Marketing and Household Spending:

Impact of demographic, transactional and campaign factors on spending by households

University of Southern California

Fall 2023

Team members (Core A, Section - 16234)

Tatevik Hakobyan, Aakanksha Sangwan, Rishabh Pratap Singh Sisodia, Akshat Rai, Raunak Moolchandani

Abstract

The study examines the influence of marketing campaigns on household retail purchases, considering customer demographics and transaction behaviors. The null hypothesis is that exposure to marketing campaigns does not significantly affect the amount spent by retail households considering demographic factors such as income, household size, marital status whereas the alternative hypothesis is that exposure to marketing campaigns does affect the amount spent. Using statistical methods such as analysis of variance and multivariate regression, we analyze relationships between customer demographics, campaign opt-in, and sales. Additionally, an exploratory data analysis has been conducted to explore the different variables across demographic, transaction and campaign datasets. Existing retail marketing literature primarily focuses on campaign effects on individual customer segments for sales growth (Bas van den Putte, 2009). This dataset has also been used to analyze one category of food purchases to improve dietary (Guan, 2018). However, this study uniquely explores household consumption behavior, linking demographics to marketing impact across multiple product categories. This paper finds that factors such as type of coupon which was redeemed, distinct product bought, retail discount, distinct product manufacturer, distinct department, have significant impact on household purchases.These findings will enhance the understanding of how customer profiles relate to the effectiveness of marketing interventions. In turn, it aids decision-makers in devising personalized discount structures and campaigns, ultimately optimizing retail marketing strategies.

1. Introduction

Marketers frequently employ coupons as a significant promotional strategy across various consumer product sectors. 77% of the United States population uses coupons and shoppers saved more than $3 billion last year by doing so. Coupon users report an average of 11.5% savings on their grocery bill with coupons, prompting manufacturers to respond by offering more than $250 billion in coupon savings in 2003 (Somjit Barat, Lilly Ye, 2015). From a managerial perspective, it is very important to figure out factors that affect coupon redemption, so that marketers can design effective coupons in accordance with their promotional objectives. Many researchers have been trying to identify critical factors in coupon redemption (Boon Young Lee, 2004). Through empirical analysis and visual representation of data, this paper aims to show if marketing campaigns affect household purchases. The hypothesis are as follows:

H_0: Exposure to marketing campaigns does not significantly affect the amount spent by retail households considering demographic factors such as income, household size, marital status.

H_A: Exposure to marketing campaigns significantly affects the amount spent by retail households considering demographic factors such as income, household size, marital status.

The structure of this paper is as follows. We examine the relationship between household purchases and factores. We also explore the relationship between the campaign received and income of the households. After that we define our hypothesis followed by discussion of our data and the design of our empirical testing. Afterwards we provide our analysis and findings. Last part summarizes key findings and offers recommendations.

2. Literature Review

In the dynamic landscape of consumer behavior, the role of marketing campaigns emerges as a pivotal force shaping the choices and spending patterns of individuals. A comprehensive exploration of the literature reveals that effective marketing strategies wield considerable influence over consumer spending behavior. Existing retail marketing literature predominantly concentrates on the effects of campaigns on individual customer segments for sales growth (Bas van den Putte, 2009), type of coupon to be recommended (Mutanen & Nousiainen, 2010) to the customers and churn prediction. In contrast, this study delves into household consumption behavior, establishing links between demographics (Tasnim AlHelali, 2023), marketing impact, and diverse product categories. Notably, Guan et al.'s research (2018) reveals that coupon exposure significantly increases category-level purchase rates. This study complements this insight, aiming to broaden the understanding of how marketing campaigns impact household purchases across varied product categories.

3. Methodology

3.1 Dataset

The dataset gives information about consumer purchasing behavior aggregated at a household level for one retailer and contains information across three broad categories of demographics, transaction behavior, and marketing campaign data. Each of these categories consists of key factors (KPIs) for the research. This dataset is spread across eight different tables, linked

through various common identifiers such as household_key, campaign etc. The dataset belongs to firm Dunnhumby, a retail and customer data science firm and made available through Kaggle.

For the research hypothesis, the data needs to be aggregated at household level and relevant fields combined. Tables are merged to create the sub dataset of 801 households with 47 factors under all the categories.

3.2 Statistical Analysis

For the hypothesis, this paper explores the relation between the dependent variable (amount spent) and independent variables (demographic and marketing related). Accordingly, multivariate regression is used to analyze the effect of various factors on the amount spent. The analysis uses total_sales as the proxy variable for the amount spent by the consumer. Using p-values the statistical significance of the relationship is determined and for this analysis as per (Sama, R. , 2019) and (M. Khan and Amna Tanveer, 2019) alpha is set at value 0.05. In order to test the difference in marketing factors such as received campaign and coupon redemption across different income categories, 7 way ANOVA, controlling for interaction error due to multiple income categories. The combined analysis, done using Python, provides insights at both comprehensive and individual factor level.

4. Results

4.1 Exploratory Data Analysis

Each table in the dataset represents a facet of the multifaceted dataset under study, ranging from demographic information to purchasing behaviors. The process of EDA focuses not only on

descriptives but extends to graphical representations, correlation matrices to understand the nuances in each dataset. These insights also help with handling missing data. In the current dataset, there aren't any null values, however the overall size is reduced from the list of 2500 households due to household demographic information being limited for 801 households. Summary statistics of numerical variables show that on average a household received 5

	total_campaign_rec	total_basket	total_distinct_department	total_coupon_redm
count	801.00	801.0	801.00	801.00
mean	5.26	175.2	15.63	2.32
std	3.20	130.9	2.98	5.00
min	0.00	22.0	7.00	0.00
25%	3.00	96.0	14.00	0.00
50%	5.00	139.0	16.00	0.00
75%	7.00	214.0	18.00	2.00
max	17.00	1223.0	24.00	35.00

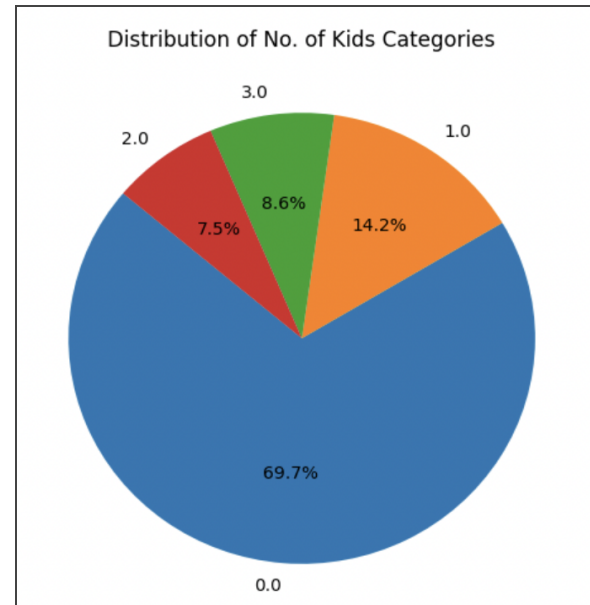*Table 1: Summary statistics for numerical independent variables*

campaigns and had an average basket size of 175.2, which had products from 15 departments.
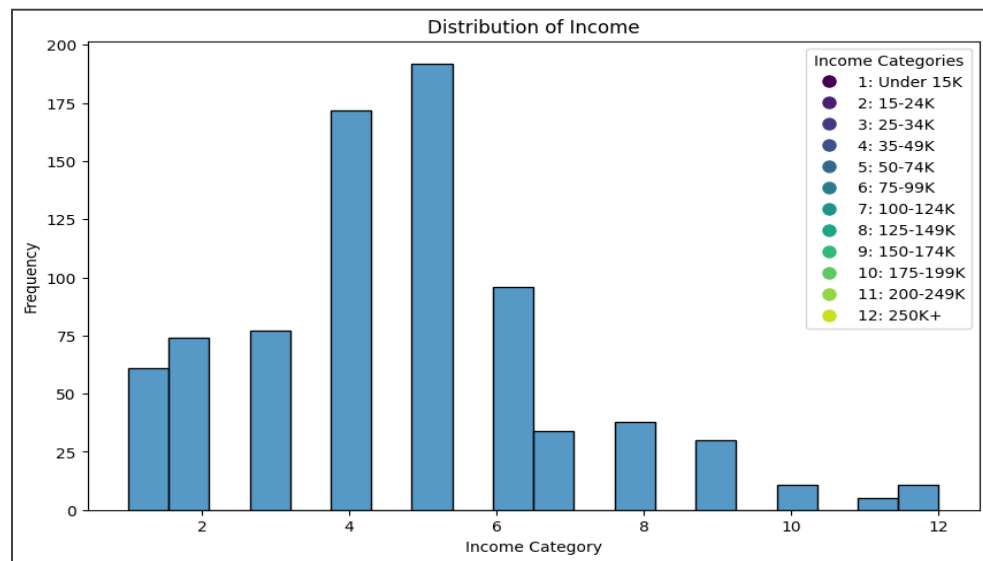
Histogram 1: Distribution of households across different income categories
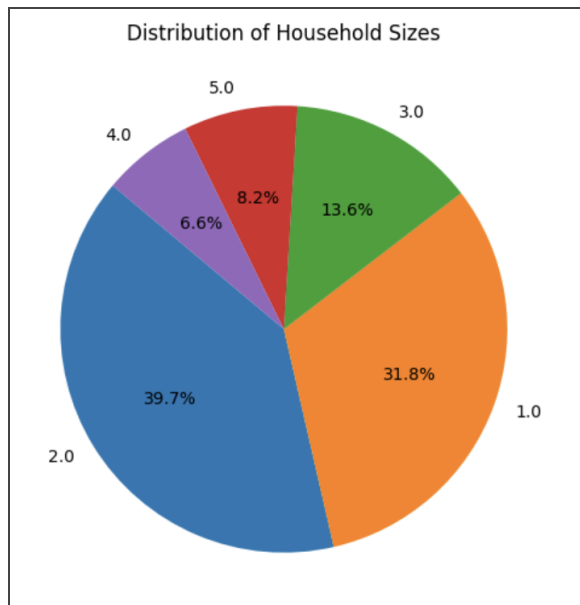
Pie chart 1

Pie chart 2

## 4.2 Analysis of Variance : Income Categories

The analysis focused on examining the relationship between income categories and two campaign-related metrics: total campaign receipts and total coupon redemption. The ANOVA results revealed statistically significant differences in both total campaign receipts across different income categories.

```
ANOVA for Total Campaign Received:
F-statistic: 2.413327
P-value: 0.00594

ANOVA for Total Coupon Redemption:
F-statistic: 2.734124
P-value: 0.001776
```

Table 2: One-way ANOVA for total campaign received and total coupon redemption across income categories

To further investigate the difference in campaign metrics, the paper uses post hoc test such as Tukey's HSD test to provide further insights. The result shows that the majority of the pairwise comparisons did not yield statistically significant differences, as indicated by the "False" entries under the "reject" column. Thus for most income groups there is no difference in the means of total campaign receipts. Exception lies for category 150-174K and 25-34K revealed a significant mean difference (meandiff = -2.31, p-adj = 0.036), indicating that these two income groups differ significantly in their total campaign receipts, with individuals in the income range of 150-174K receive fewer total campaigns compared to those in the income range of 25-34K. Doing Tukey'HSD for income categories and coupon redemption, yields no statistically significant difference across income categories except for income ranges of 35-49K , 50-74K and 15-24K, 50-74K. The detailed results of Tukey's HSD test are attached in the appendix.
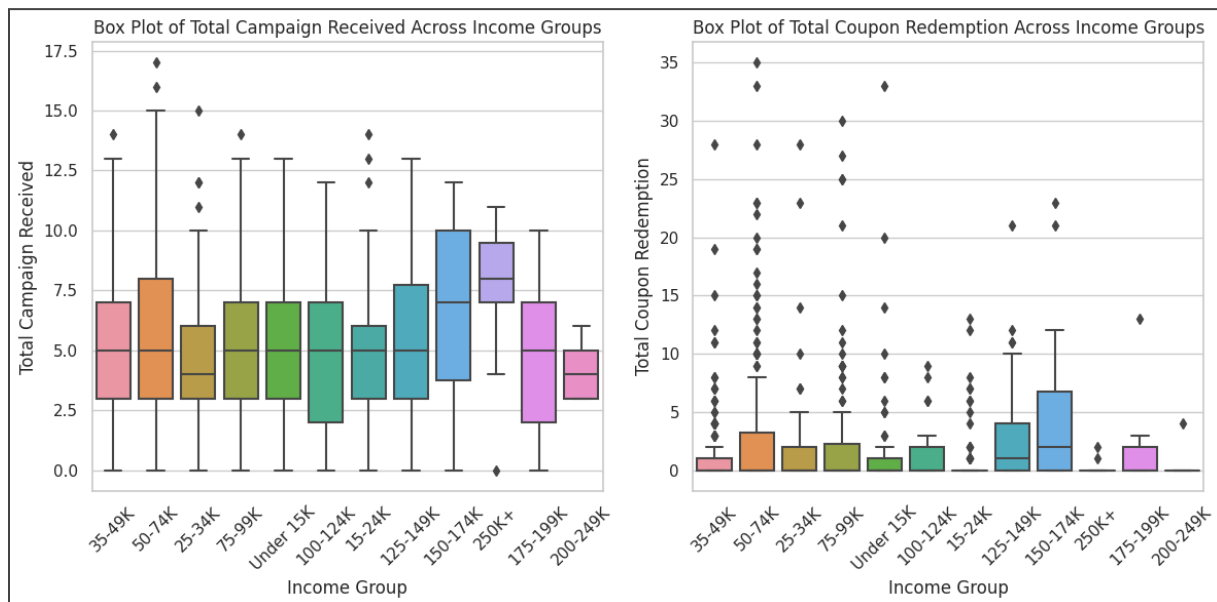
Box plot 1, 2: Campaign received and Coupon redemption variance across income categories

4.3 Correlation Matrix

To get a broad understanding of how various factors affect the total_sales or amount spent, correlation matrix is used to visualize the strength of a relationship. For the attached matrix, representative variables have been chosen across the three categories instead of taking all the variables. In the heat map we see that total_sales is more affected by campaign variables such as distinct_camapign_received (0.64), distinct_products_bought (0.85) over demographic variables such as income or household size etc.

Correlation Matrix Heatmap
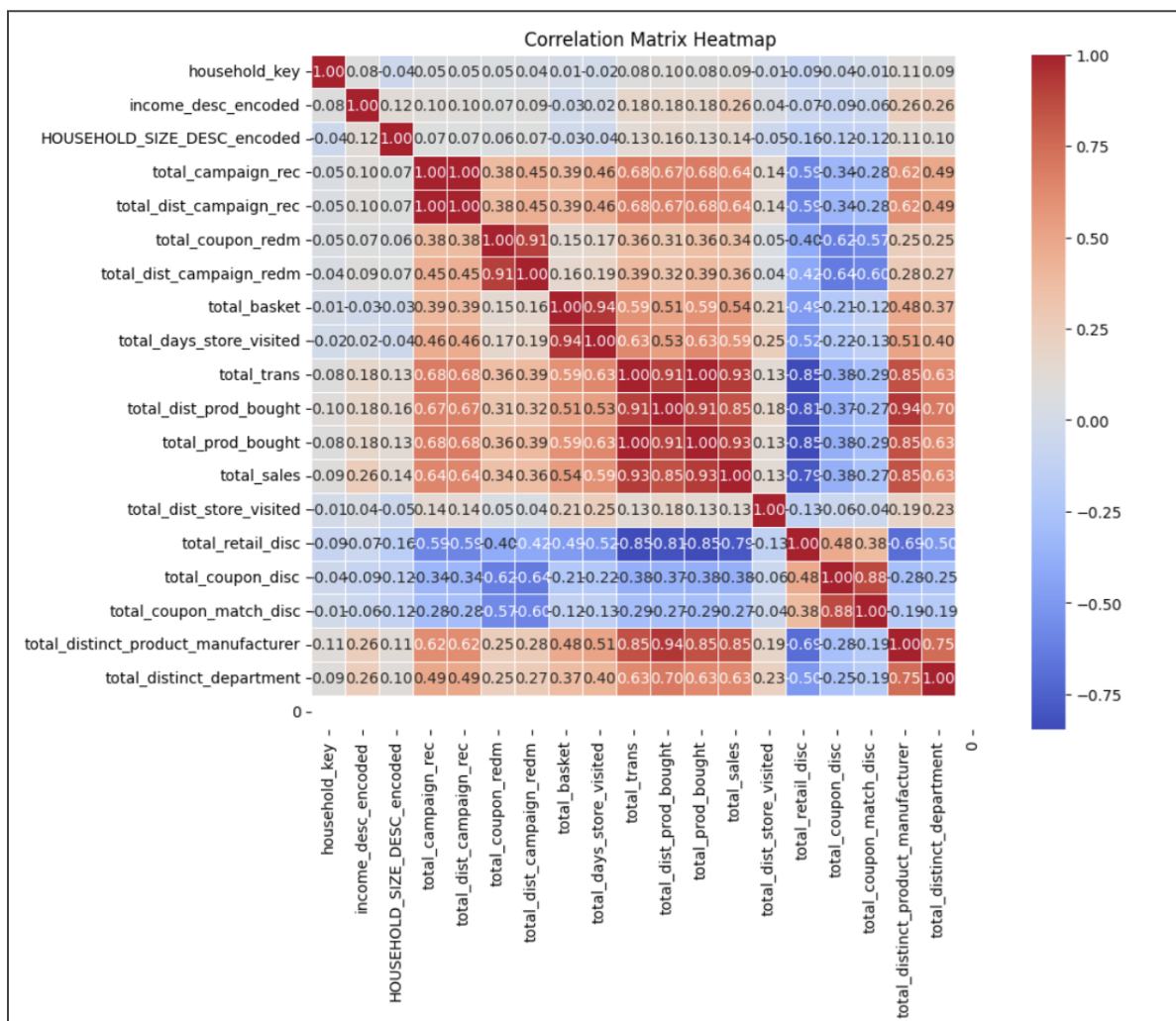
4.4 Multivariate Regression

We use the inbuilt function in python to get dummy labels for encoding the categorical variables. For example, income categories are divided across codes from 1 to 12. This is a necessary step as numerical inputs are needed for linear regression. However, this increases the dimensionality of the dataset, thus bringing the number of independent varll the 3 broad

categories. The dependent variable for the regression is total_sales, proxy for amount spent by the consumer. In order to assure the assumptions for linear regression hold true, some of the variables were transformed. The yeo johnson transformation was used out of five other methods and helped balance out data variations, making the dataset more consistent. Power transformations stabilized variances and strengthened our analysis. The log transformation was key in making the data less skewed and more balanced. Sqrt transformations, good for reducing skewness, were flexible enough to handle instances with zero values. The square transformation, our final step, continued to make the data less skewed and more even. A correlation analysis and correlation matrix created for a complete list of independent variables to identify highly correlated factors, i.e. having a correlation index greater than 0.95 in magnitude.

The results of the regression model are attached in the table below. The p value of multiple factors shows their significance level when compared with 0.05 alpha. Analysis shows that most factors do not have a significant effect on the total_sales, however factors such as total_typeB_coupon_redm,   total_distinct_product_manufacturer,   total_distinct_department,

yeojoohnson_total_coupon_disc, total_dist_campaign_redm, had significant impact with

```
                          OLS Regression Results
===============================================================================
Dep. Variable:               total_sales   R-squared:                     0.916
Model:                               OLS   Adj. R-squared:                0.912
Method:                    Least Squares   F-statistic:                   224.8
Date:                   Fri, 24 Nov 2023   Prob (F-statistic):             0.00
Time:                           17:03:25   Log-Likelihood:              -6736.3
No. Observations:                    801   AIC:                        1.355e+04
Df Residuals:                        763   BIC:                        1.373e+04
Df Model:                             37
Covariance Type:               nonrobust
===============================================================================
```

p-values less than 0.05.

Table 2: Multivariate regression results

Based on results, see that campaign_received variable is not significant for amount spent and thus, just having a marketing intervention does not impact the amount spent. However, the total different campaigns that were redeemed, having distinct products in their basket, and having products from different departments were effective in impacting the amount spent.

5. Discussion

The results of the analysis help businesses decide the marketing strategies and answer questions such as what number of campaigns should be launched? Should a single campaign/coupon cover multiple products or different campaigns for different products? Regression provides insights on impact of various factors and this can help businesses decide by taking into account an individual's firm's context. The aim should be to increase the redemption of unique campaigns so sending out a high frequency of campaigns to begin with should be considered. Additional analysis such as ANOVA also shows that there are certain areas where campaign metrics vary

depending on the income categories, thus indicating a more tailored approach for different customer segments.

6. References

Barat, S., Ye, L. (2015). Effects of Coupons on Consumer Purchase Behavior: A Meta-Analysis. In: Spotts, H. (eds) Marketing, Technology and Customer Commitment in the New Economy. Developments in Marketing Science: Proceedings of the Academy of Marketing Science. Springer, Cham. https://doi.org/10.1007/978-3-319-11779-9_15

Lee, B. Y. (2004). An empirical study on coupon redemption rates in service sector Focused on coupon face value and benefits (Doctoral dissertation, KDI School of Public Policy and Management).

Reibstein, D. J., & Traver, P. A. (1982). Factors Affecting Coupon Redemption Rates. Journal of Marketing, 46(4), 102–113. https://doi.org/10.2307/1251366

Bas van den Putte (2009) What matters most in advertising campaigns?, International Journal of Advertising, 28:4, 669-690, DOI: 10.2501/S0265048709200813

Guan, X., Atlas, S.A. & Vadiveloo, M. Targeted retail coupons influence category-level food purchases over 2-years. Int J Behav Nutr Phys Act 15, 111 (2018). https://doi.org/10.1186/s12966-018-0744-7

Sama, R. (2019). Impact of Media Advertisements on Consumer Behaviour. Journal of Creative Communications, 14(1), 54-68. https://doi.org/10.1177/0973258618822624

Khan, Mukaram and Tanveer, Amna and Zubair, Dr Syed Sohaib, Impact of Sales Promotion on Consumer Buying Behavior: A Case of Modern Trade, Pakistan (June 30, 2019). Impact of sales promotion on consumer buying behavior: a case of modern trade, Pakistan. Governance and Management Review, 2019, Vol 4(1), 38-53, https://ssrn.com/abstract=3441058