**PROJECT**

**ON**

**BIG MART SALES PREDICTION**

**STUDY OF BIG MART SALES PREDICTION**

**About project -:**

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 of this data science project is to build a predictive model and find out the sales of each product at a particular store.

Using this model, BigMart will try to understand the properties of products and stores which play a key role in increasing sales.

**All the Lifecycle In A Data Science Projects -:**

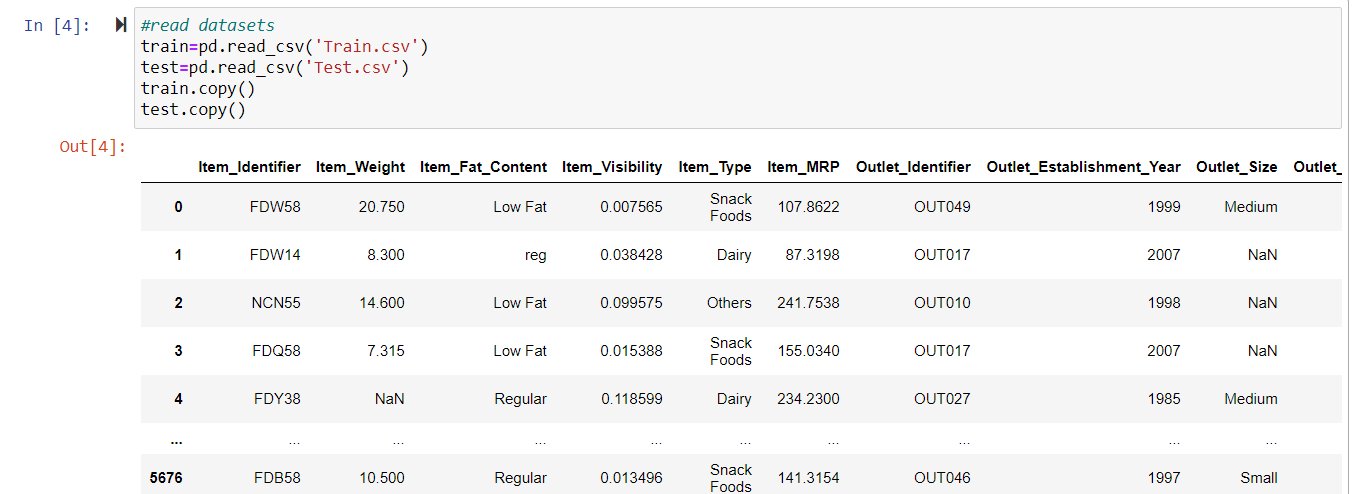
* Data Analysis
* Feature Engineering
* Feature Selection
* Model Building
* Model Deployment

**Tasks -:**

* Problem Statement
* Hypothesis Generation
* Loading Packages and Data
* Data Structure and Content
* Exploratory Data Analysis
* Univariate Analysis
* Bivariate Analysis
* Missing Value Treatment
* Feature Engineering
* Encoding Categorical Variables
* Label Encoding
* One Hot Encoding
* PreProcessing Data
* Modeling
* Linear Regression
* Regularized Linear Regression
* RandomForest
* XGBoost
* Summary

**Data Analysis Phase -:**

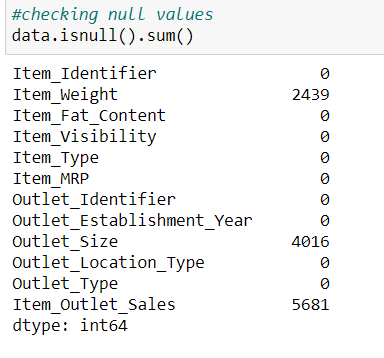
* + The main aim is to understand about the data and display all the columns of a dataframe



#### **In Data Analysis We will Analyze To Find out the below** **stuff-:**

* Missing Values
* All The Numerical Variables
* Distribution of the Numerical Variables
* Categorical Variables
* Cardinality of Categorical Variables
* Outliers
* Relationship between independent and dependent feature(SalePrice)

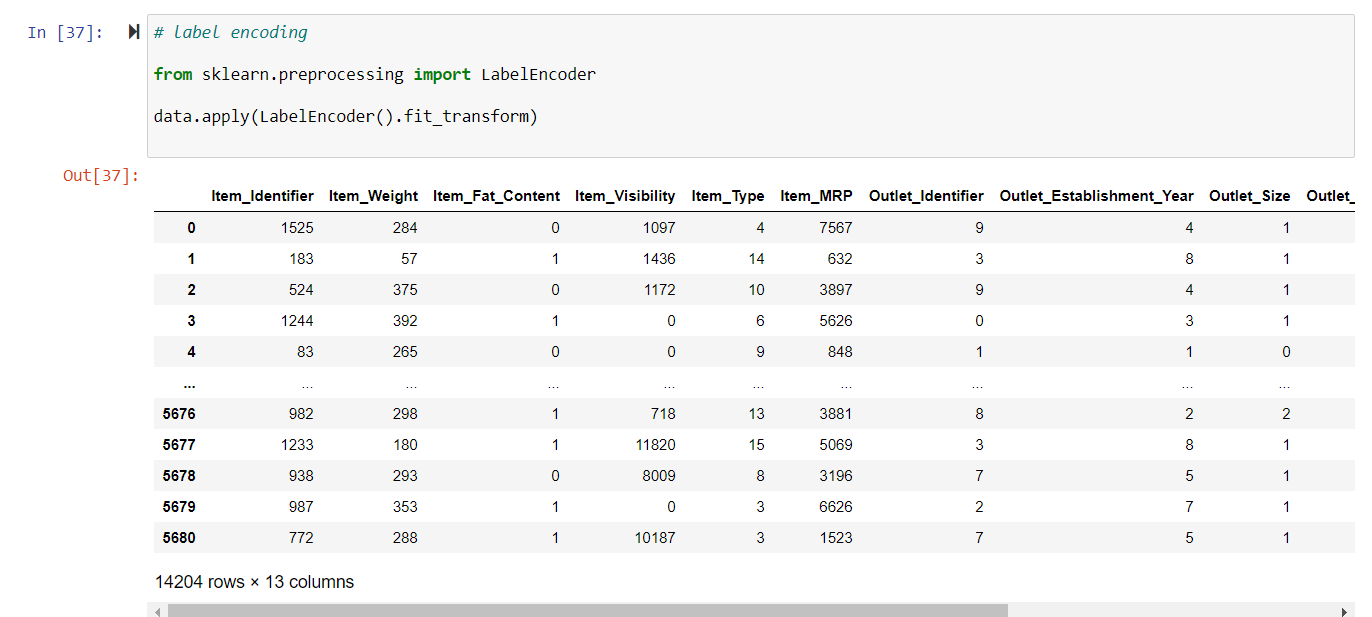
**Missing values -:**



* After finding the missing values we can fill these null values with some methods.

**Converting Categorical variables to numerical variables -:**

* Firstly I am doing label encoding and also making some dummy variables.





**Train Test Split -:**

* Splitting the data into dependent and independent variables.

**Results -:**

* The model is good fit model. Coefficient obtained is greater than 0 and p value is less than 0.05 for the model. So, the model is a good fit.

**Conclusion -:**

* We see that the outlet Sale at Bigmart is affected by factors like Item Visibility, Item MRP, Outlet Location and Establishment year. The analysis on Outlet sale at BigMart is done successfully.

.