



# Customer personality Analysis

2021144276 비즈니스인포매틱스 장은준



# 목차

1. 개요
2. EDA
3. Preprocessing
4. 군집화
5. 그룹화 결과 Insight
6. 정리 및 결론

---

# 1. 개요

## 1.1 개요 - 비즈니스 시나리오



소비자 군집화 결과에 따라 각 군집에  
맞는 마케팅 실행 혹은 서비스 개선.

---

## 2. EDA



## 2.1 데이터 유형 확인 및 NULL 값 확인

### People

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



## 2.1 데이터 유형 확인 및 NULL 값 확인

### **Products**

MntWines: Amount spent on wine in last 2 years

MntFruits: Amount spent on fruits in last 2 years

MntMeatProducts: Amount spent on meat in last 2 years

MntFishProducts: Amount spent on fish in last 2 years

MntSweetProducts: Amount spent on sweets in last 2 years

MntGoldProds: Amount spent on gold in last 2 years



## 2.1 데이터 유형 확인 및 NULL 값 확인

### Promotion

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

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

Response: 1 if customer accepted the offer in the last campaign, 0 otherwise





## 2.1 데이터 유형 확인 및 NULL 값 확인

### Place

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

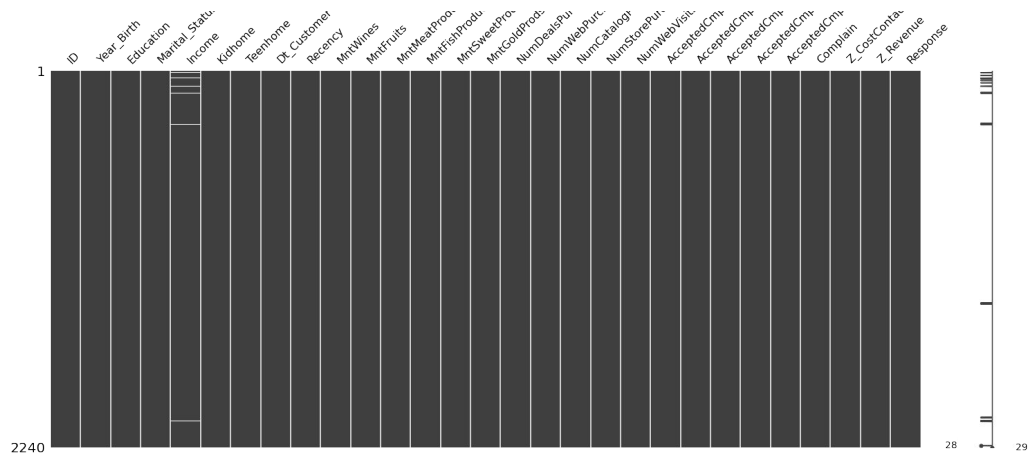
NumWebVisitsMonth: Number of visits to company's web site in the last month

## 2.1 데이터 유형 확인 및 NULL 값 확인

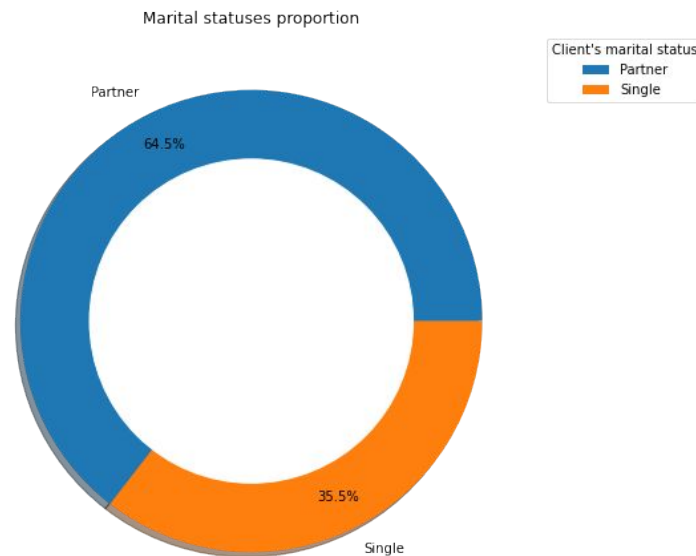
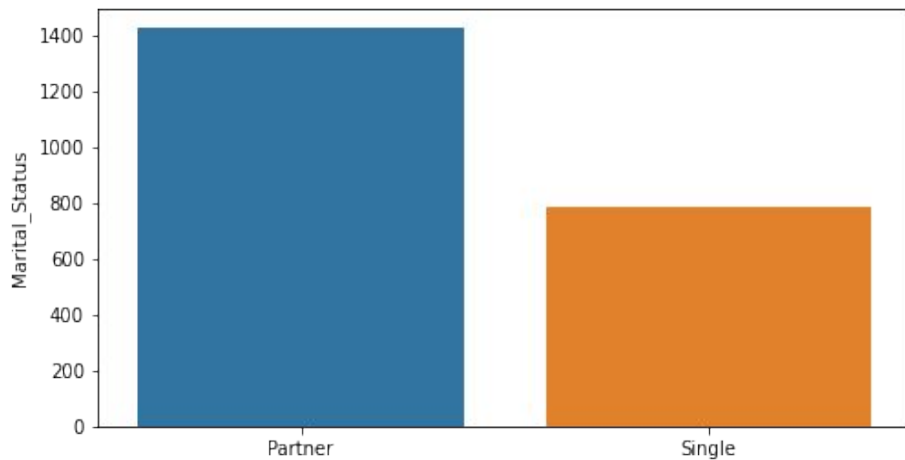
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 29 columns):
```

#	Column	Non-Null Count	Dtype
0	ID	2240 non-null	int64
1	Year_Birth	2240 non-null	int64
2	Education	2240 non-null	object
3	Marital_Status	2240 non-null	object
4	Income	2216 non-null	float64
5	Kidhome	2240 non-null	int64
6	Teenhome	2240 non-null	int64
7	Dt_Customer	2240 non-null	object
8	Recency	2240 non-null	int64
9	MntWines	2240 non-null	int64
10	MntFruits	2240 non-null	int64
11	MntMeatProducts	2240 non-null	int64
12	MntFishProducts	2240 non-null	int64
13	MntSweetProducts	2240 non-null	int64
14	MntGoldProds	2240 non-null	int64
15	NumDealsPurchases	2240 non-null	int64
16	NumWebPurchases	2240 non-null	int64
17	NumCatalogPurchases	2240 non-null	int64
18	NumStorePurchases	2240 non-null	int64
19	NumWebVisitsMonth	2240 non-null	int64
20	AcceptedCmp3	2240 non-null	int64
21	AcceptedCmp4	2240 non-null	int64
22	AcceptedCmp5	2240 non-null	int64
23	AcceptedCmp1	2240 non-null	int64
24	AcceptedCmp2	2240 non-null	int64
25	Complain	2240 non-null	int64
26	Z_CostContact	2240 non-null	int64
27	Z_Revenue	2240 non-null	int64
28	Response	2240 non-null	int64

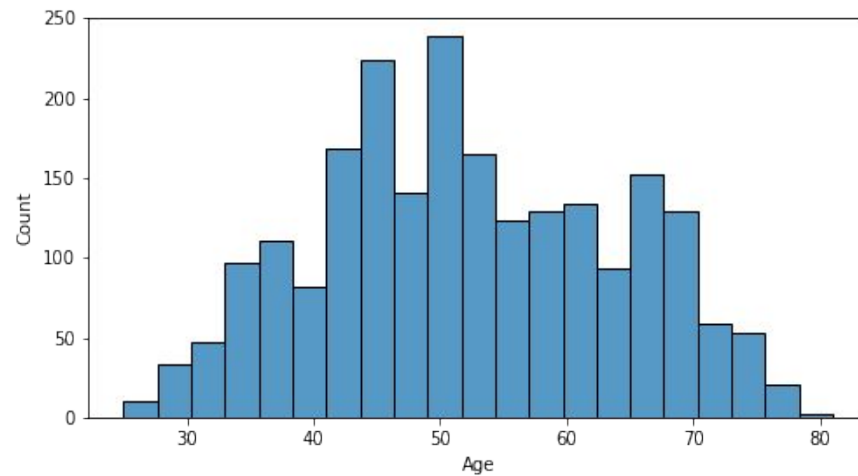
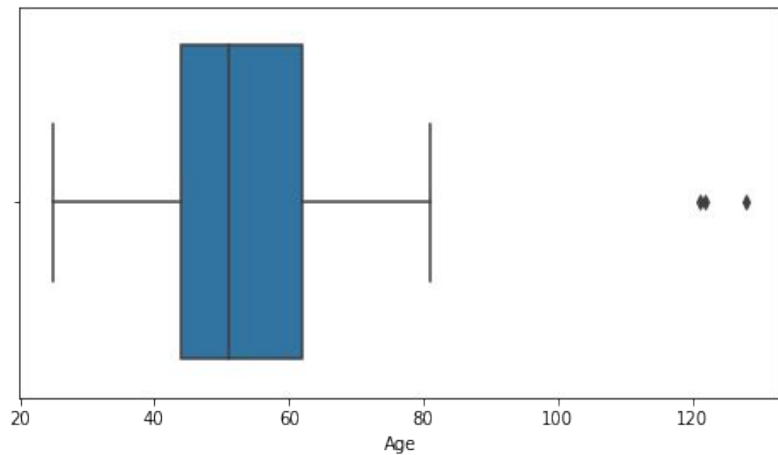
```
dtypes: float64(1), int64(25), object(3)
memory usage: 507.6+ KB
```



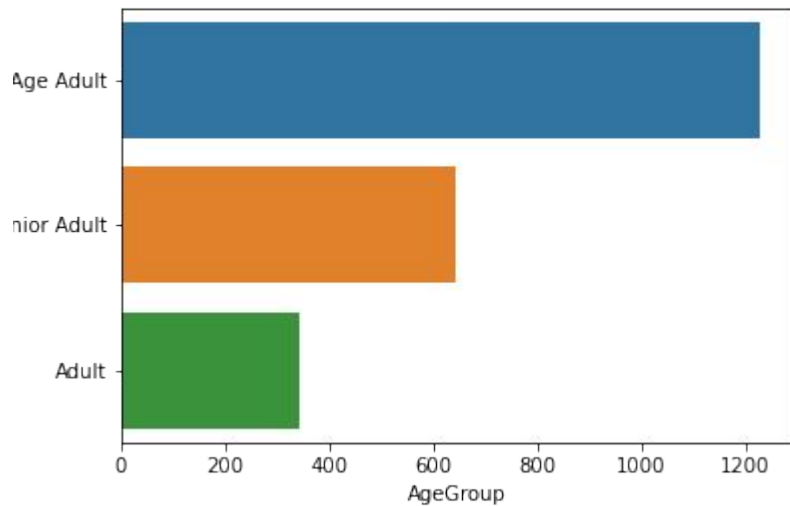
## 2.2 칼럼 확인 - Marital\_Status



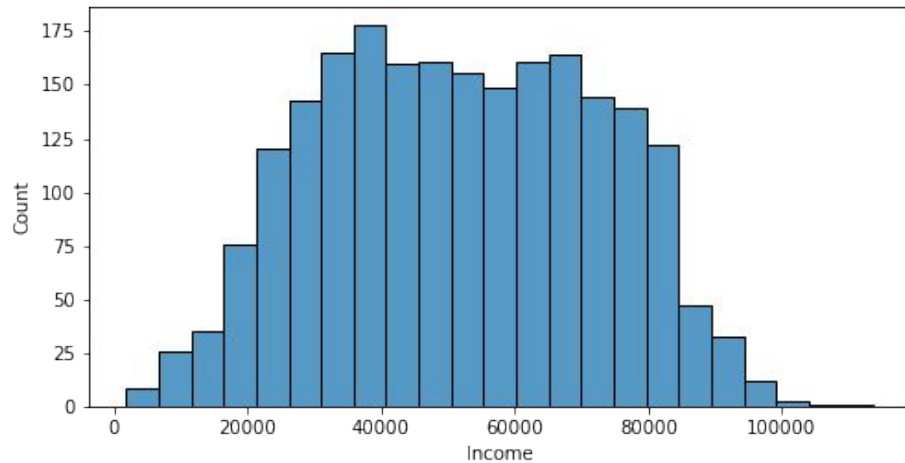
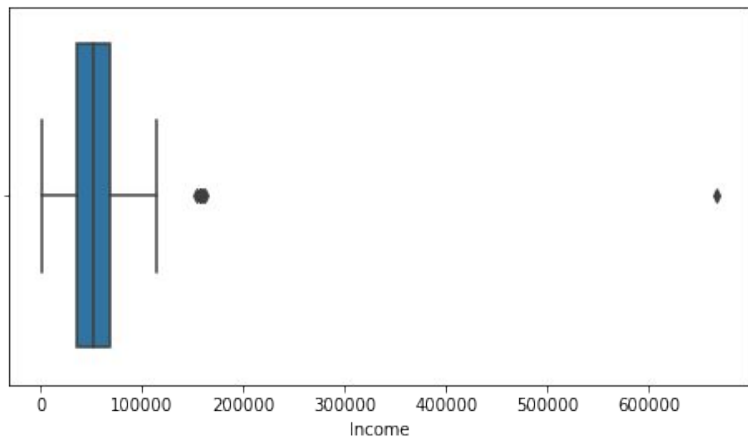
## 2.2 칼럼 확인 - Age



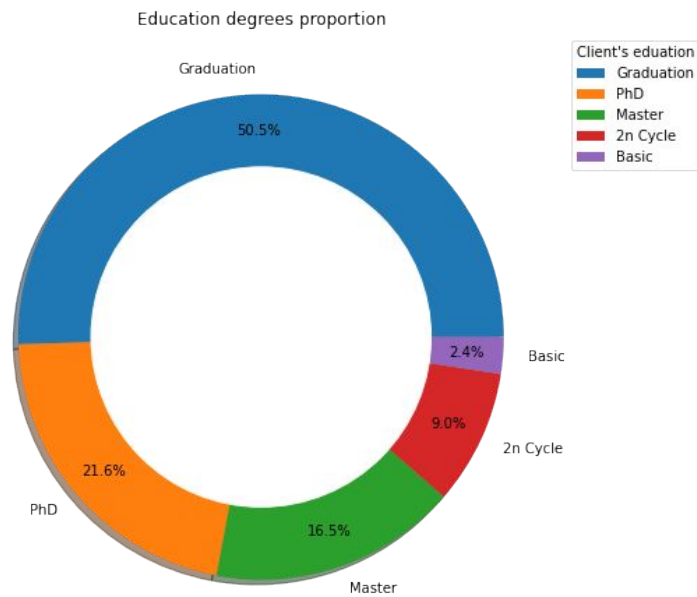
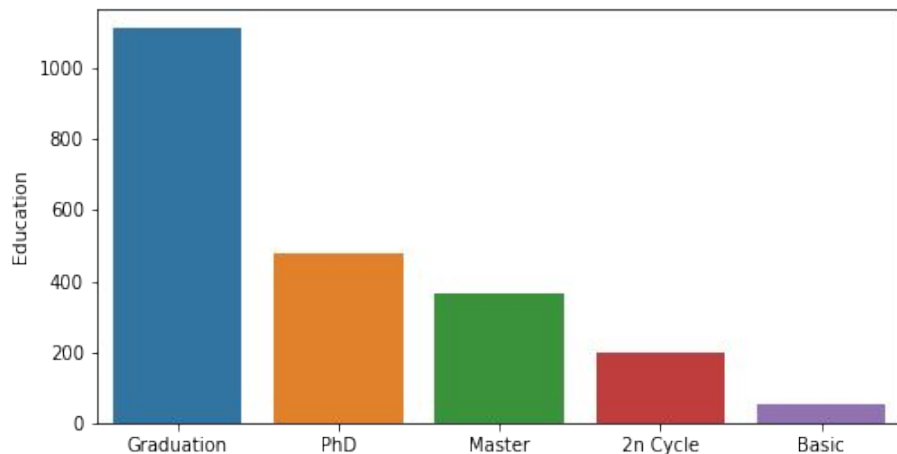
## 2.2 칼럼 확인 - AgeGroup



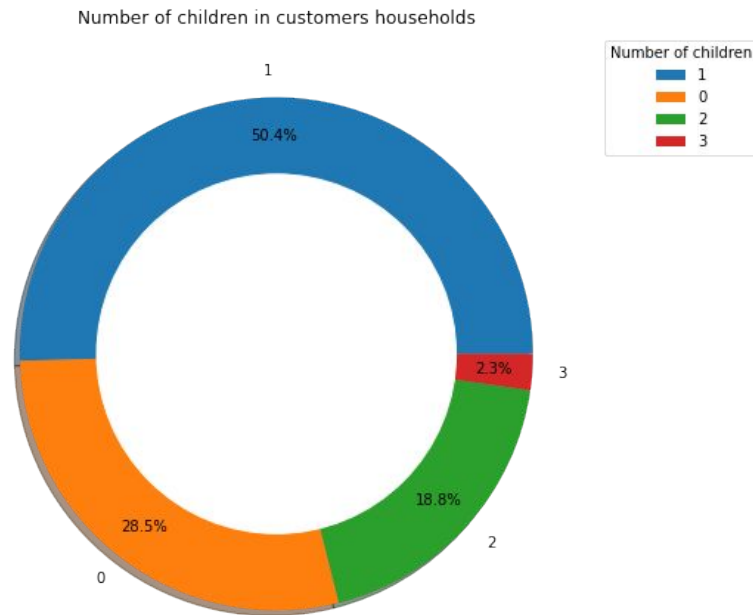
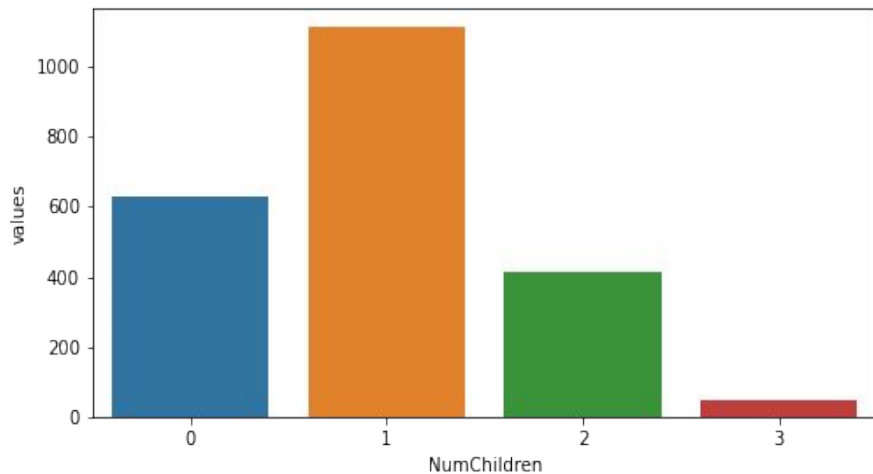
## 2.2 칼럼 확인 - Income



## 2.2 칼럼 확인 - Education

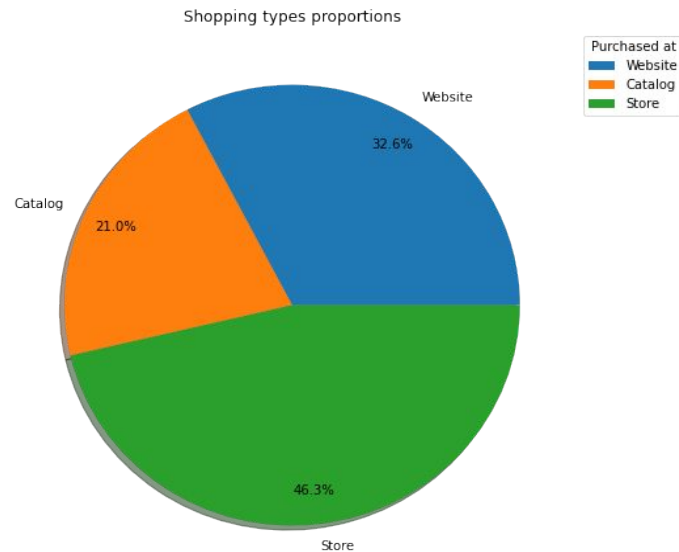
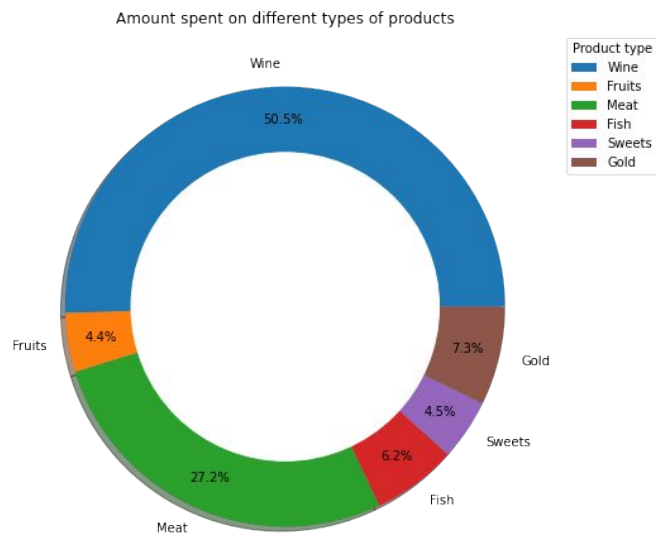


## 2.2 칼럼 확인 - NumChildren

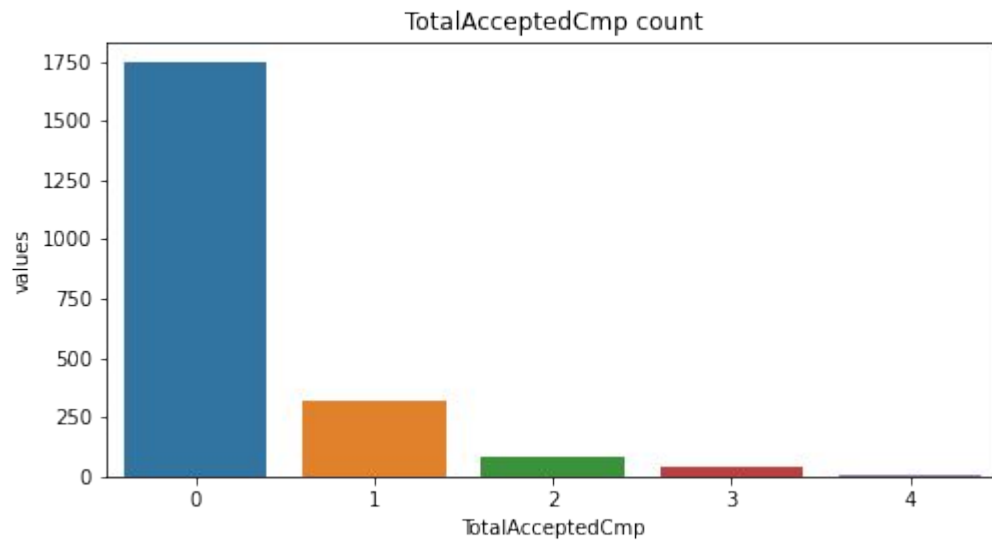




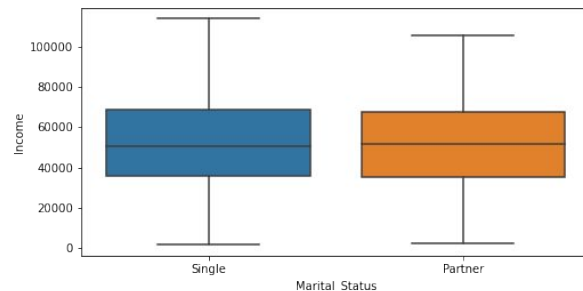
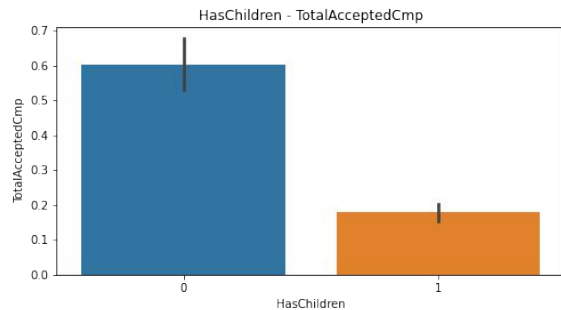
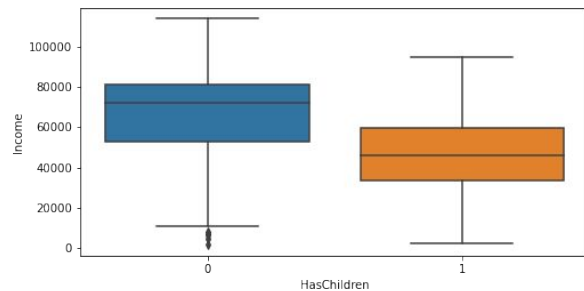
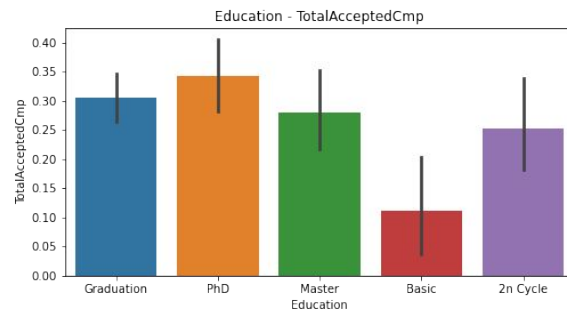
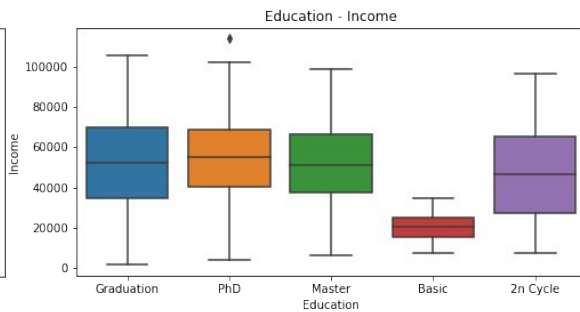
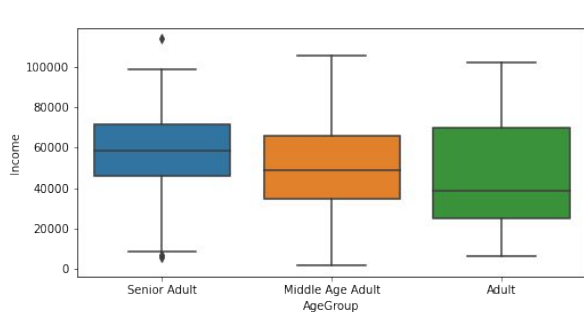
## 2.2 칼럼 확인 - Mnt, how to purchase



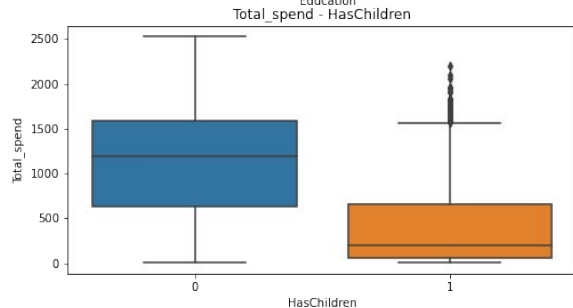
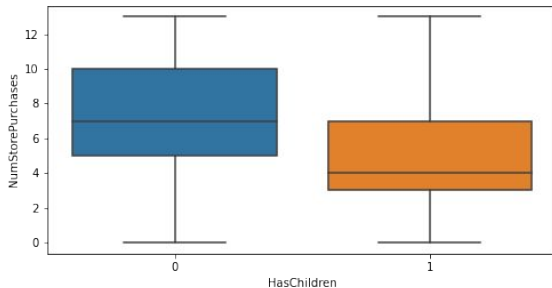
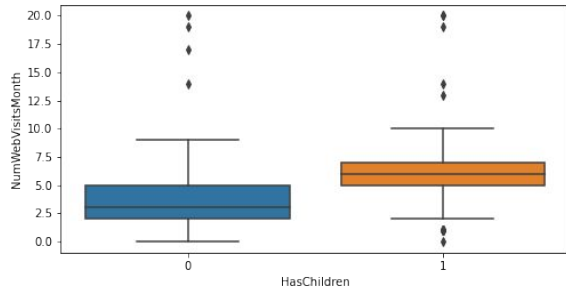
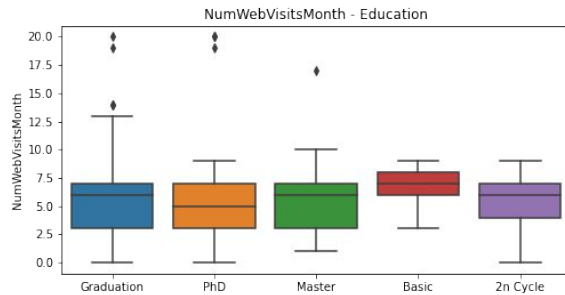
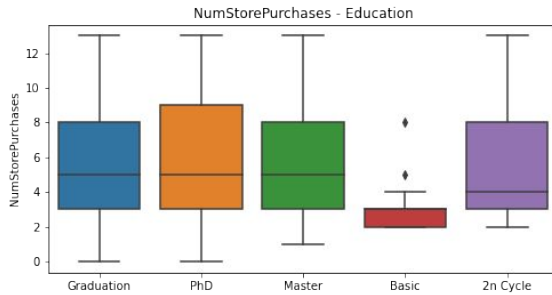
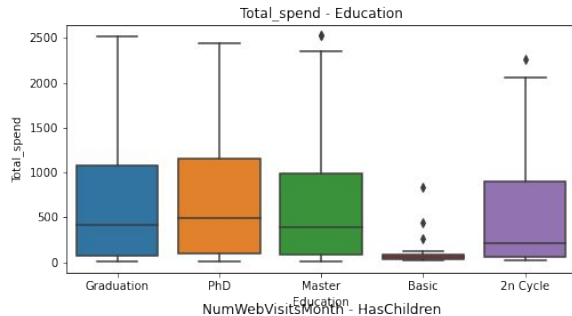
## 2.2 칼럼 확인 - Cmp



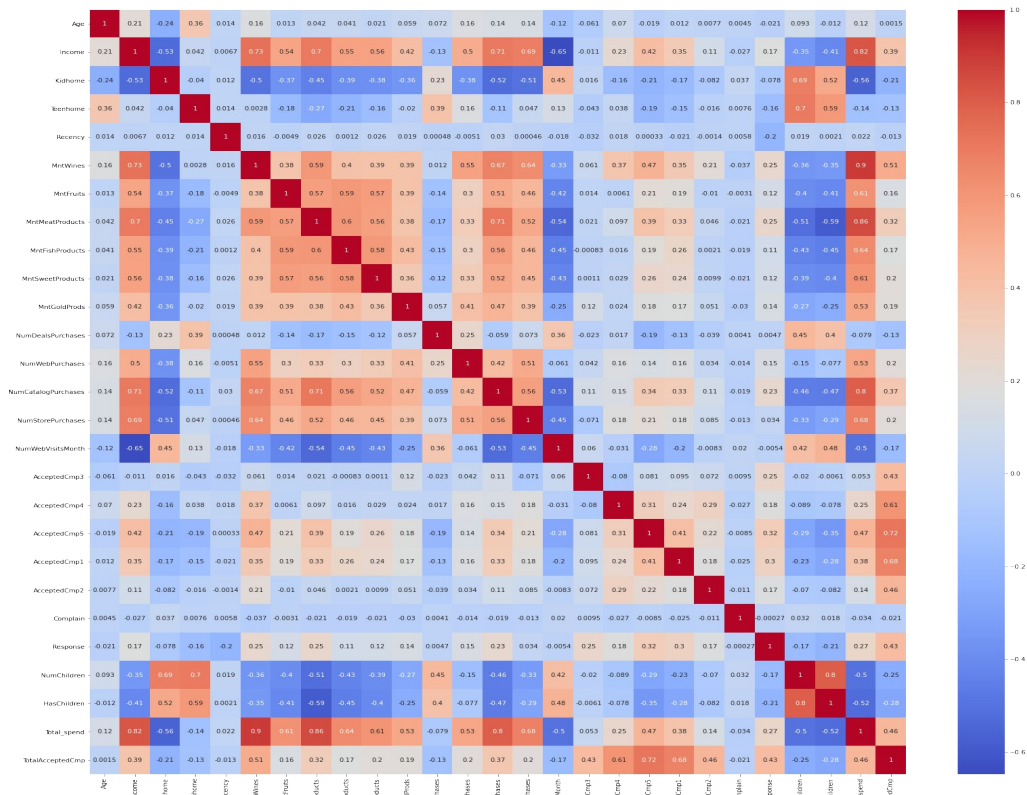
## 2.2 칼럼 확인 (칼럼간 관계)



## 2.2 칼럼 확인 (칼럼간 관계)



## 2.3 변수간 상관관계



---

## 3. Preprocessing



## 3.1 데이터 label encoder , standard scaler 적용

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
scaled_data = scaler.fit_transform(df_result)
scaled_df = pd.DataFrame(data = scaled_data, columns = df_result.columns)
```

```
from sklearn.preprocessing import LabelEncoder
lb_encoder = LabelEncoder()
df_result['Education'] = lb_encoder.fit_transform(df_result['Education'])

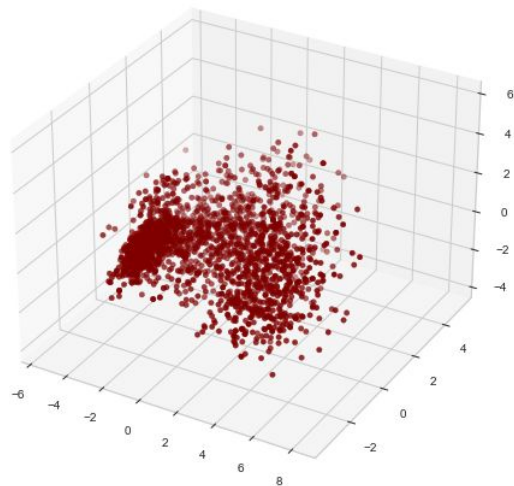
lb_encoder2 = LabelEncoder()
df_result['Marital_Status'] = lb_encoder2.fit_transform(df_result['Marital_Status'])
```

## 3.2 PCA TRANSFORM

```
pca = PCA(n_components=3)
pca.fit(scaled_df)
pca_df = pd.DataFrame(pca.transform(scaled_df), columns=["columns1", "columns2", "columns3"])

x = pca_df["columns1"]
y = pca_df["columns2"]
z = pca_df["columns3"]

fig = plt.figure(figsize=(12,8))
ax = fig.add_subplot(111, projection="3d")
ax.scatter(x, y, z, c="maroon", marker="o")
plt.savefig('3d_pca')
plt.show()
```

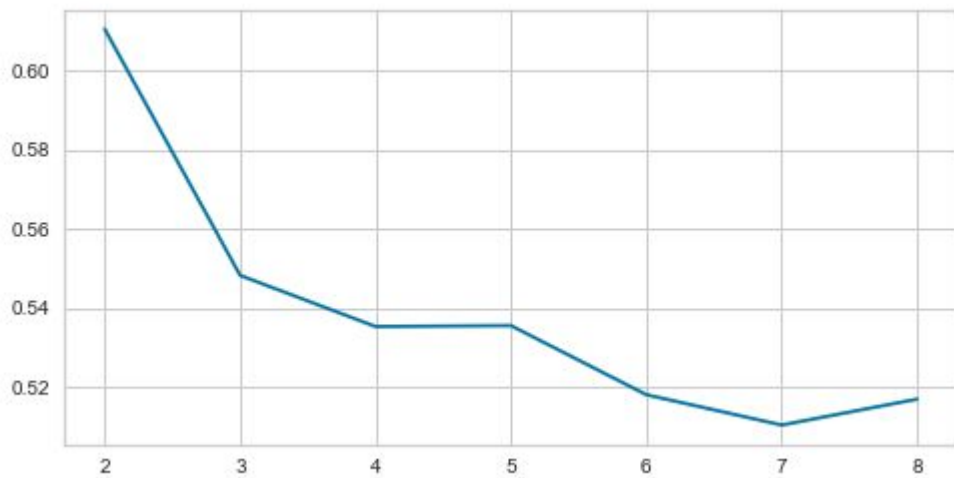




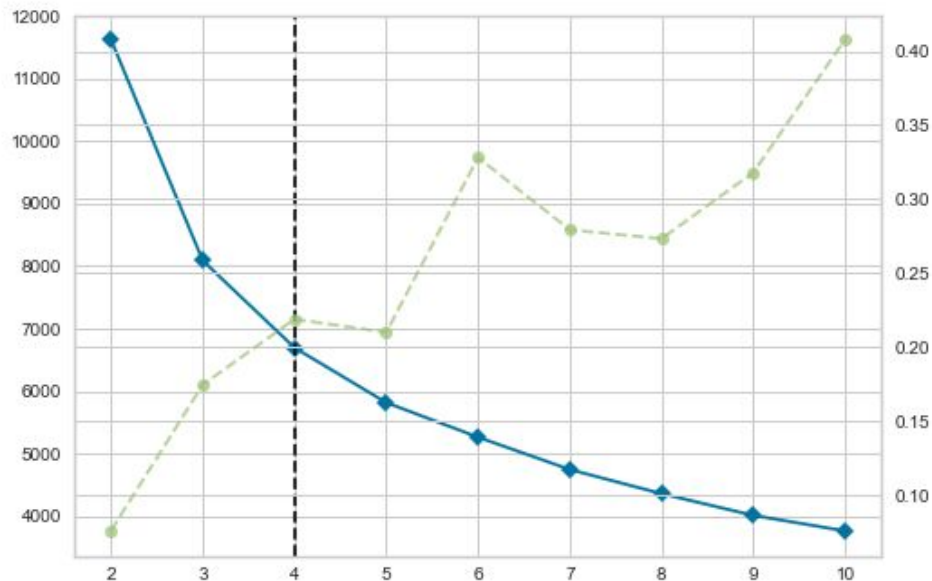
---

## 4. Clustering

## 4.1 Silhouette Score 로 군집 개수 찾기



## 4.2 Elbow Method 로 군집 개수 찾기



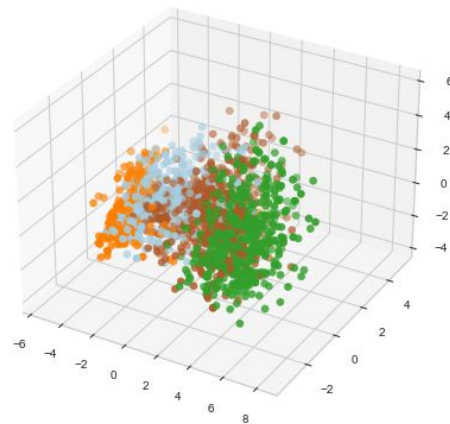
## 4.3 PCA VISUALIZATION

Cluster 1 의 개수 505

Cluster 2 의 개수 641

Cluster 2 의 개수 613

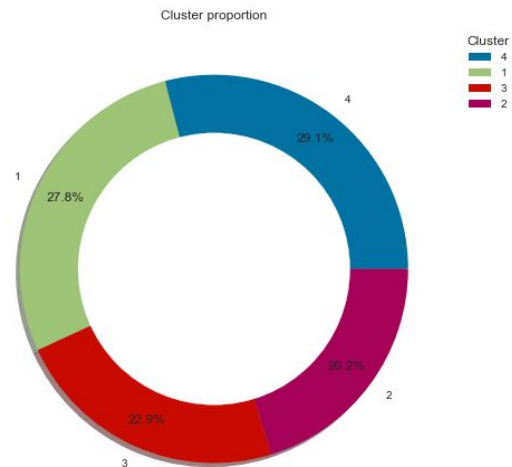
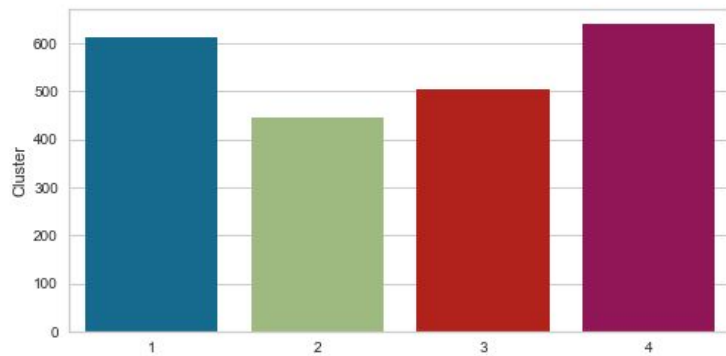
Cluster 2 의 개수 446



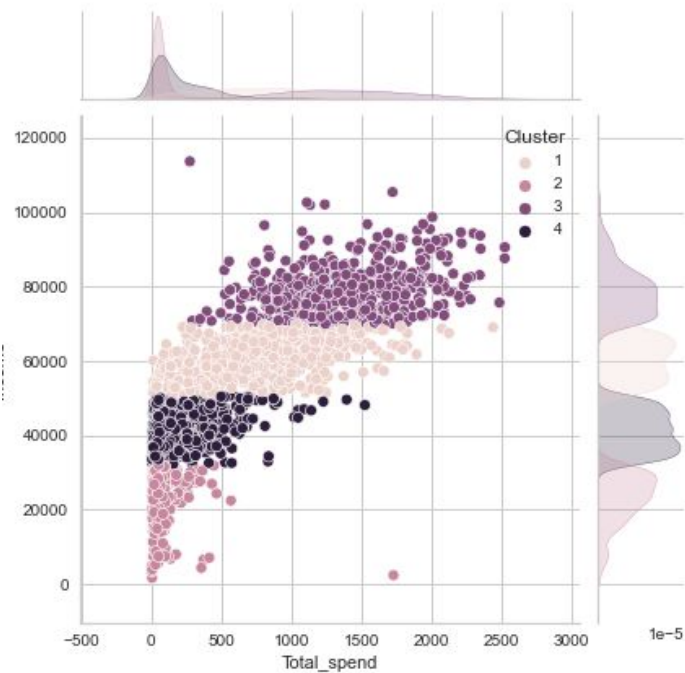
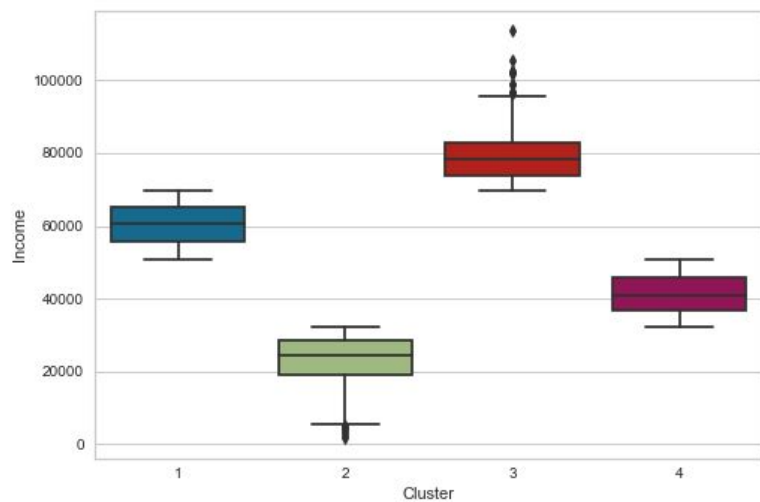
---

## 5. 그룹화 결과 insight 시각화

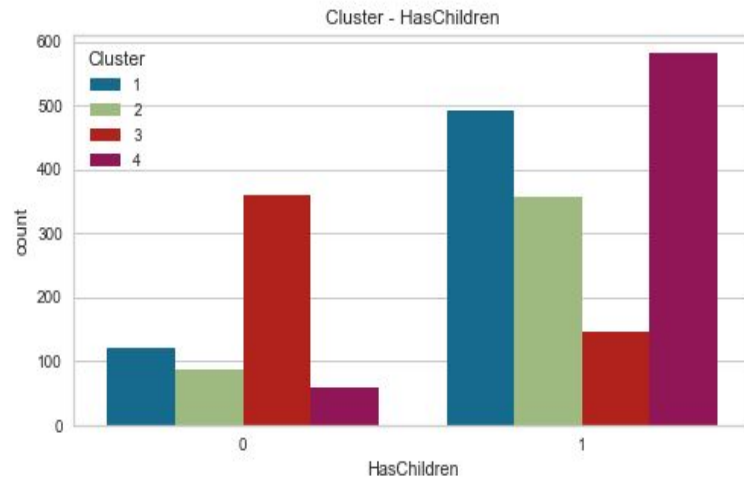
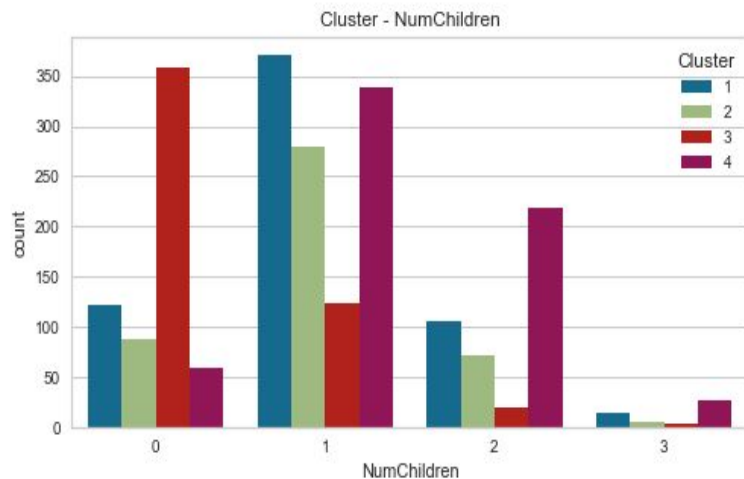
## 5.1 그룹화 결과



## 5.2 Cluster - Income

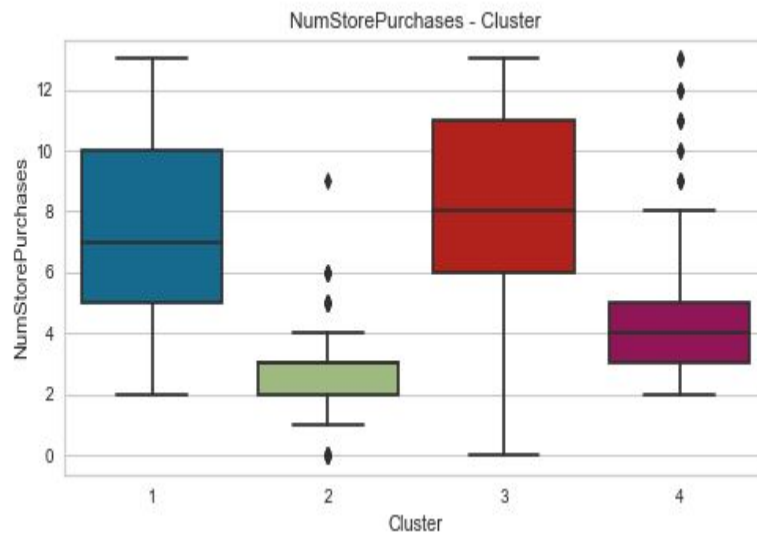
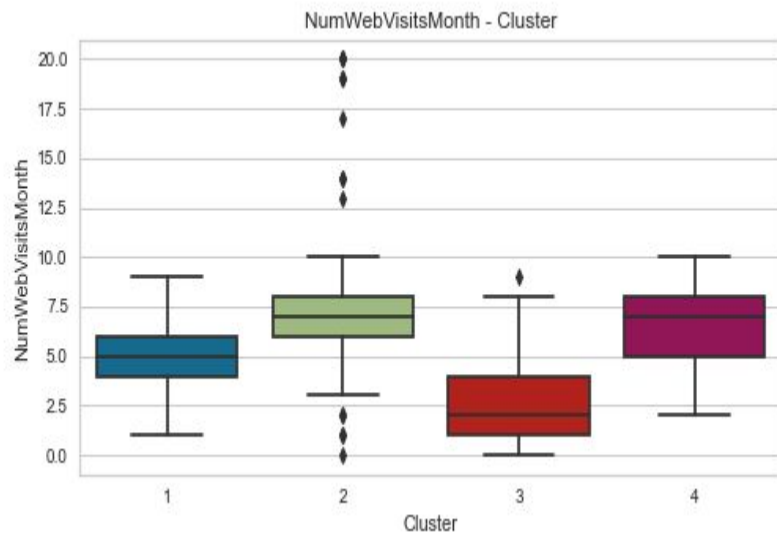


## 5.3 Cluster - NumChildren(HasChildren)





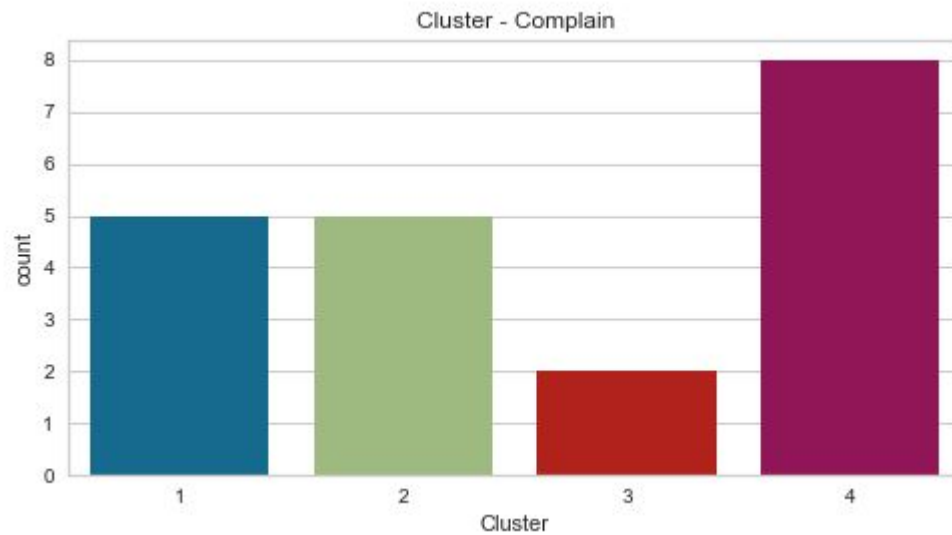
## 5.4 Cluster-NumStorePurchases or NumWebVisit



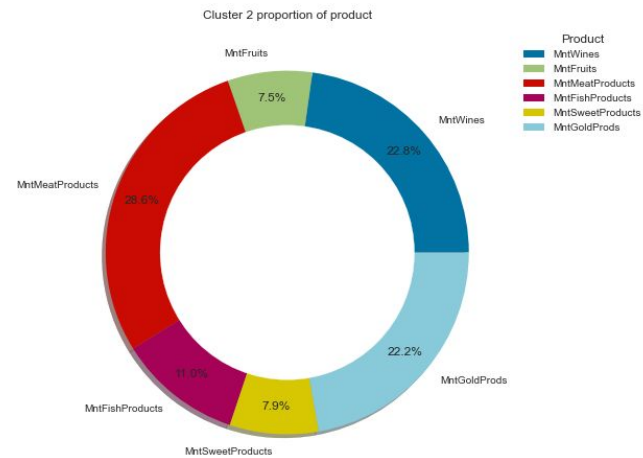
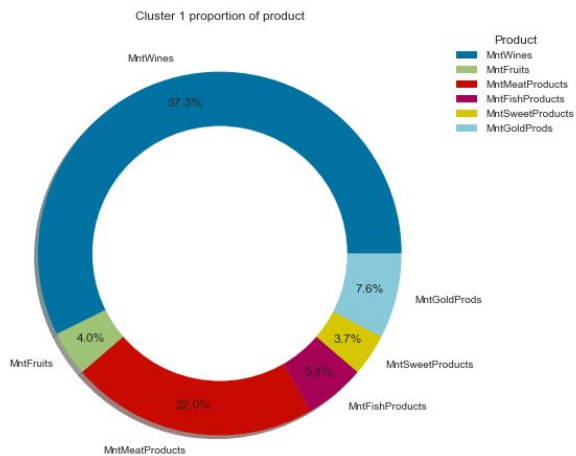
## 5.5 Cluster - Cmp



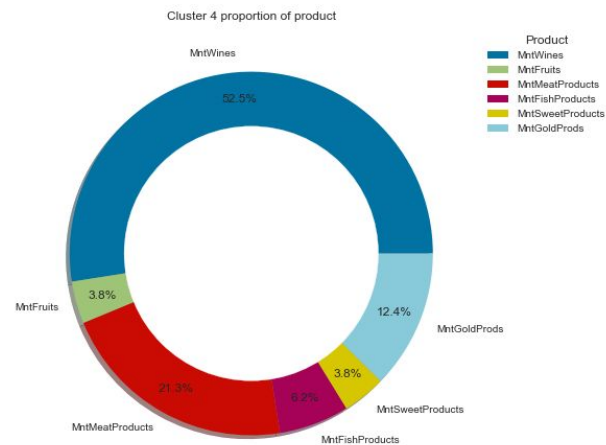
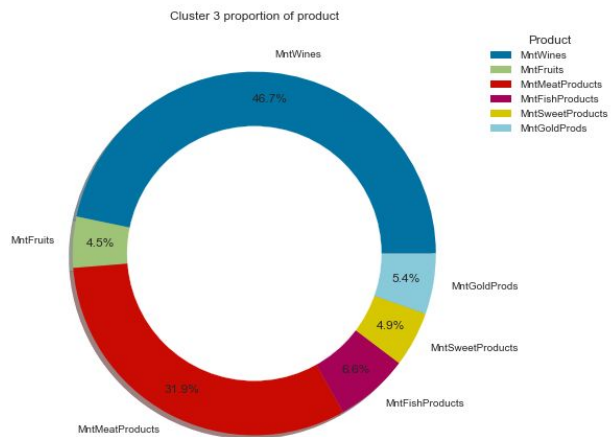
## 5.6 Cluster - Complain



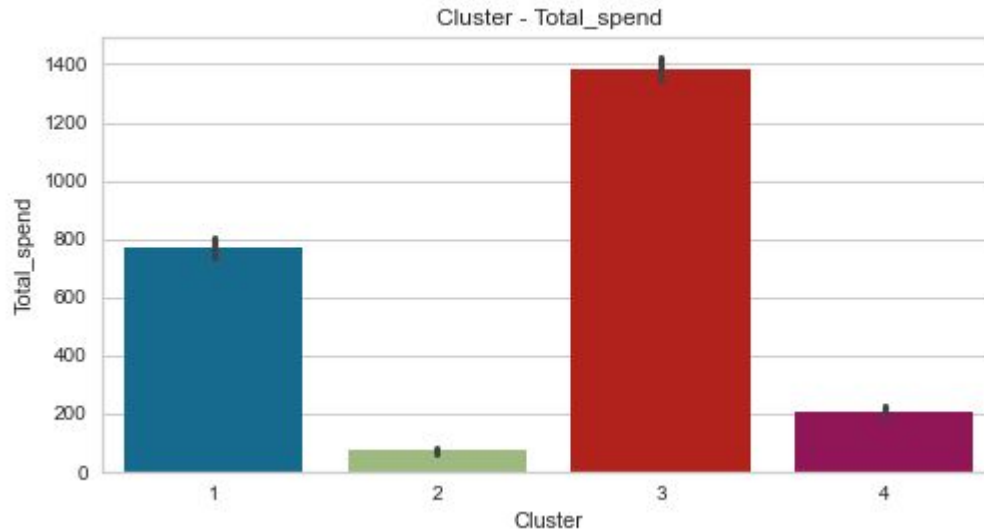
## 5.7 Cluster - Mnt



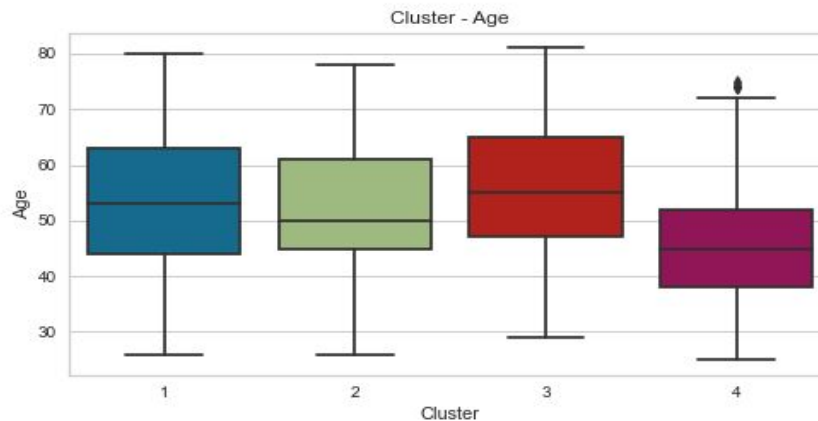
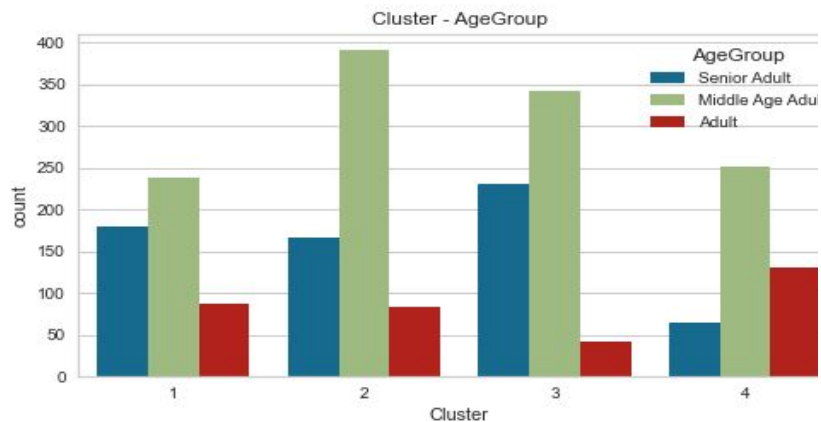
## 5.7 Cluster - Mnt



## 5.8 Cluster - Total\_spend



## 5.9 Cluster - Age(AgeGroup)



---

## 6. 정리 및 결론



## 6. 정리

### 그룹1 :

**Income** : 평균

**totalspend** : 평균

**HasChildren** : 아이가 없는 것보다 아이가 있는 경우가 많음

**store-web** : **web**사이트에서 구입보다 **store** 방문구매가 더 많음

**cmp** : 최근 캠페인 뿐만 아니라 캠페인 3을 많이 참여함, 또한 전반적으로

다른 그룹보다 더 많이 캠페인을 참여함

**complain** : 평균적인 **complain** 기록 중

**Mnt** : 와인 > 고기 > 금 > 생선 > 과일 > 설탕류

### 그룹3 :

**Income** : 제일 높은 수입을 기록하는 그룹

**totalspend** : 제일 많은 소비를 보이고 있음

**HasChildren** : 유일하게 아이가 있는 경우보다 아이가 없는 경우가 더 많음

**store - web** : 웹사이트에서 구매한 경우보다 **store** 에서 구매한 경우가 훨씬 많음

**cmp** : 캠페인 참여가 전반적으로 높음, 제일 높은 캠페인 참여도 보임

**complain** : 제일 적은 **complain** 을 보이고 있음

와인> 고기 > 그 이외에는 거의 비슷

### 그룹2:

**Income** : 제일 낮은 수입을 기록하는 그룹

**totalspend** : 제일 낮은 소비를 보이고 있음

**HasChildren** : 아이가 없는 경우보다 있는 경우가 많음

**store-web** : **web**사이트에서 구입한 경우가 **store** 방문보다 더 많음

**cmp** : 최근 캠페인도 많이 참여하지 않았으며 전반적으로 참여 하지 않음

그러나 캠페인 3번은 그나마 참여한 경우로 보임

**complain** : 전반적으로 **complain** 도 평균적임

**Mnt** : 고기 > 와인 > 금 > 생선 > 설탕 > 과일

### 그룹4 :

**income** : 3번째의 수입을 보이고 있음

**total\_spend** : 3번째의 소비를 보이고 있음

**haschildren** : 마찬가지로 아이가 있는 경우가 많음

또한 아이가 없는 경우가 가장 적고, 아이가 있는 경우의 수가 가장 많음

**web-store** : **web**사이트 방문이 **store** 방문보다 더 많음

**cmp** : 캠페인 참여도도 3번째임, 그룹2와 마찬가지로

캠페인 3번을 그나마 가장 많이 참여함

**complain** : 컴플레인 이 가장 많은 그룹임, **mnt** : 와인 > 고기 > 금

## 6. 결론

### - 이 회사의 major 소비 그룹 (3그룹):

- -> 각 상품 유형별 소비 비율이 타 그룹과 비슷하면서 전체 소비량이 많기 때문에 타 그룹에 대한 서비스를 강화하면 소비량이 저절로 유지가 될 듯.
- -> 캠페인 참여율이 가장 높은 그룹이면서 web사이트 방문이 가장 적은 그룹임. web사이트 방문을 유도하는 캠페인을 많이 시행하여 web 구매도 촉진시키면 더 많은 소비를 할 것으로 보임.

### - 이 회사의 worst 소비 그룹 (2그룹):

- -> 그룹 3과는 모든 점이 반대인 그룹
- -> 그룹 2의 소비량을 늘리기 위해서는 그룹 1번을 참조하면 좋을 듯. 그룹 1번은 수입이 평균적이지만, 소비량이 두 번째로 많은 그룹.
- -> 그룹 1번은 캠페인을 많이 참여하는 모습을 보임, 그룹 2번은 캠페인을 거의 참여하지 않는데, 유일하게 3번째 캠페인은 많이 참여.
- 따라서 3번째 캠페인과 유사한 캠페인을 앞으로 기획하여 그룹 2번의 캠페인 참여를 유도.
- -> 그룹 전체를 살펴보면 store방문하여 소비를 하는 그룹이 소비량이 많고 web에서 구매하는 그룹이 소비량이 적음. 위에서 캠페인 참여를 독려할 때, store 방문을 유도하는 캠페인으로 꾸리면 좋을 듯.
- -> 그룹 2가 다른 그룹에 비해서 gold 에 대한 소비가 높다.wine 소비량은 다른 그룹의 절반 비율밖에 되지 않는다. wine 상품 같은 경우에는 다른 그룹에서 이미 많이 사고 있기 때문에
- gold 상품에 대한 마케팅을 강화해서 그룹 2의 높은 소비를 유도한다.

### - 이 회사의 가장 많은 불만이 있는 그룹(4그룹):

- -> 나이대가 가장 어린 그룹.
- -> 아이가 없는 경우가 가장 적고 아이가 있는 비율이 가장 높음.
- -> 컴플레인인 아이가 있기 때문에 많이 있을 수 있겠다는 유추가 가능.
- -> 아이와 관련된 서비스를 강화하여 컴플레인을 줄이고 소비 독려

---

감사합니다.

code 및 자료는 따로 파일 첨부 했습니다.