INDEX

Study on OTT Platforms

1. Introduction

Did you hear about Netflix or Hotstar, or even better, ever watched something in those? Figure 1 shows the number of subscribers of Netflix and Disney+Hotstar over the years...

Well, these along with many others like Amazon Prime Video, MX Player, Hoichoi and so on are called OTT platforms.

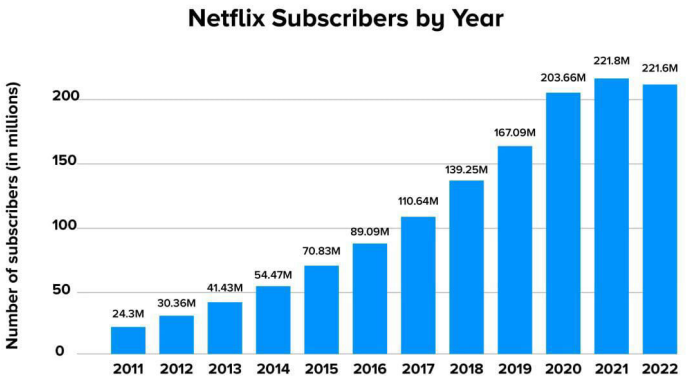
OTT stands for over-the-top as the contents are provided directly to viewers with Internet connection, with or without subscriptions bypassing telecommunications, broadcast television platforms and so on which are usually used as a distributor of such content.

Viewers have multiple options for content and genre to watch in OTT platforms which have added to the increasing popularity of the same. They just need an Internet connection and maybe a subscription to watch in OTT platforms irrespective of time and place.

But all these have had a huge impact on cinema halls. Nowadays many movies are either being released only in OTT platforms or just after a few days of their release in cinema halls they are being released in the OTT platforms. Also, apart from movies OTT platforms provide us with numerous other contents like series and documentaries.

So, our study is about learning the influence of a number of factors on the OTT platforms, the variety of contents and genres those offer, impact of OTTs on cinema halls and many more.

Figure 1: No. of subscribers over the years



2. Research Questions

The questionnaire was designed based on the following questions:

1. What are the factors influencing the use of OTT platforms?
2. What are the popular OTT platforms, genres and content preferences over them?
3. What is the average daily time spent on OTT platforms?
4. What is the impact of OTTs on cinema halls?

3. Data Collection

Let us have a look at the various steps of our data collection process.

3.1. Plan of Enquiry

Regarding our first research question about the factors influencing the use of OTT platforms, we asked people about their age to have a fair idea about the age group which mostly watches in OTT platforms, gender to know about the gender among which OTT platforms are most popular , pin code to know whether people from the rural areas are also interested in watching in OTT platforms as much as urban people are. We also asked about profession and total annual income of the family to have a fair enough idea about which profession has the most people watching in OTT platforms and whether they can afford the subscriptions of the same.

For our second research question about the most popular OTT platforms and the genre preferences, we asked about the OTT platforms in which they mostly watch, what they mostly watch in those and also the genres they prefer to know which ones are the most popular. We also questioned about the time spent on internet daily, and further the time spent on OTT platforms everyday on an average. We also wanted to know if people have subscription for the OTT platforms and if yes, which are the ones they have subscription for to cross check whether they really have the subscription or they watch on third party apps or websites and also the annual expenditure on the OTT platforms. We also questioned people about whether they watch in OTTs in their regional languages to see whether those are popular among people or not.

Lastly, for our research questions regarding the impact of OTT platforms on cinema halls we asked people whether they would prefer OTT platforms to cinema halls and the number of times they visited the cinema halls in the last three months.

3.2. Target Population and Sampling Scheme

Target population of this study is all the people aged more than 15 years. Here, we have used Simple Random Sampling scheme, i.e. the sample is selected from the target population randomly. Now to collect the data, we circulated the questionnaire among our family, friends, other acquaintances and requested them to fill it and further circulate it among people they know.

3.3. Design of Questionnaire

Our final questionnaire was not created at one go. We made questionnaires with flaws, corrected those and could at last decide on a final questionnaire. Let us share with you some of the initial questionnaires that we had made.

3.3.1. First Questionnaire

QUESTIONNAIRE (OTT PLATFORM)

1. What is your age?

   _____

2. What is your gender?
   a) Male
   b) Female
   c) Others
3. What is your profession?

   _____

4. What is the total annual income of your family?
   a) <50,000
   b) 50,000-2 lakhs
   c) 2 lakhs-6 lakhs
   d) More than 6 lakhs
5. Which OTT platforms content do you usually watch in?
   a) Netflix
   b) Amazon Prime
   c) Mx player
   d) Disney + Hotstar
   e) Hoichoi
   f) Others_____
6. Do you have a subscription for the OTT platform?
   a) Yes
   b) No
7. Mention the OTT platform which you have subscribed?

   _____

8. What is the subscription plan? (Duration with price)

   _____

9. What do you watch in OTT platform most often?
   a) Movies
   b) Series
   c) Sports
   d) Documentaries
   e) Others_____
10. What is the genre that you prefer the most?
    a) Crime & Mystery
    b) Action

c) Romance
d) Horror
e) Others _____

11. Time spent on OTT platform on a daily basis?

a) Less than 2 hours
b) 2-4 hours
c) 4-6 hours
d) More than 6 hours

12. If a movie is released in cinema hall and is scheduled to be streamed in the OTT platforms in a few week time, where will you prefer to watch it?

a) Movie hall
b) OTT platform

13. How frequently do you visit the movie halls in 3 months?

_____

In this questionnaire, we had not included any question on pincode which would have stopped us from studying about the way areas affect the views of OTT platforms. Also, the question about subscription plan would have had more continuity in the answers if we could change it to annual expenditure on the subscriptions. Along with the time spent on OTT platforms on a daily basis, we also needed to know the time of daily usage of Internet to have a better idea about the matter. Further, to have more quantitative variables, we needed to remove the options from the question related to daily average watch time of the OTTs and let people fill it up on their own. Also we wanted to cross check whether people are legally watching in the OTT platforms or are watching using third party apps and websites. All these flaws were removed and the 2nd questionnaire was created.

3.3.2. Second Questionnaire

QUESTIONNAIRE (OTT PLATFORM)

1. What is your age?

   _____

2. What is your gender?
   a) Male
   b) Female
   c) Others
3. What is your PIN code?

   _____

4. What is your profession?

   _____

5. What is the total annual income of your family?
   a) Less than 50,000
   b) 50,000 to 2 lakhs
   c) 2 lakhs to 6 lakhs
   d) More than 6 lakhs
6. Do you have a subscription for OTT platform?
   a) Yes
   b) No
7. Which OTT platforms do you have subscription for?
   a) Netflix
   b) Amazon Prime
   c) Mx player
   d) Disney + Hotstar
   e) Hoichoi
   f) Others_____
8. Which OTT platforms do you usually watch in?
   a) Netflix
   b) Amazon Prime
   c) Mx player
   d) Disney + Hotstar
   e) Hoichoi
   f) Others_____
9. What is your annual expenditure on OTT platforms?

   _____

10. What do you watch in OTT platform most often?
    a) Movies

b) Series
c) Sports
d) Documentaries
e) Others_____

11. What is the genre that you prefer the most on OTTs?

a) Crime & Mystery
b) Action
c) Sports
d) Romance
e) Horror
f) Others _____

12. How many hours do you spend on an average in OTT platforms daily?

_____

13. How many hours do you use internet daily?

_____

14. If a movie is released in cinema halls and is scheduled to be streamed in OTT platforms in a few weeks, where will you prefer to watch it?

a) Cinema hall
b) OTT platform

15. How many times did you visit cinema halls in the last three months?

_____

To cross check whether people are watching legally in the OTT platforms or watching using third party apps or websites we added a question in the second questionnaire asking about the OTT platforms the people have subscription for. Also, the question numbers 6 to 8 needed to be rearranged among themselves so that people don't feel as if they are answering the same question again. We also added a question about watching OTTs in regional languages in the final questionnaire. In this way after correcting all these flaws we got our final questionnaire.

3.3.3. Final Questionnaire

A Study Related to OTT Platforms

Provided you are aware about OTT platforms, please help us with your valuable and **honest** response through these few questions.

* Required

1. Age *

_____

2. Gender *

Mark only one oval.

- ◯ Male
- ◯ Female
- ◯ Others

3. PIN code *

(PIN code of your present address)

_____

4. Occupation *

Mark only one oval.

- ◯ Student
- ◯ Educator
- ◯ Doctor
- ◯ Homemaker
- ◯ Corporate worker
- ◯ Business
- ◯ Other: _____

5. Annual family income *

Mark only one oval.

- ◯ Less than 50,000
- ◯ 50,000 to 2 lakhs
- ◯ 2 lakhs to 6 lakhs
- ◯ more than 6 lakhs

6. Which OTT platforms do you usually watch in? *

Check all that apply.

- ☐ Netflix
- ☐ Amazon Prime
- ☐ MX Player
- ☐ Disney + Hotstar
- ☐ Other: _____

7. Do you have subscription for OTT platforms? *

Mark only one oval.

- ◯ Yes
- ◯ No

8. Which OTT platforms do you have subscription for? *
(Simply write NA in other option when have no OTT subscription)

Check all that apply.

- ☐ Netflix
- ☐ Amazon Prime
- ☐ MX player
- ☐ Disney + Hotstar
- ☐ Other: _____

9. Do you watch OTTs in your regional language? *

Mark only one oval.

- ◯ Yes
- ◯ No

10. What is your annual expenditure on OTT platforms? *
(Simply write NA when have no OTT subscription)

_____

11. What do you watch in OTT platform most often? *

Check all that apply.

- ☐ Movies
- ☐ Series
- ☐ Sports
- ☐ Documentaries
- ☐ Other: _____

12. What is the genre that you prefer the most on OTTs? *

Check all that apply.

- ☐ Thriller, Crime & Mystery
- ☐ Action
- ☐ Sports
- ☐ Romance
- ☐ Horror
- ☐ Other: _____

13. How many hours do you spend on an average in OTT platforms daily? *

_____

14. How many hours do you use internet daily? *

15. If a movie is released in cinema halls and is scheduled to be streamed in OTT platforms in a few weeks, where will you prefer to watch it? *

Mark only one oval.

- Cinema hall
- OTT platform

16. How many times did you visit cinema halls in the last three months? *

3.4. Follow-up of the Process

The final questionnaire was circulated among people, and they were requested to fill it up and to further circulate it. Although some carried out our request in the very first time we had requested , some had to be requested again before they filled it up. We also shared it in social media platforms with a request to fill it up and further share it so that our questionnaire could reach more people, than us personally sending it to them.

4. Scrutiny of the Data

After collecting the answers from people, we first scrutinized the data and removed those which seemed to be incorrect or absurd entry of data (like age is 209 years, or time spent on OTT platforms daily is 56 hrs) and modified those where people appeared to have made some mistakes while answering the questions (like maybe a zero less in the pincode). Then, we sorted the given pincodes into urban and rural areas. In case, someone gave a range for the questions related to time, that is time spent on OTT platforms on a daily basis or number of hours of daily Internet usage, we took the average. We even removed those answers which said that they don't watch in OTT platforms. Further, depending on the given annual expenditure on OTT platforms, we also removed those data which seemed to be not in accordance with the given annual income of the family. We removed those data

which had more daily watch time in OTT platforms than the time of Internet usage and also those data which although having zero annual expenditure, still had given yes as an answer to the question regarding whether they had a subscription.

5. Variable Preprocessing

We had allowed multiple answers for the OTT platforms in which the people usually watch, the ones they have subscription for, the contents they usually watch and the genres they prefer. We calculated the total numbers for each option individually for all these questions.

Thus, by making similar suitable modifications to the data, we finally got the final data for our study.

6. Exploratory Data Analysis

Firstly let us calculate summary measures of the variables.

```
##  Subscription Region.lang Annual.expend        OTTtime         Watching.status
##  No :100      No : 84     Min.   :    0.0   Min.   :0.000   Cinema hall: 66
##  Yes:100      Yes:116     1st Qu.:    0.0   1st Qu.:1.000   OTT        :134
##                           Median :   74.5   Median :2.000
##                           Mean   : 1032.0   Mean   :1.814
##                           3rd Qu.: 1500.0   3rd Qu.:2.500
##                           Max.   :10000.0   Max.   :7.000
##    Hall.visit
##  Min.   : 0.00
##  1st Qu.: 0.00
##  Median : 1.00
##  Mean   : 1.65
##  3rd Qu.: 2.00
##  Max.   :13.00
```

From these summary measures we observed that, the number of people having subscription for the OTT platforms is 100, while those who do not is 100.

The number of people watching OTTs in regional language is 116, while those who do not is 84.

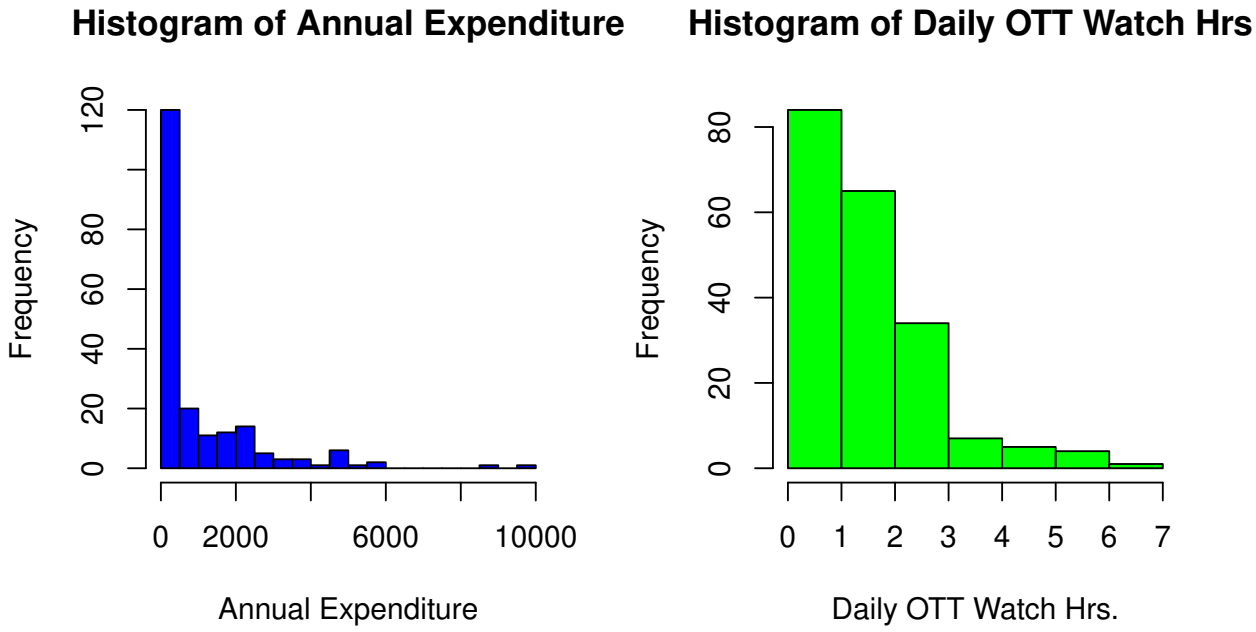
In case of annual expenditure on the subscription of OTT platforms, we found out that the maximumum expenditure was Rs 10000 and the mean expenditure was Rs 1032. The median was found out to be Rs 74.5, meaning that 50% of the people who subscribe OTT have annual expenditure less than or equal to Rs 74.5 and the third quantile came out to be Rs 1500, implying that 75% of the people who subscribe OTT have annual expenditure less than or equal to Rs 1500.

The maximum number of hours someone spends in OTT platforms is 7 hrs, while the mean is 1.814 hrs. Since the median came out to be 2, it means that 50% of the people having watch time less than or equal to 2 hrs. The third quantile came out to be 2.5 implying that 75% of the people having watch time less than or equal to 2.5 hrs.

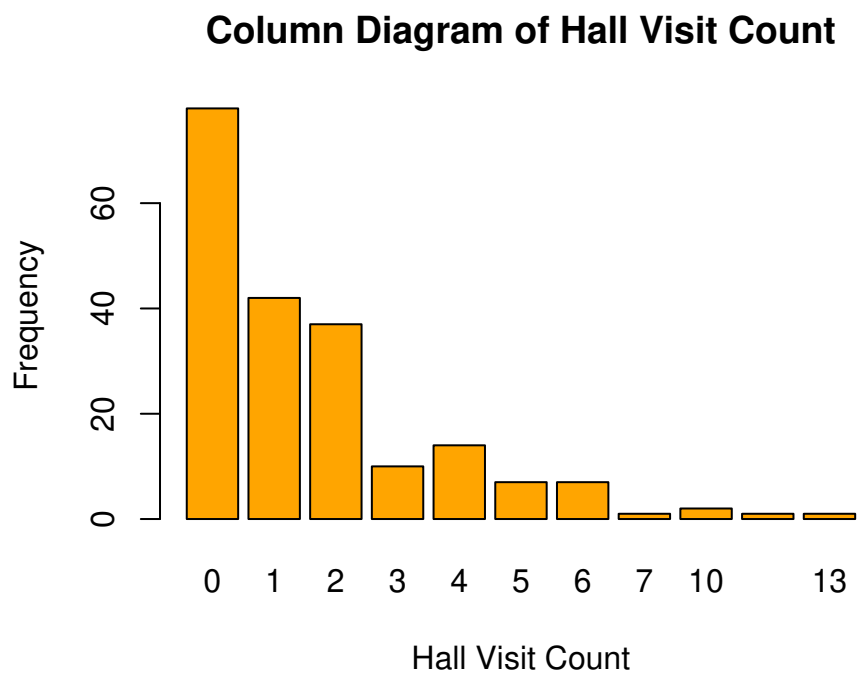
The number of people who prefer watching newly released movies in cinema halls than in OTT platforms is 66, while that of the opposite is 134.

The maximum number of times someone visited the cinema halls in the last 3 months is 13. The median is 1, implying that 50% of the people visited cinema hall less than or equal to 1 time and the third quantile came out to be 2, meaning that 75% of the the people visited cinema hall less than or equal to 2 times. The mean came out to be 1.65.

In our data set age, annual expenditure in OTT, daily OTT wach hour, daily internet use hour and hall visit frequency are numerical variable. But histogram will be meaningful only for annual expenditure in OTT and daily OTT watch hour.



Here the frequency of annual expenditure between Rs.0-Rs.500 is the maximum. That means most of the user spend less than Rs.500 annually in OTT. From the second histogram we can see that most of the consumers watch in OTTs for less than one hour daily and a significant number of users spend 1-3 hours daily in OTTs.



We can observe that most of the people have not visited cinema halls in last three months and few people visited one or two times.

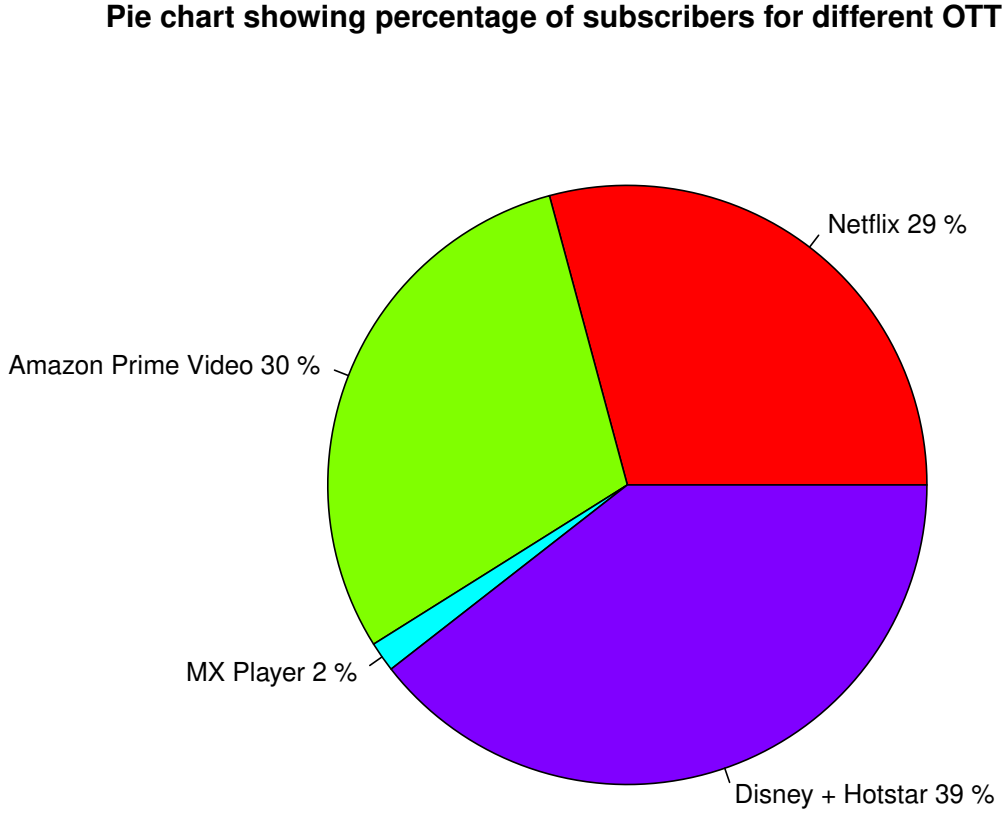
Let us check which OTT platform is most popular:

```
## [1] "Disney.Hotstar"
```

Let us check which one has maximum subscribers:

```
## [1] "Disney.Hotstar"
```

Now let us have a look at the percentage of subscribers of different OTT platforms through a pie chart.

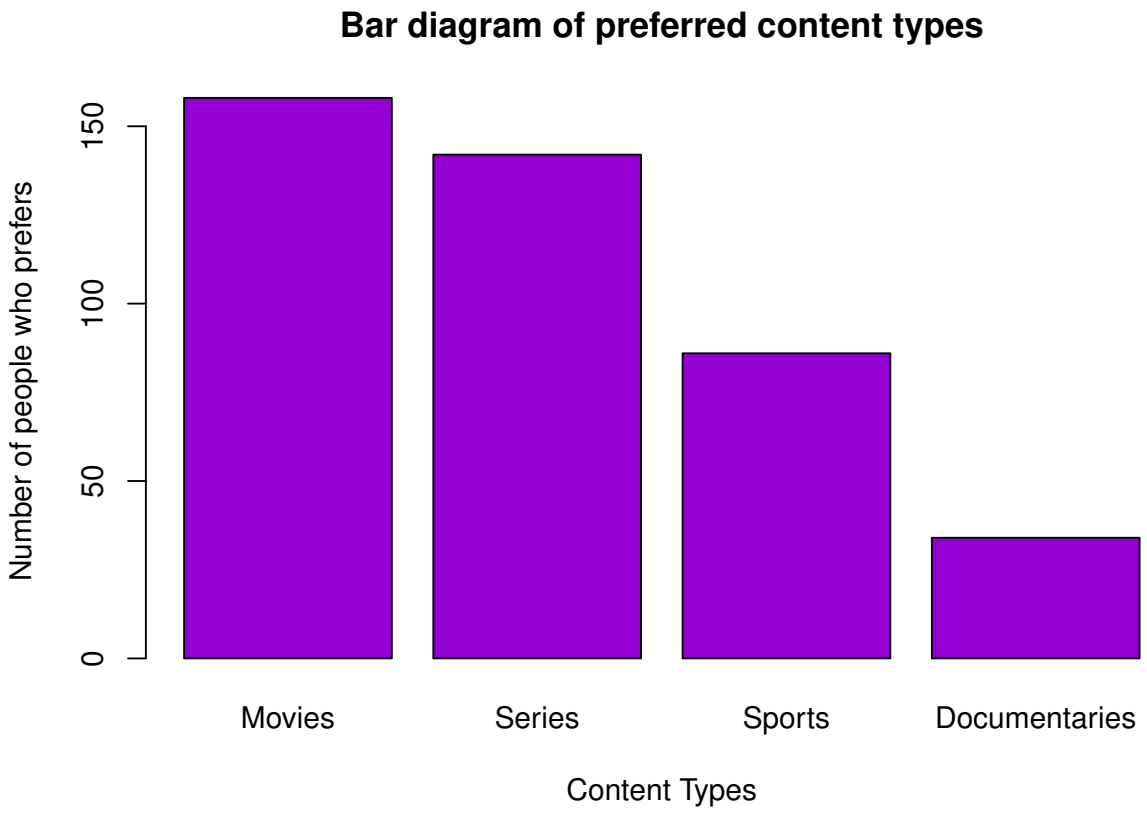


From the above pie chart we can say that Disney+Hotstar has maximum subscribers. Also Netflix and Amazon Prime Video have significant subscribers. However MX Player has few number of subscribers.

Type of contents people prefer the most:

```
## [1] "Movies"
```

Let us take a look at the bar diagram showing the number of people preferring different content types in OTTs:



Clearly from the above bar diagram most of the users prefer to watch movies and series in OTTs. Sports also has significant number of viewers.

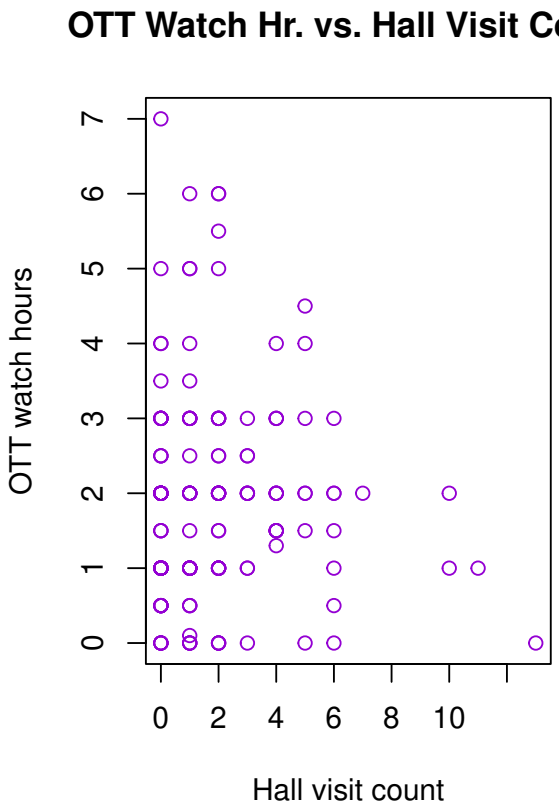
Type of genre people prefer the most:

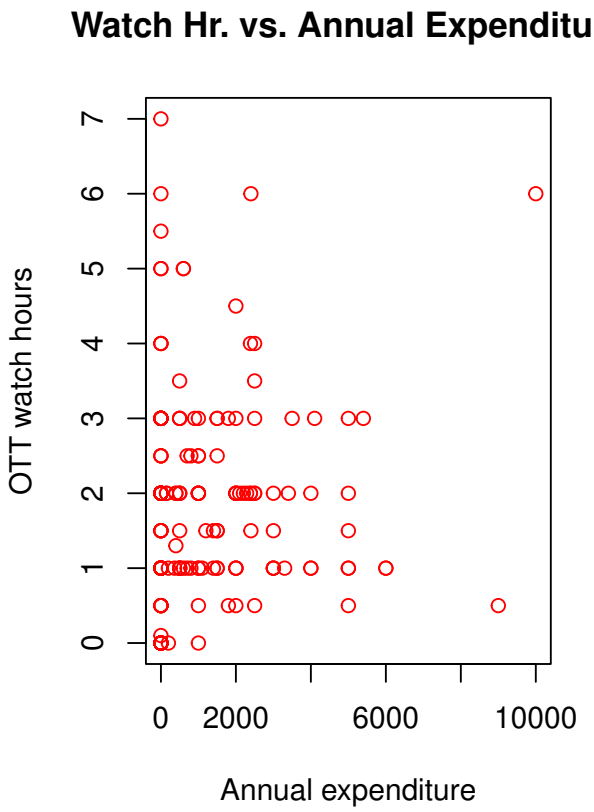
```
## [1] "Thriller.Crime.Mystery"
```

Checking whether there is any association among the numerical variables:

```
##                  Age Annual.expend OTTtime Int.time Hall.visit
## Age            1.000        -0.030  -0.051   -0.307     -0.123
## Annual.expend -0.030         1.000   0.097    0.148      0.033
## OTTtime       -0.051         0.097   1.000    0.272      0.010
## Int.time      -0.307         0.148   0.272    1.000      0.038
## Hall.visit    -0.123         0.033   0.010    0.038      1.000
```

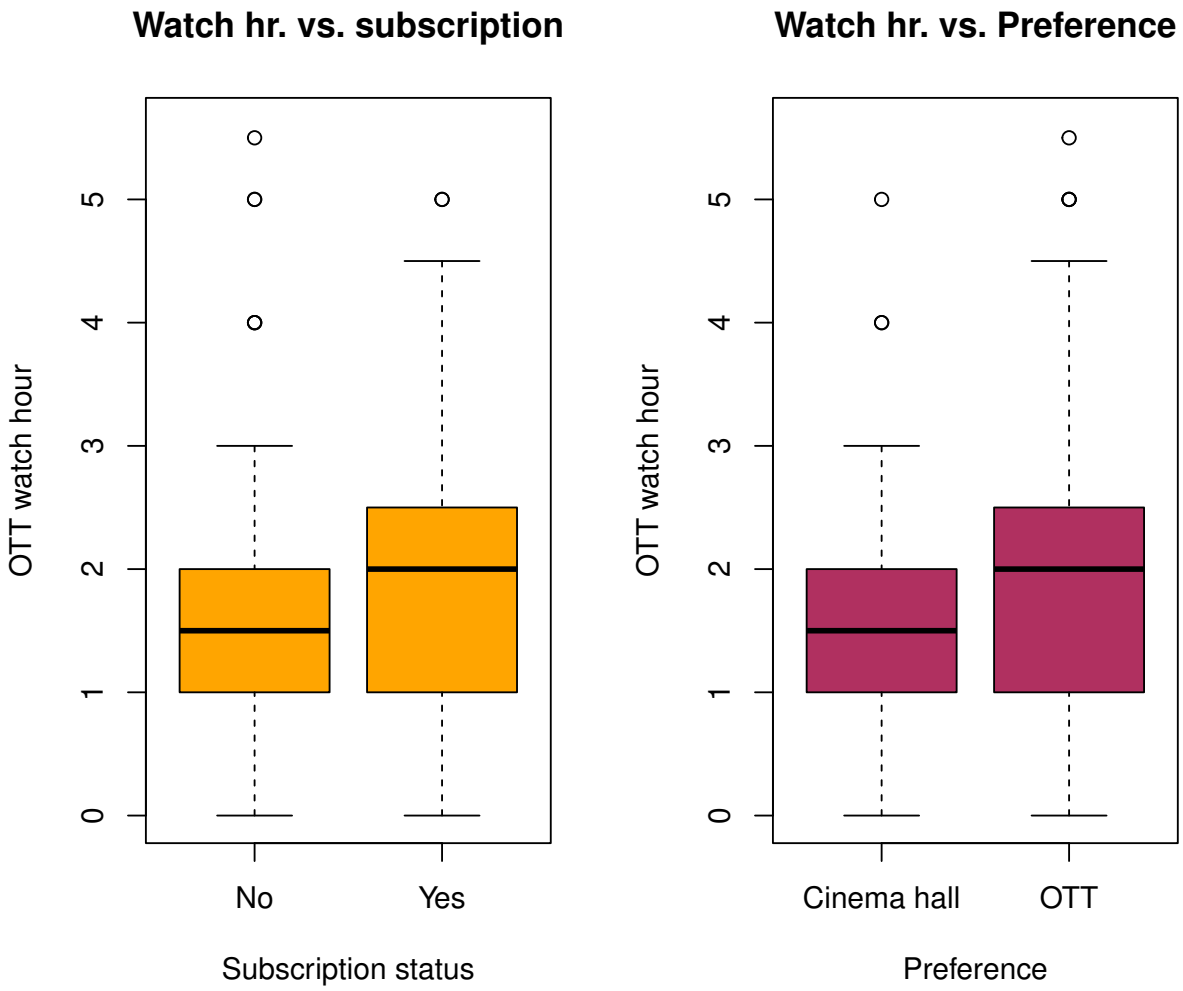
We find that the correlation between hall visit count and OTT watch hour and that between annual expenditure and OTT watch hour are both less than 0.1. But we were expecting a significant association between them. Let us do the scatterplots for more clarification.





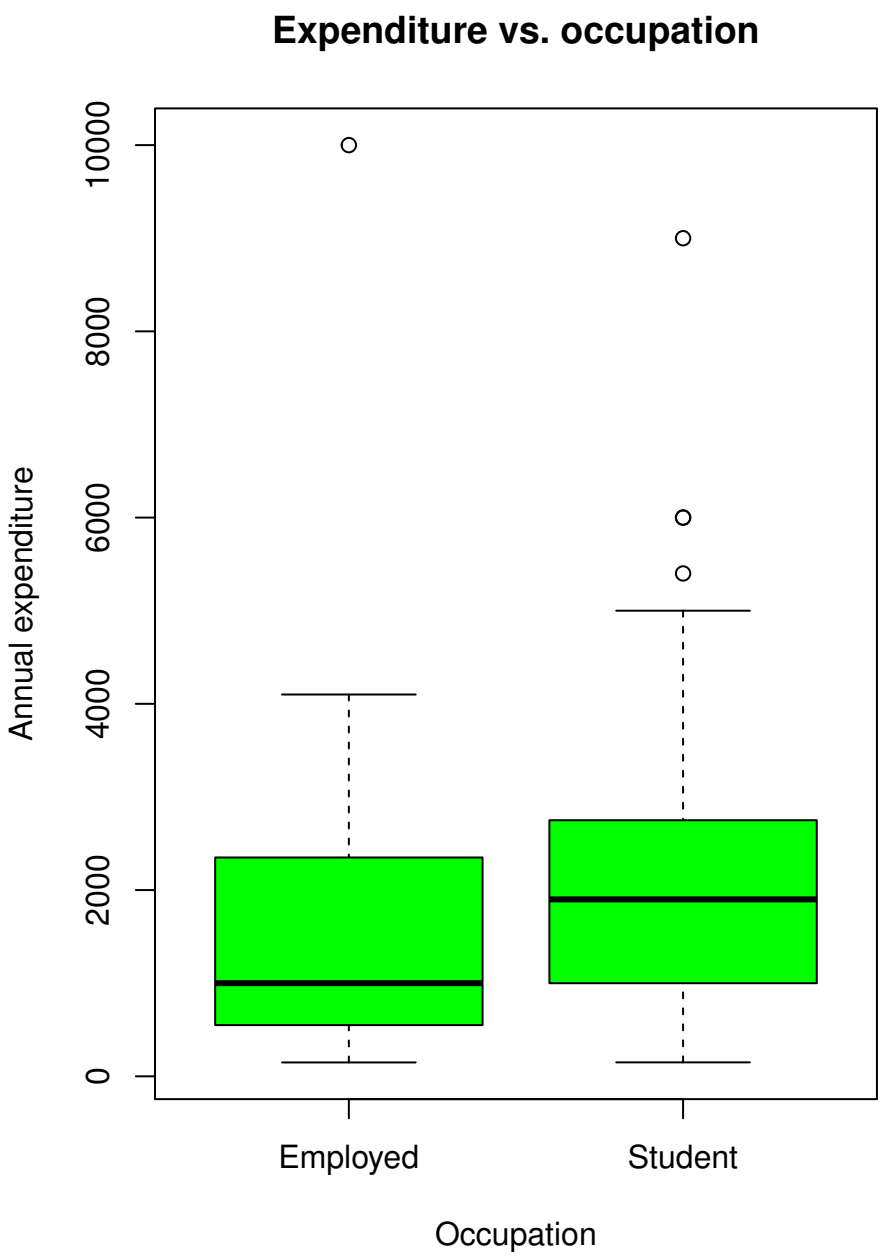
Clearly hall visit and OTT watch hour are not correlated and also there is no correlation between annual expenditure and OTT watch hour as per the concerned data.

Checking whether there is any influence of categorical variable on quantitative variable:



From the boxplot of watch hour vs. subscription status we can see that people having subscription in OTT platform have more daily watch hours as compared to the people who do not have subscription. Also 50% of people have more than 2 hours of daily watch time out of the people who have subscriptions whereas 50% of the people have approximately less than 1.5 hours of daily watch time out of the people who do not have subscription.

Again from the boxplot of watch hour vs. watching status we can conclude that people who prefer to watch any newly released cinema on OTT platforms have more daily OTT watch hour as compared to the people who prefer to watch newly released cinema in cinema halls as per the concerned data. Also 50% of people have more than 2 hours of daily watch time among the people who prefer to watch newly released cinema on OTTs whereas 50% people have approximately less than 1.5 hours of daily watch time out of those who prefer to watch newly released cinema in the cinema halls.



The boxplot shows that students have more annual expenditure than the employed people on OTT platform.

Checking whether there is any association between two categorical variables:

We want to study the association between residential status and language preference. For this purpose we create a 2×2 contingency table as follows.

```
##
##         Yes No
```

```
##   Rural  39 17
##   Urban  77 67
```

Also we calculate odds ratio to find the association.

Odds Ratio:

	A	α	
B	f_{AB}	$f_{\alpha B}$	f_B
β	$f_{A\beta}$	$f_{\alpha\beta}$	f_β
Total	f_A	f_α	n

n=$f_A + f_\alpha$=$f_B + f_\beta$

Here, categorical variable A has two categories A and α and categorical variable B has two categories B and β

The formula of Odds Ratio is given by

Odds Ratio=$\frac{f_{AB} \cdot f_{\alpha\beta}}{f_{A\beta} \cdot f_{\alpha B}}$

Decision:

If Odds Ratio>1,then we say that A and B are positively associated.

If Odds Ratio<1, then we say that A and B are negatively associated.

If Odds Ratio=1,then we say that A and B are independent.

```
## odds ratio
##   1.989438
```

Since Odds Ratio is more than 1, we can say that the residential status and language preference are positively associated. Therefore on the light of the data, we can conclude that people from rural areas prefer to watch in OTTs in regional language more than those in urban areas.

7. Inferential Data Analysis

7.1. Testing-1

We want to test whether the average annual expenditure of the viewers belonging to the lower income group (annual income is less than 2 lakhs) is same as that of the viewers belonging to higher income group (annual income is more than 2 lakhs). And here we are considering only those viewers who have subscriptions for OTT platforms.

Suppose X denote the annual expenditure of a viewer from a lower income group and Y denote the annual expenditure of a viewer from a higher income group. Clearly X and Y are independent.

Let the population mean of X and Y be μ_X and μ_Y respectively and population standard deviation of X and Y be σ_X and σ_Y respectively.

Here, we perform two sample t-test and for that test the two samples should come from normal distributions.

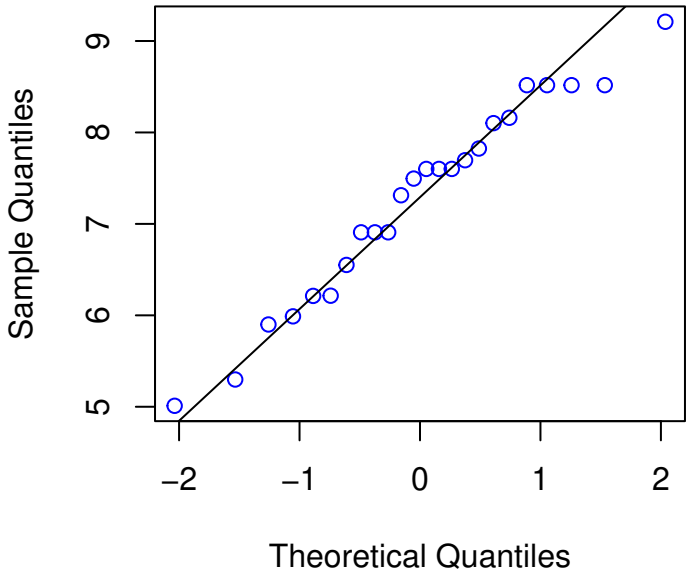


Figure 2: Normal Q-Q plot of U

So firstly we should check whether the samples are coming from normal distributions or not. But, since X and Y take only positive values so we need to take the logarithm of X and Y so as to get some negative observations as well.
Let $U = logX$ and $V = logY$
Clearly Figure(2) and Figure(3) shows that U and V both follow normal distribution.

Now let population mean of U and V be μ_U and μ_V respectively and population standard deviation of U and V be σ_U and σ_V respectively.
So $U \sim N(\mu_U, \sigma_U^2)$ and $V \sim N(\mu_V, \sigma_V^2)$ independently.
We have to test the null hypothesis,

$$H_0 : \mu_U = \mu_V$$

Against the alternative hypothesis,

$$H_1 : \mu_U \neq \mu_V$$

For this test homoscedasticity assumption is required. i.e. we have to assume that $\sigma_U = \sigma_V = \sigma$ (say).
The test-statistic is,

$$t = \frac{(\bar{U} - \bar{V}) - (\mu_U - \mu_V)}{s' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

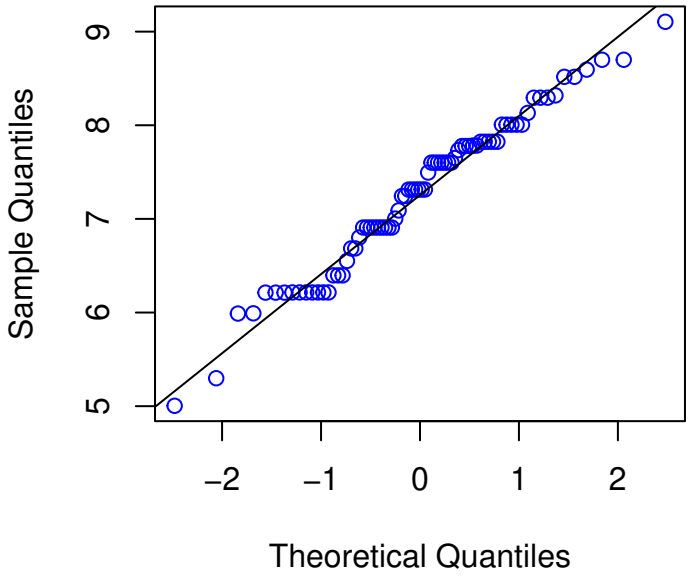


Figure 3: Normal Q-Q plot of V

where n_1 and n_2 are the number of observations in the two samples respectively and s' is an unbiased estimator of σ.

Defined as, $s'^2 = \frac{1}{n_1+n_2-2}\left(\sum_{i=1}^{n_1}(U_i-\bar{U})^2+\sum_{j=1}^{n_2}(V_j-\bar{V})^2\right)$

Where U_i 's are random observations of the variable U and V_j 's are random observations of the variable V.

So under H_0 the test-statistic is,

$$t = \frac{(\bar{U}-\bar{V})}{s'\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Now, $p-value$ of the test

$$= 2 \times min\left\{P_{H_0}\left(t > observed\ t\right), P_{H_0}\left(t < observed\ t\right)\right\}$$

If we consider α to be the level of significance of the test, then we reject H_0 at the level α when, $p-value < \alpha$; Otherwise we accept H_0 .

Also a $100(1-\alpha)\%$ confidence interval of $\mu_U - \mu_V$ is given by,

lower confidence limit

$$= (\bar{U}-\bar{V}) - s'\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\frac{\alpha}{2};n_1+n_2-2}$$

and upper confidence limit

$$= (\bar{U} - \bar{V}) + s' \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} t_{\frac{\alpha}{2}; n_1+n_2-2}$$

Where $t_{\alpha;n_1+n_2-2}$ is the $(1-\alpha)th$ quantile of $t_{n_1+n_2-2}$ distribution.
If the p-value is less than α, we reject H_0at the level α, otherwise, we accept H_0.
Here, we are taking $\alpha = 0.05$.
Now the ***R*** output of the test is,

```
##
##  Two Sample t-test
##
## data:  u and v
## t = 0.01513, df = 98, p-value = 0.988
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4178701  0.4242909
## sample estimates:
## mean of x mean of y
##   7.27385   7.27064
```

Clearly, $p - value > 0.05$.
So we accept the null hypothesis at the level 0.05 .
i.e. we conclude that the population mean of the log of annual expenditure of the viewers from a lower income group is equal to that of the viewers of a higher income group as per the concerned data. That means annual income is not an influencing factor of OTT expenditure of a viewer.
And a 95% confidence interval for difference of the population mean of the log of annual expenditure of the viewers of a lower annual income group and the population mean of the log of annual expenditure of the viewers of a higher annual income group is given by $[-0.4178701, 0.4242909]$.

7.2. Testing-2

We want to study whether there is any association between the residential status and subscription status of the viewers. Chi-square test for independence will be a useful method for this inference.
So we have to test the null hypothesis,
H_0 : Residential status and subscription status are not associated.
Against the alternative hypothesis,
H_1 : Residential status and subscription status are associated.
Two categories of residential status are "Rural" and "Urban" and the categories of subscription status are "No" and "Yes".

Residence	Subscription status: No	Yes	Total
Rural	a	b	$a+b$
Urban	c	d	$c+d$
Total	$a+c$	$b+d$	n

Table 1: Contingency table

Suppose the contingency table of the two variable is shown in the Table(1).
Then the test-statistic is,

$$\chi^2 = \frac{(|ad-bc| - \frac{n}{2})^2 n}{(a+b)(c+d)(a+c)(b+d)} \sim \chi_1^2$$

Finding the contingency table of the test using ***R***,

```
##
##         No Yes
##   Rural 26  30
##   Urban 74  70
```

Now, $p-value$ of the test is

$$= P_{H_0}(\chi_1^2 > observed\ \chi^2)$$

If we consider α to be the level of significance of the test. Then we reject H_0 at the level α when, $p-value < \alpha$; Otherwise we accept H_0 .
Here we are taking $\alpha = 0.05$.
The ***R*** output of the test is,

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(OTT$Resid.status, OTT$Subscription)
## X-squared = 0.22321, df = 1, p-value = 0.6366
```

Clearly, $p-value > 0.05$.
So we accept the null hypothesis at the level 0.05 .
i.e. as per the concerned data, residential status and subscription status are not associated.

8. Conclusion

The study enables us to verify our belief on the population characteristics statistically. We accept the fact that our sample is not so representative of the targeted population as mainly we got responses from the students. Except for this fact, the sample does not have any major flaws. Based on the sample we have studied we are going to take the following key conclusions.
Most of the viewers are watching only the free content available. Though a significant number of users have preferred content type as movie on OTTs, still people prefer to watch newly

released movies in cinema halls. Majorly preferred genres are Thriller, crime and mystery. Disney+ Hotstar has the maximum market share. Students are paying more on an average than employed people on OTTs. We found that people from rural areas prefer to watch OTTs in regional languages more than those in urban areas. Annual expenditures of the viewers from the lower and higher income groups are not significantly different. And finally, we see that subscription status is also not significantly differentiated by the residential status of the consumer.

9. Appendix

All the R codes used in the project are given below....

To calculate summary measures of the variables:

```
options(width=80)
OTT=read.csv("C:\\Users\\DELL\\Downloads\\OTT_new")
Platforms=read.csv("C:\\Users\\DELL\\Downloads\\Platforms_new")
Contype=read.csv("C:\\Users\\DELL\\Downloads\\Contype_new")
Genre=read.csv("C:\\Users\\DELL\\Downloads\\Genre_New")
for(i in c(2:7,11)){OTT[,i]=factor(OTT[,i])}
summary(OTT[,c(6,7,8,9,11,12)])
```

Histogram of annual expenditure in OTT platforms and daily OTT watch hour:

```
par(mfrow=c(1,2),new=T)
hist(OTT$Annual.expend,main="Histogram of Annual Expenditure",
xlab="Annual Expenditure",col="Blue",breaks=16)
hist(OTT$OTTtime,main="Histogram of Daily OTT Watch Hrs.",
xlab="Daily OTT Watch Hrs.",col="Green",breaks=8)
#for(i in c(8,9,12)){hist(OTT[,i],main=paste("Histogram of",colnames(OTT)[i]),
xlab=colnames(OTT)[i],col="Blue",breaks=8)}
```

Barplot of number of visits to the cinema halls in the last 3 months:

```
par(mfrow=c(1,1))
barplot(table(OTT$Hall.visit),main="Column Diagram of Hall Visit Count",
xlab="Hall Visit Count",ylab="Frequency",col="Orange")
```

To check which OTT platform is most popular:

```
Platforms$Plat.name[which(Platforms$OTTplat==max(Platforms$OTTplat))]
```

To check which one has maximum subscribers:

```
Platforms$Plat.name[which(Platforms$Subscribe.plat==max(Platforms$Subscribe.plat))]
```

Pie chart showing the percentage of subscribers of different OTT platforms:

```
data=Platforms$Subscribe.plat
percntg=round(data/sum(data)*100)
lbls=c("Netflix","Amazon Prime Video","MX Player","Disney + Hotstar")
pie(data,main="Pie chart showing percentage of subscribers for different OTT platforms",
labels=paste(lbls,percntg,"%"),
col=rainbow(length(data)))
```

Type of contents people prefer the most:

```
Contype$Content.type[which(Contype$Frequency==max(Contype$Frequency))]
```

Bar diagram showing the number of people preferring different content types in OTTs:

```
barplot(Contype$Frequency,names.arg=Contype$Content.type,ylab="Number of people who prefers",
main="Bar diagram of preferred content types",col="Darkviolet",xlab="Content Types")
```

Type of genre people prefer the most:

```
Genre$Genre.name[which(Genre$Frequency==max(Genre$Frequency))]
```

Checking whether there is any association among the numerical variables:

```
OTT_cor=cor(OTT[,c(1,8,9,10,12)])
round(OTT_cor,3)
```

Scatterplots between (a) the number of times the cinema halls were visited in the last 3 months and daily watch time of OTTs (b) Annual expenditure on subscriptions for OTT platforms and daily watch time of OTTs:

```
par(mfrow=c(1,2))
plot(OTT$Hall.visit,OTT$OTTtime,main="OTT Watch Hr. vs. Hall Visit Count",
xlab="Hall visit count",ylab="OTT watch hours",col="Darkviolet")
plot(OTT$Annual.expend,OTT$OTTtime,main="Watch Hr. vs. Annual
Expenditure",xlab="Annual expenditure",ylab="OTT watch hours",col="Red")
```

Checking whether there is any influence of categorical variable on quantitative variable:

```
par(mfrow=c(1,2))
boxplot(OTT$OTTtime~OTT$Subscription,xlab="Subscription status",ylab="OTT watch hour",
main="Watch hr. vs. subscription",col="orange",ylim=c(0,5.6))
boxplot(OTT$OTTtime~OTT$Watching.status,xlab="Preference",ylab="OTT watch hour",
main="Watch hr. vs. Preference",col="maroon",ylim=c(0,5.6))
```

```
library(forcats)
Occupation_new=fct_collapse(subset(OTT,Subscription=="Yes")$Occupation, Employed=
c("App. Dev. associate",
"Assistant director","Beautitian","Business", "Corporate worker","Govt. Service",
"Homemaker","Teacher"))
par(mfrow=c(1,1))
boxplot(subset(OTT,Subscription=="Yes")$Annual.expend~Occupation_new,
main="Expenditure vs. occupation",xlab="Occupation",ylab="Annual expenditure",col="Green")
```

.

Checking whether there is any association between two categorical variables:
2×2 contingency table

```
data=table(OTT$Resid.status,factor(OTT$Region.lang,levels=c("Yes","No")))
data
```

```
fisher.test(data)$estimate
```

.

Testing 1

```
nonzero.expend=subset(OTT,Annual.expend!=0)
x=nonzero.expend$Annual.expend[which(nonzero.expend$Annual.income==
"Less than 50,000" | nonzero.expend$Annual.income=="50,000 to 2 lakhs")]
u=log(x)
qqnorm(u,col="blue",main="")
qqline(u)
```

```
y=nonzero.expend$Annual.expend[which(nonzero.expend$Annual.income==
"2 lakhs to 6 lakhs" | nonzero.expend$Annual.income=="more than 6 lakhs")]
v=log(y)
qqnorm(v,col="blue",main="")
qqline(v)
```

```
t.test(u,v,var.equal=T)
```

.

Testing 2

```
table(OTT$Resid.status,OTT$Subscription)
```

```
chisq.test(table(OTT$Resid.status,OTT$Subscription))
```