MA4240 - APPLIED STATISTICS

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

AVERAGE AMOUNT OF TIME SPENT USING SOCIAL MEDIA PER DAY

GROUP MEMBERS

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

1. Ayush Agarwal - MA21MSCST11007@IITH.AC.IN
2. Santosh Kumar - MA21MSCST11015@IITH.AC.IN
3. Madhav Yadav-MA21MSCST11003@IITH.AC.IN
4. Tushar Kumar Biswas-MA21MSCST11014@IITH.AC.IN
5. Shrishti Agarwal-MA21MSCST11006@IITH.AC.IN
6. Supriya Rawat-MA21MSCST11004@IITH.AC.IN
7. Divyanshi -MA21MSCST11009@IITH.AC.IN
8. Archana Kumari-MA21MSCST11011@IITH.AC.IN

FLOW OF CONTENT

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- Motivation
- Data Collection
- Summarizing and Visualizing Data
- Analyzing average amount of time spent on Social Media
- Confidence Interval Estimator
- Hypothesis Testing
  - Rejection region Approach
  - p-value Approach

MOTIVATION

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

- We wish to study the average amount of time spent using social media amongst students .
- Clearly ,it is an exponentially growing field and the rise in internet usage has made social media users a potential customer base.
- The main reason we use social media is to keep in touch with one another,for example students, in particular ,interact and express their thoughts through various social media channels, regardless of their geographical location.
- It opens up new research avenues by inspiring students to be innovative and think beyond the box.

- You can get degrees , certifications ,information and entertainment in the same place.
- The right use of these numbers can benefit many businesses.
- Studying platforms, clothing brands, health and wellness business and potential startups can flourish using these numbers through right advertisements.
- So we must have a general idea about the trends in using social media example-what time is preferred for using social media, how often users log in their accounts and what do they browse and what is the impact on them?

Do we have any idea about how far this usage is going?

DATA COLLECTION

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

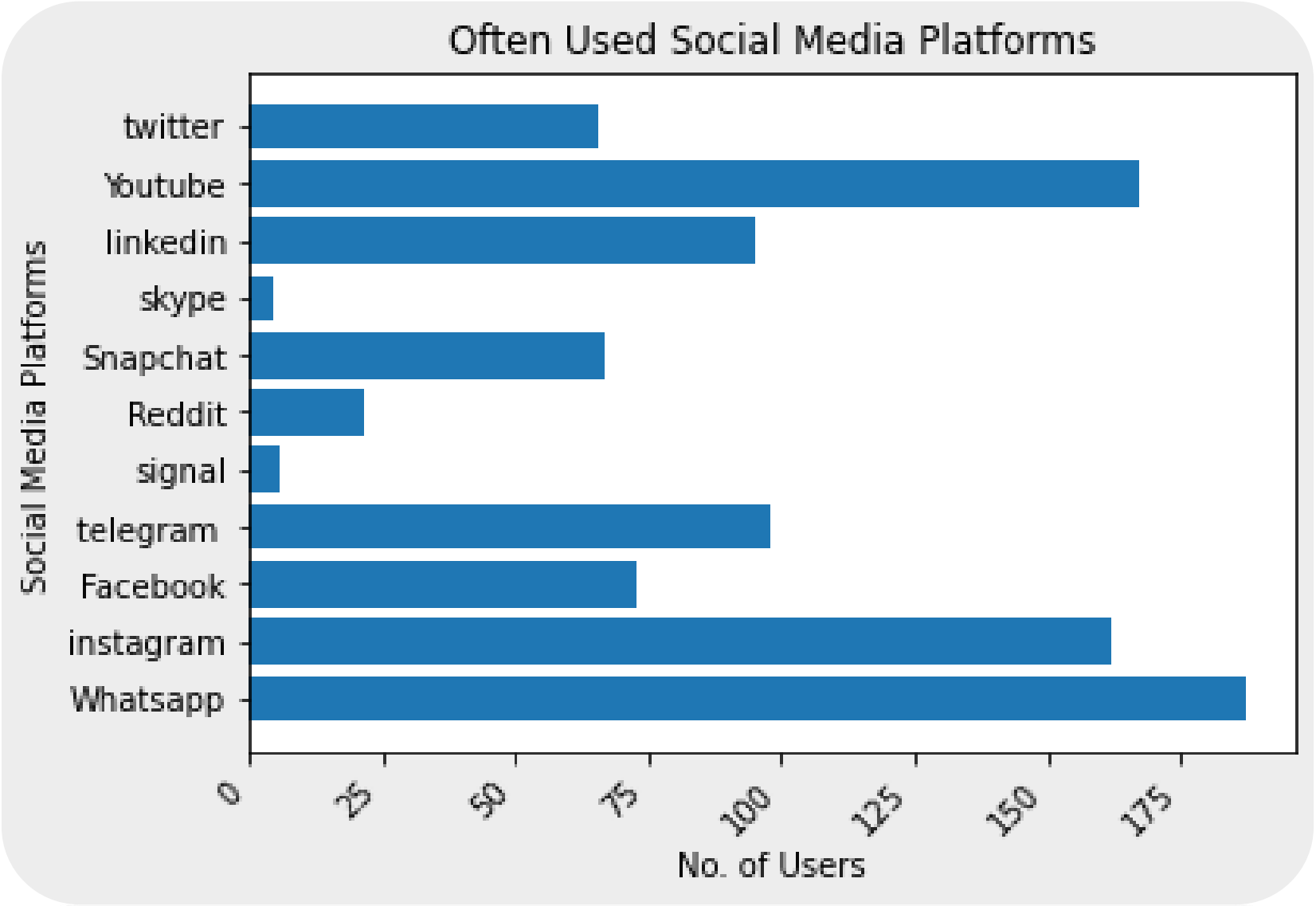
- We decided to perform an experimental study using self-administered questionnaire where we randomly targeted college/university students.
- The method we used for data collection is a non-probabilistic method i.e. Gathering Volunteers, where we invited people to participate in a survey.
- Our data set includes 192 rows and 35 columns.

CHALLENGES:-

- Form was floated through few Social Media Platforms(WhatsApp, Facebook, Instagram, G-Mail).
- Lack Of Participation amongst students caused biased data.
- Participants are generally not genuine in reviewing themselves in surveys.

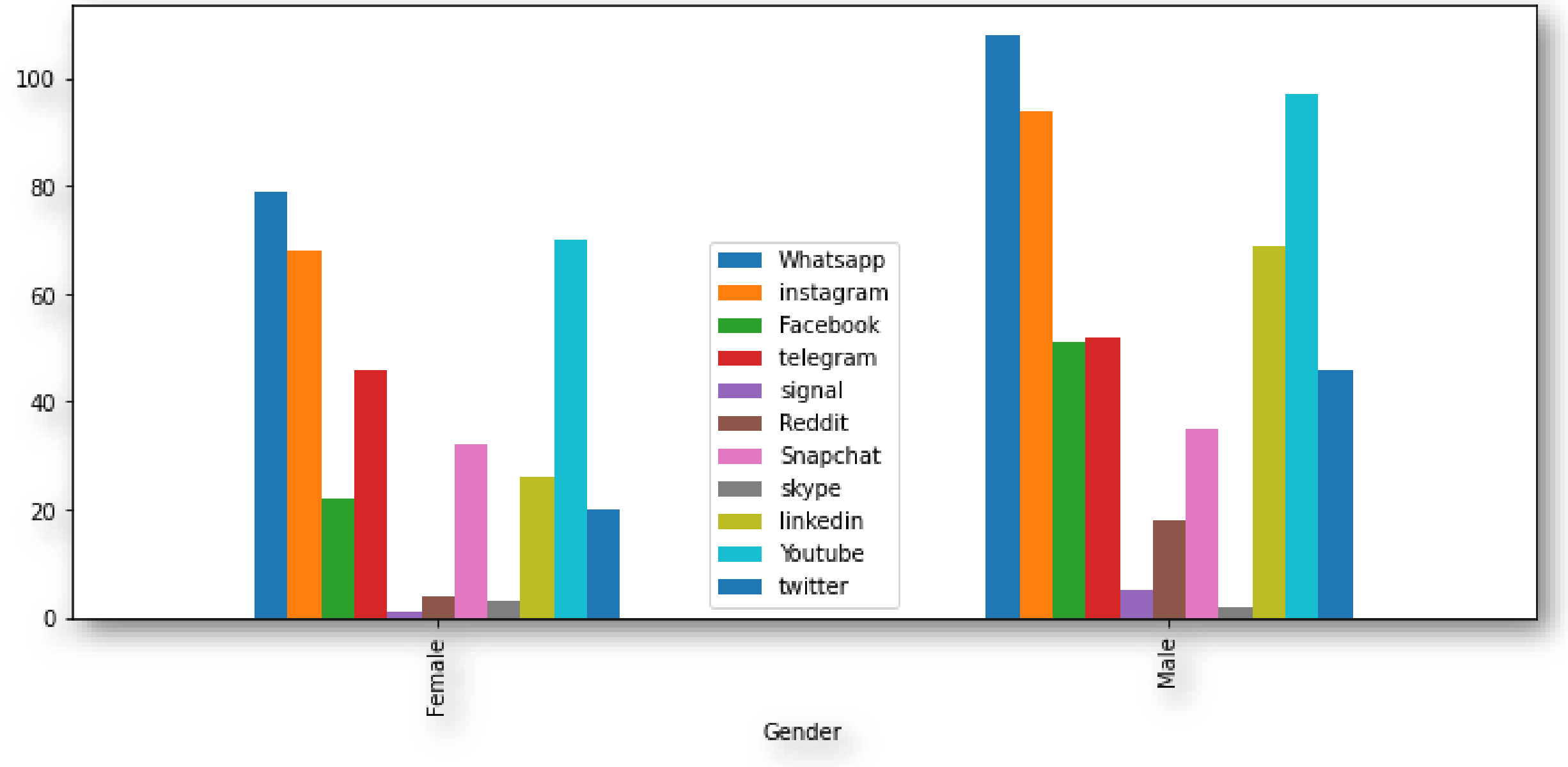
The survey can be found at https://freeonlinesurveys.com/r/k3ZI3XTX.

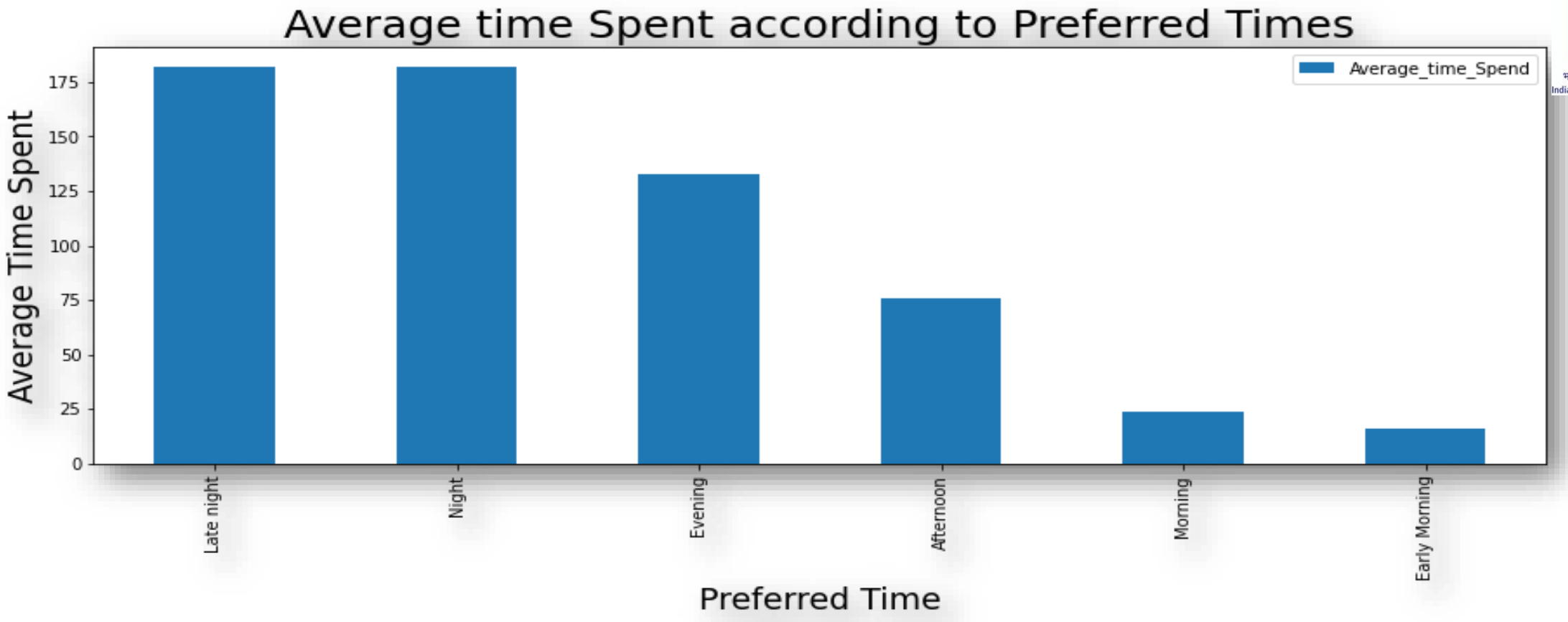
SUMMARIZING AND VISUALIZING DATA



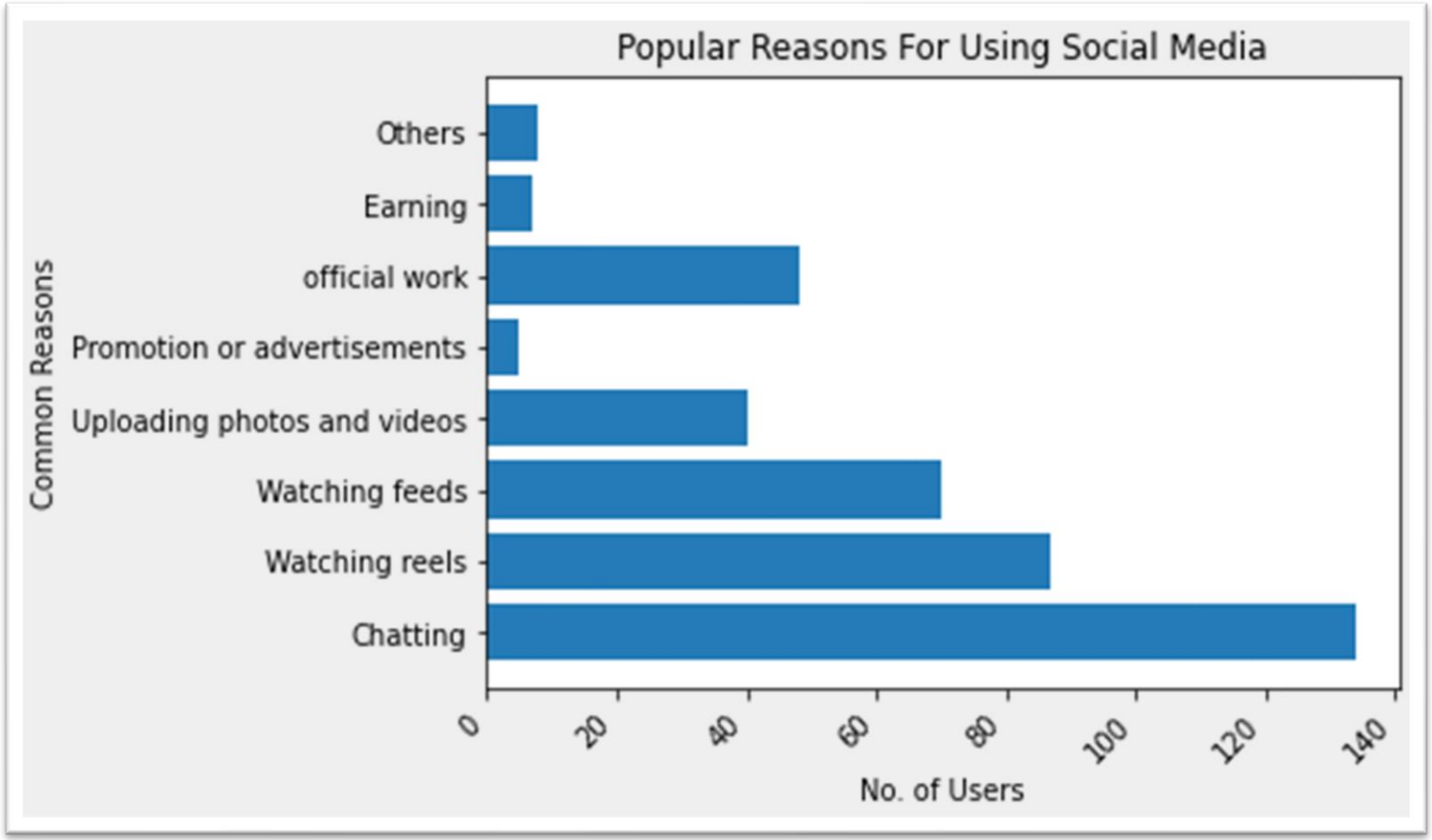
WhatsApp is the most used social media platform followed by YouTube whereas Signal and Skype are the least used social media among sample!

Which social media is most famous Male V/S Female?

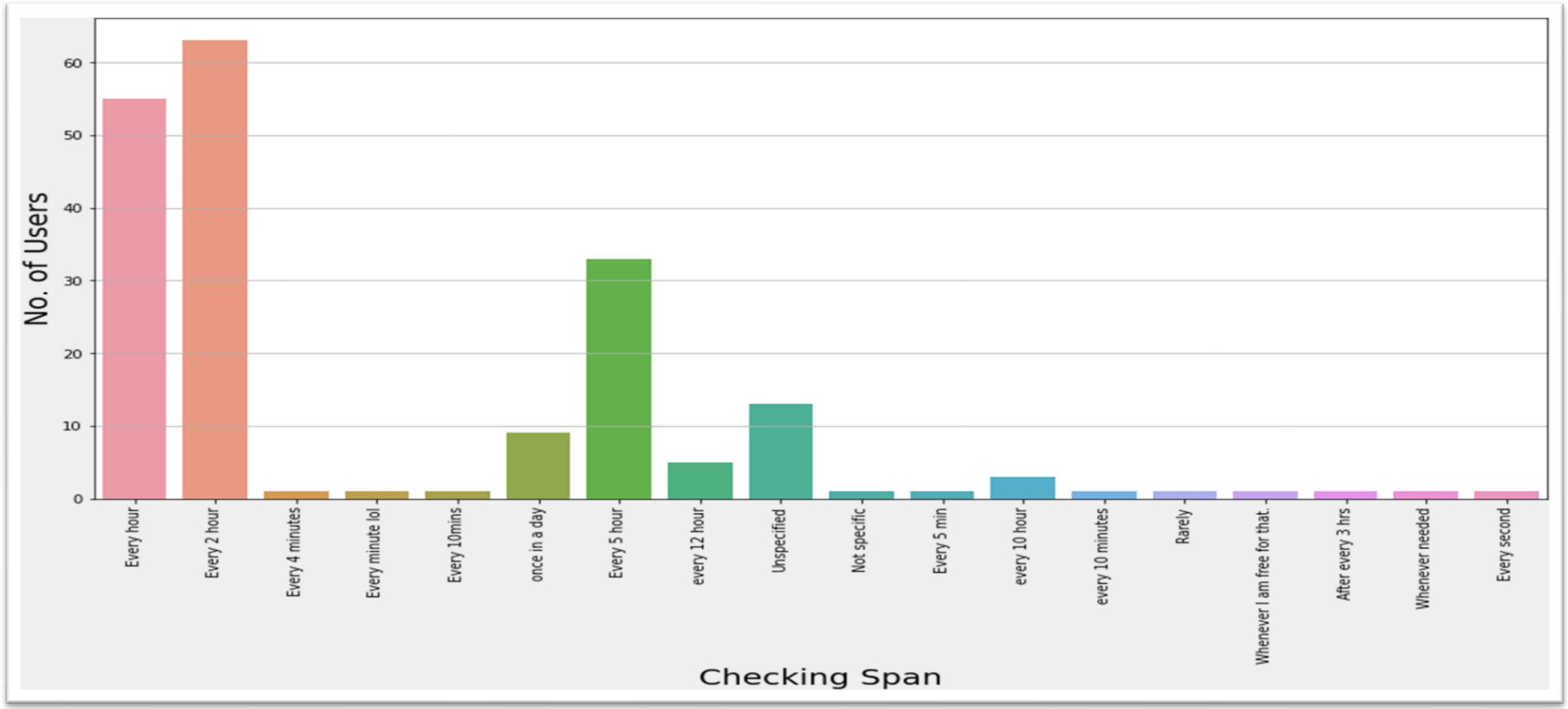




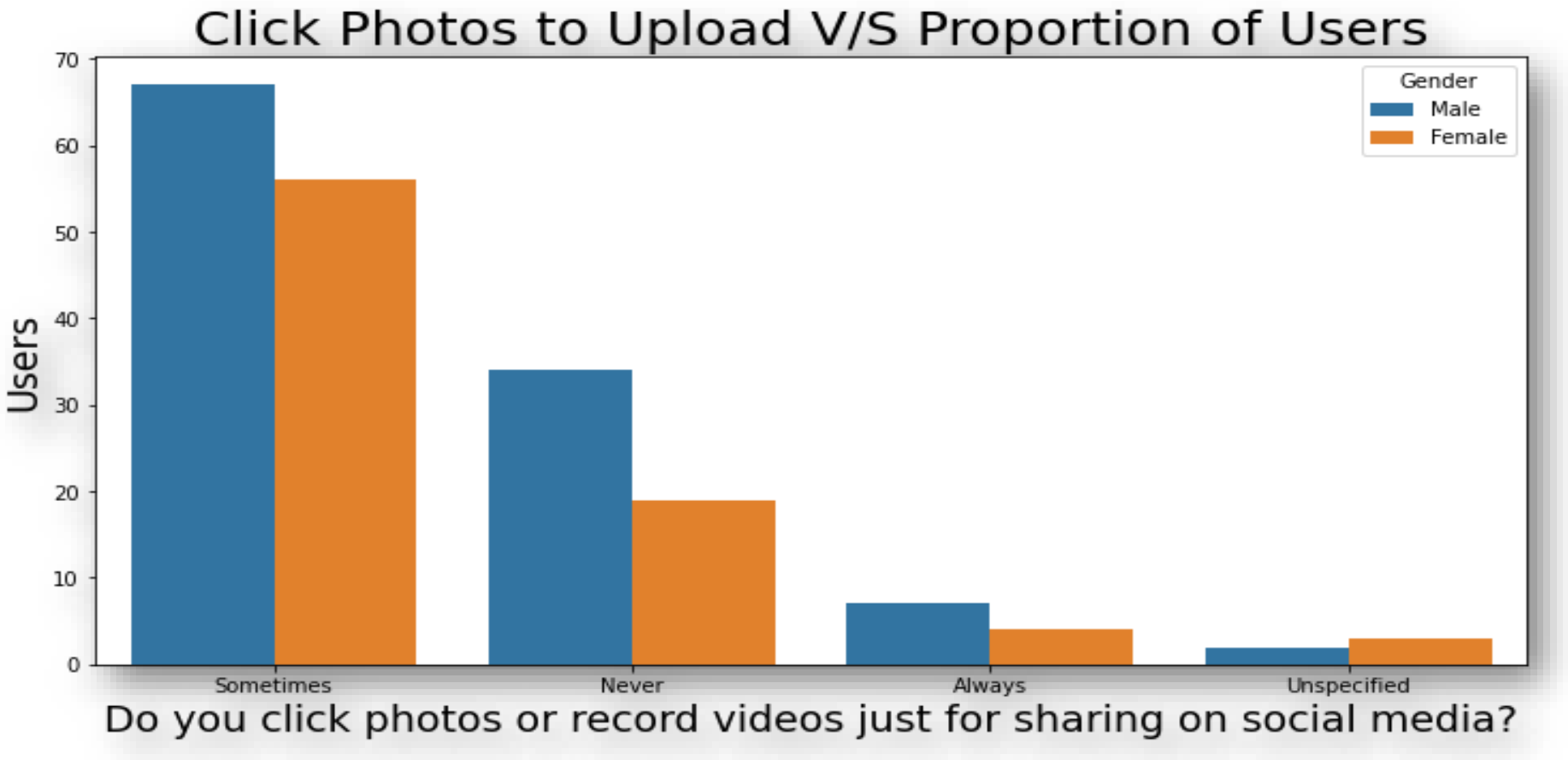
- Most of the students tend to use social media in night and late night.
- Less media used during day-time can be owed to the fact that it iss work time for every generation.



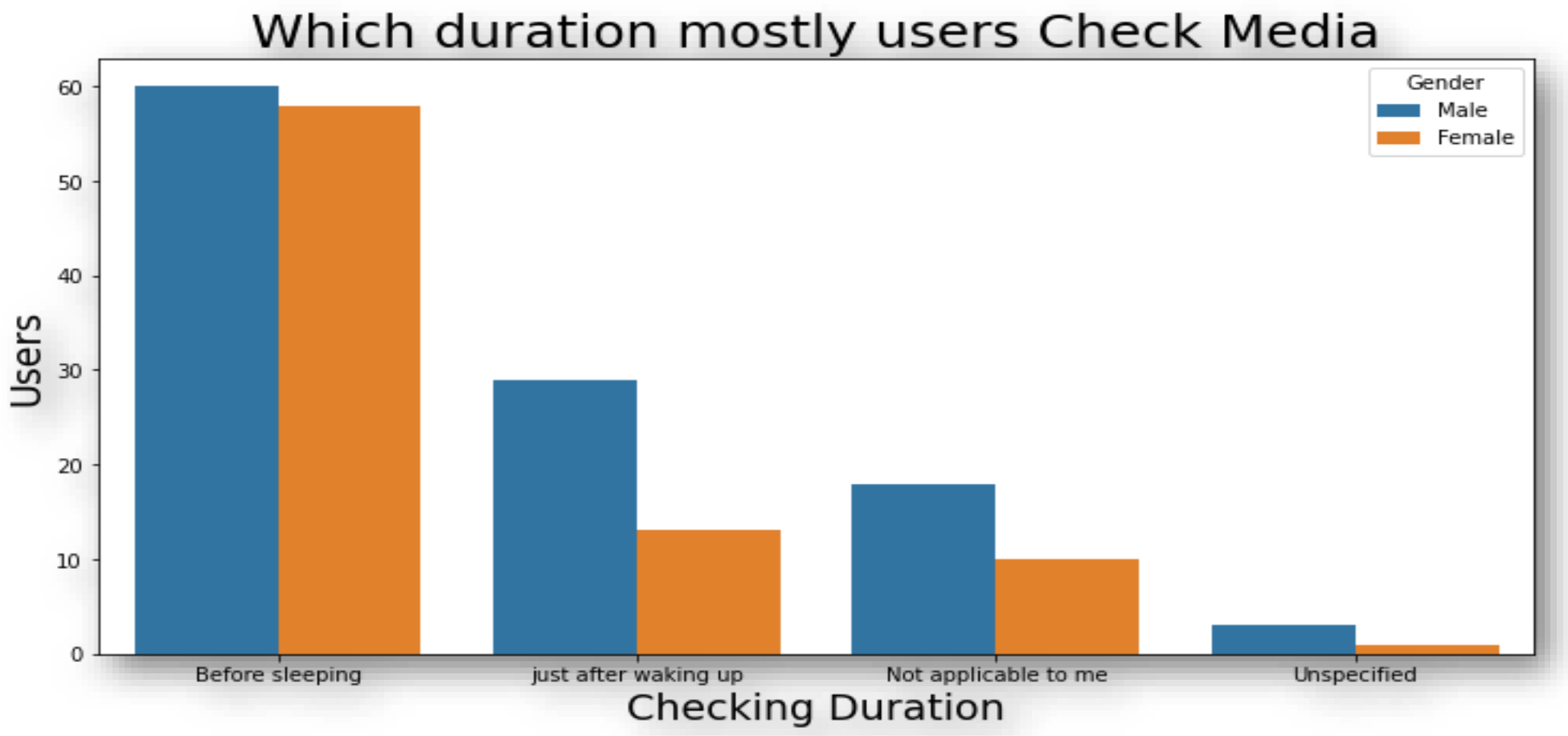
- Chatting (staying connected) is the primary reason of students using social media.
- Watching videos is the 2nd most popular reason of using social media among students.
- Earning isn't a primary reason of using social media among students!



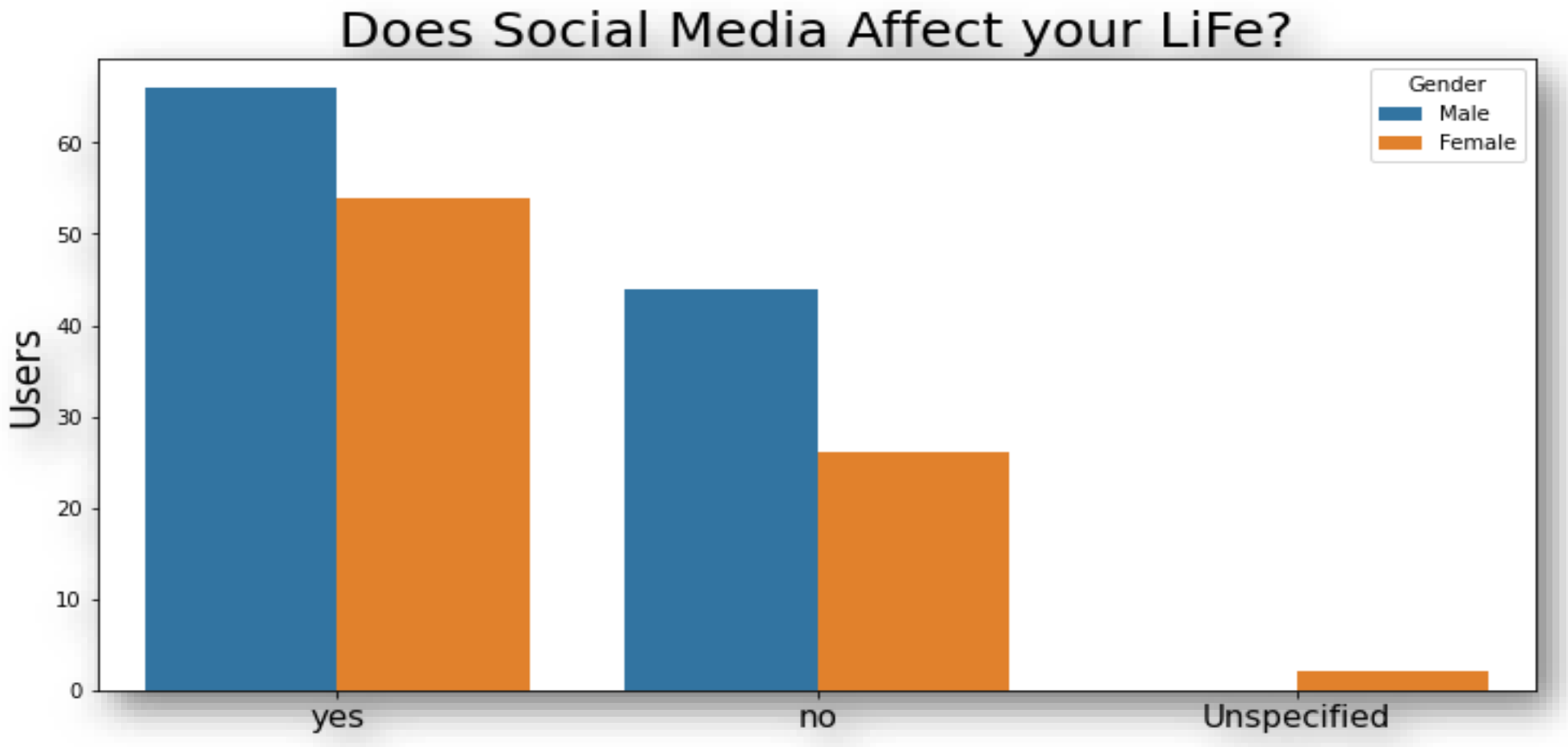
Checking social media every hour or two hours is common among half of the population!!.



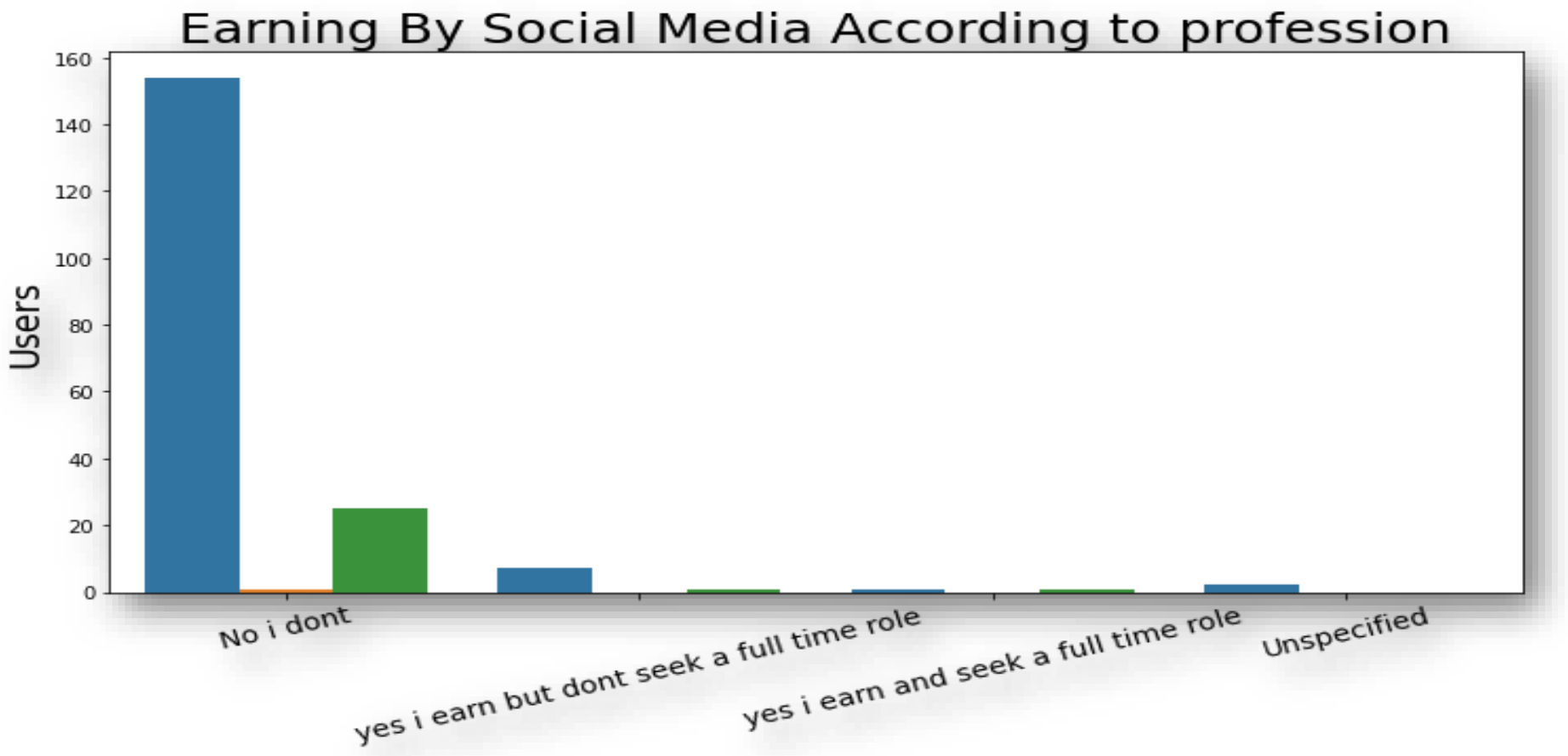
- Clicking photos just for uploading on social media is very common among both males and females!
- Very few students click photos to just upload on social media.
- Almost half of the students sometimes click photos just to upload on social media!



- More than Half of the students check their social media before sleeping.
- Checking social media after waking up is more common among males!
- Checking social media before sleeping is more common among females.

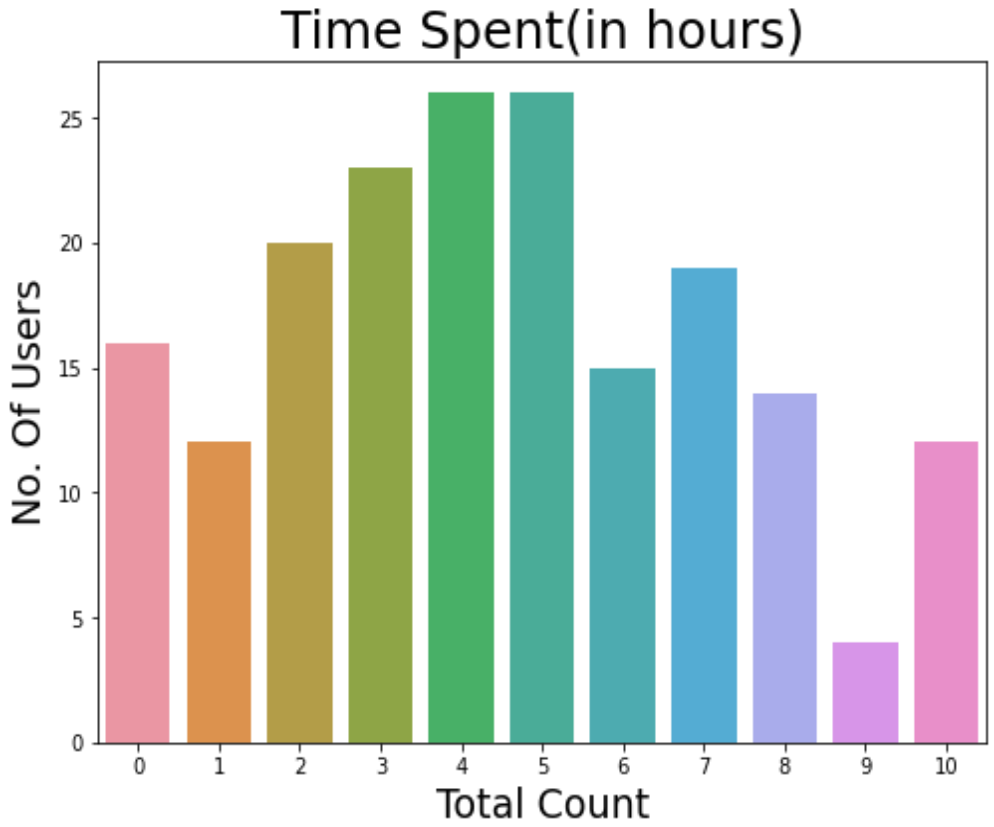


- Majority of students think that their professional life is affected by social media!
- The proportion of females who thinks that social media affects their professional life is higher than that of males!



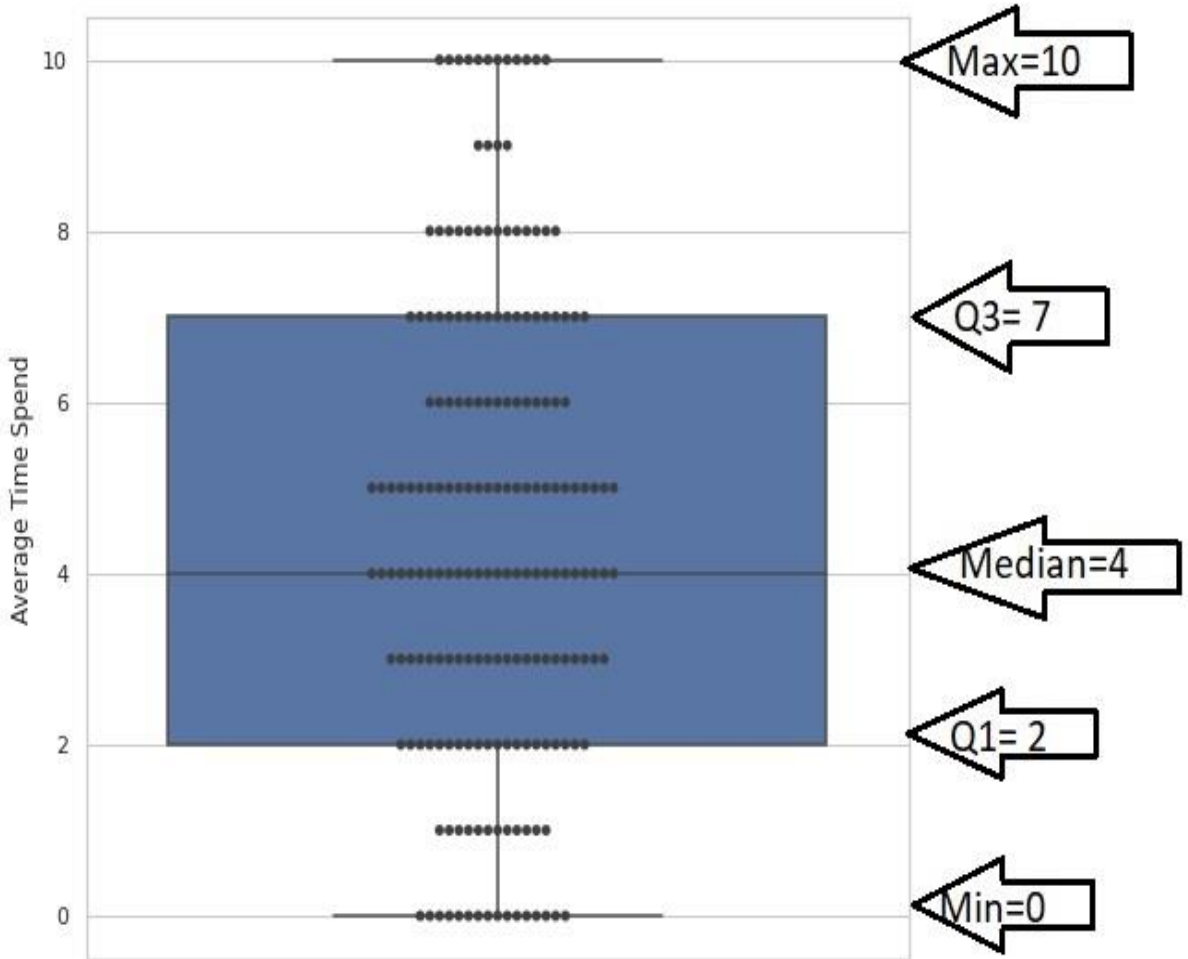
Majority of students don’t earn using social-media, some do earn and also seek full-time role.

ANALYSIS OF AVERAGE AMOUNT OF TIME SPENT USING SOCIAL MEDIA



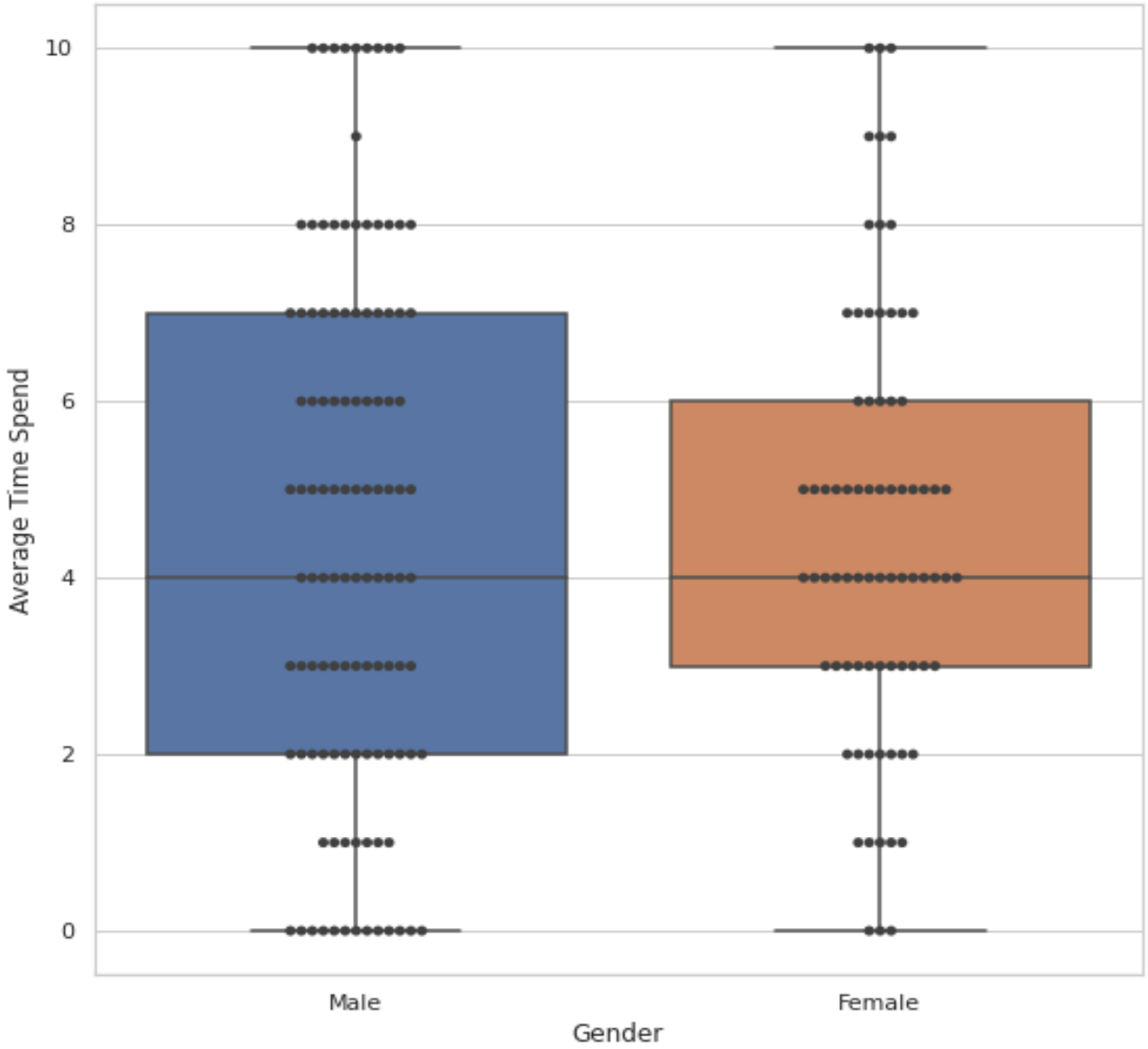
Attribute	Values
Count	191.00
Mean	3.38
Standard deviation	2.46
Minimum	0.00
Maximum	18
25^{th} percentile	2
50^{th} percentile	4
75^{th} percentile	7

BOX PLOT FOR AVERAGE TIME SPEND USING SOCIAL MEDIA



The following are the attributes shown in box-plot:

- upper-whisker = 10
- lower-whisker = 0
- Q1 = 2
- Q3 = 7
- median = 4



Males

Attributes	Value
Upper-Whiskers	10
Lower-Whiskers	0
Q1	2
Median	4
Q3	7

Females

Attributes	Value
Upper-Whiskers	10
Lower-Whiskers	0
Q1	3
Median	4
Q3	6

IQR Of male(5) is greater than of female(3)
which means there is more variability in time spent by male than in female.

CONFIDENCE INTERVAL ESTIMATION

We assume that the population is normal and the population variance is unknown. Now we wish to find a confidence interval for the population mean (the average amount of time spent by the population using social media). The confidence interval can be given as:

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{S}{\sqrt{n}}\right)$$

$\bar{X}$ is the sample mean
S is the sample standard deviation
n is the sample size
α is the confidence level in $100(1-\alpha)$ % confidence

CALCULATIONS:-

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} = 3.24\ ,\ S = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2 = 2.46\ ,\ N = 192$$

FOR 90% CONfiDENCE INTERVAL

1-α=0.90, so α=0.10 and α/2=0.05 , N-1=191.

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{S}{\sqrt{n}}\right) = 3.24 \pm t_{0.05,191}\left(\frac{2.46}{\sqrt{191}}\right) = 3.24 \pm 1.64 * 0.18 = 3.24 \pm 0.2928$$

FOR 95% CONfiDENCE INTERVAL

1-α=0.95, so α=0.05 and α/2=0.025 , N-1=191.

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{S}{\sqrt{n}}\right) = 3.24 \pm t_{0.025,191}\left(\frac{2.46}{\sqrt{191}}\right) = 3.24 \pm 1.96 * 0.18 = 3.24 \pm 0.3489$$

FOR 99% CONFIDENCE INTERVAL

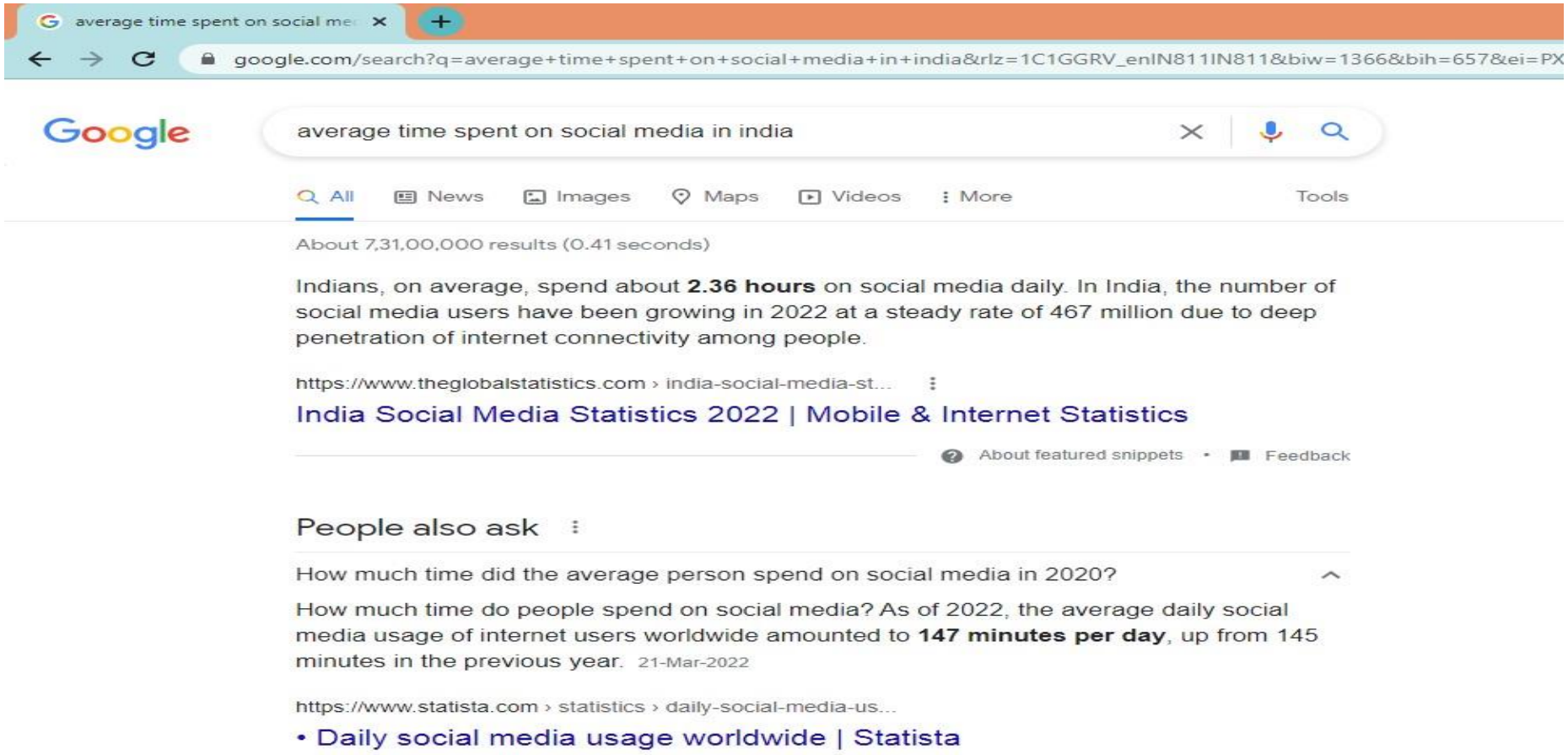
1-α=0.99, so α=0.01 and α/2=0.005, N-1=191.

$$\bar{X} \pm t_{\frac{\alpha}{2},n-1}\left(\frac{S}{\sqrt{n}}\right) = 3.24 \pm t_{0.005,191}\left(\frac{2.46}{\sqrt{191}}\right) = 3.24 \pm 2.58 * 0.18 = 3.24 \pm 0.4585$$

HYPOTHESIS TESTING

The following are the hypothesis that we have constructed:

$H_0 : \mu \leq 2.4 \; vs \; H_\alpha : \; \mu > 2.4$ (right tailed test) with support of the following Google results:



REJECTION REGION APPROACH

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

The test statistic is defined as:

$$t_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$

We already computed $\bar{X} = 3.24$, S=2.46, N=192

Also we have computed μ_0 therefore we compute the test statistic as follows:

$$t_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \frac{3.24-2.4}{2.46/\sqrt{191}} = 4.66$$

We will reject Ho if $t_0 > t_{\frac{\alpha}{2},n-1}$ where α is the level of significance.

FOR 5% LEVEL OF SIGNIfiCANCE

For 5% level of significance we have $\alpha = 0.05$ so $\frac{\alpha}{2} = 0.025$

$t_{\frac{\alpha}{2},n-1}$=$t_{0.025,190}$ =1.645

Since t_0 lines in the rejection region so we reject Ho and conclude that at the 0.05 level of significance that the average amount of time spent per day exceeds 2.4

FOR 1% LEVEL OF SIGNIfiCANCE

For 1% level of significance we have $\alpha = 0.01$ so $\frac{\alpha}{2} = 0.005$

$t_{\frac{\alpha}{2},n-1}$=$t_{0.005,190}$ =2.326

Since t_0 lines in the rejection region so we reject Ho and conclude that at the 0.01 level of significance that the average amount of time spent per day exceeds 2.4.

p-Value Approach

The test statistic is defined as:

$$t_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$$

We already computed $\overline{X} = 3.24$, S=2.46, N=192.

Also we have computed μ_0 therefore we compute the test statistic as follows:

$$t_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \frac{3.24-2.4}{2.46/\sqrt{191}} = 4.66$$

Since our H_α is right-tailed so we compute the p´value as follows:

$$p - value = P(t > |t_0|) = P(t > |4.66|)$$

At 191 degrees of freedom, the value 4.66 is obtained at 0.0000000017.

So we can approximate $P(t > |4.66|)$ with 0.0000000017.

Therefore the smallest level of significance at which the null hypothesis would be rejected is 0.0000000017 and hence we reject Ho on both 1% and 5% level of significance.

REFERENCES:

- www.theglobalstatistics.com

- Ross S.M., (2014) , Introduction to Probability and Statistics for engineers and scientists, Academic Press.


भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

THANK YOU