**Seoul Retail Case**

Date: 2019-05-09

EXECUTIVE SUMMARY

We analyzed all of company’s Q five franchise stores and conclude that Store B is up to par with its competitors. However, based on the total amount of sales, Store A outperforms all. Furthermore, our analysis focuses on the factors that attribute to Store A’s performance.

**Factors largely affecting all stores:**

According to our analysis, the number of customers that visit the store play a significant role in determining the total sales that the store makes. Consequently, the average sales per customer and the average sales per item, which is obvious and has been evident across all the stores. Also, we were informed that 95% of the store sales come from Japanese tourists, and we have indeed identified Yen-Won ratio and Japanese tourists as the top significant factors apart from the more obvious ones. Store B in particular has a significant impact of following factors:

* Yen-Won Ratio: The higher the ratio, the more the Japanese tourists can buy with the Yen currency. Hence, sales are higher at the store.
* Japanese Tourists: When more tourists visit, the sales have been higher although we require more information to be able to distinguish the Japanese tourists that actually visited the store from the generic number of the customers that have visited the store.
* Weekday Wednesday: This is an interesting observation and there is a chance that people want to shop while they are half way through the hectic workweek.
* Weekday Sunday (Weekend): It is a universal fact that families and friends plan to spend their Sundays outdoors.

**Similarity of factors with other Stores:**

* Yen Won Ratio influences all the stores and is ranked highly for Store A and Store D just like for Store B whereas Store C and Store E is impacted by it a little less significantly.
* Japanese Tourists have a similar impact on Store A and a slightly lower impact on Store D.
* Customers visiting Store D and Store E as well on Wednesdays but the impact is at a comparatively lesser extent than that for Store B.
* Sunday is also an impact factor for Store A and Store C.

**Dissimilarities of factors with other Stores:**

* Store A experience greater sales in the last three months of the year, hence Month is one of its significant factors.
* Discounts affect Store C, Store D and Store E, which makes us think that these stores put out items frequently on discount and hence they contribute towards the total sales.
* Distance from Station X influences the sales at Store C, which is interesting as it is the second farthest store from the Station X.
* Store E has Holiday as one of its prominent impact factors.

**Recommendations for Store B:**

* Increase in Avg. sales per customer as well as per item should be a top priority for Store B to match with Store A.
* Plan some exciting discounted offers especially in the last three months of the year and on long weekend holidays.
* The outlook impact is minimal across the stores and so it is perfectly fine to remain focused in terms of full-blown store operations throughout the year irrespective of the season.
* The Yen-Won conversion ratio is quite a good lead indicator to capitalize on sales and can be monitored closely. The fluctuations can really aid in deciding the timings of discount, inclusion of fresh stocks and offers.
* Distance from Metro and Main street doesn’t seem to have come out as that a significant influencing factor and so can remain least bothered about the store location per say.

TECHNICAL SUMMARY

Our analysis was conducted by using a combination of SAS E-miner, R and Excel. We were given a dataset with the following variables: Store ID, Store Name, Number of Customers, Number of Items Sold, Total Sales, Discount, Average Sales per Customer, Average Sales per Item, Date, Day of week, Distances from Metro Stations X and Y (in meters and feet), Distance from the nearest main thoroughfare (in meters and feet).

We removed the variables that were were highly correlated to Total Sales: Avg.Sales.per.Customer, Avg.Sales.per.item, \_\_of Items. We let go of some redundant variables duplicating the information as inputs to Regression and Decision Tree inputs: Distance.from.Main.Street.Feet, Distance.from.Main.X.Feet, Distance.from.Main.Y.Feet

We followed a 3-fold approach in terms of building hypothesis, validating it through feature selections and validating the selections through Regression analysis, Decision Tree (with Interactive simulation) in SAS and some basic Excel graphs. Below are the variables in the order of significance after running multiple linear regression in R. The Decision Tree default run gave us the obvious splits like No. of customers and items purchased. We pruned the tree and attempted exploring the interactive capabilities of the tree for a few nodes like No. of customers, Currency Ratio, items etc. The variations in splitting points seemed not so encouraging in a sense that we observed the alterations in the order of variable importance, to a very insignificant extent. Therefore, we ended up choosing to stick to default behavior of the tree output for most of our subsequent analyses.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Importance** | **All Stores** | **Store A** | **Store B** | **Store C** | **Store D** | **Store E** |
| 1(Highest) | Discount | YenWonRatio | YenWonRatio | Discount | YenWonRatio | Discount |
| 2 | YenWonRation | Japanese.Tourists | Japanese.Tourists | YenWonRatio | Discount | YenWonRatio |
| 3 | Distance.from.Main.Street.Meter. | Month | WeekdayWednesday | WeekdaySunday | Japanese.Tourists | Holiday |
| 4 | Japanese.Tourists | WeekdaySunday | WeekdaySunday | Distance.fromStation.X.Meter | WeekdayWednesday | WeekdayWednesday |

Store A experiences greater sales in the last three months of the year, hence Month is one of the significant factors. Discounts affect Store C, Store D and Store E, which makes us think that these stores put out items frequently on discount and hence they contribute towards the total sales. Distance from Station X impacts the sales at Store C, which is interesting at it’s the second farthest store from the Station X.Store E has Holiday as one of its prominent impact factors.

We attempted exploring the various features from the dataset to gauge the potential impact on sales. The relative p-values served as the guidance parameter for arriving at the following conclusions.

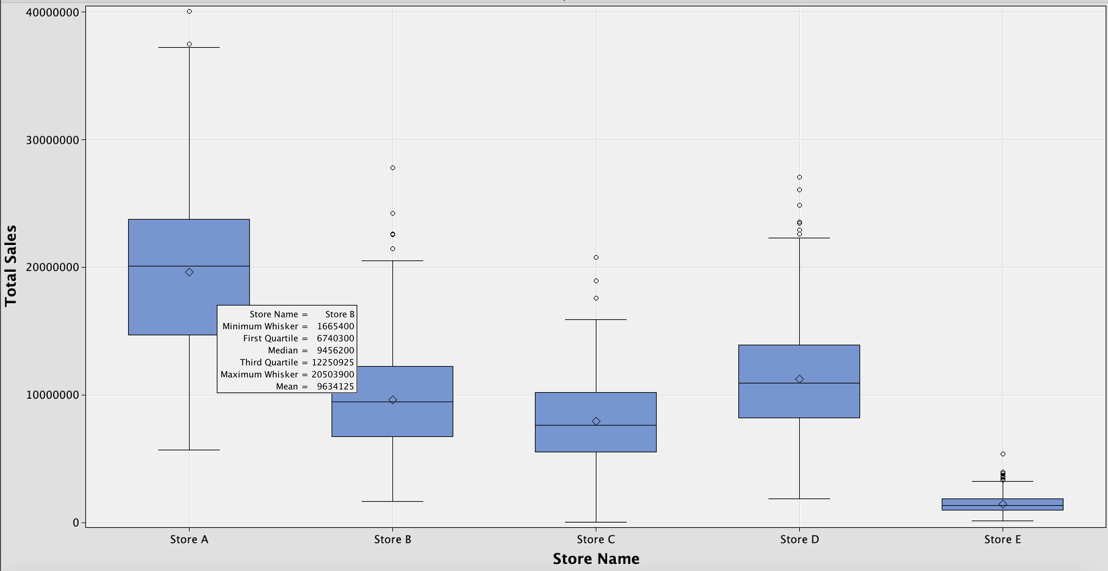
* How the stores fair on Per Customer and Per Item sales?
  + We do not place significance on average store per items due to all stores selling the same products. Thus, the output of this metric for all stores will not have significant difference. Furthermore, we weigh the total sales more.
* Do holidays really influence the sales? If so, for which all stores and by how much?
  + We see all the stores having a pretty much a positive impact on sales on Holidays and so is a definite driving force for all of them.
* Do discounts really influence the sales? If so, for which all stores and by how much?
  + Yes, it has a positive impact for almost all stores. A&D show quite a significant one.
* Does the distance from Main Street influence the sales? If so, for which all stores and by how much?
  + Not really. Not much a differentiating factor. Because the farter stores have been performing quite well. In fact, the store C performed better after re-opening at farter location. The same goes with distance from station X too.
* Do the Japanese tourists influence sales? If so, for which all stores and by how much?
  + We do not get to see any conclusive evidence to make any inference about the relationship between the influx of tourist and sales volume.

 CONTRIBUTING ANALYSTS

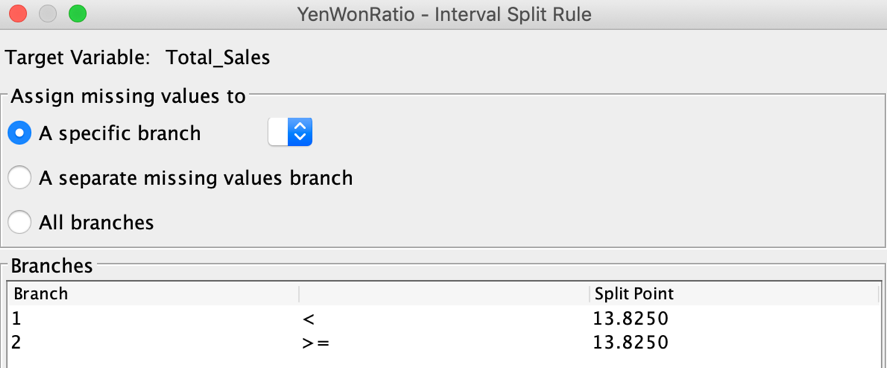
Varshila Redkar, Destiney Plaza, Sarang Ramesh Dani, Liyang Liu, Swati

APPENDIX

Sample SAS Graphs

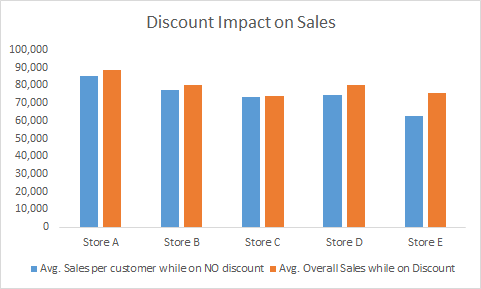


Boxplot of stores against Total Sales



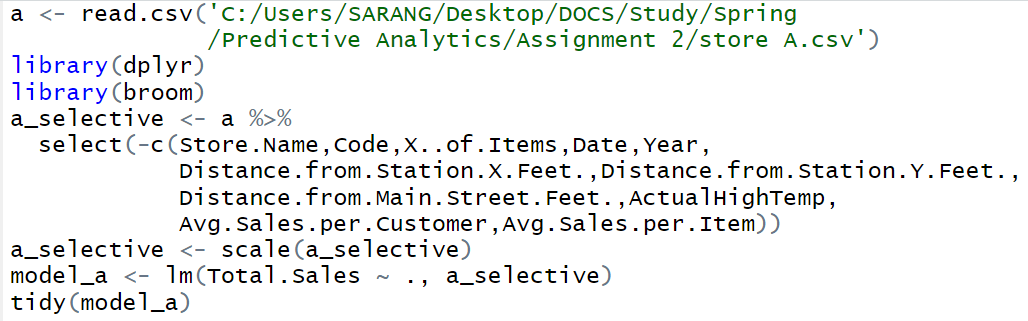
Default split rule for Store A

Sample Excel Graph



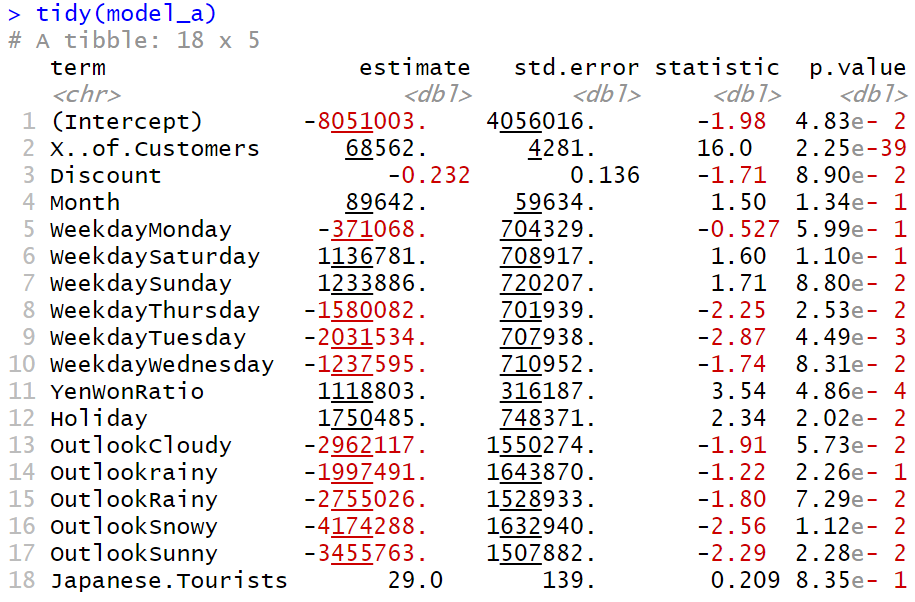
No Discount to Discount ratio for each store

R Code used to identify prominent variables for Store A as sample



Analyzing for Store A

R Output as a sample



Store A variable output