**Veri Analizi ve Manipülasyonu 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 Manipülasyonu**

* 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.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **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.

metin, diyagram, ekran görüntüsü, dikdörtgen içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Birleştirme:** Satış ve müşteri verileri musteri\_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.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, ekran görüntüsü, menü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, menü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, yazı tipi, ekran görüntüsü, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, ekran görüntüsü, yazı tipi, menü içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Grafikler:**

ekran görüntüsü, metin, renklilik, diyagram içeren bir resim

Açıklama otomatik olarak oluşturulduekran görüntüsü, metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Açıklama otomatik olarak oluşturulduHaftalık 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.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldumetin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturulduAylık, yıllık ve haftalık olarak toplam satış trendleri benzer şekilde analiz edilmiştir. metin, yazı tipi, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, diyagram, ekran görüntüsü, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, diyagram, ekran görüntüsü, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu

Ürünlerin adet bazında satılma oranları için pasta grafiğine bakıldığında en yüksek satış Kalem, en düşük satış Bilgisayardadır. 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.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Yaş Gruplarına Göre Satış Eğilimleri:**

**metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu**

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

diyagram, metin, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Cinsiyetlere Göre Harcama Davranışları:**

metin, ekran görüntüsü, yazı tipi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

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.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Görev 4: İleri Düzey Veri Manipülasyonu**

* **Şehir Bazında Harcama:**

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

metin, ekran görüntüsü, yazı tipi, menü içeren bir resim

Açıklama otomatik olarak oluşturuldu metin, ekran görüntüsü, çizgi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Ürünlerin Ortalama Satış Artışı Oranı:**

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

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Kategorilerin Aylık Toplam Satış Değişim Oranı:**

Her bir kategori için değişim oranları grafikle gösterilmiştir.

metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**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.

metin, ekran görüntüsü, dikdörtgen, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Cohort Analizi:**

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

metin, ekran görüntüsü, sayı, numara, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

* **Tahmin Modeli:**

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

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Modelin başarım metrikleri:

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

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