

EDA AND CUSTOMER SEGMENTATION

CUSTOMER PERSONALITY ANALYSIS

USING K-MEANS

COSTANZA BRUSUTTI

JULY 2022

CUSTOMER SEGMENTATION

UNSUPERVISED CLUSTERING

Helps defining and better understanding the business' target audiences and ideal customers.

To scale efficiently and effectively, companies need to focus their efforts on a specific subset of customers who are most similar to their best current customers, not a broad universe of potential customers.



TABLE OF CONTENT

- 01 PROBLEM STATEMENT
- 02 DATA CLEANING
- 03 FEATURE ENGINEERING
- 04 EDA & PLOTTING
- 05 CLUSTERING
- 06 CONSIDERATIONS



SCOPE OF ANALYSIS

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



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.





TARGET

- Find Customer Segments within the Customer Dataset.
- Within each Customer segment, analyse income, spending and relationship and family status with regards to their overall purchasing trends



INITIAL ANALYSIS

DATASET

Customer's Information

- ID
- Year_Birth
- Education
- Marital_Status
- Income
- Kidhome
- Teenhome
- Dt_Customer
- Recency
- Complain

Products

Amount spent on different products in the last 2 years

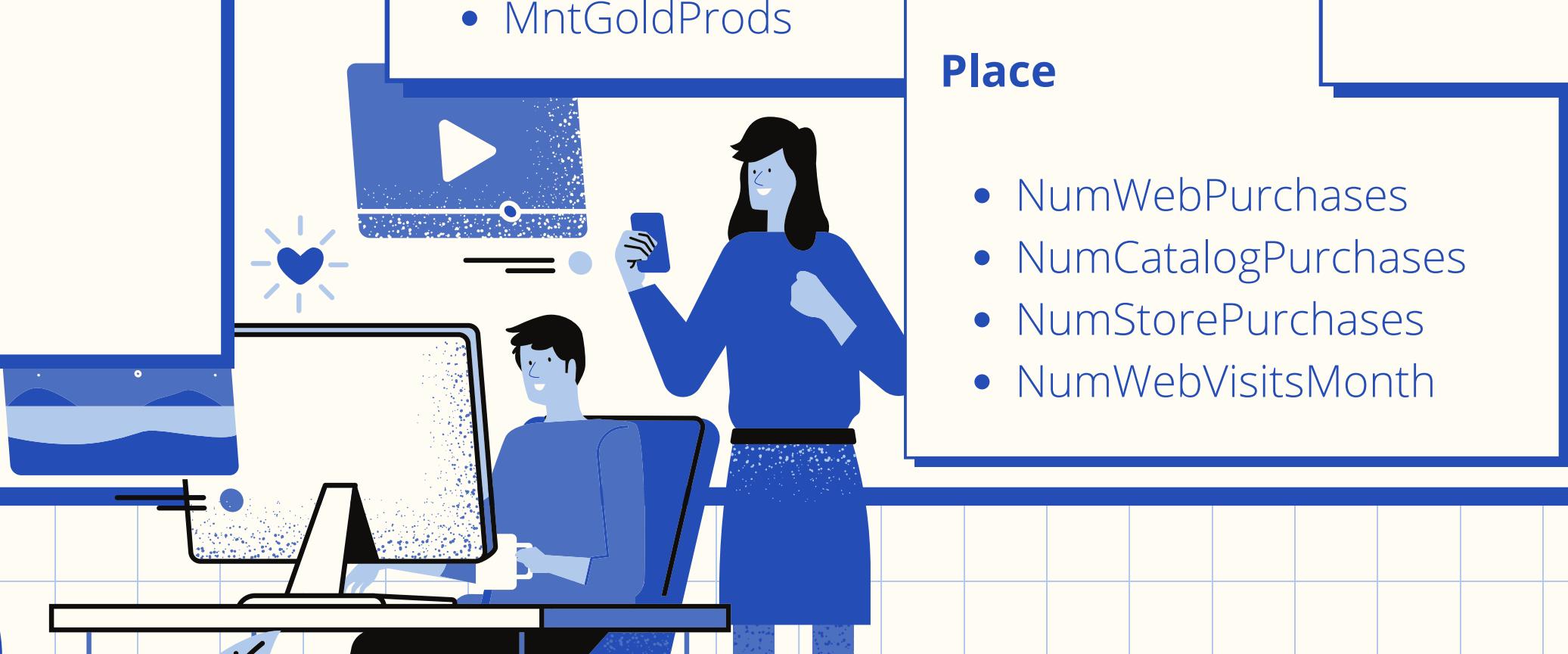
- MntWines
- MntFruits
- MntMeatProducts
- MntFishProducts
- MntSweetProducts
- MntGoldProds

Promotion

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

Place

- NumWebPurchases
- NumCatalogPurchases
- NumStorePurchases
- NumWebVisitsMonth



2.3.1 Adjust Features

```
#look into a column's unique values  
df[ "Marital_Status" ].value_counts()
```

Married	864
Together	580
Single	480

click to scroll output

Widow	
Alone	
Absurd	
YOLO	
Name: Ma	

DATA CLEANING

- Converting datetime
- NaN, dropping & filling with mean
- Feature Transformation
- Ordinal Encoding
- Standardization

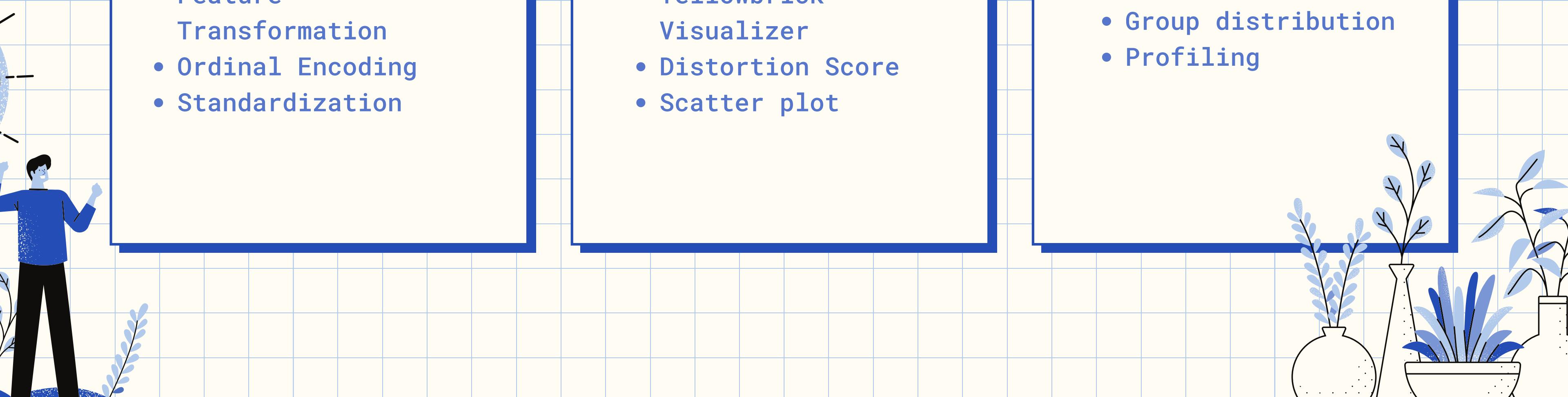
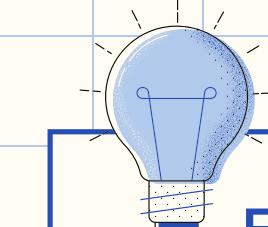
PROCESSING FLOW

CLUSTERING

- K-Means
- Elbow Method
- Silhouette
- Yellowbrick Visualizer
- Distortion Score
- Scatter plot

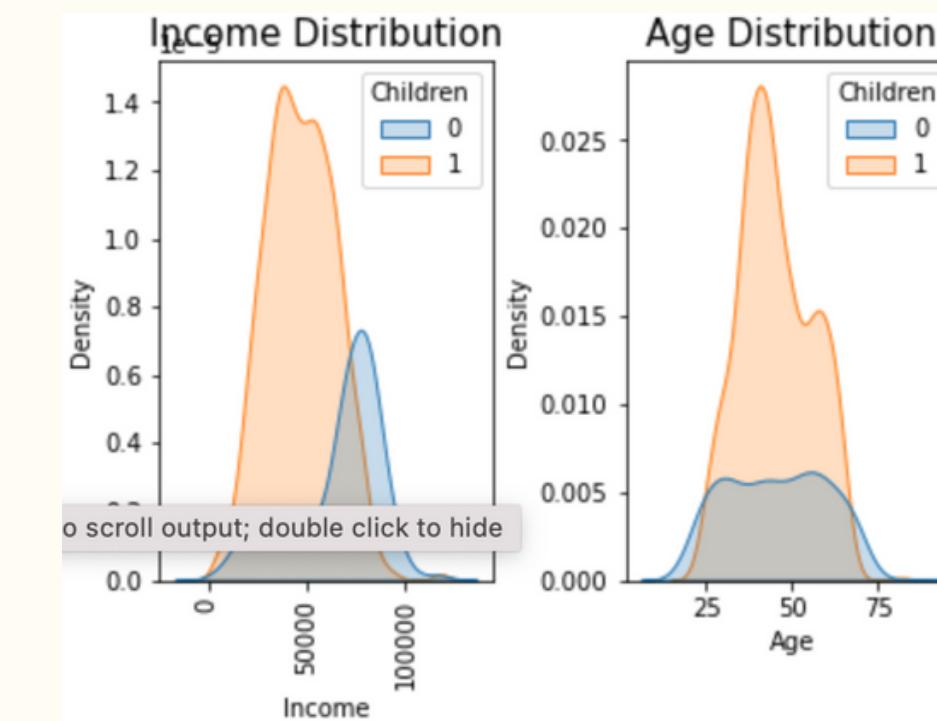
EVALUATING

- EDA after clustering
- Recognising patterns
- Group distribution
- Profiling

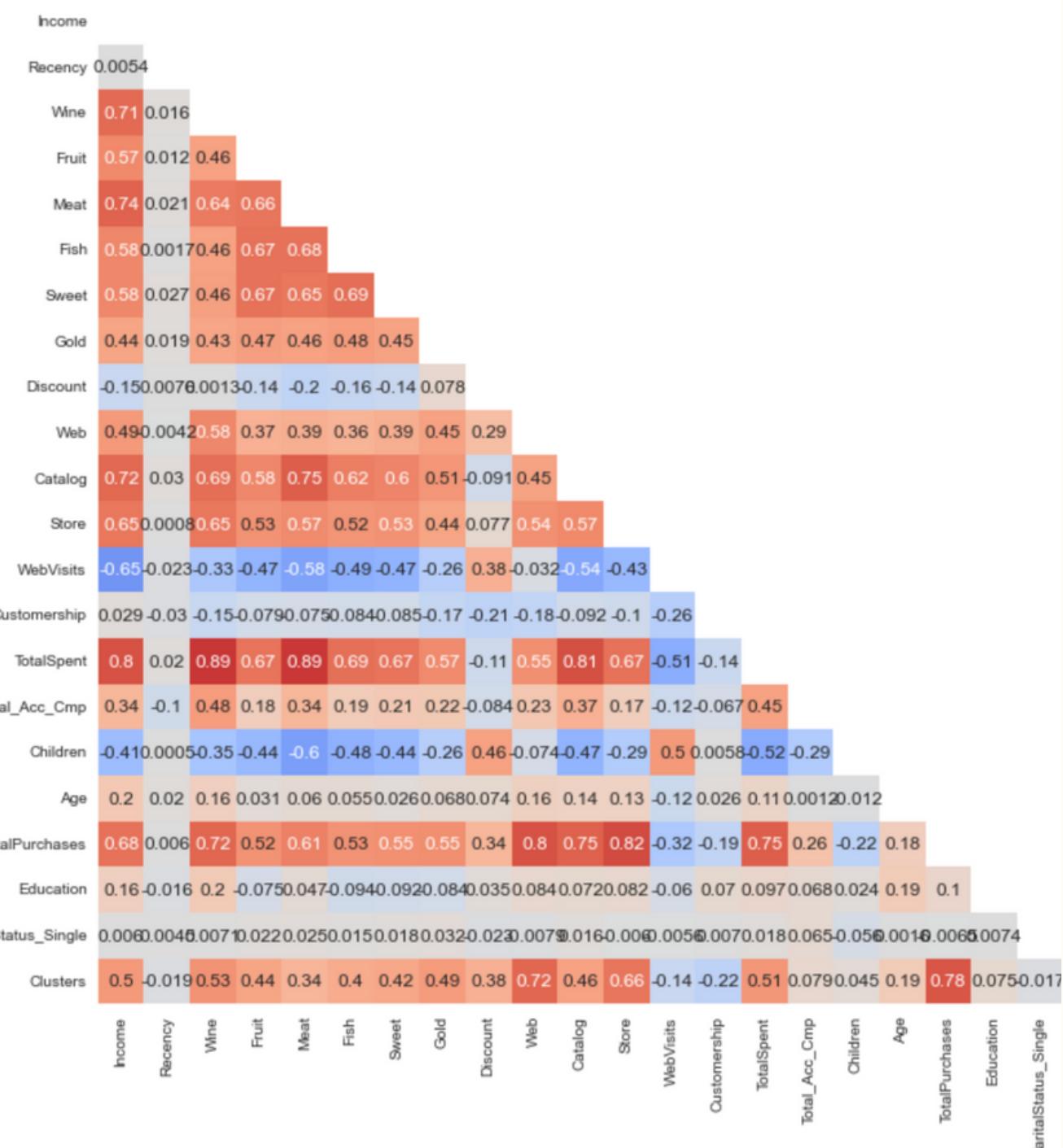


EDA BEFORE CLUSTERING

Because they have significant role in spending I will check age and income columns

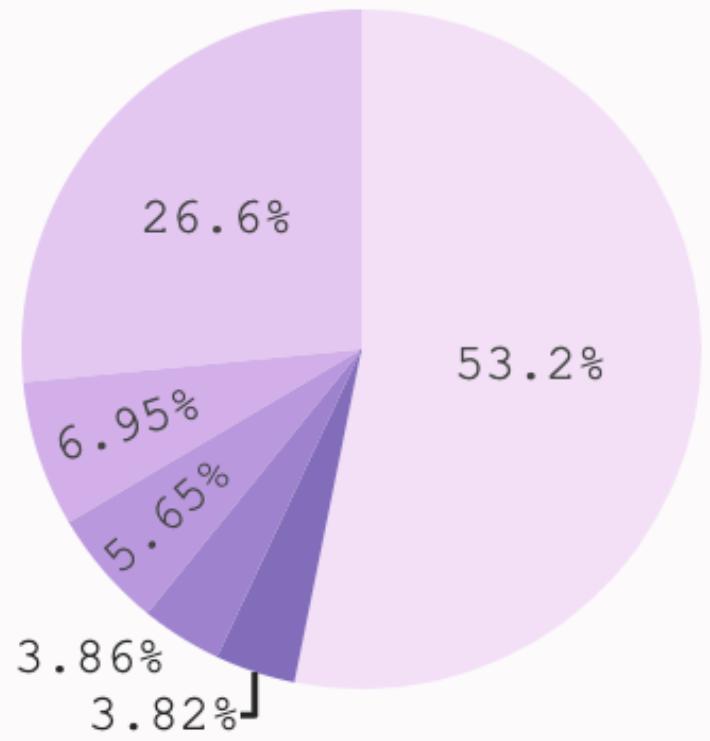


- Most of our customers are 35+
- People with children have higher income



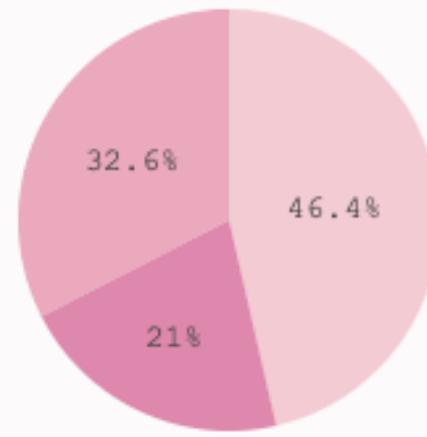
SPENDING AMONG DIFFERENT CATEGORIES

Sales by Category



Wine has the highest percentage of sales (almost half), followed by meat products, and then the percentages of sales for the rest of the products are almost the same.

Ways of Purchases

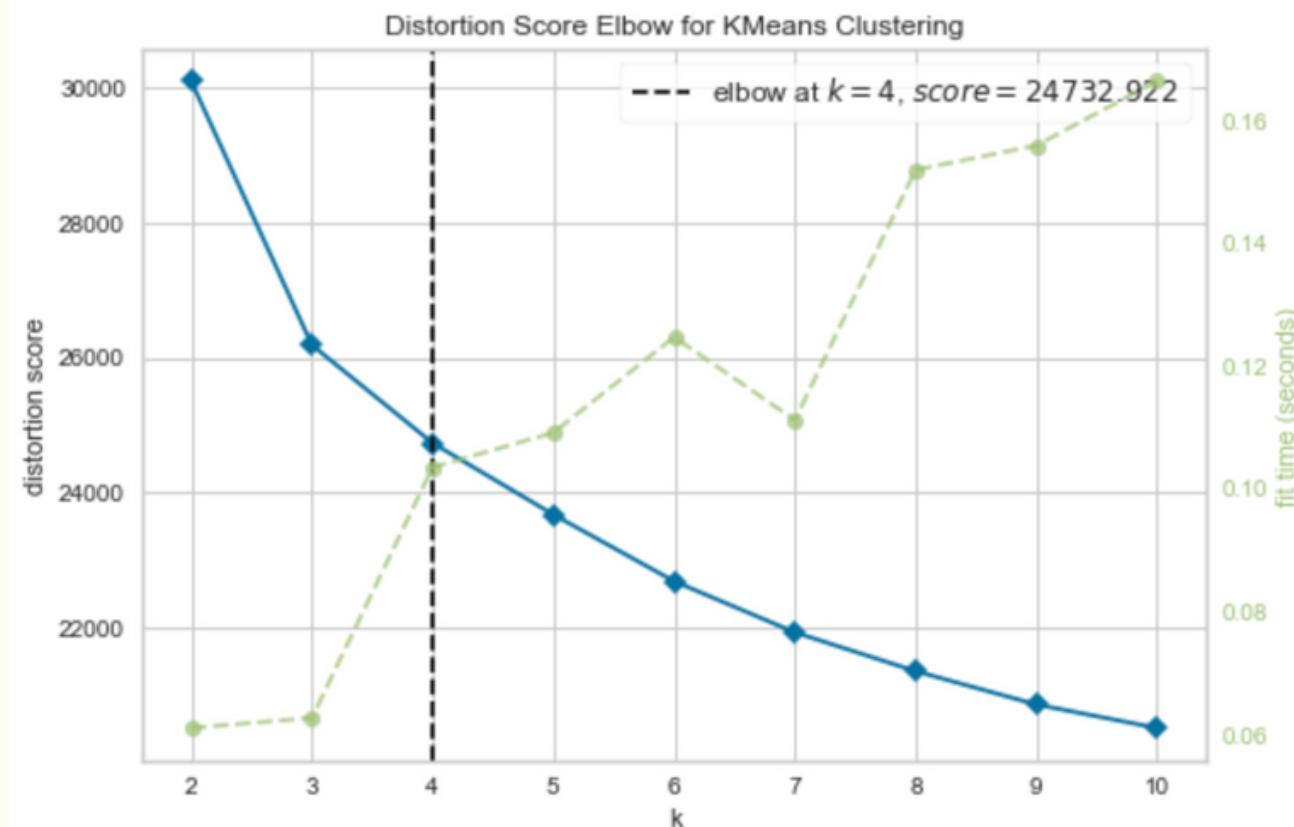


Most of our sales come from our stores, so we need to work on a marketing strategy for our web store

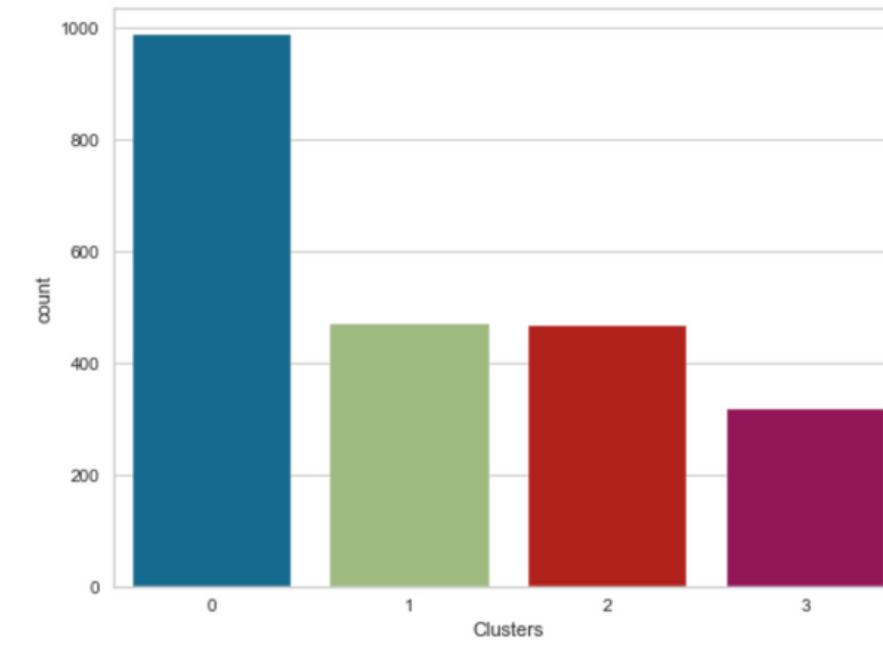


CLUSTERING

K-MEANS

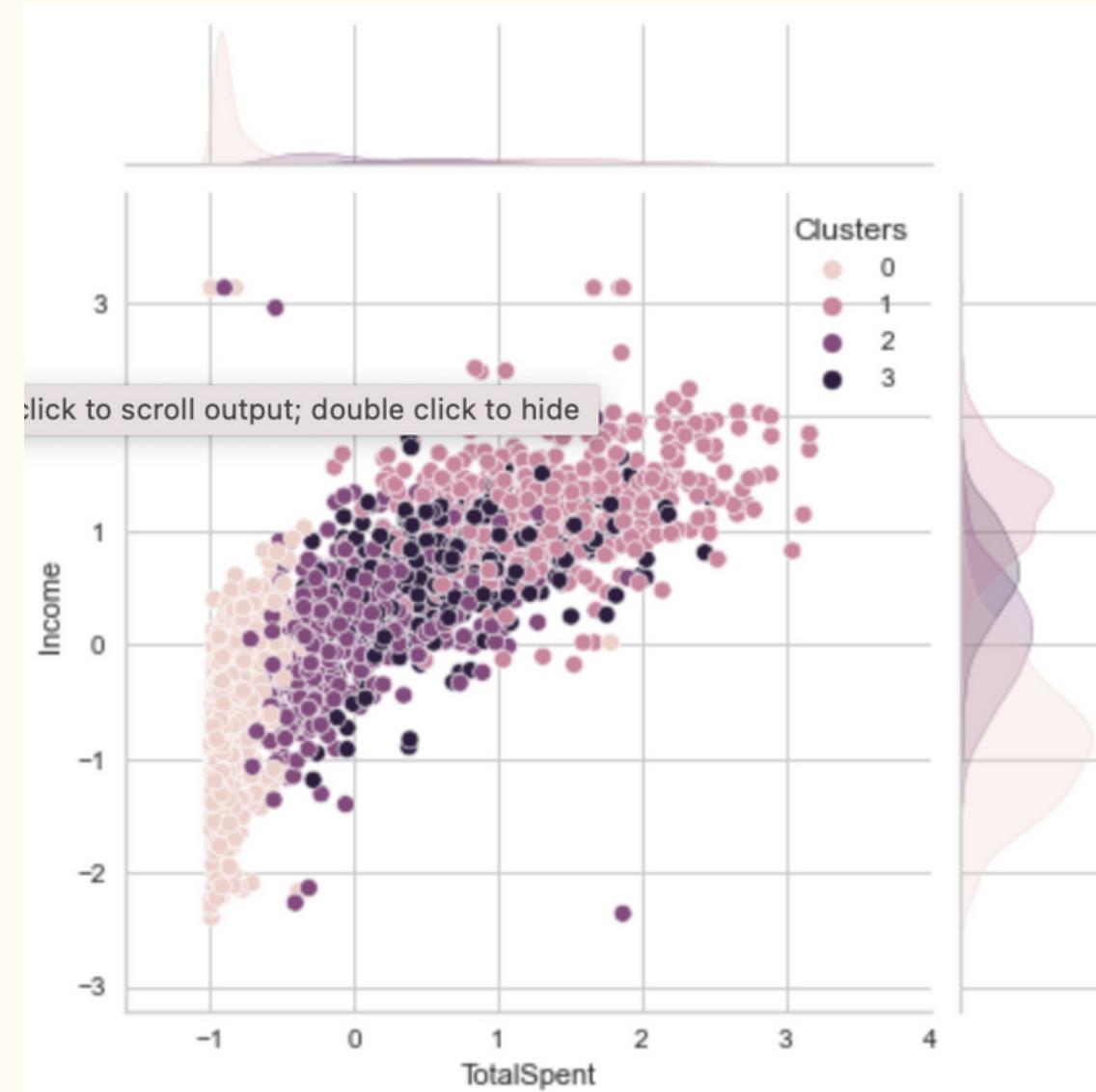


- Distance based algorithm
- Scaled data
- Elbow Method and Silhouette Score for validation



- Distance based algorithm
- Scaled data
- Elbow Method and Silhouette Score for validation

Cluster's profile based
on Income and Spending



Income vs spending plot shows the clusters pattern

group 0: high spending & high income

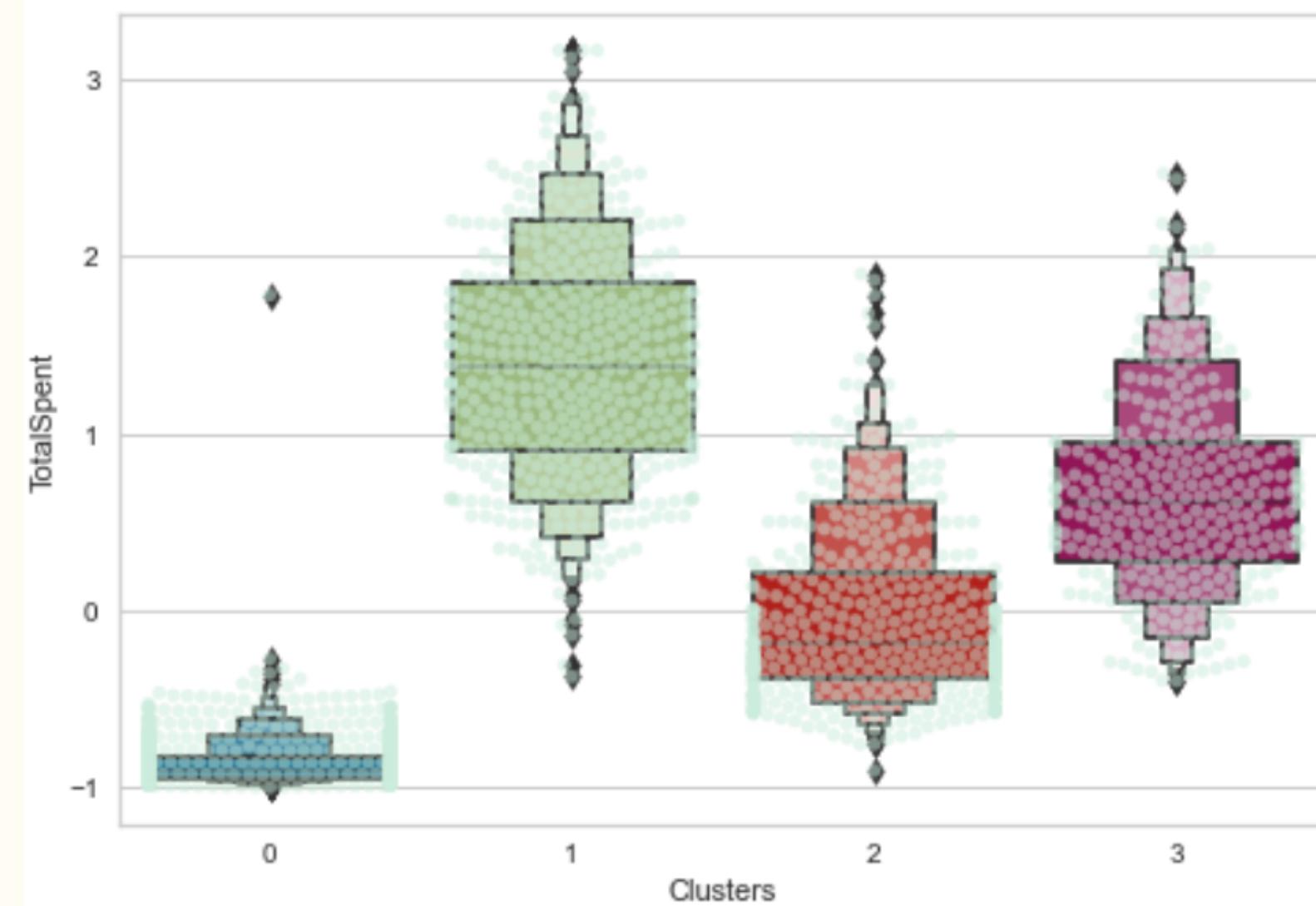
group 1: high spending & average income

group 2: low spending & low income

group 3: high spending & low income

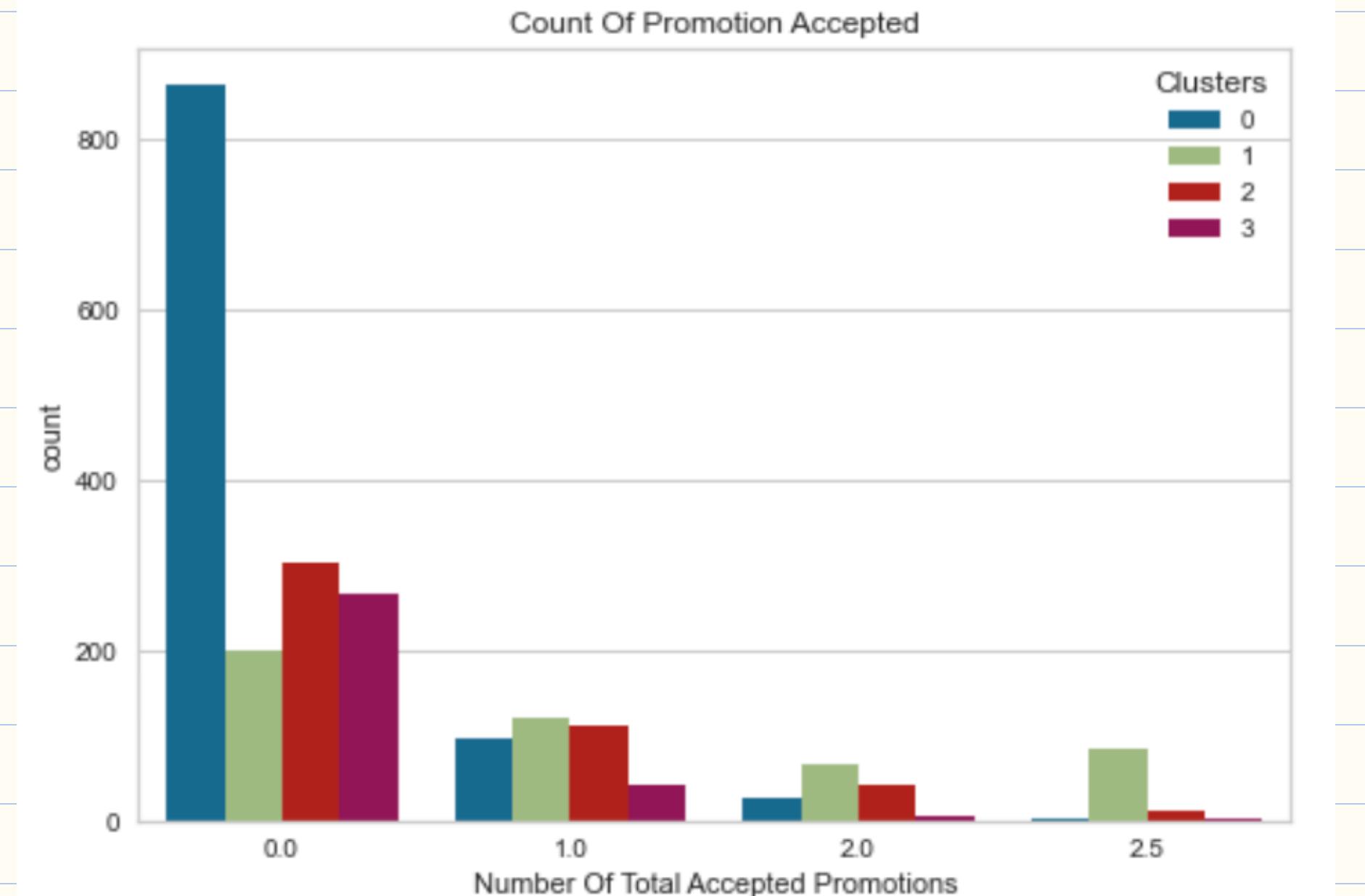


BOX+SWARMPLOT



- Cluster 1 is our biggest set of customers followed by 3
- What each cluster is spending on
- Targeted marketing strategies

How did the business' campaigns do in the past?

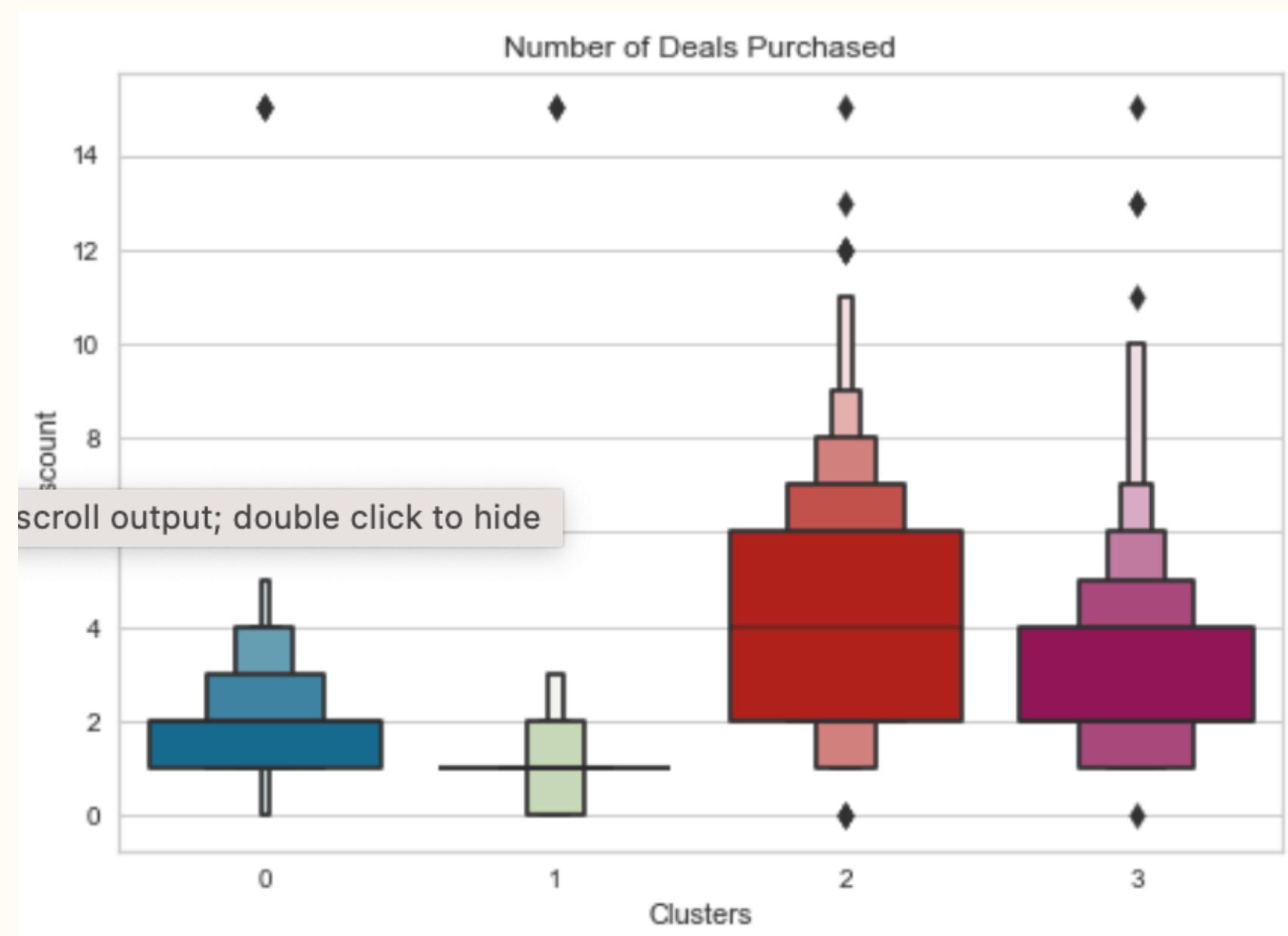


**THERE HAS NOT BEEN AN
OVERWHELMING RESPONSE TO THE
CAMPAIGNS SO FAR.**

**VERY FEW PARTICIPANTS OVERALL.
MOREOVER, NO ONE PART TAKE IN ALL
5 OF THEM.**

Perhaps better-targeted and well-planned campaigns are required to boost sales.

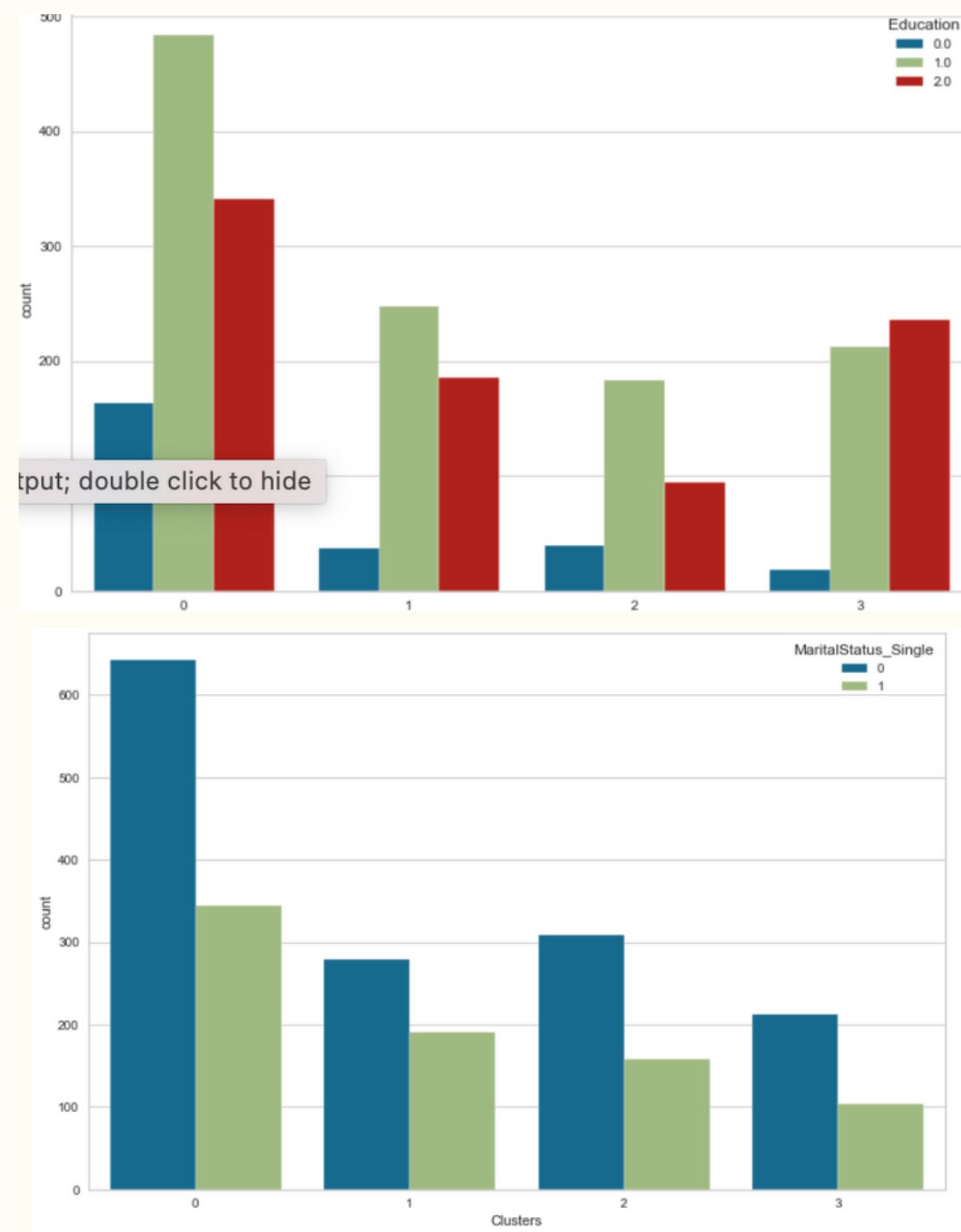
PLOTTING THE NUMBER OF DEALS PURCHASED



- Unlike campaigns, the deals offered did well.
- It has best outcome with cluster 2 and cluster 3.
- However, our star customers cluster 1 are not much into the deals.
- Nothing seems to attract cluster 0 overwhelmingly

PROFILING

- Looking at purchasing habits
- Who is our star customer
- Who needs more attention
- Plot the features that are indicative of the customer's personal traits in light of the cluster they are in



PROFILING

CLUSTER NO. 0

- Most number of customers
- Most likely a parent
- Relatively older
- Mostly graduates
- High expenditure
- Positive reply to campaigns
- Not into deals and discounts

CLUSTER NO. 1

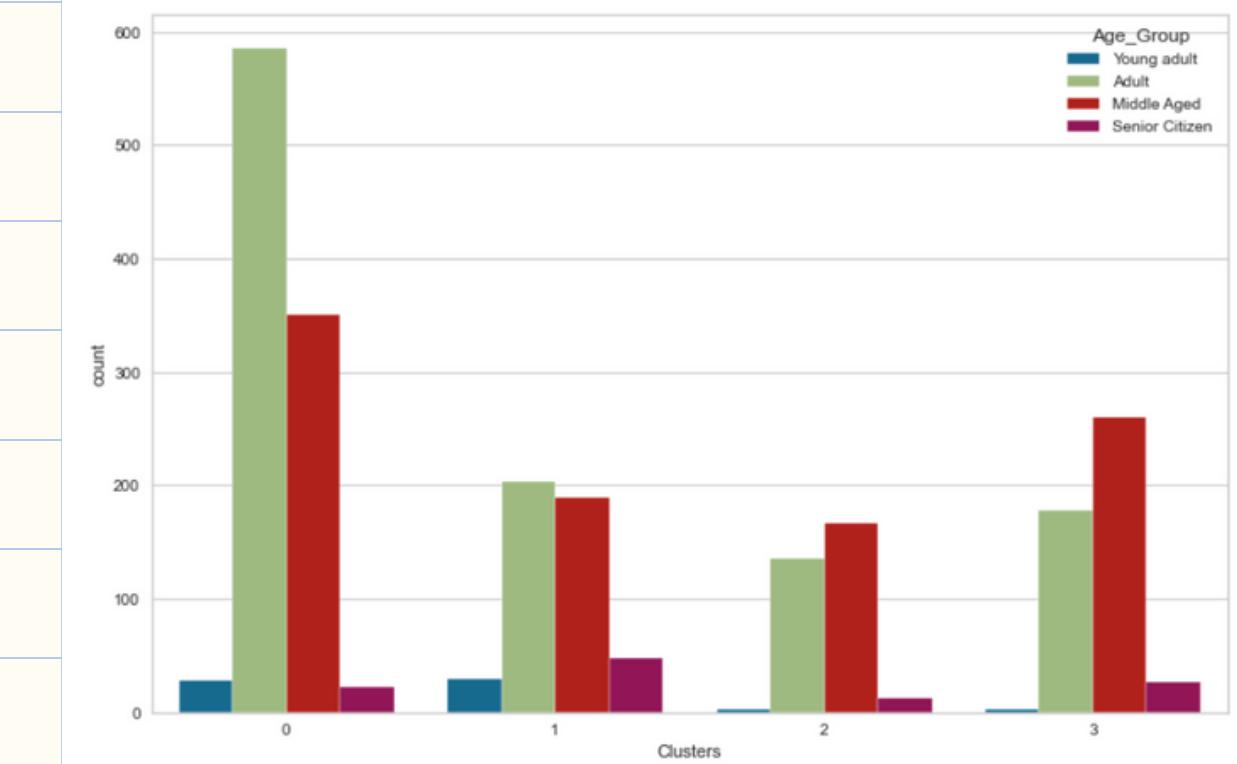
- Most likely not a parent
- Slight majority of couples over single
- Span all ages
- High income group
- Have accepted few campaigns and deals

CLUSTER NO.3

- Least number of customers
- Parents
- Relatively older
- Low income group
- Low expenditure

CLUSTER NO. 2

- Majority are parents
- Majority have children
- Second highest Income
- Relatively younger
- Above average expenditure





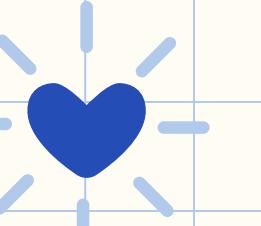
CONCLUSION

Unsupervised clustering to segment customer

4 Clusters used to profile customers according to their family structures and income/spending

Created the base to be used in planning more effective marketing strategies





THANK YOU

in linkedin.com/in/costanza-brusutti