

New Mind AI Bootcamp Ödevde Elde Edilen Bulguların

Yorumlanması

Hazırlayan: Altan Topbaş

Veri setleri yüklenip incelendiğinde satış verisinde yer alan bazı sütunlarının veri tipi, ilgili sütunla uyumlu olmadığı gözlemlenmiştir.

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	5000 non-null	int64
1	tarih	5000 non-null	object
2	ürün_kodu	5000 non-null	object
3	ürün_adi	5000 non-null	object
4	kategori	5000 non-null	object
5	fiyat	5000 non-null	object
6	adet	5000 non-null	int64
7	toplam_satis	5000 non-null	object
8	musteri_id	5000 non-null	int64

Şekil 1: Satış Verisi Bilgi Tablosu

Satış veri seti incelendiğinde **şekil 1**'de de görüldüğü üzere “fiyat” ve “toplam_satis” sütunlarındaki ifadelerin veri tipi “object” olarak gözükmemektedir. Bu sütunlarda yer alan ifadeler numerik ifadeler olduğu için int64 veya float64 olarak gözükmesi beklenmektedir. Bu sütunları float veri tipine dönüştürmek için çeşitli işlemler yapılmak mecburiyetindedir. Bu işlemler aşağıda verilen kodlarla gerçekleştirilmiştir.

```
satis_veri["fiyat"] = pd.to_numeric(satis_veri["fiyat"],
errors="coerce")
satis_veri["toplam_satis"] =
pd.to_numeric(satis_veri["toplam_satis"], errors="coerce")
```

Veri tipi dönüşümünden sonra her iki veri seti için eksik verinin olup olmadığının kontrolü **şekil 2**'de yapılmıştır.

```
Müşteri Veri Setinde, Sütunlardaki Toplam Eksik Veri Sayısı
musteri_id      0
isim            0
cinsiyet        0
yas             0
sehir           0
harcama_miktari 0

Satış Veri Setinde, Sütunlardaki Toplam Eksik Veri Sayısı
Unnamed: 0      0
tarih           0
ürün_kodu       0
ürün_adi        0
kategori        0
fiyat           42
adet            0
toplam_satis    4
musteri_id      0
```

Şekil 2: Veri setleri için yapılan eksik veri kontrolü

Veri setleri için yapılan eksik veri kontrolünde müşteri veri setinde bir eksik görülmezken satış veri setinde yer alan “fiyat” ve “toplam_satis” sütunlarında sırasıyla 42 ve 4 adet eksik değer gözükmemektedir. Verinin analizinin doğru olabilmesi için bu eksik veri değerlerinin doldurulması gerekmektedir.

Eksik verilerin doldurulması için median() metodu kullanılmıştır. Medyan ile doldurulma işlemi sonrası veri setinde eksik verinin olup olmadığı tekrar incelenmiştir. İnceleme sonucunda eksik veri olmadığı gözlemlenmiştir. Bu durum **şekil 3**'te gösterilmektedir.

```

>>> satis_veri.isnull().sum()
Unnamed: 0      0
tarih           0
Ürün_kodu       0
Ürün_adi        0
kategori        0
fiyat           0
adet           0
toplam_satis    0
musteri_id      0

```

Şekil 3: Satış verisinin eksik değerler doldurulduktan sonra incelenmesi

Bir sonraki aşamada satış veri setindeki aykırı değerlerin tespiti ve belirli bir sınır aralığına çekilmesi işlemi yapılmıştır. Numerik değerler incelendiğinde elde edilen sonuç **Şekil 4**'te gösterilmiştir.

```

Harcama Miktarı Sütunu Aykırı Değerlerin Sayısı:
0
-----
Harcama Miktarı Sütunu Aykırı Değerler:
Empty DataFrame
Columns: [musteri_id, isim, cinsiyet, yas, sehir, harcama_miktari]
Index: []
-----
Fiyat Sütunu Aykırı Değerlerin Sayısı:
0
-----
Fiyat Sütunu Aykırı Değerler:
Empty DataFrame
Columns: [Unnamed: 0, tarih, Ürün_kodu, Ürün_adi, kategori, fiyat, adet, toplam_satis, musteri_id]
Index: []
-----
Toplam Satış Sütunu Aykırı Değerlerin Sayısı:
467
-----

```

Şekil 4: Aykırı değerler sayısı

Aykırı değerler incelendiğinde yalnızca toplam_satis sütununda aykırı değerler olduğu görülmüştür. Veri tipi dönüşümünden sonra oluşan bu aykırı değerlerin belirli bir sınır aralığına sokulması işlemi yapılmıştır. Sınırlama işlemi, alt sınırın altındaki değerler alt sınıra, üst sınırın üstündekiler üst sınıra çekilme olarak yapılmıştır.

Bu ifadelerin ne olduğu ve sınırlandırıldıktan sonra nasıl bir ifadeye dönüştüğü **Şekil 5**'te gösterilmiştir.

Toplam Satış Sütunu Aykırı Değerler (İlk 10 Veri):

Unnamed: 0	tarih	ürün_kodu	...	adet	toplam_satis	musteri_id
3	2024-04-11	P063	...	9	1.330092e+16	2869
15	2024-06-08	P082	...	6	7.328220e+15	4666
27	2024-03-31	P082	...	3	1.262700e+16	7110
31	2023-11-08	P063	...	3	3.426450e+16	9142
40	2024-05-24	P010	...	18	6.339960e+15	7099
47	2024-02-18	P032	...	13	1.604967e+16	4787
66	2023-11-19	P097	...	5	6.222200e+15	1521
81	2023-06-17	P087	...	14	4.227860e+15	9526
82	2023-07-16	P087	...	6	8.189220e+15	8900
88	2024-06-28	P063	...	3	1.450800e+16	3142

TOPLAM SATIŞ SÜTUNUNUN AYKIRI DEĞERLERİ SINIR DEĞERLERİNE GETİRİLMİŞ HALİ (İLK 10 VERİ):

Unnamed: 0	tarih	ürün_kodu	...	adet	toplam_satis	musteri_id
0	2023-07-02	P032	...	13	15694.9000	9225
1	2024-02-11	P011	...	16	3794.2400	3955
2	2023-03-06	P024	...	9	9218.6100	2367
3	2024-04-11	P063	...	9	31771.2575	2869
4	2024-01-08	P063	...	19	18039.1700	9724
5	2023-06-20	P037	...	10	2664.0000	7948
6	2023-08-22	P082	...	8	10322.4000	4184
7	2024-05-11	P063	...	17	10672.2600	1522
8	2024-07-26	P010	...	8	5995.1200	5504
9	2023-03-15	P082	...	16	3098.4000	1722

Şekil 5: Aykırı değerlerin ve düzeltilmiş değerlerin ilk 10 verisi

Bir sonraki işlemde müşteri veri seti ile satış veri seti müşteri_id sütununa göre inner ile birleştirilip ayrı bir veri seti elde edilme işlemi yapılmıştır. Bu işlemin sonucu ilk 20 veri olacak şekilde **Şekil 6**'da gösterilmiştir.

```

>>> satis_veri = satis_veri_capped.copy()
>>> merged_df = pd.merge(musteri_veri, satis_veri, on='musteri_id', how='inner')
>>> merged_df.head(20)

```

	musteri_id	isim	cinsiyet	...	fiyat	adet	toplam_satis
0	3902	Candeniz	Erkek	...	898.53	18	16173.5400
1	3902	Candeniz	Erkek	...	508.58	6	3051.4800
2	2080	Gözel	Erkek	...	1454.80	5	7274.0000
3	2080	Gözel	Erkek	...	369.09	4	1476.3600
4	7901	Erdogan	Kadın	...	913.23	3	2739.6900
5	7901	Erdogan	Kadın	...	865.58	10	31771.2575
6	8958	Sonad	Kadın	...	173.90	13	31771.2575
7	8958	Sonad	Kadın	...	935.69	2	1871.3800
8	8441	Merdî	Kadın	...	522.53	3	1567.5900
9	9843	Ünübol	Kadın	...	147.99	19	31771.2575
10	9843	Ünübol	Kadın	...	233.91	6	1403.4600
11	9843	Ünübol	Kadın	...	142.07	17	2415.1900
12	3142	Özpetek	Erkek	...	48.36	3	31771.2575
13	3142	Özpetek	Erkek	...	1357.56	16	21720.9600
14	3142	Özpetek	Erkek	...	92.65	4	370.6000
15	9590	Derkay	Erkek	...	1188.23	14	16635.2200
16	9590	Derkay	Erkek	...	518.99	1	518.9900
17	4281	Zekayi	Kadın	...	860.08	7	6020.5600
18	4281	Zekayi	Kadın	...	54.52	15	31771.2575
19	4281	Zekayi	Kadın	...	1444.58	11	15890.3800

Şekil 6: müşteri_id sütununa göre birleştirilme işlemi

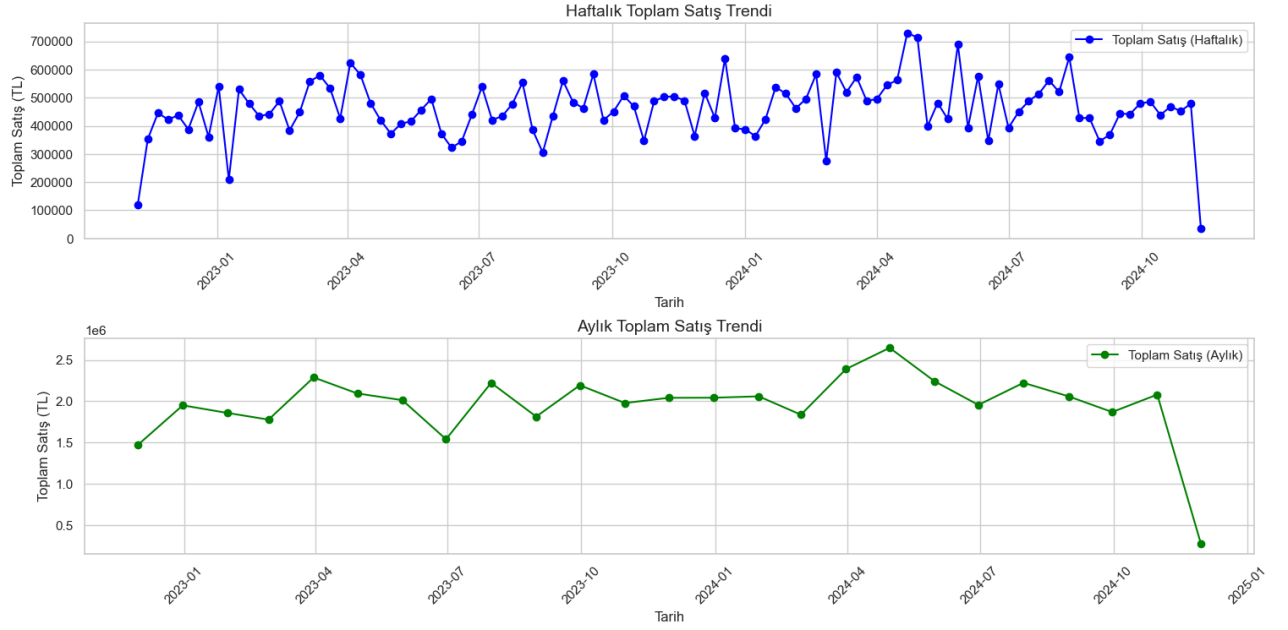
Bir sonraki aşamada, satış verisi üzerinde haftalık ve aylık bazda toplam satış ve ürün satış trendleri analiz edilmiştir. Elde edilen verilerin gösteriminde okuma kolaylığı olması için format değişikliği yapılmıştır. Sonucun yazı olarak gösterilmesi **şekil 7**'de yapılmıştır.

```
Haftalık Toplam Satışlar:
      tarih  adet  toplam_satis
0  2022-11-07  129    118,656.37
1  2022-11-14  437    353,029.33
2  2022-11-21  450    445,590.03
3  2022-11-28  444    422,481.71
4  2022-12-05  502    437,332.05
..      ...    ...      ...
101 2024-10-14  581    438,573.36
102 2024-10-21  445    468,354.52
103 2024-10-28  615    452,138.53
104 2024-11-04  433    478,638.86
105 2024-11-11   41     34,410.89

[106 rows x 3 columns]
Aylık Toplam Satışlar:
      tarih  adet  toplam_satis
0  2022-11-30  1618  1,470,211.86
1  2022-12-31  2189  1,950,656.51
2  2023-01-31  2217  1,857,791.51
3  2023-02-28  2003  1,778,120.85
4  2023-03-31  2210  2,287,971.71
5  2023-04-30  2050  2,094,617.64
6  2023-05-31  2009  2,013,054.77
7  2023-06-30  1507  1,540,298.20
```

Şekil 7: Haftalık ve Aylık Toplam Satışlar

Trend analizi yapılabilmesi için grafikleştirme işlemi yapılmıştır. **Şekil 8**'de bu grafik gösterilmiştir.



Şekil 8: Haftalık ve Aylık Toplam Satışlar Grafiği

Grafik incelendiğinde belirli dönemler de yükseliş gözükürken belirli dönemlerde yükseliş gözükmemektedir. Grafik genel olarak yorumlandığında bir artış söz konusu olmaktadır. Özellikle 2024-04 tarihinde zirve noktasına ulaşan toplam satışların 2024-11 tarihinde en düşük noktaya düştüğü dikkat çekmektedir. 2024-11 Tarihinde satış verilerinin az olması sebebiyle ani bir düşüşün yaşandığı gözlemlenmiştir.

Bir sonraki aşamada, tarih sütununu kullanarak, her ayın ilk ve son satış günlerini bulunmuştur. Ayrıca, her hafta kaç ürün satıldığını hesaplanmıştır. Elde edilen sonuçlar **şekil 9** ve **şekil 10**'da gösterilmiştir.

Ayın İlk Satış Günü ve Ayın Son Satış Günü			
	tarih	ayin_ilk_satis_gunu	ayin_son_satis_gunu
0	2022-11	2022-11-06	2022-11-30
1	2022-12	2022-12-01	2022-12-31
2	2023-01	2023-01-01	2023-01-31
3	2023-02	2023-02-01	2023-02-28
4	2023-03	2023-03-01	2023-03-31
5	2023-04	2023-04-01	2023-04-30
6	2023-05	2023-05-01	2023-05-31
7	2023-06	2023-06-01	2023-06-30
8	2023-07	2023-07-01	2023-07-31
9	2023-08	2023-08-01	2023-08-31
10	2023-09	2023-09-01	2023-09-30
11	2023-10	2023-10-01	2023-10-31
12	2023-11	2023-11-01	2023-11-30
13	2023-12	2023-12-01	2023-12-31
14	2024-01	2024-01-01	2024-01-31
15	2024-02	2024-02-01	2024-02-29
16	2024-03	2024-03-01	2024-03-31
17	2024-04	2024-04-01	2024-04-30

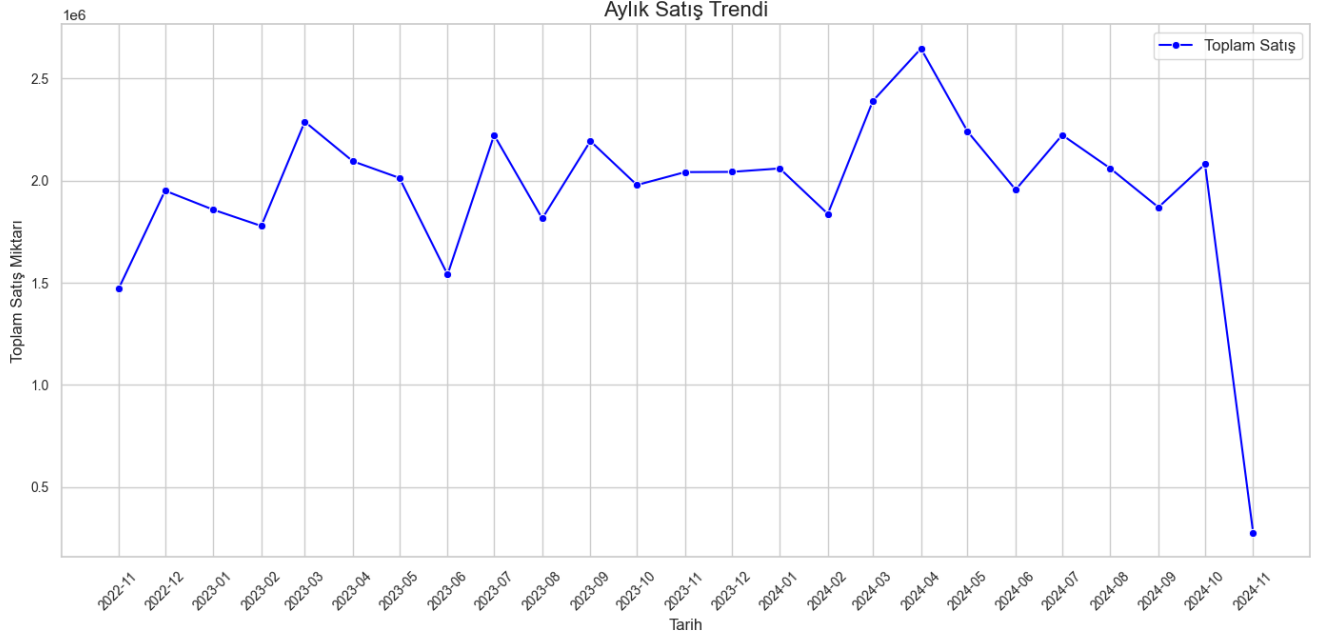
Şekil 9: Her ayın ilk ve son satış günleri

Hafta Başlangıcı ve Haftalık Satış Adedi		
	hafta_baslangici	haftalik_satis_adi
0	2022-11-07	129
1	2022-11-14	437
2	2022-11-21	450
3	2022-11-28	444
4	2022-12-05	502
..
101	2024-10-14	581
102	2024-10-21	445
103	2024-10-28	615
104	2024-11-04	433
105	2024-11-11	41

Şekil 10: Haftalık olarak satılan ürün adedi

Şekil 10'da gösterilen değerler incelendiğinde genellikle haftanın başlangıcında gerçekleşen satış adetlerinin bir önceki haftaya kıyasla artış gösterdiği söylenebilmektedir.

Bir sonraki aşamada, zaman serisindeki trendleri tespit etmek için grafikler çizdirilmiştir (Örneğin: aylık satış artışı veya düşüşü).



Şekil 11: Aylık satış artışı ve düşüşü grafiği

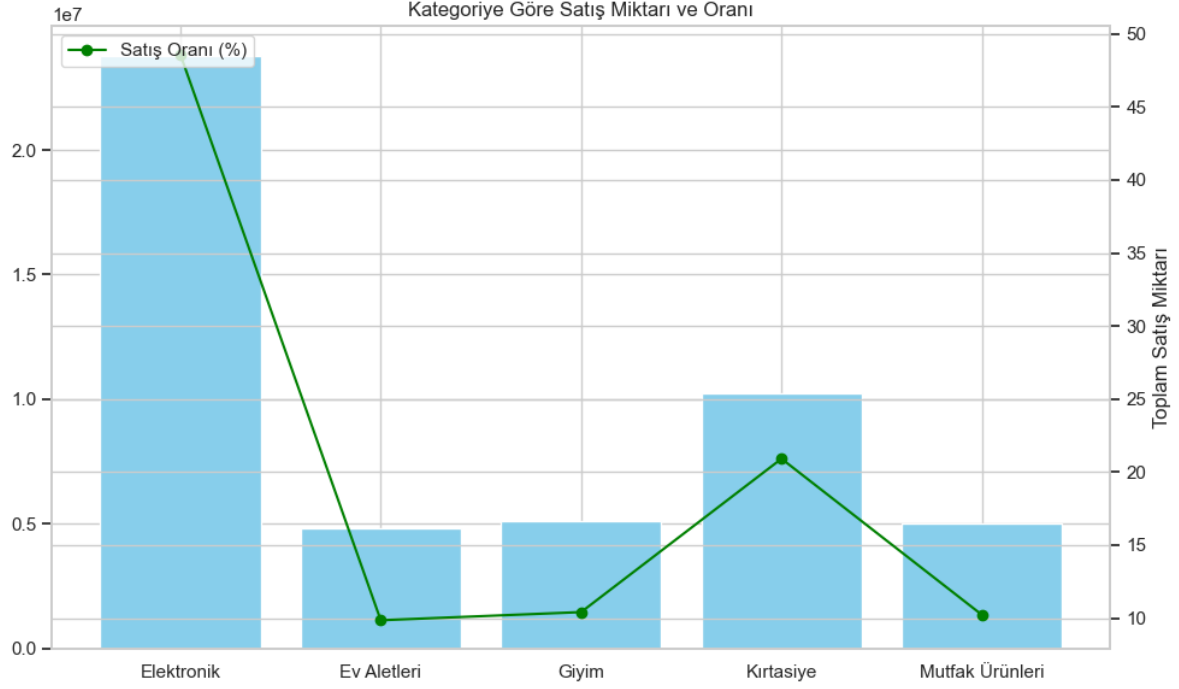
Şekil 11'de aylık satış trendinin grafiği gösterilmiştir. Bu grafik incelendiğinde, 2022 yılının son ayları ve 2023 yılının ilk ayları arasında bir artış gözlemlenirken 2023 yılının orta aylarına kadar biraz düşüş gerçekleştirdiği görülmektedir. Burada satışın az olduğu çıkarımı yapılabilir. Haziran ayından temmuz ayına geçerken büyük bir yükseliş olduğu görülmüştür. Bu yükselişin sebebi, yaz aylarında evlilik, tatil gibi özel durumlar için yapılan alışverişin fazlalığından dolayı olduğu düşünülebilir. Aynı durum 2024'ün bahar ayı olan Nisan ayı için de geçerli olmaktadır. 2024 Yılında kasım ayındaki satış verisinin azlığından dolayı ani bir düşüş gerçekleşmektedir.

Bir sonraki aşamada, ürün kategorilerine göre toplam satış miktarının ve her kategorinin tüm satışlar içindeki oranının hesaplanması yapılmıştır. **Şekil 12**'de toplam satış miktarı ve satış oranı (yüzdesel) olarak ifade edilmiştir.

	Toplam Satış Miktarı	Satış Oranı (%)
kategori		
Elektronik	23,773,443.51	48.59
Ev Aletleri	4,820,961.40	9.85
Giyim	5,094,093.19	10.41
Kırtasiye	10,233,168.01	20.92
Mutfak Ürünleri	5,003,604.24	10.23

Şekil 12: Toplam satış miktarı ve satış oranı (%)

Şekil 13'de ise bu durum grafik çizilerek incelenmiştir.



Şekil 13: Toplam satış miktarı ve satış oranı (%) grafiği

Şekil 13’de gösterilen yeşil noktalı çizgi satış oranını ifade ederken mavi sütun çubukları toplam satış miktarını ifade etmektedir.

Elde edilen bulgular incelendiğinde toplam satışların neredeyse yarısını “Elektronik” kategorisinin oluşturduğu gözlemlenmiştir. Bu kategorinin içeriğinin oldukça geniş oluşu ve tüm kitleye hitap edişini göz önüne alacak olursak bu durumun beklenen bir durum olduğunu söyleyebiliriz.

Bir sonraki aşamada, müşterilerin yaş gruplarına göre satış eğilimleri analiz edilmiştir. (Örnek yaş grupları: 18-25, 26-35, 36-50, 50+)

Analizin sonucu **şekil 14**’te gösterilmiştir.

```
>>> print(yas_grubu_satis_analizi)
```

	Toplam Harcama Miktarı	Harcama Oranı (%)
yas_grubu		
18-25	1,733,410.01	13.57
26-35	2,381,243.95	18.64
36-50	3,696,869.33	28.94
50+	4,963,271.78	38.85

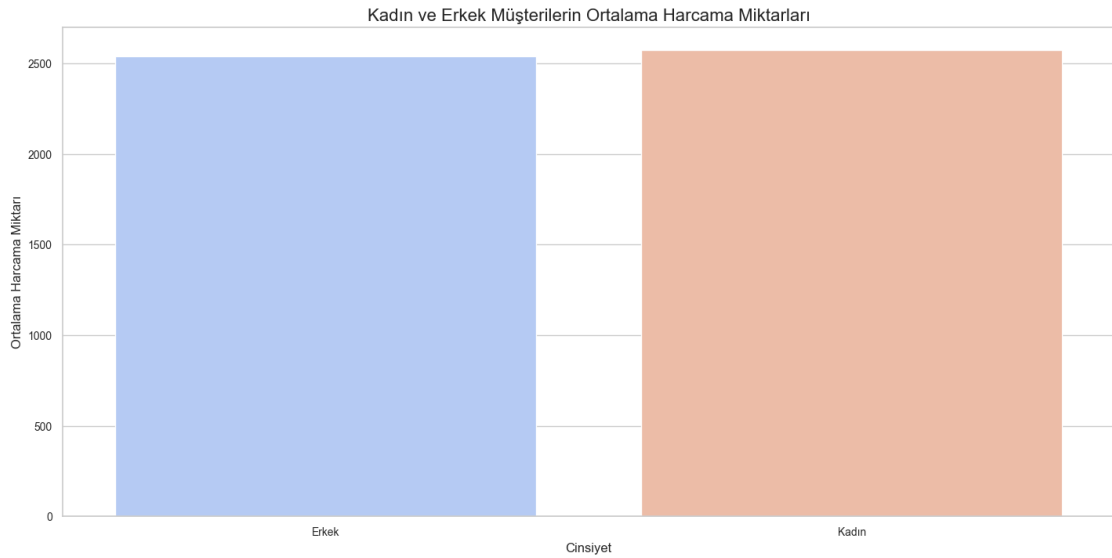
Şekil 14: Yaş grubuna göre toplam harcama miktarı ve harcama oranı

Elde edilen bu bulguya göre bu veri setinde en çok harcama yapanların yaşlarının 50 yaş üstü olduğu görülmüştür. Tüm veri setinin %38.85’ini oluşturan 50 yaş üstü müşterilerin en çok harcama miktarına sahip oldukları gözlemlenmiştir.

Bir sonraki aşamda, kadın ve erkek müşterilerin harcama miktarları karşılaştırılıp harcama davranışları arasındaki fark tespit edilmiştir. Elde edilen çıktı, **şekil 15**'te gösterilmiştir. Bunun yanı sıra bu sonucun grafik üzerinden gösterilmesi de **şekil 16**'da gösterilmiştir. Şekil 16'da ki mavi çubuk erkekler iken diğer çubuk kadınları göstermektedir.

cinsiyet	Ortalama Harcama Miktarı
0 Erkek	2,537.61
1 Kadın	2,572.35

Şekil 15: Kadın ve erkek müşterilerin harcama miktarları



Şekil 16: Kadın ve erkek müşterilerin harcama miktarları grafiği

Elde edilen sonuçlara göre kadınlar ve erkeklerin arasında harcama miktarı olarak çok bir fark olduğu söylenememektedir. Kadınların az da olsa erkeklere göre harcama miktarının biraz yüksek olduğu görülmüştür.

Bir sonraki aşamada, müşterilerin şehir bazında toplam harcama miktarı bulunmuş ve şehirleri en çok harcama yapan müşterilere göre sıralanmıştır. **Şekil 17**'de elde edilen sonuç görüntüsü verilmiştir. En çok satış gerçekleşen şehirler çoktan aza olacak şekilde yukardan aşağıya doğru sıralanmıştır.

	sehir	harcama_miktari
0	İzmir	1,719,480.53
1	Gaziantep	1,680,765.15
2	Ankara	1,632,071.24
3	Konya	1,604,023.46
4	Adana	1,567,156.27
5	İstanbul	1,544,511.23
6	Bursa	1,543,966.94
7	Antalya	1,482,820.25

Şekil 17: En çok satış yapılan şehirler

Şekil 17 incelendiğinde en çok İzmir ilinde satışın gerçekleştiği görülmüştür. Ancak diğer kıyaslamalarda olduğu gibi farklar görülmemektedir. Veri seti içerisindeki tüm illerin hemen hemen birbirine yakın miktarda harcama miktarına sahip oldukları görülmüştür. Bu 8 ilin de yaklaşık olarak eşit dağılımda harcama yaptıkları söylenebilmektedir.

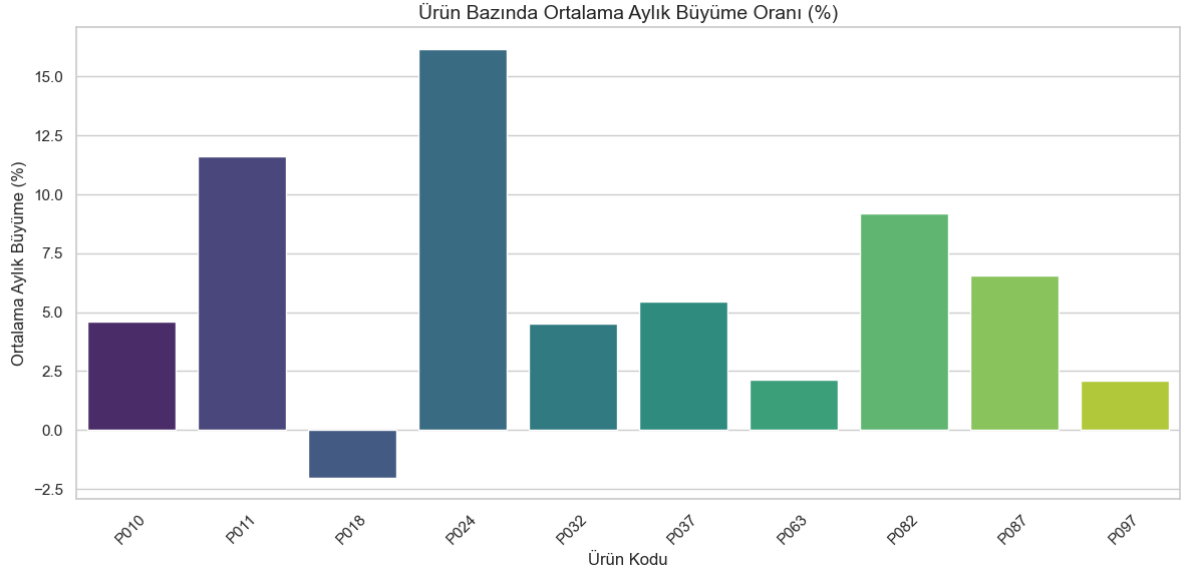
Bir sonraki aşamada satış verisinde her bir ürün için ortalama satış artışı oranı hesaplanmıştır. Bu oranı hesaplamak için her bir üründe önceki aya göre satış değişim yüzdesi kullanılmıştır. Ortalama olarak satış artışı veya azalışının sonucu **şekil 18**'de normal olarak verilirken **şekil 19**'da grafik olarak verilmiştir.

	ürün_kodu	ortalama_buyume
0	P010	4.62
1	P011	11.63
2	P018	-2.02
3	P024	16.17
4	P032	4.52
5	P037	5.44
6	P063	2.13
7	P082	9.21
8	P087	6.58
9	P097	2.10

Şekil 18: Ortalama satış artışı veya azalışı

Değişim oranının nasıl bulunduğu aşağıdaki formülde gösterilmiştir.

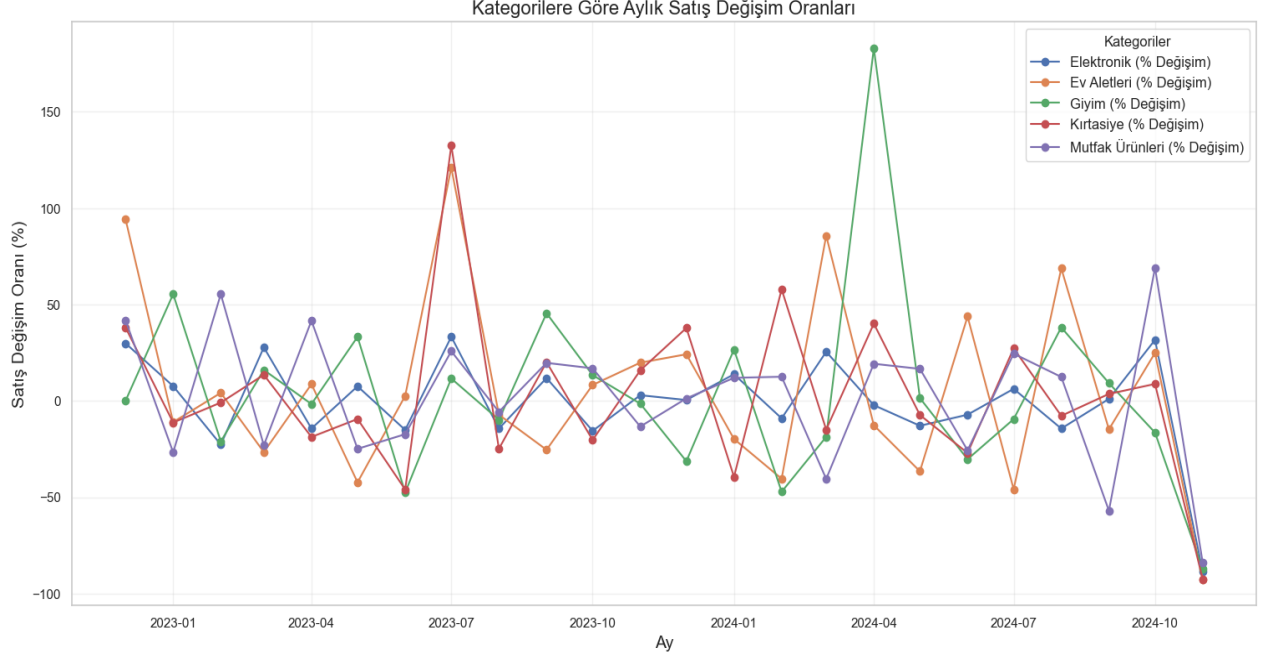
$$\text{Değişim Oranı (\%)} = \frac{\text{Bu Ayın Satışı} - \text{Önceki Ayın Satışı}}{\text{Önceki Ayın Satışı}} \times 100$$



Şekil 19: Ortalama satış artış veya azalış grafiği

Değişim oranı formülü uygulandığında P018 değerinin eksi bir değer geldiği gözlemlenmiştir. Satış oranında azalma meydana gelebileceği düşünülmektedir. Bunun yanı sıra P024 kodlu ürünün satış oranının en fazla satış oranına sahip olduğu gözlemlenmiştir. İki veri de elektronik kategorisine ait olmasına rağmen büyük bir fark gözlemlenmiştir. Bu durum için çeşitli yorumlamalar yapılabilir. Kitle gibi kriter göz önünde bulundurulduğunda P024 kodlu elektronik cihazın diğer elektronik cihazlara kıyasla çok daha ihtiyaç duyulduğu gözlemlenmiştir.

Bir sonraki aşamada, pandas groupby ile her bir kategorinin aylık toplam satışlarını hesaplayıp değişim oranlarının grafikte gösterilmesi yapılmıştır. Elde edilen grafik **şekil 20**'de gösterilmektedir.



Şekil 20: Kategorilere göre aylık satış değişim oranları

İlgili grafik incelendiğinde en çok değişimin giyim kategorisinde gerçekleştiği görülmüştür. Elektronik kategorisi en çok satışa sahip bir kategori olsa bile stabil bir artış söz konusu olmaktadır. Ancak giyim kategorisinde gözle görülür bir biçimde azalış ve artış söz konusu olmaktadır. Özellikle bahar ayının gelmesiyle beraber giyim kategorisindeki azalış, bir anda zirve yaparak tüm kategoriler arasında yüzdesel olarak en yüksek artışa dönüşmüştür. 2023 Yılı'nın temmuz ayında kırtasiye kategorisinde çok yüksek bir değişim oranı mevcut iken 2024 yılının temmuz ayında bu yüksekliğini koruyamamış ve sıradan bir yükselişe sahip kategori olduğu gözlemlenmiştir. 2024 yılının kasım ayında ise satış verisinin az olması sebebiyle tüm kategorilerde düşüş meydana gelmektedir.

Bir sonraki aşamada ise aylık olarak toplam satış verilerinin tahmin modeli yapılmıştır. Tahmin sırasında model olarak, LinearRegression modeli kullanılmıştır. Test_train_split işleminde test edilmesi gereken verinin boyutu %30 olarak belirlenirken eğitim için ayrılan verilerin boyutu %70 olarak belirlenmiştir. Yapılan testin sonucunda modelin skoru, 0.88 olarak belirlenmiştir. Bu durumun çıktısı **şekil 21**'de verilmiştir.

```
>>> model.score(X_test, y_test)
0.8841071626818521
```

Şekil 21: Oluşturulan modelin skoru