

Data Description - Test and Train Data Set

1. Purpose of Data set – Record the information of the available products and store properties.
2. Source of Data set – The data set is available on Analytics Vidhya Hackathon, downloaded from https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/?utm_source=auto-email
3. Time window – Year 2013
4. Cost of Data – Downloadable for free.
5. Collection technique – Update the database as the new products gets available on the store and even update the database as soon as any attribute of the product gets change.
6. Collection tools – Big Mart database and buyer's transactional history.
7. Quality – The data is genuine since it has been provided by the internal source and thus there is a transparency in the data been available.
8. Completeness – The data isn't complete, there are some missing fields in the Test data set.

Variables that are used in the Big Mart Data set

| Name | Definition | Variable Classification |
|---------------------------|--|-------------------------|
| Item_Identifier | Unique product ID | Character |
| Item_weight | Weight of product | Numeric |
| Item_Fat_content | Whether product is low fat or not | Character |
| Item_visibility | The percentage of total display area of all products in a store allocated to the particular product. | Numeric |
| Item_Type | The category to which the product belongs. | Character |
| Item_MRP | Maximum retail price of the product. | Numeric |
| Outlet_Identifier | Unique store ID | Character |
| Outlet_Establishment_Year | The year in which store was established | Character |

| | | |
|----------------------|---|-----------|
| Outlet_Size | The size of the store in terms of the ground area covered. | Character |
| Outlet_Location_Type | The type of the city in which the store is located. | Character |
| Outlet_Type | Whether the outlet is a grocery store or a supermarket | Character |
| Item_Outlet_Sales | Sales of the product in the particular store. This is the outcome variable to be predicted. | Numeric |