



Side Projects (Kaggle Dataset)

# Customer Segmentation (Agglomerative Clustering)

Moch Nabil Farras Dhiya  
Bandung Institute of Technology

June 11th, 2022

# Agenda

---



## Context

## Model (Agglomerative Clustering)

## Analysis

## Summary and Recommendations

# Context

---



## Background

Customer Personality Analysis is a detailed analysis of a company's ideal customers. It helps a business to better understand its customers and makes it easier for them to modify products according to the specific needs, behaviors and concerns of different types of customers. For example, instead of spending money to market a new product to every customer in the company's database, a company can analyze which customer segment is most likely to buy the product and then market the product only on that particular segment.

## Main Objectives

1. **Analyze the characteristics of each product's potential customers** to build focused business and marketing strategies to minimize cost and maximize profit.
2. **Find the most used platform** by the potential customers, then upgrade that particular platform to attract more customers with (but not limited to) the same characteristics.
3. **Examine the discount and campaign performance** to build better strategies in the future.

## Dataset

The dataset used in this project are from Kaggle public dataset - [Customer Personality Analysis](#)

# Agenda

---



Context

**Model (Agglomerative Clustering)**

Analysis

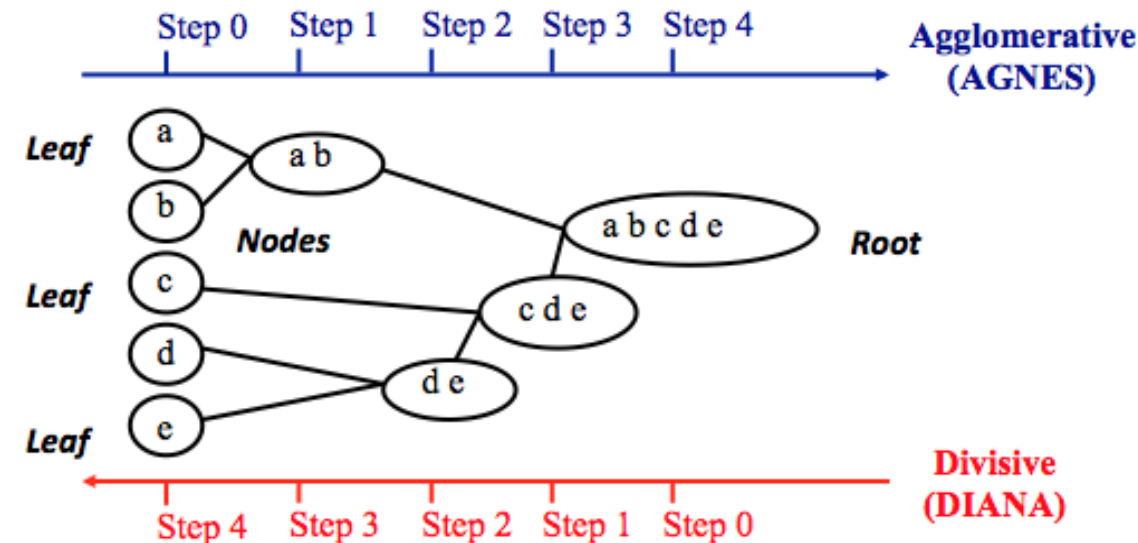
Summary and Recommendations

# Agglomerative Clustering short overview

## Definition

The agglomerative clustering is a type of **hierarchical clustering** used to group objects in clusters based on their **similarity**, and also known as AGNES (Agglomerative Nesting).

## Algorithms



1. Treating each object as a **singleton** cluster.
2. Pairs of clusters are successively **merged** **until** all clusters have been merged into **one big cluster** containing all objects.
3. The result is a **tree-based representation** of the objects, named dendrogram.

We will use this algorithm to find the characteristics of potential customers for each product to build suitable marketing strategies and maximize our profit.

Source:  
[www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering](http://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering)

# Agenda

---



Context

Model (Agglomerative Clustering)

**Analysis**

Summary and Recommendations

In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer



Top 8 **potential customers** segments for **WINE**

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Late 40s – early 70s	Single	Graduate	High	0	0	114
	Relationship	Graduate	High	0	0	65
	Single	PhD	Upper-middle	0	1	62
			High	0	0	57
		Master	High	0	0	51
	Relationship	Master	High	0	0	36
	Single	Graduate	High	0	1	34
	Relationship	PhD	High	0	0	32
	...					

**Income distribution**

**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD

In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer



Top 8 **potential customers** segments for **FRUITS**

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Early 40s – late 60s	Single	Graduate	High	0	0	123
	Relationship	Graduate	High	0	0	80
	Single	PhD	High	0	0	57
		Master	High	0	0	52
		Graduate	High	0	1	43
	Relationship	Master	High	0	0	38
		PhD	High	0	0	32
		Graduate	Low	0	0	23
	...					

**Income distribution**

**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD



In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer



Top 8 **potential customers** segments for **MEAT**

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Late 30s – late 60s	Single	Graduate	High	0	0	120
	Relationship	Graduate	High	0	0	79
	Single	PhD	High	0	0	56
		Master	High	0	0	52
		Graduate	High	0	1	41
	Relationship	Master	High	0	0	38
		PhD	High	0	0	31
		Graduate	High	0	1	21
	...					

**Income distribution**

**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD



In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer

Top 8 **potential customers** segments for **FISH**

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Late 40s – mid 70s	Single	Graduate	High	0	0	111
	Relationship	Graduate	High	0	0	61
	Single	PhD	High	0	0	56
		Master	High	0	0	48
	Relationship	Master	High	0	0	33
		PhD	High	0	0	31
	Single	Graduate	High	0	1	28
	Relationship	Graduate	High	0	1	19
	...					

**Income distribution**

**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD

**In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer**

Top 8 **potential customers** segments for **SWEETS**

'Same' segments as meat

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Late 40s – mid 70s	Single	Graduate	High	0	0	103
	Relationship	Graduate	High	0	0	60
	Single	PhD	High	0	0	57
		Master	High	0	0	50
		Graduate	High	0	1	40
	Relationship	Master	High	0	0	33
		PhD	High	0	0	32
		Graduate	High	0	1	22
	...					

**Income distribution**

**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD

In summary, people who are at least a graduate, have high income and do not have a child at home are our target customer



Top 8 **potential customers** segments for **GOLD**

'Same' segments as fruits

Age	Marital Status	Education	Income	# of kid at home	# of teen at home	# of Customer(s)
Early 30s – late 60s	Single	Graduate	High	0	0	123
	Relationship	Graduate	High	0	0	82
	Single	PhD	High	0	0	57
		Master	High	0	0	52
		Graduate	High	0	1	41
	Relationship	Master	High	0	0	38
		PhD	High	0	0	32
		Graduate	Low	0	0	25
	...					

**Income distribution**

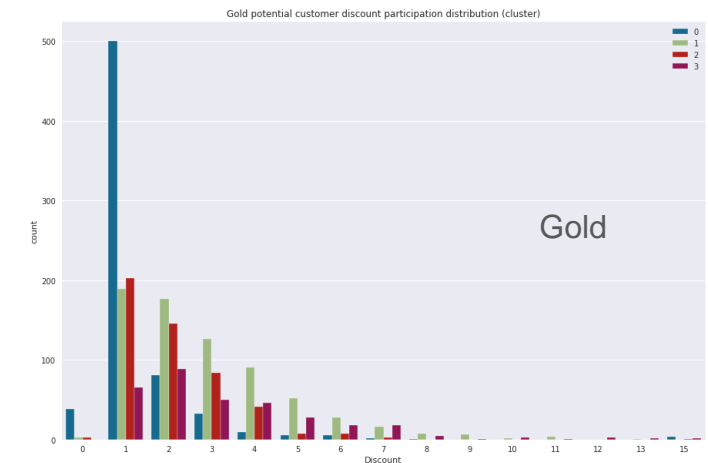
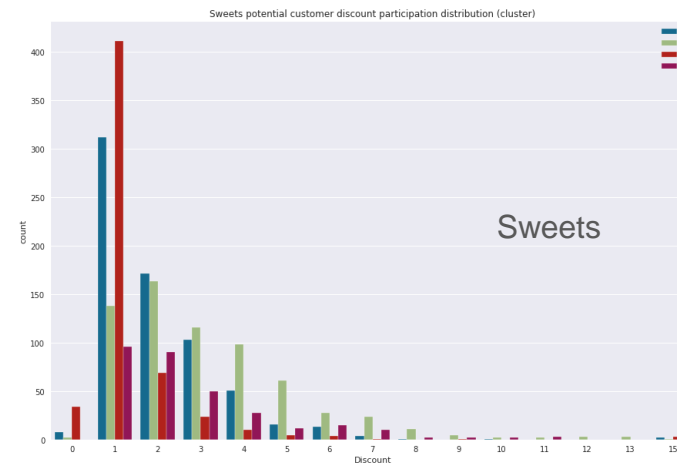
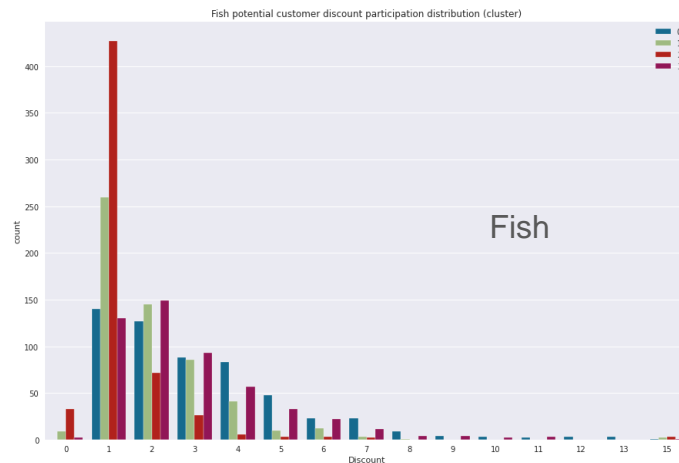
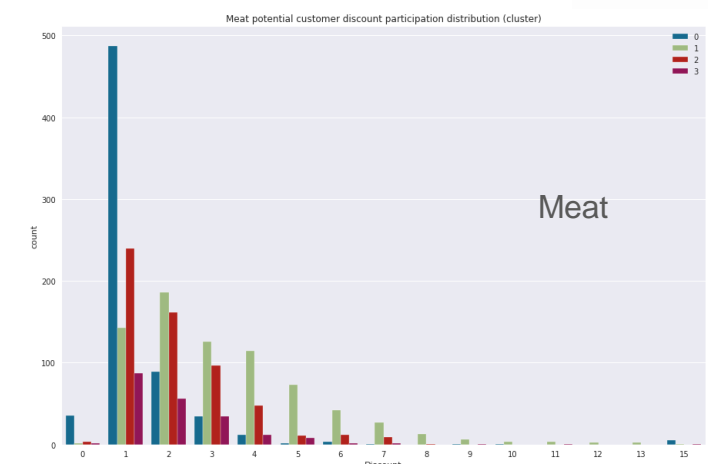
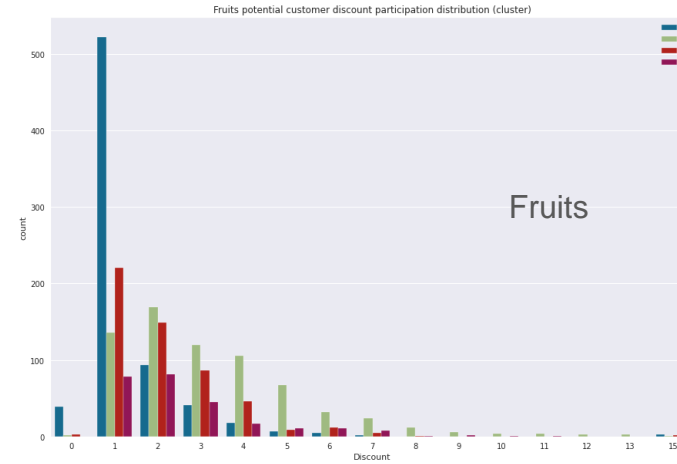
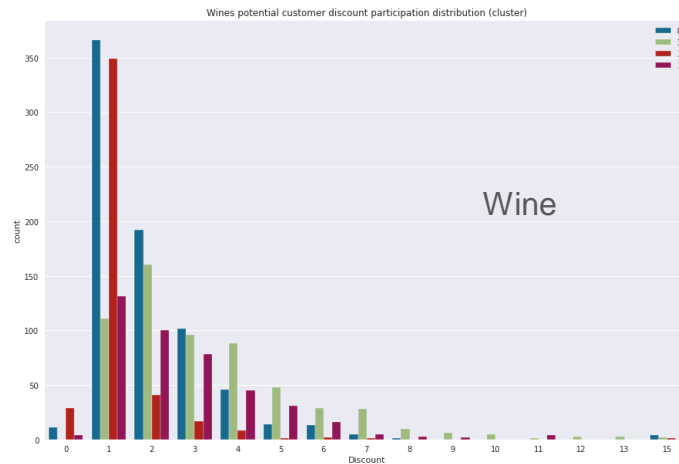
**Low** : 1,7K to 35K USD

**Lower-middle**: 35K USD to 51K USD

**Lower-middle**: 35K USD to 51K USD

**High**: 68K++ USD

# Most potential customers only participating in the discount event once. Why?



Is it because the discount given is not much? Or is there other factors which make customers are not interested by the discount given?

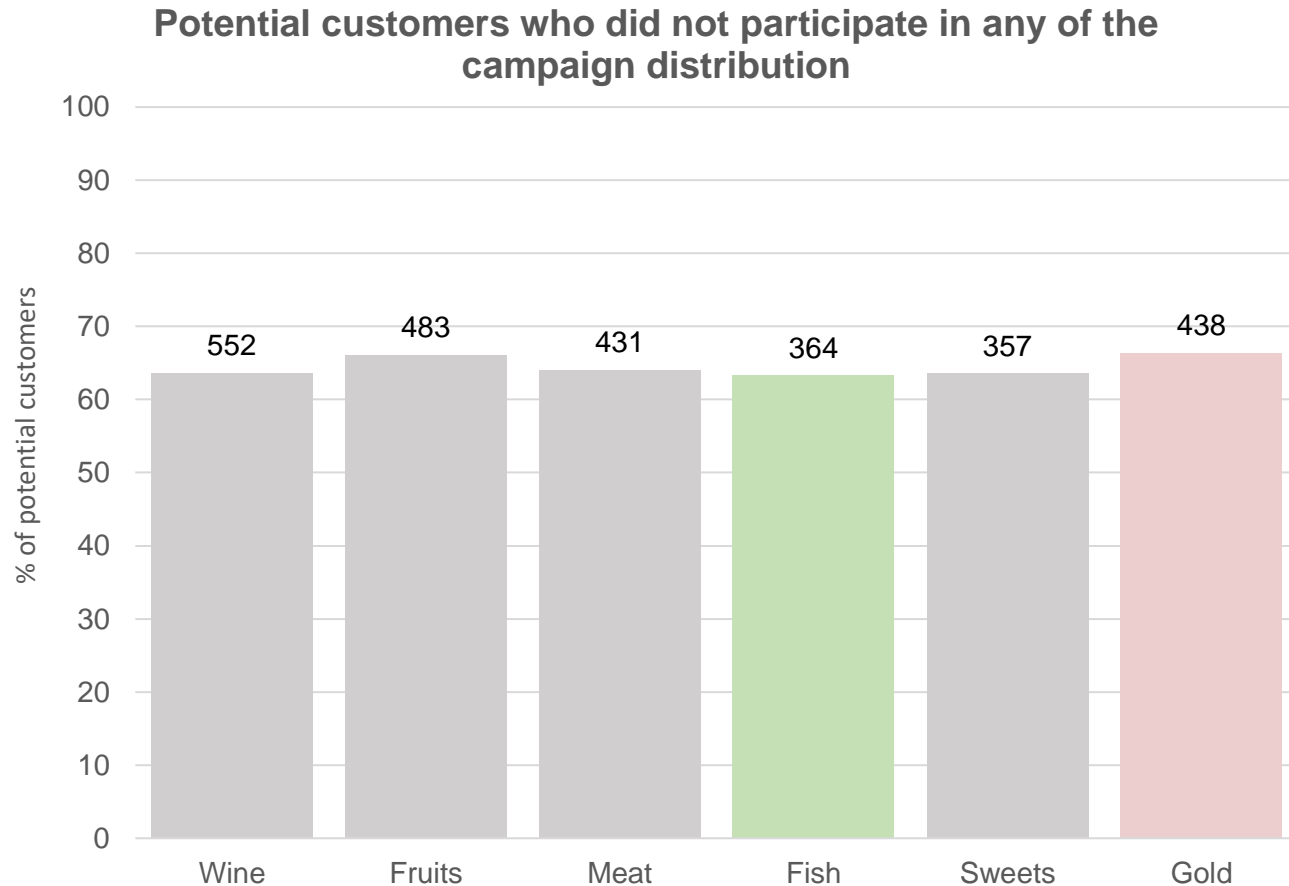
## Offline transaction is the most popular method among the potential customers (98.56%), followed by Web (93.83%).

**Approach:** calculating the % of **potential customers** who used certain **platform(s)** more than once.

**Validation:** when certain potential customers are **satisfied** by the performance of certain **platform(s)**, then the **possibilities** of them to use it again in the **future** as the media transaction is **high**.

Product	Platform usage more than 1 times (%)		
	Web	Catalogue	Store
Wines	96.31	86.54	99.54
Fruits	91.79	82.90	98.49
Meat	93.31	86.33	98.51
Fish	93.39	85.74	98.43
Sweets	93.59	87.01	98.22
Gold	91.91	80.74	98.38
<b>AVERAGE</b>	<b>93.83</b>	<b>84.88</b>	<b>98.56</b>

# The campaign participation rate is very low, including among the potential customers. Why?



## Key Findings

- Overall, there are 1754 customers (79.29%) who did not participate in any of the campaign.
- The % of potential customers for each product who did not participate in any of the campaign exceeded 60%

Is it because the campaign is **not engaging**? Or is the timing of the campaign is **ill-timed**?

# Agenda

---



Context

Model (Agglomerative Clustering)

Analysis

**Summary and Recommendations**



# Overview



## Analyses' Summary

- **Target Customer Criteria:** Mostly **early 40s – late 60s** who is **at least** an Undergraduate, has **high-income**, and **do not** have any child at home.
- **Platform:** The **most popular** platform for potential customers are **store** (98.56%), **web** (93.83%), and **catalogue** (84.88%), respectively.
- **Discount & Promo:** The participation rate is **very low**.

## Recommendations

1. Further sales and marketing **strategies** should be **focused** towards **Undergraduates** with **high-income**, and **do not** have any kid (and teens, if possible) at home.
2. There needs to be a **promo** to **attract customers**, especially the **potential customers**, so the amount of sold products will increase.
3. Further **investigation** are needed to find the reason why the **discount** and **campaign participation rate** is very **low**. Maybe because it is ill-timed or the content is not engaging?
4. The company **should upgrade** their **website**, as many potential customers are buying products through website, and also as a mean to attract another customer segments.