



# E-TİCARETTE VERİ MADENCİLİĞİ YAKLAŞIMI

E-TİCARET VERİLERİNİN MÜŞTERİ PROFİLİ AÇISINDAN DEĞERLENDİRİLMESİ

# ÖZET

Günümüzde her sektörde e-ticaret yoğun olarak kullanılmakta ve alışveriş verileri işletmelerin veri tabanlarında saklanmaktadır. Ancak bazı işletmeler için bunlar veri olarak kalırken, bazı işletmeler bu veriyi satış stratejilerini geliştirmek ve müşteri özelliklerini tespit etmek için anlamlı bilgiye dönüştürmektedir. Bu çalışmada, e-ticaret hizmeti veren bir işletmeye ait satış verileri istatistiksel yöntemlerle incelenmiş, alışverişe etki eden kullanıcı özellikleri tespit edilmiş ve müşteriler belirli özelliklere göre kümelemeye tabi tutularak kullanıcılar ve alışveriş alışkanlıkları hakkında bilgiler elde edilmiştir.

# ANA BAŐLIK

Bu alıřmada, IBM SPSS Modeler yazılımı ile bir e-ticaret sitesine ait verilerden iřletme müşterileri hakkında eřitli bilgilerin ve iliřkilerin ortaya ıkartılması amalanmıřtır. Bu amala, **farklı řehirlerde cinsiyete ve yařa gőre hangi űrőnlerin daha ok talep edildiđi ve ka adet űrőn satın alındıđına iliřkin veriler, ok deđiřkenli istatistik yőntemlerinden k-ortalamalar (k-means) yőntemi ile analiz edilmiř ve müşteri profilleri ortaya ıkarılmıřtır.**

# MATERYAL VE METOD

Çalışmada kullanılan veriler bir e-ticaret sitesine ait veriler olup 2013 yılının son 6 ayını kapsamaktadır. Analizde 42188 adet satış verisi, Türkiye'nin farklı şehirlerinde yaşayan kişilerin yaş ve cinsiyetine göre dağılmaktadır. Ürünlerin verilerini sağlayan işletme uygun görmediği için ismini açıklanmamıştır. İşletmeye ait verilerden sadece üç adet ürün grubuna (Kameralar, Spor Giyim, Mutfak Robotu) ilişkin veriler kullanılmıştır.

Veriler IBM SPSS Statistics programından çekilmiş, tanımlayıcı ve karşılaştırmalı istatistikler ve ilişki analizleri ortaya çıkarılmıştır. K ortalamalar yöntemi için veri madenciliği yazılımlarından IBM SPSS Modeler (Clementine) programı kullanılmıştır.

Çalışmada kullanılan değişkenlere ait frekans dağılımları, tanımlayıcı istatistikler, kikare ilişki analizleri, Cramer's v analizi, çoklu uyum analizi yapılmıştır.

**Frekans dağılımı**, verilerin düzenlenmesi, özetlenmesi ve anlaşılır hale getirilmesi için verileri gruplara ayırarak gözlenme sıklıklarını gösteren istatistiksel bir yöntemdir.

**İlişki analizleri**, kullanılan değişkenler arasındaki ilişkilerin belirlenmesinde kullanılır.

**Ki-kare analizi**, frekans dağılımları üzerinden işlem yaparak değişkenlerin bağımsızlığını ölçmede kullanılır.

**Cramer's v**, isimsel iki değişken arasındaki ilişkinin gücü hakkında bilgi verir. Bu katsayı iki değişken arasında ilişki olmadığı zaman 0'a eşit olur. Aksi durumda, Cramer v katsayısı 1'e eşit olsa da, değişkenler arasında yalnızca bir yönde mükemmel bir ilişki vardır (Bölükbaşı & Yıldız, 2009:345).

**Çoklu uyum analizi**, üç ya da daha fazla değişken arasındaki ilişki yapısını inceleyen, değişkenlerin yorumlanmasını kolaylaştıran, çapraz tablolarda satır ve sütunlarda yer alan değişkenler arasındaki ilişkileri grafiksel olarak gösteren bir yöntemdir (Suner & Çelikoğlu, 2010:43).

**K-Means kümeleme**, verilerin özelliklerine göre hiçbir küme bilgisi olmadan her verinin sadece bir kümeye ait olabilmesini sağlayan kümeleme yöntemidir (Davidson, 2002). Bu nedenle, keskin bir kümeleme algoritmasıdır. Küme içi benzerliğin yüksek, fakat kümeler arası benzerliğin düşük olması amaçlanır. Küme benzerliği bir kümedeki nesnelerin ortalama değeri ile ölçülmektedir, bu da kümenin ağırlık merkezidir (Han vd., 2001).

# BULGULAR

**Tablo 1: Yaşa Göre Ürün Dağılımı**

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
Yaş	17 altı	136	21,6	413	65,6	81	12,9
	18-24	1232	19,3	4481	70,3	664	10,4
	25-34	4091	21,5	11388	60,0	3507	18,5
	35-44	2820	24,4	6590	57,0	2145	18,6
	45-54	971	26,8	2011	55,6	638	17,6
	55+	344	33,7	451	44,2	225	22,1
	<b>Toplam</b>	<b>9594</b>	<b>22,7</b>	<b>25334</b>	<b>60,1</b>	<b>7260</b>	<b>17,2</b>

Tablo 1'e göre alışveriş sitesi üzerinden incelenen üç kategori arasından tüm yaş aralıklarında en çok satın alınan ürün spor giyimdir. 55 ve üzerindeki müşterilerin diğer yaş aralıklarına göre alışveriş sitesi üzerinden daha az spor giyim satın aldıkları anlaşılmaktadır.

# BULGULAR

**Tablo 2: Cinsiyete Göre Ürün Dağılımı**

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
Cinsiyet	Erkek	7981	24,0	19971	60,1	5279	15,9
	Bayan	1613	18,0	5363	59,9	1981	22,1
	Toplam	9594	22,7	25334	60,1	7260	17,2

Tablo 2'ye göre hem erkek hem de bayan müşterilerin her ikisinin de çok satın alınan ürün spor giyim olup erkeklerin ikinci olarak kameraları, bayanlarında mutfak robotu satın aldıkları anlaşılmaktadır.



# BULGULAR

**Tablo 3: Bölgelere Göre Ürün Dağılımı**

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
<b>Bölge</b>	Karadeniz	1152	22,9	3096	61,5	785	15,6
	Akdeniz	1055	25,1	2521	60,0	623	14,8
	Marmara	3425	21,7	9473	60,1	2854	18,1
	İç Anadolu	1470	23,2	3724	58,7	1149	18,1
	Doğu Anadolu	583	21,7	1579	58,7	526	19,6
	Güneydoğu Anadolu	351	22,7	938	60,7	256	16,6
	Ege	1516	23,6	3887	60,6	1016	15,8
	Diğer	42	20,1	116	55,5	51	24,4
	<b>Toplam</b>	<b>9594</b>	<b>22,7</b>	<b>25334</b>	<b>60,1</b>	<b>7260</b>	<b>17,2</b>

Tablo 3'e göre tüm bölgelerde de spor giyim öncelikli olarak tercih edilmektedir.

# BULGULAR

**Tablo 4: En Çok ve En Az Satış Yapılan 3 Şehre Göre Ürün Dağılımı**

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
En çok satış yapılan	İstanbul	1918	21,3	5185	57,6	1904	21,1
	Ankara	699	20,8	2023	60,2	640	19,0
	İzmir	614	20,9	1821	62,1	497	17,0
En az satış yapılan	Bayburt	13	22,0	37	62,7	9	15,3
	Siirt	11	14,5	52	68,4	13	17,1
	Ardahan	20	22,7	44	50,0	24	27,3

Tablo 4'e göre en çok satış yapılan şehir İstanbul ve en az satış yapılan şehir Bayburt'dur.

Alışveriş verilerinde yaşı'n ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapabiliriz.



# BULGULAR

**Tablo 5: Müşterilerin Ürün Tipleri ve Yaş Çapraz Tablosu**

Ürün-Yaş		17 altı	18-24	25-34	35-44	45-54	55+	Toplam
<b>Kameralar</b>	n	136	1232	4091	2820	971	344	9594
	Ürün İçindeki %	1,4	12,8	42,7	29,4	10,1	3,6	100,0
	Yaş İçindeki %	21,6	19,4	21,6	24,4	26,8	33,7	22,7
<b>Spor Giyim</b>	n	413	4481	11388	6590	2011	451	25334
	Ürün İçindeki %	1,6	17,7	45,0	26,0	8,0	1,8	100,0
	Yaş İçindeki %	65,5	70,2	60,0	57,0	55,6	44,2	60,1
<b>Mutfak Robotu</b>	n	81	664	3507	2145	638	225	7260
	Ürün İçindeki %	1,1	9,2	48,3	29,5	8,8	3,1	100,0
	Yaş İçindeki %	12,9	10,4	18,4	18,6	17,6	22,1	17,2
<b>Toplam</b>	n	630	6377	18986	11555	3620	1020	42188
	Ürün İçindeki %	1,5	15,1	45,0	27,4	8,6	2,4	100,0
	Yaş İçindeki %	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Tablo 5'e göre müşterilerin % 45'i 25-34 arası yaş aralığında olup % 60'ı spor giyim ürünlerini satın aldığı ve müşterilerin % 27,4'ü 35-44 arası yaş aralığında olup, ürün tercihinde % 57'si spor giyim ürünlerini satın aldığı görülmektedir.

Yaş ile ürün değişkenleri arasında anlamlı bir ilişki olup olmadığını anlamak için Tablo 5'ten elde edilen verilere ki-kare testi uygulayabiliriz.

# BULGULAR

**Tablo 6: Yaş-Ürün Değişkenleri Ki-Kare Testi Sonuçları**

	n	Serbestlik Derecesi	p (Önem düzeyi)
Anlamlılık	548,511	10	0,000
Olabilirlik Oranı	567,923	10	0,000
Doğrusal Bağlantı	3,495	1	0,062
Geçerli Veri Sayısı		42188	

Ki-kare tablosuna göre anlamlılık değeri  $0,000 < 0,05$  olduğu için bireylerin yaş düzeyleri ile aldıkları ürün eğilim faktörü arasında anlamlı bir ilişkinin var olduğu görülmektedir.

Tablo 6’da verilen ki-kare testi sonuçlarına göre ilişkinin anlam düzeyini belirlemek için Cramer’s v testi yapabiliriz.

# BULGULAR

**Tablo 7: Yaş-Ürün Değişkenleri Cramer's v Testi Sonuçları**

	Değer	P (Önem düzeyi)
Phi	0,114	0,000
Cramer's v	0,081	0,000
Veri Sayısı	42188	

Cramer's v testinden elde edilen değerlere göre; yaş ve ürün seçimi arasında ilişki olduğu, ancak ilişkinin derecesinin az olduğu (0,08) görülmektedir. Yaş arttıkça kamera satışlarının arttığı, spor giyim ürünleri satışının azaldığı görülmektedir

Alışveriş verilerinde cinsiyetin ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapabiliriz.

# BULGULAR

**Tablo 8: Müşterilerin Ürün Tipleri ve Cinsiyet Çapraz Tablosu**

Ürün-Cinsiyet		Erkek	Bayan	Toplam
Kameralar	n	7981	1613	9594
	Ürün İçindeki %	83,2	16,8	100,0
	Cinsiyet İçindeki %	24,0	18,0	22,8
Spor Giyim	n	19971	5363	25334
	Ürün İçindeki %	78,8	21,2	100,0
	Cinsiyet İçindeki %	60,1	59,9	60,0
Mutfak Robotu	n	5279	1981	7260
	Ürün İçindeki %	72,7	27,3	100,0
	Cinsiyet İçindeki %	15,9	21,1	17,2
Toplam	n	33231	8957	42188
	Ürün İçindeki %	78,8	21,2	100,0
	Cinsiyet İçindeki %	100,0	100,0	100,0

Tablo 8'e göre müşterilerin % 78,8'i erkek olup bunların % 60,1'ı spor giyim ürünü satın alırken, müşterilerin % 21,2'si bayan olup bunların % 59,9'u spor giyim ürünü satın aldığı görülmektedir.

Cinsiyet ile ürün arasında anlamlı bir ilişki olup olmadığını anlamak için Tablo 8'den elde edilen verilere ki-kare testi uygulayabiliriz.

# BULGULAR

**Tablo 9: Cinsiyet-Ürün Ki-Kare Testi Sonuçları**

	n	Serbestlik Derecesi	p (Önem düzeyi)
Anlamlılık	271,242	2	0,000
Olabilirlik Oranı	268,094	2	0,000
Doğrusal Bağlantı	266,599	1	0,000
Geçerli Veri Sayısı	42188		

Ki-kare tablosuna göre anlamlılık değeri  $0,000 < 0,05$  olduğu için cinsiyet ile aldıkları ürün arasında anlamlı bir ilişkinin var olduğu görülmektedir.

Tablo 9'da verilen ki-kare testi sonuçlarına göre ilişkinin anlam düzeyini belirlemek için Cramer's v testi yapabiliriz.

# BULGULAR

**Tablo 10: Cinsiyet-Ürün Değişkenleri Cramer's v Testi Sonuçları**

	Değer	p(Önem düzeyi)
Phi	0,080	0,000
Cramer's v	0,080	0,000
Veri Sayısı	42188	

Cramer's v testinden elde edilen değerlere göre; cinsiyet ve ürün arasında ilişki olduğu, ancak ilişkinin derecesinin az olduğu (0,08) görülmektedir. Yaş arttıkça kamera satışlarının arttığı, spor giyim ürünlerinin satışının azaldığı görülmektedir. Erkek müşteriler bayanlara göre daha fazla kamera satın alırken, bayan müşterilerin erkeklere göre mutfak robotunu daha fazla satın aldığı görülmektedir.

Alışveriş verilerinde bölgenin ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapabiliriz.



# BULGULAR

**Tablo 11: Müşterilerin Ürün Tipleri ve Yaşadığı Bölge Çapraz Tablosu**

Ürün-Cinsiyet		A	B	C	D	E	F	G	Toplam
Kameralar	n	1152	1055	3425	1470	583	351	1516	9552
	%*	12,0	11,0	35,9	15,4	6,1	3,7	15,9	100,0
	%**	22,9	25,1	21,7	23,2	21,7	22,7	23,6	22,8
Spor Giyim	n	3096	2521	9473	3724	1579	938	3887	25218
	%*	12,3	10,0	37,6	14,8	6,2	3,7	15,4	100,0
	%**	61,5	60,0	60,1	58,7	58,7	60,7	60,6	60,0
Mutfak Robotu	n	785	623	2854	1149	526	256	1016	7209
	%*	10,9	8,8	39,5	15,9	7,3	3,5	14,1	100,0
	%**	15,6	14,9	18,2	18,1	19,6	16,6	15,8	17,2
Toplam	n	5033	4199	15752	6343	2688	1545	6419	41979
	%*	12,0	10,0	37,5	15,1	6,4	3,7	15,3	100,0
	%**	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

\* Ürün İçerisinde, \*\* Bölge İçerisinde

A: Karadeniz Bölgesi, B: Akdeniz Bölgesi, C: Marmara Bölgesi, D: İç Anadolu Bölgesi, E: Doğu Anadolu Böl.  
F: Güneydoğu Anadolu Böl. G: Ege Bölgesi

Alışveriş verilerinde bölge ve ürün seçimi etkileşiminin sonuçlarına göre; müşterilerin % 37,5'i Marmara bölgesinden olup ürün tercihinde % 60,1'i spor giyim % 21,7'si kameraları tercih etmektedir. Müşterilerin % 15,3'ü Ege bölgesinden olup ürün tercihinde % 60,6'sı spor giyim % 23,6'sı kameraları tercih etmektedir.

Bölge ile ürün değişkenleri arasında anlamlı bir ilişki olup olmadığını anlamak için Tablo 11'den elde edilen verilere ki-kare testi uygulanmıştır.

# BULGULAR

**Tablo 12: Bölge-Ürün Değişkenleri Ki-Kare Testi Sonuçları**

	n	Serbestlik Derecesi	p(Önem düzeyi)
Anlamlılık	534,300	162	0,000
Olabilirlik Oranı	529,231	162	0,000
Doğrusal Bağlantı	0,314	1	0,575
Geçerli Veri Sayısı		41979	

Ki-kare tablosuna göre anlamlılık değeri  $0,000 < 0,05$  olduğu için bölge ile bireylerin aldıkları ürün eğilim faktörü arasında anlamlı bir ilişkinin var olduğu görülmektedir. Tablo 12’de verilen ki-kare testi sonuçlarına göre ilişkinin anlam düzeyini belirlemek için Cramer’s v testi yapılmış.

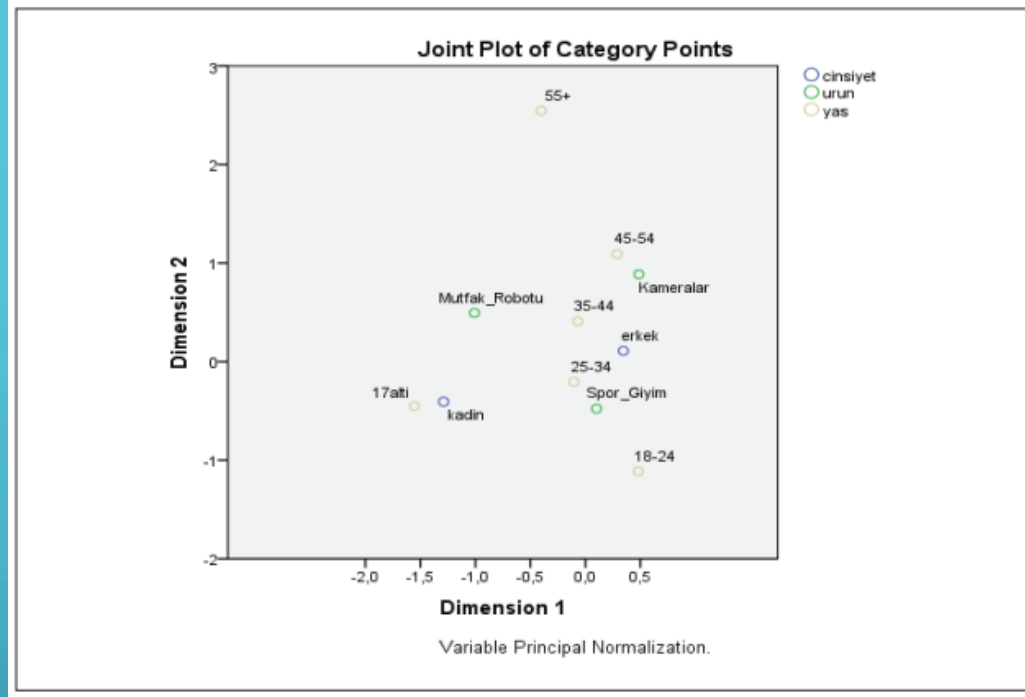
# BULGULAR

**Tablo 13: Bölge-Ürün Değişkenleri Cramer's v Testi Sonuçları**

	Değer	p (Önem düzeyi)
Phi	0,113	0,000
Cramer's V	0,080	0,000
Veri Sayısı	42188	

Cramer's v testinden elde edilen değerlere göre; bölge ve ürün seçimi arasında düşük seviyede ilişki olduğu (0,08) görülmektedir. Bölgelerin gelişmişlik düzeyi ile e-ticaret üzerinden yapılan alışveriş arasında doğrusal bir bağlantı olduğu görülmektedir. Tabi ki bunda bölgenin nüfus miktarı da ayrı bir etkindir. Ayrıca tüm bölgelerde spor giyim ürünlerinin satın alma oranı en yüksek olup, en düşük satın alma oranının mutfak robotunda olduğu görülmektedir.

**Şekil 1: Cinsiyet, Ürün ve Yaş Değişkenleri İçin Çoklu Uyum Analizi**

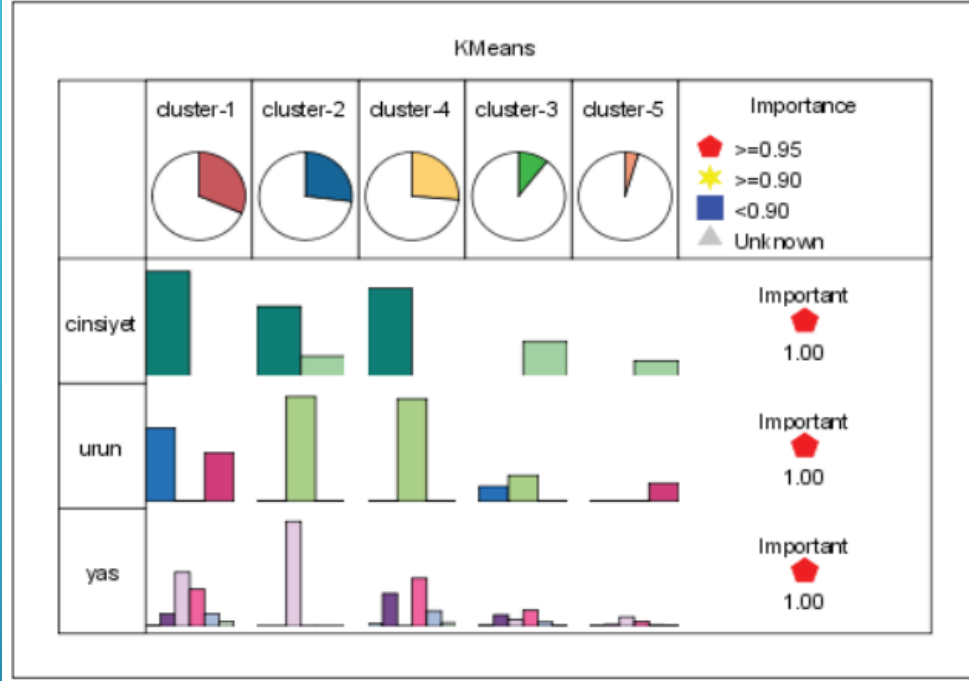


Şekil 1 incelendiğinde; ürün alışverişinde 18-24 ve 25-34 yaş arasındaki erkek müşterilerin spor giyimi tercih ettikleri, 35-44 yaş arasındaki erkeklerin ise kameraları tercih ettikleri görülmüştür. Aynı şekilde de 17 yaş altı ve 18-24 yaş arasındaki kadınlarında spor giyimi tercih ettikleri, 35-44 yaş arasındaki kadınların ise mutfak robotu tercih ettikleri görülmüştür. Ürün tercihi öncelikli spor giyim ürünlerinde, sonrasında kameralarda yoğunlaşmıştır.

Çalışma verileri müşteri ve ürün özelliklerine k-means ile kümelemeye tabi tutulduğunda Şekil 2’de gösterildiği gibi 5 küme elde edilmiştir. Yaşanılan şehrin satın alınan ürün miktarında önemli olmadığı analiz sonucunda ortaya çıkmıştır.

# BULGULAR

Şekil 2. K-Means Kümeleme Analizi Sonuçları



Tablo 14: Elde Edilen 5 Kümeye Ait Özellikler

		1 Küme	2 Küme	3 Küme	4 Küme	5 Küme
Küme Kayıt Sayısı		13260	11388	4430	11129	1981
Cinsiyet	Erkek %	100	77,64	0	100	0
	Bayan %	0	22,36	100	0	100
Ürün	Kameralar %	60,19	0	36,41	0	0
	S. Giyim %	0	100	63,59	100	0
	M. Robotu %	39,81	0	0	0	100
Yaş	17 altı %	1	0	3,36	2,78	2,02
	18-24 %	10,32	0	28,31	31,91	10,3
	25-34 %	44,35	100	16,14	0	50,58
	35-44 %	30,42	0	39,5	47	27,26
	45-54 %	10,22	0	10,5	14,87	7,32
	55+ %	3,7	0	2,19	3,44	2,52

Elde edilen kümeleme sonuçlarına göre;

1. kümede, yaş aralığı 25-34 olan erkeklerin kamera satın aldıkları,
2. kümede, yaş aralığı 25-34 olan erkeklerin spor giyim ürünleri satın aldıkları,
3. kümede, yaş aralığı 35-44 olan bayanların spor giyim ürünleri satın aldıkları,
4. kümede, yaş aralığı 35-44 olan erkeklerin spor giyim ürünleri satın aldıkları,
5. kümede, yaş aralığı 25-34 olan bayanların mutfak robotu satın aldıkları görülmüştür.



# BULGULAR

**Tablo 15: K-Means Kümeleme Analizinde Oluşan İterasyon ve Hatalar**

İterasyon	Hata
1	0,846
2	0,571
3	0,248
4	0,237
5	0,000

# SONUÇ

Elde edilen sonuçlara göre, müşterilerin büyük çoğunluğu spor giyim ürünlerini web sitesi üzerinden satın almayı tercih etmişken, en az tercih edilen ürün grubun mutfak robotu olduğu görülmüştür. Müşterilerin ürün tercihinde cinsiyet ve yaş bilgisi önem arz ederken, yaşadığı şehir bir anlam ifade etmemektedir. Şehir bilgisi ürün tercihi için anlamlı olmamasına karşın, müşteri potansiyeli açısından bilgi vermekte ve en çok satış yapılan bölge Marmara bölgesi olurken, en az satış yapılan bölge Güneydoğu Anadolu bölgesi olmuştur.



ALİ KEREM ŞİMŞEK  
9501228951