Customer Personality Analysis

Analysis of company's ideal customers

Uswatun Hasanah

Background

Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments

Data PreProcessing

About The Dataset

Data contains 29 columns with 2240 rows

```
<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
```

Subsets

People
- ID
- Year_Birt
- Education
- Marital_Status
- IncomeKidhome
- Teenhome
- Dt_Customer
- Recency
- Complain

ProductsID
- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProds

Promotion
- NumDealsPurchases
- AcceptedCmp1
- AcceptedCmp2
- AcceptedCmp3
- AcceptedCmp4
- AcceptedCmp5
- Response

Place
- NumWebPurchases
- NumCatalogPurchases
- NumStorePurchases
- NumWebVisitsMonth

Drop Missing Value and Duplicates

```
[ ] data.isna().sum()
         ID                    0
         Year_Birth            0
         Education             0
         Marital_Status        0
         Income                24
         Kidhome               0
         Teenhome              0
         Dt_Customer           0
         Recency               0
         MntWines              0
         MntFruits             0
         MntMeatProducts       0
         MntFishProducts       0
         MntSweetProducts      0
         MntGoldProds          0
         NumDealsPurchases     0
         NumWebPurchases       0
         NumCatalogPurchases   0
         NumStorePurchases     0
         NumWebVisitsMonth     0
         AcceptedCmp3          0
         AcceptedCmp4          0
         AcceptedCmp5          0
         AcceptedCmp1          0
         AcceptedCmp2          0
         Complain              0
         Z_CostContact         0
         Z_Revenue             0
         Response              0
         dtype: int64
```

```
data.isna().sum()
         ID                    0
         Year_Birth            0
         Education             0
         Marital_Status        0
         Income                0
         Kidhome               0
         Teenhome              0
         Dt_Customer           0
         Recency               0
         MntWines              0
         MntFruits             0
         MntMeatProducts       0
         MntFishProducts       0
         MntSweetProducts      0
         MntGoldProds          0
         NumDealsPurchases     0
         NumWebPurchases       0
         NumCatalogPurchases   0
         NumStorePurchases     0
         NumWebVisitsMonth     0
         AcceptedCmp3          0
         AcceptedCmp4          0
         AcceptedCmp5          0
         AcceptedCmp1          0
         AcceptedCmp2          0
         Complain              0
         Z_CostContact         0
         Z_Revenue             0
         Response              0
         dtype: int64
```

```
[ ] data = data.drop_duplicates()

    data.duplicated().sum()

    0
```
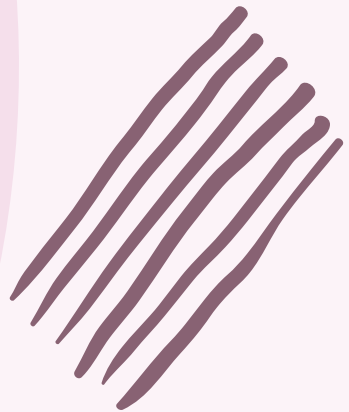
EDA

Simplifying the features

```python
data["shopping"] = data["MntWines"]+ data["MntFruits"]+ data["MntMeatProducts"]+ data["MntFishProducts"]+ data["MntSweetProducts"]+ data["MntGoldProds"]
```

```python
data['relationship']=data['Marital_Status'].replace({'Married':'in_relationship' , 'Together':'in_relationship' , 'Single':'single' , 'Divorced':'single',
                                                     'YOLO':'single' , 'Absurd':'single' , 'Widow':'single' ,'Alone':'single'})
data["Children"]=data["Kidhome"]+data["Teenhome"]
```

```python
data['members_home']= data["Children"]+data['relationship'].replace({'single':0,'in_relationship':1})
```

```python
data["Is_Parent"] = np.where(data.Children> 0, 1, 0)
```

```python
data['Purchases'] = data['AcceptedCmp1'] + data['AcceptedCmp2'] + data['AcceptedCmp3']
    + data['AcceptedCmp4'] + data['AcceptedCmp5'] + data['Response']
```

```python
data['num_purchases'] = data['NumWebPurchases'] + data['NumCatalogPurchases'] + data['NumStorePurchases'] + data['NumDealsPurchases']
```

```python
data.drop(labels=['Marital_Status','ID','Year_Birth'
                  ,'Dt_Customer',
                  'MntWines', 'MntFruits','MntMeatProducts', 'MntFishProducts',
                  'MntSweetProducts','MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases',
                  'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth',
                  'AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1',
                  'AcceptedCmp2','Z_CostContact', 'Z_Revenue',"Recency", "Complain",], axis=1, inplace=True)
```

outlier
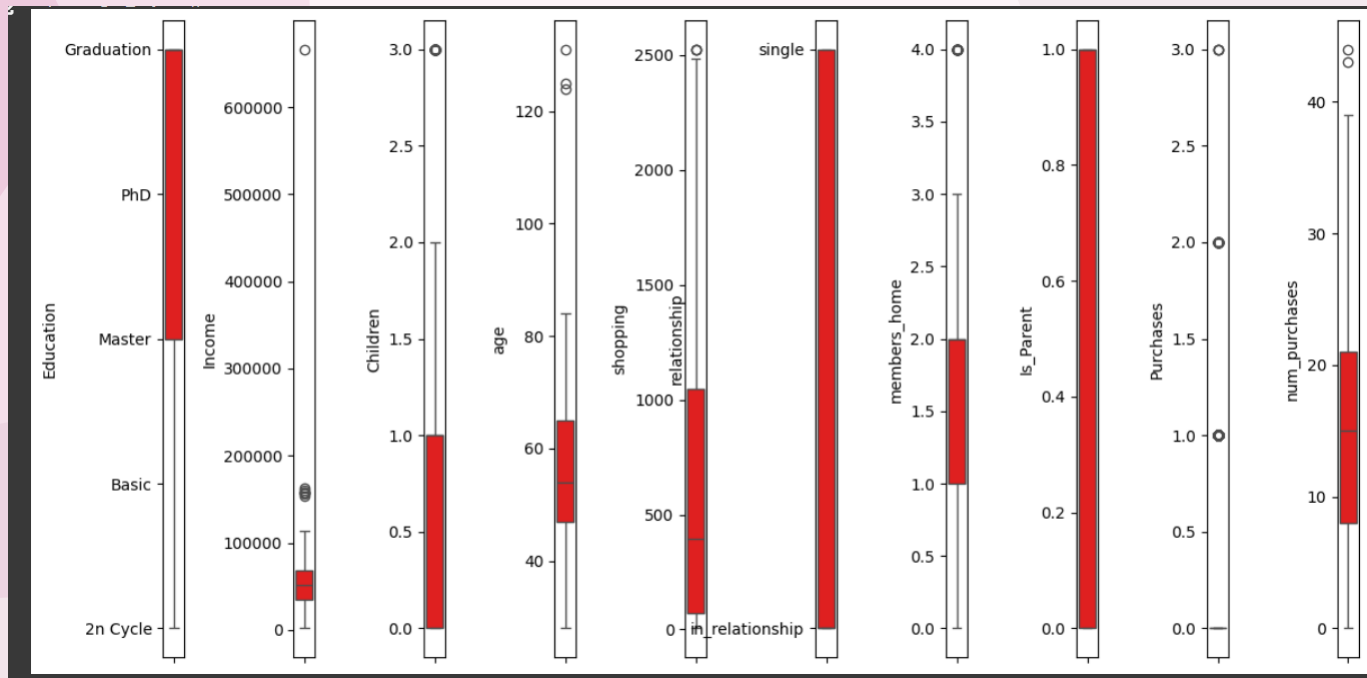
Outlier :
- Income
- Age

Outlier done drop

```
data = data[(data["age"]<80)]
data = data[(data["Income"]<150000)]
print("The total number of data-points after removing the outliers are:", len(data))

The total number of data-points after removing the outliers are: 2190
```

```
plt.figure(figsize=(12,6))

features = "age","Income"
for i in range(0, len(features)):
    plt.subplot(1, len(features), i+1)
    sns.boxplot(y=data[features[i]], color='red')
    plt.tight_layout()
```
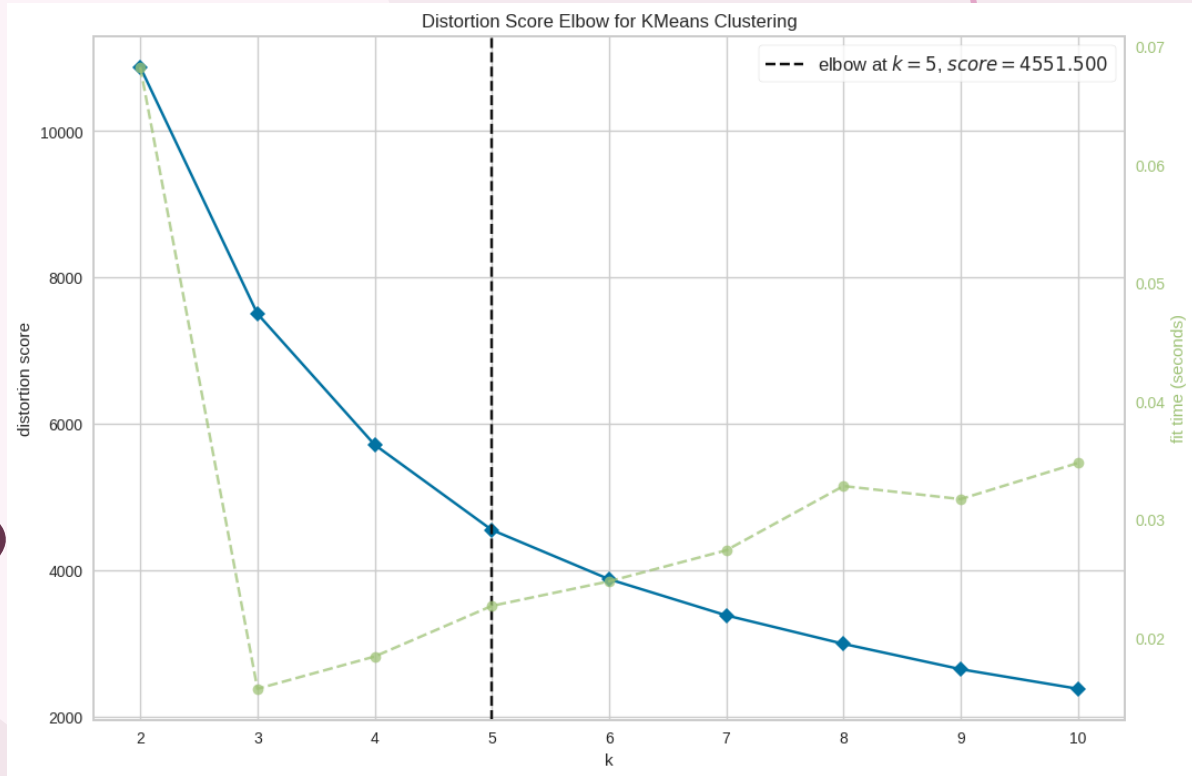

OUTLIER Done drop
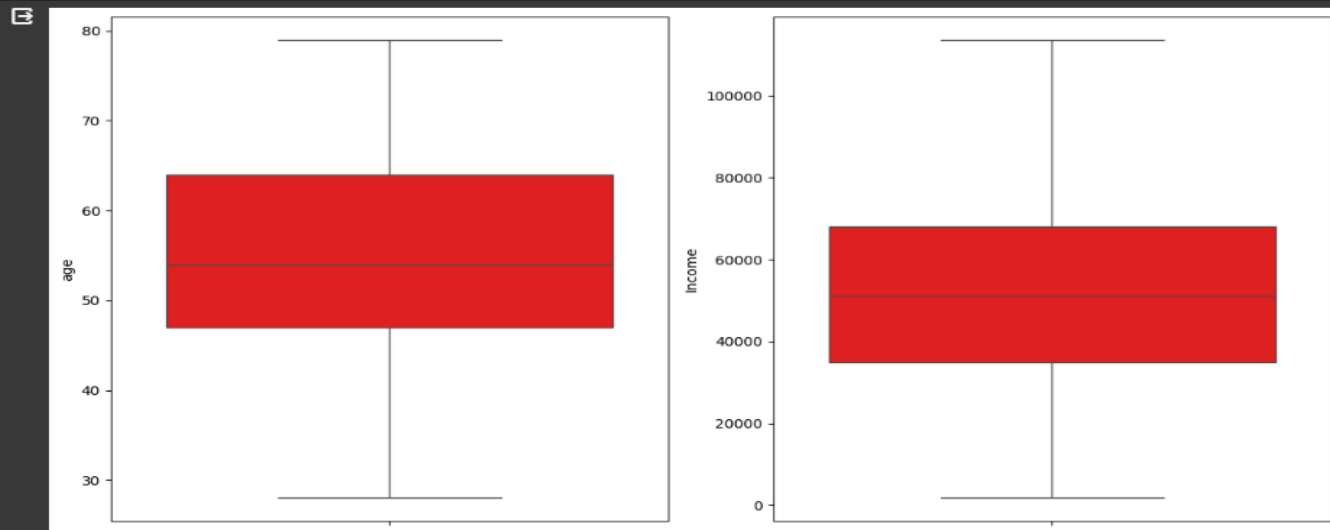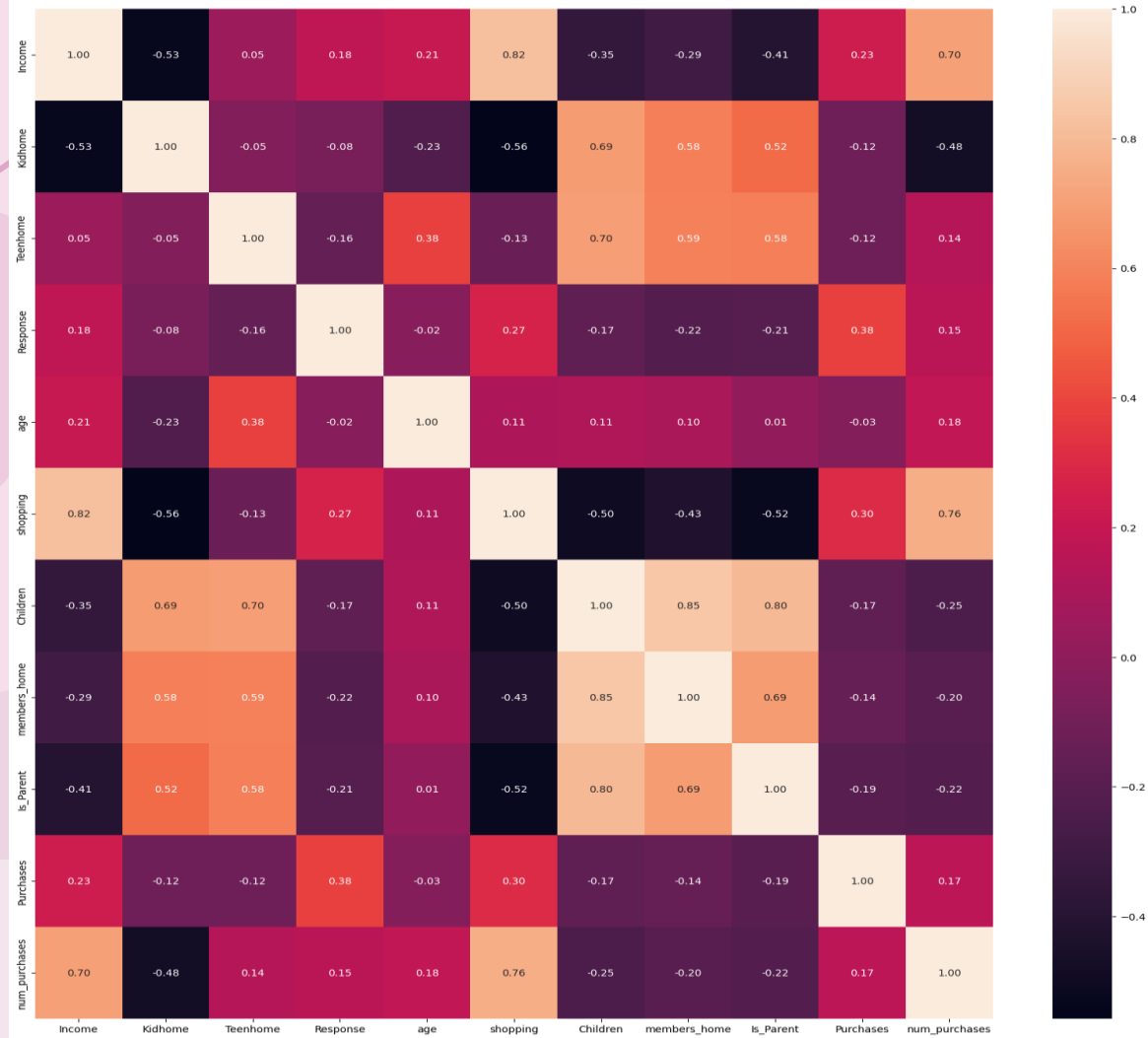
correlation amongst the features

CLUSTERING

The result of the **score plot** can be taken as the best and ideal value, which is at point 5, or **N = 5**

Distortion Score Elbow for KMeans Clustering

- - - elbow at $k = 5$, $score = 4551.500$

From the results of this customer clustering and visualized with a scatterplot as shown below.

This diagram shows the distribution of customer data which is divided into clusters according to the K-Means Clustering algorithm.

clusters in data
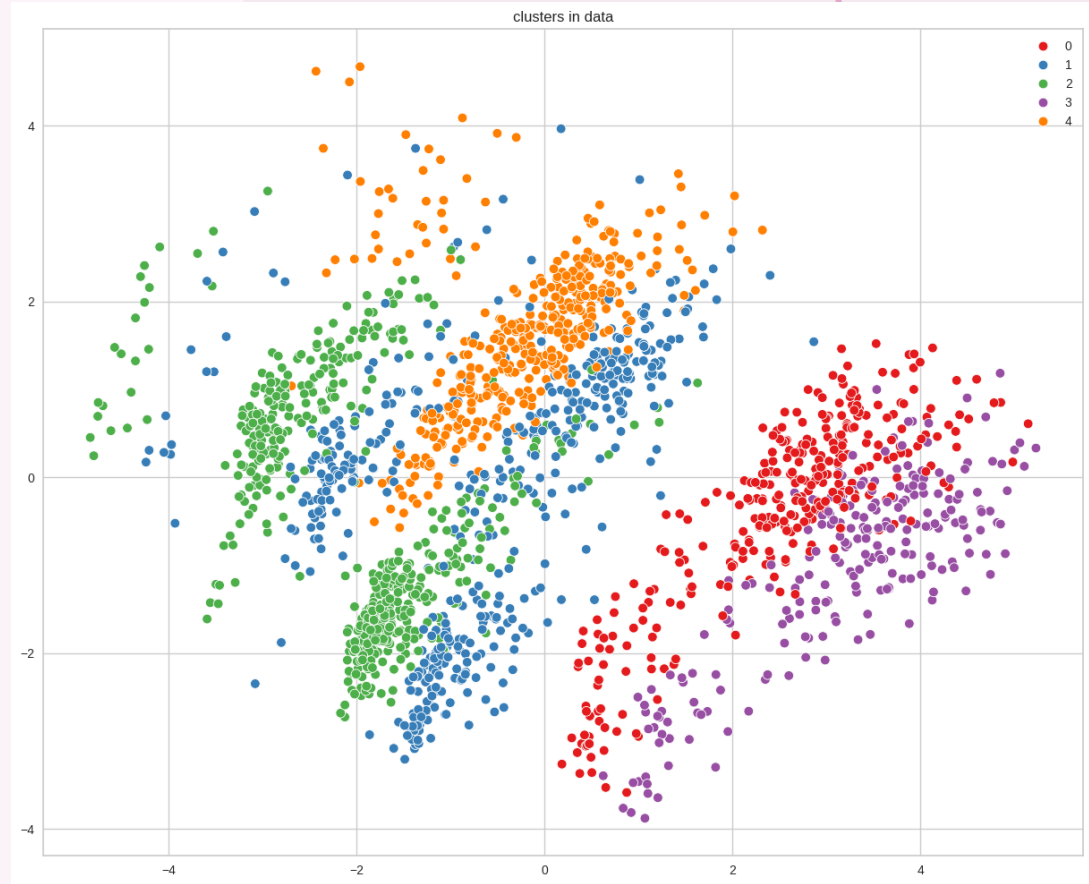
shopping capabilities and income for each cluster

Income vs shopping plot shows the clusters pattern

- cluster 1: high shopping & high income
- cluster 2: most are low shopping and income, but some are in high.
- cluster 3 & 4: most are low shopping and income
- cluster 5 : high shopping & low income

shopping and income of each cluster
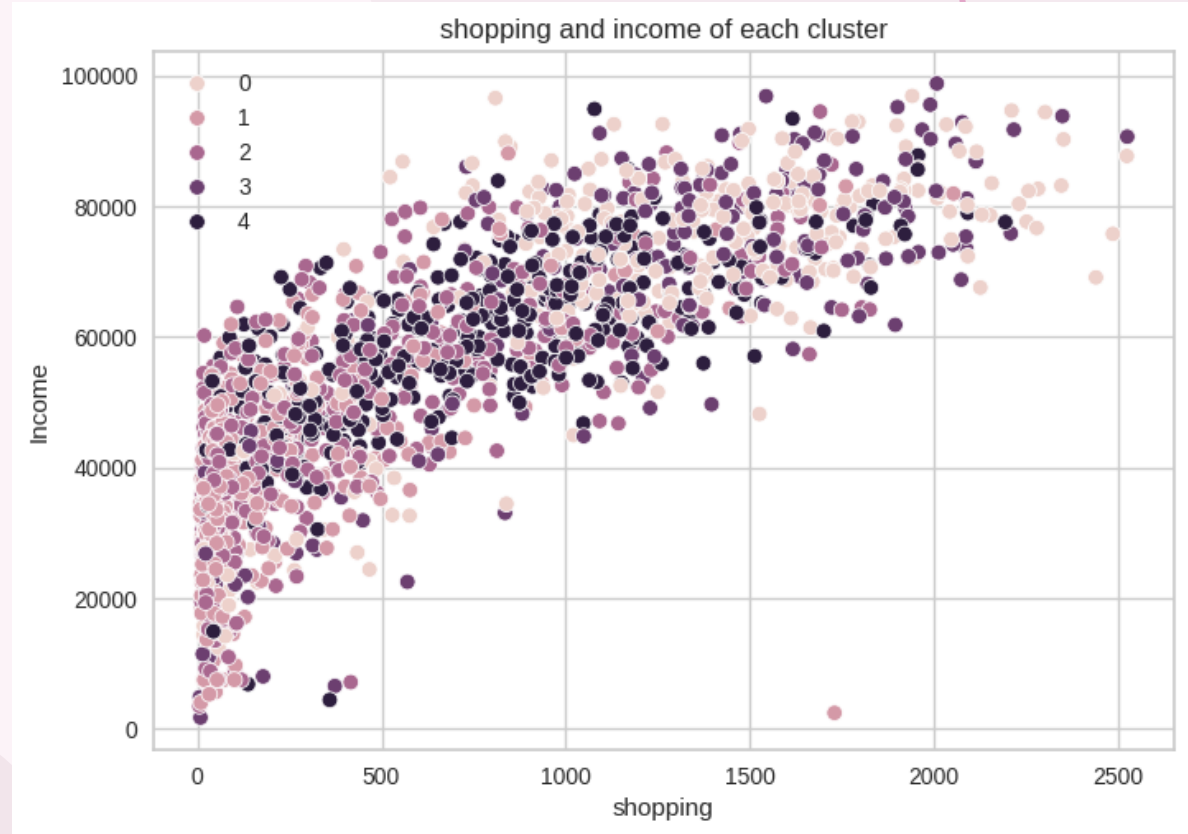
the count of each cluster
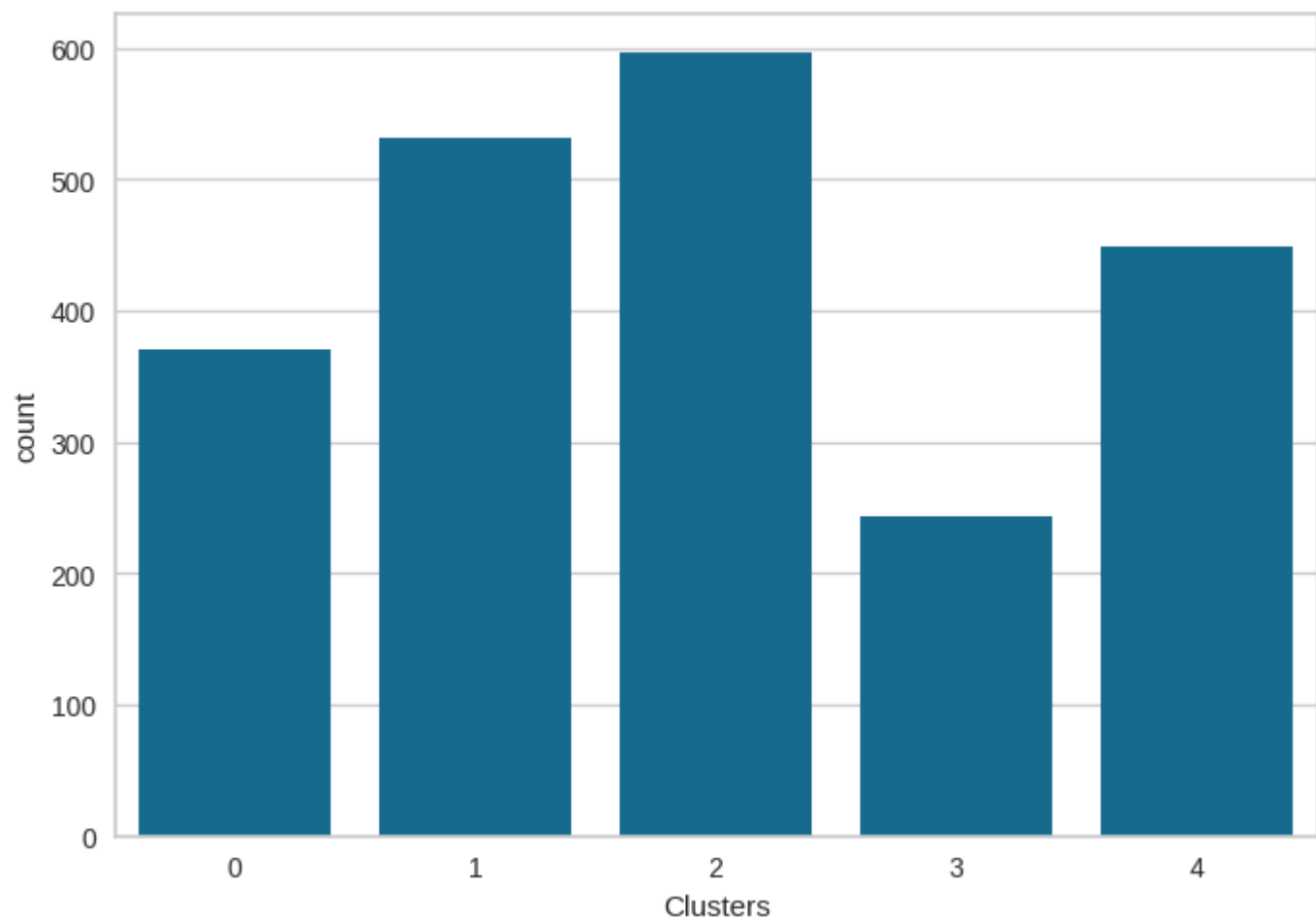
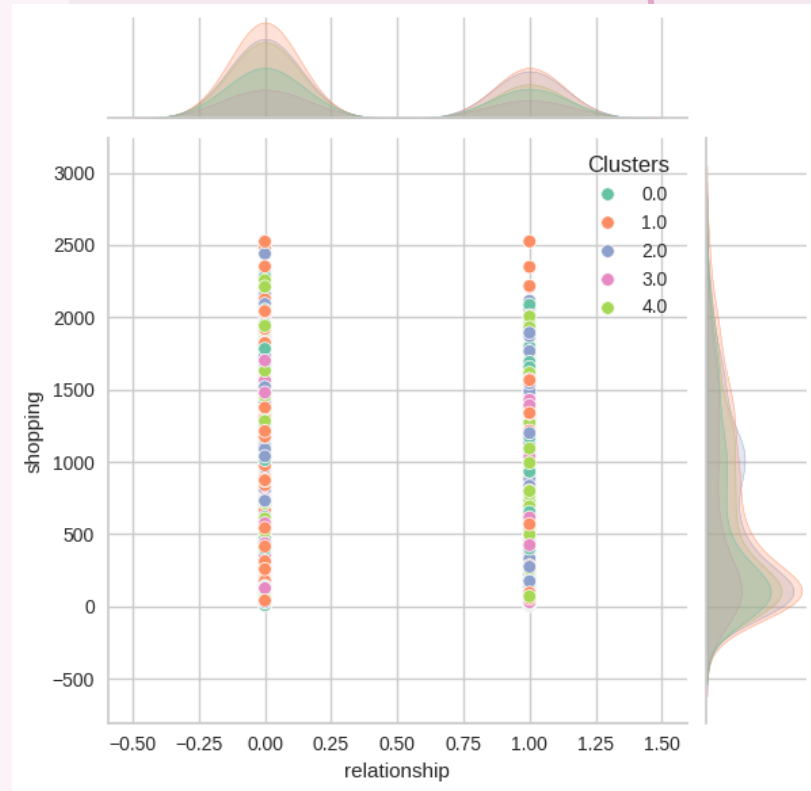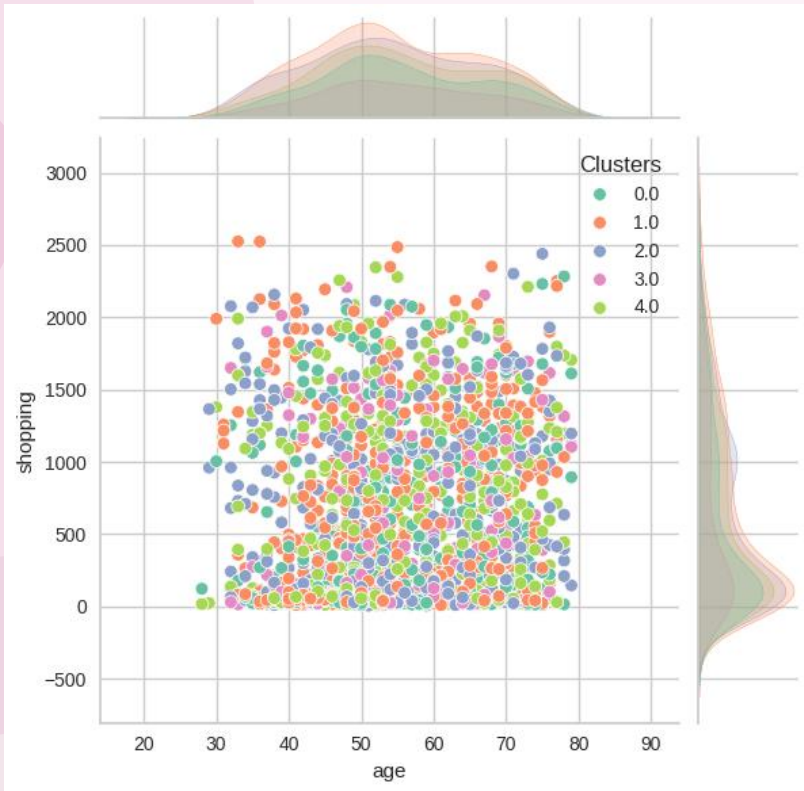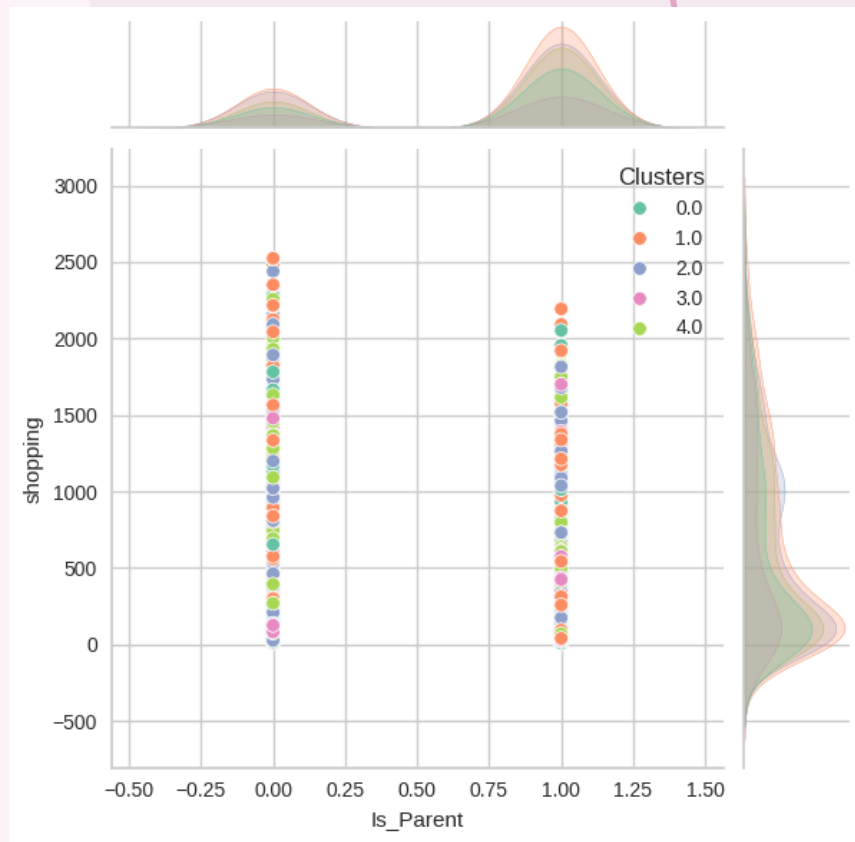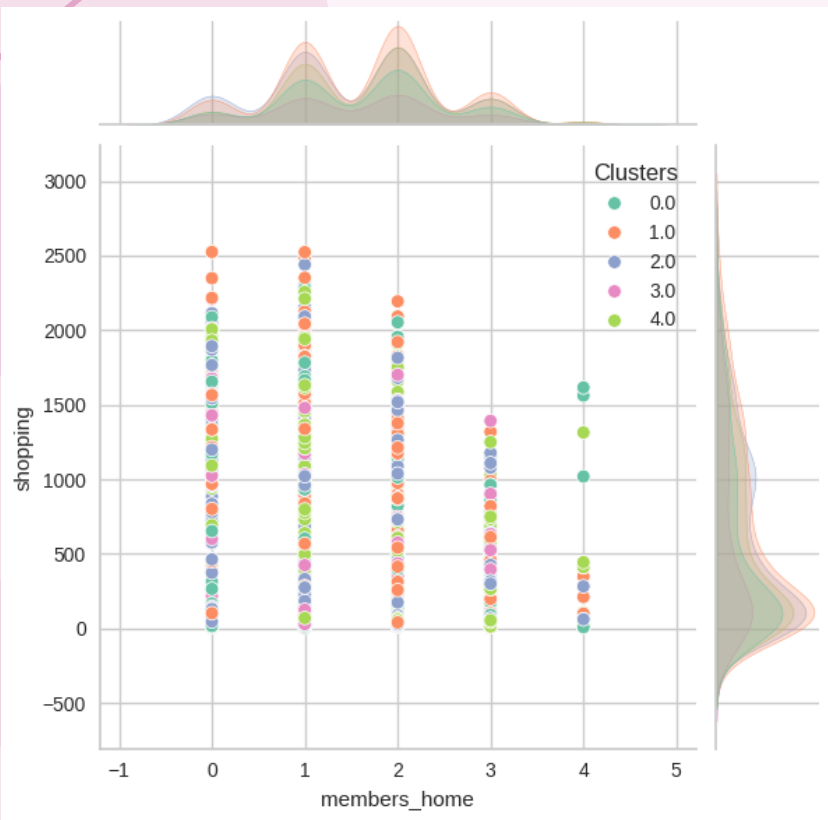shopping capabilities of each cluster
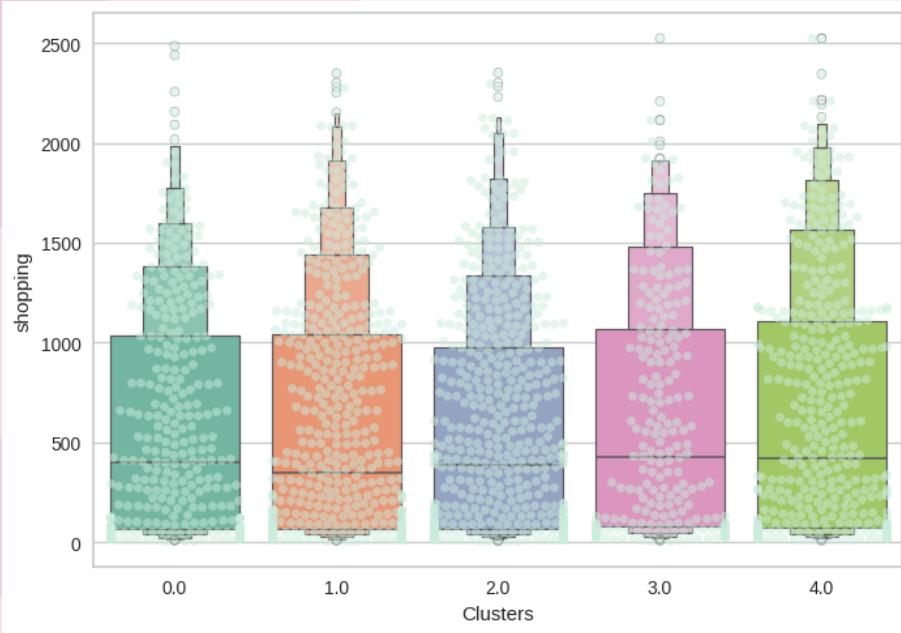
Income each cluster
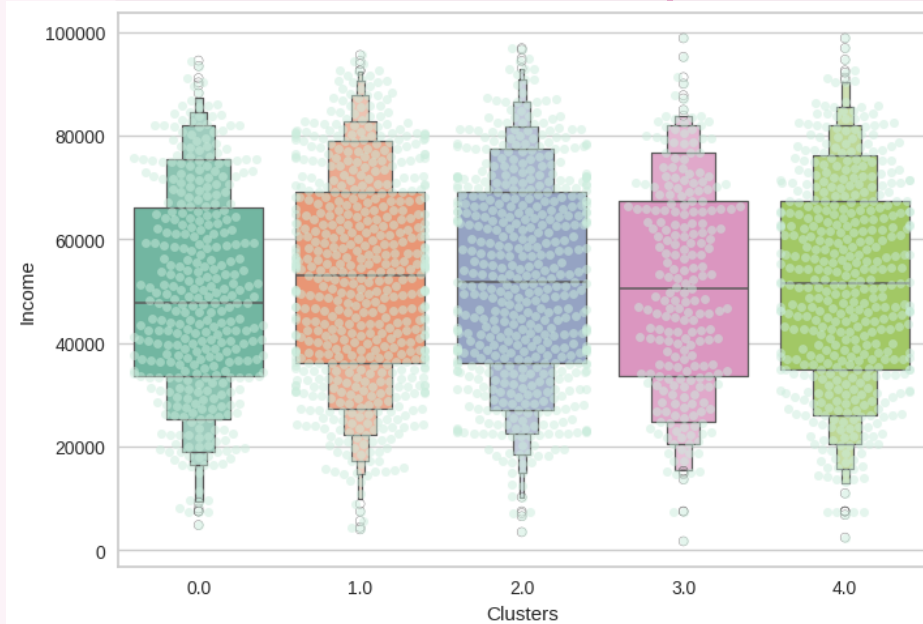
PROFILING
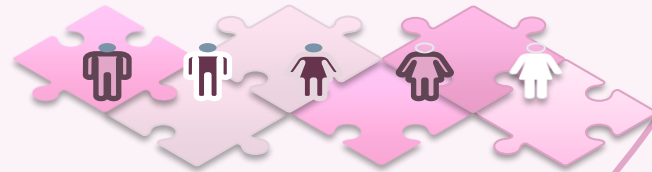
plotting some of the features that are indicative of the
customer's personal traits in light of the cluster they are in.
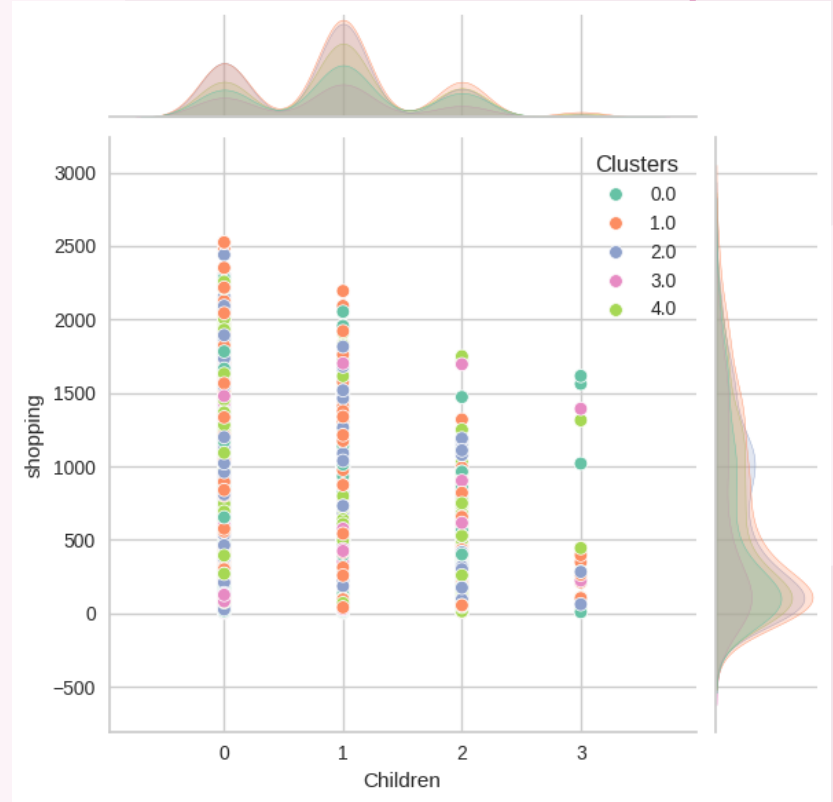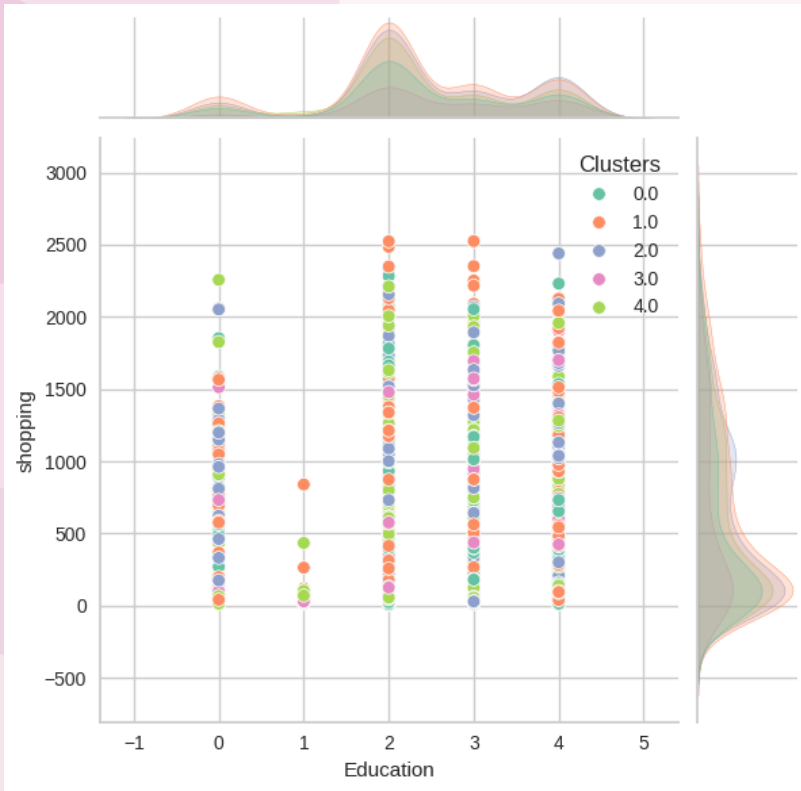
Profiling The Clusters

Cluster 1

✓ age around ‹ 30 – ‹ 80
✓ the majority of these people are single living.
✓ Some have kids and teenagers at home
✓ have 4 members in the family
✓ High income and spent

Cluster 2

✓ age around ›30- ‹ 80, but most in ‹ 70
✓ the majority of these people are in relationship not single living
✓ a highest capability shopping
✓ have 4 members in the family

Cluster 3

✓ age around ›30- ‹80
✓ more are single category
✓ the majority of these people are in relationship not single living
✓ have 4 members and 0 members in the family
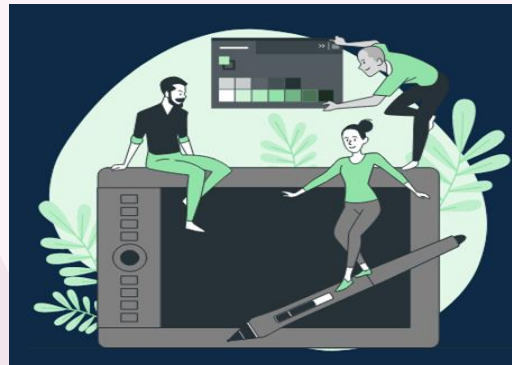✓ Largest number of cluster members

Cluster 4

- ✓ age around <30 - <80
- ✓ more are in relationship
- ✓ have 3 members in the family
- ✓ most have kids and teenager at home
- ✓ smallest number of cluster members

Cluster 5

- ✓ >40 tahun s/d < 80 tahun
- ✓ single and relationship almost the same
- ✓ most have 1 members in the family
- ✓ kid and teenager are not dominant in this group

CONCLUSION & RECOMMENDATION

memanfaatkan cluster 2 untuk memperbanyak penawaran produk, karena dari hasil penjualan terahir, cluster 2 paling tinggi diantara cluster yang lain   dan mayoritas cluster 2 mempunyai anggota keluarga yang banyak

Cluster 3, mempunyai jumlah anggota paling sedikit daripada cluster lain, dari kapabilitas pembelian produk juga paling sedikit, dari segi income juga termasuk rendah dibandingkan dengan cluster lain, ini menjadi perhatian bagi bisnis, untuk dapat meningkatkan penjualan dari pelanggan cluster 3, bisa dengan melakukan promo diskon pada barang yang sering dibeli oleh customer di cluster 3.

THANK YOU

Have any questions ?

[in] www.linkedin.com/in/uswatun-hasanah-20b84a27b

[✉] chasanah35@gmail.com

Source

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

https://slidesgo.com