study is conducted by using quantitative research methodology, using the questionnaire research method. The respondents (who watch online video content) of the questionnaire are undergraduate and postgraduate students, and young working professionals from different fields.

Keywords: OTT Platforms, Web Series, Online Streaming, Youth, Online Content, Impact, Correlation.

Methodology

In order to get the utmost from this research study, we used the quantitative method to conduct this study. This particular study is designed as exploratory research, where it aims to study the various effects of web series and online streaming content on the psychology of Indian youth.

Data collection method:

Primary Data Collection:

- An online survey was conducted by sharing a questionnaire via Google Forms.
- The form comprised of 26 questions relevant to carry out this study and was majorly circulated through cross-platform messaging, like, WhatsApp and Instagram for a couple of days.
- Questionnaire was distributed with the purpose of collecting information about individual's age, gender, educational, family income, watching time (in hours) of online content, their personal opinions on the psychological effects of OTT platforms faced (if any).
- A total of 151 responses were collected.
- Answers related to the psychological and day-to-day life questions help us in providing information of None to Extreme effects of OTT Platforms on an individual.

Primary data was collected and data cleaning was done using MS Excel and Python programming. The data was sorted depending on different age groups ranging from 15 to 30 as that is considered to be the standard age for Indian Youth according to

WHO (World Health Organization). Also, the qualitative data collected was converted to quantitative and categorical data respectively.

This sorted data was then analyzed by using statistical tools like Correlation and Regression (model based on a relevant hypothesis) for finding the result of our problem statement “Does OTT platforms affect the mindset of Indian Youth”. The Data Visualization is done using MS Excel, R-Programming and Python.

Tools/Software requirements

- **Statistical Tools:**

1) Correlation: Correlation is a statistical measure that expresses the extent to which two variables are linearly related (meaning they change together at a constant rate). In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data.

2) Chi-Square Test: A chi-square (χ^2) statistic is a test that measures how a model compares to actual observed data. Chi-square tests are often used in hypothesis testing. The chi-square statistic compares the size of any discrepancies between the expected results and the actual results, given the size of the sample and the number of variables in the relationship.

3) Shapiro Test (Normality test for ANOVA): In statistics, normality tests are used to determine if a data set is well-modelled by a normal distribution and to compute how likely it is for a random variable underlying the data set to be normally distributed. The two well-known tests of normality, namely, the Kolmogorov–Smirnov test and the Shapiro–Wilk test are most widely used methods to test the normality of the data. The Shapiro–Wilk test is a test of normality in frequentist statistics.

4) ANOVA: Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among means. ANOVA is a statistical method that separates observed variance data into different components to use for additional tests. A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

- **Software:**

- MS Excel
- R Software
- Python

Data Cleaning

The collected data of 151 individuals was cleaned and sorted by using MS Excel and Python. The qualitative data was mostly converted into Categorical data comprising of both Nominal and Ordinal types of data.

1) MS Excel:

Most of our questions comprised of ordinal data, for example:

“Do you think binge watching web-series and online videos cause insomnia or depression?”

“Options: ‘Not at all’, ‘Slightly’, ‘Moderately’, ‘Very Much’, ‘Extremely’ ”

These options were converted to 1, 2, 3, 4 and 5 respectively.

2) Python:

Few other questions comprising of nominal data were sorted using Python programming.

Example: For the question, “What do you prefer to watch the most on online platforms?”

“Options: ‘Web-Series’, ‘TV Shows’, ‘Movies’ ”

The frequencies of these options were measured as 0 or 1 by using the command: **pd.get_dummies()** from the Python Pandas library.

Which gave us a result like:

Web Series	Tv Shows	Movies
1	0	0
1	0	0
1	0	0
0	0	1
1	0	0

Also, for questions like “Which top two Genres do you watch the most?”
The complex data:

Which top two Genres do you watch the most?
Adventure,Thriller/Horror
Crime,Drama
Romance,Drama
Adventure,Sci-Fi/Fantasy
Romance,Comedy

It was hard to calculate frequency of each genre independently so a simple code was used:

```
data1=pd.read_excel(r'E:\Study\Sem 6\Project\Data_New_1\Categorical_new_Genre.xlsx')
print(data1)
for row in data1['8) What genre of Movies/Web Series/Tv Shows do you watch the most?']:
    for platform in row.split(','):
        data1[platform]=0

counter=0
for row in data1['8) What genre of Movies/Web Series/Tv Shows do you watch the most?']:
    for platform in row.split(','):
        data1[platform][counter]=1
    counter+=1
data1.to_excel(r'E:\Study\Sem 6\Project\Data_New_1\Categorical_new_Genre_1.xlsx')
```

And the data was converted to the from:

Adventure	Thriller/Horror	Crime	Romance	Sci-Fi/Fantasy	Drama	Comedy
1	1	0	0	0	0	0
0	0	1	0	0	1	0
0	0	0	1	0	1	0
1	0	0	0	1	0	0
0	0	0	1	0	0	1

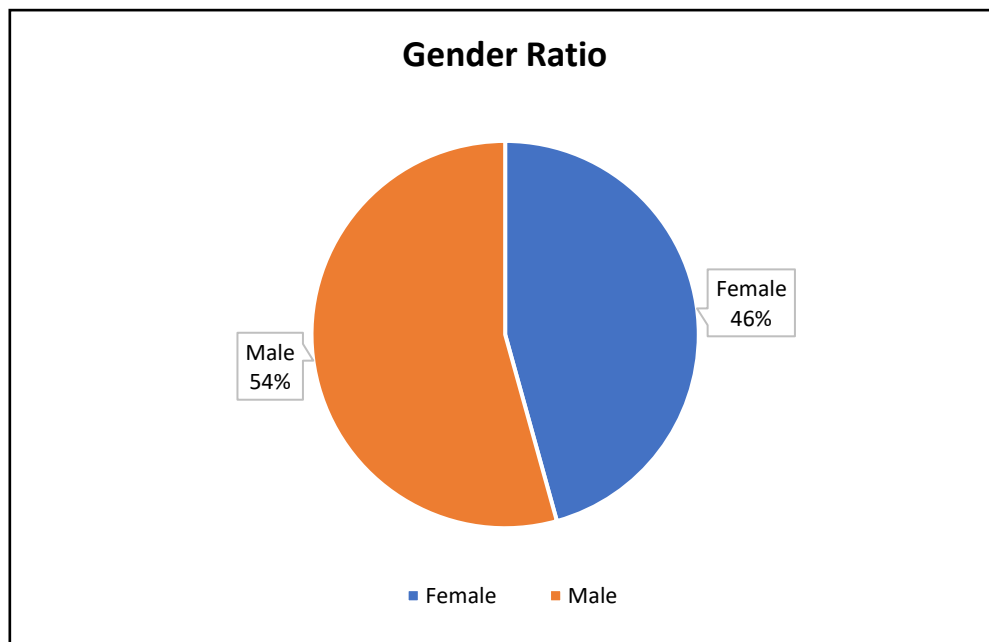
This made it easier to calculate individual frequencies of each genre.

Analysis

The analysis of all the problem questions is done using R-Software (RStudio) and visualization is done using R, MS Excel and Python.

- Gender Ratio:

Gender	Frequency
Female	69
Male	82
Grand Total	151

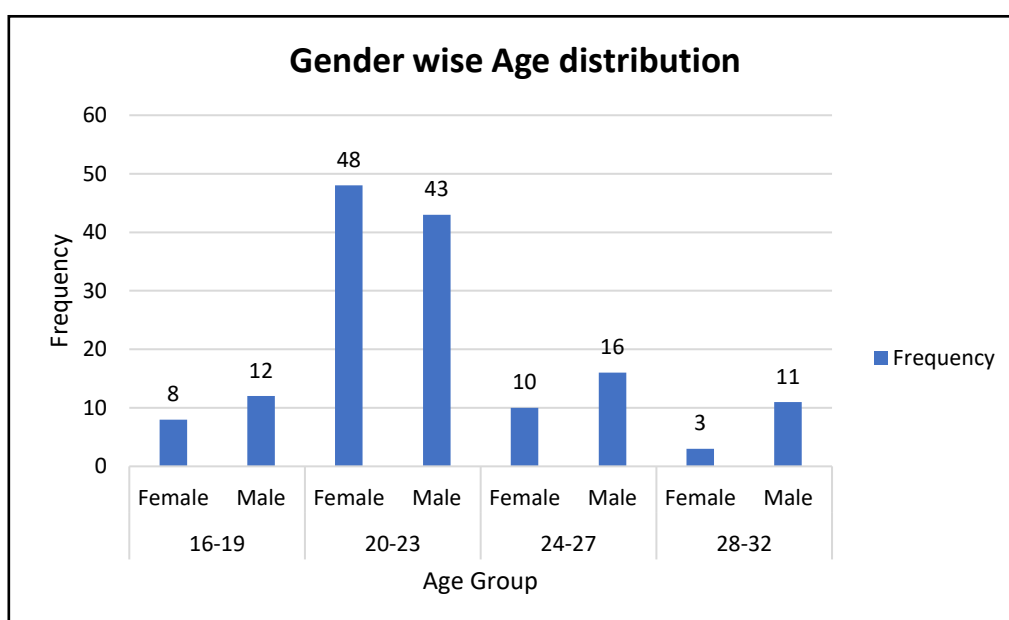


Inference:

Out of 151 responses collected in the survey, there are 82 (54%) Males, 69 (45%) Females.

- Age Groups:

Age Group	Gender	Frequency	Total
16-19	Female	8	20
	Male	12	
20-23	Female	48	91
	Male	43	
24-27	Female	10	26
	Male	16	
28-32	Female	3	14
	Male	11	
Grand Total			151

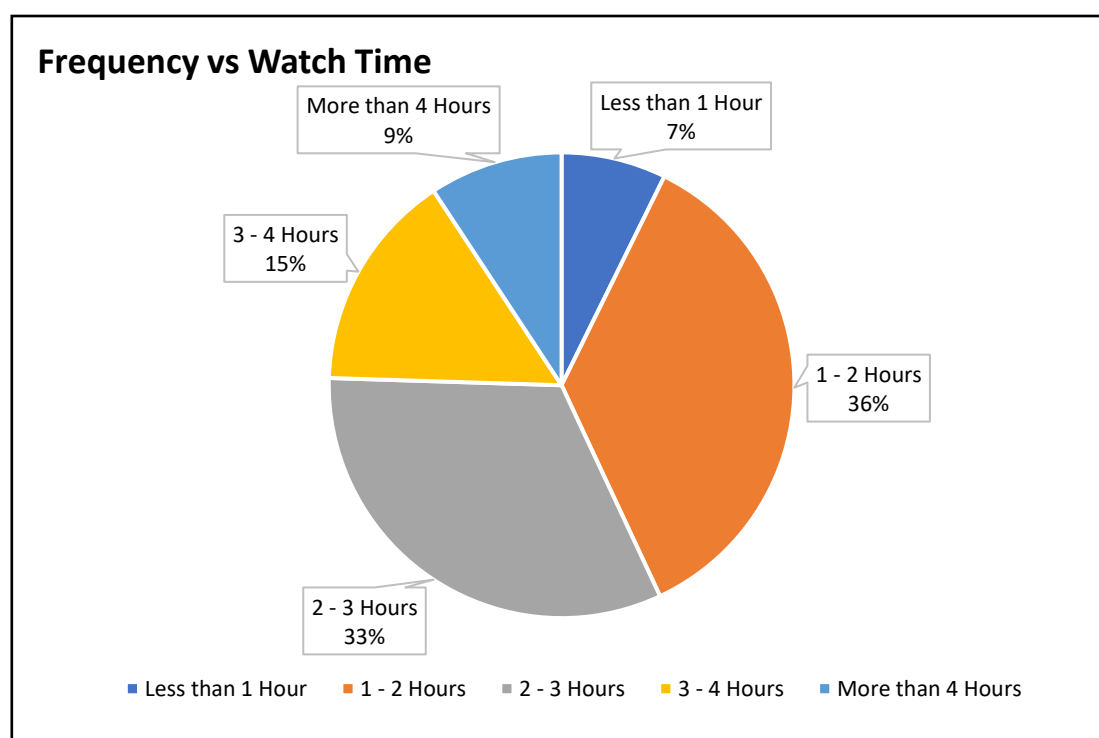


Inference:

Majority (60%) of the data collected belong to the age group ranging between 20 to 23 years which mostly includes students perceiving their UG and PG, the remaining 40% data was distributed between age groups 24 to 27 years (17%) with 26 respondents comprising of working professionals, then students from age 16 to 19 years (13%) and finally 10% people from 28 to 32 years of age.

- Watching Time of online content:

No. of Hours	Frequency
Less than 1 Hour	11
1 - 2 Hours	54
2 - 3 Hours	49
3 - 4 Hours	23
More than 4 Hours	14
Grand Total	151

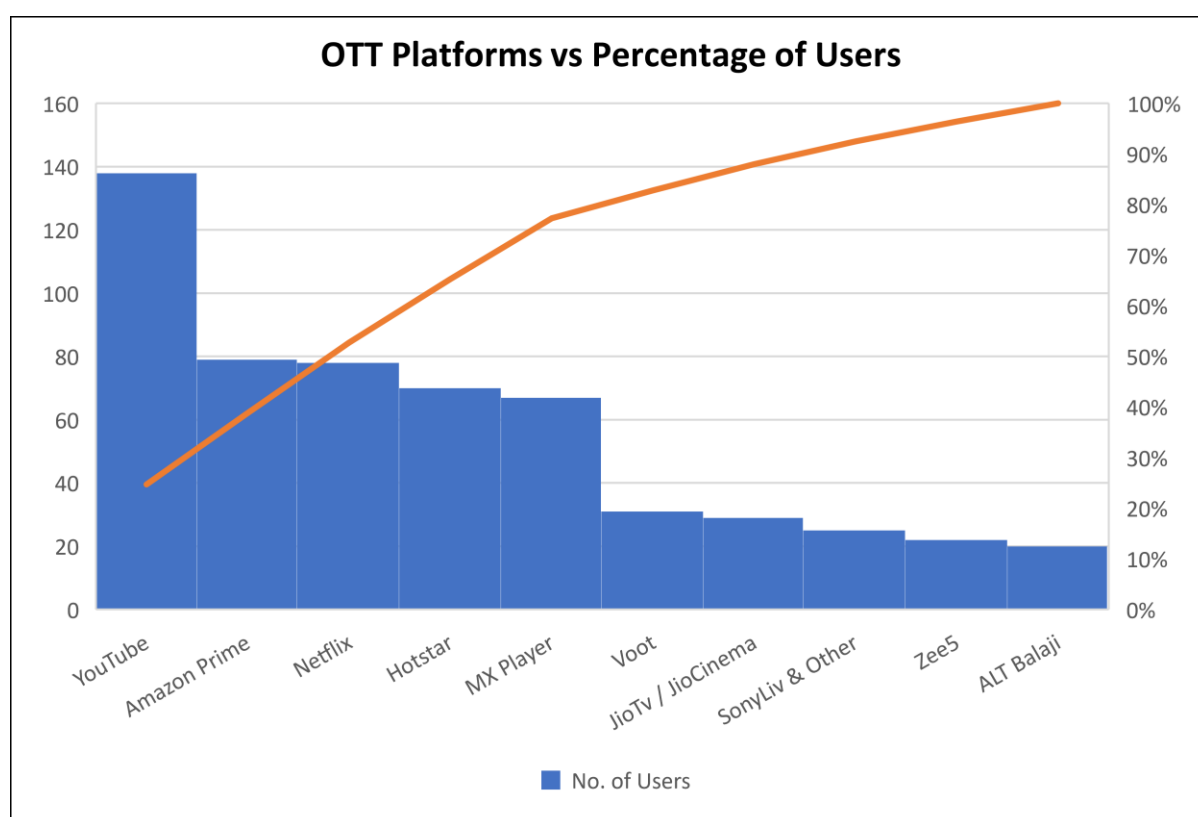


Inference:

The highest frequency i.e., 36% is of people watching OTT content for 1 to 2 Hours and the second highest frequency is for 2 to 3 hours (33%), both these time intervals have almost 70% of the total population. The remaining 31% of the respondents watching 3 to 4 hours, more than 4 hours and Less than an hour are 15%, 9% and 7% respectively.

- Preference of OTT Platforms:

Platform	No. of Users
YouTube	138
Netflix	78
Amazon Prime	79
Hotstar	70
Zee5	22
MX Player	67
ALT Balaji	20
JioTv / JioCinema	29
Voot	31
SonyLiv & Other	25



Inference:

Here, we have ranked all the potential OTT platforms that Indian Youths prefer according to their scores.

Youtube has the top score of 138 (91%) followed by Amazon Prime(52%), Netflix(51%), Hotstar(46%), MX Player(44%) and others have values less than 20%.

The Pareto chart shows that the first five platforms cover more than 80% of the viewing audience and the remaining platforms namely Voot, JioTv/JioCinema, Zee5, ALT Balaji and SonyLIV&Others have a low audience.

For carrying out the analysis based on our objective, we formed Four questions correspondingly:

Q.1) Is there any relationship between the psychological effects like Depression or Aggression and the factors like gender of the person along with the time they spend while watching different genres?

Q.2) Is the academic or professional work of a person affected by the amount of time they spend watching online content depending on their gender?

Q.3) Does the time spent on OTT Platforms have any significant effect on the physical activities and social interaction of a male or female?

Q.4) What kind of content does the youth find more appealing to watch and does it vary based on their gender?

The analysis and results are as follows:

➤ Analysis of Question 1:

• Relationship Between ‘Gender’ and ‘Watching Time’ of Youth:

Gender	1	2	3	4	5	Total
Female	9	32	19	6	3	69
Male	2	22	30	17	11	82
Total	11	54	49	23	14	151

Here, we have represented watch time as:

1 = ‘Less than an hour’

2 = ‘1-2 Hours’

3 = ‘2-3 Hours’

4 = ‘3-4 Hours’

5 = ‘More than 4 Hours’

Hypothesis:

H0: The ‘Watching time’ of programs is independent of ‘Gender’.

H1: The ‘Watching time’ of programs is dependent on ‘Gender’.

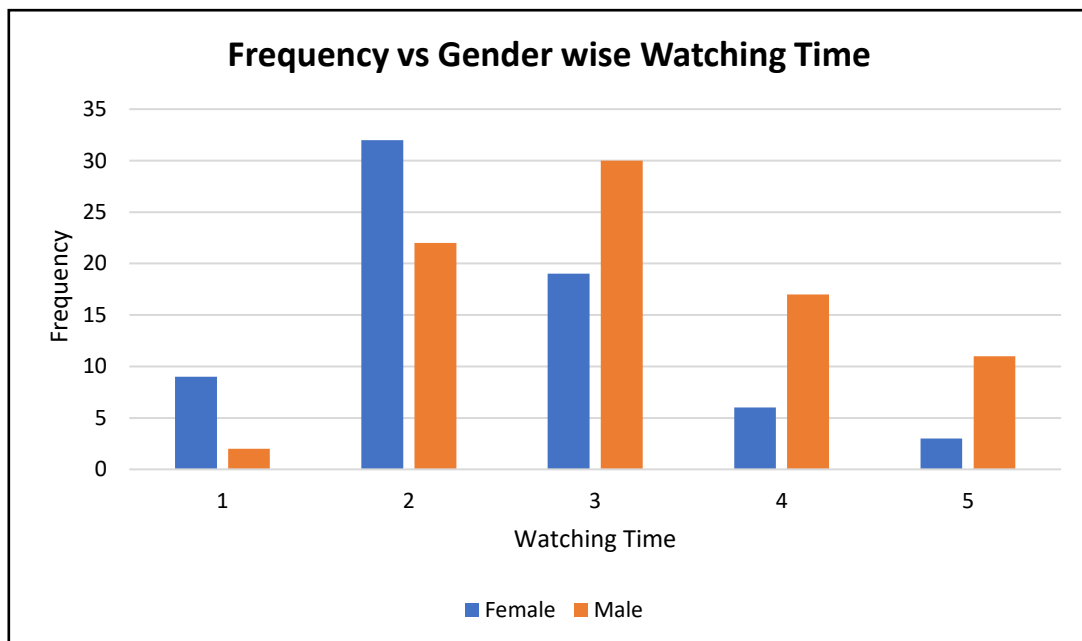
Applied Chi-Square Test in R:

```
> Data_N=data.frame(Gender=c('Female','Male'),'1'=c(9,2),'2'=c(32,22),'3'=c(19,30),
'4'=c(6,17),'5'=c(3,11))
> Data_N
  Gender x1 x2 x3 x4 x5
1 Female  9 32 19  6  3
2 Male   2 22 30 17 11
> Data_N1=data.matrix(Data_N)
> chisq.test(Data_N1)

Pearson's Chi-squared test

data:  Data_N1
X-squared = 17.816, df = 5, p-value = 0.003186
```

Here, p-value (0.003186) < alpha (0.05), also the observed Chi-Squared value (17.816) > Critical Chi-Squared value (11.07) for 5 degrees of freedom. So, we can reject the H₀. Therefore, there is significant relationship between the two variables 'Gender' and 'Time'. Hence, they are dependent variables.



Inference:

We can say that the two genders have different preferences for watching time of OTT platforms.

- Relationship between ‘Watching Time’ and ‘Depression Level’ of Youth:

Hypothesis:

H0: The mean depression levels for different watching times are equal.

H1: The mean depression levels for different watching times are not equal.

ANOVA Test with normality check in R:

```
> Data_GTD=data.frame(read.csv('Gender_Time_Depression.csv'))
> colnames(Data_GTD)=c('Gender','Time','Depression','Aggression')
> apply(Data_GTD,2,shapiro.test) # Normality test
$Gender
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.6337, p-value = 0.084
$Time
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.89387, p-value = 0.1092
$Depression
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.89722, p-value = 0.0954
$Aggression
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.87715, p-value = 0.071
> #ANOVA
> AOV_GTD=aov(Depression~Time,Data_GTD)
> summary(AOV_GTD)
              Df Sum Sq Mean Sq F value    Pr(>F)
Time              1  12.44   12.443     7.568 0.00668 **
Residuals       149 244.96    1.644
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, we checked the normality of our data using ‘Shapiro Test’ for normality. We can see that the p-values for each corresponding column are greater than Alpha (0.05) indicating that we can assume the normality.

For ANOVA, p-value (0.00668) < Alpha (0.05). So, we can reject H0. Therefore, at least two mean depression levels corresponding to watching time are different.

Inference:

We can say that the mean depression levels change with number of hours watched.

- Relationship between ‘Watching Time’ and ‘Aggression Level’ of Youth:

Hypothesis:

H0: The mean aggression levels for different watching times are equal.

H1: The mean aggression levels for different watching times are not equal.

ANOVA Test in R:

```
> # Gender time depression
> Data_GTD=data.frame(read.csv('Gender_Time_Depression.csv'))
> colnames(Data_GTD)=c('Gender','Time','Depression','Aggression')
> AOV_GTD=aov(Aggression~Time,Data_GTD)
> summary(AOV_GTD) # ANOVA
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Time	1	17.15	17.147	21.91	6.38e-06 ***
Residuals	149	116.63	0.783		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Here, the data is normal (with reference to the Shapiro Test performed in the previous question).

For ANOVA, p-value (6.38e-06) < Alpha (0.05). So, we can reject H0. Therefore, at least two mean aggression levels corresponding to watching time are different.

Inference:

We can say that the mean aggression levels change with number of hours watched.

- Relation between ‘Gender’ and ‘Genres watched’ of Youth:

Gender	Adventure	Thriller/Horror	Crime	Romance	Sci-Fi/Fantasy	Drama	Comedy
Female	26	22	5	34	23	14	27
Male	38	32	20	21	24	5	11
Grand Total	64	54	25	55	47	19	38

Hypothesis:

H0: The preference of ‘Genre watched’ is independent of ‘Gender’.

H1: The preference of ‘Genre watched’ is dependent on ‘Gender’.

Applied Chi-Square Test in R:

```
> Data_GG=data.frame(Gender=c('Female','Male'),Adventure=c(26,38),Thriller=c(22,32),
+ Crime=c(5,20),Romance=c(34,21),Sci_Fi=c(23,24),Drama=c(14,5),Comedy=c(27,11))
> Data_GG
  Gender Adventure Thriller Crime Romance Sci_Fi Drama Comedy
1 Female      26      22      5      34      23      14      27
2  Male      38      32     20      21      24       5      11
> Data_GG1=data.matrix(Data_GG)
> chisq.test(Data_GG1)
```

Pearson's Chi-squared test

```
data: Data_GG1
X-squared = 27.526, df = 7, p-value = 0.0002679
```

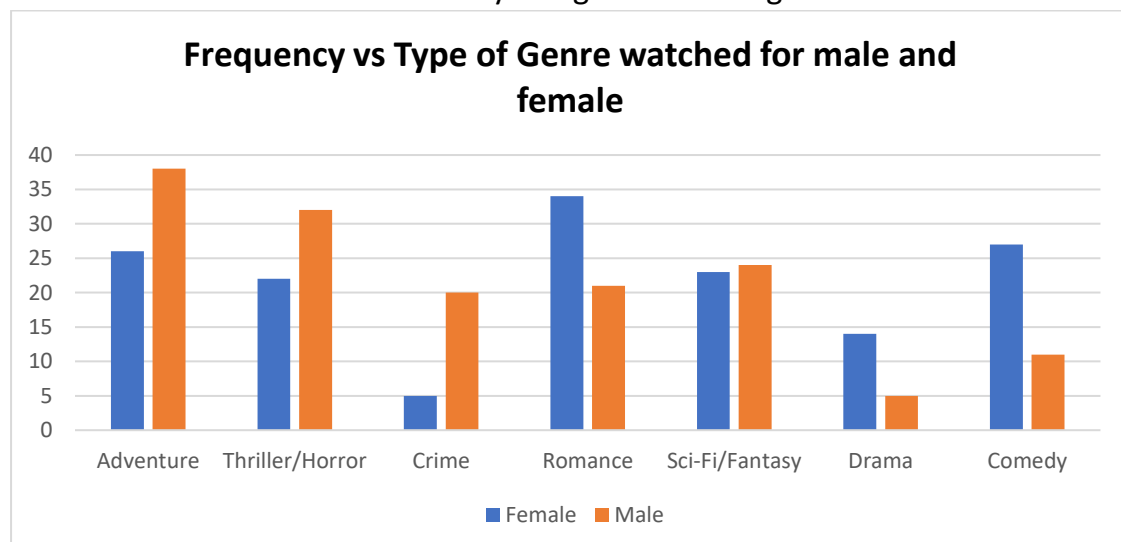
```
Warning message:
In chisq.test(Data_GG1) : Chi-squared approximation may be incorrect
```

Here, $p\text{-value} < (0.0002679) < \alpha (0.05)$, also the observed Chi-Squared value (27.526) > Critical Chi-Squared value (14.067) for 7 degrees of freedom. So, we can reject the H0. Therefore, there is significant relationship between the two variables ‘Gender’ and ‘Genre watched’. Hence, they are dependent variables.

Inference:

We can say that the two genders prefer to watch different genres on OTT platforms.

This result can also be visualized by using the following bar chart:



➤ Analysis of Question 2:

- Correlation between ‘Effects on exam and other important work’ and ‘Watching time’ of Youth:

Hypothesis:

H0: Correlation between the two variables is 0.

H1: Correlation between the two variables is not 0.

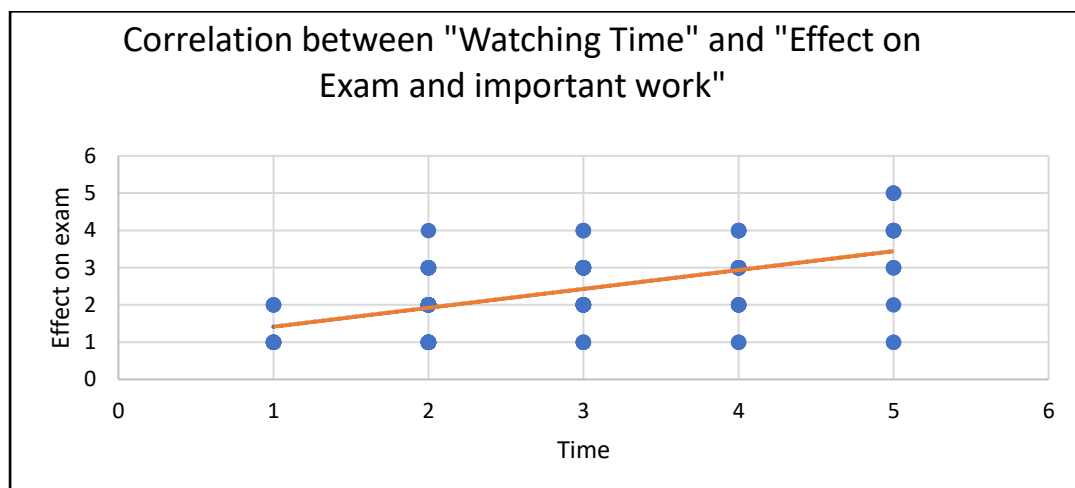
Applied Correlation Test in R:

```
> colnames(Data_GE2)=c('Time','Exam')
> cor.test(Data_GE2$Time,Data_GE2$Exam)

Pearson's product-moment correlation

data:  Data_GE2$Time and Data_GE2$Exam
t = 8.4032, df = 149, p-value = 3.19e-14
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4478808 0.6664128
sample estimates:
cor
0.5670434
```

Here, p-value ($3.19e-14$) < alpha (0.05) and Correlation coefficient (r) = 0.567. So, we can reject H0. Therefore, there is correlation between the two variables.



Also, through visualization from this scatter plot and the value of $r=0.567$, we can tell that there is positive correlation between these variables.

Inference:

We can say that as the watching time of online content increases, it is more likely to affect the exams or other important work of youth, thus affecting their working and learning efficiency.

- Relation between ‘Effects on exam and other important work’ and ‘Gender’ of Youth:

Hypothesis:

H0: The view of ‘Effect on Exam’ is independent of ‘Gender’.

H1: The view of ‘Effect on Exam’ is dependent on ‘Gender’.

Applied Chi-Square Test in R:

```
> Data_GE=data.frame(Gender=c('Female','Male'),'1'=c(16,13),'2'=c(31,30),'3'=c(16,28),
'4'=c(5,9),'5'=c(1,2))
> Data_GE
  Gender x1 x2 x3 x4 x5
1 Female 16 31 16  5  1
2  Male 13 30 28  9  2
> Data_GE1=data.matrix(Data_GE)
> chisq.test(Data_GE1)

Pearson's Chi-squared test

data:  Data_GE1
X-squared = 4.1707, df = 5, p-value = 0.5251

Warning message:
In chisq.test(Data_GE1) : Chi-squared approximation may be incorrect
```

Here, p-value (0.5251) > alpha (0.05), also the observed Chi-Squared value (4.1707) > Critical Chi-Squared value (11.07) for 5 degrees of freedom. So, we cannot reject the H0. Therefore, there is no significant relationship between the two variables ‘Gender’ and ‘Effect on exam’. Hence, they are independent variables.

Inference:

We can say that watching online content can affect any person equally irrespective of their gender. Meaning that males and females can show similar effects of watching online content depending on only the number of hours they spend watching (relevance from previous ‘Time’ and ‘Effect on Exam’ inference).

➤ Analysis of Question 3:

- Relation between ‘Physical Activities’ and ‘Watching time’ of Youth:

Hypothesis:

H0: The mean lack in physical activities for different watching times are equal.

H1: The mean lack in physical activities for different watching times are not equal.

Normality Check of Data:

```
> apply(Data_Si,2,shapiro.test) # Normality test
$Time
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.89387, p-value = 5.54e-09
$Physical_Activity
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.79349, p-value = 2.544e-13
$Friends_Family
      Shapiro-wilk normality test
data:  newX[, i]
W = 0.73596, p-value = 3.618e-15
```

As p-value < Alpha(0.05), we can say that the data is not Normal.

Performing Non-Parametric Test (Kruskal-Wallis) in R:

```
> # Non-Parametric Test:
> kruskal.test(Physical_Activity~Time,Data_Si)

      Kruskal-wallis rank sum test
data:  Physical_Activity by Time
Kruskal-wallis chi-squared = 2.7759, df = 4, p-value = 0.596
```

Here, p-value (0.596) > Alpha (0.05) for Physical Activity w.r.t Time. So, we cannot reject H0. Therefore, the mean lack of physical activities for different watching time are not equal.

Inference:

We can say, we have enough evidence that Physical activity is not related to the Watching Time. Which means that avoiding physical activities while watching OTT shows is not dependent on the watching time.

- Relation between ‘Interaction with friends and family’ and ‘Watching time’ of Youth:

Hypothesis:

H0: The mean lack in Interaction with friends and family for different watching times are equal.

H1: The mean lack in Interaction with friends and family for different watching times are not equal.

Here, the data is normal (with reference to the Shapiro Test performed in the previous question).

Performing Non-Parametric Test (Kruskal-Wallis) in R:

```
> # Non-Parametric Test:  
> kruskal.test(Friends_Family~Time,Data_Si)  
  
Kruskal-wallis rank sum test  
  
data: Friends_Family by Time  
Kruskal-wallis chi-squared = 2.2257, df = 4, p-value = 0.6943
```

Here, p-value (0.6943) > Alpha (0.05) for Friends_Family w.r.t Time. So, we cannot reject H0. Therefore, there is no significant relationship between these variables. Therefore, the mean lack of interaction with friends and family for different watching time are not equal.

Inference:

We can say, we have enough evidence that the factor of interaction with friends and family is not related to watching time of online content. Which means that avoiding friends and family while watching OTT shows is not dependent on the watching time.

- Relation between ‘Watching Time’ and ‘Data connection type’ of Youth:

	Time					
Connection Type	1	2	3	4	5	Total
Mobile Data	8	32	23	6	4	72
Broadband/WiFi	3	22	26	17	10	79
Grand Total	11	54	49	23	14	151

Hypothesis:

H0: The variables ‘Watching Time’ and ‘Data connection type’ are independent.

H1: The variables ‘Watching Time’ and ‘Data connection type’ are dependent.

Applied Chi-Square Test in R:

```
> Data_Si1=data.frame(Gender=c('Mobile Data','Broadband/wiFi'),'1'=c(8,3),
'2'=c(32,22),'3'=c(23,26),'4'=c(6,17),'5'=c(4,10))
> Data_Si1
      Gender x1 x2 x3 x4 x5
1  Mobile Data 8 32 23  6  4
2 Broadband/wiFi 3 22 26 17 10
> Data_Si2=data.matrix(Data_Si1)
> chisq.test(Data_Si2)

      Pearson's Chi-squared test

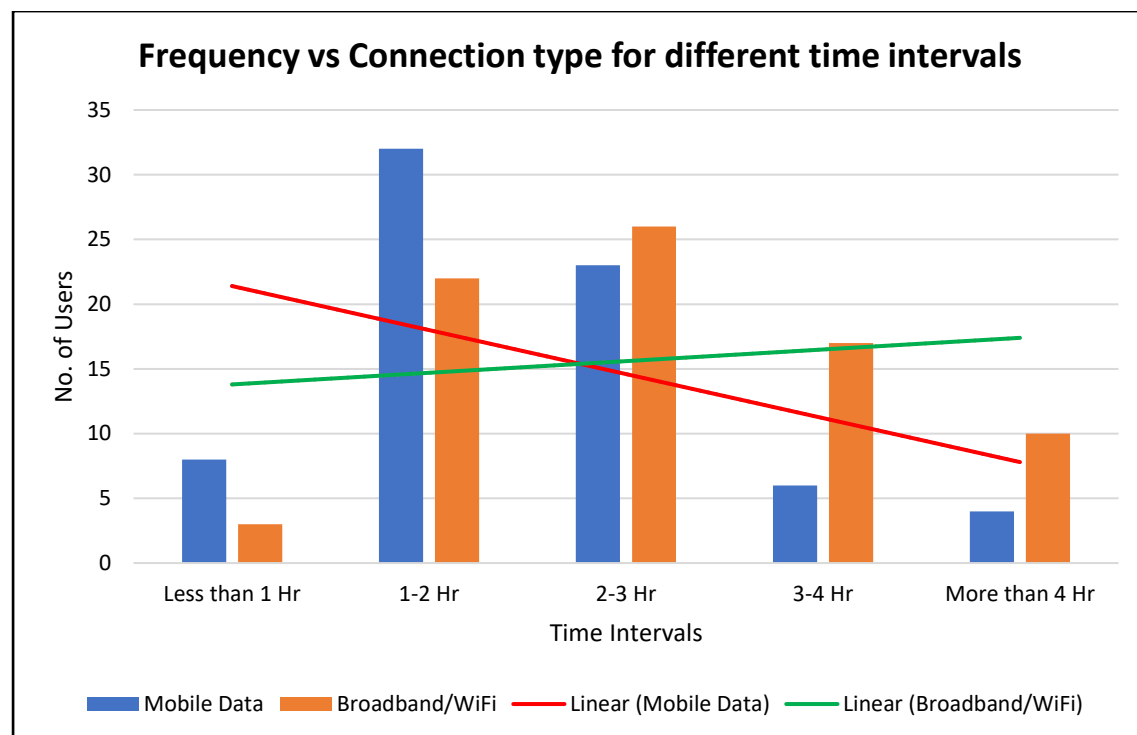
data:  Data_Si2
X-squared = 12.378, df = 5, p-value = 0.02996

Warning message:
In chisq.test(Data_Si2) : Chi-squared approximation may be incorrect

>
```

Here, p-value (0.02996) < alpha (0.05), also the observed Chi-Squared value (12.378) > Critical Chi-Squared value (11.07) for 5 degrees of freedom. So, we can reject the H0.

Therefore, there is significant relationship between the two variables ‘Watching time’ and ‘Data connection type’. Hence, they are dependent variables.



Inference:

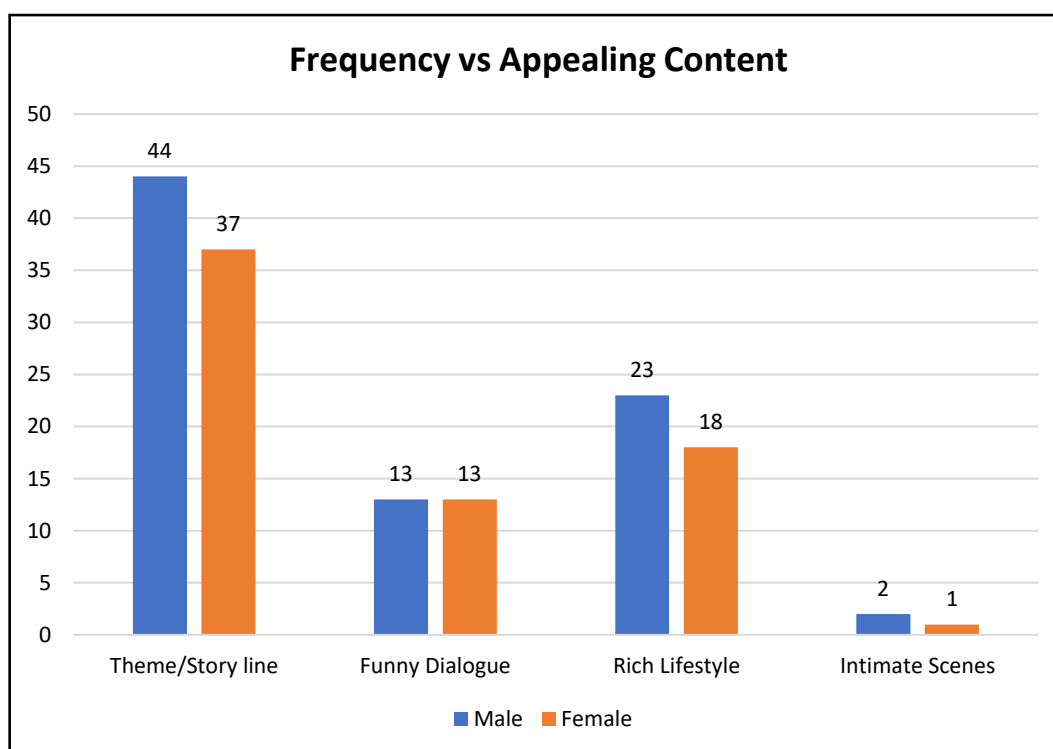
We can say that the people with different data connection types (namely mobile data connection and Broadband/Wi-Fi connection) have a difference in their watching pattern.

Also, we can clearly see the trend lines in the above bar chart. Here, the trend for Mobile data connection is decreasing indicating that people using mobile data have higher frequencies in the lower time intervals and for Broadband/Wi-Fi users the trend is increasing, indicating that these users tend to watch online content for more time and hence, have higher frequency in the higher time intervals.

➤ Analysis of Question 4:

- Relation between ‘Gender’ and ‘Appealing Content’ Of Youth:

Gender	Theme/Story line	Funny Dialogue	Rich Lifestyle	Intimate Scenes	Total
Male	44	13	23	2	82
Female	37	13	18	1	69
Grand Total	81	26	41	3	151



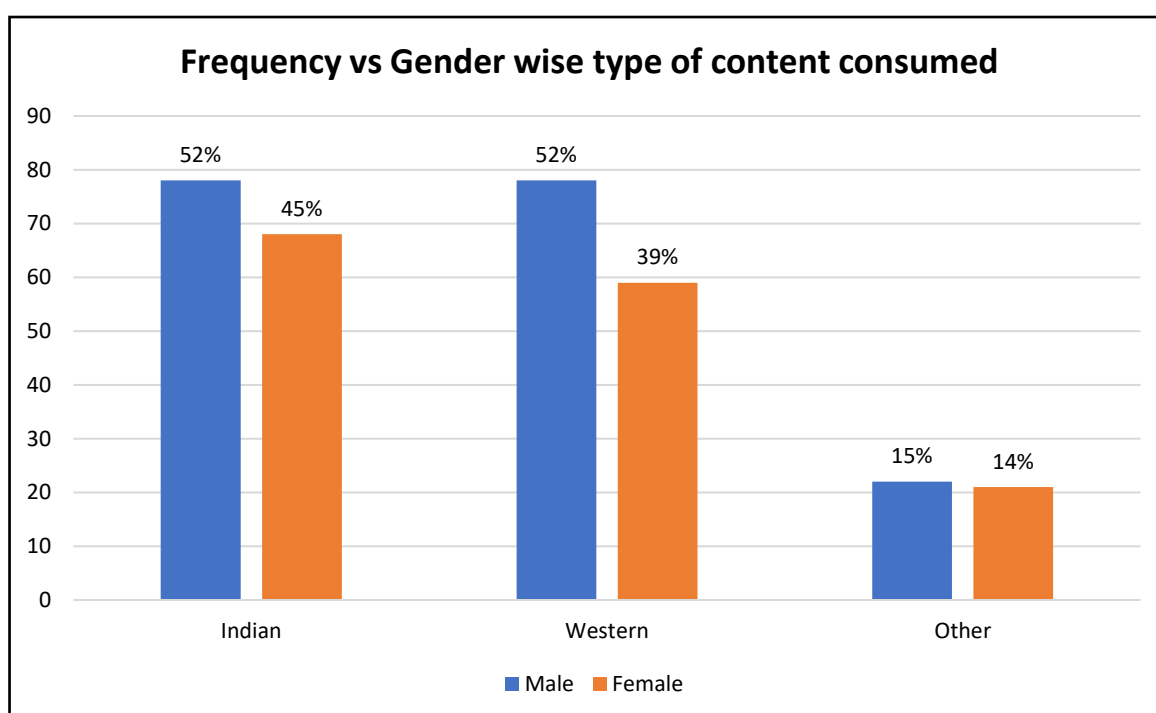
Inference:

Theme/Story Line (54%) are the most preferred content in a movie/web-series that the users watch of which males count 44(29%) females are 37(25%).

The second preference is Rich Lifestyle (27%), followed by Funny Dialogues (17%) and lastly Intimate Scenes (2%).

- Relation between ‘Gender’ and ‘Types of content consumed’:

Types	Male	Female
Indian	78	68
Western	78	59
Other	22	21

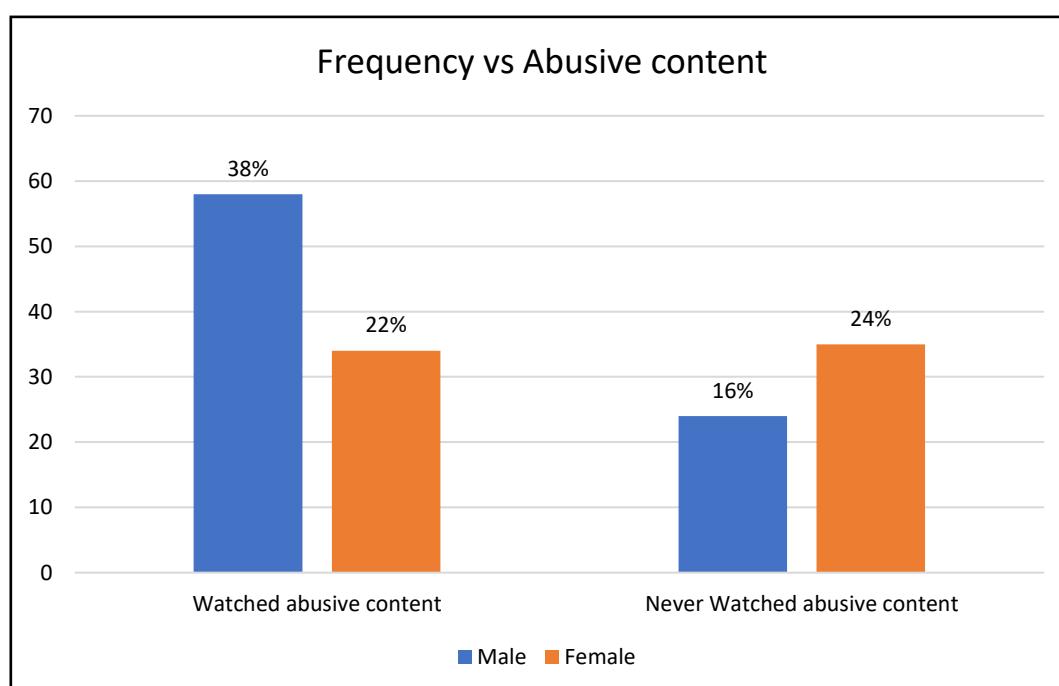


Inference:

From this graph we can say that the respondents prefer Indian and Western content more than Other (Anime, K-Drama, Arabic, etc.)

- Relation between ‘Gender’ and ‘Watching abusive content’:

Gender	Watched abusive content	Never Watched abusive content	Total
Male	58	24	82
Female	34	35	69
Grand Total	92	59	151



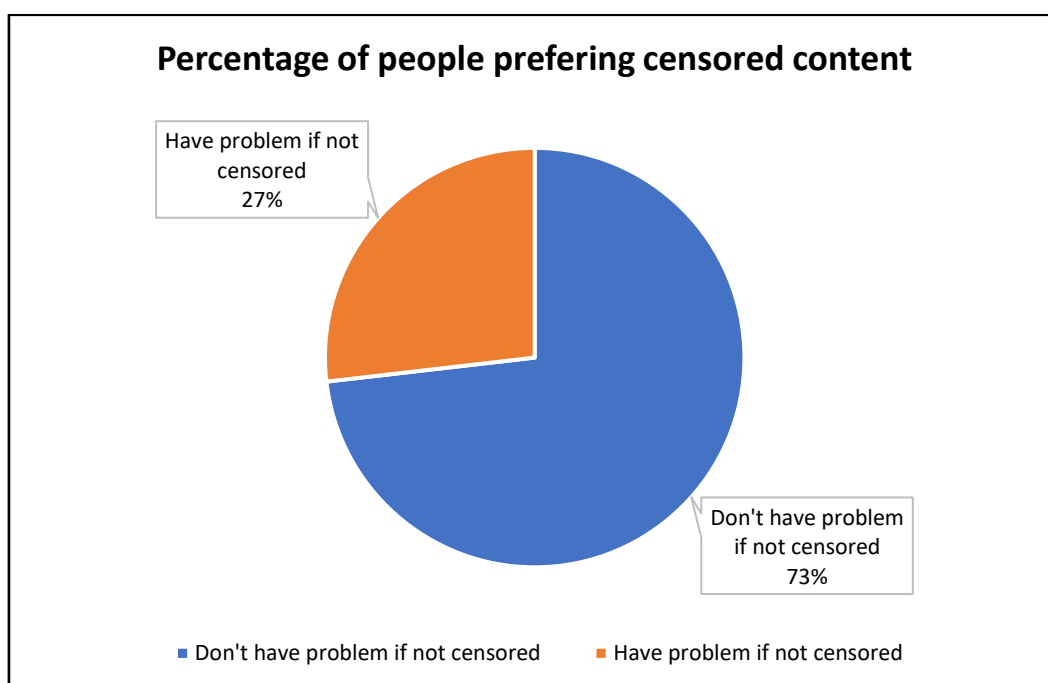
Inference:

We can conclude that a higher proportion of Indian Youth watch abusive content (60%) in which 38% are males and 22% are females.

And the remaining 40% of respondents 16% males and 24% females tend to avoid watching abusive content.

- Relation between ‘Gender’ and ‘Viewing censored content’:

Gender	Don't have problem if not censored	Have problem if not censored	
Male	60	22	82
Female	41	28	69
Grand Total	101	50	151

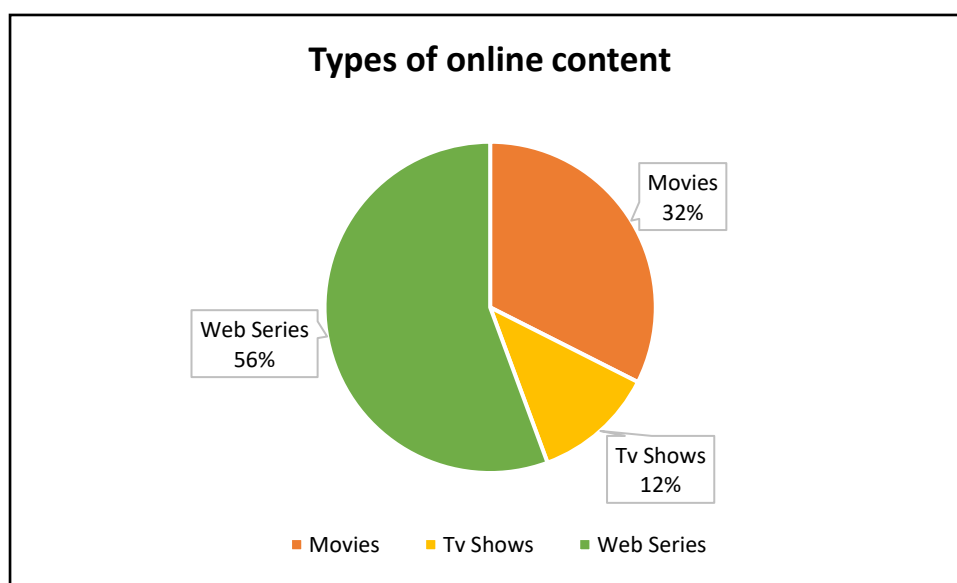


Inference:

73% of the respondents do not have problem with uncensored content on OTT Platforms, whereas 27% of individuals have some problem.

- What today's youth like more:

Movies	TV Shows	Web Series	Grand Total
49	18	84	151



Inference:

More than half of the youth (56%) like web-series, 32% people like Movies and 12% people like TV Show.

This could be because the new rapidly emerging trend and fondness for web-series since 2016 which have a variety of genres and are appreciated by the audience from different age groups.

Conclusion

After analysing several aspects, it can be clearly stated that web series and online streaming content have a huge impact on the Indian youth.

This new media platform is growing rapidly and is gaining momentum as the day passes. The content being produced and showcased on the online platforms have been successful in grabbing youth's attention, and moving them away from the traditional television soap operas.

The new generation has moved on from the traditional broadcasting system to the mobile gadgets. The emergence of Reliance JIO and cheaper data with uninterrupted 4G services has helped the OTT platforms grow immensely.

The content showcased on OTT platforms filled with sexual, abusive and violent content together with alcohol and drugs have caused psychological effects on the Indian youth, where they have agreed to suffer from insomnia, depression and insecurities in their life. The youth is also witnessing academic loss and are also getting prone to health issues. On the other side, the preference to binge watching is also affecting their relations with friends and family. The Data which was collected through questionnaires via digital media was analysed in MS Excel, RStudio and Python.

Analysis of the OTT platforms was done to find out the Strengths, Weakness, Opportunity and Threats. The acceptance of the platforms was remarkable signifying a greater change over in the near future.

Video streaming has become one of the most successful avenues in the content consumption space in India. Even the smaller OTT platforms are raising capital from international investors and making a significant impact on the market.

Findings:

1. The study found that YouTube is the most popular online video streaming platform amongst the respondents, followed by Netflix, Amazon Prime, Hotstar, MX Player and others.
2. The respondents agreed that web series and online video content have direct or indirect psychological impact on them and they also feel insecurities in their life as well.

3. The respondents agreed to have watched any specific web series or online content just because of intimate scenes or abusive language used in the content, where nearly half of them agreed that they had its effect on their spoken language.
4. A majority of the respondents agreed that binge watching has given a rise to insomnia and depression amongst them, and a good percentage also agreed that watching web series and online video content have affected their academic performance as well.
5. All the psychological effects like depression, insomnia, feeling insecure and negative effects on exams were directly related to the amount of time the respondents spent on watching OTT content.
6. Both male and female respondents had equal chance of having all the above-mentioned psychological effects as they are gender independent.
7. People having a broadband/WiFi connection tend to stream OTT platforms for a greater number of hours than people with mobile data.
8. A high number of respondents showed interest in Western content along with Indian content which could be enough to explain the growing fondness for western culture in Indian youth.
9. More than 70% of the youth (both male and female) do not have any problem whether or not the content they watch is censored.

Future scope of study

India, the fastest growing **OTT** market.

According to a recent PwC report, **India** is currently the world's fastest growing **OTT** market, set to become the sixth largest around the globe by 2024.

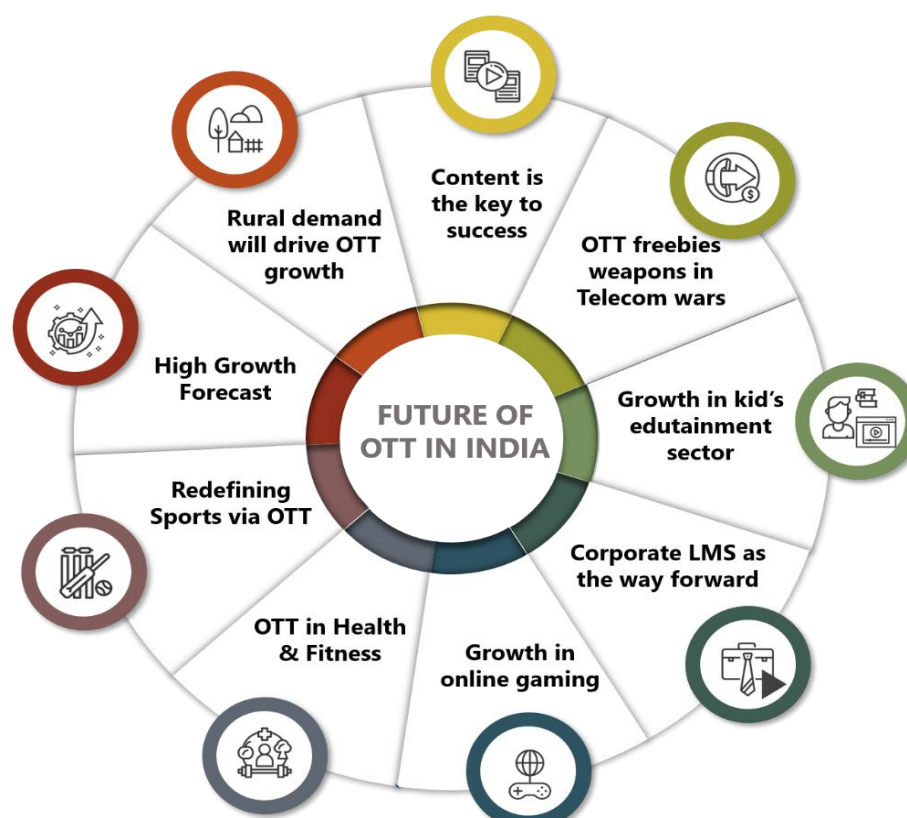
Industry trends indicate that with access to better networks, digital connectivity and smartphones, OTT platforms in India have been increasingly attracting subscribers on a daily basis. Apart from top favourites Disney+ Hotstar, Amazon Prime Video and Netflix, the space is seeing a plethora of local and regional OTT players, such as SonyLIV, Voot, Zee5, ALTBalaji, JioTV etc

The next decade in this space will very possibly belong to not only those with the best original content, but also the most advanced technology, analytics and user-friendly features. Ultimately, the streaming wars will empower viewers and revolutionise how they consume content.

Future scope of study on this topic could be:

1. Age-wise study of factors like Aggression level, preferred Genres or types of programs the youth like to watch, etc.
2. Study of online gaming addiction.
3. Analysing the growing online education culture on various platforms along with documentaries and other knowledgeable content on OTT Platforms.

Etc.



References

- 1) Savita Shah & V.K. Satya Prasad's (2021), Consumption Pattern of OTT Platforms in India
Link: <http://www.modern-journals.com/index.php/ijma/article/view/780/676>
- 2) Chun-Mei-Chen (2019), Evaluating the efficiency change and productivity progress of the top Global telecom operators since OTT prevalence.
- 3) CRITICAL REVIEW OF BINGE-WATCHING BEHAVIOUR THROUGH THE PRISM OF VROOM'S EXPECTANCY THEORY
Link: https://www.researchgate.net/publication/342170874_CRITICAL_REVIEW_OF_BINGE_WATCHING_BEHAVIOUR_THROUGH_THE_PRISM_OF_VROOM'S_EXPECTANCY_THEORY
- 4) Link: https://images.assettype.com/afaqs/2020-05/e249842c-880c-4048-acdf-34a0eb514b05/OTT_AudienceSegments_OrmaxMedia_1_.pdf
- 5) Link: <https://www.statista.com/statistics/1132381/india-preferred-online-media-platforms-by-genre/>