The Relationship Between Cigarette Prices and Smoking Prevalence

Parbati Belbase

MS in Data Science, Bellevue University

DSC530-T302 Data Exploration and Analysis (2251-1)

Matthew Metzger

November 17, 2024

# The Relationship Between Cigarette Prices and Smoking Prevalence

# Introduction

Globally, smoking is one of the leading causes of untimely deaths and prolonged health implications and has been a global healthcare burden. To address this issue, governments often make policies to increase taxation on goods like cigarettes and tobacco hoping that the increased prices would motivate smokers to quit smoking. But how effective are these price based strategies? This project aims to analyze the relationship between cigarette prices and smoking prevalence across different countries and find out whether higher prices of cigarettes are associated with lower smoking rates.

# Research Question/Theories:

Does increased prices of cigarettes lead to decrease in smoking prevalence?

**Null Hypothesis (H0):**

There is no significant relationship between cigarette prices and smoking prevalence.

**Alternate Hypothesis (H1):**

Higher cigarette prices lead to lower smoking prevalence.

# Research Summary

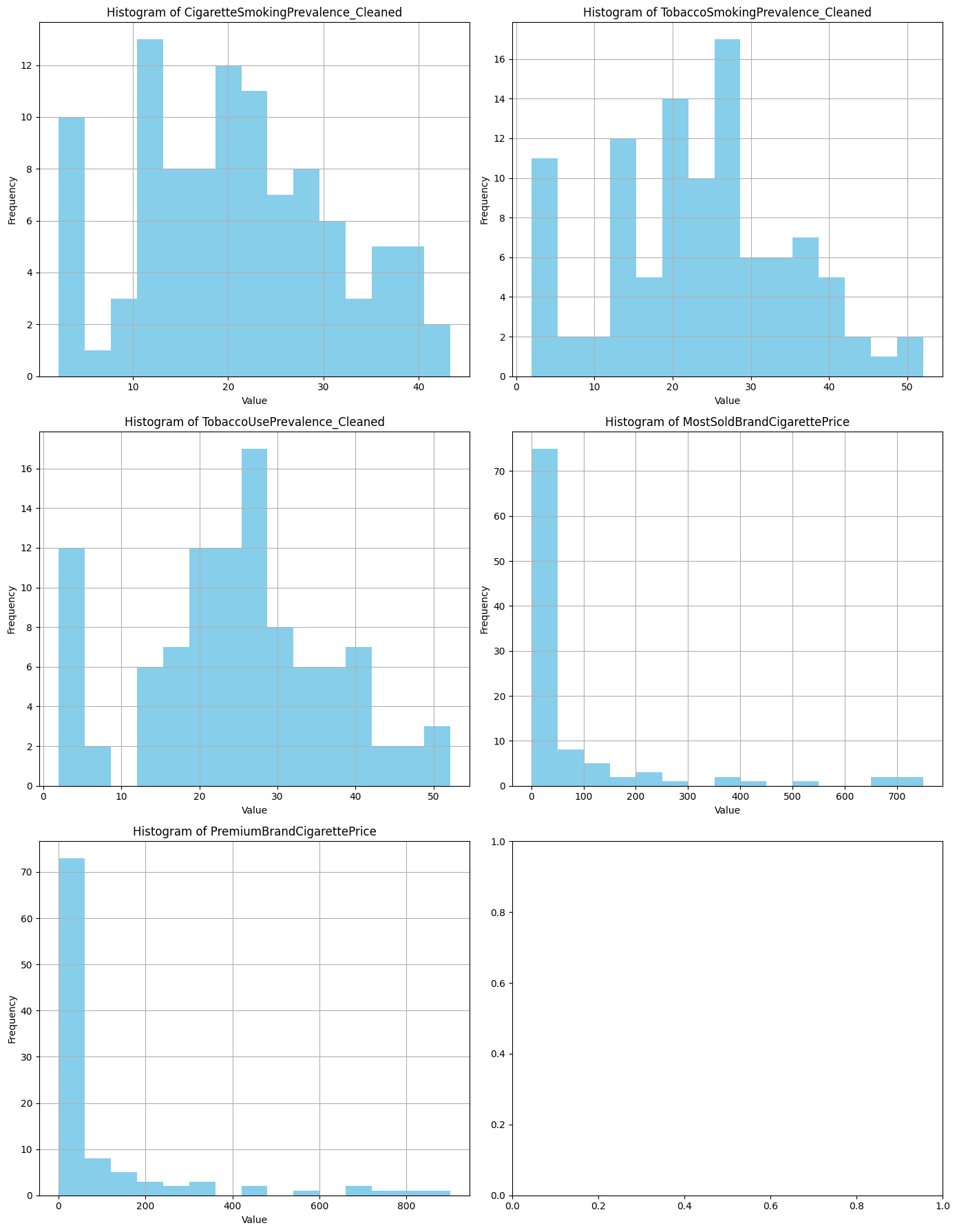
The datasets obtained from Kaggle, includes information on cigarette prices and smoking prevalence from 2008 to 2018. From the dataset, five quantitative variables were selected : Cigarette Smoking Prevalence, Tobacco Smoking Prevalence, Tobacco Use Prevalence, Most Sold Brand Cigarette Price and Premium Brand Cigarette Price. The raw datasets were cleaned, merged and prepared for further analysis.

Exploratory Data Analysis (EDA) was performed on the cleaned dataset to identify the trends and outliers. After analyzing EDA, further statistical techniques such as Probability Mass Function (PMF), Cumulative Distribution Function (CDF), and regression analysis were performed to test the hypothesis.

# Key Findings

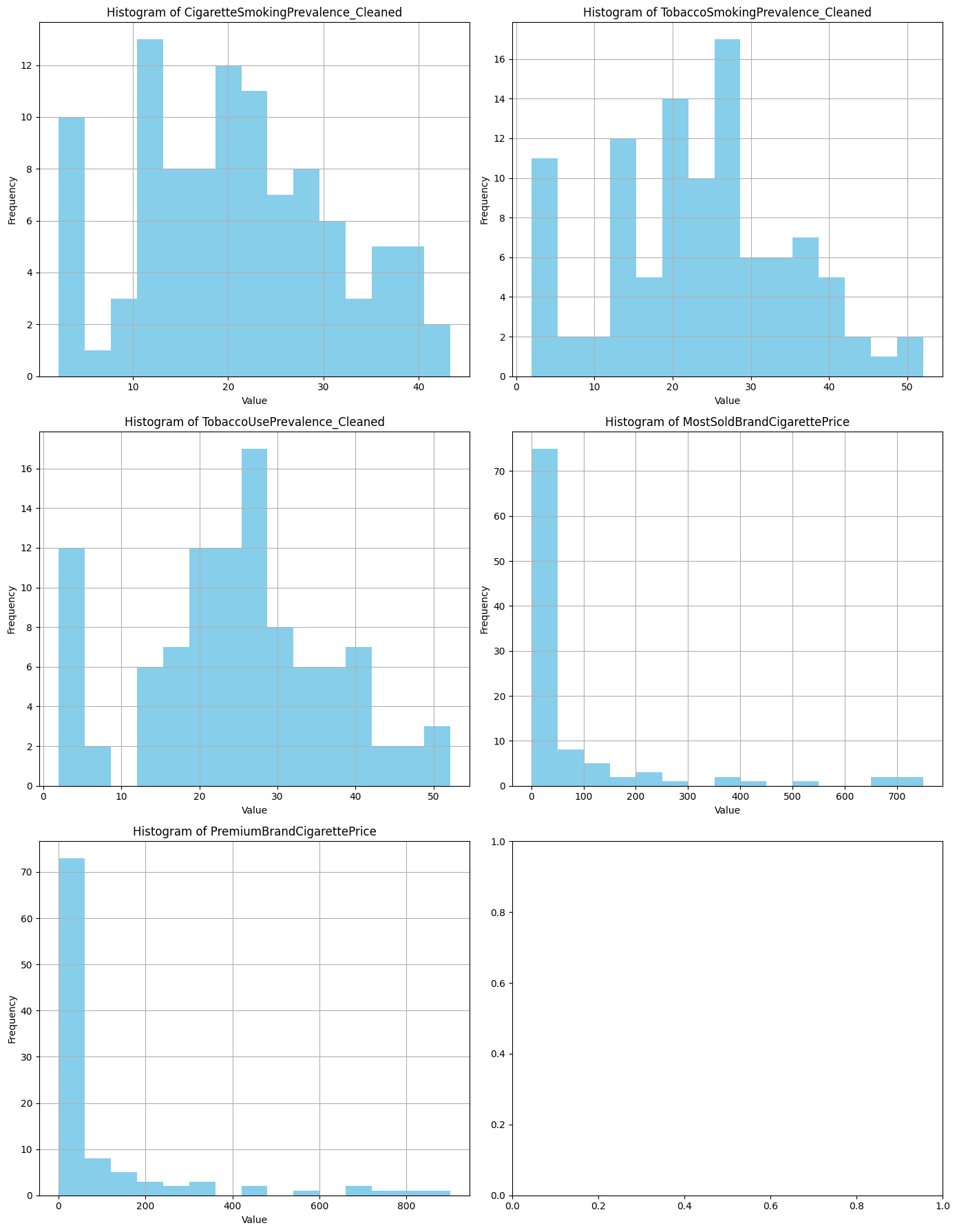
## EDA Results

**Histogram of Cigarette Smoking Prevalence**

****

It can be seen that smoking prevalence is distributed moderately as most countries have prevalence rate below 30% except for a few outliers.

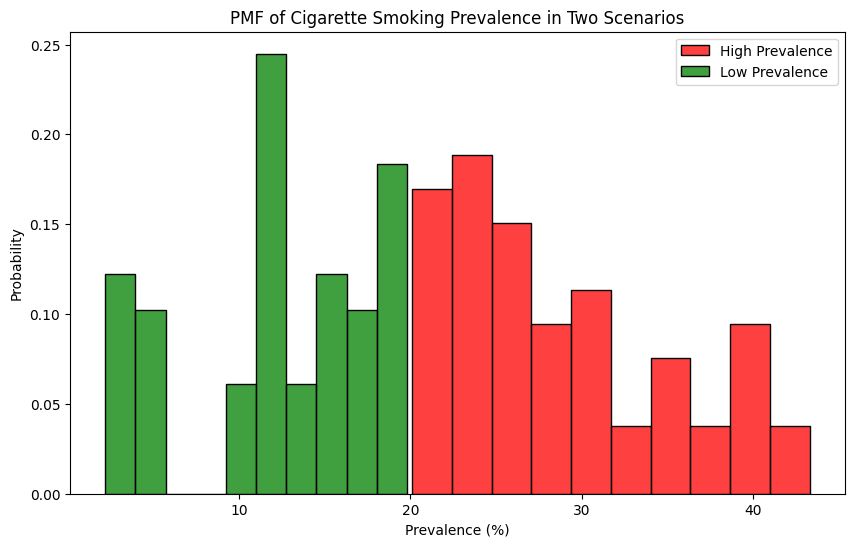
**Histogram of Most Sold Brand Cigarette Price**

****

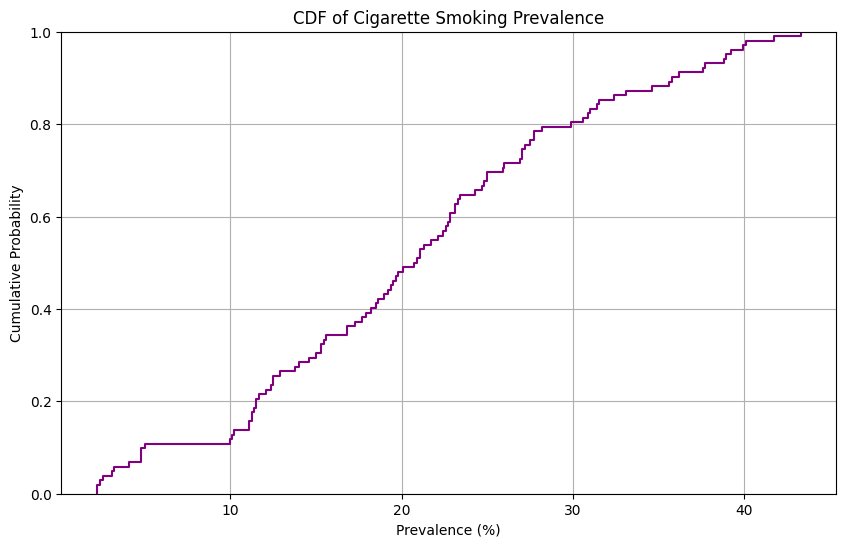
This shows a diverse distribution in cigarette prices among countries with most countries having lower price range i.e below $50. However several also have prices exceeding $200. These outliers show the potential difference in taxation.

## 

## PMF and CDF Analysis

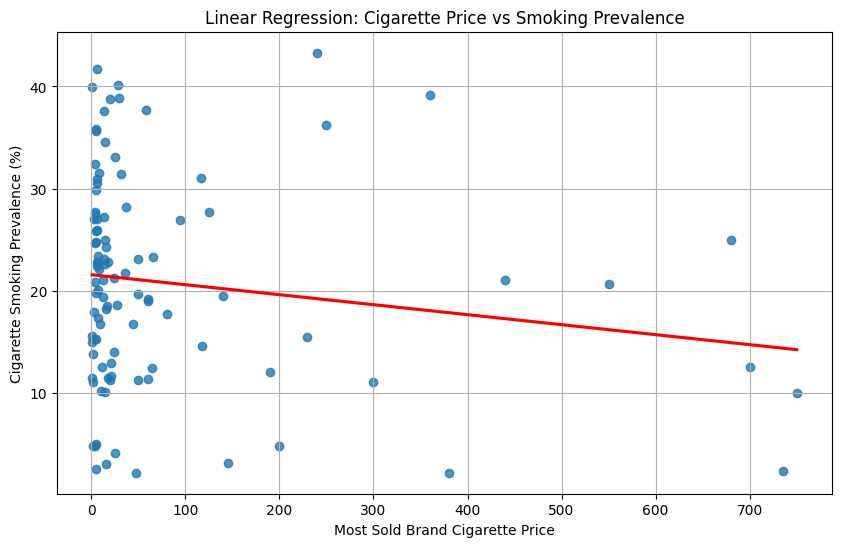


The PMF analysis shows that the distributions for countries with high and low smoking prevalence are distinct. Countries with lower rates have more concentrated distribution compared to the ones with higher rates. This shows different behaviors across different regions.



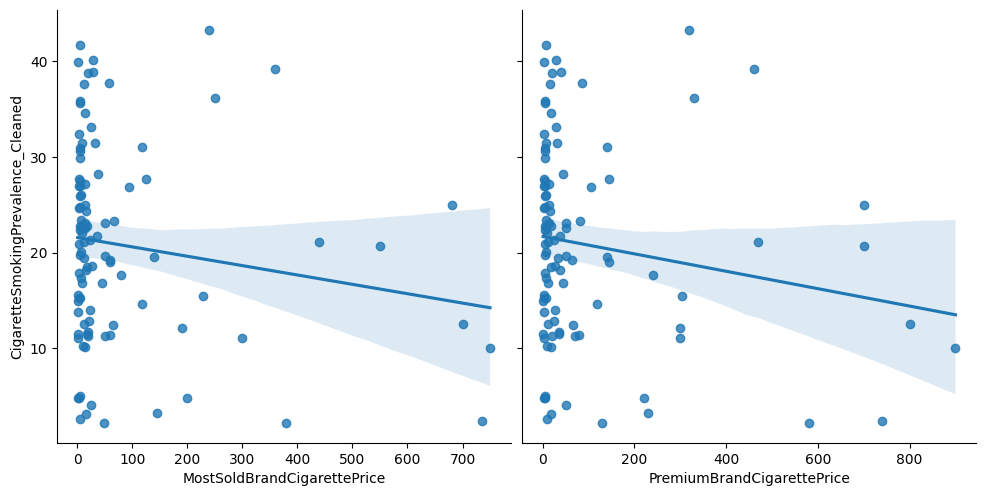
Similarly, the CDF analysis shows that around 50% countries have smoking prevalence below 21%.

## Regression Analysis



#### **Linear Regression Results**

* **R-squared**: 0.024, this indicates only 2.4% of the variation in smoking prevalence is caused by cigarette prices.
* **P-value**: 0.121 suggests that no there is no statistically significant relationship.
* **Coefficient for Most Sold Brand Price**: -0.0098, indicates that there is minimal decrease in smoking prevalence due to increase in prices.



#### **Multivariate Regression Results**

* **Adjusted R-squared**: 0.032, this is a slight improvement but not significant
* **P-values**: Both predictors (Most Sold Brand Price and Premium Brand Price) were statistically insignificant i.e. p-value is less than 0.05.
* **Coefficients**:
  + Most Sold Brand Price: 0.0228
  + Premium Brand Price: -0.0284

From the regression analysis, we can draw a conclusion that cigarette prices alone cannot determine the smoking prevalence. Thus, we fail to reject the null hypothesis (H₀).

# Limitations

The dataset used in the analysis did not include other factors such as public campaigns, bans on smoking etc. which may influence the smoking habits of people. Variables such as government expenditure on awareness programs, access to healthcare and level of restrictions could help in better analysis. A linear relationship was assumed between cigarette prices and smoking prevalence. However, since it is a complex behavior, it may not follow a linear pattern.

# Conclusion

This analysis could not find a significant relationship between cigarette prices and smoking prevalence. After conducting EDA and Regression analysis, it failed to reject the null hypothesis. We can conclude that the price based strategies might not be sufficient to reduce the smoking prevalence across the globe as there might be other factors impacting the people’s behavior of smoking cigarettes.

# References

Le, T. T. T., & Jaffri, M. A. (2022). The association between smoking behaviors and prices and taxes per cigarette pack in the United States from 2000 through 2019. *BMC Public Health*, *22*(1). <https://doi.org/10.1186/s12889-022-13242-5>

World Health Organization. (2024). *WHO global report on trends in prevalence of tobacco use 2000–2030* [Report]. <https://iris.who.int/bitstream/handle/10665/375711/9789240088283-eng.pdf?sequence=1>

*WHO tobacco and smoking data 2008-2018*. (2021, September 6). Kaggle. <https://www.kaggle.com/datasets/ozgurdogan646/who-tobacco-and-smoking-data-20082018>