FORECASTING FUTURE SPENDS & CAMPAIGN CONVERSION OF IFOOD CUSTOMERS

Course 52500 : Marketing Analytics

Group 3

Sankarsan Gautam

Juan Martin Cadena

Keerthana Nemili

Carlyse Wallace

Udyog Pati

PURDUE UNIVERSITY

Motivation for Pursuing the Problem

- Most consumer-driven businesses increasingly face the challenge of understanding the data they collect, in order to find business opportunities & actionable insights.

- They use this knowledge to drive strategy such as optimizing campaigns results and increasing customer spends on their platform to generate value to the company.

- We chose a real-life dataset shared by a food delivery app in Brazil, with the intention of helping it better understand its customer segments and create targeted strategies for different segments.

Purpose & Context

Our purpose is to help our client which is a growing Food Delivery company, with the following 3 goals:

- To forecast and quantify the total revenue a new customer will generate on the platform over the next 2 years.
- To determine lucrative customer segments, allow the client to optimize and redirect its revenue strategy towards this its most valuable customers.
- To predict which customers will convert in the client's next campaign
- To achieve this, we have been provided customer level information for 2204 customers over the past 2 years, including their responses to the last 5 campaigns

PURDUE UNIVERSITY®

Data Description

People

ID: Customer's unique identifier
Year_Birth: Customer's birth year
Education: Customer's education level
Marital_Status: Customer's marital status
Income: Customer's yearly household income
Kidhome: Number of children in customer's household
Teenhome: Number of teenagers in customer's household
Dt_Customer: Date of customer's enrollment with the company
Recency: Number of days since customer's last purchase
Complain: 1 if customer complained in the last 2 years, 0 otherwise
Country: Customer's location

Products

MntWines: Amount spent on wine in the last 2 years
MntFruits: Amount spent on fruits in the last 2 years
MntMeatProducts: Amount spent on meat in the last 2 years
MntFishProducts: Amount spent on fish in the last 2 years
MntSweetProducts: Amount spent on sweets in the last 2 years
MntGoldProds: Amount spent on gold in the last 2 years

The data has a set of 28 columns describing characteristics of the customers, their purchase volumes, spending habits, channel preferences and their response to earlier campaigns

PURDUE UNIVERSITY

Data Description

Place

NumWebPurchases: Number of purchases made through the company's web site
NumCatalogPurchases: Number of purchases made using a catalogue
NumStorePurchases: Number of purchases made directly in stores
NumWebVisitsMonth: Number of visits to company's web site in the last month

Promotion

NumDealsPurchases: Number of purchases made with a discount
AcceptedCmp3: 1 if customer accepted the offer in the 3rd campaign, 0 otherwise
AcceptedCmp4: 1 if customer accepted the offer in the 4th campaign, 0 otherwise
AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
AcceptedCmp1: 1 if customer accepted the offer in the 1st campaign, 0 otherwise
AcceptedCmp2: 1 if customer accepted the offer in the 2nd campaign, 0 otherwise
Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

Feature Engineering

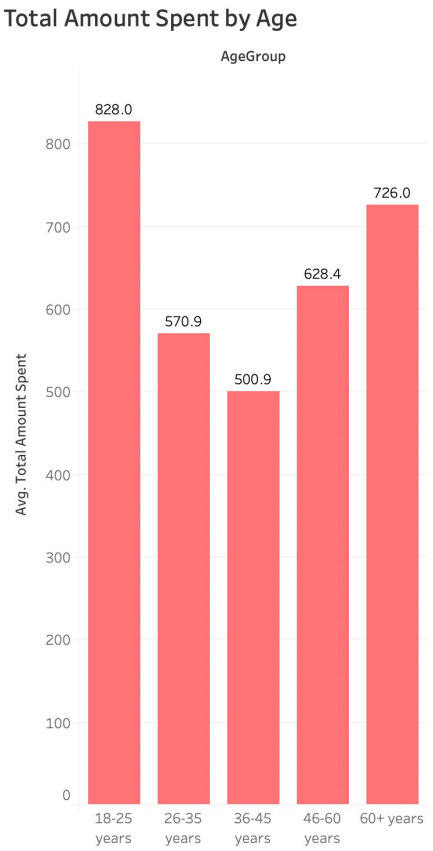
Features created from Existing Fields:

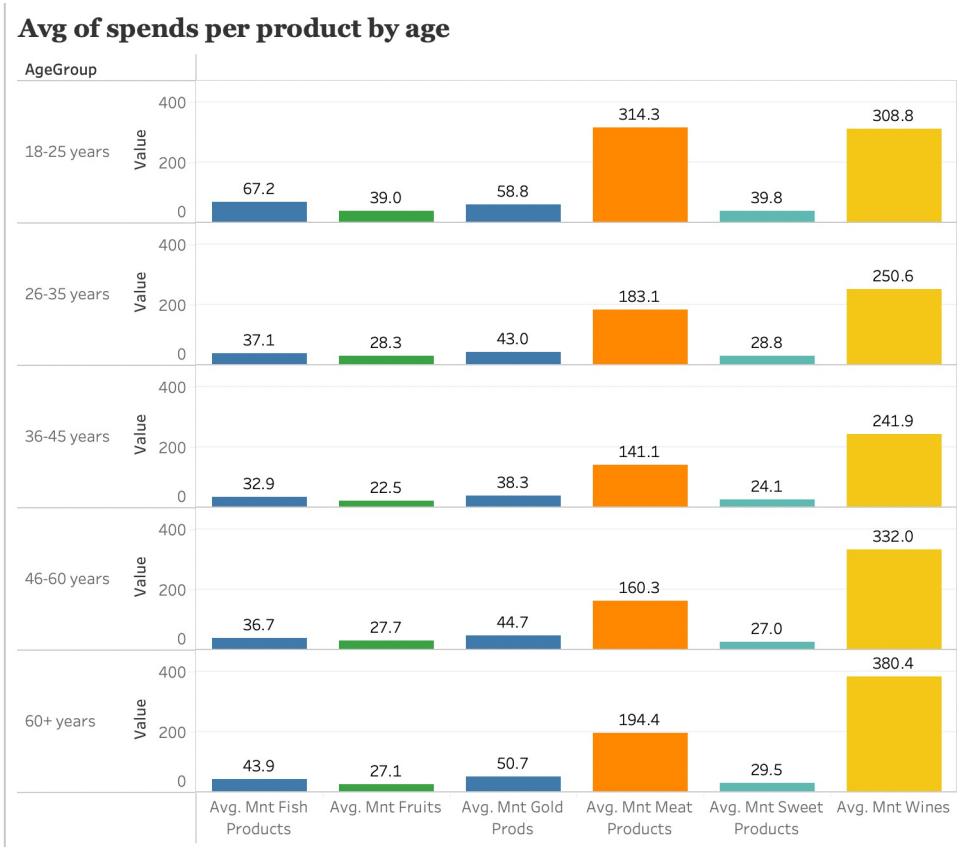
- **Customer Age (months):** The age of the customer on the platform since their day of joining
- **Total Spending:** Total Amount spent by the customer across the multiple product categories
- **Dependents:** Sum of Total teens and kids for each customer.
- **Total Purchases:** Total amount spent by customer across multiple channels and deals.
- **Conversion Rate:** Ratio of number of times customer responded positively to total number of campaigns (6).

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis

How does age affect spending (in USD)?



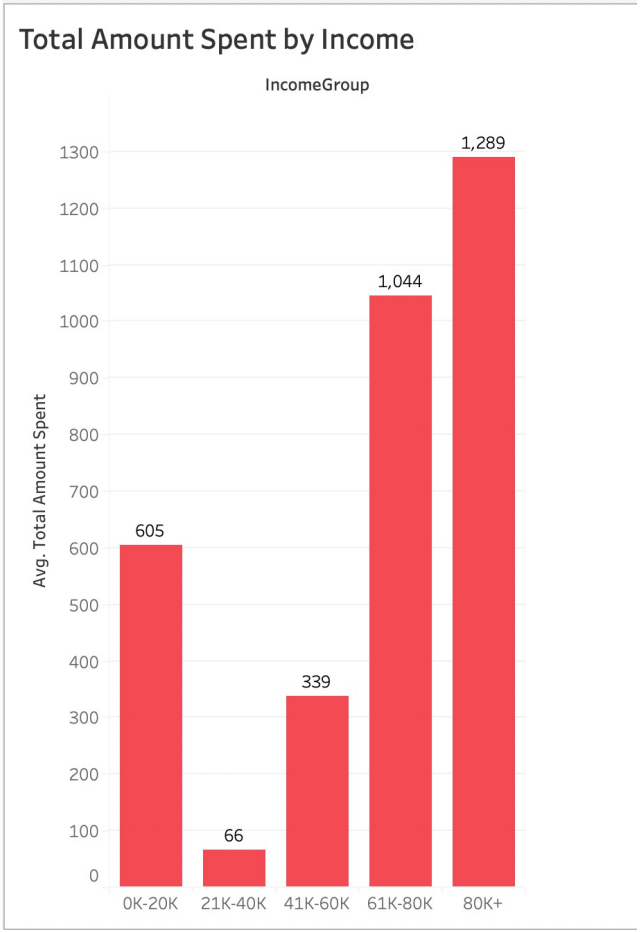


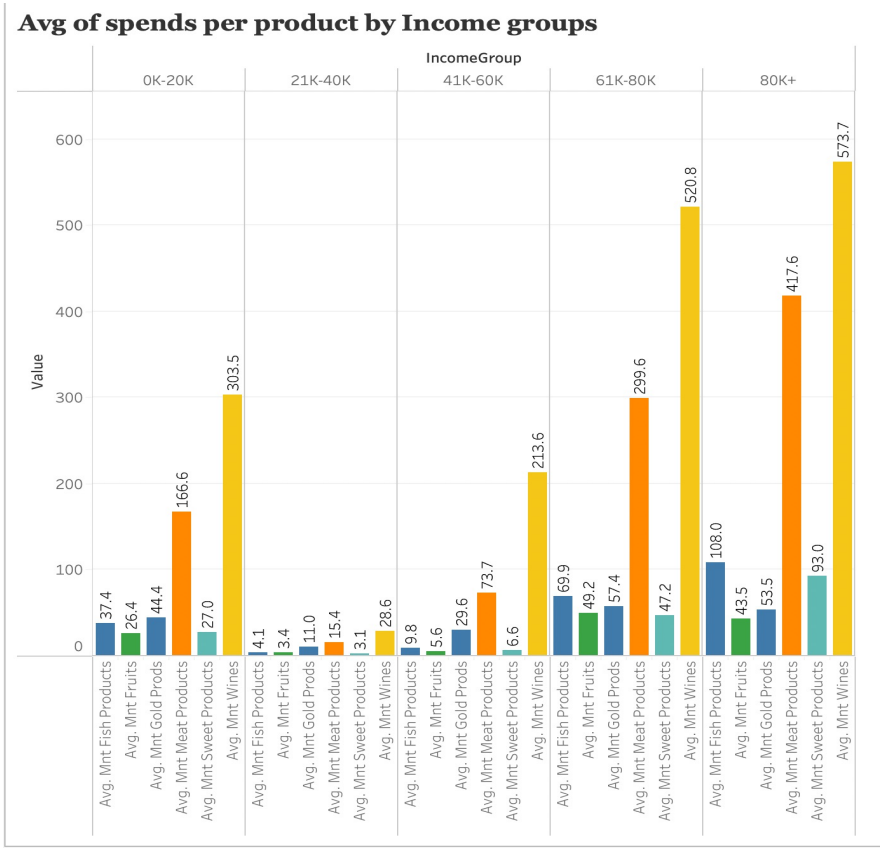
- The youngest age group (18-25 years) are the highest spenders in all product groups followed by the oldest age group (60+ years).

PURDUE UNIVERSITY

Exploratory Data Analysis

How does Income affect spending (in USD)?



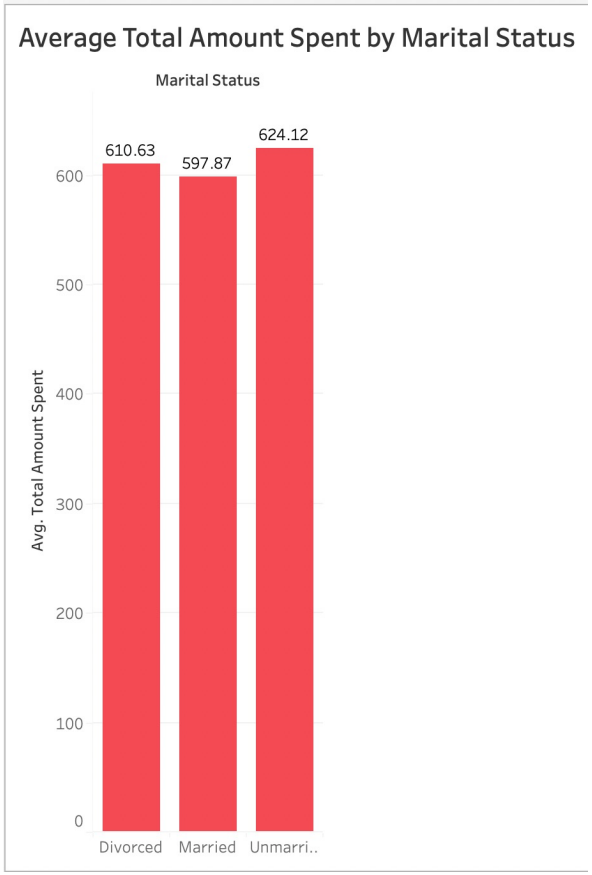


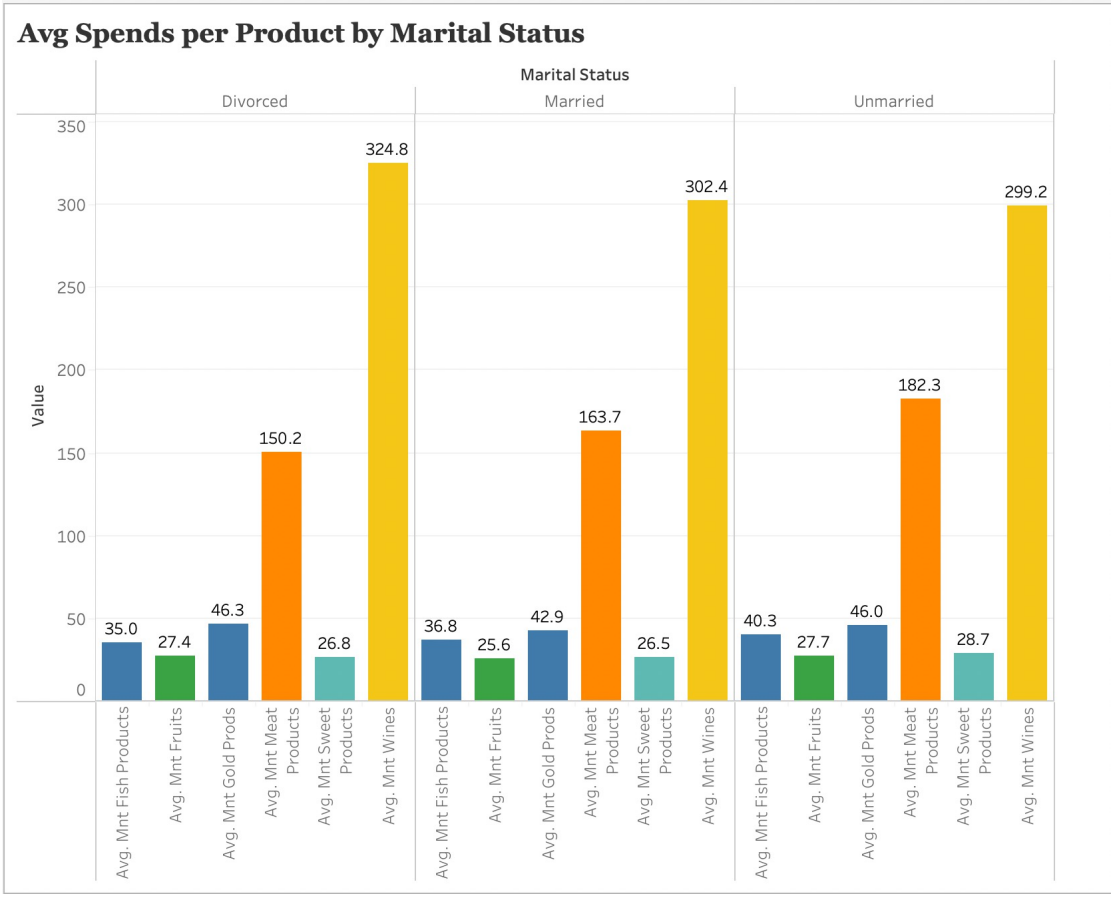
- The highest income group spend the most in each of the product category.
- Income group of 21K-40K spend the least in all the categories.

PURDUE UNIVERSITY

Exploratory Data Analysis

How does Marital Status affect spending (in USD)?



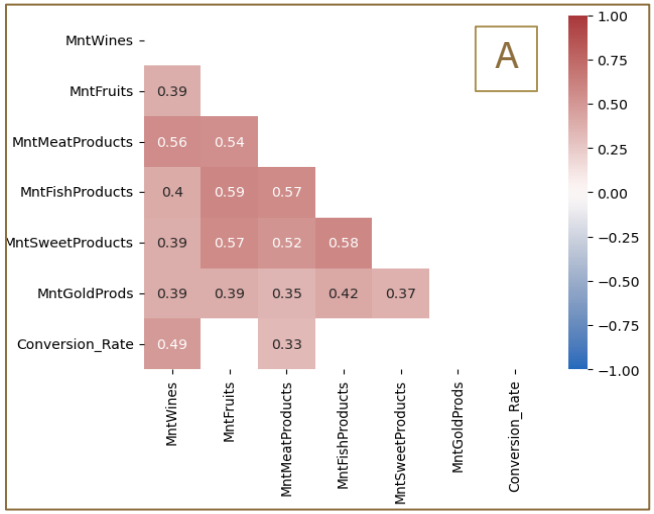


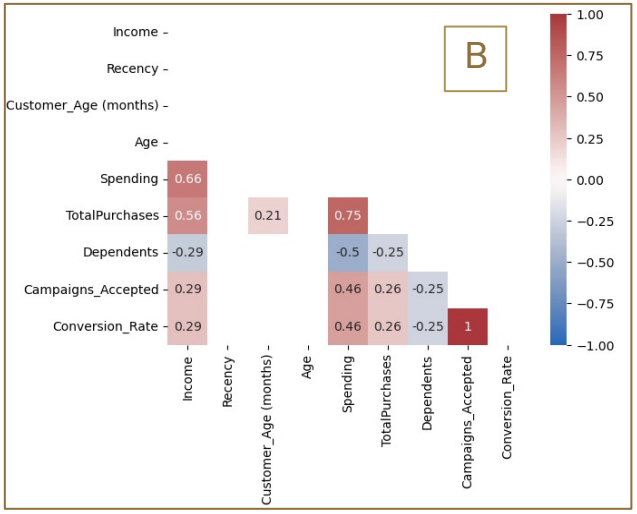
- There seems to be a consistent trend with wines, meat, and gold products which rank as the top 3.

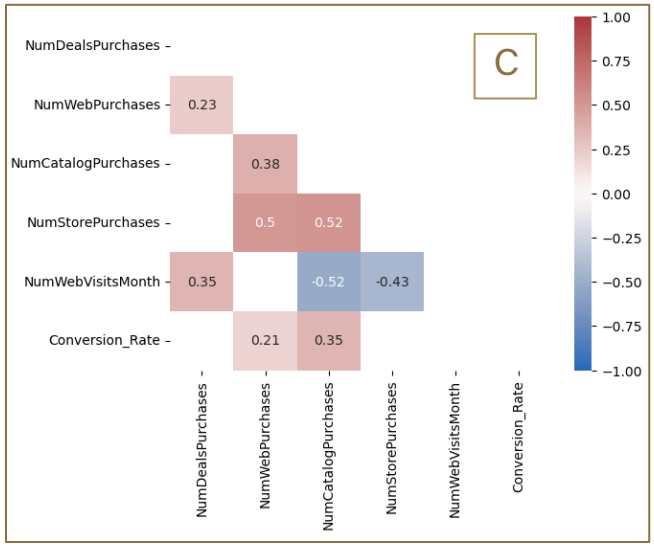
PURDUE UNIVERSITY

Exploratory Data Analysis

Variable Dependencies and Effect on Conversion Rate



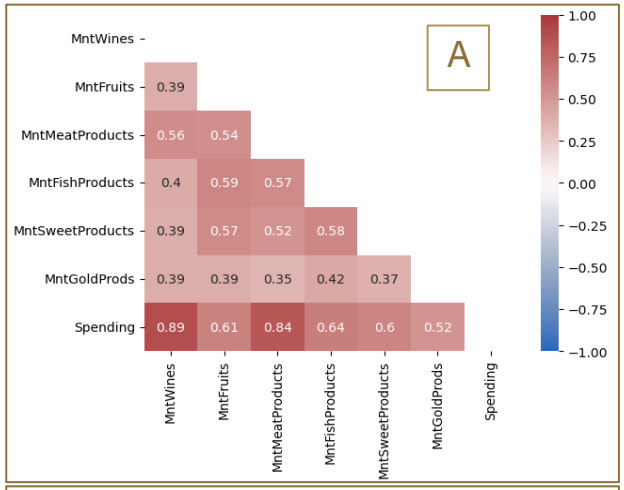


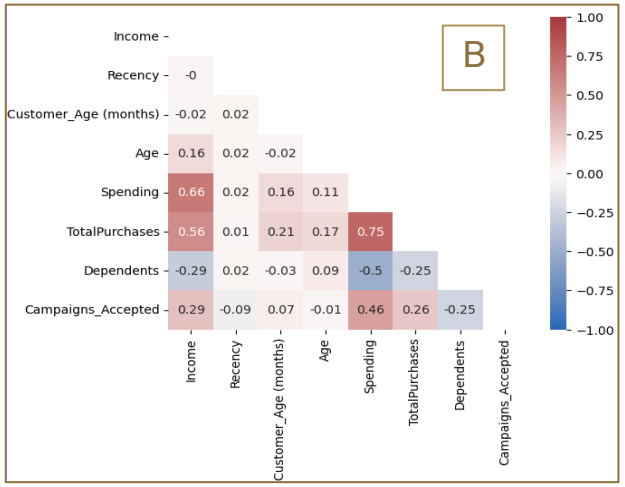


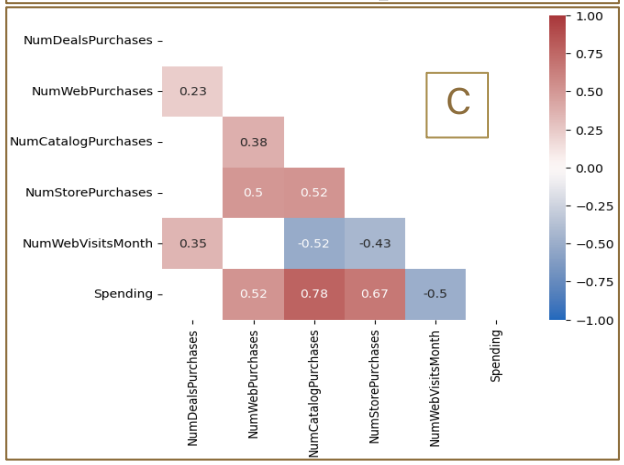
- In figure A, we see that amount spent on meats is moderately correlated with conversion rate while amount spent on wines is significantly correlated.
- In figure B, we see that Income, Spending, Total Purchases are positively correlated with Conversion Rate while Dependents is negatively correlated
- In figure C, number of purchases made on the website and number of purchases made through catalogs is positively correlated with conversion rate
- These variables will be significant when predicting conversion rate

Exploratory Data Analysis

Variable Dependencies and Effect on Total Spending







- In figure A, we see that spending is highly correlated with all product categories but meat and wine spendings are most correlated with spending.
- In figure B, we see that Income, Total Purchases and Campaigns Accepted are positively correlated with Spending while Dependents is negatively correlated.
- In figure C, number of purchases made on the website, store and catalogs is positively correlated with spending, while the number of webvisits is negatively correlated
- These variables will be significant when predicting customer spending over next 2 years

PURDUE UNIVERSITY

CUSTOMER SEGMENTATION

PURDUE UNIVERSITY®

Customer Segmentation by Conversion Rate

Age_Group	
18 to 25 years	11.90
26 to 35 years	9.70
36 to 45 years	6.40
46 to 60 years	6.85
Over 60 years	8.23

Name: Conversion_Rate, dtype: float64

Education	
2n Cycle	5.99
Basic	2.47
Graduation	7.32
Master	7.21
PhD	9.09

Name: Conversion_Rate, dtype: float64

The youngest age group has the highest conversion rates, which decline with age. Over-60s have a higher conversion rate. More educated the customers more are they receptive to campaigns.

Marital_Situation	
Divorced	8.33
Married	6.79
Unmarried	8.78

Name: Conversion_Rate, dtype: float64

Marital_Situation	Dependents	
Divorced	No	18.68
	Yes	4.89
Married	No	13.06
	Yes	4.51
Unmarried	No	15.82
	Yes	5.03

Name: Conversion_Rate, dtype: float64

The conversion rate does not appear to be significantly affected by marital status, but the situation changes when we segment by whether they have children or not.

PURDUE UNIVERSITY

Customer Segmentation by Conversion Rate

Age_Group	Income_Group	
18 to 25 years	0-20000 dollars	8.33
	20000-40000 dollars	0.00
	40000-60000 dollars	0.00
	60000-80000 dollars	10.42
	80000+ dollars	25.00
26 to 35 years	0-20000 dollars	2.92
	20000-40000 dollars	4.04
	40000-60000 dollars	4.92
	60000-80000 dollars	15.08
	80000+ dollars	32.48
36 to 45 years	0-20000 dollars	4.26
	20000-40000 dollars	3.59
	40000-60000 dollars	3.89
	60000-80000 dollars	6.28
	80000+ dollars	30.67
46 to 60 years	0-20000 dollars	1.67
	20000-40000 dollars	3.35
	40000-60000 dollars	5.02
	60000-80000 dollars	7.75
	80000+ dollars	22.38
Over 60 years	0-20000 dollars	4.17
	20000-40000 dollars	1.88
	40000-60000 dollars	5.37
	60000-80000 dollars	8.72
	80000+ dollars	24.00

Name: Conversion_Rate, dtype: float64

In all groups, conversion rate is significantly depending on whether the consumer has dependents or not.

Marital_Situation	Income_Group	
Divorced	0-20000 dollars	5.56
	20000-40000 dollars	3.33
	40000-60000 dollars	6.25
	60000-80000 dollars	8.69
	80000+ dollars	28.03
Married	0-20000 dollars	3.37
	20000-40000 dollars	2.54
	40000-60000 dollars	4.45
	60000-80000 dollars	7.42
	80000+ dollars	27.56
Unmarried	0-20000 dollars	2.69
	20000-40000 dollars	5.66
	40000-60000 dollars	4.86
	60000-80000 dollars	11.11
	80000+ dollars	24.10

Name: Conversion_Rate, dtype: float64

Age_Group	Dependents	
18 to 25 years	No	13.19
	Yes	4.17
26 to 35 years	No	16.67
	Yes	4.14
36 to 45 years	No	15.50
	Yes	4.48
46 to 60 years	No	12.71
	Yes	4.93
Over 60 years	No	14.01
	Yes	4.82

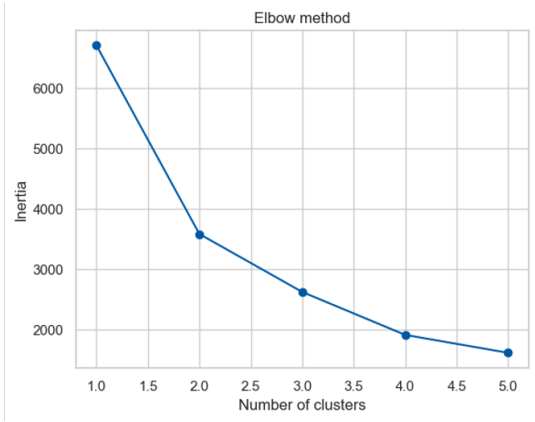
Name: Conversion_Rate, dtype: float64

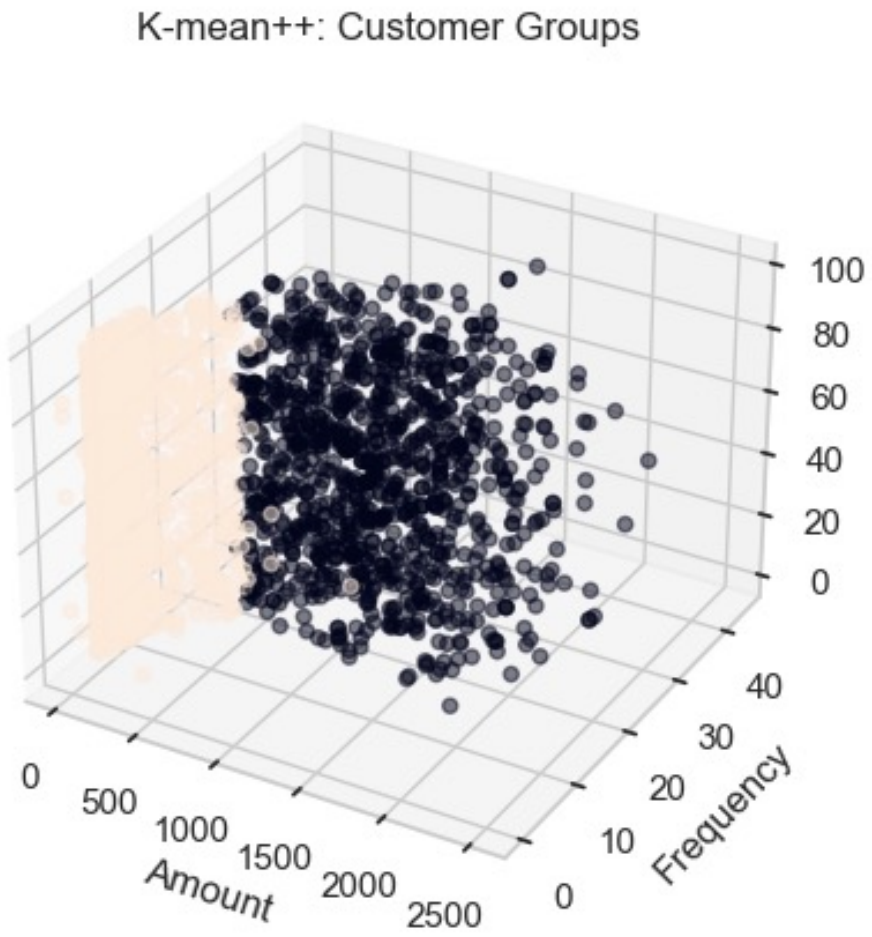
PURDUE UNIVERSITY

Customer Segmentation

K-means++ algorithm

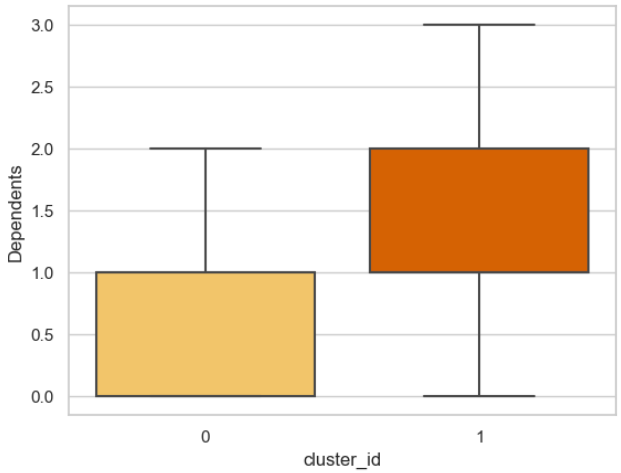
- 2 clusters selected
- RMF basis for segmentatio

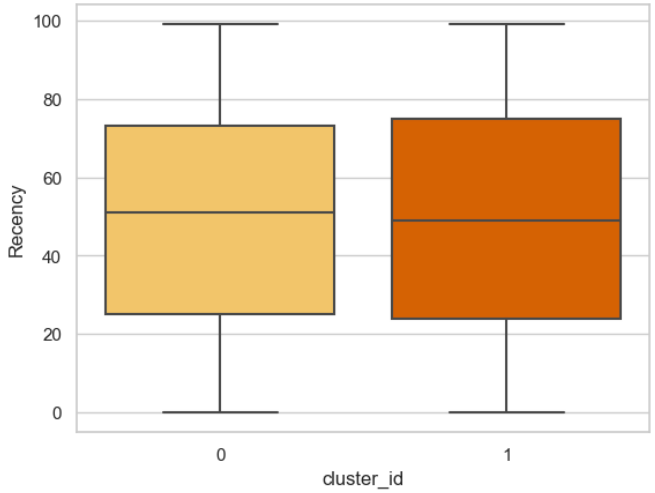


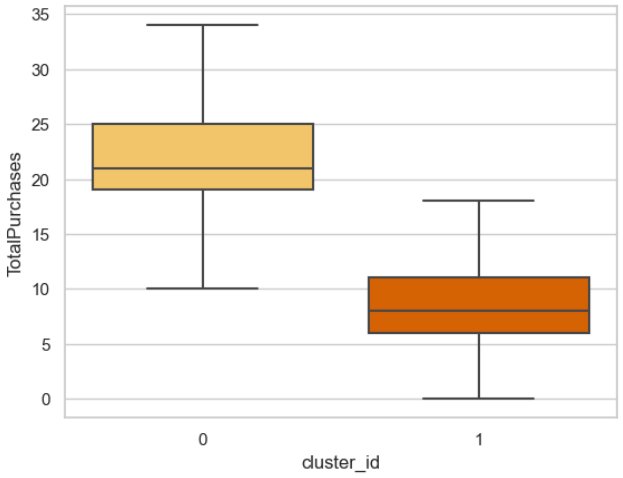


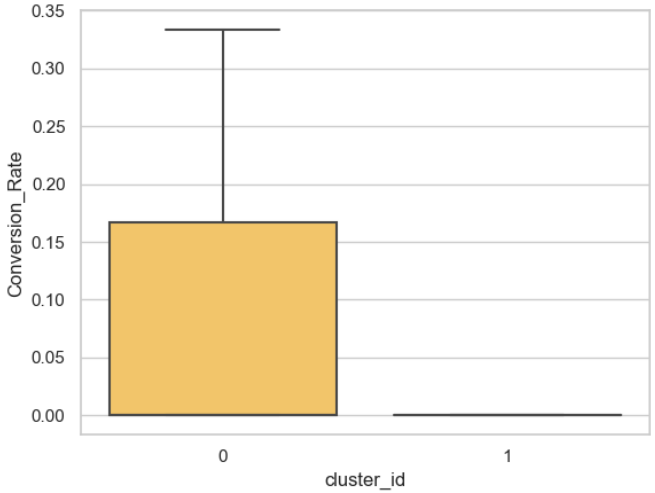
PURDUE UNIVERSITY

Segmentation Results



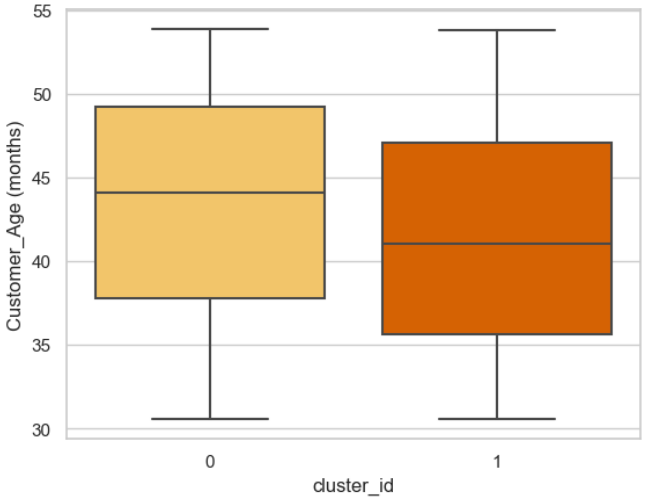


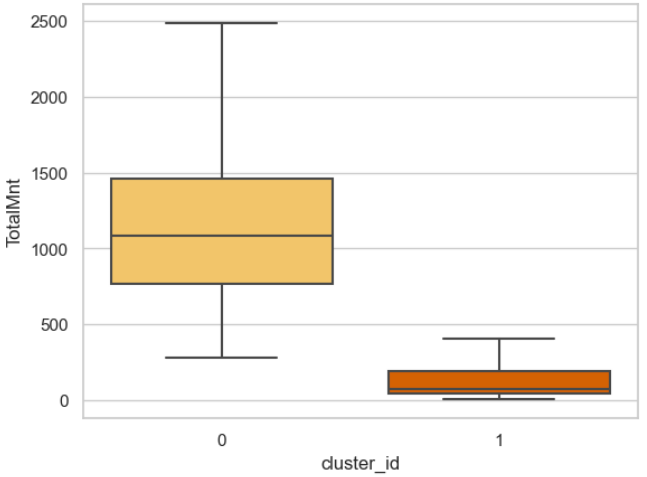


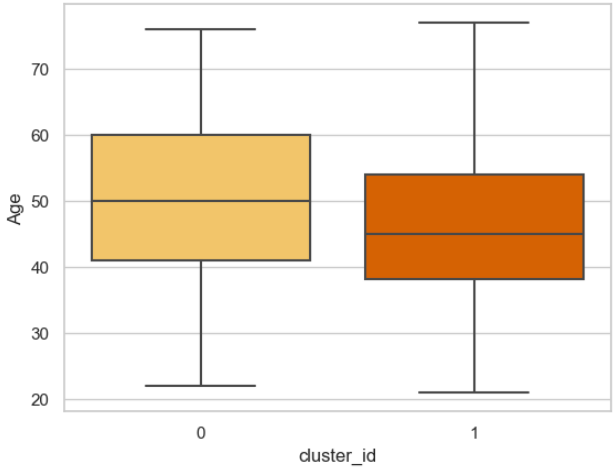


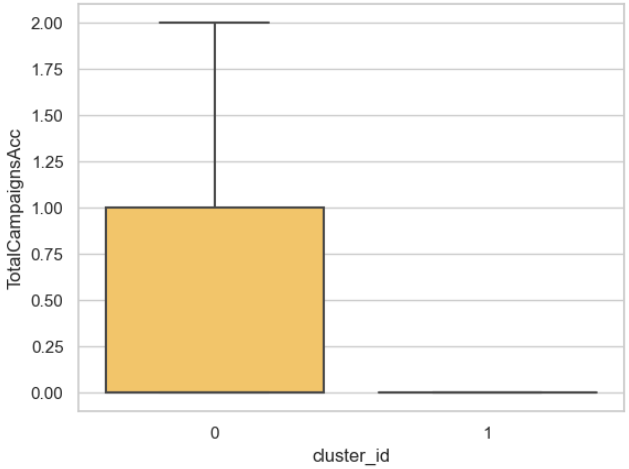
PURDUE UNIVERSITY®

Segmentation Results









PURDUE UNIVERSITY

Segmentation Results

	Cluster 0	Cluster 1
Count	1055	1182
Dependents	0.65	1.23
Income [$]	$68,558	$37,650
Recency [days]	49	48
Amount [$]	$1,136	$132
Purchases	21	8
Camp. Accept.	0.75	0.18
Conversion Rate	0.12	0.03
Age [years]	50.08	46.33
Loyalty [months]	43.36	41.47

CAMPAIGN PREDICTION

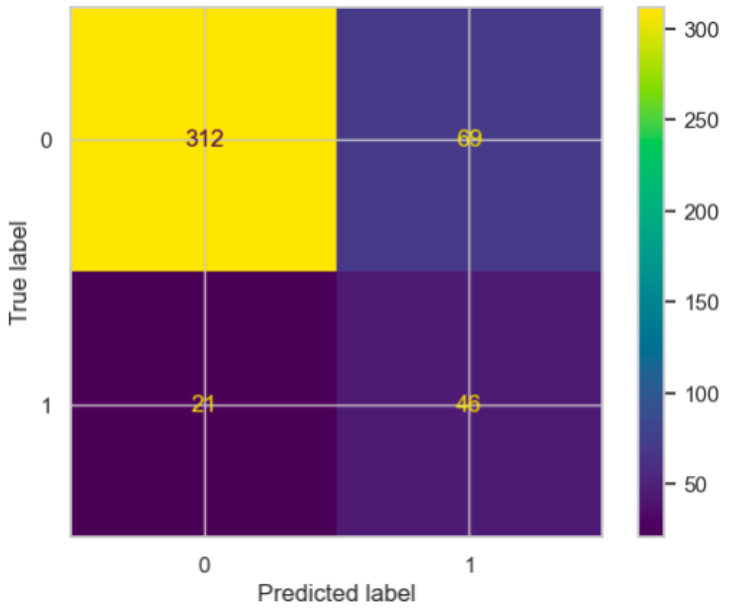
Campaign performance prediction of the subsequent marketing campaign based on past performance.

PURDUE UNIVERSITY

Campaign Prediction

Logistic Regression

Complaint	# of Complaints
Education	Education level (categorical value)
Deals Purchases	# deals purchased
Amount	Total amount spent
Marital Status	Marital status (categorical value)
Web visits	# of web visits
Dependents	# of children (kids & teens)
Channel purchases	# purchases per channel
Past Campaign	Reponses (y/n) on previous campaigns (5 in total)
Income	Customer income



Results	
Accuracy	79.91%
Sensitivity	40.00%
Specificity	81.88%

PURDUE UNIVERSITY

SPENDING PREDICTION

Prediction of client expenditures based on several customer factors, including demographic and behavioral information.

PURDUE UNIVERSITY

Spending Prediction

Linear Regression

- **Regressors:** Marital Status, Education, # Dependents, Previous Marketing Campaign Acceptance, Recency Age, Customer Loyalty, Deals Purchases, Different channels purchases, Number of Web Visits per Month, Complaints
- **Regressed Outcome:** Spending
- **Test split:** 0.8/0.2

PURDUE UNIVERSITY

Spending Prediction

Linear Regression Results

```
Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -6.284e+02  1.935e+02  -3.248 0.001184 **
Income                         2.790e-03  3.383e-04   8.247 3.14e-16 ***
factor(Marital_Status)Alone    2.540e+02  2.404e+02   1.057 0.290870
factor(Marital_Status)Divorced 2.305e+02  1.873e+02   1.230 0.218743
factor(Marital_Status)Married  2.083e+02  1.865e+02   1.117 0.264271
factor(Marital_Status)Single   2.290e+02  1.867e+02   1.227 0.220144
factor(Marital_Status)Together 2.187e+02  1.866e+02   1.172 0.241491
factor(Marital_Status)Widow    2.340e+02  1.894e+02   1.235 0.216867
factor(Marital_Status)YOLO     9.451e+01  3.221e+02   0.293 0.769252
factor(Education)Basic        -1.957e+00  4.494e+01  -0.044 0.965265
factor(Education)Graduation    4.814e+01  2.217e+01   2.171 0.030040 *
factor(Education)Master        5.966e+01  2.554e+01   2.336 0.019623 *
factor(Education)PhD           7.191e+01  2.481e+01   2.899 0.003790 **
Dependents                    -9.033e+01  1.104e+01  -8.179 5.44e-16 ***
AcceptedCmp1                   1.804e+02  2.920e+01   6.177 8.08e-10 ***
AcceptedCmp2                   5.595e+01  5.850e+01   0.956 0.338973
AcceptedCmp3                   2.483e+01  2.428e+01   1.023 0.306508
AcceptedCmp4                   9.185e+01  2.710e+01   3.389 0.000716 ***
AcceptedCmp5                   3.198e+02  2.902e+01  11.018  < 2e-16 ***
Recency                        5.340e-01  2.151e-01   2.483 0.013123 *
Age                           -3.469e-01  5.768e-01  -0.601 0.547600
Customer_Age..months.          7.807e+00  1.033e+00   7.560 6.47e-14 ***
NumDealsPurchases             -1.463e+00  4.185e+00  -0.349 0.726765
NumCatalogPurchases            8.088e+01  3.372e+00  23.988  < 2e-16 ***
NumWebPurchases                2.997e+01  2.921e+00  10.258  < 2e-16 ***
NumStorePurchases              4.230e+01  2.768e+00  15.280  < 2e-16 ***
NumWebVisitsMonth             -1.024e+01  3.969e+00  -2.581 0.009930 **
Complain                      -1.005e+02  6.816e+01  -1.475 0.140471

Residual standard error: 261.7 on 1761 degrees of freedom
Multiple R-squared:  0.8128,    Adjusted R-squared:  0.8099
F-statistic: 283.1 on 27 and 1761 DF,  p-value: < 2.2e-16
```

Results	
In of Sample R2	0.8128
Out of Sample R2	0.7802

With a significance value of 0.05 the following variables are statistically significant: Income, Education, Previous Campaigns Acceptance, Recency, Customer Loyalty, Purchases, and Web visits per Month

PURDUE UNIVERSITY

Summary of Findings and Key Achievements

The variables found to most influence a higher conversion rate are the total amount spent by customers, total number of purchases made across different avenues, number of dependents and others.

Meat and wine are the products that most loyal customers buy and these customer seem to spend a lot more and also respond to campaigns more.

The customer segmentation achieved through k-means was very effective in identifying customer segments drastically different from each other In terms of spends and campaign response.

This will allow the client to spend its dollars most effectively, targeting the right customers for the campaigns and understanding how much revenue each customer will generate over the next 2 years.

The forecasting model developed to determine new customer spending over the next 2 years has performed very well with an out-of-sample accuracy of 78%. This should help the client identify valuable customers and target them.

The logistic regression model developed for identifying customer conversion probability on a subsequent campaigns has achieved an accuracy of ~80% with a sensitivity of 40% on test data, which we hope to increase with a more balanced dataset.

THANK YOU

PURDUE UNIVERSITY®