

**A  
PROJECT ON  
STATISTICAL ANALYSIS ON  
AWARENESS ON CHATGPT IN Z. B. PATIL COLLEGE  
STUDENTS**



*Presented By*

**Mr. OM J. BADGUJAR**

**Ms. SWATI A. BADGUJAR**

**Mr. ABHITAY S. MORE**

**Ms. MADHURI S. PATIL**

**Mr. HITESH G. DEORE**

**Submitted To**

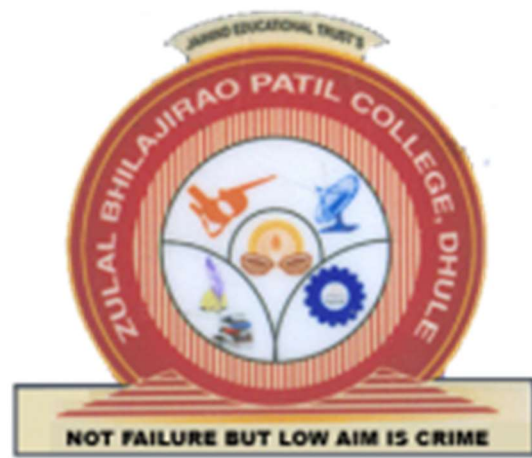
**DEPARTMENT OF STATISTICS**

**Jai Hind Education Trust's  
Z. B. PATIL COLLEGE, DHULE**

**Under the guidance of  
PROF. P. U. PATIL**

**Department of Statistics  
In Partial Fulfilment of Degree Course  
B.Sc. (Statistics)**

**Academic Year 2024-2025**



**CERTIFICATE**

**This is to certify that a project report on**  
**“AWARENESS ON CHATGPT IN Z. B. PATIL COLLEGE STUDENTS”.**

**Presented By**

**Mr. OM J. BADGUJAR**

**Ms. SWATI A. BADGUJAR**

**Mr. ABHITAY S. MORE**

**Ms. MADHURI S. PATIL**

**Mr. HITESH G. DEORE**

**Under our guidance in partial fulfilment of requirement for award  
of “Bachelor of Science (statistics)” of North Maharashtra  
University, Jalgaon, in academic year of 2024-2025.**

**Project Preceptor**

**Head of Department**

**Prof. P. U. PATIL**  
**(Project Guide)**

**Vice Principal Prof.Dr.N.S.Chavan**

## **ACKNOWLEDGEMENT**

**It is with great pleasure and effort that we are able to present this project.**

**We are very thankful to Prof. P.U. PATIL sir, Department of statistics, for this his great support and encouraging us during the course of development of this project.**

**We also express our pleasure towards,  
Vice Principal Prof. Dr. N.S. CHAVAN sir, (Head of Department) of Statistics, Z. B. College, Dhule. For their valuable guidance.**

**We are also thankful to our technical Lab assistant Mr. P.R. Jadhav who helps us while doing the project work.**

**I whole heartedly thank my parents without whose support I could never have completely this project.**

**Lastly, we would like to thanking to all our friends who given us their kind co-operation and valuable suggestion while making of project.**

**“THANKS A LOT”**

**Mr. OM J. BADGUJAR**

**Ms. SWATI A. BADGUJAR**

**Mr. ABHITAY S. MORE**

**Ms. MADHURI S. PATIL**

**Mr. HITESH G. DEORE**

**T. Y. B. Sc. (2024-2025)**

**Dept. of Statistics**

• **Manifestation:-**

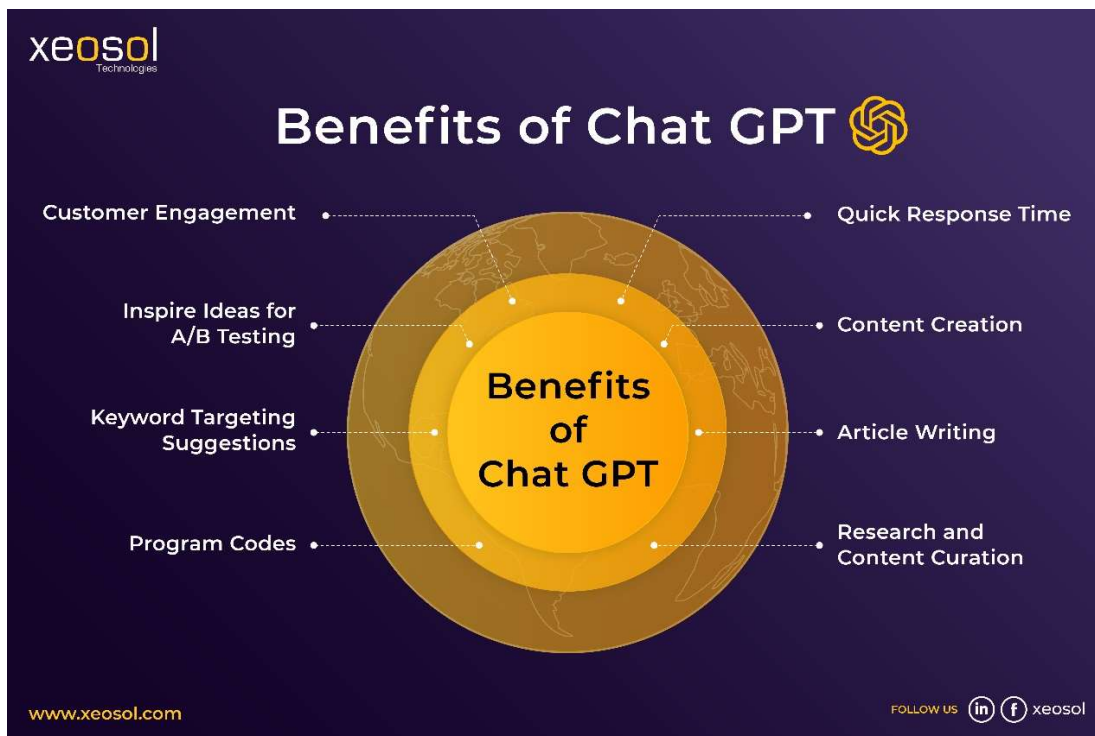
1. Introduction
2. Objective
3. Questionnaire
4. Determination of Sample Size
5. Test Of Equality Of Two Proportion
6. Test of Randomness
7. Selection of Random Sample
8. Chi-Square Test
9. Paired t-test
10. Anova
11. Graphical Representation
12. Conclusion

- **Introduction to chat GPT:-**

ChatGPT was created by a San Francisco-based startup called [OpenAI](#). When OpenAI was founded in late 2015, it received \$1 billion in funding from various investors. At the time, it was designed to operate as a non-profit AI research firm. By 2019, however, the company transitioned to a for-profit model. This allowed OpenAI to accept investments from venture capitalist firms and offer larger employee incentives.

ChatGPT enables users to "converse" with it in a way that mimics natural conversation. As a user, you can ask questions or make requests in the form of prompts, and ChatGPT will respond. The intuitive, easy-to-use, and free tool has already gained popularity as an alternative to traditional search engines and as a tool for AI writing, among Chat GPT, GPT is short for [generative pre-trained transformer](#). In the field of AI, *training* other things.

The "GPT" in ChatGPT refers to the process of teaching a computer system to recognize patterns and make decisions based on input data, much like how a teacher gives information to their students, then tests their understanding of that information.



- **Methods:-**

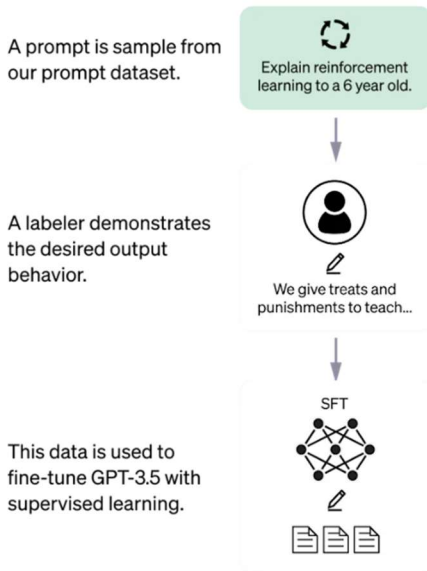
We trained this model using Reinforcement Learning from Human Feedback (RLHF), using the same methods as [InstructGPT](#), but with slight differences in the data collection setup. We trained an initial model using supervised fine-tuning: human AI trainers provided

conversations in which they played both sides—the user and an AI assistant. We gave the trainers access to model-written suggestions to help them compose their responses. We mixed this new dialogue dataset with the InstructGPT dataset, which we transformed into a dialogue format.

To create a reward model for reinforcement learning, we needed to collect comparison data, which consisted of two or more model responses ranked by quality. To collect this data, we took conversations that AI trainers had with the chatbot. We randomly selected a model-written message, sampled several alternative completions, and had AI trainers rank them. Using these reward models, we can fine-tune the model using [Proximal Policy Optimization](#). We performed several iterations of this process.

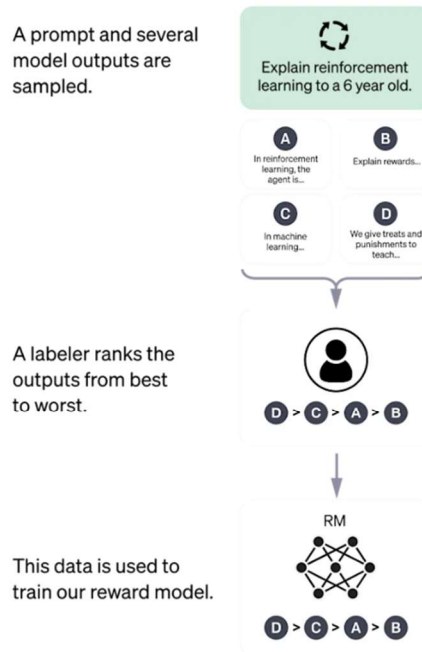
### Step 1

Collect demonstration data and train a supervised policy.



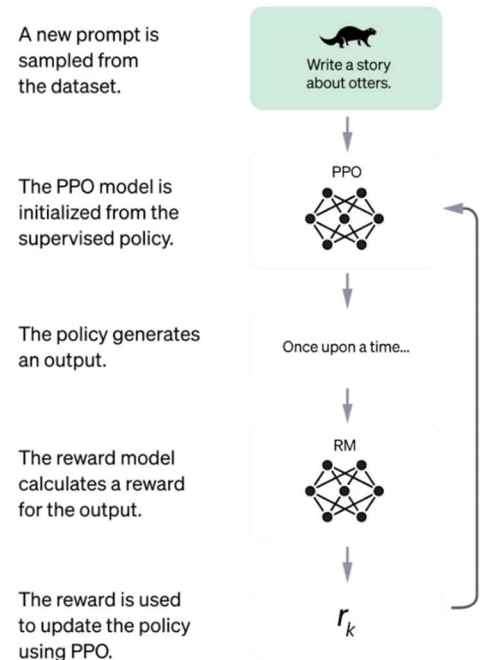
### Step 2

Collect comparison data and train a reward model.



### Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.



## • What can you do with data in ChatGPT?

When analyzing data with ChatGPT, you can create static and interactive tables and charts from your uploaded data.

- ChatGPT will automatically create an interactive table view, allowing you to scroll through your data and view all of your rows and columns.
- After uploading a file, ChatGPT can determine the ideal chart type for the dataset, or you can specify one of our supported chart types in your prompt.
- You can customize the graphics of your interactive charts and create summaries explaining your findings.



- **What file types are supported?**

ChatGPT can analyze data uploaded in a variety of file formats, including:-

1. Excel (.xls / .xlsx),
2. Comma-separated values (.csv),
3. PDF (.pdf),
4. JSON.

After the above summary on Chat GPT, We decided to check the Awareness on Chat GPT Among the T.Y. B.Sc. Students of Z.B. Patil College, Dhule.

- **Objectives:-**

To Study the Awareness on Chat GPT in T.Y.Bsc Students of Z.B.Patil College.

- Area Under Consideration:-T.Y.Bsc Students of Z.B.Patil College.
- Duration of Project:-December2024-February2025.
- Population Size:-Total N=226
- Sampling Technique:-Simple Random Sampling Without Replacement.
- Sample Sizes:-
  1. Group A:  
N1=110
  2. Group B:  
N2=116
- Data Type:-Primary Data
- Method of Collecting Data:-Questionnaire

• **Questionnaire:-**

1. What is your name :-
2. Gender :-
3. Subject Of Graduation:-
4. Which Search Engine Do you use The Most?
5. What is Full Of Chat GPT?
6. Who Invented Chat GPT?
7. In Which Year Chat GPT was Founded?
8. Is Chat GPT Free to use?
9. How Often Do You search Things?
10. How do you Get to Know about Chat GPT
11. How much time do you spend on Chat GPT
12. In Which Field you use chat GPT Most
13. How Chat GPT Is useful in your Subject?
14. Before Chat GPT Which search engine have you used most?
15. Does Chat GPT maintain Accuracy in giving answers to particular question
16. Do you think Chat GPT can replace Google In Future?
17. Feedback on Chat GPT

- **Determination of Sample Size:-**

Here we assume that,

Total Population=N=226

Desire margin of error=d=0.2

$\alpha=0.05$

$$n = \frac{N * P_i * (1 - P_i) * Z^2 \alpha / 2}{P_i * (1 - P_i) * Z^2 \alpha / 2 + (N - 1) * d^2}$$

1.Group A:-

N1=110

Pi=0.48

Qi=0.52

$$n = \frac{226 * 0.48 * 0.52 * 3.814}{(0.48 * 0.52 * 3.814) + (225 * 0.04)}$$

n=21.61

2.Group B:-

N2=116

Pi=0.53

Qi=0.47

$$n = \frac{226 * 0.53 * 0.47 * 3.814}{(0.53 * 0.47 * 3.814) + (225 * 0.04)}$$

n=21.57

- **Test Of Equality of Two Proportion:-**

A Test of Equality of Two Proportions is used to determine if there is a significant difference between the proportions of two independent groups. It is commonly used in situations where you want to compare the success rates or proportions of two groups.

Null Hypothesis (Ho): The two proportions are equal.

$$H_0: p_1 = p_2$$

Alternative Hypothesis (H1): The two proportions are not equal (two-tailed test).

$$H_1: p_1 \neq p_2$$

for the one-tailed test:

$$H_1: p_1 > p_2 \text{ or } H_1: p_1 < p_2$$

Calculate the sample proportions:

Sample Proportion for group A:

$$\widehat{p}_1 = \frac{x_1}{n_1} = 0.6818$$

Sample Proportion for group B:

$$\widehat{p}_2 = \frac{x_2}{n_2} = 0.4090$$

Where:

- $x_1$  is the number of Successes in group A,
- $n_1$  is the sample size of group A,
- $x_2$  is the number of successes in group B,
- $n_2$  is the sample size of group B.

To Calculate the pooled sample proportion: If the null hypothesis assumes that the two proportions are equal, you can combine the successes and sample sizes from both groups to form a pooled proportion:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \hat{p} = \frac{15 + 9}{22 + 22}$$
$$\hat{p} = 0.5454$$

To Compute the test statistic (Z): The test statistic is given by:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$Z = 1.817$$

If we use significance level of  $\alpha = 0.05$

$$Z_{\text{tab}} = 1.96$$

Make a decision, This means we Accept Null Hypothesis  $H_0$

Conclusion:-

There is No Significant Difference Between The proportion of Two Groups i.e. they are equal.

- **TEST OF RANDOMNESS:-**

$H_0$  = Arrangement of Y and N is random.

$H_1$  = Arrangement of Y and N is not random.

To test above hypothesis first we find the value of  $r$  i.e. no. of runs.

Y: No. of student who are aware about Chat GPT.

N: No. of student who are not aware about Chat GPT.

The sample is:-

YNNYYYYNNYYNNYYNNYYYYNNYYYYYYYYYYYYNNNNYY  
YNNYNNYYYYNNYYNNNNYYNNYYYYYYYYNNNNYYNY  
YYNNYYYYNNYYNNYYYYYYYYNNNNNNYYYYNNNNYYYY  
YYNNYYNNNNNNYYYYNNNNNN

Sample size  $150 > 20$

So we will apply normal approximation

$r = \text{no. of runs} = 59$

$n_1$  = No. of student who think that effect of Chat GPT on their study.

$n_2$  = No. of student who think that effect of Chat GPT is not on their study.

Test of statistics:-

$$Z_{cal} = \frac{(r - E(r))}{\sqrt{Var(r)}}$$

Consider,

$$E(r) = 1 + \frac{(2 * n_1 * n_2)}{n_1 + n_2}$$

$$E(r) = 68.32$$

Now,

$$Var(r) = \frac{2 * n_1 * n_2 * (2 * n_1 * n_2 - n_1 - n_2)}{(n_1 + n_2)^2 * (n_1 + n_2 - 1)}$$



$$\text{Var}(r) = 29.9641$$

$$Z = -1.7026$$

$$|Z| = 1.7026$$

$$\text{At } \alpha = 0.05, \quad Z_{\text{tab}} = Z_{\alpha/2}$$

$$Z_{\alpha/2} = 1.96$$

$$|Z| < Z_{\alpha/2}$$

$$1.7026 < 1.96$$

Accept  $H_0$ .

**Conclusion:-** Arrangement of Y & N is Random.

- **SELECTION OF RANDOM SAMPLE:-**

1) For Group A:-

```
>a =1:110
```

```
>sample(a,22)
```

```
[1] 12 5 10 1 22 14 11 18 17 13 3 15 7 2 9 4 20 21  
16 19 6 8
```

2) For Group B:-

```
>a =1:116
```

```
>sample(a,22)
```

```
[1] 19 20 9 10 18 5 6 7 15 3 1 17 14 16 21 11 22 12  
4 13 2 8
```

- **Chi - Square Test:-**

$H_0$  : Chat GPT Awareness is independent on different Group.

$H_1$  : Chat GPT Awareness is not independent on different Group.

Effect	Faculty		Total
	Group A	Group B	
Aware	15	9	24
Not Aware	7	13	20
Total	22	22	44

• **Expected Frequency Table:-**

Effect	Faculty		Total
	Group A	Group B	
Aware	12	12	24
Not Aware	10	10	20
Total	22	22	44

O <sub>i</sub>	E <sub>i</sub>	(O <sub>i</sub> -E <sub>i</sub> ) <sup>2</sup> /E <sub>i</sub>
15	12	0.75
9	12	0.75
7	10	0.9
13	10	0.9
		3.3

Test Statistics:-

$\chi^2_{\text{cal}} =$	3.3
$\chi^2_{\text{tab}} =$	7.81

i.e.  $\chi^2_{\text{cal}} < \chi^2_{\text{tab}}$

then  $H_0$  is Accept.

**Conclusion:-** Awareness on Chat GPT is independent on different Group.

- **Paired - t Test:-**

**R-Coding:-****# For Group A:-**

```
> # H0:  $\mu_1 = \mu_2$  Vs H1:  $\mu_1 \neq \mu_2$ 
```

```
>x=c(7,6,1,2,19,12,17,18,10,8,21,20,4,16,14,11,13,5,22,9,15,3);
```

```
>y=c(22,1,18,20,6,3,21,13,4,5,10,12,8,7,2,14,15,16,19,11,9,17);
```

```
> n=length(x);n;
```

```
[1] 22
```

```
> d = x-y ; d;
```

```
[1] -15 5 -17 -18 13 9 -4 5 6 3 11 8 -4 9 12 -3 -2 -11 3  
-2 6 -14
```

```
> dbar=sum(d)/n;dbar;
```

```
[1] 0
```

```
> var = var(d);var;
```

```
[1] 95.42857
```

```
> tcal = dbar/sqrt(var/n); tcal;
```

```
[1] 0
```

```
> ttab = qt(0.05,21); ttab;
```

```
[1] -1.720743
```

```
> abs (ttab);
```

```
[1] 1.720743
```

```
> # tcal < ttab
```

```
> # Accept Ho.
```

```
> # Conclusion:- There is no significant difference between  
two mean.
```

### **# For Group B:-**

```
> #  $H_0: \mu_1 = \mu_2$  Vs  $H_1: \mu_1 \neq \mu_2$ 
```

```
>x=c(18,12,14,9,21,8,4,2,3,15,20,16,19,10,6,7,11,22,13,1,5,  
17);
```

```
>y=c(7,12,15,8,3,4,13,18,17,22,16,20,1,21,10,6,5,9,11,19,1  
4,2);
```

```
> n=length(x);n;
```

```
[1] 22
```

```
> d = x-y ; d;
```

```
[1] 11 0 -1 1 18 4 -9 -16 -14 -7 4 -4 18 -11 -4 1 6 13  
2 -18 -9 15
```

```
> dbar=sum(d)/n;dbar;
```

```
[1] 0
```

```
> var = var(d);var;
```

```
[1] 113.2381
```

```
> tcal = dbar/sqrt(var/n); tcal;
```

```
[1] 0
```

```
> ttab = qt(0.05,12); ttab;
```

```
[1] -1.720743
```

```
> abs (ttab);
```

```
[1] 1.720743
```

```
> # tcal < ttab
```

```
> # Accept Ho.
```

```
> # Conclusion:- There is no significant difference between  
two mean.
```

- **ANOVA:-**

T.Y.Bsc	Faculty	
	Group A	Group B
Aware	6	11
Not Aware	7	4
Total	13	15

$H_{01}$ : Class do not differ significantly.

$H_{02}$ : Faculty do not differ significantly.

### **Anova: Two-Factor Without Replication**

<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>
Row 1	2	17	8.5	12.5
Row 2	2	11	5.5	4.5
Column 1	2	13	6.5	0.5
Column 2	2	15	7.5	24.5

ANOVA						
<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Rows	9	1	9	0.5625	0.590334	161.4476
Columns	1	1	1	0.0625	0.844042	161.4476
Error	16	1	16			
Total	26	3				

#### **1) For Rows:-**

**Here,  $F - \text{cal} < F - \text{crit}$**

**Then, Accept  $H_{01}$  for treatment at 5% l.o.s.**

**Conclusion:- Class do not differ significantly.**



## 2) For Columns:-

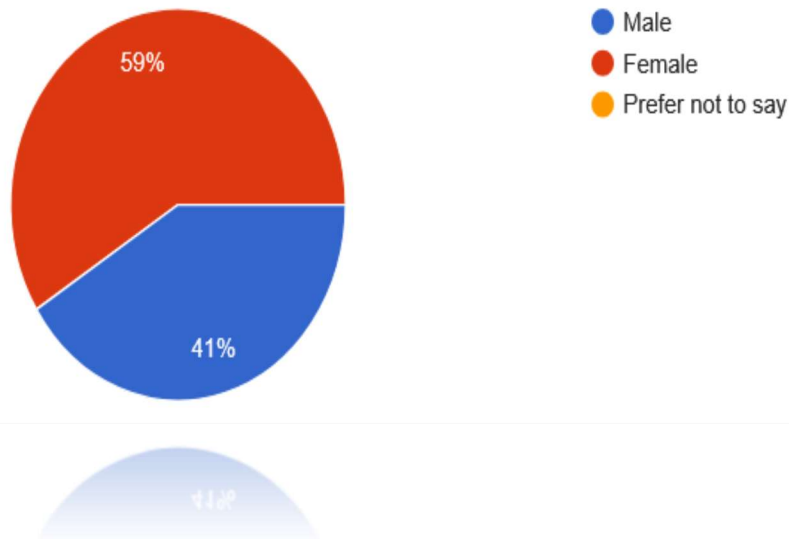
Here,  $F - \text{cal} < F - \text{crit}$

Then, Accept  $H_0$  for treatment at 5% l.o.s.

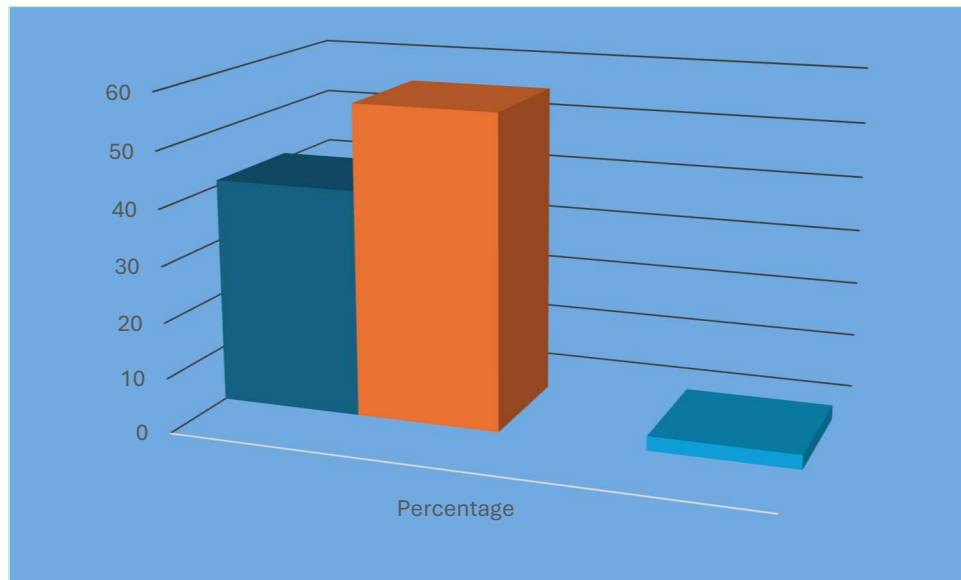
**Conclusion:-** Faculty do not differ significantly.

- **GRAPHICAL REPRESENTATION:-**

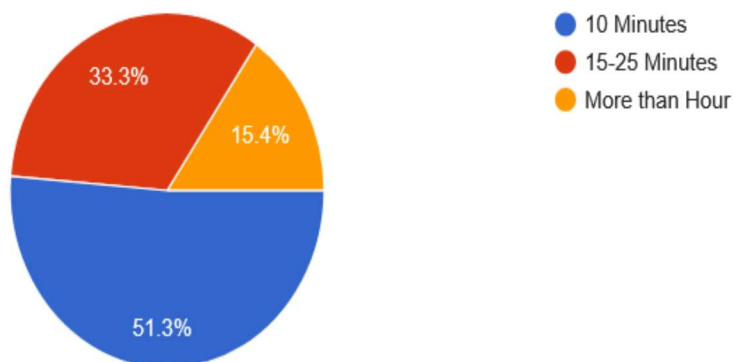
### **1)Gender wise Chat GPT Users:-**



### **2)Most Used Search Engine :-**



### **3)Time Spend On Chat GPT:-**



## • **Conclusion:-**

- **We know that, the analysis of awareness on chat GPT. The applying statistical tool say Proportion Test, Paired t-test, Chi-square test, ANOVA & Graphical representation.**

- **From Proportion Test:-**

**There is No Significant Difference Between The proportion of Two Groups i.e. they are equal.**

- **From test of randomness:-**

**From test of randomness (Run test) we conclude that, our sample is random.**

- **From Chi-square test:-**

**We conclude that, awareness on chat GPT is independent of different Groups.**

- **From Paired t-test:-**

**There is no significant differ between two means.**

**Hence there is no effect of Chat GPT on student study.**

- **From ANOVA:-**

**Both Group wise Chat GPT users do not differ significantly.**

➤ **From Graphical Representation:**

- 1. From first graph we conclude that there are more female users of Chat GPT than males.**
- 2. From second graph we conclude that there are more users of Chat GPT than any other Search Engine.**
- 3. From third graph we conclude that there is average 10 minutes times spend on Chat GPT.**