I imported the data from local file. Then I used pre-processing work which contain finding missing values, and I dropped the columns which doesn’t contains more than 80% of data. And then I fill the missing values which contains numerical columns with mean and the column which contains categorical with random categorical values which are present in that columns. In this data there is a customer email column and addresses columns which contains hidden pattern which is like mail id, city codes so I split the columns using the split operation to get the mail id & city codes. Finally I converted all columns in to zero null values (missing values).

I split the data into attributes & response where ‘x’ contains attributes and ‘y’ contains response. I used decision tree algorithm to built a model where I can fit the x & y

I check the feature importance using correlation, in correlation each attribute acquires an importance and in summation of each attributes where I get 100% I consider them as important attributes and remaining are dropped which has zero importance.

I did normalization to scale the values by using minmaxscalar. In order to scale the values in equal format

Using decision tree algorithm I fond the accuracy of the model is 75%