Decision Tree & Random Forest

About the data:

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

The attributes are as follows:

 Sales

 Competitor Price

 Income

 Advertising

 Population

 Price

 Shelf Location at stores

 Age

 Education

 Urban

 US

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 –

Step 1: Understand the data by doing the EDA with different visualizations to understand the patterns. Pre process the data like Handling outliers & Null values, applying one hot encoding or label encoding etc.

Step 2: Split the data to train and test.

Step 3: Build a basic model using the Decision tree.

Step 4: Do hyperparameter tuning on the model.

Step 5: Apply Random Forest, Bagging and boosting methods.

Ste6 6: Create a chart which shows the accuracy of all the models you have build

Step 7: Write conclusions.