MA4240

Laptop Specifications

Name	Roll Number
Siddhant Chandorkar	MA19BTECH11003
Haritha .R	AI20BTECH11010
Vijay Varma	AI20BTECH11012
Sachin Karumanchi	AI20BTECH11013
Manikanta Vallepu	AI20BTECH11014
Guguloth Hruday	MA20BTECH11020
Chintalapudi Abhiroop	AI20BTECH11005
Mohd Saad	MA20BTECH11009

Motivation

For many students laptops are becoming daily part of their life, before they join in the required stream they approach their respective seniors or batchmates regarding the specifications of a laptop that is comfortable to use, so for such students these analysis will help to buy the laptop according to their requirement.

DATA PROCESSING

Data Collected

We have collected 239 random samples of student's Laptop Specifications of IITH students

Columns in the Data collected:
- timestamp
- email
- stream
- brand
- price_range
- display
- Display_type
- ram_size
- cpu_clock_speed
- cpu
- cpu_model
- Graphic_card
- storage_type
- hdd_size
- ssd_size
- operating_sys
- avg_watch_time

Format of Samples Collected:

timestamp	email	stream	brand	price_range	display	display_type	ram_size	cpu_clock_speed	cpu	cpu_model	graphic_card	storage_type	hdd_size	ssd_size	operating_sys	avg_watch_time
3/29/2022 17:15:21	ep18btech11011@iith.ac.in	UG	Lenovo	Rs. 70,000 - Rs. 80,000	14 inch - 15 inch	Touch	8 GB	2 GHZ - 3 GHZ	Intel	Intel i7	None	HDD	500 GB	None	Windows	180
3/29/2022 17:15:29	ee21resch01003@iith.ac.in	PhD	HP	Rs. 50,000 - Rs. 60,000	13 inch - 14 inch	Non-Touch	16 GB	2 GHZ - 3 GHZ	Intel	Intel i5	Intel Graphics	SSD	None	512 GB	Windows	212
3/29/2022 17:16:12	ai21mtech13006@iith.ac.in	PG	HP	Rs. 40,000 - Rs. 50,000	14 inch - 15 inch	Non-Touch	4 GB	2 GHZ - 3 GHZ	Intel	Intel i3	None	HDD	500 GB	None	Windows	164
3/29/2022 17:16:50	ai20btech11022@iith.ac.in	UG	Apple	Rs. 90,000 and above	13 inch - 14 inch	Non-Touch	8 GB	2 GHZ - 3 GHZ	Apple	M1	None	SSD	None	256 GB	MacOS	469
3/29/2022 17:18:23	ns20mtech11004@iith.ac.in	PG	Lenovo	Rs. 60,000 - Rs. 70,000	14 inch - 15 inch	Non-Touch	8 GB	3 GHZ and above	AMD	AMD Ryzen 7	Nvidia Graphics	SSD	None	512 GB	Linux	413
3/29/2022 17:18:23	ch20btech11040@iith.ac.in	UG	Dell	Rs. 70,000 - Rs. 80,000	14 inch - 15 inch	Non-Touch	8 GB	3 GHZ and above	Intel	Intel i5	Nvidia Graphics	SSD	None	512 GB	Windows	324
3/29/2022 17:19:02	ch18btech11005@iith.ac.in	UG	HP	Rs. 50,000 - Rs. 60,000	14 inch - 15 inch	Non-Touch	8 GB	3 GHZ and above	Intel	Intel i7	Nvidia Graphics	HDD	1 TB	None	Windows, Linux	841
3/29/2022 17:19:12	ch20resch11003@iith.ac.in	PhD	Lenovo	Rs. 30,000 - Rs. 40,000	14 inch - 15 inch	Non-Touch	4 GB	1 GHZ - 2 GHZ	Intel	Intel i3	None	HDD	500 GB	None	Windows	647
3/29/2022 17:24:27	cs20btech11003@iith.ac.in	PhD	Dell	Rs. 60,000 - Rs. 70,000	14 inch - 15 inch	Non-Touch	8 GB	2 GHZ - 3 GHZ	Intel	Intel i7	AMD Radeon Graphics	HDD	1 TB	None	Windows, Linux	626
3/29/2022 17:24:53	me19btech11008@iith.ac.in	UG	HP	Rs. 90,000 and above	15 inch - 16 inch	Touch	16 GB	3 GHZ and above	Intel	Intel i7	Intel Graphics	HDD	500 GB	None	Windows	683
3/29/2022 17:25:25	ew21mtech11005@iith.ac.in	PG	HP	Rs. 40,000 - Rs. 50,000	14 inch - 15 inch	Non-Touch	4 GB	2 GHZ - 3 GHZ	Intel	Intel i5	AMD Radeon Graphics	HDD	1 TB	None	Windows	672
3/29/2022 17:25:34	ep19btech11007@iith.ac.in	UG	Dell	Rs. 60,000 - Rs. 70,000	14 inch - 15 inch	Non-Touch	16 GB	3 GHZ and above	Intel	Intel i5	Intel Graphics	SSD	None	512 GB	Windows	262
3/29/2022 17:26:22	cs20btech11012@iith.ac.in	UG	HP	Rs. 50,000 - Rs. 60,000	13 inch - 14 inch	Non-Touch	8 GB	1 GHZ - 2 GHZ	Intel	Intel i5	Intel Graphics	SSD	None	512 GB	Windows, Linux	524
3/29/2022 17:27:49	ch19btech11010@iith.ac.in	UG	HP	Rs. 70,000 - Rs. 80,000	14 inch - 15 inch	Non-Touch	8 GB	3 GHZ and above	Intel	Intel i5	Nvidia Graphics	Hybrid (Both SSD and HDD)	1 TB	256 GB	Windows	807
3/29/2022 17:28:47	me18btech11020@iith.ac.in	UG	Lenovo	Rs. 40,000 - Rs. 50,000	14 inch - 15 inch	Touch	8 GB	2 GHZ - 3 GHZ	Intel	Intel i3	Intel Graphics	HDD	500 GB	None	Linux	341

Data Processing

Our final data is prepared by following the steps below:

- Extracting Department and Joined Year of Student from their Email Address
- Converting Ram size to int data type
- Dropping Email Address and Time stamp

Data Processed

Format of Data after processing:

stream	brand	price_range	display	display_type	ram_size	cpu_clock_speed	cpu	cpu_model	graphic_card	storage_type	hdd_size	ssd_size	operating_sys	avg_watch_time	department	year_join
UG	Lenovo	Rs. 70,000 - Rs. 80,000	14 inch - 15 inch	Touch	8	2 GHZ - 3 GHZ	Intel	Intel i7	None	HDD	500 GB	None	Windows	180	ep	18
PhD	HP	Rs. 50,000 - Rs. 60,000	13 inch - 14 inch	Non-Touch	16	2 GHZ - 3 GHZ	Intel	Intel i5	Intel Graphics	SSD	None	512 GB	Windows	212	ee	21
PG	HP	Rs. 40,000 - Rs. 50,000	14 inch - 15 inch	Non-Touch	4	2 GHZ - 3 GHZ	Intel	Intel i3	None	HDD	500 GB	None	Windows	164	ai	21
UG	Apple	Rs. 90,000 and above	13 inch - 14 inch	Non-Touch	8	2 GHZ - 3 GHZ	Apple	M1	None	SSD	None	256 GB	MacOS	469	ai	20
PG	Lenovo	Rs. 60,000 - Rs. 70,000	14 inch - 15 inch	Non-Touch	8	3 GHZ and above	AMD	AMD Ryzen 7	Nvidia Graphics	SSD	None	512 GB	Linux	413	ns	20
UG	Dell	Rs. 70,000 - Rs. 80,000	14 inch - 15 inch	Non-Touch	8	3 GHZ and above	Intel	Intel i5	Nvidia Graphics	SSD	None	512 GB	Windows	324	ch	20
UG	HP	Rs. 50,000 - Rs. 60,000	14 inch - 15 inch	Non-Touch	8	3 GHZ and above	Intel	Intel i7	Nvidia Graphics	HDD	1 TB	None	Windows, Linux	841	ch	18
PhD	Lenovo	Rs. 30,000 - Rs. 40,000	14 inch - 15 inch	Non-Touch	4	1 GHZ - 2 GHZ	Intel	Intel i3	None	HDD	500 GB	None	Windows	647	ch	20
PhD	Dell	Rs. 60,000 - Rs. 70,000	14 inch - 15 inch	Non-Touch	8	2 GHZ - 3 GHZ	Intel	Intel i7	AMD Radeon Graphics	HDD	1 TB	None	Windows, Linux	626	cs	20
UG	HP	Rs. 90,000 and above	15 inch - 16 inch	Touch	16	3 GHZ and above	Intel	Intel i7	Intel Graphics	HDD	500 GB	None	Windows	683	me	19

DATA ANALYSIS AND VISUALISATION

Data used for Analysis

- ❖ Stream
- ❖ Department
- ❖ Year join
- ❖ Laptop Brand
- ❖ Operating System
- ❖ Display Type
- ❖ Display size
- ❖ Graphic Card
- ❖ Price Range

- ❖ RAM Size
- ❖ CPU
- ❖ CPU Model
- ❖ CPU Clock Speed
- ❖ Storage Type
- ❖ HDD Size
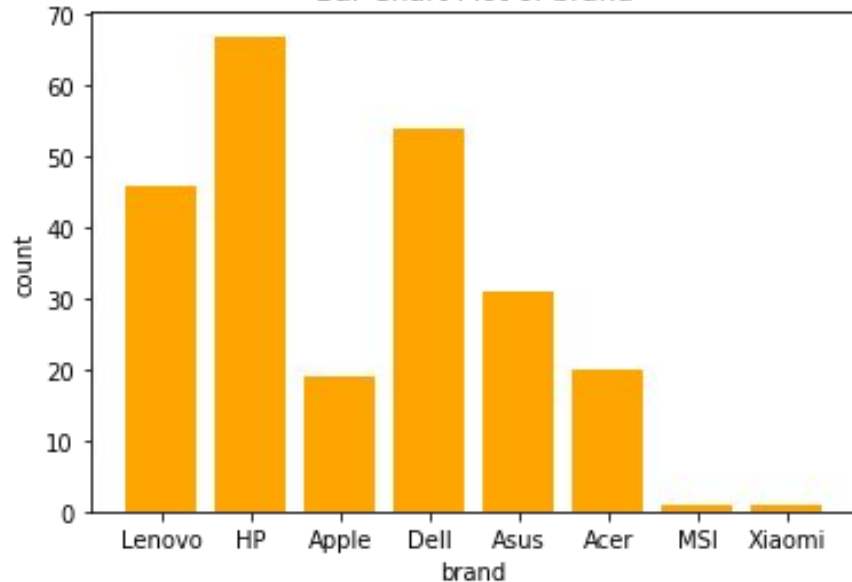- ❖ SSD Size
- ❖ Average Watch Time

❖ Regular categorical
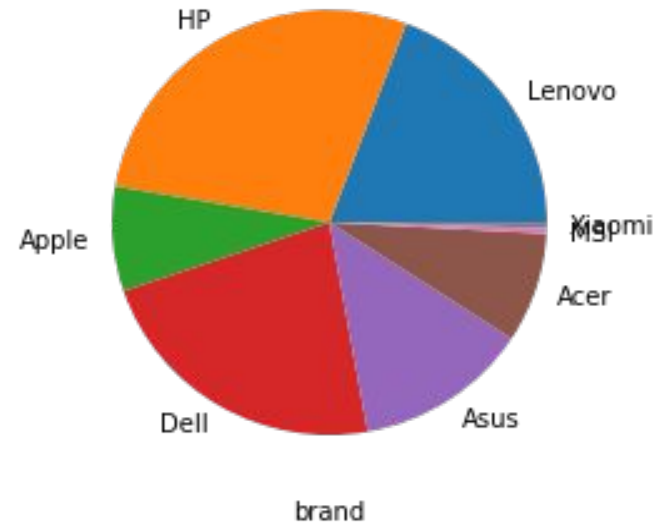❖ Ordinal categorical
❖ Discrete numerical
❖ Continuous numerical

Laptop Brand

❏ The data for the type of Laptop brand used by the students are represented in **Bar chart** and **Pie chart** as shown below.

Laptop Brand

- Type of data - Regular Categorical
- Mode - HP

Count

- HP   67
- Dell   54
- Lenovo   46
- Asus   31

- Acer   20
- Apple   19
- MSI   1
- Xiaomi   1

Operating System

Data collected

Operating systems available in student's laptop

Windows   -148
Windows, Linux   -64
MacOS   -19
Linux - 5
Windows, Linux, MacOS - 3

Data Processed

Count of students having a particular operating system

Windows - 215
Linux - 72
MacOS - 22

Operating System

❏ The data for the Operating System of the Laptop used by the students are represented in **Bar Chart** and **Pie chart** as shown below.

Bar Chart Plot of Operating System
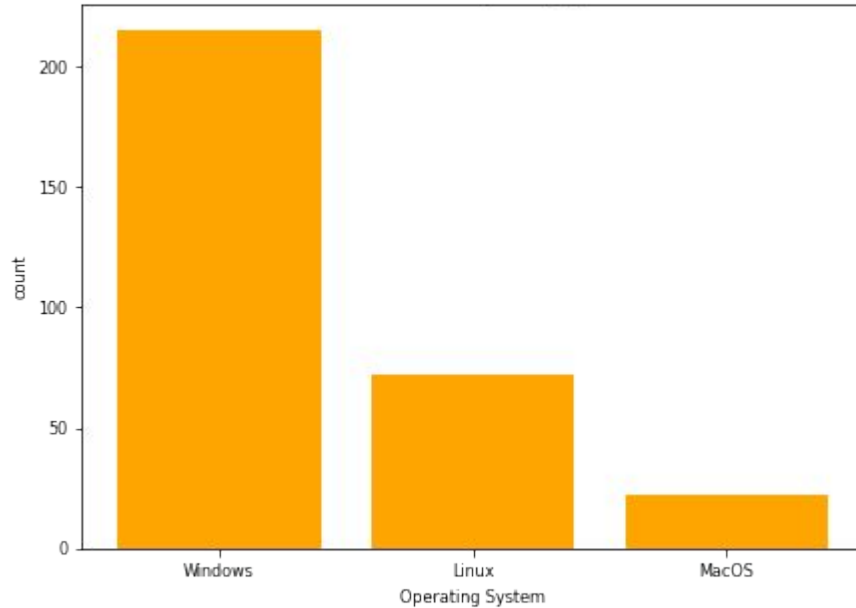
Pie Chart Plot of Operating System
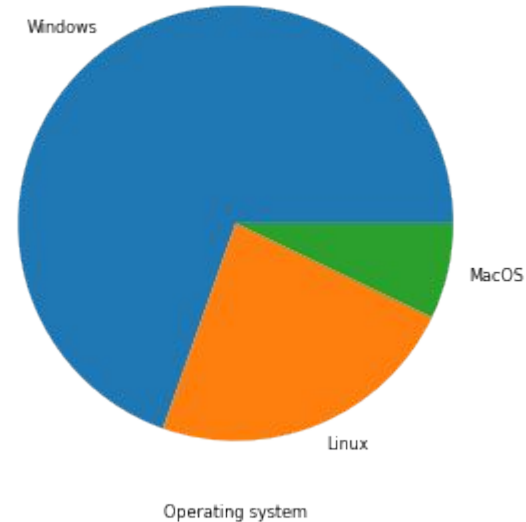
Operating System

- ❏ Type of data - Regular Categorical
- ❏ Mode - Windows

Count
- ➤ Windows   215
- ➤ Linux   72
- ➤ MacOS   22

Stream

❏ The data for the type of stream the students are graduating are represented in **Bar chart** and **Pie chart** as shown below.
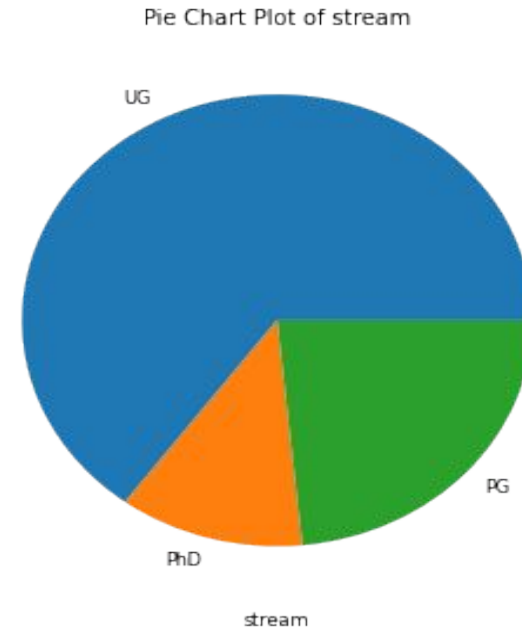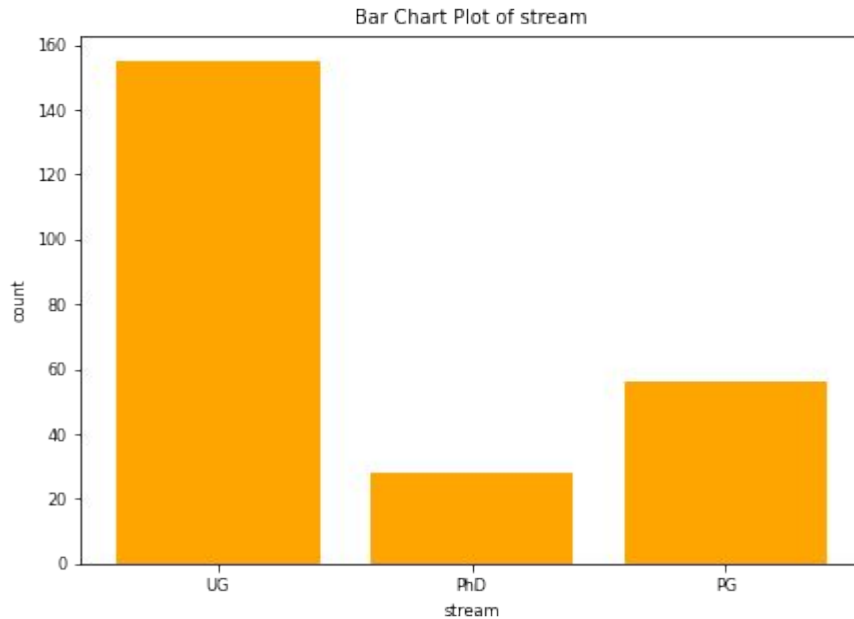
Stream

- ❏ Type of data - Regular Categorical
- ❏ Mode - UG

Count

➢ UG   155

➢ PG   56

➢ PHD   28

Department

Bar Chart Plot of department
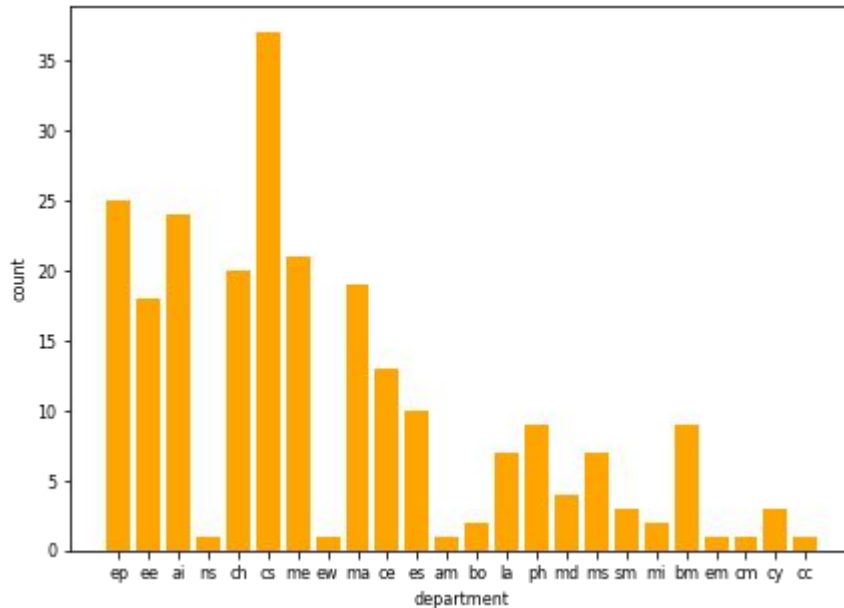
Pie Chart Plot of department
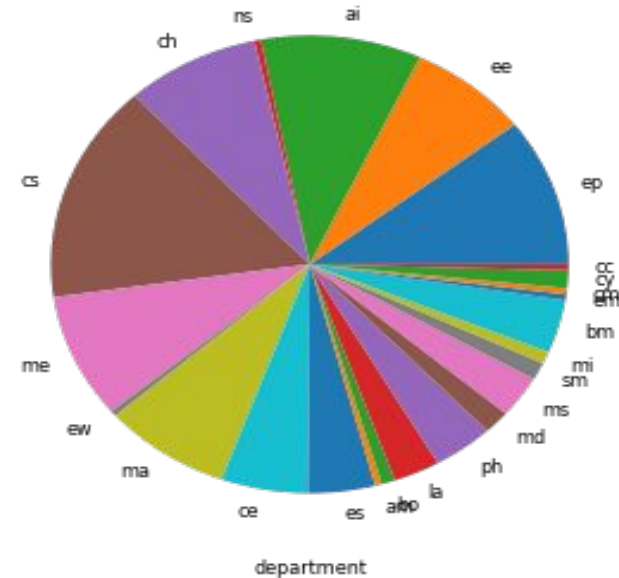
Department

- ❏ Type of data - Regular Categorical
- ❏ Mode - CS

Count

➢ CS	37	➢ ES	10	➢ MI	2
➢ EP	25	➢ BM	9	➢ BO	2
➢ AI	24	➢ PH	9	➢ AM	1
➢ ME	21	➢ LA	7	➢ EW	1
➢ CH	20	➢ MS	7	➢ EM	1
➢ MA	19	➢ MD	4	➢ CM	1
➢ EE	18	➢ CY	3	➢ NS	1
➢ CE	13	➢ SM	3	➢ CC	1

Display Type

❏ The data for the type of Display of the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.
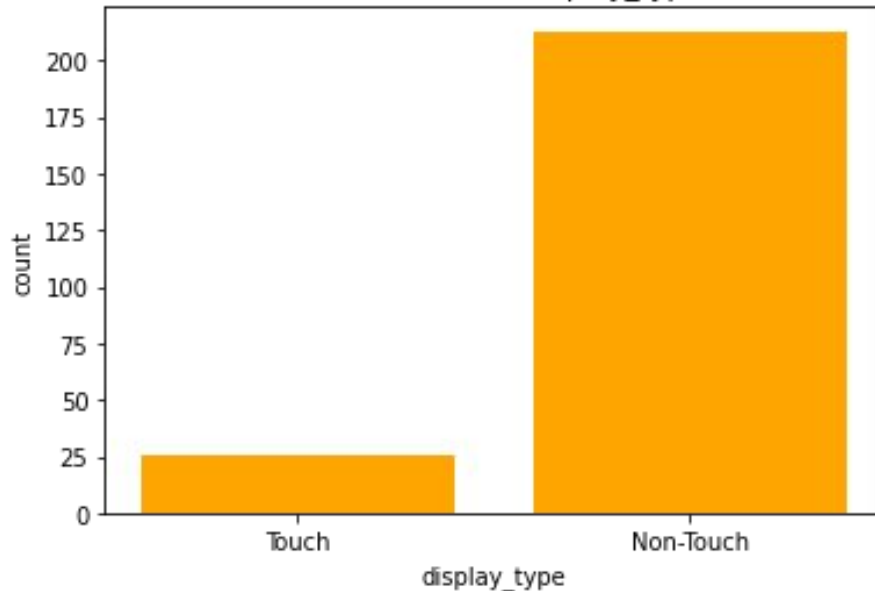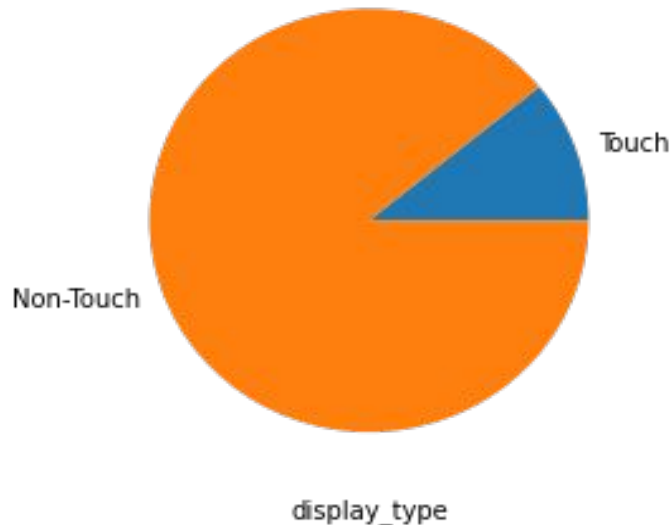
Display Type

- ❏ Type of data - Regular Categorical
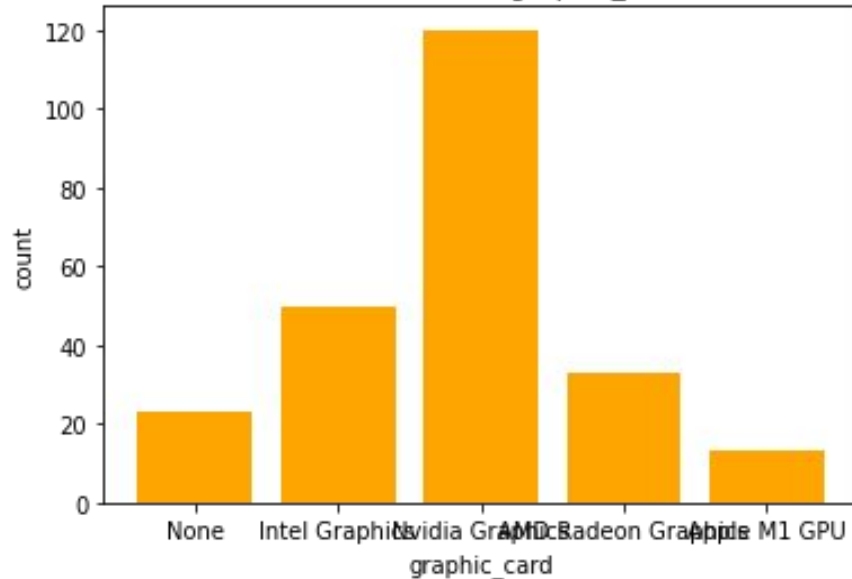- ❏ Mode - Non - Touch
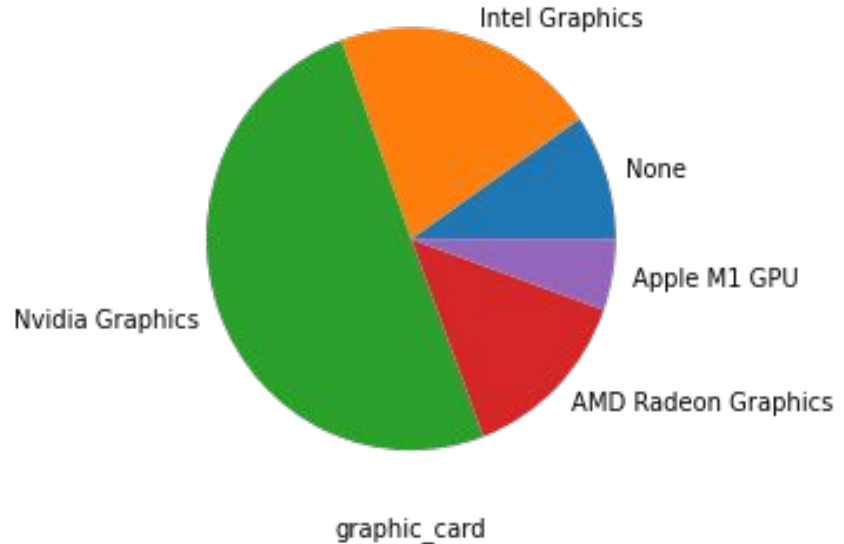
Count

➢ Non-Touch 213
➢ Touch   26

Graphic Card

❑ The data for the type of Graphic card available in the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.

Graphic Card

- ❏ Type of data - Regular Categorical
- ❏ Mode - Nvidia Graphics
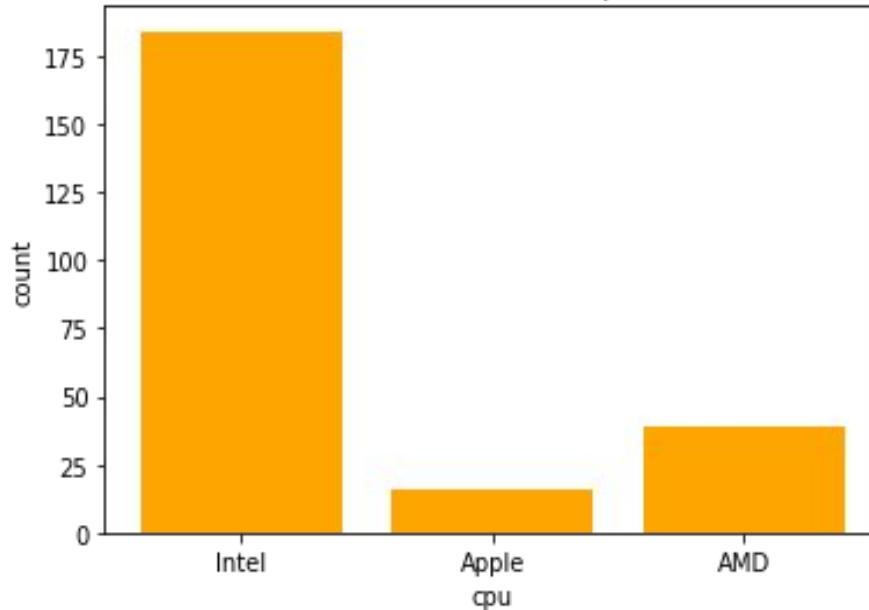
Count

- ➢ Nvidia Graphics   120
- ➢ Intel Graphics   50
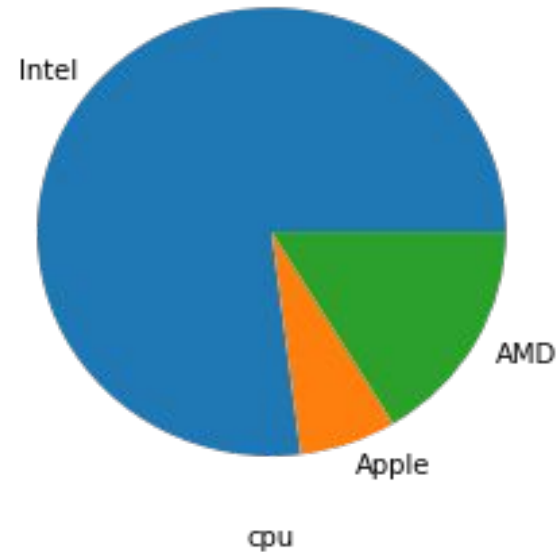- ➢ AMD Radeon graphics   33
- ➢ None   23
- ➢ Apple M1 GPU   13

CPU

❏ The data for the type of CPU available in the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.

CPU

- ❏ Type of data - Regular Categorical
- ❏ Mode - Intel

	Count
➢ Intel	184
➢ AMD	39
➢ Apple	16

CPU Model

❏ The data for the type of CPU model available in the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.
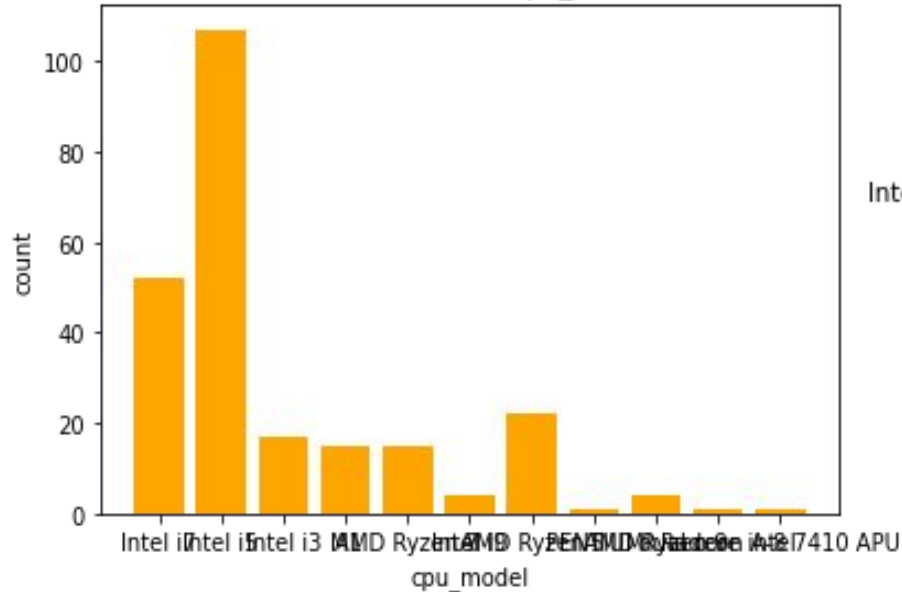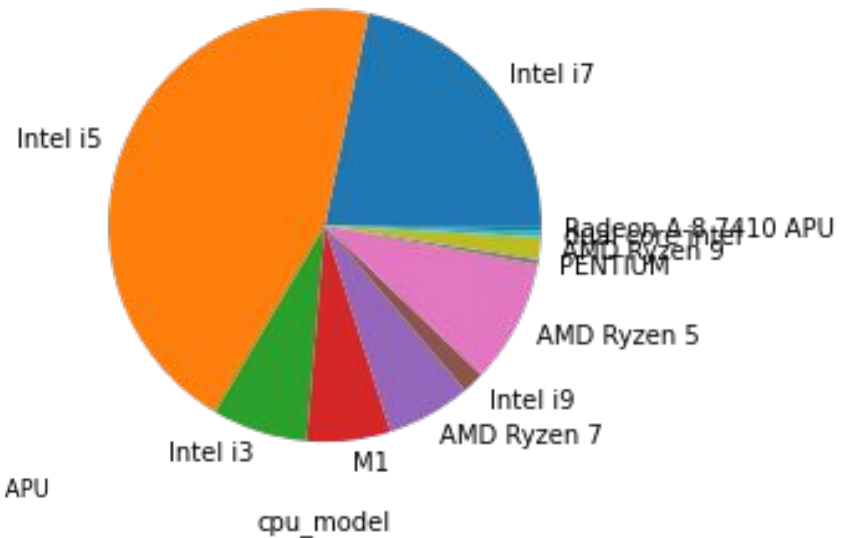
CPU Model

❏ Type of data - Regular Categorical
❏ Mode - Intel i5

Count

- ➢ Intel i5   107
- ➢ Intel i7   52
- ➢ AMD Ryzen 5   22
- ➢ Intel i3   17
- ➢ M1   15
- ➢ AMD Ryzen 7   15

- ➢ Intel i9   4
- ➢ AMD Ryzen 9   4
- ➢ PENTIUM   1
- ➢ Dual Core Intel   1
- ➢ Radeon A-8 7410 APU   1

CPU Clock Speed

❏ The data for the CPU clock speed of the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.
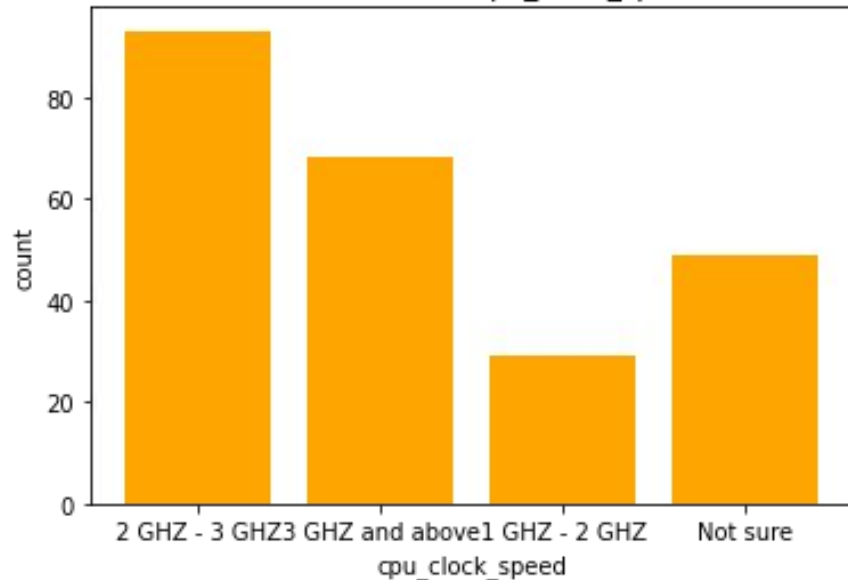
CPU Clock Speed
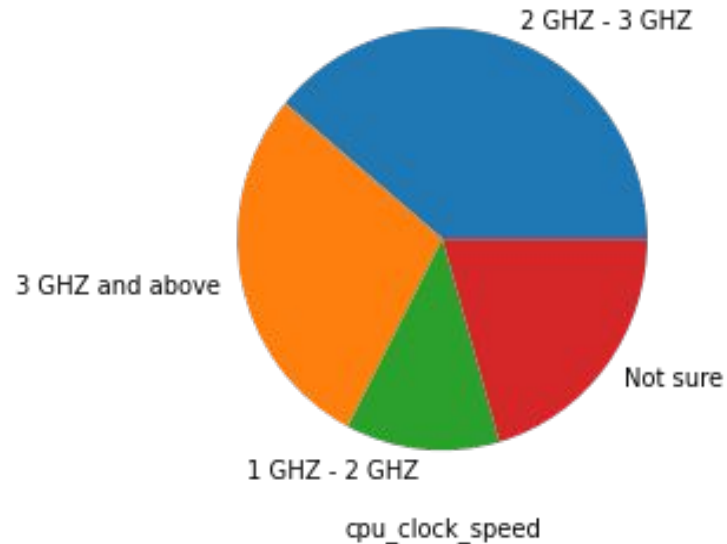
- ❏ Type of data - Regular Categorical
- ❏ Mode - (2 GHZ - 3 GHZ)

Count

- ➢ 2 GHZ - 3 GHZ   93
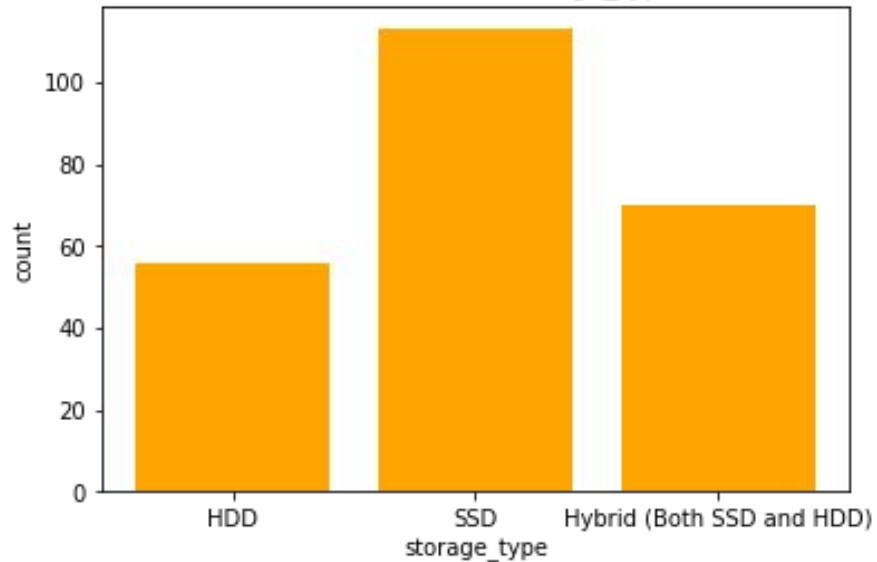- ➢ 3 GHZ and above   68
- ➢ Not Sure   49
- ➢ 1 GHZ - 2 GHZ   29

Storage Type

❏ The data for the Storage type of the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.

Storage Type

- ❏ Type of data - Regular Categorical
- ❏ Mode - SSD

	Count
➢ SSD	113
➢ Hybrid	70
➢ HDD	56

HDD Size

❏ The data for the HDD size of the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.

Bar Chart Plot of hdd_size

Pie Chart Plot of hdd_size

HDD Size

- ❏ Type of data - Regular Categorical
- ❏ Mode - None

	Count
➢ None	110
➢ 1 TB	95
➢ 500 GB	25
➢ 2 TB	9

SSD Size
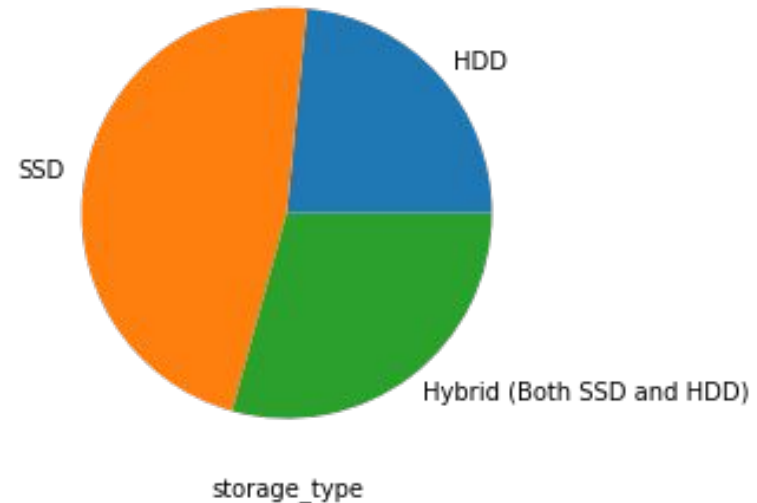
❏ The data for the SDD size of the Laptop used by the students are represented in **Bar chart** and **Pie chart** as shown below.
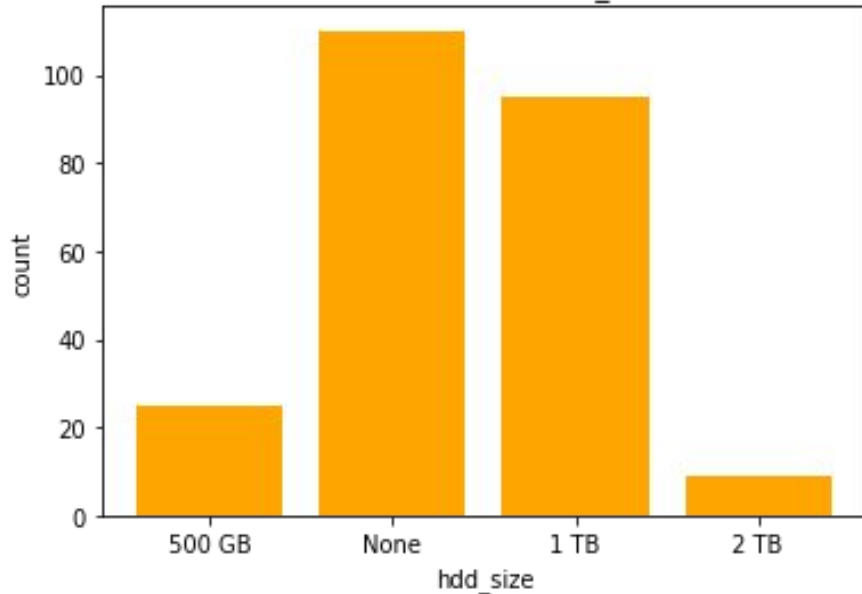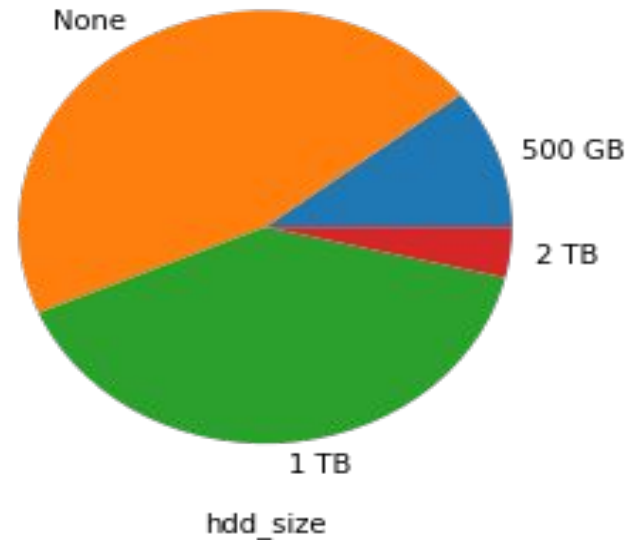

Bar Chart Plot of ssd_size
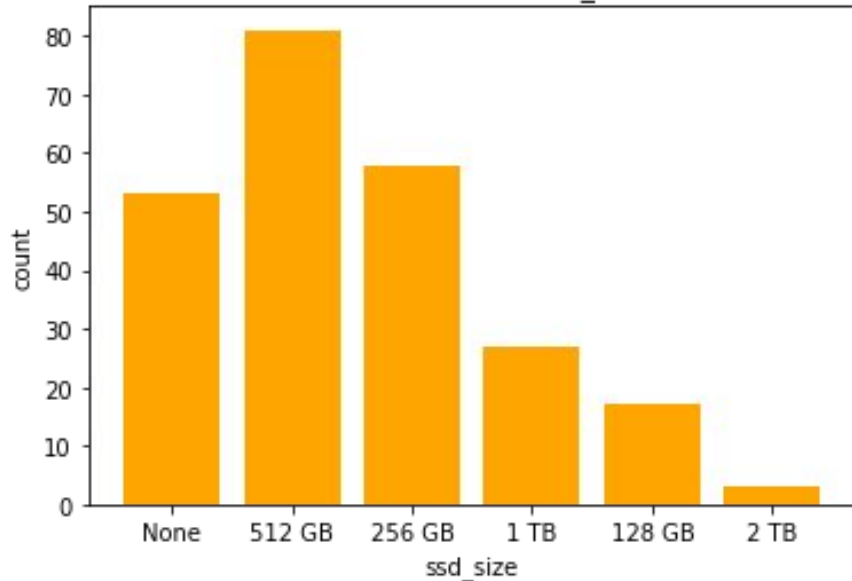
Pie Chart Plot of ssd_size
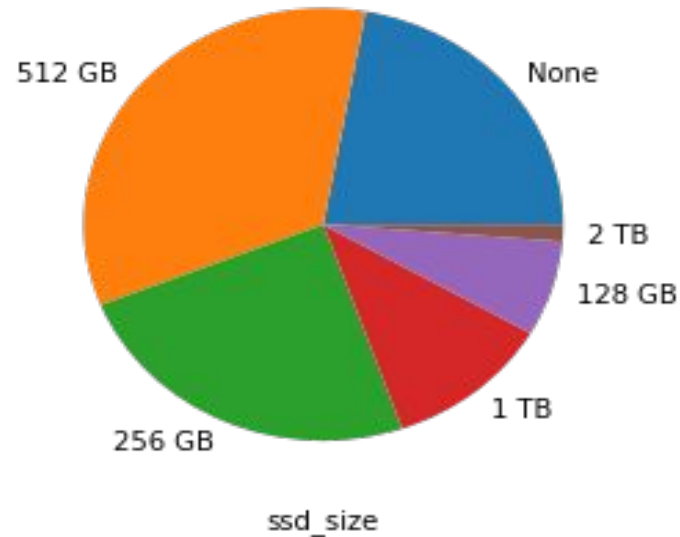
SSD Size

- ❏ Type of data - Regular Categorical
- ❏ Mode - 512 GB

Count

- ➢ 512 GB   81
- ➢ 256 GB   58
- ➢ None   53
- ➢ 1 TB   27
- ➢ 128 TB   17
- ➢ 2 TB   3

Price Range

❏ The data for the price of the Laptop used by the students are represented in **Histogram plot** and **Pie chart** as shown below.

Histogram Plot of price range
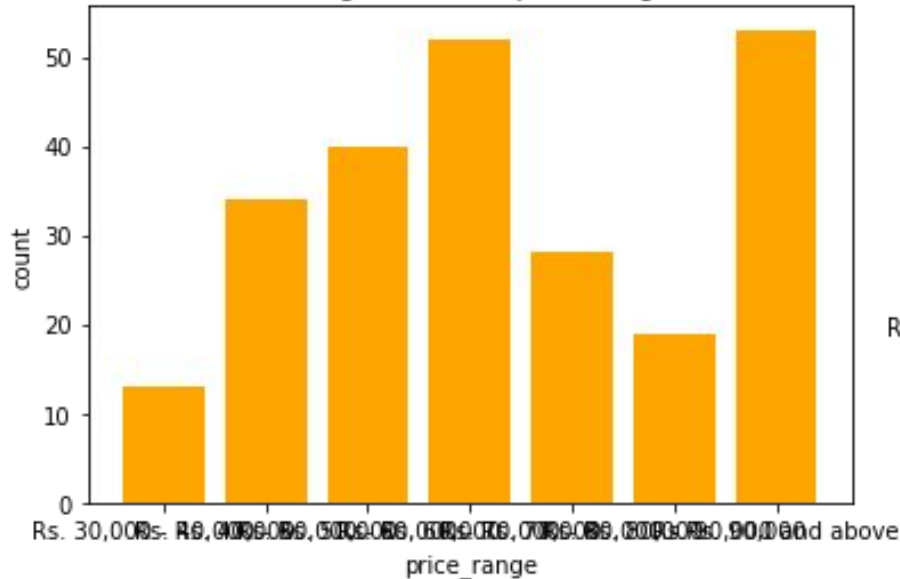
Pie Chart Plot of price_range

Price Range

- Type of data - Ordinal Categorical
- Mode - (Rs. 90,000 and above)

Count

- Rs. 90,000 and above — 53
- Rs. 60,000 - Rs. 70,000 — 52
- Rs. 50,000 - Rs. 60,000 — 40
- Rs. 50,000 - Rs. 50,000 — 34
- Rs. 70,000 - Rs. 80,000 — 28
- Rs. 80,000 - Rs. 90,000 — 19
- Rs. 30,000 - Rs. 40,000 — 13

★ The Distribution of Price Range is almost Left Skewed.
★ The Histogram is bimodal

Display size

❏ The data for the Size of Display of the Laptop used by the students are represented in **Histogram plot** and **Pie chart** as shown below.

Display size

❏ Type of data - Ordinal Categorical

❏ Mode - (14 inch - 15 inch)

Count

➢ 14 inch - 15 inch   100
➢ 15 inch - 16 inch   74
➢ 13 inch - 14 inch   45
➢ 12 inch - 13 inch   15
➢ 16 inch - 17 inch   5

★ The Distribution of Display is almost Symmetric.
★ The Histogram is Unimodal

Year of Joining

❑ The data for the year of joining of the students in the college are represented in **Histogram plot** and **Box plot** as shown below.
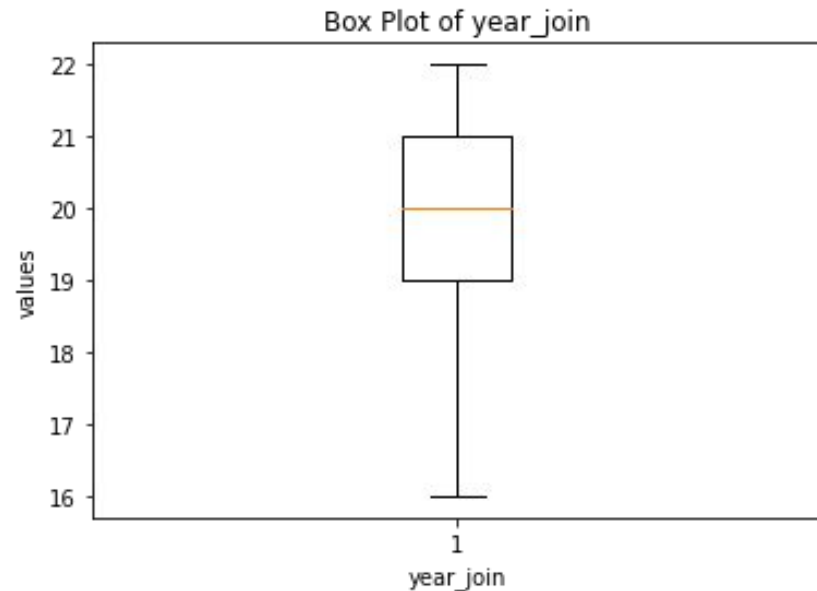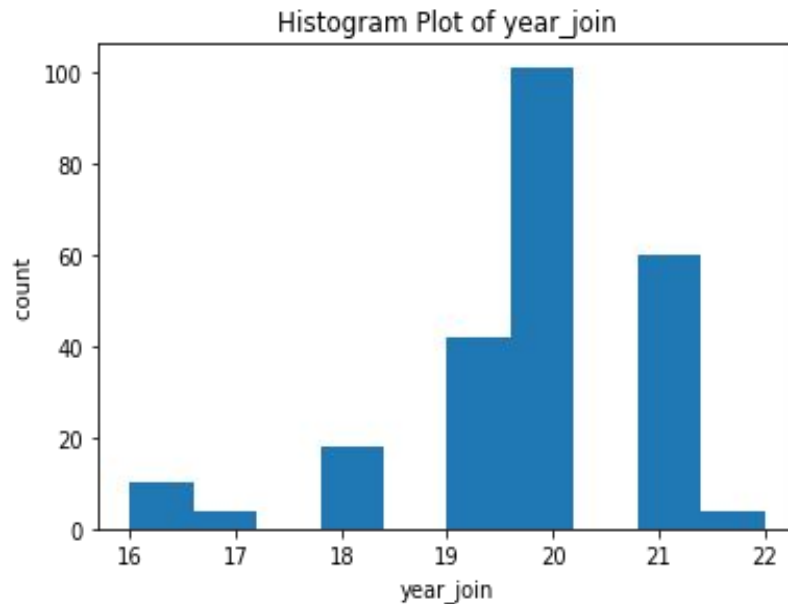
Year of Joining

- Type of data - Discrete Numerical
- Mode of Year of Joining is 20
- Mean of Year of Joining is 19.7405885774058577
- Median of Year of Joining is 20.0
- Range of Year of Joining = 6
- 0th of Year of Joining quartile = 16.0
- 1th of Year of Joining quartile = 19.0
- 2nd of Year of Joining quartile = 20.0
- 3rd of Year of Joining quartile = 21.0
- 4th of Year of Joining quartile = 22.0
- IQR of Year of Joining = 2.0
- Variance of Year of Joining = 1.5477670208854886
- Standard Deviation of Year of Joining = 1.2440928505885276
- The data is Left Skewed
- The Histogram is Bimodal

Year of Joining

Count

- 2020   101
- 2021   60
- 2019   42
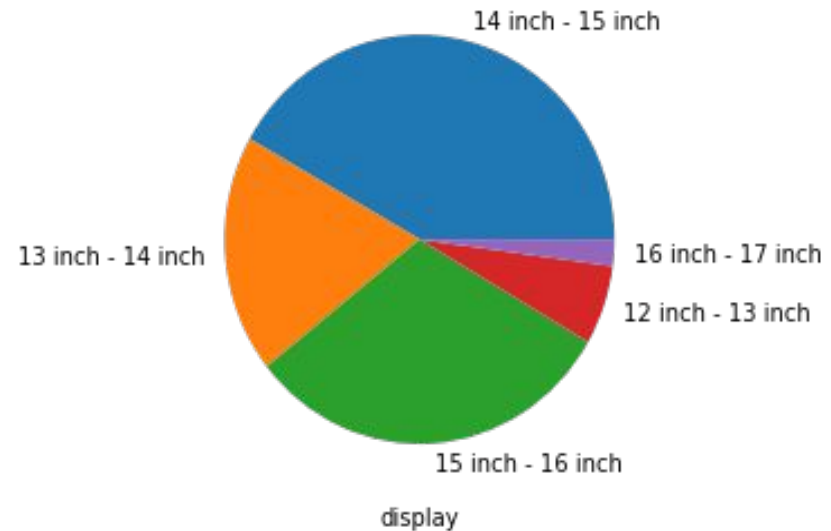- 2018   18

- 2016   10
- 2017   4
- 2022   4

RAM Size

❑ The data for the Size of RAM of the Laptop used by the students are represented in **Histogram plot** and **Box plot** as shown below.
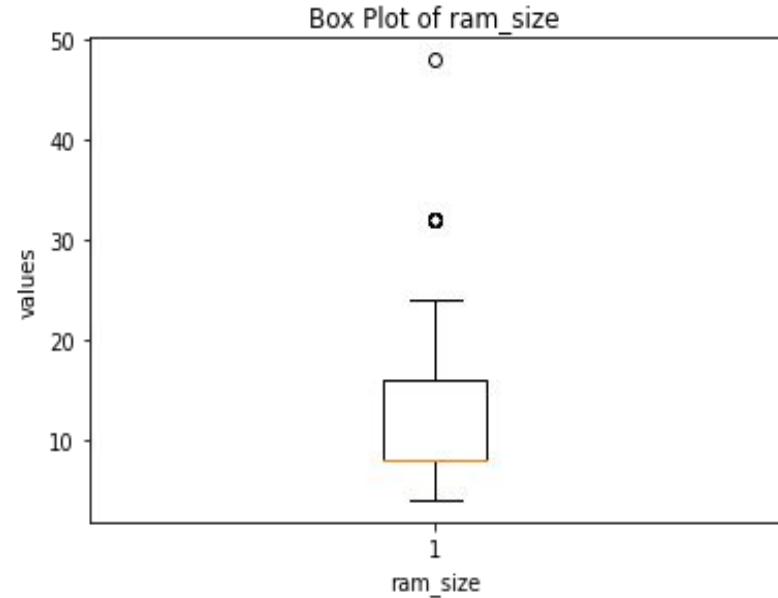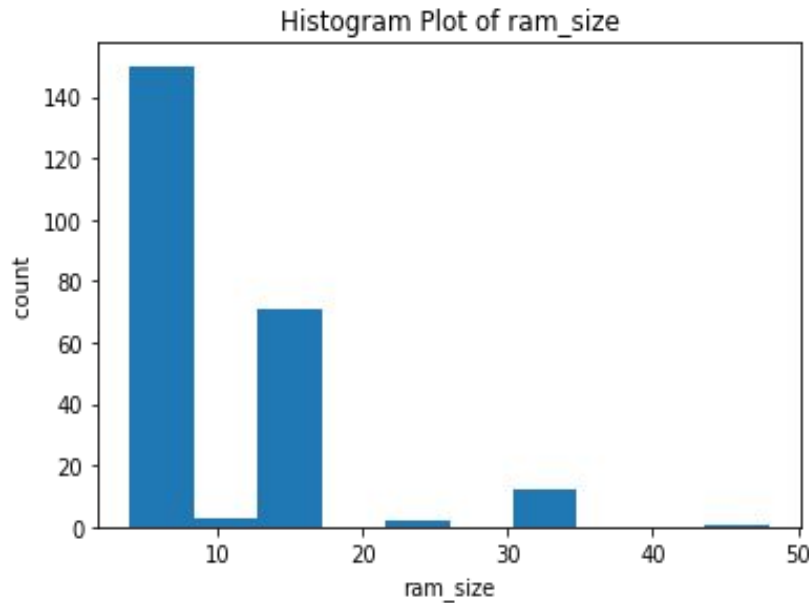
RAM Size

- ❏ Type of data - Discrete Numerical
- ❏ Mode of RAM Size is 8
- ❏ Mean of RAM Size is 11.6820083368200837
- ❏ Median of RAM Size is 8.0
- ❏ Range of RAM Size = 44
- ❏ 0th of RAM Size quartile = 4.0
- ❏ 1th of RAM Size quartile = 8.0
- ❏ 2nd of RAM Size quartile = 8.0
- ❏ 3rd of RAM Size quartile = 16.0
- ❏ 4th of RAM Size quartile = 48.0
- ❏ IQR of RAM Size = 8.0
- ❏ Variance of RAM Size = 44.417709773988555
- ❏ Standard Deviation of RAM Size = 6.664661264759714
- ❏ The data is Right Skewed
- ❏ The Histogram is Unimodal

RAM Size

Count

- 8 GB 135
- 16 GB 71
- 4 GB 15
- 32 GB 12
- 12 GB 3
- 24 GB 2
- 48 GB 1

Average Watch Time

❏ The data for the Average watch time of the Laptop used by the students are represented in **Histogram plot** and **Box plot** as shown below.
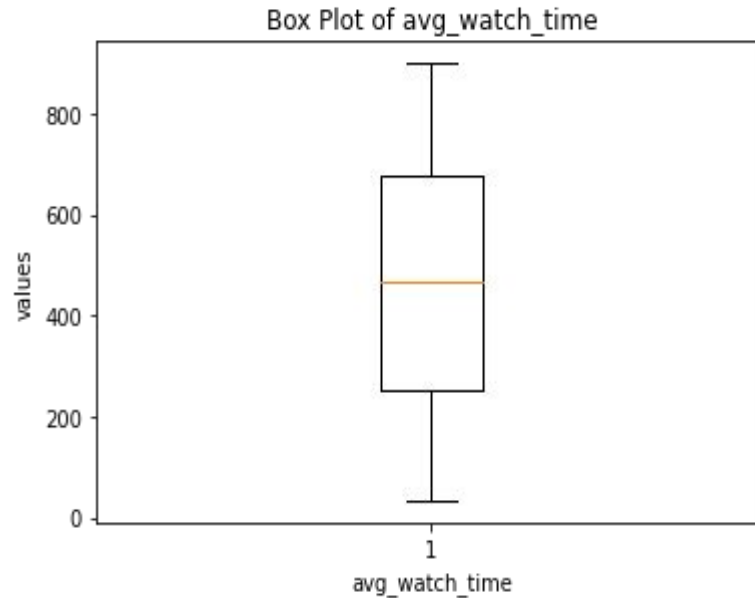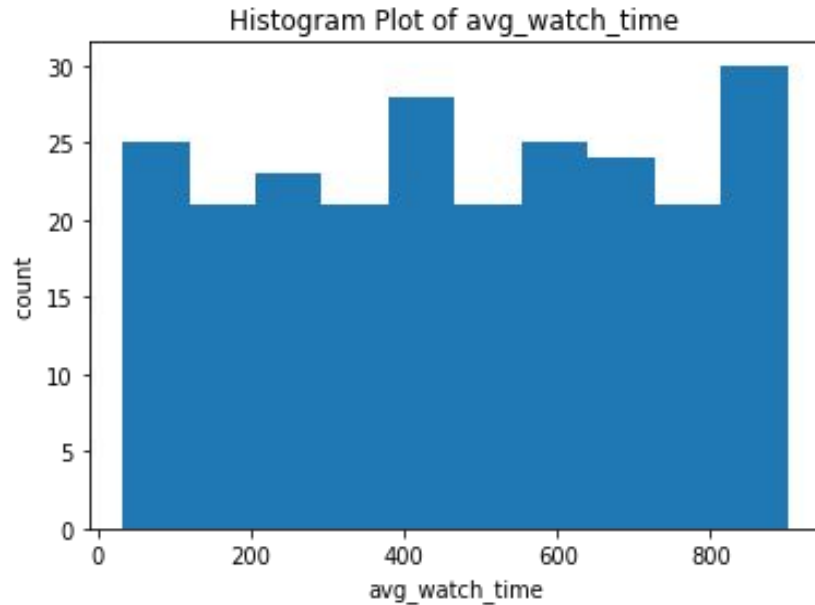
Average Watch Time

- Type of data - Continuous Numerical
- Modes of Average watch time are 451, 844, 883
- Mean of Average watch time is 474.092050209205
- Median of Average watch time is 467.0
- Range of Average watch time = 867
- 0th of Average watch time quartile = 33.0
- 1th of Average watch time quartile = 251.0
- 2nd of Average watch time quartile = 467.0
- 3rd of Average watch time quartile = 679.0
- 4th of Average watch time quartile = 900.0
- IQR of Average watch time = 427.5
- Variance of Average watch time = 64306.15052257488
- Standard Deviation of Average watch time = 253.5865740187656
- The data is Left Skewed
- The Histogram is Uniform

SAMPLING DISTRIBUTION

Sampling Distribution

   Sampling the given Data is of two types. One for Numerical Data and One for Categorical Data. The two parts are sampled in different ways.

❏   So for that we created a function which takes inputs as DataFrame, Sample Size, Number of Combinations.

❏   The function separates the Columns of the DataFrame into either Categorical or Numerical Columns. On classifying them, the function further analyses the column based on its type as follows:

Analysing Numerical Data

If the data is Numerical:

❏ For every combination, we calculated sample parameters like sample mean, sample variance, standard error of the randomly selected samples (of size sample size).

❏ Plotted the Frequency Plots (Histograms) of Sample Mean, Sample Variance for every data column from the above combinations.

❏ Found the Expectation of Sample Mean(E(Xbar)), Variance of Sample Mean (Var(Xbar)) and Expectation of Sample Variance(E(S^2)) and verified the above values with population parameters.

❏ Also found Expectation of Standard Error.

❏ To compare sample distribution with population distribution, we also plotted population Frequency Plots (histograms). As the Central Limit Theorem says the Sample Distribution tends towards Normal Distribution.

Analysing Categorical Data

If the data is Categorical:

❏ For every combination, we calculated sample parameters like sample mode(since mean, variance are not suitable for categorical data) of the randomly selected samples (of size sample size).

❏ From the above different combinations, we have plotted the Frequency Plots (Bar Charts) for different columns of the data.

❏ We also mentioned Sample Mode of each Column along with Count of each category of every column in the Data obtained from above combinations.

❏ We also mentioned Sample Proportions of each category in every column for Sampled Data and Population Data.

Results

- Our Data Frame consists of 17 attributes.
- We have 3 Numerical Attributes and 14 Categorical attributes out of these 17 attributes.
- We have generated 9 Histograms for Numerical Attributes(Population Distribution, Sample Mean Distribution, Sample Variance Distribution) and 14 Bar Graphs for Categorical Attributes(Frequency Plots of Sampled Data).
- From the above values of proportions, we found the proportions of Sampled Data are close to that of Population Data.
- We obtained 12 parameter values for Numerical Attributes(E(Xbar), Var(Xbar), E(S^2), E(s)).

Results(Contd.)

The Input for the above function is {final_df (after cleaning the data frame), 10 (Sample Size), 300(Number of Combinations)}.

Our function classified the columns as follows:

1. Numerical Columns - ['ram_size', 'year_join', 'avg_watch_time']
2. Categorical Columns - ['stream', 'brand', 'display_type', 'cpu', 'cpu_model', 'graphic_card', 'storage_type', 'operating_sys', 'department', 'cpu_clock_speed', 'hdd_size', 'ssd_size', 'price_range', 'display']

1. year_join

Population Distribution Plot

Sample Distribution of Mean Plot

Population Distribution for column year_join

Sampling Distribution(Sample Size = 10) of Mean of 300 random samples for column year_join
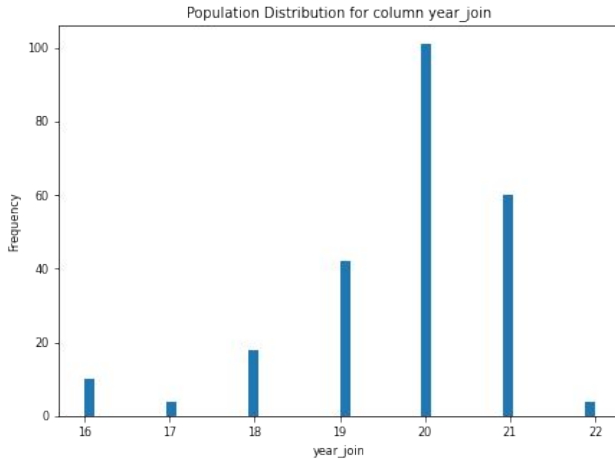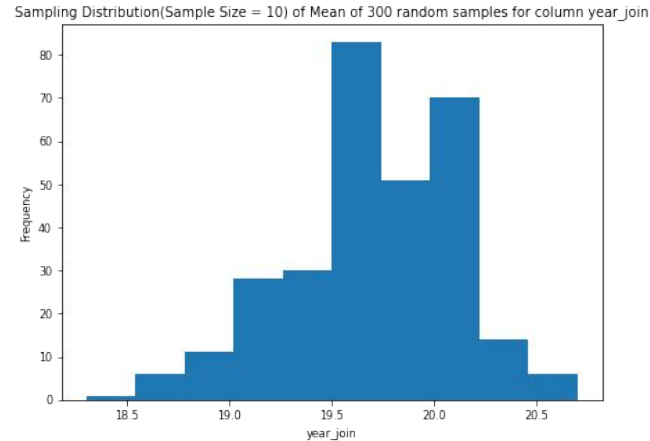
Numerical Analysis Findings

1. year_join

Sample Statistics Values:

- Expectation of Sample Mean(Sample Size = 10) of year_join for 300 random combinations is 19.699333333333332
- Variance of Sample Mean(Sample Size = 10) of year_join for 300 random combinations is 0.1457345555555558
- Expectation of Sample Variance(Sample Size = 10) of year_join for 300 random combinations is 1.5580740740740742
- Standard Error(Sample Size = 10) for Sample Mean of year_join for 300 random combinations is 0.37461535495142007

Numerical Analysis Findings

2. avg_watch_time

Population Distribution Plot
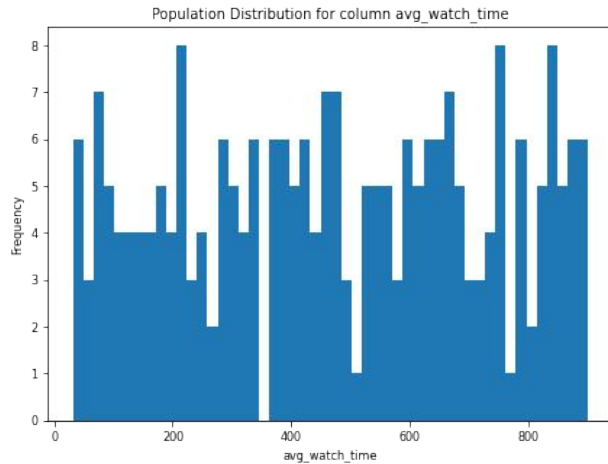
Sample Distribution of Mean Plot
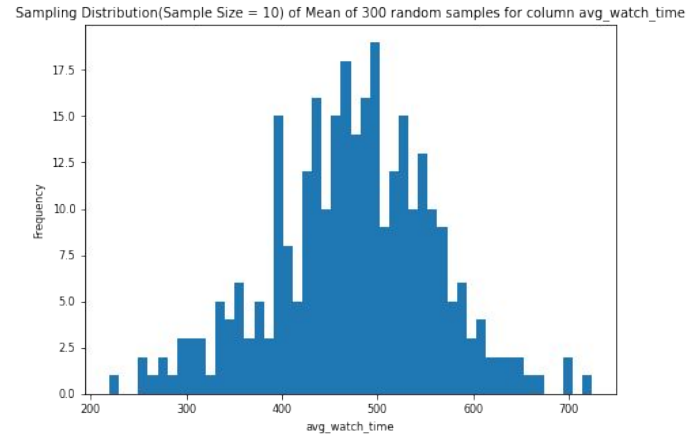
Numerical Analysis Findings

2. avg_watch_time

<u>Sample Statistics Values:</u>

- ❏ Expectation of Sample Mean(Sample Size = 10) of avg_watch_time for 300 random combinations is 473.9263333333331
- ❏ Variance of Sample Mean(Sample Size = 10) of avg_watch_time for 300 random combinations is 6458.584670666668
- ❏ Expectation of Sample Variance(Sample Size = 10) of avg_watch_time for 300 random combinations is 63282.939814814774
- ❏ Standard Error(Sample Size = 10) of avg_watch_time for 300 random combinations is 78.60697175381785

3. ram_size

Population Distribution Plot
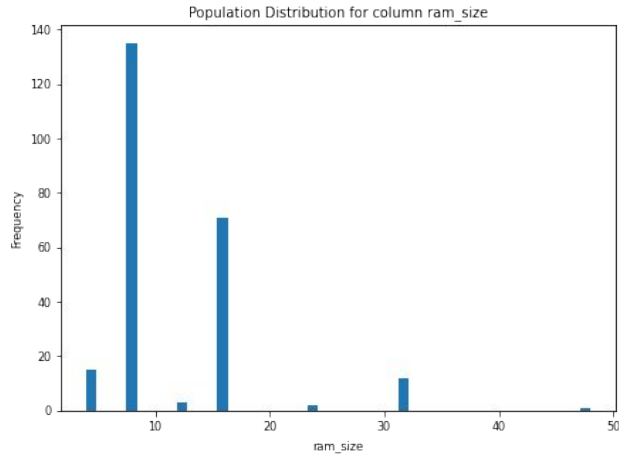
Sample Distribution of Mean Plot
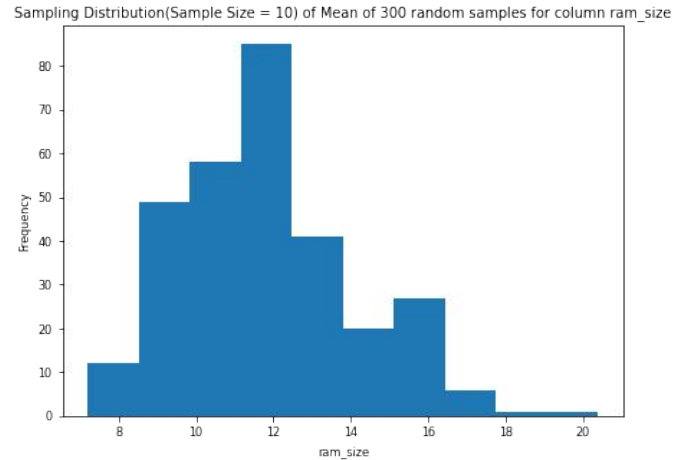
Numerical Analysis Findings

3. ram_size

Sample Statistics Values:

- Expectation of Sample Mean(Sample Size = 10) of ram_size for 300 random combinations is 11.811999999999998
- Variance of Sample Mean(Sample Size = 10) of ram_size for 300 random combinations is 4.511822222222225
- Expectation of Sample Variance(Sample Size = 10) of ram_size for 300 random combinations is 47.664
- Standard Error(Sample Size = 10) of ram_size for 300 random combinations is 1.9946963871174948

Theorem

Cross Verifying with below theorem from above calculated parameters.

Theorem

Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

(i) $E\left(\overline{X}\right) = \mu$

(ii) $Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$

(iii) $E\left(S^2\right) = \sigma^2$

Cross Verification

1. year_join

Calculated Parameters from Sampling Distribution

Population Parameters

$$n = 10$$
$$E(\bar{X}) = 19.699333333333332$$
$$Var(\bar{X}) = 0.1457345555555558$$
$$E(S^2) = 1.5580740740740742$$

$$\mu = 19.740585774058577$$
$$\sigma^2 = 1.5477670208854886$$
$$\sigma^2/n = \sigma^2/10 = 0.15477670208854886$$

Cross Verification

2. avg_watch_time

Calculated Parameters from Sampling Distribution

$n = 10$

$E(\bar{X}) = 473.9263333333331$

$Var(\bar{X}) = 6458.584670666668$

$E(S^2) = 63282.939814814774$

Population Parameters

$\mu = 474.092050209205$

$\sigma^2 = 64306.15052257488$

$\sigma^2/n = \sigma^2/10 = 6430.615052257488$

3. ram_size

Calculated Parameters from Sampling Distribution

$n = 10$

$E(\bar{X}) = 11.811999999999998$

$Var(\bar{X}) = 4.511822222222225$

$E(S^2) = 47.664$

Population Parameters

$\mu = 11.682008368200837$

$\sigma^2 = 44.417709773988555$

$\sigma^2/n = \sigma^2/10 = 4.4417709773988555$

Categorical Analysis Findings

1. brand

Sampling Distribution of 'brand'

Sampling Distribution of brand for sample size 10
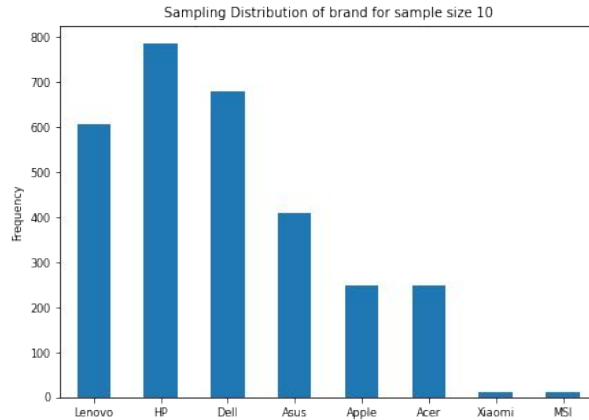
Mode of Sample Data:

The Mode of the Column brand is HP and its sample proportion is 0.262

Categorical Analysis Findings

1. brand

Value Counts of each category in 'brand'

Frequency of column brand of Sampling Distribution(Sample Size = 10) is

Brand	Count
Lenovo	607
HP	786
Dell	680
Asus	409
Apple	248
Acer	248
Xiaomi	11
MSI	11

Comparing Population Proportions and Sample Proportions

Sample Proportions:

Brand	Proportion
HP	0.280335
Apple	0.088000
Dell	0.236000
Lenovo	0.192000
Acer	0.086333
Asus	0.129000
Xiaomi	0.002667
MSI	0.001667

Population Proportions:

Brand	Proportion
HP	0.280335
Dell	0.225941
Lenovo	0.192469
Asus	0.129707
Acer	0.083682
Apple	0.079498
MSI	0.004184
Xiaomi	0.004184

Categorical Analysis Findings

2. price_range

Sampling Distribution of 'price_range'

Sampling Distribution of price_range for sample size 10
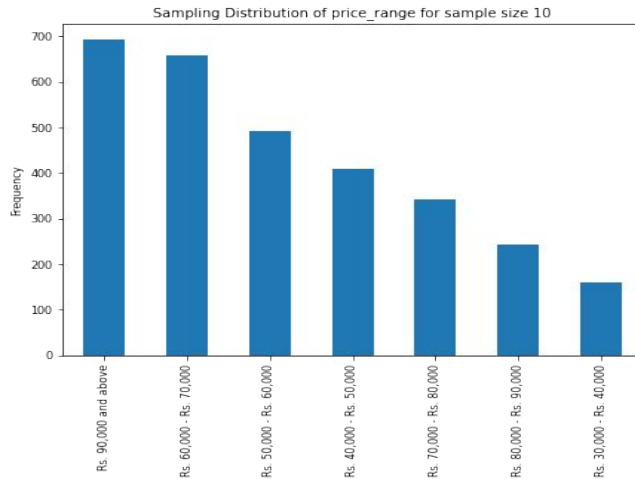
Mode of Sample Data:

The Mode of the Column price_range is Rs. 90,000 and above and its sample proportion is 0.231

Categorical Analysis Findings

2. price_range

Value Counts of each category in 'price_range'

Frequency of column price_range of Sampling Distribution(Sample Size = 10) is

Rs. 90,000 and above	693
Rs. 60,000 - Rs. 70,000	658
Rs. 50,000 - Rs. 60,000	492
Rs. 40,000 - Rs. 50,000	410
Rs. 70,000 - Rs. 80,000	343
Rs. 80,000 - Rs. 90,000	243
Rs. 30,000 - Rs. 40,000	161

Comparing Population Proportions and Sample Proportions

Sample Proportions:

Rs. 90,000 and above	0.224000
Rs. 60,000 - Rs. 70,000	0.219000
Rs. 50,000 - Rs. 60,000	0.168000
Rs. 40,000 - Rs. 50,000	0.143000
Rs. 70,000 - Rs. 80,000	0.118000
Rs. 80,000 - Rs. 90,000	0.080333
Rs. 30,000 - Rs. 40,000	0.047667

Population Proportions:

Rs. 90,000 and above	0.221757
Rs. 60,000 - Rs. 70,000	0.217573
Rs. 50,000 - Rs. 60,000	0.167364
Rs. 40,000 - Rs. 50,000	0.142259
Rs. 70,000 - Rs. 80,000	0.117155
Rs. 80,000 - Rs. 90,000	0.079498
Rs. 30,000 - Rs. 40,000	0.054393

Categorical Analysis Findings

3. stream

Sampling Distribution of 'stream'

Sampling Distribution of stream for sample size 10
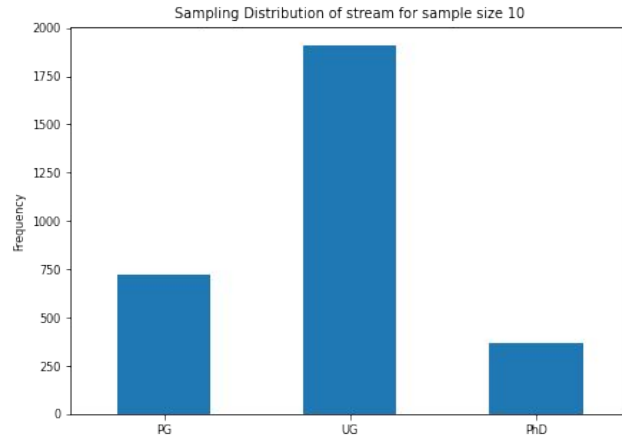
Mode of Sample Data:

The Mode of the Column stream is UG and its sample proportion is 0.6363333333333333.

Categorical Analysis Findings

3. stream

Value Counts of each category in 'stream'

Frequency of column stream of Sampling Distribution(Sample Size = 10) is

PG   723
UG   1909
PhD   368

Comparing Population Proportions and Sample Proportions

Sample Proportions:
PG   0.241000
UG   0.636333
PhD   0.122667

Population Proportions:
UG   0.648536
PG   0.234310
PhD   0.117155

Categorical Analysis Findings

4. display

Sampling Distribution of 'display'
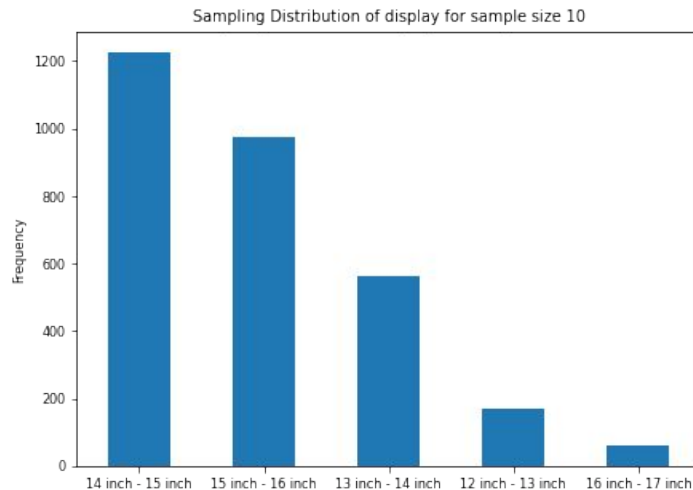
Mode of Sample Data:

The Mode of the Column display is 14 inch - 15 inch and its sample proportion is 0.4086666666666667

Categorical Analysis Findings

4. display

Value Counts of each category in 'display'

Frequency of column display of Sampling Distribution(Sample Size = 10) is

14 inch - 15 inch	1226
15 inch - 16 inch	976
13 inch - 14 inch	565
12 inch - 13 inch	171
16 inch - 17 inch	62

Comparing Population Proportions and Sample Proportions

Sample Proportions:

14 inch - 15 inch	0.426333
15 inch - 16 inch	0.297333
13 inch - 14 inch	0.190000
12 inch - 13 inch	0.064333
16 inch - 17 inch	0.022000

Population Proportions:

14 inch - 15 inch	0.418410
15 inch - 16 inch	0.309623
13 inch - 14 inch	0.188285
12 inch - 13 inch	0.062762
16 inch - 17 inch	0.020921

Categorical Analysis Findings

5. display_type

Sampling Distribution of 'display_type

Sampling Distribution of display_type for sample size 10
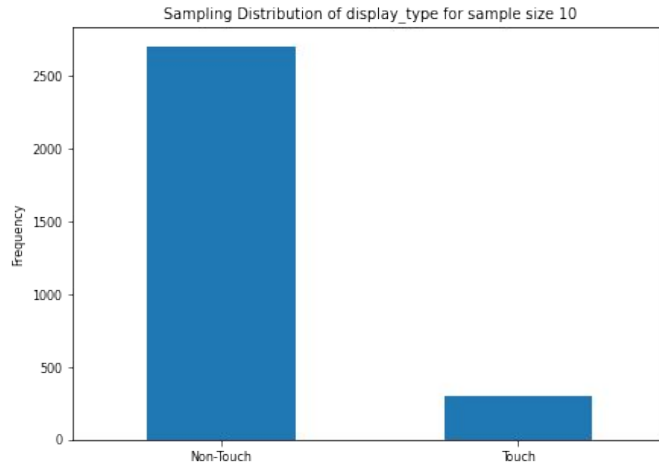
Mode of Sample Data:

The Mode of the Column display_type is Non-Touch and its sample proportion is 0.899

Categorical Analysis Findings

5. display_type

Value Counts of each category in 'display_type'

Frequency of column display_type of Sampling Distribution(Sample Size = 10) is

Non-Touch   2697
Touch   303

Comparing Population Proportions and Sample Proportions

Sample Proportions:
Non-Touch   0.887667
Touch   0.112333

Population Proportions:
Non-Touch   0.891213
Touch   0.108787

Categorical Analysis Findings

6. cpu_clock_speed

Sampling Distribution of 'cpu_clock_speed'

Sampling Distribution of cpu_clock_speed for sample size 10
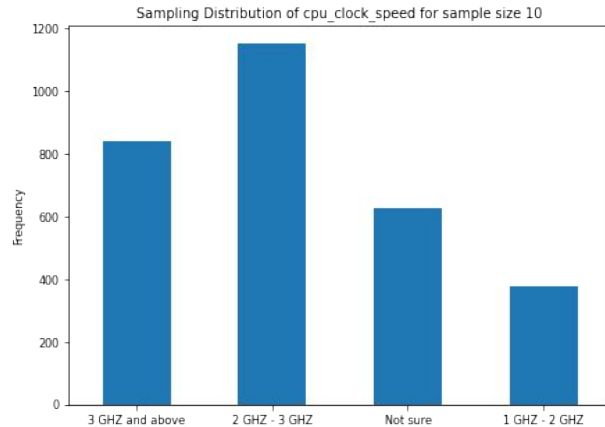
Mode of Sample Data:

The Mode of the Column cpu_clock_speed is 2 GHZ - 3 GHZ and its sample proportion is 0.38433333333333336

6. cpu_clock_speed

Value Counts of each category in 'cpu_clock_speed'

Frequency of column cpu_clock_speed of Sampling Distribution(Sample Size = 10) is

3 GHZ and above	841
2 GHZ - 3 GHZ	1153
Not sure	627
1 GHZ - 2 GHZ	379

Comparing Population Proportions and Sample Proportions

Sample Proportions:

3 GHZ and above	0.283333
Not sure	0.203000
2 GHZ - 3 GHZ	0.389333
1 GHZ - 2 GHZ	0.124333

Population Proportions:

2 GHZ - 3 GHZ	0.389121
3 GHZ and above	0.284519
Not sure	0.205021
1 GHZ - 2 GHZ	0.121339

Categorical Analysis Findings

7. cpu

Sampling Distribution of 'cpu'

Sampling Distribution of cpu for sample size 10
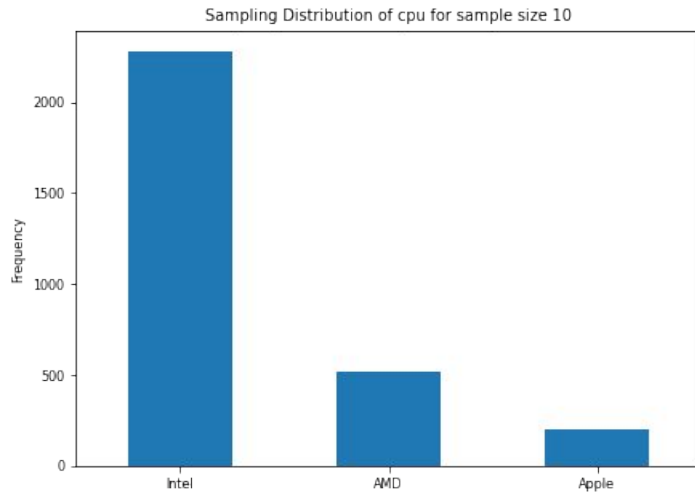
Mode of Sample Data:

The Mode of the Column cpu is Intel and its sample proportion is 0.76

Categorical Analysis Findings

7. cpu

Value Counts of each category in 'cpu'

Frequency of column cpu of Sampling Distribution(Sample Size = 10) is

Intel   2280
AMD   515
Apple   205

Comparing Population Proportions and Sample Proportions

Sample Proportions:
Intel   0.768333
Apple   0.075333
AMD   0.156333

Population Proportions:
Intel   0.769874
AMD   0.163180
Apple   0.066946

Categorical Analysis Findings

8. cpu_model

Sampling Distribution of 'cpu_model'
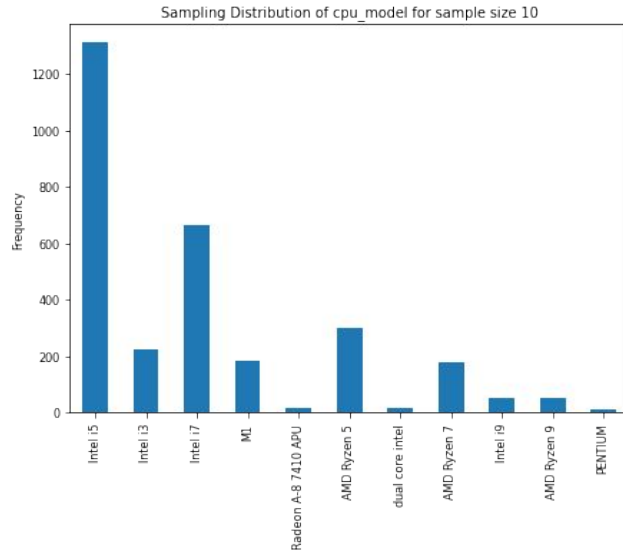
Mode of Sample Data:

The Mode of the Column cpu_model is Intel i5 and its sample proportion is 0.43666666666666665

Categorical Analysis Findings

8. cpu_model

Value Counts of each category in 'cpu_model'

Frequency of column cpu_model of Sampling
Distribution(Sample Size = 10) is

Intel i5	1346
M1	218
Intel i7	646
AMD Ryzen 9	41
AMD Ryzen 5	259
AMD Ryzen 7	189
Intel i3	206
Intel i9	64
dual core intel	12
PENTIUM	10
Radeon A-8 7410 APU	9

Comparing Population Proportions and Sample Proportions

Sample Proportions:

Intel i5	0.448667
M1	0.072667
Intel i7	0.215333
AMD Ryzen 9	0.013667
AMD Ryzen 5	0.086333
AMD Ryzen 7	0.063000
Intel i3	0.068667
Intel i9	0.021333
dual core intel	0.004000
PENTIUM	0.003333
Radeon A-8 7410 APU	0.003000

Population Proportions:

Intel i5	0.447699
Intel i7	0.217573
AMD Ryzen 5	0.092050
Intel i3	0.071130
M1	0.062762
AMD Ryzen 7	0.062762
Intel i9	0.016736
AMD Ryzen 9	0.016736
PENTIUM	0.004184
dual core intel	0.004184
Radeon A-8 7410 APU	0.004184

9. graphic_card

Sampling Distribution of 'graphic_card'

Sampling Distribution of graphic_card for sample size 10

Mode of Sample Data:

The Mode of the Column graphic_card is Nvidia Graphics and its sample proportion is 0.49766666666666665

Categorical Analysis Findings

9. graphic_card

Value Counts of each category in 'graphic_card'

Frequency of column graphic_card of Sampling Distribution(Sample Size = 10) is

Intel Graphics	616
Nvidia Graphics	1493
AMD Radeon Graphics	409
Apple M1 GPU	179
None	303

Comparing Population Proportions and Sample Proportions

Sample Proportions:		Population Proportions:	
Intel Graphics	0.205333	Nvidia Graphics	0.502092
Nvidia Graphics	0.497667	Intel Graphics	0.209205
AMD Radeon Graphics	0.136333	AMD Radeon Graphics	0.138075
Apple M1 GPU	0.059667	None	0.096234
None	0.101000	Apple M1 GPU	0.054393

Categorical Analysis Findings

10. storage_type

Sampling Distribution of 'storage_type'

Sampling Distribution of storage_type for sample size 10
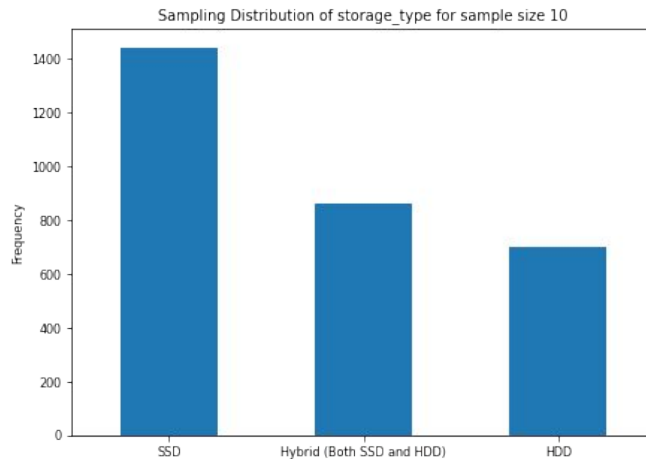
Mode of Sample Data:

The Mode of the Column storage_type is SSD and its sample proportion is 0.4723333333333333

Categorical Analysis Findings

10. storage_type

Value Counts of each category in 'storage_type'

Frequency of column storage_type of Sampling Distribution(Sample Size = 10) is

SSD	1417
Hybrid (Both SSD and HDD)	878
HDD	705

Comparing Population Proportions and Sample Proportions

Sample Proportions:

SSD	0.476667
HDD	0.238333
Hybrid (Both SSD and HDD)	0.285000

Population Proportions:

SSD	0.472803
Hybrid (Both SSD and HDD)	0.292887
HDD	0.234310

Categorical Analysis Findings

Sampling Distribution of 'hdd_size'

Sampling Distribution of hdd_size for sample size 10
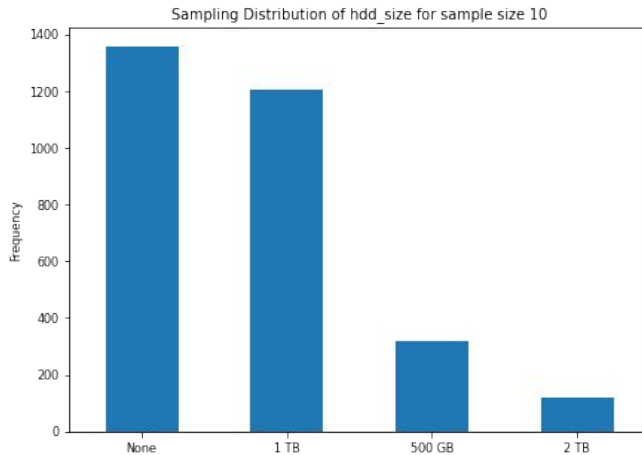
Mode of Sample Data:

The Mode of the Column hdd_size is None and its sample proportion is 0.4523333333333333

Categorical Analysis Findings

11. hdd_size

Value Counts of each category in 'hdd_size'

Frequency of column hdd_size of Sampling Distribution(Sample Size = 10) is

None	1357
1 TB	1207
500 GB	317
2 TB	119

Comparing Population Proportions and Sample Proportions

Sample Proportions:		Population Proportions:	
None	0.462333	None	0.460251
1 TB	0.392667	1 TB	0.397490
500 GB	0.107000	500 GB	0.104603
2 TB	0.038000	2 TB	0.037657

Categorical Analysis Findings

12. ssd_size

Sampling Distribution of 'ssd_size'

Sampling Distribution of ssd_size for sample size 10
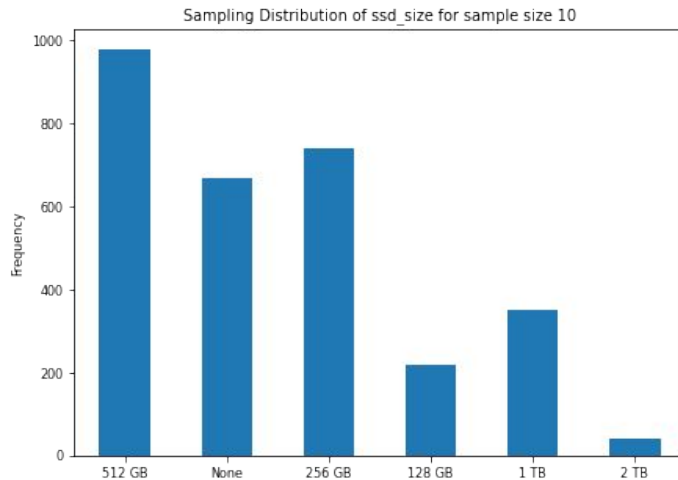
Mode of Sample Data:

The Mode of the Column ssd_size is 512 GB and its sample proportion is 0.32566666666666666

Categorical Analysis Findings

12. ssd_size

Value Counts of each category in 'ssd_size'

Frequency of column ssd_size of Sampling Distribution(Sample Size = 10) is

512 GB	977
None	669
256 GB	740
128 GB	220
1 TB	352
2 TB	42

Comparing Population Proportions and Sample Proportions

Sample Proportions:		Population Proportions:	
512 GB	0.342000	512 GB	0.338912
None	0.227667	256 GB	0.242678
128 GB	0.073667	None	0.221757
1 TB	0.106667	1 TB	0.112971
256 GB	0.237000	128 GB	0.071130
2 TB	0.013000	2 TB	0.012552

13. operating_sys

Sampling Distribution of 'operating_sys'

Sampling Distribution of operating_sys for sample size 10
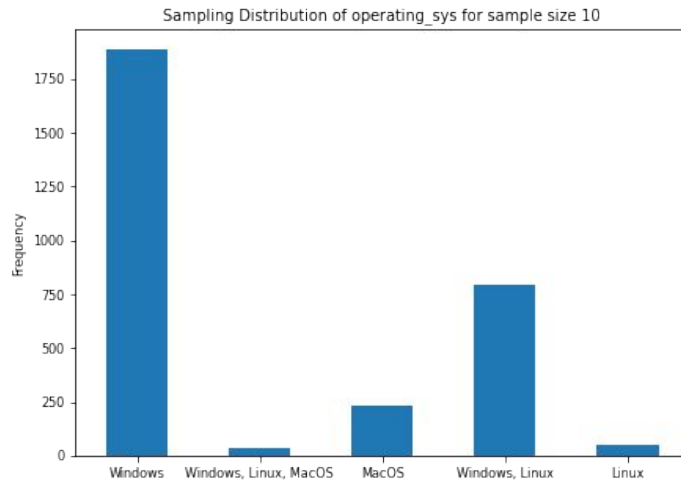
Mode of Sample Data:

The Mode of the Column operating_sys is Windows and its sample proportion is 0.629

Categorical Analysis Findings

13. operating_sys

Value Counts of each category in 'operating_sys'

Frequency of column operating_sys of Sampling Distribution(Sample Size = 10) is

Windows	1887
Windows, Linux, MacOS	38
MacOS	233
Windows, Linux	795
Linux	47

Comparing Population Proportions and Sample Proportions

Sample Proportions:

Windows	0.619000
MacOS	0.088000
Windows, Linux	0.261333
Linux	0.020667
Windows, Linux, MacOS	0.011000

Population Proportions:

Windows	0.619247
Windows, Linux	0.267782
MacOS	0.079498
Linux	0.020921
Windows, Linux, MacOS	0.012552

14. department

Sampling Distribution of 'department'

Sampling Distribution of department for sample size 10
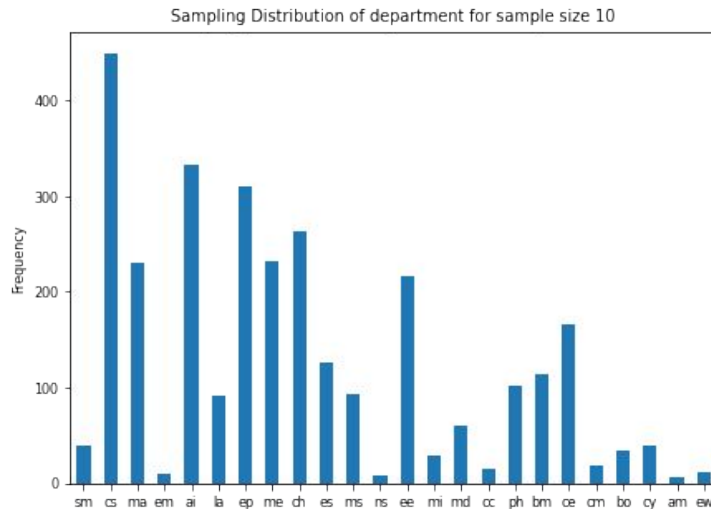
Mode of Sample Data:

The Mode of the Column department is cs and its sample proportion is 0.14966666666666667

Categorical Analysis Findings

14. department

Value Counts of each category in 'department'

Frequency of column department of Sampling Distribution(Sample Size = 10) is

sm	40	ch	263	ph	102
cs	449	es	126	bm	114
ma	230	ms	93	ce	167
em	10	ns	8	cm	18
ai	333	ee	217	bo	34
la	91	mi	30	cy	40
ep	310	md	60	am	6
me	232	cc	15	ew	12

Categorical Analysis Findings

14. department

Comparing Population Proportions and Sample Proportions

Sample Proportions:

ma	0.079333	ch	0.082333	cm	0.007000
ep	0.102667	ph	0.036667	ew	0.005667
ai	0.103333	es	0.045000	am	0.004333
ee	0.074000	ce	0.058333	bo	0.009000
me	0.084333	md	0.019667	mi	0.012000
la	0.030667	ms	0.025333	em	0.004000
cs	0.148667	sm	0.011000	ns	0.006000
bm	0.030333	cy	0.014333	cc	0.006000

Population Proportions:

cs	0.154812	es	0.041841	mi	0.008368
ep	0.104603	bm	0.037657	bo	0.008368
ai	0.100418	ph	0.037657	am	0.004184
me	0.087866	la	0.029289	ew	0.004184
ch	0.083682	ms	0.029289	em	0.004184
ma	0.079498	md	0.016736	cm	0.004184
ee	0.075314	cy	0.012552	ns	0.004184
ce	0.054393	sm	0.012552	cc	0.004184

Laptop Specifications

POINT ESTIMATION

Point Estimation

❏ For Point Estimation, we took 100 samples randomly from the population and calculated the Expectation of that random Sample.

❏ Point estimators are functions that are used to find an approximate value of a population parameter from random samples of the population. They use the sample data of a population to calculate a point estimate or a statistic that serves as the best estimate of an unknown parameter of a population.

❏ A point estimate of the mean of a population is determined by calculating the mean of a sample drawn from the population. The calculation of the mean is **the sum of all sample values divided by the number of values**.

CONFIDENCE INTERVAL

Confidence Interval

- ❏ We found the Confidence Interval of the Mean with 97.5% Confidence using the above Point Estimates.
- ❏ We found the value of z_0.125 and Standard Deviation of the Population Data.
- ❏ Based on the below formula, we estimated the Confidence Interval of the Mean and Margin of the Error of the Mean where x_bar is the Point Estimator of Mean.

$$\overline{x} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}} \right)$$

Confidence Interval

1. Ram Size
❑ For n = 100 samples, Point Estimator of Mean is 11.96, $z_{0.125}$ = 1.96.
❑ 97.5% Confidence Interval of the Mean is given by (11.451009049878326, 14.068990950121673)
❑ Margin of Error is 1.3089909501216725

2. Average Watch Time
❑ For n = 100 samples, Point Estimator of Mean is 460.34
❑ 97.5% Confidence Interval of the Mean is given by (415.86363816461096, 515.4763618353891)
❑ Margin of Error is 49.80636183538905

3. Year Join
❑ For n = 100 samples, Point Estimator of Mean is 19.75
❑ 97.5% Confidence Interval of the Mean is given by (19.465650554793775, 19.954349445206226)
❑ Margin of Error is 0.24434944520622548

Conclusion

From the analysis we can conclude that
- ❖ Many students preferred HP laptop
- ❖ Most of them are using Windows
- ❖ Display Type - Non-Touch
- ❖ Graphic Card - Nvidia Graphics
- ❖ CPU - Intel (Intel i5)
- ❖ CPU Clock Speed - 2-3GHz
- ❖ Storage Type - SSD
- ❖ SSD - 512GB
- ❖ HDD - 1TB
- ❖ Price range between 60,000-70,000 or 90,000 and above
- ❖ Display size - 14-15 inches or 15-16 inches
- ❖ RAM - 8GB
- ❖ Average Watch Time is around 400 - 500 minutes.