Histogram ve Yoğunluk Çaprazlamalar

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Bir sürekli değişkenin yoğunluğunu incelediğimiz zaman grafikte karşımıza çıkan şeyler bu şekilde

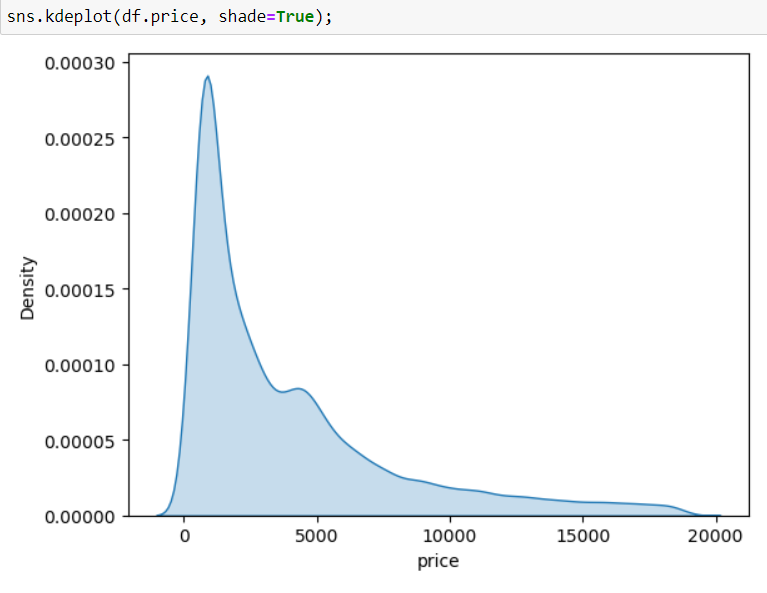
Fakat grafiğin bizden gizlediği şeyler de var…

Çaprazlamalarla bu gizliliği ortaya çıkaracağız (Neden? Nasıl?)

Bu yoğunluk neyden kaynaklanıyor? Hedeflediğimiz ana değişkene “price” detaylı bakmalıyız

Hedefledğimiz Ana Değişken = Bağımlı Değişken = Price

**NOT**: Bu yanıltıcılık aynı Medyan ve Ortalama arasındaki gibidir! Farklı yerlerden de bakmalıyız dağılıma



**NOT**: Facetgrid, Grafik üzerine eklenen boyutları bölerek bir arada göstermek için kullanılmaktadır.

Hue = Renk Tonu (Boyut Ekleme Metodu)

Shade = Gölge

Hight = Yükseklik

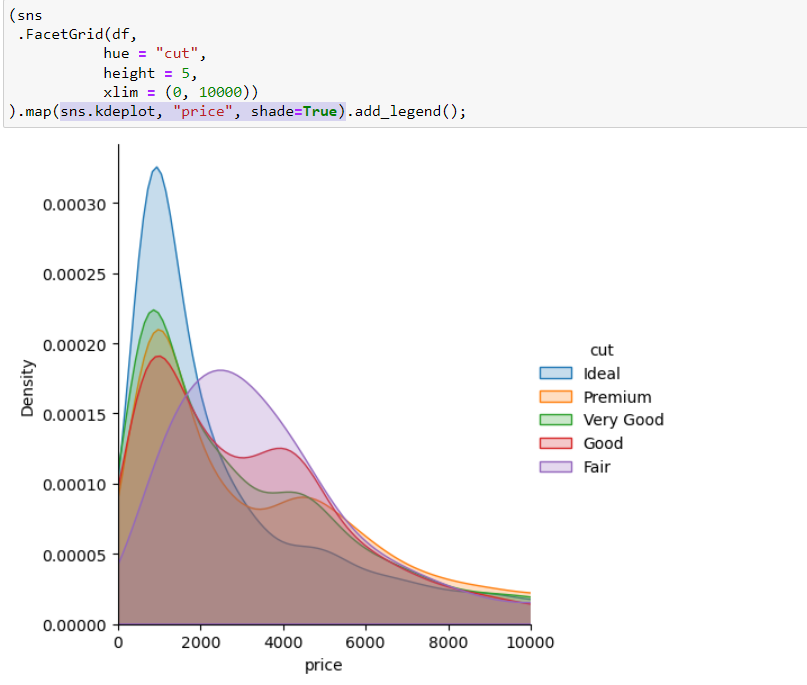
Xlim = X ekseni (0, 10000) olmasının sebebi, bizim için önemli olan kısım 0-10.000 aralığı demek oluyor!

Ylim = Y ekseni

Add\_legend = Bilgi eklemek için kullanılıyor

**Çaprazlama İşlemi**

Aslında sns.kdeplot üzerine “FacetGrid (Boyut Bölerek Gösterme)” ‘yi map ettik yani ekledik



**Grafik Yorumu**

Görüyoruz ki aslında grafik tahmin ettiğimiz gibi değilmiş (Kalite arttıkça fiyatta artmıyormuş)

Ideal kalite demek en üst kalite demekti fakat fiyatlara baktığımızda görüyoruz ki İdeal kalitedeki diamonds 0-2000 Price değerinde yoğunlaşmakta

Ve grafiğin sağ kısmına doğru gittikçe yani fiyatlar arttıkça diamondsların kalitesine bakılmaksızın price değerlerinin eşitlendiğini görüyoruz

Yani “cut(kalite)” ve “price(fiyat)” arasında bir ilişki yokmuş aslında

ÖNEMLİ

Bunu çıkarımsal olarak anlamanın diğer bir yoluda ilk grafikte dağılımın iki tepeden meydana gelmesi

Bu demek oluyor ki içerisindeki yapıyı oluşturan birden fazla faktör var! (cut ve price)

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

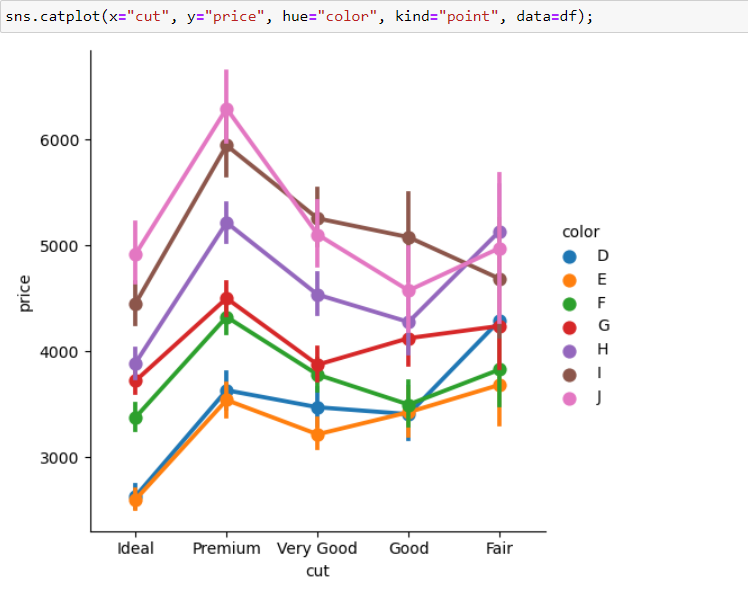
Açıklama otomatik olarak oluşturuldu

Şimdi daha da farklı bir yerden bakmak için “cut” – “price” ve “color” değişkenleri bir arada ele almak istiyoruz.

ÖNEMLİ

Price (sürekli) ve colot (kategorik) kısacası nitel ve nicel degiskenler bir arada olduğu için catplot kullanıyoruz

Kind = “point” grafik stilidir



Point ile baktığımız zamanda görüyoruz ki “color” aracılığıyla bilgiler taşınmakta.

Çünkü her rengin farklı bir değeri, boyutu mevcut. Demek ki “color” burada etkileyici durumunda

**NOT:** Değerlerin yine 3000-6000 olduğuna da dikkat et! (Aggregate edilmiştir!)

**NOT**:

Makine öğrenmesinde amacımız, hedeflediğimiz bağımlı değişkende var olan bilgiyi (değişimi) farklı değişkenlerce açıklamaya çalışmaktır

Bu bilgiyi farklı değişkenlerce açıklamaya çalışmanın en büyük hedefi veri setinin içindeki hedef değişken hedefiyle, değişkenliklerin bulunmaya çalışılmasıdır.

Yani örneğin; elimizde 3 değişkenli bir veri seti olduğunu düşünürsek (price, cut, color)

Bunu biz makine öğrenmesi ile modellemeye başladığımızda arka taraftaki algoritmaların temel amacı, price içerisindeki bilgiyi diğer değişkenlerce ortaya çıkarmaktır.

Bunun amacı ise price’ın içerisindeki bu değişkenlerin ayırt edici bilgi taşıması oluyor

Şu an bu işlemi biz yapmış olduk.

Özetle, çaprazlama yaparken değişkenlerin birbirleri üzerlerindeki etkilerini görmeye çalışıyoruz!

Bunu yaparken de değişkenlere, grafiklere başka başka açılardan bakmaya çalışarak bunu yapıyoruz