**Big\_Mart Sales prediction**

**Problem statement:** The data scientists at BigMart 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 was to build a predictive model and predict the sales of each product at a particular outlet.

**Solution/Approach:**

There can be multiple possibilities for the sales:

1.On basis of item:

a.Item visibility in store: The location of product in a store will impact sales. Ones which are right at entrance will catch the eye of customer first rather than the ones in back.

b.Product Frequency: More frequent products will have high Sales.

2.On basis of store:

a.City type: Stores located in Tier1 cities should have higher sales because of the higher income levels of people there.

b.Store capacity: Stores which are very big in size should have higher sales as they act like one-stop-shops for people[**¶**](http://localhost:8888/notebooks/ABB_case_study/Market%20prediction.ipynb#b.Store-capacity:-Stores-which-are-very-big-in-size-should-have-higher-sales-as-they-act-like-one-stop-shops-for-people)

EDA:

Initial approach was to first improve the data quality by working on the null values and improving data consistency.

For null value treatment we tried various Univariate and Multivariate approaches based upon analysis we adjusted the nulls.

We checked if feature value complexity could be improved or not once adjusted we did the categorical values transformation for machine use.

Modelling:

This is a classical continuous value prediction problem and we have tested various regressor model XGboost, Linear Regression and various regularization techniques   
The XGBRFregressor stood out to be the champion model as it produced the highest r2 score.