**Analysis of Ice Cream and Customers**

Section 02

Nan LIANG

**Executive Summary**

Using a dataset of 21974 transactions and 15 variables about customers and ice cream products, the Analysis team was able to identify key variables such as household income and coupons, influencing customers to spend on ice cream. In further study, analyst declined the assumption that coupons cast stronger influence on Asian customers in purchasing. Although exposed to the risks of considerable proportion of false positives and unreliable data, the Analysis team carefully drew inferences from selected variables, and made recommendations to future marketing and product development strategies.

**Introduction**

To increase the ice cream sales in our company and better target customers, the Analysis team conducted an observational study of purchases of our ice cream products.

**Problem Formulation & Data Characteristics**

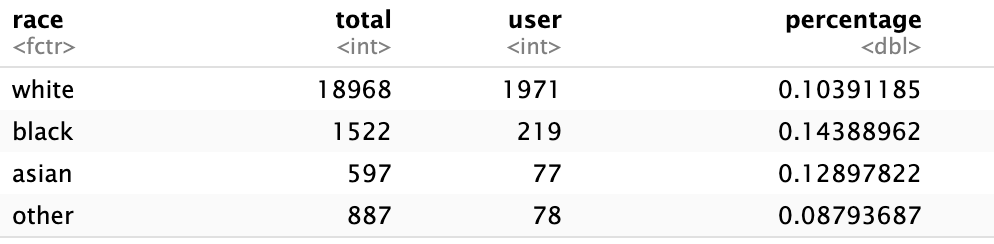
The Analysis team collected data from 21974 transactions and generated 15 variables of interest to explore their relationships with price paid per item of ice cream. The variables mostly concern customers’ lifestyle whilst a couple of them concerning the characteristics of products. The Analysis team’s objective was to find out what factors contribute to customers’ purchasing behavior and what kind of products are of their interests.

**Model Development**

The analysis started with comprehensive multivariate regression with all the 15 variables. Through statistical analysis, the analysts were able to identify variables with significant ability to predict price per item. In the end, the analyst reached an effective multiple regression model with predictors the company should focus on. (Appendix 01, Appendix 02)

Based on Pew Research, Asian population has been the fastest growing group in the country

in the past 20 years (Reference 01). Further study was based on racial groups’ preference of coupons. Analyst had assumptions about Asians are more likely to purchase when coupons were available. However, analysis result suggested otherwise (Appendix 03).



(*Table 1. Percentage of Coupon usage based on racial groups*)

**Results**

The final reduced model still contains 13 variables, including customer traits and product traits (Appendix 02). Among all these statistically significant variables, customers’ household income, coupons, and Southern Region are particularly outstanding. In products traits, product size and certain product flavors such as Pumpkin CSK, have relatively better influence on the price customer pay for products (Appendix 04).

In further studies, through method of Difference in Difference analysis to find out true efficacy of household income and race (divided into Asian and Non-Asian ), analyst was surprised to discover there was no significant evidence for coupon usage stimulating Asian customers to purchase ice cream (Appendix 05).

**Recommendation & Implication**

Extended from statistical inferences, the Analysis team recommend 1) marketing team focus more on sending out coupons in Southern region to households with disposable income. Targeted marketing schemes are expected, 2) product development team invest resources more on developing new flavors, 3) since 32oz of size has not been favored in the market, product designer could experiment on finding other sizes that interest customers.

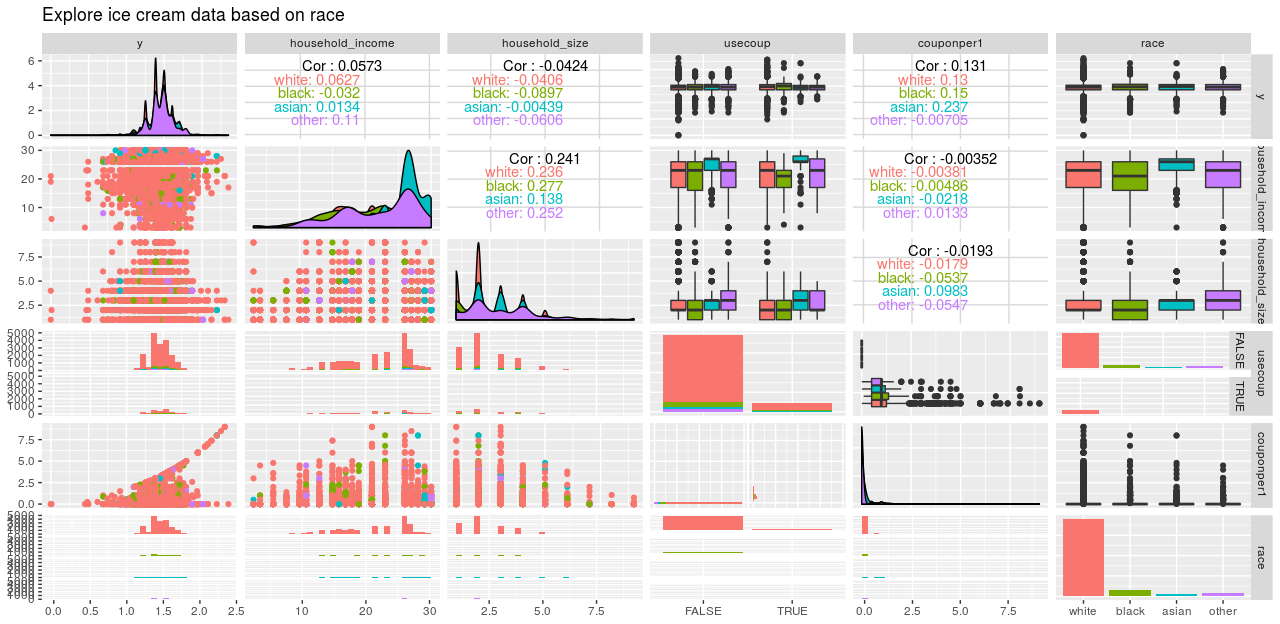
The case did not end without concerns. 1) the proportion of falsely identified significant variables (FDR) correlates with the number of variables used in the formula and the levels of them. In the reduced model, we have 13 variables with multiple levels. Solution such as rank the significance of each then decide a cutoff to find the truly significant should be applied in further analysis, 2) since there was no background information about when and how the data was collected, influence of seasonality and data credibility were put to questions, 3) the significant efficacy of Southern Region might just be another form of temperature influences ice cream sales. Difference in Difference method might be applicable, studying region and temperature influencing ice cream sales.

**Conclusion**

Ice cream is a special product with distinctive traits. Without reliable data or scientific analysis, the company would not be able to deploy resources to the right place or reach out to the right customers.

**Appendix**

1. Full model: y ~ flavor\_descr + size1\_descr + household\_income + household\_size + usecoup + couponper1 + region + married + race + hispanic\_origin + microwave + dishwasher + sfh + internet + tvcable , (*y = log(1 + (price\_paid\_deal + price\_paid\_non\_deal) / quantity)*)
2. Reduced model: y ~ flavor\_descr + size1\_descr + household\_income + household\_size + usecoup + couponper1 + region + married + race + microwave + sfh + internet + tvcable
3. Relationship exploration among selected variables including race.



1. R markdown output. line 45

A close up of text on a black background

Description automatically generated

1. R markdown output. Line 56

A screenshot of a cell phone

Description automatically generated

**Reference**

1. Lopez G. et, 2017, *Key facts about Asian Americans, a diverse and growing population*, Pew Research Center, https://www.pewresearch.org/fact-tank/2017/09/08/key-facts-about-asian-americans/