

MTS Müşteri Segmentasyonu: Kümeleme

Hazırlayan: Ümit EGBATAN



Bu projede, MTS sistemi veritabanındaki müşteri kayıtları üzerinde denetimsiz bir veri kümelemesi gerçekleştirdim.

Müşteri segmentasyonu, müşterileri her kümedeki müşteriler arasındaki benzerlikleri yansıtan gruplara ayırma uygulamasıdır.


Her müşterinin işletme için önemini optimize etmek için müşterileri segmentlere ayıracağım. Ürünleri müşterilerin farklı ihtiyaç ve davranışlarına göre değiştirmek. Ayrıca işletmenin farklı müşteri türlerinin endişelerini karşılamasına yardımcı olur.

Müşteri Veri setimiz 4075 satır ve 23 kolondan oluşuyor.

Number of datapoints: 4075

	CUST_ORD_ITEM_ID	CUST_ID	CUST_TYPE	FRST_NAME	LST_NAME	BRTH_DATE	AGE	AGE_CATEGORY	GENDR	MARITAL_STATUS	...	ACCPT_PRM	CUST_ORD_ID	SALE_CHANNEL	SIPARIS	CDATE	UDATE	AKTIVASYON_SERVIS	DEAKTIVASYON_SERVIS	CMPG_NAME	CMPG_TYPE
0	1.673220e+10	1.346319e+10	Özel	KÜBRA	ÇİL	1993-11-28 00:00:00.000	31.0	1.0	Kadın	Evli	...	1.0	3.719669e+09	TT OFIS	Kampanya Katılımı/Değişikliği	2024-08-01 09:03:07.000	2024-08-01 09:07:25.000	VDSL İnternet Servis	VDSL İnternet Servis	4 Mevsim Kampanyası	0.0
1	1.673110e+10	1.346321e+10	Özel	EMRE	GÜRBÜZ	1994-07-25 00:00:00.000	30.0	1.0	Erkek	Bekar	...	1.0	3.719252e+09	TT OFIS	NAKİL	2024-02-05 16:25:06.000	2024-02-05 16:38:00.000	VDSL İnternet Servis	VDSL İnternet Servis	4 Mevsim Kampanyası	0.0
2	1.672990e+10	1.346272e+10	Özel	GÜVEN	ÖZMERSİN	1982-01-08 00:00:00.000	42.0	1.0	Erkek	Evli	...	1.0	3.718784e+09	TTM	NAKİL	2023-07-07 09:28:00.000	2023-07-07 11:03:08.000	ADSL İnternet Servis	VDSL İnternet Servis	Bizden Kampanyası	0.0
3	1.672985e+10	1.346265e+10	Özel	MUHARREM ZEKİ	SEN	1959-09-15 00:00:00.000	65.0	2.0	Erkek	Bekar	...	0.0	3.718765e+09	CALL CENTER	NAKİL	2023-06-14 16:14:29.000	2023-06-14 16:22:37.000	VDSL İnternet Servis	VDSL İnternet Servis	Emekliye İnternet Kampanyası	1.0
4	1.673002e+10	1.346281e+10	Özel	MÜMİN	YÜKSEL	1978-01-14 00:00:00.000	46.0	1.0	Erkek	Evli	...	1.0	3.718826e+09	TTM	NAKİL	2023-07-31 14:55:32.000	2023-07-31 15:26:01.000	VDSL İnternet Servis	VDSL İnternet Servis	4 Mevsim Kampanyası	0.0

5 rows × 23 columns




```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4075 entries, 0 to 4074
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CUST_ORD_ITEM_ID      4073 non-null   float64
1   CUST_ID               4073 non-null   float64
2   CUST_TYPE             4073 non-null   object
3   FRST_NAME             4073 non-null   object
4   LST_NAME              4073 non-null   object
5   BRTH_DATE            4073 non-null   object
6   AGE                   4073 non-null   float64
7   AGE_CATEGORY         4073 non-null   float64
8   GENDR                 4073 non-null   object
9   MARITAL_STATUS       4073 non-null   object
10  CHILDREN              4073 non-null   float64
11  INCOME                4073 non-null   float64
12  EDUCATION             4073 non-null   float64
13  ACCPT_PRM            4073 non-null   float64
14  CUST_ORD_ID          4073 non-null   float64
15  SALE_CHANNEL         4073 non-null   object
16  SIPARIS              4073 non-null   object
17  CDATE                4073 non-null   object
18  UDATE                4073 non-null   object
19  AKTIVASYON_SERVIS    4073 non-null   object
20  DEAKTIVASYON_SERVIS  3968 non-null   object
21  CMPG_NAME            4073 non-null   object
22  CMPG_TYPE            4073 non-null   float64
dtypes: float64(10), object(13)
memory usage: 732.4+ KB
```

DATA TEMİZLİĞİ VE YENİ VERİ SETİNE YENİ ÖZELLİKLER EKLEME

```
[ ]  
#Dropping some of the redundant features  
to_drop = ["CUST_ORD_ITEM_ID", "CUST_ID", "CUST_TYPE", "FRST_NAME", "LST_NAME", "BRTH_DATE", "CUST_ORD_ID", "AGE", "DEAKTIVASYON_SERVIS", "CDATE", "UDATE"]  
data = data.drop(to_drop, axis=1)
```

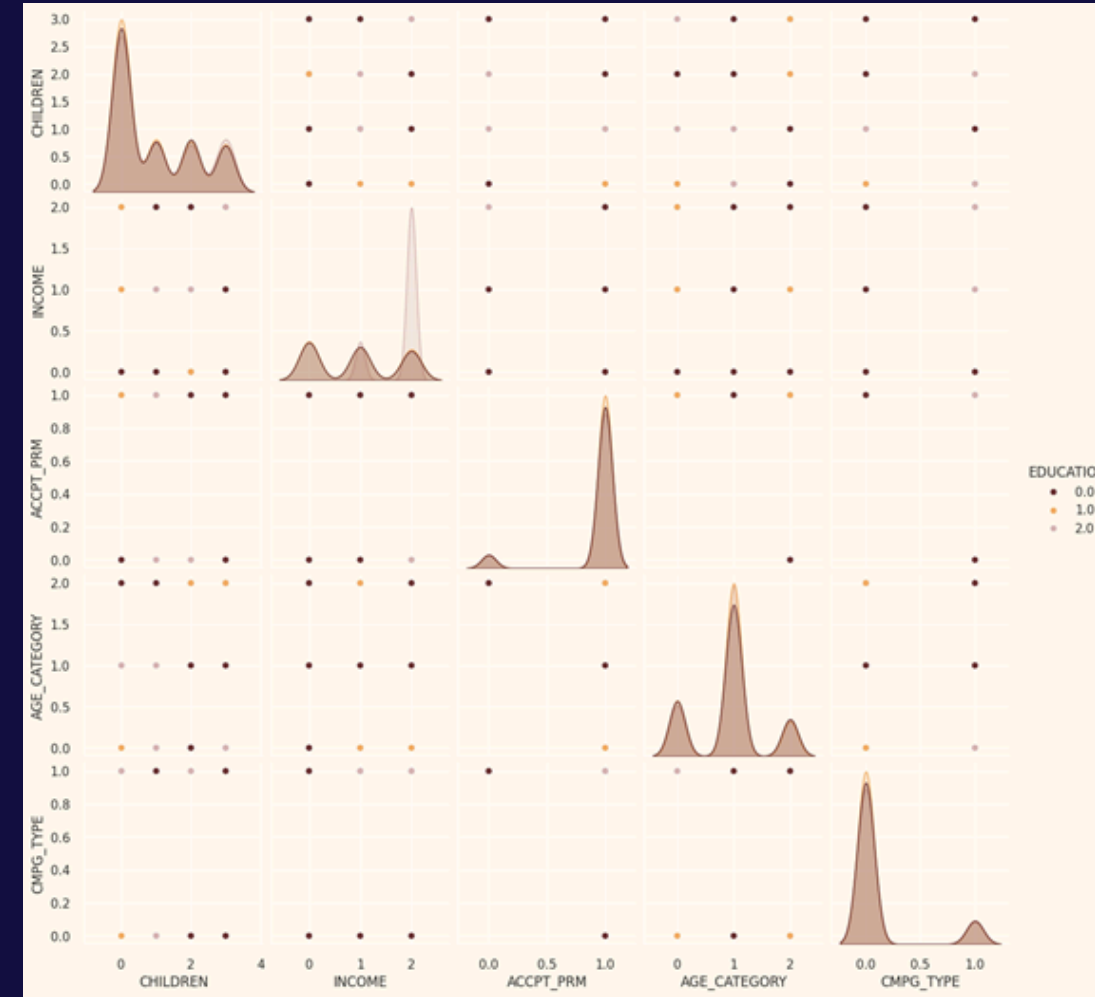
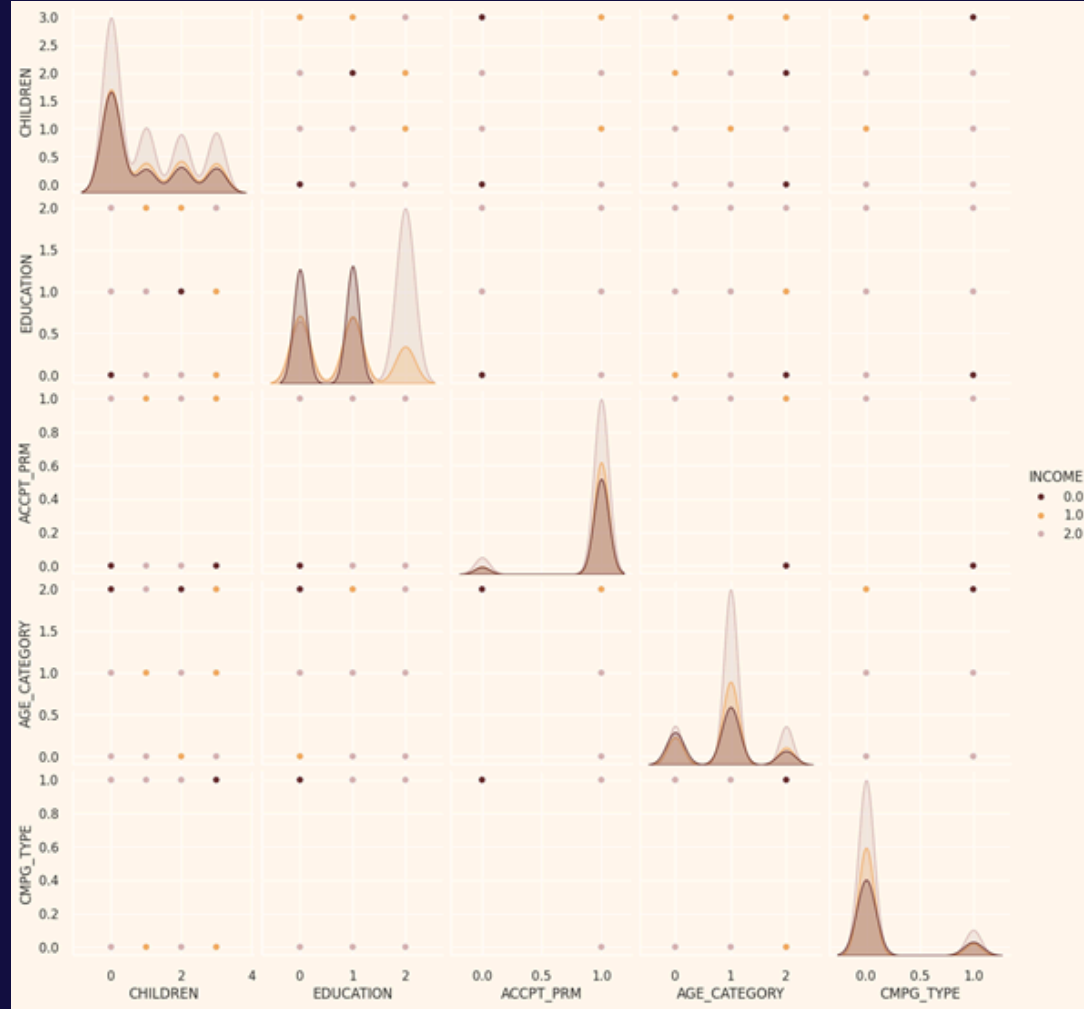
Veri setimizi describe ettim.

```
[ ] data.describe()
```

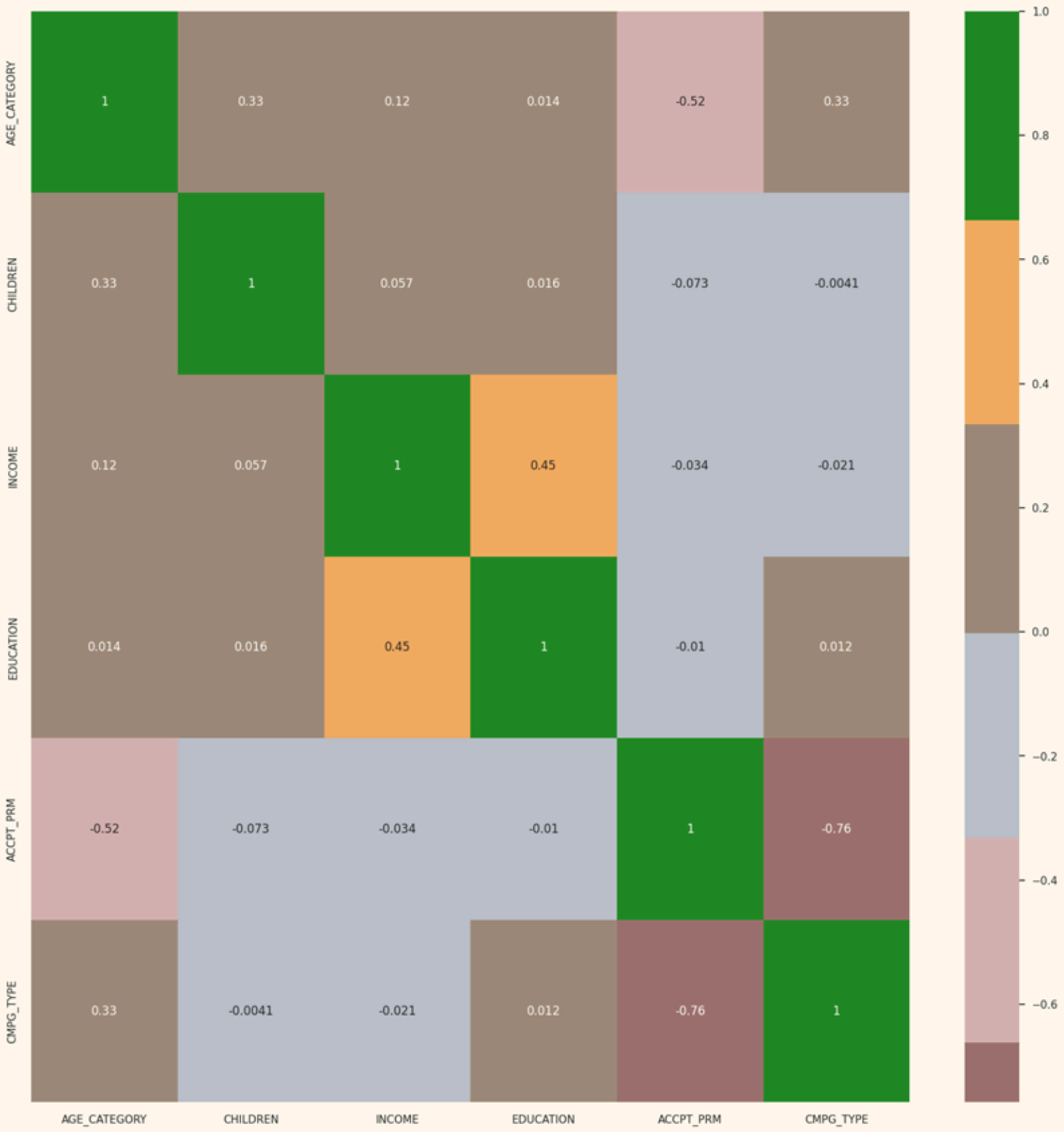


	AGE_CATEGORY	CHILDREN	INCOME	EDUCATION	ACCPT_PRM	CMPG_TYPE
count	4073.000000	4073.000000	4073.000000	4073.000000	4073.000000	4073.000000
mean	0.930764	0.947459	1.216302	1.000491	0.924626	0.124478
std	0.589794	1.142237	0.821560	0.813383	0.264027	0.330167
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.000000	0.000000	0.000000	0.000000	1.000000	0.000000
50%	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000
75%	1.000000	2.000000	2.000000	2.000000	1.000000	0.000000
max	2.000000	3.000000	2.000000	2.000000	1.000000	1.000000

Verilerin daha geniş görünümünü görmek için seçilen özelliklere göre grafiklerini çıkardım



Kolonlar arasındaki korelasyon



Veri setimizde tüm kolonları numerik hale geldikten sonra veriyi scale ettik.

```
#Scaled data to be used for reducing the dimensionality  
print("Dataframe to be used for further modelling:")  
scaled_ds.head()
```

Dataframe to be used for further modelling:

	AGE_CATEGORY	GENDR	MARITAL_STATUS	CHILDREN	INCOME	EDUCATION	ACCPT_PRM	SALE_CHANNEL	SIPARIS	AKTIVASYON_SERVIS	CMPG_NAME	CMPG_TYPE
0	0.117405	1.334452	0.808797	0.921587	-0.263315	-0.000604	0.285515	0.364730	-0.606399	0.650933	-0.637313	-0.377062
1	0.117405	-0.727469	-1.310360	-0.829579	0.954032	1.228981	0.285515	0.364730	0.092092	0.650933	-0.637313	-0.377062
2	0.117405	-0.727469	0.808797	-0.829579	-0.263315	-1.230189	0.285515	0.673511	0.092092	-1.055148	-0.359342	-0.377062
3	1.813120	-0.727469	-1.310360	-0.829579	0.954032	1.228981	-3.502442	-3.340640	0.092092	0.650933	0.037760	2.652080
4	0.117405	-0.727469	0.808797	0.046004	0.954032	-1.230189	0.285515	0.673511	0.092092	0.650933	-0.637313	-0.377062

[] scaled_ds

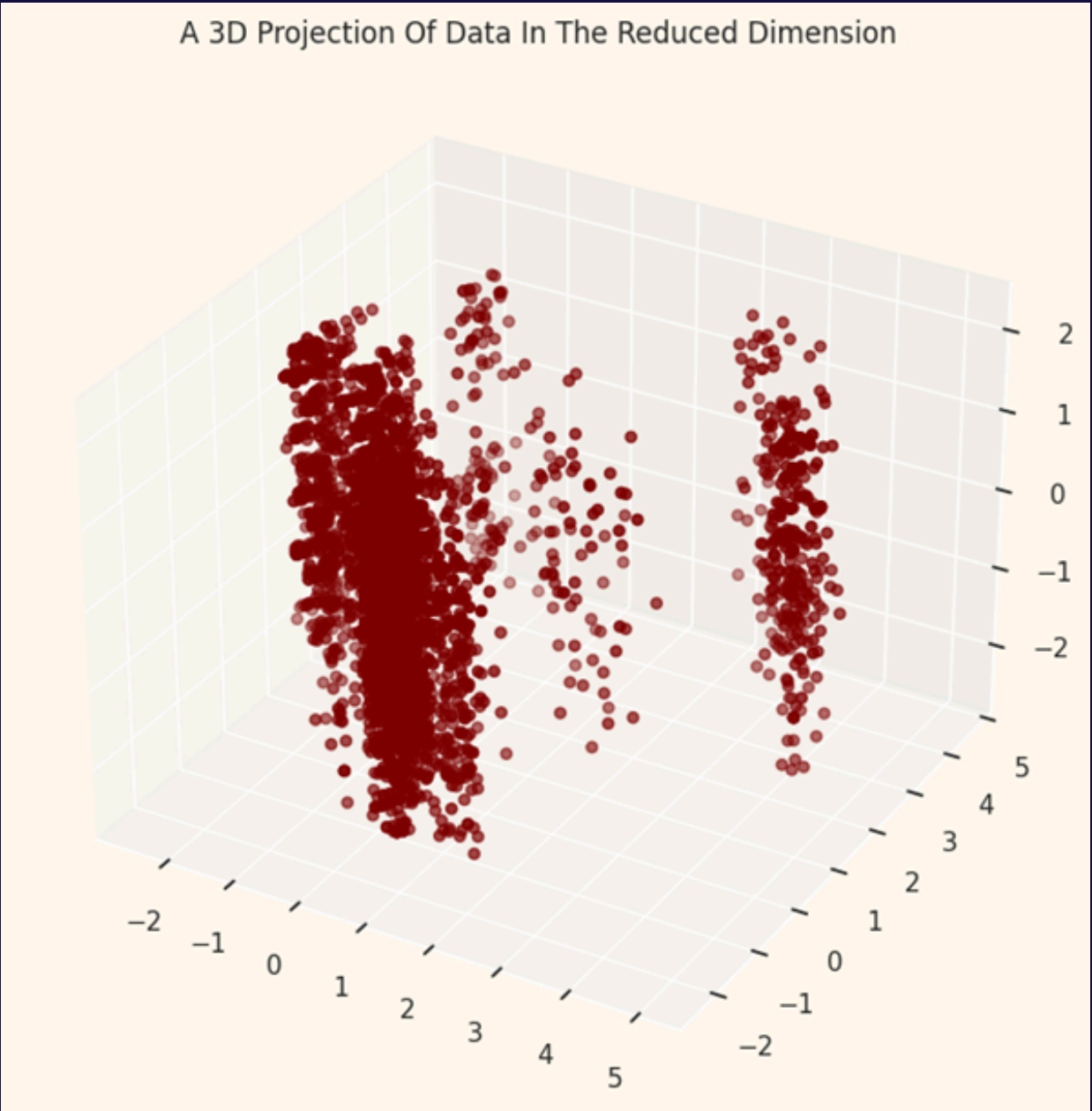
	AGE_CATEGORY	GENDR	MARITAL_STATUS	CHILDREN	INCOME	EDUCATION	ACCPT_PRM	SALE_CHANNEL	SIPARIS	AKTIVASYON_SERVIS	CMPG_NAME	CMPG_TYPE
0	0.117405	1.334452	0.808797	0.921587	-0.263315	-0.000604	0.285515	0.364730	-0.606399	0.650933	-0.637313	-0.377062
1	0.117405	-0.727469	-1.310360	-0.829579	0.954032	1.228981	0.285515	0.364730	0.092092	0.650933	-0.637313	-0.377062
2	0.117405	-0.727469	0.808797	-0.829579	-0.263315	-1.230189	0.285515	0.673511	0.092092	-1.055148	-0.359342	-0.377062
3	1.813120	-0.727469	-1.310360	-0.829579	0.954032	1.228981	-3.502442	-3.340640	0.092092	0.650933	0.037760	2.652080
4	0.117405	-0.727469	0.808797	0.046004	0.954032	-1.230189	0.285515	0.673511	0.092092	0.650933	-0.637313	-0.377062
...
4068	0.117405	1.334452	0.808797	1.797170	-0.263315	-1.230189	0.285515	0.364730	2.187565	0.650933	-0.637313	-0.377062
4069	-1.578309	-0.727469	-1.310360	-0.829579	0.954032	1.228981	0.285515	0.364730	2.187565	0.082239	1.149646	2.652080
4070	0.117405	-0.727469	-1.310360	-0.829579	0.954032	1.228981	0.285515	0.364730	2.187565	0.650933	-0.637313	-0.377062
4071	-1.578309	-0.727469	-1.310360	-0.829579	0.954032	-0.000604	0.285515	0.364730	2.187565	-0.486454	-0.677023	-0.377062
4072	0.117405	1.334452	0.808797	0.921587	0.954032	-1.230189	0.285515	0.364730	2.187565	1.219627	-0.637313	-0.377062

4073 rows × 12 columns

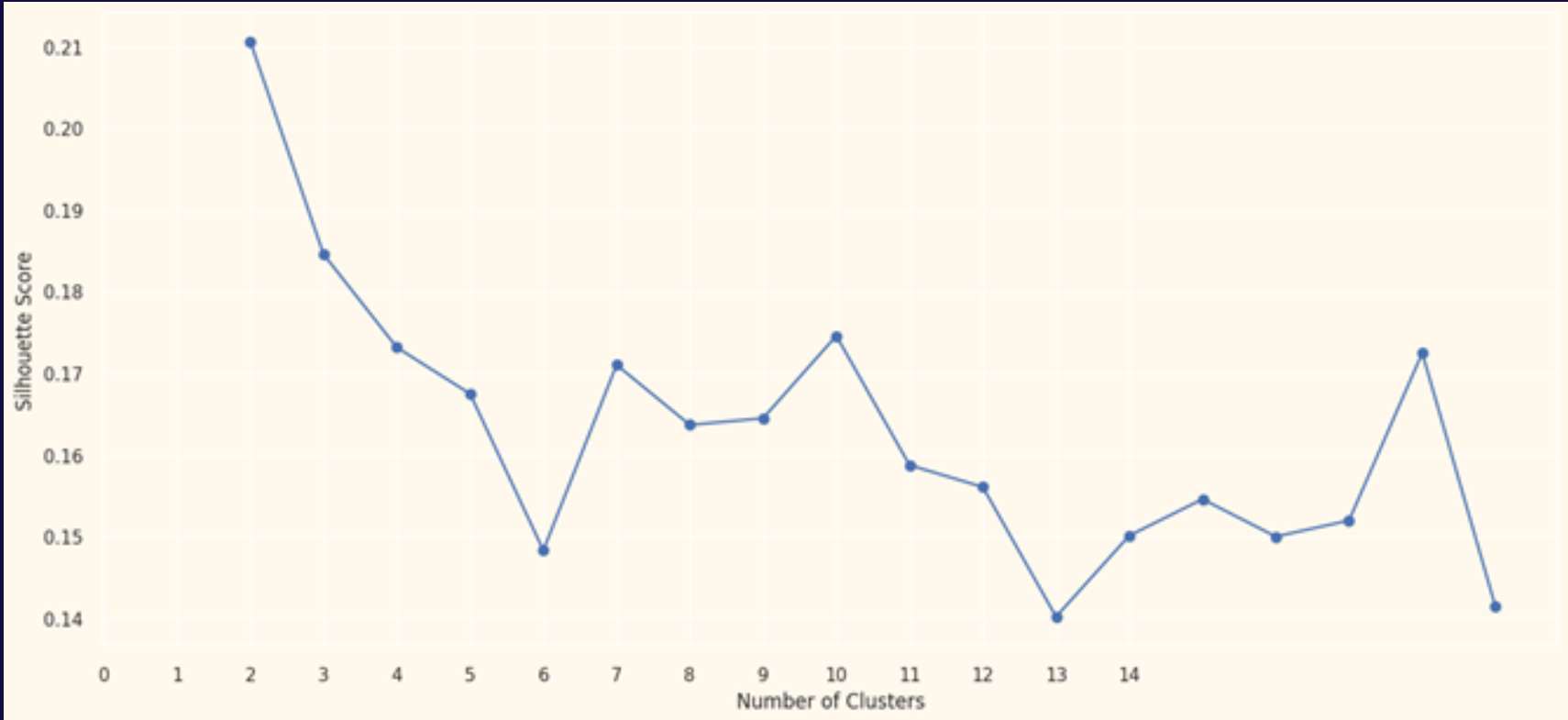
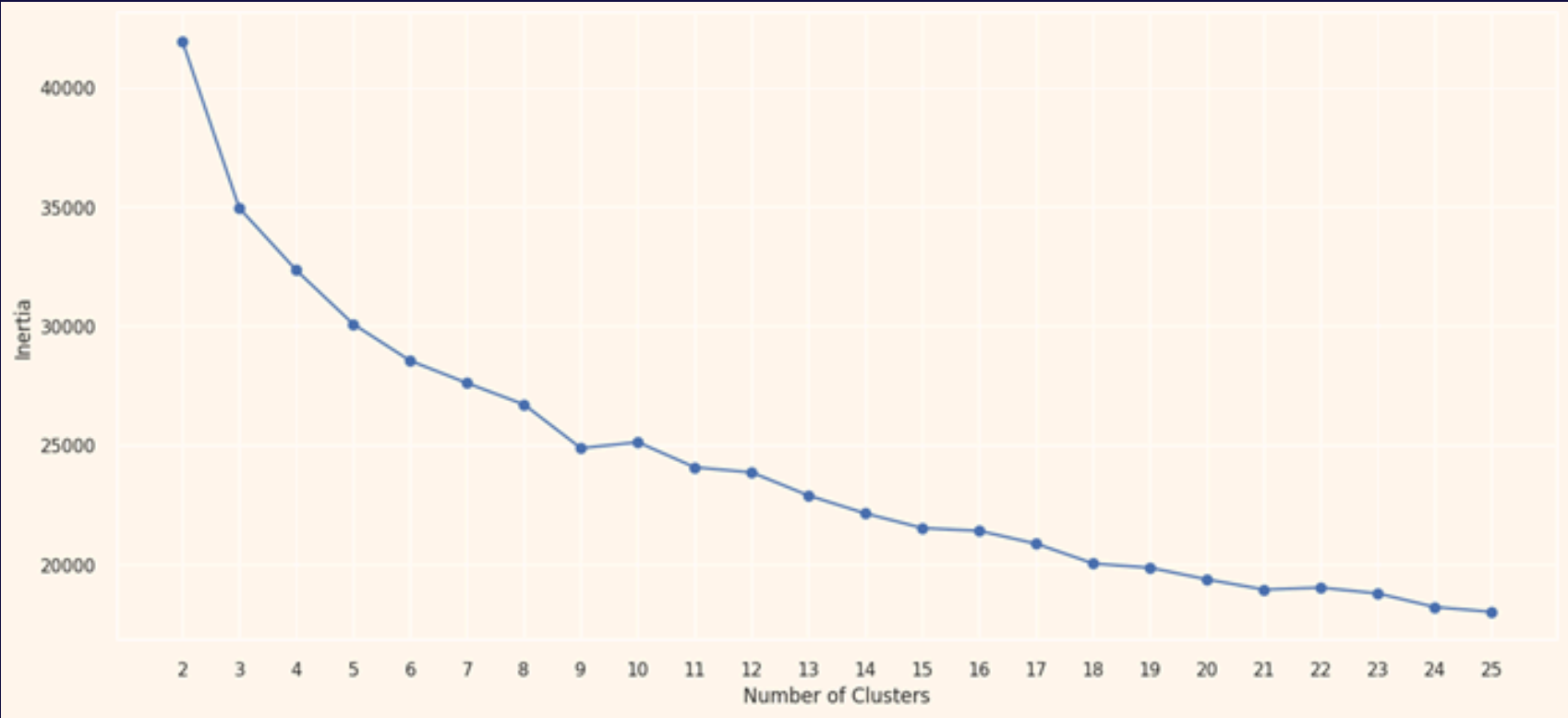
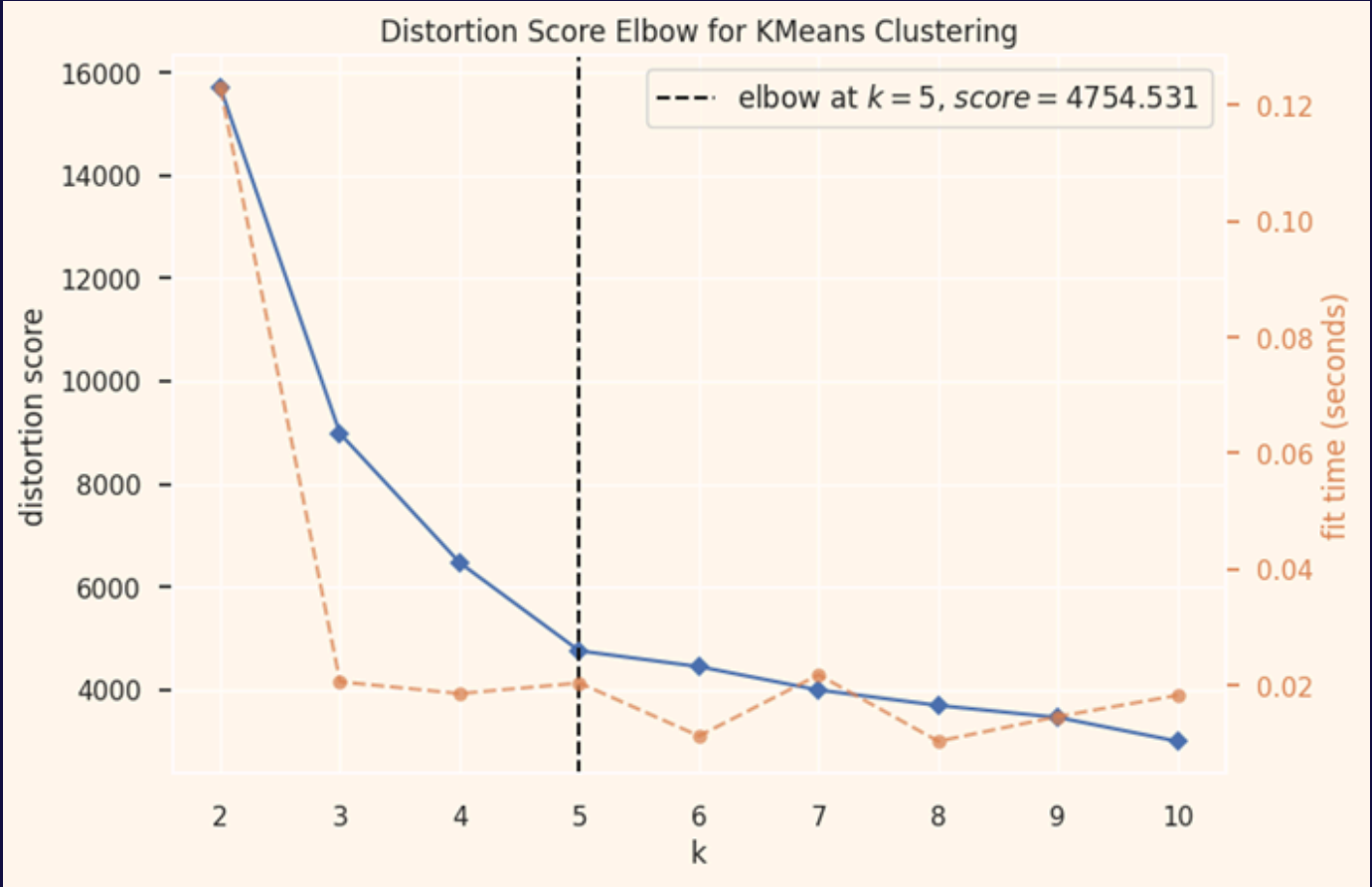
Temel bileşen analizi (PCA), bu tür veri kümelerinin boyutluluğunu azaltmak, yorumlanabilirliği artırmak ve aynı zamanda bilgi kaybını en aza indirmek için kullanılan bir tekniktir. Bu kısımda verileri 3 boyuta indirgedim

```
#Initiating PCA to reduce dimentions aka features to 3
pca = PCA(n_components=3)
pca.fit(scaled_ds)
PCA_ds = pd.DataFrame(pca.transform(scaled_ds), columns=([col1", "col2", "col3"]))
PCA_ds.describe().T
```

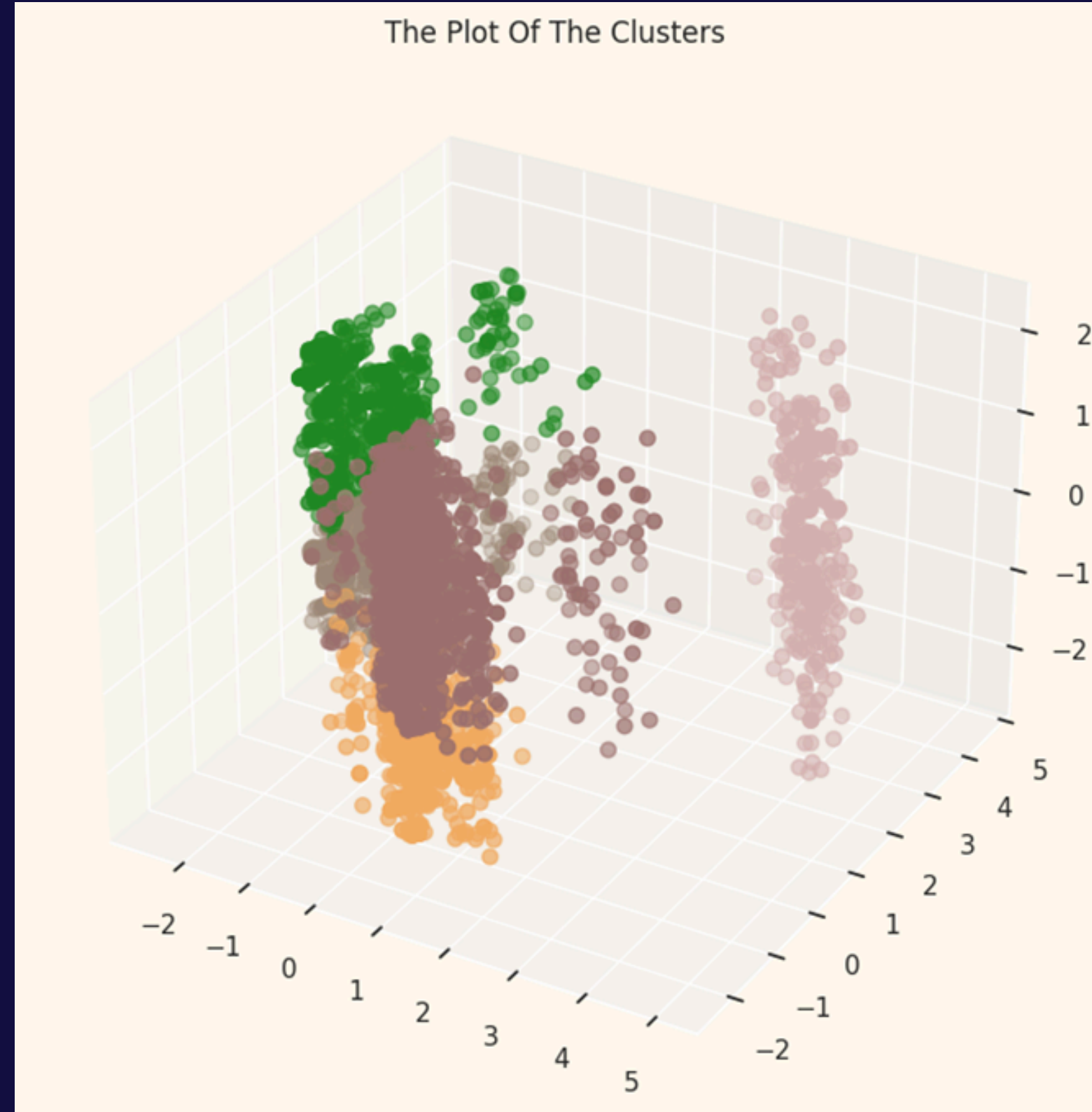
	count	mean	std	min	25%	50%	75%	max
col1	4073.0	0.000000e+00	1.538914	-2.534899	-1.000717	-0.007635	0.605405	5.068152
col2	4073.0	3.489039e-18	1.307544	-2.325567	-1.032166	-0.152730	0.810504	4.653397
col3	4073.0	-2.093423e-17	1.194178	-2.537231	-0.910123	-0.057051	1.084945	2.230568



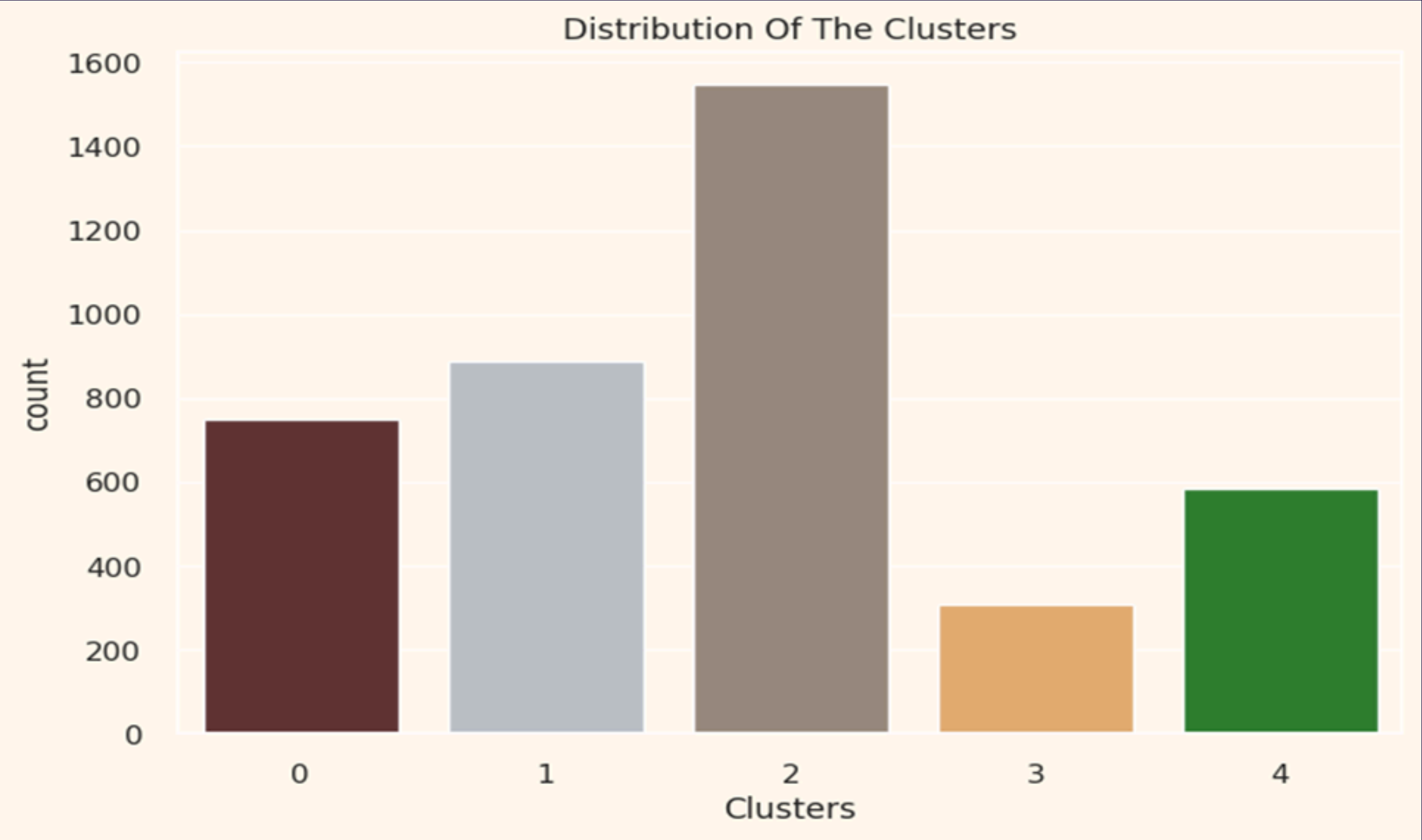
Elbow method, Inertia ve Silhouette Değerlerine göre sınıf sayısı 5 olarak görülmekte



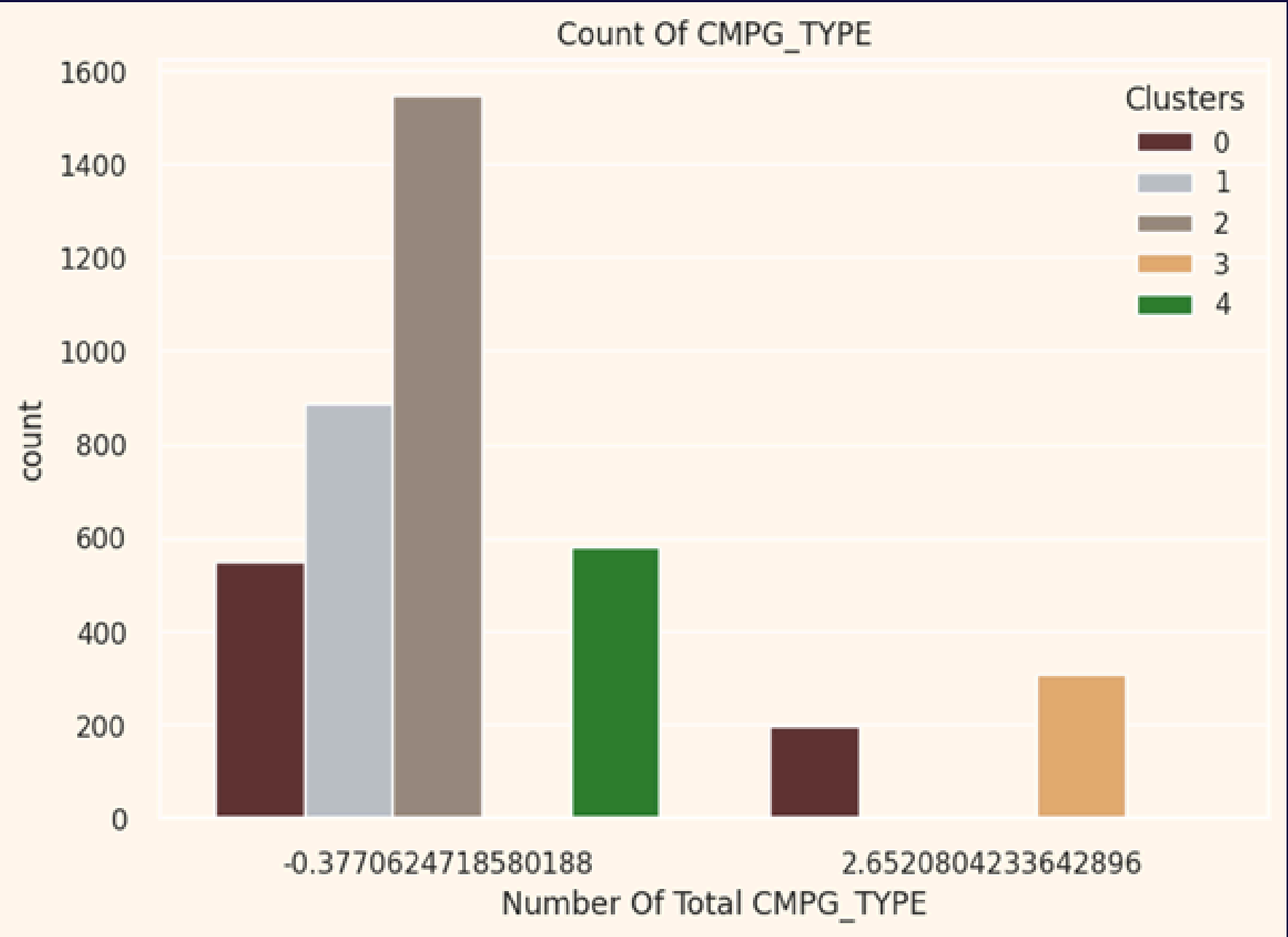
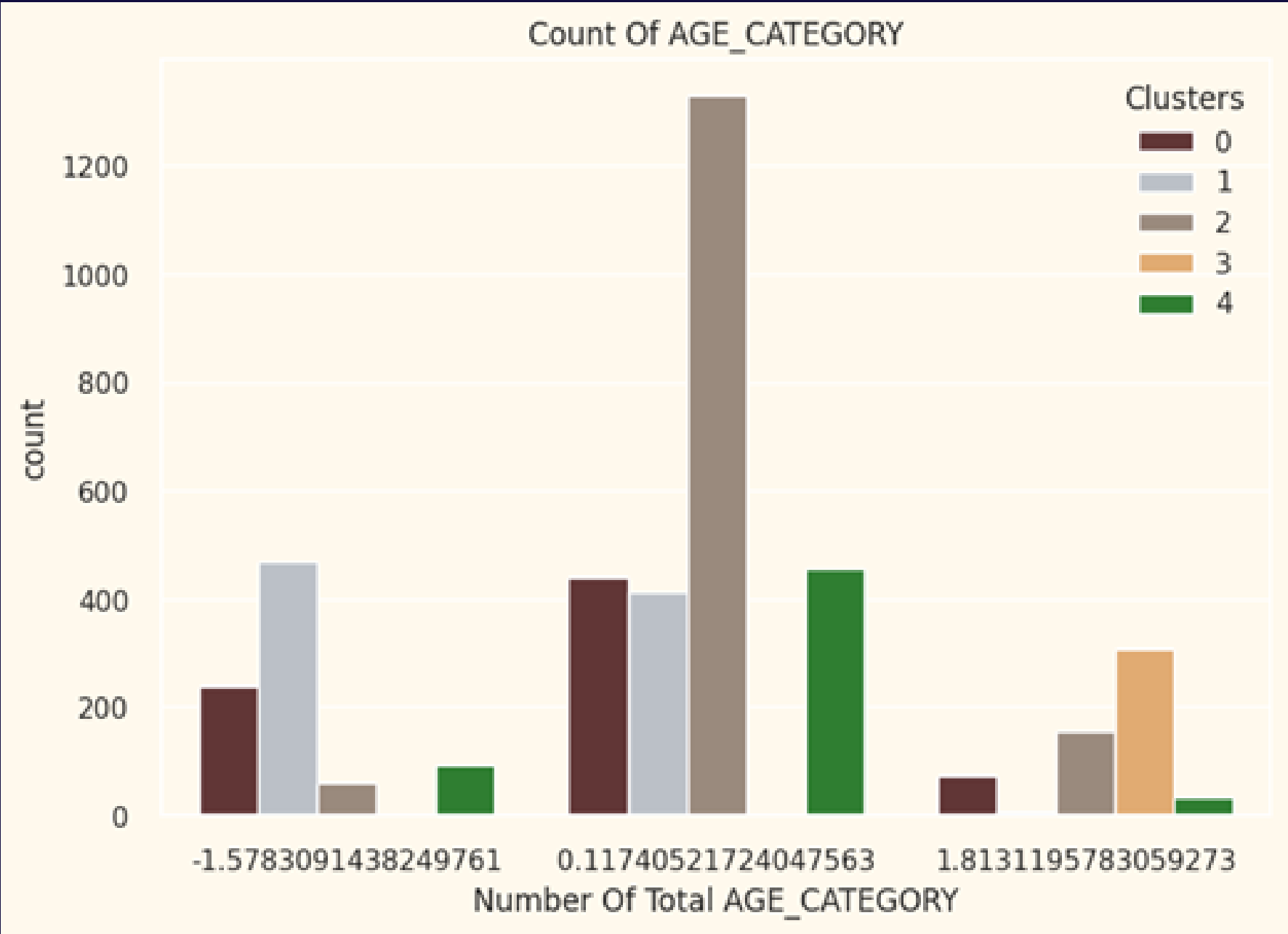
Oluřan kmeleri incelemek iin kmelerin 3 boyutlu daėılımı



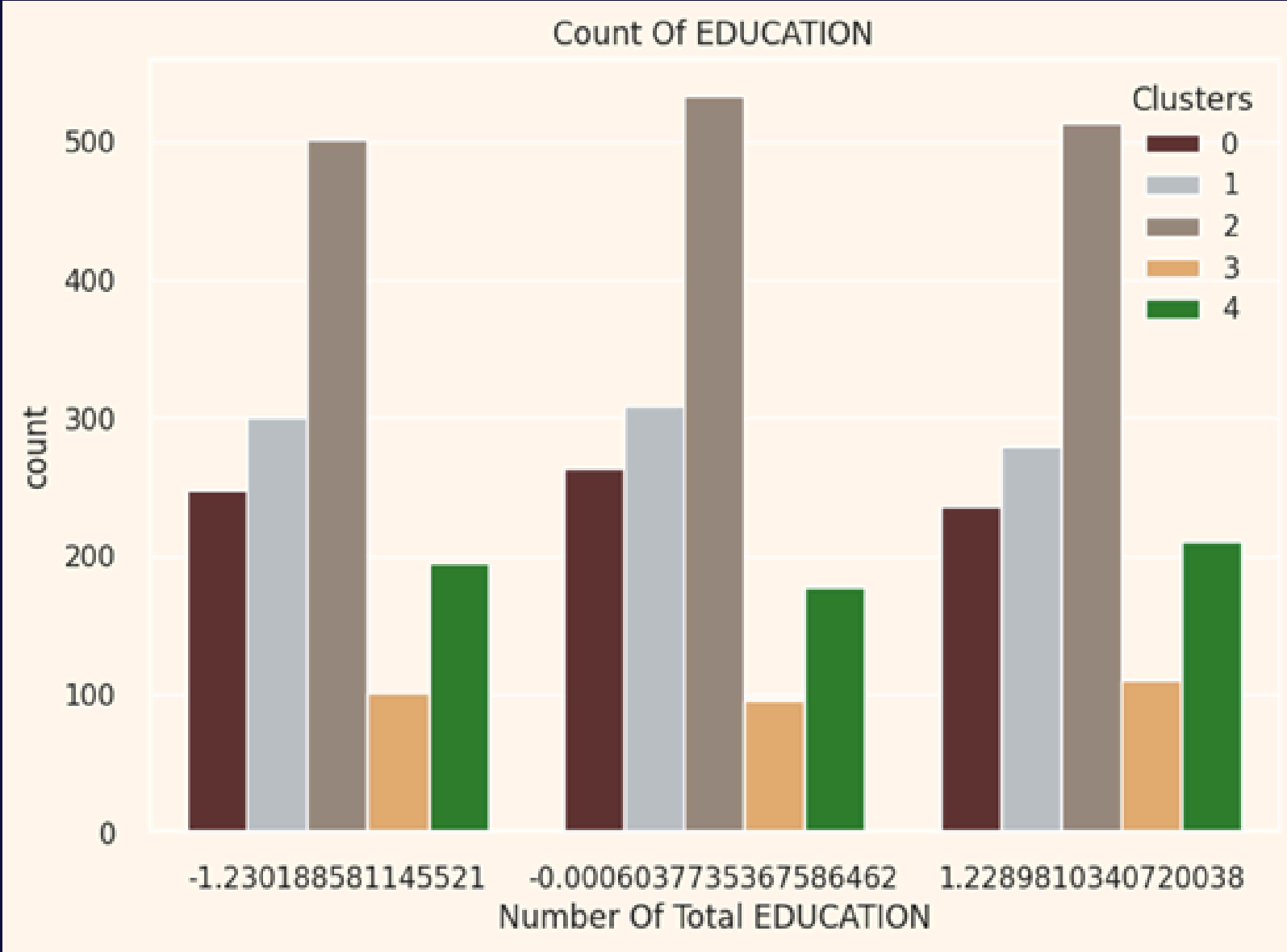
Kümelerin Dağılımı



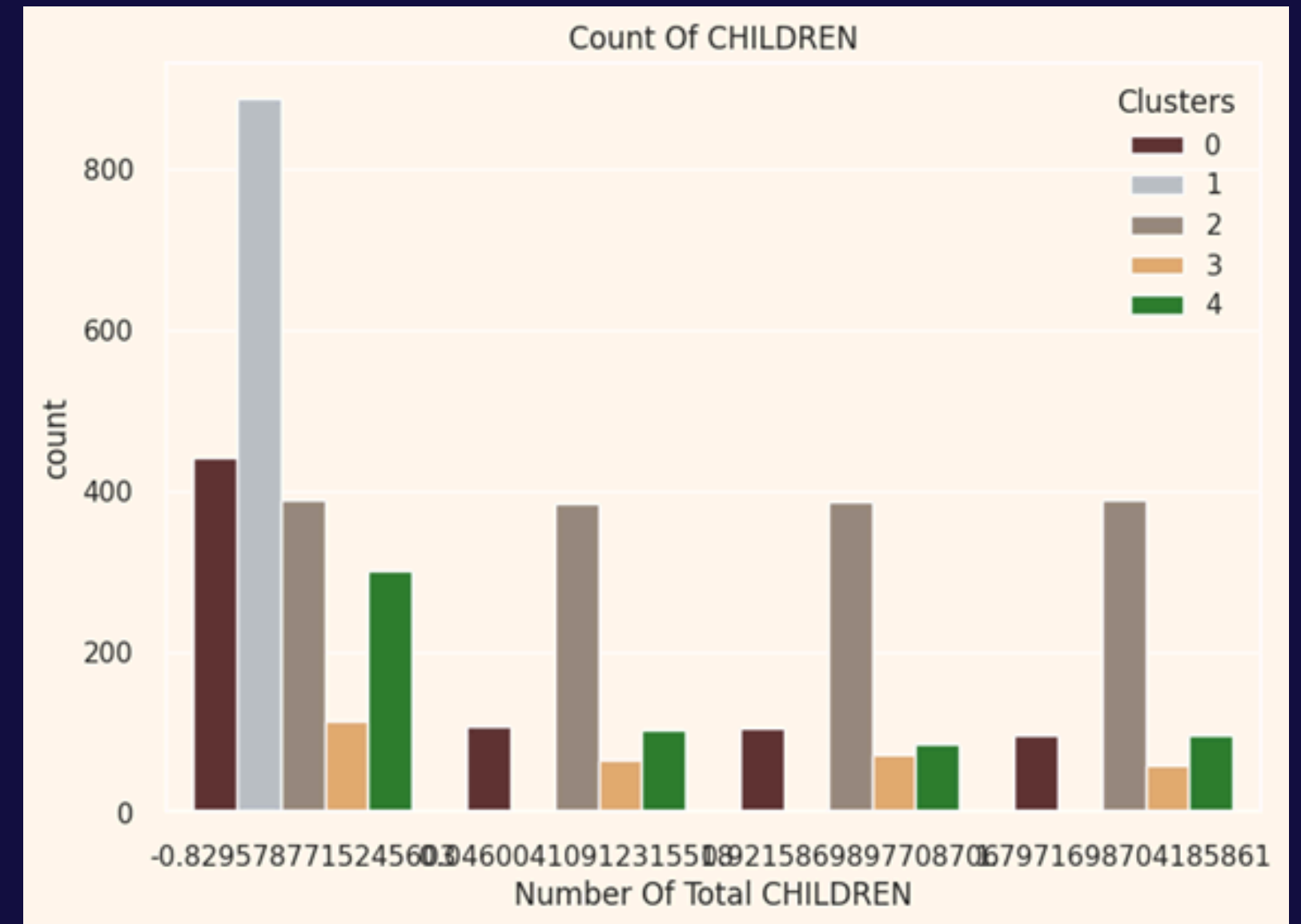
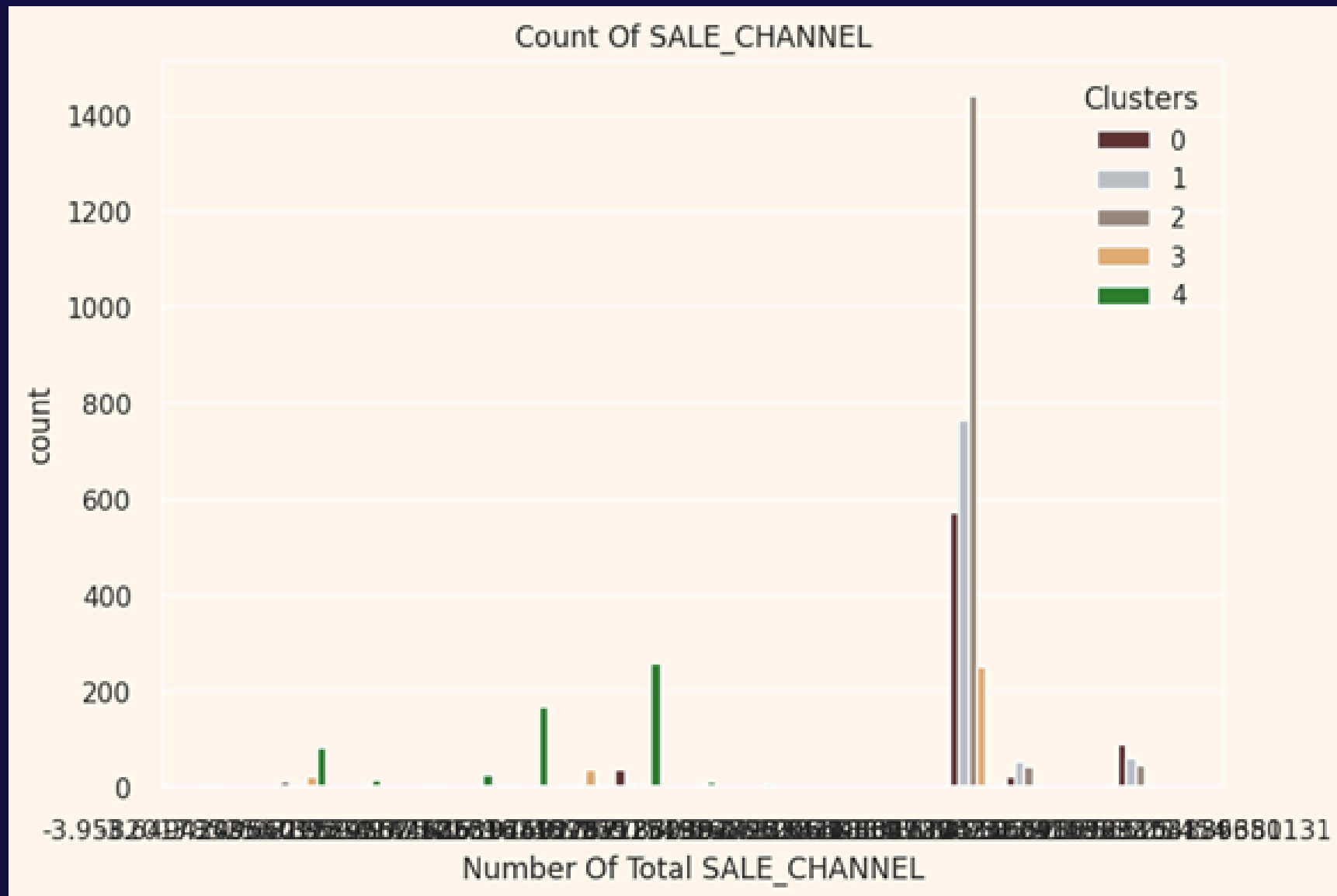
AGE_CATEGORY ve CMPG_TYPE için Kümlerin dağılımı



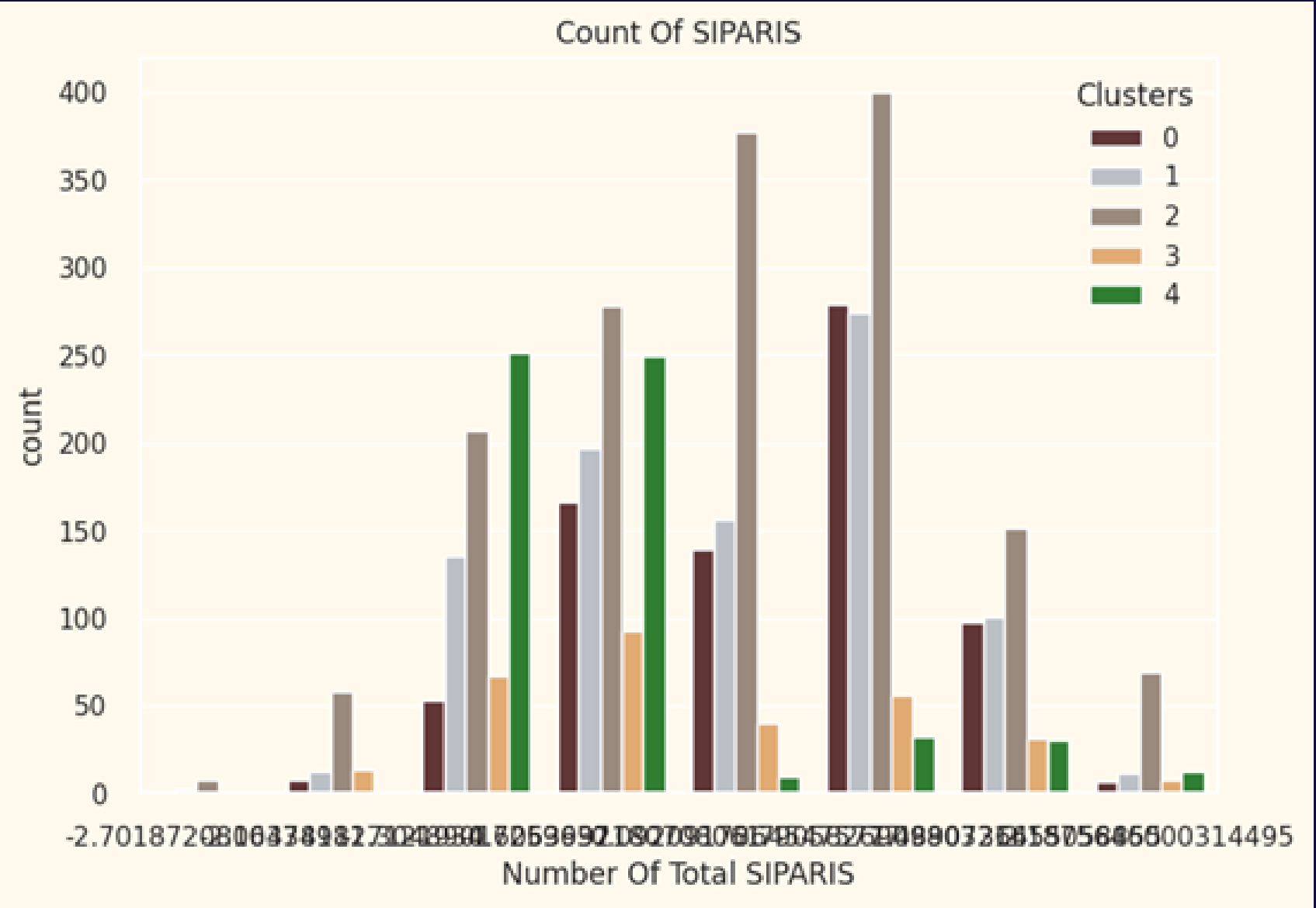
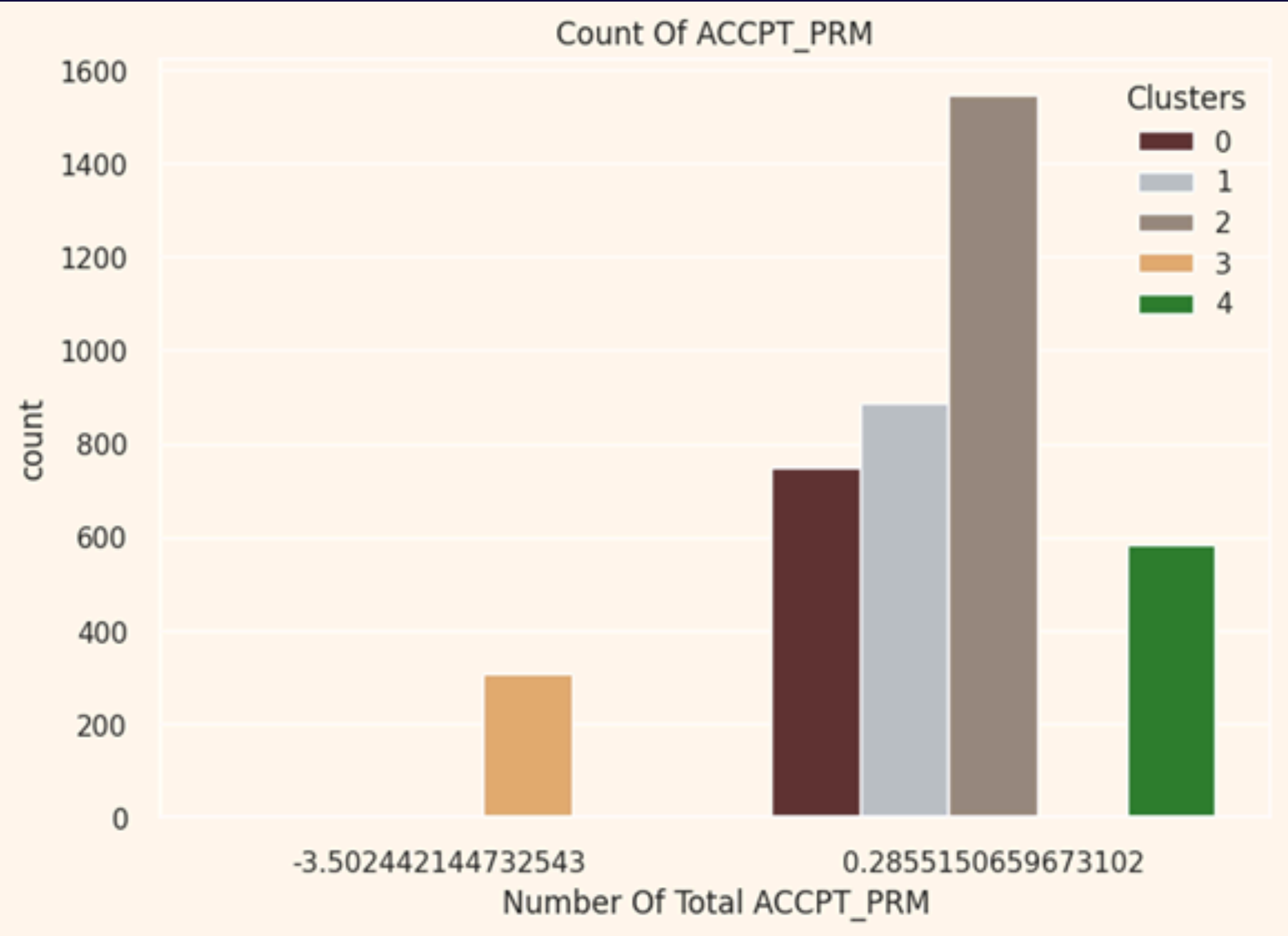
INCOME ve EDUCATION için Kümlerin dağılımı



SALE_CHANNEL ve CHILDREN için Kümlerin dağılımı



ACCPT_PRM ve SIPARIS için Kümlerin dağılımı



Kümelerin Özellikleri

Cluster0

- Genel kampanyalardan faydalanma: Özel kampanyalardan ziyade genel kampanyaları tercih ediyorlar.
- Çeşitli satış kanallarını kullanma: Şube (ttofis) dahil olmak üzere farklı satış kanallarını kullanıyorlar.
- Promosyonları kabul etme: Sunulan promosyonlardan yararlanıyorlar.
- Nakil işlemlerinde aktif: Özellikle nakil işlemleri konusunda daha yoğun bir şekilde hareket ediyorlar.

Cluster1

- Çocuğu olmayanların yoğun olduğu grup: Diğer gruplara göre çocuğu olmayan bireylerin oranı daha yüksek.
- Genel kampanyalardan faydalanma: Özel kampanyalardan ziyade genel kampanyaları tercih ediyorlar.
- Çeşitli satış kanallarını kullanma: Şube (ttofis) dahil olmak üzere farklı satış kanallarını kullanıyorlar.
- Promosyonları kabul etme: Sunulan promosyonlardan yararlanıyorlar.
- Nakil işlemlerinde aktif: Özellikle nakil işlemleri konusunda daha yoğun bir şekilde hareket ediyorlar.

Cluster2

- Orta yaşlıların ağırlıklı olduğu grup: Yaş dağılımı, orta yaşlı bireylerin yoğun olduğu şekilde.
- Yüksek gelirlilerin yoğun olduğu grup: Diğer gruplara göre yüksek gelirli bireylerin oranı daha yüksek.
- Şubeyi tercih etme: Diğer gruplara göre şube (ttofis) üzerinden daha fazla işlem yapıyorlar.
- Promosyonları kabul etme: Sunulan promosyonlardan en fazla yararlanan grup.
- Kampanya katılımında aktif: Özellikle kampanya katılımı ve değişiklik işlemleri konusunda daha yoğun bir şekilde hareket ediyorlar.

Cluster3

- Yaşlıların ağırlıklı olduğu grup: Yaş dağılımı, yaşlı bireylerin yoğun olduğu şekilde.
- Özel kampanyaları tercih etme: Genel kampanyalardan ziyade özel kampanyaları tercih ediyorlar.
- Promosyonları kabul etmeme: Sunulan promosyonlardan yararlanmıyorlar.
- Şubeyi tercih etme: Diğer gruplara göre şube (ttofis) üzerinden işlem yapıyorlar.

Cluster4

- Gençlerin bulunmadığı grup: Yaş dağılımı, genç bireylerin bulunmadığı şekilde.
- Alternatif satış kanallarını tercih etme: Şube (ttofis) yerine online, çağrı merkezi gibi alternatif kanalları tercih ediyorlar.
- Fiber dönüşümünde aktif: Özellikle fiber dönüşüm işlemleri konusunda daha yoğun bir şekilde hareket ediyorlar.

Dinlediğiniz
İçin
Teşekkürler !