Retail Chain- Stores Sales Analysis

This dataset has information of sales of a retail chain for 1559 products across 10 stores in different cities.

The aim of this project is to analyze sales in these 10 different stores, look at the best and worst performing stores, and build necessary recommendations for the retail chain to improve their sales through two main metrics –

1.Which type of store to open in which type of city? &

2.Which type of outlet returns the best sales?

The data is at two levels: Level 1: Items (Item_ID, Item_MRP, Item_Desc), and Level 2: Outlets (Outlet_ID, Outlet_Type). Hence, we have to use multi-level analysis for this analysis.

EDA and Visualizations

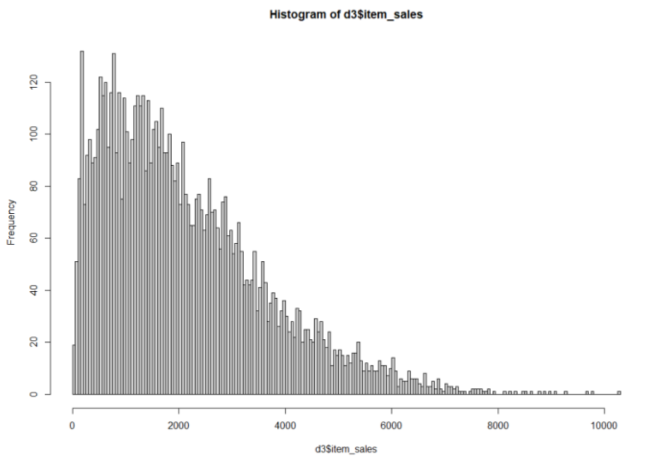


Figure 1.Sales Distribution

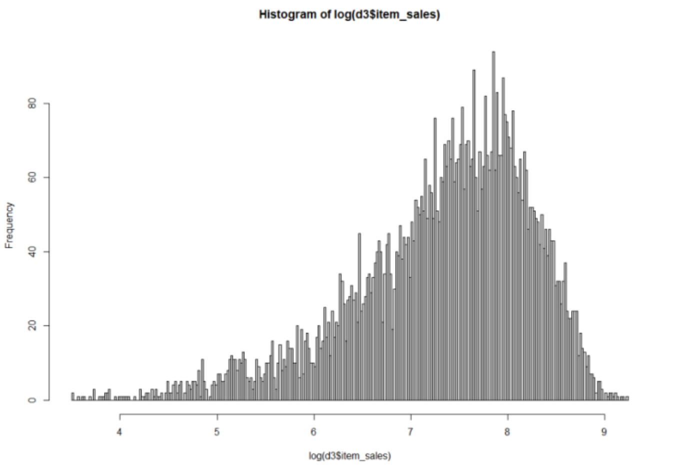


Figure 2.Log Transformation of Sales Distribution

Log transformed variable looks closer to normal distribution although lightly skewed.

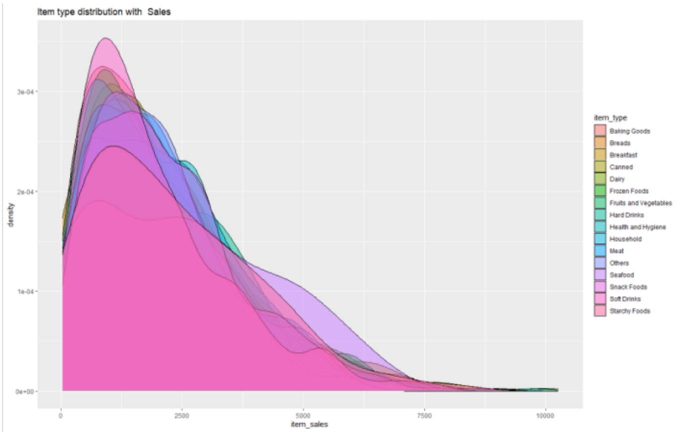


Figure 3.Sales by Item Type

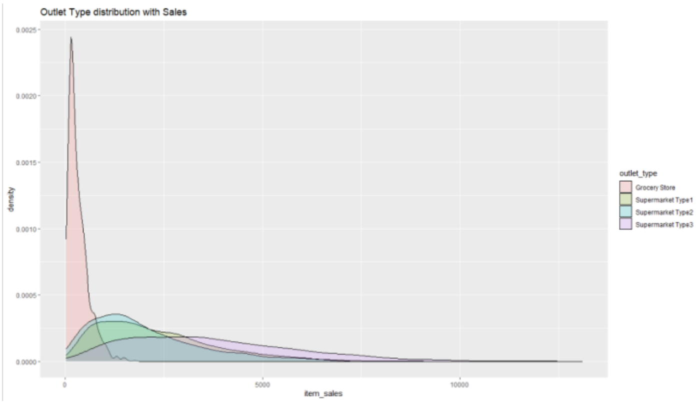


Figure 4.Sales by Outlet Type

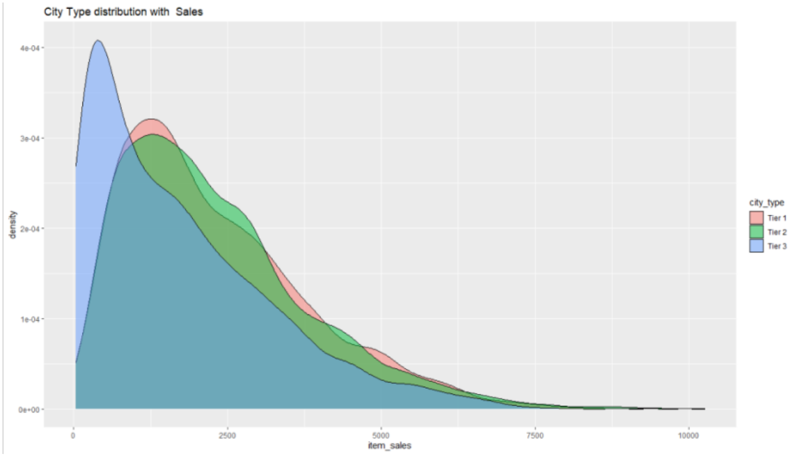


Figure 5.Sales by City Type

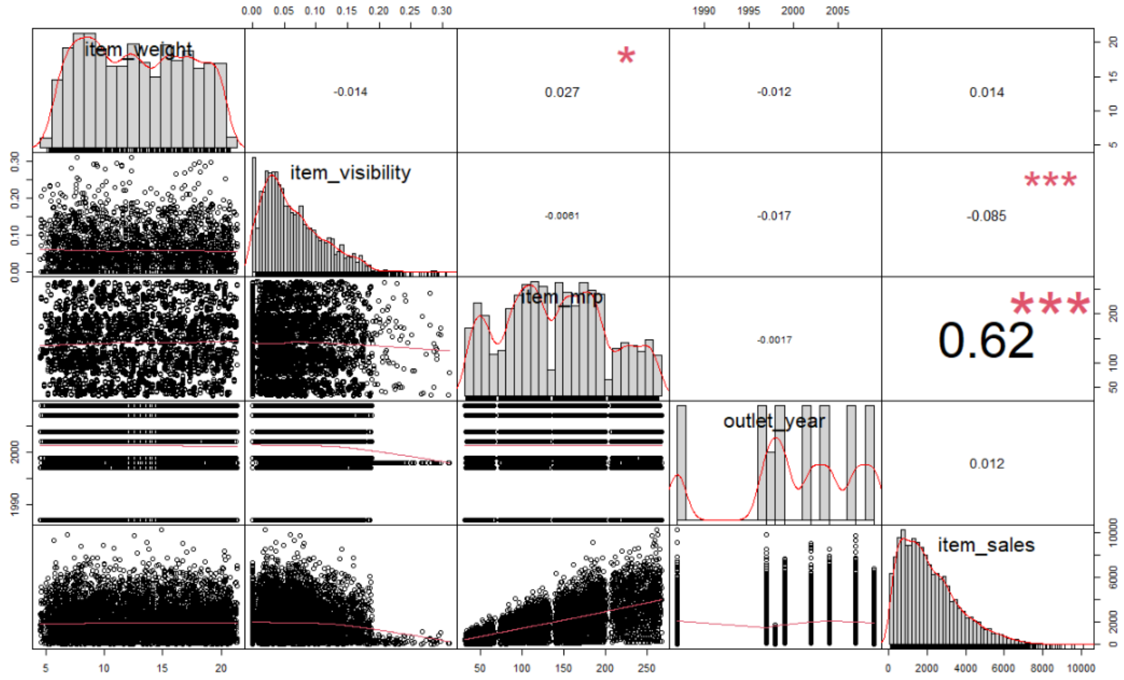
Simple takeaways from EDA and Data Distribution-

- Soft Drinks and Snack Foods had highest sales.
- Tier 3 City Type Stores had higher sales than others.
- Grocery Stores had higher sales than others.

Predictor Table & Reasoning

Y Variable : Item_Sales		
Predictor Variables	**Direction of Effect**	**Rationale**
Item_Weight	+	Products with higher weight will have more selling price which will influence sales.
Item_Fat_Content	"+/-"	Customers prefer low fat food with regards to health choices/ food habits which could influence sales.
Item_Visibility	+	Higher product visibility will attract higher sales.
Item_Type	?	Included to understand which type of items sell best in which stores and which city type.
Item_MRP	"+/-"	Price is generally a crucial factor in a customer's decision whether to purchase a product/sales and will influence sales.
Outlet_ID	"+/-"	Included to understand which stores perform better than others.
Outlet_Year	"+/-"	Older stores have better reputation & attract more customers.
City_Type	"+/-"	Included to understand which city type stores perform better than others.
Outlet_Type	"+/-"	Included to understand which outlet type stores perform better than others.
Excluded Variables:		
Item_ID	None	Unique product ID, not useful for analysis.
Outlet_Size	None	Size of outlet depends on the Unique Store ID which we already have, so this variable can be excluded.

Performance Analytics Chart



MODELS & OUTPUT

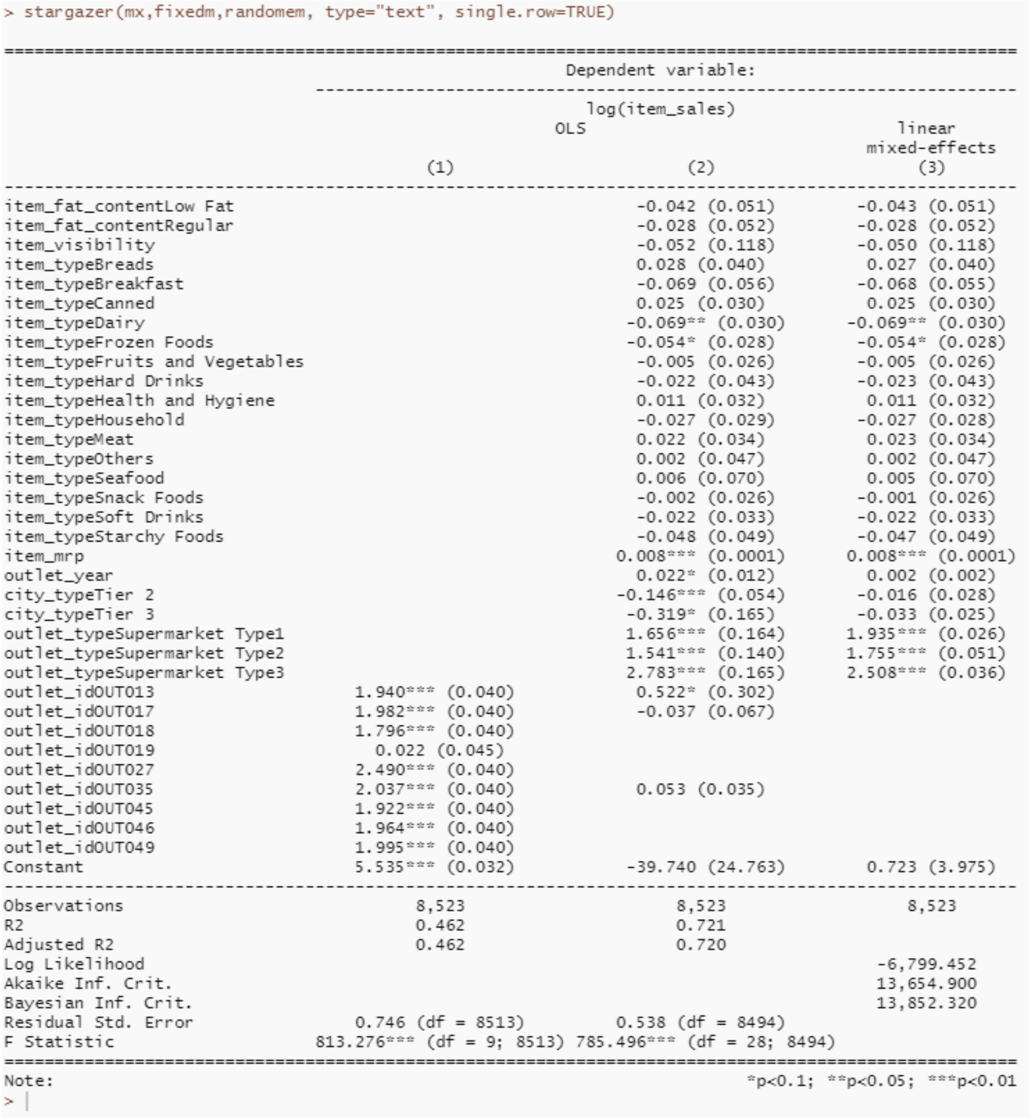
In order to control for the multi-level data, we can use a mixed-effects model to check marginal effect on sales as a function of required input variables, whilst controlling for outlet_id.

mx <- lm(log(item_sales)~ outlet_id, data=d2) fixedm <- lm(log(item_sales)~ item_fat_content + item_visibility + item_type +item_mrp + outlet_year + city_type + outlet_type + outlet_id, data=d2)

randomem <- lmer(log(item_sales)~ item_fat_content + item_visibility + item_type +item_mrp + outlet_year + city_type + outlet_type +(1 | outlet_id), data=d2, REML=FALSE)

1st model uses only the outlet_id variable to understand the difference in item sales between the 11 types of outlets. 2nd model considers outlet_id as a fixed effect (no pooling) whereas 3rd model considers outlet_id as a random effect (partial pooling).

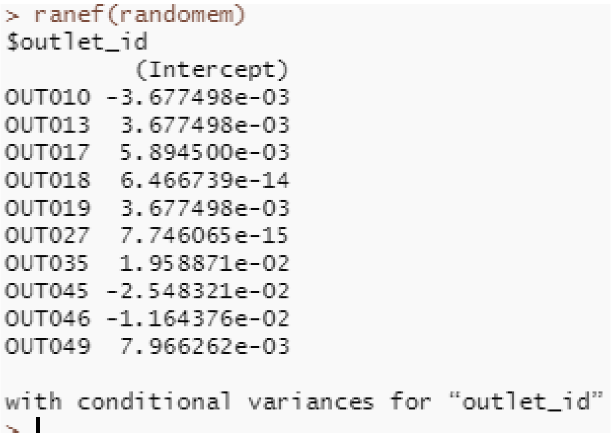
The model output coefficients and the random effect coefficients are presented in the below images and their interpretations follow.



```
> stargazer(mx,fixedm,randomem, type="text", single.row=TRUE)

=====================================================================================================
                                                        Dependent variable:
                               ----------------------------------------------------------------------
                                                          log(item_sales)
                                                       OLS                               linear
                                                                                      mixed-effects
                                          (1)                       (2)                    (3)
-----------------------------------------------------------------------------------------------------
item_fat_contentLow Fat                                        -0.042 (0.051)        -0.043 (0.051)
item_fat_contentRegular                                        -0.028 (0.052)        -0.028 (0.052)
item_visibility                                                -0.052 (0.118)        -0.050 (0.118)
item_typeBreads                                                0.028 (0.040)          0.027 (0.040)
item_typeBreakfast                                             -0.069 (0.056)        -0.068 (0.055)
item_typeCanned                                                0.025 (0.030)          0.025 (0.030)
item_typeDairy                                                -0.069** (0.030)      -0.069** (0.030)
item_typeFrozen Foods                                         -0.054* (0.028)        -0.054* (0.028)
item_typeFruits and Vegetables                                 -0.005 (0.026)        -0.005 (0.026)
item_typeHard Drinks                                           -0.022 (0.043)        -0.023 (0.043)
item_typeHealth and Hygiene                                    0.011 (0.032)          0.011 (0.032)
item_typeHousehold                                             -0.027 (0.029)        -0.027 (0.028)
item_typeMeat                                                  0.022 (0.034)          0.023 (0.034)
item_typeOthers                                                0.002 (0.047)          0.002 (0.047)
item_typeSeafood                                               0.006 (0.070)          0.005 (0.070)
item_typeSnack Foods                                           -0.002 (0.026)        -0.001 (0.026)
item_typeSoft Drinks                                           -0.022 (0.033)        -0.022 (0.033)
item_typeStarchy Foods                                         -0.048 (0.049)        -0.047 (0.049)
item_mrp                                                     0.008*** (0.0001)      0.008*** (0.0001)
outlet_year                                                    0.022* (0.012)         0.002 (0.002)
city_typeTier 2                                              -0.146*** (0.054)       -0.016 (0.028)
city_typeTier 3                                               -0.319* (0.165)        -0.033 (0.025)
outlet_typeSupermarket Type1                                  1.656*** (0.164)      1.935*** (0.026)
outlet_typeSupermarket Type2                                  1.541*** (0.140)      1.755*** (0.051)
outlet_typeSupermarket Type3                                  2.783*** (0.165)      2.508*** (0.036)
outlet_idOUT013                    1.940*** (0.040)            0.522* (0.302)
outlet_idOUT017                    1.982*** (0.040)            -0.037 (0.067)
outlet_idOUT018                    1.796*** (0.040)
outlet_idOUT019                      0.022 (0.045)
outlet_idOUT027                    2.490*** (0.040)
outlet_idOUT035                    2.037*** (0.040)            0.053 (0.035)
outlet_idOUT045                    1.922*** (0.040)
outlet_idOUT046                    1.964*** (0.040)
outlet_idOUT049                    1.995*** (0.040)
Constant                           5.535*** (0.032)           -39.740 (24.763)        0.723 (3.975)
-----------------------------------------------------------------------------------------------------
Observations                             8,523                     8,523                  8,523
R2                                       0.462                     0.721
Adjusted R2                              0.462                     0.720
Log Likelihood                                                                         -6,799.452
Akaike Inf. Crit.                                                                      13,654.900
Bayesian Inf. Crit.                                                                    13,852.320
Residual Std. Error                0.746 (df = 8513)         0.538 (df = 8494)
F Statistic                    813.276*** (df = 9; 8513) 785.496*** (df = 28; 8494)
=====================================================================================================
Note:                                                                     *p<0.1; **p<0.05; ***p<0.01
> |
```

In order to analyze the sales at the granularity level of the outlets, we can look at the beta coefficients of the outlet id's against the percentage effect on item sales since we are using a log transformation.



```
> ranef(randomem)
$outlet_id
         (Intercept)
OUT010 -3.677498e-03
OUT013  3.677498e-03
OUT017  5.894500e-03
OUT018  6.466739e-14
OUT019  3.677498e-03
OUT027  7.746065e-15
OUT035  1.958871e-02
OUT045 -2.548321e-02
OUT046 -1.164376e-02
OUT049  7.966262e-03

with conditional variances for "outlet_id"
> 
```

Interpretation & recommendations

Using the random effects model and its beta coefficients, we can draw the below interpretations and recommendations for some business questions that we can think of from the data-

1. What type of outlet will return the best sales: Grocery store or Supermarket Type 1, 2, or 3?

Compared to grocery stores (baseline),

- Supermarket Type 3 has 250.8% more sales.
- Supermarket Type 2 has 175.5% more sales.
- Supermarket Type 1 has 193.5% more sales.

This also means that, Supermarket Type 3 had 75.3% more sales than Supermarket Type 2 and 56.8% higher sales than Supermarket Type 1. (based on the difference in beta coefficients/ percentages above)

Hence, Supermarket Type 3 has the best sales.

2. What type of city will return the best sales: Tier 1, 2 or 3?

We can observe from the beta coefficients of the city type, we can say that City Tier 1 had-

- 1.6% higher sales than City Tier 2.
- 3.3% higher sales than City Tier 3.

Hence, City Tier 1 will return the best sales.

Therefore, it is recommended to open **Supermarket Type 3** in **City Tier 1 to get even better sales.**

3. What are the top 3 highest performing and lowest performing stores in the sample?

Using the ranef conditional means derived from the final random effects model shown in the image above Q1, we can say that the **best performing stores were OUT049,OUT027, OUT018,OUT017 (in that order)** whereas the **worst performing stores were OUT010, OUT045** and **OUT046 (in that order).**

BUSINESS RECOMMENDATIONS TO INCREASE SALES AND REVENUE

It is recommended to open **Supermarket Type 3** in **City Tier 1 to get improved sales. If Tier 1 is not possible due to business constraints, Tier 3 is the next best possible choice.**

Conduct a survey across best performing stores vs worst performing stores for KPI metrics and conduct further analysis to improve sales at the ground level post-analysis.