



Customer Analysis at Supermarket with Clustering

Project produced by:
Mirto Gasparinatou
Iason Loxos-Andreou
Maryna Bilash

WORKFLOW

1. Customer profile

Providing briefing to the client and doing research on the client's precise needs

2. Exploratory data analysis

Explore the dataset, clean the data and check for outliers

3. Clustering

Define the clusters (groups) of customers

Customer Profile

Age

Income

Spend

Recency

Customer
Duration

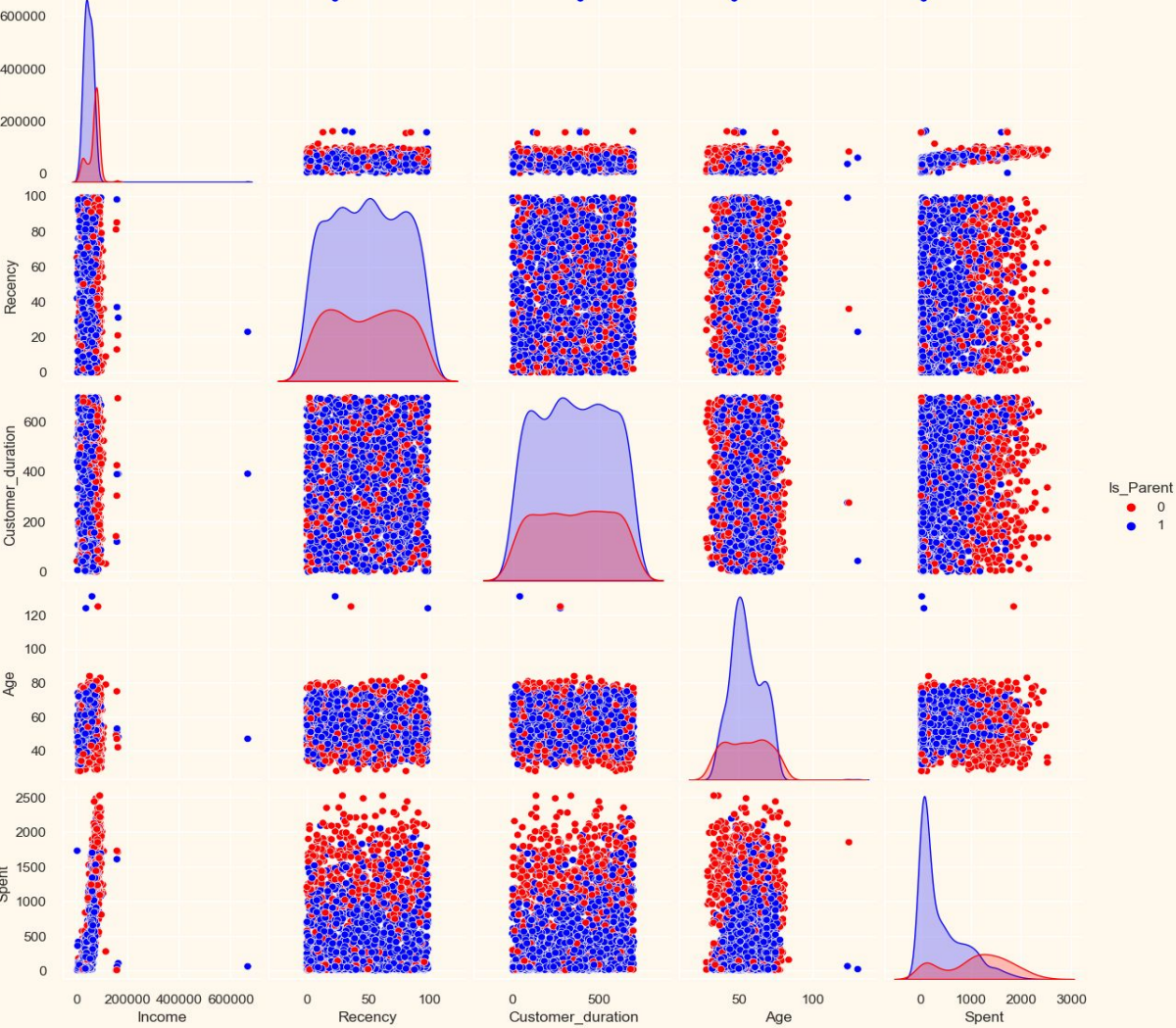
Marital Status



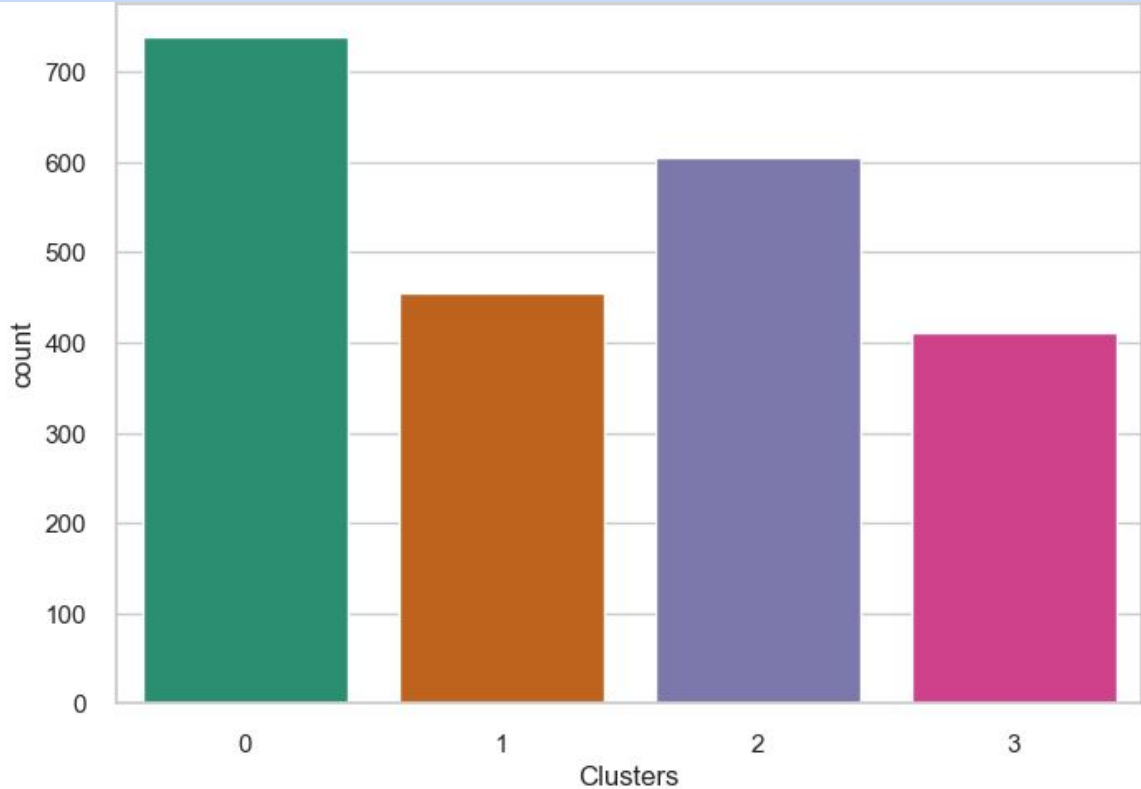
Outliers Check

Clearly, there are a few outliers in the Income and Age features. We will be deleting the outliers in the data.

- High Income and Age value
- Remove Ages > 90 (max age 131 years)
- Remove Income > 600000



Distribution Of the Clusters

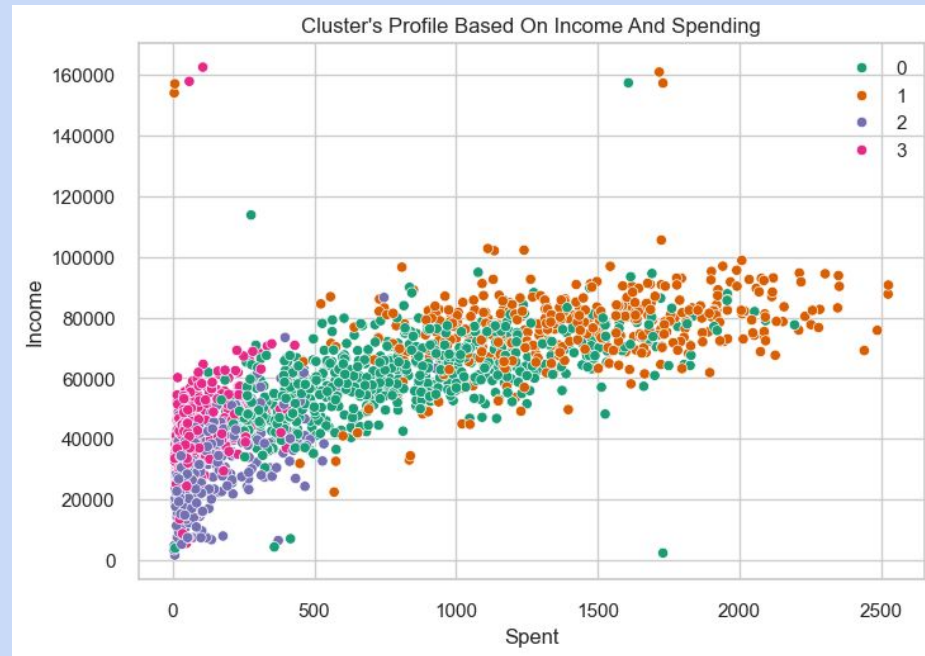


- Cluster 0: high spending & average income
- Cluster 1: high spending & high income
- Cluster 2: low spending & low income
- Cluster 3: high spending & low income

Clusters with deals and promotions



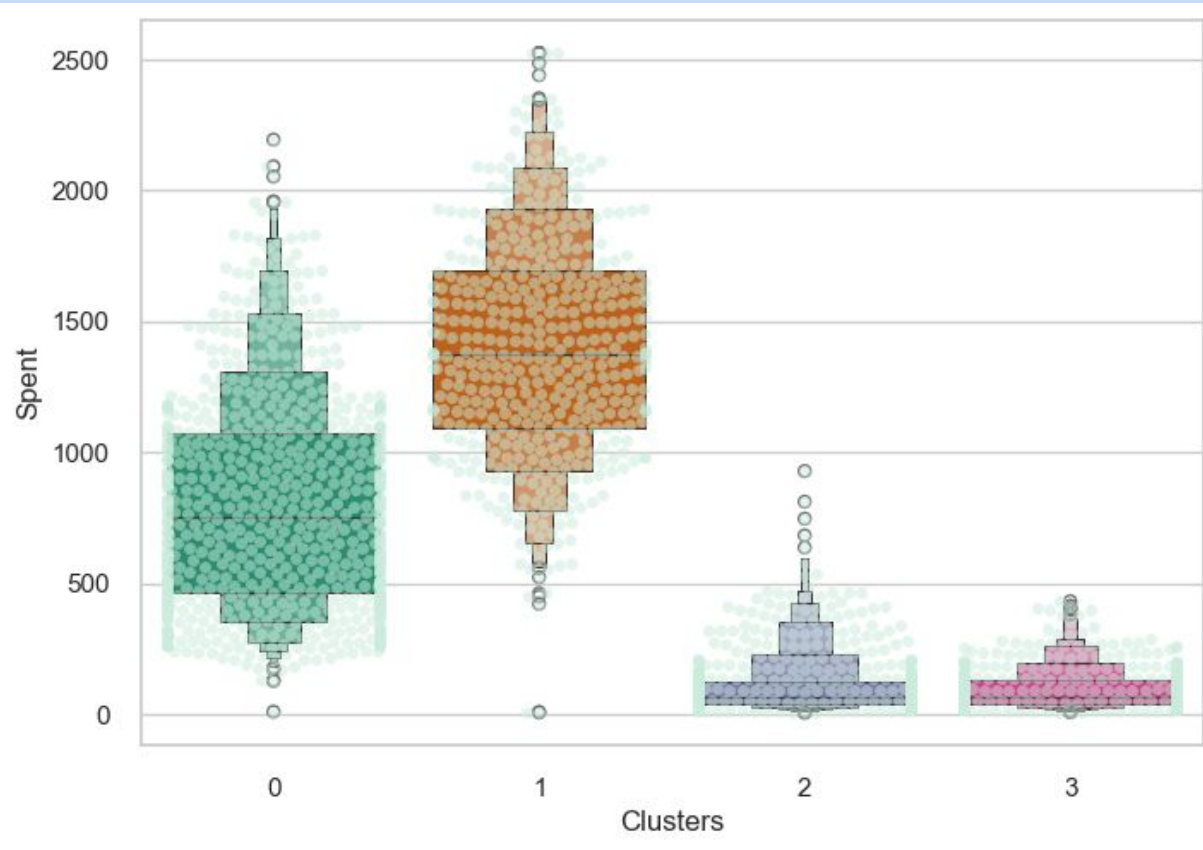
Clusters without deals and promotions





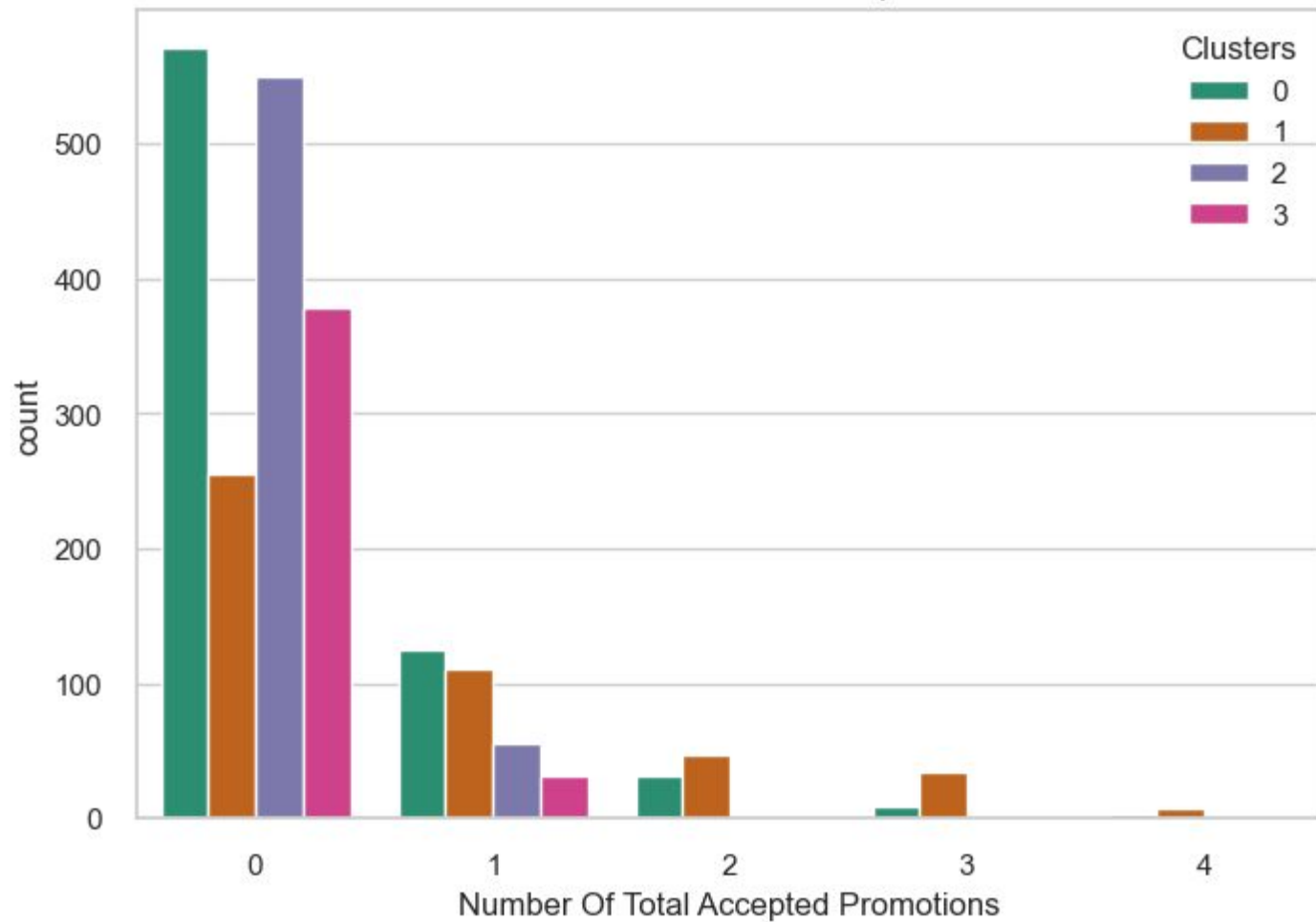
Distribution of Clusters based on every feature

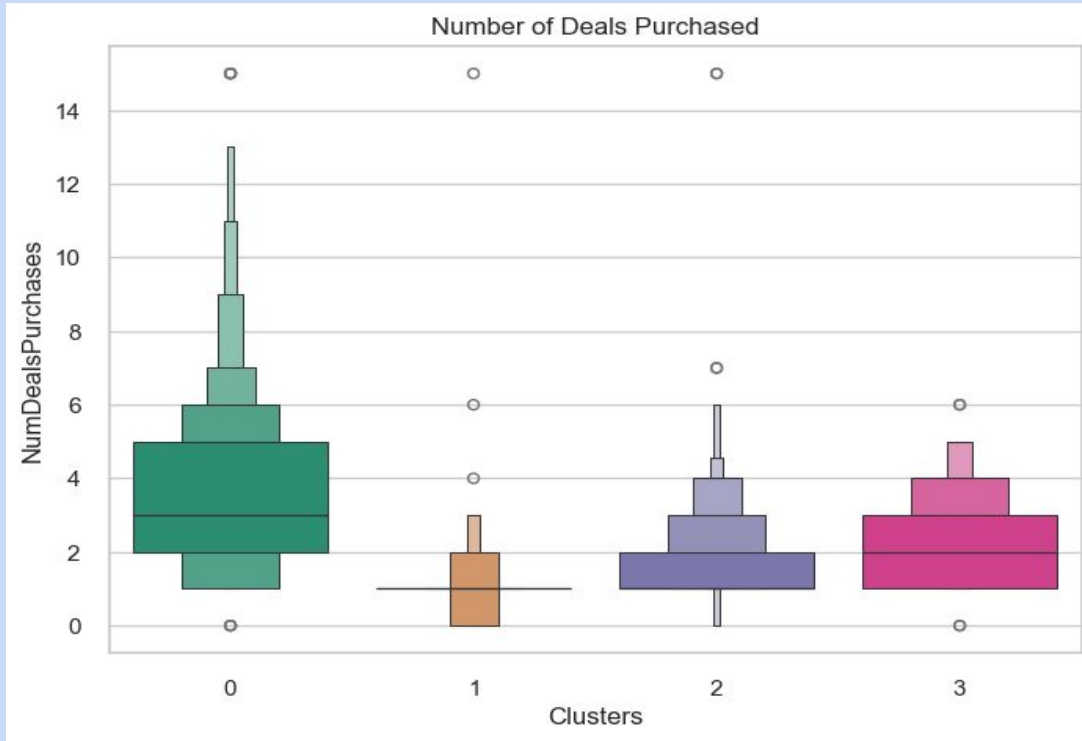
We pay attention on feature that have big variance between clusters



Cluster 1 has the biggest spendings of customers closely followed by cluster 0

Count Of Promotion Accepted





Unlike campaigns, the deals offered did well.

- ❖ Best outcome: clusters 0 & 3.
- ❖ Customers of cluster 1 are not much into deals.
- ❖ Nothing seems to attract cluster 2 overwhelmingly

Summary

Cluster Number 0:

- Not parents (Average 0.3 children)
- Highest Income
- Biggest Spenders

Cluster Number 1:

- Not parent
- Lowest Income
- Low Spenders
- Mean age = 50 (younger of all clusters)
- Max 2 members in the family

Cluster Number 2:

- The majority are parents
- Average less than 1 children
- The majority have one kid (and not teenagers, typically)

Cluster Number 3:

- Parents
- The majority have teenager at home
- Relatively older
- A lower-income group