

Veri Analizi ve Manipölasyonu Ödev Raporu

MERVE TUTAR

Giriş

Bu rapor, veri analizi sürecinde gerçekleştirilen adımları, kullanılan yöntemleri ve elde edilen sonuçları detaylı bir şekilde açıklamaktadır. Satış ve müşteri verileri üzerinde veri temizleme, analiz ve görselleştirme adımlarını içermektedir. Ödev kapsamında dört ana görev ve ek olarak bir bonus görev gerçekleştirilmiştir.

Görev 1: Veri Temizleme ve Manipölasyonu

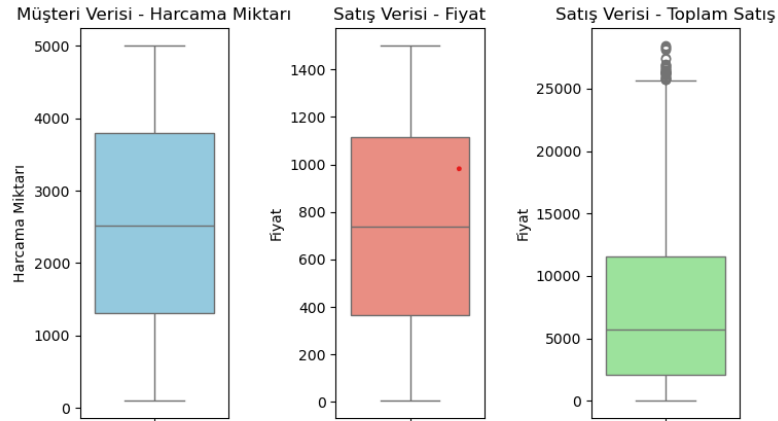
- Satış ve müşteri veri setlerine genel bir bakış yapılmış ve değişkenlerin tip bilgileri, satır sütun sayıları, eksik değerler, aykırı değerler incelenmiştir.
- Eksik Veriler:** Veri setlerinde eksik değer bulunmamıştır.

```
##### NA #####  
tarih          0  
ürün_kodu      0  
ürün_adi       0  
kategori       0  
fiyat          0  
adet          0  
toplam_satis   0  
musteri_id     0
```

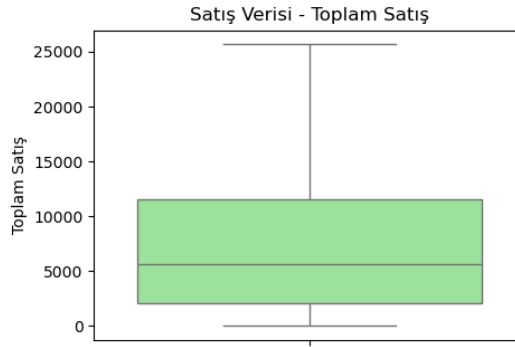
```
##### NA #####  
musteri_id     0  
isim          0  
cinsiyet       0  
yas           0  
sehir         0  
harcama_miktari 0
```

- Aykırı Değerler:**

Aykırı değerlerin tespiti için boxplot ve IQR (çeyreklik aralığı) yöntemi kullanılmıştır. Harcama miktarı ve fiyat değişkenlerinde aykırı değer bulunmamış; ancak toplam satış değişkeninde 35 aykırı değer tespit edilmiştir.



Aykırı değerler alt ve üst sınırların dışına çıkmayacak şekilde sınır değerleriyle değiştirilmiştir. Winsorization yöntemi ile düzeltilmiştir. İşlemden sonra yeniden aykırı değerlerin kontrolü yapılmış ve yöntem doğrulanmıştır.



- **Birleştirme:** Satış ve müşteri verileri müşteri_id anahtarına göre birleştirilmiş ve yeni bir veri seti oluşturulmuştur.

Sonuç: Aykırı değerlerin düzeltilmesi ve veri setinin birleştirilmesi, daha tutarlı bir analiz yapılması için önemli bir adımdır.

Görev 2: Zaman Serisi Analizi

- **Trend Analizleri:**

Haftalık ve aylık bazda toplam satış ve ürün adet trendleri analiz edilmiştir. Bu analizin daha kapsamlı olması için hem kategori bazında hem ürün bazında incelemeler gerçekleştirilmiştir. Bu analizin görselleştirilmesi Grafikler kısmında bulunmaktadır.

>>> print(haftalik_kategori_trend.head(10))					>>> print(aylik_kategori_trend.head(10))				
	tarih	kategori	toplam_satis	adet		tarih	kategori	toplam_satis	adet
0	2022-11-06	Giyim	7597.08	18	0	2022-11-30	Elektronik	122967.7700	182
1	2022-11-06	Kırtasiye	15166.89	39	1	2022-11-30	Ev Aletleri	224270.6400	337
2	2022-11-13	Elektronik	32978.25	57	2	2022-11-30	Giyim	194517.4300	292
3	2022-11-13	Ev Aletleri	41221.23	64	3	2022-11-30	Kozmetik	98075.4800	144
4	2022-11-13	Giyim	91152.90	94	4	2022-11-30	Kırtasiye	301804.9800	408
5	2022-11-13	Kozmetik	28540.98	42	5	2022-11-30	Mutfak Ürünleri	180047.7700	255
6	2022-11-13	Kırtasiye	60555.16	106	6	2022-12-31	Elektronik	270735.5600	407
7	2022-11-13	Mutfak Ürünleri	29381.55	65	7	2022-12-31	Ev Aletleri	341231.6400	412
8	2022-11-20	Elektronik	38762.63	51	8	2022-12-31	Giyim	319756.7225	366
9	2022-11-20	Ev Aletleri	107404.81	143	9	2022-12-31	Kozmetik	294354.9200	390

Haftalık ve aylık bazda kategorilere ait toplam satış ve adet bilgilerinin ilk 10 verisi bu şekildedir. Veriyi yorumlayacak olursak 2022 Kasım ayının 3. Haftasında Ev aletleri satışı bir önceki haftaya göre ani şekilde artmış, aynı şekilde Kırtasiye kategorisindeki satışlar da Kasım ayının 2. Haftası ani şekilde artmıştır. Aylık veriyi incelediğimizde de bu artışların yansıdığını kasım ayında en çok ev aletleri ve kırtasiye kategorisinde satış yapıldığını gözlemleyebilmekteyiz.

```
>>> print(haftalik_urun_trend.head(15))
      tarih  ürün_adi  toplam_satis  adet
0  2022-11-06  Bilgisayar    5332.44    12
1  2022-11-06      Kalem    7597.08    18
2  2022-11-06      Çanta    9834.45    27
3  2022-11-13  Bilgisayar    4816.92    12
4  2022-11-13      Defter    48192.32    62
5  2022-11-13      Fırın    15637.65    14
6  2022-11-13      Kalem    15377.69    47
7  2022-11-13      Klima    30073.80    32
8  2022-11-13  Kulaklık    42530.27    46
9  2022-11-13      Mouse    18223.89    28
10 2022-11-13  Su Şişesi     9735.03    24
11 2022-11-13      Telefon    52936.23    97
12 2022-11-13      Çanta    46306.27    66
13 2022-11-20  Bilgisayar    22344.51    45
14 2022-11-20      Defter    29476.68    54

>>> print(aylik_urun_trend.head(15))
      tarih  ürün_adi  toplam_satis  adet
0  2022-11-30  Bilgisayar    85284.7800    116
1  2022-11-30      Defter   158835.9300    211
2  2022-11-30      Fırın   100436.1100    140
3  2022-11-30      Kalem    69107.7700    149
4  2022-11-30      Klima   125215.7500    149
5  2022-11-30  Kulaklık    96697.9300    120
6  2022-11-30      Mouse   115329.1500    139
7  2022-11-30  Su Şişesi   102384.6000    153
8  2022-11-30      Telefon   166140.7100    273
9  2022-11-30      Çanta   102251.3400    168
10 2022-12-31  Bilgisayar   127810.4500    170
11 2022-12-31      Defter   244378.2200    279
12 2022-12-31      Fırın   222149.7825    272
13 2022-12-31      Kalem   157456.3400    218
14 2022-12-31      Klima   123579.2100    199
```

Haftalık ürün trendlerine ve aylık ürün trendlerine baktığımızda 2022 Kasım ayında telefon satışlarının en yüksek, bilgisayar satışlarının en düşük olduğunu görmekteyiz.

Her ayın ilk ve son satış günleri bulunmuş ve haftalık satış adetleri hesaplanmıştır.

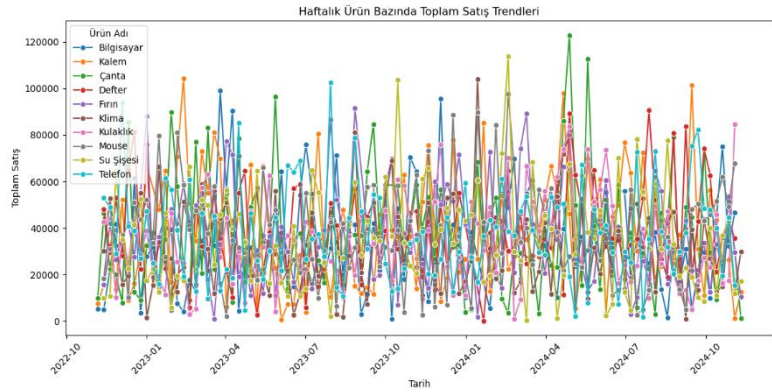
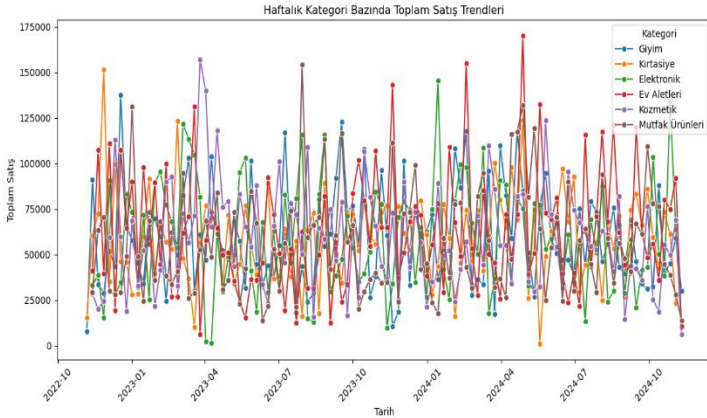
Verilerin ilk 5'er satırı bu şekildedir. Kasım ayında ürün satışı en yüksek 4. Haftadadır. İlk satış günleri incelendiğinde 2023 Şubat ayının ilk satış günü diğer 4 aya göre daha fazla satışa sahipken, 2022 Aralık en düşüktür.. 2023 Mart ayının ise son satış günü en yüksek satışa sahipken, 2023 Şubat en düşüktür.

```
Her ayın ilk satış günleri:
      tarih  toplam_satis
0  2022-11-06      8957.25
1  2022-12-01     2520.00
2  2023-01-01    13890.33
3  2023-02-01    21381.15
4  2023-03-01    13499.37
Her ayın son satış günleri:
      tarih  toplam_satis
0  2022-11-30      3828.54
1  2022-12-31     3793.57
2  2023-01-31     6891.69
3  2023-02-28     3480.12
4  2023-03-31    20343.39
```

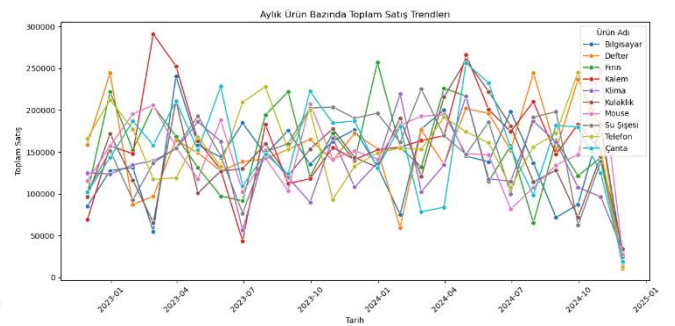
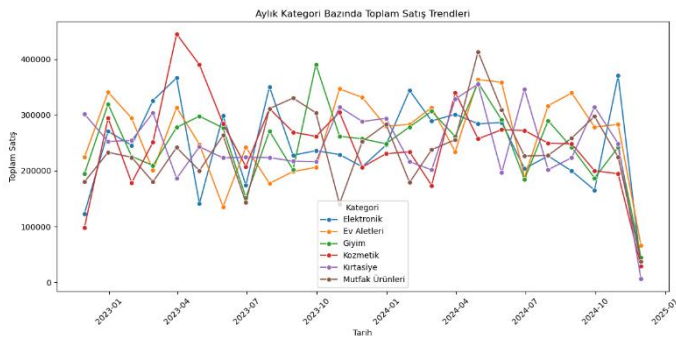
```
Haftalık Ürün Satışları:
      tarih  adet
0  2022-11-06     57
1  2022-11-13    428
2  2022-11-20    455
3  2022-11-27    459
4  2022-12-04    518
```

- **Grafikler:**

Haftalık ve aylık bazda toplam satış ve ürün adet trendlerinin sonuçları çizgi grafiği ile görselleştirilmiştir.

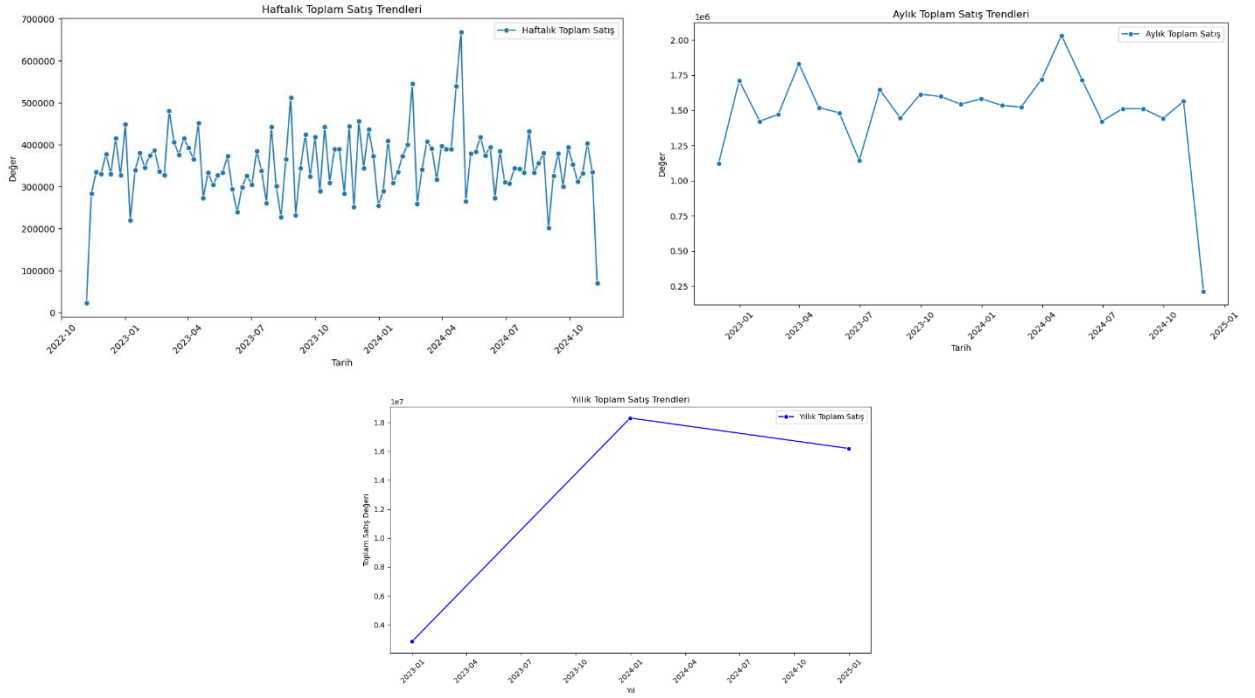


Haftalık ürün ve kategori bazında toplam satış trendleri incelendiğinde ürün olarak Kalem, Çanta, Su Şişesi ve Telefonun diğerlerine göre daha sık peak yaptığı görülmektedir. Haftalık kategori bazında baktığımızda ise ev aletleri kategorisinin en çok peak yapan kategori olduğu görülmektedir. Normalde her iki grafik arasında bir ilişki olması beklenir satış trendi yüksek olan ürünün ait olduğu kategorinin de satış trendi yüksek olmalıdır. Ancak verilen random oluşturulmuş veri setinde ürünler kategorilere aidiyet göstermediği için analizlerde ürünler ve kategoriler arasında bir ilişki beklemiyoruz.

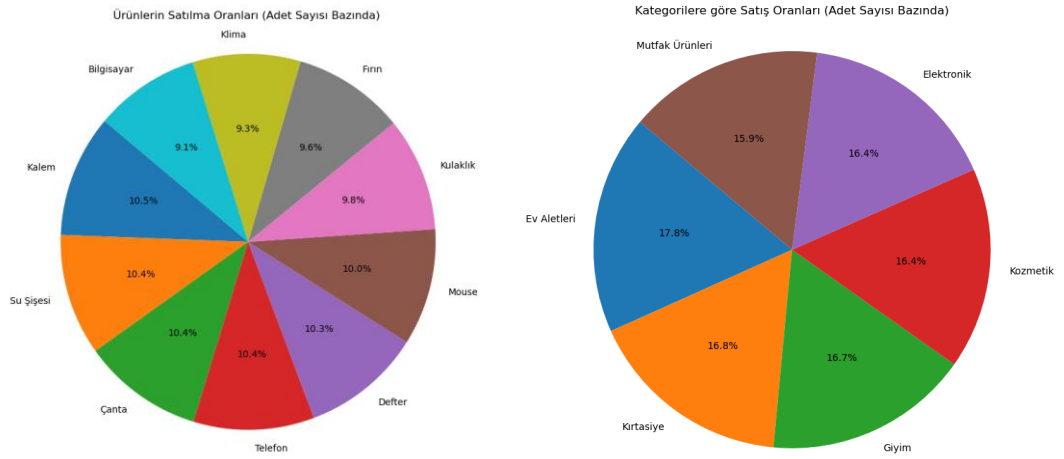


Aylık Kategori bazında satış trendleri incelendiğinde 2023 Nisan ayında Kozmetikte, Eylül ayında Giyimde, Mayıs ayında Mutfak Ürünlerinde, Ekim ayında Elektronikte en yüksek satışın gerçekleştiği görülüyor. Aylık ürün bazında baktığımızda ise Kalem satışlarının Şubat 2023'te en yüksek satışa sahip ürün olduğu görülmektedir. Haziran ayında da bir çok üründe en düşük satış miktarına ulaşılmıştır.

Aylık, yıllık ve haftalık olarak toplam satış trendleri benzer şekilde analiz edilmiştir.



Haftalık ve Aylık toplam satış trendleri grafikleri incelendiğinde 2024 Nisan ayında belirgin bir yükseliş trendi vardır ve bazı haftalarda satışlarda belirgin dalgalanmalar mevcuttur. Yıl bazında ise toplam satış değeri 2024 te en yüksek değerine ulaşmıştır. Aylık, haftalık, sezonluk dalgalanmalar tüketicilerin belirli dönemlerde yaşadığı özel günler, bayramlar, tatiller, kampanya dönemleri, mevsimsel etkiler ve ekonomik değişikliklerin alışveriş alışkanlıklarını değiştirmesinden kaynaklanabilir.



Ürünlerin adet bazında satılma oranları için pasta grafiğine bakıldığında en yüksek satış Kalem, en düşük satış Bilgisayardır. Kategorik olarak incelendiğinde ise en yüksek satış Ev Aletlerinde, en düşük Mutfak Ürünlerindedir.

Görev 3: Kategorisel ve Sayısal Analiz

- Kategorilere Göre Satışlar:**

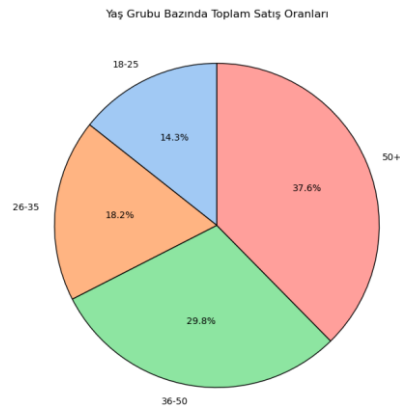
Ürün kategorilerinin toplam satışlar içindeki oranları hesaplanmıştır. Ev aletleri en yüksek orana, Mutfak Ürünleri en düşük orana sahiptir.

Kategori Bazında Toplam Satış ve Oranlar:			
	kategori	toplam_satis	oran
0	Elektronik	6.143434e+06	16.456960
1	Ev Aletleri	6.565546e+06	17.587709
2	Giyim	6.272433e+06	16.802523
3	Kozmetik	6.208006e+06	16.629937
4	Kırtasiye	6.183385e+06	16.563981
5	Mutfak Ürünleri	5.957503e+06	15.958890

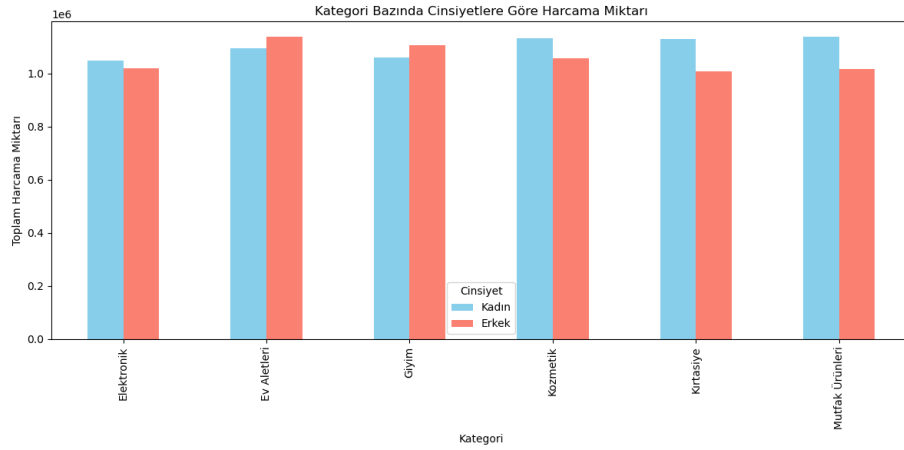
- Yaş Gruplarına Göre Satış Eğilimleri:**

Yaş Gruplarına Göre Toplam Satış ve Oranlar:			
	yas_grubu	toplam_satis	oran
0	18-25	5.354201e+06	14.342772
1	26-35	6.787118e+06	18.181255
2	36-50	1.113863e+07	29.838030
3	50+	1.405036e+07	37.637943

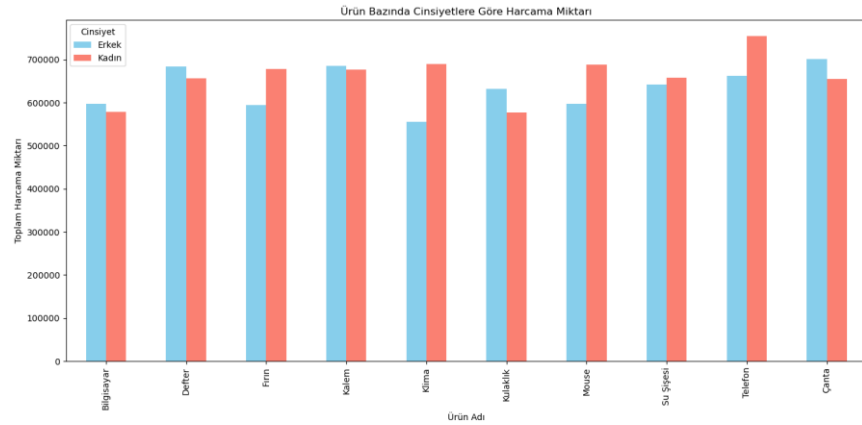
En yüksek satış oranı **50+ yaş grubu** müşterilerinden elde edilmiştir. En düşük satış oranı **18-25 yaş grubu** müşterilerinden elde edilmiştir.



- Cinsiyetlere Göre Harcama Davranışları:**



Kadın ve erkek müşterilerin harcama davranışları analiz edilmiş ve cinsiyete göre belirgin farklılıklar tespit edilmiştir. Erkekler yalnızca Ev aletleri ve Giyim kategorilerinde kadınlardan daha fazla harcama yapmıştır, diğer kategorilerde Kadınlar daha öndedir.



Ürün bazında ise Bilgisayar, Defter, Kalem, Kulaklık, Çanta ürünlerine Erkekler daha fazla harcama yapmıştır. Fırın, Klima, Mouse, Su Şişesi, Telefon ürünlerine ise Kadınlar daha fazla harcama yapmıştır. Sayısal veriler aşağıdaki gibidir.

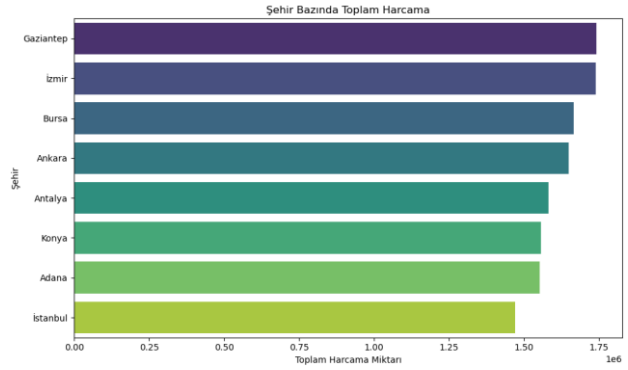
Kategori Bazında Cinsiyet Harcama Analizi:					Ürün Bazında Cinsiyet Harcama Analizi:		
cinsiyet	Erkek	Kadın	Kadın (%)	Erkek (%)	cinsiyet	Erkek	Kadın
kategori					ürün_adi		
Elektronik	1021228.79	1049575.63	50.684440	49.314353	Bilgisayar	596478.29	578440.43
Ev Aletleri	1137550.48	1094238.58	49.029660	50.969221	Defter	683293.38	655946.62
Giyim	1105768.97	1062010.99	48.990719	51.008128	Fırın	594728.70	677770.35
Kozmetik	1056958.99	1131808.95	51.709865	48.288994	Kalem	685387.80	676031.98
Kırtasiye	1009478.31	1129560.34	52.806916	47.191919	Klima	554705.42	689303.90
Mutfak Ürünleri	1017614.43	1139035.37	52.815036	47.183808	Kulaklık	632414.30	576419.14
					Mouse	596774.18	687285.15
					Su Şişesi	642283.68	657208.84
					Telefon	661386.93	753556.73
					Çanta	701147.29	654266.72

Görev 4: İleri Düzey Veri Manipülasyonu

- Şehir Bazında Harcama:

Şehirler toplam harcama miktarına göre sıralanmıştır. En çok harcama yapan şehirler belirlenmiştir.

Şehir Bazında Toplam Harcama		
	şehir	harcama_miktari
0	Gaziantep	1740073.60
1	İzmir	1739145.40
2	Bursa	1665417.27
3	Ankara	1649167.77
4	Antalya	1581436.54
5	Konya	1556038.91
6	Adana	1552729.00
7	İstanbul	1470821.34



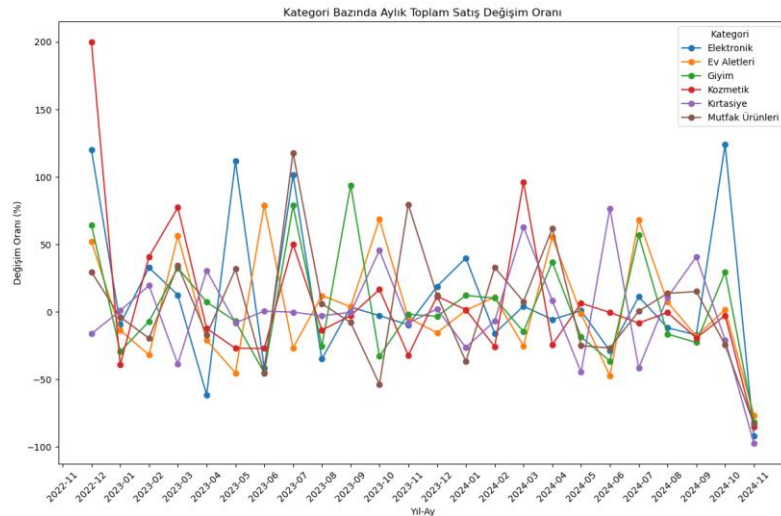
- Ürünlerin Ortalama Satış Artış Oranı:

Aylık satış artış oranları hesaplanmış ve ortalama değerler ürün bazında raporlanmıştır.

Ürünlerin Ortalama Satış Artış Oranı:		
	Ürün_adi	ortalama_artis_orani
0	Bilgisayar	15.746606
3	Kalem	14.042587
5	Kulaklık	10.172540
2	Fırın	9.938462
9	Çanta	8.058814
7	Su Şişesi	7.246460
1	Defter	6.907569
4	Klima	6.407863
6	Mouse	5.221097
8	Telefon	0.468948

- Kategorilerin Aylık Toplam Satış Değişim Oranı:

Her bir kategori için değişim oranları grafikte gösterilmiştir.



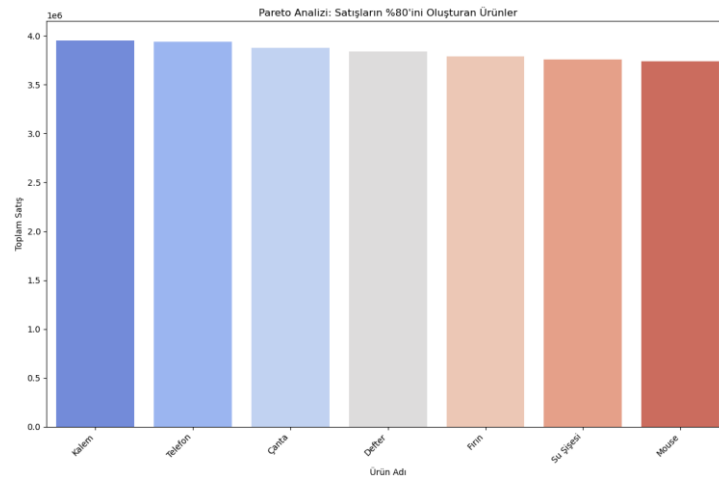
Sonuç:

- Şehir bazında harcama analizleri, bölgesel satış stratejilerinin geliştirilmesine yönelik içgörüler sunabilir.
- Ürün ve kategori bazlı değişim oranları, hangi alanlarda daha fazla odaklanılması gerektiğini göstermiştir.

Görev 5: Ekstra (BONUS)

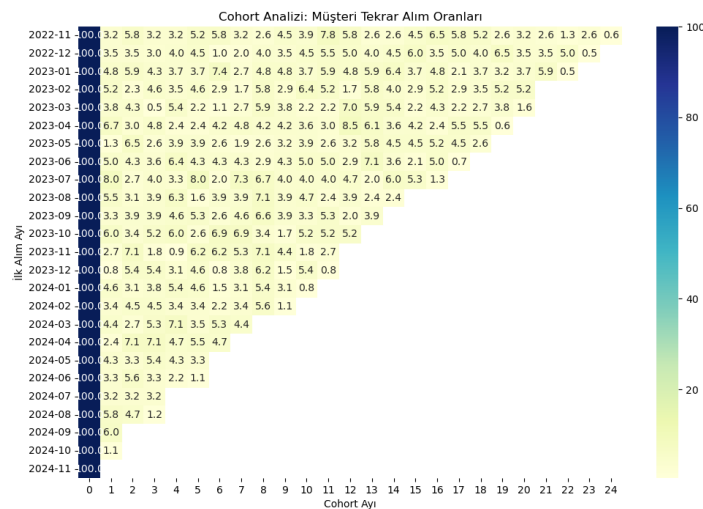
- **Pareto Analizi:**

Satışların %80'ini oluşturan ürünler belirlenmiş ve bir bar grafikte görselleştirilmiştir.



- **Cohort Analizi:**

Müşterilerin ilk alışveriş aylarına göre tekrar alım oranları analiz edilmiştir. Cohort analizi, müşteri davranışlarının sürekliliği hakkında bilgi vermiştir.



- **Tahmin Modeli:**

Aylık toplam satış tahmini için **Linear Regression** kullanılmıştır. Modelde kullanılan kategorik verilere one-hot encoding uygulanmıştır.

```
>>> print(f"Mean Absolute Error: {mae:.2f}")
Mean Absolute Error: 1783.73
>>> print(f"Mean Squared Error: {mse:.2f}")
Mean Squared Error: 5567930.31
>>> print(f"R^2 Score: {r2:.2f}")
R^2 Score: 0.87
```

Modelin başarımlık metrikleri:

- **MAE:** 1783.79
- **MSE:** 5567930.31
- **R²:** 0.87

Sonuç: Tahmin modeli, gelecekteki satış miktarlarını öngörmek için temel bir başlangıç sağlamıştır.