Company Data Decision Tree

# Assignment

About the data:

Let’s consider a Company dataset with around 10 variables and 400 records.

The attributes are as follows:

 Sales -- Unit sales (in thousands) at each location

 Competitor Price -- Price charged by competitor at each location

 Income -- Community income level (in thousands of dollars)

 Advertising -- Local advertising budget for company at each location (in thousands of dollars)

 Population -- Population size in region (in thousands)

 Price -- Price company charges for car seats at each site

 Shelf Location at stores -- A factor with levels Bad, Good and Medium indicating the quality of the shelving location for the car seats at each site

 Age -- Average age of the local population

 Education -- Education level at each location

 Urban -- A factor with levels No and Yes to indicate whether the store is in an urban or rural location

 US -- A factor with levels No and Yes to indicate whether the store is in the US or not

The company dataset looks like this:



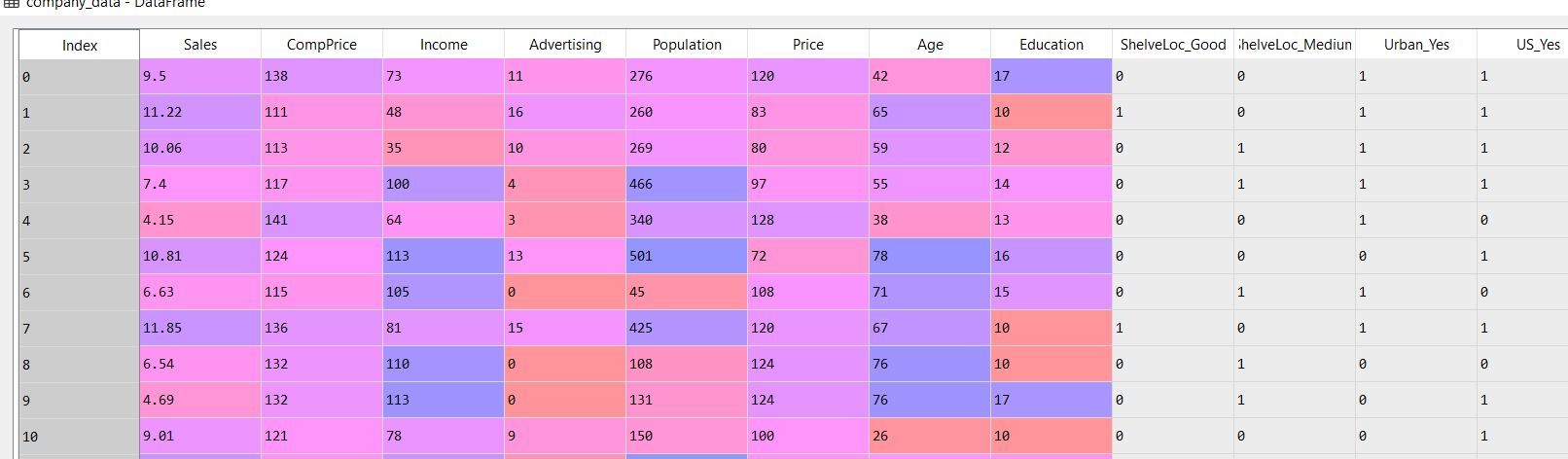
Problem Statement:

A cloth manufacturing company is interested to know about the segment or attributes causes high sale.

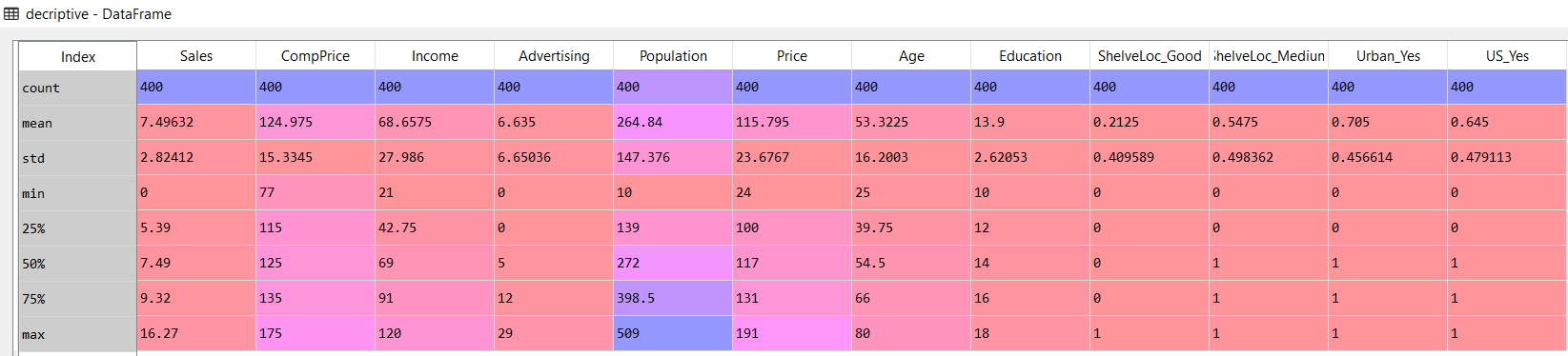
Approach - A decision tree can be built with target variable Sale (we will first convert it in categorical variable) & all other variable will be independent in the analysis.

In the above data ShelveLoc, Urban and US are categorical variables and the rest are continuous data.

After converting the categorical data the following is the result:



The following are the descriptive statics of the above data:



Looking at the quartiles, we convert the sales data in to categorical data by creating bins as Low, Medium and High.

-For Sales between 0 and 5 is considered to be Low

-For Sales between 5 and 9 is considered to be Medium

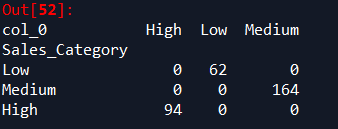
-For Sales between 9 and 17 is considered to be High

After converting the continuous Sales data to Categorical, we set this Sales variables as the target whereas the variables as the predictors for the decision tree.

Predictors:

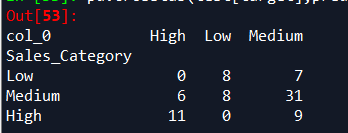


After executing the Decision tree algorithm on the training data, we get the following results:



The accuracy for the above training data is 100%

Now we execute for the training data and the results are as follows:



Accuracy for the training data is 0.625