

T.C

KARADENİZ TEKNİK ÜNİVERSİTESİ

İKTİSADİ VE İDARİ BİLİMLER FAKÜLTESİ

YÖNETİM BİLİŞİM SİSTEMLERİ

ÖDEV KONUSU;

DİAMONDS VERİ SETİ İNCELEME

HAZIRLAYAN;

SEDANUR GEBİÇ 368702

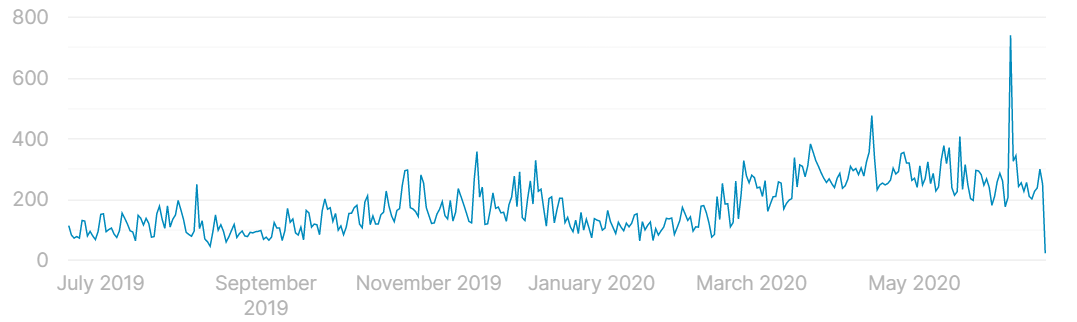
DANIŞMAN;

EKREM BAHÇEKAPILI

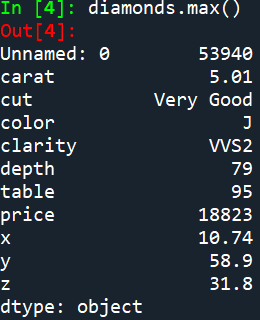
TRABZON/2020

**DİAMONDS VERİ SETİ**

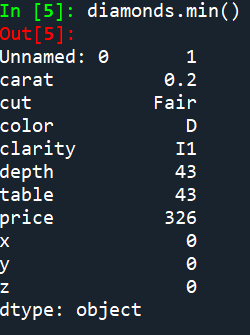
Bu klasik veri seti, yaklaşık 54.000 elmasın fiyatlarını ve diğer özelliklerini içerir. Veri seti 10 sütun ve 53940 satırdan oluşmaktadır. İlk “carat” sütunu elmasların ağırlığını ifade etmektedir. Carat sütunu 0,2- 5.01 arasında değerler almaktadır. İkinci sütun “cut” elmasın kesim kalitesini ifade etmektedir. Cut sütunu Adil, İyi, Çok İyi, Premium, İdeal olmak üzere beş farklı değer almaktadır. Üçüncü sütun “color” elmasın rengini ifade etmektedir. Dördüncü sütun “clarity” elmasın ne kadar berrak olduğunu ifade etmektedir. Beşinci sütun “depth” elmasın toplam derinliğini ifade etmektedir. Altıncı sütun “table” elmasın en geniş noktasına göre genişliğini ifade etmektedir. Table sütunu 43 – 95 arasında değerler almaktadır. Yedinci sütun “price” elmasın fiyatını ifade etmektedir. Price sütunu 18823-326 arasında değerler almaktadır. Sekizinci sütun “x” sütunu elmasın mm cinsinden uzunluğunu ifade etmektedir. X sütunu 0 – 10,74 arasında değerler yer almaktadır. Dokuzuncu sütun “y” sütunu elmasın mm cinsinden genişliğini ifade etmektedir. Y sütunu 0-58,9 arasında değerler yer almaktadır. Son sütun “z” sütunu elmasın mm cinsinden derinliğini ifade etmektedir. Z sütunu 0- 31,8 arasında değerler almaktadır.

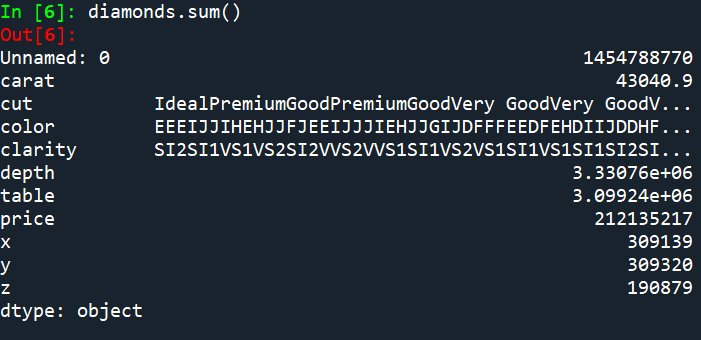


Diamonds veri seti önceki aylara göre son aylarda daha çok görüntülenmeye sahiptir. Veri setine en çok Amerika Birleşik Devletleri ve Hindistan katkıda bulunmuştur.

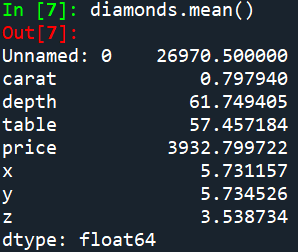
**SAYISAL TEMEL İSTATİSTİKLER**

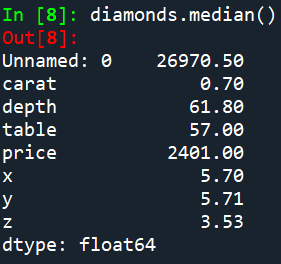
Diamonds veri setinde en yüksek değerleri diamonds.max() kodu ile bulduk. Carat sütununun en yüksek değeri 5.01, Cut sütunun en yüksek değeri Very Good, Color sütunun en yüksek değeri J, Calarity sütununun en yüksek değeri VVS2, Depth sütununun en yüksek değeri 79, Table sütununun en yüksek değeri 95, Price sütununun en yüksek değeri 18823, X sütununun en yüksek değeri 10.74, Y sütununun en yüksek değeri 58.9, z sütununun en yüksek değeri 31.8’dir.

Diamonds veri setinde en düşük değerleri diamonds.min() kodu ile bulduk. Carat sütununun en düşük değeri 0.2, Cut sütunun en düşük değeri Fair, Color sütunun en düşük değeri D, Calarity sütununun en düşük değeri I1, Depth sütununun en düşük değeri 43, Table sütununun en düşük değeri 43, Price sütununun en düşük değeri 326, X sütununun en düşük değeri 0, Y sütununun en düşük değeri 0, z sütununun en düşük değeri 0’dır.



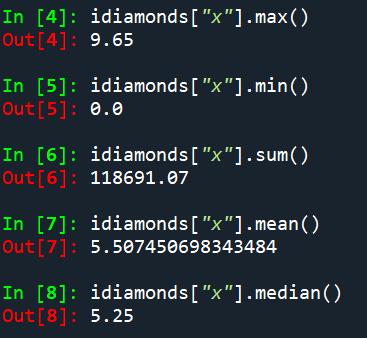
Diamonds veri setinde ortalama değerleri diamonds.sum() kodu ile bulduk. Carat sütununun ortalama değeri 43040.9, Depth sütununun ortalama değeri 3330762.9000000004, Table sütununun ortalama değeri 3099240.5, Price sütununun ortalama değeri 212135217, X sütununun ortalama değeri 309139, Y sütununun ortalama değeri 309320, z sütununun ortalama değeri 190879’dur.

Diamonds veri setinde mod değerleri diamonds.mean() kodu ile bulduk. Carat sütununun mod değeri 0.797940, Depth sütununun mod değeri 61.749405, Table sütununun mod değeri 57.457184, Price sütununun mod değeri 3932.799722, X sütununun mod değeri 5.7311157, Y sütununun mod değeri 5.734526, z sütununun en düşük değeri 3.538734 ’dür.

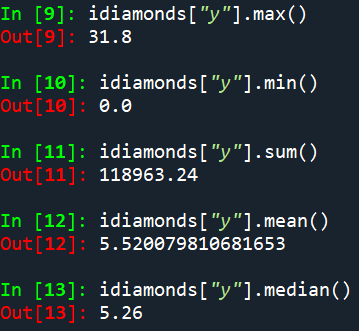
Diamonds veri setinde medyan değerleri diamonds.median() kodu ile bulduk. Carat sütununun medyan değeri 0.70, Depth sütununun mod değeri 61.80, Table sütununun mod değeri 57.00, Price sütununun mod değeri 2401.00, X sütununun mod değeri 5.70, Y sütununun mod değeri 5.71, z sütununun en düşük değeri 3.53 ’dür.

**CUT SÜTUNU İDEAL KATEGORİSİ TEMEL İSTATİSTİKLER**

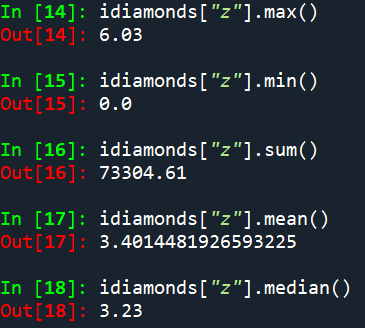
Aşağıdaki tabloda ideal kategorisinde bulunan elmasların, x yani elmasın uzunluğunun en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

Aşağıdaki tabloda ideal kategorisinde bulunan elmasların, y yani elmasın genişliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

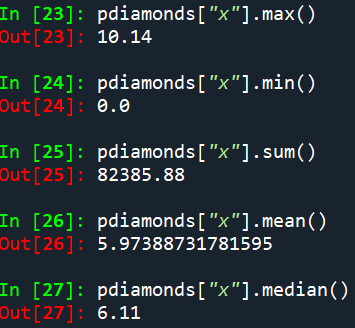


Aşağıdaki tabloda ideal kategorisinde bulunan elmasların, z yani elmasın derinliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

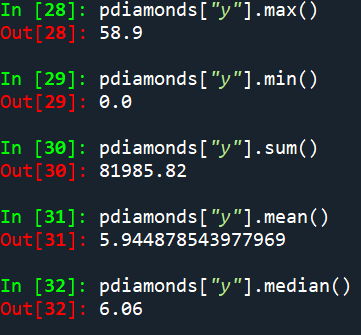
****

**CUT SÜTUNU PREMİUM KATEGORİSİ TEMEL İSTATİSTİKLER**

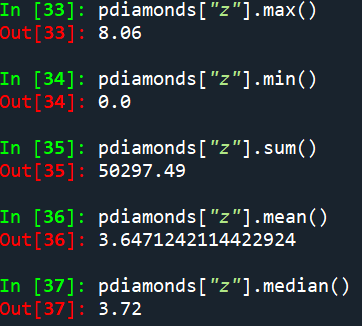
Aşağıdaki tabloda premium kategorisinde bulunan elmasların, x yani elmasın uzunluğunun en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

Aşağıdaki tabloda premium kategorisinde bulunan elmasların, y yani elmasın genişliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

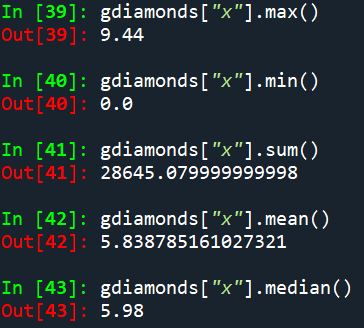
****

Aşağıdaki tabloda premium kategorisinde bulunan elmasların, z yani elmasın derinliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

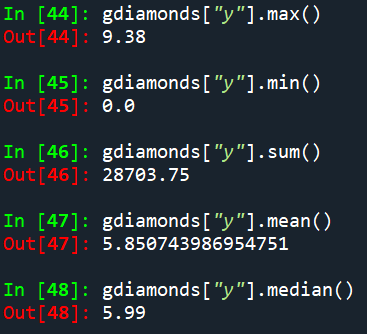
****

**CUT SÜTUNU GOOD KATEGORİSİ TEMEL İSTATİSTİKLER**

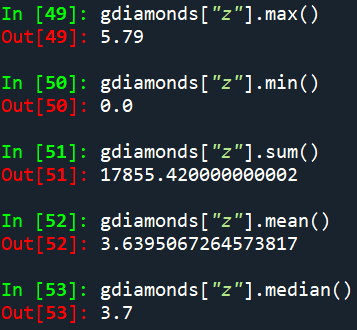
Aşağıdaki tabloda good kategorisinde bulunan elmasların, x yani elmasın uzunluğunun en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

Aşağıdaki tabloda good kategorisinde bulunan elmasların, y yani elmasın genişliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

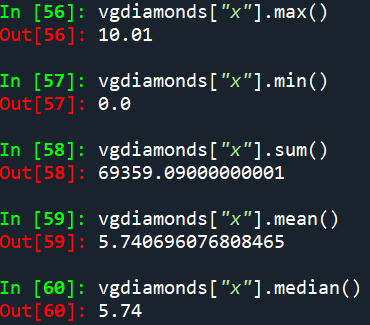
****

Aşağıdaki tabloda good kategorisinde bulunan elmasların, z yani elmasın derinliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

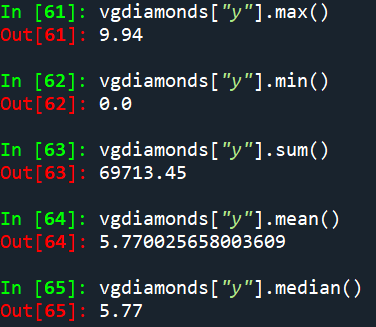
****

**CUT SÜTUNU VERY GOOD KATEGORİSİ TEMEL İSTATİSTİKLER**

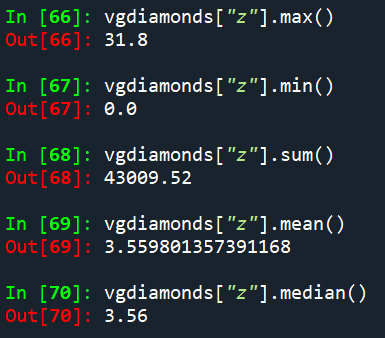
Aşağıdaki tabloda very good kategorisinde bulunan elmasların, x yani elmasın uzunluğunun en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

Aşağıdaki tabloda very good kategorisinde bulunan elmasların, y yani elmasın genişliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

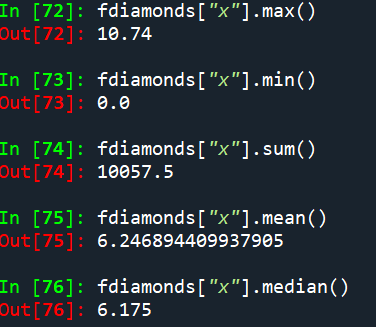
****

Aşağıdaki tabloda very good kategorisinde bulunan elmasların, z yani elmasın derinliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

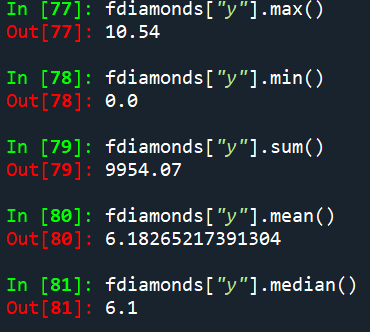
****

**CUT SÜTUNU FAİR KATEGORİSİ TEMEL İSTATİSTİKLER**

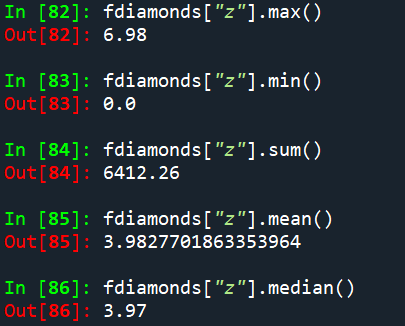
Aşağıdaki tabloda fair kategorisinde bulunan elmasların, x yani elmasın uzunluğunun en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

Aşağıdaki tabloda fair kategorisinde bulunan elmasların, y yani elmasın genişliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

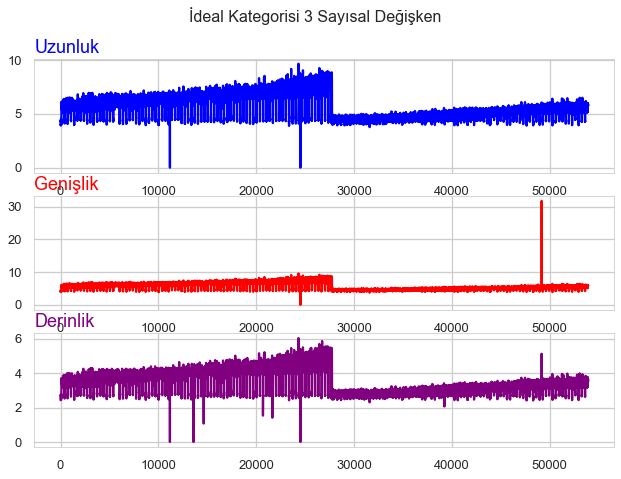
****

Aşağıdaki tabloda fair kategorisinde bulunan elmasların, z yani elmasın derinliğinin en yüksek değeri, en düşük değeri, ortalaması, mod ve medyan değeri bulunmaktadır.

****

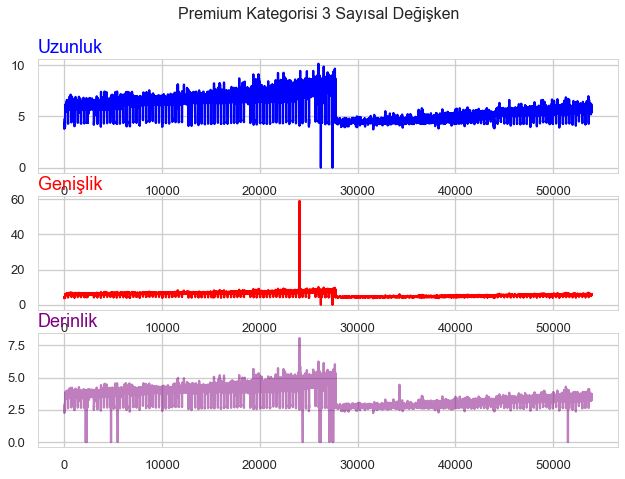
**PLOT GRAFİKLER**

**1-İdeal Kategorisi**

****

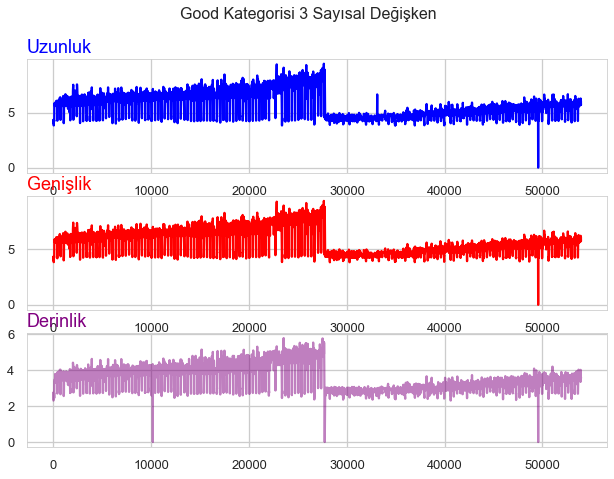
İdeal kategorisinde uzunluk değerini gösteren x sütunun da en düşük 0, en yüksek 10 değerlerini almaktadır. Ağırlıklı olarak 5-6 arasında değerler almaktadır. İdeal kategorisinde genişlik değerini gösteren y sütununda en yüksek 31.8, en düşük 0 değerini almaktadır. Ağırlıklı olarak 8-9 arasında değerler almaktadır. İdeal kategorisinde derinlik değerini gösteren z sütununda en yüksek 6, en düşük 0 değerlerini almaktadır. Ağırlıklı olarak 2-3 arasında değerler almaktadır. Elmasların ideal kategorisinde elmasın uzunluğuna ve derinliğine göre farklılık göstermektedir. Elmasın ideal kategorisinde olması elmasın genişliği üzerinde büyük bir etki yaratmaz.

**2- Premium Kategorisi**



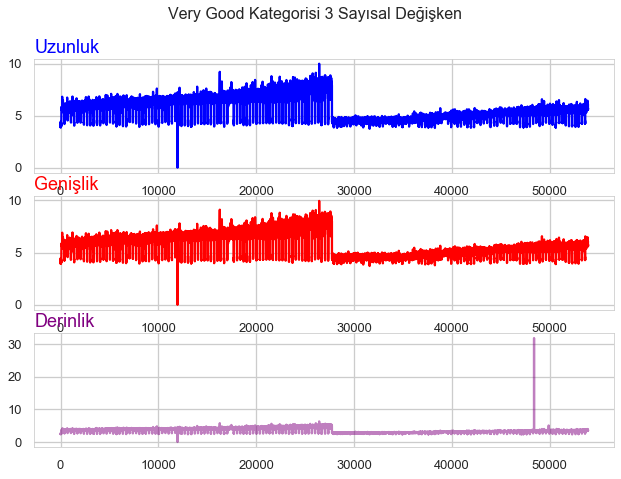
Premium kategorisinde uzunluk değerini gösteren x sütunun da en düşük 0, en yüksek 10 değerlerini almaktadır. Ağırlıklı olarak 5-6 arasında değerler almaktadır. Premium kategorisinde genişlik değerini gösteren y sütununda en yüksek 60, en düşük 0 değerini almaktadır. Ağırlıklı olarak 7-8 arasında değerler almaktadır. Premium kategorisinde derinlik değerini gösteren z sütununda en yüksek 8, en düşük 0 değerlerini almaktadır. Ağırlıklı olarak 2-4 arasında değerler almaktadır. Elmasın Premium kategorisinde bulunması elmasın uzunluğu ve derinliğini etkilemektedir. Fakat genişliği olarak premium kategorisinde büyük bir etkiye sahip değildir.

**3- Good Kategorisi**



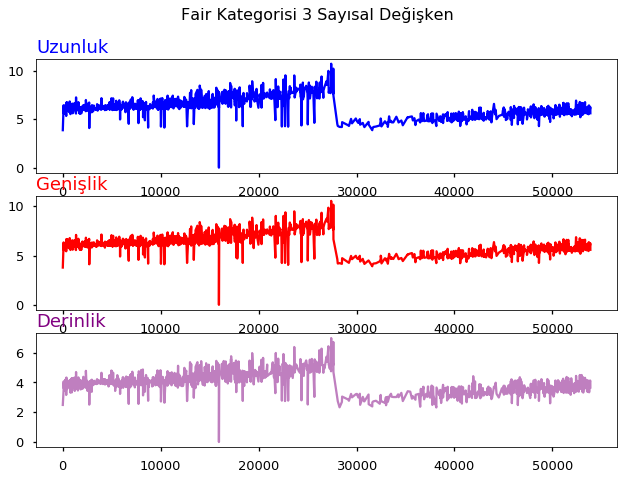
Good kategorisinde uzunluk değerini gösteren x sütunun da en düşük 0, en yüksek 9.44 değerlerini almaktadır. Ağırlıklı olarak 5-6 arasında değerler almaktadır. Good kategorisinde genişlik değerini gösteren y sütununda en yüksek 9.38, en düşük 0 değerini almaktadır. Ağırlıklı olarak 5-6 arasında değerler almaktadır. Good kategorisinde derinlik değerini gösteren z sütununda en yüksek 5.79, en düşük 0 değerlerini almaktadır. Ağırlıklı olarak 3-4 arasında değerler almaktadır. Elmasın good kategorisinde bulunması elmasın uzunluğu, genişliği ve derinliği üzerinde etkilidir.

**4- Very Good Kategorisi**



Very Good kategorisinde uzunluk değerini gösteren x sütunun da en düşük 0, en yüksek 10.01 değerlerini almaktadır. Ağırlıklı olarak 5-6 arasında değerler almaktadır. Very Good kategorisinde genişlik değerini gösteren y sütununda en yüksek 9.94, en düşük 0 değerini almaktadır. Ağırlıklı olarak 4-5 arasında değerler almaktadır. Very Good kategorisinde derinlik değerini gösteren z sütununda en yüksek 31.8, en düşük 0 değerlerini almaktadır. Ağırlıklı olarak 2-3 arasında değerler almaktadır. Elmasın Very Good kategorisinde bulunması elmasın uzunluğunu, genişliğini etkilemektedir. Fakat elmasın Very Good kategorisinde oluşu elmasın derinliğini etkilememektedir.

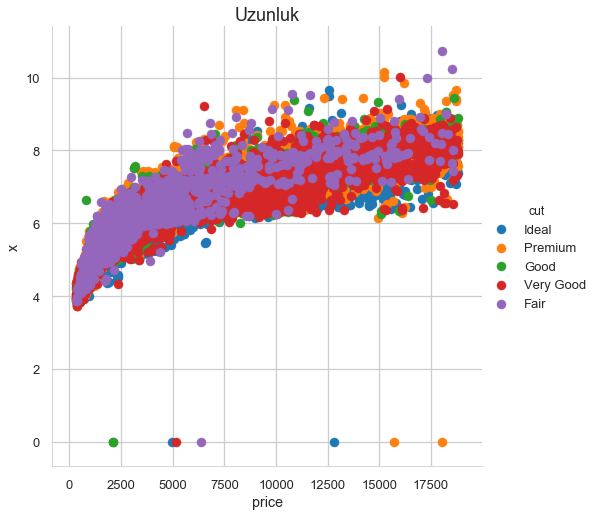
**5- Fair Kategorisi**



Fair kategorisinde uzunluk değerini gösteren x sütunun da en düşük 0, en yüksek 10.74 değerlerini almaktadır. Ağırlıklı olarak 4-5 arasında değerler almaktadır. Fair kategorisinde genişlik değerini gösteren y sütununda en yüksek 10.54, en düşük 0 değerini almaktadır. Ağırlıklı olarak 4-5 arasında değerler almaktadır. Fair kategorisinde derinlik değerini gösteren z sütununda en yüksek 6.98, en düşük 0 değerlerini almaktadır. Ağırlıklı olarak 2-4 arasında değerler almaktadır. Elmasın Fair kategorisinde bulunması elmasın uzunluğunu, genişliğini ve derinliğini etkilemektedir.

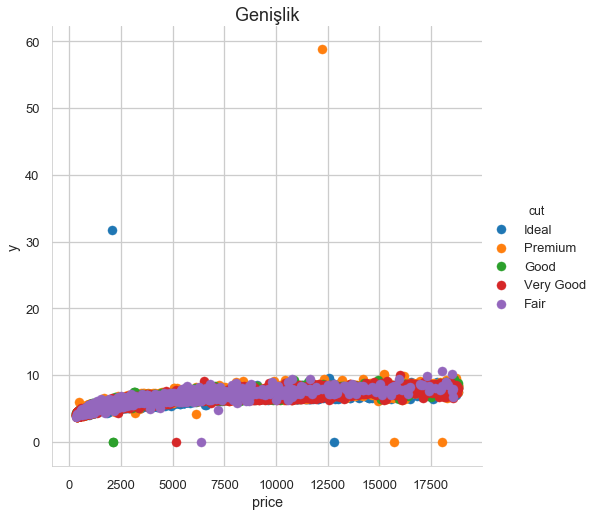
**SCATTER PLOT GRAFİKLER**

1. **Uzunluk (x)- Price**



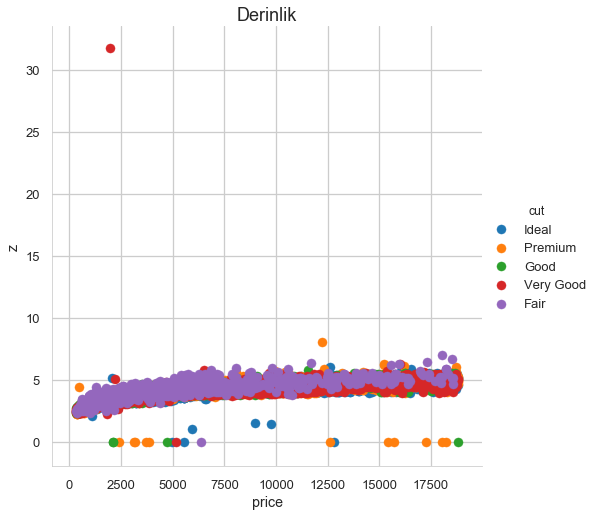
Diamonds veri setinde x sütunu yani elmasların uzunluklarına göre price sütunu olarak belirtilen elmas fiyatlarını veri setinde belirlediğimiz kategoriler olan cut sütununda bulunan elmas kesim kalitesi olan Ideal, Premium, Good, Very Good ve Fair kategorilerine göre grafiği oluşmaktadır. Tüm kategoriler en düşük olarak 0 değerinde dağılım göstermişlerdir. Genellikle uzunluk arttıkça fiyatta da artış görülmektedir. Uzunluk artışına göre fiyat artışı tüm kategorilerde aynı şekilde artış göstermektedir. Fiyat artışı genellikle 3 ile 11 arasında değişiklik göstermektedir. Sonuç olarak elmasların uzunlukları arttıkça fiyatları da artar.

1. **Genişlik (y)- Price**



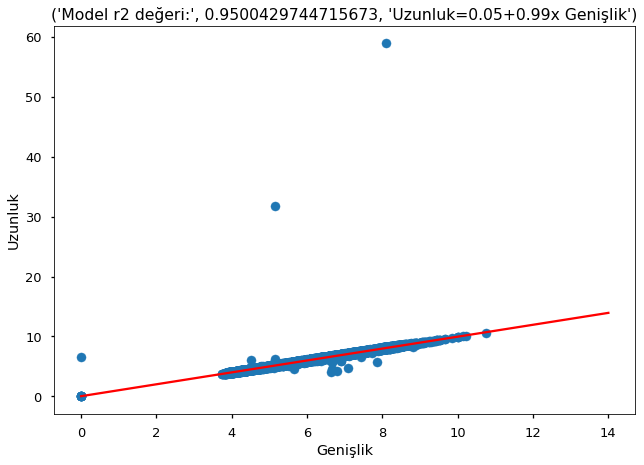
Diamonds veri setinde y sütunu yani elmasların genişliğine göre price sütunu olarak belirtilen elmas fiyatlarını veri setinde belirlediğimiz kategoriler olan cut sütununda bulunan elmas kesim kalitesi olan Ideal, Premium, Good, Very Good ve Fair kategorilerine göre grafiği oluşmaktadır. Tüm kategoriler en düşük olarak 0 değerinde dağılım göstermişlerdir. Genellikle genişlik arttıkça fiyatta da gözle görülür bir artış vardır fakat artış sadece 0-11 arasındadır. Ağırlıklı olarak bir değerde toplandığı için elmasların genişliğinin fiyata pek fazla etki etmediğini söyleyebiliriz. Sadece bir nokta olarak Ideal kategorisinin 30-40 arasında, Premium kategorisinin de 60’ yakın bir değerde olduğu gözlemlenmektedir.

1. **Derinlik (z)- Price**



Diamonds veri setinde z sütunu yani elmasların derinliklerine göre price sütunu olarak belirtilen elmas fiyatlarını veri setinde belirlediğimiz kategoriler olan cut sütununda bulunan elmas kesim kalitesi olan Ideal, Premium, Good, Very Good ve Fair kategorilerine göre grafiği oluşmaktadır. Tüm kategoriler en düşük olarak 0 değerinde dağılım göstermişlerdir. Derinlik arttıkça fiyatta da artış vardır. Ama bu da elmasların genişliğinin bulunduğu grafikteki gibi gözle görülür bile olsa pek fazla bir artış olmamaktadır. Yine ağırlıklı olarak bir değerde toplandığı için elmasların derinliğinin fiyata pek fazla etki etmediğini söyleyebiliriz. Sadece bir nokta olarak Very Good kategorisinin 30-40 arasında, bir değerde olduğu gözlemlenmektedir. Geriye kalan diğer değerler ise 0 ile 10 arasında farklılık göstermektedir.

**REGRESYON**



Genişlik değeri uzunluk değerlerinin tahmin modeli nasıl olabileceği bu modelde gösterilmiştir. R2 değeri 1’e yaklaştıkça oluşturulan tahmin modeli ile değerlerin ilişkisi arttığı bilinmektedir. Modelin R2 değeri 0.95’ dir. Yani çok yüksek bir ilişki bulunmaktadır. Bununla birlikte genişlik artarsa uzunlukta artar, genişlik azalırsa uzunlukta azalır diyebiliriz. Uzunluk değerlerinde genişlik 6-8 olduğu zaman dağılım gözlenmektedir. Uzunluk değeri en fazla genişlik 8 iken 60’ta dağılım göstermektedir. Uzunluk genel olarak 0 ile 10 arasında dağılım göstermektedir. Genişlik genel olarak 3 ile 12 arasında dağılım göstermektedir.