Customer Personality Analysis

Agenda

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Clustering result

Retail Industry

Business Problem

- During COVID-19, online channel raise in popularity. Therefore, entering the market is considered as ***Opportunity and Inevitable action***
- Trend in ***Personalization***
  - Customer more likely to make purchase that fits with their lifestyle
  - Over 70% of consumers expect company to should have one
  - 78 % more likely to repurchase
- Issues of ***Advertising to incorrect customer*** segment leading to overspending marketing campaign

Project objective

- With existing products, we would like to create a campaign that maximize marketing spending and revenue by
  - Creating persona to right target customer
  - Require to update persona every 2 weeks, to fit with campaign schedule
  - Target only those who are likely to buy provided products

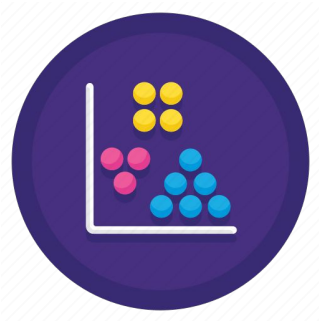
Action After Data Processing

- Create the right campaign for each customer group with high potential for purchases
- Utilizing Online channel (i.e., facebook) to advertise directly to each customer group
  - select specific characteristic according to selected persona

Aims

- Adjust to new customer behavior & understand the need of each customer group
- Encourage more revenue growth while trying to save cost through efficient excellency

- Customer Segmentation
- Accuracy and efficiency in marketing campaign through prediction



Data understaning

Detail data

Data 2,240 Record

ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProduct
5524	1957	Graduation	Single	58138	0	0	04-09-2012	58	635	88	546	17
2174	1954	Graduation	Single	46344	1	1	08-03-2014	38	11	1	6	
4141	1965	Graduation	Together	71613	0	0	21-08-2013	26	426	49	127	11
6182	1984	Graduation	Together	26646	1	0	10-02-2014	26	11	4	20	1
5324	1981	PhD	Married	58293	1	0	19-01-2014	94	173	43	118	4
7446	1967	Master	Together	62513	0	1	09-09-2013	16	520	42	98	
965	1971	Graduation	Divorced	55635	0	1	13-11-2012	34	235	65	164	5
6177	1985	PhD	Married	33454	1	0	08-05-2013	32	76	10	56	
4855	1974	PhD	Together	30351	1	0	06-06-2013	19	14	0	24	
5899	1950	PhD	Together	5648	1	1	13-03-2014	68	28	0	6	
1994	1983	Graduation	Married		1	0	15-11-2013	11	5	5	6	
387	1976	Basic	Married	7500	0	0	13-11-2012	59	6	16	11	1
2125	1959	Graduation	Divorced	63033	0	0	15-11-2013	82	194	61	480	22
8180	1952	Master	Divorced	59354	1	1	15-11-2013	53	233	2	53	
2569	1987	Graduation	Married	17323	0	0	10-10-2012	38	3	14	17	
2114	1946	PhD	Single	82800	0	0	24-11-2012	23	1006	22	115	5
9736	1980	Graduation	Married	41850	1	1	24-12-2012	51	53	5	19	
4939	1946	Graduation	Together	37760	0	0	31-08-2012	20	84	5	38	15

Attributes

People (10 columns)

- ID: Customer's unique identifier
- Year_Birth: Customer's birth year
- Education: Customer's education level
- ……..

Products (6 columns)

- MntWines: Amount spent on wine in last 2 years
- MntWines: Amount spent on wine in last 2 years
- MntFruits: Amount spent on fruits in last 2 years
- …….

Promotion (6 columns)

- NumDealsPurchases: Number of purchases made with a discount
- 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
- …….

Place (4 columns)

- NumWebPurchases: Number of purchases made through the company's website
- NumCatalogPurchases: Number of purchases made using a catalogue
- NumStorePurchases: Number of purchases made directly in stores
- ……..

People

Data Dictionary			
No.	Attribute	Description	Type
1	ID	รหัสประจำตัวลูกค้า	Integer
2	Year_Birth	ปี ค.ศ.เกิด	Integer
3	Education	ระดับการศึกษา	Nominal
4	Marital_Status	สถานภาพการสมรส	Nominal
5	Income	รายได้ครัวเรือน	Nominal
6	Kidhome	จำนวนเด็กในครัวเรือน	Integer
7	Teenhome	จำนวนวัยรุ่นในครัวเรือน	Integer
8	Dt_Customer	วันที่ลูกค้าลงทะเบียนกับบริษัท	Nominal
9	Recency	จำนวนวัน ตั้งแต่การซื้อครั้งล่าสุดของลูกค้า	Nominal

Products

Data Dictionary			
No.	Attribute	Description	Type
10	MntWines	จำนวนเงินที่ใช้ไปกับไวน์ใน 2 ปีที่ผ่านมา	Integer
11	MntFruits	จำนวนเงินที่ใช้ไปกับผลไม้ใน 2 ปีที่ผ่านมา	Integer
12	MntMeatProducts	จำนวนเงินที่ใช้ไปกับเนื้อสัตว์ใน 2 ปีที่ผ่านมา	Integer
13	MntFishProducts	จำนวนเงินที่ใช้ไปกับปลาใน 2 ปีที่ผ่านมา	Integer
14	MntSweetProducts	จำนวนเงินที่ใช้ไปกับขนมหวานใน 2 ปีที่ผ่านมา	Integer
15	MntGoldProds	จำนวนเงินที่ใช้ไปกับทองคำใน 2 ปีที่ผ่านมา	Integer

Place

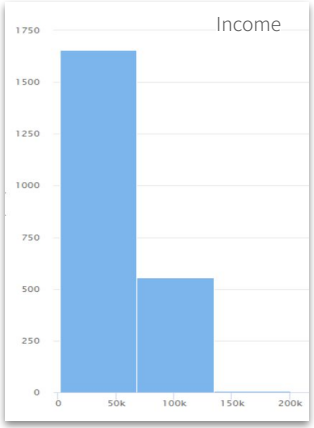
Data Dictionary			
No.	Attribute	Description	Type
16	NumDealsPurchases	จำนวนการซื้อที่มีส่วนลด	Integer
17	NumWebPurchases	จำนวนการซื้อผ่านเว็บไซต์ของบริษัท	Integer
18	NumCatalogPurchases	จำนวนการซื้อโดยใช้แคตตาล็อก	Integer
19	NumStorePurchases	จำนวนการซื้อโดยตรงในร้านค้า	Integer
20	NumWebVisitsMonth	จำนวนการเข้าเยี่ยมชมเว็บไซต์ของบริษัทในเดือนที่ผ่านมา	Integer

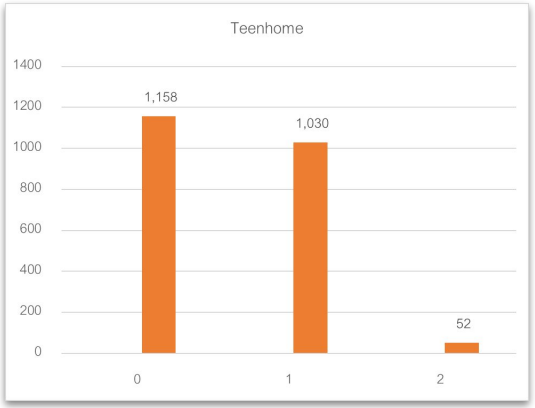
Promotion

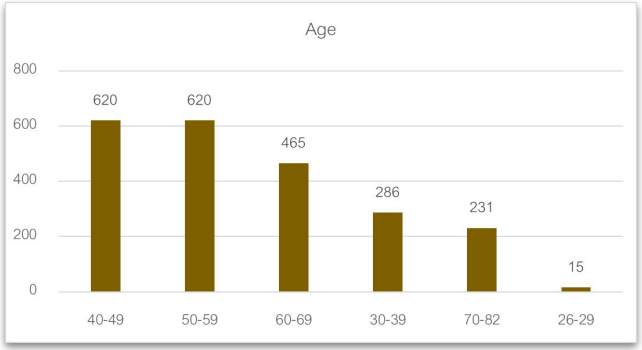
No.	Attribute	Description	Type
21	AcceptedCmp1	ลูกค้ายอมรับข้อเสนอในแคมเปญที่ 1 (ใช่/ไม่ใช่)	Integer
22	AcceptedCmp2	ลูกค้ายอมรับข้อเสนอในแคมเปญที่ 2 (ใช่/ไม่ใช่)	Integer
23	AcceptedCmp3	ลูกค้ายอมรับข้อเสนอในแคมเปญที่ 3 (ใช่/ไม่ใช่)	Integer
24	AcceptedCmp4	ลูกค้ายอมรับข้อเสนอในแคมเปญที่ 4 (ใช่/ไม่ใช่)	Integer
25	AcceptedCmp5	ลูกค้ายอมรับข้อเสนอในแคมเปญที่ 5 (ใช่/ไม่ใช่)	Integer
26	Response	ลูกค้ายอมรับข้อเสนอในแคมเปญที่แล้ว (เคย/ไม่เคย)	Integer

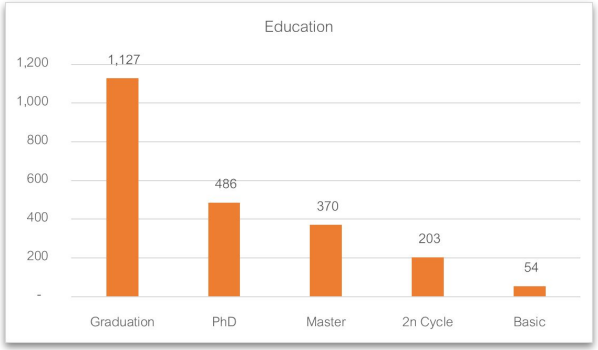
Data Exploration & Data Preparation

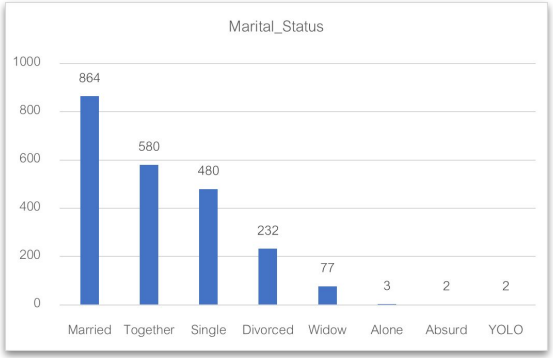
People







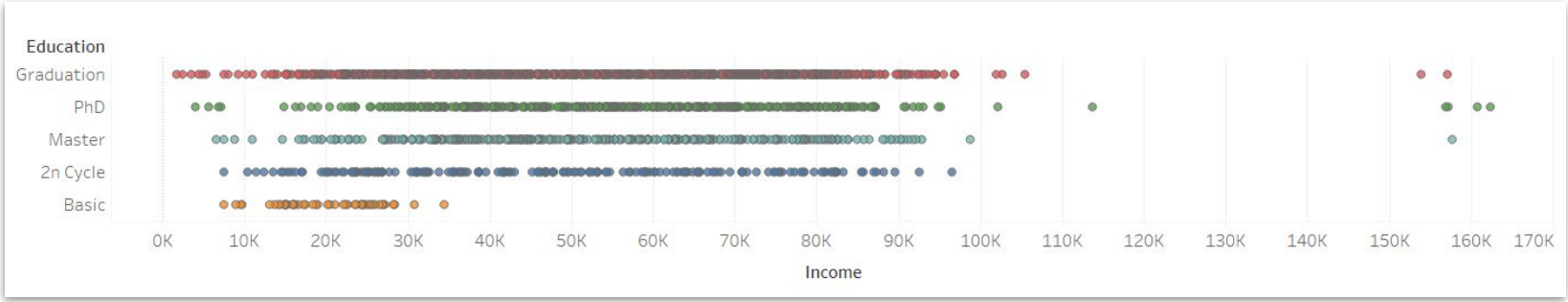




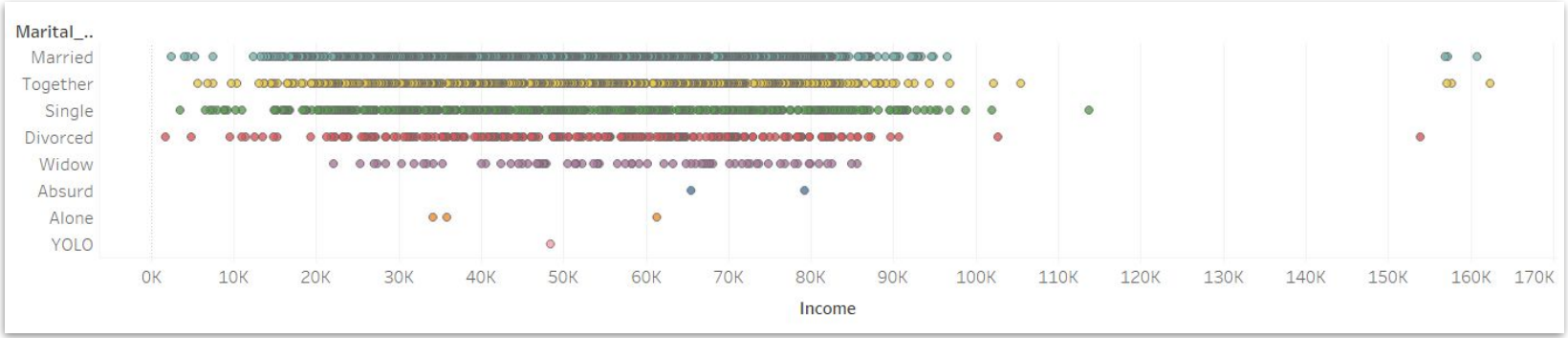
Data Exploration

Income Feature

Education:

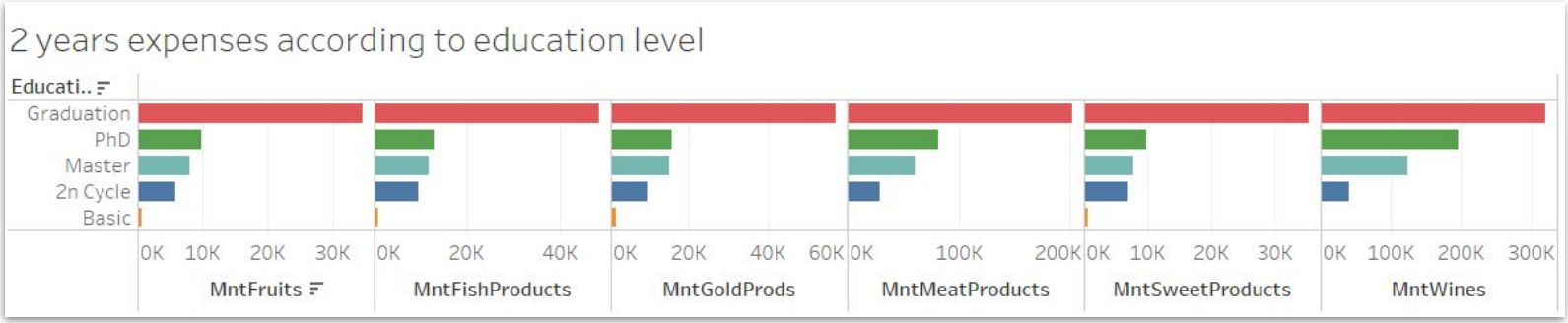


Marital Status:

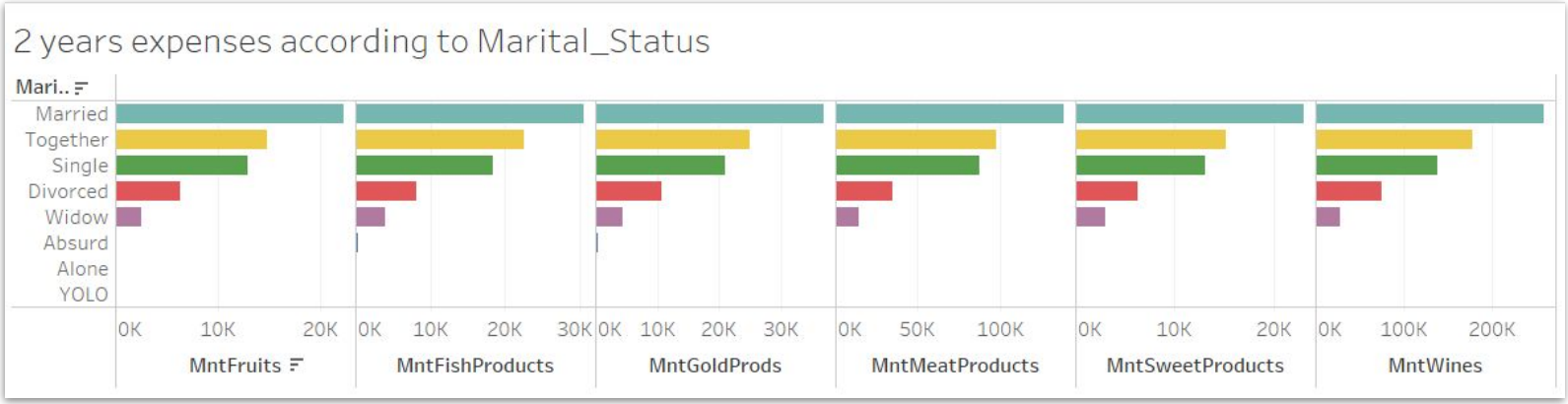


Product Feature

Education:

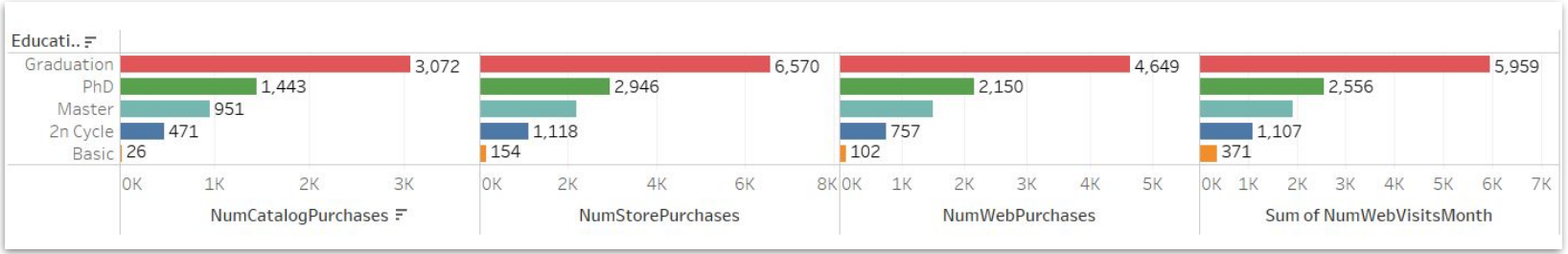


Marital Status:

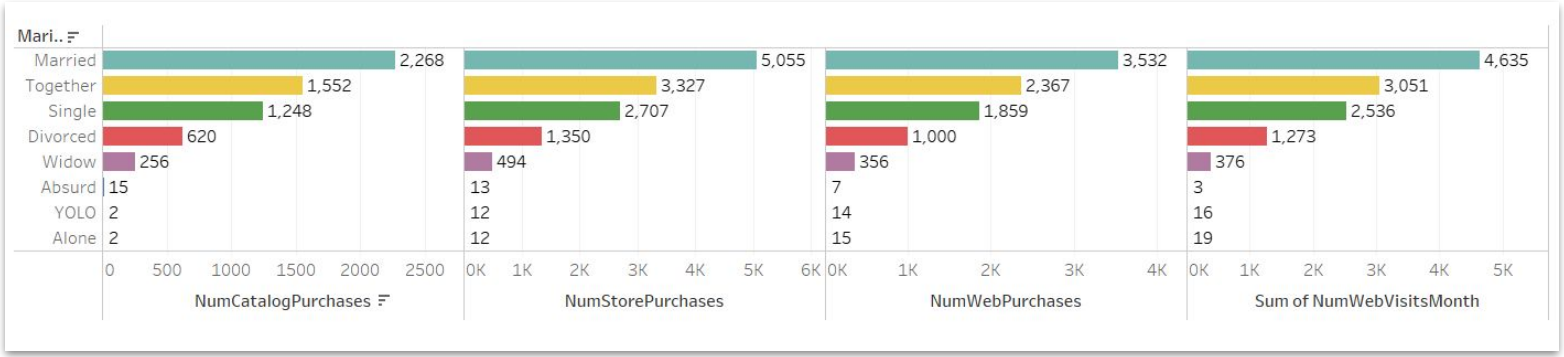


Place Feature

Education:

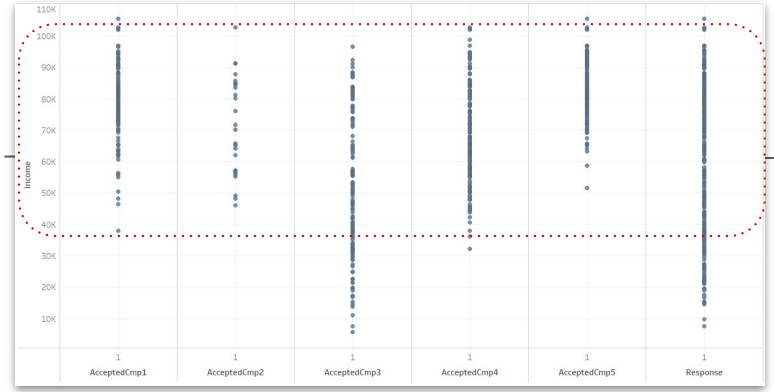


Marital Status:

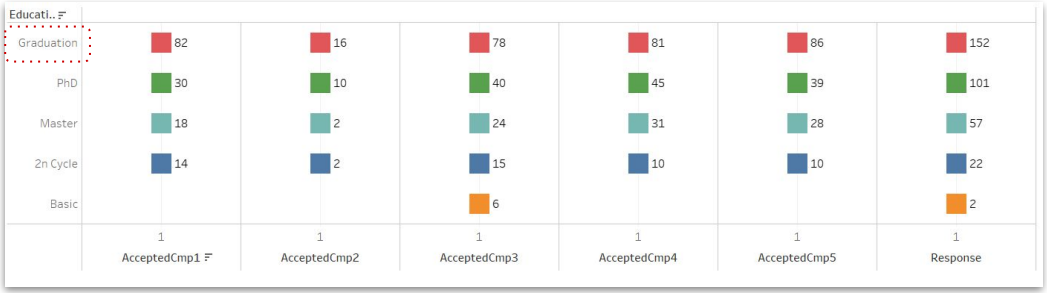


Campaign Feature

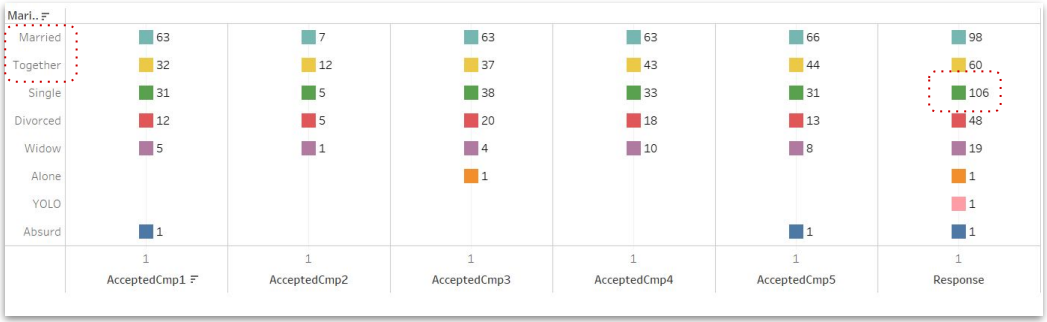
Income



Education

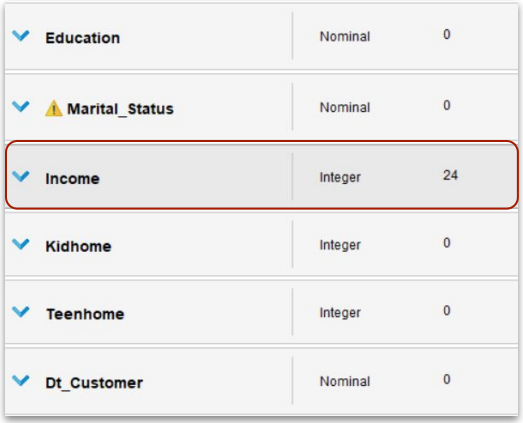


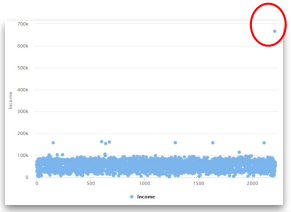
Marital Status

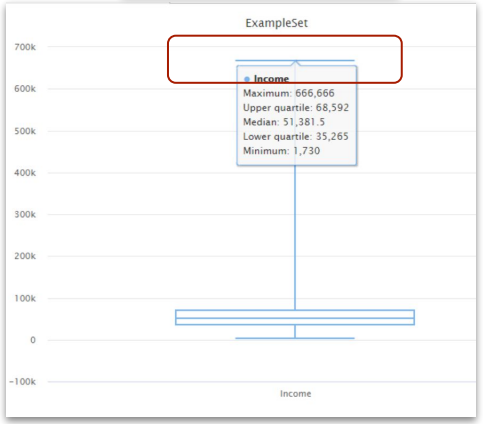


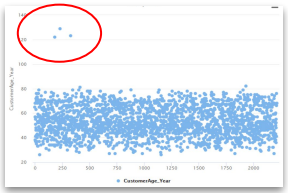
Data Exploration

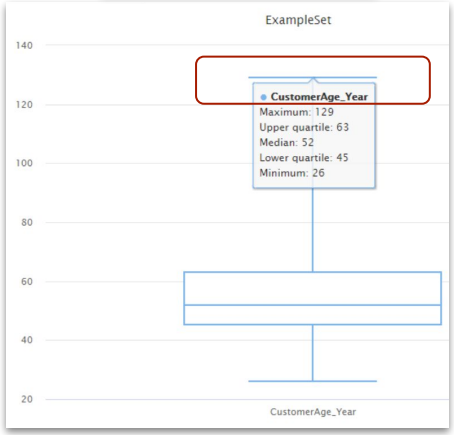
Finding Data







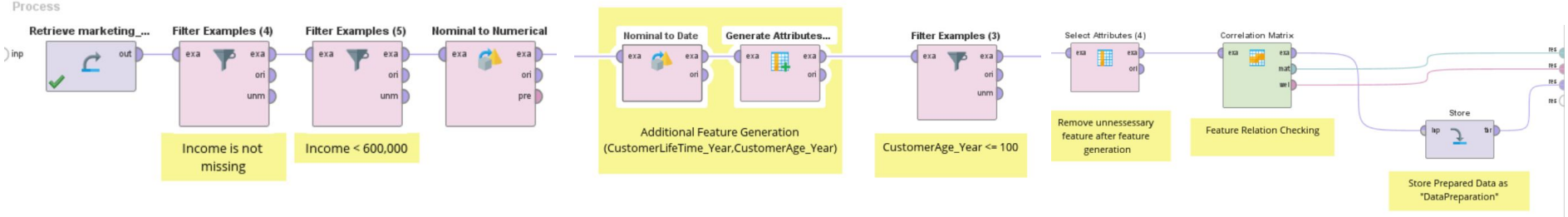




Remove Missing Data	Remove Outlier	Remove Outlier
Income = 24 Attribute	Income > 600,000 (1 Attribute)	Age of customer: >100 (3 Attribute)

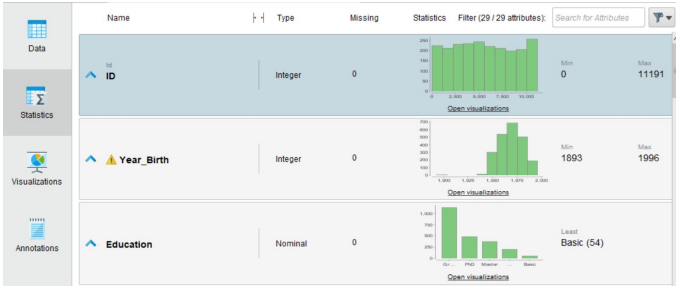
Step of Work

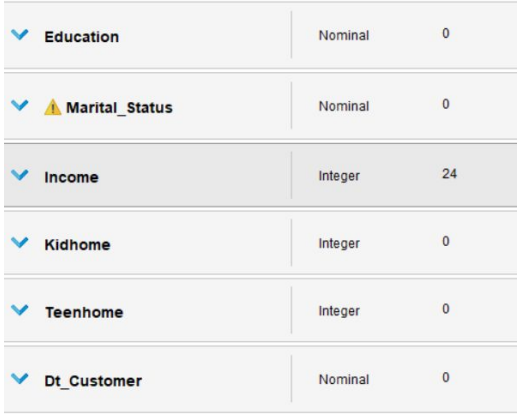
- Data Exploration
- Missing Data Cleansing
- Nominal to Numerical Transformation
- Additional Feature Generation (DT_customer, Year_Birth)
- Outlier Management (Income > 600,000, Age of customer: >100)
- Feature Relation Checking

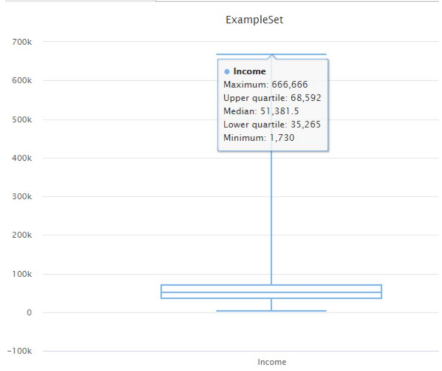


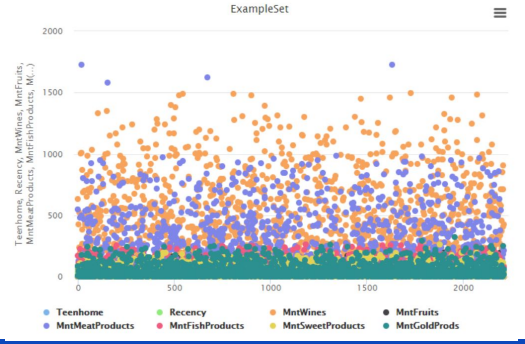
Data Exploration

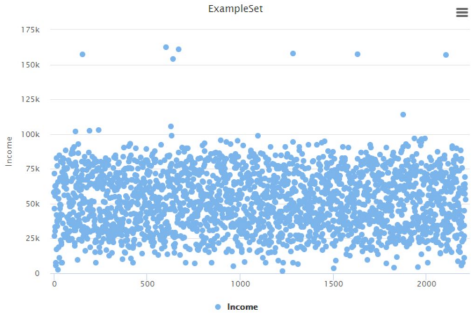
➢ **Tools**

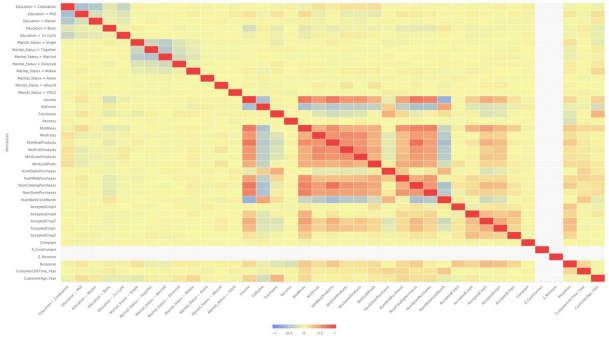












Data Exploration

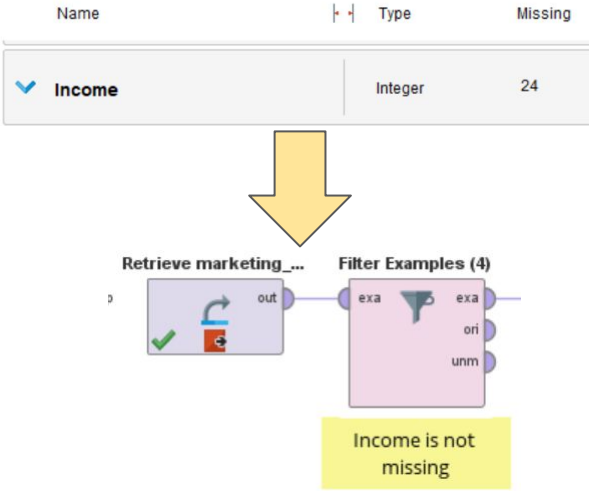
➢ **Finding**

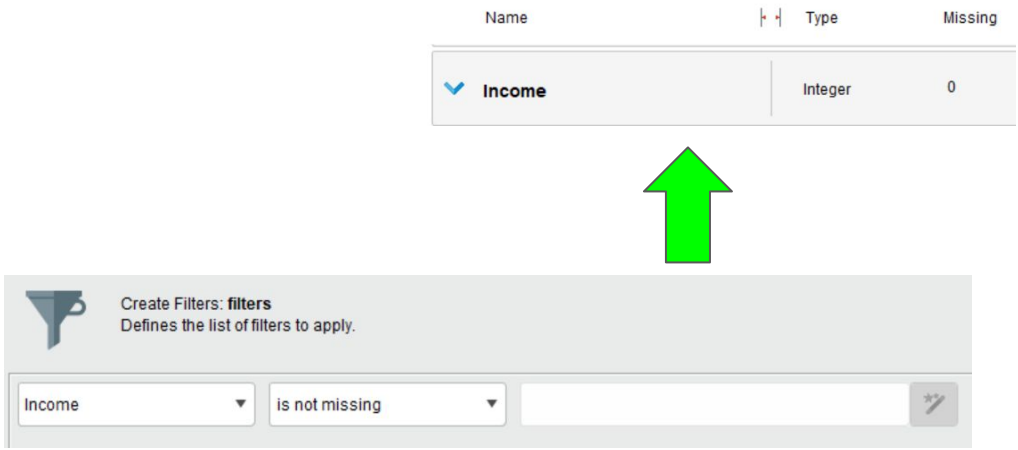
Finding	Feature	Detail	Action
1) Missing Data	Income	1.07 % of missing data (24 missing from 2,240) Remark: < 2%	Remove missing value
2) Incompatible feature type with model	Education Merital_Status	Education: Nominal type Marital_Status: Nominal type	Nominal to Numerical Transformation
3) Improper Meaning related with real situation	DT_customer Year_Birth	DT_customer: Year that customers start registration Year_Birth: Customer birth year	DT_customer: transform to age of customer Year_Birth: transform to customer life time
4) Outlier Detection	Income Age of customer	Income: > 600,000 Age of customer: >100	Remove outlier

Data Exploration

➢ **Finding: 1) Missing Data**

Finding	Feature	Detail	Action
1) Missing Data	Income	1.07 % of missing data (24 missing from 2,240) Remark: < 2%	Remove missing value

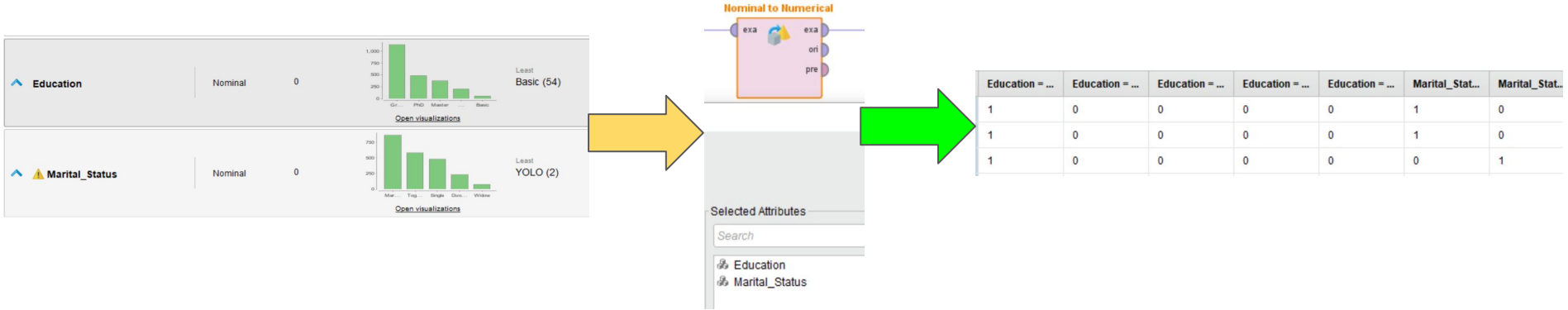




Data Exploration

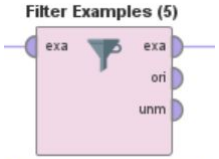
➢ **Finding: 2) Incompatible feature type with model**

Finding	Feature	Detail	Action
2) Incompatible feature type with model	Education Merital_Stage	Education: Nominal type Marital_Stage: Nominal type	Nominal to Numerical Transformation

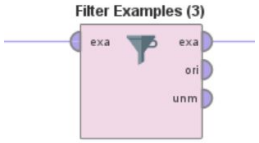


Data Exploration

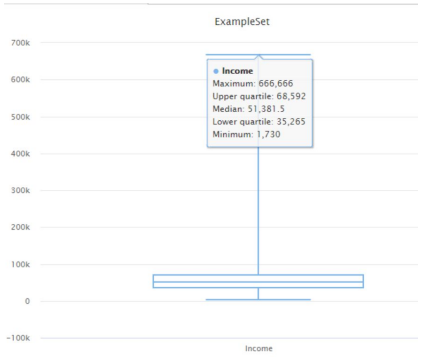
➢ **Finding: 4) Outliner**

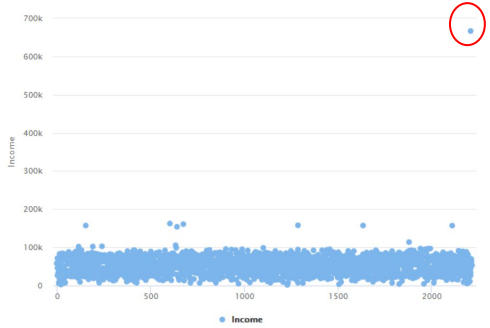


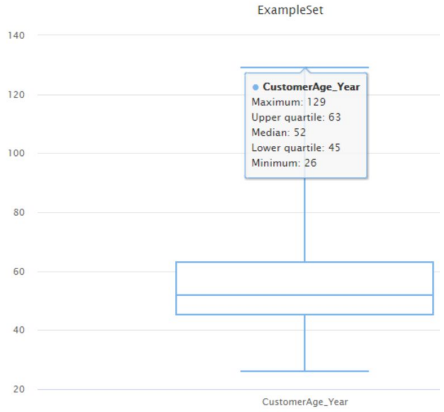
Income < 600,000

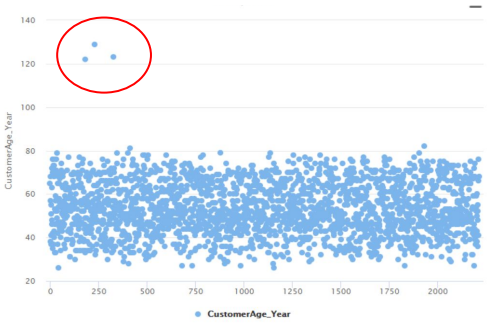


CustomerAge_Year <= 100



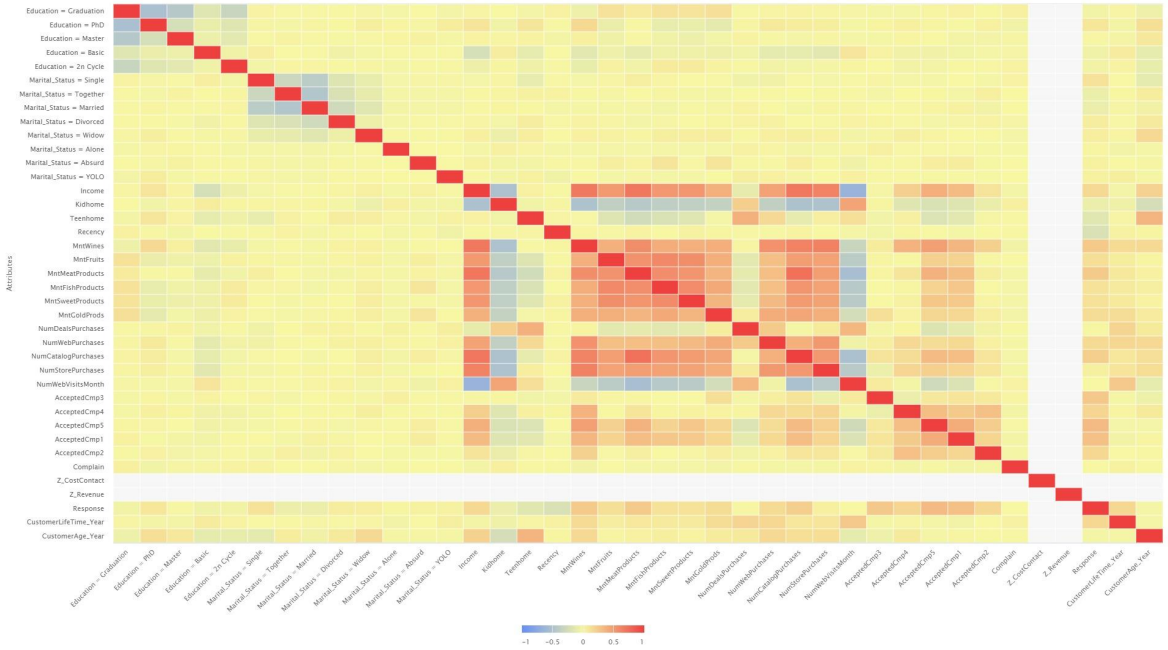






Data Exploration

➢ **Feature Relation Checking**



Correlation Coefficient (r)

\ r /	Strength
0.8 to 1.0	Very strong
0.6 to 0.8	Strong
0.4 to 0.6	Moderate
0.2 to 0.4	Weak
0.0 to 0.2	Very weak

abs < 0.8 : Not Very Strong

First Attribute	Second Attribute	Correlation ↓
MntMeatProducts	NumCatalogPurch...	0.734
Income	NumCatalogPurch...	0.697
Income	MntMeatProducts	0.692
Income	MntWines	0.688
MntWines	NumStorePurchas...	0.640
MntWines	NumCatalogPurch...	0.634

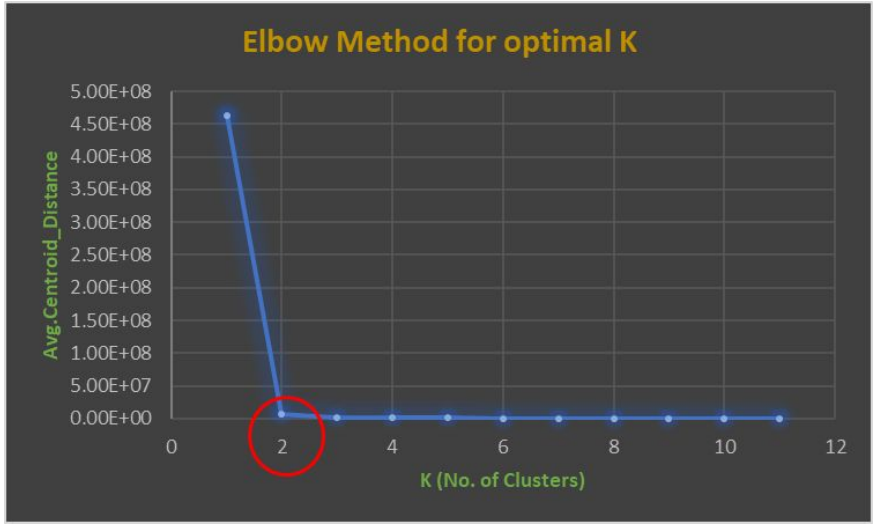
First Attribute	Second Attribute	Correlation ↑
Income	NumWebVisitsMo...	-0.650
MntMeatProducts	NumWebVisitsMo...	-0.539
Education = Gradua...	Education = PhD	-0.531
NumCatalogPurcha...	NumWebVisitsMo...	-0.522

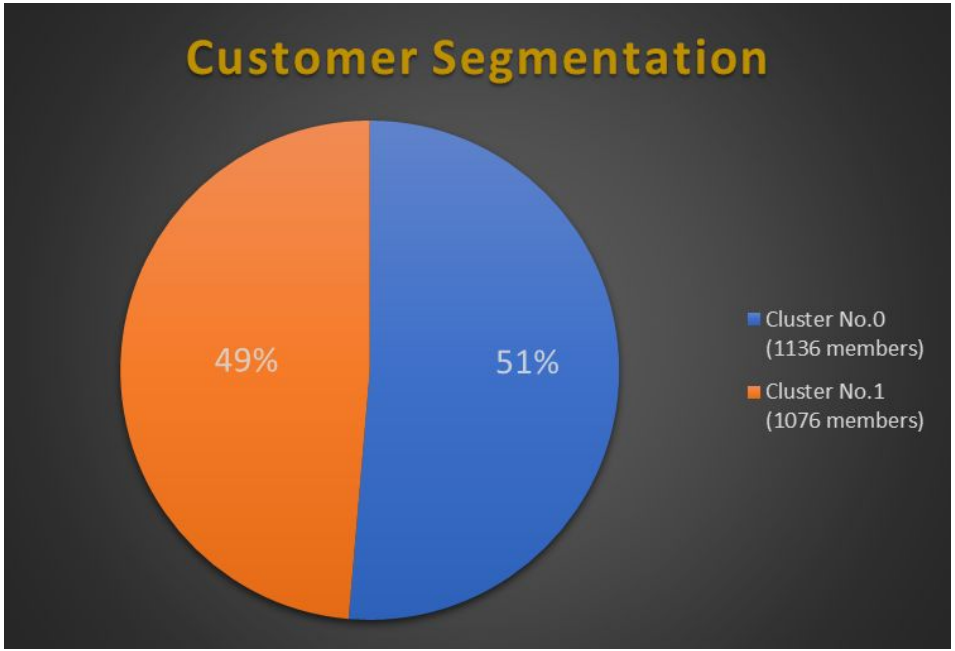
abs < 0.8 : Not Very Strong

Keep all feature in this stage

Modeling

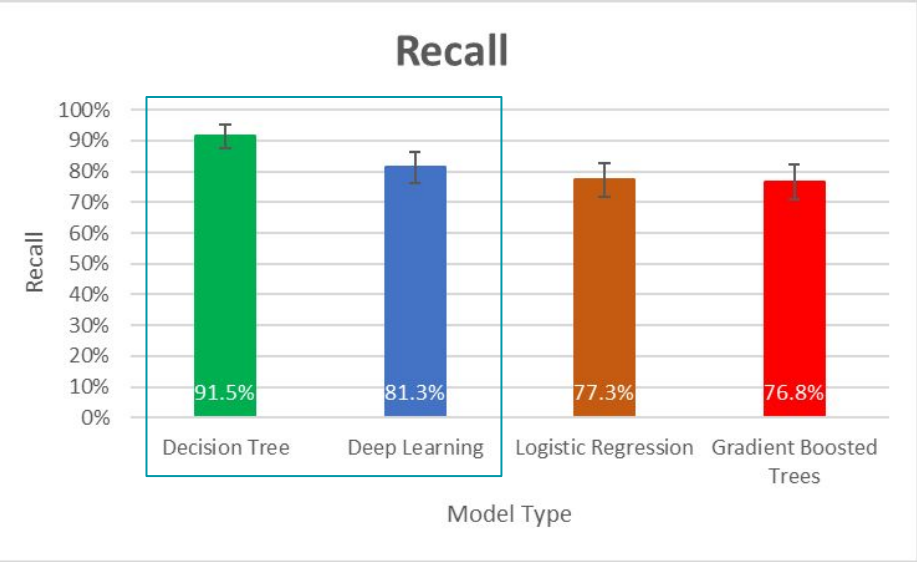
Clustering Model

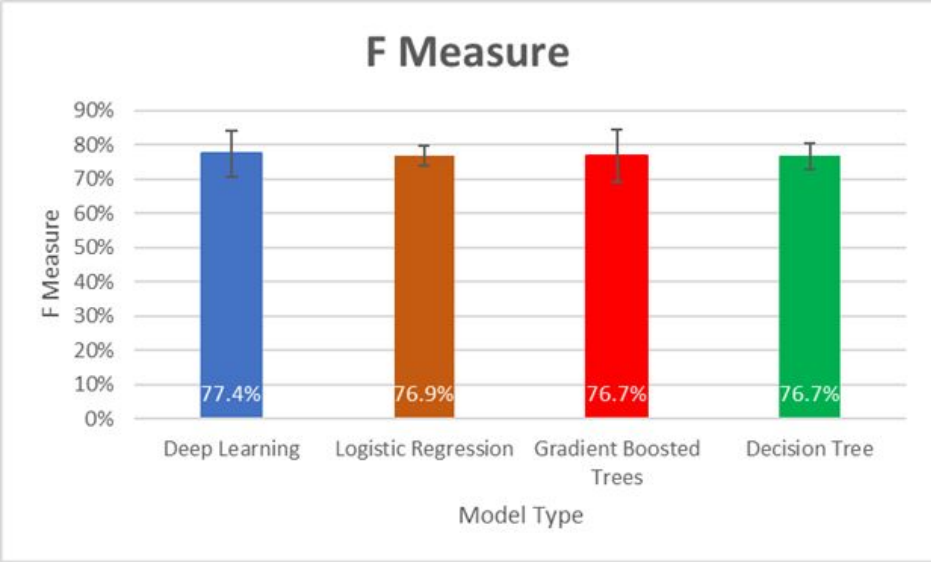




Classification Model

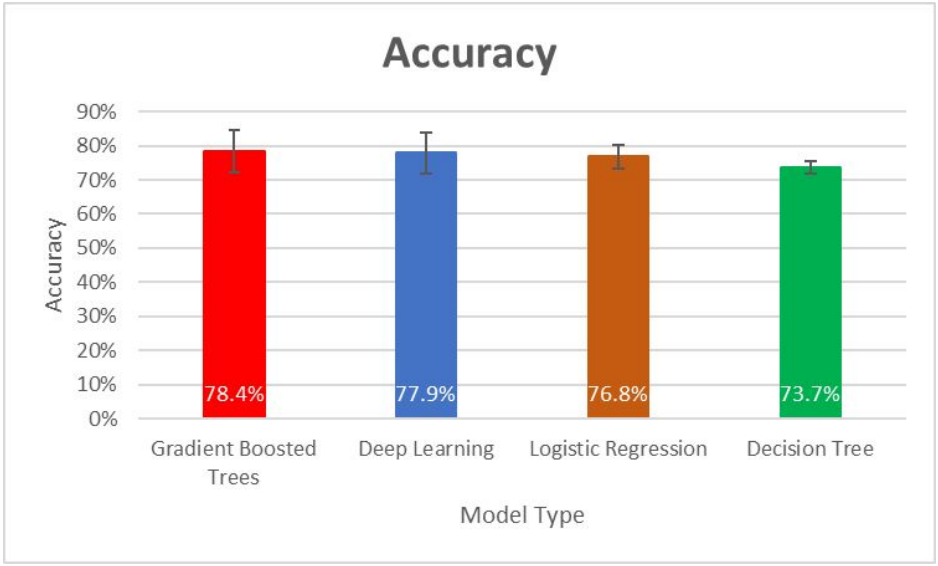
➢ **Pre-model Selection (by Auto Model)**

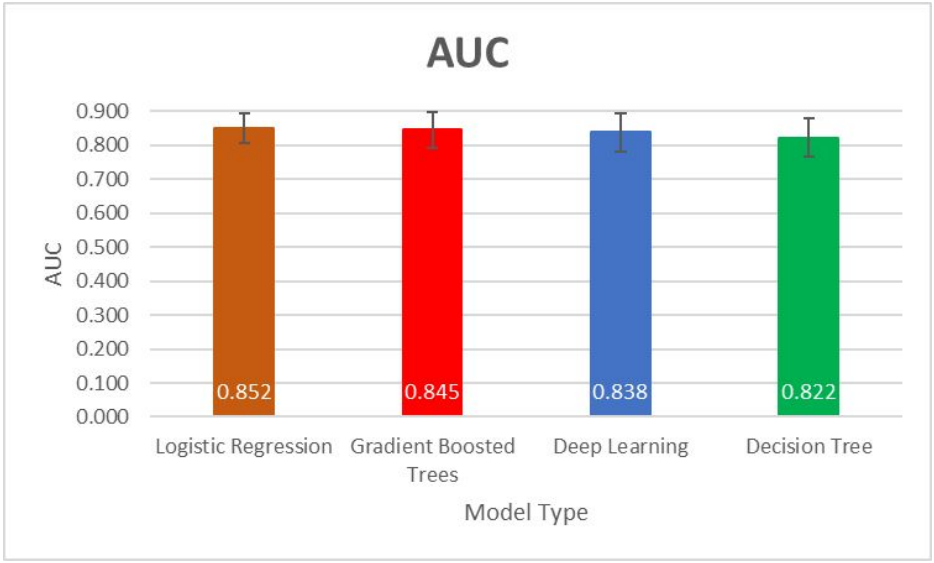




Classification Model

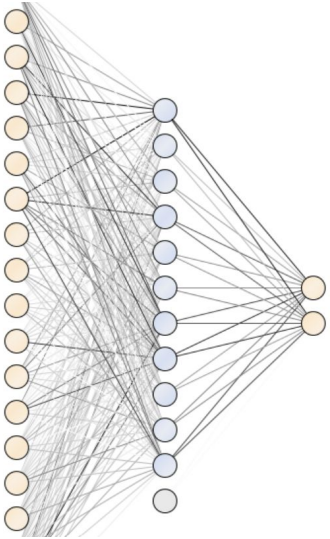
➢ **Pre-model Selection (by Auto Model)**





Classification Model

- **Production Model**



Neural Network

Accuracy : 86.94 %

True Precision : 86.18 %

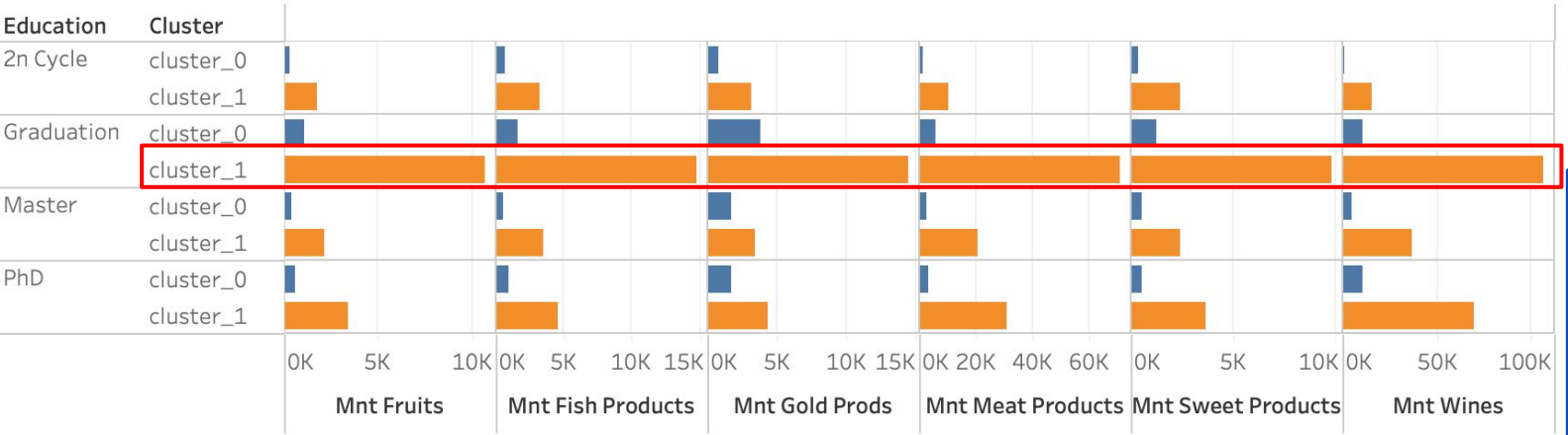
True Recall : 87.99 %

Impact

Clustering Result

Characteristic

Prod Edu



Sum of Mnt Fruits, sum of Mnt Fish Products, sum of Mnt Gold Prods, sum of Mnt Meat Products, sum of Mnt Sweet Products and sum of Mnt Wines for each Cluster broken down by Education. Color shows details about Cluster. The view is filtered on Education, which keeps 2n Cycle, Graduation, Master and PhD.

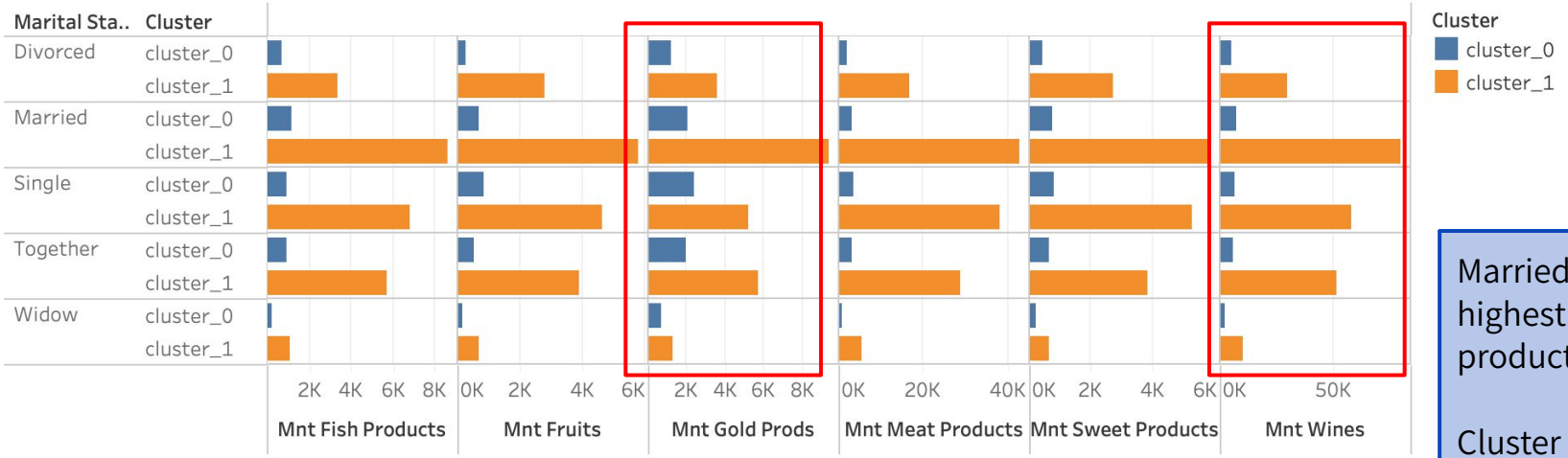
Graduated group are our main target customer.

Cluster 1 has higher in numbers for all products.

Clustering Result

Characteristic

Prod Marital



Sum of Mnt Fish Products, sum of Mnt Fruits, sum of Mnt Gold Prods, sum of Mnt Meat Products, sum of Mnt Sweet Products and sum of Mnt Wines for each Cluster broken down by Marital Status. Color shows details about Cluster. The view is filtered on Marital Status, which keeps Divorced, Married, Single, Together and Widow.

Married group are the highest buyer of all products.

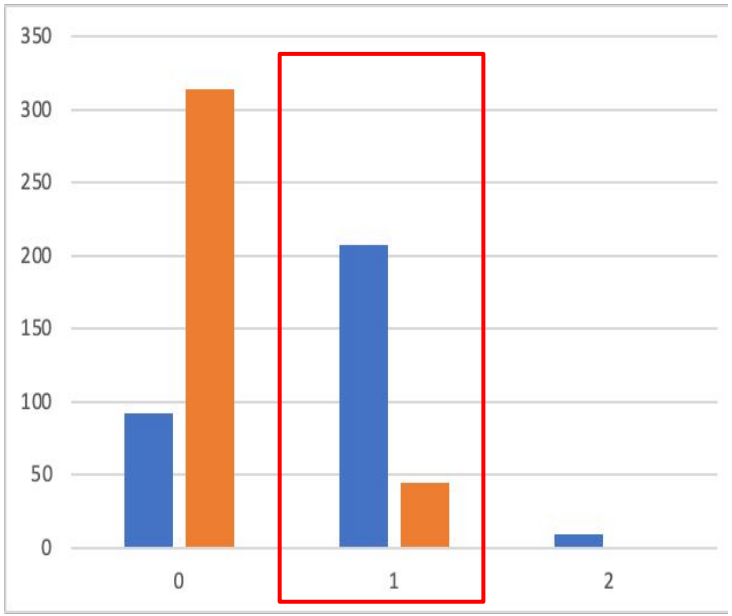
Cluster 0 mostly spend on gold.

Cluster 1 are highly spend on wines

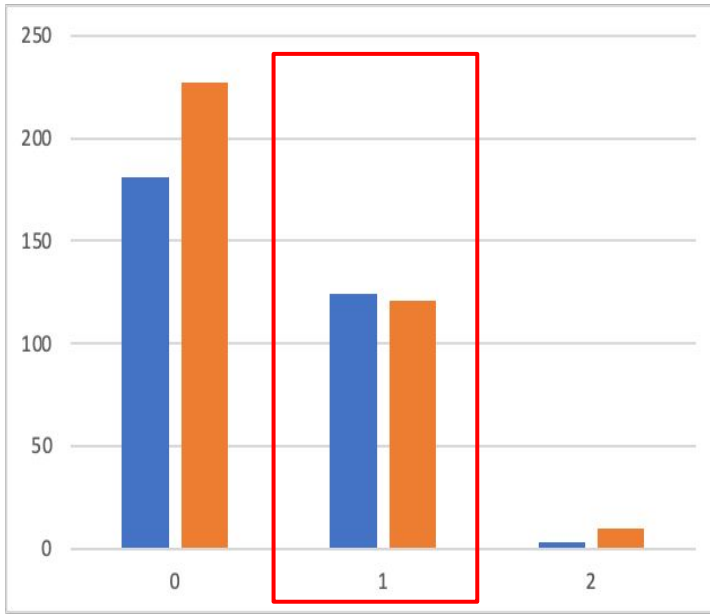
Clustering Result

Characteristic

Cluster by count of kids



Cluster by count of Teens



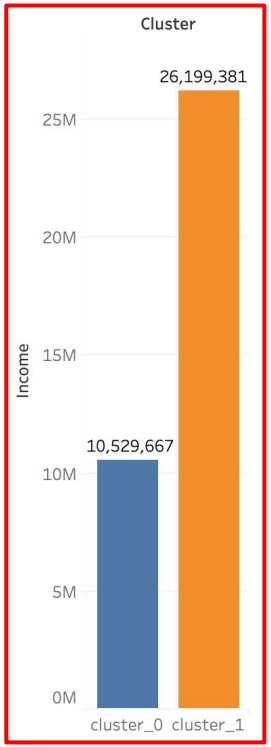
Cluster 1 are more like to have 0 child both on kids and teens.

Where cluster 0 are more popular with 1 or more kids, and slightly higher percentage of 1 teen

Clustering Result

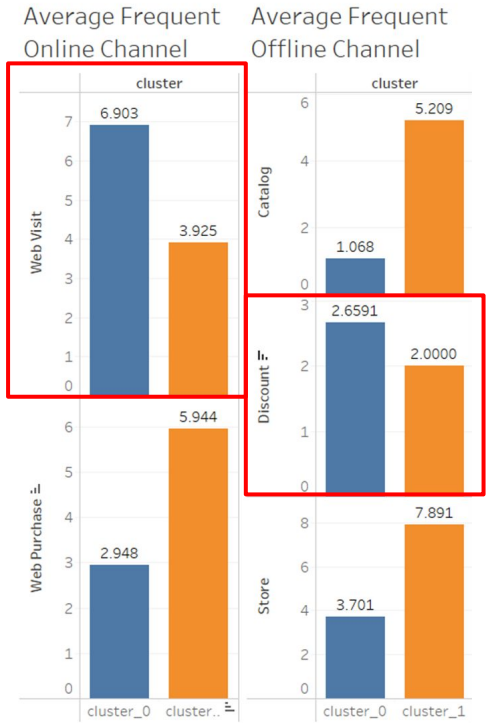
Characteristic

Cluster by income



Sum of Income for each Cluster.
Color shows details about Cluster.

Cluster by channel

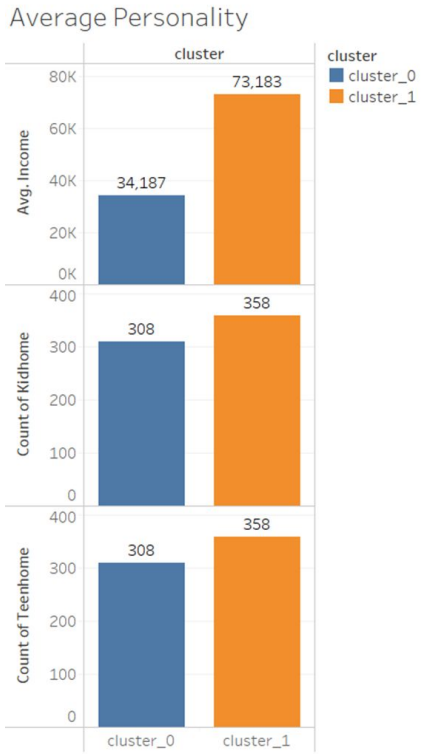


For the data, Cluster 1 have higher income and larger basket size.

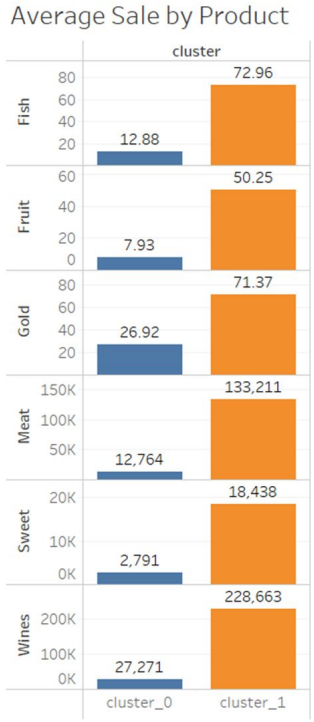
Cluster 0 has interest in purchasing from discount.

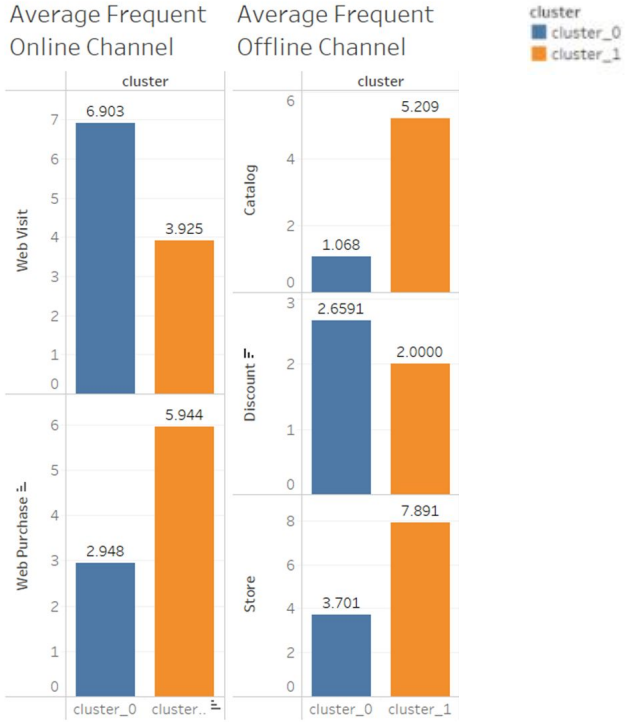
Clustering Result

Characteristic



Average of Income, count of Kidhome and count of Teenhome for each cluster. Color shows details about cluster.





Clustering Result

Characteristic

Cluster 0

- **Income:** Approximately 0-50K
- **Kid Home:** More likely to have >= 1 kids
- **Teenhome:** More likely to have teen at home
- **Spending:** Buy in small value or likely not to buy products
  - Fish/Fruit/ Gold/ Meats/ Sweet/ Wine
- **Frequent Channel:**
  - Discount
  - High Website but low avg purchase (3 time/month)

Cluster 1

- **Income:** approximate 50 - 100K
- **Kid Home:** Are likely to have no kids
- **Teenhome:** No teen at home
- **Spending:** Large spending range with higher value
  - Fish/Fruit/ Gold/ Meats/ Sweet/ Wine
- **Frequent Channel:**
  - Catalog, Store, Web
  - Low website visit but higher avg purchase (6 time/month)

Both Cluster purchase through Web more often than other channel