

# Regressa Örnek Veri Kalitesi Kontrol Raporu

## 1. Raporun Amacı

Bu örnek rapor, ham veri setlerinde analiz öncesinde tespit edilmesi gereken eksik, hatalı, tekrarlı ve tutarsız kayıtların nasıl raporlanabileceğini göstermek amacıyla hazırlanmıştır. Rapor kapsamında örnek bir işletmeye ait satış verisi ele alınmış; eksik değerler, tekrar eden kayıtlar, yazım farklılıkları, tarih formatı sorunları, sayısal değişkenlerdeki tutarsızlıklar ve mantıksız değerler veri kalitesi açısından değerlendirilmiştir.

Bu raporun temel amacı, veri setinin doğrudan analize uygun olup olmadığını ortaya koymak ve analiz öncesinde düzeltilmesi gereken veri kalitesi sorunlarını açık biçimde göstermektir. Böylece yapılacak istatistiksel analizlerin, dashboard çalışmalarının veya işletme raporlarının daha güvenilir bir veri yapısı üzerine kurulması hedeflenmektedir.

Bu çalışma gerçek kişi, kurum veya müşteri verisi içermemektedir. Kullanılan veriler yalnızca örnek veri kalitesi kontrol raporu formatını göstermek amacıyla kurgulanmıştır.

## 2. Veri Seti Hakkında

Örnek veri seti, bir işletmenin satış kayıtlarından oluşmaktadır. Ham veri dosyasında tarih, ürün adı, kategori, satış adedi, birim fiyat, müşteri türü ve ödeme yöntemi gibi değişkenler yer almaktadır.

Analizde kullanılan örnek değişkenler şunlardır:

Değişken	Açıklama
Tarih	Satış işleminin gerçekleştiği tarih
Ürün Adı	Satışı yapılan ürünün adı
Kategori	Ürünün ait olduğu ürün grubu
Satış Adedi	Satılan ürün miktarı
Birim Fiyat	Ürünün satış fiyatı
Toplam Tutar	Satış adedi ile birim fiyatın çarpımı
Müşteri Türü	Yeni müşteri / mevcut müşteri
Ödeme Yöntemi	Nakit, kredi kartı, havale/EFT vb.

Ham veri setinde farklı kaynaklardan gelen kayıtlar bir araya getirildiği için bazı biçim ve tutarlılık sorunları bulunmaktadır. Bu nedenle veri seti doğrudan analize alınmadan önce veri kalitesi açısından kontrol edilmelidir.

### 3. Tespit Edilen Veri Sorunları

Veri kalitesi kontrol sürecinin ilk aşamasında ham veri seti incelenmiş ve analiz kalitesini etkileyebilecek temel sorunlar belirlenmiştir.

**Tablo 1. Ham Veri Setinde Tespit Edilen Sorunlar**

Sorun Türü	Açıklama	Örnek
Eksik değer	Bazı hücrelerde bilgi bulunmaması	Boş satış adedi veya boş ödeme yöntemi
Tekrar eden kayıt	Aynı satış işleminin birden fazla kez yer alması	Aynı tarih, ürün ve tutarın iki kez girilmesi
Yazım farklılıkları	Aynı kategorinin farklı biçimlerde yazılması	"Dış Giyim", "Dis giyim", "Dışgiyim"
Tarih formatı sorunu	Tarihlerin farklı biçimlerde girilmesi	12.03.2024, 2024/03/12, 12-03-24
Sayısal format sorunu	Sayısal değerlerde nokta/virgül veya metin karışıklığı	"1.250 TL", "1250", "1,250"
Negatif veya hatalı değer	Satış adedi veya fiyat gibi alanlarda hatalı giriş	Satış adedi = -2
Kategori uyumsuzluğu	Ürün ile kategori eşleşmesinde hata	Ayakkabı ürününün "Üst Giyim" kategorisinde yer alması

Bu sorunlar giderilmeden yapılacak analizlerde hatalı toplamlar, yanlış kategori dağılımları, eksik satış hesaplamaları ve güvenilir olmayan rapor sonuçları ortaya çıkabilir.

### 4. Veri Kalitesi Kontrol Özeti

Ham veri setinin genel yapısını değerlendirmek amacıyla temel veri kalitesi kontrolleri yapılmıştır.

**Tablo 2. Veri Kalitesi Kontrol Özeti**

Kontrol Alanı	Tespit Edilen Kayıt Sayısı	Açıklama
---------------	----------------------------	----------

Eksik deęer ieren kayıtlar	18	En sık deme yntemi ve satıř adedi alanlarında grlmřtr
Tekrar eden kayıtlar	7	Aynı satıř iřleminin birden fazla kaydedildięi belirlenmiřtir
Tarih formatı hatası	11	Farklı tarih yazım biimleri tespit edilmiřtir
Kategori yazım tutarsızlıęı	14	Aynı kategori farklı řekillerde yazılmıřtır
Sayısal format sorunu	9	Fiyat ve tutar alanlarında metin/sayı karıřıklıęı grlmřtr
Negatif veya mantıksız deęer	3	Satıř adedi ve birim fiyat alanlarında hatalı deęerler bulunmuřtur

Kontrol sonuları, veri setinin analiz ncesinde standartlařtırılması gerektięini gstermektedir. zellikle kategori adları, tarih formatı ve sayısal deęiřkenlerdeki tutarsızlıklar satıř, stok ve krlılık analizlerinin doęruluęunu doęrudan etkileyebilir.

## 5. Veri Kalitesi Sorunlarının Olası Etkileri

Tespit edilen veri sorunları, yapılacak analizlerin doęruluęunu ve rapor sonularının gvenilirlięini etkileyebilir. Bu nedenle her sorun trnn analiz srecinde nasıl bir etki oluřturabileceęi ayrıca deęerlendirilmelidir.

**Tablo 3. Veri Sorunlarının Analiz Srecine Olası Etkileri**

Sorun Tr	Olası Etki
Eksik deęer	Toplam satıř, ortalama, yzde ve frekans hesaplamalarında eksik veya hatalı sonulara yol aabilir
Tekrar eden kayıt	Satıř adedi, ciro ve kr deęerlerinin olduęundan yksek grnmesine neden olabilir
Yazım farklılıkları	Aynı kategori veya rnn farklı gruplar gibi raporlanmasına yol aabilir
Tarih formatı sorunu	Aylık trend, dnemsel karıřılařtırma ve zaman serisi analizlerini bozabilir
Sayısal format sorunu	Toplam, ortalama ve oran hesaplamalarında hata oluřturabilir
Negatif veya mantıksız deęer	Satıř performansı ve krlılık analizlerinde yanlış yorumlara neden olabilir

Kategori uyumsuzluğu Ürün/kategori bazlı performans değerlendirmelerini hatalı hâle getirebilir

Bu değerlendirme, veri kalitesi kontrolünün yalnızca teknik bir işlem olmadığını; analiz sonuçlarının geçerliliği açısından temel bir aşama olduğunu göstermektedir.

## 6. Kontrol Öncesi ve Sonrası Karşılaştırma

Veri kalitesi kontrolü sonrasında veri setinde düzeltilmesi gereken temel alanlar belirlenmiştir. Bu bölümde, kontrol öncesi veri yapısı ile kontrol sonrasında raporlanan sorunların durumu karşılaştırmalı olarak sunulmuştur.

**Tablo 4. Kontrol Öncesi ve Sonrası Veri Kalitesi Durumu**

Gösterge	Kontrol Öncesi	Kontrol Sonrası / Raporlanan Durum
Toplam kayıt sayısı	1.250	1.250
Eksik değer içeren kayıt sayısı	Bilinmiyor	18
Tekrar eden kayıt sayısı	Bilinmiyor	7
Farklı kategori yazım sayısı	Bilinmiyor	12
Tarih formatı türü	Bilinmiyor	4
Sayısal format sorunu	Bilinmiyor	9
Hatalı negatif değer	Bilinmiyor	3

Karşılaştırma sonuçları, veri setinin kontrol öncesinde sorun alanlarının net biçimde bilinmediğini; kontrol sonrasında ise eksik, tekrarlı ve tutarsız kayıtların açık biçimde raporlandığını göstermektedir. Bu aşama, veri temizleme ve düzenleme sürecinin daha planlı yürütülmesini sağlar.

## 7. Öncelikli Düzeltmesi Gereken Alanlar

Veri kalitesi kontrolü sonucunda bazı sorunların analiz süreci açısından daha kritik olduğu belirlenmiştir. Bu nedenle sorunların öncelik düzeyine göre ele alınması önerilir.

**Tablo 5. Önceliklendirilmiş Veri Kalitesi Sorunları**

Öncelik Düzeyi	Sorun Alanı	Gerekçe
----------------	-------------	---------

Yüksek	Eksik satış adedi ve fiyat bilgileri	Ciro, satış adedi ve kârlılık hesaplamalarını doğrudan etkiler
Yüksek	Tekrar eden satış kayıtları	Toplam satışların olduğundan yüksek görünmesine neden olabilir
Yüksek	Sayısal format sorunları	Hesaplama ve toplama işlemlerinde hata oluşturabilir
Orta	Tarih formatı farklılıkları	Aylık ve dönemsel analizleri etkileyebilir
Orta	Kategori yazım tutarsızlıkları	Kategori bazlı raporların doğruluğunu bozabilir
Düşük	Ödeme yöntemi eksikleri	Ana satış analizinden çok ödeme davranışı analizlerini etkiler

Bu önceliklendirme, veri temizleme ve düzenleme işlemlerinin hangi sırayla yapılması gerektiğine ilişkin yol gösterici bir çerçeve sunmaktadır.

## 8. Analize Uygunluk Değerlendirmesi

Veri kalitesi kontrolü sonrasında veri setinin hangi analizler için doğrudan uygun olduğu, hangi analizler için ise ön düzenleme gerektiği değerlendirilmiştir.

**Tablo 6. Veri Setinin Analize Uygunluk Durumu**

Analiz Türü	Uygunluk Durumu	Açıklama
Satış analizi	Ön düzenleme gerekli	Eksik satış adedi ve sayısal format sorunları giderilmelidir
Kategori performans analizi	Ön düzenleme gerekli	Kategori adları standartlaştırılmalıdır
Ürün bazlı ciro analizi	Ön düzenleme gerekli	Ürün ve tutar değişkenleri kontrol edilmelidir
Aylık trend analizi	Ön düzenleme gerekli	Tarih formatı tek biçime dönüştürülmelidir
Ödeme yöntemi analizi	Kısmen uygun	Eksik ödeme yöntemi kayıtları tamamlanmalıdır
Dashboard raporlama	Ön düzenleme gerekli	Veri yapısı görsel raporlama için standartlaştırılmalıdır

Bu değerlendirme, veri setinin mevcut hâliyle bazı analizler için risk taşıdığını göstermektedir. Verinin daha güvenilir analiz çıktıları üretebilmesi için veri temizleme ve düzenleme aşamasından geçirilmesi önerilir.

## 9. Genel Değerlendirme

Bu örnek çalışma, ham veri setlerinin analiz öncesinde veri kalitesi açısından mutlaka kontrol edilmesi gerektiğini göstermektedir. Eksik değerler, tekrar eden kayıtlar, tarih formatı farklılıkları ve kategori yazım tutarsızlıkları küçük gibi görünse de analiz sonuçlarını önemli ölçüde etkileyebilir.

Özellikle işletme verilerinde farklı kişiler tarafından girilen Excel kayıtları, satış sistemlerinden alınan dosyalar veya manuel tutulan tablolar sıklıkla düzensiz olabilir. Bu nedenle veri kalitesi kontrolü, güvenilir rapor üretiminin ilk adımıdır.

Veri kalitesi kontrol raporu sayesinde veri setindeki temel sorunlar açık biçimde ortaya konur. Bu aşamadan sonra veri temizleme ve düzenleme işlemleri daha planlı biçimde yapılabilir; satış, stok, kategori, müşteri ve kârlılık analizleri için daha güvenilir bir veri altyapısı oluşturulabilir.

## 10. Regressa'nın Veri Kalitesi Kontrol Yaklaşımı

Regressa, analiz sürecine başlamadan önce veri setinin güvenilirliğini ve analize uygunluğunu değerlendirir. Bu kapsamda eksik değerler, tekrar eden kayıtlar, hatalı formatlar, tutarsız kategori yazımları, sayısal değişken sorunları ve mantıksız değerler sistematik olarak kontrol edilir.

Veri kalitesi kontrolü, yalnızca teknik bir ön inceleme değil; analiz sonuçlarının doğruluğunu doğrudan etkileyen temel bir aşamadır. Bu nedenle Regressa, veri setinde tespit edilen sorunları tablo yapısıyla raporlayarak müşterinin veri yapısını açık biçimde görmesini sağlar.

Bu yaklaşım sayesinde müşteriler, analiz veya dashboard sürecine geçmeden önce verilerindeki temel sorunları görebilir; hangi alanların temizlenmesi, düzenlenmesi veya yeniden kontrol edilmesi gerektiğini net biçimde anlayabilir.

---

### Regressa

Veri Analizi | İstatistiksel Modelleme | Raporlama

E-posta: regressanalysis@gmail.com