Given Data:

1. 2013 sales
2. 1559 products- each product’s attributes are given
3. 10 stores in different cities

AIM: Predict sales of each product at a particular store.

Purpose: Using this prediction, we increase the sales of each product by understanding the properties of product and store which impact the product’s sales.

Hypothesis Generation:

Product Types: i) liquid

* Dairy
* chemicals
* Cool drinks
* Pharmacy (syrups)
* Deodorants
* Body care

ii) solid

* Food
* Electrical Instruments
* Kitchen ware
* Sports
* Clothes
* Stationary
* Body care

iii) gas

* Chemicals

Store Types: i) small store

ii) big store

iii) very big store

iv) gas station

Store location Types: i) urban

ii) rural

Inferences:

Location impacts the product sales:

* urban areas have more number people, therefore more scales regardless of the size of the store.
* In gas station cool drinks and chips will be on hot sales.

Product type also impacts the sales:

* Daily households include yogurt, milk, and other food (meat, vegetables and fruits). Trash bags, kitchen and bathroom requirements.
* Products which cause less impact on the sales: electronics, clothes, etc.

Data Exploration:

Observations:

1. # of features= 11
2. # of features that can affect the sales= 9
3. # of useless features= 2 (Item Identifier and outlet identifier)
4. Output variable= 1 (Item Outlet Sales)
5. # of categorical variables= 5
6. Total # of datapoints= 14204
7. # of datapoints in training set= 8523
8. # of datapoints in test set= 5681