

Marketing Response

Project by: Erlando Febrian

Table of Contents

Section 1: Project Background

Section 2: Exploratory Data Analysis

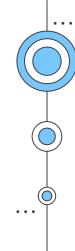
Section 3: Data Pre-Processing

Section 4: Modeling

Section 5: Clustering

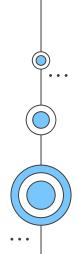
Section 6: Business Insights & Recommendation





Section 1

Project Background



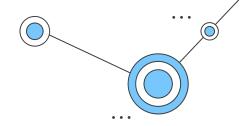
Objective

The objective is of the team is to develop a model that predicts customer behavior and to apply it to the rest of the customer base.

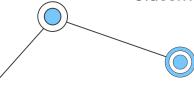
Moreover, other than maximizing the profit of the campaign, the CMO is interested in understanding to study the characteristic features of those customers who are willing to buy the gadget.

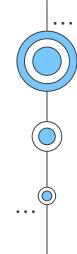
The steps are:

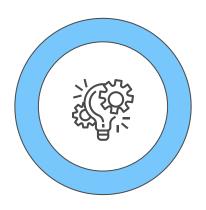
- Data Exploration;
- Segmentation;
- Classification Model;







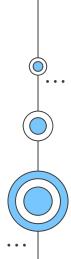


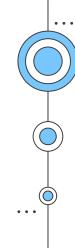


Our Goal

Increase the company's **Conversion Rate** up to **20**% and **Revenue** up to **15**% in 2023

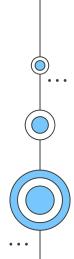
• • •





Section 2

Exploratory Data Analysis



Dataset Overview

2 Year Historical Data, contains 2240 rows, 1 row mean 1 unique customer

People

- ID
- Year_Birth
- Education
- Marital_Status
- Income
- Kidhome
- Teenhome
- Dt_Customer
- Recency
- Complain
- Country

Products

- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProds

Promotions

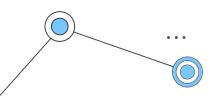
- NumDealsPurchases
- AcceptedCmp1
- AcceptedCmp2
- AcceptedCmp3
- AcceptedCmp4
- AcceptedCmp5

Place

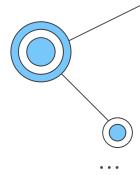
- NumWebPurchases
- NumCatalogPurchases
- NumStorePurchases
- NumWebVisitsMonth







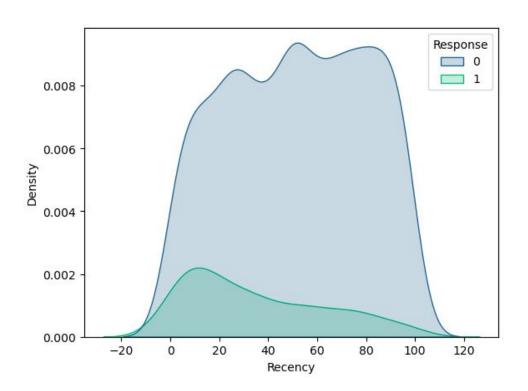
Target Feature

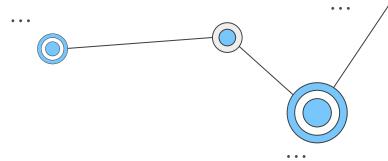




^{*)} Response :1 if customer accepted the offer in the last campaign, 0 otherwise

Recency





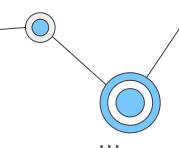
Avg Recency for customer who response is 35 days and for customer who didn't response is 51 days. This feature is normal, contains no extreme/strage values.

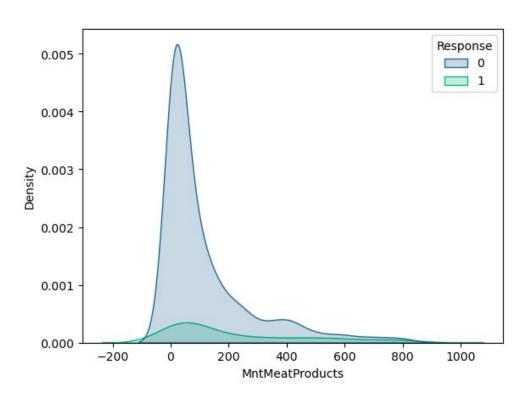
- We need to explore the cause of high recency
- Need to know what made recency values stay high

^{*)} Recency is Number of days since customer's last purchase

MntMeatProducts







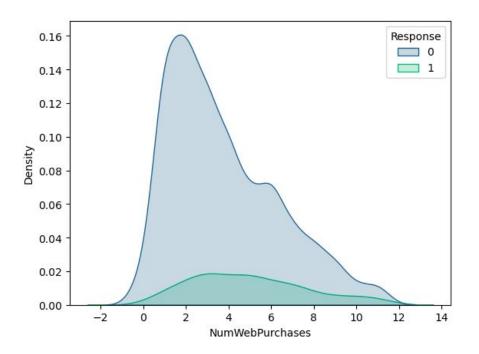
Avg Amount of Meat Products bought for customer who response is 230 pcs and for customer who didn't response is 116 pcs. This feature is normal, contains no extreme/strage values.

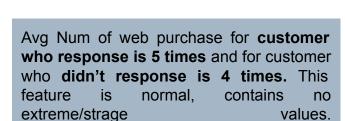
- We need to explore the cause of low amount of meat products bought
- Need to know what made amount of meat products bought values stay high

^{*)} MntMeatProducts Amount spent on meat in the last 2 years

NumWebPurchases ©





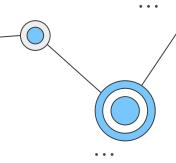


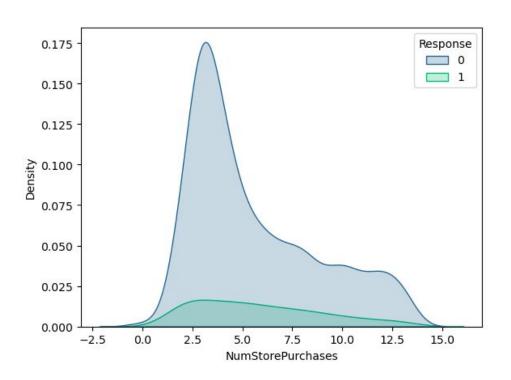
- We need to explore the cause of low num of web purchases
- Need to know what made num of web purchases values stay high

^{*)} NumWebPurchases is Number of purchases made through the company's web site

NumStorePurchases



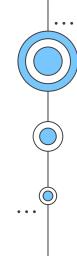




Avg Num of store purchase for customer who response is 6 times and for customer who didn't response is 5 times. This feature contains normal, no extreme/strage values.

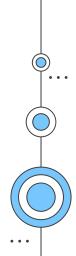
- We need to explore the cause of high num of store purchases
- Need to know what made num of store purchases values stay low

^{*)} NumStorePurchases is Number of purchases made directly in store



Section 3

Data Pre Processing



Data Pre Processing

Feature Extraction

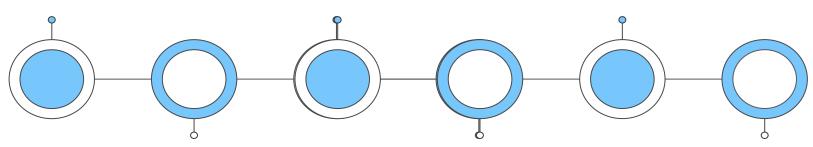
Age, Tenure, Day Registered, Month Registered, and Register on Weekend

Outlier Handling

Using Z-score method, we dropped 229 outlier values

Feature Selection

Using Quasi Constant, Chi Square, and Univariate Feature Selection Method



Missing Values Handling

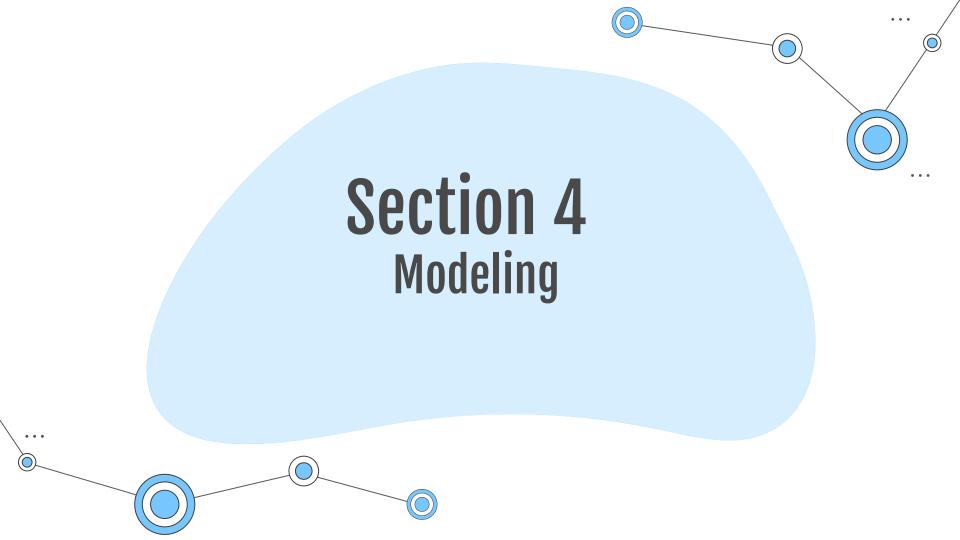
Impute 24 median values on Income feature

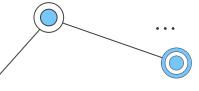
Feature Encoding & Transformation

Using One Hot Encoder and RobustScaler Method

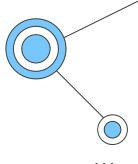
Handling Imbalaced

- SMOTE-NC (sampling strategy: 0.6)
- SMOTEENN





Basic Classification Model (SMOTE-NC)



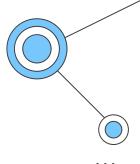
| | Model | Accuracy | Precision | Recall | F1 Score | AUC |
|---|---------------------|----------|-----------|----------|----------|----------|
| 0 | Logistic Regression | 0.866005 | 0.420290 | 0.674419 | 0.517857 | 0.781654 |
| 1 | Decision Tree | 0.826303 | 0.315068 | 0.534884 | 0.396552 | 0.697997 |
| 2 | Random Forest | 0.888337 | 0.479167 | 0.534884 | 0.505495 | 0.732720 |
| 3 | Ada Boost | 0.866005 | 0.420290 | 0.674419 | 0.517857 | 0.781654 |
| 4 | Gradient Boost | 0.900744 | 0.529412 | 0.627907 | 0.574468 | 0.780620 |
| 5 | XG Boost | 0.903226 | 0.550000 | 0.511628 | 0.530120 | 0.730814 |

We decided to do hyperparameter tuning on Gradient Boost and XGBoost, because:

- High AUC score
- Pretty much balance on Recall and Precision



Basic Classification Model (SMOTEENN)



| 195 | Model | Accuracy | Precision | Recall | F1 Score | AUC | 21 |
|-----|---------------------|----------|-----------|----------|----------|----------|----|
| 0 | Logistic Regression | 0.866005 | 0.420290 | 0.674419 | 0.517857 | 0.781654 | |
| 1 | Decision Tree | 0.826303 | 0.315068 | 0.534884 | 0.396552 | 0.697997 | |
| 2 | Random Forest | 0.888337 | 0.479167 | 0.534884 | 0.505495 | 0.732720 | |
| 3 | Ada Boost | 0.866005 | 0.420290 | 0.674419 | 0.517857 | 0.781654 | |
| 4 | Gradient Boost | 0.900744 | 0.529412 | 0.627907 | 0.574468 | 0.780620 | |
| 5 | XG Boost | 0.903226 | 0.550000 | 0.511628 | 0.530120 | 0.730814 | |

We decided to do hyperparameter tuning on Random Forest, Gradient Boost, and XGBoost, because:

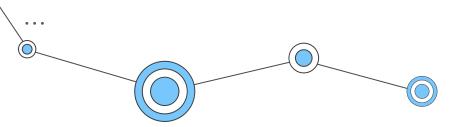
- High AUC score
- Pretty much balance on Recall and Precision

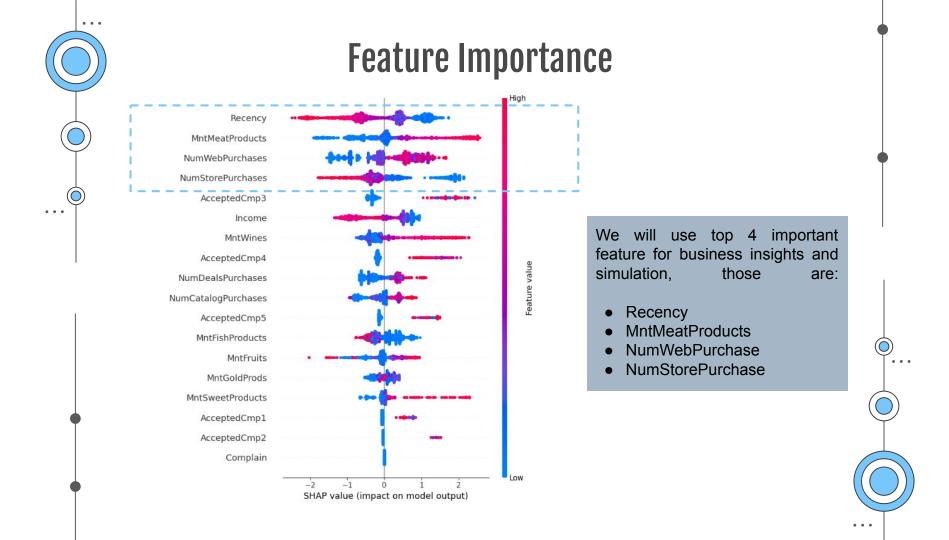
Hyperparameter Tuning

| Model | Train AUC | Test AUC |
|---------------------------------------|-----------|----------|
| Random Forest SMOTEENN | 0.823564 | 0.801938 |
| Gradient Boosting SMOTEENN | 0.859136 | 0.801776 |
| XGB SMOTEENN | 0.849670 | 0.825032 |
| Gradient Boosting Classifier SMOTE-NC | 0.803604 | 0.828521 |
| XGBClassifier SMOTE-NC | 0.970898 | 0.755975 |

We decided to do interprete **XGB** - **SMOTEENN**, because:

- High AUC score
- Less Overfitting





Section 5 Clustering



Recency

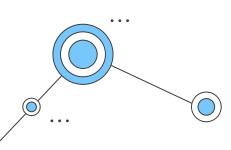
Refers to when the customer did the most recent transaction using our product (we already have Recency feature)

Frequency

Refers to how often customers do transactions using our product (sum all the feature placement type (eg: NumWebPurchases, NumCatalogPurchases, and NumStorePurchases)

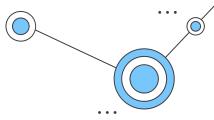
Monetary Value

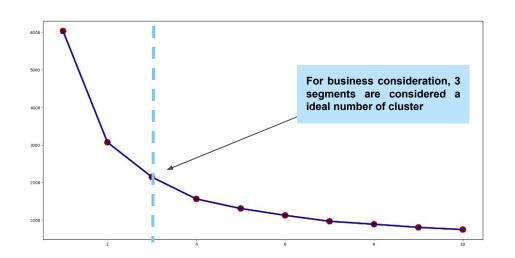
Refers to how much does a customer spend in our product (add up all the items purchased (eg: MntWines & MntFruits) by the customer and multiply by the price of each item based on additional data)¹



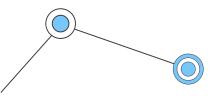
. . .

Feature Standardization and Elbow Method •





We used **StandardScaler** to standardize our data, then we tried to find the proper number of cluster, Elbow method shows 2 is the proper number of cluster, but we consider to **make 3 cluster due to business consideration**







Segment 1

Active customers with low recency, low spending amount and less frequent

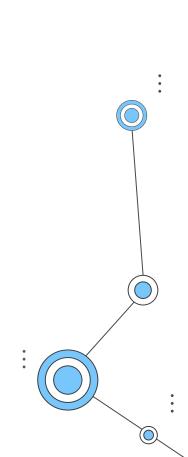
Segment 2

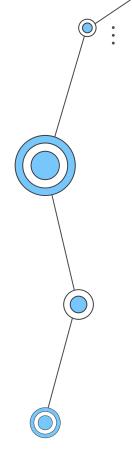
Inactive customers with high recency, low spending amount and less frequent transaction

Segment 3

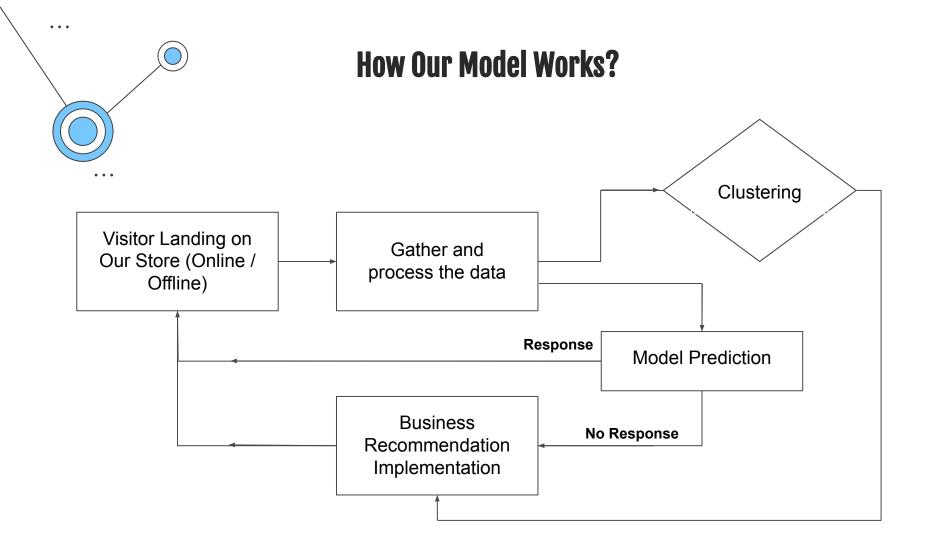
Quite active in shopping, high spending, and frequent transaction

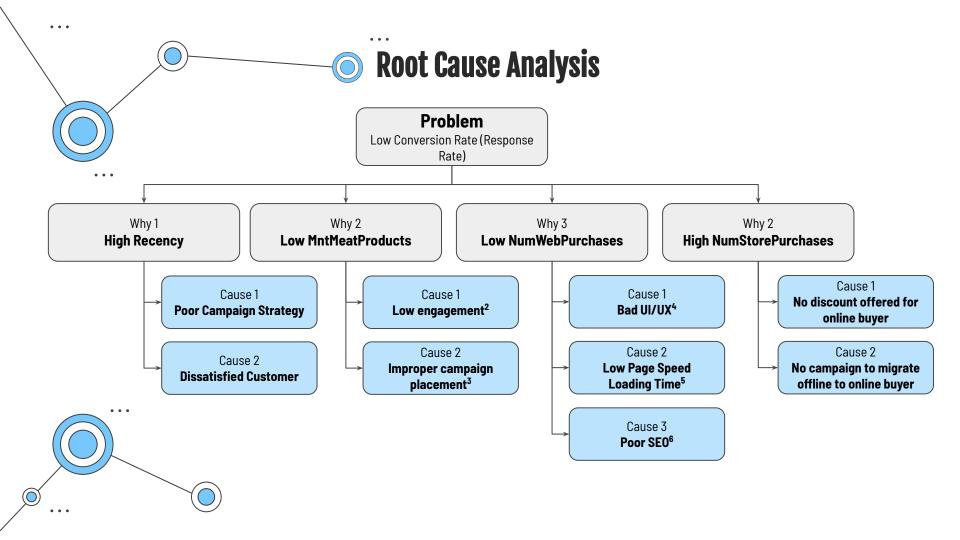
*) On python notebook, cluster 0 = segment 1, cluster 1 = segment 2, and cluster 2 = segment 3



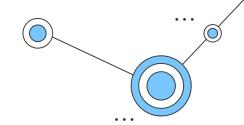


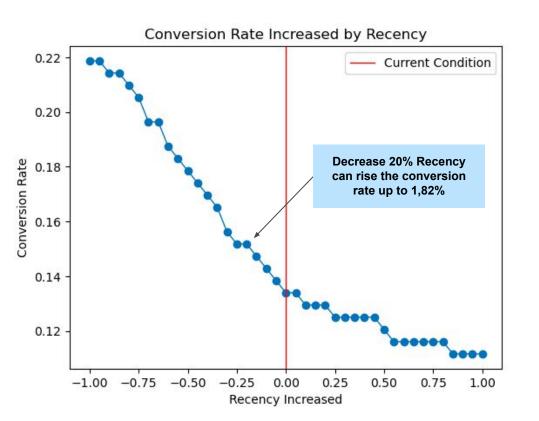
Section 6 Business Insights and Recommendation





Recency Sensitivity Analysis





Note

If the existing average Recency is 49 days, then after decreasing 20% Recency it becomes 39 days, it means that we will only target customer with Recency 39 days or less.

^{*)} Recency is Number of days since customer's last purchase

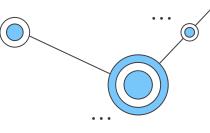
Recommendation 1

How to Deal with High Recency Values?

| Root Cause | Recommendation | How to do | Pros | Cons | |
|---|--|---|--|---|--|
| Poor Campaign Strategy: After conduct exploratory data analysis on campaign type, we know that campaign4 and campaign5 have highest average recency | Analyze and Improve campaign strategy | Analyze why campaign4 and campaign5 have bad performace Modify and implement existing successfuly campaign to campaign5 and campaign 4 | Improving campaign means improving conversion rate indirectly. Improve only 2 campaign means save marketing cost rather than improve all campaign | It keep needs cost Need long working time to make a good campaign | |
| Dissatisfied Customers | Give loyalty membership | Make loyalty membership program based on RFM segmentation. Member: • Gold: Segement 3 • Silver: Segement 1 • Bronze: Segment 2 | Increase conversion rate indirectly Quite effective method to retain customers | it costs money to make this program depending on how many members join | |
| | Audit all matters directly related to customers and improve it | Audit product quality, customer service, and marketing campaign | We can know the problem of our customer | Need cost based on the problemNeed long working time | |

Simulation 1 - Target Customer who have low Recency

Using 20% Recency Reduction



Before

After

49.10 Days

Avg Recency / Year



39.28 Days

Avg Recency / Year

13.53 % Conversion Rate / Year



15.35 %

Conversion Rate / Year

2.12 M



2.22 M

Revenue / Year

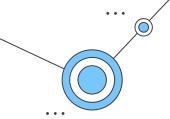
Assumption

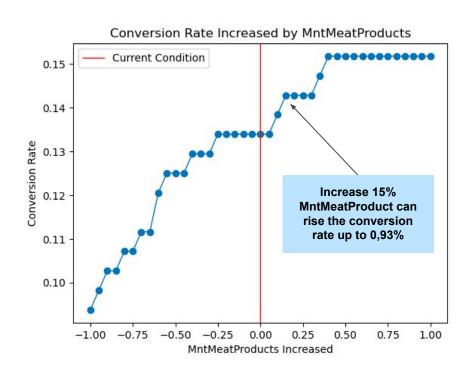
- 1. Num of Customers = 2011 / Year
- 2. Treatment Efficiency Rate = 1%
- Treatment Cost = 10% of AverageMonetary

Recommendation

For remarketing strategy, target only customer in Segment 1 who have Average Recency less than 39 Days

MntMeatProduct Sensitivity Analysis





Note

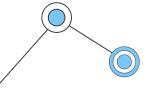
15% increment of MntMeatProduct can increase the conversion rate (response rate). It could be offer best deals for customer to stimulate them to buy more on Meat Products.

*) MntMeatProducts Amount spent on meat in the last 2 years

Recommendation 2

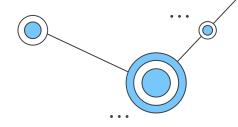
How to Deal with Low Amount of Meat Product Bought?

| Root Cause | Recommendation | How to do | Pros | Cons |
|--------------------------------|---|---|--|---|
| Low Product Engagement | Create product based campaign and content about the product | Analyze what interesting about the product Create blog post, social media post, and video post about the product | Improving campaign means improving conversion rate indirectly. Improve only 2 campaign means save marketing cost rather than improve all campaign | It keep needs cost Need long working time to make a good campaign |
| Improper Campaign Placement | Find the proper placement about this product (either online or offline) | Find the place on our platform using heatmap software which placement is better for this product Choose proper placement for product offline placement | We can know our customer behavior when surfing our product on our platform | Need cost and time to do an experimentation design Need additional working time to do ab testing |



Simulation 2 - Increase Meat Products Sales

Using 15% MntMeatProducts Increment



Before

After

13.53 %
Conversion Rate / Year



14.46 %

Conversion Rate / Year

2.12 M Revenue / Year



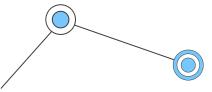
2.22 M

Revenue / Year

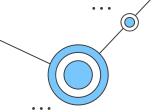
Assumption

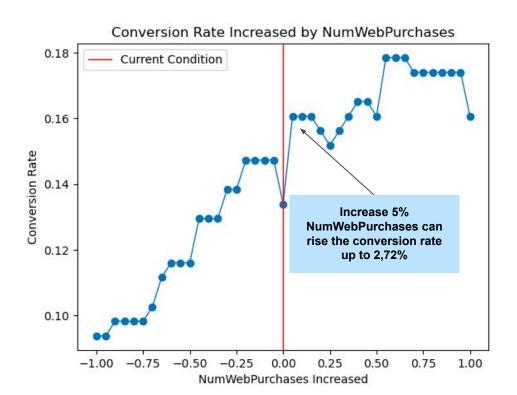
- 1. Num of Customers = 2011 / Year
- 2. Treatment Efficiency Rate = 1%
- Treatment Cost = 10% of AverageMonetary

Second recommendation can potentially increase 0.93% conversion rate (response rate) and increase 4.6% Revenue



NumWebPurchases Sensitivity Analysis





Note

Based on the result, we are looking for optimal increment of Num Web Purchases for cost efficiency, we decided to increase it 5% and it potentially lead up to 2,72% increment of conversion rate.

^{*)} NumWebPurchases is Number of purchases made through the company's web site

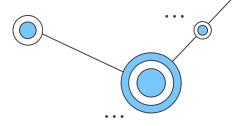
Recommendation 3

How to Deal with Low Num Web Purchases?

| How to Bear with Low Num web ruichases: | | | | | |
|---|---|--|---|--|--|
| Root Cause | Recommendation | How to do | Pros | Cons | |
| Bad UI/UX | Improve UI/UX Design ⁸ | Conduct Through ResearchSimplicity is a MustExperimental Design | Have a long-term effectRelatively cheap | Requires long working time Need additional working time to do A/B testing | |
| Low Page Speed Loading Time | Improve Page Loading Time ⁹ | Upgrade web hosting Optimize image (compression) Upgrade CMS and it's plugin | Potentially increase conversion rate higher than other recommendation | Requires good hosting and it's quite expensive Need high technical skill team and it is costly to get qualified employees | |
| Poor SEO | SEO Improvement ⁷ | Create content based on keyword7 Optimize on page SEO Go after featured snippets | Relatively cheap Have a long term effect | Need regularly improvement It needs longer time to work on than other recommendation | |

Simulation 3 - Increase Num Web Purchase

Using 5% NumWebPurchase Increment



Before

After

13.53 %
Conversion Rate / Year



14.46 %

Conversion Rate / Year

2.12 M



2.50 M

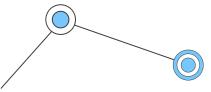
Revenue / Year

Assumption

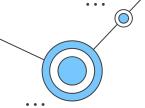
- 1. Num of Customers = 2011 / Year
- 2. Treatment Efficiency Rate = 1%
- 3. Treatment Cost = 10% of Average

 Monetary

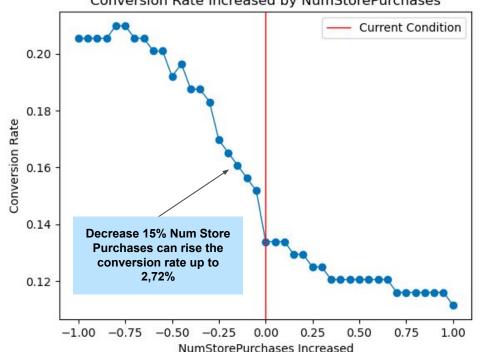
Third recommendation can potentially increase 2.72% conversion rate (response rate) and increase 17.8 % Revenue



NumStorePurchases Sensitivity Analysis







Note

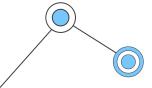
Based on the result, we are looking for optimal increment of Num Store Purchases for cost efficiency, we decided to decrease it 15% and it potentially lead up to 2,72% increment of conversion rate.

*) NumStorePurchases is Number of purchases made directly in store

Recommendation 4

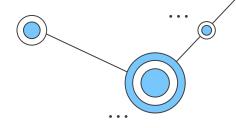
How to Deal with High Num of Store Purchases?

| | | | 7(()) | | |
|--|---|--|---|---|--|
| Root Cause | Recommendation | How to do | Pros | Cons | |
| No Discount Offered for Online Buyer | Offer best deal and discount to offline buyer | Conduct Through Research type of discount is proper for migrating customer | Decrease Num of Store Purchase cause increment of conversion rate Deal with online customer is quite easier to analyze rather than offline customers | Need budget/cost to build campaign and discount offer | |
| No Campaign is Running in purpose to Migrate Offline to Online Buyer | Running campaign on this purpose | Analyze the what campaign is suitable for this purpose | | | |



Simulation 4 - Decrease Num Store Purchase

Using 15% Num Store Purchase Reduction



Before

After

13.53 %
Conversion Rate / Year



14.46 %

Conversion Rate / Year

2.12 M Revenue / Year



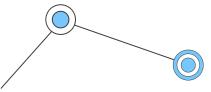
2.50 M

Revenue / Year

Assumption

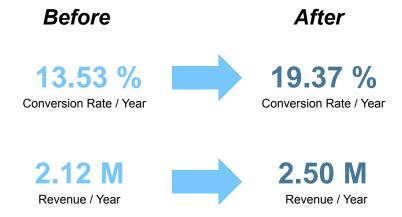
- 1. Num of Customers = 2011 / Year
- 2. Treatment Efficiency Rate = 1%
- Treatment Cost = 10% of AverageMonetary

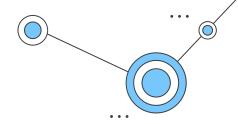
Fourth recommendation can potentially increase 2.72% conversion rate (response rate) and increase 17.8 % Revenue



Combine Simulation 2,3, and 4

- Using 15% MntMeatProducts Increment
- Using 5% NumWebPurchase Increment
- Using 15% Num Store Purchase Reduction



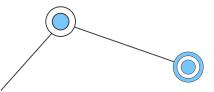


Assumption

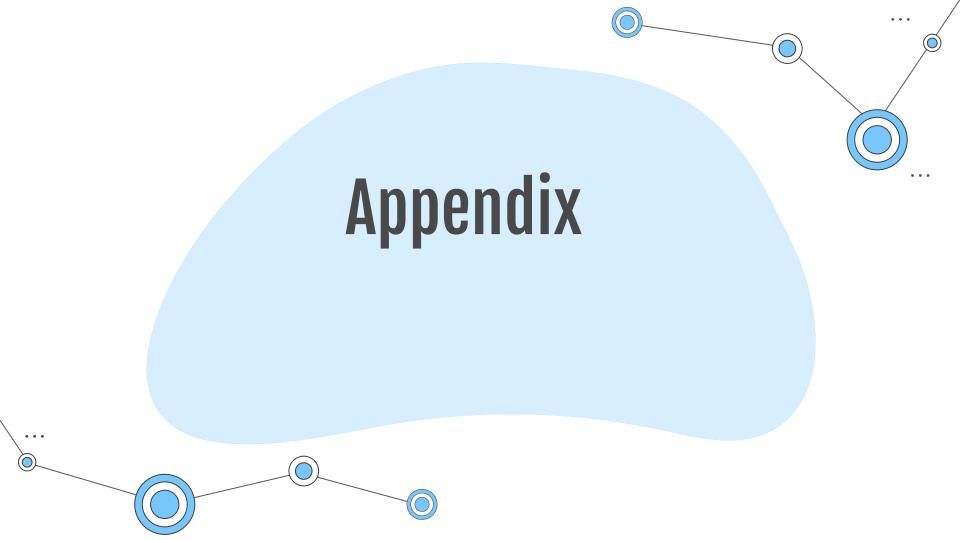
- 1. Num of Customers = 2011 / Year
- 2. Treatment Efficiency Rate = 1%
- 3. Treatment Cost = 10% of Average

 Monetary

Fourth recommendation can potentially increase 5.84% conversion rate (response rate) and increase 40.91 % Revenue



Thank You



Data Dictionary



• 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



Data Dictionary

Promotions:

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

Places:

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



References



- a. https://www.vivino.com/wine-news/how-much-does-a-good-bottle-of-wine-cost
- b. https://www.statista.com/statistics/643582/average-unit-price-candy-and-chocolate-us-by-segmen
- c. https://harga-emas.org/
- 2. https://www.benchmarkone.com/blog/marketing-campaign-engagement/
- 3. https://www.instorindia.com/7-product-placement-tips-for-retailers/
- 4. https://www.google.com/url?q=https://bizeducator.com/10-benefits-and-importance-of-ux-design/& sa=D&source=docs&ust=1656576670066302&usg=A0vVaw2JNpY_F90EM0ycc4p4ls_N
- 5. https://designwebkit.com/web-and-trends/slash-ecommerce-website-load-time/9
- 6. https://www.crazyegg.com/blog/benefits/
- 7. https://www.semrush.com/blog/improve-seo/
- 8. https://bizeducator.com/10-benefits-and-importance-of-ux-design/
- 9. https://designwebkit.com/web-and-trends/slash-ecommerce-website-load-time/9/

