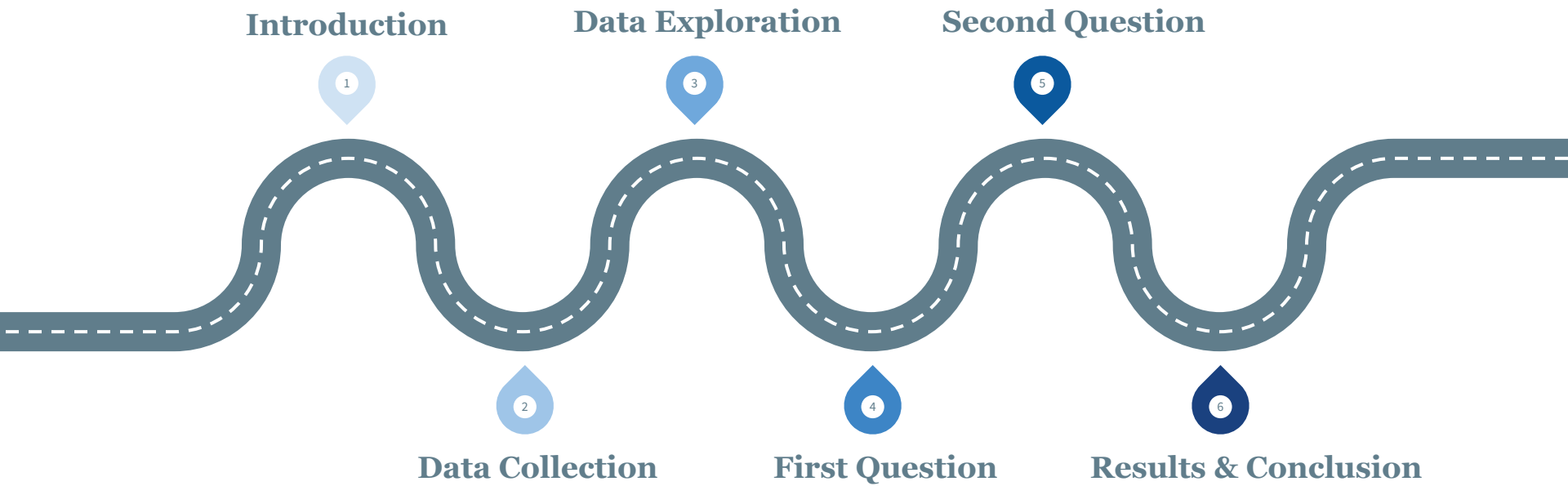
Customer Personality Analysis

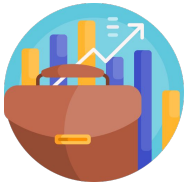
Amber Shao

Road Map


Introduction
1
Data Collection
2
Data Exploration
3
First Question
4
Second Question
5
Results & Conclusion
6

Introduction

Customer Personality Analysis is a detailed analysis of a company's **ideal customers**



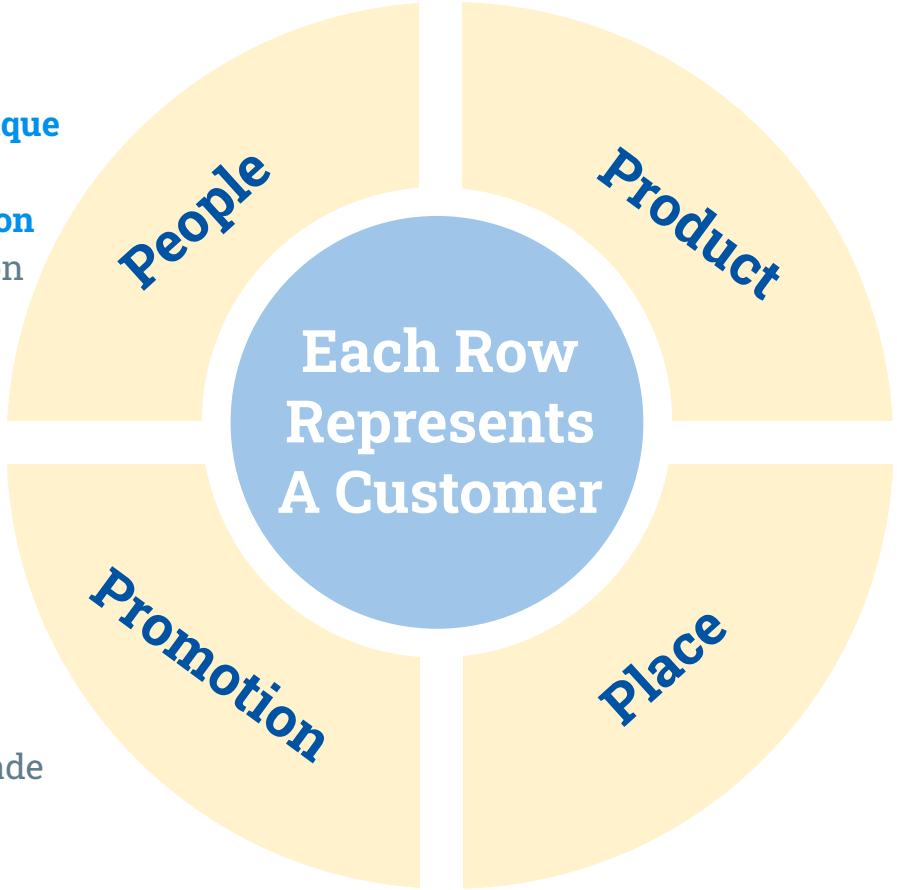
Helps a business to:

- **better understand its customers**
- **modify its product** based on its target customers from different types of **customer segments**

Data Collection - Kaggle

- 10 Features
- Include customer's **unique identifier number** and **background information**
- E.g birth year, education level, marital status, income, # of child, complain history

- 6 Features
- Include customer's **spent distribution** in different categories of products
- E.g Amount spent on wine/fruits/meat/fish/sweet/gold



- 7 Features
- Include customer's **responses to** in total 6 rounds of **campaigns**
- And # of purchases made with discounts

- 4 Features
- Include customer's **spent on different channels**
- E.g Number of purchases made through websites/ offline stores

Question:

Using the Data provided, Can we generate a **user portraits**, that is, different clusters of consumers, based on their personal information and purchasing behavior?

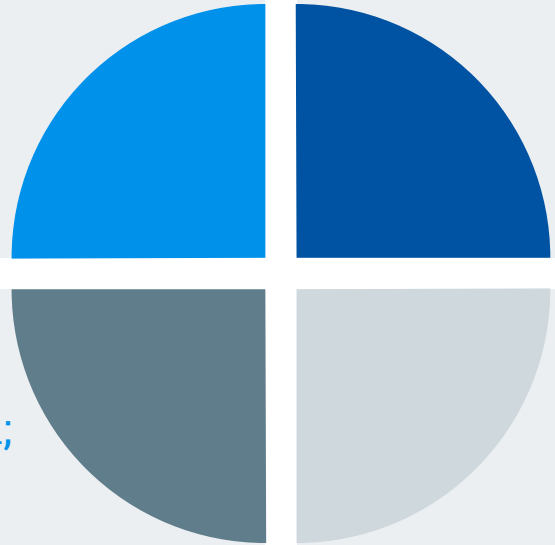
Data Exploration - Feature Engineering

MntLuxury

Sum of wine, sweets, and gold purchased

MntNecessity

Sum of fruit, meat, and fish purchased

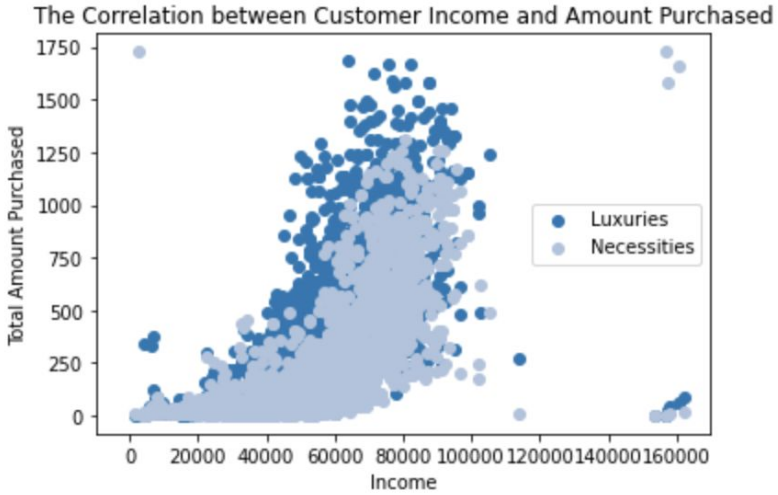


1 if the customer accept the marketing campaign; 0 otherwise

Accepted

Sum of childrens and teenagers in customer's household

House Size

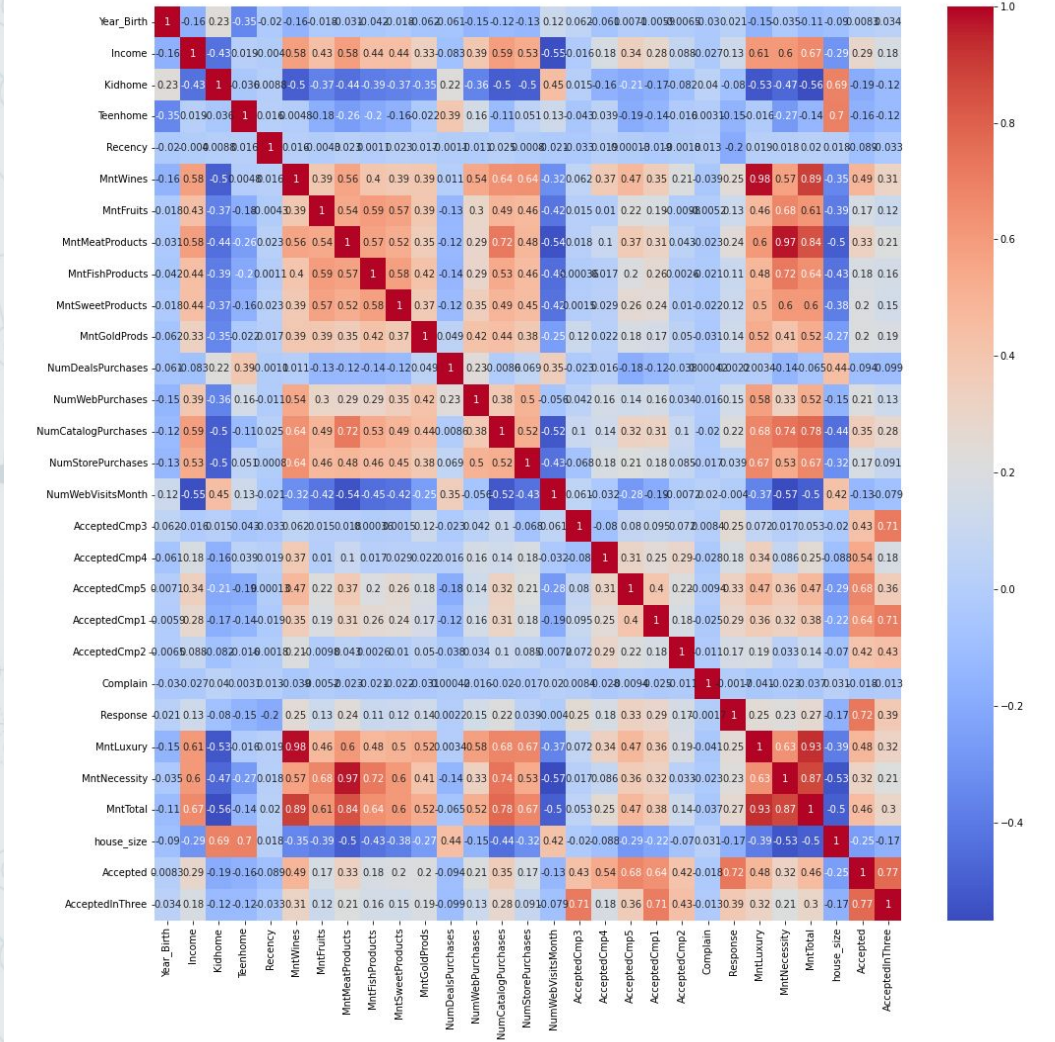


- **Positive relationship** between Income and Amount Purchased
  - Applicable to **both luxuries and necessities**
  - Can be a meaningful feature
  - Not linear but **exponential**

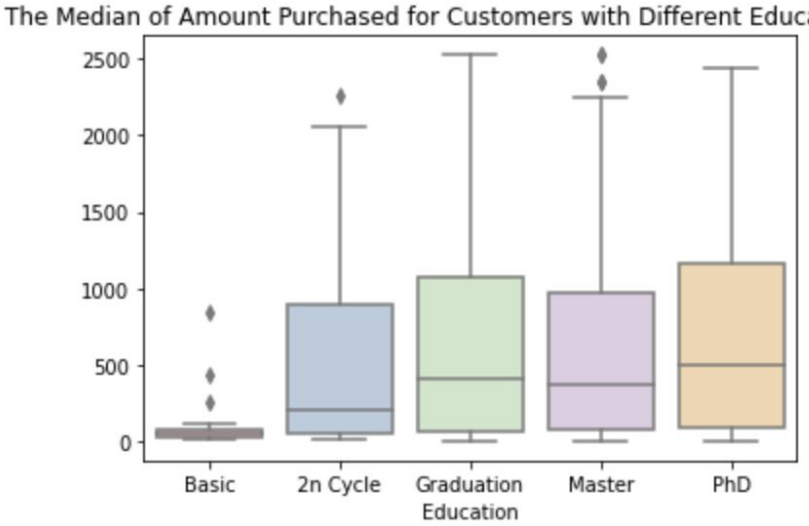
Data Exploration - Visualization

Correlation Heatmap

- There's a **moderate correlation** between **house size and number of purchases** made with a discount
- We can also see that there's a **correlation** between number of kids and the number of visits to the website per month
- The **correlation** between number of kids at home and income is negative

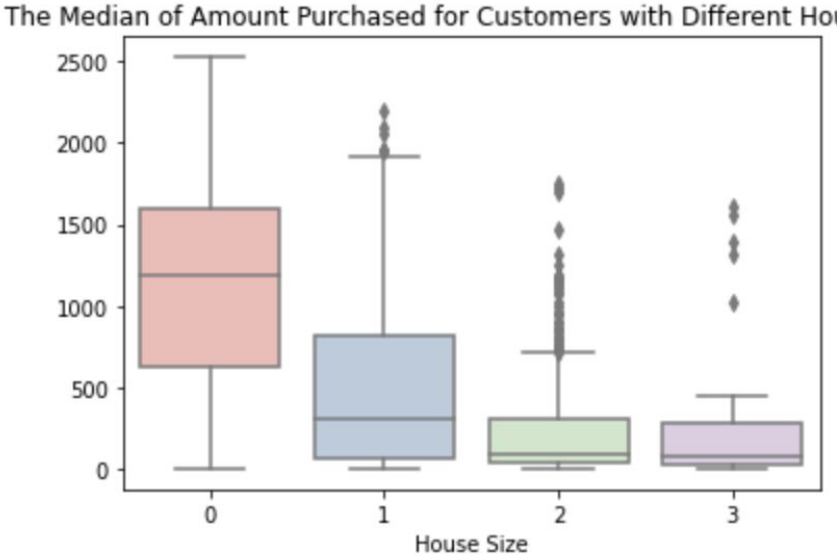


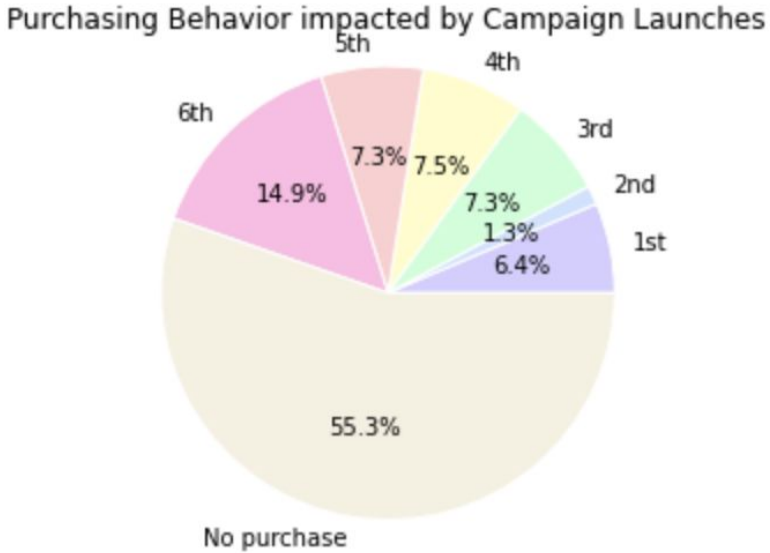
Data Exploration - Visualization



- **Difference on amount purchased** -> can be a meaningful feature
- **Education** falls **under our expectation**
- **Surprised** on the distribution of **marital status**

Data Exploration - Visualization



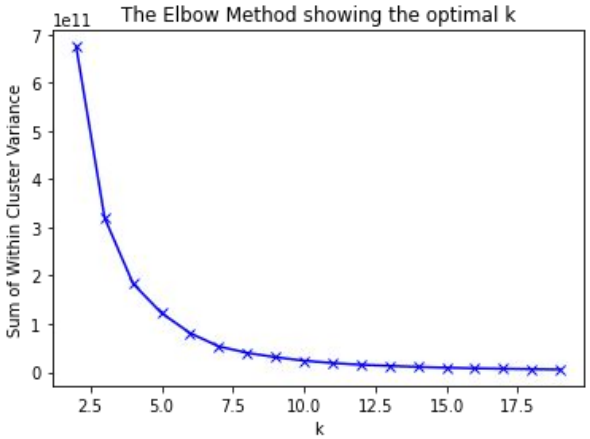


Customer Segmentation Process

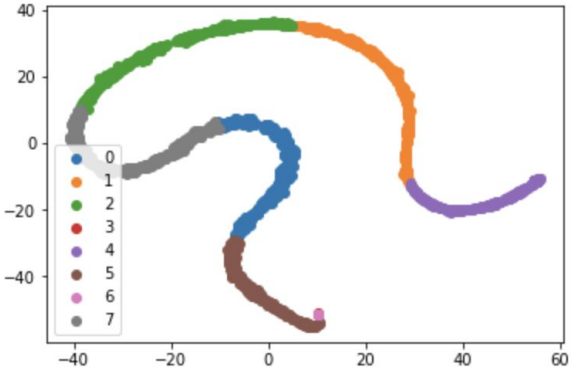
- We can divide customers into different clusters based on their behaviors and traits
- Identifies the distinction between each segment
- Helps the company to target specific groups of customers in promotions or developing sales strategies

Classification: K-Means

- **K-Means: k = 8**
- Using TSNE to **reduce high dimensional vectors** of customer features down to 2 dimensions to visualize
  - **Direct take:** Year_Birth, Income, Recency, MntLuxury, MntNecessity, House_size
  - **One hot code:** Marital_Status and Education

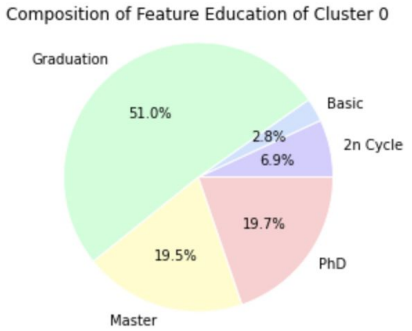


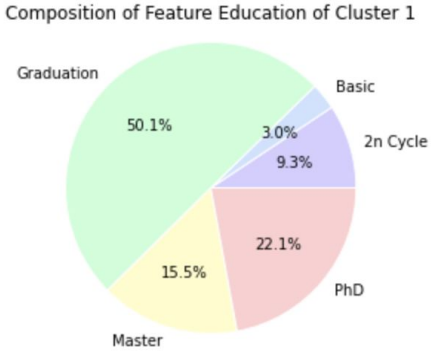
	x	y	ID	label
0	29.810543	-0.533634	5524.0	7
1	-45.343758	-5.444921	2174.0	2
2	-31.872677	12.125150	4141.0	0
3	9.377102	34.827084	6182.0	1
4	28.291882	-0.749628	5324.0	7

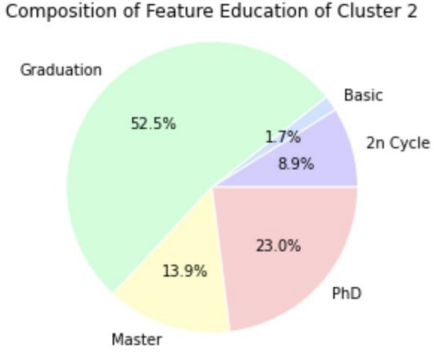


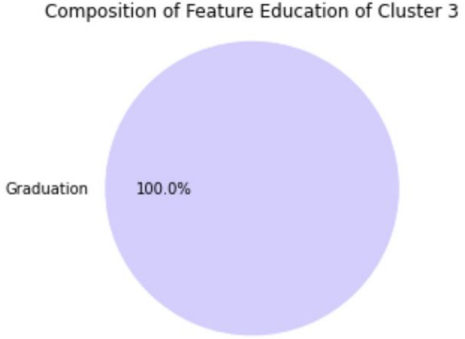
Exploring Each Clusters: Education

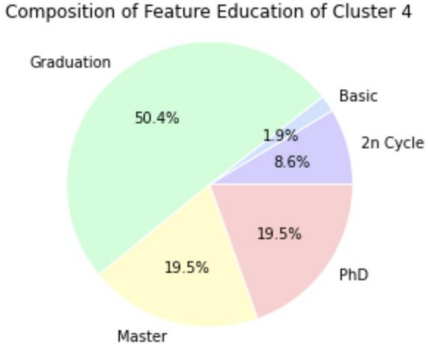
- **Similar distribution** among cluster 0,1,2,4,5,7
- **Cluster 6:** No Basic; **Cluster 3:** Only Graduation

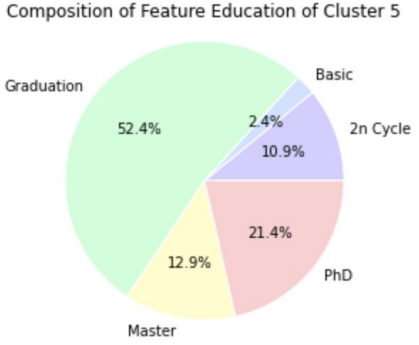


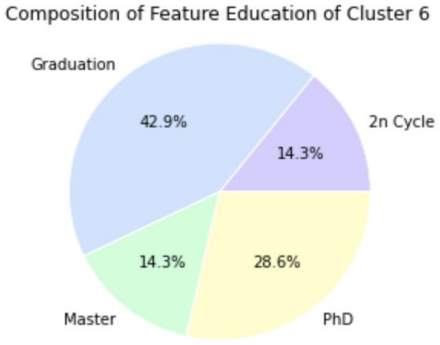


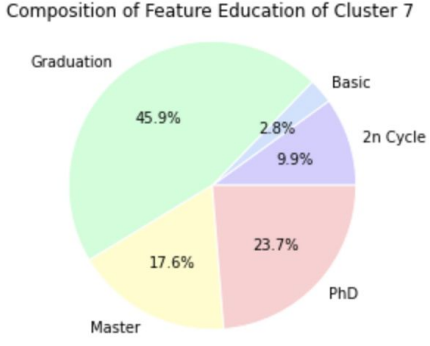


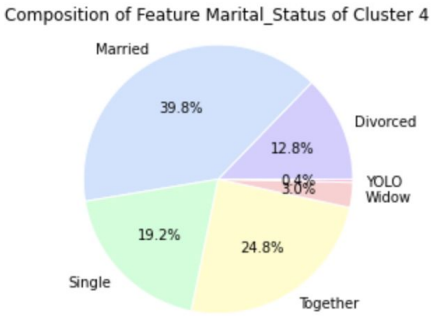


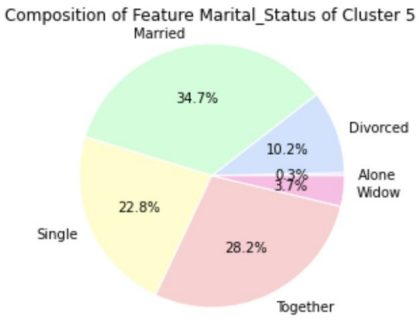


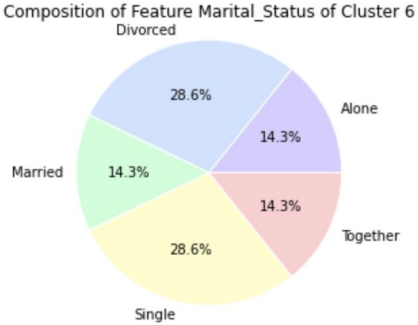


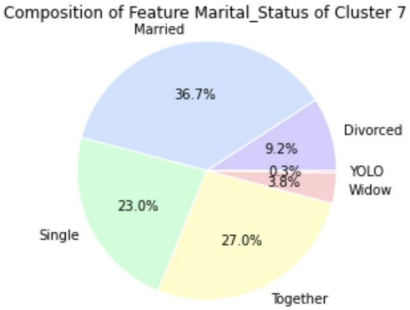


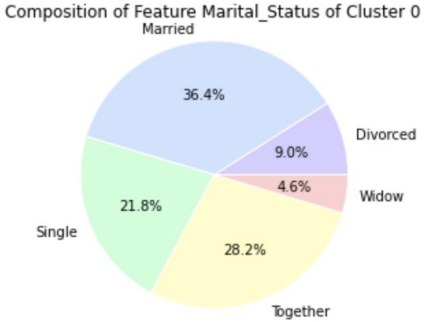


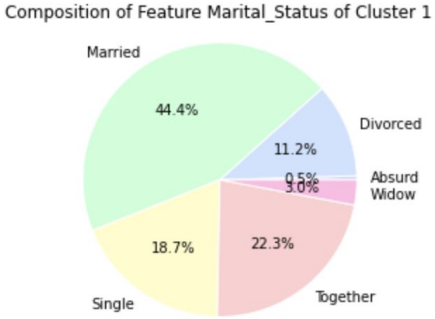


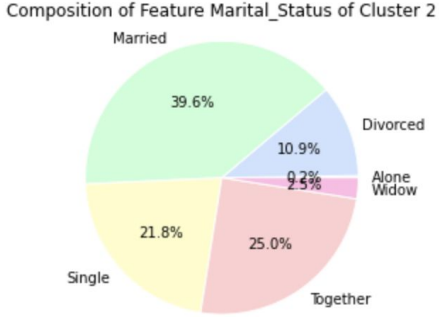


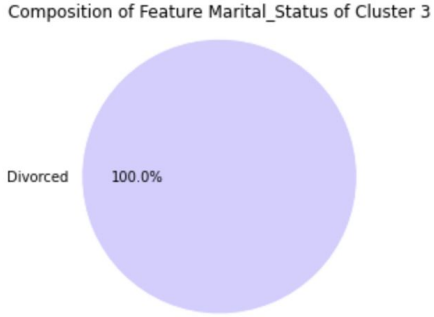












Exploring Each Clusters: Marital Status

- **Similar distribution** among cluster 0,1,2,4,5,7
- **Cluster 3:** Only Divorced ; **Cluster 1:** only cluster that contains absurd
- **Cluster 6:** contains the majority of consumers with Alone marital status

Exploring Each Clusters

- **Income**
  - **2** >> **7** > 40156 >> 3
- **Recency**
  - **6** >> 1720 > 54 >> 3
- **MntLuxury**
  - **6** >> **2** > 74015 >> 3
- **MntNecessity**
  - 20 > **67** > 541 > 3

Income

	Cluster	Average	Median
2	2	53179.076733	54327.0
7	7	52360.451531	51546.0
4	4	50942.221805	50727.0
0	0	52117.635897	50699.5
1	1	51384.829157	50300.0
5	5	51746.993197	50124.5
6	6	54061.714286	50002.0
3	3	45146.000000	45146.0

Recency

	Cluster	Average	Median
6	6	54.714286	63.0
1	1	48.640091	51.0
7	7	50.951531	51.0
2	2	49.099010	50.5
0	0	49.335897	50.0
5	5	47.969388	47.0
4	4	47.240602	46.0
3	3	28.000000	28.0

MntLuxury

	Cluster	Average	Median
6	6	514.285714	541.0
2	2	411.131188	304.0
7	7	384.734694	263.5
4	4	372.101504	251.5
0	0	375.346154	238.0
1	1	359.697039	226.0
5	5	345.581633	196.0
3	3	48.000000	48.0

MntNecessity

	Cluster	Average	Median
2	2	231.398515	104.5
0	0	234.446154	100.5
6	6	233.857143	97.0
7	7	249.535714	90.5
5	5	227.248299	86.0
4	4	216.394737	82.5
1	1	226.977221	80.0
3	3	5.000000	5.0

- **Cluster 267 = major consumers**
  - Distinguish between them is their income
  - Also top 3 for recency -> losing them
- **Cluster 3 = minor consumers**
  - Can ignore when design marketing strategies

Exploring Each Clusters

- **267 complain comparatively less**
  - Doesn't affect people's decision
  - Can sometimes ignore complains
- **267 accept campaign offer -> effective**

Complain

	Cluster	Average
1	1	0.013667
5	5	0.013605
7	7	0.012755
0	0	0.007692
2	2	0.004950
4	4	0.003759
3	3	0.000000
6	6	0.000000

Accepting

	Cluster	Average
6	6	0.142857
2	2	0.113861
7	7	0.109694
0	0	0.107692
5	5	0.105442
1	1	0.102506
4	4	0.101504
3	3	0.000000