

Exploratory Data Analysis

Topic: Tobacco Usage in Indian States

Submitted by DS Batch of July 2022

Ashutosh Pradhan – A22011

Nishant Kr. Todi – A22023

Sahil Dhankhar – A22029

CONTENTS

Sl. No.	Topic	Page No.
1	Motivation Behind the Project	2
2	Introduction	3
3	Methodology	4
4	Analysis & Interpretation	5
5	Observation & Conclusion	24

MOTIVATION BEHIND THE PROJECT

The objective of this project study is to comprehensively encompass the different aspects of tobacco usage with reference to the Indian scenario. The information on prevalent tobacco habits in India, passive smoking and its impact, legislation to control and the awareness created amongst the different sections of the society.

Tobacco is a leading preventable cause of death, killing nearly six million people worldwide each year. Reversing this entirely preventable manmade epidemic should be a top priority. This global tobacco epidemic kills more people than tuberculosis, HIV/AIDS and malaria combined and can be resolved by becoming aware of the devastating effects of tobacco and learning about the proven effective tobacco control measures.

The tobacco problem in India is peculiar, with consumption of variety of smokeless and smoking forms tobacco. Understanding the tobacco consumption pattern in India, focusing more efforts on what works and investigating the impact of sociocultural diversity and cost-effectiveness of various modalities of tobacco control is the agenda of this project.

We would like to express our sincere gratitude to **Sayantani Roy Choudhary Ma'am** for providing us an opportunity to do this project and help us practically understand the various things while doing exploratory data analysis of a dataset. The guidance and support we received from her was valuable for the completion of this project report.

INTRODUCTION

DATASET DESCRIPTION

The data has been extracted open Government Data Platform India. data.gov.in is a platform for supporting open data initiative of Government of India. This portal is a single-point access to datasets, documents, services, tools and applications published by ministries, departments and organisations of the Government of India.

The data is about a survey conducted by the Global youth Tobacco Survey (GYTS). In India, GYTS-4 was conducted in 2019 by the International Institute of Population Sciences (IIPS) under the Ministry of health and Family Welfare (MoHFW). The overall response rate was 96.9%. A total of 97,302 students from 987 schools participated in the survey.

The questionnaire covered the following topics: tobacco use (smoking and smokeless), cessation, second-hand smoke (SHS), pro- and anti-tobacco media messages and advertisements, access to and availability of tobacco products, and knowledge and attitudes regarding tobacco use. The questionnaire was self-administered; using paper sheets, it was anonymous to ensure confidentiality.

AIM OF THE PROJECT

- To determine the incidence of tobacco by different state/UTs and by rural/urban split.
- To determine the usage of various forms of tobacco by different states/UTs and by rural/urban split.
- To estimate the age of initiation of cigarette and bidi smoking and smokeless tobacco.
- To estimate the exposure to second-hand smoking (SHS).
- To estimate the exposure to tobacco and anti-tobacco advertising.
- To assess the reach of major prevention programmes to school-based population.

METHODOLOGY

The Exploratory Data Analysis was conducted in the following steps:

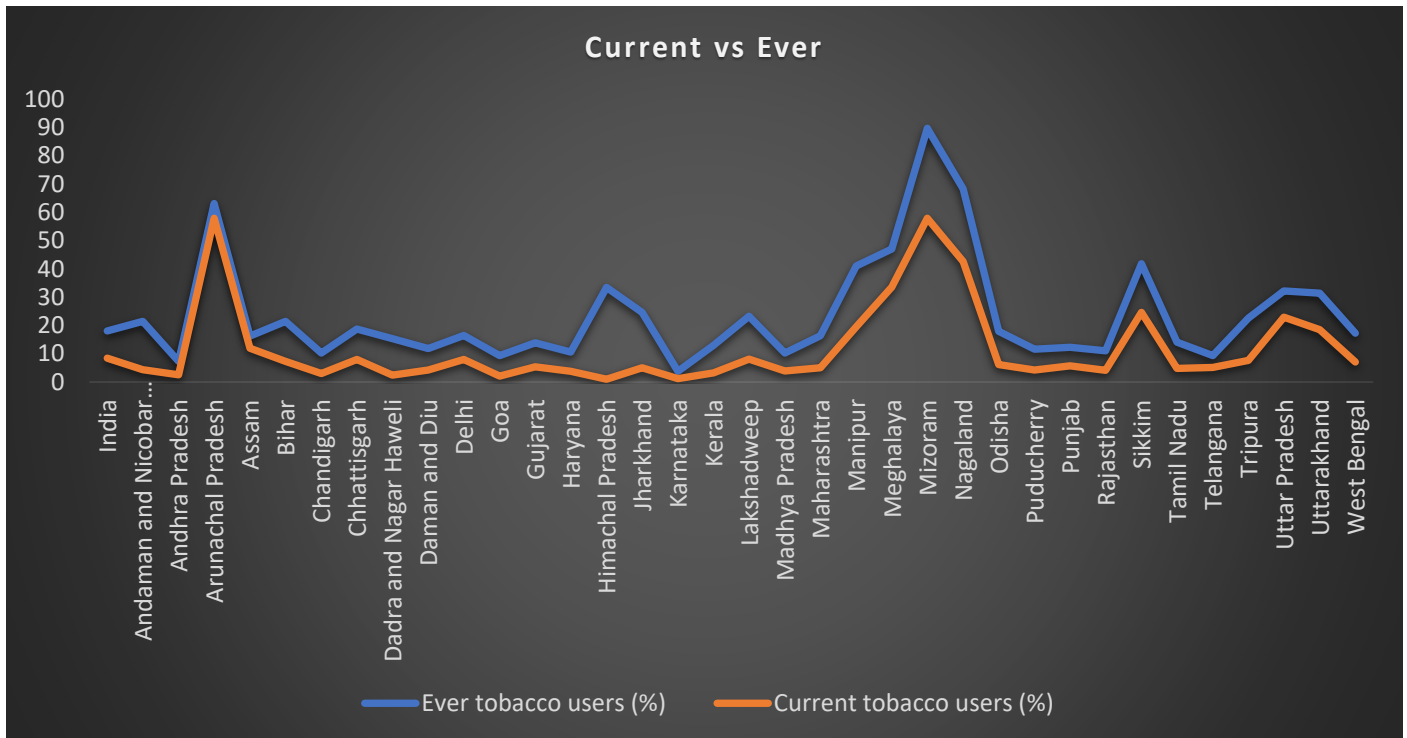
1. After the dataset was extracted from the aforesaid government website, we looked for discrepancies in the data.
2. Some cells in the 'Median Ages' columns had input as ">7". This would not be counted as numerical value and hence would create issue while analysis of these columns as numerical columns. Hence, they were replaced as value 7.
3. There were some cells with value 0. They have not been tempered with as they are possibly legitimate.
4. After pre-processing in **MS-Excel**, the data was imported into **Tableau**.
5. There upon many graphs were made to understand the data and draw some inferences from it.
6. MS-Excel was also used to find correlation between some variables.
7. We split our analysis into two sections: **State-Wise Analysis and Rural-Urban Analysis**.
8. In state-wise analysis, inferences about how different states compare among themselves on various aspects are shown. While in rural-urban analysis, a basic differentiation about how tobacco usage is prevalent in these two areas in analysed.
9. Thereafter, appropriate conclusions have been drawn.



ANALYSIS & INTERPRETATION

STATE-WISE ANALYSIS

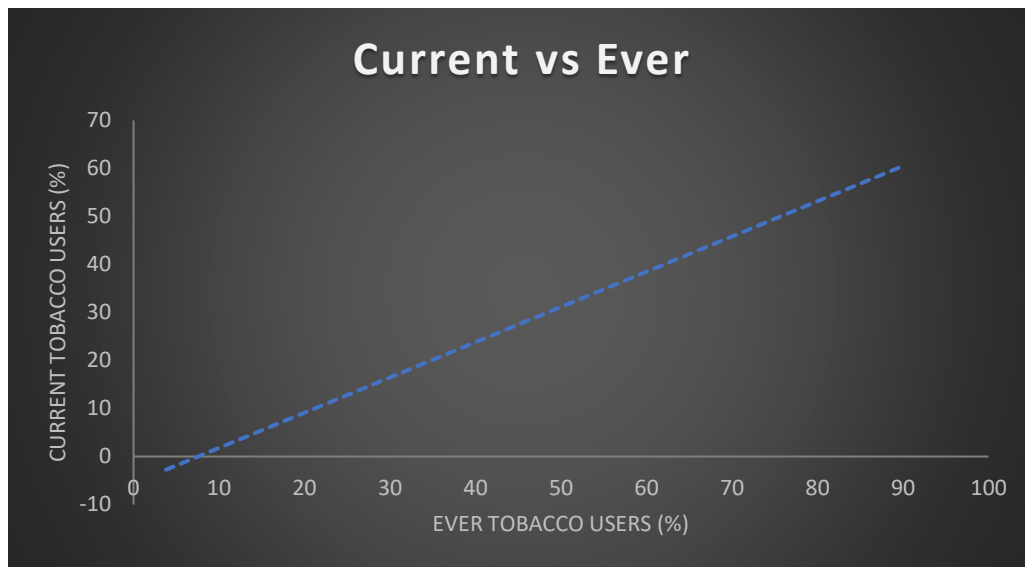
1. Current Tobacco Users VS Ever Tobacco Users



INTERPRETATION

- Tobacco users have decreased over time.
- High tobacco consumption in north-eastern states.
- Himachal Pradesh seems to have improved the most among high tobacco users.
- Karnataka has the lowest ever as well as current tobacco users.

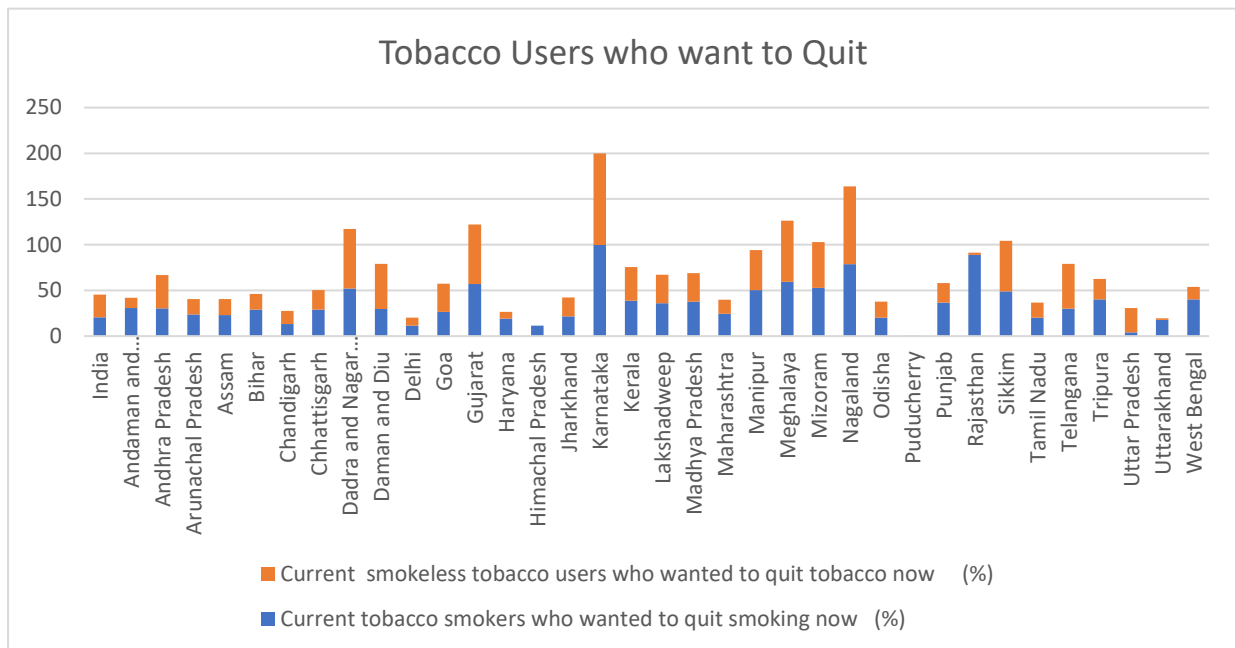
	<i>Ever tobacco users (%)</i>	<i>Current tobacco users (%)</i>
<i>Ever tobacco users (%)</i>	1	0.936081319
<i>Current tobacco users (%)</i>	0.936081319	1



INTERPRETATION

- High correlation between current and ever users, that is, states with high ever tobacco users have high current users.

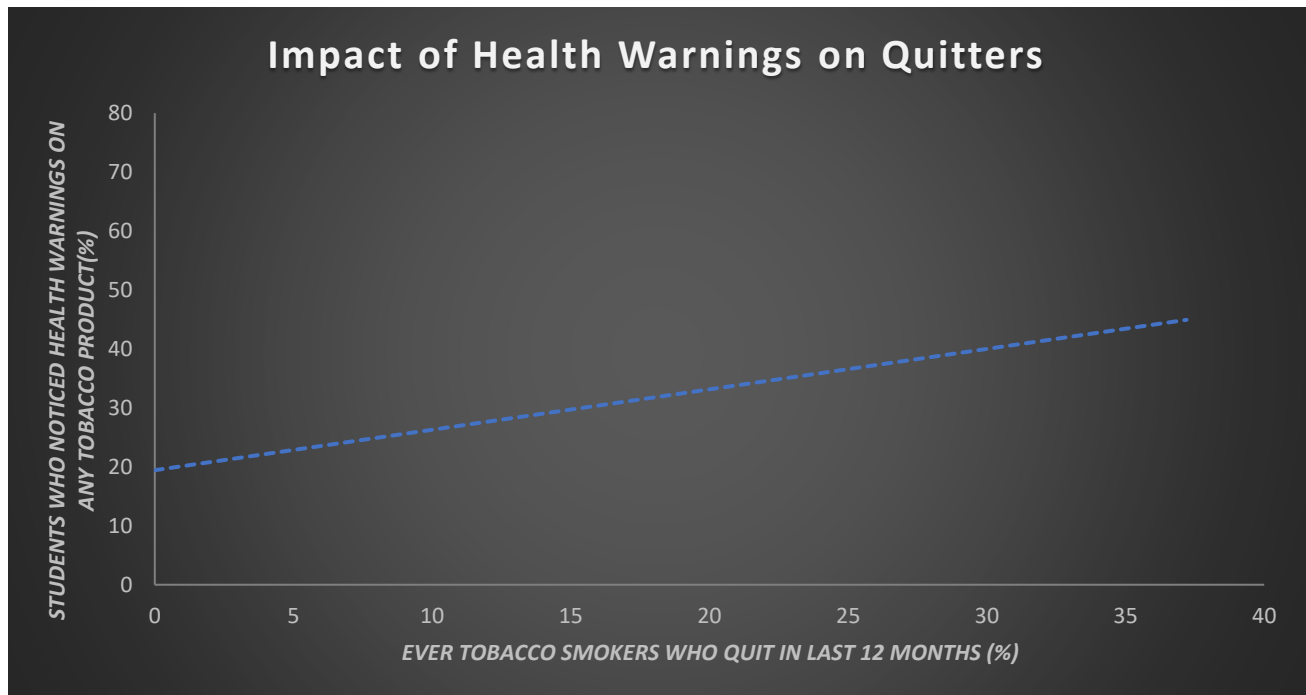
2. Users who want to Quit



INTERPRETATION

- Despite having the lowest proportion of tobacco users, Karnataka has the highest number of people who want to quit both smokeless tobacco as well as smoking tobacco.
- North-Eastern States of Nagaland, Meghalaya, Mizoram and Sikkim also have high proportion of people who want to quit.

3. Impact of Health Warnings on Quitters

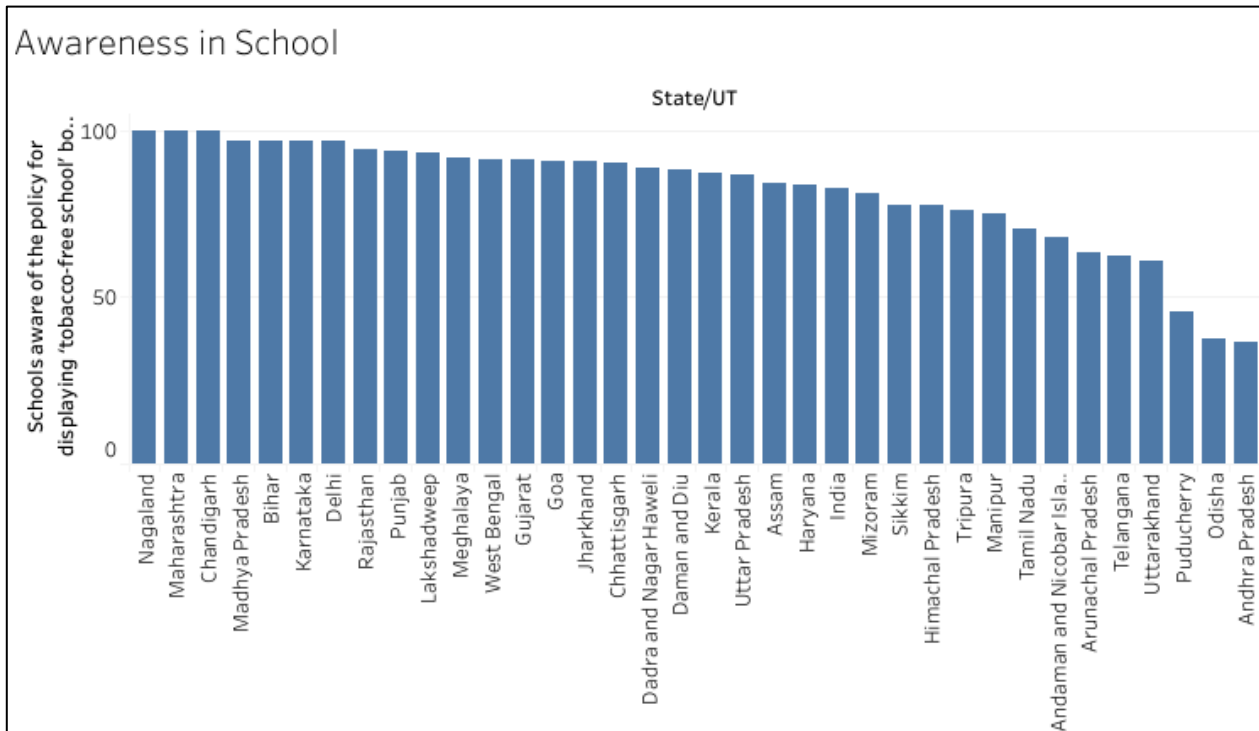


	Ever tobacco smokers who quit in last 12 months (%)	Students who noticed health warnings on any tobacco product(%)
Ever tobacco smokers who quit in last 12 months (%)	1	0.538412997
Students who noticed health warnings on any tobacco product(%)	0.538412997	1

INTERPRETATION

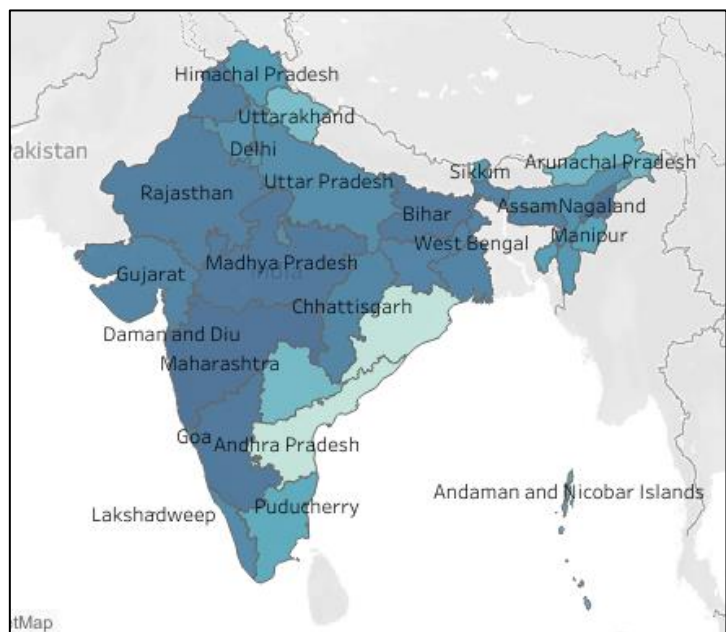
- States where more students noticed health warnings on tobacco product have more quitters as correlation value is 0.538.

4. Awareness in School

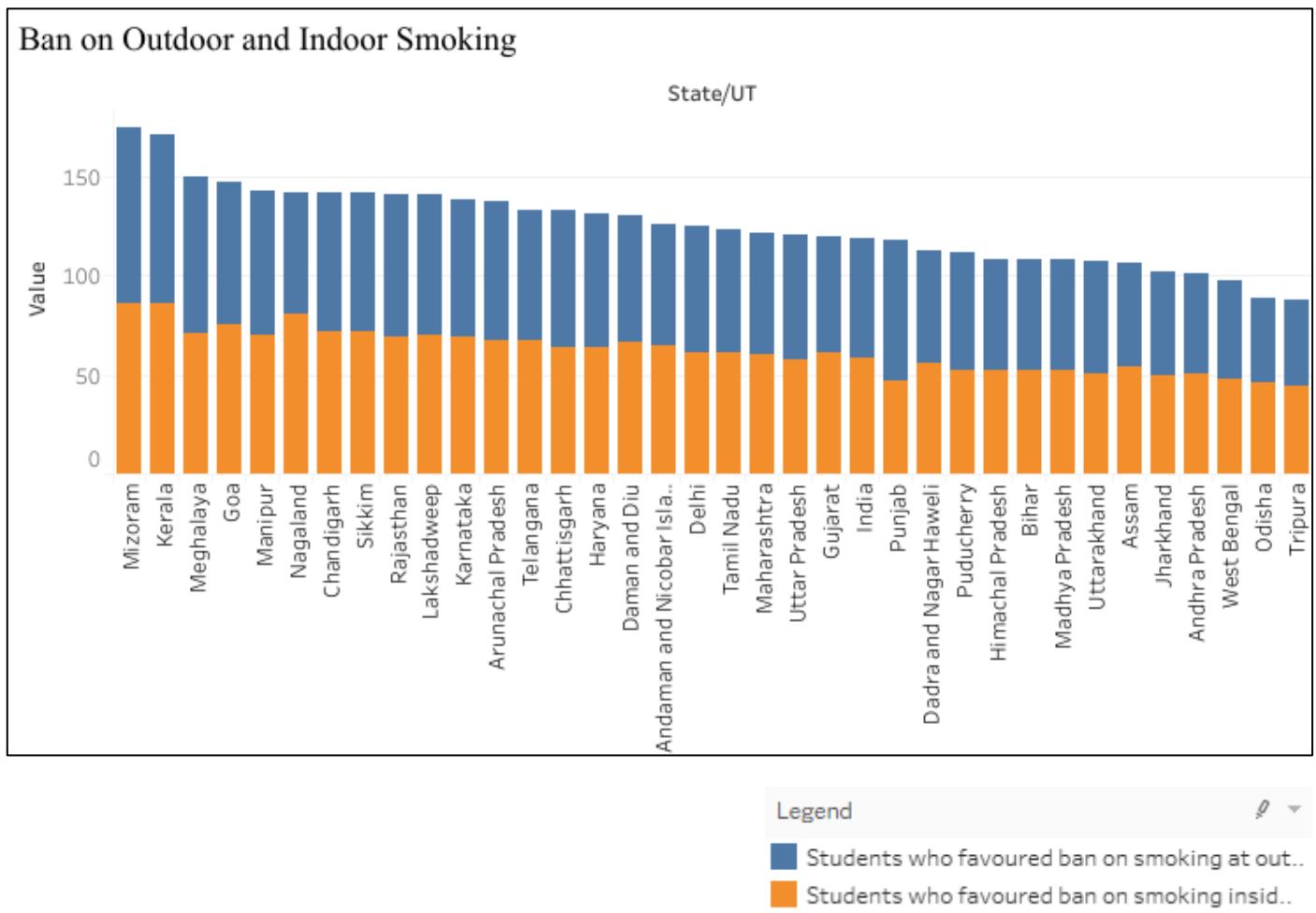


INTERPRETATION

- Around 97% of the schools in Karnataka are aware of the policy for displaying tobacco-free school board.
- North-Eastern States of Nagaland and Meghalaya also display good proportion of school who are creating awareness.



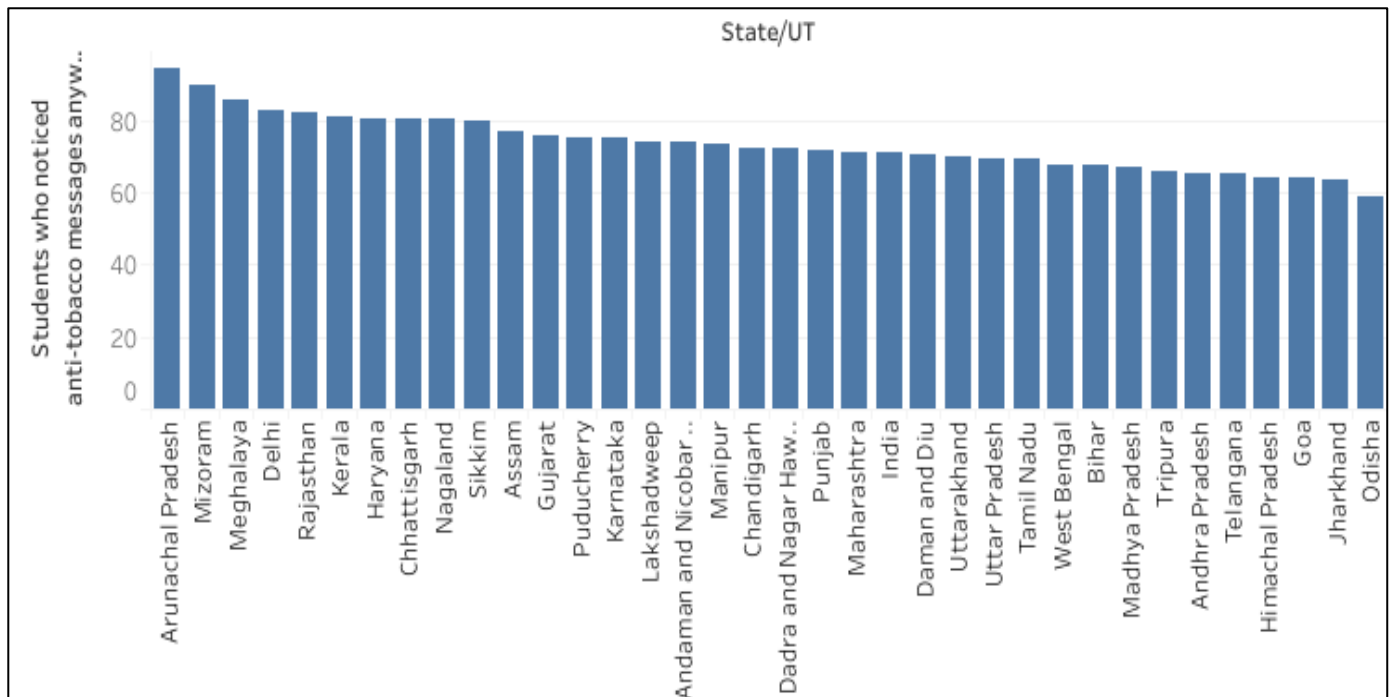
5. Ban on Outdoor and Indoor Smoking



INTERPRETATION

- North-east states of Mizoram, Meghalaya, Manipur and Nagaland display a high percentage of students who are saying to not to smoke outside/inside closed public places.
- The state of Karnataka also has high portion of students wanting ban on both outdoor and indoor smoking.

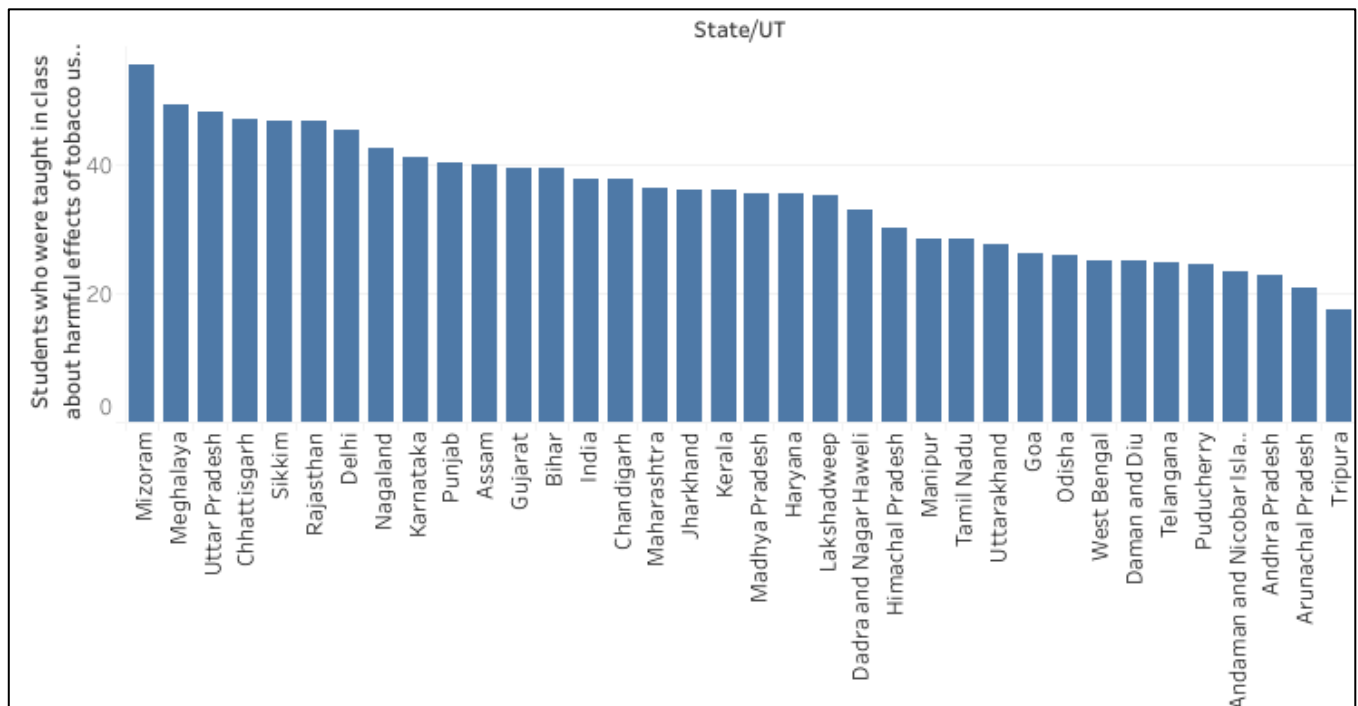
6. Anti-Tobacco Warnings



INTERPRETATION

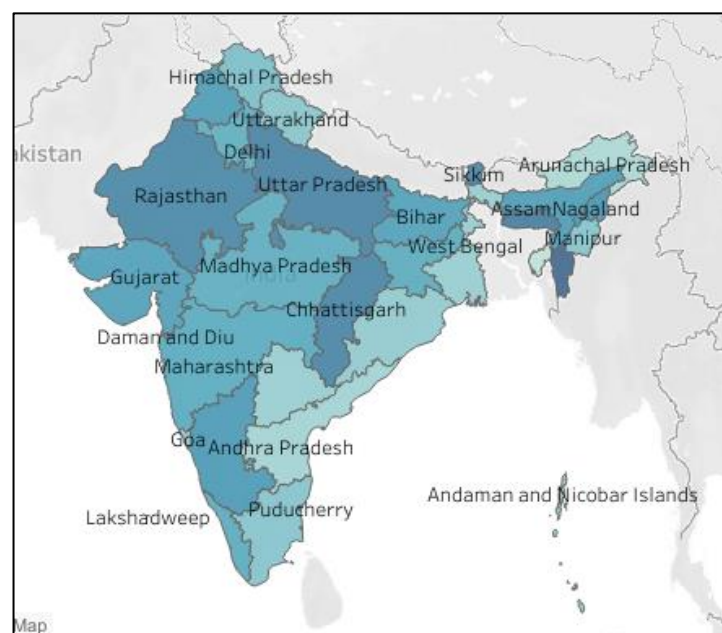
- North-Eastern states have a percentage of students who have noticed anti-tobacco warnings.
- The states of Arunachal Pradesh, Mizoram and Meghalaya all have at least 85% of students who have seen anti-tobacco warnings.

7. Teachings in Class

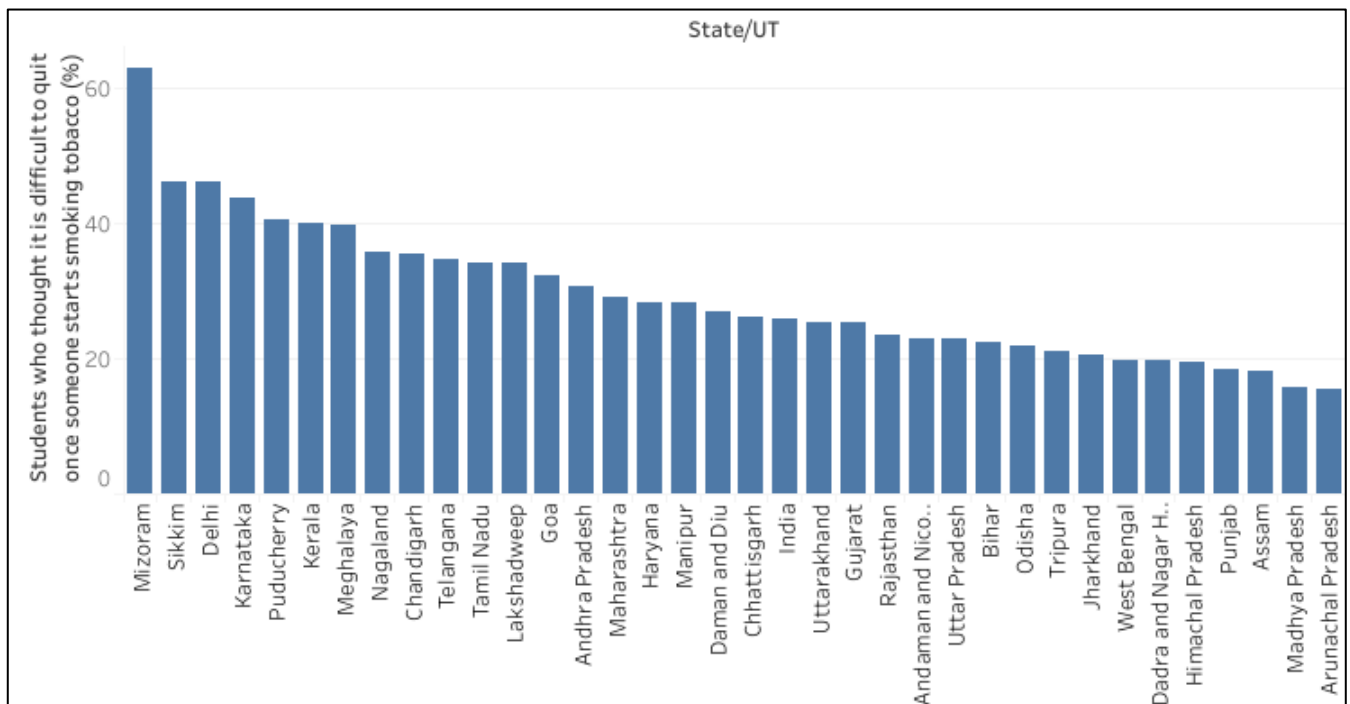


INTERPRETATION

- Again, 42%+ students in Mizoram, Meghalaya, Sikkim and Nagaland were taught about the harmful effect of tobacco use.



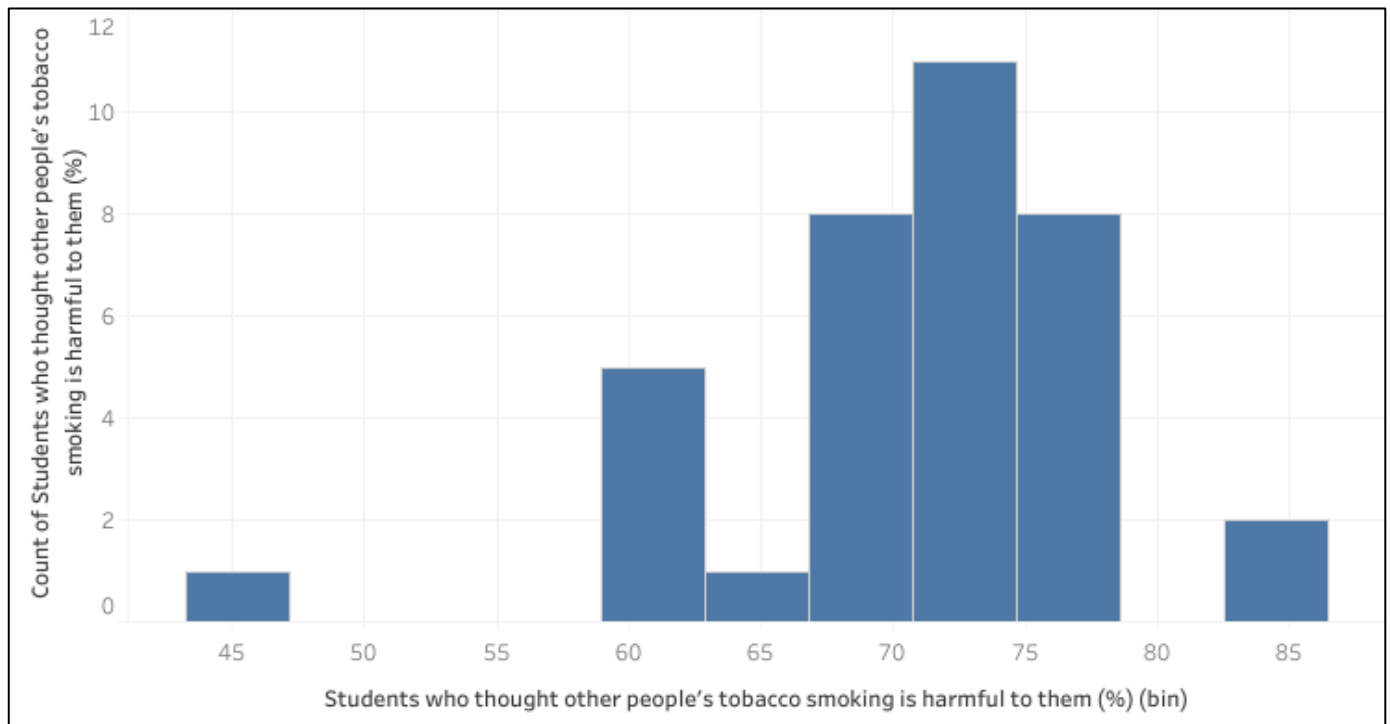
8. Difficult to Quit



INTERPRETATION

- North-eastern states of Mizoram and Sikkim along with Meghalaya and Nagaland show high percentage of students who think it's difficult to quit smoking tobacco once someone starts it.
- Even Karnataka is among the top 5 states with students believing it's difficult to quit once someone starts smoking.

9. Exposure to Second-Hand Smoking

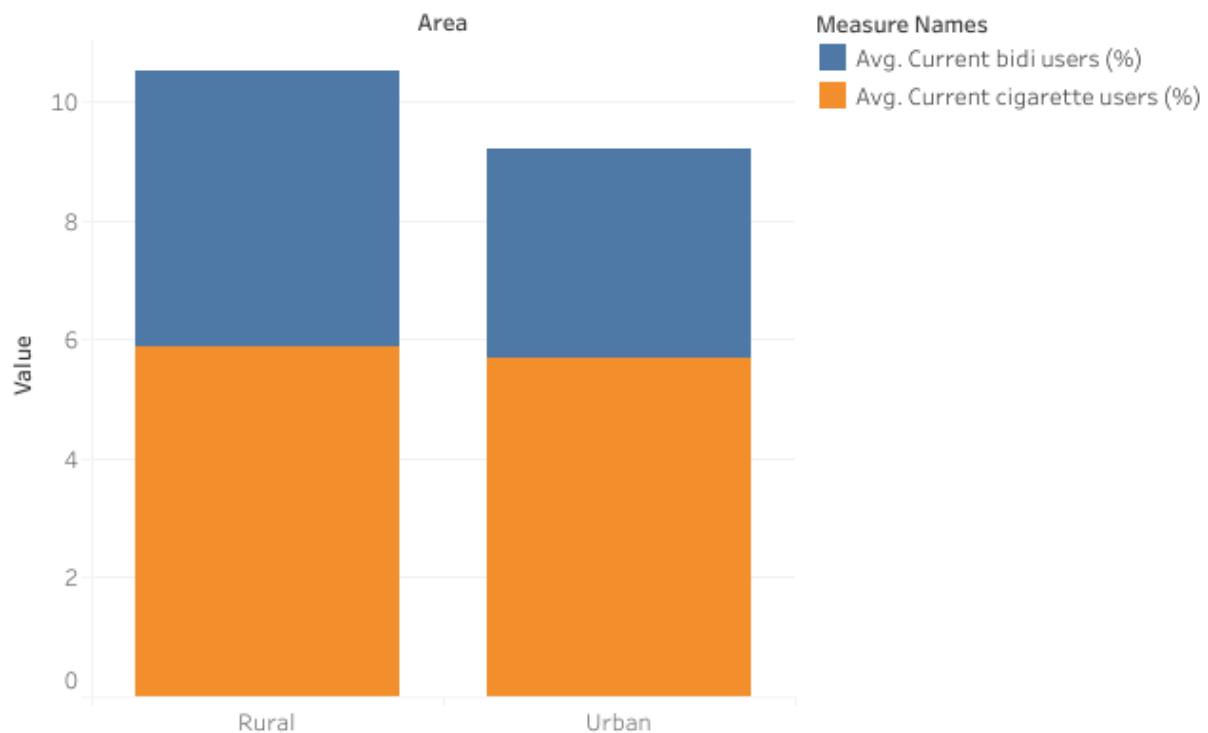


INTERPRETATION

- A very percentage of students agree that other's people smoking is harmful to them i.e., they believe exposure to second-hand smoking can deteriorate their own health.
- The only outlier state here is Arunachal Pradesh where lesser number of students think that SHS is a serious issue.

RURAL-URBAN ANALYSIS

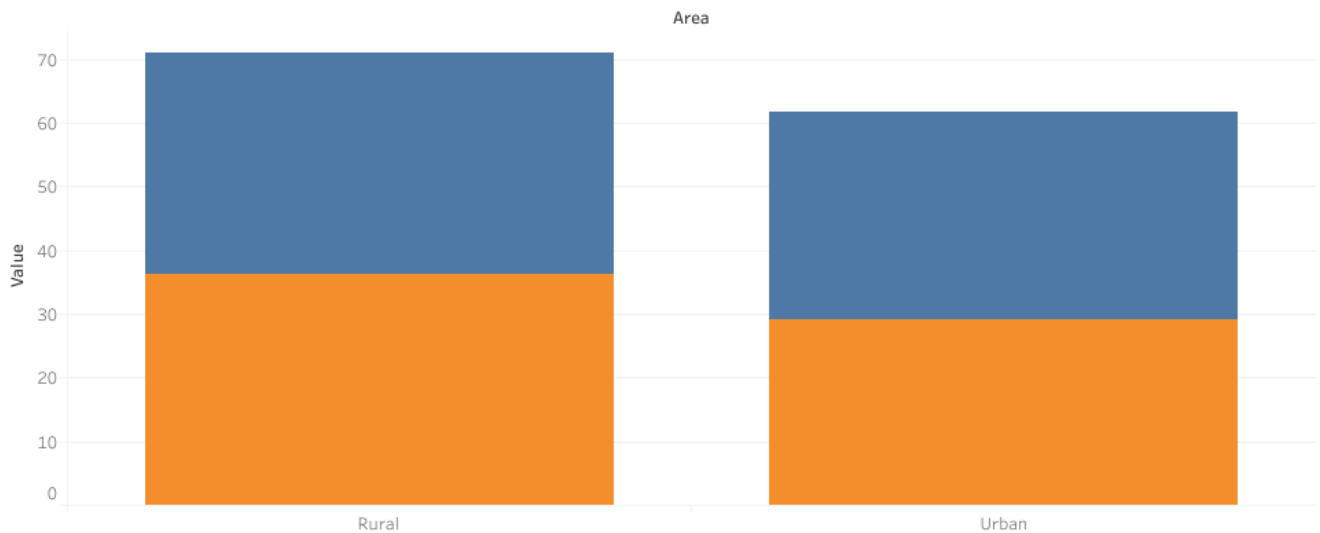
1. Bidi vs Cigarette



INTERPRETATION

- Proportion of cigarette smokers is somewhat same in both rural and urban areas. However, proportion of bidi users is more in rural areas as compared to urban areas. This goes with general belief that bidi being cheaper tobacco product than cigarette, is used more in rural areas as they have lower income levels.

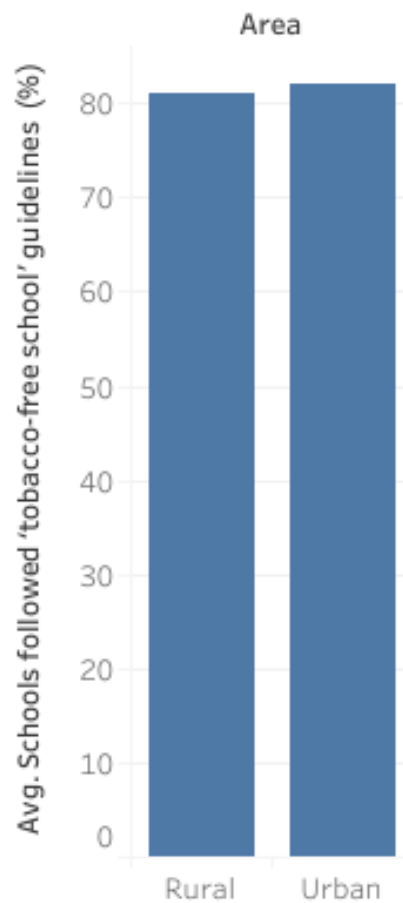
2. Quitters



INTERPRETATION

- Interestingly, rural areas have high number of people who want to quit smoking. Rural areas are expected to have lower awareness about harmful effects of tobacco than the urban population, and yet they are ahead in the quitting race.

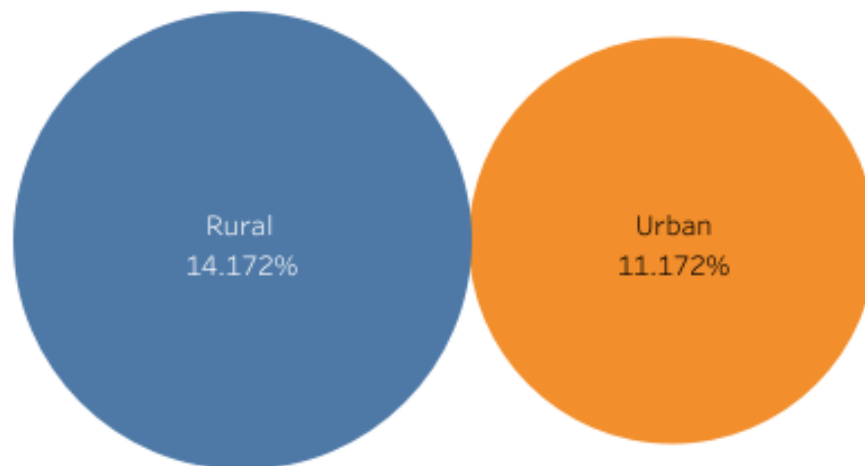
3. Awareness in Schools



INTERPRETATION

- On an average, 81% of the schools in both Rural and Urban follow the guidelines for being tobacco free school.
- Hence, schools in both urban and rural areas of India display similar level of enthusiasm in creating awareness.

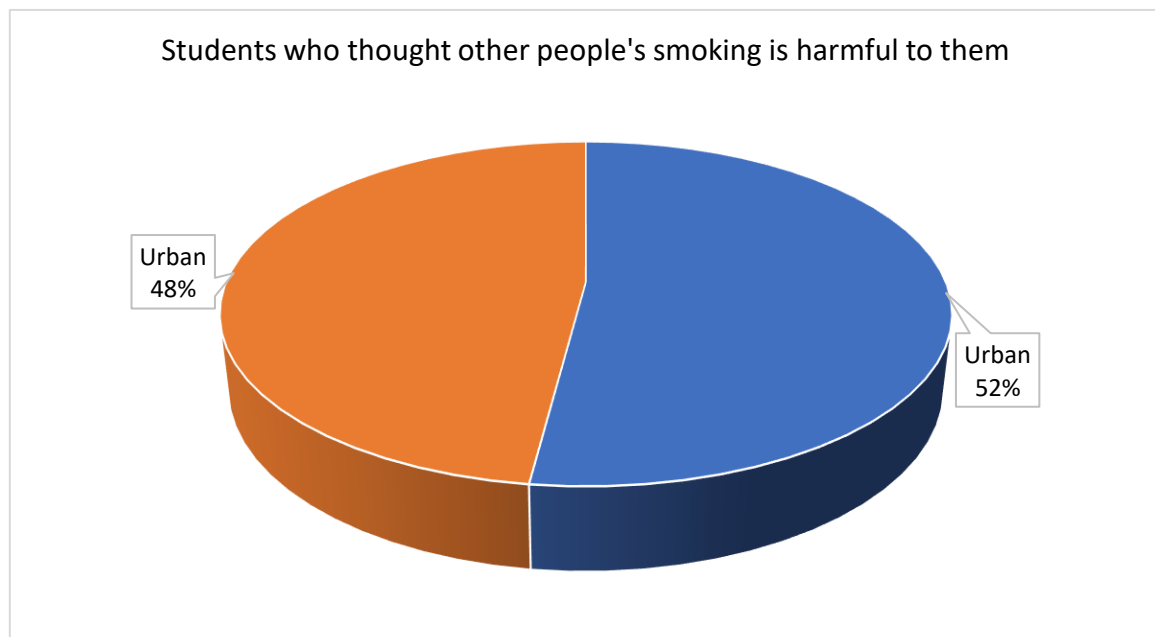
4. Exposure at Home



INTERPRETATION

- On an average, rural areas have higher exposure of tobacco smoke at home than in urban areas.
- As per the above chart, 14.17% of the students have exposure to smoke at home in rural areas.
- The same number is 11.17% for students in urban areas.

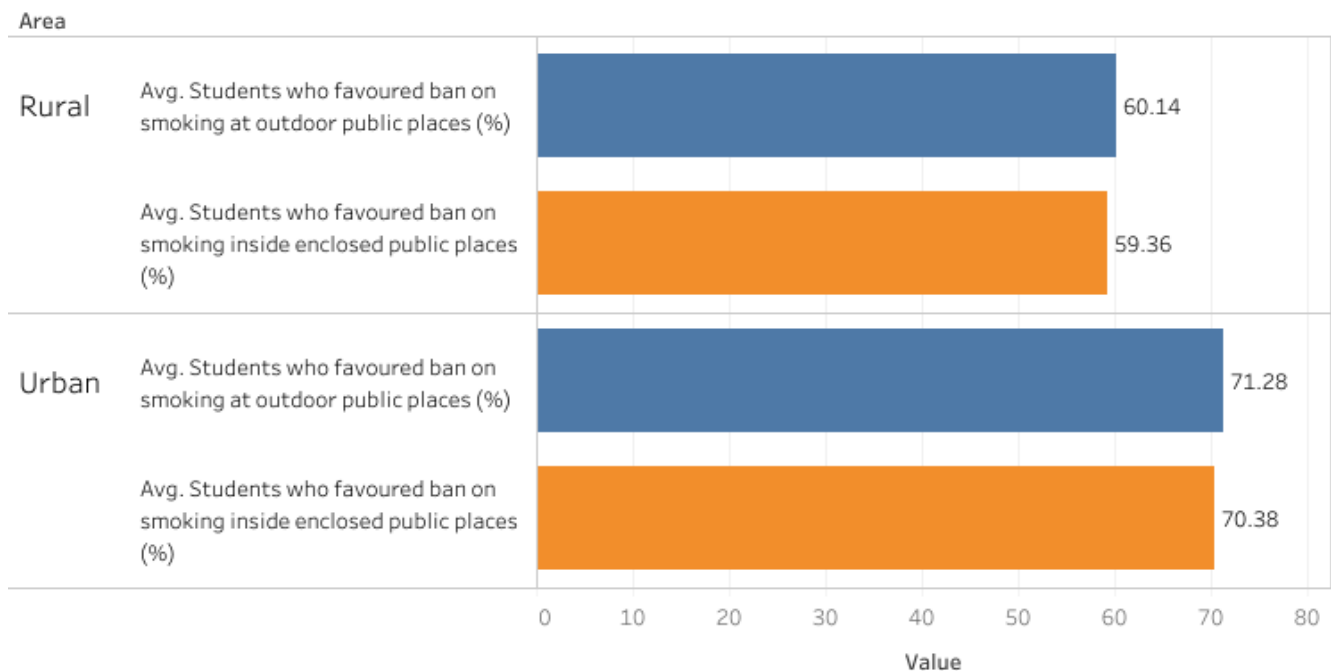
5. Harm to Others (SHS)



INTERPRETATION

- On an average, more students in urban areas think that other's people smoking is harmful for them.
- Around 70% of the total students in Rural areas think that other people's smoking is harmful for them.
- Whereas 75% of the total students in Urban areas think that other people's smoking is harmful for them.

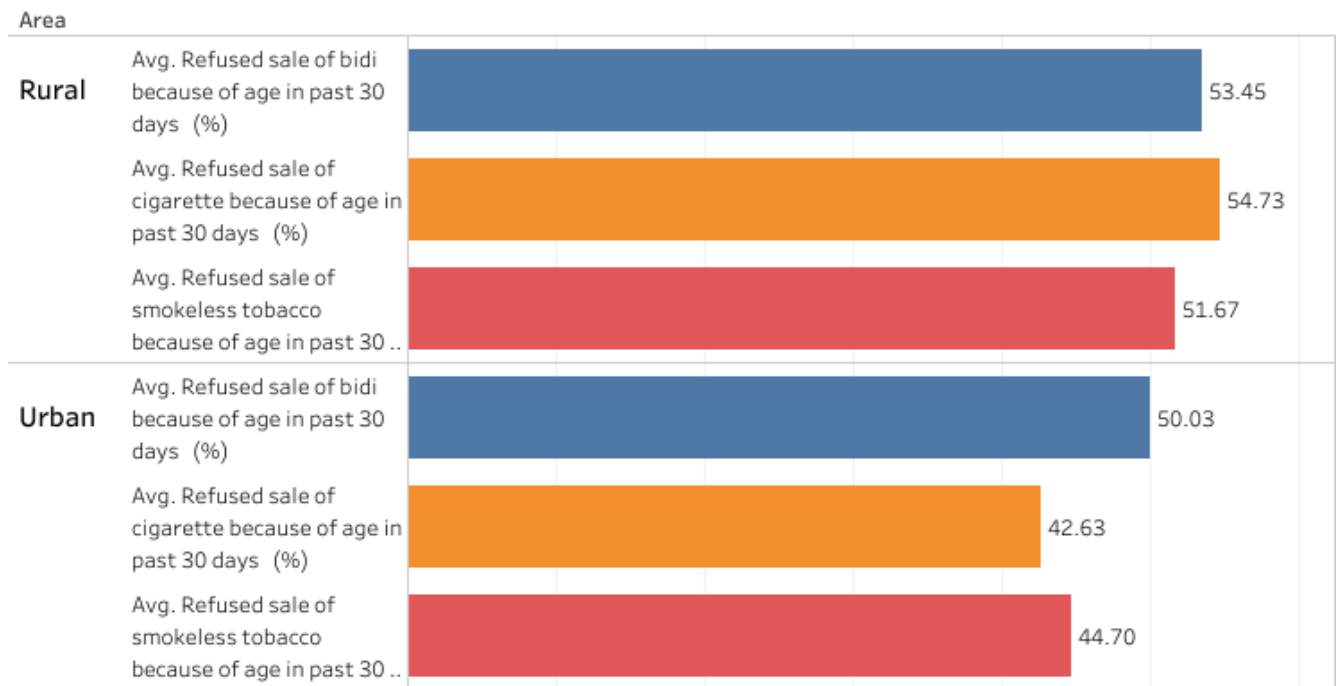
6. Ban on Smoking in Public Places



INTERPRETATION

- On an average, more students in urban areas favour ban on smoking in public places than in rural areas.
- For outdoor places, on an average 71% of the students favour ban on smoking in outdoor public places for urban areas whereas 60% for rural areas.
- For enclosed places, on an average 70% of the students favour ban on smoking inside enclosed public places for urban areas whereas 59% for rural areas.

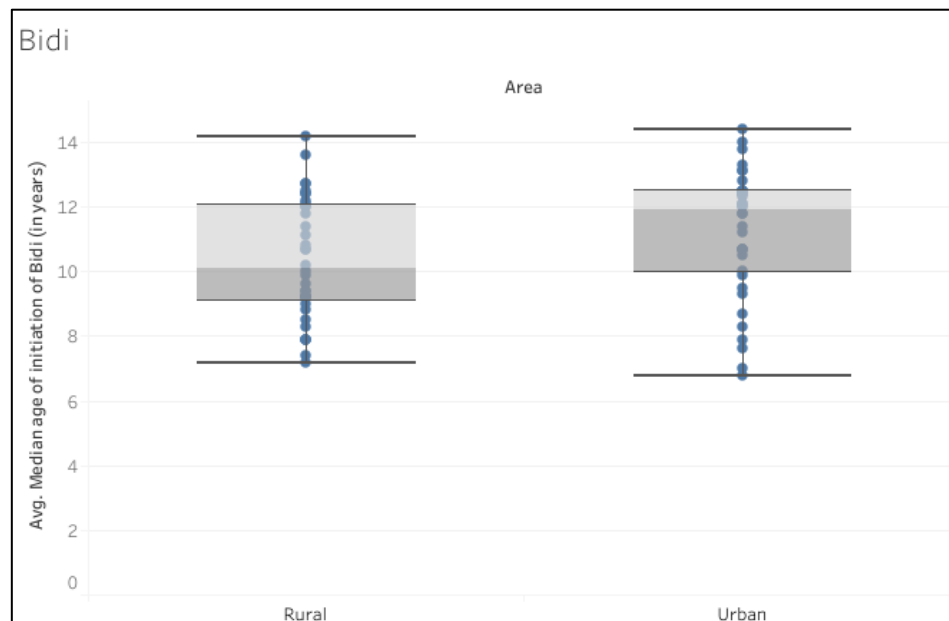
7. Refusal Rate



INTERPRETATION

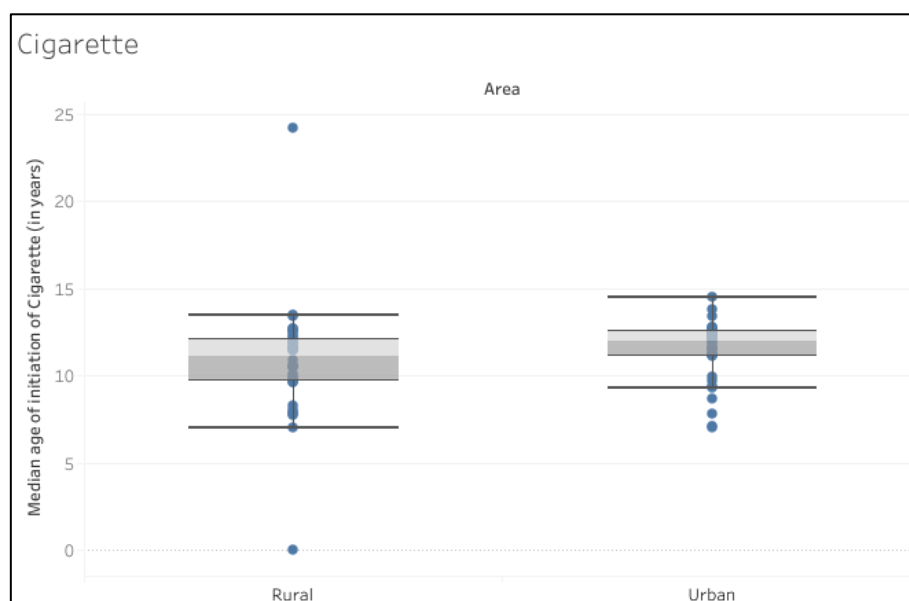
- On an average, rural areas have higher refusal rates than urban areas.
- As seen in the chart for all the tobacco products viz. bidi, cigarette and smokeless tobacco, the percentage of students being refused sale because of age is higher in rural areas with 53%, 54% and 51% respectively for the different products as compared to 50%, 42% and 44% in urban areas.

8. Median Age of Initiation



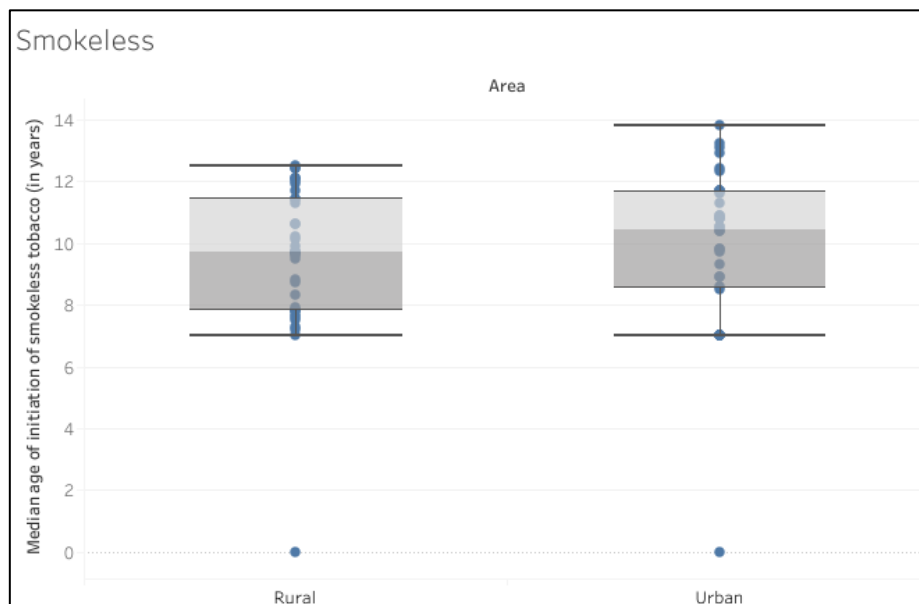
INTERPRETATION

- Median age of bidi initiation in rural areas is right-skewed whereas for urban areas it's left skewed.
- This means that rural areas have more data-points in higher value ages and urban areas have more data-points in lower value ages.



INTERPRETATION

- Median age of cigarette initiation in rural areas is generally lower as compared to urban areas.
- The outlier here is Manipur with a Median Age of 24.20 years in rural areas.



INTERPRETATION

Median age of initiation of smokeless tobacco products in rural areas is somewhat normally distributed and in lower range as compared to urban areas.

OBSERVATION & CONCLUSION

STATE-WISE

➤ Karnataka

- It has the lowest ever and current tobacco users.
- Despite that, it has the highest number of people who want to quit tobacco.
- One reason could be that 97% schools in Karnataka want to be tobacco free schools.
- This could also be the reason why students in Karnataka want a ban on outdoor and indoor smoking.
- Here, the students think that it's difficult to quit once someone starts smoking tobacco.

➤ North-Eastern States

- These states have high tobacco consumption rates.
- North-eastern states also have high proportion of people who want to quit.
- Here, the schools also want to become tobacco free schools.
- Same as Karnataka, students here want a ban on outdoor and indoor smoking.
- Here, high percentage of students who have noticed anti-tobacco warnings.
- Even students here are taught about the harmful effects of tobacco.

While it makes sense for Karnataka students to say this as we saw in earlier graphs that Karnataka has a low incidence of tobacco users, and have a high number of schools creating awareness for being tobacco free. But it's interesting to see that the students in North-eastern states also want a ban on smoking tobacco products.

Also, interestingly, despite having high number of schools educating students about the harmful effects of tobacco and high percentage of anti-tobacco messages in North Eastern States, these states have high percentage of tobacco users. This leads us to an interesting question. Do north-eastern states have high incidence of people who consume tobacco products because they see more of the

anti-tobacco warnings than other states? Does reverse psychology comes into play here?

Humans are notoriously curious creatures, so when you label something as the forbidden fruit and then say someone can't have it, you ultimately make them want it even though they might not have even wanted it to begin with.

RURAL-URBAN

Bidi being cheaper tobacco product than cigarette, is used more in rural areas as they have lower income levels.

Schools in both urban and rural areas of India display similar level of enthusiasm in creating awareness. Still, rural areas have high number of people who want to quit smoking.

Rural areas have higher exposure of tobacco smoke at home. Lesser students in rural areas think that other's people smoking is harmful for them. This could lead to serious issues for children who don't consume tobacco. Hence, more education should be imparted on this matter.