**Time and Data Consumption on OTT Applications**

G. Sri Sai Adarsh

**ABSTRACT**

This paper discusses the usage of data and time on Over The Top applications (OTT) among college students. The discussion includes various plans students use for their internet consumption and different OTT applications they use to spend their free time.

This paper also discusses the consumption of data by using various applications of media.

**Table of Contents:**

1. Introduction
2. Graphical Representation
3. Analysis & Interpretation
4. Conclusion & Limitations
5. References

**Introduction**

This paper looks into the Time and Data Consumption on OTT (Over the top) media apps. E.g. Netflix, Amazon Prime, Disney+ Hotstar, Voot, others: Sony Liv, MX Player etcetera.

The data is primary. This data contains the usage of different OTT apps and data consumed by them in Gigabytes (GB).

The data contains the information of 63 student’s time and internet consumption. It is collected with the help of google forms.

In this paper I am going to take help of various statistical tools, I learned, and use them to make meaningful observations from the data.

The different statistical methods used in this project are measures of dispersion, measures of central tendency, regression analysis and hypothesis testing.

This study, data and time consumption, mainly focuses on how the lockdown has an impact on the usage of OTT apps and daily consumption of data using them.

The data considered only the students who are using mobile data because people who use Wi-Fi, mostly, watch videos in high definition. Which tends to higher data consumption and as a result leads to outliers in the data. Since it is only taken from students it might not be a prefect sample to predict the amount of time and data consumed for the whole population.

**Graphical Representation**

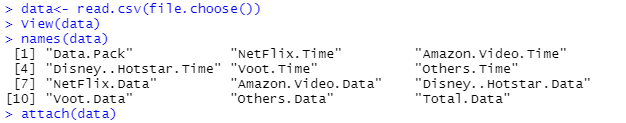
Among the 63 observations 57% percent of the students use 1.5Gb data pack, the next highest is 2Gb pack with 22% and 1Gb with 21%. However, the service provider might be different but this study focuses only on the data consumption and not the service providers.

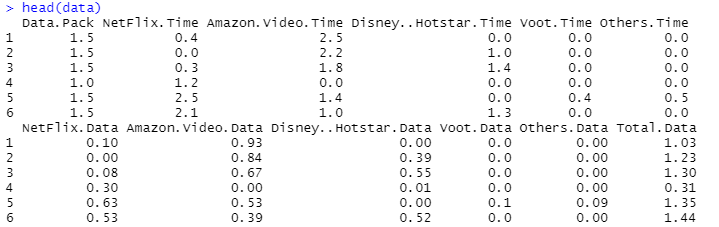
From the data almost 38% percent use Netflix, 31% use Prime video and 27% use Disney+ Hotstar and only 3% and 1% use Voot and Other apps (Sony Liv, MX Player etc) respectively.

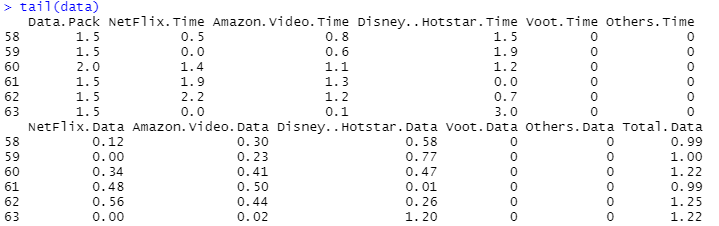
In the sample almost 19 people use OTT apps for 5 hours and 18 people use it for 3 hours. 13 people use it for 4 hours and 7,6 people use it for 2 and 6 hours respectively.

**Analysis**

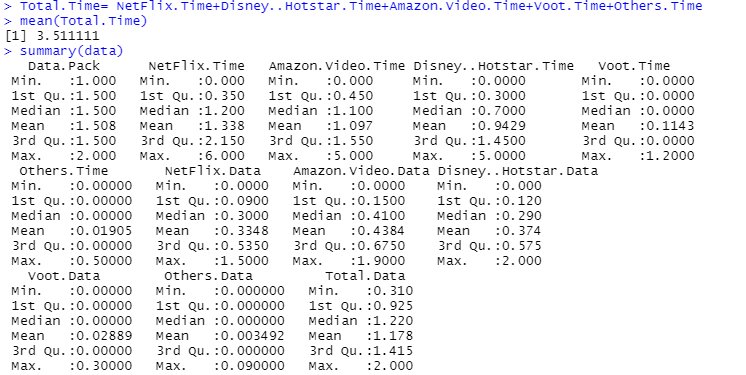
This sample data has 12 variables. Netflix Time, Amazon video time, Disney + Hotstar time, Voot time and Other apps time in hours and their Data consumptions in gigabytes (GB) respectively.







**Inferential Statistics**

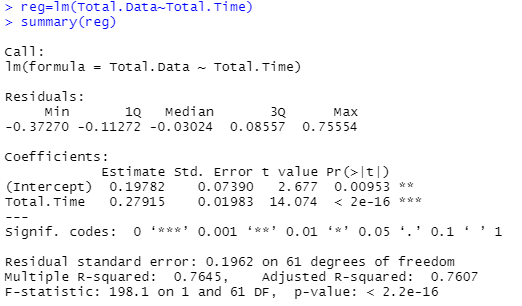




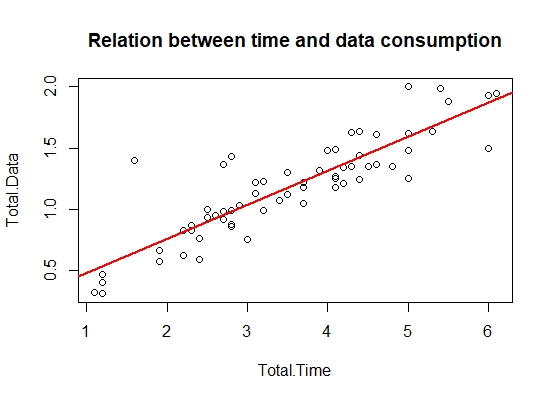




**Regression Analysis**



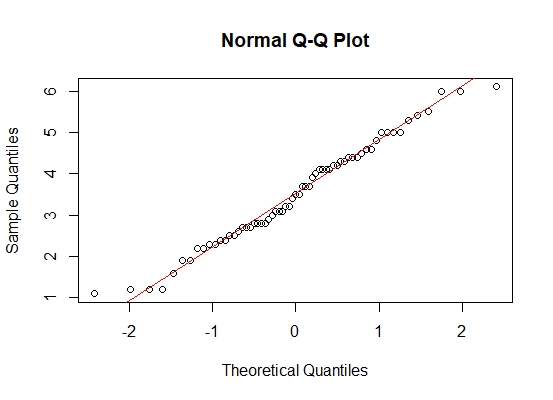


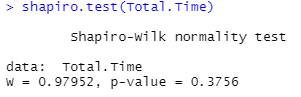


**Hypothesis Testing**

**Testing for normality**

****

****

****

From the Quantile – Quantile chart we can say that the data points fall approximately along the reference line and from Sharipo Wilk test for normality, since P-value > 0.05 we have very less proof to reject null hypothesis H₀: Sample is normally Distributed. Therefore, from both the visual representation and Sharipo Wilk test we can conclude that the data is not significantly different from the normal distribution. Hence, we can assume Normality.

Aim of the hypothesis test is to see whether the lockdown has affected the usage of internet and time.

A study Conducted in the month of Feb,2020 by The Economic Times, said that a person spends average of 70mins(≈1.2Hrs)

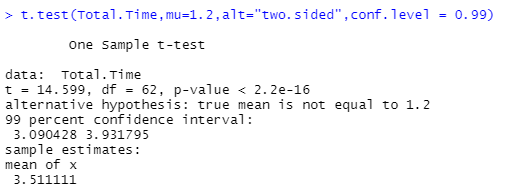
Since the study by The Economic Times is conducted in the month of February, which is before lockdown. I will try to test whether this claim still holds or not using Hypothesis test.

Step 1: Formulation of the Hypothesis.

H₀: The average mean time is equal to 1.2.

H₁: The average mean time is not equal to 1.2.

Step 2: Calculations of T-Statistic at α=0.05 i.e. 95% Confidence Level.



Step 3: Interpretation of the results.

Since the p-value is 2.2e-16 which is less than 0.05(α) there is strong evidence to reject the null hypothesis.

Step 4: Conclusion.

Therefore, we accept, H₁: The average mean is not equal to 1.2.

Therefore, we can say that after the lockdown time spent on OTT apps per day is increased.

**Interpretation**

* Summary of Statistics:

Average Time consumption on OTT (Total. Time) in a day = 3.51Hr (≈211Mins)

Average Data Consumption on OTT (Total. Data) in a day= 1.178Gb

From the above statistics it is clear that there is very small difference between Median and mean.

The deviation of Time consumption (Total. Time) from mean is 1.256481 and deviation of Data consumption of (Total. Data) is 0.4011321.

* Correlation:

We can say from above Pearson correlation that there is a positive relationship between the time spent and data consumed.

* Regression:
* Aim here is to study the relation between data (dependent) and time consumption (independent).
* The p-value of time consumption (Total. Time) is <2e-16 which is less than 0.05(α). The variable is very highly significant for the fit of the data.
* R- Squared is a goodness-of-fit measure. In my case it is 0.7645. My model can explain approximately 76% of the data.
* Residual Standard error is the difference between the actual values and the predicted values. In my case it is 0.1962.
* The F-statistic indicates whether our linear regression model provides a better fit to the data than a model that contains no independent variables.

In my case H₀: Model with no independent variable fits better.

H₁: Model with no independent variable does not fit better.

Since the p-value is 2.2e-10 < 0.05 we have strong evidence to reject the null hypothesis and we can say that my linear regression model fits data better than a model with no independent variables.

* Fitting the regression line

Total. Data = (0.27915) \*TTC+ 0.19782

We can use this equation to predict the data consumption (TDC).

Example: If we’re interested in predicting the data consumption after 8 hours of watching movies on an OTT app.

Total. Data= (0.27915) \*8 + 0.19782

Total. Data= 2.43102 (≈2.43GB)

Therefore, the data consumed after watching movies for 8 hours is 2.43 GB.

* The best Fit line is represented by the scatter plot with red line. We can see that not all the points lie on that line but it passes through the points representing the linearity of the model, making it a good fit.

**Conclusion**

Lockdown has affected the lifestyle of many in the past couple of months. The most damage was done for the students. Their exams were postponed indefinitely. The universities and colleges are shut down. The stress in them was building up because of the postponed exams. The only thing that kept them sane was their mobile and OTT apps. They blew their steam off using the help of movies or Television shows, et cetera.

From the data analysis performed above we can say that there was a significant change in the amount of time and data spent on Over the top media watching movies, TV shows, etc.

The average time consumed on OTT apps jumped from 1.2Hrs to almost 3.6Hrs after the lockdown. I conclude that lockdown had played a significant role in the rise of Time and data consumption.

**Limitations**

There are certain limitations to this study. Such as,

* Only considers who use mobile data.
* Data obtained is only taken from students.
* Data was collected before the start of online classes.

**References**

THE ECONOMIC TIMES: https://economictimes.indiatimes.com/magazines/panache/always-connected-indians-consume-over-11gb-data-daily-spend-70-mins-a-day-on-ott-platforms/articleshow/74333875.cms