Predict Customer Personality to boost marketing campaign by using Machine Learning

Rakamin
Academy

Created by:
Lana Fuadi
lanafuadi@gmail.com
linkedin.com/lanafuadi

"I am a results-driven data scientist with a passion for leveraging data to drive business success. Through rigorous training at Rakamin Bootcamp, I have honed my skills in Python, ML, SQL, and Business Intelligence. I am committed to continuous learning and have completed additional certifications and course to further enhance my expertise. With strong collaboration, critical thinking, and problem solving skills, I am confident in my ability to help your company make data-driven decisions that propel your business to new heights."

Overview

"Sebuah perusahaan dapat berkembang dengan pesat saat mengetahui perilaku customer personality nya, sehingga dapat memberikan layanan serta manfaat lebih baik kepada customers yang berpotensi menjadi loyal customers. Dengan mengolah data historical marketing campaign guna menaikkan performa dan menyasar customers yang tepat agar dapat bertransaksi di platform perusahaan, dari insight data tersebut fokus kita adalah membuat sebuah model prediksi kluster sehingga memudahkan perusahaan dalam membuat keputusan "
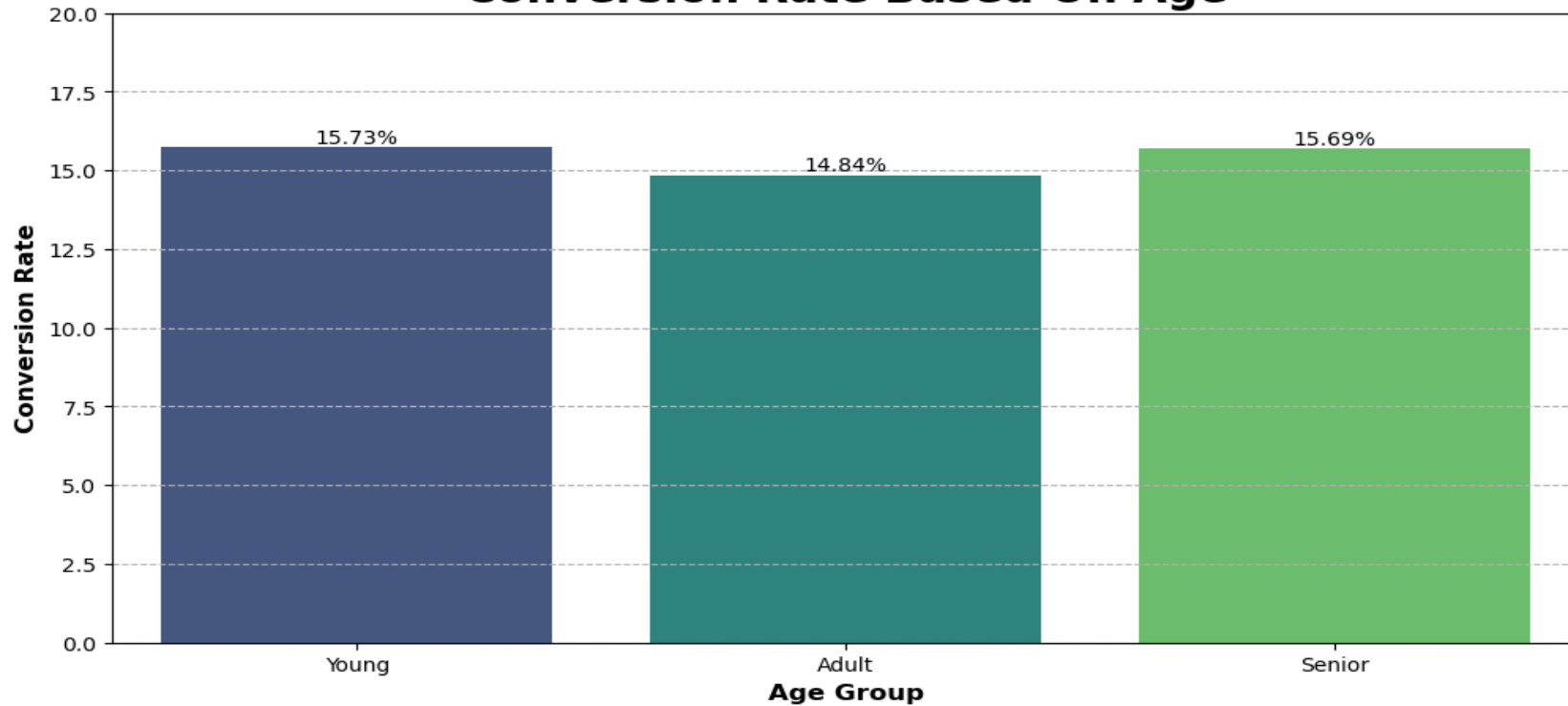
EXPLORATORY DATA ANALYSIS

Rakamin Academy

Simplilearn

Conversion Rate Based On Age

Based on the analysis, all group have relative low conversion rates. it indicates that the existing strategies are not that impactful. this is either because of misaligned marketing or bad website experience. to solve it, we should:
1. improve UI UX design and try to implementing A/B testing to ensure a positive website experience.
2. retargeting strategies to ensure that that is align with customer preference.
3. young is a potential customer, try to add gamification purchase.

For more detail, access jupyter notebook here

Campaign Success Rate

Conversion Rate (y-axis)
Age Group (x-axis)

Age Group	Conversion Rate
Young	54.55%
Adult	40.21%
Senior	38.81%
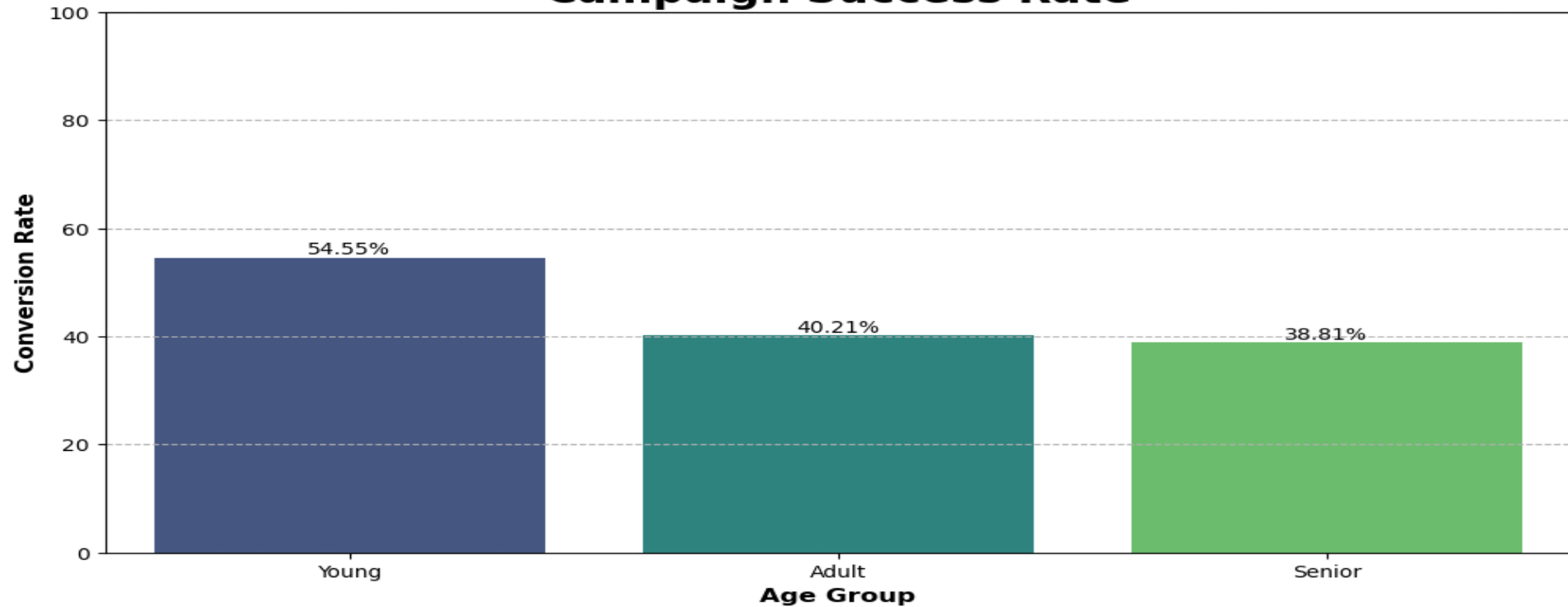
Young group are more likely to make a purchase when receiving a campaign. To maximize campaign success, it is crucial to focus on targeting and tailoring marketing efforts towards young individuals. This could involve creating appealing and engaging campaigns that resonate with their preferences and interests. Both Adult and Senior group has similar rates, it is important to evaluate the existing strategies and identify potential improvement.

	Missing Values	Missing Percentage
Income	24	1.071429
Unnamed: 0	0	0.000000
NumDealsPurchases	0	0.000000
Z_Revenue	0	0.000000
Z_CostContact	0	0.000000

- **There are 24 missing values in column Income.**
- **It is considerable very few and insignificant since it is only around 1% of data.**
- **To handle it we can drop it or do imputation.**
- **For the sake of assignment, we will do imputation using mean fillna.**

- **There is no any duplicated value in our dataset.**
- **Then we will not do anything to handle it.**
- **Even after we do mean imputation, we do not find any duplicate.**

	Duplicated Values	Duplicated Percentage
Stats	0	0.0

- We extract new feature from similar features and drop them to reduce dimension.
- Age is from Dt_Customer – Year_Birth.
- Children is from Teenhome + Kidhome.
- AcceptedCampaign is from AcceptedCmp1 + AcceptedCmp2 + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5.
- We did label encoding on Education because it is considered as ordinal.
- We did one-hot encoding on Marital_Status because it is considered as nominal.

```python
data['Dt_Customer'] = data['Dt_Customer'].apply(lambda x : x.split('-')[-1])
data['Age']         = data['Dt_Customer'].astype(int) - data['Year_Birth']
```
✓ 0.0s

```python
data['Children']         = data['Kidhome'] + data['Teenhome']
data['AcceptedCampaign'] = data['AcceptedCmp1'] + data['AcceptedCmp2'] + data['AcceptedCmp3'] + data['AcceptedCmp4'] + data['AcceptedCmp5']
```
✓ 0.0s

```python
# Perform label encoding
data['Education'] = data['Education'].replace({'SMA':0, 'D3': 1, 'S1': 2, 'S2': 3, 'S3': 4})
# Perform one-hot encoding
data              = pd.get_dummies(data, columns=['Marital_Status'], prefix='Status')
```
✓ 0.0s

FEATURE STANDARDIZATION

```python
from sklearn.preprocessing import StandardScaler

standardize_features = ['Income', 'Recency', 'MntCoke', 'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
                        'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases', 'NumCatalogPurchases', 'NumStorePurchases']
scaler = StandardScaler()

data[standardize_features] = scaler.fit_transform(data[standardize_features])
```
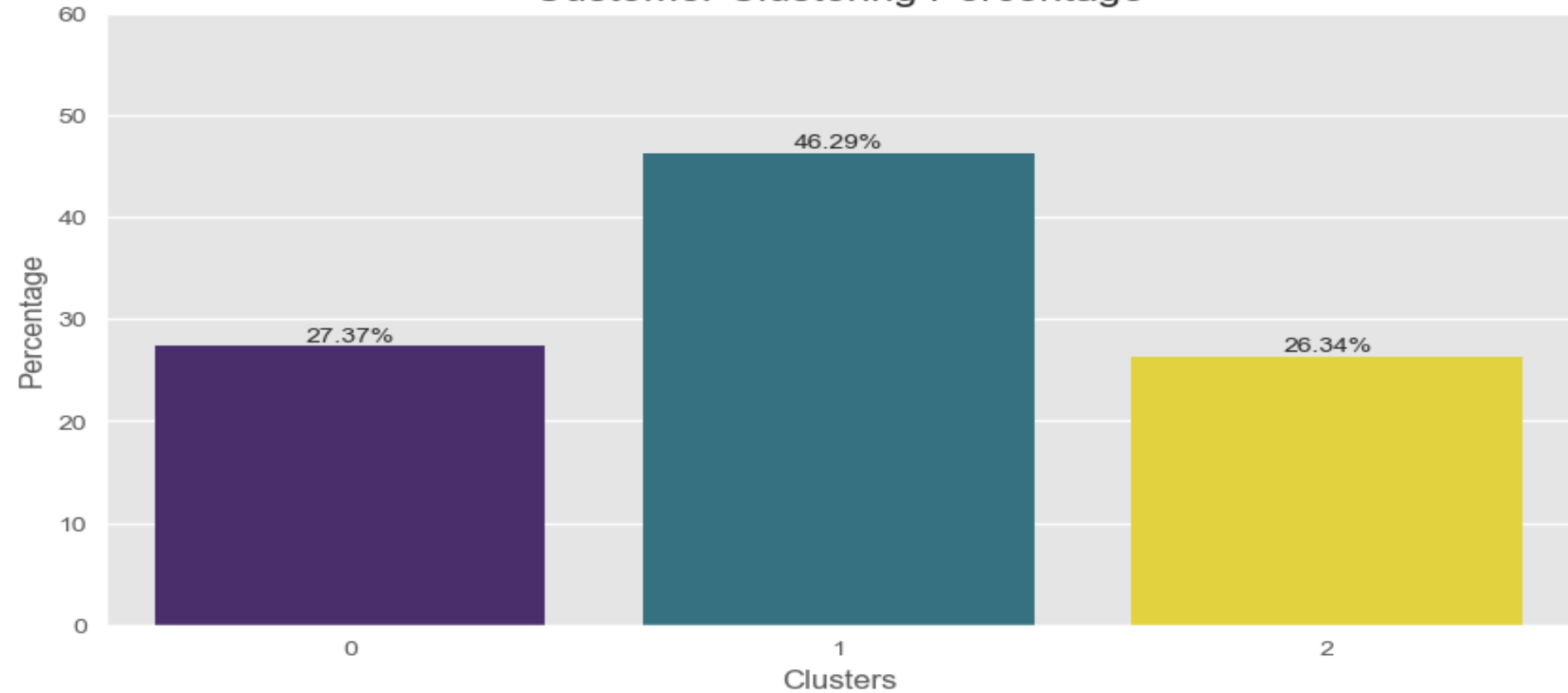✓ 0.1s

- We use StandarScaler to transform numerical features in a dataset to a common scale.
- We use it due to its simplicity and effectiveness in many scenarios.

INTERPRETATION
AND
BUSINESS
RECOMMENDATION

Customer Clustering Percentage

- **Almost half of our customer are cluster 1 (low customers).**
- **Followed by cluster 0 (medium customers) and cluster 2 (high customers).**

Clusters Insight

1. Cluster 0:
- Income 58,275,000/Year
- Spend 589,000
- Made 7 times store purchases
- Made 6 times web purchases
- Made 3 times catalog and deals purchases
- Age around 47

2. Cluster 1:
- Income 34,704,000/Year
- Spend 52,000
- Made 3 times store purchases
- Made 2 times web and deals purchases
- Never made catalog purchase
- Age around 41

3. Cluster 2:
- Income 75,774,000/Year
- Spend 1,236,500
- Made 8 times store purchases
- Made 6 times catalog purchases
- Made 5 times web purchases
- Only made 1 deals purchase
- Age around 44

| clusters		0	1	2
Income	median	58275000.0	34704000.0	75774000.0
MntCoke	median	395000.0	22000.0	546000.0
MntFruits	median	12000.0	2000.0	61000.0
MntMeatProducts	median	113000.0	14000.0	417000.0
MntFishProducts	median	16000.0	3000.0	93000.0
MntSweetProducts	median	12000.0	2000.0	61000.0
MntGoldProds	median	41000.0	9000.0	58500.0
NumDealsPurchases	median	3.0	2.0	1.0
NumWebPurchases	median	6.0	2.0	5.0
NumCatalogPurchases	median	3.0	0.0	6.0
NumStorePurchases	median	7.0	3.0	8.0
Age	median	47.0	41.0	44.0

Parallel Coordinates

- **Cluster 2 has the most spending for various products.**
- **Catalog purchase is more popular in cluster 2.**
- **While deals and web purchase are more popular in cluster 0.**
- **There is no any significant trend among cluster 1.**
- **Age distribution is considerable balance, means has no correlation with clustering behaviour.**
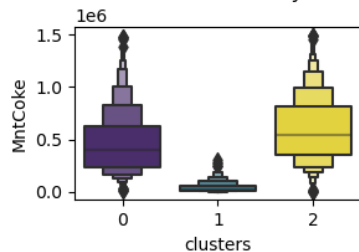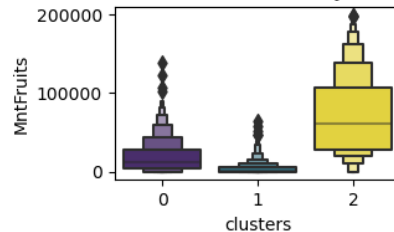

Clusters Parallel Coordinates
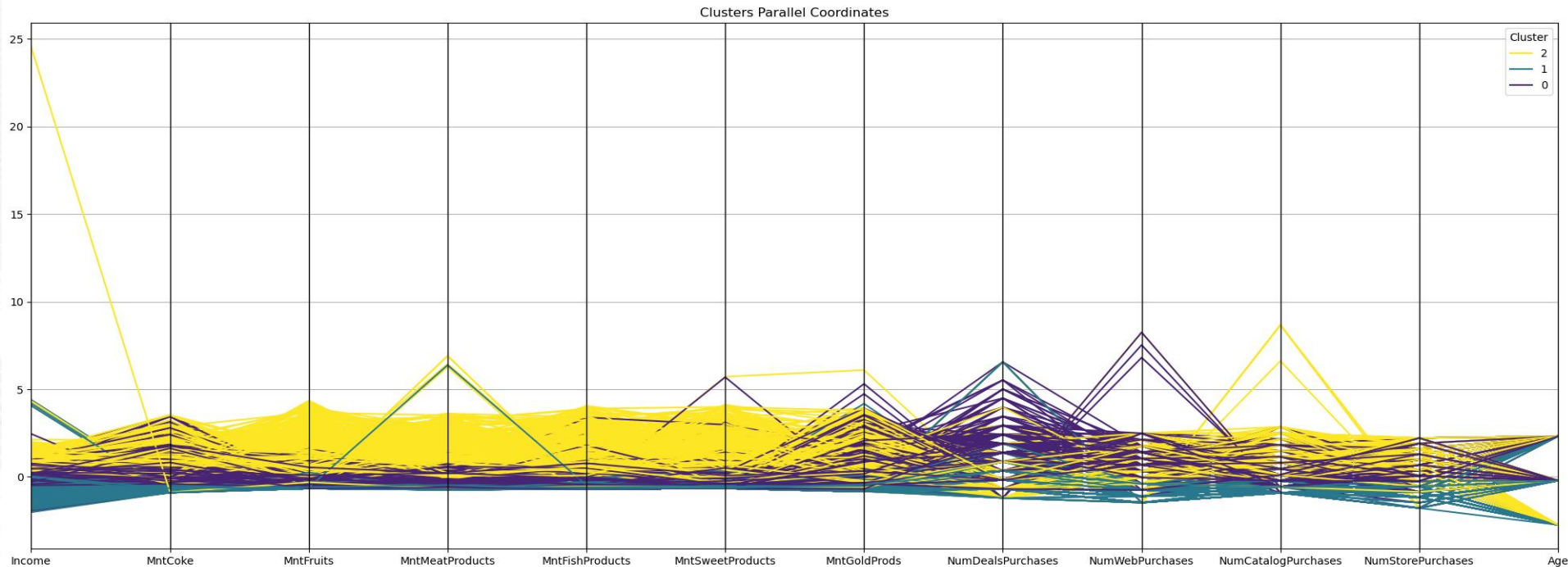
Boxen Plot Distribution
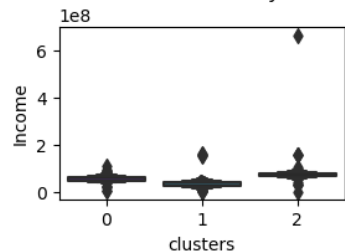
Cluster 0 Recommendation
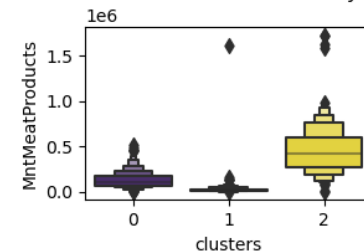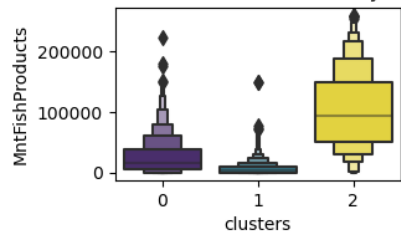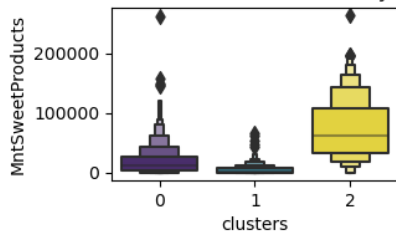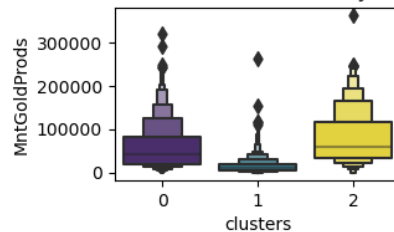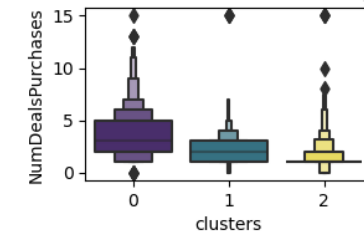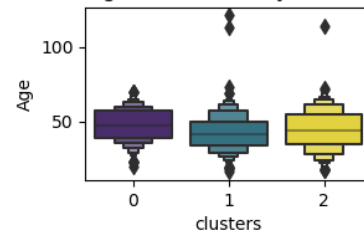
To further engage customers in Cluster 0, who have medium spending habits, it is important to focus on personalized deals campaigns and specific product categories. We can try these recommendations:

- Develop and launch different types of deals, such as limited-time discounts, buy-one-get-one offers, or bundle promotions. Analyze the campaign results to identify the most optimal deals that resonate with this customer segment.
- Create personalized deals and targeted marketing campaigns featuring Coke and Gold products. Tailor the offers based on their past purchase history and preferences to increase the likelihood of conversion.
- Develop a robust loyalty program that rewards customers for their continued patronage. Offer exclusive discounts, early access to sales, and special promotions to incentivize repeat purchases and foster customer loyalty.

To address the challenge of low spending habits among customers in Cluster 1, we have to refine the marketing approach and focus on their slight interest in deals. We can try these recommendations:

- Focus on digital marketing channels and leverage social media platforms, search engine ads, and targeted email campaigns to showcase products or services that align with their preferences and interests.
- Implement Upselling and Cross-selling strategies to increase the average order, provide personalized recommendations during the checkout process or through targeted product suggestions to encourage additional purchases.
- Encourage customers to try catalog purchases by offering exclusive incentives, such as limited-time discounts, free shipping, or personalized catalogs tailored to their interests. Highlight the benefits of catalog shopping, such as convenience or a curated selection of products.
- Re-evaluate our marketing strategies, analyze customer feedback and conduct surveys to gather insights into their needs, preferences, and pain points.
- Capitalize on the slight interest in deals within Cluster 1 by optimizing our deals campaigns. Experiment with different types of deals and offer structures to find the most compelling ones. Monitor customer response and adjust the campaigns based on the data-driven insights gathered during the experimentation phase.

Cluster 2 Recommendation

To leverage the potential of Cluster 2, the highest-spending customer segment, it is important to focus on catalog campaigns and improving in-store services. We can try these recommendations:

- Invest in creating visually appealing and engaging catalog campaigns targeted specifically at Cluster 2 customers. Highlight premium products, exclusive collections, and limited editions to capture their attention. Tailor the catalog content to their preferences and use data-driven insights to curate personalized recommendations.
- Train our staff to deliver top-notch customer service and create a welcoming and luxurious ambiance in our stores. Consider offering exclusive benefits, such as priority access to new arrivals or dedicated personal shopping assistants, to enhance the overall shopping experience.
- Provide personalized VIP treatments, exclusive access to new products, and premium experiences. Offering a dedicated account manager or priority customer support can help nurture loyalty and increase customer satisfaction.
- Form strategic partnerships with luxury brands or high-end product suppliers to cater to the preferences of this affluent customer segment. Collaborative promotions or exclusive product collaborations can attract their attention and reinforce the perception of your business as a high-end destination.

More detail is in here

GitHub

Gmail

Linkedin

Let's Connect !