

*IBM SPSS Data Preparation 29*



**Not**

Bu belgeyi ve desteklediđi ürünü kullanmadan önce řu kısımdaki bilgileri okuyun: [“Özel Notlar” sayfa 23.](#)

**Ürün Bilgileri**

Bu basım, yeni basımlarında tersi belirtilmediđi sürece, IBM® SPSS Statistics sürüm 29, yayın düzeyi 0, deđişiklik 1 ve sonraki tüm yayın düzeyleri ve deđişiklikler için geçerlidir.

© Copyright International Business Machines Corporation .

# İçindekiler

<b>Bölüm 1. Veri hazırlığı.....</b>	<b>1</b>
Veri Hazırlıklarına Giriş.....	1
Veri Hazırlama yordamlarının kullanımı.....	1
Doğrulama Kuralları.....	1
Önceden Tanımlı Doğrulama Kurallarını Yükle.....	1
Doğrulama Kurallarını Tanımla.....	2
Verileri Doğrula.....	3
Veri Temel Denetimlerini Doğrula.....	4
Veri Tek Değişken Kurallarını Doğrula.....	5
Veri Çapraz Değişkenini Doğrula-Değişken Kuralları.....	5
Veri Çıkışını Doğrula.....	5
Veri Kaydetmeyi Doğrula.....	6
Otomatikleştirilmiş Veri Hazırlığı.....	6
Otomatik Veri Hazırlığı Elde Etmek İçin.....	7
Etkileşimli Veri Hazırlığı Elde Etmek İçin.....	7
Alanlar Sekmesi .....	7
Ayarlar Etiketleri.....	8
Çözümleme Etiketleri .....	12
Geri Dönüştürme Puanları.....	18
Olağan Dışı Durumları Tanımlama.....	18
Olağan Dışı Vakalar çıkışını Tanımla.....	19
Olağan Dışı Vakaların Kaydedilmesini Tanımla.....	19
Olağan Dışı Vakaların Eksik Değerlerini Belirleme.....	20
Olağan Dışı Durumlar Seçeneklerini Belirleyin.....	20
DETECTANOMALY Komut Ek Özellikleri.....	20
En Uygun Vuruş.....	21
En Uygun Başlangıç Çıktısı.....	21
En Uygun Başlangıç Kaydı.....	22
Eksik Değerleri En Uygun şekilde Sabitle.....	22
En Uygun Vuruş Seçenekleri.....	22
EN İYİ SABITLEME KOMUTU Ek Özellikler.....	22
<b>Özel Notlar.....</b>	<b>23</b>
Ticari markalar.....	24
<b>Dizin.....</b>	<b>27</b>



# Bölüm 1. Veri hazırlığı

Base Edition 'da aşağıdaki veri hazırlama özellikleri yer almaktadır.

## Veri Hazırlıklarına Giriş

Bilgi işlem sistemleri güçte artış gösterirken, bilgi iştahı orantılı olarak artar ve daha fazla veri toplama hataları, daha çok değişken ve daha fazla veri girdisi hatasına yol açmasını sağlar. Bu hatalar, veri ambarı oluşturma nihai hedefi olan tahmine dayalı model tahminlerinin banadır, bu nedenle "clean" (temizleme) verilerini saklamanız gerekir. Ancak, veri ambarı miktarı, vakaları el ile doğrulamanın ötesinde, verilerin doğrulanmasına yönelik otomatikleştirilmiş süreçleri uygulamak için hayati önem taşıyan bir yetenekten daha fazla büyümektedir.

Veri Hazırlama, etkin veri kümenize olağan dışı durumları ve geçersiz vakaları, değişkenleri ve veri değerlerini tanımlamanızı ve modelleme için verileri hazırlamanızı sağlar.

## Veri Hazırlama yordamlarının kullanımı

Veri Hazırlama prosedürleri kullanımınız, sizin özel gereksinimlerinize bağlıdır. Tipik bir rota, verilerinizi yükledikten sonra aşağıdaki gibi olur:

- **Meta veri hazırlığı.** Veri dosyanızın değişkenlerini gözden geçirin ve bunların geçerli değerlerini, etiketlerini ve ölçüm düzeylerini belirleyin. İmkansız ancak yaygın olarak yanlış kodlanan değişken değerleri bileşimlerini tanımlayın. Bu bilgilere dayalı olarak geçerlilik denetimi kurallarını tanımlayın. Bu, zaman alıcı bir görev olabilir, ancak benzer özniteliklere sahip veri dosyalarının düzenli olarak doğrulanabilmesi için gereken çabaya değer.
- **Veri doğrulaması.** Temel denetimleri çalıştırın ve geçersiz vakaları, değişkenleri ve veri değerlerini tanımlamak için tanımlı doğrulama kurallarına göre denetler. Geçersiz veriler bulunduğunda, nedeni araştırın ve düzeltin. Bu, meta veri hazırlığı yoluyla başka bir adım gerektirebilir.
- **Model hazırlığı.** Model oluşturmayı iyileştirecek özgün alanların dönüşümleri elde etmek için otomatik veri hazırlama olanağını kullanın. Birçok tahmine dayalı model için sorun oluşturabilecek olası istatistiksel aykırı değerleri belirleyin. Bazı aykırı değerler, tanımlanmamış geçersiz değişken değerlerinin sonuçlarıdır. Bu, meta veri hazırlığı yoluyla başka bir adım gerektirebilir.

Veri dosyanız "temiz" olduğunda, diğer eklenti modüllerinden modeller oluşturmaya hazırsınız.

## Doğrulama Kuralları

Bir vakanın geçerli olup olmadığını belirlemek için kural kullanılır. İki tip geçerlilik denetimi kuralı vardır:

- **Tek değişkenli kurallar.** Tek değişkenli kurallar, tek bir değişken için geçerli olan ve aralık dışı değerler için denetim gibi, sabit bir denetim kümeninden oluşur. Tek değişkenli kurallarda, geçerli değerler bir değer aralığı ya da kabul edilebilir değerler listesi olarak ifade edilebilir.
- **Çapraz değişken kuralları.** Çapraz değişken kuralları, tek bir değişkene ya da değişkenlerin birleşimine uygulanabilen kullanıcı tanımlı kurallardır. Çapraz değişken kuralları, geçersiz değerleri işaretleyen bir mantıksal ifadeye göre tanımlanır.

Doğrulama kuralları, veri dosyanızın veri sözlüğüne kaydedilir. Bu, bir kuralı bir kez belirtmenize ve sonra yeniden kullanmanıza olanak sağlar.

## Önceden Tanımlı Doğrulama Kurallarını Yükle

Kuruluma dahil edilen bir dış veri dosyasından önceden tanımlanmış kuralları yükleyerek, kullanıma hazır bir doğrulama kuralları kümesini hızla edinebilirsiniz.

Önceden Tanımlanmış Geçerlilik Denetimi Kurallarını Yükleme İçin

1. Menülerden şunları seçin:

**Veri > Geçerlilik Denetimi > Önceden Tanımlı Kuralları Yükle ...**

Diğer bir seçenek olarak, kuralları herhangi bir veri dosyasından yüklemek için Veri Özelliklerini Kopyala Sihirbazı'nı kullanabilirsiniz.

## Doğrulama Kurallarını Tanımla

Doğrulama Kurallarını Tanımla iletişim kutusu, tek değişkenli ve çapraz değişken geçerlilik denetimi kuralları yaratmanızı ve görüntülemenizi sağlar.

Doğrulama kuralları yaratmak ve görüntülemek için

1. Menülerden şunları seçin:

**Veri > Geçerlilik Denetimi > Kuralları Tanımla ...**

İletişim kutusu, veri sözlüğünden okunan tek değişkenli ve değişken çapraz geçerlilik denetimi kuralları ile doldurulur. Kural yoksa, amaçlarınıza uygun olarak değiştirebileceğiniz yeni bir yer tutucu kural otomatik olarak yaratılır.

2. Kendi özelliklerini görüntülemek ve değiştirmek için Tek Değişken Kuralları ve Çapraz Değişken Kuralları sekmeleri üzerinde tek tek kuralları seçin.

## Tek Değişkenli Kuralları Tanımla

Tek Değişken Kuralları sekmesi, tek değişkenli doğrulama kurallarını oluşturmanıza, görüntülemenize ve değiştirmenize olanak sağlar.

**Kurallar.** Liste, ada göre tek değişkenli geçerlilik denetimi kurallarını ve kuralın uygulanabileceği değişken tipini gösterir. İletişim kutusu açıldığında, veri sözlüğünde tanımlanan kuralları gösterir ya da şu anda herhangi bir kural tanımlanmadıysa, "Tek Değişken Kural 1" adlı bir yertutucu kuralı gösterilir. Kural listesinin altında aşağıdaki düğmeler yer alır:

- **Yeni.** Kural listesinin alt kısmına yeni bir giriş ekler. Kural seçilir ve "SingleVarRule  $n$ " adı atanır; burada  $n$ , yeni kuralın adının tek değişkenli ve çapraz değişken kuralları arasında benzersiz olması için bir tamsayıdır.
- **Çoğalt.** Seçilen kuralın bir kopyasını Kural listesinin alt kısmına ekler. Kural adı, tek değişkenli ve çapraz değişken kuralları arasında benzersiz olacak şekilde ayarlanır. Örneğin, "SingleVarRule 1" yineleniyorsa, ilk yinelenen kuralın adı "SingleVarRule 1'in Kopyası" olur, ikincisi "Copy (2) of SingleVarRule 1," vb. olur.
- **Sil.** Seçilen kuralı siler.

**Kural Tanımlaması.** Bu denetimler, seçilen bir kurala ilişkin özellikleri görüntülemenizi ve ayarlamanızı sağlar.

- **Ad.** Kuralın adı tek değişkenli ve çapraz değişken kuralları arasında benzersiz olmalıdır.
- **Tip.** Bu, kuralın uygulanabileceği değişken tipidir. **Sayısal, Dizgi ve Tarih** arasından seçim yapın.
- **Biçim.** Bu, tarih değişkenlerine uygulanabilen kurallar için tarih biçimini seçmenize olanak sağlar.
- **Geçerli Değerler.** Geçerli değerleri bir aralık ya da değer listesi olarak belirleyebilirsiniz.

Aralık Tanımı

Aralık tanımlaması denetimleri, geçerli bir aralık belirtmenizi sağlar. Aralığın dışındaki değerler geçersiz olarak işaretlendi.

Bir aralık belirtmek için, alt sınır ya da üst sınır değerlerini girin ya da her ikisini birden girin. Onay kutusu denetimleri, aralık içinde etiketlenmemiş ve tamsayı olmayan değerleri işaretlemenize olanak tanır.

Liste Tanımı

Liste tanımı denetimleri, geçerli değerlerin bir listesini tanımlamanızı sağlar. Listede yer almayan değerler geçersiz olarak işaretlenir.

Izgaradaki liste değerlerini girin. Bu onay kutusu, dizgi veri değerleri, kabul edilebilir değerler listesine göre işaretlendiğinde büyük/küçük harf durumu önem olup olmadığını belirler.

- **Kullanıcının eksik değerlerine izin ver.** Kullanıcı eksik değerlerinin geçersiz olarak işaretlenip işaretlenmediğini denetler.
- **Sistem eksik değerlere izin ver.** Sistem eksik değerlerinin geçersiz olarak işaretlenip işaretlenmediğini denetler. Bu, dizgi kuralı tipleri için geçerli değildir.
- **Boş değerlere izin ver.** Boş (tamamen boş olan) dizgi değerlerinin geçersiz olarak işaretlenip işaretlenmediğini denetler. Bu, dizgi dışı kural tipleri için geçerli değildir.

## Çapraz Değişken Kuralları Tanımla

Çapraz Değişken Kuralları sekmesi, değişken doğrulama kurallarını oluşturmanıza, görüntülenize ve değiştirmenize olanak sağlar.

**Kurallar.** Liste, ada göre değişken geçerlilik denetimi kurallarını gösterir. İletişim kutusu açıldığında, "CrossVarRule 1" adlı bir yertutucu kuralı gösterir. Kural listesinin altında aşağıdaki düğmeler yer alır:

- **Yeni.** Kural listesinin alt kısmına yeni bir giriş ekler. Kural seçilir ve "CrossVarRule n" adı atanır; burada n , yeni kuralın adının tek değişkenli ve çapraz değişken kuralları arasında benzersiz olması için bir tamsayıdır.
- **Çoğalt.** Seçilen kuralın bir kopyasını Kural listesinin alt kısmına ekler. Kural adı, tek değişkenli ve çapraz değişken kuralları arasında benzersiz olacak şekilde ayarlanır. Örneğin, "CrossVarRule 1" yinelenirse, ilk yinelenen kuralın adı "CrossVarRule 1 'in Kopyası" olur, ikincisi "Copy (2) of CrossVarRule 1," vb. olur.
- **Sil.** Seçilen kuralı siler.

**Kural Tanımlaması.** Bu denetimler, seçilen bir kurala ilişkin özellikleri görüntülenizi ve ayarlamınızı sağlar.

- **Ad.** Kuralın adı tek değişkenli ve çapraz değişken kuralları arasında benzersiz olmalıdır.
- **Mantıksal İfade.** Bu, temel olarak, kural tanımlar. İfadeyi geçersiz vakalar 1 olarak değerlendirecek şekilde kodlamalısınız.

İfadeler oluşturuluyor

1. Bir ifade oluşturmak için, bileşenleri İfade alanına yapıştırın ya da doğrudan İfade alanına yazın.
- İşlev grubu listesinden bir grup seçerek ve İşlevler ve Özel Değişkenler listesindeki işlevi ya da değişkeni çift tıklatarak, işlevleri ya da yaygın olarak kullanılan sistem değişkenlerini yapıştırabilirsiniz (ya da işlevi ya da değişkeni seçip **Araya Ekle'** yi tıklatın). Soru işaretleriyle belirtilen parametrelere ilişkin değerleri girin (yalnızca işlevler için geçerlidir). **All** (Tümü) etiketli işlev grubu, kullanılabilir tüm işlevlerin ve sistem değişkenlerinin bir listesini sağlar. Seçilen işleve ya da değişkene ilişkin kısa bir açıklama, iletişim penceresindeki ayrılmış bir alanda görüntülenir.
  - Dizilim değişmezleri tırnak işaretleri ya da kesme imi içine alınmalıdır.
  - Değerler ondalık değer içeriyorsa, ondalık gösterge olarak nokta (.) kullanılmalıdır.

## Verileri Doğrula

Verileri Doğrula iletişim kutusu, etkin veri kümesinde şüpheli ve geçersiz durumları, değişkenleri ve veri değerlerini tanımlamanıza olanak sağlar.

**Örnek.** Bir veri analisti, müşterisine aylık bir müşteri memnuniyeti raporu sağlamalıdır. Her ay aldığı veriler, eksik müşteri kimlikleri, aralık dışında değişen değişken değerleri ve hata olarak sıkça girilen değişken değerlerinin birleşimleri için kalite kontrol edilmesi gerekir. Verileri Doğrula iletişim kutusu, çözümleyicinin müşterileri benzersiz bir şekilde tanımlayan değişkenleri belirtmesini, geçerli değişken aralıkları için tek değişkenli kuralları tanımlamasını ve imkansız birleşimleri yakalamak için değişken kurallar tanımlamanızı sağlar. Yordam, sorun senaryolarının ve değişkenlerin raporunu döndürür. Ayrıca, veriler her ay aynı veri öğelerine sahiptir, bu nedenle analist, kuralları yeni veri dosyasına gelecek ay uygulayabilmiştir.

**İstatistikler.** Bu prosedür, çeşitli çekler, tek değişkenlik ve çapraz değişken kurallarının ihlal sayımları ve analiz değişkenlerinin basit açıklayıcı özetleri gibi çeşitli denetimlerde başarısız olan değişkenler, durumlar ve veri değerleri listelerini üretir.

**Ağırlıklar.** Yordam, ağırlık değişkeni belirtimini yoksayar ve bunun yerine bunu başka bir çözümleme değişkeni olarak değerlendirir.

Verilerin Geçerliliğini Denetlemek İçin

1. Menülerden şunları seçin:

**Veri > Geçerlilik Denetimi > Verilerin Geçerliliğini Denetle ...**

2. Temel değişken denetimlerine göre ya da tek değişkenli geçerlilik denetimi kurallarına göre geçerlilik denetimi için bir ya da daha çok çözümleme değişkeni seçin.

Diğer bir seçenek olarak, şunları yapabilirsiniz:

3. **Çapraz Değişken Kuralları** sekmesini tıklatın ve bir ya da daha çok çapraz değişken kuralı uygulayın.

İsteğe bağlı olarak şunları yapabilirsiniz:

- Yinelenen ya da tamamlanmamış tanıtıcıları denetlemek için bir ya da daha çok vaka tanımlaması değişkeni seçin. Vaka tanıtıcısı değişkenleri, kasebilly çıkışı etiketlemek için de kullanılır. İki ya da daha çok vaka tanıtıcısı değişkeni belirtilirse, değerlerinin birleşimi vaka tanıtıcısı olarak kabul edilir.

Bilinmeyen Ölçüm Düzeyine Sahip Alanlar

Veri kümesindeki bir ya da daha fazla değişkene (alanlar) ilişkin ölçüm düzeyi bilinmiyorsa Ölçüm Düzeyi uyarısı görüntülenir. Ölçüm düzeyi, bu yordama ilişkin sonuçların hesaplamasını etkilediğinden, tüm değişkenlerin tanımlanmış bir ölçüm düzeyine sahip olması gerekir.

**Verileri Tara.** Etkin veri kümesindeki verileri okur ve şu anda bilinmeyen bir ölçüm düzeyiyle herhangi bir alana varsayılan ölçü düzeyi atar. Veri kümesi büyükse, bu işlem biraz zaman alabilir.

**El ile ata.** Bilinmeyen ölçüm düzeyine sahip tüm alanları listeleyen bir iletişim kutusu açar. Bu alanlara ölçüm düzeyi atamak için bu iletişim kutusunu kullanabilirsiniz. Ayrıca, Veri Düzenleyici 'nin Değişken Görünümü 'nde ölçüm düzeyi de atayabilirsiniz.

Ölçüm düzeyi bu yordam için önemli olduğundan, tüm alanlar tanımlı bir ölçüm düzeyine sahip oluncaya kadar bu yordamı çalıştırmak için iletişim kutusuna erişemezsiniz.

## Veri Temel Denetimlerini Doğrula

Temel Denetimler sekmesi, çözümleme değişkenleri, vaka tanıtıcıları ve tüm vakalar için temel denetimleri seçmenize olanak tanır.

**Çözümleme Değişkenleri.** Değişkenler sekmesinde herhangi bir analiz değişkeni seçtiyseniz, bunların geçerliliğinin denetlenmesini istediğiniz herhangi bir denetlemeyi seçebilirsiniz. Onay kutusu, çekleri açık ya da kapalı olarak açmanızı sağlar.

- **Eksik değer yüzdesi üst sınırı.** Belirtilen değerden büyük eksik değerlerin yüzdesine sahip raporlar analiz değişkenleri. Belirtilen değer, 100 'den küçük ya da 100 'e eşit bir pozitif sayı olmalıdır.
- **Tek bir kategorindeki vakaların yüzdesi üst sınırı.** Herhangi bir analiz değişkeni kategorik durumdaysa, bu seçenek, eksik olmayan tek bir kategoriyi belirtilen değerden büyük olan vakaların yüzdesine sahip kategorik analiz değişkenlerini bildirir. Belirtilen değer, 100 'den küçük ya da 100 'e eşit bir pozitif sayı olmalıdır. Yüzdeler, değişkenin eksik olmayan değerleri olan vakalara dayalıdır.
- **Kategori sayısı üst sınırı 1 olan kategori yüzdesi.** Herhangi bir analiz değişkeni kategorik durumdaysa, bu seçenek, yalnızca bir vakayı içeren değişken kategorilerinin yüzdesinin belirtilen değerden büyük olduğu kategorik analiz değişkenlerini bildirir. Belirtilen değer, 100 'den küçük ya da 100 'e eşit bir pozitif sayı olmalıdır.
- **Varyasyon katsayısı alt sınırı.** Herhangi bir analiz değişkeni ölçekse, bu seçenek ölçek çözümleme değişkenlerini, varyasyon katsayısının mutlak değerinin belirtilen değerden küçük olduğu raporlar ölçeklendirir. Bu seçenek yalnızca, ortanın sıfır olmayan değişkenler için geçerlidir. Belirtilen değer eksi olmayan bir sayı olmalıdır. 0 değerinin belirlenmesi, varyasyon katsayısını geri çevirir.



- **Standart sapma alt sınırı.** Herhangi bir analiz değişkeni ölçekse, bu seçenek, standart sapması belirtilen değerden küçük olan ölçek çözümlene değişkenlerini bildirir. Belirtilen değer eksi olmayan bir sayı olmalıdır. 0 belirtildiğinde standart sapma denetimi kapatılıyor.

**Vaka Tanıtıcıları.** Değişkenler sekmesinde herhangi bir vaka tanıtıcısı değişkeni seçtiyseniz, bunların geçerliliğinin denetlenmesini istediğiniz herhangi bir denetim ögesini seçebilirsiniz.

- **Tamamlanmamış tanıtıcıları işaretle.** Bu seçenek, eksik vaka tanıtıcılarına sahip vakaları bildirir. Belirli bir vaka için, herhangi bir tanıtıcı değişkeninin değeri boş ya da eksikse, tanıtıcı eksik olarak kabul edilir.
- **Yinelenen tanıtıcıları işaretle.** Bu seçenek, yinelenen vaka tanıtıcılarına sahip vakaları bildirir. Eksik tanıtıcılar olası yinelemeler kümesinden dışlanır.

**Boş vakaları işaretle.** Bu seçenek, tüm değişkenlerin boş ya da boş olduğu durumları bildirir. Boş vakaları tanımlamak amacıyla, dosyadaki tüm değişkenleri (herhangi bir tanıtıcı değişkeni dışında) ya da yalnızca Değişkenler sekmesinde tanımlanan analiz değişkenleri kullanmayı seçebilirsiniz.

## Veri Tek Değişken Kurallarını Doğrula

Tek Değişken Kuralları sekmesinde, var olan tek değişkenli doğrulama kuralları görüntülenir ve bunları çözümlene değişkenlerine uygulamanıza olanak sağlanır. Ek tek değişkenli kuralları tanımlamak için **Kuralları Tanımladüğmesini** tıklatın. Ek bilgi için "[Tek Değişkenli Kuralları Tanımla](#)" sayfa 2 konusuna bakın.

**Çözümlene Değişkenleri.** Liste, çözümlene değişkenlerini gösterir, dağıtımlarını özetler ve her bir değişkene uygulanan kural sayısını gösterir. Kullanıcı ve sistem eksik değerlerin özetlere dahil edilmediğini unutmayın. Görüntüleme açılan listesi, hangi değişkenlerin gösterildiğini denetler; **Tüm değişkenler, Sayısal değişkenler, Dizgi değişkenler** ve **Tarih değişkenleri**'nden seçim yapabilirsiniz.

**Kurallar.** Çözümlene değişkenlerine kurallar uygulamak için, bir ya da daha çok değişken seçin ve Kurallar listesinde uygulamak istediğiniz tüm kuralları işaretleyin. Kurallar listesi, yalnızca seçilen çözümlene değişkenlerine uygun kuralları gösterir. Örneğin, sayısal çözümlene değişkenleri seçildiyse, yalnızca sayısal kurallar gösterilir; bir dizgi değişkeni seçildiyse yalnızca dizgi kuralları gösterilir. Herhangi bir çözümlene değişkeni seçilmezse ya da karma veri tipleri varsa, hiçbir kural gösterilmez.

**Değişken Dağıtımları.** Çözümlene Değişkenleri listesinde gösterilen dağıtım özetleri, tüm vakalara ya da Vakalar metin kutusunda belirtildiği gibi ilk  $n$  vakanın taranmasını temel alır. **Yeniden Tara** düğmesinin tıklatılması, dağıtım özetlerini günceller.

## Veri Çapraz Değişkenini Doğrula-Değişken Kuralları

Çapraz Değişken Kuralları sekmesi, kullanılabilir çapraz değişken kurallarını görüntüler ve bunları verilerinize uygulamanıza olanak sağlar. Ek çapraz değişken kuralları tanımlamak için **Kuralları Tanımladüğmesini** tıklatın. Ek bilgi için "[Çapraz Değişken Kuralları Tanımla](#)" sayfa 3 konusuna bakın.

## Veri Çıkışını Doğrula

**Kasa Raporu.** Tek değişkenli ya da çapraz değişken geçerlilik denetimi kuralları uyguladıysanız, her bir vaka için geçerlilik denetimi kural ihlallerini listeleyen bir rapor isteyebilirsiniz.

- **Minimum İhlal Sayısı.** Bu seçenek, bir vaka için raporda yer almak için gereken kural ihlali sayısı alt sınırını belirtir. Artı bir tamsayı belirtin.
- **Vaka Sayısı Üst Sınırı.** Bu seçenek, vaka raporunda yer alan vaka sayısı üst sınırını belirtir. 1000 'den küçük ya da 1000 'e eşit bir artı tamsayı belirtin.

**Tek Değişken Geçerlilik Denetimi Kuralları.** Tek değişkenli geçerlilik denetimi kuralları uyguladıysanız, sonuçların nasıl görüntüleneceğini ya da bunların tümünü görüntüleyip görüntülemeyeceğini seçebilirsiniz.

- **Analiz değişkenine göre ihlalleri özetleyin.** Her analiz değişkeni için bu seçenek, ihlal edilen tüm tek değişkenli doğrulama kurallarını ve her kuralı ihlal eden değer sayısını gösterir. Ayrıca, her bir değişken için tek değişkenli kural ihlallerinin toplam sayısını da bildirir.

- **Kurallara göre ihlalleri özetleyin.** Her bir tek değişkenli doğrulama kuralı için bu seçenek, kuralı ihlal eden değişkenleri ve değişken başına geçersiz değer sayısını bildirir. Ayrıca, değişkenler arasında her bir kuralı ihlal eden toplam değer sayısını da bildirir.

**Analiz değişkenlerine ilişkin açıklayıcı istatistikleri görüntüleyin.** Bu seçenek, çözümleme değişkenlerine ilişkin açıklayıcı istatistikler istemenize olanak tanır. Her kategorik değişken için bir sıklık tablosu oluşturulur. Ölçek değişkenleri için ortalama, standart sapma, alt sınır ve üst sınır değerleri de dahil olmak üzere bir özet istatistikleri tablosu oluşturulur.

**Doğrulama kuralı ihlallerini içeren vakaları etkin veri kümesinin en üstüne taşıyın.** Bu seçenek, kolay nedeni için tek değişkenli ya da çapraz değişken kural ihlalleriyle etkin veri kümesinin üst kısmına taşınır.

## Veri Kaydetmeyi Doğrula

Sakla etiketi, kural ihlallerini etkin veri kümesine kaydetmiş olan değişkenleri kaydetmenizi sağlar.

**Özet Değişkenleri.** Bunlar, saklanabilen tek tek değişkenlerdir. Değişkeni kaydetmek için bir kutuyu işaretleyin. Değişkenlere ilişkin varsayılan adlar sağlanır; bunları düzenleyebilirsiniz.

- **Vaka göstergesi boş.** Boş vakalar 1 değerine atanır. Diğer tüm durumlar 0 kodlanmıştır. Değişken değerleri, Temel Denetimler sekmesinde belirtilen kapsamı yansıtır.
- **Yinelenen Tanıtıcı Grubu.** Aynı vaka tanıtıcısına sahip olan (tamamlanmamış tanıtıcılara sahip vakalardan başka) vakalar aynı grup numarasına atanır. Benzersiz ya da eksik tanıtıcılara sahip olan vakalar 0 kodlu kodlardır.
- **Eksik tanıtıcı göstergesi.** Boş ya da eksik vaka tanıtıcılarına sahip vakalara 1 değeri atanır. Diğer tüm durumlar 0 kodlanmıştır.
- **Doğrulama kuralı ihlalleri.** Bu, tek değişkenli ve değişken arası geçerlilik denetimi kuralı ihlallerinin kaseblik toplam sayısıdır.

**Var olan özet değişkenleri değiştirin.** Veri dosyasına kaydedilen değişkenlerin benzersiz adları olmalı ya da aynı adı taşıyan değişkenleri değiştirmelidir.

**Gösterge değişkenlerini sakla.** Bu seçenek, geçerlilik denetimi kuralı ihlallerinin eksiksiz bir kaydını saklamanızı sağlar. Her değişken bir doğrulama kuralının uygulanmasına karşılık gelir ve durum kuralı ihlal ederse 1 değerini ve değilse de 0 değerini içerir.

## Otomatikleştirilmiş Veri Hazırlığı

Analiz için veri hazırlama, herhangi bir projeden en önemli adımlardan biridir ve geleneksel olarak en çok zaman tüketimlerinden biridir. Otomatikleştirilmiş Veri Hazırlama (ADP), sizin için görevi işler, verilerinizi analiz eder ve düzeltmeleri tanımlar, sorunlu alanları tarar ya da yararlı olma olasılığı yüksek olan alanları tarar, uygun olduğunda yeni öznelikler üretir ve akıllı tarama teknikleri sayesinde performansı geliştirir. Algoritmayı tam olarak **otomatik** modasında kullanabilir, düzeltmeleri seçebilir ve uygulayabilir ya da **etkileşimli** modasında kullanabilir, önce değişiklikleri önceden görüntüleyerek ve bunları istediğiniz gibi kabul edebilir ya da reddedebilirsiniz.

Bu koşullarla, ilgili istatistiksel kavramların önceden bilgisine ihtiyaç duymadan, verilerinizi model oluşturma için hızlı ve kolay bir şekilde hazırlamaya hazır hale getirmenizi sağlar. Models will tend to build and score more quickly; in addition, using ADP improves the robustness of automated modeling processes.

*Not:* Ve çözümleme için bir alan hazırlandığında, eski alanın var olan değerlerini ve özelliklerini değiştirmek yerine, ayarlamaları ya da dönüştürmeleri içeren yeni bir alan yaratır. Eski alan daha fazla çözümlemede kullanılmaz; bu alan rolü Yok olarak ayarlanır. Ayrıca, kullanıcı eksik değer bilgilerinin bu yeni yaratılan alanlara aktarılmadığına ve yeni alanda eksik değerlerin sistem eksik olduğuna dikkat edin.

**Örnek.** Ev sahibinin sigorta poliçesini araştırmak için sınırlı kaynaklara sahip bir sigorta şirketi şüpheli ve potansiyel olarak dolandırıcı iddialar için bir model oluşturmak istiyor. Modeli oluşturmadan önce, otomatikleştirilmiş veri hazırlama kullanarak modelleme için verileri hazırlar. Dönüşümler uygulanmadan önce önerilen dönüşümleri gözden geçirebilmeleri için etkileşimli kipte otomatik veri hazırlama olanağını kullanacaklardır.

Bir otomotiv sanayi grubu, çeşitli kişisel motorlu taşıtlar satışlarını takip ediyor. Aşırı ve az performans gösteren modelleri tanımlayabilmek için, araç satış ve araç özellikleri arasında bir ilişki kurmak istiyorlar. Bunlar, çözümlenmeye ilişkin verileri hazırlamak için otomatikleştirilmiş veri hazırlığı ve sonuçların nasıl farklılık göstereceğini görmek için "önce" ve "sonra" hazırlanmasını kullanarak modeller oluşturmaları için kullanacaklardır.

**Amacınız nedir?** Otomatikleştirilmiş veri hazırlığı, diğer algoritmaların modelleri oluşturabileceği ve bu modellerin tahmine dayalı gücünü geliştirebileceği hızı etkileyecek veri hazırlama adımlarını önerir. Bu, özellikleri dönüştürmeyi, oluşturmayı ve seçmeyi içerebilir. Hedef de dönüştürülebilir. Veri hazırlama işleminin yoğunlaştıracağı model oluşturma önceliklerini belirtebilirsiniz.

- **Hız ve doğruluğu dengeleme.** Bu seçenek, veriler, model oluşturma algoritmaları ve öngörülerin doğruluğu ile işlendiği hızı eşit öncelik vermek için verileri hazırlar.
- **Hız için optimize edin.** Bu seçenek, verileri, model oluşturma algoritmaları tarafından hangi verilerin işlendiği hızı öncelik verecek şekilde hazırlar. Çok büyük veri kümeleriyle çalışırken ya da hızlı bir yanıt aradığınızda bu seçeneği belirleyin.
- **Doğruluk için optimize edin.** Bu seçenek, veriler, model oluşturma algoritmaları tarafından üretilen öngörülerin doğruluğuna öncelik verecek şekilde hazırlar.
- **Özel analiz.** Ayarlar sekmesinde algoritmayı el ile değiştirmek istediğinizde bu seçeneği belirleyin. Daha sonra, diğer hedeflerden biriyle uyumsuz olan Ayarlar sekmesinde seçeneklerde değişiklik yaparsanız, bu ayarın otomatik olarak seçildiğini unutmayın.

## Otomatik Veri Hazırlığı Elde Etmek İçin

Menülerden şunları seçin:

1. Menülerden şunları seçin:

**Dönüştürme > Modelleme İçin Veri Hazırla > Otomatik ...**

2. **Çalıştır**'ı tıklatın.

İsteğe bağlı olarak şunları yapabilirsiniz:

- Hedef sekmesinde bir hedef belirtin.
- Alanlar sekmesinde alan atamalarını belirtin.
- Ayarlar sekmesinde uzman ayarları belirtin.

## Etkileşimli Veri Hazırlığı Elde Etmek İçin

1. Menülerden şunları seçin:

**Dönüştürme > Modelleme İçin Veri Hazırla > Etkileşimli ...**

2. İletişim kutusunun üst kısmındaki araç çubuğunda **Çözümle** 'yi tıklatın.
3. Çözümleme sekmesini tıklatın ve önerilen veri hazırlama adımlarını gözden geçirin.
4. Memnunsanız, **Çalıştır** 'ı tıklatın. Aksi takdirde, **Analizi Temizle**'yi tıklatın, istediğiniz ayarları değiştirin ve **Analiz Et** 'i tıklatın.

İsteğe bağlı olarak şunları yapabilirsiniz:

- Hedef sekmesinde bir hedef belirtin.
- Alanlar sekmesinde alan atamalarını belirtin.
- Ayarlar sekmesinde uzman ayarları belirtin.
- Save the suggested data preparation steps to an XML file by clicking **XML 'i Sakla**.

## Alanlar Sekmesi

Alanlar sekmesi, ek çözümlenmeler için hangi alanların hazırlanabileceğini belirtir.

**Önceden tanımlanmış rolleri kullan.** Bu seçenek, varolan alan bilgilerini kullanır. Hedef olarak bir role sahip tek bir alan varsa, hedef olarak kullanılır; tersi durumda hedef olmaz. Giriş olarak önceden tanımlanmış bir role sahip tüm alanlar giriş olarak kullanılır. En az bir giriş alanı gerekli.

**Özel alan atamalarını kullanın.** Alanları varsayılan listelerinden taşıyarak alan rollerini geçersiz kılarken, iletişim kutusu otomatik olarak bu seçeneğe geçer. Özel alan atamaları yapılırken, aşağıdaki alanları belirtin:

- **Hedef (isteğe bağlı).** Hedef gerektiren modeller oluşturmayı planlıyorsanız, hedef alanı seçin. Bu, alan rolünü Hedefe ayarlamaya benzer.
- **Girişler.** Bir ya da daha çok giriş alanı seçin. Bu, alan rolünün Giriş olarak ayarlanmasını benzer bir şekilde sağlar.

## Ayarlar Etiketi

Settings (Ayarlar) sekmesi, algoritmanın verilerinizi nasıl işlediğini ince ayarlamaya göre değiştirebileceğiniz çeşitli ayarlar gruplarından oluşur. If you make any changes to the default settings that are incompatible with the other objectives, the Objective tab is automatically updated to select the **Çözümlemeyi özelleştirme** option.

## Tarihleri ve Tarihleri Hazırla

Birçok modelleme algoritması, tarih ve saat ayrıntılarını doğrudan işleyemez; bu ayarlar, var olan verilerinizdeki tarih ve saatlerden model girişleri olarak kullanılacak yeni süre verilerini üretmenizi sağlar. Tarih ve saat içeren alanların tarih ya da saat depolama tipleriyle önceden tanımlanmış olması gerekir. Otomatik veri hazırlığını izleyen model girişleri olarak özgün tarih ve saat alanları önerilmez.

**Modellemeye ilişkin tarih ve saatleri hazırlayın.** Bu seçeneğin seçimini kaldırırken, seçimleri korurken diğer tüm Prepare Tarihler & Times denetimleri devre dışı bırakılır.

**Başvuru tarihine kadar geçen süre hesaplanması.** Bu, tarihleri içeren her bir değişken için başvuru tarihinden bu yana yıl/ay/gün sayısını oluşturur.

- **Başvuru Tarihi.** Girişin, giriş verilerindeki tarih bilgileriyle ilgili olarak hesaplanacağı tarihi belirtin. **Bugün tarihi** 'nin seçilmesi, yürürlükteki sistem tarihinin her zaman ve ADP yürütüldüğünde kullanıldığı anlamına gelir. Belirli bir tarihi kullanmak için, **Sabit tarih** 'i seçin ve gerekli tarihi girin.
- **Tarih Süresi için Birim Sayısı.** Koşulun otomatik olarak tarih süre biriminde karar verip vermeyeceğini belirtin ya da Yıllar, Ay ya da Gün sayısı **Sabit birimler** 'den seçin.

**Başvuru saatine kadar geçen süre hesapla.** Bu, her değişken için bir başvuru sürelerinin bulunduğu saatler/dakika/dakika/saniye sayısını üretir.

- **Başvuru Zamanı.** Girişin, giriş verilerindeki saat bilgileriyle ilgili olarak ne zaman hesaplanacağı belirtilir. **Yürürlükteki saat** ögesinin seçilmesi, yürürlükteki sistem saatinin her zaman ve ADP yürütüldüğünde kullanıldığı anlamına gelir. Belirli bir zamanı kullanmak için, **Sabit saat** 'i seçin ve gerekli ayrıntıları girin.
- **Zaman Süresi için Birim Sayısı.** Koşulun otomatik olarak zaman aralığı üzerinde karar verip vermeyeceğini belirtin ya da Saat, Dakika ya da Saniye **Sabit birimler** 'den seçim yapın.

**Döngüsel Zaman Ögelerini Çek.** Tek bir tarih ya da saat alanını bir ya da daha çok alana bölmek için bu ayarları kullanın. Örneğin, üç tarih onay kutusunu işaretlediyseniz, "1954-05-23" giriş tarihi alanı üç alana bölünür: 1954, 5 ve 23, her biri **Alan Adları** panosunda tanımlanan soneki kullanır ve özgün tarih alanı yoksayıdır.

- **Tarihlerden çıkar.** Herhangi bir tarih girişi için, yıl, ay, gün ya da herhangi bir birleşimin çekilmesini isteyip istemediğinizi belirtin.
- **Zaman tan çıkarın.** Herhangi bir zaman girişi için, saat, dakika, saniye ya da herhangi bir birleşimin ayıklamayı isteyip istemediğinizi belirtin.

## Alanları Dışla

Düşük kaliteli veriler öngörülerinizin doğruluğunu etkileyebilir; bu nedenle, giriş özellikleri için kabul edilebilir kalite düzeyini belirleyebilirsiniz. Sabit ya da %100 eksik olan tüm alanlar otomatik olarak dışlanır.

**Düşük kaliteli giriş alanlarını dışla.** Bu seçeneğin seçimini kaldırarak, seçimleri korurken diğer tüm Dışlama Alanları denetim öğelerini devre dışı bırakır.

**Çok fazla eksik değer içeren alanları dışla.** Belirtilen yüzdeden fazla eksik değer içeren alanlar, ek çözümlenmelerden kaldırılır. 0 'dan büyük ya da 0 'a eşit bir değer belirleyin; bu seçenek, tüm eksik değerlere sahip alanlar otomatik olarak dışlansa da, bu seçeneğin belirlenmesine eşit ve 100 'den küçük ya da 100 'e eşit bir değer belirleyin. Varsayılan değer 50 'dir.

**Çok fazla benzersiz kategori içeren nominal alanları dışlayın.** Belirtilen sayıda kategori sayısından fazla olan nominal alanlar, ek çözümlenmelerden kaldırılır. Artı bir tamsayı belirtin. Varsayılan değer 100 değeridir. Tanıtıcı, adres ya da ad gibi, modelmeden kayıt benzersiz bilgileri içeren alanların otomatik olarak kaldırılması için bu olanak yararlı olur.

**Tek bir kategoride çok fazla değer içeren kategorik alanları dışla.** Belirtilen yüzdelerden daha fazla kaydın bulunduğu kategori içeren sıralı ve nominal alanlar, ek çözümlenmelerden kaldırılır. Değişmez alanlar otomatik olarak dışlansa da, bu seçeneğin belirlenmesine eşit ya da 0 'dan büyük bir değer ve 100 'den küçük ya da 100 'e eşit bir değer belirleyin. Varsayılan değer 95 'tür.

## Ölçümü Ayarla

**Ölçüm düzeyini ayarlayın.** Bu seçeneğin seçimini kaldırırken, seçimleri korurken diğer tüm Ayarlama Ölçü denetimleri devre dışı bırakılır.

**Ölçüm Düzeyi.** "Çok az" değerine sahip sürekli alanların ölçü düzeyinin, ordinal olarak ayarlanabileceğini ve "çok fazla" değerine sahip sıra alanlarının sürekli olarak ayarlanıp ayarlanabileceğini belirtin.

- **Sıra alanları için değer sayısı üst sınırı.** Belirtilen sayıda kategori sayısından daha fazla sıralı alan, sürekli alan olarak yeniden yayınlanır. Artı bir tamsayı belirtin. Varsayılan değer 10 'dur. Bu değer, sürekli alanlar için değer sayısı alt sınırından büyük ya da bu değere eşit olmalıdır.
- **Sürekli alanlar için değer sayısı alt sınırı.** Belirtilen sayıda benzersiz değere sahip sürekli alanlar, sıra alan olarak yeniden yayınlanır. Artı bir tamsayı belirtin. Varsayılan 5'tir. Bu değer, sıra alanları için değer sayısı üst sınırından küçük ya da bu değere eşit olmalıdır.

## Veri Kalitesini İyileştirin

**Veri kalitesini iyileştirmek için alanları hazırlayın.** Bu seçeneğin seçimini kaldırırken, seçimleri korurken diğer tüm Veri Kalitesi İyileştirme denetimleri devre dışı bırakılır.

**Dış İşleme.** Girişler ve hedef için aykırı değerler değiştirilip değiştirilmeyeceğini belirtin; eğer öyleyse, standart sapmalar ile ölçülen bir aykırı kesme ölçütü ve aykırı değerleri değiştirmek için bir yöntem belirtin. Aykırı değerler, kırpmaya ayarlanarak (kesme değerine ayarlanarak) ya da bunları eksik değerler olarak tanımlanarak değiştirilebilir. Eksik değerler için ayarlanan aykırı değerler, aşağıda seçilen eksik değer işleme ayarlarını izler.

**Eksik Değerleri Değiştir.** Sürekli, nominal ya da sıralı alanların eksik değerlerinin değiştirilip değiştirilmeyeceğini belirtin.

**Nominal Alanları Yeniden Sipariş Edin.** Nominal (set) alanların değerlerini en küçük (en az sıklıkla oluşan) kategoriden en büyük (en sık oluşan) kategoriye yeniden kodlamak için bu seçeneği belirleyin. Yeni alan değerleri, en az sık kullanılan kategori olarak 0 ile başlar. Özgün alan bir dizgi olsa da yeni alanın sayısal olacağını unutmayın. Örneğin, nominal bir alanın veri değerleri "A", "A", "A", "B", "C", "C" ise, otomatikleştirilmiş veri hazırlığı "B" değerini 0, "C", 1 'e ve "A" ile 2 'ye yeniden kodlayacaktır.

## Rescale Alanları

**Alanları yeniden ölçekle.** Bu seçeneğin seçimini kaldırırken, seçimleri korurken diğer tüm Rescale Fields denetimleri devre dışı bırakılır.

**Çözümleme Ağırlığı.** Bu değişken çözümleme (regresyon ya da örnekleme) ağırlıkları içerir. Analiz ağırlıkları, hedef alanın düzeylerindeki varyans farklarını dikkate almak için kullanılır. Sürekli bir alan seçin.

**Sürekli Giriş Alanları.** Bu, **z-score dönüştürme** ya da **min/max dönüştürme** kullanarak sürekli giriş alanlarının normalleşmesini sağlar. Seçilen ve Yapıt ayarlarında **Özellik yapımını gerçekleştir** seçeneğini belirlediğinizde, girişleri yeniden ölçekleme özelliği yararlı olur.

- **Z-score dönüştürmesi.** Gözlemlenen ortalama ve standart sapmayı, nüfus parametre tahminleri olarak kullanarak, alanlar standartlaştırılır ve daha sonra z puanları, belirtilen **Son ortalama** ve **Son standart sapma** ile normal bir dağılımın karşılık gelen değerlerine eşlenir. **Son ortalama** için bir sayı ve **Son standart sapma** için pozitif bir sayı belirtin. Varsayılan değerler, standartlaştırılmış yeniden ölçekleme için karşılık gelen 0 ve 1 'tür.
- **Dönüştürme alt sınırı/üst sınırı.** Popülasyon parametresi tahminleri olarak gözlenen minimum ve maksimum değerleri kullanarak alanlar, belirtilen **Alt Sınır** ve **Üst Sınır** değerlerine sahip bir tek tip bir dağılımın karşılık gelen değerleriyle eşleştirilir. Specify numbers with **Üst Sınır** greater than **En Az**.

**Sürekli Hedef.** Bu, Box-Cox dönüşümünü kullanarak sürekli bir hedefi, belirtilen **Son ortalama** ve **Son standart sapma** ile normal bir dağılıma sahip bir alana dönüştürür. **Son ortalama** için bir sayı ve **Son standart sapma** için pozitif bir sayı belirtin. Varsayılan değer, sırasıyla 0 ve 1 'tür.

*Not:* Hedef, dönüştürülen hedef puanı kullanılarak oluşturulan sonraki modeller ve koşullarıyla dönüştürülmüşse, dönüştürülen birimleri kullanın. Sonuçları yorumlamak ve kullanmak için, tahmin edilen değeri özgün ölçeğe dönüştürmeniz gerekir. Ek bilgi için konusuna bakın. Ek bilgi için "[Geri Dönüştürme Puanları](#)" sayfa 18 konusuna bakın.

## Alanları Dönüştür

Verilerinizin tahmine dayalı gücünü artırmak için, giriş alanlarını dönüştürebilirsiniz.

**Modelleme için alanı dönüştürün.** Bu seçeneğin seçimini kaldırırken, seçimleri korurken diğer tüm Dönüştürme Alanları denetim öğeleri de geçersiz kılınır.

**Kategorik Giriş Alanları** Aşağıdaki seçenekler kullanılabilir:

- **Hedefle ilişkilendirmeyi en üst düzeye çıkarmak için seyrek kategorileri birleştirin.** Hedefle ilişkili olarak işlenecek alan sayısını azaltarak, daha çok parsimonious bir model yapmak için bu seçeneği belirleyin. Benzer kategoriler, giriş ve hedef arasındaki ilişkiye dayalı olarak tanımlanır. Önemli ölçüde farklı olmayan (yani, belirtilen değerden daha büyük bir *p* değeri olması) kategoriler birleştirilir. 0 'dan büyük ve 1 'den küçük ya da 1 'e eşit bir değer belirtin. Tüm kategoriler birleştirilir ise, alanın özgün ve türetilmiş sürümleri, bir karşılaştırma belirtimi olarak herhangi bir değeri olmadıkları için ek çözümler kapsamından çıkarılır.
- **Hedef olmadığında, sayılara göre seyrek kategorileri birleştirin.** Veri kümesinin hedefi yoksa, sıra ve nominal alanların seyrek kategorilerini birleştirmeyi seçebilirsiniz. Eşit sıklık yöntemi, kategorileri, toplam kayıt sayısı alt sınırından daha küçük bir yüzdeyle birleştirmek için kullanılır. 0 'dan büyük ya da 0 'a eşit ve 100 'den küçük ya da 100 'e eşit bir değer belirtin. Varsayılan değer 10 'dur. Belirtilen toplam vaka sayısı alt sınırından daha küçük olan kategoriler yoksa ya da yalnızca iki kategori kaldığında birleştirme durdurulur.

**Sürekli Giriş Alanları.** Veri kümesi bir kategorik hedef içeriyorsa, işlem performansını artırmak için güçlü ilişkilendirmeler ile sürekli girişleri bölebilirsiniz. Kutular, türdeş altkümelerin belirlenmesi için kritik değerin alfa olarak belirtilen *p* değerini kullanarak Scheffe yöntemiyle tanımlanan "türdeş altkümeler" özelliklerine dayalı olarak oluşturulur. 0 'dan büyük ve 1 'den küçük ya da 1 'e eşit bir değer belirtin. Varsayılan değer 0.05 'tür. Sabitleme işlemi belirli bir alan için tek bir bölmede sonuçlanırsa, bu alanın özgün ve ikili sürümleri, bir karşılaştırma belirtimi olarak herhangi bir değeri olmadıkları için dışlanır.

*Not:* Sabitleme, en uygun şekilde sabitleme açısından farklılık gösterir. Optimal binning, sürekli bir alanı kategorik alana dönüştürmek için entropi bilgilerini kullanır; bu, verileri sıralamaya ve belleğe tüm belleğe

saklamaya gerek duyar. Özdeş altkümelerin, sürekli bir alanı bölmek için kullanılır. Bu, verileri sıralamak zorunda kalmaması ve tüm verileri belleğiyle saklamaması anlamına gelen sürekli bir alan anlamına gelir. Türdeş altküme yönteminin sürekli bir alan için kullanılması, biniş sonrası kategori sayısının her zaman hedefteki kategori sayısına eşit ya da daha küçük olduğu anlamına gelir.

## Seç ve Yapıt

Verilerinizin tahmine dayalı gücünü artırmak için, var olan alanlara dayalı olarak yeni alanlar oluşturabilirsiniz.

**Özellik seçimini gerçekleştir.** Hedef ile ilintisinin  $P$ değeri, belirtilen  $P$ -değerden büyükse, çözümden sürekli bir giriş kaldırılır.

**Özellik yapımını gerçekleştir.** Varolan bazı özelliklerin birleşiminden yeni özellikler elde etmek için bu seçeneği belirleyin. Eski özellikler, ek çözümlemede kullanılmaz. Bu seçenek yalnızca hedefin sürekli olduğu ya da hedefin olmadığı sürekli giriş özellikleri için geçerlidir.

## Alan Adları

Yeni ve dönüştürülen özellikleri kolayca tanımlamak, temel yeni adları, önekleri ya da sonekleri yaratır ve uygular. Bu adların, kendi gereksinimlerinize ve verilerinize daha uygun olması için, bu adları değiştirebilirsiniz.

**Dönüştürülen ve Yapılanmış Alanlar.** Dönüştürülen hedef ve giriş alanlarına uygulanacak ad uzantılarını belirtin.

Ayrıca, Seç ve Yapıt ayarları aracılığıyla oluşturulan herhangi bir özelliğe uygulanacak önek adını da belirtin. Yeni ad, bu önek kök adına sayısal bir sonek eklenerek yaratılır. Sayının biçimi, türetilmiş yeni özelliklerin sayısına göre değişir. Örneğin:

- 1-9 oluşturulan özellikler feature1 - feature9olarak adlandırılacaktır.
- 10-99 yapılı özellikler: feature01 - feature99olarak adlandırılır.
- 100-999 oluşturulan özellik şu adı taşır: feature001 - feature999, vb.

Bu, oluşturulan özelliklerin ne kadar fazla olursa olsun, mantıklı bir düzende sıralanmasını sağlar.

**Tarihler ve Saatlerden Hesaplanan Süreler.** Her iki tarih ve saat tarafından hesaplanan sürelere uygulanacak ad uzantılarını belirtin.

**Tarihler ve Saatlerden Alınan Döngüsel Öğeler.** Her iki tarih ve saatten çıkartılan döngüsel öğelere uygulanacak ad uzantılarını belirtin.

## Dönüşüm Uygulama ve Kaydetme

Etkileşimli ya da Otomatik Veri Hazırlama iletişim kutularını kullanıp kullanmamanıza bağlı olarak, dönüştürmeleri uygulama ve kaydetme ayarları biraz farklı olur.

Etkileşimli Veri Hazırlık Uygulama Dönüştürmeleri Ayarları

**Dönüştürülen Veriler.** Bu ayarlar, dönüştürülen verilerin nereye kaydedileceğini belirtir.

- **Etkin veri kümesine yeni alanlar ekleyin.** Otomatik veri hazırlığı tarafından oluşturulan her alan, etkin veri kümesine yeni alanlar olarak eklenir. **Analiz edilen alanlarla ilgili rolleri güncelle**, otomatikleştirilmiş veri hazırlığı tarafından daha fazla analiz kapsamı dışında bırakılan alanlar için rolü Yok olarak ayarlar.
- **Dönüştürülen verileri içeren yeni bir veri kümesi ya da dosya yaratın.** Otomatikleştirilmiş veri hazırlığı tarafından önerilen alanlar, yeni bir veri kümesine ya da dosyaya eklenir. **Çözümlememiş alanları içer** seçeneği, özgün veri kümesine Alanlar etiketinde belirtilmeyen alanları yeni veri kümesine ekler. Tanıtıcı ya da adres gibi, modellemede kullanılmayan bilgileri içeren alanların ya da adın yeni veri kümesine aktarılmasına olanak sağlar.

Otomatik Veri Hazırlık Uygulama ve Saklama Ayarları

Dönüştürülen Veri grubu, Etkileşimli Veri Hazırlığı 'nda olduğu gibi. Otomatik veri hazırlığında aşağıdaki ek seçenekler kullanılabilir:

**Dönüşümleri uygulayın.** Otomatik Veri Hazırlığı iletişim kutularında, bu seçeneğin belirlenmesi, seçimleri korurken diğer tüm Apply ve Save denetim öğelerini devre dışı bırakır.

**Dönüşümleri sözdizimi olarak kaydedin.** Bu işlem, önerilen dönüşümleri komut sözdizimi olarak bir dış dosyaya kaydeder. Etkileşimli Veri Hazırlık iletişim kutusu, **Yapıştırdüğmesini** tıklatırsanız, dönüşümleri komut sözdizimi penceresine yapıştıracağı için bu denetim öğesi yoktur.

**Dönüşümleri XML olarak kaydedin.** Bu, önerilen dönüşümleri XML olarak bir dış dosyaya kaydeder; bu, TMS MERGE kullanılarak model PMML ile birleştirilebilir ya da TMS IMPORT kullanılarak başka bir veri kümesine uygulanabilir. İletişim kutusunun üst kısmındaki araç çubuğunda **XML Kaydet** ' i tıklatırsanız, Etkileşimli Veri Hazırlık iletişim kutusunda bu denetim öğesi yoktur. XML olarak dönüştürmeleri kaydedecektir.

## Çözümleme Etiketi

*Not:* Çözümleme sekmesi, önerilen dönüşümleri gözden geçirmenize olanak sağlamak için Etkileşimli Veri Hazırlığı iletişim kutusunda kullanılır. Otomatik Veri Hazırlık iletişim kutusu bu adımı içermez.

1. Hedef, Alanlar ve Ayarlar sekmelerinde yapılan değişiklikler de dahil olmak üzere, ADP ayarlarından memnun olduğunuz zaman, **Verileri Çözümle** ' i tıklatın; algoritma veri girişlerine ayarlar uygular ve Çözümleme sekmesinde sonuçları görüntüler.

Çözümleme etiketi, verilerinizin işlenmesini özetleyen sekmeli ve grafik çıkışı içerir ve verilerin puanlama için nasıl değiştirilebileceğini ya da iyileştirilebileceğini özetleyen öneriler görüntüler. Daha sonra bu önerileri gözden geçirebilir ve reddedebilir ya da reddedebilirsiniz.

Çözümleme sekmesi iki panodan oluşur, soldaki ana görünüm ve bağlantılı ya da yardımcı, sağ görünümde yer alan görünüm. Üç ana görünüm vardır:

- Alan İşleme Özeti (varsayılan). Ek bilgi için [“Alan İşleme Özeti ” sayfa 12](#) başlıklı konuya bakın.
- Alanlar. Ek bilgi için [“Alanlar ” sayfa 13](#) başlıklı konuya bakın.
- İşlem Özeti. Ek bilgi için [“İşlem Özeti ” sayfa 14](#) başlıklı konuya bakın.

Dört bağlantılı/yardımcı görünüm vardır:

- Tahmine Dayalı Güç (varsayılan). Ek bilgi için [“Tahmine Dayalı Güç ” sayfa 14](#) başlıklı konuya bakın.
- Alanlar Tablosu. Ek bilgi için [“Alanlar Tablosu ” sayfa 14](#) başlıklı konuya bakın.
- Alan Ayrıntıları. Ek bilgi için [“Alan Ayrıntıları ” sayfa 15](#) başlıklı konuya bakın.
- İşlem Ayrıntıları. Ek bilgi için [“İşlem Ayrıntıları ” sayfa 16](#) başlıklı konuya bakın.

Görünümler arasındaki bağlantılar

Ana görünümde, tablolaradaki altı çizili metin, bağlantılı görünümdeki görüntüyü kontrol eder. Metnin tıklatılması, belirli bir alana, alan kümesine ya da işleme adımına ilişkin ayrıntıları almanıza olanak sağlar. Son seçtiğiniz bağlantı daha koyu bir renkte gösterilir; bu, iki görünüm panelinin içeriği arasındaki bağlantıyı belirlemenize yardımcı olur.

Görünümlerin ilk durumuna getirilmesi

Özgün çözümleme önerilerini yeniden görüntülemek ve Analiz görünümünde yaptığınız değişiklikleri iptal etmek için ana görünüm panosunun altındaki **İlk Durumuna Getir** düğmesini tıklatın.

## Alan İşleme Özeti

Field Processing Summary (Alan İşleme Özeti) çizelgesi, özelliklerin durumuna ve oluşturulan özellik sayısına ilişkin değişiklikler de içinde olmak üzere, işlenmenin genel olarak yansıtılması için bir anlık görüntü verir.



Hiçbir modelin gerçekte oluşturulmadığına dikkat edin, bu nedenle veri hazırlamadan önce ve sonra genel tahmine dayalı güçte değişikliğe ilişkin bir ölçü ya da grafik yoktur; bunun yerine, önerilen tek tek tahmin edicilerin tahmine dayalı gücünün grafiklerini görüntüleyebilirsiniz.

Tabloda, aşağıdaki bilgiler görüntülenir:

- Hedef alan sayısı.
- Özgün (input) tahmin edicilerinin sayısı.
- Öngörüler, çözümlemede ve modellemede kullanılmak üzere önerilir. Bu, önerilen toplam alan sayısını içerir; özgün, dönüştürülemez, önerilen alan sayısı; önerilen dönüştürülen alan sayısı (herhangi bir alanın ara sürümleri hariç olmak üzere, tarih/saat önyüklerinden türetilen alanlar ve oluşturulan tahmin ediciler); tarih/saat alanlarından türetilen alanların sayısı ve önerilen oluşturulan karşılaştırma belirtileri sayısı.
- Özgün formlarında türetilmiş bir alan olarak ya da oluşturulan bir öngörüleme girişi olarak, herhangi bir biçimde kullanılmak üzere önerilmeyen giriş karşılaştırma belirtilerinin sayısı.

**Alanlar** bilgilerinde altı çizili olarak, bağlantılı bir görünümde daha fazla ayrıntı görüntülemek için tıklanın. Details of the **Hedef**, **Giriş özellikleri**, and **Giriş özellikleri kullanılmadı** are shown in the Fields Table linked view. Ek bilgi için “[Alanlar Tablosu](#)” sayfa 14 konusuna bakın. **Çözümlemede kullanılmak üzere önerilen özellikler** are displayed in the Predictive Power linked view. Ek bilgi için “[Tahmine Dayalı Güç](#)” sayfa 14 konusuna bakın.

## Alanlar

Alanlar ana görünümü, işlenmiş alanları ve aşağı akış modellerinde bunları kullanıp kullanmayacağınızı gösterir. Herhangi bir alan için öneriyi geçersiz kılabilirsiniz; örneğin, oluşturulan özellikleri dışlamak ya da kapsam dışı olarak önerilebilecek özellikleri içeren özellikleri içermek. Bir alan dönüştürülürse, önerilen dönüştürmeyi kabul edip etmemeye karar verebilirsiniz ya da özgün sürümü kullanabilirsiniz.

Alanlar görünümü, biri hedef, diğeri işlenmiş ya da yaratılan tahmin ediciler için olmak üzere iki çizelgeden oluşur.

Hedef çizelge

**Hedef** çizelge yalnızca, verilerde bir hedef tanımlandıysa gösterilir.

Çizelgede iki kolon bulunur:

- **Ad.** Bu, hedef alanın adı ya da etiketidir; alan dönüştürülmüş olsa da, özgün ad her zaman kullanılır.
- **Ölçüm Düzeyi.** Bu, ölçüm düzeyini gösteren simgeyi görüntüler; verileri açıklayan bir etiketi (sürekli, sıra, nominal, vb.) görüntülemek için fareyi simgenin üzerinde gezdirin.

Hedef, **Ölçüm Düzeyi** sütunu dönüştürülmüş ise, dönüştürülen son sürümü yansıtır. *Not:* Hedef için dönüştürmeleri kapatamazsınız.

Tahmin ediciler tablosu

**Öngörüler** tablosu her zaman gösterilir. Çizelgenin her satırı bir alanı temsil eder. Varsayılan olarak, satırlar tahmine dayalı gücün azalan sırasına göre sıralanır.

Olağan özellikler için, özgün ad her zaman satır adı olarak kullanılır. Tarih/saat alanlarının hem özgün hem de türetilmiş sürümleri çizelgede (ayrı satırlarda) görüntülenir; tablo, oluşturulan karşılaştırma göstericileri de içerir.

Tabloda gösterilen alanların dönüştürülmüş sürümlerinin her zaman son sürümleri temsil ettiğini unutmayın.

Varsayılan olarak, Predictors (Önericiler) çizelgesinde yalnızca önerilen alanlar gösterilir. Kalan alanları görüntülemek için, tablonun üstündeki **Tablo içinde önerilmeyen alanları ekle** kutusunu işaretleyin; daha sonra bu alanlar tablonun alt kısmında görüntülenir.

Tablo aşağıdaki sütunları içerir:

- **Kullanılacak Sürüm.** Bu liste, bir alanın aşağı akışın kullanılıp kullanılmayacağını ve önerilen dönüşümlerin kullanılıp kullanılmayacağını denetleyen bir açılan liste görüntüler. Varsayılan olarak, açılan liste önerileri yansıtır.

Açılan listeyi dönüştüren sıradan tahmin ediciler için üç seçenek vardır: **Dönüştürülen, Özgün ve Kullanmayın.**

Dönüştürülemeyen sıradan tahmin ediciler için seçenekler şunlardır: **Özgün ve Kullanmayın.**

Türetilmiş tarih/saat alanları ve oluşturulan karşılaştırma belirtileri için seçenekler şunlardır: **Transformed ve Do not use.**

Özgün tarih alanları için açılan liste devre dışı bırakılır ve **Kullanmayın** olarak ayarlanır.

*Not:* Hem özgün hem de dönüştürülmüş sürümlere sahip tahmin edilenler için, **Özgün** ile **Dönüştürülen** sürümleri arasında değişen özellikler, bu özelliklere ilişkin **Ölçüm Düzeyi** ve **Tahmine dayalı** ayarlarını otomatik olarak güncelleştirir.

- **Ad.** Her alanın adı bir bağlantıdır. Bağlantı oluşturulan görünümde alana ilişkin daha fazla bilgi görüntülemek için bir adı tıklatın. Ek bilgi için "[Alan Ayrıntıları](#)" sayfa 15 başlıklı konuya bakın.
- **Ölçüm Düzeyi.** Bu, veri tipini temsil eden simgeyi görüntüler; verileri açıklayan bir etiketi (sürekli, sıra, nominal, vb.) görüntülemek için fareyi simgenin üzerinde gezdirin.
- **Tahmine dayalı Güç.** Tahmine dayalı güç yalnızca koşullarla ilgili olarak önerilen alanlar için görüntülenir. Hedef tanımlı bir hedef yoksa bu kolon görüntülenmez. Tahmine dayalı güç aralığı 0 'dan 1 'e, daha büyük değerler ise "daha iyi" tahmin edicileri belirtir. Genel olarak, tahmine dayalı güç, karşılaştırma belirtilerini bir ADP analiziyle karşılaştırmak için kullanışlıdır, ancak tahmine dayalı güç değerleri çözümler arasında karşılaştırılmamalı.

## İşlem Özeti

Otomatik veri hazırlığı tarafından gerçekleştirilen her işlem için, giriş karşılaştırma belirtileri dönüştürülür ve/veya süzgeçten geçirilir; bir eylemden sağ kurtulan alanlar bir sonrakinde kullanılır. En son adıma kadar devam eden alanlar, dönüştürülen ve oluşturulan öngörülebilirlerin girişleri dışarı süzölmüş olarak, modellemede kullanılmak üzere önerilir.

İşlem Özeti, koşullarla alınan işlem işlemlerini listeleyen basit bir tablodur. Herhangi bir **Eylem** ' in altı çizildiği durumlarda, yapılan işlemlerle ilgili bağlantılı bir görünümde daha fazla ayrıntı görüntülemek için tıklatın. Ek bilgi için "[İşlem Ayrıntıları](#)" sayfa 16 konusuna bakın.

*Not:* Çözümleme sırasında kullanılan ara düzey sürümleri değil, yalnızca her alanın özgün ve son dönüştürülen sürümleri gösterilir.

## Tahmine Dayalı Güç

Çözümleme ilk kez çalıştırıldığında ya da Alan İşleme Özeti ana görünümünde **Çözümlemede kullanılmak üzere öneriliciler önerilir** seçeneğini belirlediğinizde varsayılan olarak görüntülenen grafik, önerilen karşılaştırma belirtilerinin tahmine dayalı gücünü görüntüler. Alanlar, tahmine dayalı güce göre sıralanır ve en yüksek değere sahip alana sahip olarak en üstte yer alan alana göre sıralanır.

Olağan öngörülebilirlerin dönüştürülmüş sürümleri için alan adı, Ayarlar sekmesinin Alan Adları panosunda sonek seçiminizi yansıtır; örneğin: *\_dönüştürülen*.

Ölçüm düzeyi simgeleri, ayrı alan adlarından sonra görüntülenir.

Her bir önerilen önermenin tahmine dayalı gücü, hedefin sürekli veya kategorik olup olmadığına bağlı olarak doğrusal regresyon veya naif Bayes modelinden hesaplanır.

## Alanlar Tablosu

Alan İşleme Özeti ana görünümünde **Hedef, Önericiler** ya da **Kullanılamaz Öngörüler** ' i tıklattığınızda görüntülenen Alanlar Çizelgesi görünümünde, ilgili özellikleri listeleyen basit bir çizelge görüntülenir.

Çizelgede iki kolon bulunur:

- **Ad.** Tahmin ediliçi adı.

Hedefler için, hedef dönüştürülmüş olsa da, alanın özgün adı ya da etiketi kullanılır.

Olağan öngörülebilirlerin dönüştürülmüş sürümleri için, ad, Ayarlar sekmesinin Alan Adları panosunda sonek seçiminizi yansıtır; örneğin: *\_dönüştürülen*.

Tarih ve saatlerden türetilen alanlar için, dönüştürülen son sürümün adı kullanılır; örneğin: *bdate\_years*.

Oluşturulan tahmin ediliciler için, oluşturulan karşılaştırma belirtimlerinin adı kullanılır; örneğin: *Predictor1*.

- **Ölçüm Düzeyi.** Bu, veri tipini temsil eden simgeyi görüntüler.

Hedef için, **Ölçüm Düzeyi** her zaman dönüştürülen sürümü (hedef dönüştürülmüş ise) yansıtır; örneğin, sıra (sıralı küme) dan sürekli (aralık, ölçek) ya da tam tersi şekilde değiştirilir.

## Alan Ayrıntıları

Alanlar ana görünümünde herhangi bir **Ad** ' ı tıklattığınızda, Alan Ayrıntıları görünümünde seçilen alan için dağıtım, eksik değerler ve tahmine dayalı güç grafikleri (varsa) yer alır. Ayrıca, alana ilişkin işleme geçmiş ve dönüştürülen alanın adı da gösterilir (geçerliyse).

Her bir grafik kümesi için iki sürüm yan yana gösterilir; dönüştürmeler uygulanmadan ve dönüşümsüz olarak karşılaştırılır; alanın dönüştürülmüş bir sürümü yoksa, yalnızca özgün sürüm için bir grafik gösterilir. Türetilmiş tarih ya da saat alanları ve oluşturulan karşılaştırma belirtimleri için, grafikler yalnızca yeni önyükleyici için gösterilir.

*Not:* Yalnızca işleme geçmişinin gösterileceği çok sayıda kategori olması nedeniyle bir alan dışlanırsa.

### Dağıtım Grafiği

Sürekli alan dağılımı bir histogram olarak, normal bir eğri yerleştirilmiş olarak gösterilir ve ortalama değer için dikey bir başvuru çizgisi; kategorik alanlar bir çubuk grafik olarak görüntülenir.

Histogramlar standart sapma ve çarpıklık göstermek için etiketlenir; ancak, değerlerin sayısı 2 ya da daha azsa ya da özgün alanın varyansı 10-20 'den küçükse çarpıklık görüntülenmez.

Histogramlar için ortalama ve ya da çubuk grafiklerdeki kategorilere ilişkin toplam kayıt sayısının sayısını ve yüzdesini görüntülemek için fareyi grafiğin üzerinde gezdirin.

### Değer Grafiği Eksik

Pasta grafikler, dönüşümler uygulanmış olan ve içermeyen eksik değerlerin yüzdesini karşılaştırır; grafik etiketleri yüzdeyi gösterir.

Eksik değer ele alma koşullarıyla birlikte, dönüştürme sonrası pasta grafiği, yerine koyma değerini bir etiket olarak da içerir; yani, eksik değerlerin yerine kullanılan değer.

Eksik değer sayısını ve toplam kayıt sayısının yüzdesini görüntülemek için fareyi grafiğin üzerinde gezdirin.

### Tahmine dayalı Güç Grafiği

Önerilen alanlar için çubuk grafikler, tahmine dayalı gücü dönüştürmeden önce ve sonra görüntüler. Hedef dönüştürülürse, hesaplanan tahmine dayalı güç dönüştürülen hedefe göre olur.

*Not:* Hedef tanımlı bir hedef yoksa ya da hedef ana görünüm panosunda tıklatıldıysa, tahmine dayalı güç grafikleri gösterilmez.

Tahmine dayalı güç değerini görüntülemek için fareyi grafiğin üzerine getirin.

### İşleme Geçmiş Tablosu

Çizelge, bir alanın dönüştürülen sürümünün nasıl türetilceğini gösterir. Koşullarla alınan işlemler, gerçekleştirildikleri sırayla listelenir; ancak, belirli adımlar için, belirli bir alan için birden çok işlem gerçekleştirilmiş olabilir.

*Not:* Bu çizelge dönüştürülmemiş alanlar için gösterilmez.

Çizelgedeki bilgiler iki ya da üç sütuna bölünebilir:

- **İşlem.** İşlemin adı. Örneğin, Continuous Predictors. Ek bilgi için “İşlem Ayrıntıları ” sayfa 16 başlıklı konuya bakın.
- **Ayrıntılar.** Yürütülen işleme listesi. Örneğin, standart birimlere dönüştürün.
- **İşlev.** Yalnızca oluşturulan önericiler için gösterilen bu, giriş alanlarının doğrusal birleşimini görüntüler; örneğin,  $.06*age + 1.21*height$ .

## İşlem Ayrıntıları

Eylem Özeti ana görünümünde altı çizili **Eylem** seçeneğini belirlediğinizde görüntülenen Eylem Ayrıntıları bağlantılı görünümü, yürütülen her bir işleme adımına ilişkin hem eyleme özgü hem de genel bilgileri görüntüler; eyleme özgü ayrıntılar önce görüntülenir.

Her işlem için, tanım, bağlantılı görünümün üst kısmında başlık olarak kullanılır. Eyleme özgü ayrıntılar, başlığın altında görüntülenir ve türetilen karşılaştırma belirtileri, alanlar yeniden dönüşümü, hedef dönüşümler, birleştirilen ya da yeniden sıralanan kategoriler ve öngörülebilirlerin oluşturduğu ya da dışlanan tahmin ediciler sayısı ile ilgili ayrıntıları içerebilir.

Her işlem işlendiği gibi, örneğin, tahmin ediciler dışlanmış ya da birleştirildiği gibi, işlemde kullanılan öngörülebilirlerin sayısı değişebilir.

*Not:* Bir işlem kapatıldıysa ya da herhangi bir hedef belirtilmediyse, İşlem Özeti ana görünümünde eylem tıklatıldığında yapılacak işlem ayrıntılarının yerine bir hata iletisi görüntülenir.

Dokuz olası eylem vardır; ancak, her analiz için her şey tam olarak etkindir.

Metin Alanları Tablosu

Tabloda aşağıdakilerin sayısı görüntülenir:

- Öngörüler analiz kapsamı dışında bırakılıyor.

Tarih ve Saat Tahminleri Tablosu

Tabloda aşağıdakilerin sayısı görüntülenir:

- Tarih ve saat karşılaştırma belirtilerinden türetilen süreler.
- Tarih ve saat öğeleri.
- Toplam olarak, türetilmiş tarih ve saat koşullanıcıları.

Herhangi bir tarih süresine hesaplandıysa, başvuru tarihi ya da saati bir dipnot olarak görüntülenir.

Tahmin Edilebilir Tarama Tablosu

Tabloda, işlenmekten dışlanan aşağıdaki öngörülebilirlerin sayısı görüntülenir:

- Sabitler.
- Çok sayıda eksik değeri olan öngörülebilirler.
- Çok sayıda vakayı tek bir kategoride olan tahmin ediciler.
- Çok fazla kategori içeren nominal alanlar (kümeler).
- Tahmin ediciler toplamda, toplamda gösterilmiş.

Ölçüm Düzeyi Çizelgesini Denetle

Tabloda, yeniden yayınların sayısı, aşağıda belirtilen sayılarla bölünen sayılar görüntülenir:

- Sıralı alanlar (sıralı kümeler) sürekli alan olarak yeniden yayınlar.
- Sürekli alanlar sıra alan olarak yeniden çevrim eder.
- Toplam yeniden dönüştürme sayısı.

Hiçbir giriş alanı (hedef ya da koşullanıcılar) sürekli ya da sıra numarası yoksa, bu bir dipnot olarak gösterilir.

## Aykırı Tablo

Tabloda, herhangi bir aykırılığın nasıl ele alınmalarıyla ilgili sayıları görüntülenir.

- Ayarlar sekmesinde Hazırla Girintiler ve Hedef panosunda ayarlarınıza bağlı olarak, aykırı alanların bulunduğu ve kırılan sürekli alan sayısı ya da atlanacak sürekli alan sayısı ya da eksik olarak ayarlanan sürekli alan sayısı.
- Sürekli alanların sürekli olarak işlenmesinden sonra, sürekli olarak dışlanan sürekli alan sayısı.

Bir dipnot, aykırı kesme değerini gösterir; hiçbir giriş alanı (hedef ya da karşılaştırma belirtimi) sürekli değilse, başka bir dipnot gösterilirken.

## Eksik Değerler Tablosu

Tabloda, eksik değerleri olan alanların sayıları görüntülenir ve bu sayılar aşağıdaki şekilde bölünerek aşağıdaki sayılar açılır:

- Hedef. Hedef belirtilmediyse bu satır gösterilmez.
- Tahmin ediciler. Bu, nominal (set), ordinal (sıralı küme) ve sürekli olarak bölünen sayılara daha da bölünmektedir.
- Değiştirilen eksik değerlerin toplam sayısı.

## Hedef Çizelge

Hedef, hedefin dönüştürülmüş olup olmadığını görüntüler; aşağıdaki gibi gösterilir:

- Box-Cox 'un normalliğe dönüşümü. Bu, belirtilen ölçütleri (ortalama ve standart sapma) ve Lambda 'yı gösteren sütunlara daha da bölünmektedir.
- Kararlılığı artırmak için hedef kategoriler yeniden sıralanacak.

## Kategorik Öngörüler Tablosu

Tabloda, kategorik karşılaştırma belirtilerinin sayısı görüntülenir:

- Stabilitiyi geliştirmek için, kategorileri en düşükten en yükseğe doğru yeniden sıralandı.
- Hedefle ilişkilendirmeyi en üst düzeye çıkarmak için kategorileri birleştirildi.
- Kategorilerin seyrek kategorilerle işlenmesi için birleştirilen kategoriler.
- Hedefle birlikte düşük ilişki nedeniyle dışlandı.
- Birleştikten sonra sabit oldukları için dışlandı.

Herhangi bir kategorik karşılaştırma belirtimi yoksa, bir dipnot gösterilir.

## Sürekli Önericiler Tablosu

İki masa var. İlk olarak, aşağıdaki dönüşümlerden biri görüntülenir:

- Tahmin edici değerler standart birimlere dönüştürüldü. Buna ek olarak, dönüştürülen öngörülebilirlerin sayısını, belirtilen ortalama ve standart sapmayı gösterir.
- Karşılaştırma belirtimi değerleri ortak bir aralıkla eşlendi. Buna ek olarak, bir min-max dönüşümünün yanı sıra, belirtilen alt sınır ve üst sınır değerlerinin de dönüştürülmüş öngörülebilirlerinin sayısını gösterir.
- Tahmin edici değerler, binned ve öngörülebilirlerin sayısı.

İkinci tabloda, tahmin edici sayısı olarak gösterilen tahmin alanı alanı oluşturma ayrıntıları görüntülenir:

- Oluşturuldu.
- Hedefle birlikte düşük bir ilişki nedeniyle dışlandı.
- Vurulmasından sonra sabit oldukları için hariç tutuldular.
- İnşaattan sonra sabit oldukları için dışlanmış.

Hiçbir sürekli karşılaştırma belirtimi girilmezse, bir dipnot gösterilir.

## Geri Dönüştürme Puanları

Bir hedef, dönüştürülen hedef puanı kullanılarak oluşturulan sonraki modellerle dönüştürülebiliyorsa, dönüştürülen birimleri de kullanın. Sonuçları yorumlamak ve kullanmak için, tahmin edilen değeri özgün ölçüğe dönüştürmeniz gerekir.

1. Puanları geri dönüştürmek için menülerden şunları seçin:

**Dönüştürme > Modelleme için Veri Hazırla > Geri Dönüştürme Scores ...**

2. Geri dönüşümü yapmak için bir alan seçin. Bu alan, dönüştürülen hedefin model tahmini değerlerini içermelidir.
3. Yeni alan için bir sonek belirtin. Bu yeni alan, dönüştürülemez hedefin özgün ölçüğünde model tahmini değerleri içerir.
4. ADP dönüştürmelerini içeren XML dosyasının yerini belirtin. Bu, Etkileşimli ya da Otomatik Veri Hazırlık iletişim kutularından saklanmış bir dosya olmalıdır. Ek bilgi için [“Dönüşüm Uygulama ve Kaydetme” sayfa 11](#) başlıklı konuya bakın.

## Olağan Dışı Durumları Tanımlama

Anomali Algılama yordamı, küme gruplarının normlarından sapmaları temel alarak olağan dışı durumları arar. Bu yordam, herhangi bir çıkarımsal veri analizinden önce, keşif veri analizi adımıyla veri denetleme amacıyla olağan dışı durumları hızla saptamak için tasarlanmıştır. Bu algoritma, genel anomali algılama için tasarlanmıştır; yani, anormal bir davanın tanımı, sağlık sektöründeki alışılmadık ödeme şekillerinin saptanması ya da finans endüstrisinde para aklama tespiti gibi herhangi bir özel uygulamaya özgü değildir. Bu durum, bir anomali tanımının iyi tanımlanabileceği bir şekilde tanımlanabilir.

**Örnek.** İnme tedavisi sonuçları için tahmine dayalı modeller oluşturmak üzere işe alınan bir veri analisti, veri kalitesinden kaygı duymaktadır, çünkü bu tür modeller alışılmadık gözlemlere karşı duyarlı olabilir. Bu dışlanmış gözlemlerin bazıları gerçekten benzersiz vakaları temsil eder ve bu nedenle öngörü için uygun olmamakla birlikte, diğer gözlemler, değerlerin teknik olarak "doğru" olduğu ve dolayısıyla veri doğrulama yordamlarıyla yakalanamadığı veri girişi hatalarından kaynaklanır. Olağandışı Durumlar Tanımlanması yordamı, analist tarafından nasıl işleneceğine karar verebilmesi için bu aykırı değerleri bulur ve bildirir.

**İstatistikler.** Bu yordam, eş grupları, sürekli ve kategorik değişkenler için eş grup normları, eş grubu normlarındaki sapmaları temel alan anomali endeksleri ve bir vakaya en çok katkıda bulunan değişkenlere ilişkin değişken etki değerlerini olağandışı olarak kabul edilen değişkenlere üretir.

Veri İle İlgili

**Veri.** Bu yordam hem sürekli hem de kategorik değişkenlerle çalışır. Her satır ayrı bir gözlemi temsil eder ve her bir sütun, eş grupların dayalı olduğu ayrı bir değişkeni temsil eder. Bir vaka tanıtıcısı değişkeni, çıktıyı işaretlemek için veri dosyasında kullanılabilir, ancak çözümlemede kullanılmayacaktır. Eksik değerlere izin verilir. Ağırlık değişkeni belirtilirse, yoksayılır.

Algılama modeli, yeni bir test verileri dosyasına uygulanabilir. Test verilerinin öğeleri, eğitim verisinin öğeleriyle aynı olmalıdır. Algoritma ayarlarına bağlı olarak, modeli yaratmak için kullanılan eksik değer işleme, puanlama işleminden önce test verileri dosyasına uygulanabilir.

**Vaka sırası.** Çözümün, vakaların sırasına bağlı olabileceğini göz önünde bulundurun. Sipariş etkilerini en aza indirmek için vakaları rasgele sipariş edin. Belirli bir çözümün istikrarını doğrulamak için farklı rasgele siparişlerde sıralanan vakalarla farklı çözümler elde etmek isteyebilirsiniz. Çok büyük dosya büyüklüklerine sahip durumlarda, birden çok çalıştırma farklı rasgele siparişlerde sıralanan bir vaka örneği ile gerçekleştirilebilir.

**Varsayımlar.** Algoritma, tüm değişkenlerin sabit olmayan ve bağımsız olduğunu ve giriş değişkenlerinin herhangi biri için hiçbir vakanın eksik değerleri olmadığını varsayar. Her bir sürekli değişkenin normal (Gauss) dağılımına sahip olduğu varsayılır ve her bir kategori değişkeninde çok terimli bir dağılıma sahip olduğu varsayılır. Ampirik iç test, bu yordamın hem bağımsızlık hem de dağıtımsal varsayımların ihlalleri için oldukça sağlam olduğunu, ancak bu varsayımların ne kadar iyi karşılanabildiğini dikkate aldığından emin olun.

Olağan dışı durumları tanımlamak için

1. Menülerden şunları seçin:

**Veri > Olağan Olmayan Vakaları Tanımla ...**

2. En az bir çözümlene değişkeni seçin.

3. İsteğe bağlı olarak, çıktının etiketlenmesinde kullanılacak bir vaka tanıtıcısı değişkeni seçin.

Bilinmeyen Ölçüm Düzeyine Sahip Alanlar

Veri kümesindeki bir ya da daha fazla değişkene (alanlar) ilişkin ölçüm düzeyi bilinmiyorsa Ölçüm Düzeyi uyarısı görüntülenir. Ölçüm düzeyi, bu yordama ilişkin sonuçların hesaplamasını etkilediğinden, tüm değişkenlerin tanımlanmış bir ölçüm düzeyine sahip olması gerekir.

**Verileri Tara.** Etkin veri kümesindeki verileri okur ve şu anda bilinmeyen bir ölçüm düzeyiyle herhangi bir alana varsayılan ölçü düzeyi atar. Veri kümesi büyükse, bu işlem biraz zaman alabilir.

**El ile ata.** Bilinmeyen ölçüm düzeyine sahip tüm alanları listeleyen bir iletişim kutusu açar. Bu alanlara ölçüm düzeyi atamak için bu iletişim kutusunu kullanabilirsiniz. Ayrıca, Veri Düzenleyici 'nin Değişken Görünümü 'nde ölçüm düzeyi de atayabilirsiniz.

Ölçüm düzeyi bu yordam için önemli olduğundan, tüm alanlar tanımlı bir ölçüm düzeyine sahip oluncaya kadar bu yordamı çalıştırmak için iletişim kutusuna erişemezsiniz.

## Olağan Dışı Vakalar Çıkışını Tanımla

**Olağandışı olarak kabul edilme nedenlerinin ve nedenlerin listesi.** Bu seçenek üç çizelge üretir:

- Anomali durum dizini listesi, olağan dışı olarak tanımlanan durumları görüntüler ve bunlara karşılık gelen anomali dizin değerlerini görüntüler.
- Anomali vaka eşdüzey kimliği listesi, ilgili eş gruplarıyla ilgili olağan dışı durumları ve bilgileri görüntüler.
- Anomali neden listesi, vaka numarasını, neden değişkenini, değişken etki değerini, değişkenin değerini ve her neden için değişkenin normunu görüntüler.

Tüm çizelgeler, alçalan düzende anomali dizinine göre sıralanır. Ayrıca, Değişkenler sekmesinde vaka tanıtıcısı değişkeni belirtilirse, vakaların tanıtıcıları görüntülenir.

**Özetler.** Bu gruptaki denetimler, dağıtım özetleri üretir.

- **Eş grup normları.** Bu seçenek, sürekli değişken normları tablosunu (analizde sürekli değişken varsa) ve kategorik değişken normları tablosunu (analizde herhangi bir kategorik değişken kullanılırsa) görüntüler. Sürekli değişken normları tablosu, her bir eş grubu için her bir sürekli değişkenin ortalama ve standart sapmasını görüntüler. Kategorik değişken normları tablosu, her bir eş grubu için her kategorik değişkenin kipini (en sık kullanılan kategori), sıklığını ve sıklık yüzdesini görüntüler. Sürekli değişken ve bir kategorik değişkenin kipi, analizde norm değerleri olarak kullanılır.
- **Anomali endeksleri.** Anomali dizin özeti, en olağan dışı durum olarak tanımlanan durumların anomali indeksi için açıklayıcı istatistikleri görüntüler.
- **Çözümlene değişkenine göre durum ortaya çıktı.** Her neden için tabloda, her bir değişkenin oluşumunun sıklık ve sıklık yüzdesi bir neden olarak görüntülenir. Çizelge ayrıca, her değişkenin etkisinin açıklayıcı istatistiklerini de bildirir. Seçenekler etiketinde neden sayısı üst sınırı 0 olarak ayarlandıysa, bu seçenek kullanılamaz.
- **Vakalar işlendi.** Vaka işleme özeti, etkin veri kümesindeki tüm vakalara ilişkin sayıları ve sayı yüzdelerini, analiz edilen ve çözümlenmede hariç tutulan durumları ve her bir eş grubundaki vakaları görüntüler.

## Olağan Dışı Vakaların Kaydedilmesini Tanımla

**Değişkenleri Kaydedin.** Bu gruptaki denetimler, model değişkenlerini etkin veri kümesine kaydetmenize olanak tanır. Ayrıca, adları saklanacak değişkenlerle çakışan var olan değişkenleri değiştirmeyi de seçebilirsiniz.

- **Anomali dizini.** Her vaka için anomali dizininin değerini, belirtilen ada sahip bir değişkene kaydeder.
- **Eşdüzey gruplar.** Eşdüzey grup kimliğini, vaka sayısını ve boyutu, her bir vaka için belirtilen kök ada sahip değişkenlere yüzde olarak kaydeder. Örneğin, kök adı *Eşdüzey* belirtilirse, *Peerid*, *PeerSize* ve *PeerPctSize* değişkenleri oluşturulur. *Peerid* , vakanın eşdüzey grup tanıtıcısıdır; *PeerSize* , grubun boyudur ve *PeerPctSize* , grubun büyüklüğü yüzdesindedir.
- **Nedenler.** Belirtilen kök adla, muhakeme değişkenlerinin kümelerini kaydeder. Bir muhakeme değişkenleri kümesi, değişkenin adı, değişken etki ölçümü, kendi değeri ve norm değeri gibi değişkenlerden oluşur. Kümelerin sayısı, Seçenekler sekmesinde istenen nedenlerin sayısına bağlıdır. Örneğin, kök adı *Neden* belirtilirse, *ReasonVar\_k*, *NedenMeasure\_k*, *NedenValue\_k* ve *Neden Norm\_k* değişkenleri oluşturulur; burada *k* , *k'* in neden olduğu bir nedendir. Nedenlerin sayısı 0 olarak ayarlandığında bu seçenek kullanılamaz.

**Model Dosyasını Dışa Aktar.** Modeli XML biçiminde kaydetmenizi sağlar.

## Olağan Dışı Vakaların Eksik Değerlerini Belirleme

Eksik Değerler sekmesi, kullanıcı eksik ve sistem eksik değerlerinin işlenmesini denetlemek için kullanılır.

- **Eksik değerleri çözümlenmekten dışla.** Eksik değerleri olan durumlar çözümlenmeyle dışlanır.
- **Analizde eksik değerleri içer.** Sürekli değişkenlerin eksik değerleri, karşılık gelen büyük anlamlarıyla değiştirilir ve kategorik değişkenlerin eksik kategorileri gruplanır ve geçerli bir kategori olarak değerlendirilir. Daha sonra, işlenen değişkenler çözümlenmede kullanılır. İsteğe bağlı olarak, her vakada eksik değişkenlerin oranını gösteren ek bir değişken yaratılmasını ve bu değişkenin çözümlenmede kullanılmasını isteyebilirsiniz.

## Olağan Dışı Durumlar Seçeneklerini Belirleyin

**Sıra dışı durumları tanımlayan ölçütler.** Bu seçimler, anomali listesinde kaç vaka eklendiğini belirler.

- **En yüksek anomali dizin değerlerine sahip vakaların yüzdesi.** 100 'den küçük ya da 100 'e eşit bir pozitif sayı belirtin.
- **En yüksek anomali dizin değerlerine sahip vakaların sayısı değişmez.** Çözümlenmede kullanılan etkin veri kümesindeki vakaların toplam sayısından az ya da ona eşit bir artı tamsayı belirtin.
- **Yalnızca, anomali dizin değeri bir değer alt sınırını karşılayan ya da bu değeri aşan vakaları belirleyin.** Negatif olmayan bir sayı belirtin. Bir vaka, anomali dizin değeri, belirtilen kesme noktasına eşit ya da daha büyük olduğunda olağandışı bir durum olarak kabul edilir. Bu seçenek, **Vakalar yüzdesi** ve **Değişmez vaka sayısı** seçenekleri ile birlikte kullanılır. Örneğin, değişmez sayıda 50 vaka ve 2 kesme değeri belirlerseniz, anomali listesi en çok 50 durumda, her biri 2 'den büyük ya da 2 olan bir anormallik dizin değerine sahip olur.

**Eşdüzey Grup Sayısı.** Yordam, belirtilen alt sınır ve üst sınır değerleri arasındaki en iyi eş grubu sayısını arar. Değerler pozitif tamsayılar olmalı ve alt sınır değeri üst sınırı aşmamalıdır. Belirtilen değerler eşit olduğunda, yordama göre sabit sayıda eşdüzey grup vardır.

*Not:* Verinizdeki varyasyon miktarına bağlı olarak, verilerin destekleyebileceği eşdüzey grupların sayısının alt sınır olarak belirtilen sayıdan az olduğu durumlar olabilir. Böyle bir durumda, yordam daha az sayıda eşdüzey grup üretebilir.

**Reasons Sayısı Üst Sınırı.** Bir neden, değişken etki ölçümünden, bu nedenle değişken adını, değişkenin değerini ve karşılık gelen eşdüzey grubun değerini içerir. Negatif olmayan bir tamsayı belirtin; bu değer, çözümlenmede kullanılan işlenen değişkenlerin sayısını eşitse ya da aşıyorsa, tüm değişkenler gösterilir.

## DETECTANOMALY Komut Ek Özellikleri

Komut sözdizimi dili, aşağıdaki özellikleri de sağlar:

- Tüm analiz değişkenlerini açık bir şekilde belirtmeden, etkin veri kümesindeki birkaç değişkeni çözümlenmeden kaldırın ( EXCEPT altkomutunu kullanın).



- Sürekli ve kategorik değişkenlerin etkisini dengelemek için bir ayarlama belirtin ( CRITERIA altkomutundaki MLWEIGHT anahtar sözcüğünü kullanarak).

Tam sözdizimi bilgileri için *Command Syntax Reference* belgesine bakın.

## En Uygun Vuruş

The Optimal Binning procedure discretizes one or more scale variables (referred to henceforth as **giriş değişkenlerini sabitleme**) by distributing the values of each variable into bins. Bölme oluşumu, benning işlemini "üst düzey" yapan bir kategori kılavuzu değişkenine göre optimal (optimal) bir değişkendir. Daha sonra, daha fazla analiz için özgün veri değerleri yerine çöp kutusu kullanılabilir.

**Örnekler.** Bir değişkenin aldığı ayrı değer sayısının azaltılması, aşağıdakiler de içinde olmak üzere bir dizi kullanıma sahip olur:

- Diğer yordamların veri gereksinimleri. Uyumlaştırılmış değişkenler, kategorik değişkenler gerektiren yordamlarda kullanılmak üzere kategorik olarak değerlendirilebilir. Örneğin, Çapraz tabloların yordamı, tüm değişkenlerin kategorik olmasını gerektirir.
- Veri gizliliği. Gerçek değerler yerine binned değerlerin raporlanması, veri kaynaklarınızın gizliliğini korumaya yardımcı olabilir. En Uygun Sabitleme yordamı, çöp kutusu seçmesine yol gösterebilirler.
- Hız performansı. Bazı yordamlar, azaltılmış ayrı değerler sayısı ile çalışırken daha verimlidir. Örneğin, Multinomial Logistic Regression hızının, uyumlaştırılmış değişkenler kullanılarak iyileştirilebilir.
- Tam ya da quasi-tam veri ayırımını ortaya çıkarın.

**Optimal-Visual Benning.** Görsel Benleme iletişim kutuları, bir kılavuz değişkeni kullanılmadan çöp kutusu oluşturmak için birkaç otomatik yöntem sunar. Bu "denetimsiz" kurallar, sıklık tabloları gibi tanımlayıcı istatistikler üretmek için kullanışlıdır, ancak nihai hedefiniz tahmine dayalı bir model üretirken Optimal Barning üstün olur.

**Çıktı.** Bu yordam, binler için kesme noktalarının çizelgeleri ve her bir bining giriş değişkeni için açıklayıcı istatistikler üretir. Ayrıca, yeni değişkenleri etkin veri kümesine saklayabilir ve bining giriş değişkenlerinin bind değerlerini içeren etkin veri kümesine saklayabilir ve yeni verileri, kullanıcı sözdizimine uygun olarak kullanmak için sabitleme kurallarını komut sözdizimi olarak saklayabilirsiniz.

En Uygun Veri Sabitleme Verileri Dikkate Alınması

**Veri.** Bu yordam, girilen giriş değişkenlerinin ölçeklendirilmelerini, sayısal değişkenlerin olmasını bekler. Kılavuz değişkeni kategorik olmalı ve dizgi ya da sayısal olabilir.

En uygun sabitleme elde etmek için

1. Menülerden şunları seçin:

**Dönüştürme > Optimal Benning ...**

2. Bir ya da daha çok çekişme giriş değişkeni seçin.
3. Bir kılavuz değişkeni seçin.

Binned veri değerlerini içeren değişkenler varsayılan olarak oluşturulmaz. Bu değişkenleri kaydetmek için [Kaydet](#) sekmesini kullanın.

## En Uygun Başlangıç Çıktısı

Çıkış etiketi, sonuçların görüntülenmesini denetler.

- **Bölmeler için uç noktalar.** Her bir bining giriş değişkenine ilişkin uç nokta kümesini görüntüler.
- **Binned olan değişkenlere ilişkin açıklayıcı istatistikler.** Her bir bining giriş değişkeni için, bu seçenek geçerli değer içeren vakaların sayısını, eksik değerleri olan vaka sayısını, ayrı geçerli değerlerin sayısını ve alt sınır ve üst sınır değerlerini görüntüler. Kılavuz değişkeni için bu seçenek, ilgili her bir bining giriş değişkeni için sınıf dağılımını görüntüler.
- **Bind olan değişkenler için model entropisi.** Her bir benning giriş değişkeni için, bu seçenek, kılavuz değişkenine göre değişkenin tahmine dayalı doğruluğunun bir ölçüsünün görüntülenmesini sağlar.

## En Uygun Başlangıç Kaydı

**Değişkenleri Etkin Veri Kümesine Kaydet.** İned veri değerlerini içeren değişkenler, daha ayrıntılı çözümlenmede özgün değişkenlerin yerine kullanılabilir.

**Sabitleme Kurallarını Sözdizimi Olarak Kaydet.** Diğer veri kümelerini bölmek için kullanılacak komut sözdizimini oluşturur. Kurtarma kuralları, sabitleme algoritmasının belirlediği kesme noktalarına dayalıdır.

## Eksik Değerleri En Uygun şekilde Sabitle

Eksik Değerler sekmesi, eksik değerlerin liste bilgisiyle mi, yoksa çiftli silme kullanılarak mı işleneceğini belirtir. Kullanıcı-eksik değerler her zaman geçersiz olarak işlenir. Özgün değişken değerleri yeni bir değişkene geri çevrildiğinde, kullanıcı eksik değerler sistem eksik olarak dönüştürülür.

- **Akıllıca.** Bu seçenek, her bir kılavuz ve benlik giriş değişkeni çiftinde çalışır. Yordam, kılavuzda eksik olmayan değerleri olan tüm vakaları kullanır ve giriş değişkenini sabitler.
- **Listwise** Bu seçenek, Değişkenler sekmesinde belirtilen tüm değişkenlerde çalışır. Bir vaka için herhangi bir değişken eksikse, tüm vaka dışlanır.

## En Uygun Vuruş Seçenekleri

**Ön işleme.** "Önceden Yerleştirme" giriş değişkenlerinin pek çok farklı değere sahip olması, son çöp kutularının kalitesinde büyük bir fedakarlık olmadan işlem süresini iyileştirebilir. Kutu sayısı üst sınırı, oluşturulan kutu sayısına bir üst sınır verir. Bu nedenle 1000 değerini üst sınır olarak belirtirseniz, ancak bir binyerleştirme giriş değişkeninin 1000 'den az ayrı değeri varsa, bining giriş değişkeni için oluşturulan önceden işlenen sepetlerin sayısı, bining giriş değişkenindeki ayrı değer sayısına eşit olur.

**Seyrek Dolu Kutular.** Bazen, prosedür çok az sayıda vakaya sahip kutu üretebilir. Aşağıdaki strateji, bu sözde kesme noktalarını siler:

Belirli bir değişken için, algoritmanın  $n_{\text{son}}$  kesme noktalarını bulduğunu ve bu nedenle  $n_{\text{son}} + 1$  bölmeleri. Bölmeler için  $i = 2, \dots, n_{\text{final}}$  (ikinci en yüksek değerli bin ile ikinci en düşük değerli bölme), hesapla

$$\frac{\text{sizeof}(b_i)}{\min(\text{sizeof}(b_{i-1}), \text{sizeof}(b_{i+1}))}$$

burada  $\text{sizeof}(b)$ , kutudaki vaka sayısıdır.

Bu değer belirtilen birleştirme eşliğinden küçük olduğunda,  $B_i$ , yoğun şekilde doldurulmuştur ve alt sınıf bilgi entropisine sahip hangisi varsa,  $B_{i-1}$  ya da  $B_{i+1}$  ile birleştirilir.

Prosedür, çöplerden tek bir geçiş yapar.

**Bölme Uç Noktaları.** Bu seçenek, bir aralığın alt sınırının nasıl tanımlanıyor olduğunu belirtir. Yordam, kesme noktalarının değerlerini otomatik olarak belirlediğinden, bu, büyük ölçüde bir tercih meselesidir.

**İlk (En Düşük)/Son (En Yüksek) Bin.** Bu seçenekler, her bir bining giriş değişkeni için minimum ve maksimum kesme noktalarının nasıl tanımlananın tanımlarını belirler. Genellikle, yordam, bining giriş değişkenlerinin gerçek sayı çizgindeki herhangi bir değeri alabileceğini varsayar; ancak, aralığı sınırlayan bir teorik ya da pratik nedeniniz varsa, bu değeri en düşük/en yüksek değerlere bağlayabilirsiniz.

## EN İYİ SABITLEME KOMUTU Ek Özellikler

Komut sözdizimi dili, aşağıdaki özellikleri de sağlar:

- Eşit frekanslar yöntemiyle ( CRITERIA altkomutunu kullanarak) gözetimsiz olarak sabitleme işlemi gerçekleştirin.

Tam sözdizimi bilgileri için *Command Syntax Reference* belgesine bakın.

## Özel Notlar

Bu bilgiler, ABD'de kullanıma sunulan ürünler ve hizmetler için geliştirilmiştir. IBM bu bilgileri başka dillerde kullanıma sunabilir. Ancak, bu bilgilere erişebilmek için, ürünün ya da ürün sürümünün o dildeki bir kopyasına sahip olmanız gerekebilir.

IBM, bu belgede sözü edilen ürün, hizmet ya da özellikleri diğer ülkelerde kullanıma sunmayabilir. Bulduğunuz yerde kullanıma sunulan ürün ve hizmetleri yerel IBM müşteri temsilcisinden ya da çözüm ortağınızdan öğrenebilirsiniz. Bir IBM ürün, program ya da hizmetine gönderme yapılması, açık ya da örtük olarak, yalnızca o IBM ürünü, programı ya da hizmetinin kullanılabileceğini göstermez. Aynı işlevi gören ve IBM'in fikri mülkiyet haklarına zarar vermeyen herhangi bir ürün, program ya da hizmet de kullanılabilir. Ancak, IBM dışı ürün, program ya da hizmetlerle gerçekleştirilen işlemlerin değerlendirilmesi ve doğrulanması kullanıcının sorumluluğundadır.

IBM'in, bu belgedeki konularla ilgili patentleri ya da patent başvuruları olabilir. Bu belgenin size verilmiş olması, patentlerin izinsiz kullanım hakkının da verildiği anlamına gelmez. Lisansla ilgili sorularınızı aşağıdaki adrese yazabilirsiniz:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

Çift byte (DBCS) bilgilerle ilgili lisans soruları için, ülkenizdeki IBM'in Fikri Haklar (Intellectual Property) bölümüyle bağlantı kurun ya da sorularınızı aşağıda adrese yazın:

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japonya*

IBM BU YAYINI, OLDUĞU GİBİ, HİÇBİR KONUDA AÇIK YA DA ÖRTÜK GARANTİ VERMEKSİZİN SAĞLAMAKTADIR; TİCARİ KULLANIMA UYGUNLUK AÇISINDAN HER TÜRLÜ GARANTİ VE BELİRLİ BİR AMACA UYGUNLUK İDDİASI AÇIKÇA REDDEDİLİR. Bazı hukuk bölgeleri, belirli işlemlerde açık ya da zımni garantilerin reddedilmesine izin vermez, bu nedenle bu bildirim sizin için geçerli olmayabilir.

Bu yayın teknik yanlışlar ya da yazım hataları içerebilir. Buradaki bilgiler üzerinde düzenli olarak değişiklik yapılmaktadır; söz konusu değişiklikler sonraki basımlara yansıtılacaktır. IBM, önceden bildirimde bulunmaksızın, bu yayında açıklanan ürünler ve/ya da programlar üzerinde iyileştirmeler ve/ya da değişiklikler yapabilir.

Bu belgede IBM dışı web sitelerine gönderme yapılması kolaylık sağlama amacına yöneliktir ve o web siteleri için herhangi bir şekilde onay verilmesi anlamına gelmez. Bu web sitelerinin içerdiği malzeme, bu IBM ürününe ilişkin malzemenin bir parçası değildir ve bu tür web sitelerinin kullanılmasının sorumluluğu size aittir.

IBM'e bilgi ilettiğinizde, IBM bu bilgileri size karşı hiçbir yükümlülük almaksızın uygun gördüğü yöntemlerle kullanabilir ya da dağıtabilir.

(i) Bağımsız olarak yaratılan programlarla, bu program da içinde olmak üzere diğer programlar arasında bilgi değiş tokuşuna ve (ii) değiş tokuş edilen bilginin karşılıklı kullanımına olanak sağlamak amacıyla bu program hakkında bilgi sahibi olmak isteyen lisans sahipleri şu adrese yazabilirler:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119*

Armonk, NY 10504-1785  
US

Bu tür bilgiler, ilgili kayıt ve koşullar altında ve bazı durumlarda bedelli olarak edinilebilir.

Bu belgede açıklanan lisanslı program ve bu programla birlikte kullanılacak tüm lisanslı malzeme, IBM tarafından IBM Müşteri Sözleşmesi, IBM Uluslararası Program Lisansı Sözleşmesi ya da eşdeğer sözleşmelerin kayıt ve koşulları altında sağlanır.

Performans verileri ve müşteri örnekleri, örnek olarak yalnızca gösterim amaçlı olarak sunulmuştur. Gerçek performans sonuçları, belirli yapılandırmalara ve işletim koşullarına bağlı olarak değişebilir.

IBM dışı ürünlerle ilgili bilgiler, bu ürünleri sağlayan firmalardan, bu firmaların yayın ve belgelerinden ve genel kullanıma açık diğer kaynaklardan alınmıştır. IBM, bu ürünleri test etmemiştir ve performansın, uyumluluğun ya da IBM dışı ürünlerle ilgili diğer iddiaların doğruluğunu onaylayamaz. IBM dışı ürünlerin yeteneklerine ilişkin sorular, bu ürünleri sağlayan firmalara yöneltilmelidir.

IBM' in gelecekteki yönelim ve kararlarına ilişkin bildirimler değişebilir ya da herhangi bir duyuruda bulunulmadan bunlardan vazgeçilir; bu yönelim ve kararlar yalnızca amaç ve hedefleri gösterir.

Bu belge, günlük iş ortamında kullanılan veri ve raporlara ilişkin örnekler içerir. Örneklerin olabildiğince açıklayıcı olması amacıyla kişi, şirket, marka ve ürün adları belirtilmiş olabilir. Bu adların tümü gerçek dışıdır ve gerçek kişilerle ya da işletmelerle olabilecek herhangi bir benzerlik tümüyle rastlantıdır.

#### YAYIN HAKKI LİSANSI:

Bu belge, çeşitli işletim platformlarında programlama tekniklerini gösteren, kaynak dilde yazılmış örnek uygulama programları içerir. Bu örnek programları, IBM'e herhangi bir ödemede bulunmadan, örnek programların yazıldığı işletim altyapısına ilişkin uygulama programlama arabirimiyle uyumlu uygulama programlarının geliştirilmesi, kullanılması, pazarlanması ya da dağıtılması amacıyla herhangi bir biçimde kopyalayabilir, değiştirebilir ve dağıtabilirsiniz. Bu örnekler her koşul altında tüm ayrıntılarıyla sınanmamıştır. Dolayısıyla, IBM bu programların güvenilirliği, bakım yapılabilirliği ya da işlevleri konusunda açık ya da örtük güvence veremez. Örnek programlar, hiçbir türde garanti verilmeksizin "OLDUĞU GİBİ" sağlanır. IBM, örnek programları kullanmanızdan kaynaklanan hiçbir zarar nedeniyle sorumlu tutulamaz.

Örnek programların ya da bunlardan türetilmiş çalışmaların her kopyası ya da her kısmı, belirtilen biçimde bir yayın hakkı duyurusu içermelidir:

© Copyright IBM Corp. 2021. Bu kodun bazı kısımları IBM Corp.'un Örnek Programlarından türetilmiştir.

© Copyright IBM Corp. 1989-2021. All rights reserved. (Her hakkı saklıdır.)

## Ticari markalar

IBM, IBM logosu ve ibm.com, International Business Machines Corp. ' un ticari markaları ya da tescilli ticari markalarıdır. dünya çapında birçok yargı bölgesinde kayıtlı. Diğer ürün ve hizmet adları IBM'in ya da diğer firmaların ticari markaları olabilir. IBM ticari markalarının güncel bir listesini web üzerinde [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) adresindeki "Copyright and trademark information" (Telif hakkı ve ticari marka bilgileri) altında bulabilirsiniz.

Adobe, Adobe logosu, PostScript ve PostScript logosu, Adobe Systems Incorporated şirketinin ABD ve/veya diğer ülkelerdeki tescilli ticari markaları veya ticari markalarıdır.

Intel, Intel logosu, Intel Inside, Intel Inside logosu, Intel Centrino, Intel Centrino logosu, Celeron, Intel Xeon, Intel SpeedStep, Itanium ve Pentium; Intel Corporation'ın veya ABD ve diğer ülkelerdeki yan kuruluşlarının ticari markaları ya da tescilli ticari markalarıdır.

Linux, Linus Torvalds şirketinin ABD ve/veya diğer ülkelerdeki tescilli ticari markasıdır.

Microsoft, Windows, Windows NT ve Windows logosu Microsoft Corporation şirketinin ABD ve/veya diğer ülkelerdeki ticari markalarıdır.

UNIX, The Open Group şirketinin ABD ve diğer ülkelerdeki tescilli ticari markasıdır.

Java ve tüm Java tabanlı ticari markalar ve logolar, Oracle'ın ve/veya bağılı kuruluşlarının ticari markaları ya da tescilli ticari markalarıdır.



# Dizin

## Özel karakterler

- önceden sabitleme
  - Optimum Benning içinde [22](#)
- özellik yapımı
  - otomatik veri hazırlığında [11](#)

## A

- anomali dizinleri
  - Belirli Olmayan Vakaları Tanımla [19](#)

## B

- bilgi işlem süreleri
  - otomatik veri hazırlama [8](#)
- boş vakalar
  - Verilerin geçerliliğini denetle [6](#)
- Box-Cox dönüşümü
  - otomatik veri hazırlığında [10](#)
- bölmeler için uç
  - Optimum Benning içinde [21](#)

## Ç

- çapraz değişken doğrulama kuralları
  - Geçerlilik Denetimi Kurallarını Tanımla [3](#)
  - Verilerin geçerliliğini denetle [5](#)
- çözümleme ağırlığı
  - otomatik veri hazırlığında [10](#)

## D

- denetimli olarak sabitleme
  - denetimsiz vuruşlara karşı [21](#)
  - Optimum Benning içinde [21](#)
- denetimsiz sabitleme
  - denetimli sabitleme ile karşılaştırılan [21](#)
- doğrulama kuralları [1](#)
- doğrulama kuralları ihlalleri
  - Verilerin geçerliliğini denetle [6](#)
- Doğrulama Kurallarını Tanımla
  - çapraz değişken kuralları [3](#)
  - tek değişkenli kurallar [2](#)
- döngüsel zaman öğeleri
  - otomatik veri hazırlama [8](#)

## E

- eksik değerler
  - Belirli Olmayan Vakaları Tanımla [20](#)
- eksik vaka tanıtıcıları
  - Verilerin geçerliliğini denetle [6](#)
- En Uygun Vuruş
  - eksik değerler [22](#)
  - kaydet [22](#)

- En Uygun Vuruş (*devamı var*)
  - OUTPUT [21](#)
  - seçenekler [22](#)
- eş gruplar
  - Belirli Olmayan Vakaları Tanımla [19](#)
- Etkileşimli Veri Hazırlığı [6](#)

## K

- kural ihlallerini doğrulama
  - Verilerin geçerliliğini denetle [6](#)

## M

- MDLP
  - Optimum Benning içinde [21](#)
- Model Görünümü
  - otomatik veri hazırlığında [12](#)

## N

- nedenleri
  - Belirli Olmayan Vakaları Tanımla [19](#)

## O

- Olağan Dışı Durumları Tanımlama
  - değişkenleri kaydet [19](#)
  - eksik değerler [20](#)
  - model dosyasını dışa aktar [19](#)
  - OUTPUT [19](#)
  - seçenekler [20](#)
- otomatik veri hazırlama
  - ad alanları [11](#)
  - alan analizi [13](#)
  - alan ayrıntıları [15](#)
  - alan işleme özeti [12](#)
  - Alanlar [7](#)
  - alanlar tablosu [14](#)
  - dışlama alanları [9](#)
  - dönüştürme alanları [10](#)
  - dönüşümler uygula [11](#)
  - Eylem Ayrıntıları [16](#)
  - Eylem özeti [14](#)
  - görünüm arasındaki bağlantılar [12](#)
  - görünümleri sıfırlama [12](#)
  - Hedefler [6](#)
  - Model Görünümü [12](#)
  - ölçüm düzeyini ayarla [9](#)
  - Özellik Seçimi [11](#)
  - özellik yapımı [11](#)
  - Puanları geri dönüştürme [18](#)
  - rescale alanları [10](#)
  - sürekli hedefi normalleştir [10](#)
  - tahmini güç [14](#)
  - tarih ve saatleri hazırlama [8](#)

otomatik veri hazırlama (*devamı var*)  
veri kalitesini artırma [9](#)  
Otomatik Veri Hazırlığı [6](#)

## Ö

Özellik Seçimi  
otomatik veri hazırlığında [11](#)

## S

Sabitleme kuralları  
Optimum Benning içinde [22](#)  
süre hesaplaması  
otomatik veri hazırlama [8](#)  
sürekli hedefi normalleştir [10](#)

## T

tek değişkenli geçerlilik denetimi kuralları  
Geçerlilik Denetimi Kurallarını Tanımla [2](#)  
Verilerin geçerliliğini denetle [5](#)

## V

veri doğrulama  
Verilerin geçerliliğini denetle [3](#)  
Verileri Doğrula  
çapraz değişken kuralları [5](#)  
değişkenleri kaydet [6](#)  
OUTPUT [5](#)  
tek değişkenli kurallar [5](#)  
temel denetimler [4](#)

## Y

yenelene vaka tanıtıcıları  
Verilerin geçerliliğini denetle [6](#)





