



भारतीय प्रौद्योगिकी संस्थान हैदराबाद  
Indian Institute of Technology Hyderabad

## **Applied Statistics - MA4240**

### **Survey on Non-Alcoholic Sanitizer Product**

#### **Group Members (Group No. 7)**

| <b>Group Member Name:</b>    | <b>Roll Number:</b> |
|------------------------------|---------------------|
| Lokesh J. K.                 | AI21MTECH14001      |
| Maddula Sai Sunamdha Harinhi | AI21MTECH14002      |
| Kaushiki Dwivedi             | AI21MTECH14003      |
| Tejas Arya                   | AI21MTECH14004      |
| Anant Mittal                 | AI21MTECH14005      |
| Kashyap Parmar               | AI21MTECH14006      |
| Karthikeyan Mohanraj         | AI21MTECH14007      |
| Darshan Dobariya             | AI21MTECH14008      |
| Varshita Sharma              | AI21MTECH14009      |
| Bhavya Kanzariya             | AI21MTECH11002      |

## **Contribution of each member :**

### **Data Collection, Preprocessing and Sample Statistics -**

Tejas Arya, Lokesh J. K. & Varshita Sharma

### **Central Limit Theorem -**

Anant Mittal

### **Hypothesis Testing -**

Bhavya Kanzariya & Karthikeyan Mohanraj

### **Graph Plotting -**

Kaushiki Dwivedi & Maddula Sai Sunamdha Harinhi

### **Report and PPT-**

Kashyap Parmar & Darshan Dobariya

## **Table of Contents**

| <b>Sr No.</b> | <b>Topic</b>                       | <b>Page No.</b> |
|---------------|------------------------------------|-----------------|
| 1.            | Acknowledgement                    | 4               |
| 2.            | Abstract                           | 5               |
| 3.            | Introduction                       | 6               |
| 4.            | Data Description and Visualization | 7               |
| 5.            | Sample Statistics                  | 18              |
| 6.            | Testing the Central Limit Theorem  | 19              |
| 7.            | Hypothesis Testing                 | 21              |
| 8.            | Conclusions                        | 25              |
| 9.            | References                         | 26              |

## **Acknowledgement**

We want to take the opportunity to thank all the people without whom reaching this stage would not have been possible. I want to thank our mentor, Prof. SAMEEN NAQVI, who has been a great source of inspiration and provided us with the space to work in our way.

We would also like to pay our gratitude to our team for extending their help in every possible way and providing us with such an enriching and out-of-the-box project.

We would also like to express our special thanks to all the 259 people who had participated in our survey, and enlightened us with their viewpoint, which had been a crucial part of this project.

Lastly, we would like to thank our family and friends for their never-ending encouragement and reliance on us, all through this memorable journey.

# **Abstract**

The rise of COVID-19 cases worldwide has been a severe challenge as global deaths have increased exponentially. This virus can be transmitted from one person to another by inhaling an individual's infected respiratory droplets (Chu et al., 2020).

Hands are the primary transmission mode of viruses and microbes, Mondal and Kolhapure(2004). It has been estimated that there are not less than 10000 organisms per cm<sup>2</sup> of normal skin. This includes both non-pathogenic resident flora and pathogenic transient flora (Carter et al., 2000).

Before Covid-19, only 50.8% of respondents used hand sanitizer; this number grew to 86.4% during Covid-19, according to research (Vidyabharati Intl, 2021). The growth in usage of sanitizers requires awareness about the existing options of sanitizer available. A review of hand sanitizers is discussed below, followed by the comparative evaluation of the efficacy of alcoholic and non-alcoholic sanitizers as both are equally effective in fighting Covid-19 (BCIT Health Sc., 2016).

# **Introduction**

Proper hand hygiene saves lives. Throughout the COVID-19 pandemic, one of the most important health messages we've heard is the need of thoroughly washing our hands to remove any harmful germs from our skin. The Coronavirus pandemic which started in late 2019 has brought a huge demand for hygiene products, primarily sanitizers sales have grown with huge numbers and because of the new formed awareness the sales of sanitizers will continue to grow. By removing germs from our hands, we can avoid contaminating other surfaces and things that may be touched by others. However, we must consider more than only the coronavirus. Hand cleanliness is critical in preventing the transmission of infections ranging from respiratory to stomach ailments. Hence, handwashing should be a lifelong practice even after the pandemic.

## **Disadvantages of Alcohol Based Sanitizers:**

1. Alcohol Toxicity in sanitizer.
2. Dry Skin and risk of infection
3. Fire hazard

## **Benefits of Non-Alcoholic Sanitizers:**

1. Safe for skin.
2. Provides sanitation for longer duration.
3. Poses no threat to fire

## **Target Audience**

Furthermore, initially, due to the characteristics like non-flammability, skin protection, and slight higher price of the sanitizer, we narrowed down to the target audience as follows:

1. Offices and industries- safe to use because it is fireproof
2. Schools - safe for delicate skin
3. Hotel/ Restaurant kitchen- safe usage, non-flammability
4. Hospitals- solves the adverse effects on skin problem

# Data Description and Visualization

## Data Description

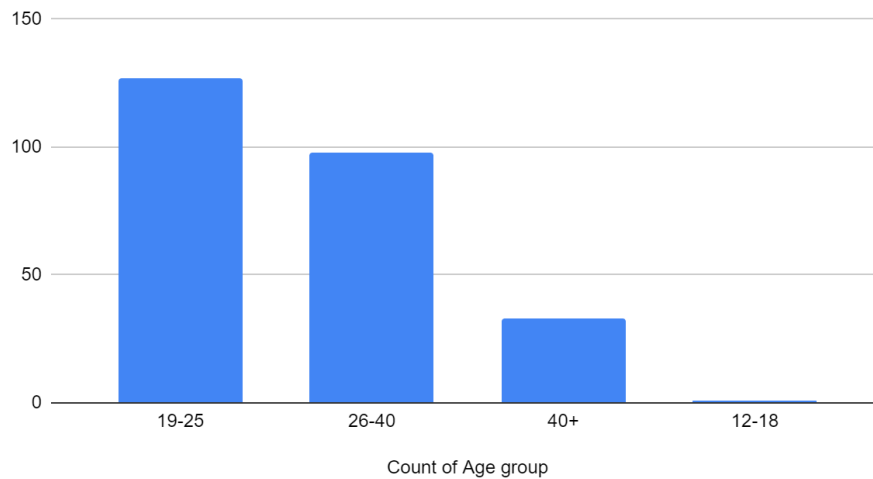
Data was primarily collected from the target audience.

We distributed our google form among our family members and friends for data collection. Each of us floated an online survey form to around 100+ people. Each of the ten group members contributed about 20 to 30 responses. We used random sampling for data collection.

## Data Visualization

- Demography of the population on which survey has been collected.

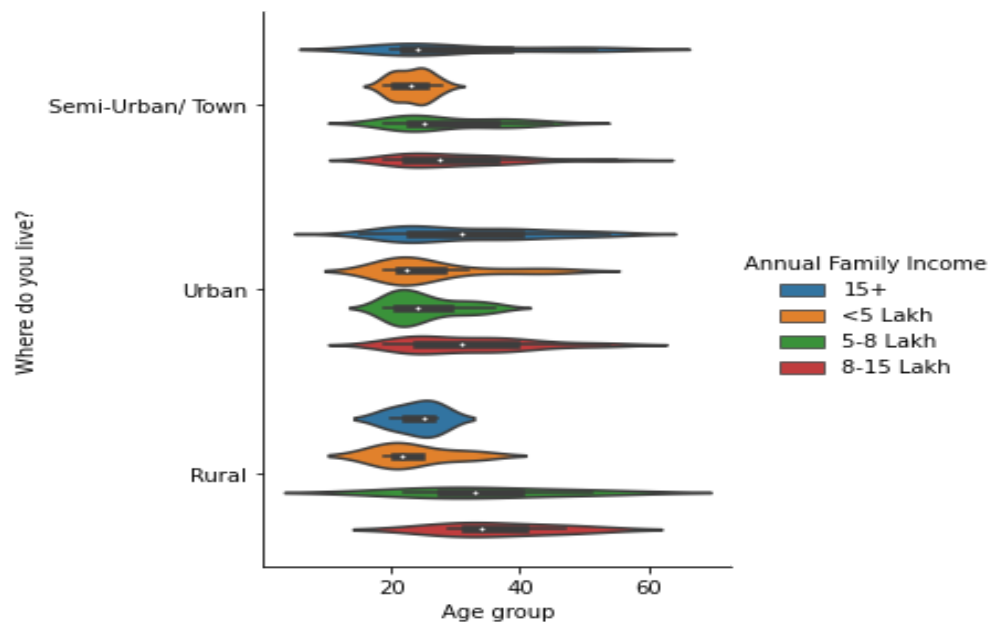
### 1.Count of Age group:



**Fig: 1**

**Observation:** In our survey around 130 people belong to the age group of 19-25, around 90 in the age group of 26-40 and the rest 40 in the age group 40+.

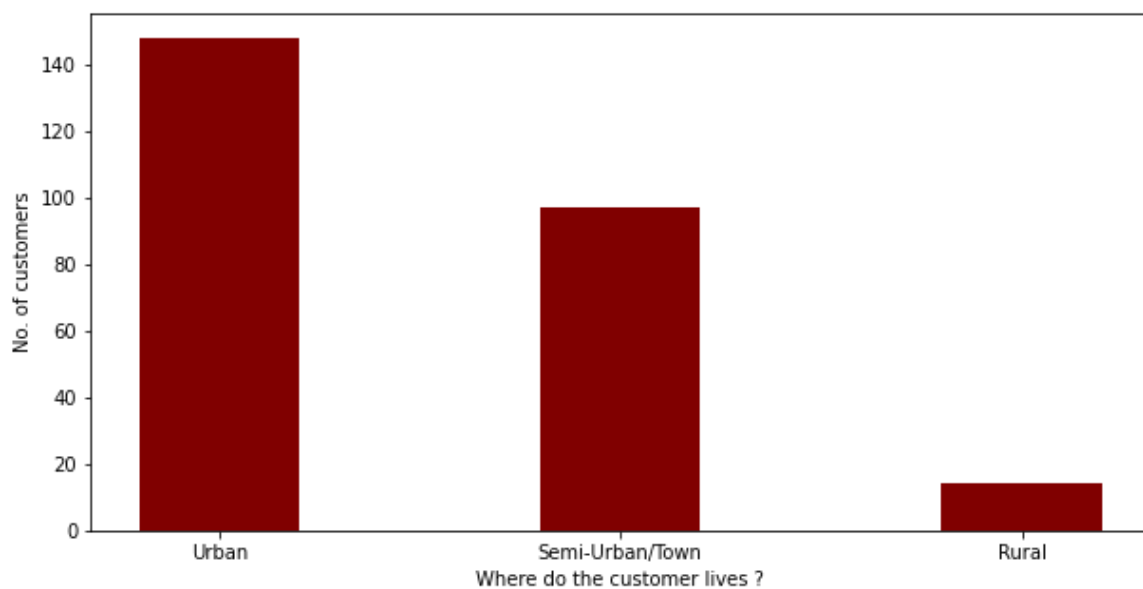
## 2. Distribution of survey of people's age group in various areas with annual family income:



**Fig: 2**

**Observation:** This graph explains the demography of the population from whom data is collected. Also it can be seen that as Age group increases Family income also increases.

## 3. Customers residence:



**Fig: 3**

**Observation:** The plot shows that more respondents live in urban areas as compared to the semi-urban/town and rural areas.



#### 4. Customers and annual family income distribution:

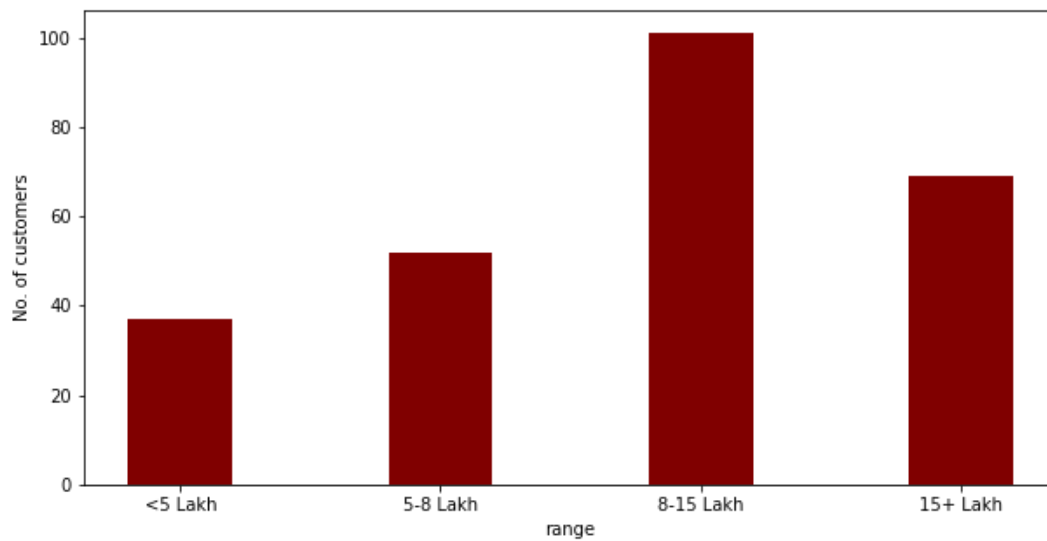


Fig: 4

#### Observation:

The plot shows that annual family income of most respondents lie within the range of 8-15 lakh, our population parameters have wide diversity in case of annual family income.

#### 5. Distribution of “Where the customer lives?”-vs-”Age group”:

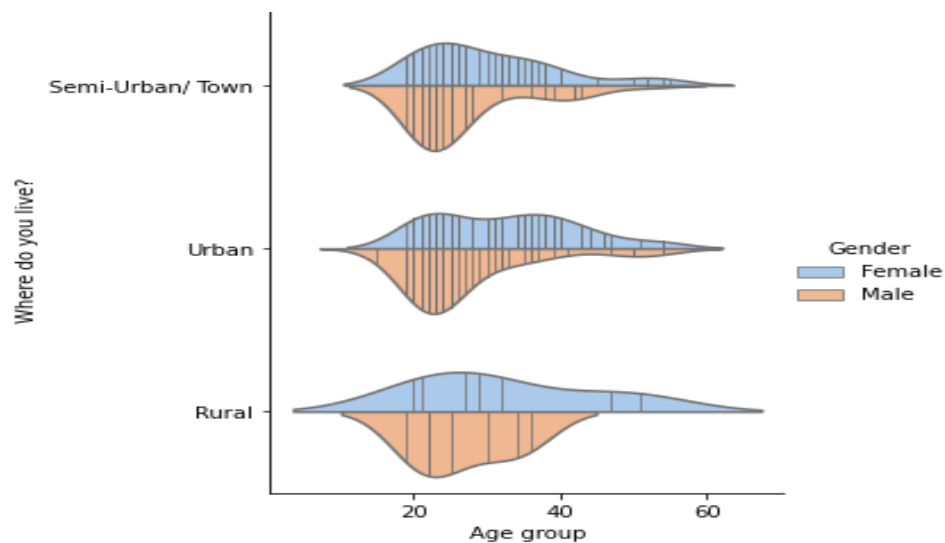
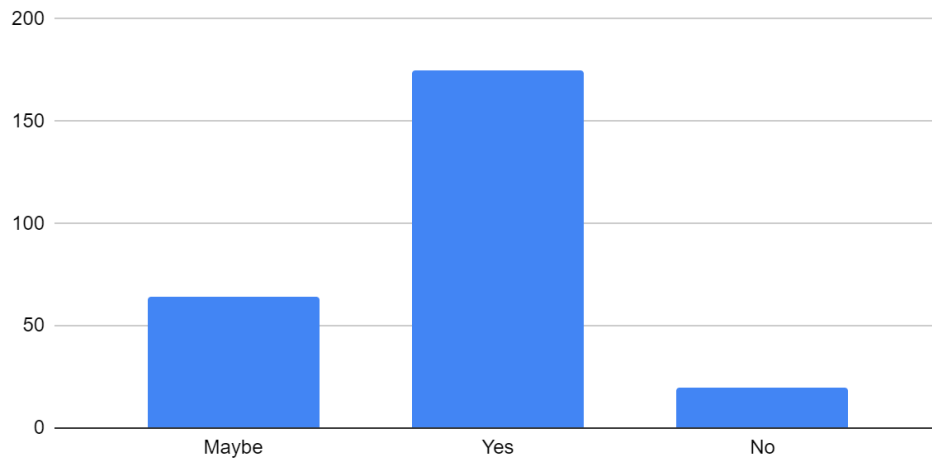


Fig: 5

**Observation:** This graph shows the complete demography of our target population from whom data is collected.

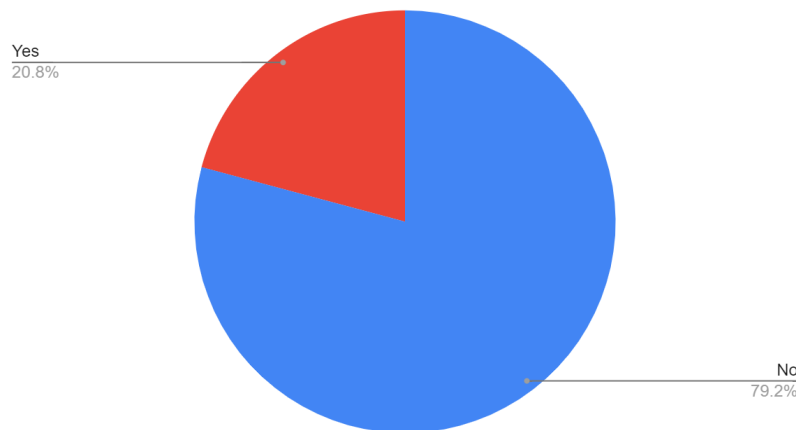
## 6.Distribution of “Would you like to try non-alcoholic sanitizer”?



**Fig: 6**

**Observation:** Amongst the survey conducted on how many people wanted to try non-alcoholic sanitizer most of the people said yes(70%), whereas around 25% people are confused and the rest 7% are not ready to use Non-alcoholic hand sanitizer.

## 7.Count of “Do you know about Non-alcoholic sanitizer”



**Fig: 7**

**Observation:** About 80% of Total population are not aware of Non-alcoholic Hand sanitizer and only 20% are aware of Non-alcoholic hand sanitizer.

### 8.Count of “Do your Sanitizer have alcohol”

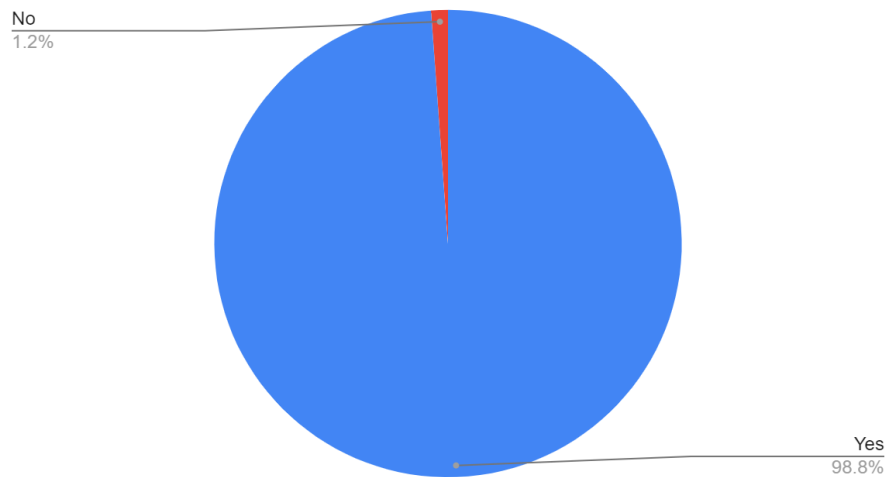


Fig: 8

**Observation:** Almost all the people are using Alcoholic Hand Sanitizer. There is an urgent need of awareness among people regarding disadvantages of alcoholic sanitizer and benefits of Non-alcoholic hand sanitizer.

### 9.Where do you prefer to buy hand sanitizer?

259 responses

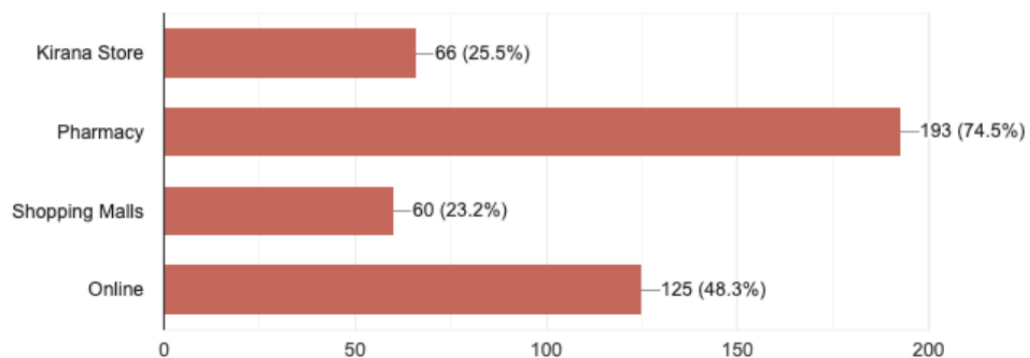
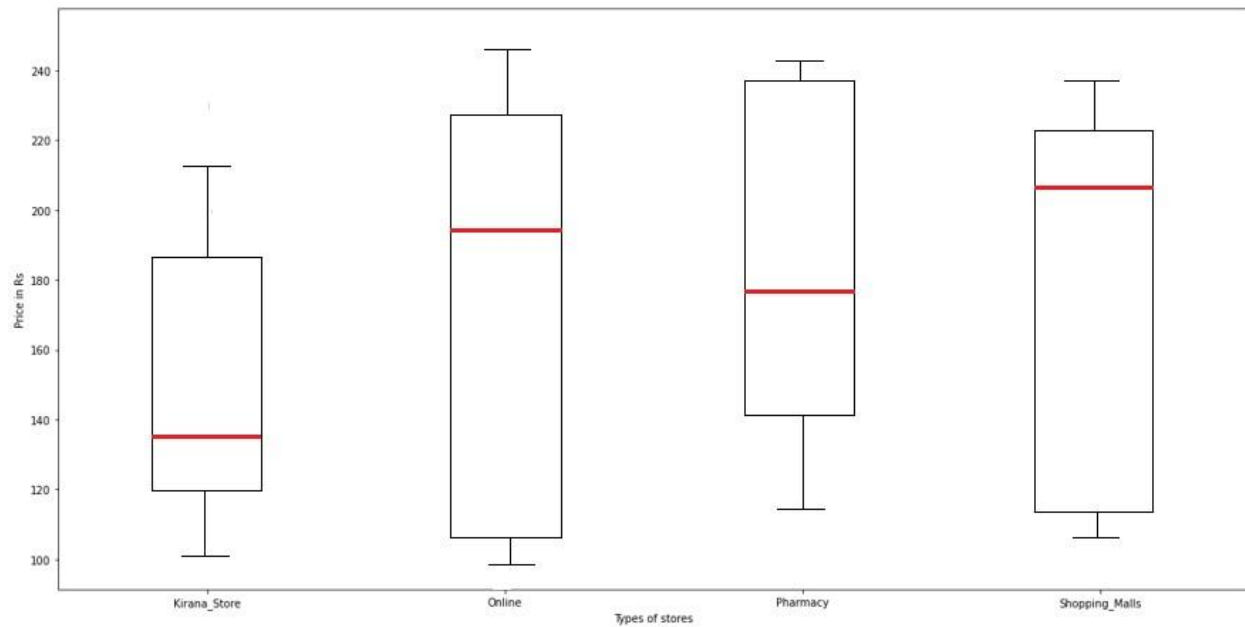


Fig: 9

**Observation:** Most of the people prefer to buy sanitizer from pharmacies followed by online mode and then by kirana store and shopping mall respectively.

## 10.Reason why people prefer to buy hand sanitizer from particular places.

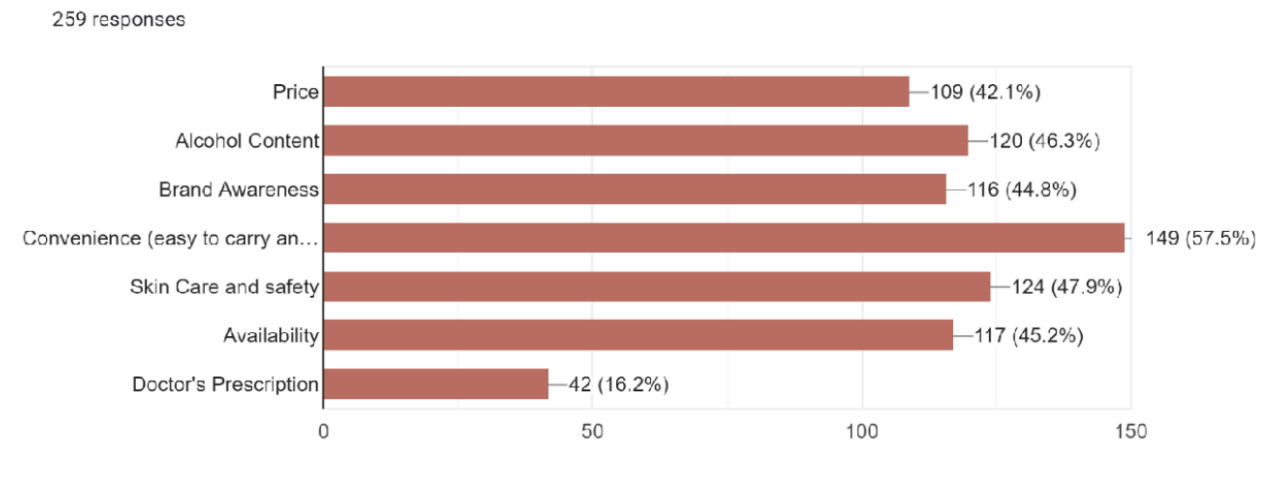


**Fig: 10**

### Observations:

- The above boxplot describes the relationship between different types of stores w.r.t the Sanitizer Price (in Rs).
- The observations made are that the shopping malls and online stores keep very high rating prices of the sanitizer.
- While it is found that the pharmacy and kirana stores keep the sanitizer price quite cheap.

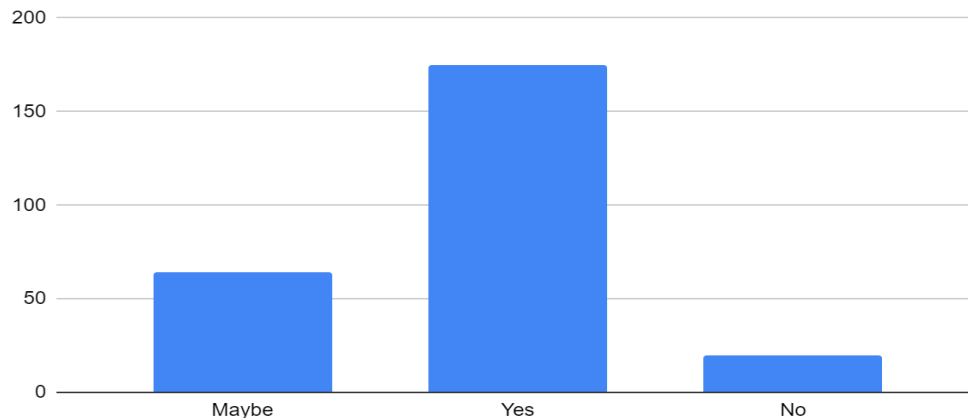
### 11.What factors do you prefer while buying sanitizer for yourself ?



**Fig: 11**

**Observation:** From the above bar plot it can be inferred that people are most focussed on convenience (easy to carry and easy to spread), so any company going to launch Non-alcoholic hand sanitizer give more focus on sanitizer Bottle shape, Bottle size, Nozzle.

### 12.Count of Would you like to try Non-alcoholic Sanitizer ?



**Fig: 12**

**Observation:** Amongst the survey conducted on how many people wanted to try non-alcoholic sanitizer most of the people said yes(70%) , whereas around 25% people are confused and the rest 7% are not ready to use Non-alcoholic hand sanitizer.

### 13. Sanitizer Price vs Would you like to try Non-Alcoholic Sanitizer :

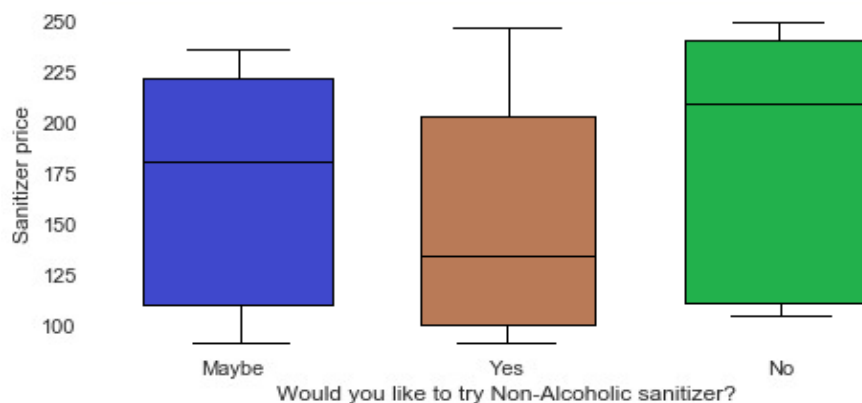


Fig: 13

**Observation:** The columns “sanitizer price” and “would you like to try non-alcoholic sanitizer” are independent of each other.

### 14. “Where do you live ?” vs “Would you like to try a non-alcoholic sanitizer ?”

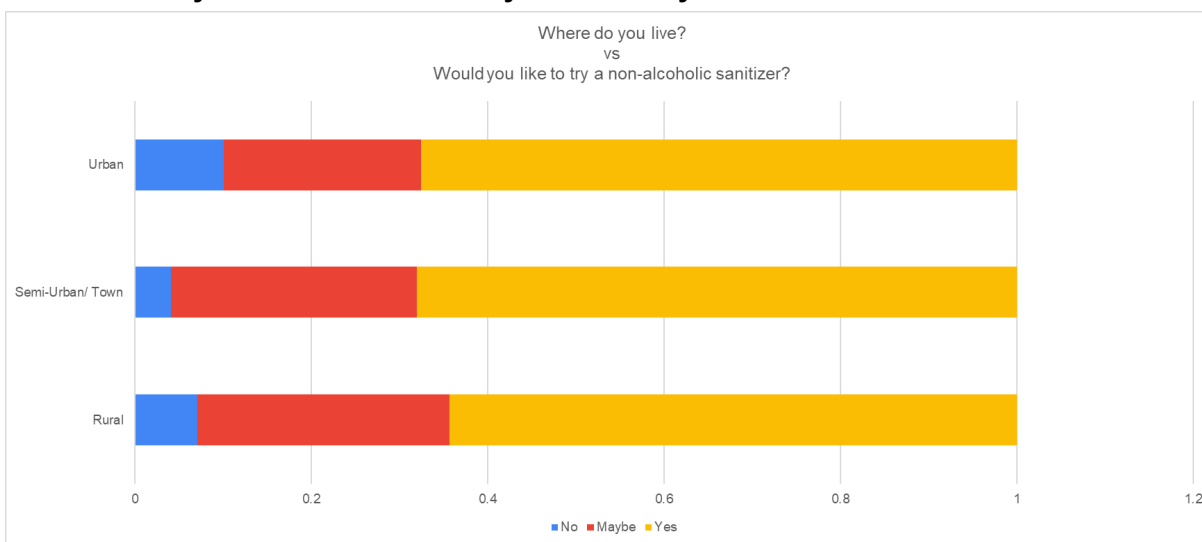


Fig: 14

**Observation:** - There is a minute difference between the rural and urban areas. Almost equal proportion of people are in favor of trying Non-alcoholic sanitizer. So any company willing to launch their product should have an equal focus in all the three regions.

15.

i) Plot of “People who have faced problems due to alcoholic sanitizer” vs “Count of people who like to try non-alcoholic sanitizer”.

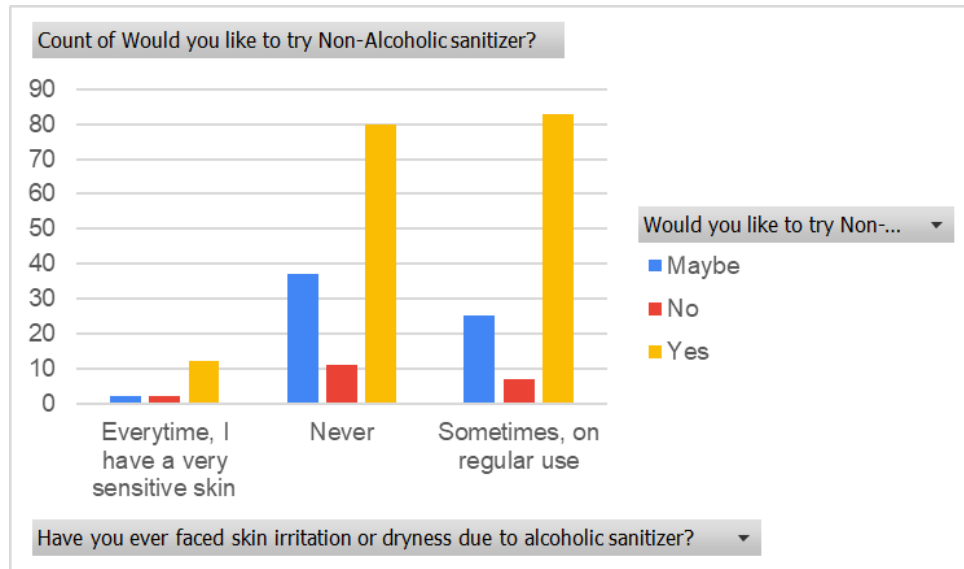


Fig: 15.a

**Observation:** - From this graph it is evident that people who have faced skin irritation and with sensitive skin are more interested to try Non-alcoholic hand sanitizer followed by people who have sometimes faced skin irritation. People who have never faced skin irritation are also interested to try Non-alcoholic hand sanitizer.

ii) Plot of “People who have faced problems due to alcoholic sanitizer” vs “Proportion of people who like to try non-alcoholic sanitizer”

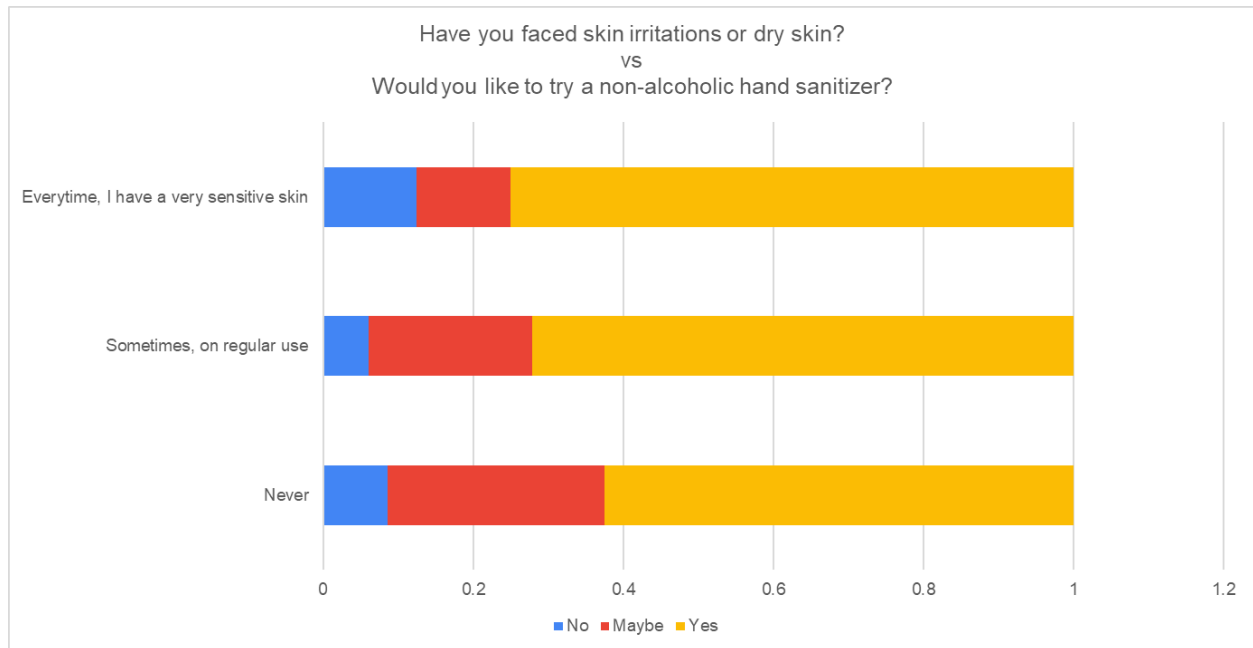


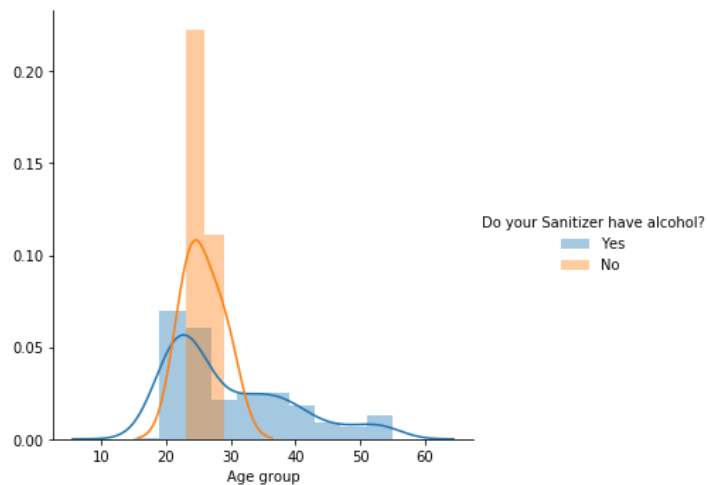
Fig: 15.b

**Observation:** - This graph also shows the same observation as above. It can be seen that people with sensitive skin have the highest interest toward trying Non-alcoholic hand sanitizer.



## Histogram & PDF:

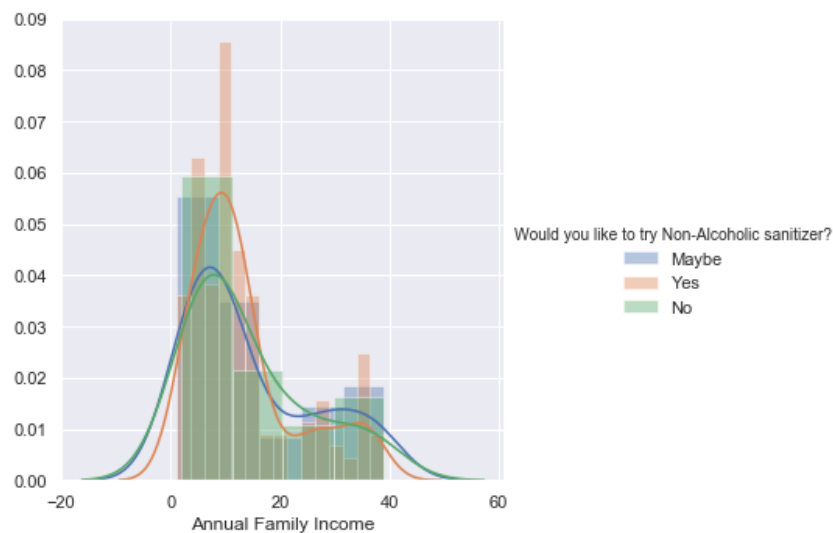
**16. Plot showing the pdf of Age group vs “Do your sanitizer have alcohol?”:**



**Fig: 16**

**Observations:** There is massive overlap between the probability distributions (pdf) of people based on age group.

**17. Probability distribution function of “Would you like to try Non-Alcoholic sanitizer?” based on Annual Family Income:**



**Fig: 17**

**Observation:**

There is massive overlap between the probability distributions (pdf) of people based on their annual family income who are willing to use/non-willing to use and may-be willing to try non-alcoholic sanitizer .

# **Sample Statistics:**

## **Confidence Interval Interpretation:**

### **Theorem:**

Assume:

- 1)  $X_1, X_2, \dots, X_n$  is a random sample from a normal population

with mean  $\mu$  and variance  $\sigma^2$ , so that,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } Z = \frac{(\bar{X} - \mu)}{(\sigma/\sqrt{n})} \sim N(0,1)$$

- 2) The population variance  $\sigma^2$  is known.

Then, a  $(1 - \alpha)$  100% CI for the mean  $\mu$  is:

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

The interval, because it depends on Z, is often referred to as the Z-interval for a mean.

### **Calculated Values:**

- Sample mean = 173.86 INR
- Sample Standard Deviation = 54.87 INR
- To have 95% confidence interval to be  $\pm 10$ ,  $n = 116$  calculated using :

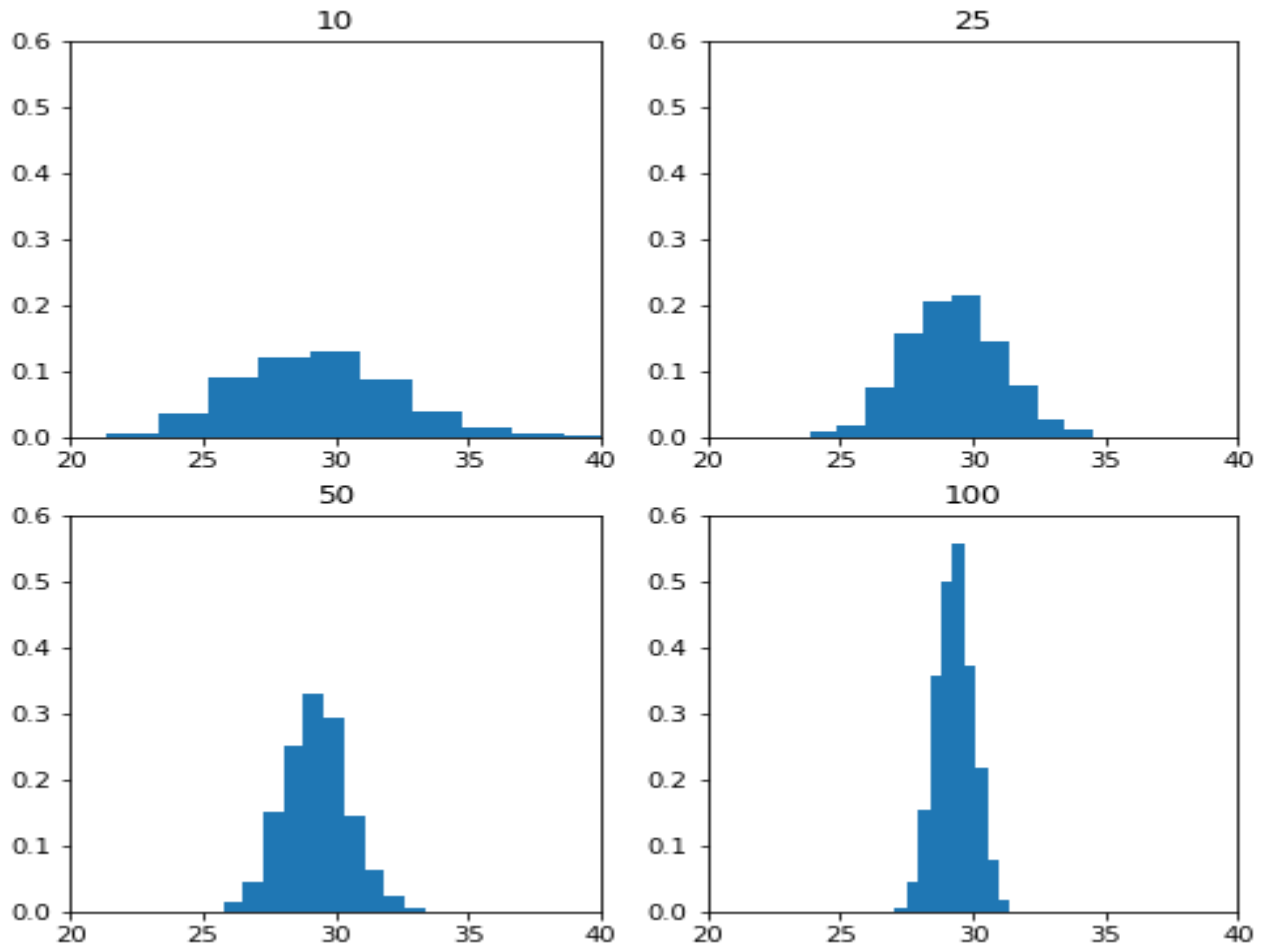
$$n = \frac{(z_{\alpha/2})^2 S^2}{E^2}, \text{ where } E = 10 \text{ (half interval), } z_{\alpha/2} = 1.96$$

|                | Age Group | Freq of using hand sanitizer in a day | Sanitizer Price |
|----------------|-----------|---------------------------------------|-----------------|
| Sample Mean    | 29.95     | 4.27                                  | 180.06          |
| Sample Std Dev | 9.85      | 3.32                                  | 54.71           |
| Std Error      | 0.58      | 0.19                                  | 3.41            |

**Table: 1**

# Testing the Central Limit Theorem

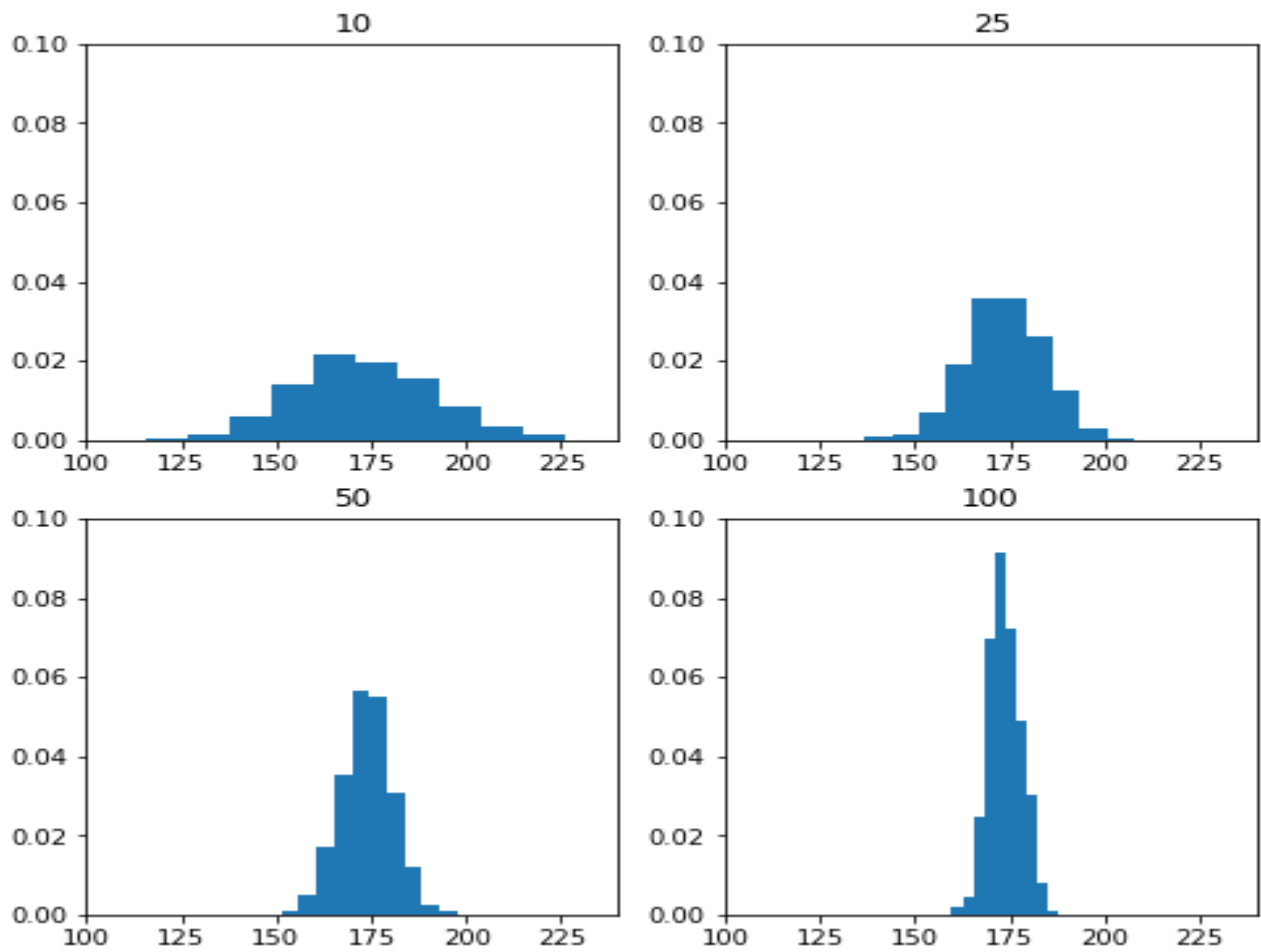
## CLT for Age group



**Fig: 18**

Distribution of Sample Mean of Age for sample sizes of 10, 25, 50 and 100.

### CLT for Sanitizer Price:



**Fig: 19**

Distribution of Sample Mean of Sanitizer Price for sample sizes of 10, 25, 50 and 100.

# Hypothesis Testing

## Hypothesis 1

Suppose we want to check whether or not people with children spend at least ₹25 more than people without children. With 95% confidence level.

The dataset is divided into two parts, Y for people without children and X for people with children.

Here, we have a difference of mean problem, Our null hypothesis and alternate hypothesis are

$$H_0 : \mu_Y - \mu_X \geq 25$$

$$H_a : \mu_Y - \mu_X < 25$$

Looking at the both sample variances, the ratio of both the variances is less than 4.

$$\frac{S_Y^2}{S_X^2} < 4$$

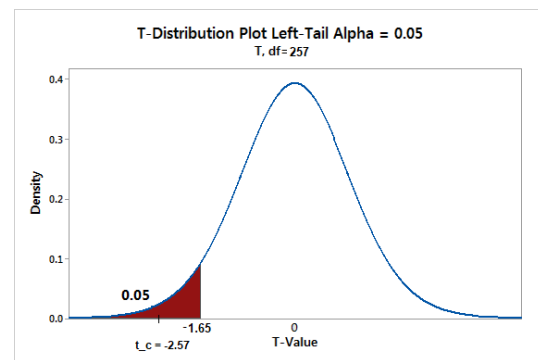
$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{m+n-2}$$

Replacing the values,  $S_X = 56.73, S_Y = 52.70$

$$S_p = 54.95$$

Since, the population variance is unknown, the t-distribution is used, the critical t value is

$$t_c = \frac{(\bar{Y} - \bar{X}) - (\mu_Y - \mu_X)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$
$$t_c = \frac{(7.35) - (25)}{54.95 \sqrt{\frac{1}{142} + \frac{1}{117}}}$$
$$t_c = \frac{-17.65}{6.86}$$
$$t_c = -2.57 < t_{\alpha, n+m-2} = -1.65$$



Since, the critical t value lies in the rejection region, the hypothesis is thus rejected.

## Hypothesis 2:

**Females(Mothers) who are buying sanitizers are looking into alcohol content before buying sanitizer.**

$P_0$  : Proportion of females who have children considering alcohol content while buying Sanitizer.

$P_1$  : Proportion of females who don't have children considering alcohol content while buying Sanitizer.

**NULL Hypothesis  $H_0$  :**  $P_0 - P_1 = 0$

**Alternative Hypothesis  $H_a$  :**  $P_0 - P_1 \neq 0$

Below table show the data of gender vs with or without children for those who consider alcohol content while buying sanitizer.

|        | No Children | With Children |
|--------|-------------|---------------|
| Female | 21          | 43            |
| Male   | 42          | 14            |
| Total  | 63          | 57            |

**Table: 2**

Number of samples:

$$n_0 = 63$$

$$n_1 = 57$$

**Checking for z distribution approximation:**

$$\begin{aligned}\hat{p}_0 n_0 &= 0.754 \times 57 = 42 > 5 \\ n_0 (1 - \hat{p}_0) &= 0.246 \times 57 = 14 > 5 \\ n_1 \hat{p}_1 &= 0.33 \times 63 \approx 21 > 5 \\ n_1 (1 - \hat{p}_1) &\approx 42 > 5\end{aligned}$$

Let us fix  $\alpha = 0.05 \Rightarrow$ critical value  $Z_{\alpha/2} = 1.96$

**Calculation:**

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{43 + 21}{57 + 63} = 0.534$$

$$z = \frac{(0.754 - 0.33) - (0)}{\sqrt{0.534 (0.466) \left(\frac{1}{57} + \frac{1}{63}\right)}}$$
$$z = 4.67$$

But since,  $z_{\alpha/2} = 1.96, z > z_{\alpha/2}$

Hence, we are rejecting the NULL hypothesis.

### **Hypothesis 3**

**People are more likely to consider Safety Factors while buying Sanitizer for their kids/ younger siblings as opposed to while buying for themselves (with 95% confidence level)**

$p_1$  : proportion of people considering safety as a factor while buying the sanitizer for their kids/younger siblings

$p_2$  : proportion of people considering safety as a factor while buying the sanitizer for themselves.

$$H_0 : (p_1 - p_2) \geq 0$$

$$H_1 : (p_1 - p_2) < 0$$

Standard Error of our distribution,

$$\begin{aligned} S.E.(p_1 - p_2) &= \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \\ &= 0.041 \end{aligned}$$

Rejection region for hypothesis is  $z < 1.645$ .

$$\begin{aligned} z_c &= \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{S.E.} \\ &= \frac{0.264 - 0}{0.041} \\ &= 6.43 \end{aligned}$$

But since critical z value is not in the rejection region, we fail to reject the null hypothesis.



## **Conclusions**

1. We can deduce that many factors affect the sale of non-alcoholic sanitizers.
2. Significant factors include age group, demography, and children count.
3. Overall, it can be concluded that there is a vast scope of Non-alcoholic sanitizers in the market as the overall sanitizer market is multiplying and will continue to grow for the next 4-5 years as well.

## **References**

- <https://www.mdpi.com/1660-4601/17/9/3326/htm>
- <https://journals.bcit.ca/index.php/ehj/article/view/102>
- <https://www.businesswire.com/news/home/20201221005262/en/Hand-Sanitizer-Market-in-India-to-Grow-by-405.31-During-2020-2024-Industry-Analysis-Market-Trends-Market-Growth-Opportunities-and-Forecast-2024-Technavio>
- <https://www.lexology.com/library/detail.aspx?g=b44ab4bf-f596-4747-84bc-ec93c28b9ee7>
- <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.300.9359&rep=rep1&type=pdf>
- [https://docs.google.com/forms/d/e/1FAIpQLScvIN7CFWsgOT07FOphVno6Pb4u\\_ms1a25U1T902xxnrl6agA/viewform](https://docs.google.com/forms/d/e/1FAIpQLScvIN7CFWsgOT07FOphVno6Pb4u_ms1a25U1T902xxnrl6agA/viewform)