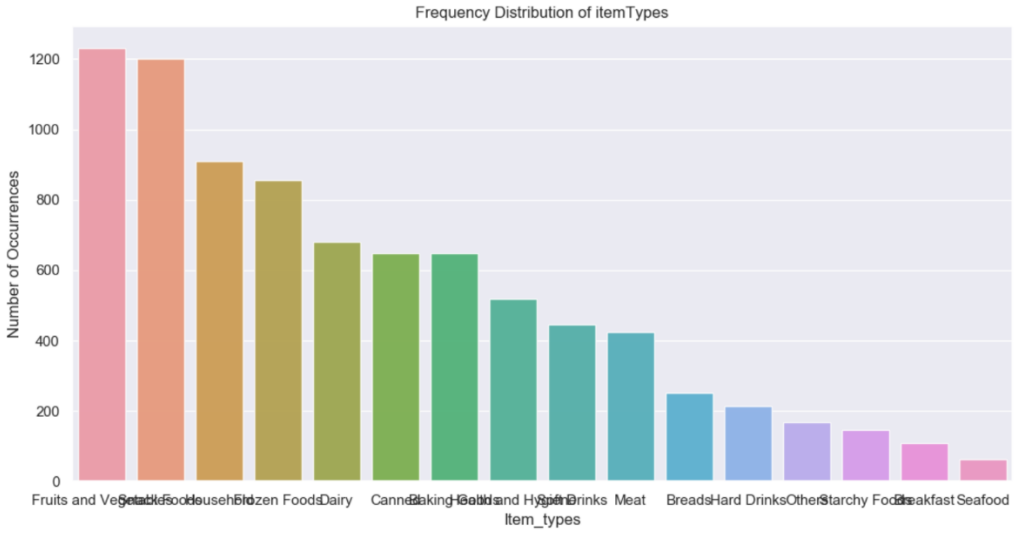
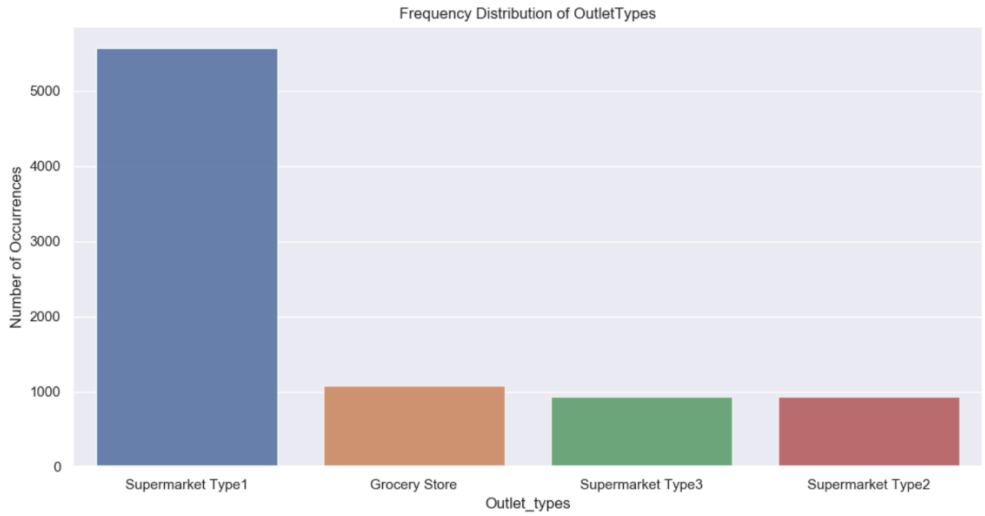
Big Mart Sales Analysis

Problem Statement - 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 is to build a predictive model and find out the sales of each product at a store. Using this model, BigMart will try to understand the properties of products and stores which play a key role in increasing sales.

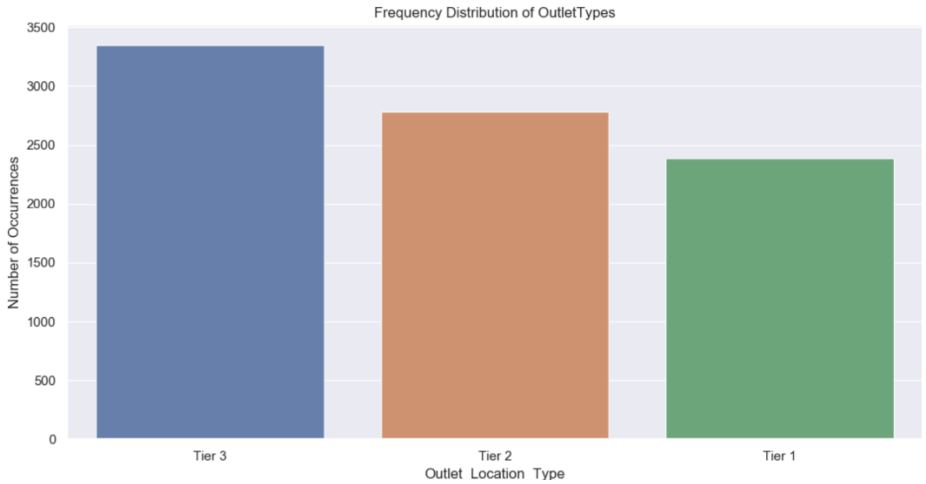
- Performed data wrangling as there were many missing values in the dataset.
- Performed some exploratory analysis on the relevant variables and derived the following conclusions:
  - We see that the maximum frequency is for the item type Fruits and Vegetables.



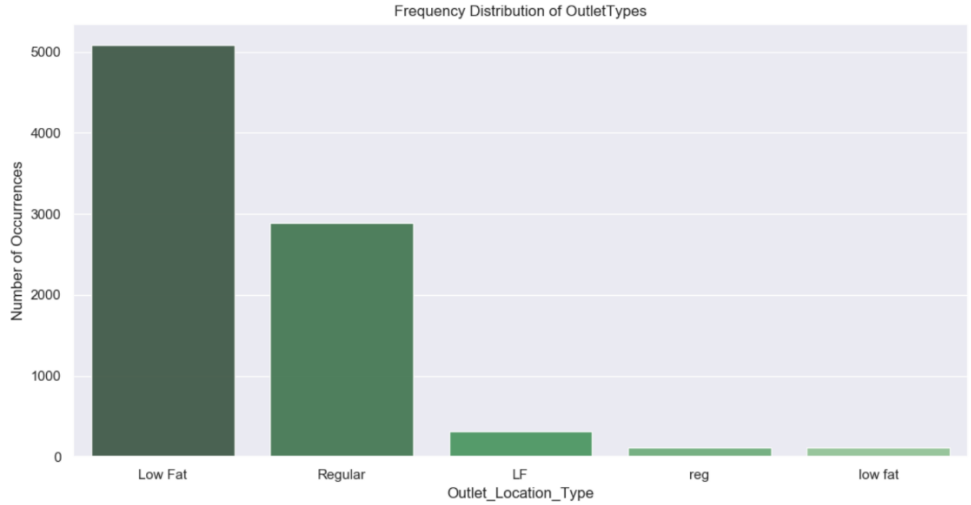
  - We see that the maximum number of transactions were made from the Supermarket Type1 outlet type.



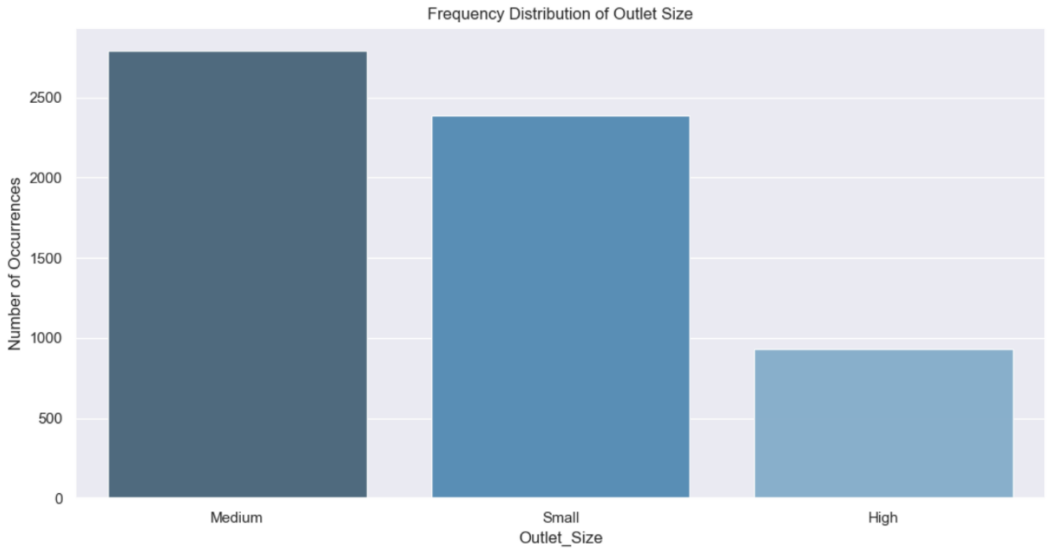
  - We see that the maximum number of transactions were made from the Tier 3outlet location type.



- We see that the maximum frequency of the item sold are the item having low fat content.



- We observe that the maximum number of items are sold from medium sized outlets.



- Performed One hot encoding to convert categorical variables to binary form so that they can be used in the model.
- Performed valid imputations for the missing values.

Modelling:

Below is the summary of the regression results:

Summary: OLS Regression Results

Dep. Variable:	Item_Outlet_Sales	R-squared:	0.563
Model:	OLS	Adj. R-squared:	0.562
Method:	Least Squares	F-statistic:	391.2
Date:	Fri, 14 Dec 2018	Prob (F-statistic):	0.00
Time:	13:07:58	Log-Likelihood:	-71993.
No. Observations:	8523	AIC:	1.440e+05
Df Residuals:	8494	BIC:	1.442e+05
Df Model:	28		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-104.7836	41.638	-2.517	0.012	-186.405	-23.162
Item_Weight	-0.4119	2.916	-0.141	0.888	-6.129	5.305
Item_Fat_Content_Low Fat	-65.1849	34.959	-1.865	0.062	-133.714	3.344
Item_Fat_Content_Regular	-23.0489	36.812	-0.626	0.531	-95.209	49.111
Item_Fat_Content_low fat	-16.5498	78.480	-0.211	0.833	-170.389	137.290
Item_Type_Baking Goods	-11.7519	45.191	-0.260	0.795	-100.337	76.833
Item_Type_Breads	-11.2068	69.079	-0.162	0.871	-146.619	124.205
Item_Type_Breakfast	-2.6871	102.871	-0.026	0.979	-204.339	198.965
Item_Type_Canned	12.5993	45.103	0.279	0.780	-75.813	101.012
Item_Type_Dairy	-52.3792	44.195	-1.185	0.236	-139.011	34.253
Item_Type_Frozen Foods	-37.2525	40.237	-0.926	0.355	-116.127	41.622
Item_Type_Fruits and Vegetables	16.6621	35.041	0.475	0.634	-52.028	85.352
Item_Type_Hard Drinks	-15.0691	75.065	-0.201	0.841	-162.215	132.077
Item_Type_Health and Hygiene	-21.3704	50.471	-0.423	0.672	-120.306	77.565
Item_Type_Household	-51.3307	40.377	-1.271	0.204	-130.480	27.819
Item_Type_Meat	-11.0373	54.687	-0.202	0.840	-118.237	96.163
Item_Type_Others	-35.0888	83.888	-0.418	0.676	-199.531	129.353
Item_Type_Seafood	167.2163	133.534	1.252	0.211	-94.542	428.975
Item_Type_Snack Foods	-23.8752	35.194	-0.678	0.498	-92.863	45.113
Item_Type_Soft Drinks	-38.9222	53.280	-0.731	0.465	-143.365	65.520
Item_Type_Starchy Foods	10.7098	88.922	0.120	0.904	-163.600	185.019
Outlet_Size_High	47.3642	86.453	0.548	0.584	-122.105	216.834
Outlet_Size_Medium	223.6496	69.344	3.225	0.001	87.719	359.580
Outlet_Size_Small	127.0786	45.405	2.799	0.005	38.075	216.083
Outlet_Location_Type_Tier 1	-119.3488	46.005	-2.594	0.009	-209.530	-29.167
Outlet_Location_Type_Tier 2	24.1829	39.207	0.617	0.537	-52.673	101.039
Outlet_Location_Type_Tier 3	-9.6177	48.654	-0.198	0.843	-104.992	85.757
Outlet_Type_Grocery Store	-1632.6965	47.687	-34.238	0.000	-1726.174	-1539.219
Outlet_Type_Supermarket Type1	256.5887	51.002	5.031	0.000	156.612	356.565
Outlet_Type_Supermarket Type2	-227.6473	56.146	-4.055	0.000	-337.707	-117.588
Outlet_Type_Supermarket Type3	1498.9716	56.027	26.754	0.000	1389.144	1608.799
Item_MRP	15.5629	0.198	78.673	0.000	15.175	15.951
Item_Visibility	-300.3685	248.872	-1.207	0.227	-788.218	187.481

Omnibus:	967.222	Durbin-Watson:	2.004
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2318.335
Skew:	0.670	Prob(JB):	0.00
Kurtosis:	5.176	Cond. No.	8.35e+16

However this model can be improved further .