Rotman

Master of Management Analytics

A GUIDE TO FINDING THE DIGI-AD PARTNER

By GMT+8

Rachel Goh

Bingtian He

Chun Wing Yuen

Pin Chien Wang

Rotman School of Management
UNIVERSITY OF TORONTO

November 9, 2022

Business Problem

How can MMA identify the manufacturers with national brands that have the capacity (revenue) and inclination to participate in their digital advertising scheme?

To answer the business problem, we broke it down to the following analytical questions

1. Which of those manufacturers have the budget to partner with MMA?
2. Which of these manufacturers have shown an inclination for marketing schemes?
3. Which of the manufacturers that are suitable to cater to the needs of the customer segments?
4. Who are the manufacturers we would recommend?

Data Sources & Methodology

Tools

R

Python

Thinkcell

Microsoft Excel

SAS Enterprise Miner

Models

Generalized Linear Regression

Cluster Analysis

Overview of Data Sources

0. Transaction/Demographic Dataset (Raw)

1. Manufacturer Dataset (6433*20)

Aggregated data with rows of unique manufacturer ID

Aggregated Variables	Description
STORE_NUM	Number of entered store
REACHED_CSTMR_NUM	Number of reached-out customers
TOTAL_REV	Total revenue
TOTAL_QTY_SOLD	Total quantity of product sold
UNIQ_PROD_NUM	Number of unique departments
UNIQ_DEPART_NUM	The department with the most number of products sold
….	……..

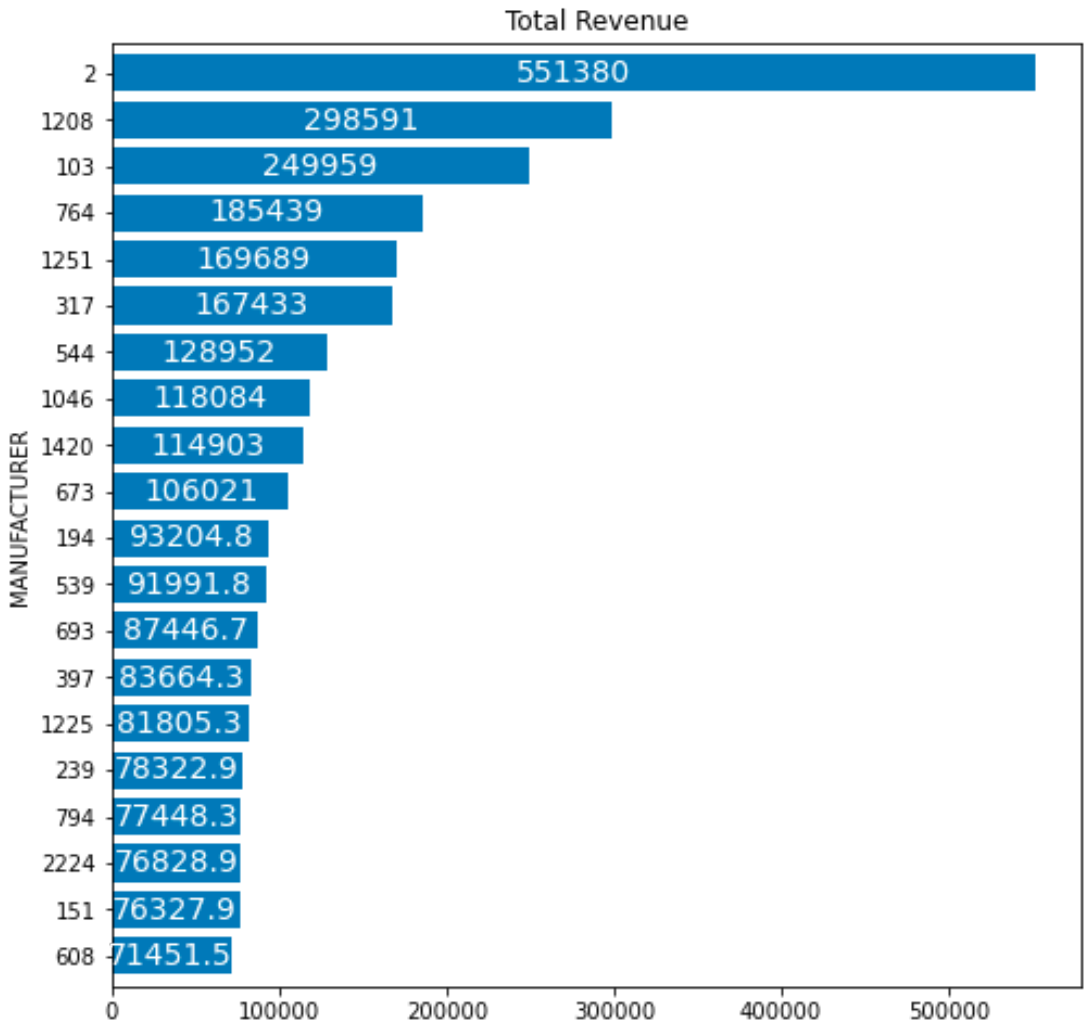
2. Customer Dataset (2500*13)

Aggregated data with rows of unique Customer ID

Aggregated Variables	Description
MEAN_BASKET_PAY	Mean payment for each purchase
MEAN_BASKET_QUANT	Mean number of products for each purchase
RECENCY	How many days until last day of purchase
FREQUENCY	Number of purchase within 2 years
MONETARY	Total purchase amount
….	……..

Which are the manufacturers that have the budget to partner with MMA?

From a total of 6433 manufacturers, we had narrowed down the Top 50 based on their revenue to approximate their financial capacity to participate in an extensive digital advertising campaign with MMA

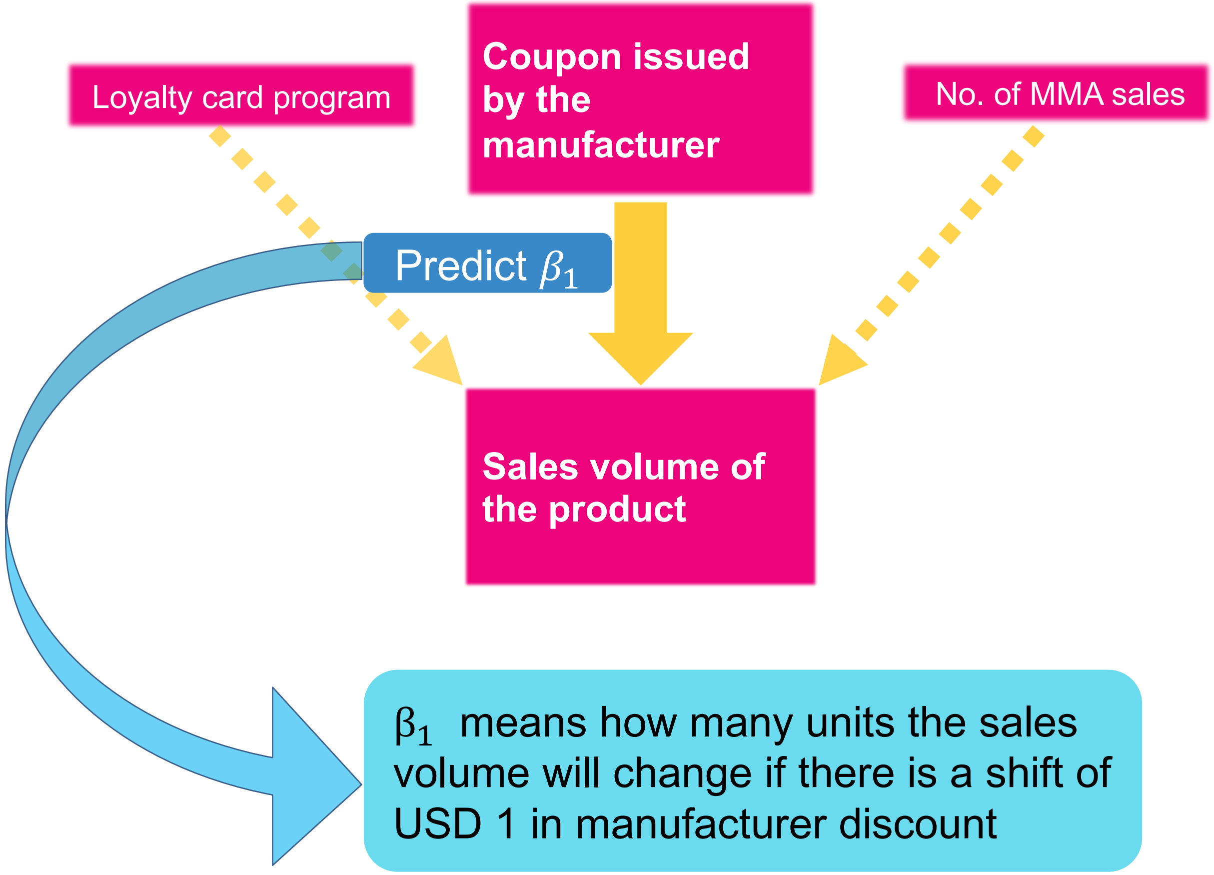


Ranking of highest budget manufacturers

Which of these manufacturers have shown an inclination for **marketing schemes**?

By running a model on the transaction data, coupon issued by the manufacturer and sales volume of the product for each manufacturer, also taking effect of discount of loyalty program and no. of MMA sales points into account

The model uses the variable coupon discount from the transaction data to proxy the manufacturers' attitude towards marketing and promotion.

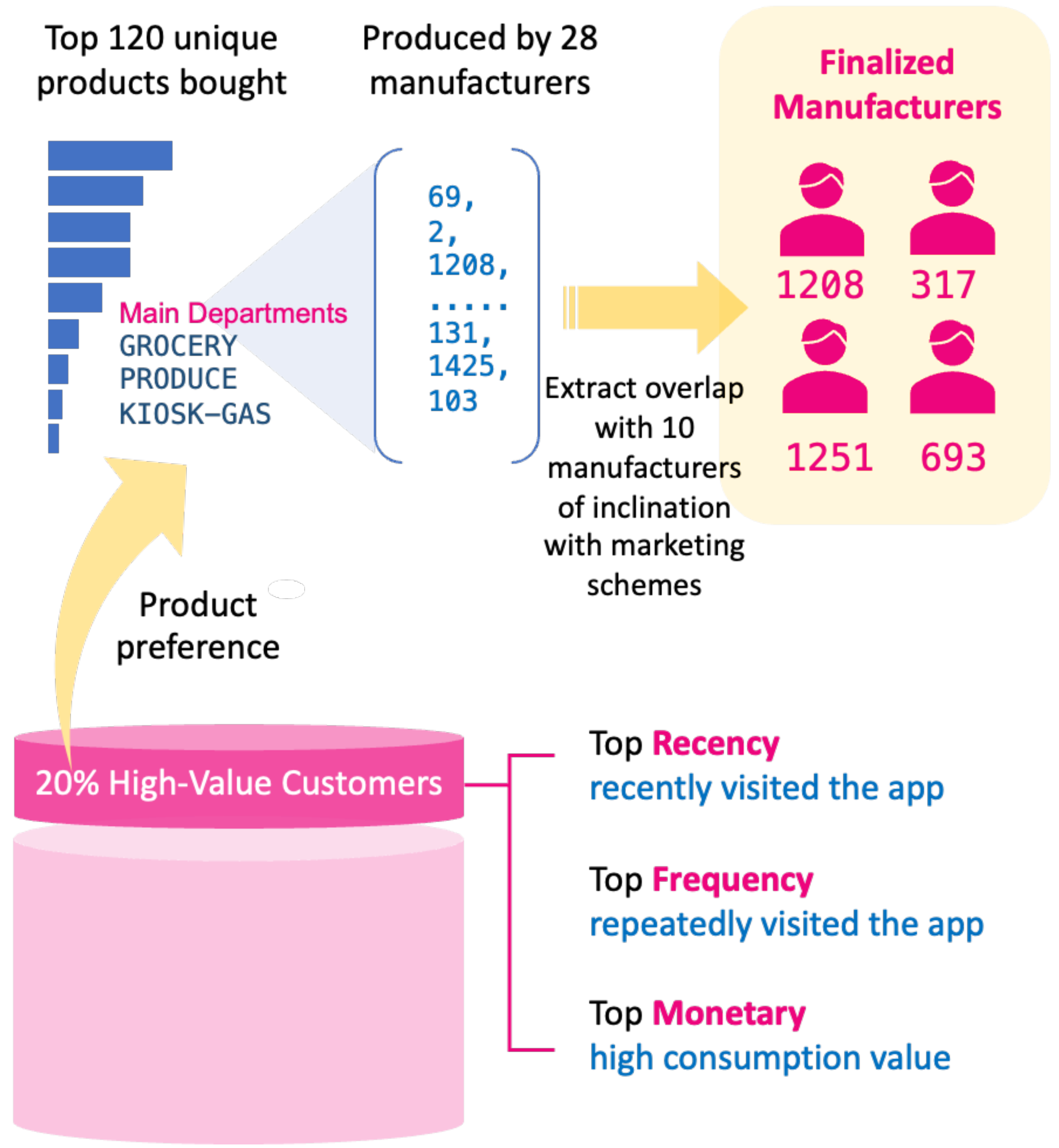


Manufacturer ID	β_1
Linear relationship	
1208	56.29
Quadratic Relationship	
1251	0.39
217	0.28
693	1.45

Which of these manufacturers are suitable to cater to the needs of the customer segments?

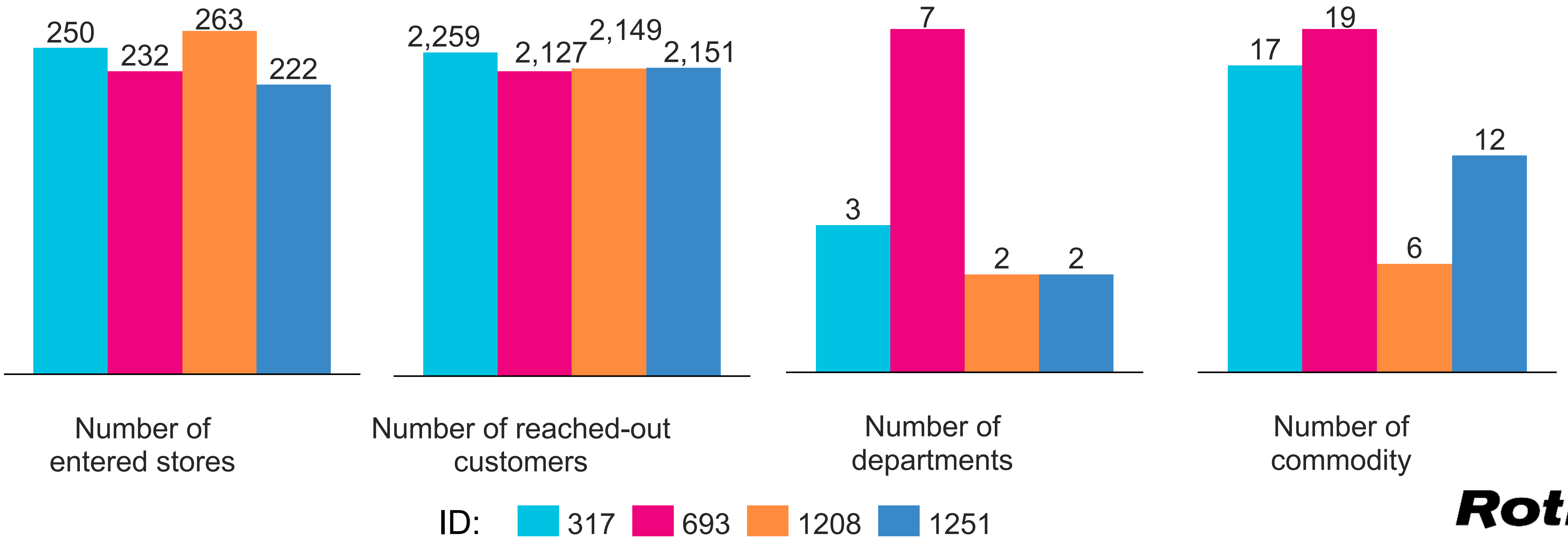
Defined the top 20% high-value customers using the RFM model and inspected their consuming preferences to find the most favored manufacturers.

Why High value?
The 80:20 Rule = 20% of the customers contribute to 80% of revenue!



WHO ARE THE FINAL MANUFACTURERS WE WOULD RECOMMEND?

- *Based on the previous analyses that had filtered the manufacturers out based on their financial capacity, their inclination towards participating in such marketing schemes and the suitability of their product offering to cater to customer needs, we had identified the four manufactures which we want to recommend, which are:*
- *Manufacturers with the following IDs: **317, 693, 1208, and 1251** respectively.*
- *Some of the key characteristics of these manufactures are displayed below:*
    - *Number of entered stores and number of reached-out customers convey the reach and influence of the manufacturers, with **1208 and 317** edging out the other two manufacturers*
    - *Number of departments provides insight into how prolific are the manufacturers' product offerings, with **693** being a clear leader and **317** second in place*
    - *Number of commodity supports the insight that **693** has a breadth of commodities to sell, with **317** being a close second*

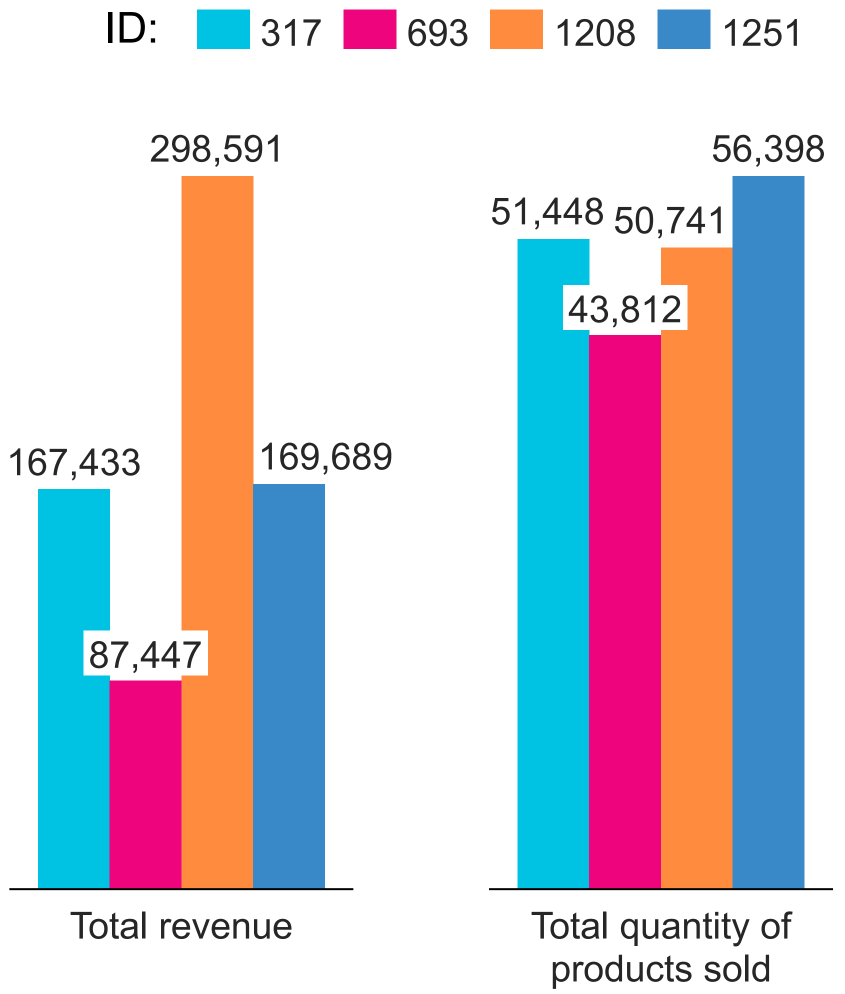


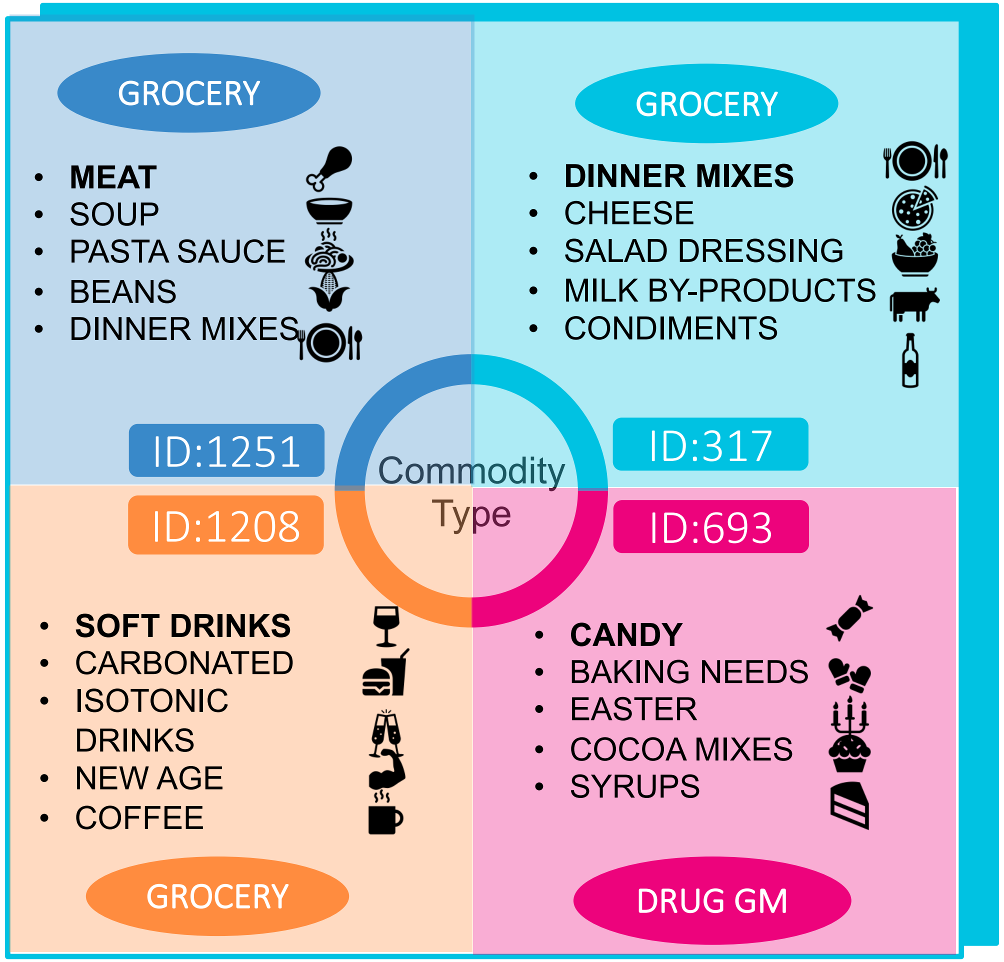
Rotman

MANUFACTURER ANALYSIS

THE DESCRIPTION OF FINAL MANUFACTURER WE FOUND

- What's more, we want tell you more details about the sales condition of four manufactures.





*The department types here are the main types.
*The commodity types here are the top5 main commodity.

Rotman

Customer Segmentation

*By segmenting customers from the demographic data, it is possible to identify **3 main customer segments** that frequent MMA grocery store, which enables MMA to personalize their digital ads*

"The Singles"

- No kids
- Spends larger amounts per purchase compared to other segments
- Not a frequent spender
- Largely from middle-income group

"The Family"

- Have at least 1 child
- Household size avg. 3-5
- Spends more and frequently
- Mostly people from 25-34, 35-44, 45-54

"The Middlings"

- Mostly without children, but some households have 1-2 children
- Spends lesser than other segments on purchases
- Consist of people from middle-income groups, with some under the middle-income threshold

RECOMMENDATIONS

Rotman

Partner with the following manufacturers with IDs: 1208, 317, 1251, 693 which have been singled out for their:

- *large trade promotion budget*
- *inclination for the exposure additional digital ad promises*
- *Existing manufacturing capacity of products that the customers usually buy*

Capitalize on the drugs (Candy, Syrup, Cocoa Mixes, etc.) as these are in high demand by customers and also command high margin

Personalize the advertisements by customer segments in the following priority:

- ***The Singles** – Primary Target Group, due to their inclination for high spending but low spending frequency. Ads targeted at this segment should be focused on frequency and personalized to motivate increased number of purchases to capture greater revenue.*
  - ✓ *Ex: promote dinner packages that is served for one and convenient to eat at home*
- ***The Fam** – Second Priority. Personalize family-oriented promotions on drugs (e.g., promotions that expound on the positive effect of a drug commodity like syrups has on family welfare).*
  - ✓ *Ex: boost sales via ad promotion in festivals, ex: easter goods, baking tools*
- ***The Middlings** – Tailor seasonal promotions or flash sales for this segment, which are composed primarily of middle-income **to** lower-income groups.*

EXPECTED OUTCOME

864.71 USD

Per manufacturer spend to participate in the digital ad program

283%

Percentage increase in expected revenue for manufacturers and MMA

Note: View Appendix for the model

APPENDIX

Subtitle

Here's
where it
changes.

Regression model

Outline

- The predictive analysis includes:
  - Considering the methodology
    - To decide on a set of candidate models (and list out the tuning parameters of each method)
    - To find the “best model” with the best tuning parameters by resampling method
  - Test-train split the available data
  - Use chosen model to make predictions
  - Calculate relevant metrics on the test data

Regression model

Overview

- Source data: transactions.csv
- Analytical tool: R
- Data processing
  - Filter for transactions of national manufacturers
  - Aggregate transaction data for each product
    - Sum of COUPON_MATCH_DISC
    - Sum of RETAIL_DISC
    - COUNT_STORE_ID
- Modelling
  - Generalized Linear Model (GLM)
  - Cross-validation
  - Link function
  - Assumption tests
  - Prediction

- Example (raw dataset)

Manufacturer	Product ID	QUANTITY	COUPON_MATCH_DISC	RETAIL_DISC	STORE_ID
4	76741	1	0.1	12	56494
6	15852	2	0.5	19	11022
6	15852	1	0.1	4	72744
1	54497	1	0.5	7	75695
7	29325	5	0.1	6	18208
6	98525	3	0.4	19	57438
8	58148	2	0	18	59317
9	43264	2	0	4	23295

- Example (processed dataset)

Manufacturer	Product no.	SUM_QUANTITY	SUM_COUPON_MATCH_DISC	SUM_RETAIL_DISC	COUNT_STORE_ID
4	1	1	0.1	12	1
6	2	6	1	42	3
1	1	1	0.5	7	1
7	1	5	0.1	6	1
8	1	2	0	18	1
9	1	2	0	4	1

Rotman

Regression

Model decision

- Fit a regression model to aggregate transaction data of all products for each manufacturer (6433 manufacturers in total)
  - To investigate the relationship of product sales of a product by the promotion expense and reach rate (number of stores that transact the products)

Dependent variables	Independent variable
SUM_QUANTITY($Y_{predict}$ – $Discrete$)	SUM_COUPON_MATCH_DISC (X_1 – $Continuous$) SUM_RETAIL_DISC (X_2 – $Continuous$) COUNT_STORE_ID (X_3 – $Discrete$)

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, where\ g(\mu) is\ the\ link\ function$$

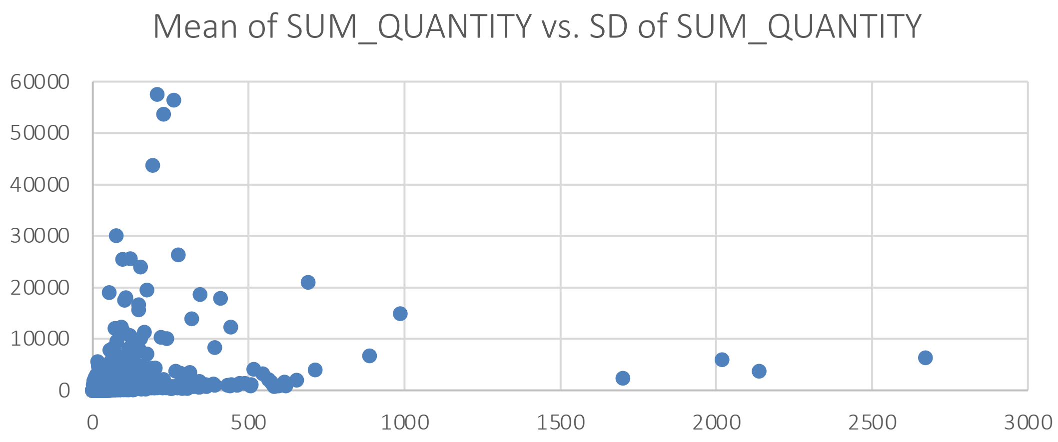
Model: Negative-binomial Generalized Linear Model

Justification:

- Count dependent variable
- Over-dispersion observed (non-linear relationship in the chart)

  → Poisson regression model (common for count data) will yield biased parameter estimation and underestimated standard error, leading to invalid conclusions

  → remediated by adding a multiplicative random effect θ to represent unobserved heterogeneity



Regression model

Modelling in R

- R is the analytical tool used for data management and data analysis of the case
- MASS and caret are the core packages that derives most of the metrics related to the regression model

Procedure:

Data import → GLM model fitting → Cross-validation → Best link function selection → Model output → Predictive analysis

- GLM model fitting: MASS.glm.nb is used to fit generalized linear models, the output is a description of the predictors and the error distribution
- Cross-validation: To partition the complete dataset into the training set and the testing set
  - 10-fold cross-validation: To partition the data into 10 subsets of similar size then conduct 10 times of training and validation.
  - In each validation, 9 subsets will be employed for model training, and the remaining one would be left for testing purpose. Repeat validation 10 times. Prediction error (NRMSE) will be generated by then
- Link function selection
  - The train function selects the link function which returns the best accuracy,
  - Link functions available for glm.nb on R:
    - Identity link $g(\mu) = \mu$
    - Square root link $g(\mu) = \sqrt{\mu}$
    - Log link $g(\mu) = \log(\mu)$ (Canonical link)

```
#Specify the parameters for cv <caret>
train_control <- trainControl(method = "cv", number=10)

#Fit the model with glm.nb <MASS>
m1 <- (try(glm.nb(sum_QUANTITY_manufacturer ~ sum_COUPON_MATCH_DISC_manufacturer
                  +sum_RETAIL_DISC_manufacturer   #Default log-link
                  +count_STORE_ID_manufacturer, data = data_final)))
#Train the model with cv and output the best model <caret>
model <- train(sum_QUANTITY_manufacturer ~., data = data_final,
               method = "glm.nb",
               trControl = train_control)
summary(m1)
print(model$finalModel)
```

Rotman

Regression model

Tests of the model

- Multiple tests will be placed against each model to check any breach of the assumptions
  - Variance inflation factor (VIF) for identify the degree of multicollinearity (Rule of thumb: Less than 5)
  - Likelihood test for statistical significance of Coupon discount invested as a predictor to the model
    - Run a model removing variable “coupon discount”, applying GLM (Negative binomial distribution)
    - Compare with the complete model (Significance: Less than 0.05)
  - Likelihood test for model assumption “the conditional means are not equal to the conditional variances”
    - Run a model remaining variables unchaged, applying GLM (Poisson distribution)
    - Compare with the complete model (Significance: Less than 0.05)

	Likelihood test For significance	VIF of the variables			Likelihood test for model assumption
Manufacturer	P-value	Coupon discount invested for the product X_1	Discount offered by Loyalty card program X_2	No. of stores completed transactions for the product X_3	P-value
1251	1.09E-08	1.33939487	1.50437781	1.83223613	0
1208	7.47E-06	1.09944852	1.23755729	1.34732379	0
317	0.04107905	1.78183421	1.9392302	1.44290026	0
693	0.00207167	1.00039269	2.03093391	2.03133569	0

Regression model

R output

Descriptive statistics

Manufacturer	Product no. (Sample size)	Sales volume of products: Sample mean	Sales volume of products: Sample S.D.	Confidence interval of t test: Lower	Confidence interval of t test: Upper
1251	685	82.3328467	259.599439	62.8579254	101.807768
1208	445	114.024719	536.924774	64.0020345	164.047404
317	669	76.9028401	278.341513	55.7727858	98.0328944
693	977	44.8433982	192.855487	32.7354177	56.9513786

- Variables
  - Y Total Sales volume of the product
  - X_1 Coupon discount invested for the product
  - X_2 Discount offered by Loyalty card program
  - X_3 No. of stores completed transactions for the product

Metrics of the regression models

		Coefficient estimate				Standard error				Z-value				P-value			
Manufacturer	Link function	(Intercept)	X_1	X_2	X_3	(Intercept)	X_1	X_2	X_3	(Intercept).	X_1	X_2	X_3	(Intercept)	X_1	X_2	X_3
1251	sqrt	1.2200	0.3854	(0.0007)	0.2419	0.0335	0.0958	0.0032	0.0050	36.3778	4.0226	(0.2331)	48.3320	9.54E-290	5.76E-05	0.81572166	0
1208	identity	(0.9859)	56.2850	0.5761	2.0107	0.1479	15.3035	0.0544	0.1030	(6.6660)	3.6779	10.5883	19.5257	2.63E-11	0.00023515	3.38E-26	6.64E-85
317	sqrt	1.1479	0.2777	0.0113	0.2118	0.0297	0.1544	0.0019	0.0040	38.6761	1.7984	5.8572	52.6319	0	0.07211838	4.71E-09	0
693	sqrt	1.2639	1.4477	0.0248	0.2234	0.0266	0.5851	0.0038	0.0044	47.5745	2.4742	6.4486	50.9938	0	0.01335255	1.13E-10	0

Regression model

Estimation

- The form of the model equation for negative binomial regression depends on the link function employed
  - Square root link function: $\sqrt{\mu} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$
  - Identity function: $\mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$
  - Canonical link: $\log_e \mu = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$

Manufacturer	Link function	Model equation
1251	sqrt	$\sqrt{\hat{Y}} = 1.219985602 + 0.385401092X_1 + -0.000744203X_2 + 0.241942128X_3$
1208	identity	$\hat{Y} = -0.985945863 + 56.28500803X_1 + 0.576124135X_2 + 2.010677011X_3$
317	sqrt	$\sqrt{\hat{Y}} = 1.147922227 + 0.277749251X_1 + 0.011319076X_2 + 0.211844968X_3$
693	sqrt	$\sqrt{\hat{Y}} = 1.263870995 + 1.447667153X_1 + 0.024815195X_2 + 0.223352136X_3$

Application Regression model

Sales quantity prediction

- Assume X_1 & X_3 & metrics related to revenue calculation remains constant
  - Predictor variable (Input) – Budget of coupon promotion to evenly allocate to the in-scope product
  - Distribution cost per coupon – $0.5[1]
  - Budget allocation for coupon promotion – 10% of total revenue[2]
  - Budget allocation for sponsorship to ad space – 10% of total revenue[2]

- ROI
  - Calculation: $$\frac{Revenue - COGS - cost\ of\ coupon\ promotion - sponsorship\ fee}{COGS + cost\ of\ coupon\ promotion + sponsorship\ fee}$$
  - ROI threshold set to 50% – No budget allocation to the products with calculated ROI less than 50%
  - COGS set to 40% – For ROI calculation

Sample data from 1 manufacturer (for demonstration only)

Estimated quantity	ROI estimation	Quantity increment	Estimated revenue	Product ID	X_2	X_3	Unit price
31	93%	29	263.19	992939	0	2	8.49
28	93%	26	237.72	1054115	-2	1	8.49
42	77%	35	264.18	941853	0	6	6.29
36	77%	4	226.44	1103550	-1.7	4	6.29
28	72%	26	153.72	278211	-5.5	1	5.49
164	67%	87	834.76	895702	-21.5	32	5.09
158	66%	113	788.42	868430	0	31	4.99
45	59%	33	224.55	5995638	-17	7	4.99
28	51%	27	106.12	9418274	-2.02	1	3.79
(omit calculation)	49% (not in-scope)	(omit calculation)	665.19	1087851	0	33	3.89
(omit calculation)	49% (not in-scope)	(omit calculation)	1147.55	904105	0	50	3.89
(omit calculation)	49% (not in-scope)	(omit calculation)	800.73	832007	-51.25	40	3.69

[1], B. of A. (2022, September 7). Cost Per Click (CPC) Rates 2022. Business of Apps. Retrieved October 17, 2022, from https://www.businessofapps.com/ads/cpc/research/cpc-rates/
[2] Lesonsky, R. (2019, July 9). How to Get the Most From Your Marketing Budget. U.S. Small Business Administration. Retrieved October 17, 2022, from https://www.sba.gov/blog/how-get-most-your-marketing-budget

Application Regression model

Sales quantity prediction

Sample data from manufacturer 1251

Total revenue	86,470.86
5% of revenue	4,323.54
Budget on coupon promotion (%)	10%
Budget on coupon promotion ($)	432.35

- Budget allocated to the products with ROI estimation >50%
- Calculation of budget allocated to each in-scope products:

$$Budget\ of\ each\ product = \frac{coupon\ promotion\ budget}{count\ of\ productswith\ ROI > 50\%}$$

- Plug budget of each product into model equation

$$\sqrt{\hat{Y}} = 1.219985602 + 0.385401092\beta_1 + -0.000744203\beta_2 + 0.241942128\beta_3$$

Sample data from 1 manufacturer (for demonstration only)

Estimated quantity	ROI estimation	Quantity increment	Estimated revenue	Product ID	X_2	X_3	Unit price
31	93%	29	263.19	992939	0	2	8.49
28	93%	26	237.72	1054115	-2	1	8.49
42	77%	35	264.18	941853	0	6	6.29
36	77%	4	226.44	1103550	-1.7	4	6.29
28	72%	26	153.72	278211	-5.5	1	5.49
164	67%	87	834.76	895702	-21.5	32	5.09
158	66%	113	788.42	868430	0	31	4.99
45	59%	33	224.55	5995638	-17	7	4.99
28	51%	27	106.12	9418274	-2.02	1	3.79
(omit calculation)	49% (not in-scope)	(omit calculation)	665.19	1087851	0	33	3.89
(omit calculation)	49% (not in-scope)	(omit calculation)	1147.55	904105	0	50	3.89
(omit calculation)	49% (not in-scope)	(omit calculation)	800.73	832007	-51.25	40	3.69

[1], B. of A. (2022, September 7). Cost Per Click (CPC) Rates 2022. Business of Apps. Retrieved October 17, 2022, from https://www.businessofapps.com/ads/cpc/research/cpc-rates/
[2] Lesonsky, R. (2019, July 9). How to Get the Most From Your Marketing Budget. U.S. Small Business Administration. Retrieved October 17, 2022, from https://www.sba.gov/blog/how-get-most-your-marketing-budget

Model for Expected Outcome

Input for outcome estimate

Link function	β_0	β_1	β_2	β_3
SQRT	1.2199856	0.38540109	-0.0007442	0.24194213
Total revenue	86,470.86		COGS (%)	40%
5% of revenue	4,323.54		ROI threshold	50%
Budget on coupon promotion (%)	10%		Budget on fee	10%
Budget on coupon promotion ($)	432.35		Fee	432.35
Coupon rate	0.5		Coupon no.	864
			Distribution rate (Unit fee to MM&A)	0.5

Results

Estimated revenue of product with ROI > threshold	13,052.56
Historical revenue of product with ROI > threshold	3,406
Percentage increase	283%
Investment on digital ad program	864.71

Rotman

Reference

- [1], B. of A. (2022, September 7). Cost Per Click (CPC) Rates 2022. Business of Apps. Retrieved October 17, 2022, from https://www.businessofapps.com/ads/cpc/research/cpc-rates/
- [2] Lesonsky, R. (2019, July 9). How to Get the Most From Your Marketing Budget. U.S. Small Business Administration. Retrieved October 17, 2022, from https://www.sba.gov/blog/how-get-most-your-marketing-budget