? Build a predictive model and predict the sales of each product

About Dataset

Sales Prediction for Mart Outlets

The data scientists at Mart have collected 2013 sales data for 1559 products across 10 stores in different cities. Also, certain attributes of each product and store have been defined. The aim is to build a predictive model and predict the sales of each product at a particular outlet.

Using this model, Mart will try to understand the properties of products and outlets which play a key role in increasing sales.

Note that the data may have missing values as some stores might not report all the data due to technical glitches. Hence, it will be required to treat them accordingly.

Data Dictionary: have a train (8523) and test (5681) data set, the train data set has both input and output variable(s). You need to predict the sales for the test data set.

Train file:

CSV containing the item outlet information with a sales value

Variable Description

Item_Identifier :Unique product ID
Item_Weight : Weight of product
Item_Fat_Content : Whether the product is low fat or not
Item_Visibility : The % of the total display area of all products in a store allocated to the particular product
Item_Type : The category to which the product belongs
Item_MRP : Maximum Retail Price (list price) of the product
Outlet_Identifier : Unique store ID
Outlet_Establishment_Year : The year in which the store was established
Outlet_Size : The size of the store in terms of ground area covered
Outlet_Location_Type : The type of city in which the store is located

*Outlet_Type : Whether the outlet is just a grocery store or some sort of supermarket

Item_Outlet_Sales : sales of the product in t particular store. This is the outcome variable to be predicted.

CSV containing item outlet combinations for which sales need to be forecasted
Variable Description

Item_Identifier ----- Unique product ID

Item_Weight ---- Weight of product

Item_Fat_Content ----- Whether the product is low fat or not

Item_Visibility ---- The % of the total display area of all products in a store allocated to the particular product

Item_Type ---- The category to which the product belongs

Item_MRP ----- Maximum Retail Price (list price) of the product

Outlet_Identifier ----- Unique store ID

Outlet_Establishment_Year ----- The year in which store store was established

Outlet_Size ----- The size of the store in terms of ground area covered

Outlet_Location_Type ---- The type of city in which the store is located

Outlet_Type ---- whether the outlet is just a grocery store or some sort of supermarket

Variable Description

Item_Identifier ----- Unique product ID
Outlet_Identifier ----- Unique store ID
Item_Outlet_Sales ----- Sales of the product in t particular store. This is the outcome variable to be predicted.

Evaluation Metric

Your model performance will be evaluated on the basis of your prediction of the sales for the test data (test.csv), which contains similar data-points as train except for the sales to be predicted. Your submission needs to be in the format as shown in the same sample submission.

We at our end, have the actual sales for the test dataset, against which your predictions will be evaluated. We will use the Root Mean Square Error value to judge your response.