



## **Why this topic-**

1. Customers are one of the top 4C's In the market.
2. Marketers have long tried to appeal to consumers in terms of their personality characteristics.
3. Consumers purchase, and when and how they consume, are likely to be influenced by personality factors.
4. To improve customer's journey.
5. To optimize the significance of each customer to the business.

Objective

Problem Statement

Past Studies

Our questions

## Objective

- The goal of the project is segmentation to foresee the needs of customers.
- Get to know their interests, lifestyles, priorities and learn their spending habits.
- Acquisition and Retention
- Increased Revenues
- Enhanced Competitiveness





## **Why this topic-**

1. Customers are one of the top 4C's In the market.
2. Marketers have long tried to appeal to consumers in terms of their personality characteristics.
3. Consumers purchase, and when and how they consume, are likely to be influenced by personality factors.
4. To improve customer's journey.
5. To optimize the significance of each customer to the business.

Objective

Problem Statement

Past Studies

Our questions

- It helps a business to better understand its customers.
- Makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers.
- Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments.





## **Why this topic-**

1. Customers are one of the top 4C's In the market.
2. Marketers have long tried to appeal to consumers in terms of their personality characteristics.
3. Consumers purchase, and when and how they consume, are likely to be influenced by personality factors.
4. To improve customer's journey.
5. To optimize the significance of each customer to the business.

Objective

Problem Statement

Past Studies

Our questions

## **Avenues explored:-**





## **Why this topic-**

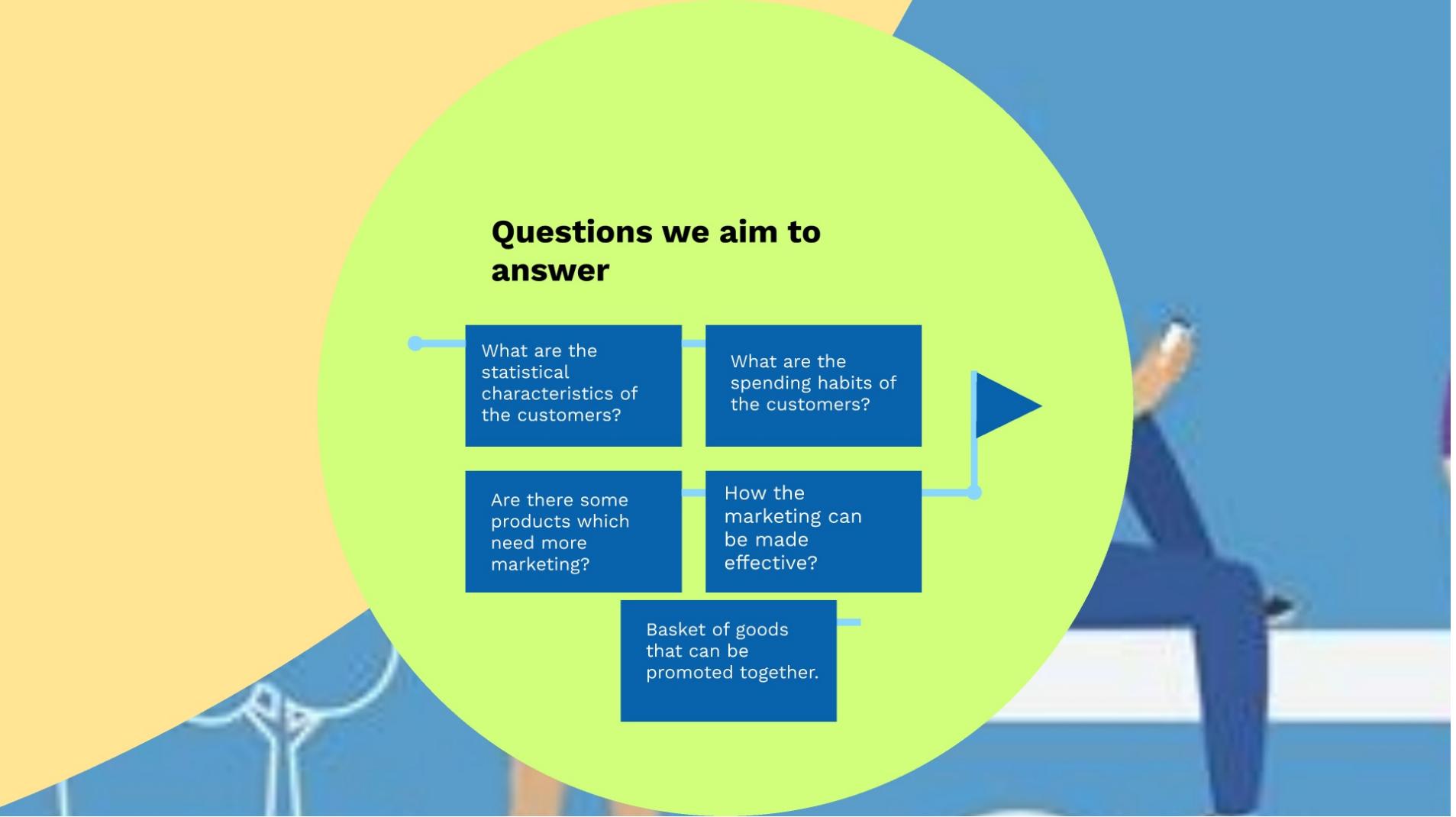
1. Customers are one of the top 4C's In the market.
2. Marketers have long tried to appeal to consumers in terms of their personality characteristics.
3. Consumers purchase, and when and how they consume, are likely to be influenced by personality factors.
4. To improve customer's journey.
5. To optimize the significance of each customer to the business.

Objective

Problem Statement

Past Studies

Our questions



## **Questions we aim to answer**

What are the statistical characteristics of the customers?

Are there some products which need more marketing?

Basket of goods that can be promoted together.

What are the spending habits of the customers?

How the marketing can be made effective?



## **Why this topic-**

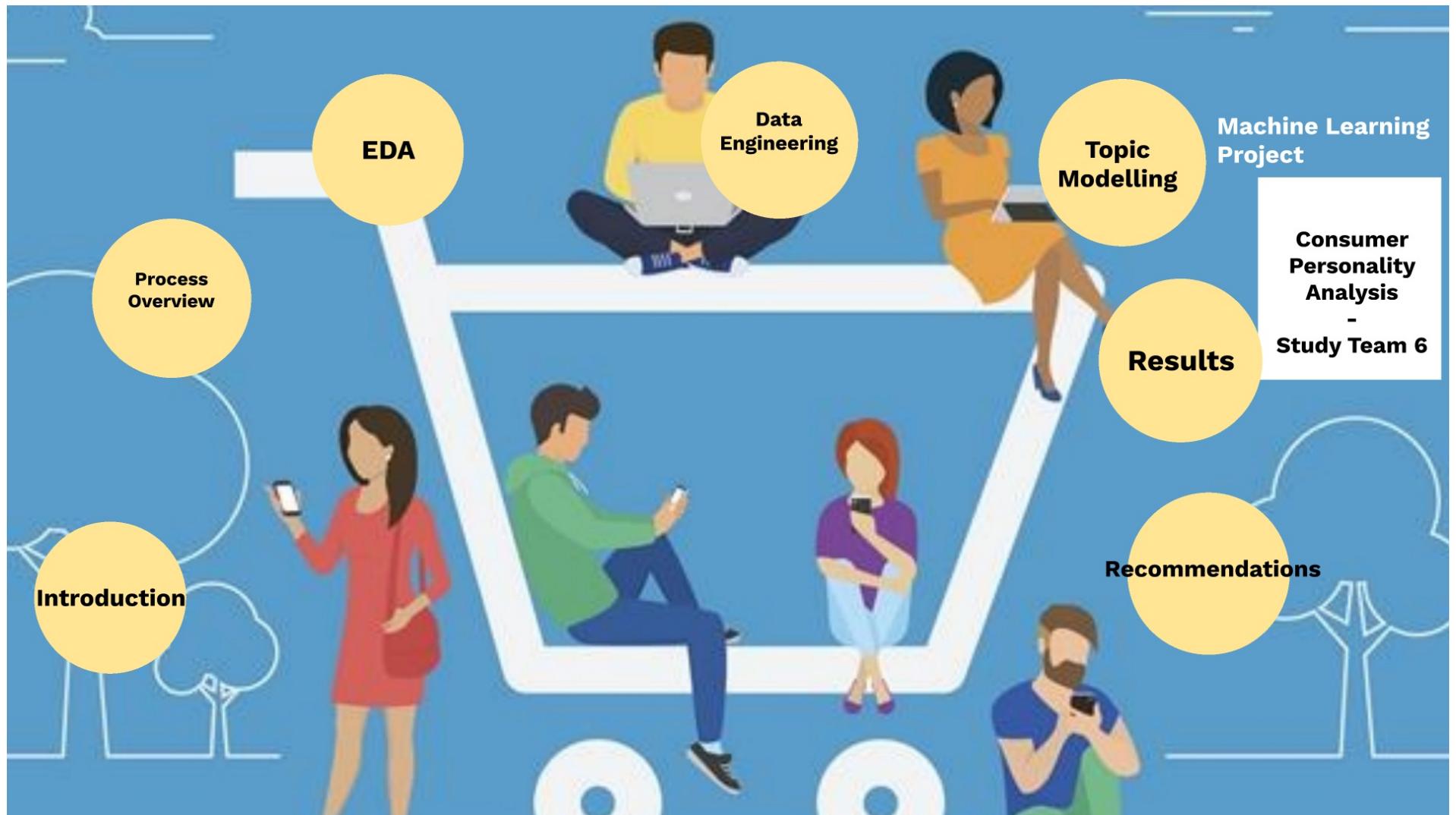
1. Customers are one of the top 4C's In the market.
2. Marketers have long tried to appeal to consumers in terms of their personality characteristics.
3. Consumers purchase, and when and how they consume, are likely to be influenced by personality factors.
4. To improve customer's journey.
5. To optimize the significance of each customer to the business.

Objective

Problem Statement

Past Studies

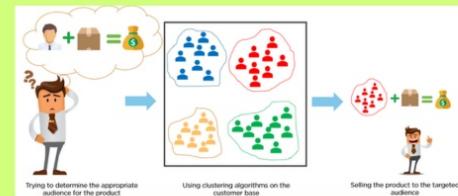
Our questions





# Roadmap

## Walk in through our project



Data Cleaning

Exploratory Data  
Analysis

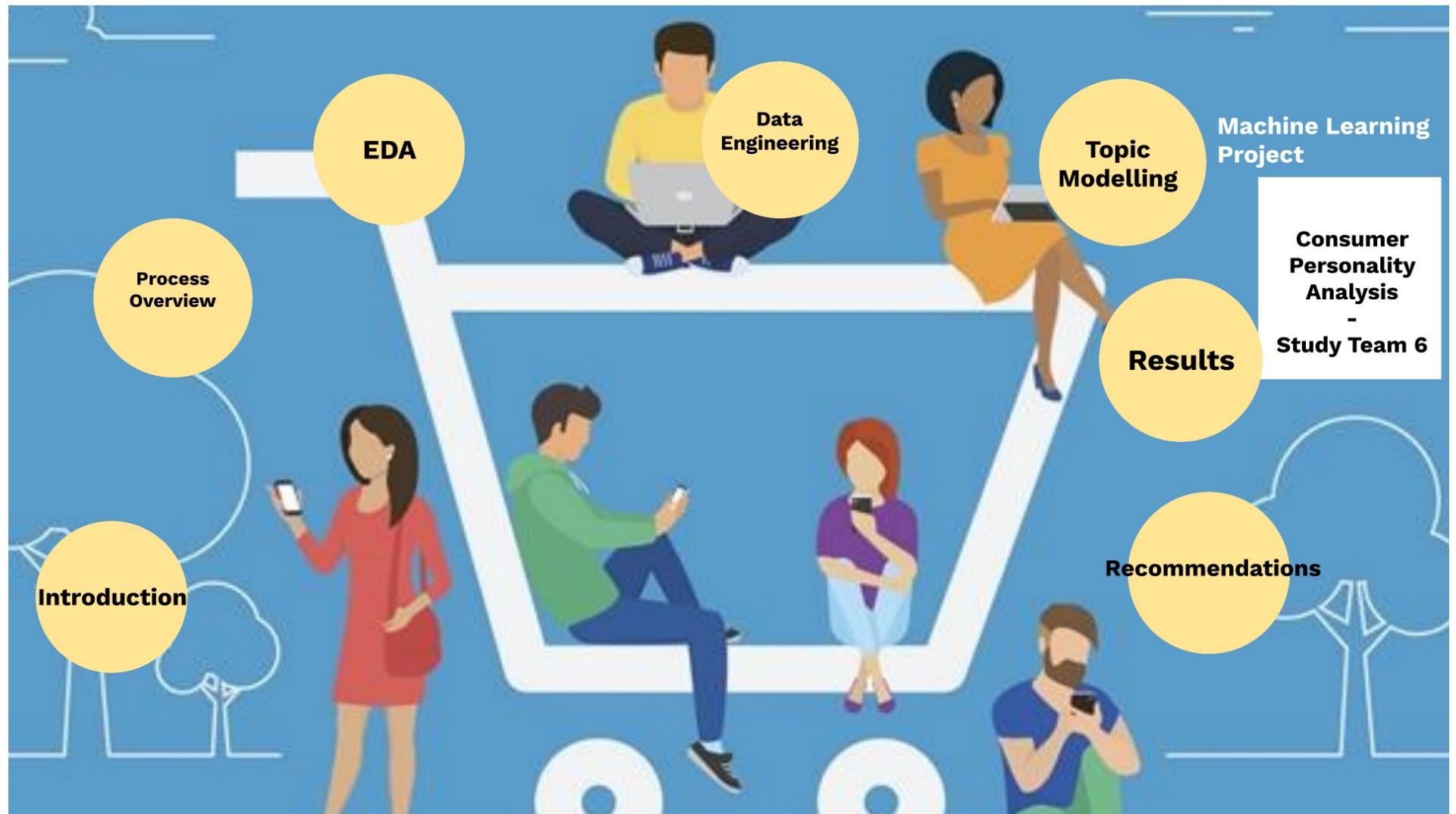
Topic modelling &  
approach

Results

Key Takeaways



# Roadmap



# Exploratory Data Analysis

## Dataset

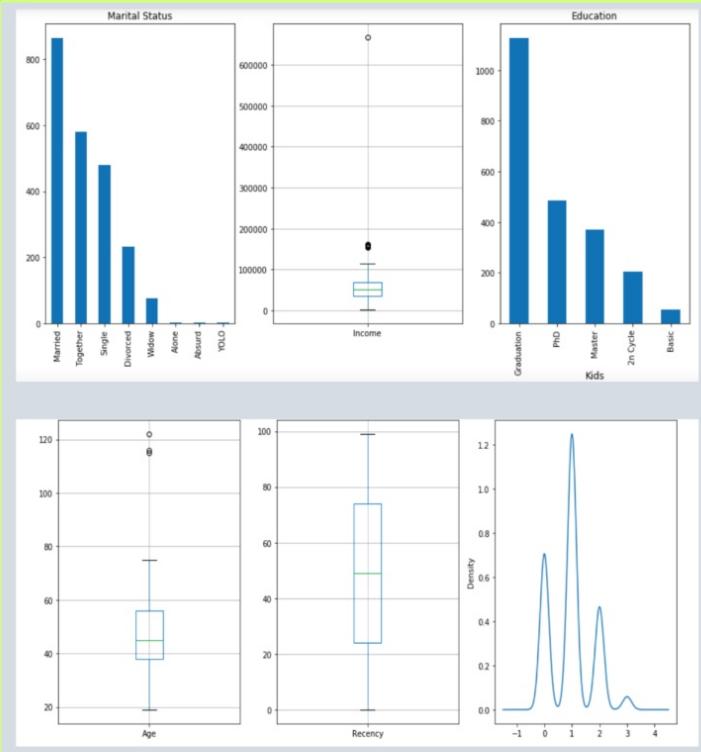
Attributes	
<b>ID</b>	Customer's unique identifier to the firm.
<b>Year_birth</b>	Customer's birth year.
<b>Education</b>	Customer's education level i.e. highschool, undergraduate or post graduate.
<b>Marital_Status</b>	Customer's marital status i.e. single, married, divorced etc.
<b>Income</b>	Customer's yearly household income
<b>Kidhome</b>	Number of children in customer's household
<b>Teenhome</b>	Number of teenagers in customer's household
<b>Dt_Customer</b>	Date of customer's enrollment with the company
<b>Recency</b>	Number of days since customer's last purchase
<b>Complain</b>	It is a dummy variable which gives 1 if customer complained in the last 2 years, 0 otherwise
<b>MntWines</b>	Amount spent on wine in last 2 years
<b>MntFruits</b>	Amount spent on fruits in last 2 years
<b>MntMeatProducts</b>	Amount spent on meat in last 2 years
<b>MntFishProducts</b>	Amount spent on fish in last 2 years
<b>MntSweetProducts</b>	Amount spent on sweets in last 2 years
<b>MntGoldProds</b>	Amount spent on gold in last 2 years
<b>NumDealsPurchases</b>	Number of purchases made with a discount
<b>AcceptedCmp1</b>	1 if customer accepted the offer in the 1st campaign, 0 otherwise
<b>AcceptedCmp2</b>	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
<b>AcceptedCmp3</b>	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
<b>AcceptedCmp4</b>	1 if customer accepted the offer in the 4th campaign, 0 otherwise
<b>AcceptedCmp5</b>	1 if customer accepted the offer in the 5th campaign, 0 otherwise
<b>Response</b>	1 if customer accepted the offer in the last campaign, 0 otherwise
<b>NumWebPurchases</b>	Number of purchases made through the company's web site
<b>NumCatalogPurchases</b>	Number of purchases made using a catalogue
<b>NumStorePurchases</b>	Number of purchases made directly in stores
<b>NumWebVisitsMonth</b>	Number of visits to company's web site in the last month

Outlier check

Correlation check

Null Value check

Other charts



# Exploratory Data Analysis

## Dataset

Attributes	
<b>ID</b>	Customer's unique identifier to the firm.
<b>Year_birth</b>	Customer's birth year.
<b>Education</b>	Customer's education level i.e. highschool, undergraduate or post graduate.
<b>Marital_Status</b>	Customer's marital status i.e. single, married, divorced etc.
<b>Income</b>	Customer's yearly household income
<b>Kidhome</b>	Number of children in customer's household
<b>Teenhome</b>	Number of teenagers in customer's household
<b>Dt_Customer</b>	Date of customer's enrollment with the company
<b>Recency</b>	Number of days since customer's last purchase
<b>Complain</b>	It is a dummy variable which gives 1 if customer complained in the last 2 years, 0 otherwise
<b>MntWines</b>	Amount spent on wine in last 2 years
<b>MntFruits</b>	Amount spent on fruits in last 2 years
<b>MntMeatProducts</b>	Amount spent on meat in last 2 years
<b>MntFishProducts</b>	Amount spent on fish in last 2 years
<b>MntSweetProducts</b>	Amount spent on sweets in last 2 years
<b>MntGoldProds</b>	Amount spent on gold in last 2 years
<b>NumDealsPurchases</b>	Number of purchases made with a discount
<b>AcceptedCmp1</b>	1 if customer accepted the offer in the 1st campaign, 0 otherwise
<b>AcceptedCmp2</b>	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
<b>AcceptedCmp3</b>	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
<b>AcceptedCmp4</b>	1 if customer accepted the offer in the 4th campaign, 0 otherwise
<b>AcceptedCmp5</b>	1 if customer accepted the offer in the 5th campaign, 0 otherwise
<b>Response</b>	1 if customer accepted the offer in the last campaign, 0 otherwise
<b>NumWebPurchases</b>	Number of purchases made through the company's web site
<b>NumCatalogPurchases</b>	Number of purchases made using a catalogue
<b>NumStorePurchases</b>	Number of purchases made directly in stores
<b>NumWebVisitsMonth</b>	Number of visits to company's web site in the last month

Outlier check

Correlation check

Null Value check

Other charts



# Exploratory Data Analysis

## Dataset

Attributes	
<b>ID</b>	Customer's unique identifier to the firm.
<b>Year_birth</b>	Customer's birth year.
<b>Education</b>	Customer's education level i.e. highschool, undergraduate or post graduate.
<b>Marital_Status</b>	Customer's marital status i.e. single, married, divorced etc.
<b>Income</b>	Customer's yearly household income
<b>Kidhome</b>	Number of children in customer's household
<b>Teenhome</b>	Number of teenagers in customer's household
<b>Dt_Customer</b>	Date of customer's enrollment with the company
<b>Recency</b>	Number of days since customer's last purchase
<b>Complain</b>	It is a dummy variable which gives 1 if customer complained in the last 2 years, 0 otherwise
<b>MntWines</b>	Amount spent on wine in last 2 years
<b>MntFruits</b>	Amount spent on fruits in last 2 years
<b>MntMeatProducts</b>	Amount spent on meat in last 2 years
<b>MntFishProducts</b>	Amount spent on fish in last 2 years
<b>MntSweetProducts</b>	Amount spent on sweets in last 2 years
<b>MntGoldProds</b>	Amount spent on gold in last 2 years
<b>NumDealsPurchases</b>	Number of purchases made with a discount
<b>AcceptedCmp1</b>	1 if customer accepted the offer in the 1st campaign, 0 otherwise
<b>AcceptedCmp2</b>	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
<b>AcceptedCmp3</b>	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
<b>AcceptedCmp4</b>	1 if customer accepted the offer in the 4th campaign, 0 otherwise
<b>AcceptedCmp5</b>	1 if customer accepted the offer in the 5th campaign, 0 otherwise
<b>Response</b>	1 if customer accepted the offer in the last campaign, 0 otherwise
<b>NumWebPurchases</b>	Number of purchases made through the company's web site
<b>NumCatalogPurchases</b>	Number of purchases made using a catalogue
<b>NumStorePurchases</b>	Number of purchases made directly in stores
<b>NumWebVisitsMonth</b>	Number of visits to company's web site in the last month

Outlier check

Correlation check

Null Value check

Other charts



# Exploratory Data Analysis

## Dataset

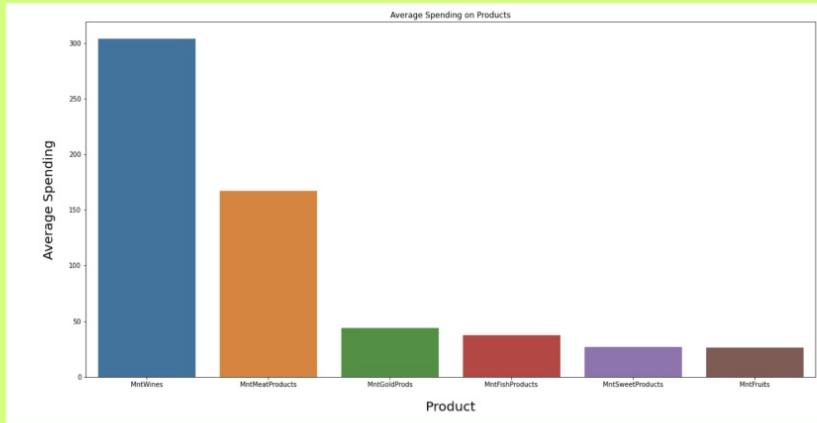
Attributes	
<b>ID</b>	Customer's unique identifier to the firm.
<b>Year_birth</b>	Customer's birth year.
<b>Education</b>	Customer's education level i.e. highschool, undergraduate or post graduate.
<b>Marital_Status</b>	Customer's marital status i.e. single, married, divorced etc.
<b>Income</b>	Customer's yearly household income
<b>Kidhome</b>	Number of children in customer's household
<b>Teenhome</b>	Number of teenagers in customer's household
<b>Dt_Customer</b>	Date of customer's enrollment with the company
<b>Recency</b>	Number of days since customer's last purchase
<b>Complain</b>	It is a dummy variable which gives 1 if customer complained in the last 2 years, 0 otherwise
<b>MntWines</b>	Amount spent on wine in last 2 years
<b>MntFruits</b>	Amount spent on fruits in last 2 years
<b>MntMeatProducts</b>	Amount spent on meat in last 2 years
<b>MntFishProducts</b>	Amount spent on fish in last 2 years
<b>MntSweetProducts</b>	Amount spent on sweets in last 2 years
<b>MntGoldProds</b>	Amount spent on gold in last 2 years
<b>NumDealsPurchases</b>	Number of purchases made with a discount
<b>AcceptedCmp1</b>	1 if customer accepted the offer in the 1st campaign, 0 otherwise
<b>AcceptedCmp2</b>	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
<b>AcceptedCmp3</b>	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
<b>AcceptedCmp4</b>	1 if customer accepted the offer in the 4th campaign, 0 otherwise
<b>AcceptedCmp5</b>	1 if customer accepted the offer in the 5th campaign, 0 otherwise
<b>Response</b>	1 if customer accepted the offer in the last campaign, 0 otherwise
<b>NumWebPurchases</b>	Number of purchases made through the company's web site
<b>NumCatalogPurchases</b>	Number of purchases made using a catalogue
<b>NumStorePurchases</b>	Number of purchases made directly in stores
<b>NumWebVisitsMonth</b>	Number of visits to company's web site in the last month

Outlier check

Correlation check

Null Value check

Other charts



# Exploratory Data Analysis

## Dataset

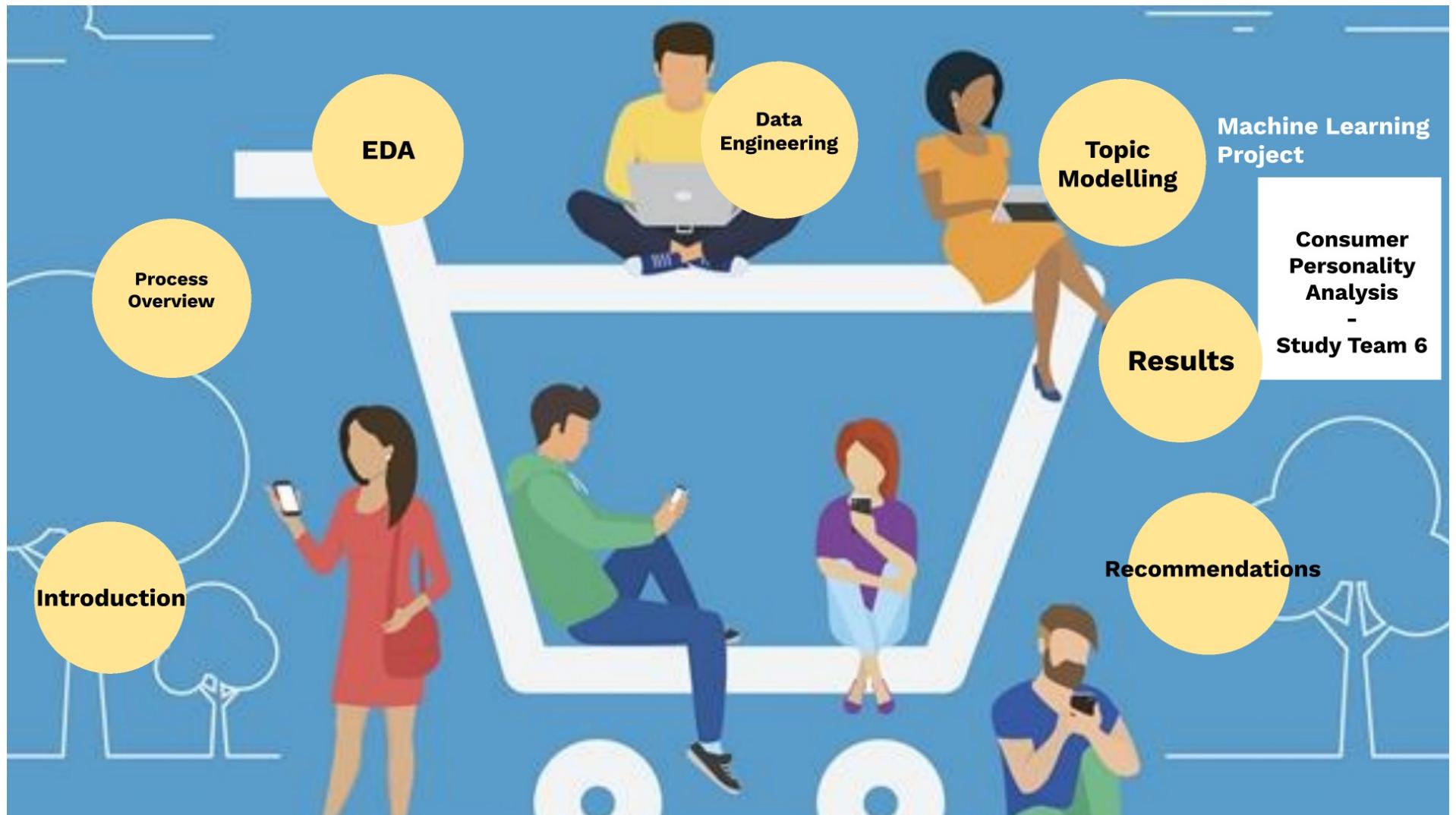
Attributes	
<b>ID</b>	Customer's unique identifier to the firm.
<b>Year_birth</b>	Customer's birth year.
<b>Education</b>	Customer's education level i.e. highschool, undergraduate or post graduate.
<b>Marital_Status</b>	Customer's marital status i.e. single, married, divorced etc.
<b>Income</b>	Customer's yearly household income
<b>Kidhome</b>	Number of children in customer's household
<b>Teenhome</b>	Number of teenagers in customer's household
<b>Dt_Customer</b>	Date of customer's enrollment with the company
<b>Recency</b>	Number of days since customer's last purchase
<b>Complain</b>	It is a dummy variable which gives 1 if customer complained in the last 2 years, 0 otherwise
<b>MntWines</b>	Amount spent on wine in last 2 years
<b>MntFruits</b>	Amount spent on fruits in last 2 years
<b>MntMeatProducts</b>	Amount spent on meat in last 2 years
<b>MntFishProducts</b>	Amount spent on fish in last 2 years
<b>MntSweetProducts</b>	Amount spent on sweets in last 2 years
<b>MntGoldProds</b>	Amount spent on gold in last 2 years
<b>NumDealsPurchases</b>	Number of purchases made with a discount
<b>AcceptedCmp1</b>	1 if customer accepted the offer in the 1st campaign, 0 otherwise
<b>AcceptedCmp2</b>	1 if customer accepted the offer in the 2nd campaign, 0 otherwise
<b>AcceptedCmp3</b>	1 if customer accepted the offer in the 3rd campaign, 0 otherwise
<b>AcceptedCmp4</b>	1 if customer accepted the offer in the 4th campaign, 0 otherwise
<b>AcceptedCmp5</b>	1 if customer accepted the offer in the 5th campaign, 0 otherwise
<b>Response</b>	1 if customer accepted the offer in the last campaign, 0 otherwise
<b>NumWebPurchases</b>	Number of purchases made through the company's web site
<b>NumCatalogPurchases</b>	Number of purchases made using a catalogue
<b>NumStorePurchases</b>	Number of purchases made directly in stores
<b>NumWebVisitsMonth</b>	Number of visits to company's web site in the last month

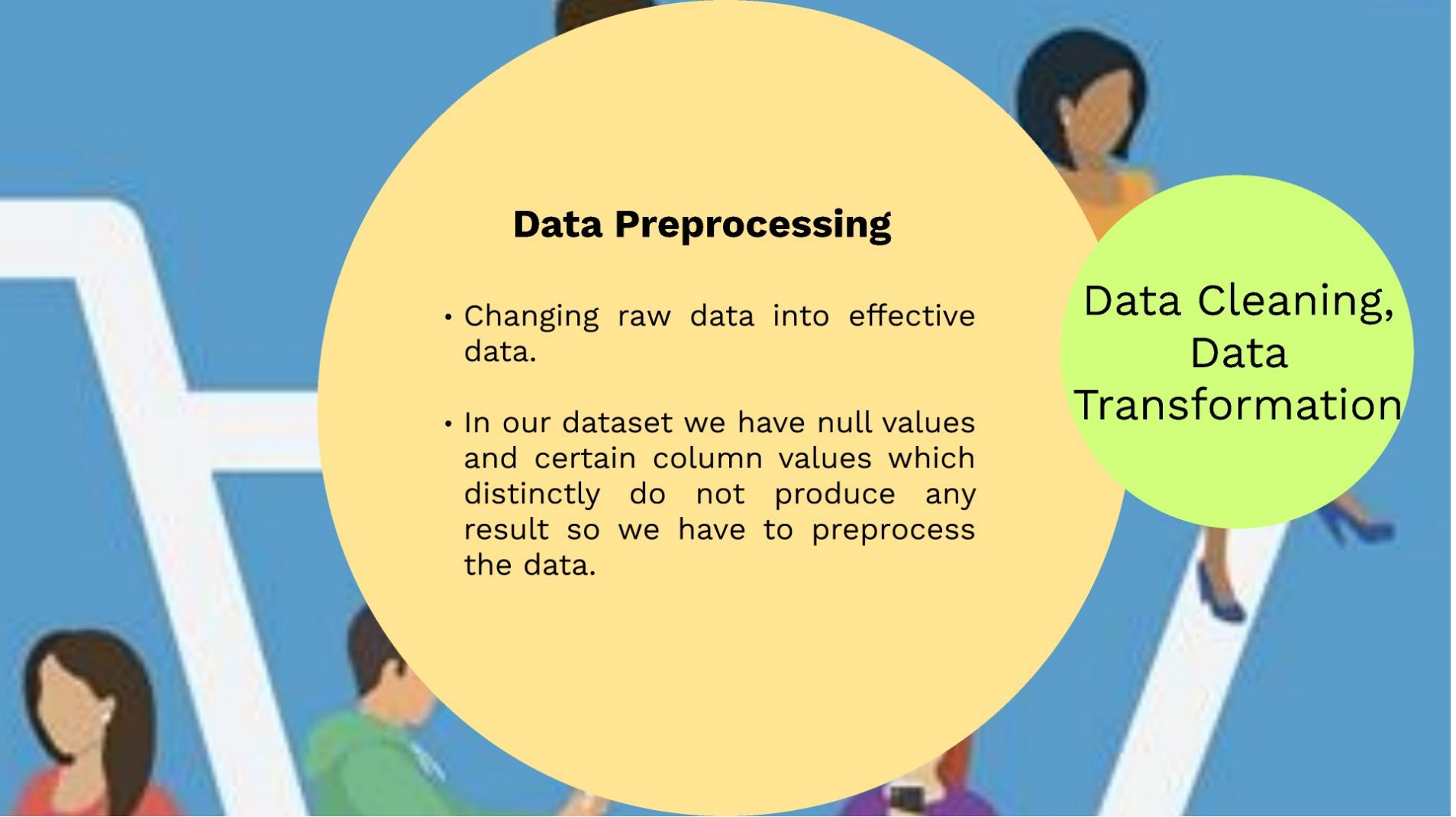
Outlier check

Correlation check

Null Value check

Other charts





## Data Preprocessing

- Changing raw data into effective data.
- In our dataset we have null values and certain column values which distinctly do not produce any result so we have to preprocess the data.

## Data Cleaning, Data Transformation

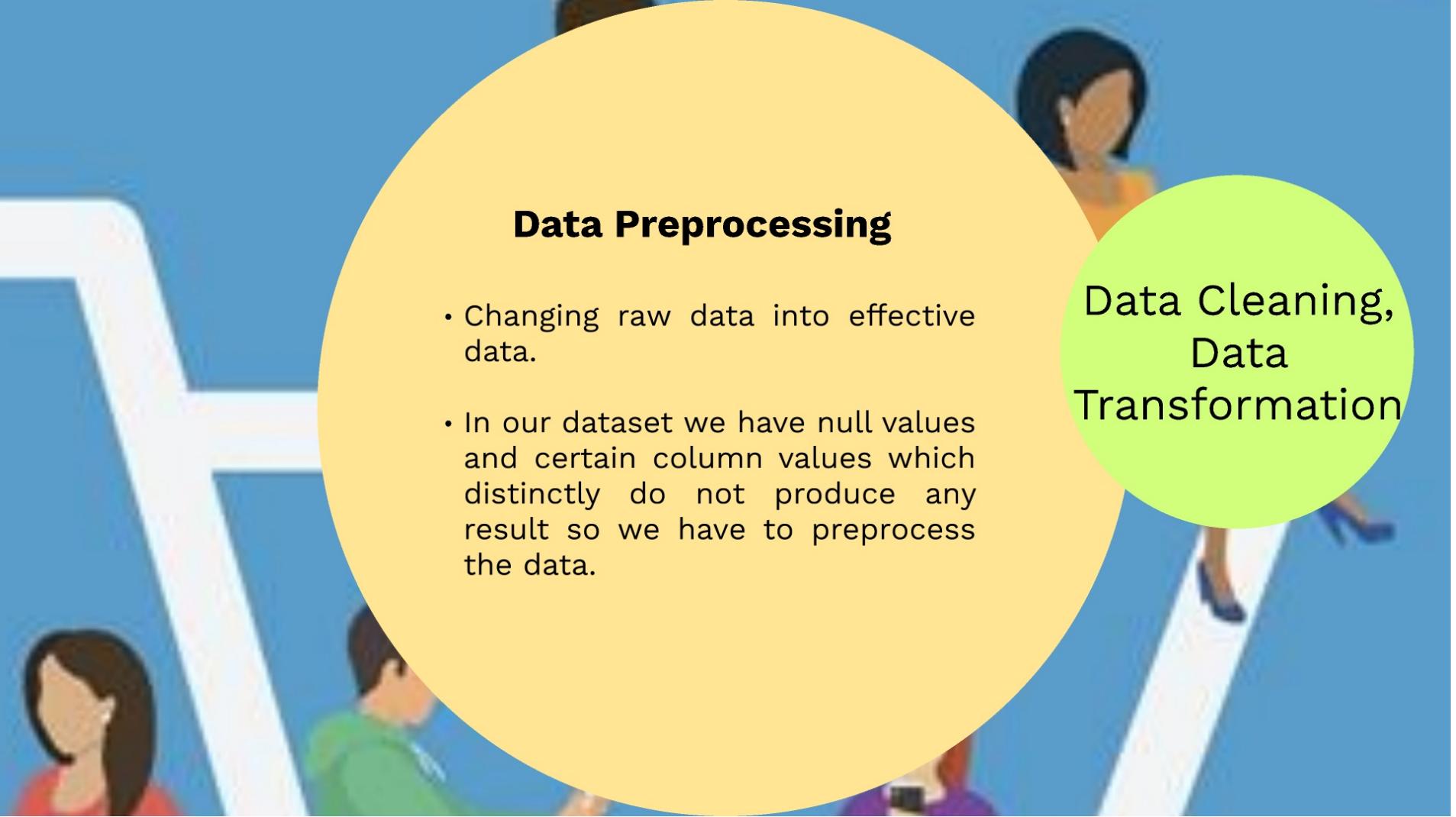
Data Cleaning	Data Transformation
Removed null values found in Income	Created 5 new columns
Removed outliers using the z score value	Rescaled the entire dataset for modelling

## For K means

	NumWebVisitsMonth	Complain	NumCatalogPurchases	NumStorePurchases	years_engaged	Age	Income	NumDealsPurchases	Response	Recency
0	0.692286	-0.097857	2.508988	-0.553716	1.501147	0.986656	0.234755	0.35145	2.380898	0.310523
1	-0.133089	-0.097857	-0.570224	-1.168833	-1.418079	1.237186	-0.233807	-0.16833	-0.420010	-0.380222
2	-0.545777	-0.097857	-0.228090	1.291634	0.041534	0.318575	0.770100	-0.68811	-0.420010	-0.794669
3	0.279599	-0.097857	-0.912359	-0.553716	-1.418079	-1.268116	-1.016386	-0.16833	-0.420010	-0.794669
4	-0.133089	-0.097857	0.114045	0.061401	-1.418079	-1.017586	0.240912	1.39101	-0.420010	1.553864

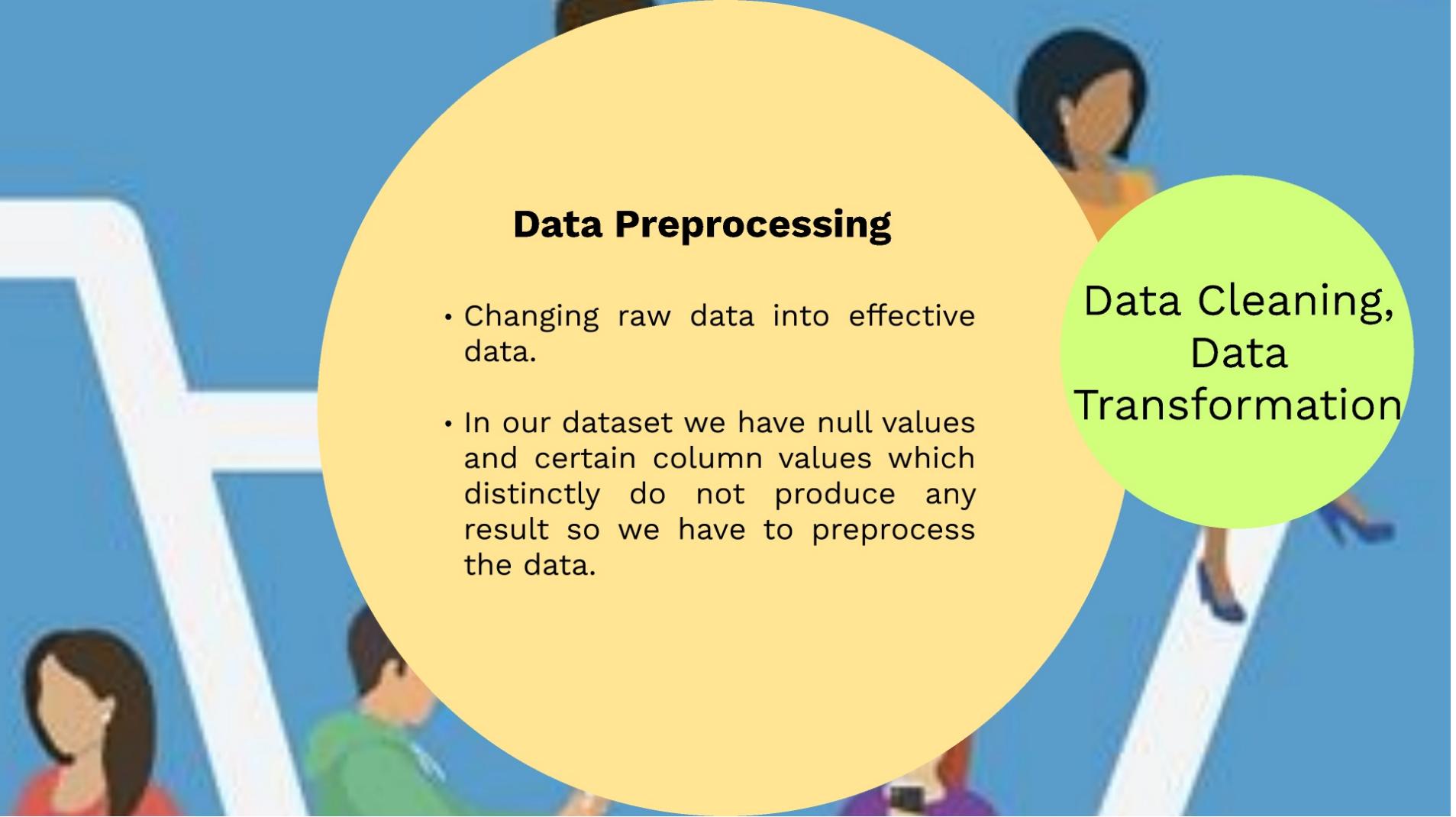
## For K prototypes

	Income	Kidhome	Teenhome	Recency	MntWines	MntFruits	MntMeatProducts	MntFishProducts	MntSweetProducts	MntGoldProd...	
kproto_labels	0	1.111037	-0.757872	-0.572718	0.070592	0.586651	1.141950	1.128633	1.200963	1.076763	0.634338
0	0.805873	-0.757872	-0.572718	0.070592	0.586651	1.141950	1.128633	1.200963	1.076763	0.634338	0.634338
1	1.111037	-0.726267	-0.665477	-0.165323	1.655301	0.659917	1.271325	0.771360	0.836439	0.625796	0.625796
2	-0.708426	0.645619	-0.107379	0.011200	-0.784710	-0.540724	-0.644081	-0.557975	-0.537406	-0.554900	-0.554900
3	0.215049	-0.323040	0.827982	-0.027857	0.472416	-0.125309	-0.135598	-0.167235	-0.125257	0.311356	0.311356

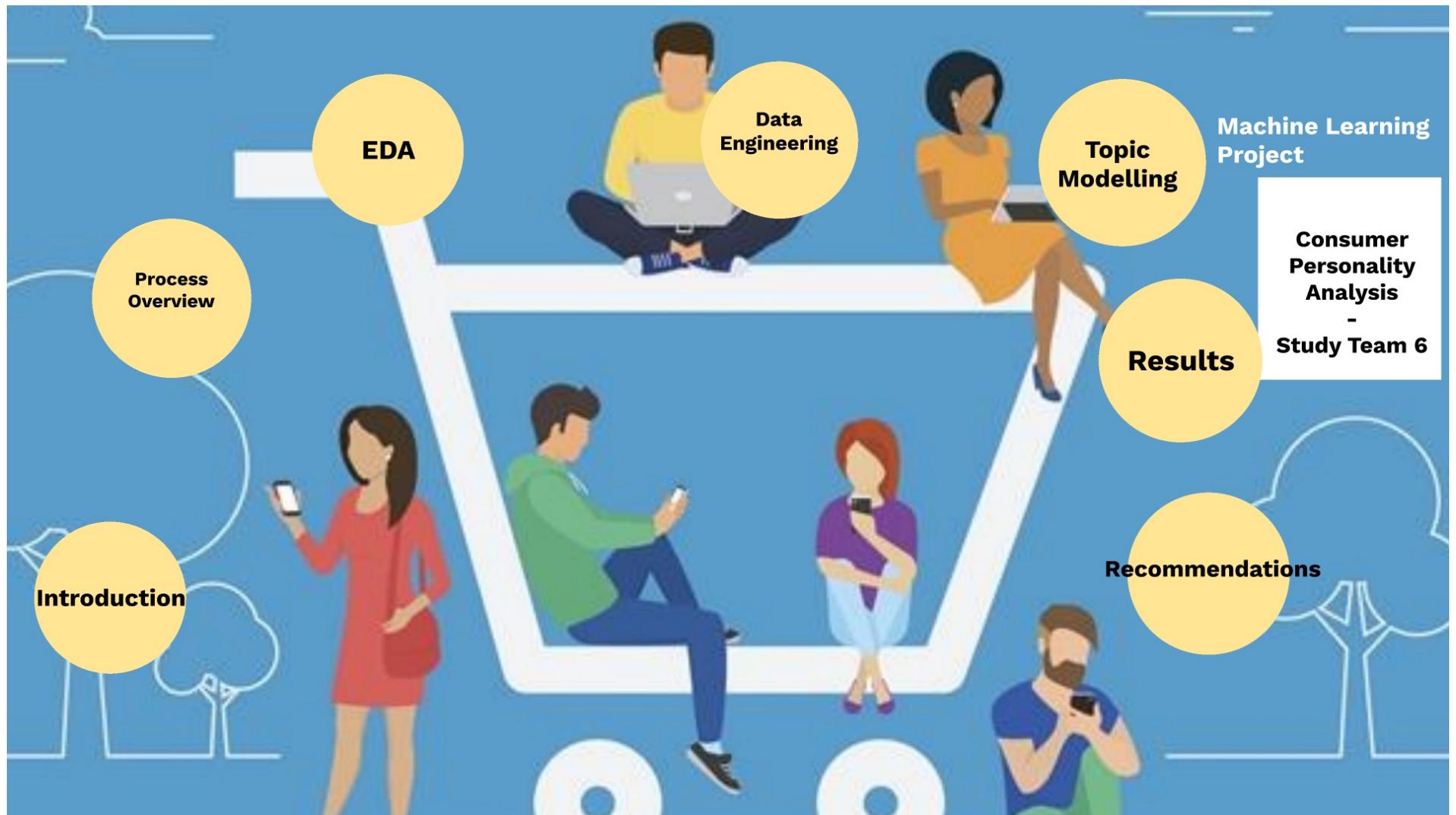


## Data Preprocessing

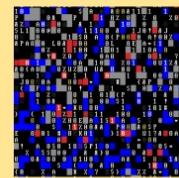
- Changing raw data into effective data.
- In our dataset we have null values and certain column values which distinctly do not produce any result so we have to preprocess the data.



## Data Cleaning, Data Transformation



## Models used

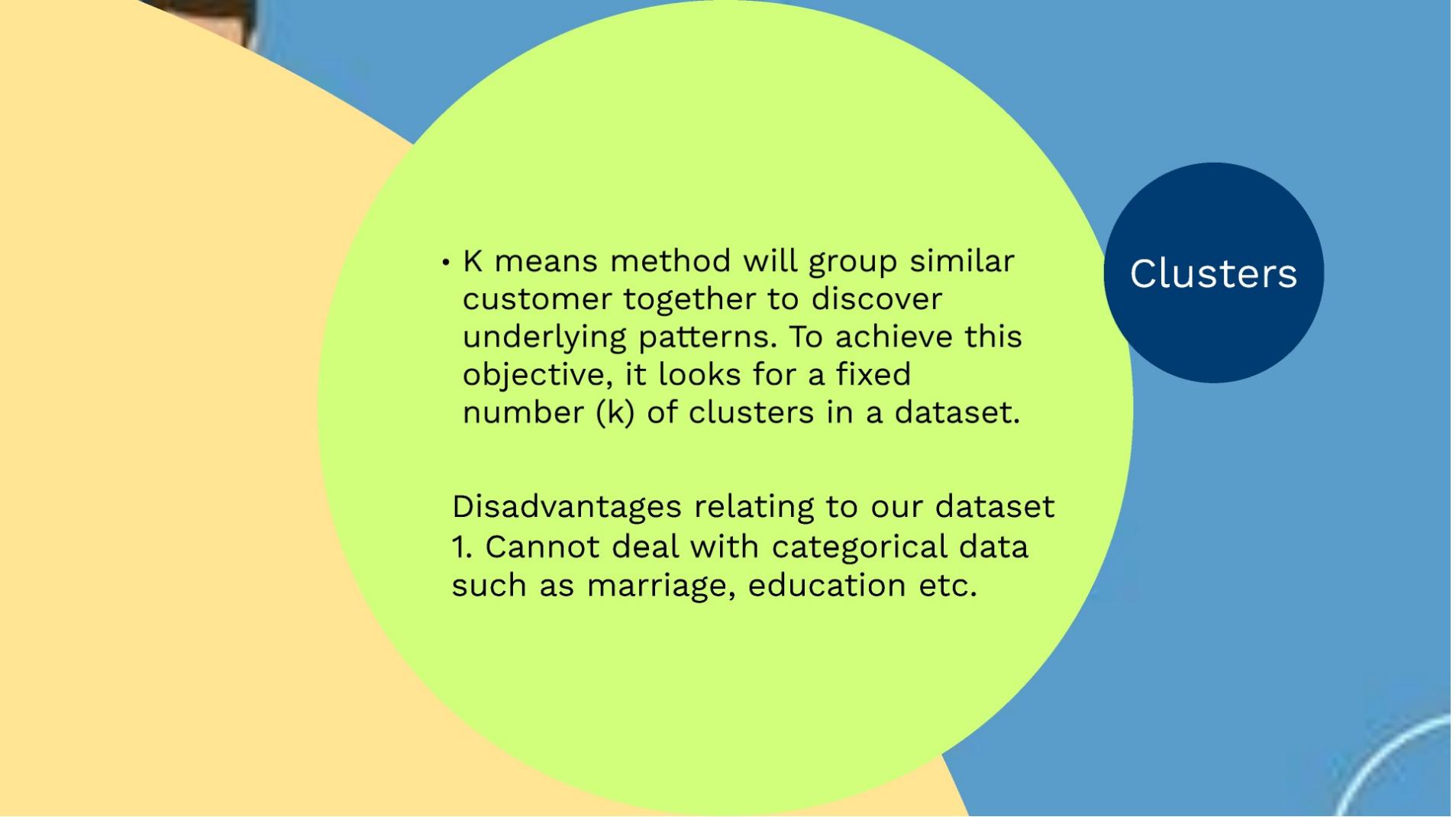


Clustering model-  
It involves automatically  
discovering natural grouping in  
data.

K  
means

K  
prototypes

Association  
rule



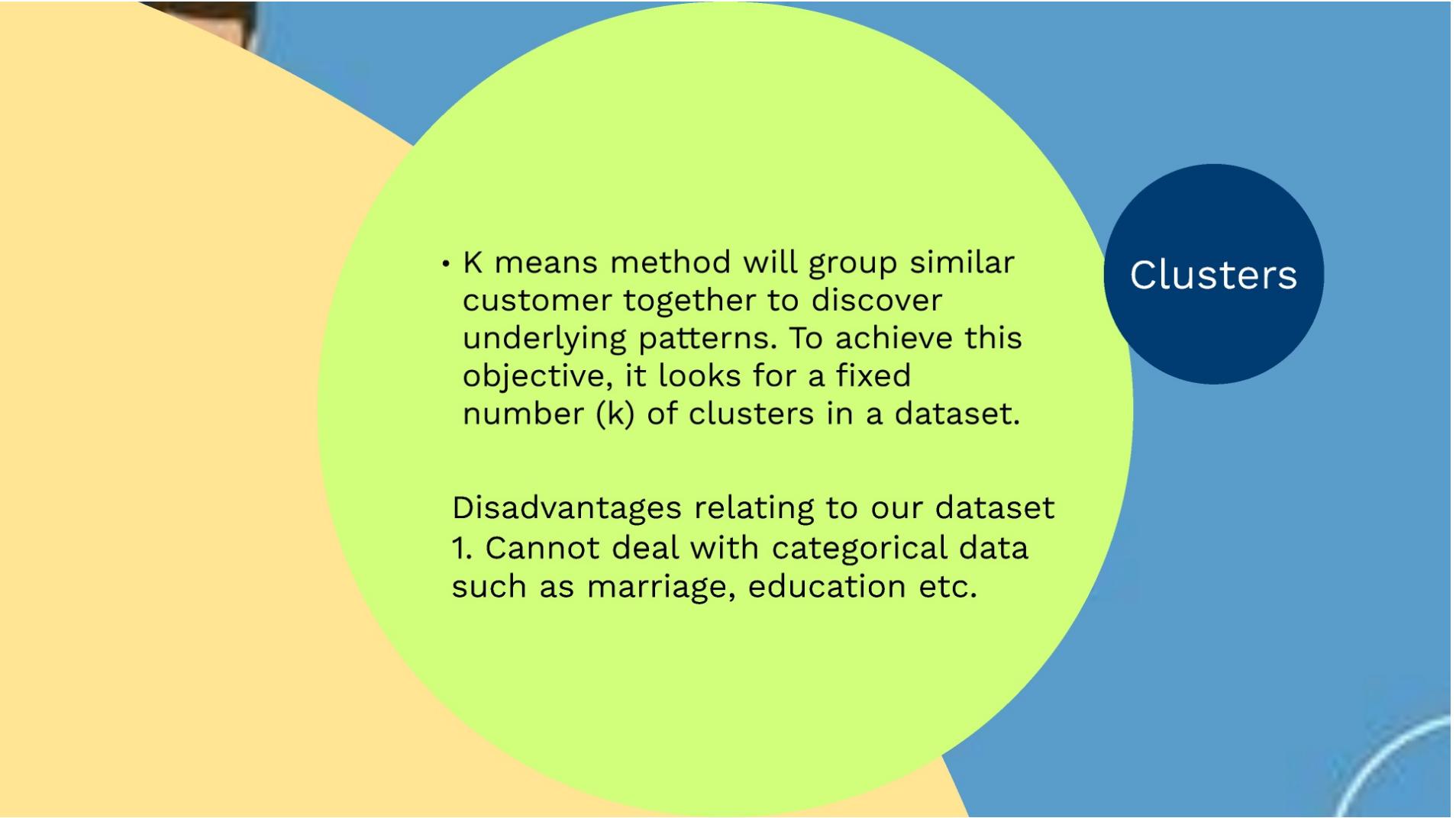
## Clusters

- K means method will group similar customer together to discover underlying patterns. To achieve this objective, it looks for a fixed number ( $k$ ) of clusters in a dataset.

Disadvantages relating to our dataset

1. Cannot deal with categorical data such as marriage, education etc.





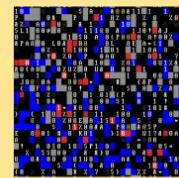
## Clusters

- K means method will group similar customer together to discover underlying patterns. To achieve this objective, it looks for a fixed number ( $k$ ) of clusters in a dataset.

Disadvantages relating to our dataset

1. Cannot deal with categorical data such as marriage, education etc.

## Models used

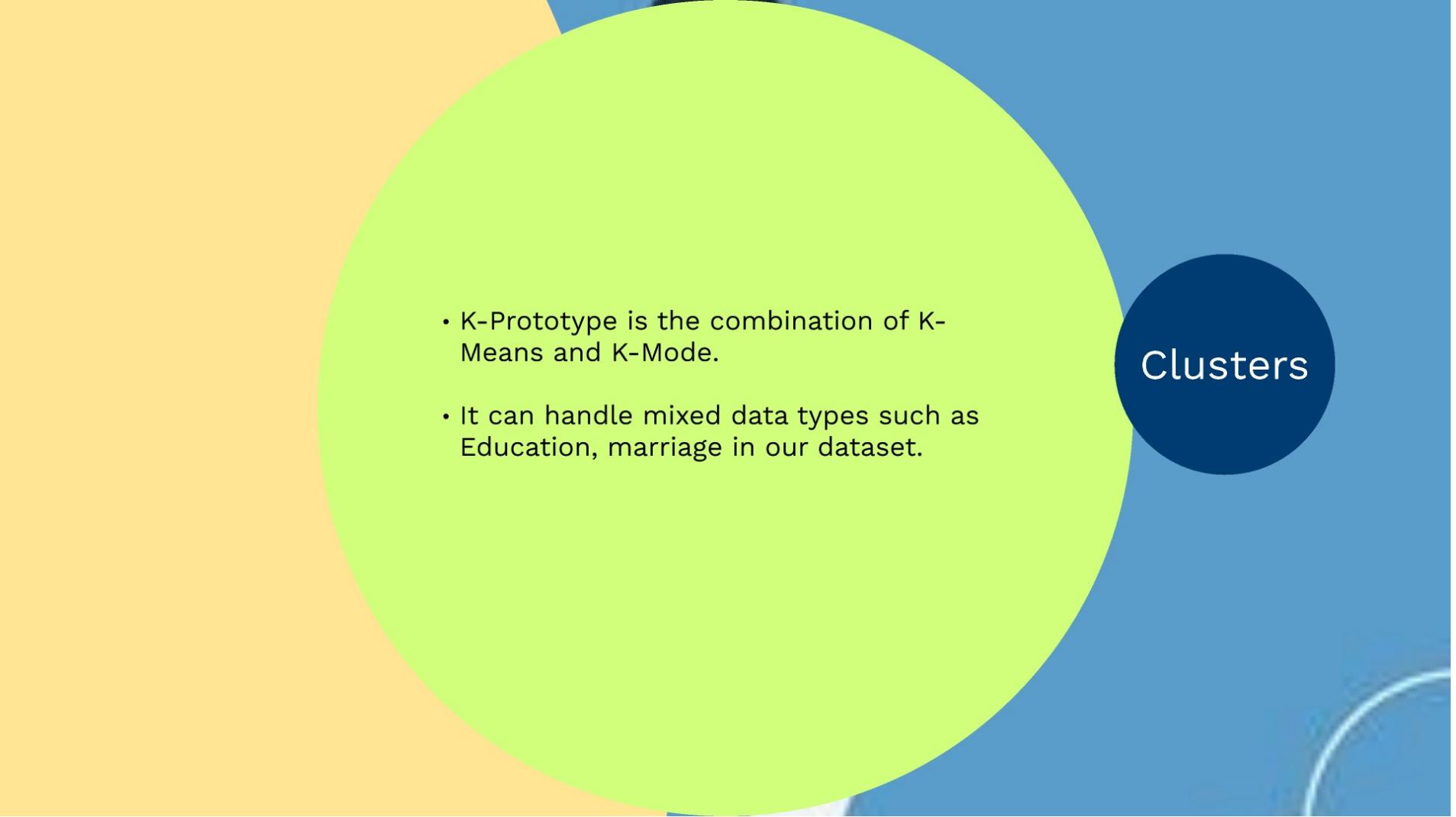


Clustering model-  
It involves automatically  
discovering natural grouping in  
data.

K  
means

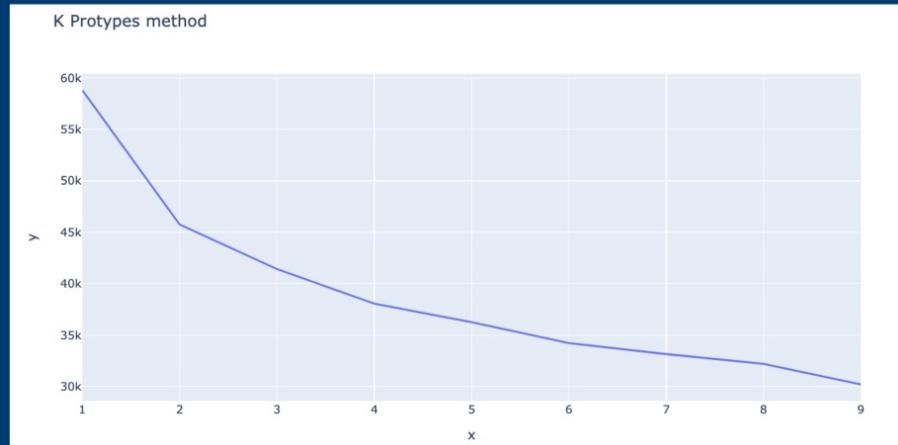
K  
prototypes

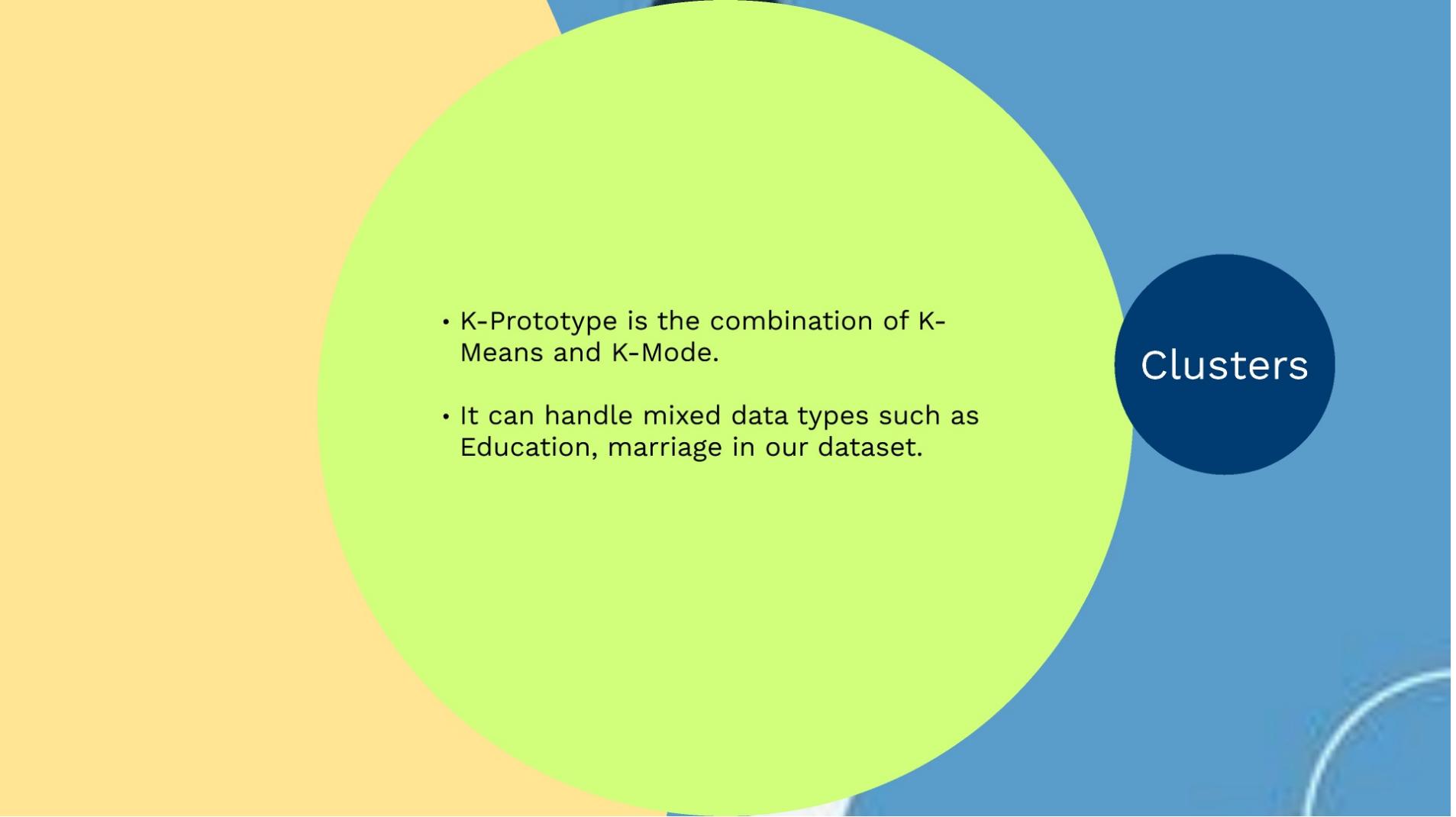
Association  
rule



## Clusters

- K-Prototype is the combination of K-Means and K-Mode.
- It can handle mixed data types such as Education, marriage in our dataset.

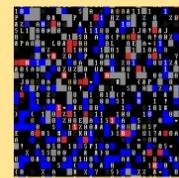




## Clusters

- K-Prototype is the combination of K-Means and K-Mode.
- It can handle mixed data types such as Education, marriage in our dataset.

## Models used



Clustering model-  
It involves automatically  
discovering natural grouping in  
data.

K  
means

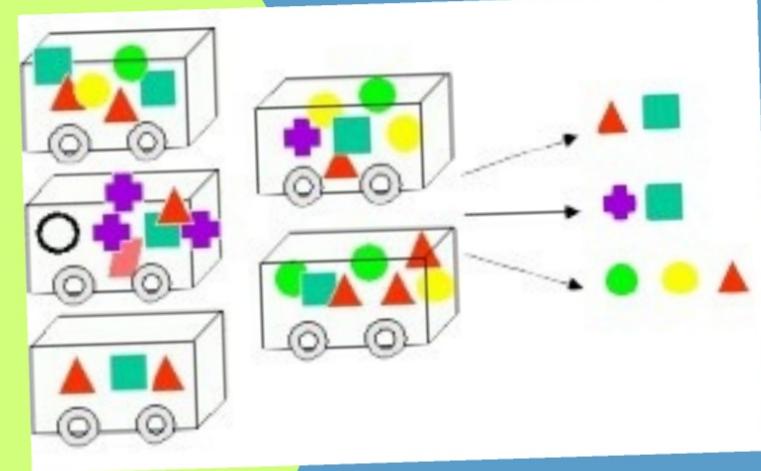
K  
prototypes

Association  
rule

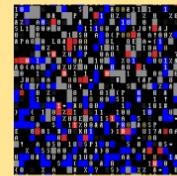
Association rules are useful for analyzing and predicting customer behavior.

This technique provides the following benefits:

1. Understand purchasing patterns.
2. Affinity Analysis. Frequent items analysis.
3. Propensity Analysis. Example: Who will buy what? Which is the profile of those customers?



## Models used

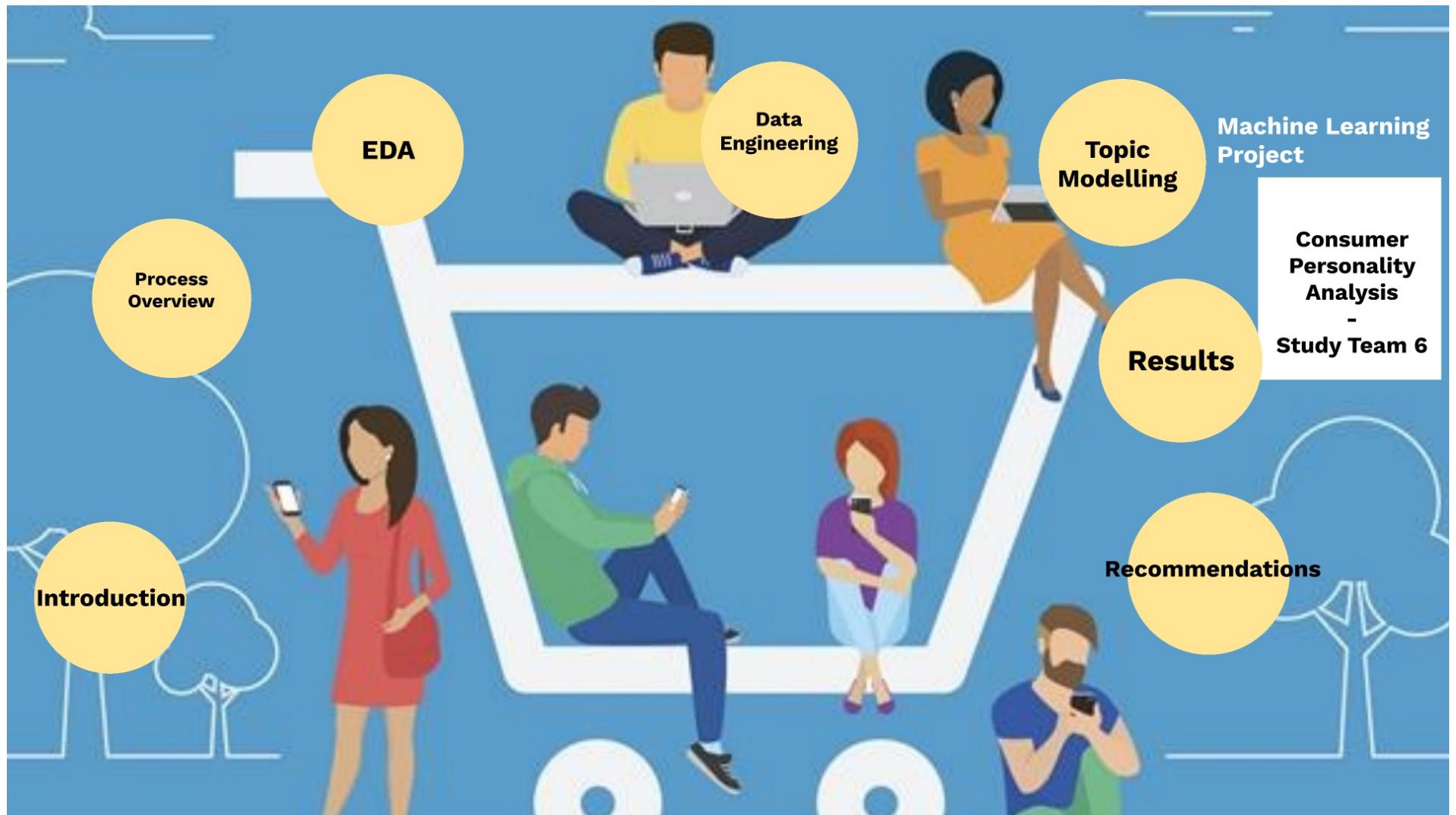


Clustering model-  
It involves automatically  
discovering natural grouping in  
data.

K  
means

K  
prototypes

Association  
rule



# Results

## Clustering models

	K means	K prototypes
Customer's Personal characteristics	Income: segment 0 < segment 2 < segment 1 < segment 3	Income: segment 2 < segment 3 < segment 0 < segment 1
Customer's buying categories	MntFruits: segment 0 < segment 2 < segment 3 < segment 1	MntFruits: segment 2 < segment 3 < segment 1 < segment 0

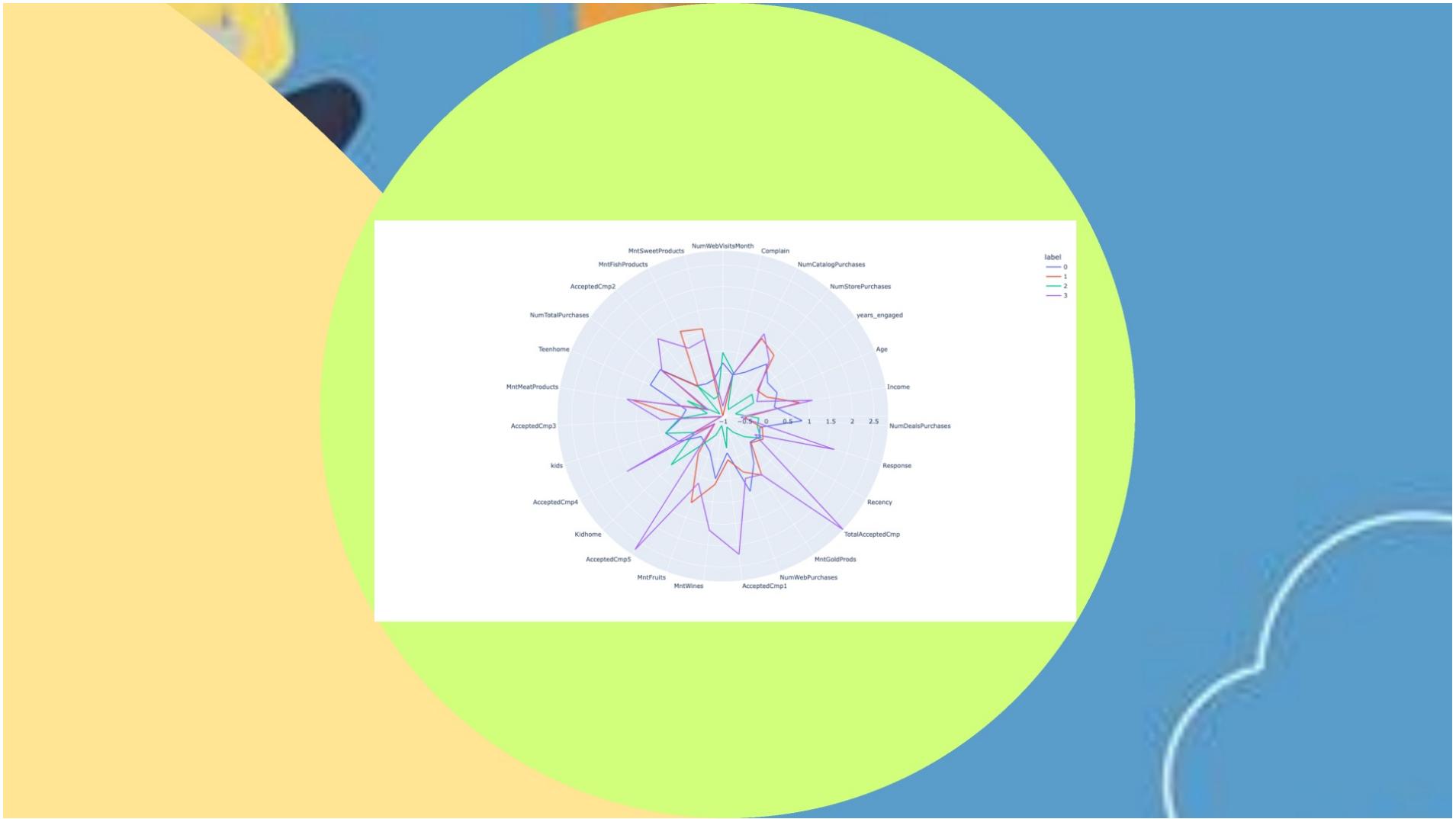
## Association rule

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
361	(MntFruits, SweetProducts, MntFishProducts, MntWines)	(MntMeatProducts)	0.129018	0.312946	0.113839	0.882353	2.819502	0.073464	5.839955
541	(MntFruits, MntGoldProds, MntFishProducts, MntWines, SweetProducts)	(MntMeatProducts)	0.082589	0.312946	0.072768	0.881081	2.815437	0.046922	5.777496
390	(MntFruits, MntFishProducts, MntWines, MntGoldProds)	(MntMeatProducts)	0.098214	0.312946	0.085714	0.872727	2.788743	0.054978	5.398278

K  
means

K  
proto  
pes

Association  
Rule



# Results

## Clustering models

	K means	K prototypes
Customer's Personal characteristics	Income: segment 0 < segment 2 < segment 1 < segment 3	Income: segment 2 < segment 3 < segment 0 < segment 1
Customer's buying categories	MntFruits: segment 0 < segment 2 < segment 3 < segment 1	MntFruits: segment 2 < segment 3 < segment 1 < segment 0

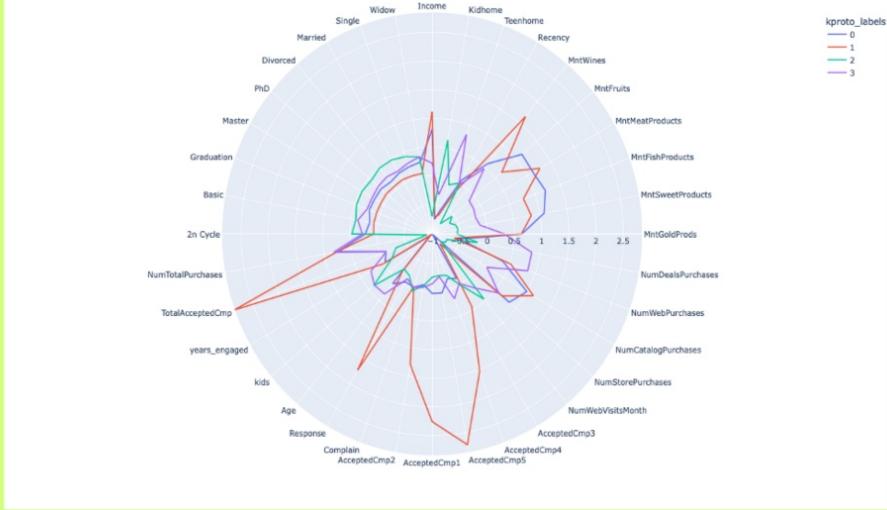
## Association rule

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
361	(MntFruits, SweetProducts, MntFishProducts, MntWines)	(MntMeatProducts)	0.129018	0.312946	0.113839	0.882353	2.819502	0.073464	5.839955
541	(MntFruits, MntGoldProds, MntFishProducts, MntWines, SweetProducts)	(MntMeatProducts)	0.082589	0.312946	0.072768	0.881081	2.815437	0.046922	5.777496
390	(MntFruits, MntFishProducts, MntWines, MntGoldProds)	(MntMeatProducts)	0.098214	0.312946	0.085714	0.872727	2.788743	0.054978	5.398278

K  
means

K  
proto  
pes

Association  
Rule



# Results

## Clustering models

	K means	K prototypes
Customer's Personal characteristics	Income: segment 0 < segment 2 < segment 1 < segment 3	Income: segment 2 < segment 3 < segment 0 < segment 1
Customer's buying categories	MntFruits: segment 0 < segment 2 < segment 3 < segment 1	MntFruits: segment 2 < segment 3 < segment 1 < segment 0

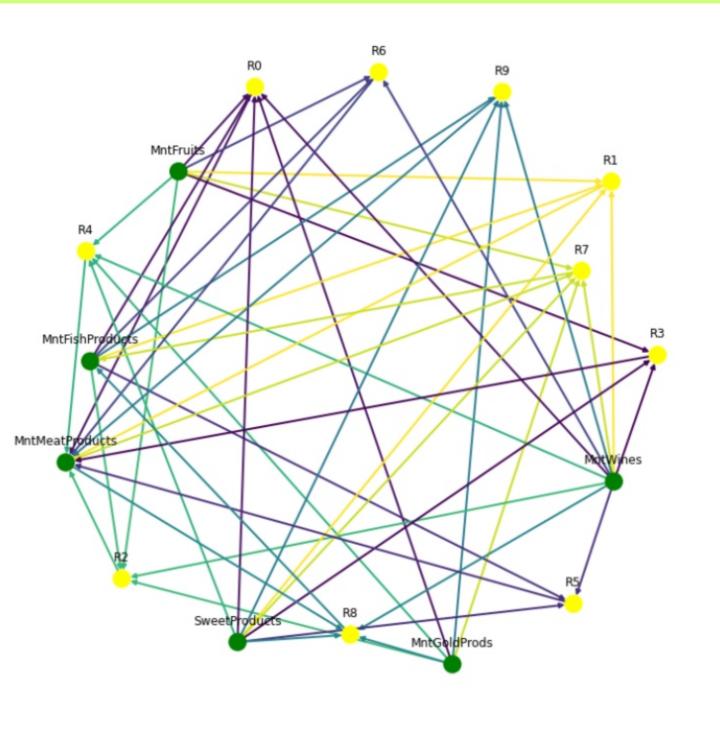
## Association rule

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
361	(MntFruits, SweetProducts, MntFishProducts, MntWines)	(MntMeatProducts)	0.129018	0.312946	0.113839	0.882353	2.819502	0.073464	5.839955
541	(MntFruits, MntGoldProds, MntFishProducts, MntWines, SweetProducts)	(MntMeatProducts)	0.082589	0.312946	0.072768	0.881081	2.815437	0.046922	5.777496
390	(MntFruits, MntFishProducts, MntWines, MntGoldProds)	(MntMeatProducts)	0.098214	0.312946	0.085714	0.872727	2.788743	0.054978	5.398278

K  
means

K  
proto  
pes

Association  
Rule



# Results

## Clustering models

	K means	K prototypes
Customer's Personal characteristics	Income: segment 0 < segment 2 < segment 1 < segment 3	Income: segment 2 < segment 3 < segment 0 < segment 1
Customer's buying categories	MntFruits: segment 0 < segment 2 < segment 3 < segment 1	MntFruits: segment 2 < segment 3 < segment 1 < segment 0

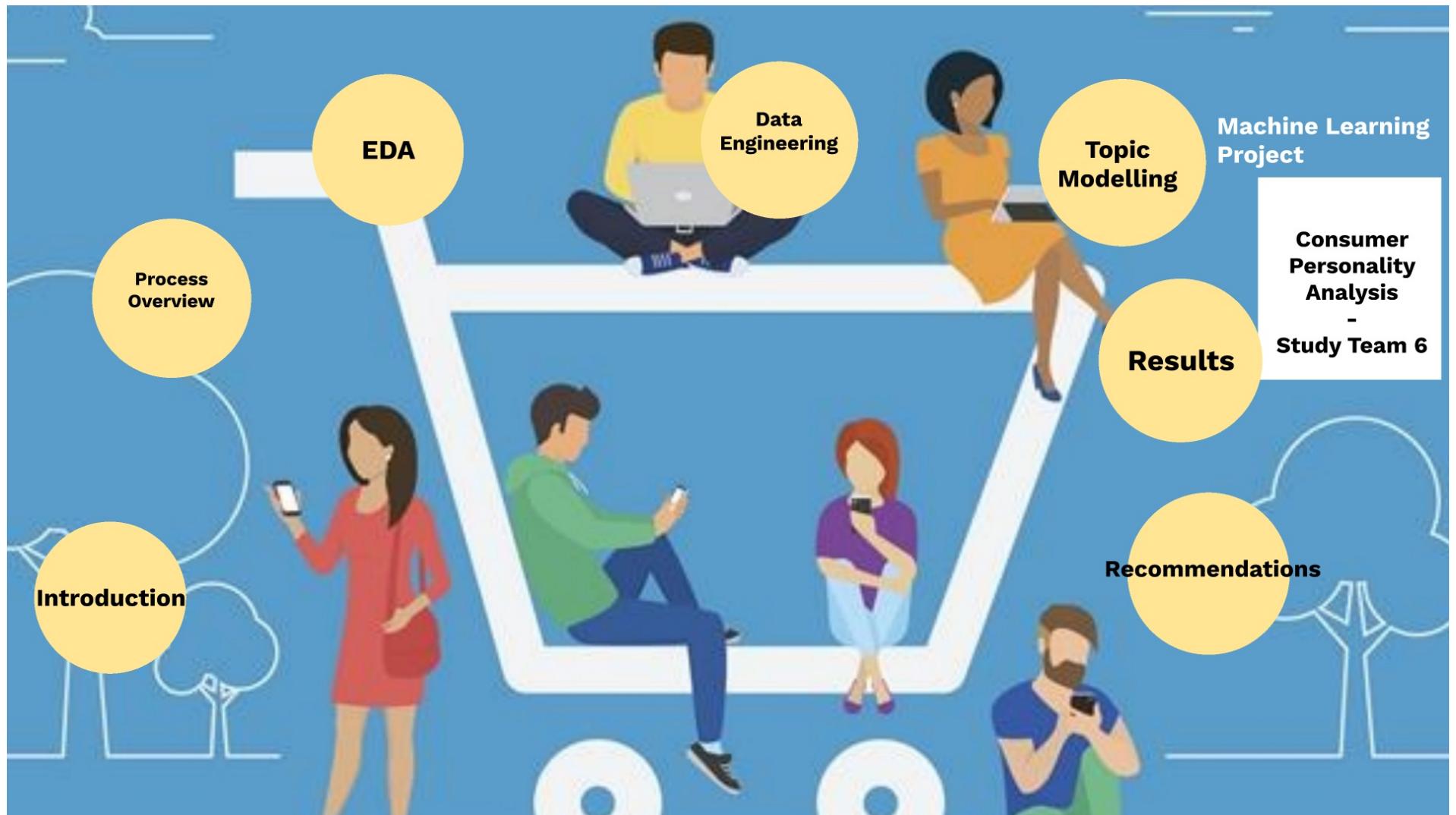
## Association rule

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
361	(MntFruits, SweetProducts, MntFishProducts, MntWines)	(MntMeatProducts)	0.129018	0.312946	0.113839	0.882353	2.819502	0.073464	5.839955
541	(MntFruits, MntGoldProds, MntFishProducts, MntWines, SweetProducts)	(MntMeatProducts)	0.082589	0.312946	0.072768	0.881081	2.815437	0.046922	5.777496
390	(MntFruits, MntFishProducts, MntWines, MntGoldProds)	(MntMeatProducts)	0.098214	0.312946	0.085714	0.872727	2.788743	0.054978	5.398278

K  
means

K  
proto  
pes

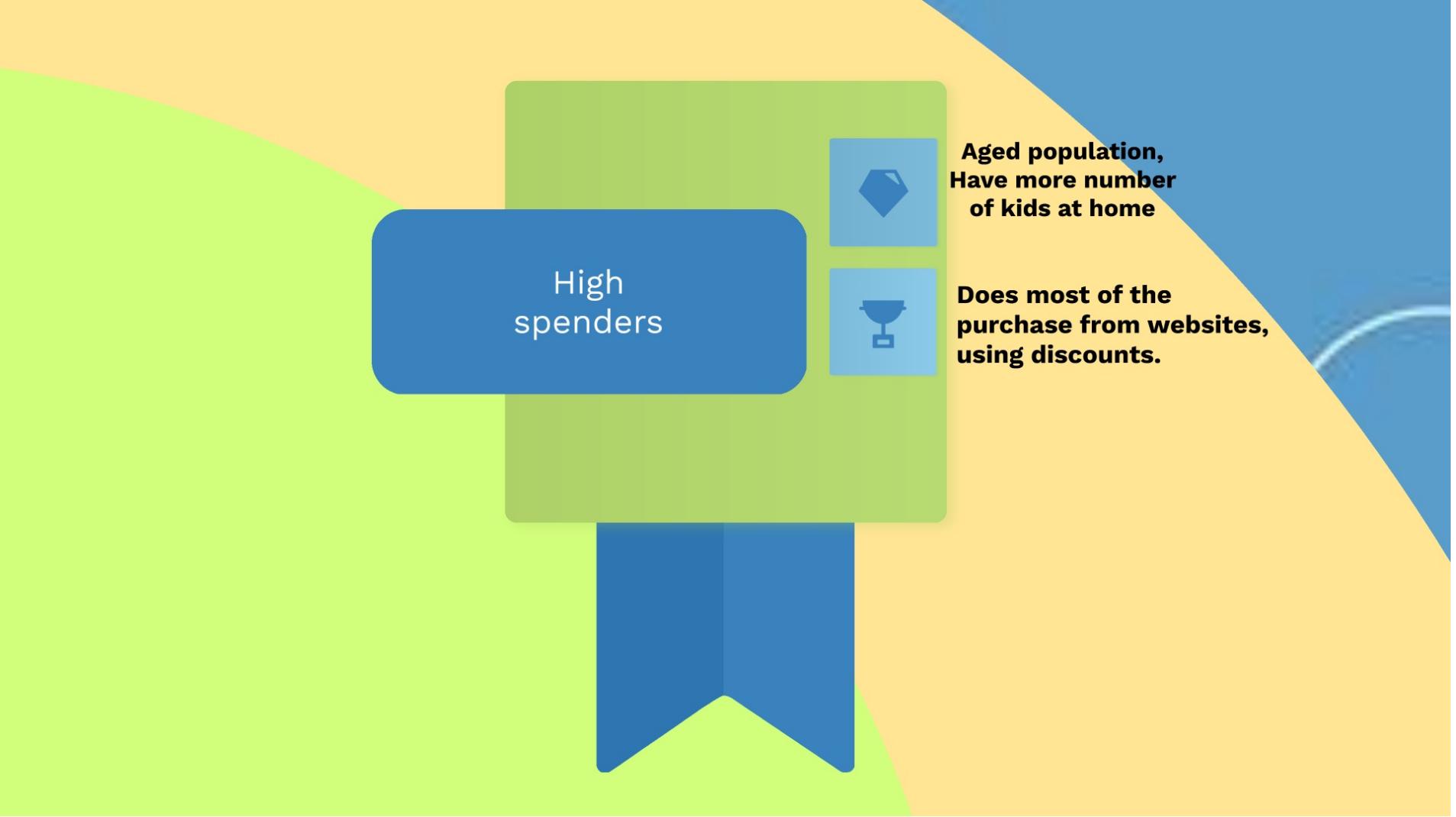
Association  
Rule





**Recommendations  
based on Cluster  
divisions**





## High spenders



**Aged population,  
Have more number  
of kids at home**



**Does most of the  
purchase from websites,  
using discounts.**

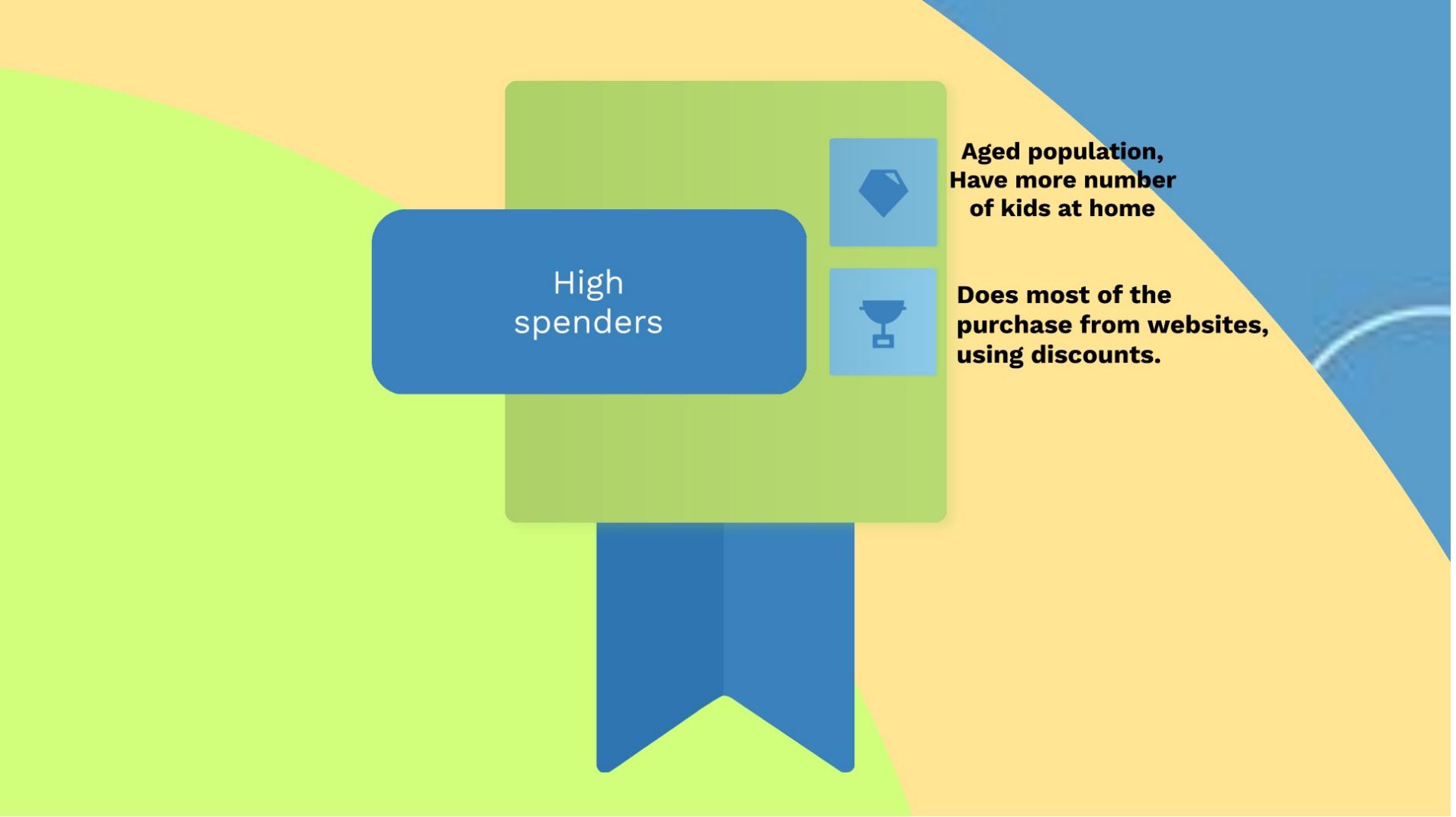
## Traits

Single/ Widow/ Divorced

Mediocre Income

Years engaged with the firm is highest

Basic education is high



## High spenders



**Aged population,  
Have more number  
of kids at home**

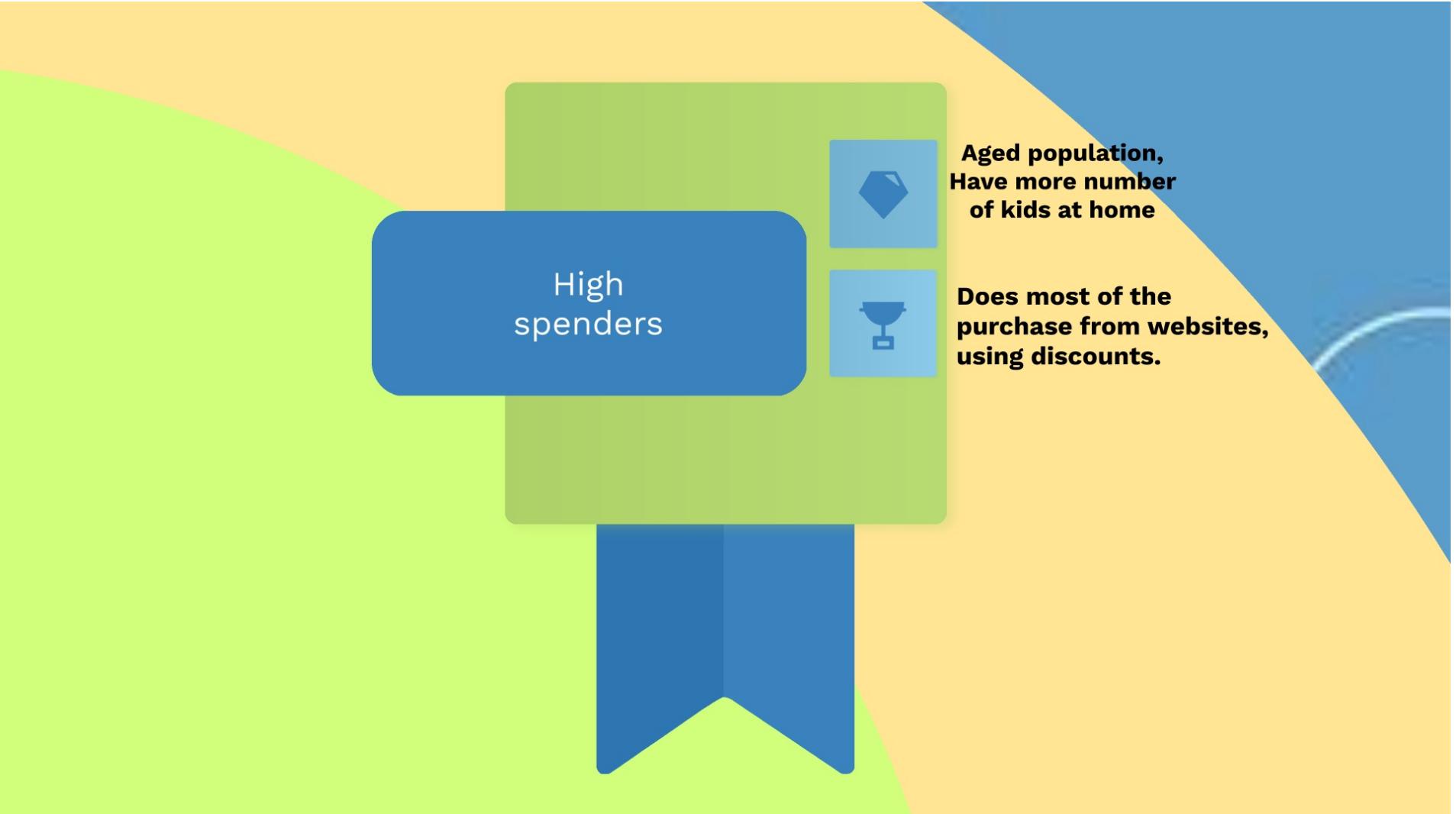


**Does most of the  
purchase from websites,  
using discounts.**

## Recommendations



1. Keep your website copy crisp and compelling.
2. Providing them with more deals.



## High spenders

**Aged population,  
Have more number  
of kids at home**

**Does most of the  
purchase from websites,  
using discounts.**



## Average Spenders

Middle aged population, Have less or no kids at home

Does all purchases from catalogs, do not care about discounts

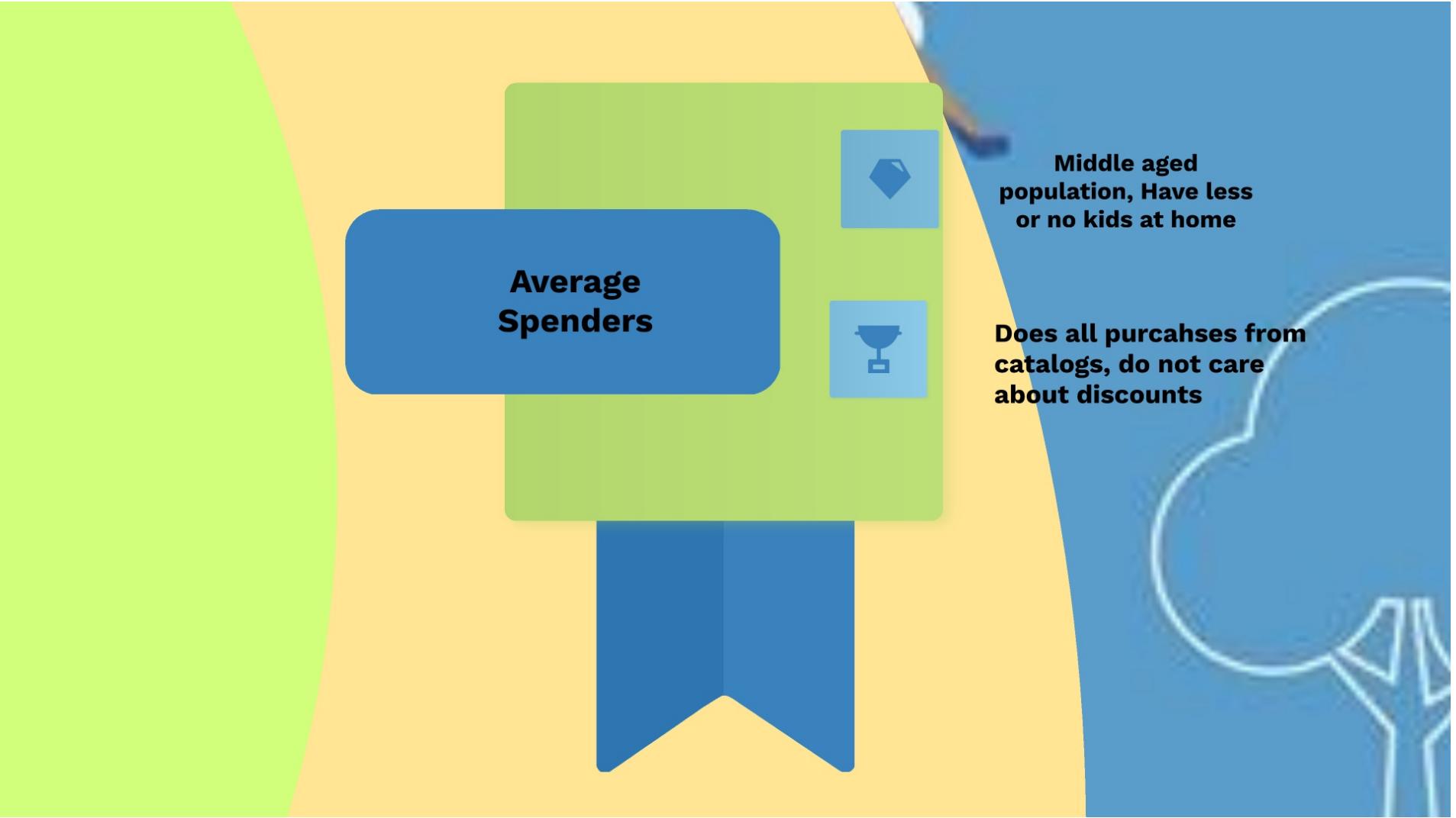
# Traits

Highest income earner

Middle aged population

Consumption of all products is the maximum.

2nd cycle, graduation degree



## Average Spenders

Middle aged population, Have less or no kids at home

Does all purchases from catalogs, do not care about discounts

## Recommendations



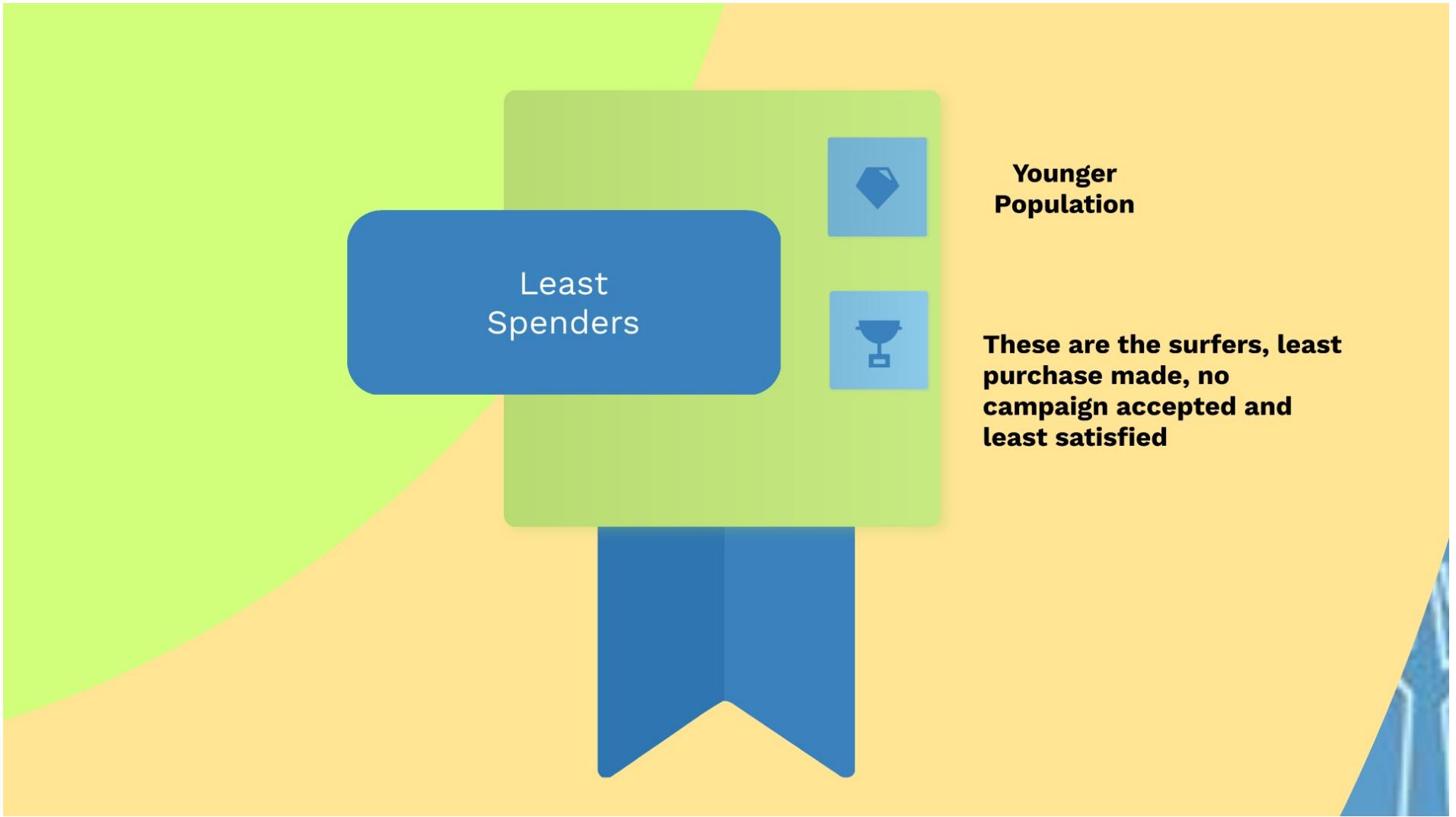
1. Provide excellent and detailed catalog content.
2. Direct more similar kind of campaigns towards them.
3. Market more wine and meat products to this category by providing discount and using association rule.

## Average Spenders

Middle aged population, Have less or no kids at home

Does all purchases from catalogs, do not care about discounts





## Traits

Lowest income earner

Have small kids at  
home

Do not buy excess  
of anything

Majority in number

## Least Spenders



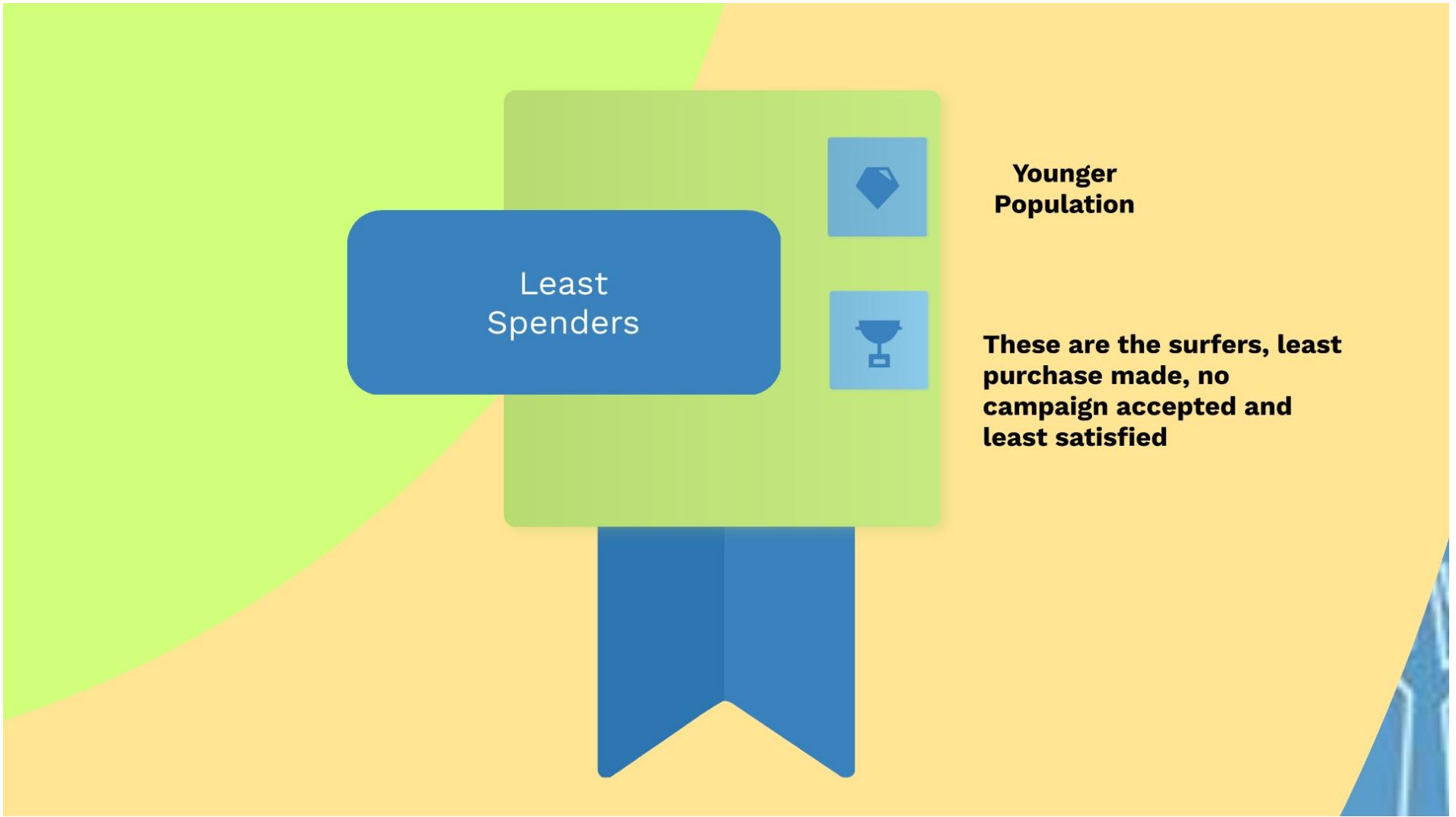
### Younger Population

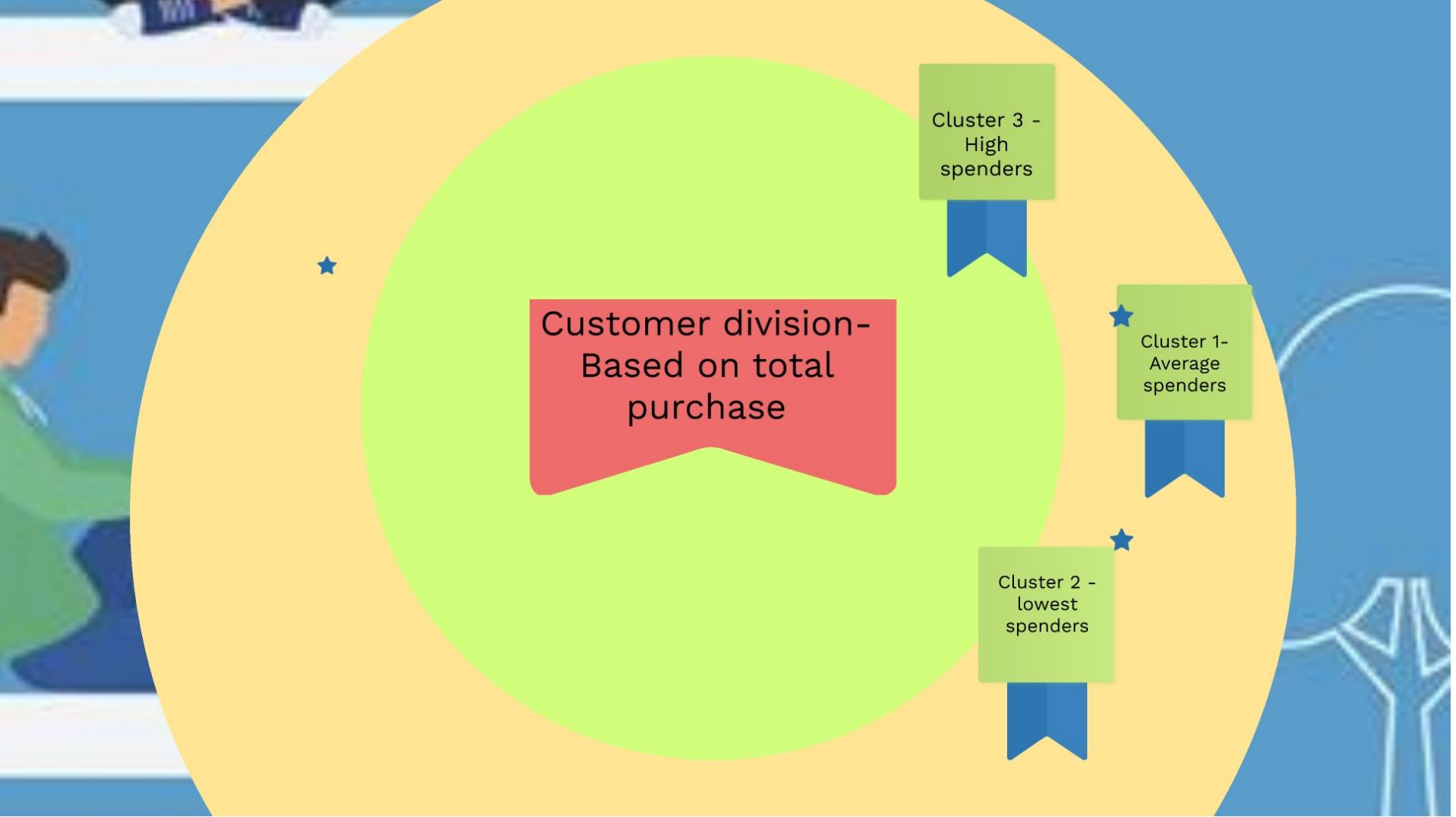
**These are the surfers, least purchase made, no campaign accepted and least satisfied**

## Recommendations

1. Providing ease of access point of contact.
2. More discounts .
3. Provide customer care support







Customer division-  
Based on total  
purchase

Cluster 3 -  
High  
spenders

Cluster 1-  
Average  
spenders

Cluster 2 -  
lowest  
spenders



**Recommendations  
based on Cluster  
divisions**

