Session Problem Statement - BigMart Sales Prediction

**Context**

In today’s world, huge shopping centers such as big malls and marts are recording data related to sales of items or products as an important step to predict the sales and get an idea about future demand that can help with inventory management. Understanding what role certain properties of an item play and how they affect their sales is imperative to any retail business.  
  
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. Using this data, BigMart is trying to understand the properties of products and stores which play a key role in increasing sales.

**Objective**

To build a predictive model that can find out the sales of each product at a particular store and then provide actionable recommendations to the BigMart sales team to understand the properties of products and stores which play a key role in increasing sales.

**Data Dictionary**

* Item\_Identifier - Unique product ID
* Item\_Weight - Weight of the 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 the particular store. This is the outcome variable to be predicted.

We have two datasets - train (8523) and test (5681) data. The training dataset has both input and output variable(s). You will need to predict the sales for the test dataset.