CUSTOMER PROFILE & BEHAVIOUR PREDICTION FOR TARGETED MARKETING

Alicja Mazur

São Paulo, 18 November 2020

AGENDA

PILOT CAMPAIGN
- SUMMARY
- CONCLUSIONS

CUSTOMER PROFILE DISCOVERY
- DATA AVAILABLE
- TWOFOLD APPROACH
- PILOT CAMPAIGN RESPONDENT PROFILE
- PROFILE GROUPS AMONG ALL CUSTOMERS

CUSTOMER BEHAVIOUR PREDICTION
- OBJECTIVE
- ASSUMPTIONS
- PREDICTIVE MODEL
- IMPORTANT QUESTIONS
- MODEL INTERPRETABILITY

KEY TAKEAWAYS
- QUESTIONS & DATA-DRIVEN ANSWERS

FUTURE WORK
- POSSIBLE IMPROVEMENTS

APPENDIX: PROFILE GROUPS AMONG CUSTOMERS
- FEATURE DISTRIBUTION ACROSS GROUPS

PILOT CAMPAIGN

PILOT CAMPAIGN

SUMMARY

Carried out on 2240 **randomly** selected customers.

- Campaign cost 6720 MU
- Campaign revenue 3674 MU
- **Campaign loss 3046 MU**

15%

success rate

PILOT CAMPAIGN

CONCLUSIONS

Objectives for the upcoming campaign:

1. Define the profile of campaign respondent based on socio-demographic data and customer purchase patterns

   → **MORE ACCURATE CAMPAIGN TARGETING**

2. Build a classifier to find all customers that are likely to respond to the upcoming campaign

   → **MAXIMIZE THE UPCOMING CAMPAIGN PROFIT**

CUSTOMER PROFILE DISCOVERY

CUSTOMER PROFILE DISCOVERY

DATA AVAILABLE

socio-demographic data

Age	Education	Marital_Status	Income	Kidhome	Teenhome
63	Graduation	Single	58138.0	0	0
66	Graduation	Single	46344.0	1	1

spending per product category

MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProds
635	88	546	172	88	88
11	1	6	2	1	6

purchase frequency per means

NumWebPurchases	NumCatalogPurchases	NumStorePurchases
8	10	4
1	1	2

other purchase behaviour

SinceEnrollment	Complain	Recency	NumDealsPurchases	NumWebVisitsMonth
2997	0	58	3	7
2447	0	38	2	5

response to previous campaigns

AcceptedCmp1	AcceptedCmp2	AcceptedCmp3	AcceptedCmp4	AcceptedCmp5
0	0	0	0	0
0	0	0	0	0

response to pilot campaign

PilotResponse
1
0

total campaign response

CampaignResponse
1
0

Categorical and binary data was not considered due to the algorithm constraints.

CUSTOMER PROFILE DISCOVERY

TWOFOLD APPROACH

- Exploratory data analysis: feature correlation with target variable

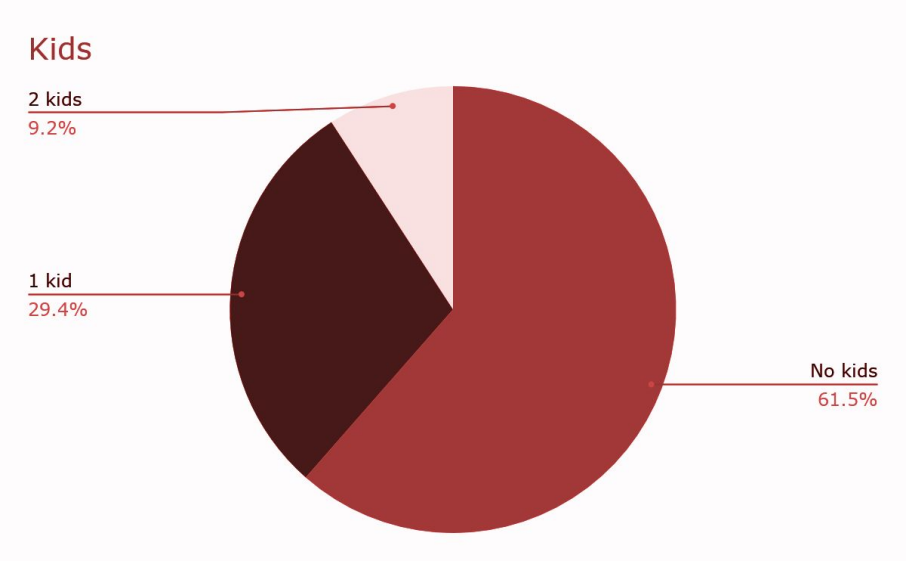
  → **DISCOVERY OF PILOT CAMPAIGN RESPONDENT PROFILE**

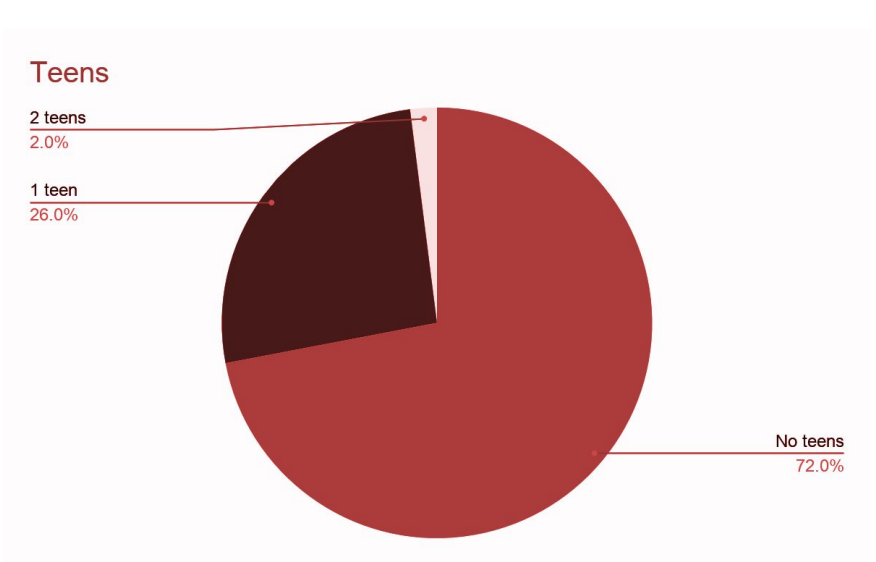
- Clustering machine learning model

  → **DISCOVERY OF CUSTOMER PROFILE GROUPS**

CUSTOMER PROFILE DISCOVERY

PILOT CAMPAIGN RESPONDENT PROFILE: OFFSPRINGS

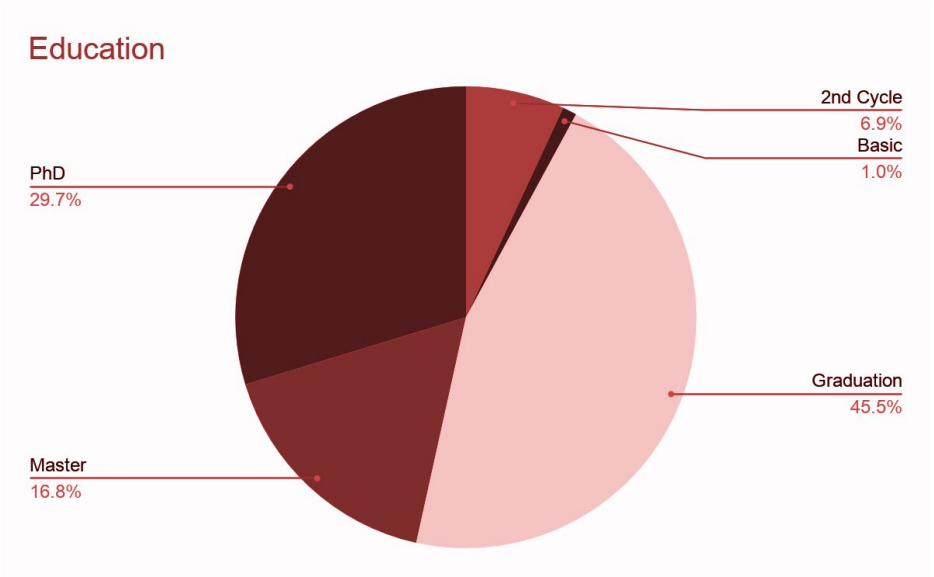




The majority of respondents do not have offsprings. Only 1 in 3 respondents have a young child and 1 in 4 have a teenager.

CUSTOMER PROFILE DISCOVERY

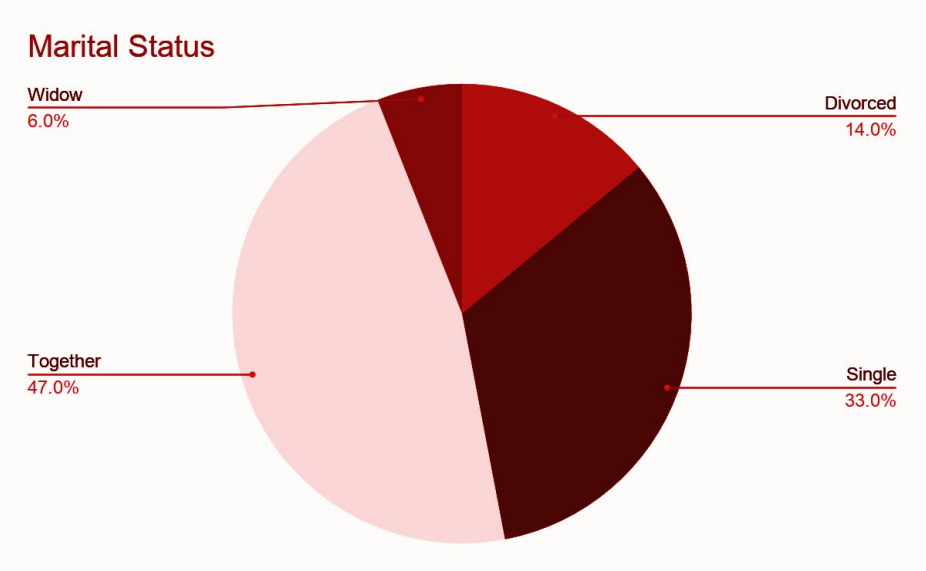
PILOT CAMPAIGN RESPONDENT PROFILE: EDUCATION



Customers with higher education, the majority have some academic degree. 1 in 3 have a PhD Degree and 1 in 5 have a Master Degree.

CUSTOMER PROFILE DISCOVERY

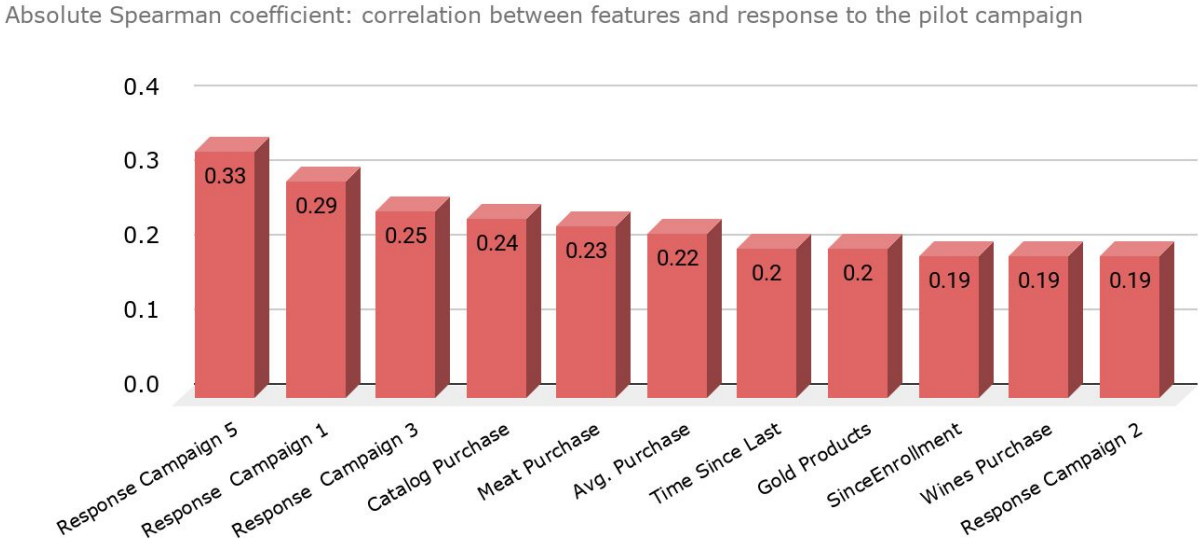
PILOT CAMPAIGN RESPONDENT PROFILE: MARITAL STATUS



Almost 50% of respondents are married or in couple. 1 in 3 is single.

CUSTOMER PROFILE DISCOVERY

PILOT CAMPAIGN RESPONDENT PROFILE: PURCHASE PATTERNS



Respondents are the old customers who responded to previous campaigns and made purchase recently. They demonstrated high spending on meat, wine, gold products and high purchase from the catalog.

CUSTOMER PROFILE DISCOVERY

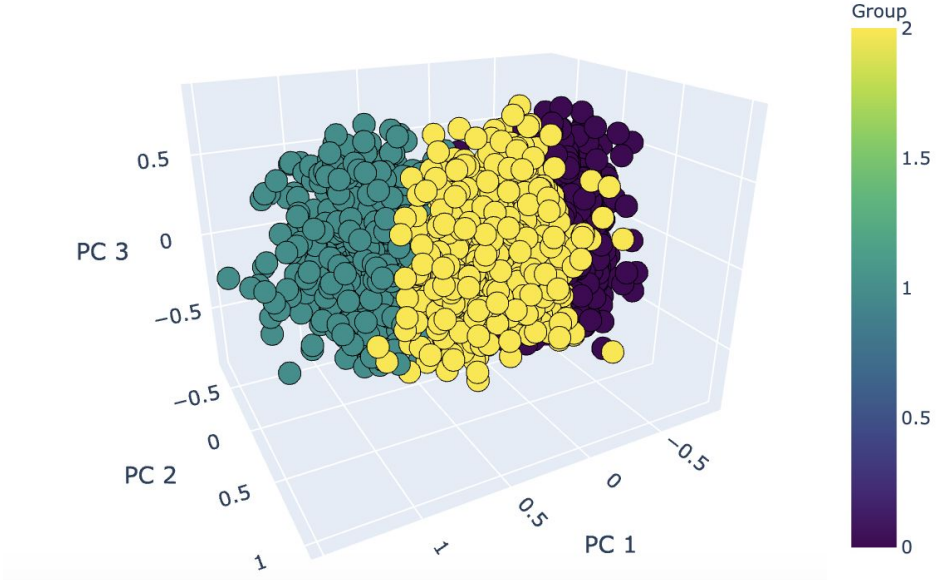
PROFILE GROUPS AMONG CUSTOMERS

GROUP 1 955 customers

GROUP 2 541 customers

GROUP 3 520 customers

Uniform distribution of marital status, history of complaints, recency (time since the last purchase) across groups. All groups include customers with some academic degree (over 80%).



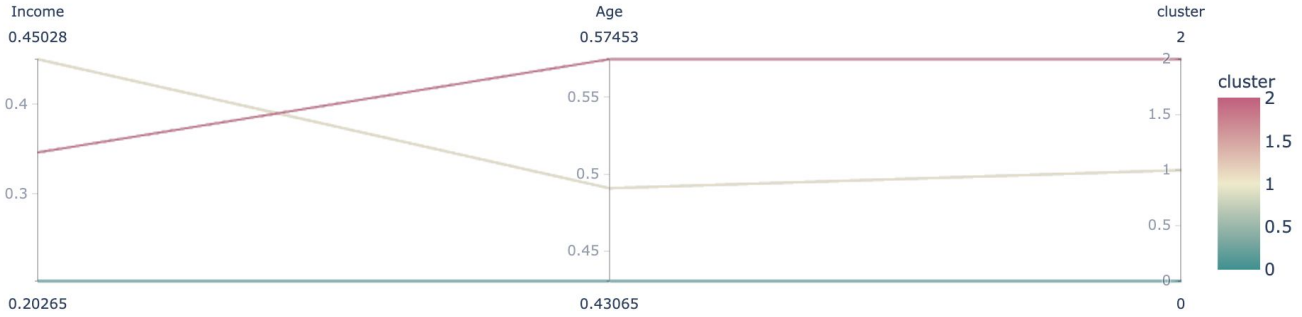
CUSTOMER PROFILE DISCOVERY

PROFILE GROUPS AMONG CUSTOMERS: CENTRAL TENDENCY

GROUP 1
Lowest income, youngest, lowest purchase

GROUP 2
Highest income, medium age, highest purchase

GROUP 3
Medium income, oldest, medium purchase



CUSTOMER PROFILE DISCOVERY

PROFILE GROUPS AMONG CUSTOMERS: CENTRAL TENDENCY

GROUP 1

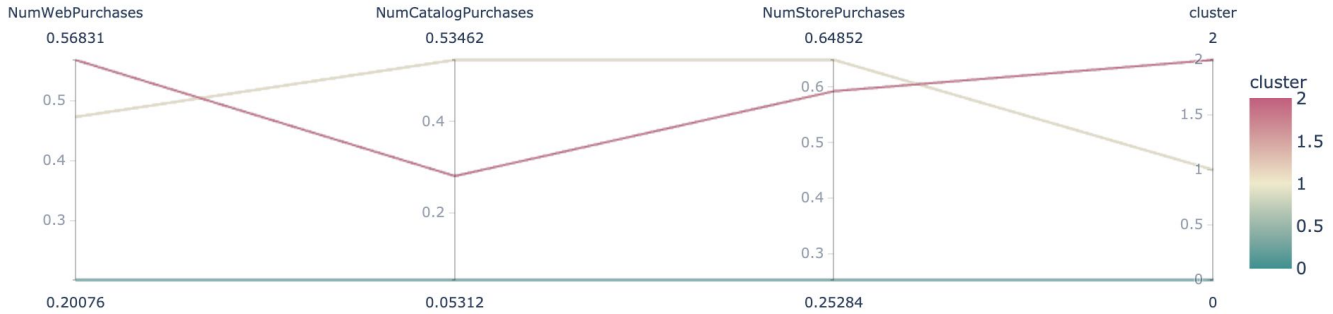
Lowest purchase, most recent customers, frequent web visits, some deal purchase

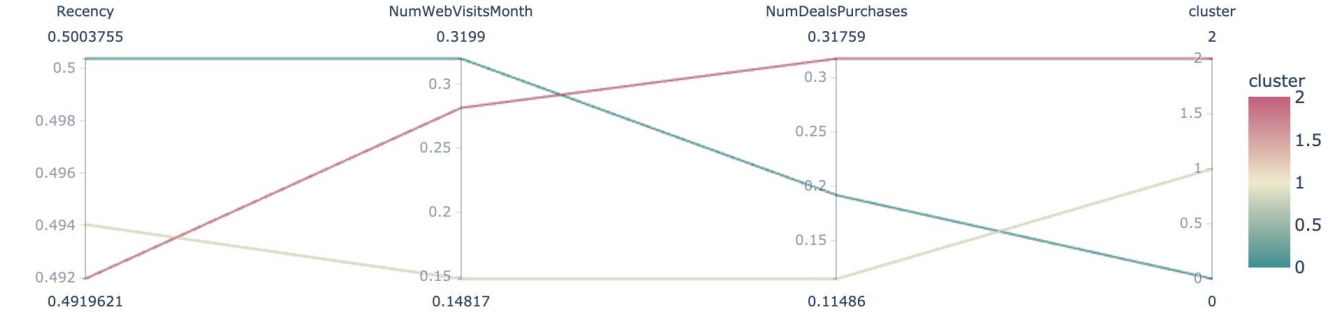
GROUP 2

High purchase, highest catalog and store purchase, not recent customers, low web visits, small deal purchase

GROUP 3

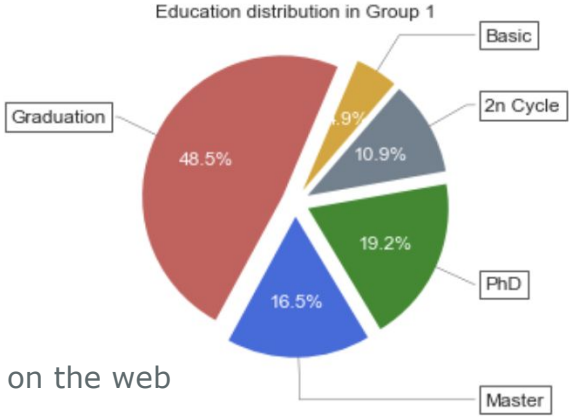
High web and store purchase, medium catalog purchase, not recent customers, some web visits, high deal purchase





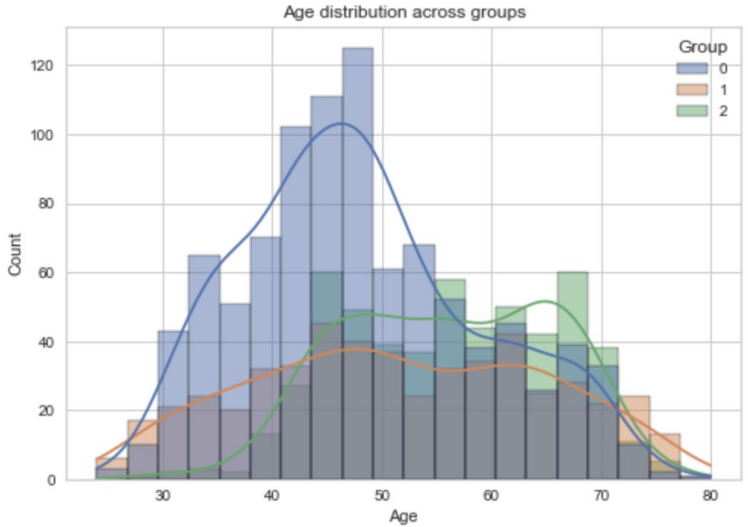
CUSTOMER PROFILE DISCOVERY

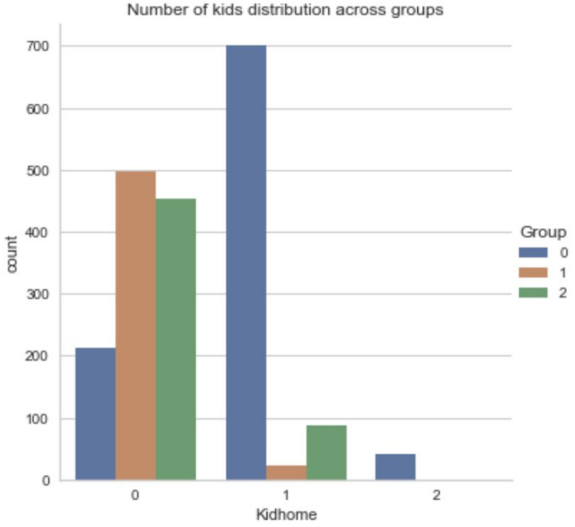
PROFILE GROUPS AMONG CUSTOMERS



GROUP 1

- the lowest age median
- present subgroup with basic education
- at least 1 kid, 50% have 1 teen
- the lowest income
- don’t spend much
- attracted by deals
- purchase either in store or on the web



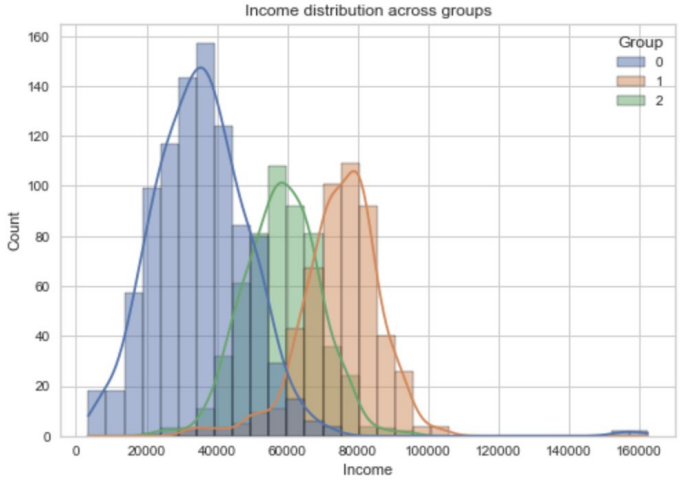


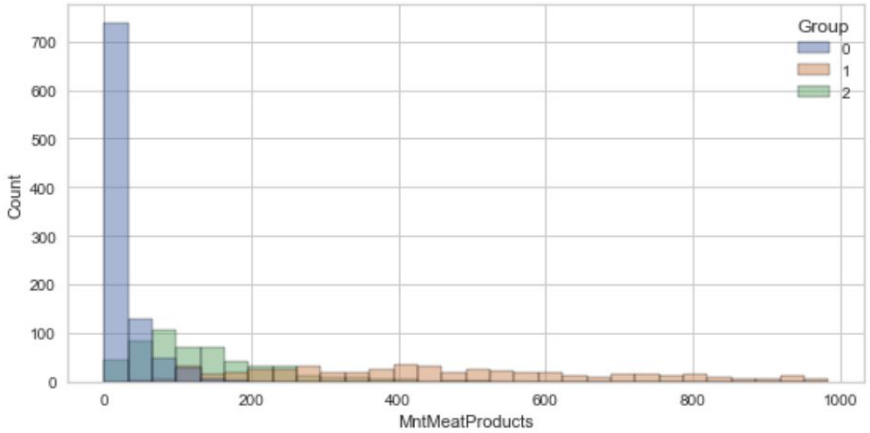
CUSTOMER PROFILE DISCOVERY

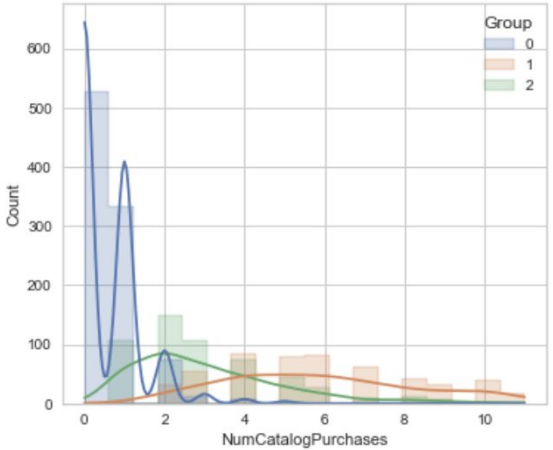
PROFILE GROUPS AMONG CUSTOMERS

GROUP 2

- no children
- the highest income
- high purchase of all products
- most likely to respond in the upcoming campaign
- often purchase from catalog
- don’t usually visit web
- not attracted by deals





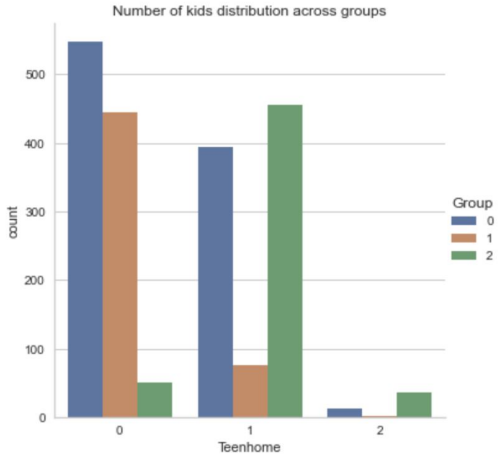


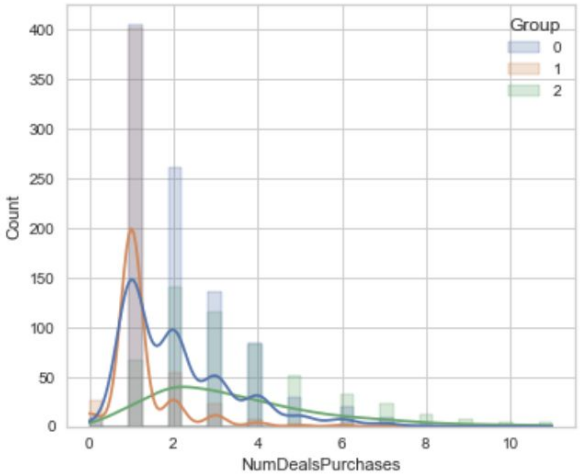
CUSTOMER PROFILE DISCOVERY

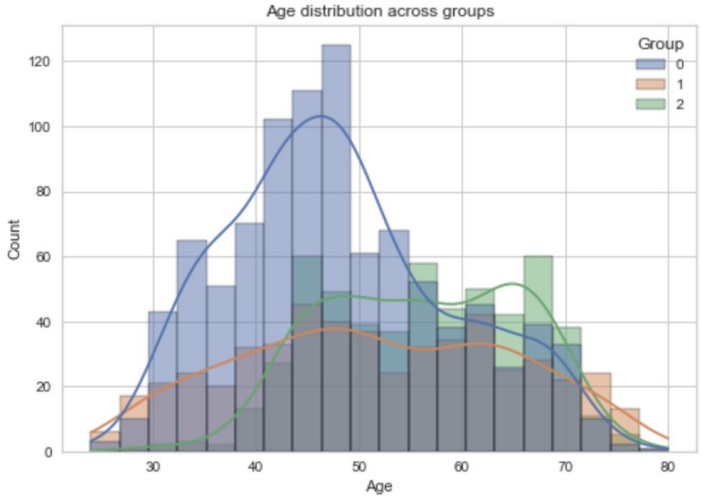
PROFILE GROUPS AMONG CUSTOMERS

GROUP 3

- the highest age median
- 1 in 4 has PhD
- have at least 1 teen
- medium income
- attracted by deals
- purchase mainly in store or on the web
- medium spending on meat, wine and gold products







CUSTOMER BEHAVIOUR PREDICTION

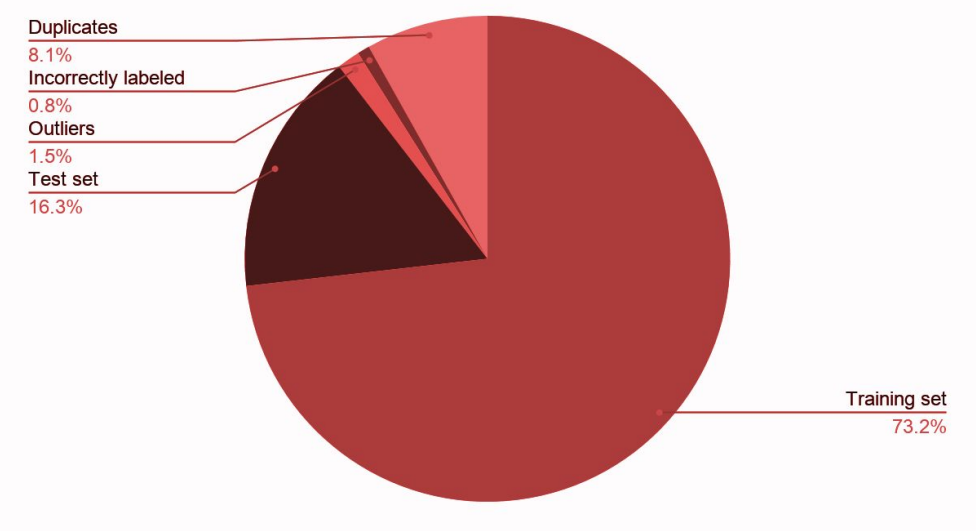
THE UPCOMING CAMPAIGN RESPONSE

CUSTOMER BEHAVIOUR PREDICTION

OBJECTIVE

MAXIMIZE THE CAMPAIGN PROFIT → predict which customers are likely to respond to the campaign

2016 useful records out of 2240 available.



CHALLENGE: imbalanced dataset

- **14% positive samples**
- 86% negative samples

CUSTOMER BEHAVIOUR PREDICTION

ASSUMPTIONS

Assumption:

the cost of FN = the cost of FP

respondents **incorrectly** classified as non-respondents

non-respondents **incorrectly** classified as respondents

- Imbalanced data problem: **accuracy as a measure of success is misleading**
- → Maximize the Area Under Curve (AUC): good balance between FN and FP

CUSTOMER BEHAVIOUR PREDICTION

PREDICTIVE MODEL

Algorithm

- **Decision tree ensemble:** gradient boosting algorithm (XGBoost) → **860 decision trees**
- Model tuning: random search with repeated stratified k-fold cross-validation
- Scaled positive class weight

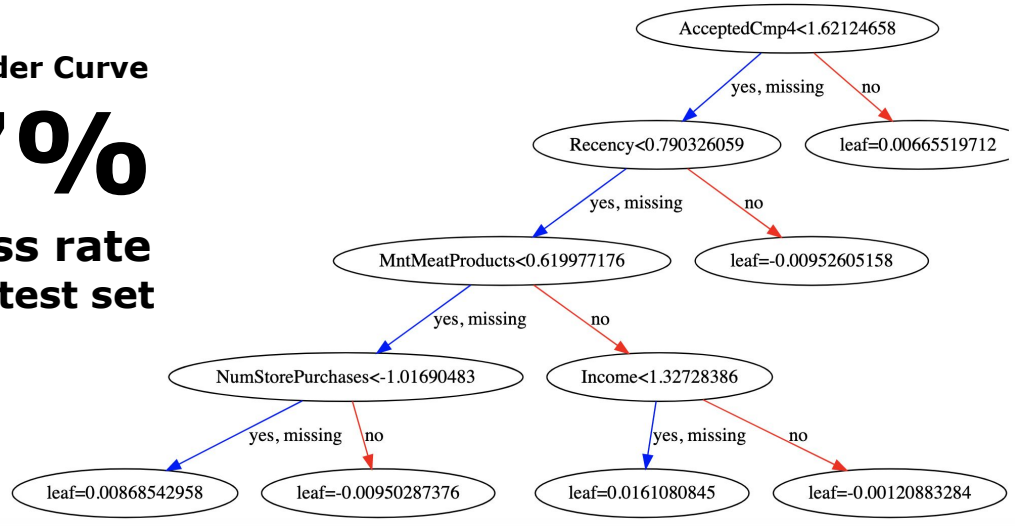
Confusion matrix:

True Negatives 294	False Positives 49
False Negatives 18	True Positives 43

Area Under Curve

87%

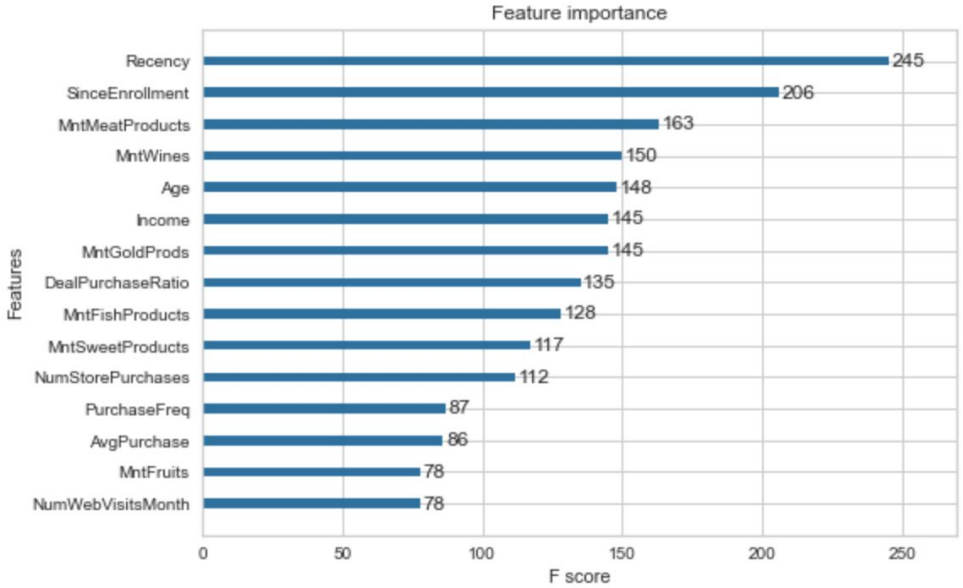
success rate on the test set



CUSTOMER BEHAVIOUR PREDICTION

IMPORTANT QUESTIONS

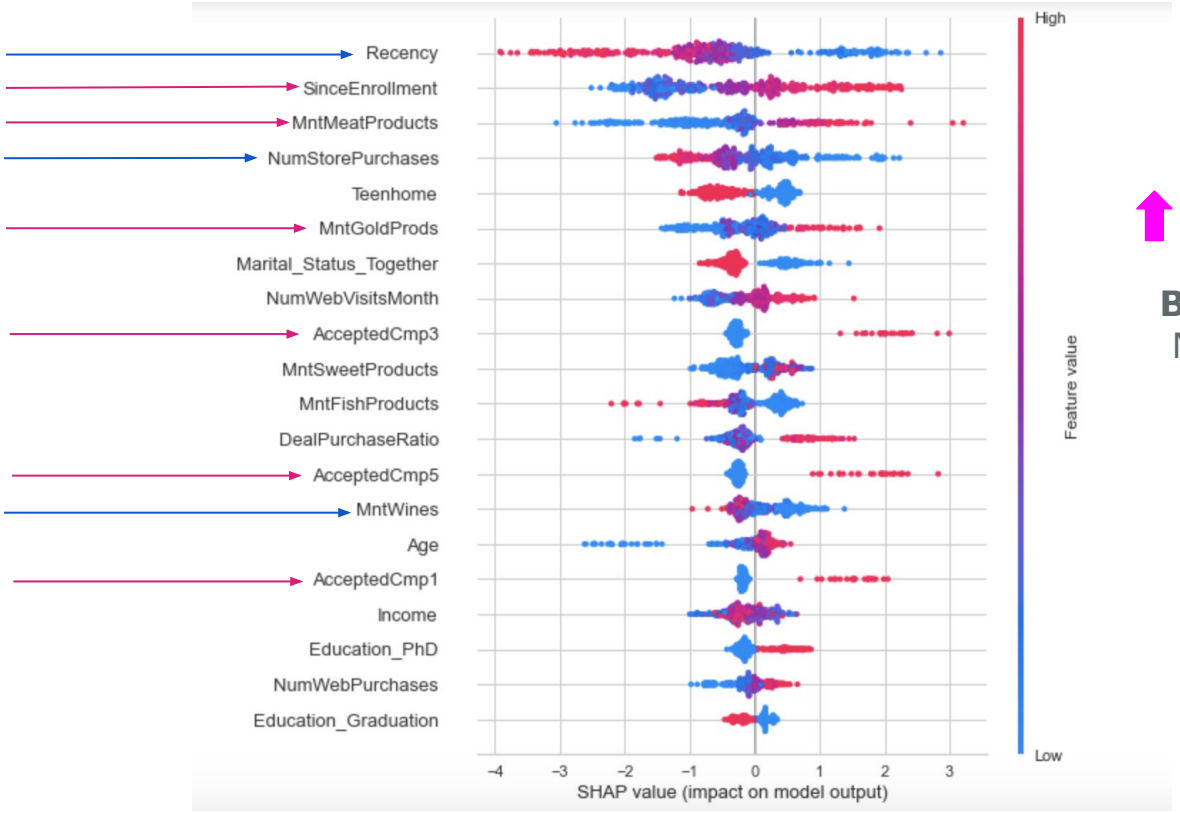
QUESTION: WILL THE CUSTOMER RESPOND TO THE UPCOMING CAMPAIGN?



- Are they our **old customer**?
- How long ago did they make their **last purchase**?
- How much **meat**, **wine**, **fish** and **gold products** do they purchase?
- What is their **age**?
- What is their **income**?
- Do they go after **deals**?

CUSTOMER BEHAVIOUR PREDICTION

INTERPRETATION OF MODEL PREDICTIONS: SHAPLEY VALUES



abs(SHAP VALUE)
=
BIGGER IMPACT ON MODEL PREDICTION

CUSTOMER BEHAVIOUR PREDICTION

INTERPRETATION OF MODEL PREDICTIONS WITH SHAPLEY VALUES - EXAMPLE

TEST SAMPLE:

Education_Graduation	Marital_Status_Together	Age	Income	Kidhome	Teenhome	SinceEnrollment	Recency
1.00	1.00	76.00	80589.00	0.00	0.00	2492.00	25.00

MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProduct	MntGoldProds	PurchaseFreq	DealPurchaseRatio	AvgPurchase
507.00	93.00	520.00	242.00	40.0	26.00	0.01	0.05	66.76

NumWebPurchases	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth
5.00	10.00	5.00	1.00

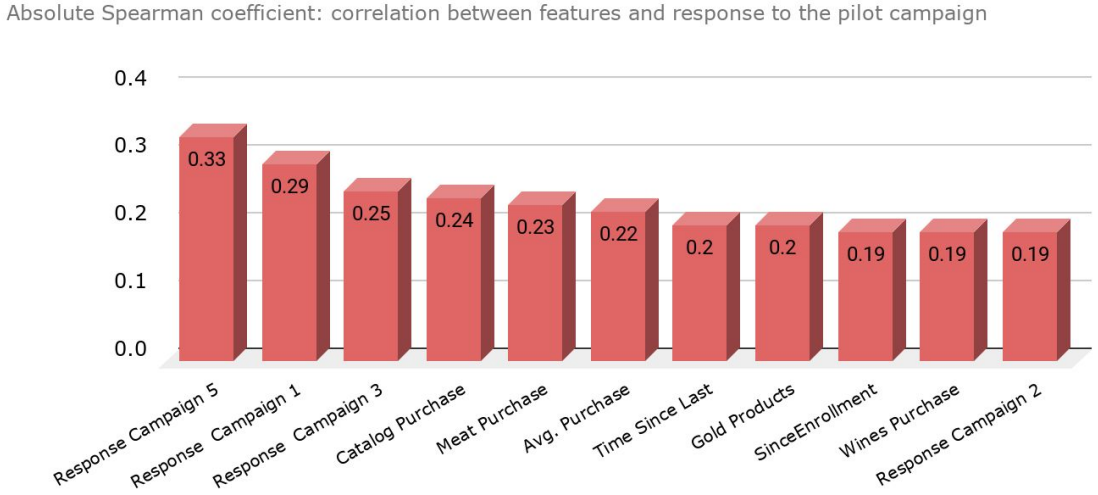
AcceptedCmp3	AcceptedCmp4	AcceptedCmp5	AcceptedCmp1	AcceptedCmp2	Complain	Group
0.00	0.00	0.00	1.00	0.00	0.00	1.00

GROUND TRUTH: 1
PREDICTION: 1

CUSTOMER BEHAVIOUR PREDICTION

INTERPRETATION OF MODEL PREDICTIONS

THE MODEL LEARNED TO IDENTIFY THE PILOT CAMPAIGN RESPONDENT PROFILE DISCOVERED DURING THE PRIOR ANALYSIS



Respondents are the old customers who responded to previous campaigns and made purchase recently. They demonstrated high spending on meat, wine, gold products and high purchase from the catalog.

KEY TAKEAWAYS

QUESTIONS AND DATA-DRIVEN ANSWERS

- **WHAT IS THE PROFILE OF CUSTOMERS THAT ARE MOST LIKELY TO PURCHASE THE NEW PRODUCT?**
  Target the upcoming campaign mainly to our old and recent customers who purchase significant amount of meat, wines, gold products and often purchase from the catalog.

- **HOW TO FIND THESE CUSTOMERS?**
  Leverage the predictive machine learning model for more effective choice of customers.

- **HOW TO BEST APPROACH THESE CUSTOMERS?**
  Leverage the knowledge on customer profile for more personalized campaign approach, e.g.:
  Group 1: target audience → under 50s, relate to children, focus on deals
  Group 2: focus on catalog offer, premium products
  Group 3: target audience → over 40s, focus on deals, relate to older offspring

FUTURE WORK

POSSIBLE IMPROVEMENTS

CUSTOMER PROFILE DISCOVERY

- Extend analysis of clusters to all customer database to better understand customer profile
- Collect data on customer purchase pattern in time

CUSTOMER BEHAVIOUR PREDICTION

- Consider the true cost of erroneous classification (FP and FN) and choose an adequate metric for the optimization problem
- Further model tuning to increase the success rate

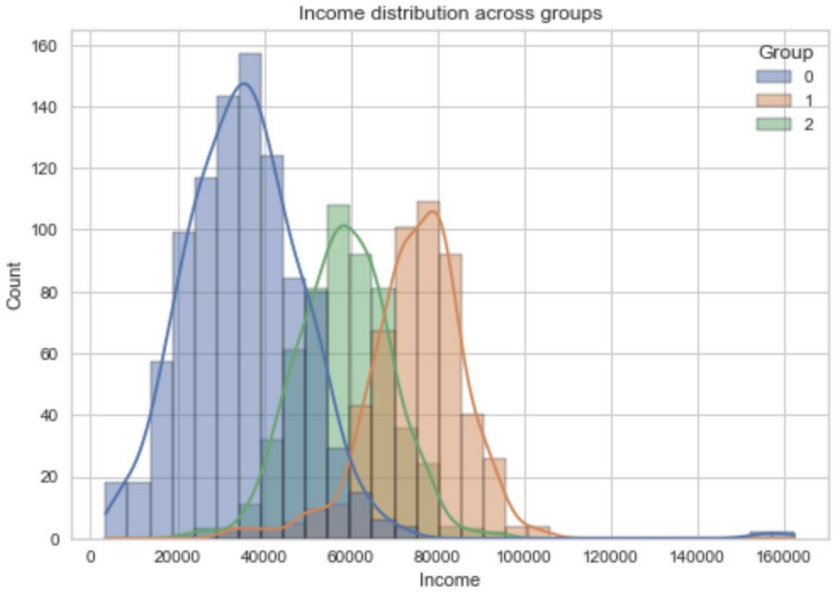
APPENDIX:

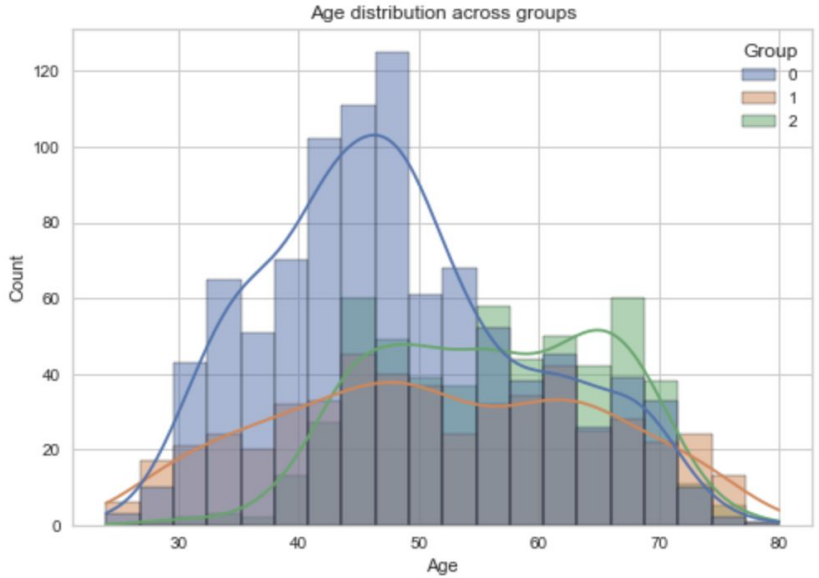
PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS

PROFILE GROUPS AMONG CUSTOMERS

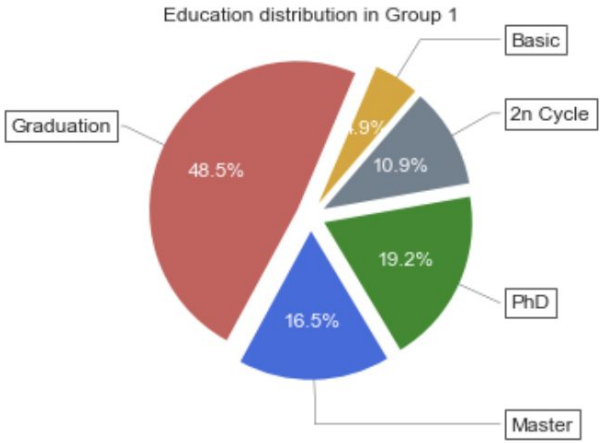
FEATURE DISTRIBUTION ACROSS GROUPS: INCOME AND AGE

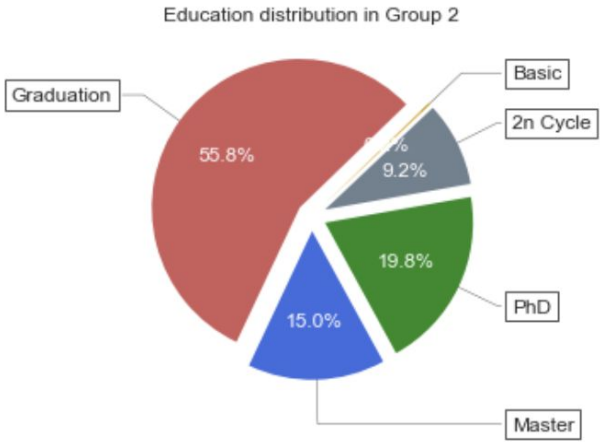


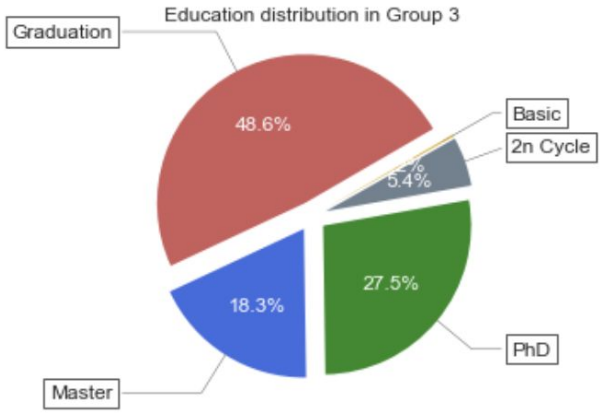


PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: EDUCATION

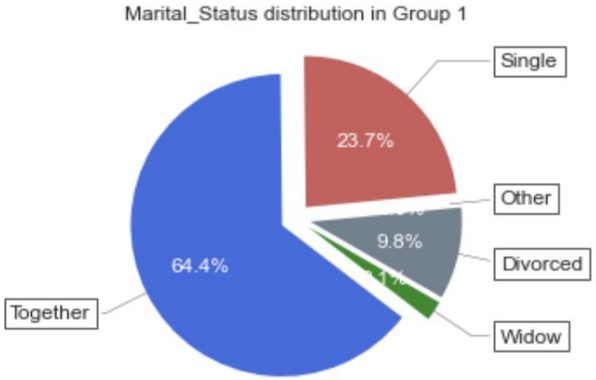


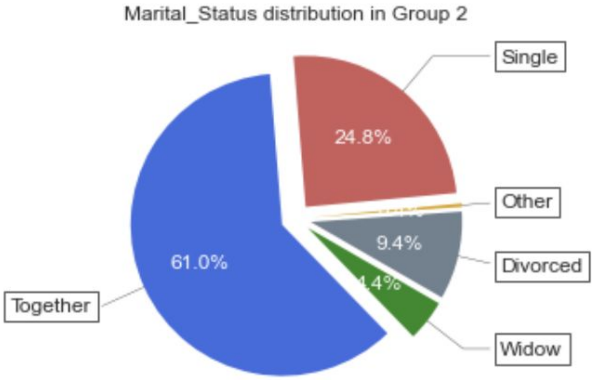


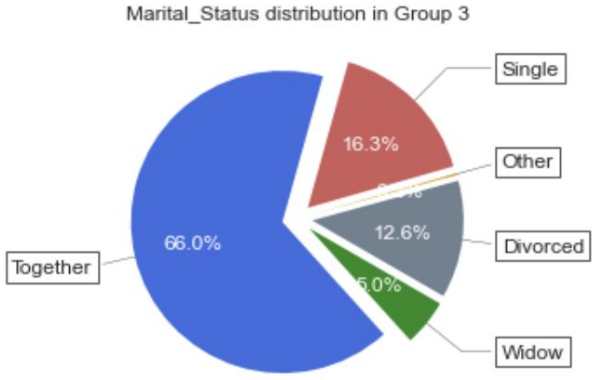


PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: MARITAL STATUS

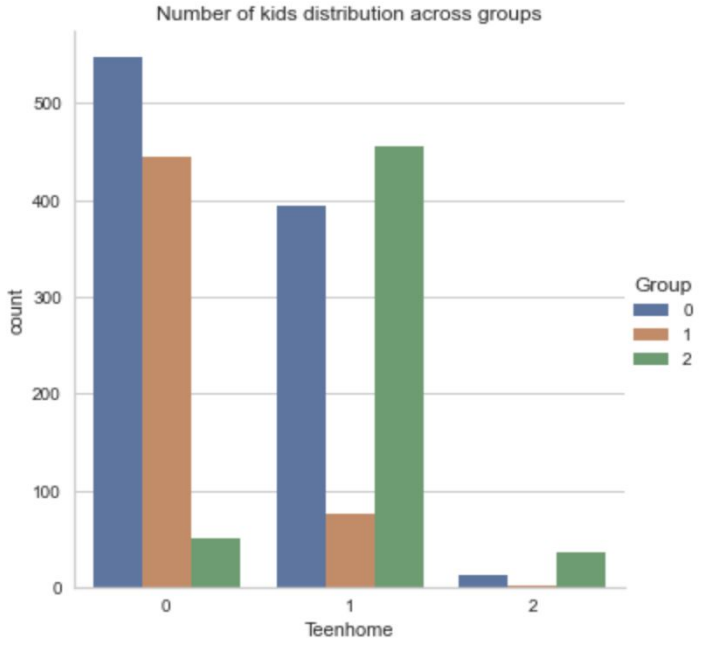


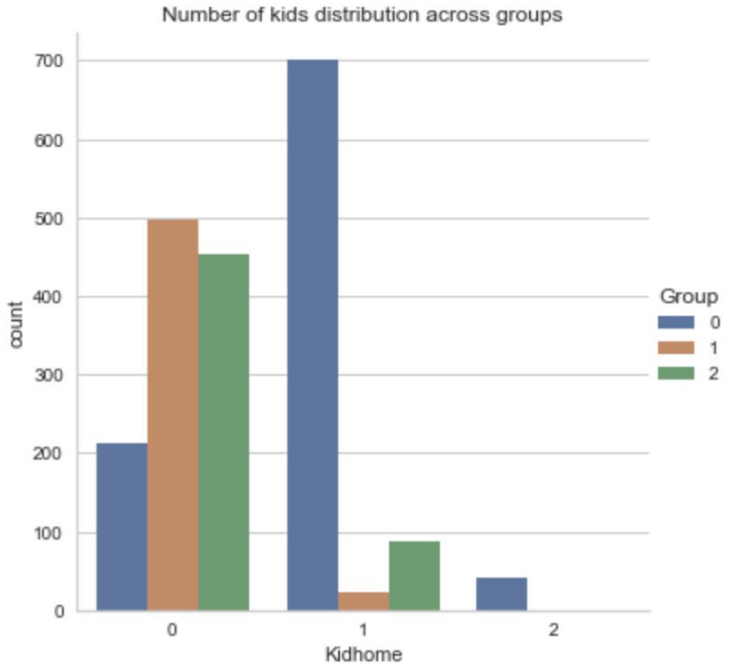




PROFILE GROUPS AMONG CUSTOMERS

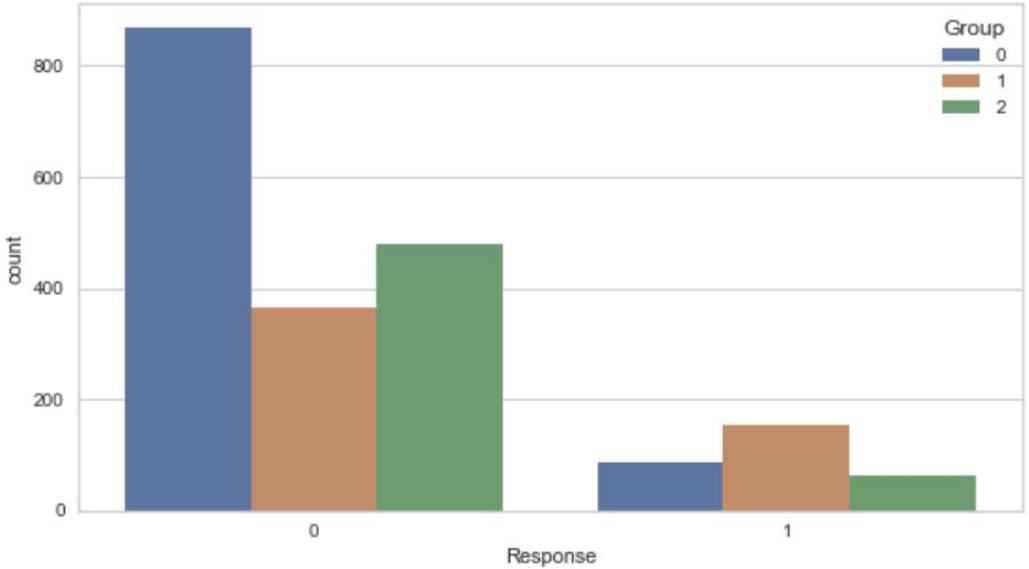
FEATURE DISTRIBUTION ACROSS GROUPS: OFFSPRINGS





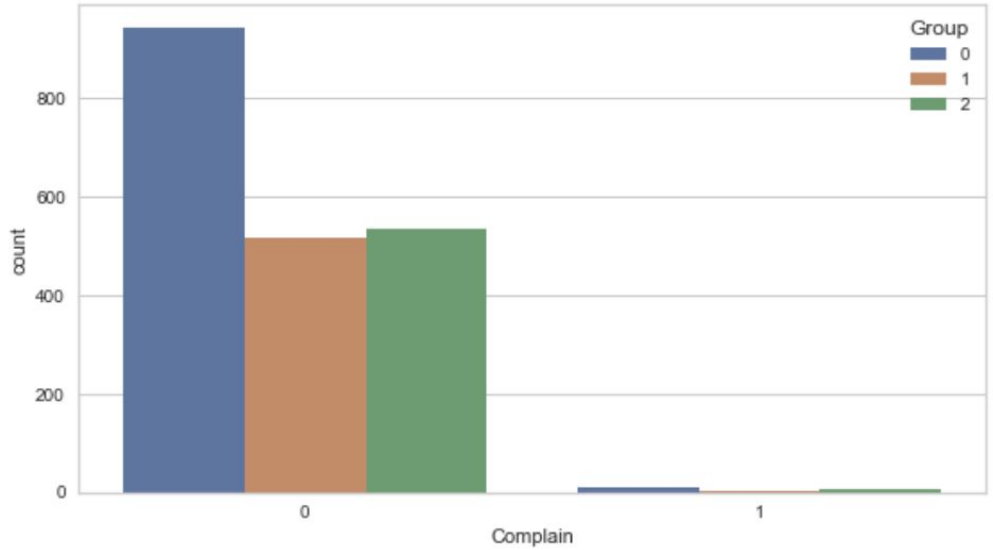
PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: PILOT CAMPAIGN RESPONSE



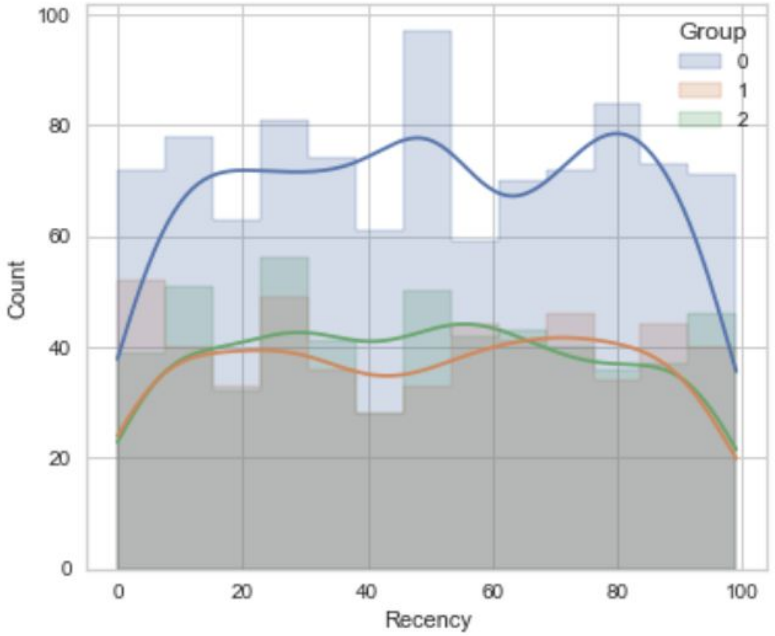
PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: HISTORY OF COMPLAINTS



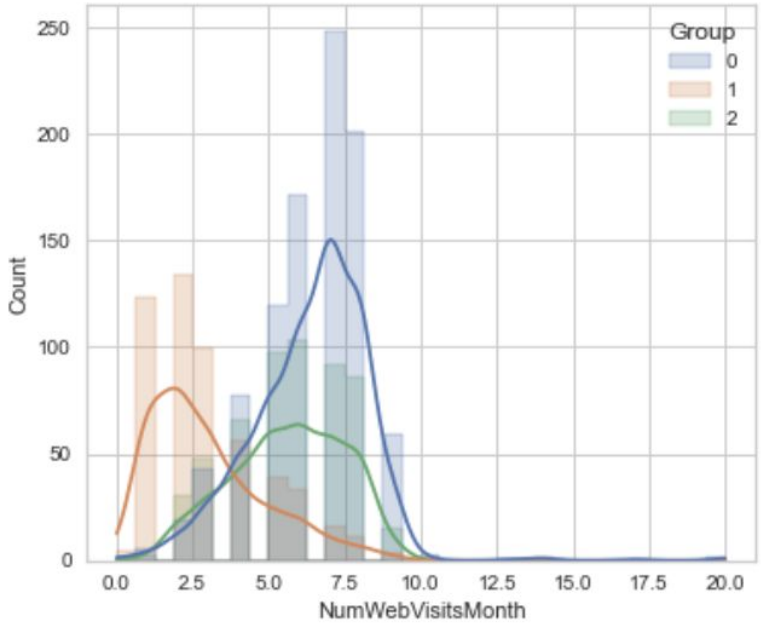
PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: TIME SINCE THE LAST PURCHASE



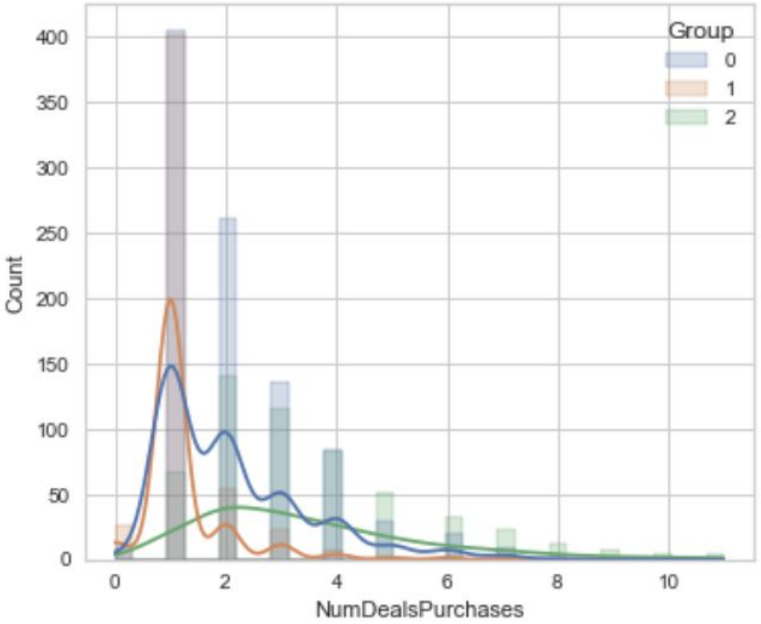
PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: WEB VISITS PER MONTH



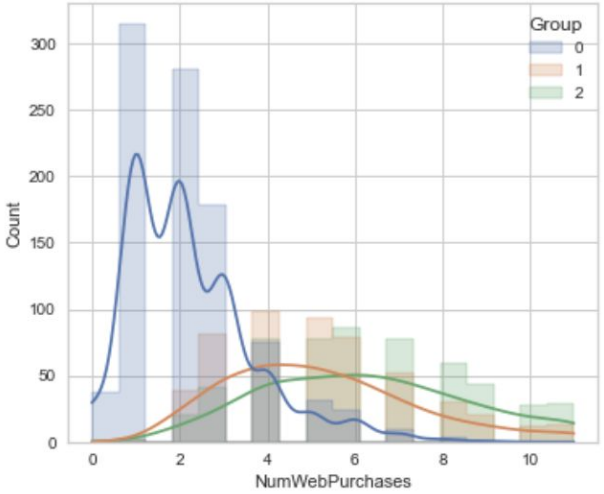
PROFILE GROUPS AMONG CUSTOMERS

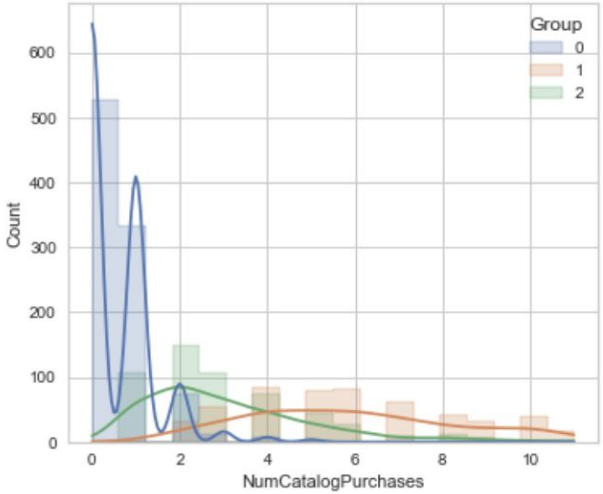
FEATURE DISTRIBUTION ACROSS GROUPS: DEALS PURCHASE

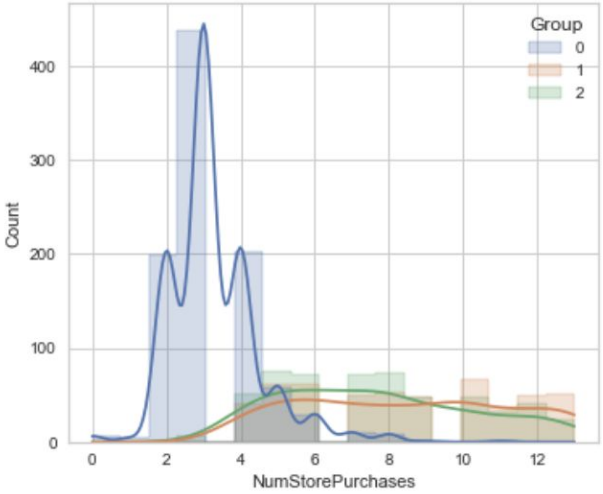


PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: MEANS OF PURCHASE

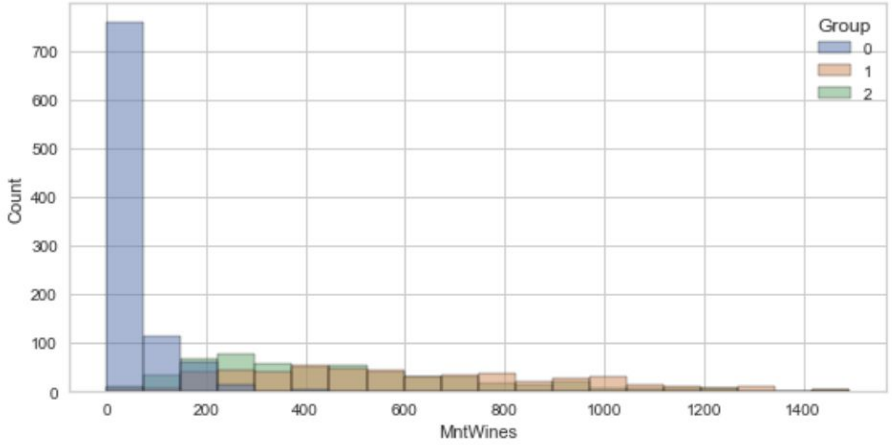


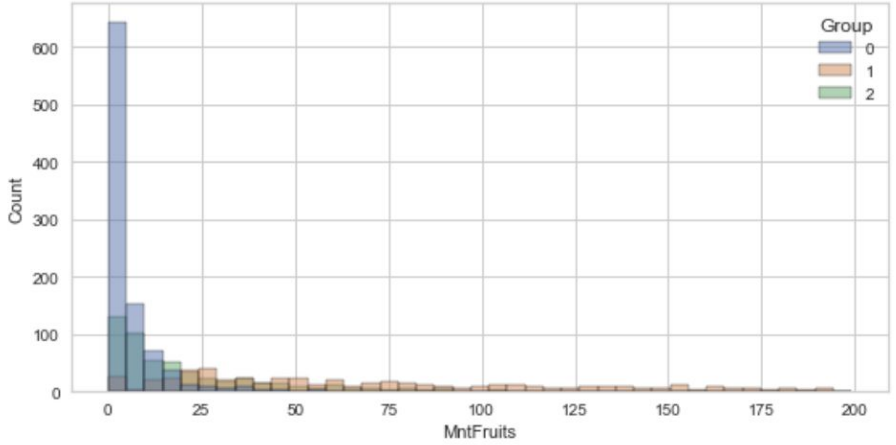




PROFILE GROUPS AMONG CUSTOMERS

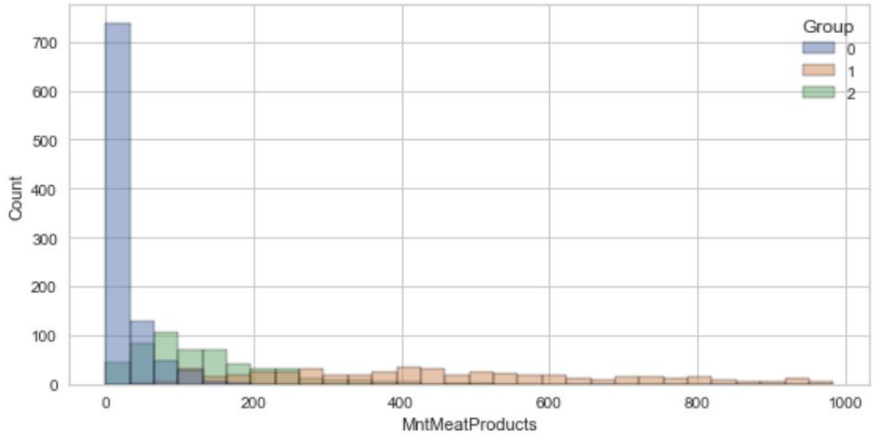
FEATURE DISTRIBUTION ACROSS GROUPS: WINES AND FRUITS PURCHASE

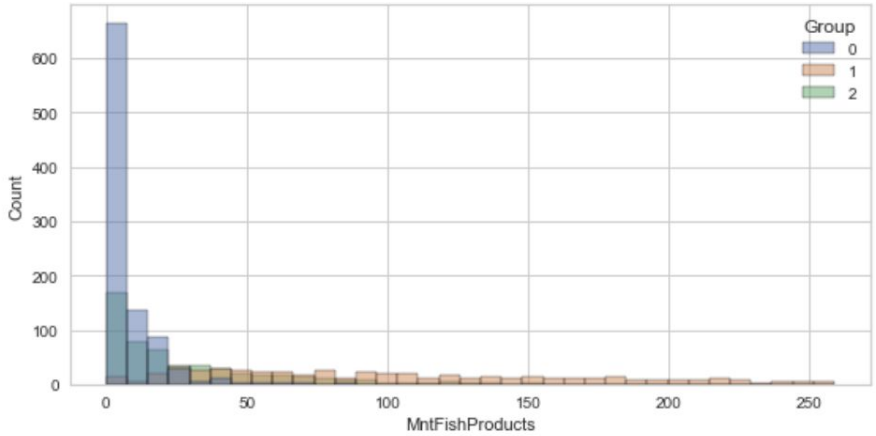




PROFILE GROUPS AMONG CUSTOMERS

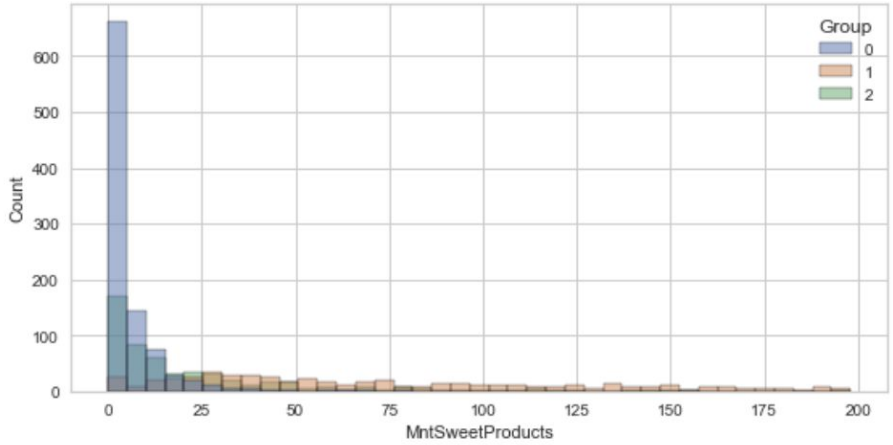
FEATURE DISTRIBUTION ACROSS GROUPS: MEAT AND FISH PURCHASE

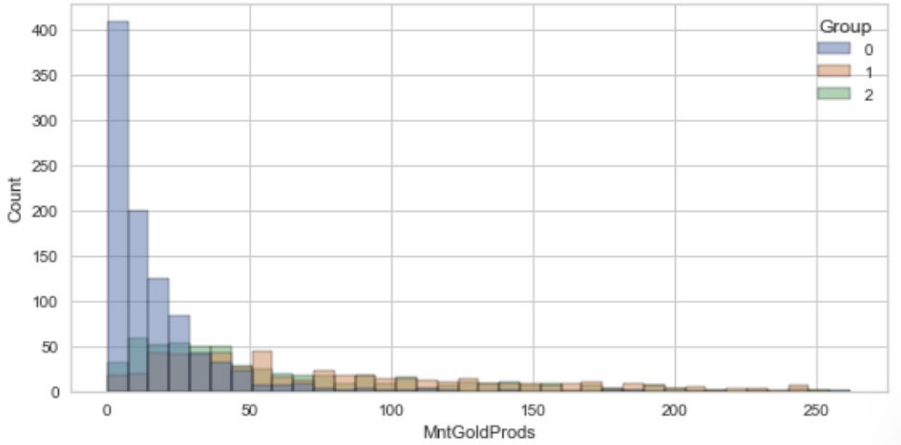




PROFILE GROUPS AMONG CUSTOMERS

FEATURE DISTRIBUTION ACROSS GROUPS: SWEETS AND GOLD PRODUCTS PURCHASE





Thank you for your feedback!

mazur.alicja4@gmail.com

linkedin.com/in/alicjamazur

+55 (11) 99539-6992

github.com/molly-moon/projects