

Regressa Örnek Veri Temizleme ve Düzenleme Çalışması

1. Çalışmanın Amacı

Bu örnek çalışma, ham ve düzensiz bir veri setinin analiz öncesinde nasıl düzenlenebileceğini göstermek amacıyla hazırlanmıştır. Çalışmada örnek bir satış veri seti üzerinden hatalı yazımların düzeltilmesi, tarih formatlarının standartlaştırılması, sayısal değerlerin analize uygun hâle getirilmesi, kategori adlarının birleştirilmesi ve hesaplanabilir değişkenlerin oluşturulması süreçleri gösterilmiştir.

Bu dosyanın amacı, yalnızca verideki sorunları raporlamak değil; ham veri ile düzenlenmiş veri arasındaki farkı açık biçimde ortaya koymaktır. Böylece veri temizleme sürecinin analiz, raporlama ve dashboard çalışmalarından önce neden gerekli olduğu somut örneklerle gösterilmektedir.

Bu çalışma gerçek kişi, kurum veya müşteri verisi içermemektedir. Kullanılan veriler yalnızca örnek veri temizleme ve düzenleme formatını göstermek amacıyla kurgulanmıştır.

2. Ham Veri Yapısı

Ham veri setleri genellikle farklı kişiler, farklı sistemler veya farklı dönemlerde oluşturulan dosyalardan geldiği için doğrudan analize uygun olmayabilir. Aynı kategori farklı biçimlerde yazılmış olabilir, tarih formatları değişebilir, sayısal alanlar metin olarak girilmiş olabilir veya bazı değerler mantıksal açıdan hatalı olabilir.

Bu örnek çalışmada ham satış verisi aşağıdaki gibi kurgulanmıştır.

Tablo 1. Örnek Ham Veri Görünümü

Tarih	Ürün Adı	Kategori	Satış Adedi	Birim Fiyat	Ödeme Yöntemi
12.03.2024	Mont A	Dış Giyim	2	1.500 TL	Kredi Kartı
2024/03/13	Ayakkabı B	ayakkabı	1	900	kredi kartı
14-03-24	Tişört E	Üstgiyim	3	300 TL	Nakit
15.03.2024	Gömlek F	Dis giyim	-1	650 TL	
16/03/2024	Çanta G	Aksesuar	2	1,250	Havale
17.03.2024	Pantolon D	Alt Giyim	1	550 TL	KREDİ KARTI

Ham veri görünümü incelendiğinde, aynı veri seti içinde farklı tarih biçimleri, kategori yazım farklılıkları, ödeme yöntemi standartlaşmaması, sayısal alanlarda metin/para birimi kullanımı ve mantıksız satış adedi gibi sorunlar bulunduğu görülmektedir.

3. Düzenleme Kuralları

Veri temizleme sürecinde her değişken için belirli standartlar tanımlanmıştır. Bu standartlar, verinin daha sonra analiz edilebilir, filtrelenebilir ve raporlanabilir hâle gelmesini sağlar.

Tablo 2. Uygulanan Düzenleme Kuralları

Alan	Ham Veri Sorunu	Düzenleme Kuralı
Tarih	Farklı tarih yazım biçimleri	Tüm tarihler GG.AA.YYYY formatına dönüştürülür
Ürün Adı	Büyük/küçük harf ve boşluk tutarsızlıkları	Ürün adları standart yazım biçimine getirilir
Kategori	Aynı kategori farklı biçimlerde yazılmıştır	Kategoriler standart kategori listesine eşleştirilir
Satış Adedi	Negatif veya mantıksız değerler olabilir	Negatif değerler kontrol edilir ve düzeltilmeden analize alınmaz
Birim Fiyat	TL ifadesi, nokta/virgöl ve metin karakterleri karışık olabilir	Sayısal formata dönüştürülür
Ödeme Yöntemi	Farklı büyük/küçük harf ve boş değerler olabilir	Standart ödeme yöntemi listesine dönüştürülür
Toplam Tutar	Bazı dosyalarda bulunmayabilir	Satış Adedi × Birim Fiyat olarak hesaplanır

Bu kurallar, veri temizleme sürecinin rastgele değil, belirli bir mantık ve standart çerçevesinde yürütülmesini sağlar.

4. Ham Veri – Düzenlenmiş Veri Karşılaştırması

Bu bölümde ham veri satırları ile düzenleme sonrası oluşturulan analize hazır veri yapısı karşılaştırmalı olarak gösterilmiştir.

Tablo 3. Ham Veri ve Düzenlenmiş Veri Karşılaştırması

Alan	Ham Veri	Düzenlenmiş Veri	Yapılan İşlem
Tarih	2024/03/13	13.03.2024	Tarih formatı standartlaştırıldı

Tarih	14-03-24	14.03.2024	Kısa tarih biçimi düzenlendi
Kategori	ayakkabı	Ayakkabı	Büyük/küçük harf standardı sağlandı
Kategori	Üstgiyim	Üst Giyim	Kategori yazımı standartlaştırıldı
Kategori	Dis giyim	Dış Giyim	Yazım hatası düzeltildi
Birim Fiyat	1.500 TL	1500	Para birimi ifadesi kaldırıldı ve sayısal değere çevrildi
Birim Fiyat	1,250	1250	Sayısal format standartlaştırıldı
Ödeme Yöntemi	kredi kartı	Kredi Kartı	Yazım standardı sağlandı
Ödeme Yöntemi	KREDİ KARTI	Kredi Kartı	Büyük harf kullanımı düzenlendi
Satış Adedi	-1	Kontrol Edilecek	Mantıksız değer olarak işaretlendi

Bu karşılaştırma, veri temizleme sürecinin yalnızca biçimsel düzenleme olmadığını göstermektedir. Bazı alanlarda doğrudan standartlaştırma yapılabilirken, bazı değerler veri sahibinden doğrulama gerektirebilir. Özellikle negatif satış adedi gibi mantıksız değerler otomatik olarak düzeltilmemeli; önce kontrol edilmelidir.

5. Kategori Standartlaştırma Örneği

Kategori adlarının farklı biçimlerde yazılması, raporlama sürecinde aynı ürün grubunun farklı kategoriler altında görünmesine neden olur. Bu durum özellikle kategori bazlı ciro, satış adedi ve kârlılık analizlerinde hatalı sonuçlara yol açabilir.

Tablo 4. Kategori Eşleştirme Tablosu

Ham Kategori Yazımı	Standart Kategori
Dış Giyim	Dış Giyim
Dis giyim	Dış Giyim
Dışgiyim	Dış Giyim
ayakkabı	Ayakkabı
Ayakkabı	Ayakkabı
Üstgiyim	Üst Giyim
Üst Giyim	Üst Giyim

Alt giyim	Alt Giyim
Aksesuar	Aksesuar

Bu eşleştirme sonucunda farklı yazımlar tek bir standart kategori altında toplanır. Böylece analizlerde kategori sayısı yapay olarak artmaz ve aynı gruba ait satışlar doğru biçimde birleştirilir.

6. Sayısal Değerlerin Düzenlenmesi

Ham veri setlerinde fiyat, tutar veya miktar alanları bazen metin olarak girilebilir. Örneğin “1.500 TL” gibi bir ifade görsel olarak fiyatı temsil etse de analiz yazılımlarında doğrudan sayısal değer olarak algılanmayabilir. Bu nedenle para birimi ifadeleri, binlik ayrıçlar ve uyumsuz nokta/virgül kullanımları düzenlenmelidir.

Tablo 5. Sayısal Format Düzenleme Örnekleri

Ham Değer	Düzenlenmiş Değer	Açıklama
1.500 TL	1500	TL ifadesi kaldırıldı, sayı formatına çevrildi
900	900	Uygun formatta olduğu için korundu
300 TL	300	Para birimi ifadesi kaldırıldı
1,250	1250	Sayısal biçim standartlaştırıldı
550 TL	550	TL ifadesi kaldırıldı

Sayısal değerlerin düzenlenmesi, toplam satış, ortalama fiyat, brüt kâr, oran ve trend hesaplamalarının doğru yapılabilmesi için gereklidir.

7. Mantıksal Kontrol ve İşaretleme

Veri temizleme sürecinde bazı değerler doğrudan düzeltilemez. Özellikle satış adedi, fiyat, tarih veya müşteri bilgisi gibi alanlarda mantıksız görünen değerler veri sahibinden doğrulama gerektirebilir.

Tablo 6. Mantıksal Kontrol Örneği

Alan	Ham Değer	Sorun	İşlem
Satış Adedi	-1	Satış adedi negatif olamaz	Kontrol edilecek olarak işaretlendi
Birim Fiyat	0	Ücretsiz ürün mü, hata mı belirsiz	Doğrulama gerekli

Tarih	31.02.2024	Geçersiz tarih	Hatalı kayıt olarak işaretlendi
Toplam Tutar	Satış adedi × fiyat ile uyumsuz	Hesaplama tutarsızlığı	Yeniden hesaplama veya kontrol gerekli

Bu yaklaşım, veri temizleme sürecinin güvenilirliğini artırır. Her sorun otomatik olarak silinmez veya değiştirilmez; gerekli durumlarda veri sahibinden doğrulama alınması önerilir.

8. Analize Hazır Veri Görünümü

Temizleme ve düzenleme işlemlerinden sonra veri seti daha standart, tutarlı ve analize uygun bir yapıya getirilir. Aşağıdaki tablo, düzenleme sonrası oluşan örnek veri görünümünü göstermektedir.

Tablo 7. Analize Hazır Veri Görünümü

Tarih	Ürün Adı	Kategori	Satış Adedi	Birim Fiyat	Toplam Tutar	Ödeme Yöntemi	Durum
12.03.2024	Mont A	Dış Giyim	2	1500	3000	Kredi Kartı	Uygun
13.03.2024	Ayakkabı B	Ayakkabı	1	900	900	Kredi Kartı	Uygun
14.03.2024	Tişört E	Üst Giyim	3	300	900	Nakit	Uygun
15.03.2024	Gömlek F	Dış Giyim	-1	650	-650	Eksik	Kontrol Edilecek
16.03.2024	Çanta G	Aksesuar	2	1250	2500	Havale	Uygun
17.03.2024	Pantolon D	Alt Giyim	1	550	550	Kredi Kartı	Uygun

Bu yapı sayesinde veri seti filtreleme, toplama, kategori bazlı raporlama, satış analizi ve dashboard çalışmaları için daha uygun hâle gelir. Kontrol gerektiren kayıtlar ise ayrı olarak işaretlenir ve analiz öncesinde veri sahibiyle değerlendirilir.

9. Temizleme Sonrası Oluşturulan Ek Değişkenler

Veri temizleme sürecinde bazı yeni değişkenler oluşturularak veri setinin analiz kapasitesi artırılabilir. Bu değişkenler, daha sonra yapılacak raporlamalarda kullanılabilir.

Tablo 8. Oluşturulabilecek Ek Değişkenler

Yeni Değişken	Nasıl Oluşturulur?	Kullanım Amacı
Toplam Tutar	Satış Adedi × Birim Fiyat	Ciro hesaplaması
Ay	Tarih değişkeninden türetilir	Aylık trend analizi
Yıl	Tarih değişkeninden türetilir	Yıllık karşılaştırma
Kontrol Durumu	Uygun / Kontrol Edilecek	Sorunlu kayıtların ayrılması
Kategori Kodu	Standart kategori listesine göre atanır	Dashboard ve filtreleme
Birim Fiyat Grubu	Düşük / Orta / Yüksek	Fiyat segmenti analizi

Bu tür ek değişkenler, veri setinin daha esnek ve analiz edilebilir bir yapıya kavuşmasını sağlar.

10. Genel Değerlendirme

Bu örnek çalışma, ham veri ile analize hazır veri arasındaki farkı göstermektedir. Ham veri setinde bulunan farklı tarih formatları, kategori yazım tutarsızlıkları, sayısal format sorunları ve mantıksız değerler analiz öncesinde düzenlenmediğinde hatalı sonuçlara neden olabilir.

Veri temizleme ve düzenleme süreci sayesinde veriler daha standart, tutarlı ve raporlanabilir hâle getirilir. Bu süreç yalnızca teknik bir hazırlık değil; güvenilir analiz üretmenin temel adıımıdır. Doğru temizlenmiş bir veri seti, daha sağlıklı istatistiksel analiz, işletme raporlaması, satış analizi ve dashboard çalışmaları yapılmasını sağlar.

11. Regressa'nın Veri Temizleme ve Düzenleme Yaklaşımı

Regressa, veri temizleme sürecinde yalnızca hatalı alanları düzeltmeye değil, veriyi analiz amacına uygun bir yapıya dönüştürmeye odaklanır. İlk aşamada ham veri yapısı incelenir, ardından değişken türleri belirlenir, standartlaştırma kuralları oluşturulur ve veri bu kurallara göre düzenlenir.

Kontrol gerektiren kayıtlar doğrudan silinmez; ayrı olarak işaretlenir ve müşteriye bildirilir. Böylece veri sahibinin bilgisi olmadan kritik kayıtlar değiştirilmez. Bu yaklaşım, hem veri güvenilirliğini hem de analiz sonuçlarının doğruluğunu artırır.

Regressa'nın veri temizleme ve düzenleme hizmeti; anket analizleri, işletme satış raporları, ekonometrik çalışmalar, Excel raporları ve dashboard projeleri için güçlü bir hazırlık aşaması sunar.

Regressa

Veri Analizi | İstatistiksel Modelleme | Raporlama

E-posta: regressanalysis@gmail.com