Data Analytics in Business and Industry

Group Project

Group 4

22-02-2022

Project Background

Market segmentation is a marketing technique which allows companies to understand customer needs and preferences and thus design strategies which can best satisfy these. Companies recognise that they can neither address all customers uniformly nor approach them uniquely. Customer segmentation sits between these two extremes. The approach facilitates consumer interaction by identifying groups with common characteristics or behaviours which can then inform marketing campaigns, pricing strategies and general consumer communication (Fonseca, 2011).

In this project we analyse customer interaction with a discount offer campaign and monitor its impact on their spending habits. Our aim is to identify groups of customers who are more or less reactive to the campaign, with the motivation that our segmentation will allow for more targeted, efficient campaigns in future.

Data Description

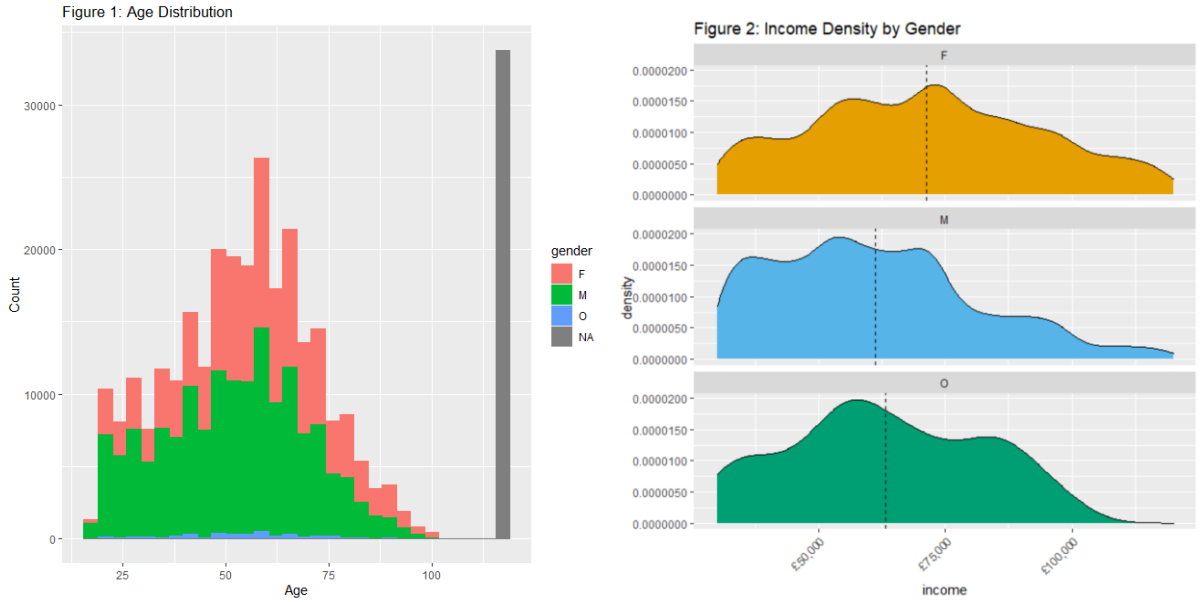
Three datasets were provided for the project:

- portfolio - describes the offers available
- profile - describes the customers
- transactions - provides the customer transactions over the promotional period

We combined them into a single customer level set which contained a combination of customer profile information, such as age and income, with their aggregated transaction data, e.g. average transaction value, frequency of transactions. All categorical variables were converted into dummy variables. We used this combined dataset as a base for our exploratory data analysis. Appendix table 1 displays the fields in our dataset.

Exploratory Data Analysis

Figures 1 and 2 display the customer income and age distributions. As is evident from figure 1, ~13% of the dataset are assigned no gender and an age of 118 – the missing data from these customers mean that they cannot be used within any customer segmentation based on these characteristics. The mean age of females is higher than the mean of males. Regarding income, the higher concentration of males is below £70,000, while females reach a peak of density at £75,000. There is a higher concentration of females with salaries greater than £75,000 compared to males and gender 'other'.



An initial analysis of the relationships between transaction value, transaction frequency, age and income in a simple, universe-wide linear context found several statistically significant relationships. Namely, the average transaction value increases with income and age, while the transaction frequency decreases with income and age. Controlling for income and age, the average transaction value increases with transaction frequency. These relationships can be viewed in table 1, which displays the coefficients, t-statistics and p-values from a simple linear regression, where ATV refers to the average transaction value and TF the transaction frequency:

$$Log(ATV) = \alpha + \beta_1 Income + \beta_2 Age + \beta_3 Gender + \beta_4 TF + \varepsilon$$

Table 1: Log average transaction value regressed onto customer characteristics

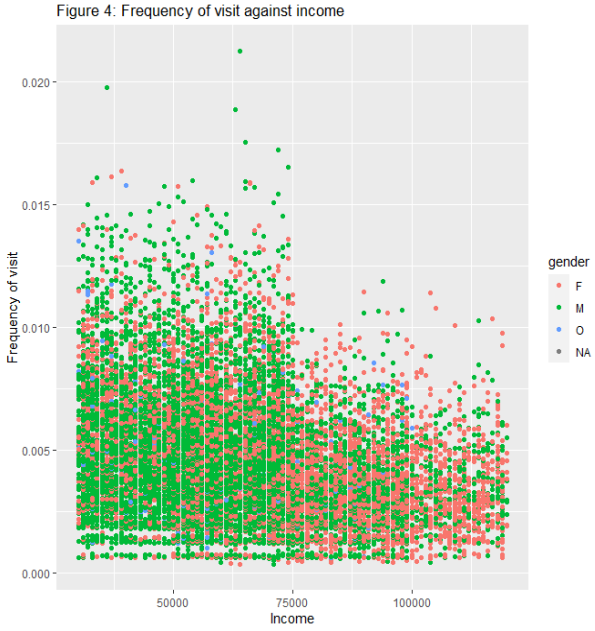
Coefficients	Estimate	Std. Error	t-stat	Pr(>|t|)
(Intercept)	0.316	0.03	10.10	<2e-16***
TF	27.070	2.37	11.42	<2e-16***
Income	0.000	0.00	92.88	<2e-16***
Age	0.005	0.00	14.20	<2e-16***
Gender: Male	-0.294	0.01	-24.24	<2e-16***
Gender: Other	0.016	0.05	0.32	0.746

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 3 plots the natural log of each customer's average transaction value against their income, displaying the aforementioned positive relationship between these variables. Figure 4 plots the frequency of each customer's average transaction value against their income, displaying a negative relationship between these variables.

Figure 3: Log average transaction against income

Figure 4: Frequency of visit against income



Regarding the relationship between age, income and the percentage of offers completed by a customer, an initial inspection suggests that both characteristics are positively correlated with a propensity to take up discount offers. This relationship can be viewed in figures 5 and 6, which plot the offer completion ratio against age and income respectively.

Figure 5: Average offer complete ratio Vs. Income Range

Data has been grouped by Gender

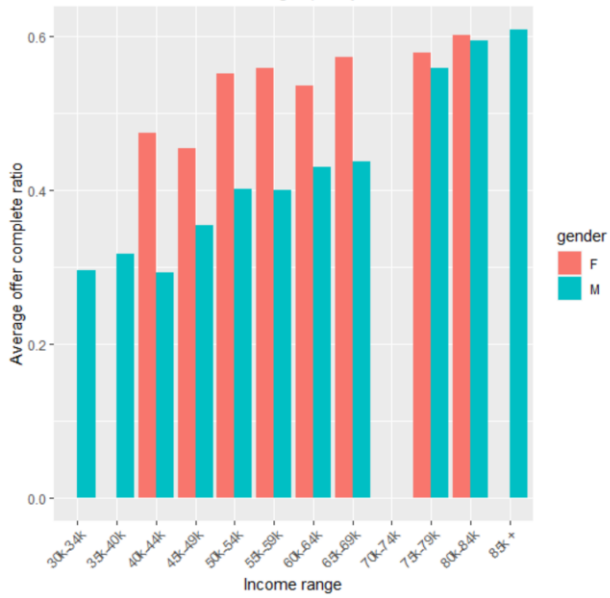
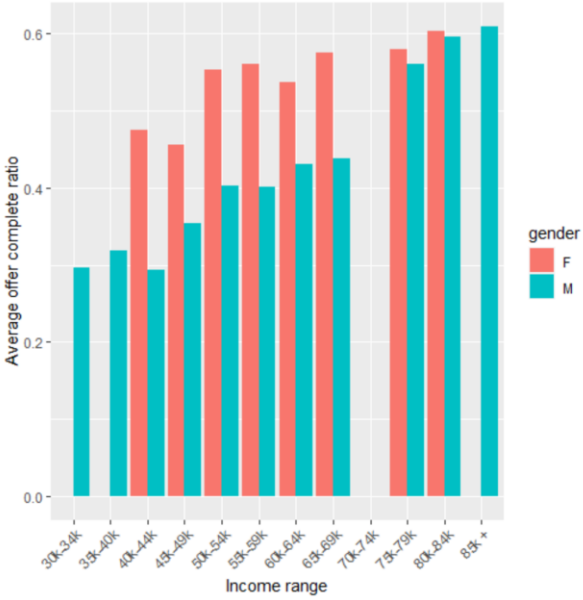


Figure 6: Average offer complete ratio Vs. Income Range

Data has been grouped by Gender



Next steps

After the exploratory analysis we have a deeper understanding of the variables that might correlate with a customer's decisions whether to join the promotion. In the next phase of the project we will choose variables, potentially using feature selection, followed by the application of clustering algorithms to find natural groups of customers.

References

Fonseca J.R. (2011) Why does Segmentation Matter? Using Mixed Methodology to Identify Market Segments. In: Morschett D., Foscht T., Rudolph T., Schnedlitz P., Schramm-Klein H., Swoboda B. (eds) European Retail Research. European Retail Research. Gabler Verlag, Wiesbaden. https://doi.org/10.1007/978-3-8349-6235-5_1

Appendix

Appendix table 1: Dataset fields

Variable	Variable Type	Description
customer_num	integer	Unique, more readable customer id
gender	character	Original gender variable
male	double	Binary - male = 1
female	double	Binary - female = 1
other_gender	double	Binary - other = 1
age_range	character	Age range variable
income	integer	Original income variable
income_range	character	A range for the income variable
month_of_membership	integer	Month when customer joined
year_of_membership	double	Year when customer joined
length_of_membership	double	Length or time since customer joined
amount_transactions	integer	Number of transactions
transaction_freq	double	Frequency of transactions since joining
total_transaction_spend	double	Total transaction spend
average_transaction_spend	double	Average transaction spend
amount_offer_received	integer	Number of offers received
amount_offer_viewed	integer	Number of offers viewed
amount_offer_completed	integer	Number of offers completed
offer_view_ratio	double	Ratio of offers viewed to recevied
offer_complete_ratio	double	Ratio of offers completed to received
total_completed_reward	integer	Total value of rewards from completed offers
average_completed_reward	double	Average value of rewards from completed offers
email_offer_view_rate	double	Email view rate
email_offer_complete_rate	double	Email completion rate
web_offer_view_rate	double	Web view rate
web_offer_complete_rate	double	Web completion rate

mobile_offer_view_rate	double	Mobile view rate
mobile_offer_complete_rate	double	Mobile completion rate
social_offer_view_rate	double	Social view rate
social_offer_complete_rate	double	Social complete rate
bogo_received	double	Bogo offers received
bogo_view_rate	double	Bogo view rate
bogo_complete_rate	double	Bogo complete rate
discount_received	double	Discount offers received
discount_view_rate	double	Discount view rate
discount_complete_rate	double	Discount complete rate
info_received	double	Info offers received
info_view_rate	double	Info view rate
info_complete_rate	double	Info complete rate
rec_bogo_diff10_rew10_dur07	double	Received Bogo_Diff10_Rew10 offers dur07
view_rate_bogo_diff10_rew10_dur07	double	View rate
comp_rate_bogo_diff10_rew10_dur07	double	Comp rate
rec_bogo_diff10_rew10_dur05	double	Received bogo_diff10_rew10_dur07
view_rate_bogo_diff10_rew10_dur05	double	View rate bogo_diff10_rew10_dur07
comp_rate_bogo_diff10_rew10_dur05	double	Comp rate bogo_diff10_rew10_dur07
rec_info_diff00_rew00_dur04	double	Received info_diff00_rew00_dur04
view_rate_info_diff00_rew00_dur04	double	View rate info_diff00_rew00_dur04
comp_rate_info_diff00_rew00_dur04	double	Comp rate info_diff00_rew00_dur04
rec_bogo_diff05_rew05_dur07	double	Received bogo_diff05_rew05_dur07
view_rate_bogo_diff05_rew05_dur07	double	View rate bogo_diff05_rew05_dur07
comp_rate_bogo_diff05_rew05_dur07	double	Comp rate bogo_diff05_rew05_dur07
rec_disc_diff20_rew05_dur10	double	Received disc_diff20_rew05_dur10
view_rate_disc_diff20_rew05_dur10	double	View rate disc_diff20_rew05_dur10
comp_rate_disc_diff20_rew05_dur10	double	Comp rate disc_diff20_rew05_dur10
rec_disc_diff07_rew03_dur07	double	Received disc_diff07_rew03_dur07
view_rate_disc_diff07_rew03_dur07	double	View rate disc_diff07_rew03_dur07
comp_rate_disc_diff07_rew03_dur07	double	Comp rate disc_diff07_rew03_dur07
rec_disc_diff10_rew02_dur10	double	Received disc_diff10_rew02_dur10
view_rate_disc_diff10_rew02_dur10	double	View rate disc_diff10_rew02_dur10
comp_rate_disc_diff10_rew02_dur10	double	Comp rate disc_diff10_rew02_dur10
rec_info_diff00_rew00_dur03	double	Received info_diff00_rew00_dur03
view_rate_info_diff00_rew00_dur03	double	View rate info_diff00_rew00_dur03
comp_rate_info_diff00_rew00_dur03	double	Comp rate info_diff00_rew00_dur03
rec_bogo_diff05_rew05_dur05	double	Received bogo_diff05_rew05_dur05
view_rate_bogo_diff05_rew05_dur05	double	View rate bogo_diff05_rew05_dur05
comp_rate_bogo_diff05_rew05_dur05	double	Comp rate bogo_diff05_rew05_dur05
rec_disc_diff10_rew02_dur07	double	Received disc_diff10_rew02_dur07
view_rate_disc_diff10_rew02_dur07	double	View rate disc_diff10_rew02_dur07
comp_rate_disc_diff10_rew02_dur07	double	Comp rate disc_diff10_rew02_dur07