Tobacco Consumption with Educational Background as Allocation Base

Linghan Zhou, Liuyu Wang, Baojia Huang

Mar 26th, 2021

Link to presentation: here

Link to code: here

Introduction

With the broader applications of data science in the business area, business is not pure business anymore. Business analytics nowadays always requires the combination of data analytics and strategic decision-making. Our topic will focus on tobacco consumption with the education background of the consumers as the allocation base. Based on the data collected, a fair recommendation can be reached for the client, US Tobacco Associate.

Basically, the five steps for our analysis are listed below:

- Define the business needs:

  The client seeks to get the improvement in budget revenue planning, allocation of raw materials and resources, and making specific marketing strategies to target customers from different education backgrounds in next year.
- Selection of the data:

  In order to obtain an exacter conclusion for the analysis, one more precise record of tobacco products and smoking supplies by education is needed. Federal Reserve Economic Data supplies the database in demand for the analysis.
- Exploration of the data:

  Exploring the data includes clearing the data, filling the vacancy, and figuring out the inner relationship of each dataset.
- Prediction by modeling:

  Applying the appropriate models to detect the relationship between the datasets, and in this procedure, machine learning methods like regression helps.
- Making a decision:

  Conclusions can be drawn based on visualization and model, and the client can take action by following the recommendations.

Data Ingestion

Data Acquisition

As mentioned before, all the data needed is from the FRED. The first step to the analysis is combining and cleaning up the data to meet the requirements for data analysis. Raw data acquired from the website are listed split in the group of educational background. To maintain the same time series of data on tobacco consumption across educational backgrounds, we select the record from 1996 to 2019, and set the frequency as an annual basis.

Parsing the data

When combining the datasets, each group of the data are separately given the labels of “High School Graduate”, “Associate Degree”, “Bachelor’s Degree”, and “Master’s, Professional, Doctoral Degree” in the csv file. Additionally, the price index of tobacco and individual expenditure of tobacco are also prepared for the forecast of next year’s price and consumption.

Exploratory Data Analysis

Expenditures: Tobacco Products and Smoking Supplies by Education

It can be seen from the bar chart (see Fig.1) that in the 23 years from 1996 to 2019, the tobacco consumed by people with high school education takes the largest proportion of total tobacco consumption, followed by the amount of tobacco consumed by groups of people with associate degree, bachelor's degree, and master's, professional, doctoral degree.

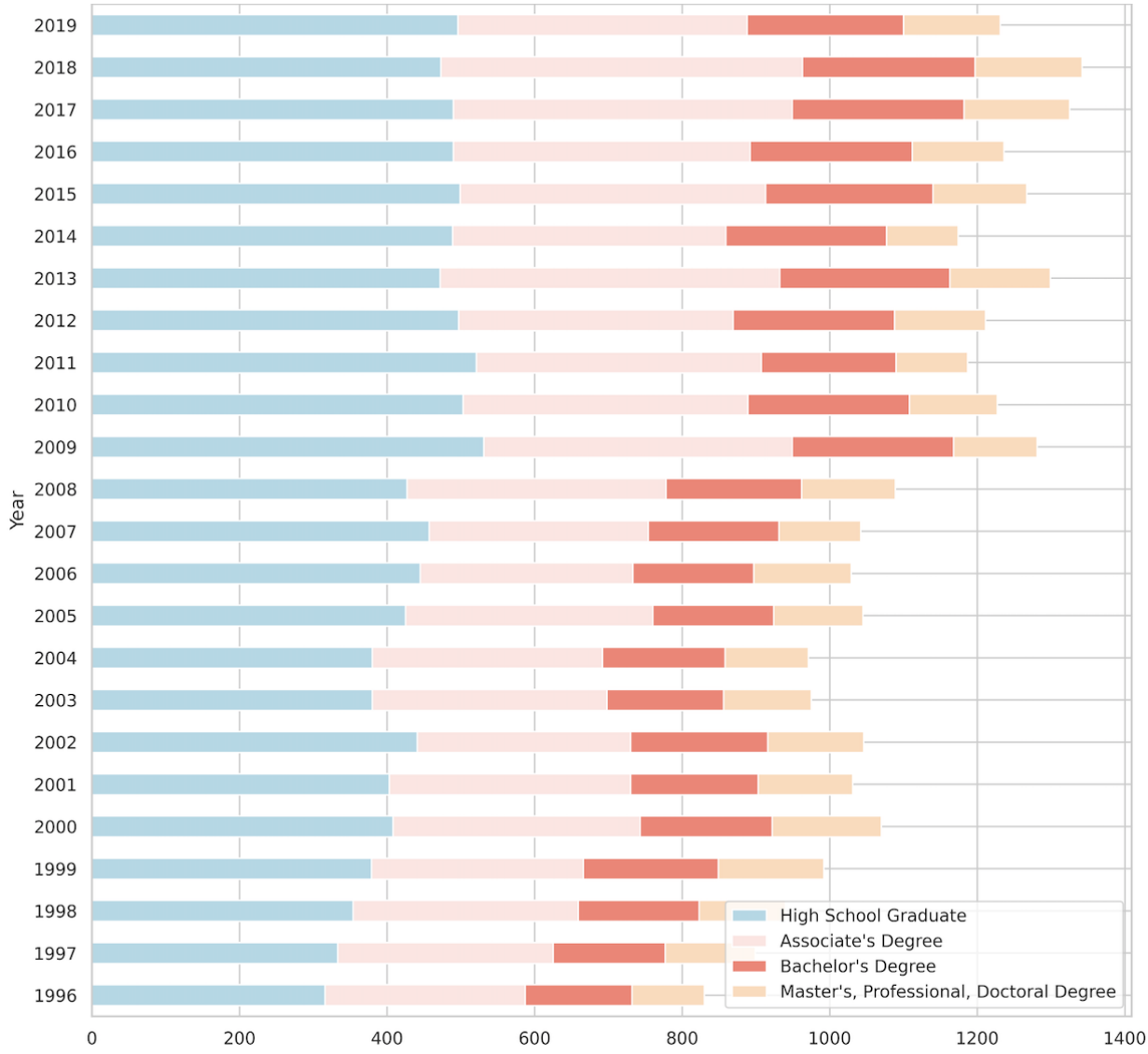

Fig.1 Tobacco products and smoking supplies by education

Same results can be drawn from the distribution and box graphs of these four groups of data. The amount of tobacco products consumed by high school graduates obtains the largest mean of 450. According to the data of people with associate's degree who consume tobacco, it has the mean around 345. While the means of data of consumption record for groups of people with bachelor's degree and master's, professional, doctoral degree both do not exceed 200, and are separately near 185 and 125.

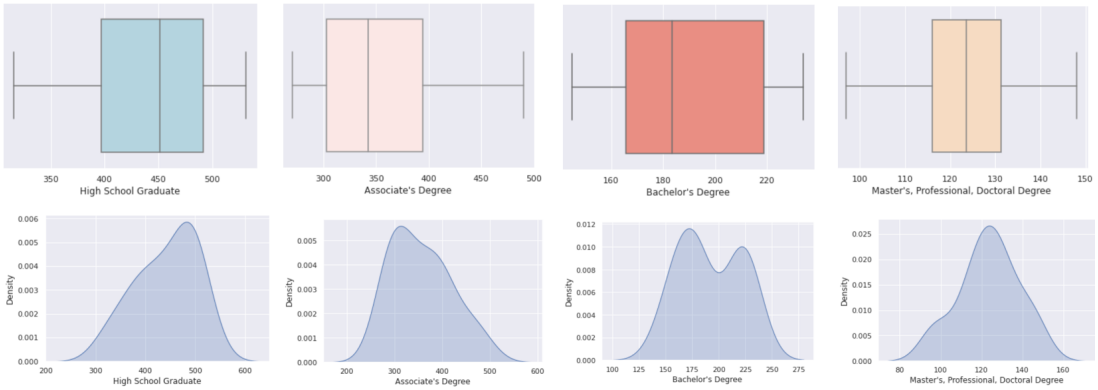

Fig.2 Boxplot and kdeplot

Consumer Price Index

The time series plot of the consumer price index of Tobacco and Smoking Products in the U.S. City Average reflects that the general trend for the consumer price index is upwards. From 1988 to 2008, the growth rate of the consumer price index slightly increased, but a sharp increase followed, and after the sharp rise, the growth rate from 2009 to nowadays remains stable. The past data recorded indicates that time series prediction can be applied to estimate the next year's consumer price index.

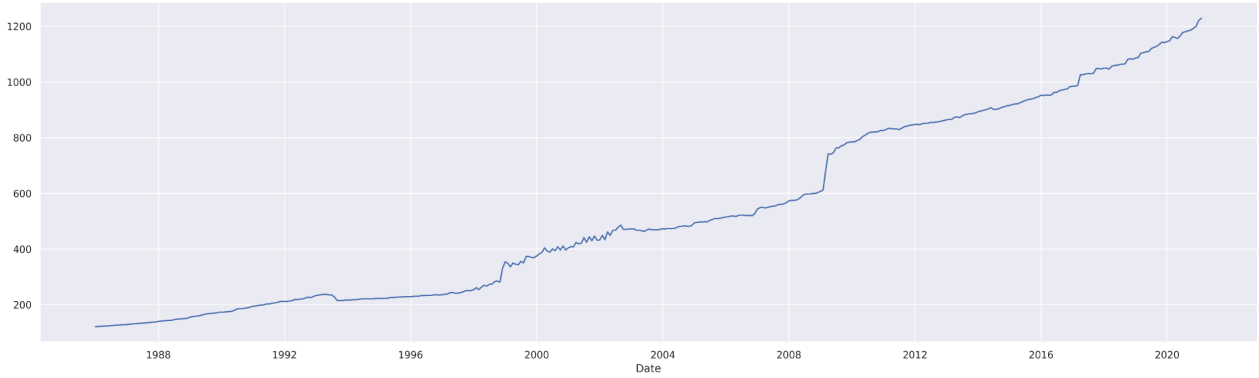

Fig.3 Time series plot of the consumer price index of tobacco and smoking products

Producer Price Index

The producer price index graph reflects the change in cost of production from 1960 to 2020. It can be seen that the producer price basically presents a stable rising trend every year. Thus, time series prediction method will also be applied to estimate the cost of production for the next year, and with this estimation, our client can have better budget cost planning.

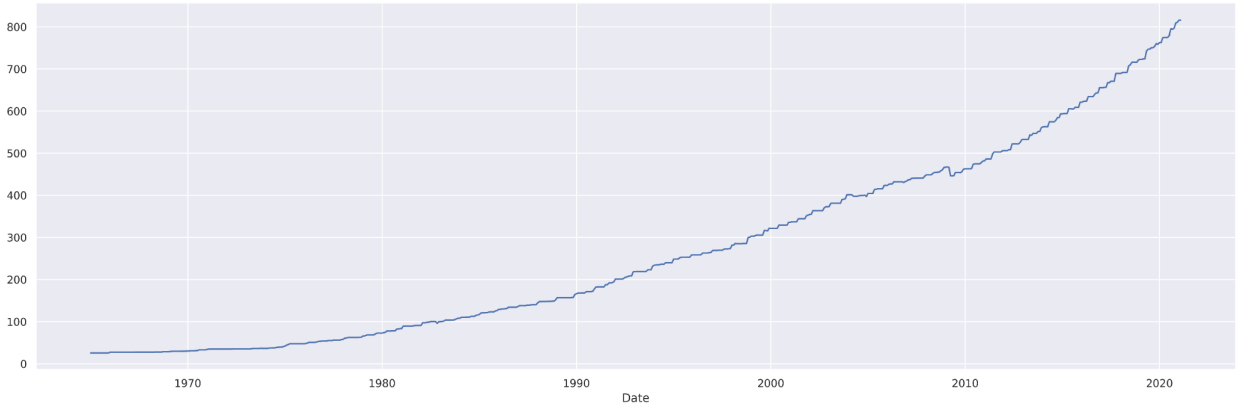

Fig.4 Time series plot of the producer price index of tobacco and smoking products

Personal Expenditures of tobacco

The following time series plot of the personal consumption expenditures of tobacco and smoking products reflects the tobacco consumption for all US citizens. The amount of consumption remains a stable growth between 1930 to 1990, and after that, the growth tends to have greater volatility. Clients can make better decisions in next year's consumption with the prediction of next year's personal expenditures.

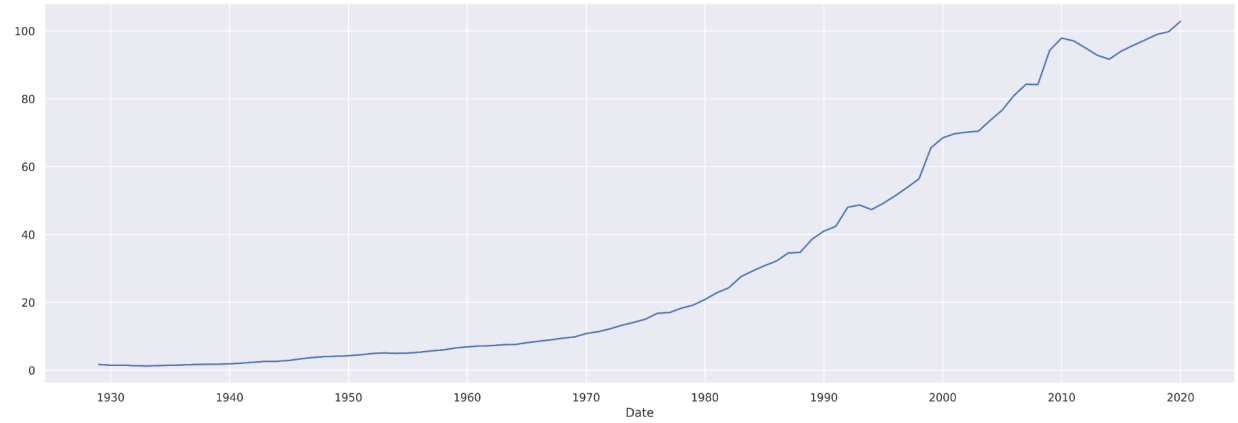

Fig.5 Time series plot of the personal consumption expenditures of tobacco and smoking products

Modeling

Tobacco Expenditures by Different Education Backgrounds

We fit four specific linear graphs that best match the four data groups, and it can be seen that the actual data points are distributed on both sides of these four lines. Similar outcomes can be drawn from the following graph by the intersections of the vertical axis of tobacco products and smoking supplies the proportion of people with different educational backgrounds who consume tobacco. While the dot plot cannot indirectly present the growth rate of tobacco consumption, the slope of the linear fit chart reflects the relationship between educational background and the growth rate of tobacco consumption: the lower the educational background, the faster the growth rate of tobacco consumption is.

Fig.6 Scatterplot of tobacco expenditures by different education backgrounds

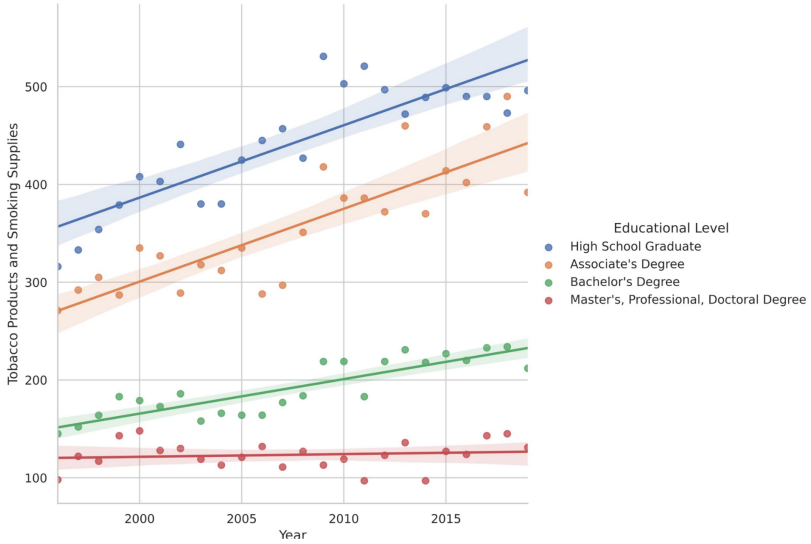

Fig.7 Regression analysis of tobacco expenditures by different education backgrounds

Consumer Price Index Prediction for All Urban Consumers

Through the time series forecasting method using SPSS, we get the next year's forecast data of the consumer price index for all urban consumers. Red line is the actual observed part, and the blue line is the estimation part. The amount of estimated consumer price index for next year is between 1250-1300, with which data we can combine the educational background of the consumers to predict the consumption of each group in the next year.

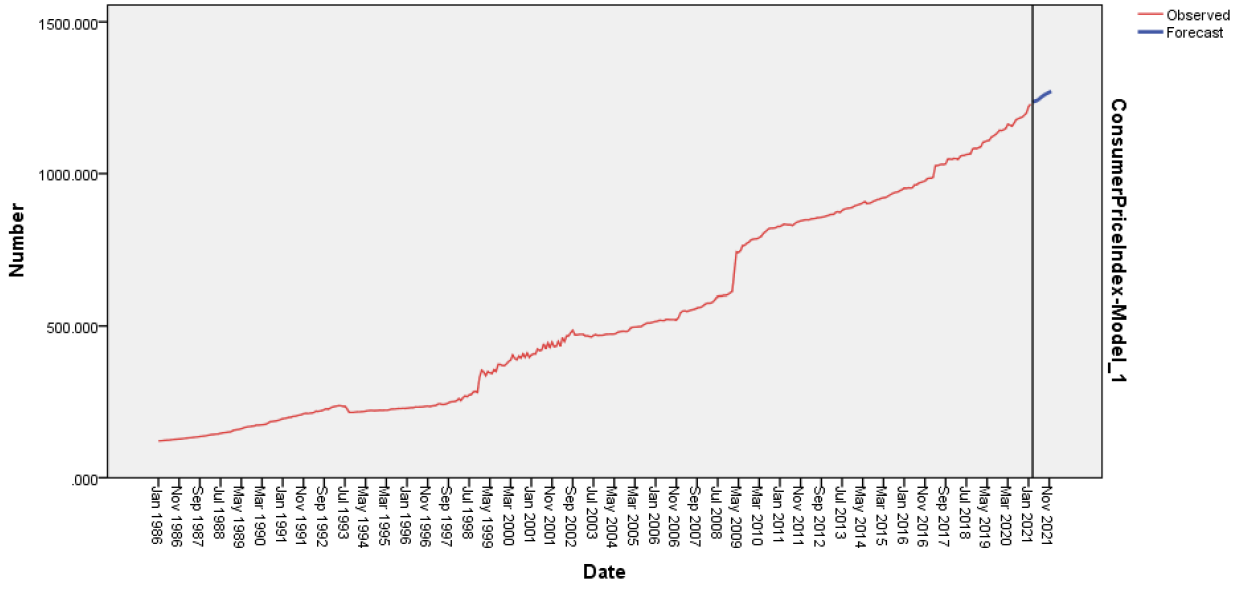

Fig.8 Time series prediction of **consumer price index**

Producer Price Index Prediction

Through the time series forecasting method using SPSS, we get the next year's forecast data of the producer price index. Red line is the actual observed part, and the blue line is the estimation part. The amount of estimated producer price index for next year is between 800 to 830, and we can calculate the estimated the budget cost of production with the estimated cost.

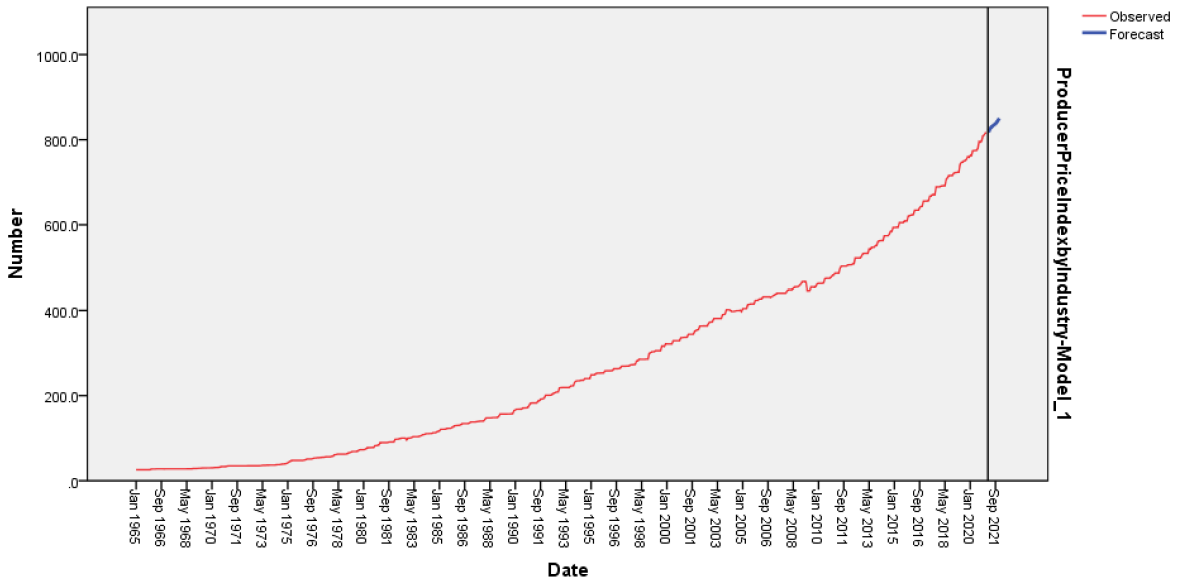

Fig.9 Time series prediction of **producer price index**

Personal Consumption Expenditures Prediction

Through the time series forecasting method using SPSS, we get the forecast data of the personal consumption expenditures for all urban consumers from 2021 to 2025. Red line is the actual observed part, and the blue line is the estimation part. The amount of estimated personal consumption expenditure for next year is between 100 to 110, and we can calculate the estimated gross margin with the estimated price.

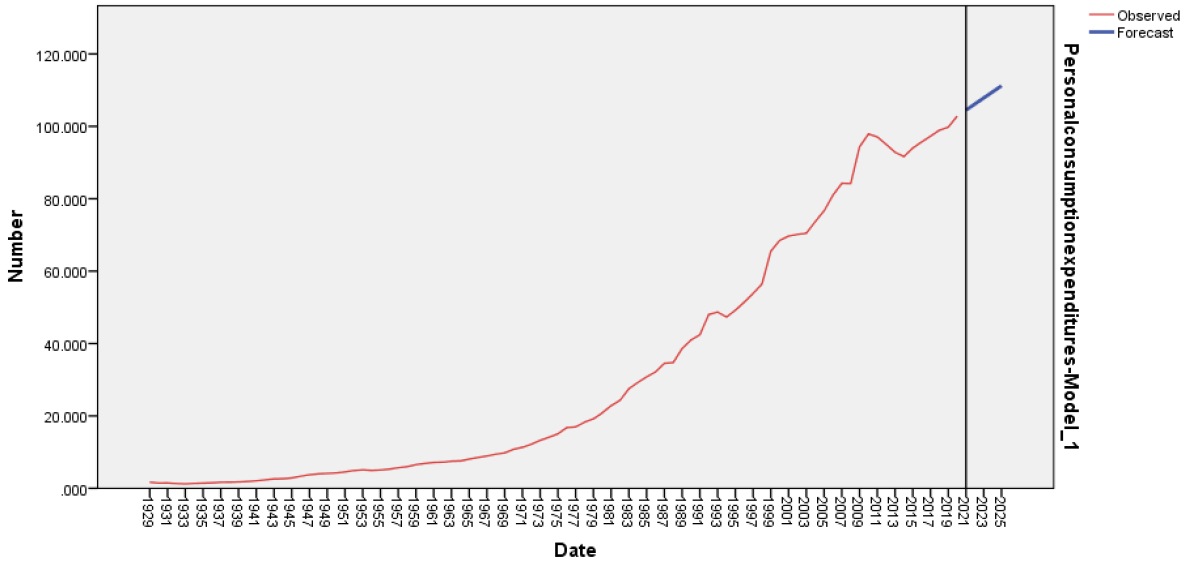

Fig.10 Time series prediction of **personal consumption expenditure**

Conclusion and Recommendation

Conclusion

- In terms of educational background, people with lower education levels consume more tobacco.
- Next year's forecasted consumption price index will be between 1250 to 1300.
- Next years' forecasted producer price index will be between 800 to 830.
- Next year's forecasted estimated personal consumption expenditure will be between 100 to 110.

Recommendation

- The US Tobacco Association should target its customers with lower education levels and target its advertisements precisely.
- The US Tobacco Association can make the budget revenue of next year with the forecast of tobacco price index and personal consumption expenditure, which is between 125,000 to 143,000 next year.
- The US Tobacco Association can make the budget cost of next year with the forecast of producer price index and personal consumption expenditure, which is between 80,000 to 91,300 next year.
- The US Tobacco Associate can make the budget gross margin of the amount between 45,000 to 77,000.