# Which Data?

The dataset contains the sales data of the year 2013 for 1559 products across 10 stores in different cities. The data has been downloaded from:

<https://code.datasciencedojo.com/tshrivas/dojoHub/tree/a152a17dee24dcfcc10bb75c77c2e88cdcf90212/Big%20Mart%20Sales%20DataSet>

Codebook:

| **Column  Position** | **Atrribute  Name** | **Definition** | **Data  Type** | **Example** | **% Null  Ratios** |
| --- | --- | --- | --- | --- | --- |
| 1 | Item\_Identifier | It is a unique product ID assigned to every distinct item. It consists of an alphanumeric string of length 5 | Alphanumeric | FDN15 | 0 |
| 2 | Item\_Weight | This field includes the wieght of the product | Numeric (float) | 17.5 | 17.16531738 |
| 3 | Item\_Fat\_Content | This attribute is categorical and describes whether the product is low fat or not. There are 2 categories of this attribute: ['Low Fat', 'Regular']. However, it is important to note that 'Low Fat' has also been written as 'low fat' and 'LF' in dataset, whereas, 'Regular' has been referred as 'reg' as well | Alpha | Low Fat | 0 |
| 4 | Item\_Visibility | This field mentions the percentage of total display area of all products in a store allocated to the particular product | Numeric (float) | 0.01676 | 0 |
| 5 | Item\_Type | This is a categorical attribute and describes the food category to which the item belongs. There are 16 different categories listed as follows: ['Dairy', 'Soft Drinks', 'Meat', 'Fruits and Vegetables', 'Household', 'Baking Goods', 'Snack Foods', 'Frozen Foods', 'Breakfast', 'Health and Hygiene', 'Hard Drinks', 'Canned', 'Breads', 'Starchy Foods', 'Others', 'Seafood'] | Alpha | Meat | 0 |
| 6 | Item\_MRP | This is the Maximum Retail Price (list price) of the product | Numeric (float) | 141.618 | 0 |
| 7 | Outlet\_Identifier | It is a unique store ID assigned. It consists of an alphanumeric string of length 6 | Alphanumeric | OUT049 | 0 |
| 8 | Outlet\_Establishment\_Year | This attribute mentions the year in which store was established | Numeric (Integer) | 1998 | 0 |
| 9 | Outlet\_Size | The attribute tells the size of the store in terms of ground area covered. It is a categorical value and described in 3 categories: ['High', 'Medium', 'Small'] | Alpha | Medium | 28.27642849 |
| 10 | Outlet\_Location\_Type | This field has categorical data and tells about the size of the city in which the store is located through 3 categories: ['Tier 1', 'Tier 2', 'Tier 3'] | Alpha | Tier 3 | 0 |
| 11 | Outlet\_Type | This field contains categorical value and tells whether the outlet is just a grocery store or some sort of supermarket. Following are the 4 categories in which the data is divided: ['Supermarket Type1', 'Supermarket Type2', 'Grocery Store','Supermarket Type3'] | Alpha | Supermarket Type2 | 0 |
| 12 | Item\_Outlet\_Sales | This is the outcome variable to be predicted. It contains the sales of the product in the particulat store | Numeric (float) | 2097.27 | 0 |