Laptop Usage Patterns -A Statistical Analysis

MA2540: APPLIED STATISTICS

Group Members

GUNTI LAHARI	AI22BTECH11008
JANAPAREDDY HIMA CHANDH	AI22BTECH11009
KONDAPARTHY ANURAGA CHANDAN	AI22BTECH11011
RUVVA SURAJ KUMAR	AI22BTECH11022
SINGA DIVIJA REDDY	AI22BTECH11026

Prof. Dr. Sameen Naqvi

2nd May, 2024

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

INTRODUCTION:

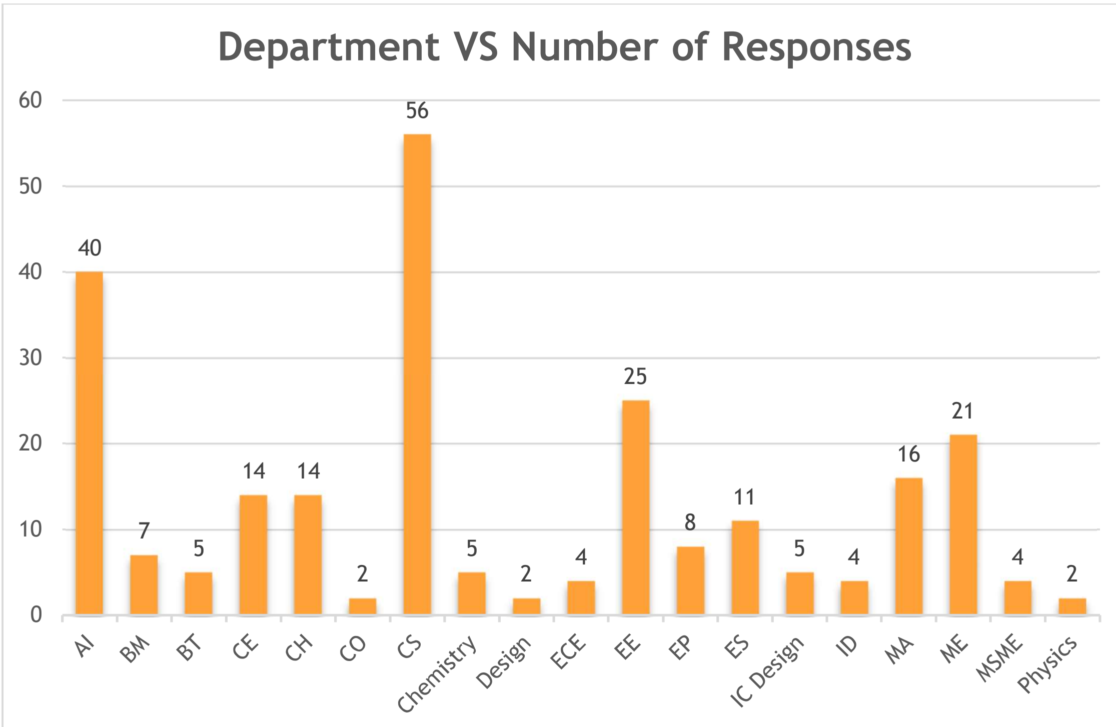
This report provides a Statistical Analysis of laptop usage patterns among students of IIT HYDERABAD. By gathering data on weekly screentime, types of usage with the laptops, preferred operating system, preferred modes of purchasing and which type of features they focus while getting a new laptop (by asking them to rate each feature out of 5, how much they consider it); through a google forms link, we seek to understand students' laptops preferences. This analysis maybe helpful if some student is planning to buy a new laptop for his studies, or other works (gaming, designing, etc.)

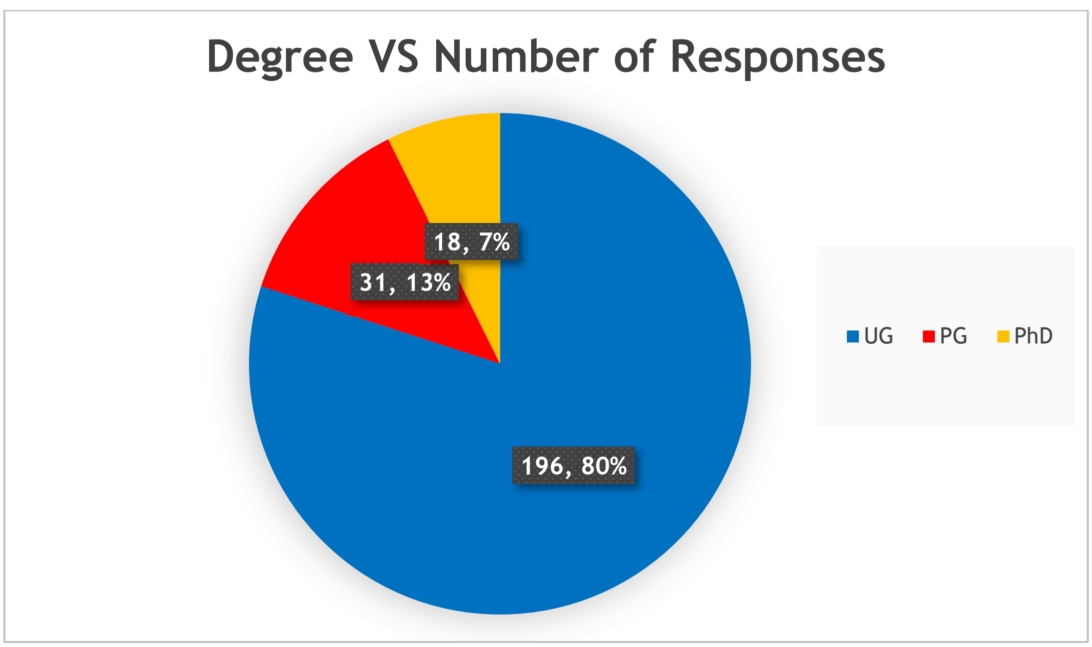
The Questions we have asked were:

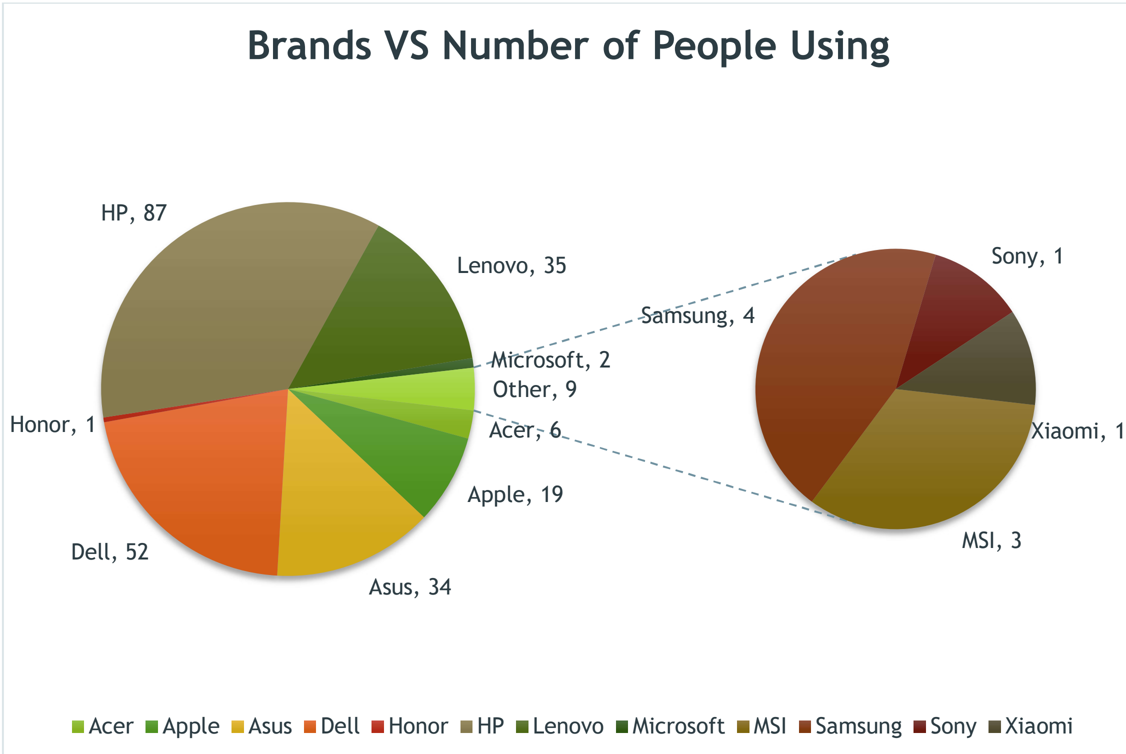
- **Which degree? (UG, PG, PhD)**
- **Department?**
- **Laptop brand?**
- **Weekly usage time?**
- **Primary use of laptop?**
- **Preferred OS?**
- **Mode of purchase?**
- **How important are the factors (like performance, battery life, price, etc.)? (with a rating out of 5)**

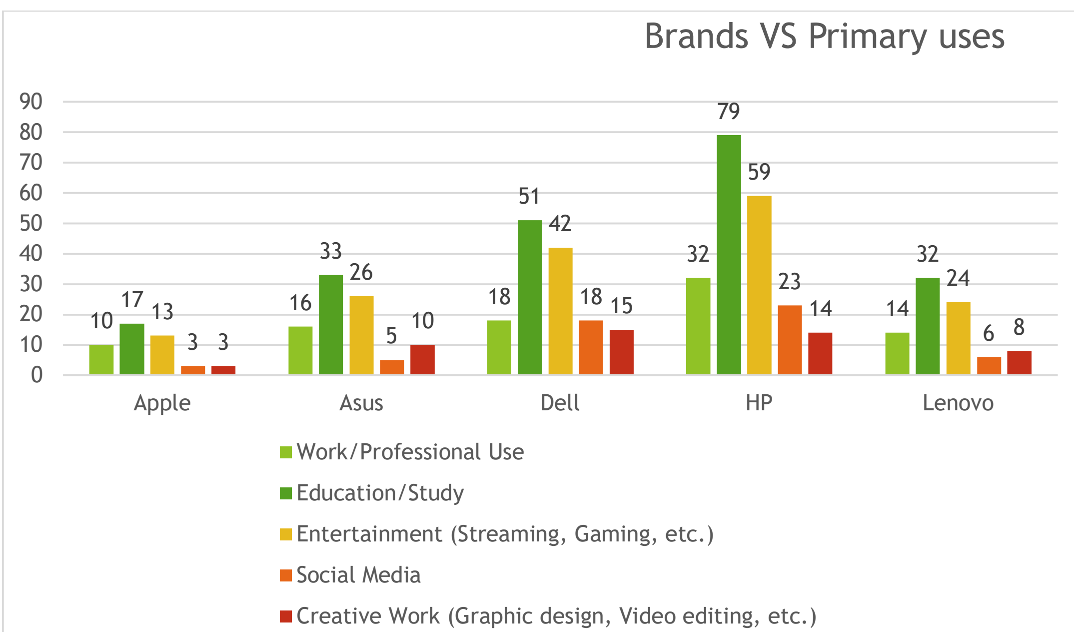
Our analysis is based only on the responses received through the google forms.
We have received 245 responses.

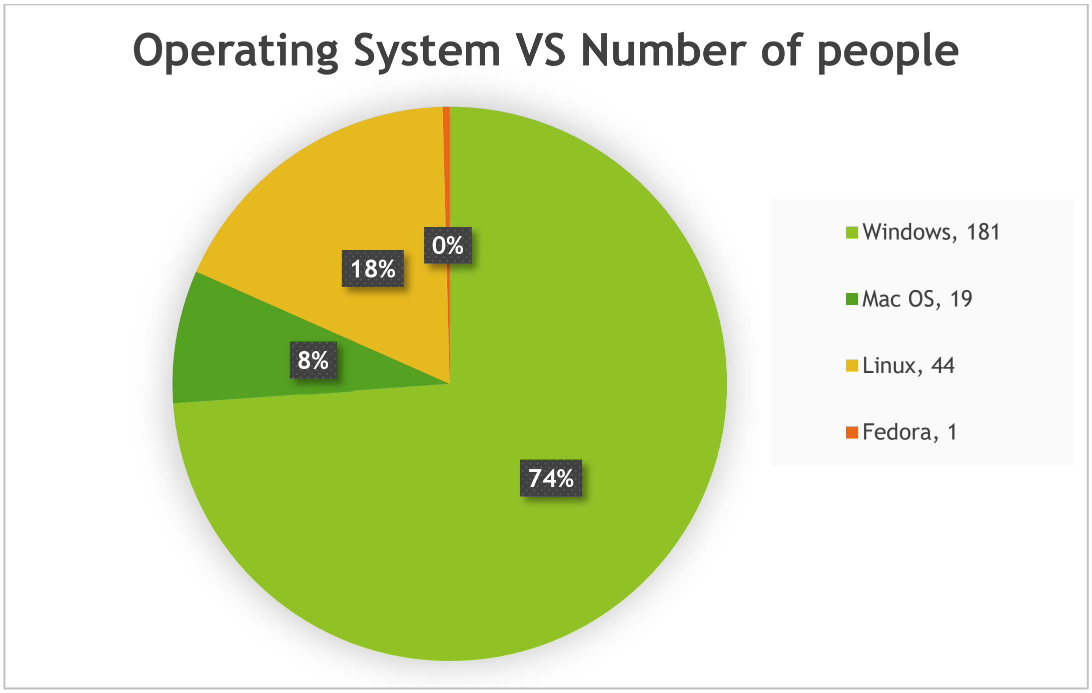
Data Visualization

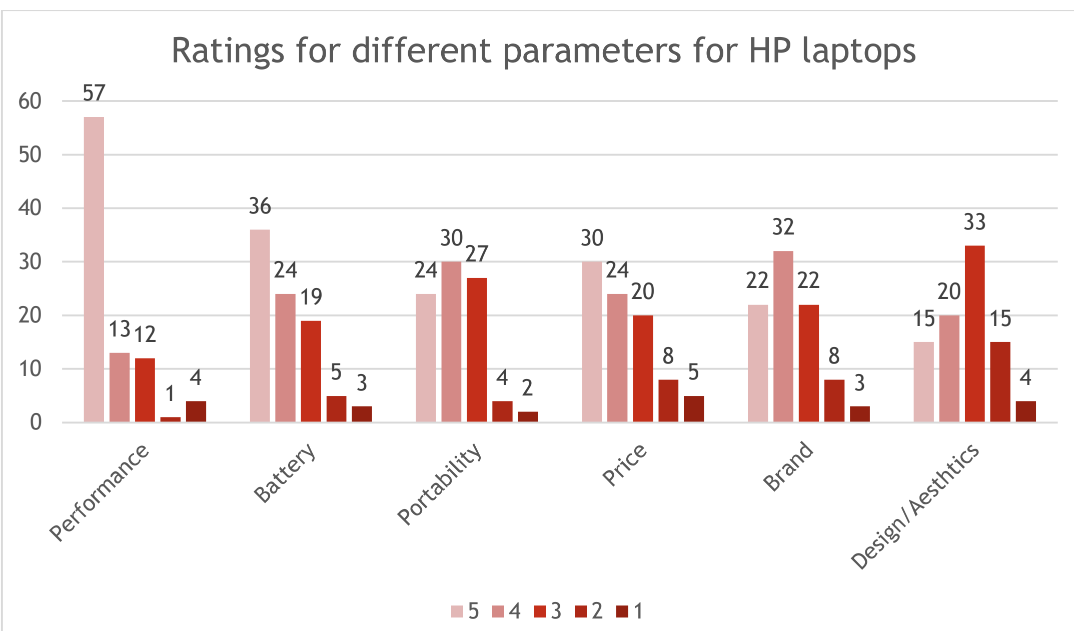


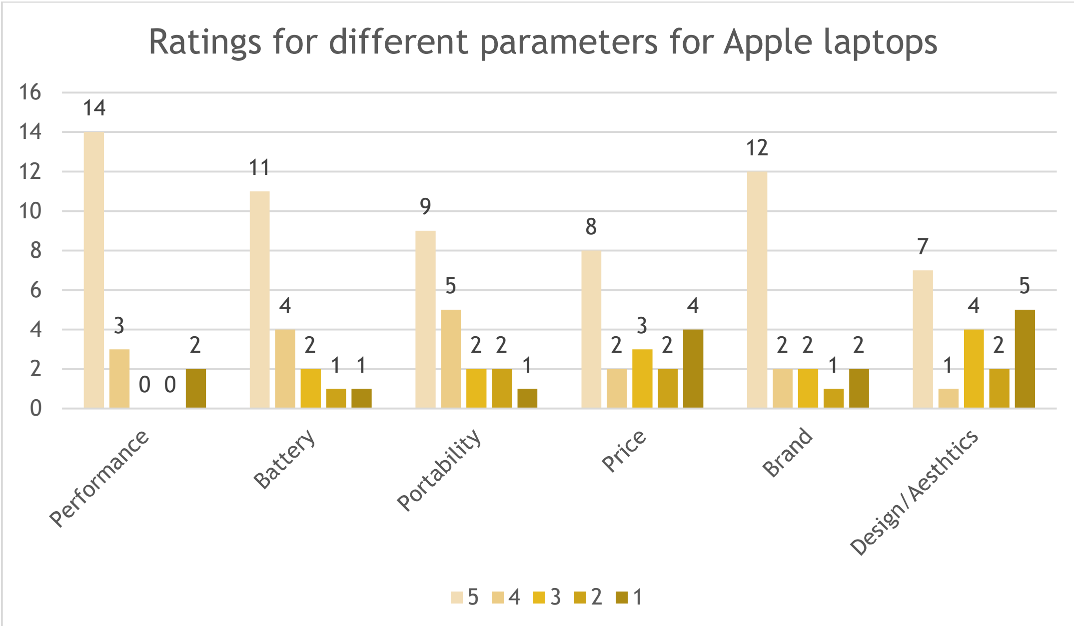


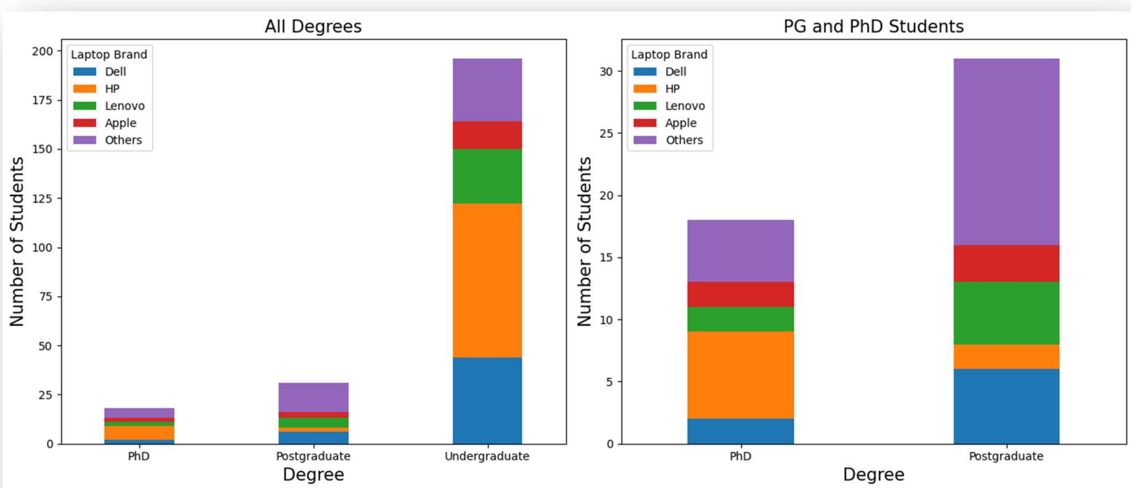

Brands VS Number of People Using
HP, 87
Lenovo, 35
Samsung, 4
Sony, 1
Microsoft, 2
Other, 9
Honor, 1
Acer, 6
Xiaomi, 1
Apple, 19
MSI, 3
Dell, 52
Asus, 34
Acer Apple Asus Dell Honor HP Lenovo Microsoft MSI Samsung Sony Xiaomi

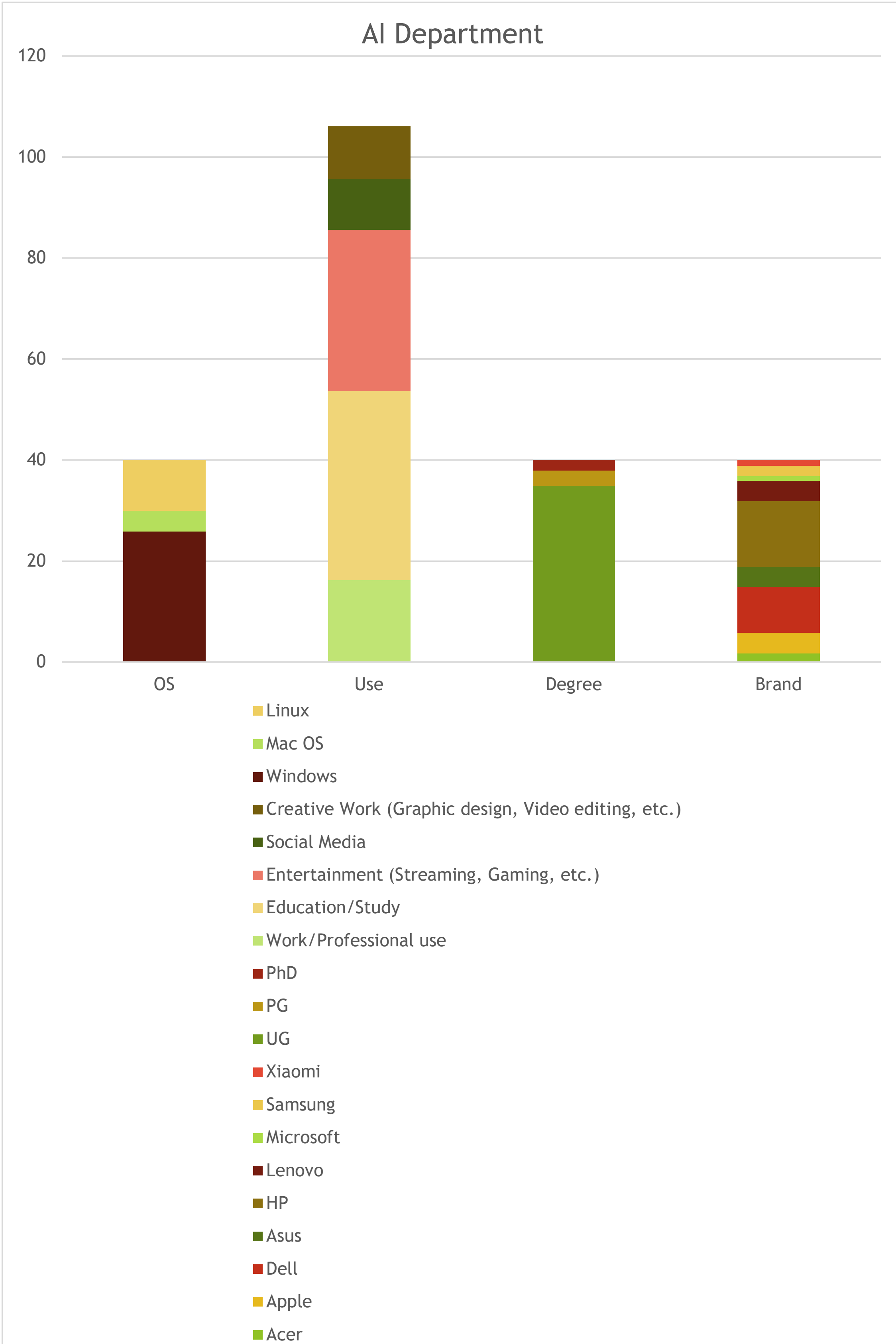

Brands VS Primary uses
90 80 70 60 50 40 30 20 10 0
Apple: 10 17 13 3 3
Asus: 16 33 26 5 10
Dell: 18 51 42 18 15
HP: 32 79 59 23 14
Lenovo: 14 32 24 6 8
Work/Professional Use
Education/Study
Entertainment (Streaming, Gaming, etc.)
Social Media
Creative Work (Graphic design, Video editing, etc.)


Operating System VS Number of people
18%
0%
8%
74%
Windows, 181
Mac OS, 19
Linux, 44
Fedora, 1


Ratings for different parameters for HP laptops
60
50
40
30
20
10
0
57
13
12
1
4
36
24
19
5
3
24
30
27
4
2
30
24
20
8
5
22
32
22
8
3
15
20
33
15
4
Performance
Battery
Portability
Price
Brand
Design/Aesthtics
5
4
3
2
1

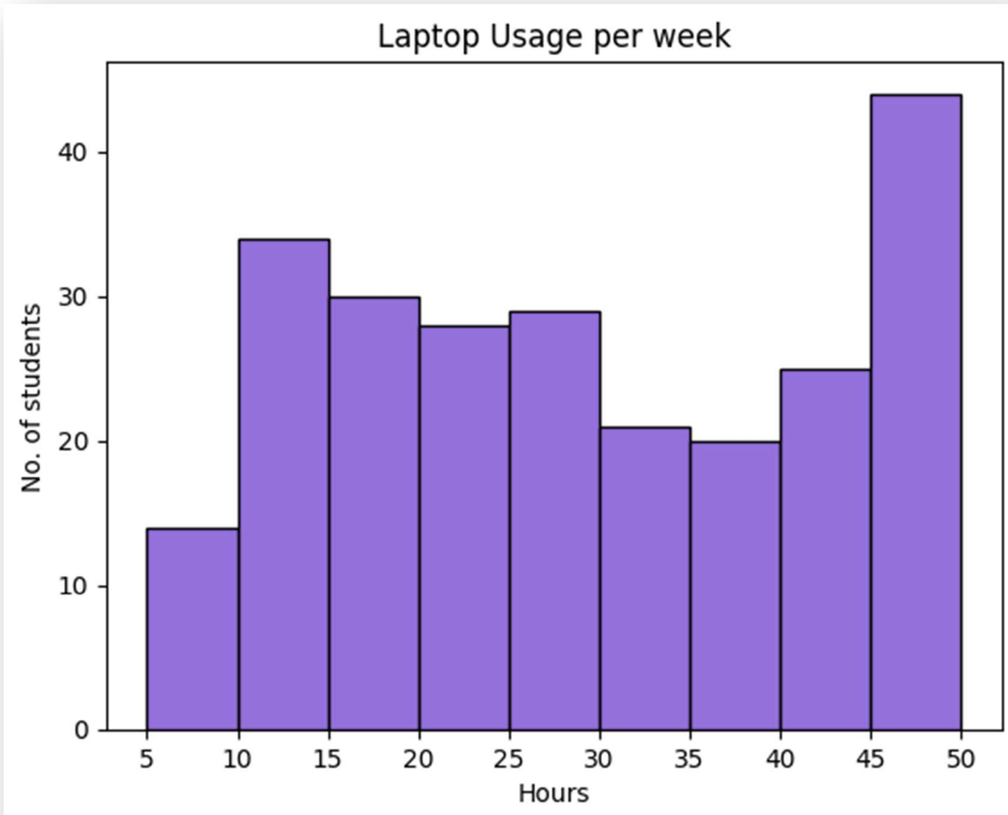

Ratings for different parameters for Apple laptops
16
14
12
10
8
6
4
2
0
14
3
0
0
2
11
4
2
1
1
9
5
2
2
1
8
2
3
2
4
12
2
2
1
2
7
1
4
2
5
Performance
Battery
Portability
Price
Brand
Design/Aesthtics
5
4
3
2
1


All Degrees
PG and PhD Students
Laptop Brand
Dell
HP
Lenovo
Apple
Others
Number of Students
Degree
PhD
Postgraduate
Undergraduate
0
25
50
75
100
125
150
175
200
5
10
15
20
25
30

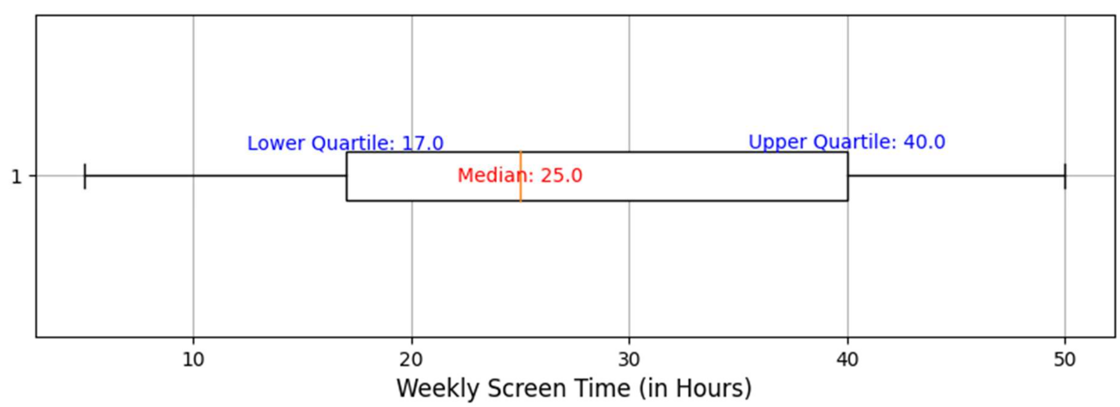

AI Department
120
100
80
60
40
20
0
OS
Use
Degree
Brand
Linux
Mac OS
Windows
Creative Work (Graphic design, Video editing, etc.)
Social Media
Entertainment (Streaming, Gaming, etc.)
Education/Study
Work/Professional use
PhD
PG
UG
Xiaomi
Samsung
Microsoft
Lenovo
HP
Asus
Dell
Apple
Acer

Analysis 1: Screen Time

Different ranges of screen time VS Number of Students



1. Considering all the students (without dividing into UG, PG, PhD)



$Mean = 28.1591$

$Median = 25$

$Standard\ Deviation = 13.5542$

$Lower\ quartile = 17$

$Upper\ quartile = 40$

- **25% of the students use their laptop between 17hrs to 25hrs per week.**

Confidence Interval Estimation

i. Mean (μ):

$$\overline{X} = 28.1591$$
$$S = 13.5542$$
$$95\%\ CI \Rightarrow \alpha = 0.05$$
$$n = 245$$
$$t_{\alpha/2,\ n-1} = 1.9697$$
$$CI \equiv \left[\overline{X} - t_{\alpha/2,\ n-1}\left(\frac{S}{\sqrt{n}}\right), \overline{X} + t_{\alpha/2,\ n-1}\left(\frac{S}{\sqrt{n}}\right)\right]$$
$$\Rightarrow CI \equiv [26.4535, 29.8647]$$

$\therefore$ *The Confidence Interval for mean μ of the weekly screen time is*

$[26.4535, 29.8647]$

ii. Variance (σ):

$$S = 13.5542 \Rightarrow S^2 = 183.7163$$
$$\alpha = 0.05$$
$$n = 245$$
$$a = \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,244} = 202.6272$$
$$b = \chi^2_{\alpha/2,n-1} = \chi^2_{0.025,244} = 289.1591$$
$$CI \equiv \left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right]$$
$$g \Rightarrow CI \equiv [155.0246, 221.2278]$$

$\therefore$ *The Confidence Interval for the variance σ^2 of the weekly screen time is* $[155.0246, 221.2278]$ *and for standard deviation σ is* $[12.45, 14.8737]$.

Hypothesis Testing

- **Is the mean of weekly screen time by student is more than 25hrs?**

Hypothesis:

$$\boldsymbol{\mu_0 = 25}$$

$$\mathbf{H_0: \mu \leq \mu_0} \qquad \mathbf{H_a: \mu > \mu_0}$$

$$\boldsymbol{\bar{X} = 28.1591}$$

$$\boldsymbol{S = 13.5542}$$

$$\boldsymbol{n = 245 \Rightarrow df = n - 1 = 244}$$

Test Statistic:

$$\boldsymbol{t^* = \frac{\bar{X} - \mu_0}{S/\sqrt{(n)}}}$$

$$\boldsymbol{\Rightarrow t^* = \frac{28.1591 - 25}{13.5542/\sqrt{244}} = 3.6408}$$

Rejection Region Approach:

$$\boldsymbol{\alpha = 0.01}$$

$$\text{Reject if } \boldsymbol{t^* \geq t_{\alpha, n-1}}$$

$$\boldsymbol{t_{\alpha, n-1} = 2.3417}$$

$$\therefore \text{Reject } \mathbf{H_0}$$

p-Value approach:

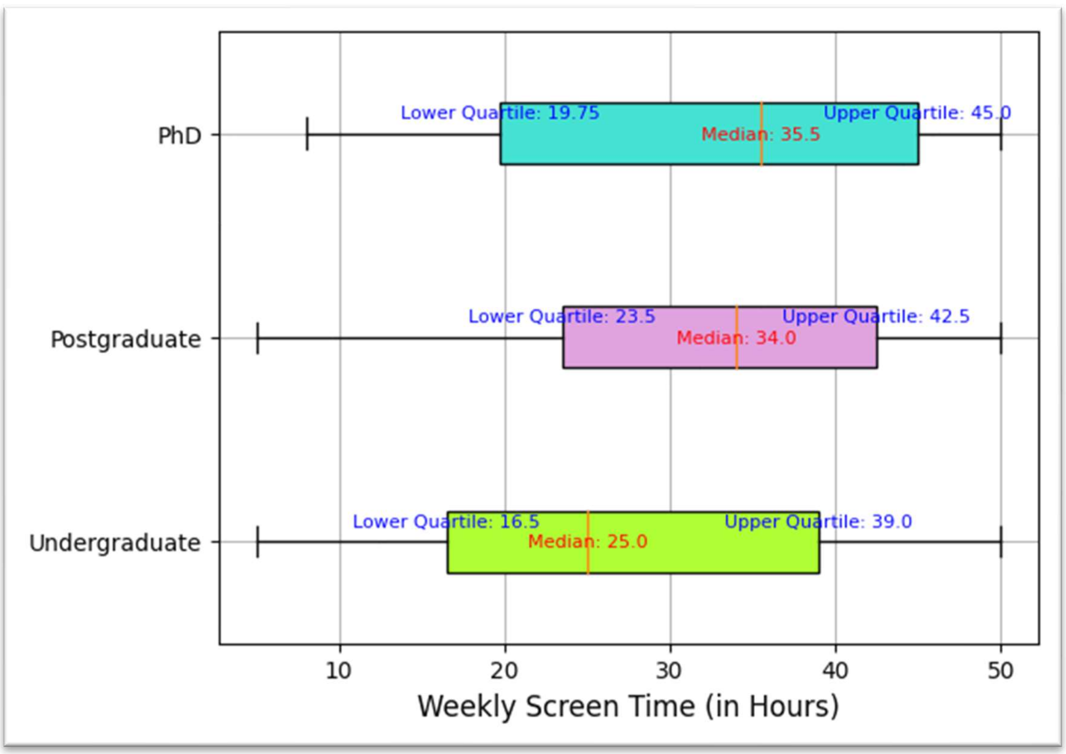
$$P(t \geq t^{*}) = 0.0003$$

$$0.0003 \leq 0.01$$

$$p \leq \alpha$$

$$\Rightarrow \text{Reject } H_0$$

2. Considering 2 samples: UG and PG



Confidence Interval Estimation

For difference of means of screen time of UG (μ_1) and PG (μ_2)

Mean of UGs' screen time: $\overline{X_1} = 27.1122$

Mean of PGs' screen time: $\overline{X_2} = 32.13$

Number of UGs in the sample: $n = 196$

Number of PGs in the sample: $m = 31$

$$S_1{}^2 = 13.2594$$

$$S_2{}^2 = 13.6326$$

$$\overline{X_1} - \overline{X_2} = -5.0178$$

$$\frac{S_1{}^2}{S_2{}^2} < 4 \Rightarrow \text{Two sample pooled interval}$$

$$S_p{}^2 = \text{pooled sample variance} = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

$$S_p{}^2 = \frac{(195 \times 13.2594) + (30 \times 13.6326)}{196 + 31 - 2} = 13.30916$$

$$S_p = 3.6482$$

$$t_{\alpha/2,m+n-2} = t_{0.025,225} = 1.97056$$

$$CI \equiv \left[\{\overline{X_1} - \overline{X_2}\} - t_{\alpha/2,n+m-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}, \{\overline{X_1} - \overline{X_2}\} + t_{\alpha/2,n+m-2} \times S_p\sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

$$\left[-5.0178 - 1.97056 \times 3.6482\sqrt{\frac{1}{196} + \frac{1}{31}}, -5.0178 + 1.97056 \times 3.6482\sqrt{\frac{1}{196} + \frac{1}{31}}\right]$$

$$CI \equiv [-6.40734, -3.6282]$$

We see that the interval is negative, so we can say that mean screen time of PGs is higher than UGs by $x \in [3.6282, 6.40734]$ with a 95% confidence level.

Hypothesis Testing

Is the difference of means of PG and PhD screen times greater than 0.5?

ASSUMPTION: population distributions are normal with unequal variances.

$$\overline{X_1} = \text{Average screen time of PGs} = 32.13$$
$$\overline{X_2} = \text{Average Screen time of PhD} = 32.7$$

$$\boldsymbol{S_1}^2 = \mathbf{185.85}$$
$$\boldsymbol{S_2}^2 = \mathbf{228.80}$$
$$\boldsymbol{n_1} = \mathbf{31}, \boldsymbol{n_2} = \mathbf{18}$$

Hypothesis (Right-tailed test):

$$\mathbf{H_0}: \boldsymbol{\mu_1} - \boldsymbol{\mu_2} \leq \mathbf{0.5} \qquad \mathbf{H_a}: \boldsymbol{\mu_1} - \boldsymbol{\mu_2} > \mathbf{0.5}$$

Test Statistic:

$$\boldsymbol{t}^* = \frac{(\overline{\boldsymbol{X_1}} - \overline{\boldsymbol{X_2}}) - \mathbf{0.5}}{\sqrt{\frac{\boldsymbol{S_1}^2}{\boldsymbol{n_1}} + \frac{\boldsymbol{S_2}^2}{\boldsymbol{n_2}}}} = \frac{-\mathbf{1.07}}{\mathbf{4.32}} = -\mathbf{0.24768}$$

$$\text{degrees of freedom } \boldsymbol{df} = \frac{\left(\frac{\boldsymbol{S_1}^2}{\boldsymbol{n_1}} + \frac{\boldsymbol{S_2}^2}{\boldsymbol{n_2}}\right)^2}{\frac{\left(\boldsymbol{S_1}^2/\boldsymbol{n_1}\right)^2}{\boldsymbol{n_1} - \mathbf{1}} + \frac{\left(\boldsymbol{S_2}^2/\boldsymbol{n_2}\right)^2}{\boldsymbol{n_2} - \mathbf{1}}} \sim \mathbf{33}$$

$$\boldsymbol{t_{\alpha,df}} = \boldsymbol{t_{0.05,33}} = \mathbf{1.69236}$$

Rejection Region Approach:

$$-\mathbf{0.24768} < \mathbf{1.69236} \Rightarrow \boldsymbol{t}^* < \boldsymbol{t_{\alpha,df}}$$

$$\therefore \text{We fail to reject } \mathbf{H_0}$$

Hypothesis Testing

Do the variances of UG (σ_1^2) and PG (σ_2^2) screen times differ significantly?

$$\text{UG: } \boldsymbol{S_1^2 = 175.81}$$

$$\text{PG: } \boldsymbol{S_2^2 = 185.85}$$

$$\boldsymbol{n_1 = 196} \text{ (UG) }, \quad \boldsymbol{n_2 = 31} \text{ (PG)}$$

$$\boldsymbol{\alpha = 0.05}$$

Hypothesis (Two-tailed test):

$$\mathbf{H_0}: \boldsymbol{\sigma_1}^2 = \boldsymbol{\sigma_2}^2 \qquad \mathbf{H_a}: \boldsymbol{\sigma_1}^2 \neq \boldsymbol{\sigma_2}^2$$

Test Statistic:

$$\boldsymbol{F = \frac{S_1{}^2}{S_2{}^2} = 0.9459}$$

$$\boldsymbol{df_1 = n_1 - 1 = 195}$$

$$\boldsymbol{df_2 = n_2 - 1 = 30}$$

$$\boldsymbol{F_{\alpha/2, df_1, df_2} = F_{0.025,195,30} = 1.837}$$

$$\boldsymbol{F_{1-\alpha/2, df_1, df_2} = 0.609}$$

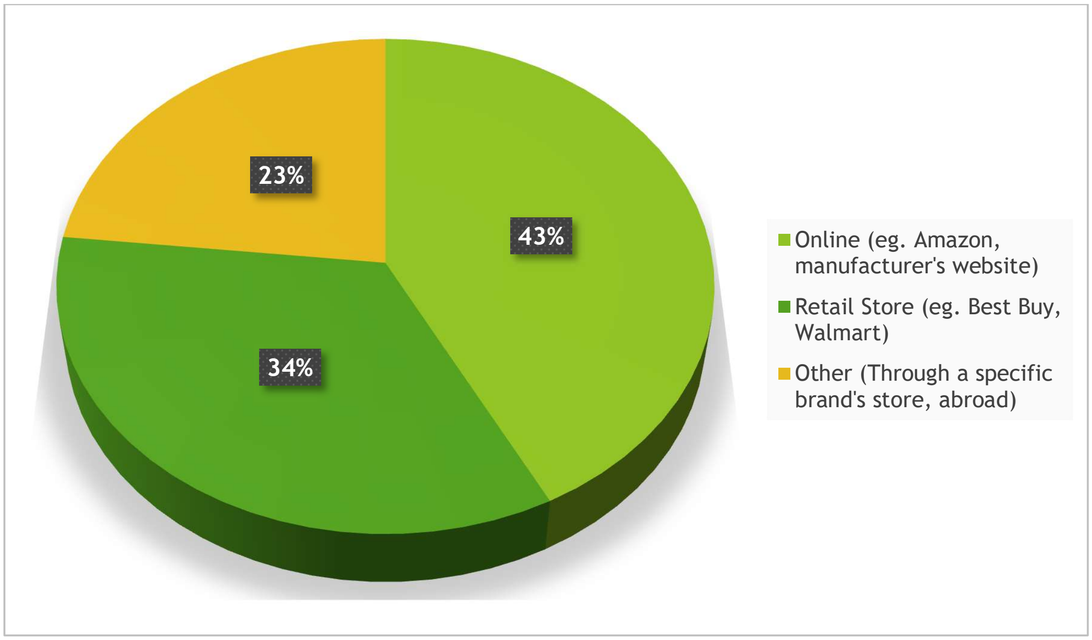
Rejection Region Approach:

$$\boldsymbol{0.9459 > 0.609} \text{ and } \boldsymbol{0.9459 < 1.837}$$

$$\Rightarrow \boldsymbol{F > F_{1-\alpha/2, df_1, df_2}} \text{ and } \boldsymbol{F < F_{\alpha/2, df_1, df_2}}$$

$$\therefore \text{ We fail to reject } \mathbf{H_0}$$

Analysis 2: Mode of Purchase



Confidence Interval Estimation

For proportion of students using online mode of purchase with 95% confidence

Total = 245

Students who chose online mode: 104

Confidence level = 95%

$$\alpha = 0.05$$

$$\hat{p} = \frac{104}{245} = 0.424$$

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

$$CI \equiv \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

$$\left[0.424 - 1.96\sqrt{\frac{0.424(0.576)}{245}}, 0.424 + 1.96\sqrt{\frac{0.424(0.576)}{245}}\right]$$

$$CI \equiv [0.3621, 0.4859]$$

$\therefore$ the confidence interval for the proportion of students choosing online mode of purchase is $[0.3621, 0.4859]$ with a 95% confidence level.

Hypothesis Testing

The proportion of students preferring online mode for purchasing laptop is less than HALF?

$\mathbf{H_0: \hat{p} \geq 0.5}$ $\mathbf{H_1: \hat{p} < 0.5}$

Total = 245

Students who chose online mode: 104

$$\text{significance level } \boldsymbol{\alpha} = \mathbf{0.01}$$

$$\widehat{\boldsymbol{p}} = \frac{\mathbf{104}}{\mathbf{245}} = \mathbf{0.424}$$

$$\boldsymbol{z}^* = \frac{\widehat{\boldsymbol{p}} - \boldsymbol{p_0}}{\sqrt{\frac{\boldsymbol{p_0}(\mathbf{1} - \boldsymbol{p_0})}{\boldsymbol{n}}}} = \frac{\mathbf{0.424 - 0.5}}{\sqrt{\frac{\mathbf{0.5 \times 0.5}}{\mathbf{245}}}} = \mathbf{-2.37}$$

$$\text{critical value } \boldsymbol{z_\alpha} = \mathbf{2.326}$$

$$\text{Reject } \mathrm{H_0} \text{ if } \boldsymbol{z}^* \leq \boldsymbol{z_\alpha}$$

$$\mathbf{-2.37 < 2.326} \Rightarrow \boldsymbol{z}^* < \boldsymbol{z_\alpha}$$

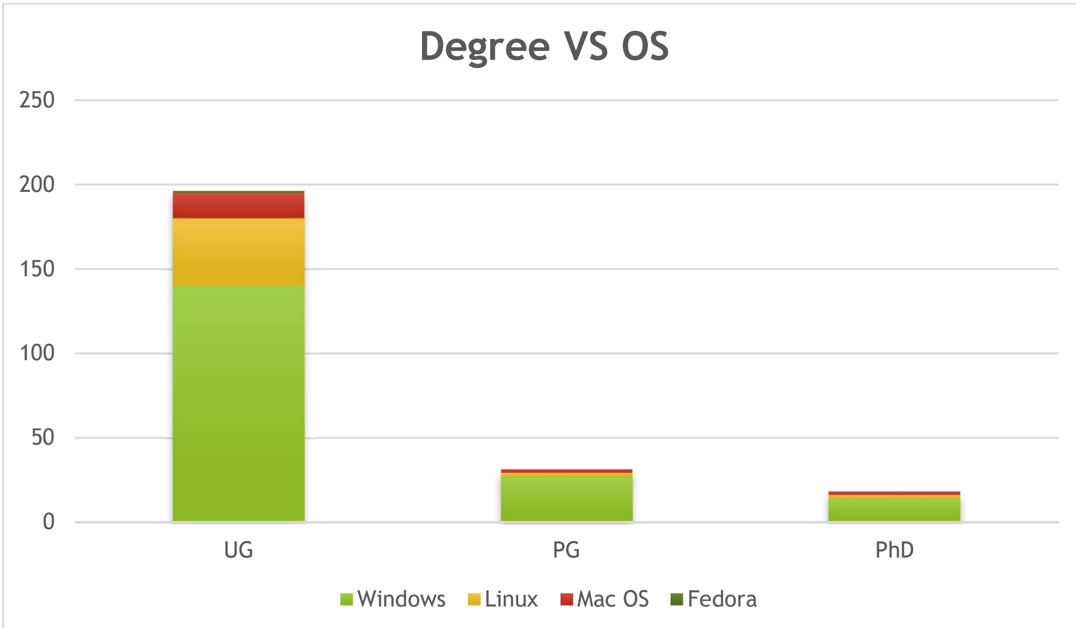
$$\therefore \text{ we reject } \boldsymbol{H_0}$$

Our hypothesis is correct.

We can even verify it from the Confidence Interval $[\mathbf{0.3621}, \mathbf{0.4859}]$ **where it is less than 0.5**

Analysis 3: Operating System

	Windows	Non-Windows
PG $(n_1 = \mathbf{31})$	**27**	**4**
PhD $(n_2 = 18)$	**14**	**4**



$$\boldsymbol{p_1} \rightarrow \text{ proportion of Windows users in PG}$$

$$\boldsymbol{p_2} \rightarrow \text{ proportion of Windows users in PhD}$$

Confidence Interval Estimation

Confidence Level = 95%

$$\boldsymbol{\alpha = 0.05}$$

$$\boldsymbol{CI \equiv \left[(\widehat{p_1} - \widehat{p_2}) - z_{\alpha/2} \times \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}},\right.}$$

$$\boldsymbol{\left.(\widehat{p_1} - \widehat{p_2}) + z_{\alpha/2} \times \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}\right]}$$

$$\widehat{p_1} = \frac{27}{31} = 0.870$$

$$\widehat{p_2} = \frac{14}{18} = 0.777$$

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

$$\text{CI for } \widehat{p_1} - \widehat{p_2} \equiv \left[0.093 - 1.96\sqrt{0.0036 + 0.0096}, 0.093 + 1.96\sqrt{0.0036 + 0.0096}\right]$$

$$\therefore \text{ The Confidence Interval for } \widehat{p_1} - \widehat{p_2} \text{ is } [-0.132, 0.318]$$

Hypothesis Testing

$$\widehat{p_1} = \frac{27}{31} = 0.870$$

$$\widehat{p_2} = \frac{14}{18} = 0.777$$

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

Hypothesis:

$$\mathbf{H_0: p_1 - p_2 = 0} \qquad \mathbf{H_a: p_1 - p_2 \neq 0}$$

Test Statistic:

$$z^* = \frac{(\widehat{p_1} - \widehat{p_2}) - p_0}{\sqrt{\frac{\widehat{\boldsymbol{p_1}}(\boldsymbol{1} - \widehat{\boldsymbol{p_1}})}{\boldsymbol{n_1}} + \frac{\widehat{\boldsymbol{p_2}}(\boldsymbol{1} - \widehat{\boldsymbol{p_2}})}{\boldsymbol{n_2}}}} = \frac{0.093}{0.115} = 0.808$$

Rejection Region Approach:

$$0.808 < 1.96$$

$$\Rightarrow z^* < z_\alpha/2$$

$$\therefore \text{ We fail to reject } \mathbf{H_0}$$

Analysis 4: Performance

$$\mu_1 = \text{Average performance of HP}$$

$$\mu_2 = \text{Average performance of Dell}$$

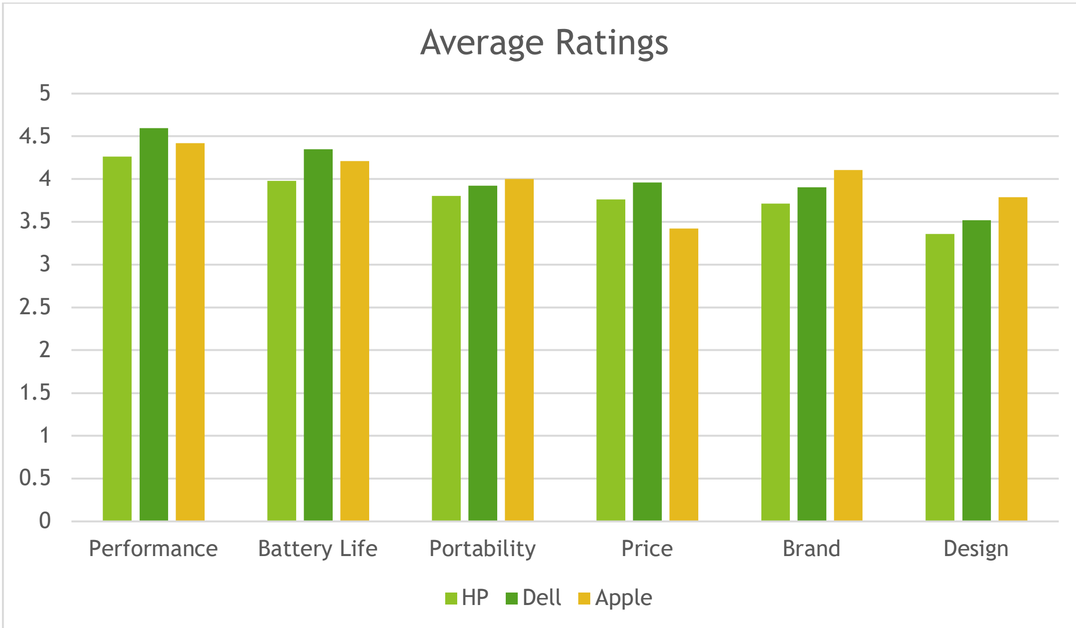
$$\overline{X_1} = 4.264$$

$$\overline{X_2} = 4.596$$

$$S_1{}^2 = 1.139$$

$$S_2{}^2 = 0.4023$$

$$n_1 = 87, n_2 = 52$$



Hypothesis Testing

Hypothesis:

$$\mathbf{H_0: \mu_1 - \mu_2 \geq 0.3} \qquad \mathbf{H_a: \mu_1 - \mu_2 < 0.3}$$

Test Statistic:

$$\boldsymbol{t^*} = \frac{(\overline{\boldsymbol{X_1}} - \overline{\boldsymbol{X_2}}) - 0.3}{S_p\sqrt{\frac{1}{\boldsymbol{n_1}} + \frac{1}{\boldsymbol{n_2}}}} = -\mathbf{0}.032$$

$$S_p = \sqrt{\frac{(\boldsymbol{n}_1 - \mathbf{1})\boldsymbol{S}_1^2 + (n_2 - \mathbf{1})\boldsymbol{S}_2^2}{\boldsymbol{n}_1 + n_2 - 2}} = 0.9299$$

$$df = n_1 + n_2 - 2 = 137$$
$$t_{\alpha,df} = 2.3538$$

Rejection Region Approach:

$$-0.032 \succ 2.3538 \Rightarrow t^* \succ t_\alpha, df$$

$$\therefore \text{ We fail to reject } \mathbf{H_0}$$

Analysis 5: Laptop Brand

Contingency Table of Degree and Laptop Brand

Original Frequencies

DEGREE	Brand			total
	ASUS	LENOVO	OTHER	
UG	20	28	148	196
PG	10	5	16	31
PhD	4	2	12	18
total	34	35	176	245

$$E_{ij} = \frac{(\text{Row Total}_i) \times (\text{Column Total}_j)}{\text{Grand Total}}$$

Expected Frequencies

DEGREE	Brand			total
	ASUS	LENOVO	OTHER	
UG	27.2	28	140.8	196
PG	4.3	4.4	22.3	31
PhD	2.5	2.6	12.9	18
total	34	35	176	245

Chi Square Test of Independence

$$\chi^{2*} = \sum_{\mathrm{i}=1}^{\mathrm{r}\times\mathrm{c}} \frac{(O_i - E_i)^2}{E_i}$$

where, r is number of rows in the Contingency Table.

c is number of columns in the Contingency Table.

$$\chi^{2*} = 12.7927$$

$$\alpha = 0.01$$

H_0**: There is no association between degree and laptop brand.**

H_a**: There is an association between degree and laptop brand.**

$$\mathbf{H_0}: \chi^{2*} \sim \boldsymbol{\chi}^2_{(r-1)(c-1)}$$

$$\boldsymbol{\chi}^2_{\boldsymbol{\alpha},4} = 13.2767$$

$$12.7927 < 13.2767 \ \Rightarrow \chi^{2*} < \boldsymbol{\chi}^2_{\boldsymbol{\alpha},04}$$

$\therefore$ We fail to reject $\mathbf{H_0}$

There is no significant evidence to suggest an association between degree and laptop brand.

....

CONTRIBUTIONS

G LAHARI ai22btech11008@iith.ac.in

- Project idea
- Confidence Interval Estimation and Hypothesis Testing
- Plotting graphs

J HIMA CHANDH ai22btech11009@iith.ac.in

- Data collection and cleaning
- Slides for presentation
- Python coding for box plots, pie charts

K ANURAGA CHANDAN ai22btech11011@iith.ac.in

- Preparing report
- Data visualization
- Slides for presentation

R SURAJ KUMAR ai22btech11022@iith.ac.in

- Slides for presentation
- Plotting graphs in Python
- Central tendencies

S DIVIJA ai22btech11026@iith.ac.in

- Confidence Interval Estimation and Hypothesis Testing
- Plotting graphs
- Project idea