E-COMMERCE
개인화 마케팅을 위한 AI 세그먼트

정다영 김시환 방제준
신현서 오종현 전진하

KHUDA 2022 2기-A조

Index

Introduction

분석 목표

개인화 마케팅

Introduction

데이터셋

Column: ID, Income, Kidhome, MntFruits, NumWebPurchases 등 29개 요소
Total values: 2240개
결측치: Income에서 24개, 제거 후 2216개 데이터

Introduction

데이터셋

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth
0	5524.0	1957.0	Graduation	Single	58138.0	0.0	0.0	2012-09-04	58.0	635.0	...	7.0
1	2174.0	1954.0	Graduation	Single	46344.0	1.0	1.0	2014-03-08	38.0	11.0	...	5.0
2	4141.0	1965.0	Graduation	Together	71613.0	0.0	0.0	2013-08-21	26.0	426.0	...	4.0
3	6182.0	1984.0	Graduation	Together	26646.0	1.0	0.0	2014-02-10	26.0	11.0	...	6.0
4	5324.0	1981.0	PhD	Married	58293.0	1.0	0.0	2014-01-19	94.0	173.0	...	5.0
...	...	...	...	...	...	...	...	...	...	...	...	...
2235	10870.0	1967.0	Graduation	Married	61223.0	0.0	1.0	2013-06-13	46.0	709.0	...	5.0
2236	4001.0	1946.0	PhD	Together	64014.0	2.0	1.0	2014-06-10	56.0	406.0	...	7.0
2237	7270.0	1981.0	Graduation	Divorced	56981.0	0.0	0.0	2014-01-25	91.0	908.0	...	6.0
2238	8235.0	1956.0	Master	Together	69245.0	0.0	1.0	2014-01-24	8.0	428.0	...	3.0
2239	9405.0	1954.0	PhD	Married	52869.0	1.0	1.0	2012-10-15	40.0	84.0	...	7.0

2240 rows × 29 columns

Introduction

분석과정 설명

EDA

1) Target feature selecting
2) Correlation analysis
3) Festure Importence
   & Feature Selection

Data Processing

1) Remove outliers
2) Data scaling

Dimensionality reduction

1) PCA (principal Comp
   -onent Analysis)

Clustering

1) K-means grouping
2) Analysis by group

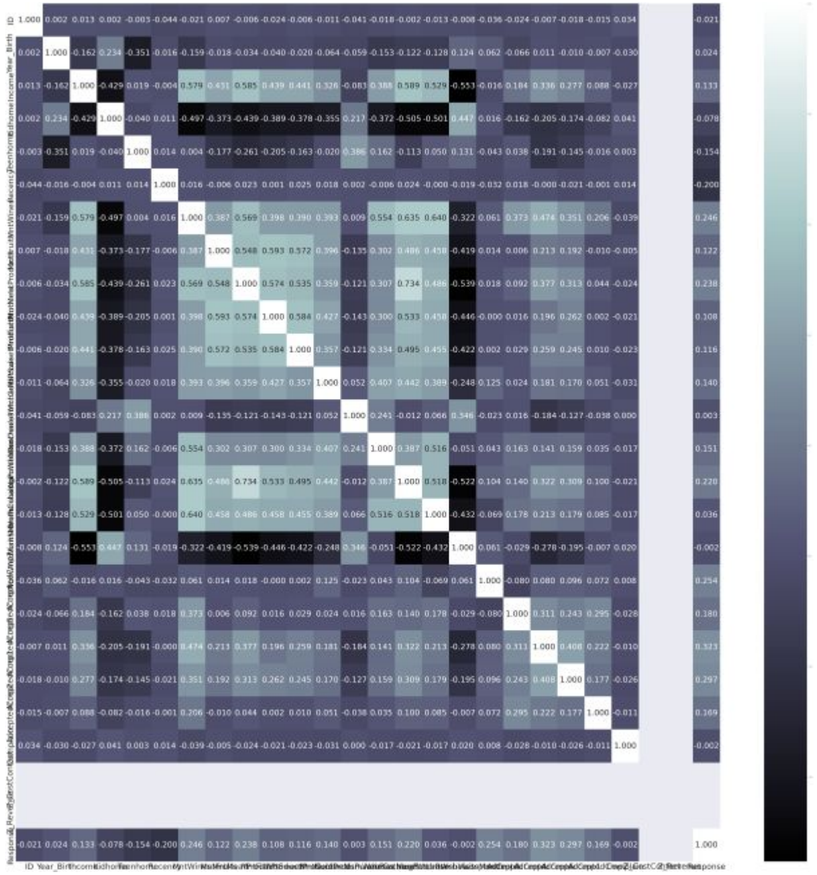
Conclusion

1) marketing strategy
   by group
2) dashboard by group

ANALYSIS

EDA

Target feature selecting & Correlation analysis



Target feature - Income

기준 - Income과의 상관 계수가 0.4 이상 혹은 -0.4 이하

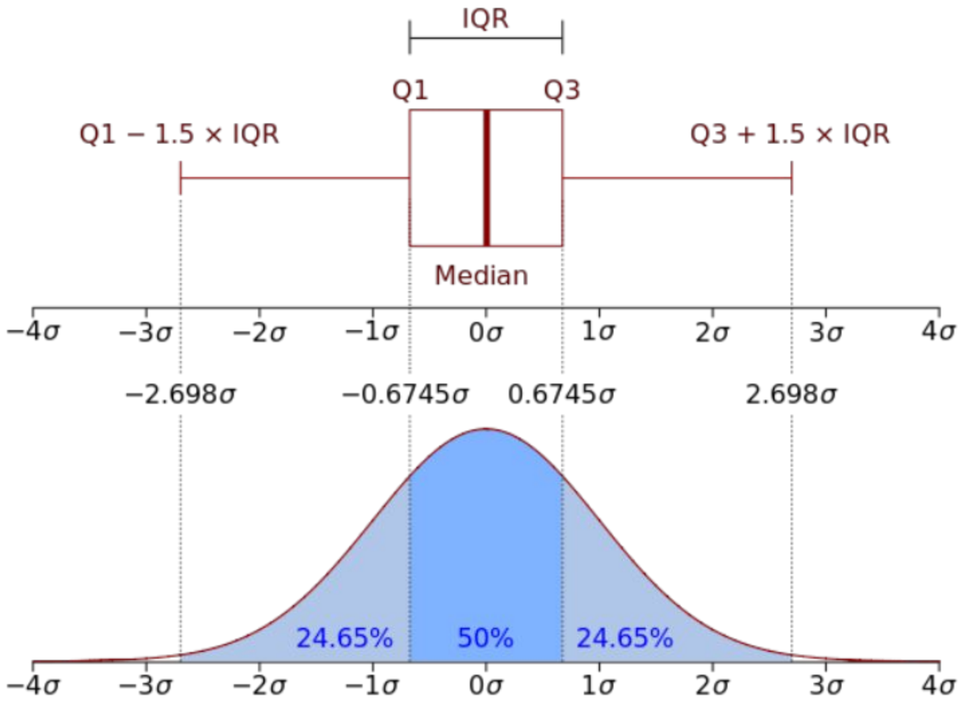
분석 결과 - 9개의 요소 발견

Kidhome, MntWines,
MntFruits, MntMeatProducts, MntFishProducts,
MntSweetProducts, NumCatalogPurchases,
NumStorePurchases, NumWebVisitsMonth

ANALYSIS

Data Processing

Remove outliers & Data scaling



이상치 제거 후: 2208개

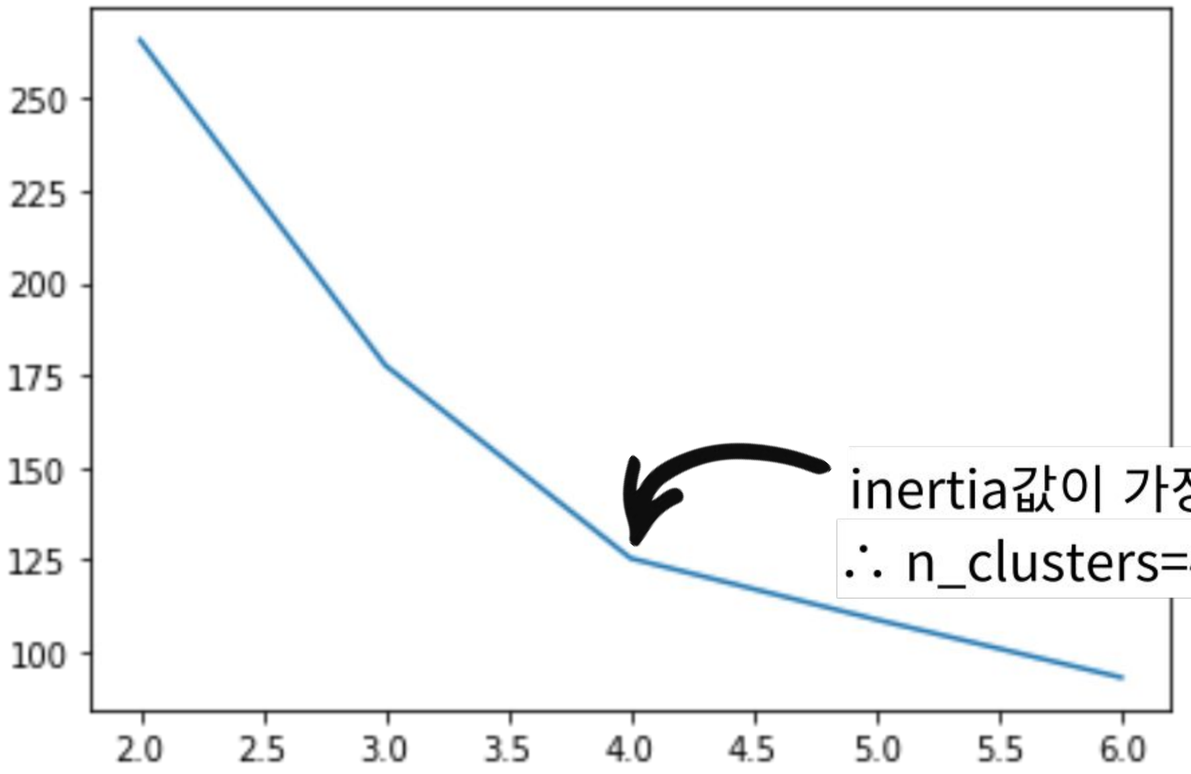
정규화: MinmaxScaler

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

ANALYSIS

Clustering

cluster 개수 결정



inertia값이 가장 크게 꺾이는 구간!

∴ n_clusters=4

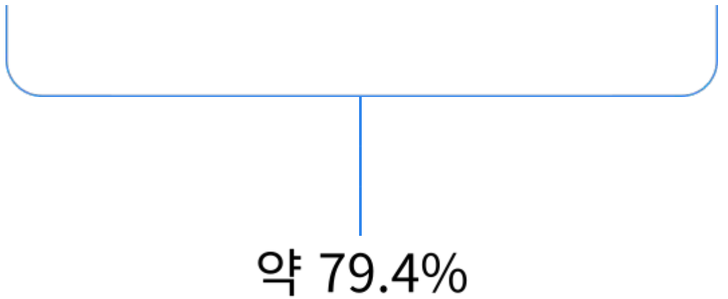
ANALYSIS

Dimensionality reduction

주성분 분석 및 시각화

```
from sklearn.decomposition import PCA
pca = PCA(n_components=5)
pca.fit(train)
```

```
pca.explained_variance_ratio_
0.54738522,  0.12334928,  0.10307418,  0.06105208,  0.05395701
```



약 79.4%

ANALYSIS

Dimensionality reduction

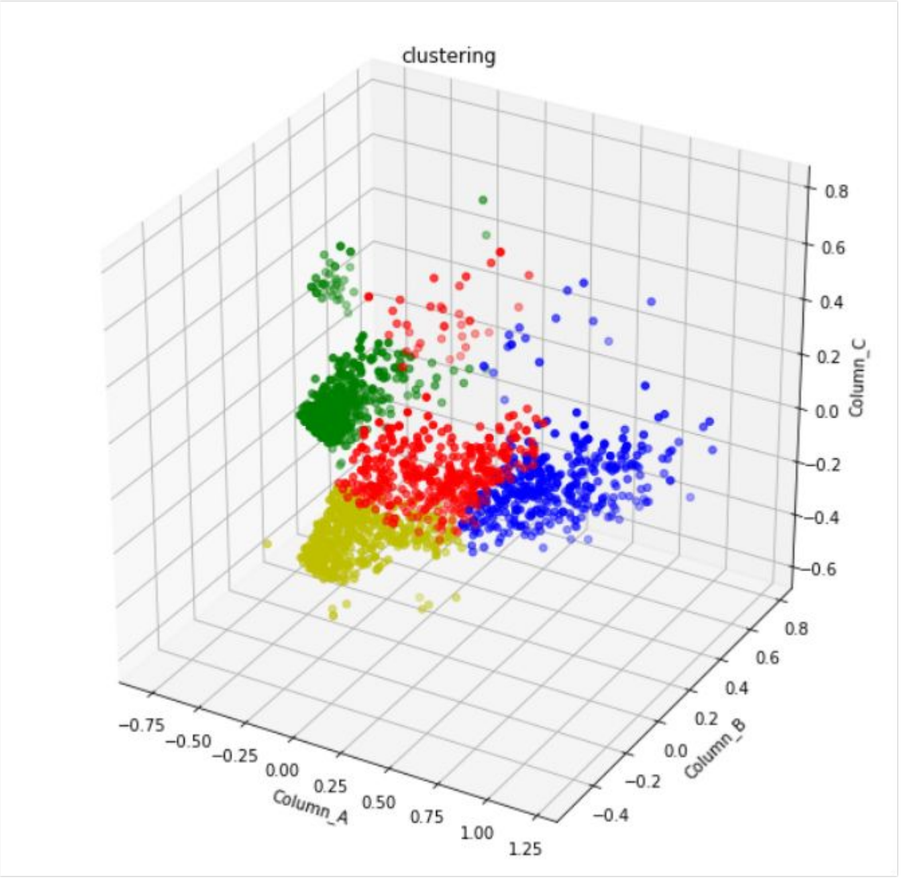
주성분 분석 및 시각화

Cluster_0

Cluster_1

Cluster_2

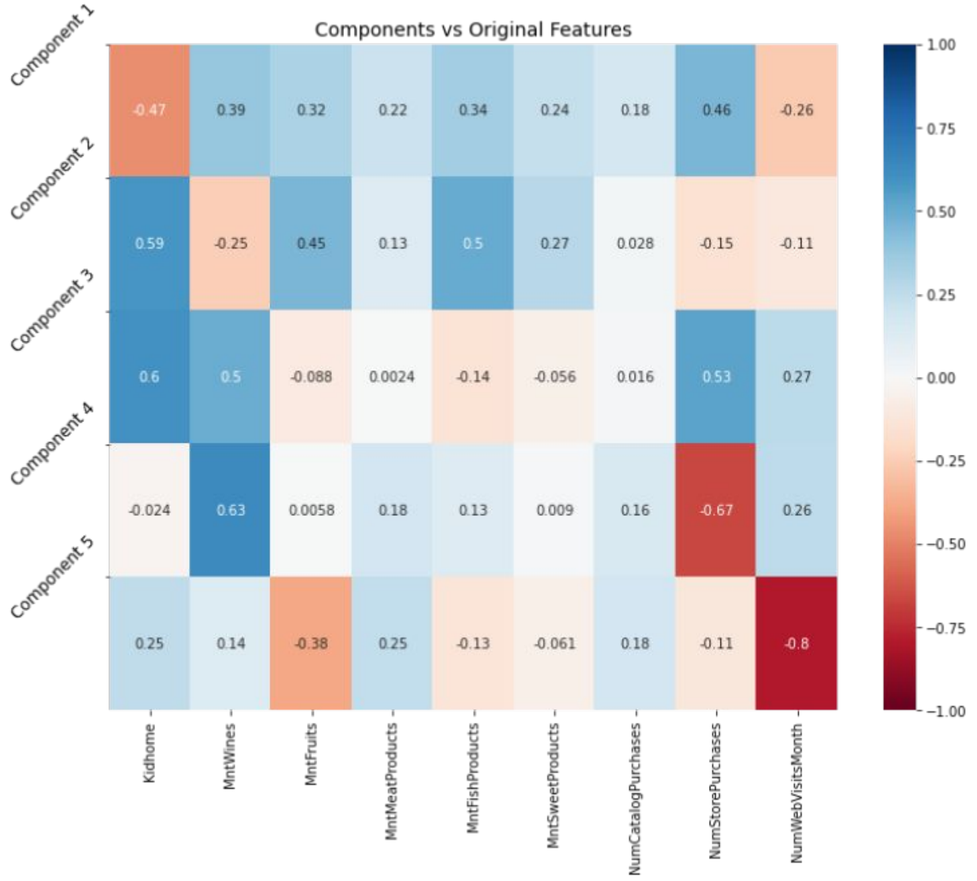
Cluster_3



ANALYSIS

Dimensionality reduction

주성분 분석 및 시각화



Component_1: kidhome(-0.47), NumStorePurcha-ses(0.46)

Component_2: kidhome(0.59), MntFruits(0.45), Mn-tFishProducts(0.5)

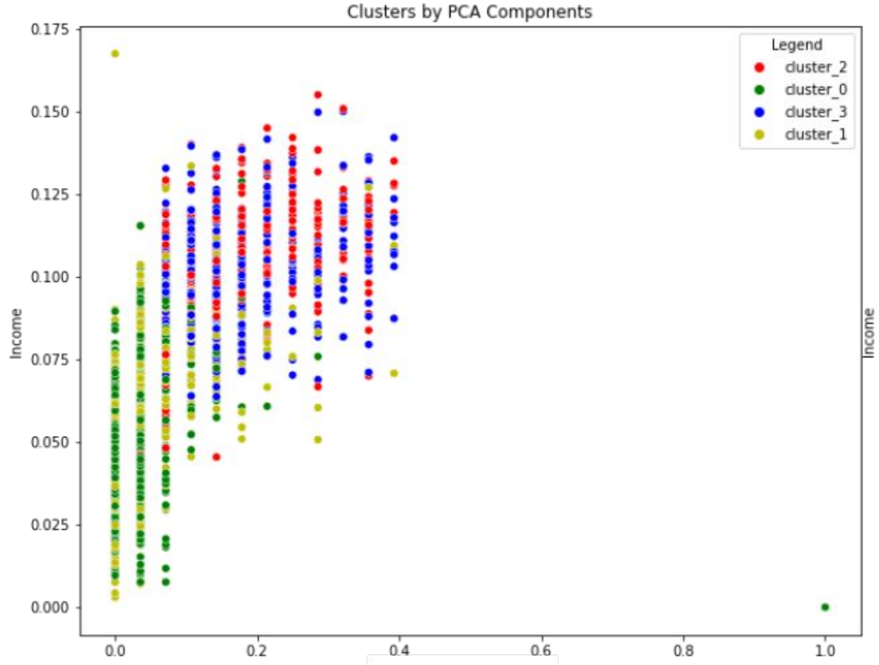
Component_3: kidhome(0.6), MntWines(0.5), Num-StorePurchases(0.53)

Component_4: MntWines(0.63), NumStorePurchas-es(-0.67)

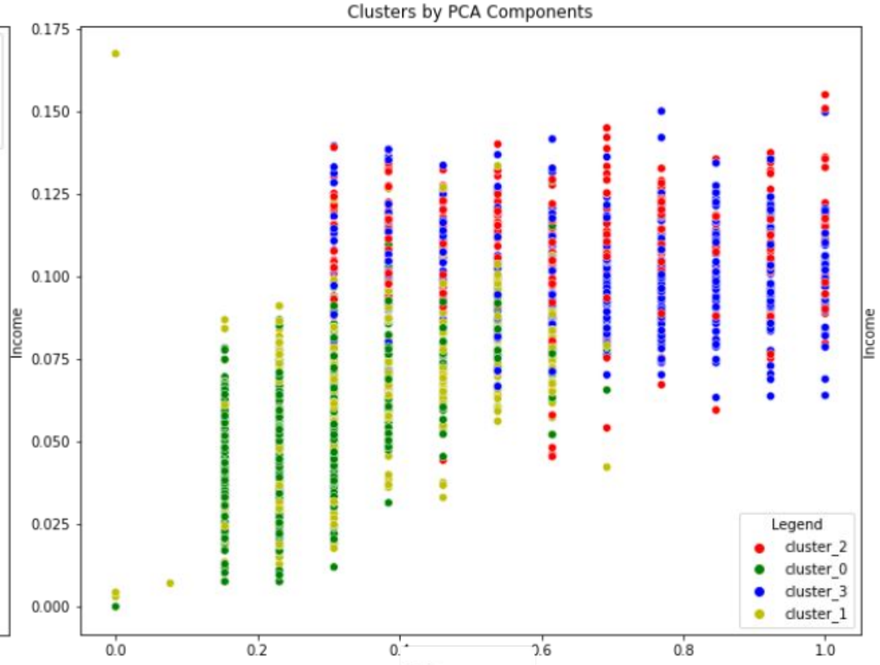
Component_5: NumWebVisitMonth(-0.8)

Conclusion

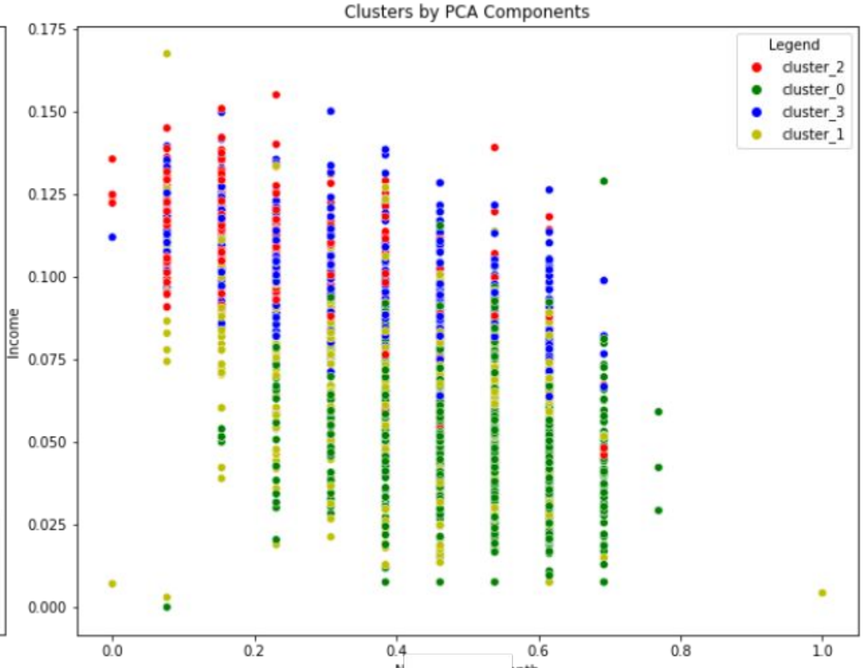
각 군집의 특성-구매경로 성향



Catalog



Store

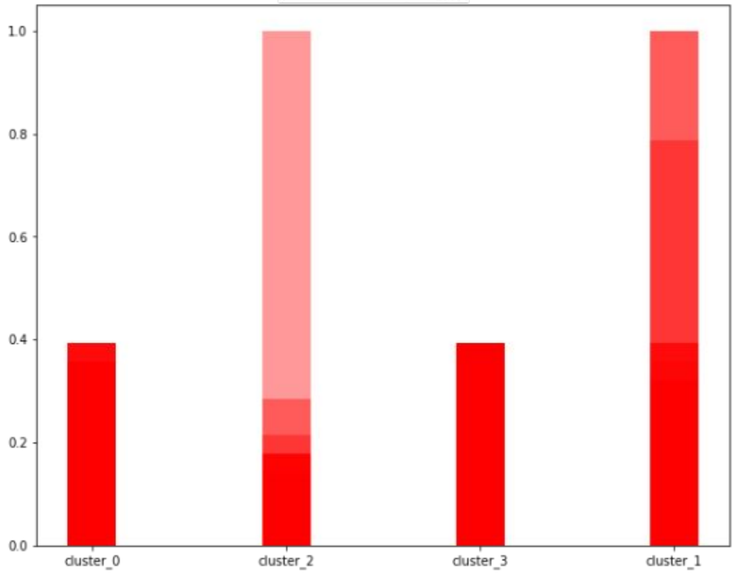


Web

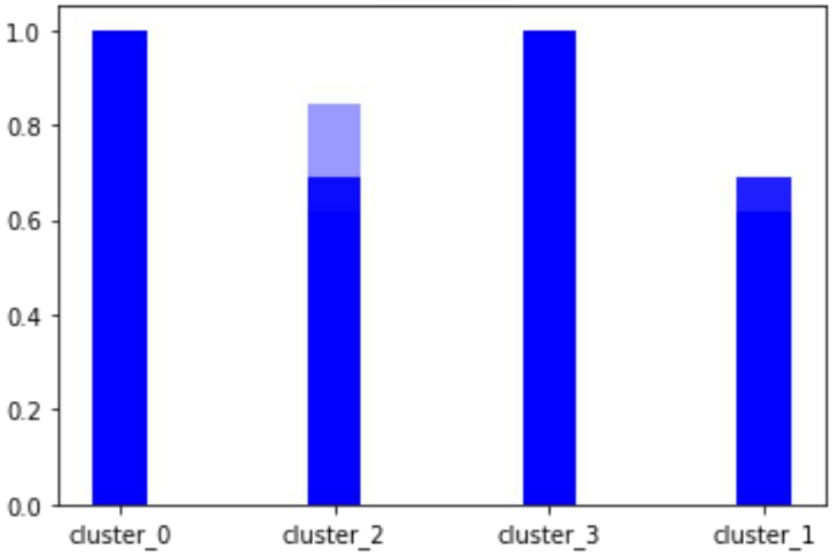
Conclusion

각 군집의 특성-구매경로 성향

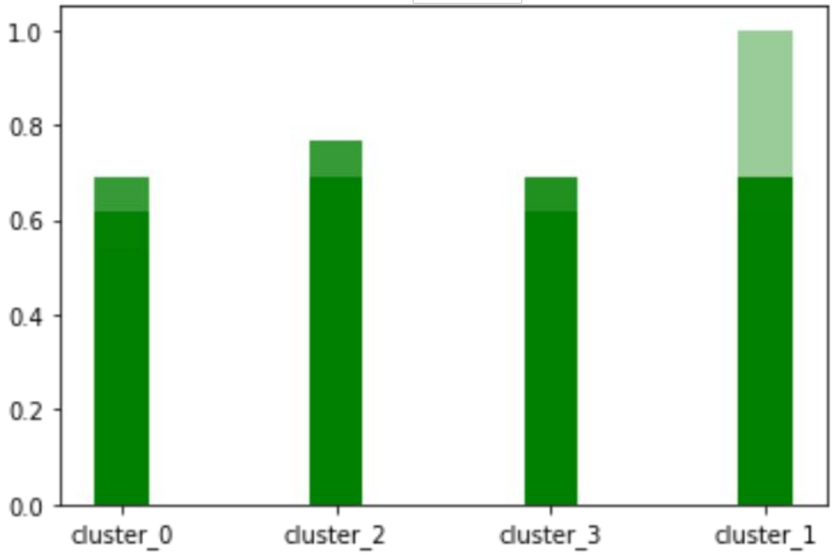
Catalog



Store



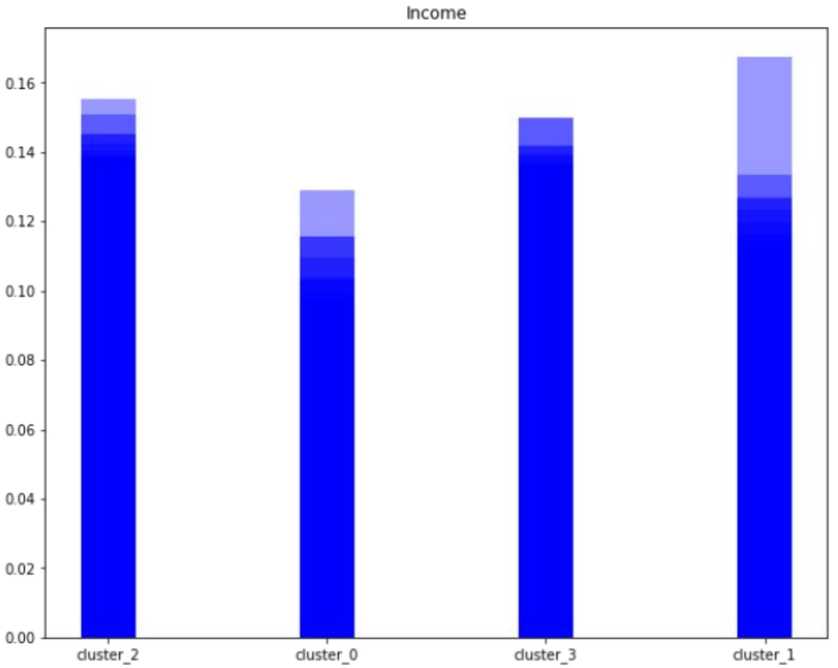
Web

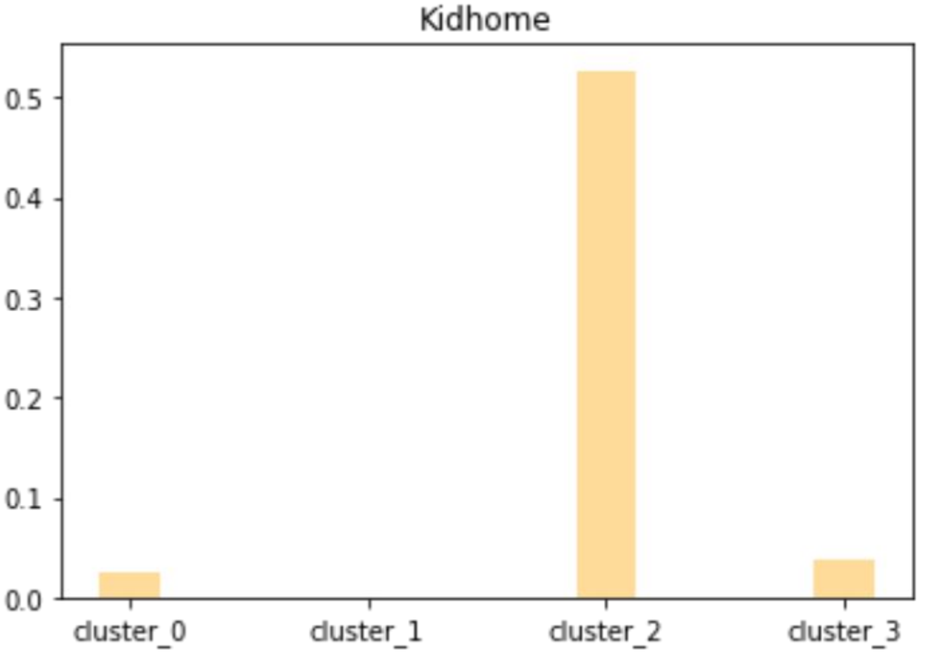


※순서 cluster_0, cluster_2, cluster_3, cluster_1

Conclusion

각 군집의 특성-구매경로 성향





Cluster_0

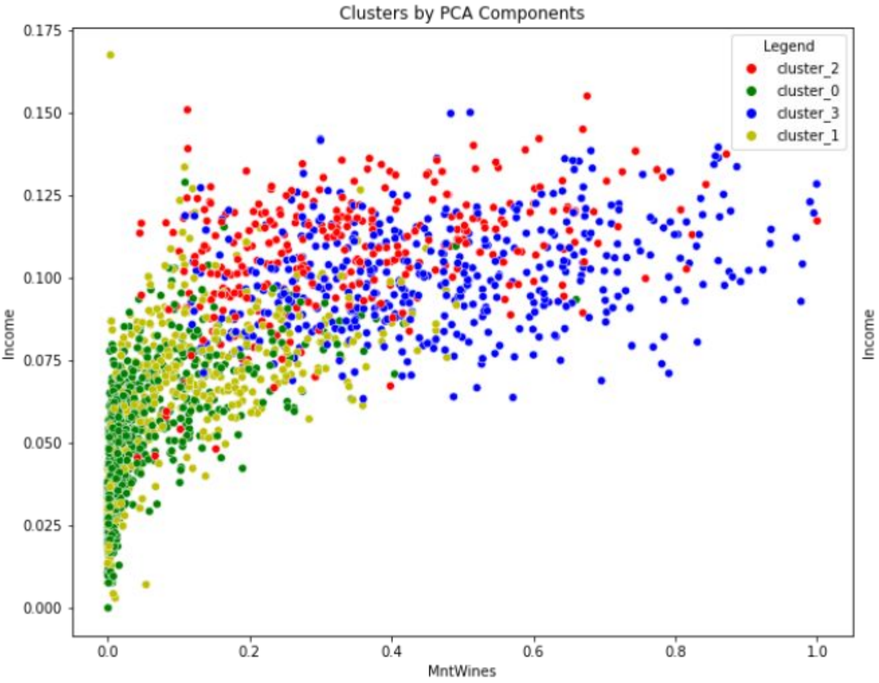
Cluster_1

Cluster_2

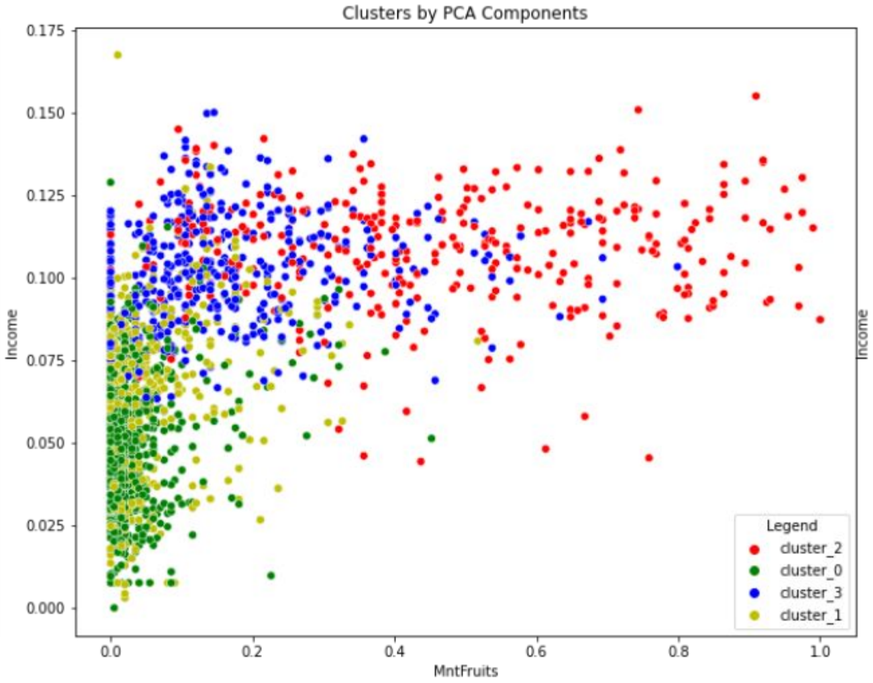
Cluster_3

Conclusion

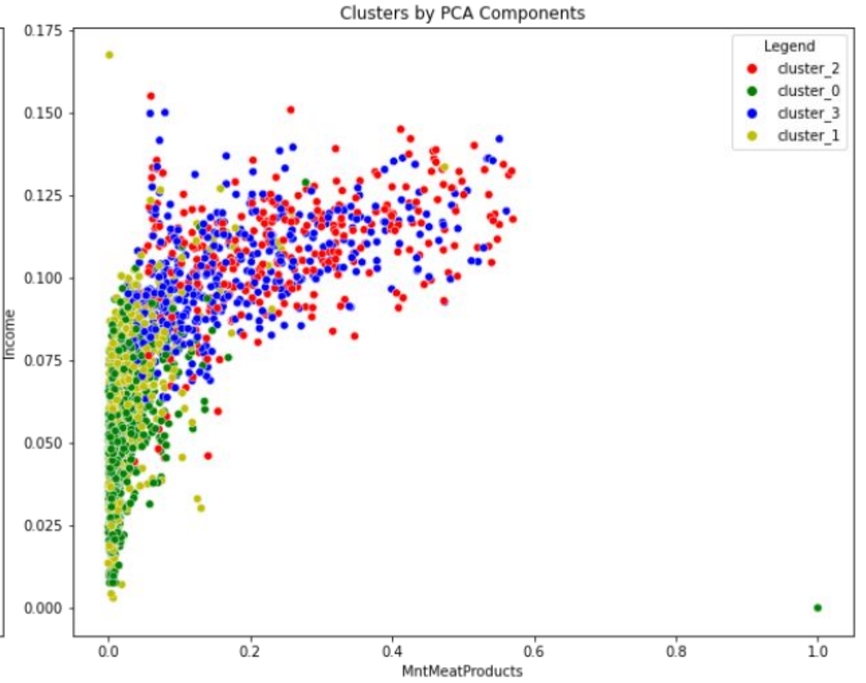
각 품목군 별 구매수량과 수입의 관계



MntWines



MntFruits



MntMeatProducts

Cluster_0

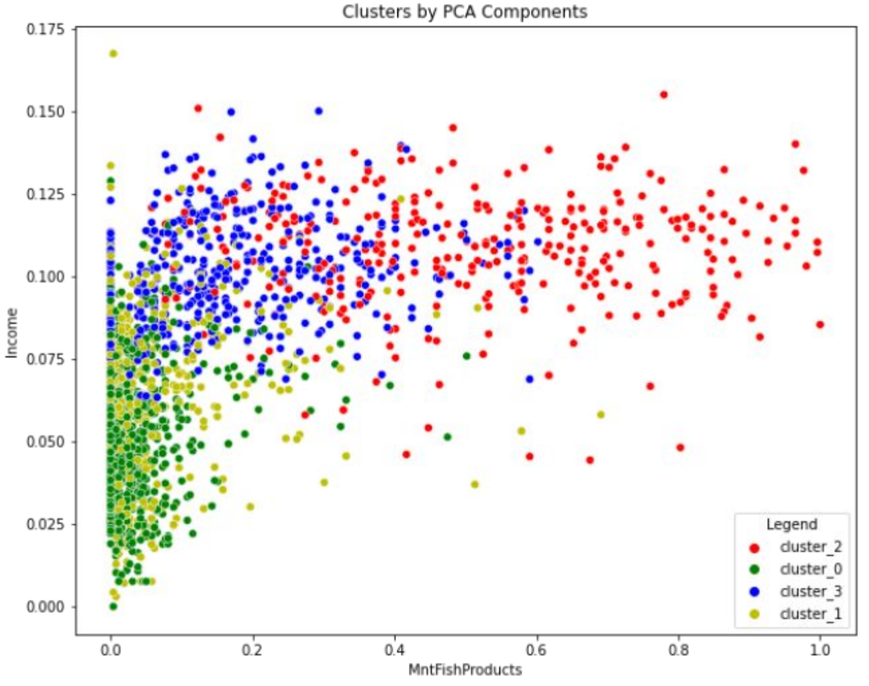
Cluster_1

Cluster_2

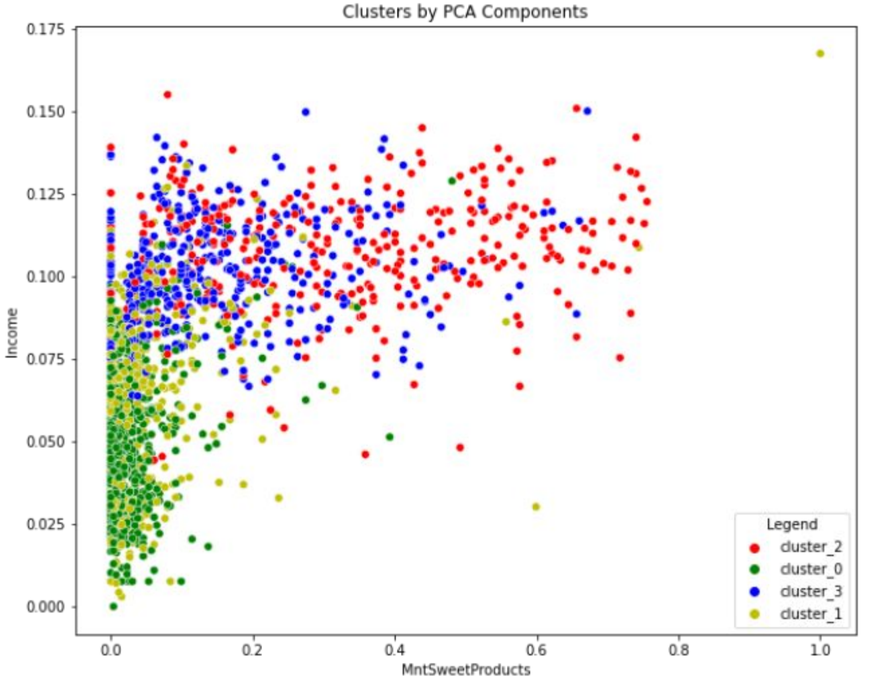
Cluster_3

Conclusion

각 품목군 별 구매수량과 수입의 관계



MntFishProducts



MntSweetProducts

Conclusion

결론 및 전략

Cluster0, Cluster1 중 수입이 낮은 집단 -> 웹사이트 이용 -> 흠과류 웹사이트에 판매

자녀가 없는 Cluster1 -> 1인 가구 품목 카탈로그에 포함

많은 자녀, 준수한 수입 Cluster2 -> 상품(上品)의 과일 배치, 근처에 아이를 위한 품목 배치

매장 구매 경험 중시 Cluster0, Cluster3 -> 와인 코너의 인력 강화

Q&A