

IS590 PR Final Project

Analysis of Tobacco consumption in the US

Sushmitha Pakanati



Introduction

- According to Centers for Disease Control and Prevention , Tobacco consumption is the major cause of preventable disease and death in the US
- It is estimated that around 34.2 million adults in the United States smoke or consume tobacco and more than 16 million americans live with a smoking-related disease
- This project analyzes tobacco consumption in the US and the aim is to predict the smoking behavior of people in all the states of the US for the next years



Hypothesis

- The tobacco consumption in the US differ by state and the behavioral trend changes over time



Dataset description

The dataset downloaded is Smoking Tobacco use from Centers for Disease Control and Prevention

Link: <https://data.cdc.gov/Smoking-Tobacco-Use/BRFSS-Prevalence-and-Trends-Data-Tobacco-Use-Four-/8zak-ewtm>



Data Cleaning

- The tobacco-use dataset which is taken for analysis and prediction shows the four level data for all the states in the US from 1995 to 2010
- This dataset is split into individual files which are taken as input files so that the output files will be showing information for every state separately

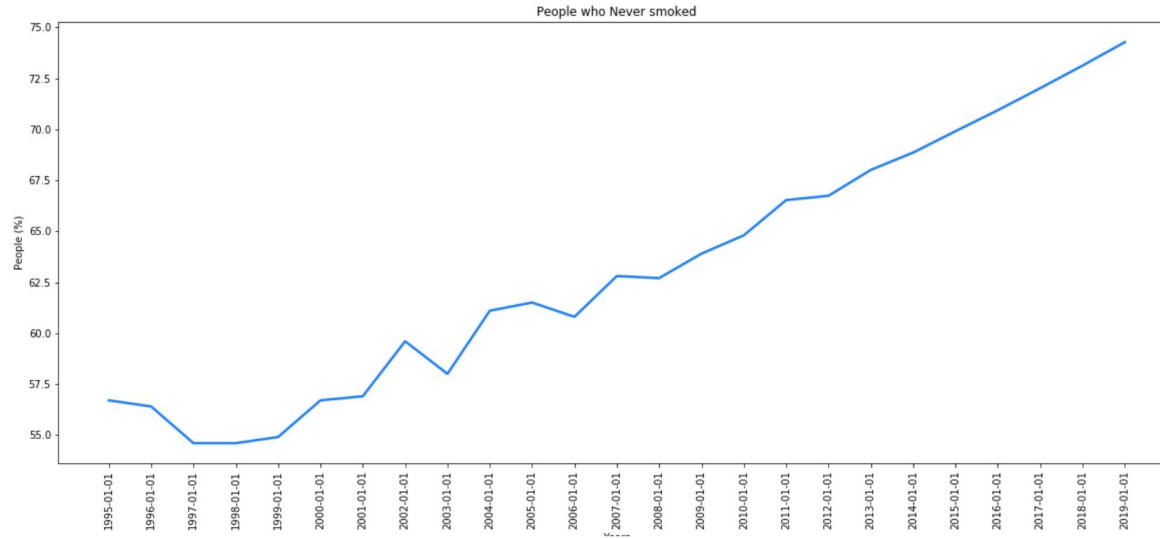


Analysis of the hypothesis

- Using Vector Auto Regression model, prediction is done for the 10 years which is till 2019, showing the percentage of people who smoke every day, people who smoke on some days, former smokers and the non smokers

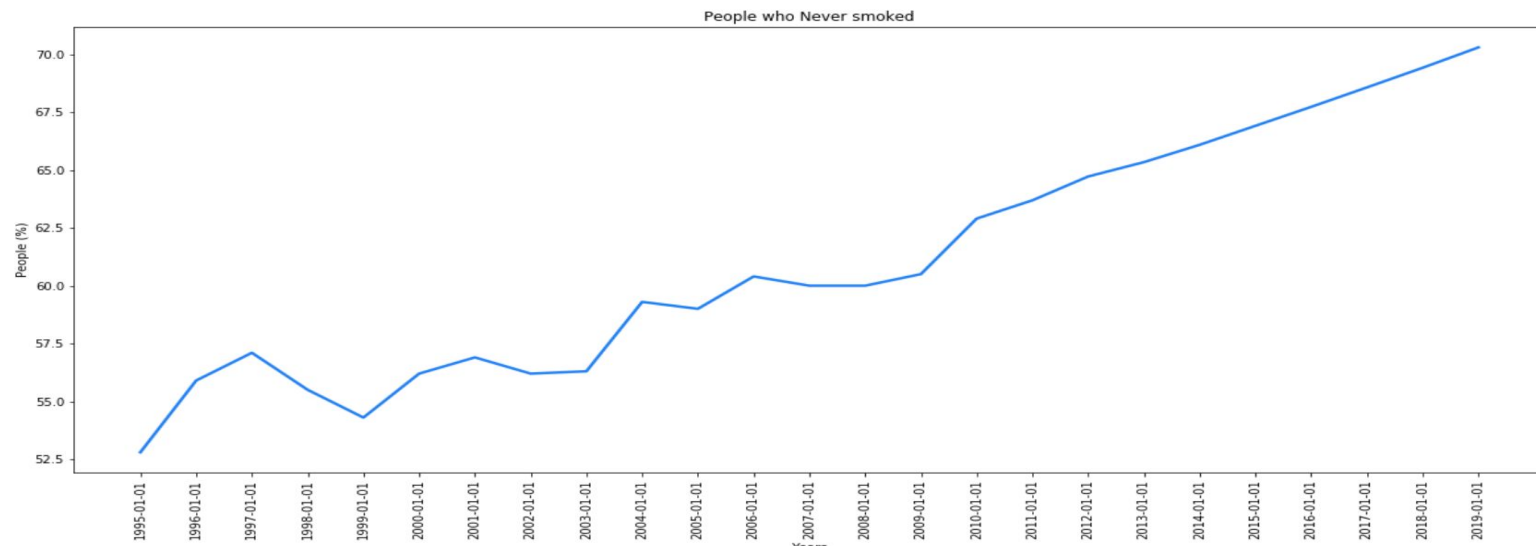
Analyzing the smoking trend in the top 10 states contributing to GDP in the US

California

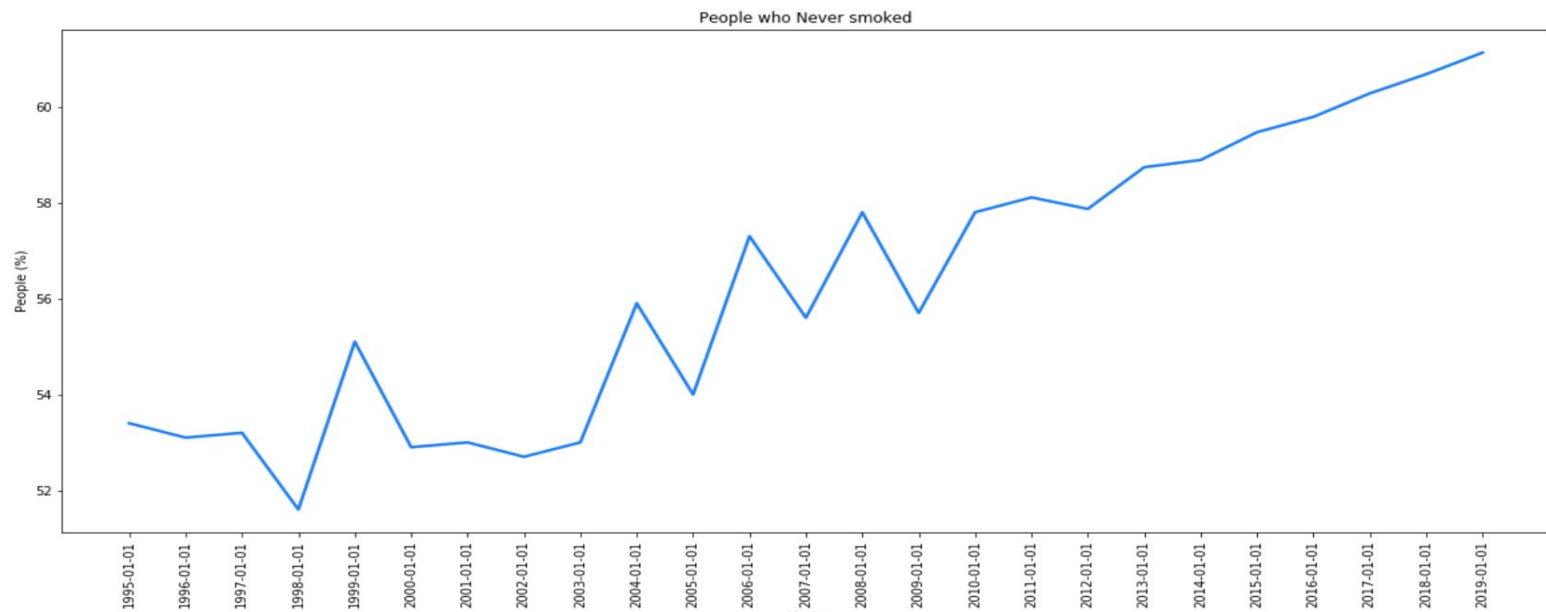




Texas

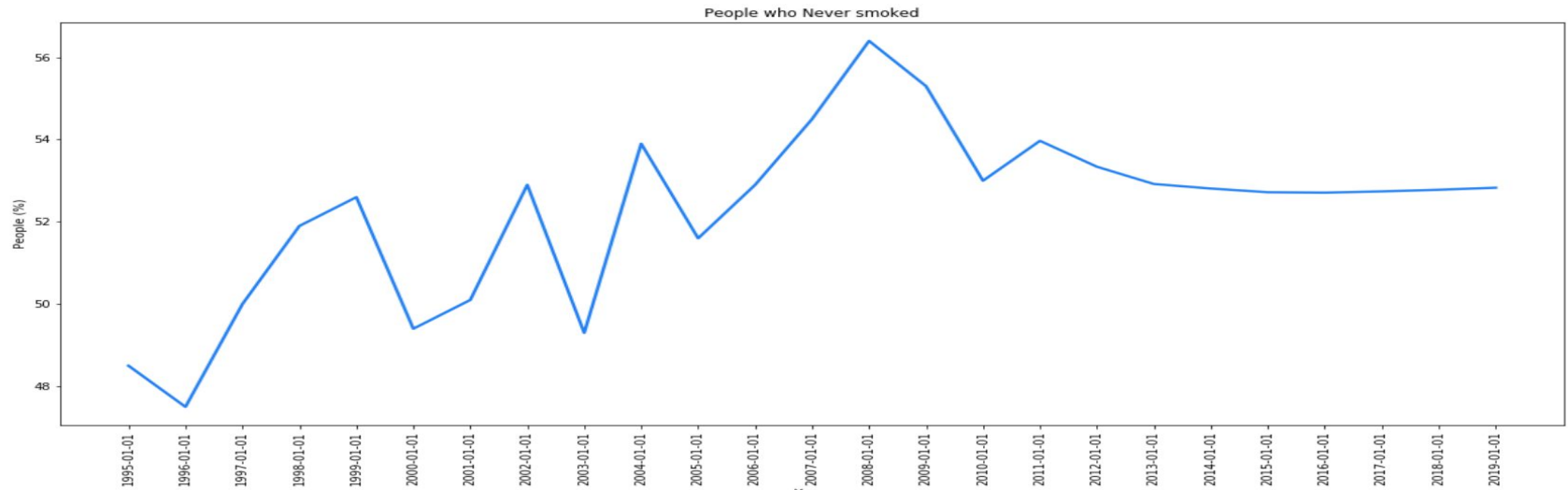


New York



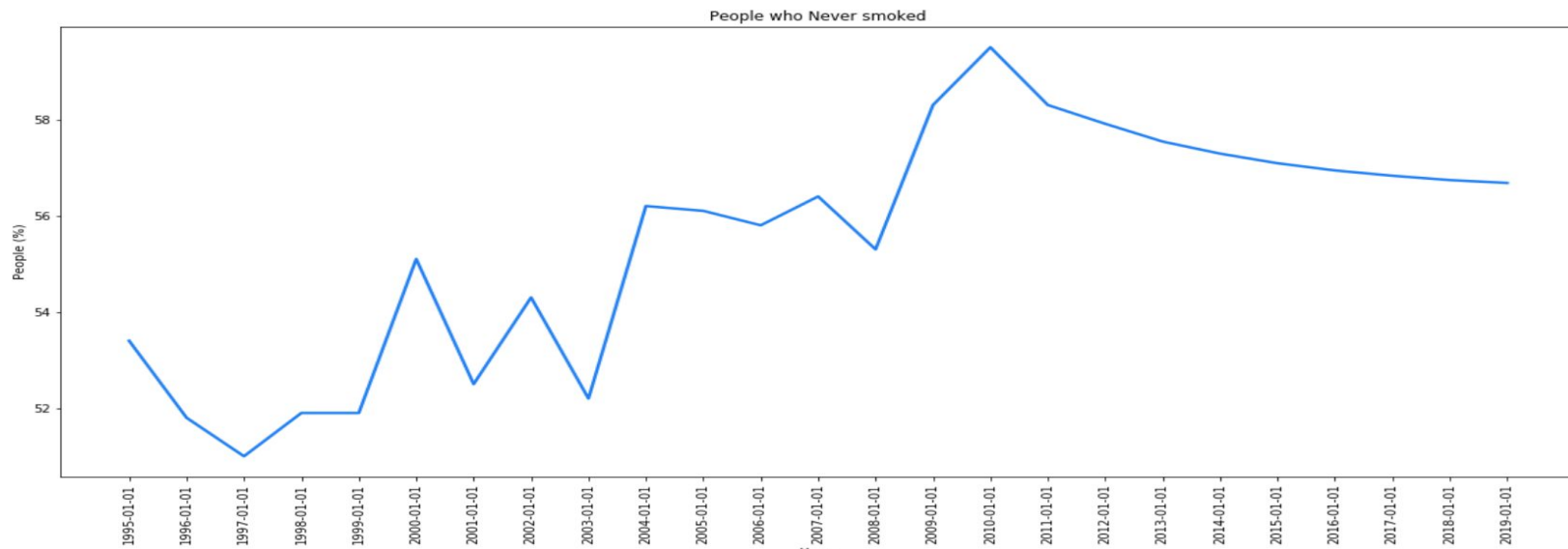


Florida

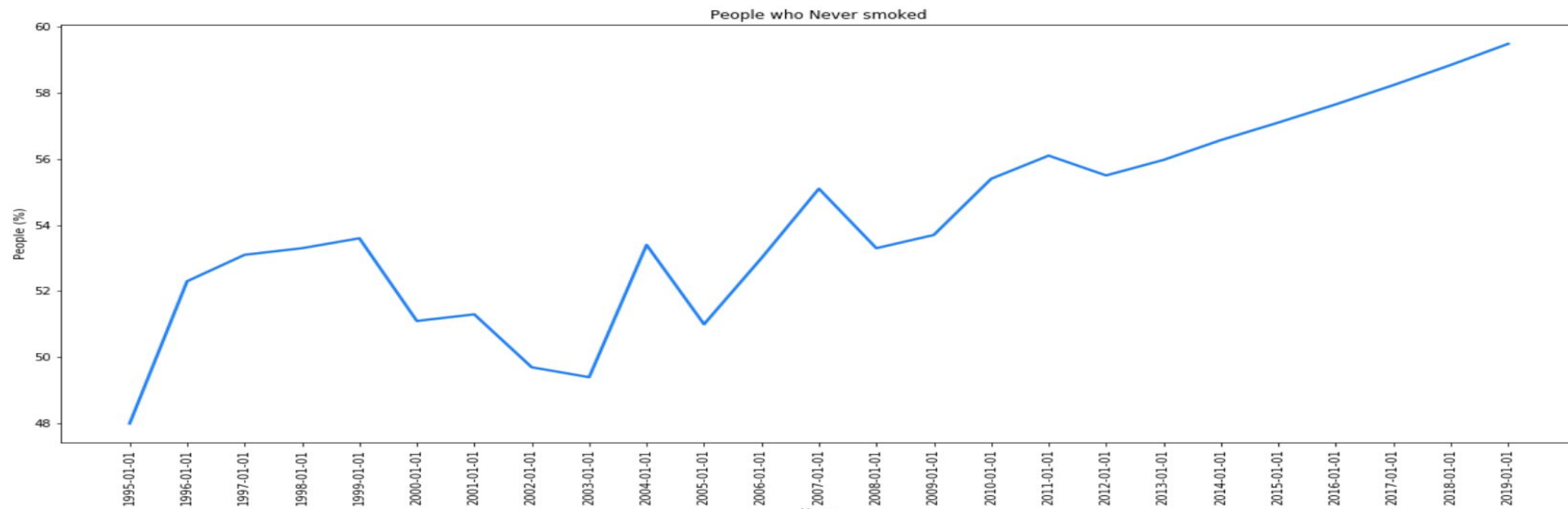




Illinois

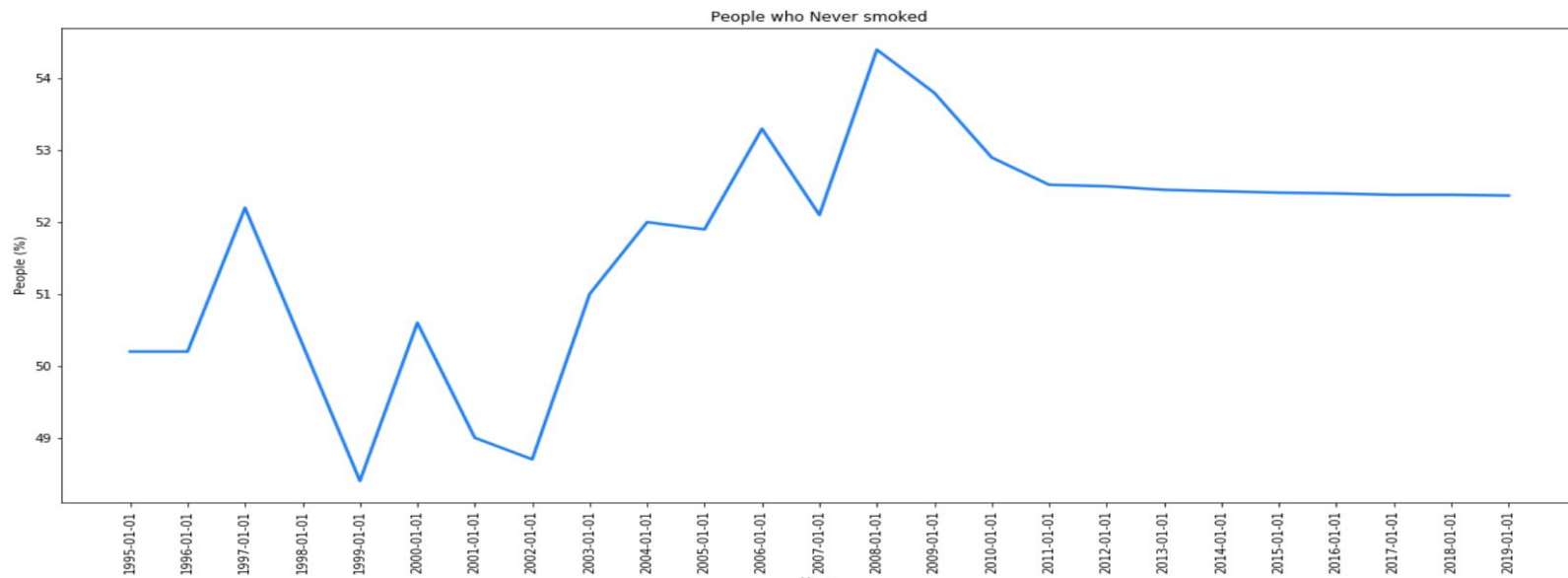


Pennsylvania



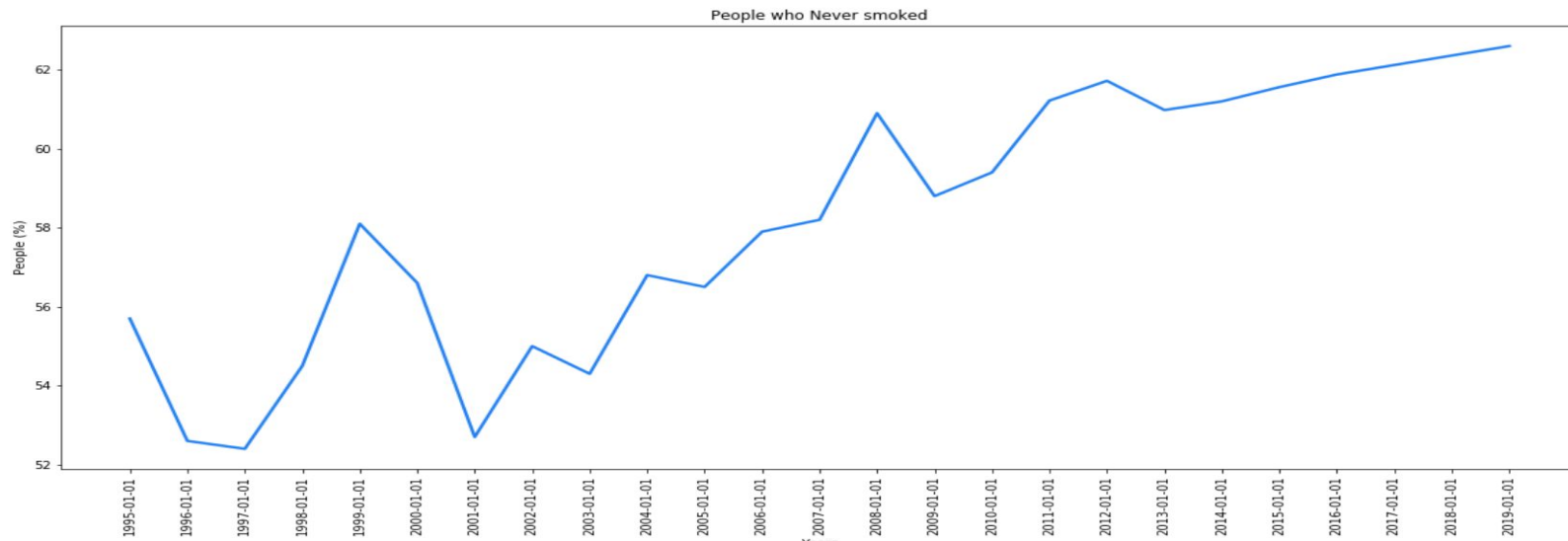


Ohio



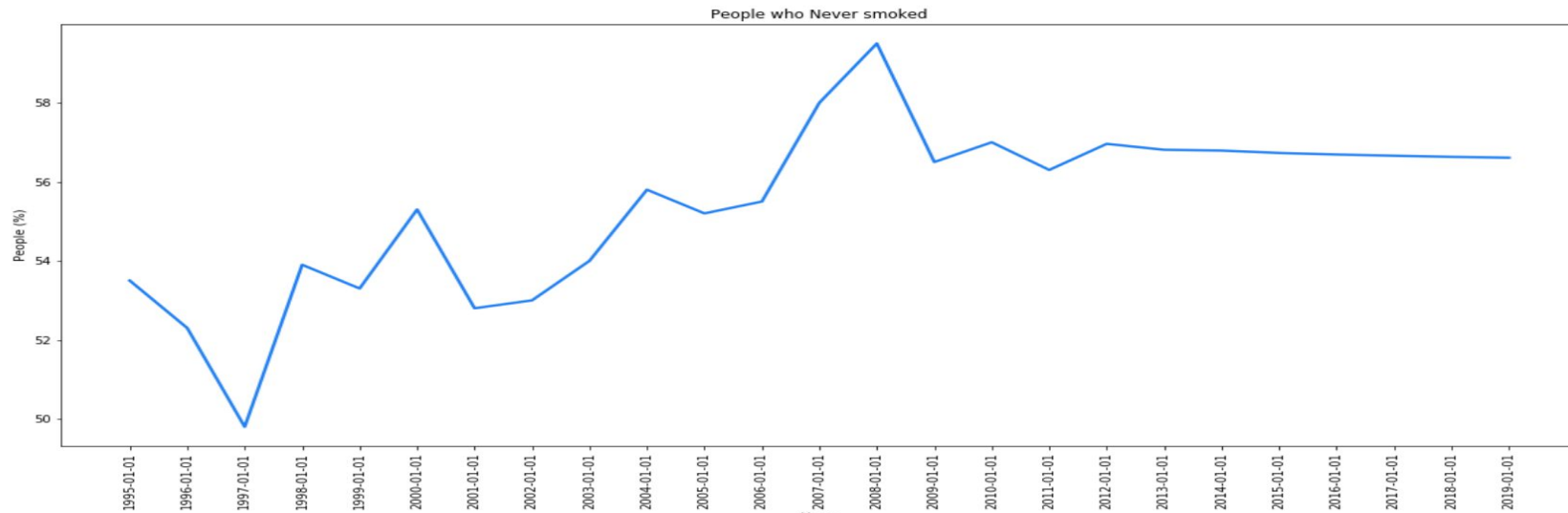


New Jersey



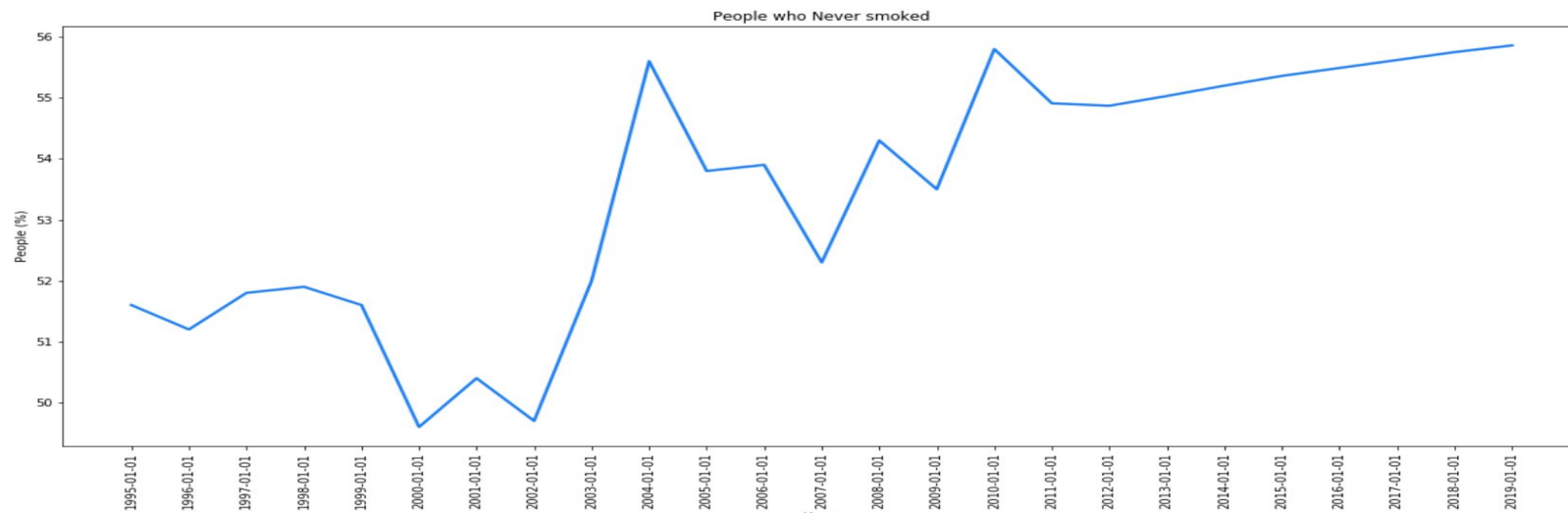


Virginia





North Carolina





Conclusion

- From all these observations we can infer that over time, with improvements in resources to help smokers quit and new control laws in place, we can see from the prediction model that more number of people may stop smoking in the future and the hypothesis is proven right



References

1. <https://stackoverflow.com/>
2. <https://www.cdc.gov/tobacco/about/osh/state-fact-sheets/california/index.html>
3. <https://chiefexecutive.net/the-10-top-states-by-gdp/>
4. <https://www.kaggle.com/cdc/tobacco-use>



Thank you!