SAMPLING THEORY

Experiential Learning - Assignment 1

USE OF SAMPLING TECHNIQUES IN DISTRIBUTION
CHANNELS

MSc in Applied Statistics

Symbiosis Statistical Institute, Pune

Report By,

Team 2 – "Sampling Error"

Amirta V - 21060641004

Anoushka Fernandes - 21060641006

Aparna Biswas - 21060641008

Shubhangi Deshmukh - 21060641048

Sudhanshu Dandriyal - 21060641050

Sumantra Kumar - 21060641051

Vidya Mallick – 21060641056

TABLE OF CONTENTS

TABLE OF CONTENTS

ACKNOWLEDGEMENT

We would like to express our sincere gratitude towards Symbiosis Statistical Institute for providing opportunities for a holistic development.

We would like to extend our gratitude to our professor Mr. Mayur Savsani, for giving us the opportunity to explore the topic, "Impact of Sampling Techniques in Distribution Channels". The experience was much of a fruitful one. The research involved in the topic would definitely help us receive a much better understanding over the subject, which is a rather important one in the field of statistics.

The Experiential Learning also helped us work on our team building skills. It also helped us gain efficiency in communication, time management and peer learning. The educational experience would definitely prove to be beneficial to all the students.

ABSTRACT

India is the second-largest producer and third largest exporter of tobacco. Nearly 30% of the Indian population consumes tobacco despite its evident ill effects on our health. The tobacco industry contributes over 11 lakh crore rupees to the economy while employing more than 45 million people.[1] The tobacco industry can be segmented into two i.e. non-smoking tobacco and smoking tobacco. Broadly speaking the distribution channels of tobacco include the manufacturers/ producers, the wholesalers, and the retailers. With the ban on direct advertising of tobacco products, it is very important for the business to analyse the market share of each and every segment and strategize accordingly. In order to optimize costs and ensure effective outreach strategies, businesses not only segment consumers based on their behaviour but also make changes in their distribution channels. Sampling techniques play a foundational role in understanding consumer behaviour and also channelizing distribution processes to enable efficient business operations.

SEGMENT 1

1.1 INTRODUCTION ON DISTRIBUTION CHANNELS

A distribution channel is a network of intermediaries that allows a product to be distributed from the manufacturer to the end user. Distributors, wholesalers, retailers, and e-commerce intermediaries are just a few of the various intermediaries. A company's overall marketing strategy for promoting, pricing, and selling a product to consumers includes distribution channels. Distribution channels are broadly categorised in two different forms such as direct channels and indirect channels. Through a direct channel, a consumer can buy directly from the manufacturer. A consumer obtains goods through an intermediary, such as a wholesaler or a retailer, in an indirect channel.

Longer distribution channels can also mean less profit each intermediary charges a manufacturer for its service.

A manufacturer should determine the distribution channel or method of distribution suited to the product and accordingly choose the channel. This needs sampling. The decision should be based on the product type, target market, and sales goals, among other factors.

1.2 IMPORTANCE OF SAMPLING IN DISTRIBUTION CHANNELS

One of the most widely accepted research techniques is sampling. A small representative segment of a target population is sampled. In order to make the best decision regarding the distribution channel, it is necessary to conduct a thorough market analysis and understand various research techniques. Sampling is more time efficient. Correct sampling gives the basis for strong decision-making. It also reduces travel cost, emphasizes on areas with greater noncompliance and allows us to compare states over different time periods.

1.3 LEVELS ON DISTRIBUTION CHANNELS

Distribution Channels are categorized into 3 levels based on the number of intermediaries between the producer and the consumer.

Level 0: Manufacturer → Consumer

Level 1: Manufacturer → Retailer → Consumer

Level 2: Manufacturer → Wholesaler → Retailer → Consumer

Level 3: Manufacturer → Wholesaler → Agent → Retailer → Consumer

1.4 TYPES OF SAMPLING TECHNIQUES

Sampling Techniques can be divided into Probability and Non-Probability Sampling. Each of the Distribution Channels involves various Sampling Techniques. The most commonly used techniques used by distributors are:

1. Simple random sampling

2. Systematic sampling

3. Stratified sampling

4. Clustered sampling

5. Convenience sampling

6. Quota sampling

7. Judgement (or Purposive) Sampling

8. Snowball sampling

1.4.1 SIMPLE RANDOM SAMPLING

Simple Random Sampling is a type of probability sampling technique which is considered to be the easiest among all the other probability sampling techniques. As the name suggests, Simple random sampling ensures that everyone in the target population has an equal and consistent chance of getting selected as a respondent of the sample. After assigning equal probabilities to each and every data point in the population, Sampling is done by selecting these data points at random so as to decrease selection bias and sampling error. More the random the process is, the lesser the sampling error associated with it.

There are 2 types of Simple Random Sampling: -

- Simple Random Sampling With Replacement (SRSWR)
- Simple Random Sampling Without Replacement (SRSWOR)

Examples:

- Used in Lottery
- Game of Bingo

Advantages:

- Lack of Bias
- Simplicity
- Free from errors in classification

Disadvantages:

- Time consuming

- Cost
- Does not work with diverse or dispersed population groups

For ex.- If the Indian government wishes to evaluate the number of immigrants living in India, they can divide it into clusters based on states such as Maharashtra, UP , Karnataka, West Bengal, Andhra Pradesh etc.

In case of Simple Random Sampling

- As the name suggests, simple random sampling gives everyone in the target population an equal and known probability of being selected as a respondent in the sample group.
- As the no. of immigrants present in every state would be a really small number as immigrants come under the minority group, This would automatically increase the sample size for the study because of the high magnitude of dispersion in the group.
- This will result in high consumption of time and money which doesn't seem to be worth risking as there are better methods which take less time, are more cost effective and provide better results.

Hence in this case, Cluster Sampling will be a much more suited sampling technique that needs to be used for this type of survey as it will be more effective and the results will be organized into states which will provide insightful immigration data.

1.4.2 SYSTEMATIC SAMPLING

It is a probability sampling in which the first unit is chosen randomly and the consecutive units are chosen in a fixed periodic interval which is also called a sample interval. The sample interval value is calculated by dividing the size of the target population by the size of the sample.

Types of systematic sampling: [2]

- Systematic random sampling
- Linear systematic sampling
- Circular systematic sampling

Example: Mall Intercept Interviewing:

In order to reduce such time sampling bias and selection bias, time based systematic sampling is used. The design of the sampling is such that the probability of sampling during the given time period is directly proportional to the number of customers expected at that time period. Samples were taken from different entrances of the mall. The number of customers entering the mall from different entrances will differ based on its location. To take an unbiased sample from each entrance, first each entrance of the mall is stratified based on its location and proportional samples are taken from each entrance using time based systematic sampling.

1.4.3 STRATIFIED SAMPLING

Stratification means, "division into layers". The method of sampling by dividing the population into mutually disjoint sub-groups (known as strata). Auxiliary information related to the character under study may be used to divide the population into various strata such that the units are as homogenous as possible within the strata and heterogeneous among strata. Stratified sampling efficiently reduces the population heterogeneity thereby increasing the precision of the dependent variable.

Advantages of Stratification: The data is more representative, more accurate, high administrative convenience and fewer sampling problems.

Disadvantages of Stratification: It is rather difficult to form homogeneous strata and decide the number of strata the population can be divided into.

Example: Stratified Sampling in Pharmaceutical Industries

In the process of making a particular drug, the pharmaceutical industry follows a particular procedure. During the Process Performance Qualification (PPQ) or product validation stage stratified sampling is used in the following manner.

Stratified Sampling of dosage units at predefined intervals and collecting samples that represent a specific location. The locations form the strata.

The samples are collected in the compressing/filling operation basically to check the homogeneity of the power mix at different intervals based on phases, time, storage temperature etc. A minimum of 7 samples are required to be collected from each of these locations.

Now based on these samples, the pharma industry compares which of these locations yield high homogeneity in the product (since a drug/tablet needs to have homogenous powder mix throughout the tablet). The differences in the tablets or test results are compared and analysed and then the best result from say location x is selected for the production and distribution of that drug.

1.4.4 CLUSTERED SAMPLING

The cluster sampling consists of forming suitable clusters of contiguous population units, and surveying all the units in a sample of clusters selected according to an appropriate sampling scheme. (Definition 10.1- Elements of Survey Sampling by Ravindra Singh and Naurang Singh Mangat Punjab Agricultural University, Ludhiana, Punjab, India)

Cluster samples are internally heterogeneous and externally homogenous.

Example: D-Mart, a retail chain of hypermarkets in India.

D-Mart's first branch was opened at Powai's Hiranandani Gardens, Mumbai in 2002.

They have used cluster sampling to excel in the retail business, middle class people are their target population.

The now have 221+ branches all over India and they've done so using the strategy of cluster-based expansion combined with EDLP/EDLC (Everyday low prices/cost).

1.4.5 CONVENIENCE SAMPLING

Convenience sampling is a type of non-probability sampling technique. In this technique, the participants of an experiment are selected at the researcher's convenience in the sense that we pick the population that is easily approachable (with the least effort possible).

Examples:
A common example of convenience sampling is the distribution of pamphlets and other promotional vouchers at a mall or crowded street.
A more relevant example for us would be circulating questionnaires for our projects. Often due to time constraints and lack of resources, we circulate questionnaires among friends and family to collect data. This is an example of convenience sampling.

Advantages :
- Convenience sampling is said to be useful for pilot testing and hypothesis generation.
- It is inexpensive as compared to other methods
- Participants are easily available
- Faster data collection

Disadvantages :

- Variability and bias cannot be controlled
- The sample may not be fully representative of the population
- Information may not be entirely reliable
- Not generalizable

Conclusion:
While convenience sampling is not highly reliable, it surely gives us a starting point for our research.

1.4.6 QUOTA SAMPLING

Quota sampling is a non-probability sampling method. In Quota sampling, researchers create a sample involving individuals that represent a population i.e., individuals with specific qualities or traits which are of interest to the researchers' study.

Example: Louis Vuitton- target population: people with annual income above $75,000.

1.4.7 JUDGEMENT (OR PURPOSIVE) SAMPLING

Judgmental sampling is also known as purposive or authoritative sampling. It is a non-probability sampling technique in which the sample members are chosen only on the basis of the researcher's knowledge and judgment. This sampling technique is most effective in situations where the number of people in a population who possess qualities that researchers expect from the target population are very few. The researcher's knowledge in this sampling technique plays the main role in creating a sample, it's very highly likely that the results obtained will be highly accurate and the margin of error will be minimal. Researchers do so only when they're confident about their knowledge to select a sample and are certain that other sampling techniques will consume more time.

Example: Consider a scenario where a panel decides to understand what are the factors which lead a person to select ethical hacking as a position.

Hiring employees for a specific position or job role.

1.4.8 SNOWBALL SAMPLING

Snowball or chain-referral sampling technique is often used in hidden populations, such as drug users or sex workers, which are difficult for researchers to access. It is a non-probability sampling technique where the samples are subject to numerous biases. Existing study subjects recruit future subjects from among their acquaintances in snowball sampling. Hence, the sample group is said to grow like a rolling snowball. Enough data are gathered to be useful for research as the sample builds up.

Example: People with a happening friend circle, i.e., with many friends are more likely to be included in the sample. Virtual snowball sampling occurs when virtual social networks are used.

Example: Jeffrey Epstein used snowball sampling by targeting young girls from low-income neighbourhoods and lured them into his international sex trafficking ring by offering them a so called part-time job of a masseuse, paying them big sums of money and asking them to refer or bring along 2-3 more girls the next time they came. He even paid them incentives for their referrals. It snowballed into an international ring sex trafficking of dozens of girls who were trafficked to many rich and elite men over the globe. Even investigations into this matter use snowball sampling.

SEGMENT 2

2.1 OVERVIEW OF THE TOBACCO INDUSTRY

India is the second largest producer, third largest exporter of Tobacco and the largest consumer in the world

It has become an integral part of the Indian-socio culture, especially in the northern, eastern and north-eastern part of the country.

The Tobacco Industry provides employment to more than 46 million people in India and almost 267 million Indians consume Tobacco in one form or another.

Indian Tobacco Company Limited (ITC Ltd.) currently holds the highest market share (~84.27%). The remaining share is distributed among minor companies like Kothari Products Ltd., Godfrey Phillips India Ltd., VST Industries Ltd., etc.

Over 80% of the world's 1.3 billion tobacco users live in low- and middle-income countries

2.2 THE VARIOUS SEGMENTS OF TOBACCO INDUSTRY

The tobacco industry is broadly categorized into two segments, Smoke and Smokeless Tobacco with a market share of 52% and 48% respectively

1) Smoking Tobacco: (Cigarettes, Bidi, Cigars, Pipes, Hookah etc.)

2) Non-smoking Tobacco: (khaini, pan masala, gutka, mawa, snuff, snus, dipping tobacco, etc.)

2.3 DISTRIBUTION CHANNELS OF THE TOBACCO INDUSTRY

2.3.1 INDIA

- **Farming** - Most factories have their own farming land or they negotiate and agree to a contract with farmers. The primary produce which is acquired from the farmers is then processed into secondary goods and sold to wholesalers.

- **Producers** - Indian tobacco company Limited (ITC Ltd.) has factories in Ratangaon, Bangalore, Hyderabad, Saharanpur and Kolkata.

- **Wholesalers** - Wholesalers stock the secondary products in their godowns after which they are supplied to Distributors, Dealers, Retailers and finally the consumers in the respectively mentioned order. i.e., Wholesalers, brokers, Retailers act as intermediators. A wholesaler keeps a regular stock of Rs.2,00,000 per week.

- **Retailers** - Tapris, grocery stores, bars, restaurants, retail stores, specialty stores, online stores, other.

- **Distributors** - 'Bombay traders' and 'Scooter wala' are the distributors of ITC products.

2.3.2 WORLDWIDE - Marlboro

- Marlboro sells their goods in more than 180 countries through retail outlets, E-commerce sites, distributors, etc.

- They have 800+ Distributors, 8 lakh Retailers and 6000 salesmen in India.

- Goods are sold mostly at small retail shops placed in a crowded locality and are promoted through advertisements by sponsoring night clubs and bars, etc.

2.4 SECONDARY DATA ANALYSIS

Get Secondary Data - compare 2 charts for channels
The table below represents the percentage of tobacco distribution in Urban and Rural areas for the two time periods 2009-10 and 2016-17

URBAN
2009-10

	Store	Street Vendor	Kiosk	Others
CIGARETTE	44.6	21.9	31.5	2.0
BIDI	44.2	15.2	40.1	0.5
SMOKELESS TOBACCO	46.9	16.9	34.5	1.7

2016-17

	Store	Street Vendor	Kiosk	Others
CIGARETTE	49.5	8.6	39.9	2.0
BIDI	53.6	7.9	38.1	0.4
SMOKELESS TOBACCO	49.9	7.1	41.5	1.5

Inference: From the above table we can infer that there are changes in the percentage of distribution of various tobacco products over the periods. The significant change is that there is an increase in the percentage of distribution of each and every tobacco product through stores. so, stores continue to be the most prominent place for the distribution of various tobacco products

RURAL

2009-10

	Store	Street Vendor	Kiosk	Others
CIGARETTE	54.5	12.7	29.9	2.9
BIDI	49.4	10.6	39.2	0.7
SMOKELESS TOBACCO	56.7	8.8	31.3	3.3

2016-17

|	Store	Street Vendor	Kiosk	Others
CIGARETTE	51.9	9.7	37.7	0.8
BIDI	62.3	6.8	30.2	0.8
SMOKELESS TOBACCO	57.5	6.8	33.9	1.8

Inference: The above table shows that over the years, the increase in the distribution of bidi and smokeless tobacco through stores shows that there wasn't much change in the consumer pattern, so the industry hasn't changed its distribution strategy. Whereas in the case of cigarettes there is a decline in percentage of distribution in stores which shows a change in consumer pattern. So, the industry has changed its distribution strategy accordingly

General Inference and Understanding of the Secondary Data
By observing the secondary data, we can infer that over the years both males and females are shifting towards smokeless tobacco which indicates that the market share of smokeless tobacco is likely to increase in the future.

2.5 SAMPLING TECHNIQUES USED IN THE TOBACCO INDUSTRY

Flowchart of 3 stages of the distribution channel and sampling technique within each level
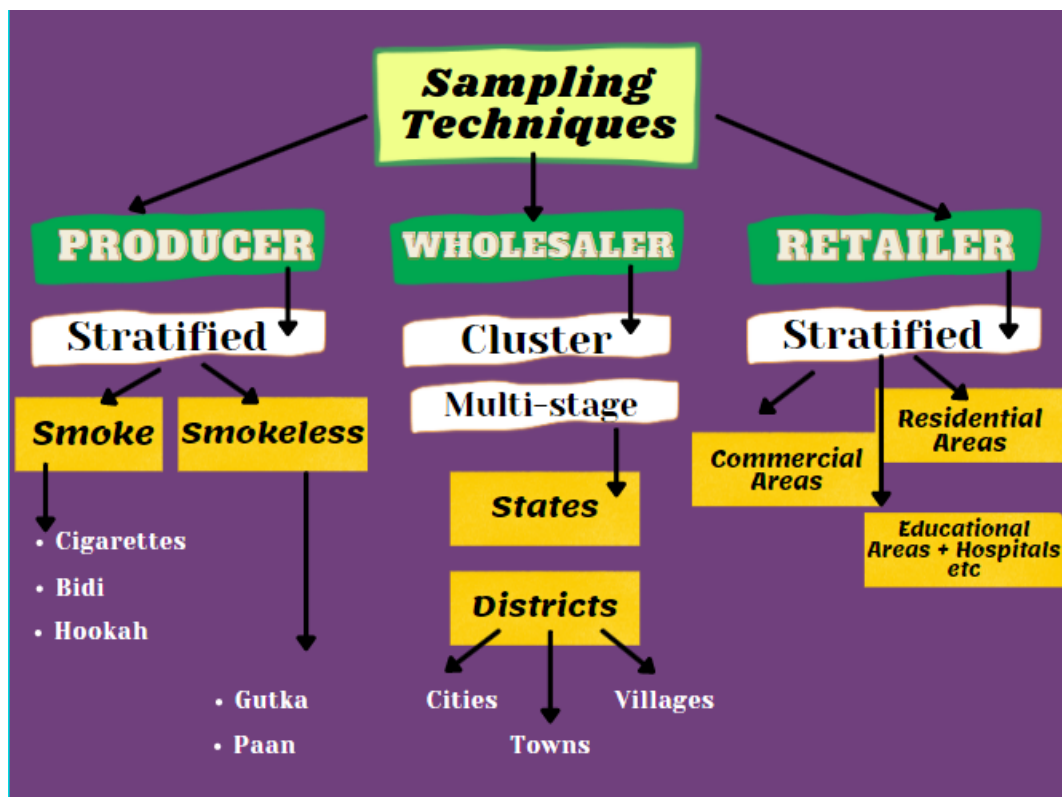
Producer Level - Stratified sampling is used. 2 Stratas are formed - Smoking and Smokeless tobacco products.
Smoking Tobacco is divided into cigarettes, bidi, hookah etc. And Smokeless Tobacco products are Paan, Gutka etc

Wholesaler Level - Cluster / Multistage Sampling - Dividing the population area wise based on States, Districts and further into Urban and Rural areas.

Retailer Level - Stratified Sampling - Dividing the population of each state based on Residential Areas, Educational Institutes + Hospitals + Parks and Commercial Areas. Each of these will be considered as an individual stratum.

Consumer Level - The final consumers purchase their products from the various distribution outlets such as Local shops, Vendors, Supermarkets etc. Tobacco Industry has increased its sales among the adolescents, middle class and especially people living below the poverty line.

2.6 FINAL STATEMENT ABOUT THE INDUSTRY

By observing the secondary data, we can infer the following:

Increase Distribution of Smokeless Tobacco: Over the years both male and female tobacco users are shifting towards smokeless tobacco which indicates that the market share of smokeless tobacco is likely to increase in the future. And thereby the distribution of the same must increase.

Preferred Outlet: People mostly buy tobacco from stores; their second most preferable source of distribution are kiosks. Whereas, Street vendors and the other sources are least preferred.

Expenditure:
1. **Urban**: The average expenditure on smokeless tobacco has increased. As per the GATS 2016-17 survey we have found that people from urban areas spend more on Smokeless tobacco (₹43.20) than on Cigarettes (₹37.89). Although, the average expenditure on cigarettes has increased over the years).

2.   **Rural**: People spend more on cigarette (₹9.60) than smokeless tobacco (₹6.00) and bidis (₹5.50) but in 2016-17 survey we have observed that people now spend more on the Smokeless tobacco (₹42.41) than cigarette (₹23.29) and bidis (₹12.83), although the expenditure on bidi is increased over these years.

SEGMENT 3

3.1 PRIMARY DATA COLLECTION:

Primary Data was collected through Google Forms shared among various age groups, primarily the younger generation. A total of <> responses were collected in a span of 3 days. To be noted: No personal information such as Name, Email id, phone number etc were collected.

The following questions were asked:

Basic Questions:

1. Age
2. Gender
3. Do you consume tobacco?

For Tobacco Consumption Details:

4. Which tobacco product do you consume the most?
5. How often do you consume?
6. Weekly Spending Amount
7. From where do you purchase the product?
8. Reason for consuming the product
9. Do you plan on quitting tobacco products?
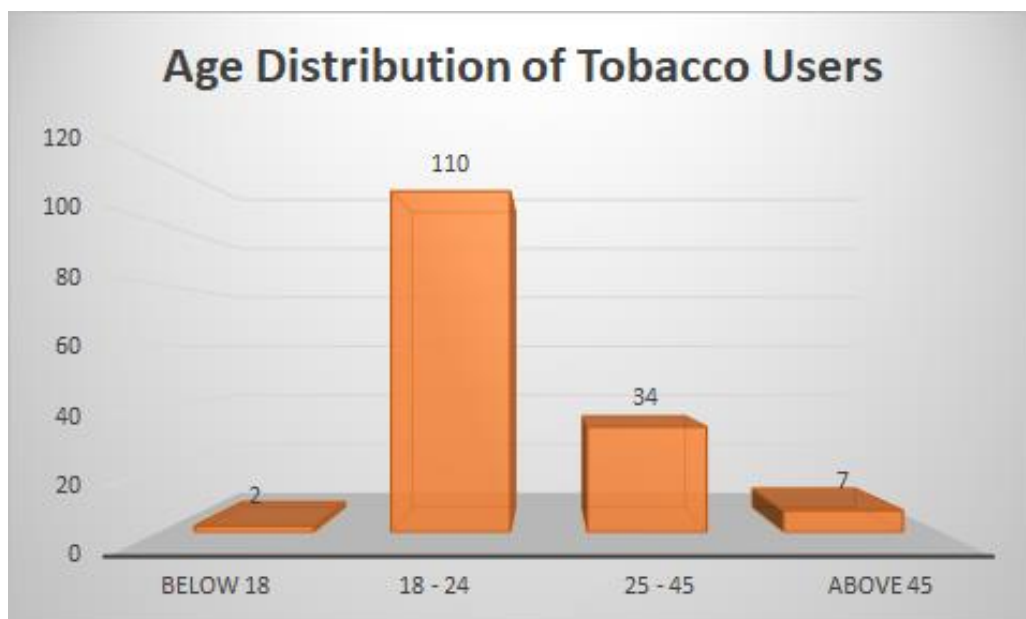10. Has consumption of tobacco products increased over time?

Behaviour towards smoking Before and During Pandemic:

1. Did Covid-19 affect your cigarette consumption?
2. Number of cigarettes consumed in a week before Covid?
3. Number of cigarettes consumed in a week ongoing Covid scenario?

3.2 ANALYSIS OF PRIMARY DATA:

3.2.1 Age, Gender and Consumption of Tobacco

Age: As seen from the Age graph, 1% are below 18, 73% belong to the 18-24 age group, 24% belong to the age group 25-45 and 4% are above the age 45. As seen the majority of our sample belongs to the 18-24 age group.

Age Distribution of Tobacco Users

Gender: 69% are Male, 30% are Female and 1% belong to the others category. Out of which 77%, 22% and 1% are Male, Female and Other Tobacco users respectively. Both the pie charts show that the maximum consumption of tobacco products is from the male population


Gender Distribution Entire Sample
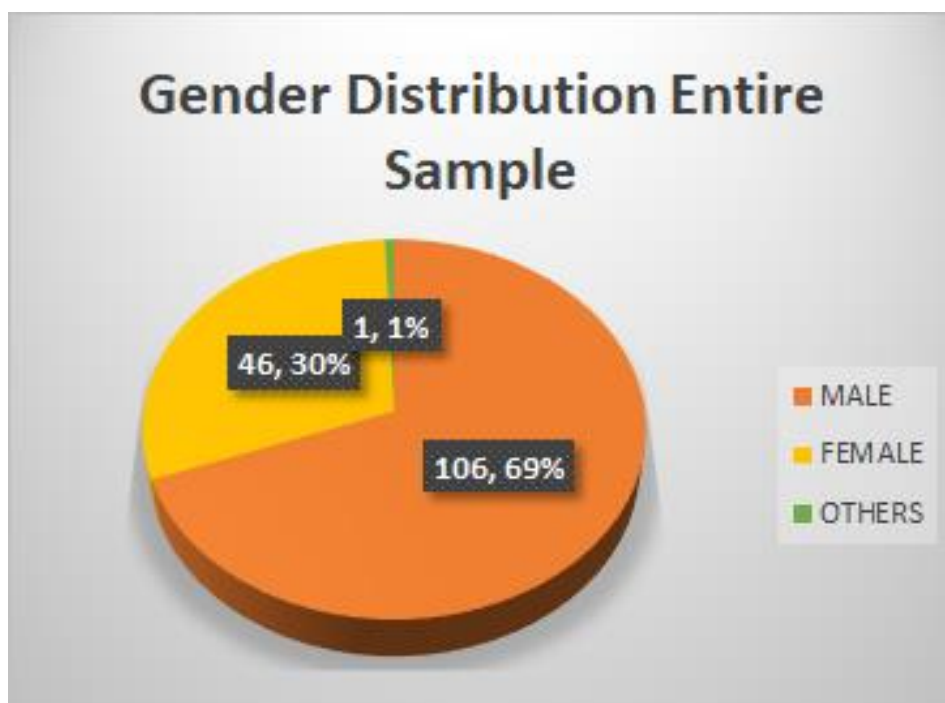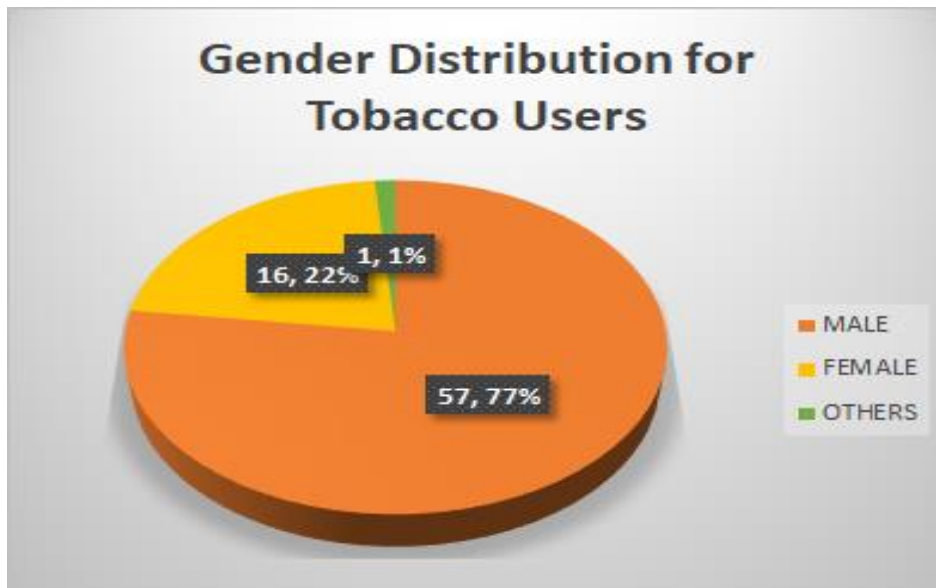
Consumption: Almost half of our sample population do not consume any tobacco products. 79 people don't consume any tobacco product and the remaining 74 do. Out of the 74, 57 are male consumers, 16 are female and 1 other consumer.
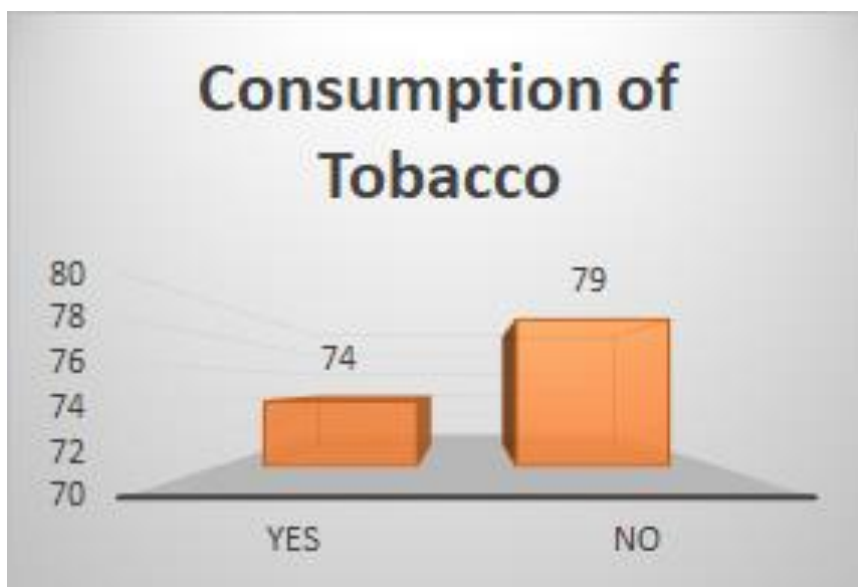
3.2.2 Segment-Wise Consumption of Tobacco Products:

Out of all the respondents who selected "Yes" for consuming tobacco products, all have selected Cigarettes as their main source of tobacco consumption. Followed by Hookah and finally Other sources. Thus, Cigarettes and Hookah are the most preferred tobacco products. Which can further imply that they are the most readily available and hence best distributed among other products.

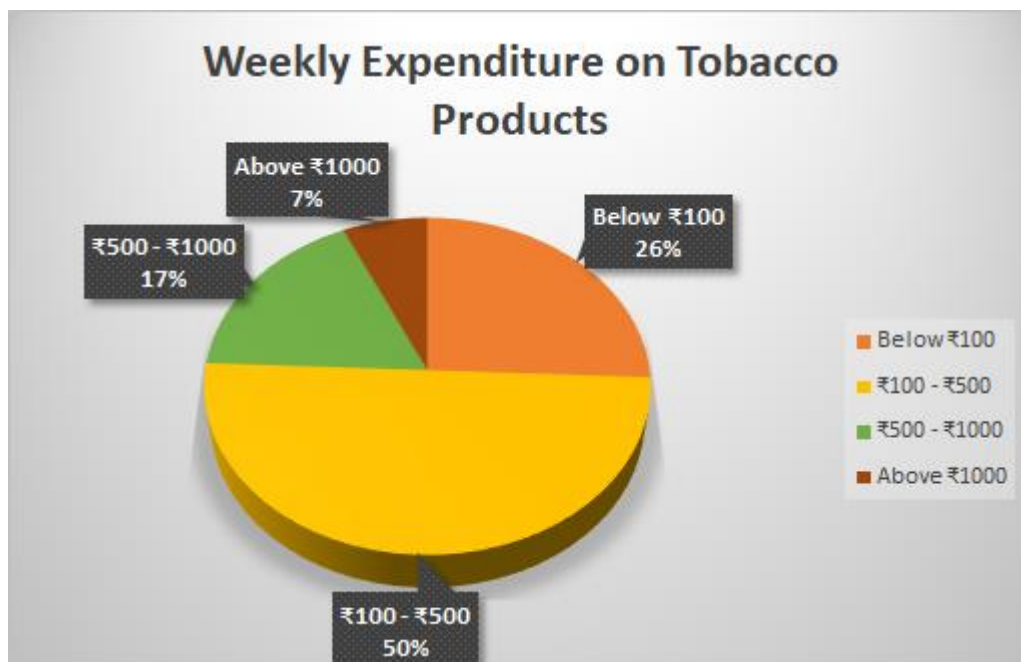**Consumption of Different Tobacco Products**
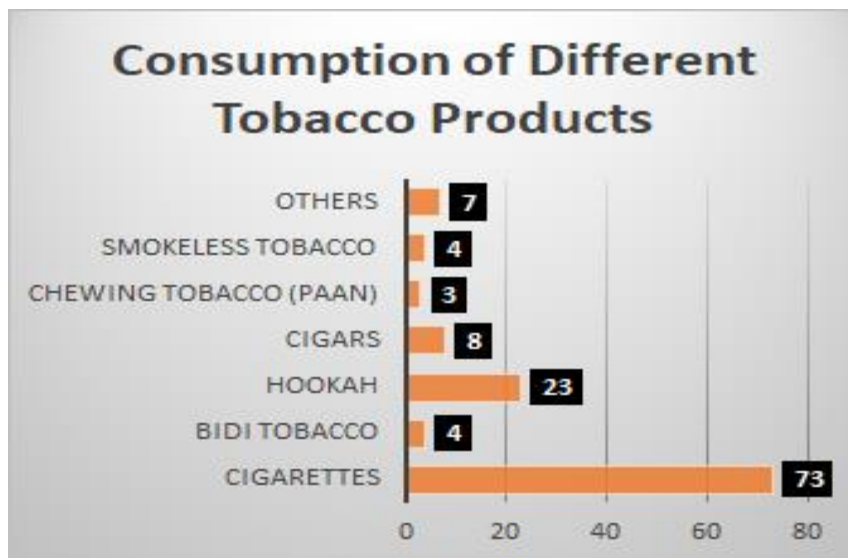
3.2.3 How often do you consume?

33% of people have agreed to have consumed tobacco products on a daily basis. 4% people have occasionally consumed and lastly 5% have responded to rarely consume any tobacco products.

Consumption Rate

3.2.4 Weekly Spending Amount?

Weekly Expenditure on Tobacco Products
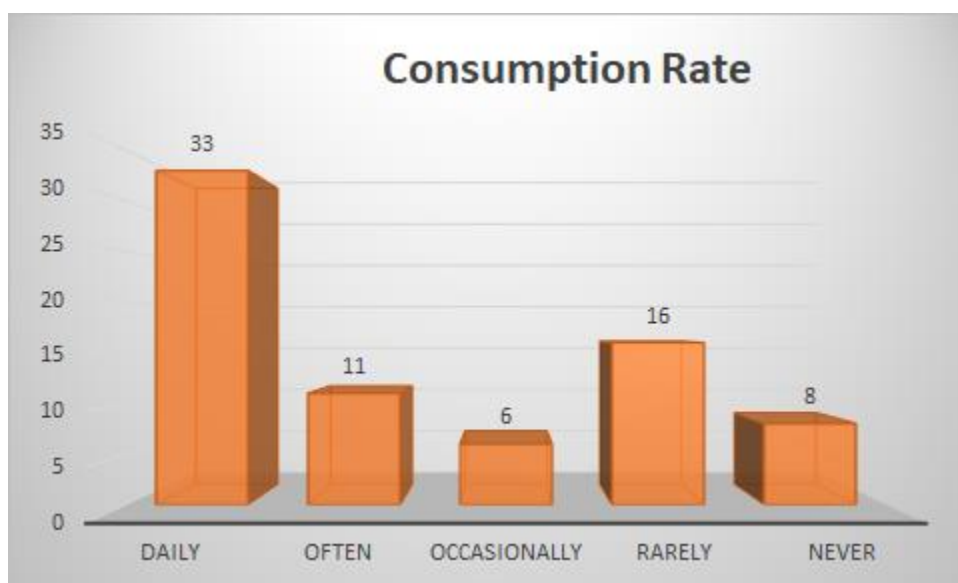
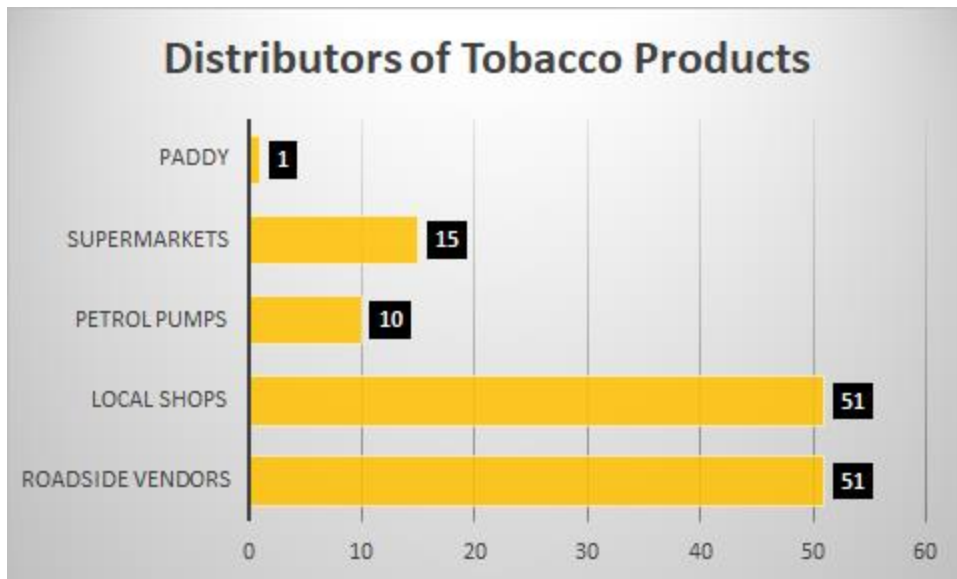Out of the 74 respondents, 37 tobacco users on an average spend about ₹300 weekly. From this we can infer that , on an yearly basis the "₹100 - ₹500" weekly spending tobacco users are ready to shell out ₹300 x 37 x 4 x 12 = ₹5,32,800.

If we extend our analysis to understand the yearly expenditure of an average tobacco user , ₹28,70,400 is contributed from 74 people yearly, who spending range is anywhere between ₹1 - ₹3000. This further shows that even with the health hazards associated with tobacco products, the contribution towards the economy is a significant one, considering the tax on most tobacco products is 28 - 49%.

3.2.5 From where do you purchase your tobacco products?

As you can see from the bar chart/pie etc. ,It is clear that roadside vendors and local shops which use a cluster based mechanism to sell their products are clearly able to sell more than any other competitor.

Distributors of Tobacco Products

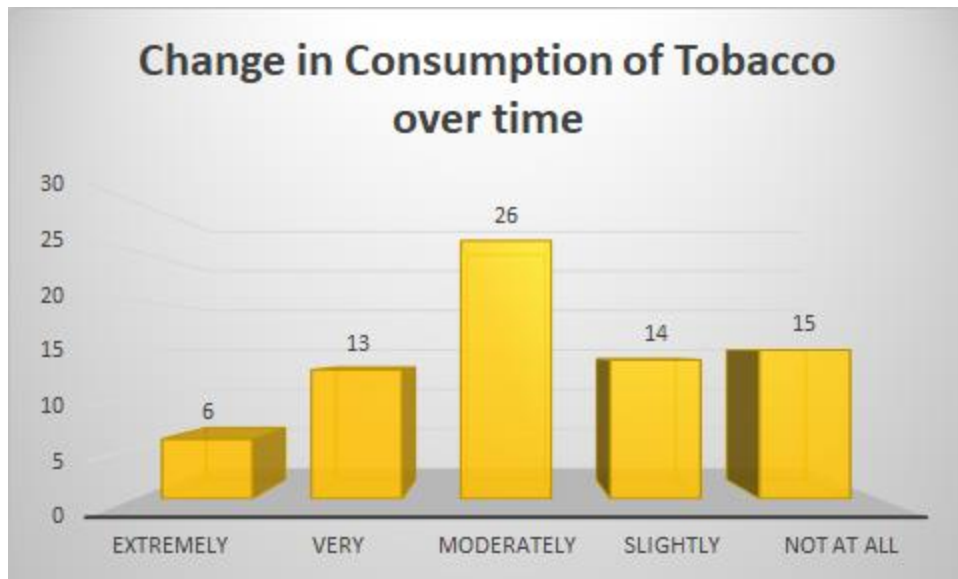| Distributor	Value
PADDY	1
SUPERMARKETS	15
PETROL PUMPS	10
LOCAL SHOPS	51
ROADSIDE VENDORS	51

3.2.6 Do you plan on quitting tobacco products?

From the 74 respondents, 19% of the respondents have claimed that they would not stop consuming tobacco products, 38% are conflicted and 43% have claimed to quit. Almost 57% of respondents will be a consumer of tobacco products even with the associated health problems. The factor "addiction" and "subtle peer pressure" plays a large role. The distribution and easy availability of tobacco products makes it very difficult for a regular consumer to quit the product and hence will always remain a sample point for the tobacco industry.

People's response towards quitting tobacco products

- Maybe 38%
- Yes 43%
- No 19%
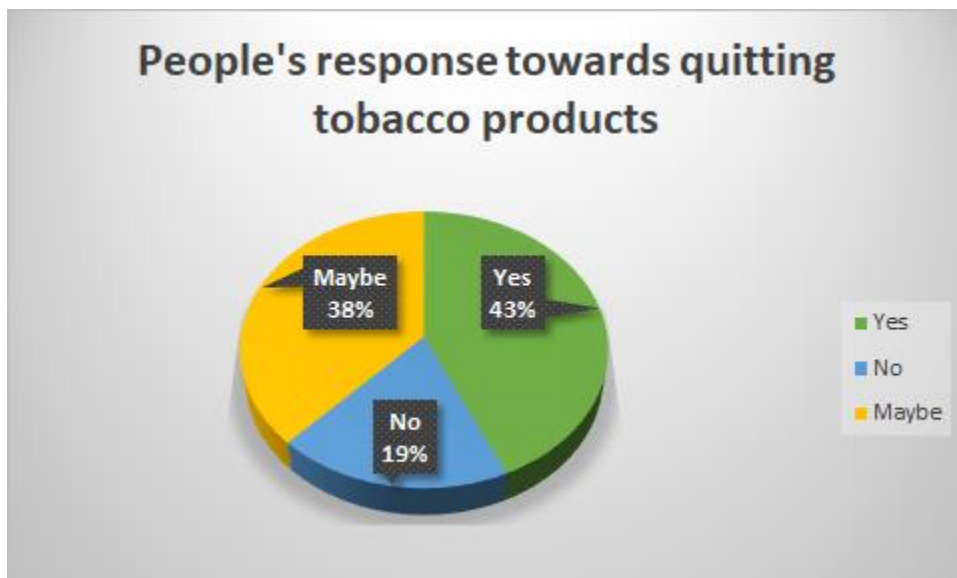
3.2.7 Has consumption of tobacco products increased over time?

3% of the respondents have admitted that their consumption has increased by a fairly high quotient over the years. 16% of the respondent's consumption has "Moderately" increased and 10% of the respondent's tobacco consumption has remained unchanged.

Change in Consumption of Tobacco over time

Category	Value
EXTREMELY	6
VERY	13
MODERATELY	26
SLIGHTLY	14
NOT AT ALL	15

3.3 HYPOTHESIS TESTING:

3.3.1 Main Objective:

Our main objective to conduct this hypothesis testing is to check if the mean weekly consumption of cigarettes has remained the same or has it been affected due to the ongoing pandemic.

  a. **Why are we doing this?**

  As we know that Covid-19 is an infectious disease caused by the SARS-CoV-2 virus. This disease can create some serious problems like chest pain, shortness of breath etc. and we also know how cigarettes can affect your lungs. Even after being fully aware of the situation, an average smoker has not quit smoking. Our study can be broken down into 2 questions,

  1. If (for example) there is no significant difference between the mean weekly consumption of cigarettes before and during Covid-19, then how did cigarette companies like Marlboro, Classic etc. adapt to a situation like this? and what all sampling techniques were used by them to maintain such demand levels? Hence, We want to understand the role of sampling techniques that were used to solve this problem.
  2. If (for example) there is a significant difference between the mean weekly consumption of cigarettes before and during Covid-19, then what were the possible reasons for the decline of weekly consumption of cigarettes?

3.3.2 Formulation of Hypothesis:

Our hypotheses will be based on the significance of difference of means of weekly consumption of cigarette before and during Covid-19. We can state our **H0** and **H1** as follows: -

H0: There is no significant difference between the mean weekly consumption of cigarettes before and during Covid-19, i.e., both the sample belong to the same population ($\mu_1 = \mu_2$).

H1: There is a significant difference between the mean weekly consumption of cigarettes before and during Covid-19, i.e. $\mu_2 > \mu_1$ or $\mu_1 < \mu_2$. (Left tailed test)

3.3.3 Which Test Are We Using?

As sample size n=73, it would be best to use Z- test for this case. A t-test would have been more appropriate for a sample size < 30.

1. **Test of Significance for Difference of Means**

In this case, Let $\bar{x}_1$ be the mean of weekly consumption of cigarettes before Covid-19 with sample size n1 from a population with mean μ_1 and variance σ_1^2 and let $\bar{x}_2$ be the mean of weekly consumption of cigarettes during Covid-19 with sample size n2 from another population with mean μ_2 and variance σ_2^2. Then, since the sample size are large,

$$\bar{x}_1 \sim N(\mu_1, \sigma_1^2/n_1) \quad \text{and} \quad \bar{x}_2 \sim N(\mu_2, \sigma_2^2/n_2)$$

Thus, Under **H0**: $\mu_1 = \mu_2$, The test statistic for large samples become,

$$Z = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)\right)}} \sim N(0,1)$$

❖ **For a left- tailed test:**

If the Zscore > $Z\alpha$, the null Hypothesis is accepted and we conclude that there is no significant difference between the sample means and they belong to the same population ($\mu_1 = \mu_2$).

If the Zscore < $Z\alpha$, the null Hypothesis is rejected and we conclude that there is significant difference between the sample means and they belong to different population ($\mu_1 < \mu_2$).

Information on both Samples (Consumption Pre and During Covid):

Based on the Primary data we collected, the information collected about both the samples have been given below,

Consumption of cigarettes during Covid-19		Consumption of cigarettes before Covid-19	
$\bar{X}_2$	28.05	$\bar{X}_1$	27.15
n_2	73	n_1	73
σ_2^2	1891.52	σ_1^2	1040.01

For the value of $Z\alpha$, we have chosen a level of significance (α) of 5% which has a critical value ($Z\alpha$) of -1.645 (for a left tailed test)

3.3.4 Conclusion and Output Sheet:

From the output sheet it is clear that our Zscore = -0.1427 which in comparison to the $Z\alpha$ give us the following conclusion:

Since -0.1427 (Zscore) > -1.645 ($Z\alpha$), it is non-significant at 5% level of significance. Hence the null hypothesis is accepted at 5% level of significance and we conclude that the mean weekly consumption of cigarette before Covid-19 is equal to the mean weekly consumption of cigarette during Covid-19.

INFO. ON THE FIRST SAMPLE			INFO. ON THE SECOND SAMPLE	
Mean	27.15		Mean	28.05
Variance	1040.0186		Variance	1891.52
Standard Deviation	32.24932		Standard Deviation	43.4917
Sample Size	73		Sample Size	73

Standard Error	6.337044

$Z\alpha$	-1.645
Z score	-0.14267

Hence, we accept the null hypothesis H0.

3.3.5 Inferences

1. According to WHO, there are approximately 182,000,000 smokers in India which is home to 16.6% of world's smoking population. From our primary data with a sample size of n = 73, the mean weekly consumption of cigarettes before Covid-19 came out to be 27.15 and that of during Covid-19 came out to be 28.05. If we divide the entire population of smokers with n = 73, we get a value of 2,493,151, i.e. the entire population has been divided into 2,493,151 stratas with each sub-group having 73 data points. The difference between the mean weekly consumption of cigarettes before and during Covid-19 came out to be 0.9. So, if we multiply 0.9 with 2,493,151, we get a value which tells us the increase in mean weekly consumption of cigarettes before and during covid which is equal to 2,243,836.

2. As Covid-19 has not affected the consumption of cigarettes, it is safe to say that the marketing campaigns and distribution techniques even during pandemic have worked. Thus, Sampling techniques used to target the population have been efficient. We can further explain this by using a timeline of events given below:

   1. During the initial stages of Covid-19 when the entire country was under a strict lockdown, it was next to impossible to buy cigarettes as all the local shops, roadside vendors were closed. This must have been a huge problem to major cigarette companies as they had to change their entire distribution channels just to meet demand. Quota sampling must have been used in this situation because only supermarkets and petrol pumps were open during such times of distress and as every supermarket and petrol pump in India were seen as a representative of all products related to cigarettes. Thus, every major player in the cigarette industry distributed their product within these quotas (supermarkets and petrol pumps) so as to meet demand.

   2. As things started to look better for India with time. Local shops and roadside vendors re-opened and this was an opportunity for all the players in this industry to re-evaluate their distribution channels and maximize their profits. In order to make the most out of this situation, Clusters of roadside vendors and local shops were formed in Commercial areas, Residential areas, Educational institutions etc. This way a massive network of shops were opened where the seller could distribute cigarettes in a greater number. Hence, Cluster sampling was the key for every big cigarette company to maximize their profits and make the most out of this situation.

3.4 MULTIVARIATE LINEAR REGRESSION MODEL:

We performed multivariate linear regression to understand the impact of age, and other factors influencing consumption of cigarettes during covid.

$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon$

Y_i: No. of cigarettes consumed per week (consumption)
β_0: slope intercept (inherent consumption of cigarettes)
β_1: coefficient (Age)
β_2: coefficient (Recreation)
β_3: coefficient (Peer pressure)
β_4: coefficient (Stress/Anxiety)
ε: error term

	Intercept	**Age**	**Recreation**	**Peer pressure**	**Stress/Anxiety**
β_i	-7.1688	0.5840	0.9458	2.0166	3.9209
p	0.5572	0.1043	0.6744	0.5235	0.0823

Of the 4 factors that we modelled; the coefficient of Stress/Anxiety is significant as one of the major reasons for consuming tobacco products @ 95% level of significance.

Multiple R-squared: 0.1064

Note: (for Multiple R-squared closer the value of Multiple R squared to 1 better the predictions of the model)
Clearly our model does not fit the data very well. This could be attributed to the fact that the relationship between the target variable is not linear. A larger dataset could also yield better results with greater accuracy.

Based on the correlation matrix above, we can say that consumption of cigarettes is positively correlated with Stress/anxiety, age and peer pressure in the order mentioned. A recent report on impact of the covid-19 pandemic on tobacco consumers also states that 4% of tobacco users observed an increase in their consumption due to stress experienced in the current context.

3.5 LIMITATIONS OF OUR STUDY

- Due to restrictions on time and resources, convenience sampling was adopted to select participants for this study. Our sample was not representative of the population of tobacco consumers

- Confidence of inferences drawn could be better with a bigger sample size

3.6 SOFTWARES USED:

1. Microsoft Excel
2. Python
3. R Studio
4. Canva

CONCLUSION

In India, the tobacco industry has a sizable influence from The Local and The Union governments. Since 1992, the Indian Tobacco Company (ITC), has made substantial contributions to the Rajiv Gandhi Foundation, the Prime Minister's, and the Chief Minister's Relief Fund of up to 27 lakhs. These are some of the ways that the industry ensures respectability and a good corporate image for the promotion and effective distribution of tobacco products.

The Tobacco Industry advertising made consumption of tobacco seem normal by glamourising tobacco consumption by creating an image of rebellion, confidence, independence and freedom. But since the ban of advertising of tobacco products, a huge share of their marketing and promotional budget has been transferred into retailing point-of-sale outlets. They began to prey on low-income, and minority smokers, by focusing a lot of tobacco advertising in those low-income areas and ensuring that these neighbourhoods had a high concentration of tobacco outlets in them. They even priced cigarettes lower in these areas to serve as an incentive for greater public spending.

Young adults are another key target market for tobacco companies especially for Marlboro and other cigarette companies. From our analysis of the secondary data, we were able to gather that at each segment of the distribution channel the industry changes its sampling technique in order to maximise their outreach. This industry practices stratified sampling in terms of addressing their customer base (the last leg of the distribution channel) by segmenting it into numerous homogeneous fractions keeping in mind their consumer behaviour and nature. Once they have identified their segments, the point of purchase for customers is spread out in clusters such that outreach to the audience is maximum. It is very important to constantly update sampling techniques in the distribution channel with reference to the consumer patterns.

Tobacco consumption has made its way into people's lifestyles and thus the market share only seems to be increasing over time. For a strong distribution it is important to understand the product and the population, thus, the sampling techniques used currently (Stratified, Cluster and Multistage Sampling) is proven effective. The fact that due to the pandemic production and distribution of tobacco products was halted, yet consumption of the products remained unaffected to an extent only speaks about the effectiveness of the distribution channels that were setup.

REFERENCES

1)https://www.researchgate.net/publication/285946644_Reducing_bias_in_shopping_mall-intercept_surveys_The_time-based_systematic_sampling_method

2) https://pqri.org/wp-content/uploads/2015/12/FDADraftGuide.pdf

3) https://www.researchandmarkets.com/reports/4757741/tobacco-market-in-india-2018-2023

4) https://www.tobaccofreekids.org/problem/toll-global/asia/india

5)https://primefeed.in/news/3745707/pan-masala-market-size-share-industry-trends-growth-insight-share-competitive-analysis-statistics-regional-and-industry-forecast-rajnigandha-rmd-pan-vilas/
https://www.imarcgroup.com/pan-masala-manufacturers

6) https://tobaccotactics.org/wiki/be-marlboro-targeting-the-worlds-biggest-brand-at-youth/

7) http://mospi.nic.in/sites/default/files/publication_reports/427_final.pdf

8)https://ntcp.nhp.gov.in/assets/document/surveys-reports-publications/Global-Adult-Tobacco-Survey-India-2009-2010-Report.pdf

9) https://finmedium.com/2020/10/cigarette-industry-in-india/

10) https://www.livemint.com/Politics/FJov7Dm38Ap2AtStLvibnO/2324-million-adults-in-India-use-tobacco-daily-report.html