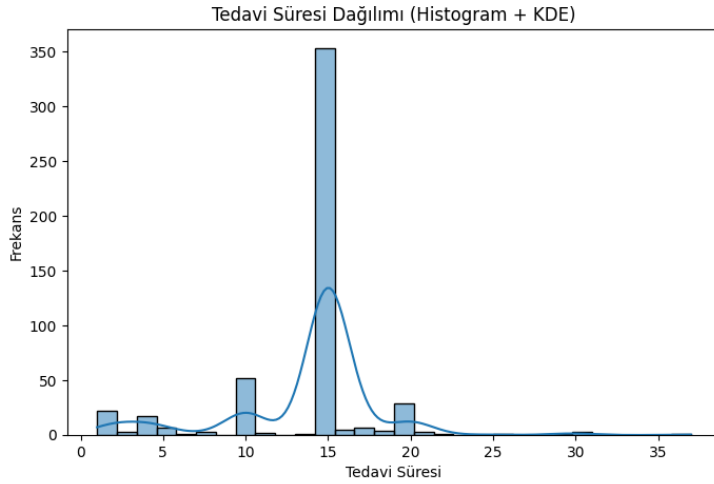


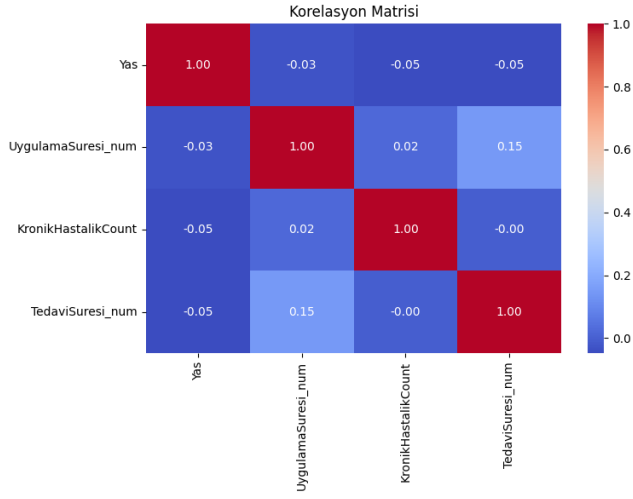
Exploratory Data Analysis

Veri setinde bir hastanın bir tedavisi için birkaç farklı seans kaydı girilmiş. İstatistik testlerinde ve grafik oluşturma aşamalarında bir tedavinin birkaç seans bilgisinin, bu tedavinin ağırlığını arttıracığından dolayı bu kayıtlar tek bir kayda indirgenmiştir. Bu işlem yapılırken sadece “HastaNo”, “TedaviAdı” ortak olan kayıtlar tek bir kayda indirgenmiştir. Bu 3 özelliği aynı olup farklı tanıya, uygulama yerine, uygulama süresine sahip kayıtlar bulunmakta. Tanı kısmındaki farklılıklar, kayıtların farklı tedavilere ait olduğunu kanıtlayacak yeterlilikte olmadığı için bu aşamada bu tarz kayıtların aynı tedaviye ait olduğu varsayılmıştır. Tek bir kayda indirgenirken de bu farklı özellikler için uygun işlemler (mean, median, en sık geçen vb.) uygulanmıştır.



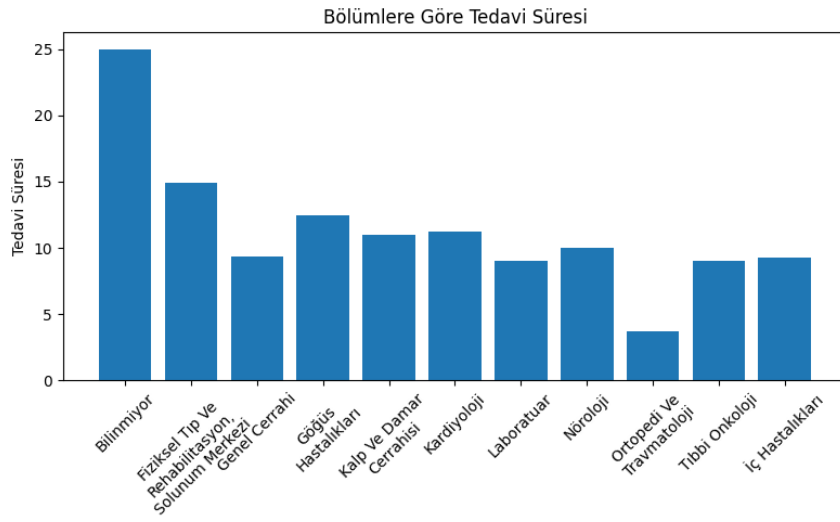
Şekil 1.1

Tedavi süresi özelliği Şekil 1.1’deki dağılıma sahip. 516 farklı tedavinin %68’i 15 seans uzunluğa sahip.

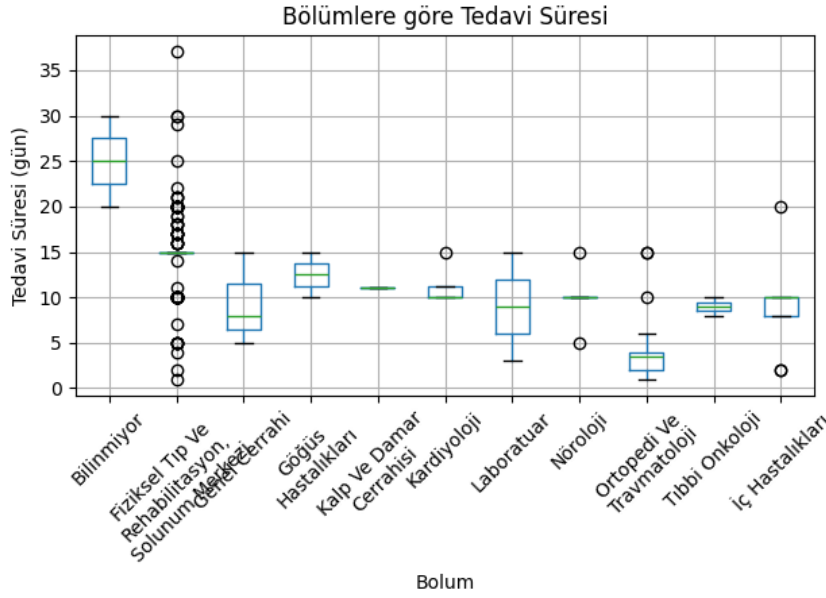


Şekil 1.11

Sadece tedavi süresi ve uygulama süresi arasında bir doğrusal ilişki var gibi görünüyor. Ancak bu ilişki güvenilir olmak için çok küçük.



Şekil 1.2



Şekil 1.6

Tedavi süresinin, tedavinin bölümlerine göre değişkenlik gösterdiği görülüyor.

Kruskal-Wallis: Bölüm

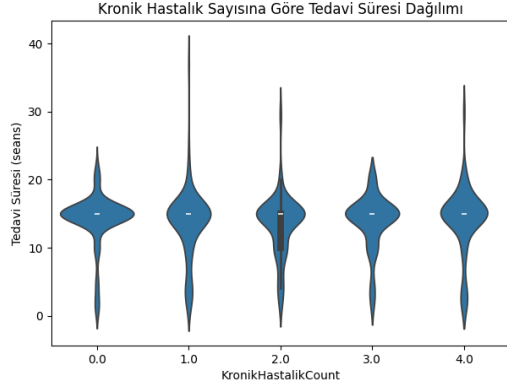
K-W Bölüm: stat=193.071, p=0.0000

Bölüm Epsilon squared (ϵ^2): 0.355477

Yapılan Kruskal-Wallis testinin çıktısı da bölümlerin tedavi süresindeki değişkenliği açıklamada etkili ve anlamlı olduğunu gösteriyor. Bu değişkenlik 1.2 ve 1.6'daki grafiklerden de görülüyor.

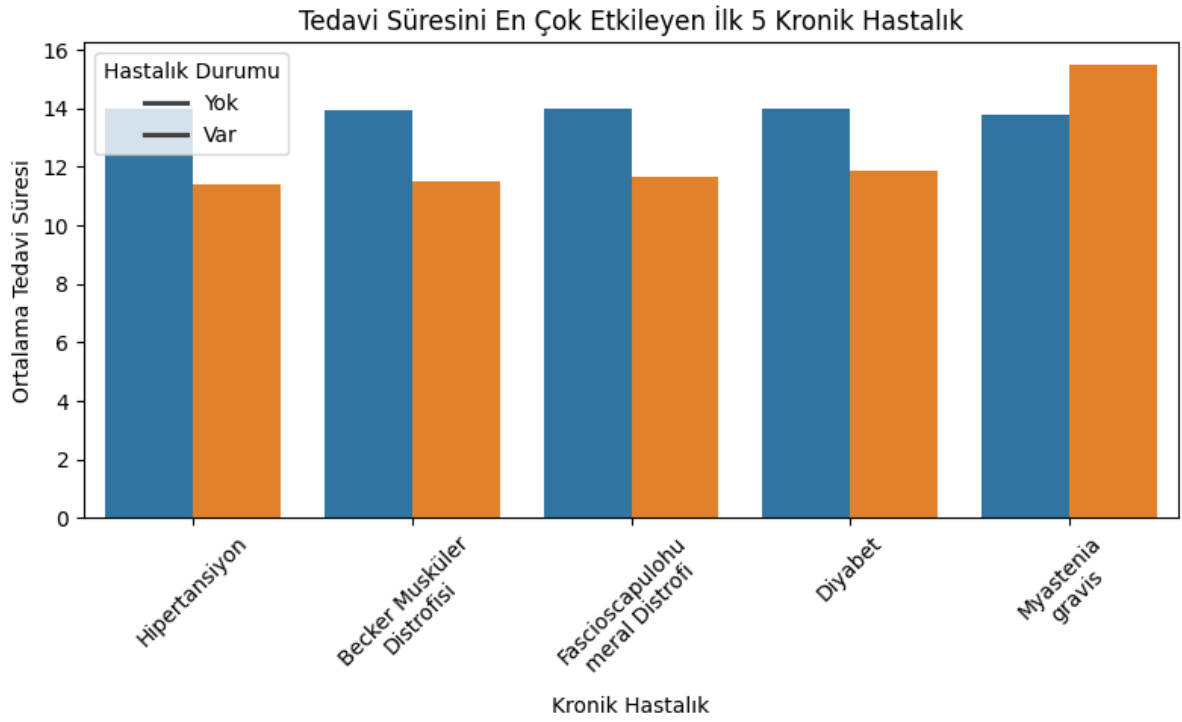
Kronik Hastalık Sayısı	Tedavi Süresi
0	14.172932
1	13.574713
2	13.411765
3	13.847059
4	14.100917

Tablo 1.1



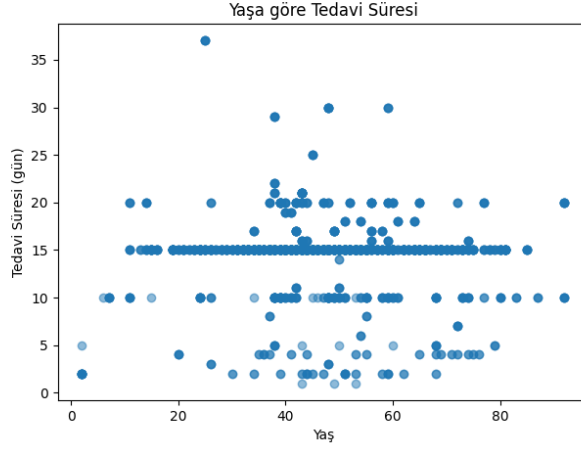
Şekil 1.8

Hastanın sahip olduğu kronik hastalık sayısının, tedavi süresine doğrudan bir etkisi olmadığı görülüyor.



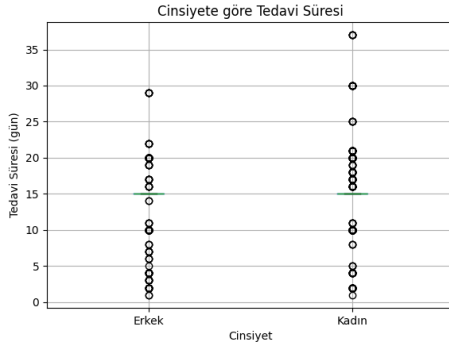
Şekil 1.3

Şekil 1.3' de varlığı ve yokluğunda tedavi süresini en çok değiştiren 5 hastalık görülmektedir. Veri setinde daha kontrollü deneye uygun sayıda ve nitelikte veri bulunmadığı için bu aşamada sadece kronik hastalığın varlığı ve yokluğu üzerinden ortalama tedavi süresi hesaplandı. Daha kesin bir çıkarım için daha çok veri üzerinden daha derin bir analiz gerekli.



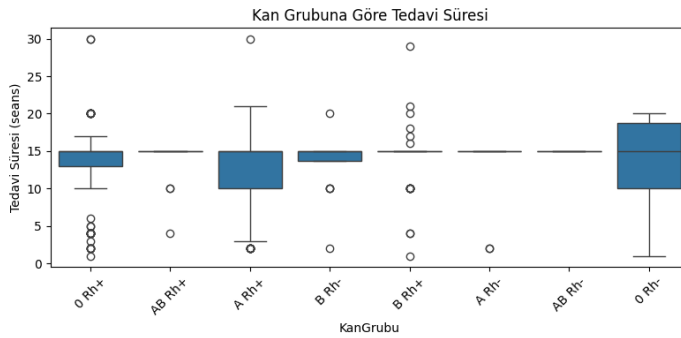
Şekil 1.4

Yaşa göre tedavi süresinde bir değişkenlik görülüyor. Sağ tarafta yoğunlaşmış bir grafik görülüyor. Ancak bunun yaşı tedavi süresi üzerindeki etkisi ile ilgisi yok. 60 yaşında üzerindeki hastalar, 40 yaşın altındaki hastalardan daha fazla olduğu için böyle bir grafik ortaya çıkıyor.



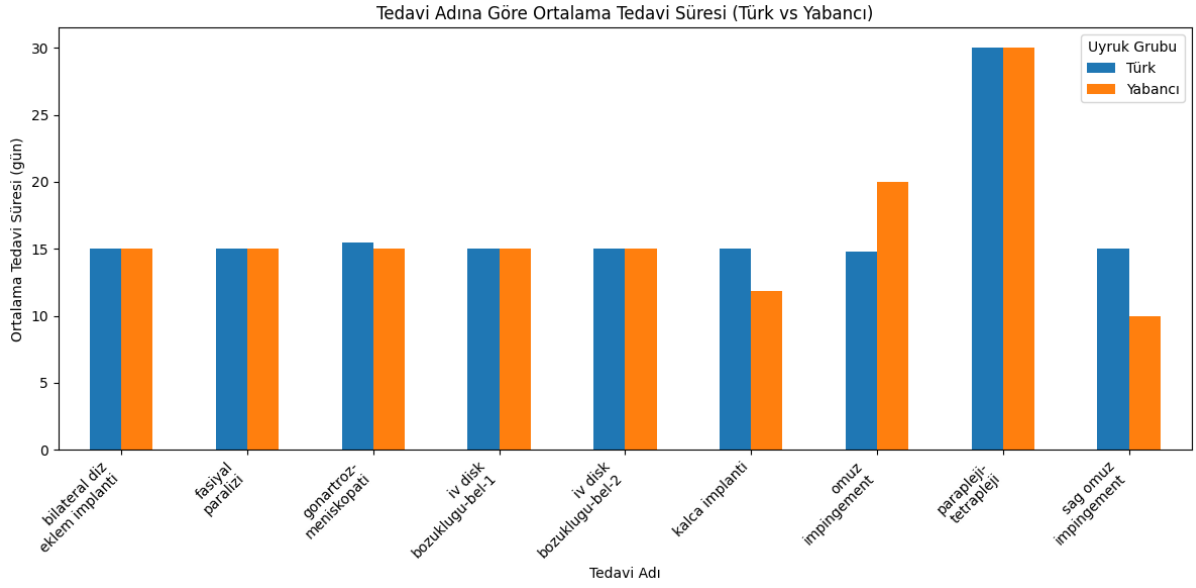
Şekil 1.5

Cinsiyetin, tedavi süresini etkilemediği Şekil 1.5’ den görülüyor.



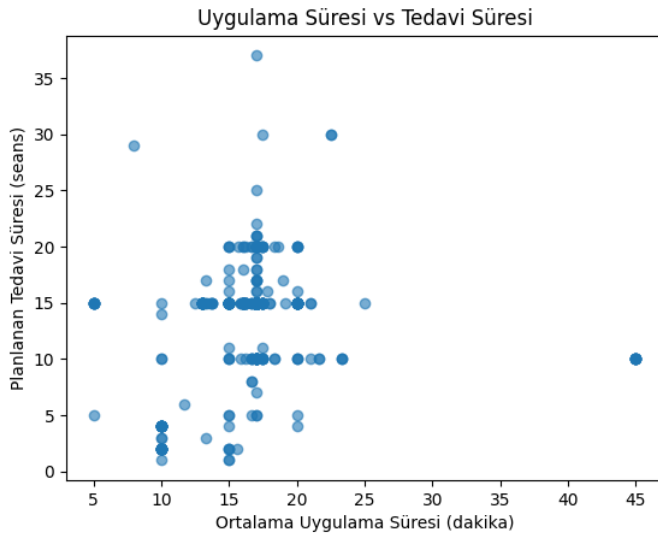
Şekil 1.7

Şekil 1.7’de kan gruplarında görünen farklılık kan gruplarının değişken yaygınlığından kaynaklı. Yani kan grubu, tedavi süresini etkilemiyor.



Şekil 1.9

Hem yabancı hem de Türklere uygulanan tedaviler incelendiğinde hastanın yabancı veya Türk olması tedavi süresini etkilemiyor.



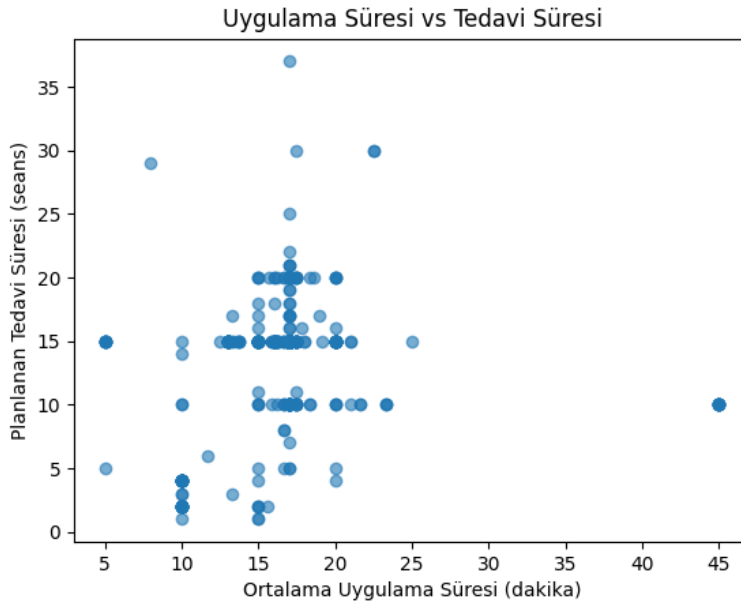
Şekil 1.10

Korelasyon matrisinde de görüldüğü gibi uygulama süresi ve tedavi süresi arasında ufak bir pozitif ilişki var gibi ancak tedavi süresindeki değişiklikleri açıklamak için yetersiz kalıyor.

UyrukGrup	Türk	Yabancı
Parapleji-Tetrapleji	30.0	30.0
Travmatik beyin yaralanması	NaN	30.0
Hemartrotik Diz-Kas Kuvvetlendirme Egzersizi	NaN	20.0
Hemartrotik Diz	NaN	20.0
Sol Omuz impingement	15.25	20.0
Serebral Palsi	NaN	20.0
Fasiyal Paralizi	15.0	15.0
Bilateral Diz Eklem İmplantı	15.0	15.0
İV DİSK BOZUKLUĞU-BEL-2	15.0	15.0
İV DİSK BOZUKLUĞU-BEL-1	15.0	15.0
Gonartroz-Meniskopati	15.5	15.0
SOL KALÇA İMPLANTI	15.0	15.0
Sağ omuz Impingement	NaN	10.0
Deneme	NaN	10.0
Sol kalça implantı	NaN	4.0
DENEME	NaN	3.0

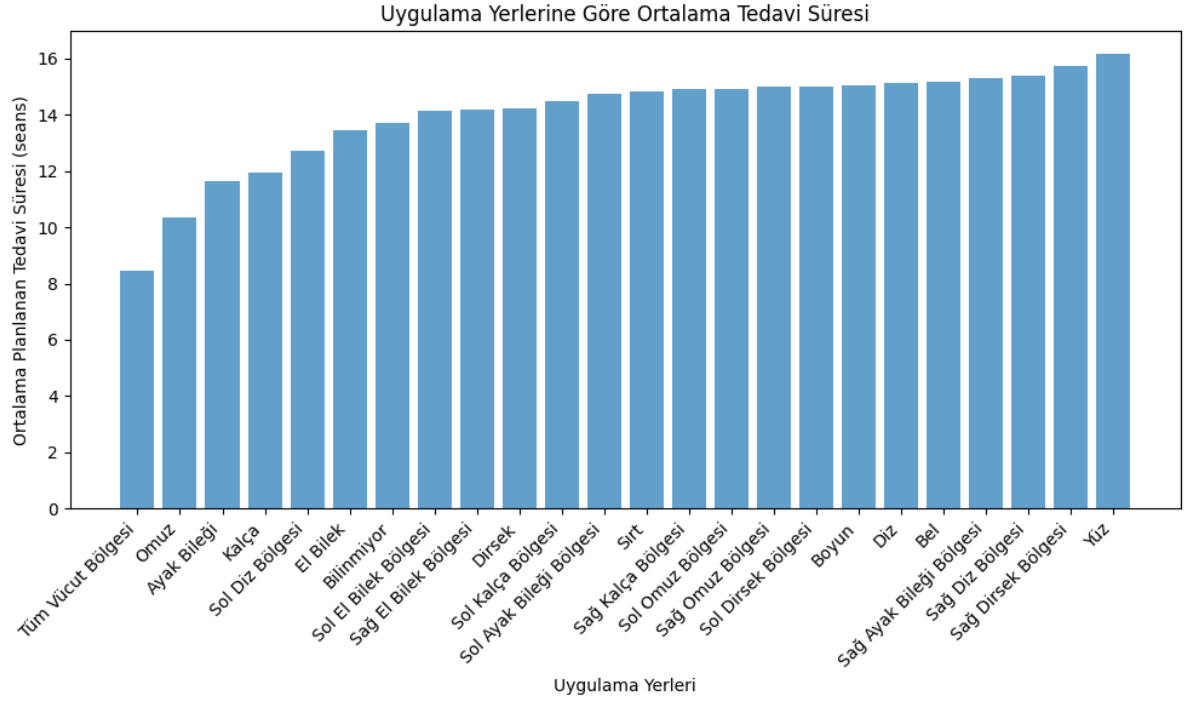
Şekil 1.12

Şekil 1.12’deki pivot tablosunda görüldüğü gibi yabancılara uygulanan tedaviler ile Türklere uygulanan tedavilerde Sadece “Sol omuz impingement” tedavisinde ortalama tedavi sürelerinde farklılık vardır, diğer tedavilerde tedavi süreleri arasında bir farklılık yoktur. “Sol omuz impingement” tedavisinde ise Türk hastalar arasında da tedavi süreleri farklılık göstermektedir. Bu yüzden bu tedavide de hastanın uyruğunun tedavi süresine bir etkisi yoktur.



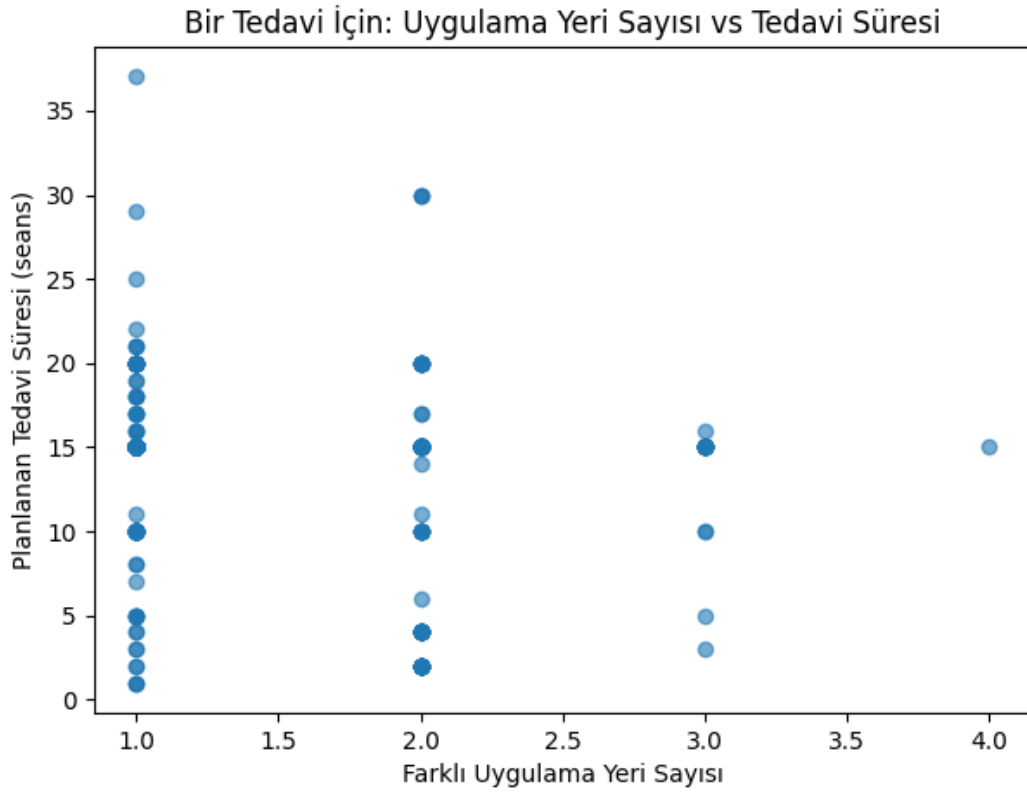
Şekil 1.13

Şekil 1.13’e göre ortalama uygulama süresi kısa olan tedavilerde tedavi süresinin de kısa olma eğilimi var. **Pearson korelasyonu: 0.147 (p=0.001)**, **Spearman korelasyonu: 0.210 (p=0.000)** katsayıları da tedavinin ortalama uygulama süresi ile tedavi süresi arasında zayıf bir pozitif ilişki olduğunu göstermektedir.



Şekil 1.14

Şekil 1.14' e göre çoğu tedavinin uygulama yerlerinde tedavi süresi için bir fark yok ancak tüm vücut bölgesi, omuz, ayak bileği, kalça bölgelerine uygulanan tedaviler, diğer tedavilere göre daha kısa tedavi süresine sahip olma eğiliminde.



Şekil 1.15

Şekil 1.15’ e göre ise tedavinin birden çok uygulama yerine yapılması ile tedavi süresi arasında doğrudan bir bağlantı görülmemektedir. Aynı şekilde farklı hastalara uygulanan aynı tedavilerde, uygulama yeri sayısının artması ile tedavi süresi arasında bir bağlantı bulunamamıştır.

HastaNo	TedaviAdi	TedaviSuresi_num	TedaviSuresi_num_mean	KronikHastalikCount	KronikHastalik_mean	UygulamaSuresi_num	UygulamaSuresi_mean	Yas	Yas_mean
145419	dorsalji - boyun+trapez	5.00	15.46	0	2.02	17.00	16.78	38	44.65
145401	dorsalji - boyun+trapez	25.00	15.46	1	2.02	17.00	16.78	45	44.65
145285	dorsalji - boyun+trapez	29.00	15.46	2	2.02	8.00	16.78	38	44.65
145530	dorsalji 1	37.00	15.89	1	2.26	17.00	17.07	25	46.56
145239	dorsalji-bel	7.00	14.67	2	1.88	17.00	17.00	72	48.88
145146	gonartroz - meniskopati	20.00	15.47	2	1.89	17.00	16.16	77	54.95
145330	iv disk bozukluğu-bel	5.00	15.03	0	1.45	17.00	16.98	68	49.73
145470	sag omuz impingement	10.00	14.69	2	1.56	17.00	13.44	38	48.75

Tablo 1.1

Anomali analizinde Tablo 1.1’deki hastaların gördükleri tedavinin normalinden en az 3 standart sapma kadar değiştikleri görülmektedir. Bulunan ortalamaların ve sapmaların yanıltıcı olmaması için en az 10 kayda sahip tedaviler incelenmiştir.

145419 numaralı hastanın ortalamadan farklı olarak herhangi bir kronik hastalığı bulunmamaktadır. Ancak veri setinde aynı tedaviyi gören ve kronik hastalığı bulunmayan başka hastalar da bulunmakta ve bu hastaların tedavi süreleri ortalama seviyededir. Yani bu tedavi kaydı anomali olabilir.

145401 numaralı hasta ortalamaya çok yakın özelliklere sahip. Bu hasta ile aynı kronik hastalığa ve aynı alerjiye sahip, aynı tedaviye gören diğer hastaların tedavi süreleri ortalamaya yakındır. Bu tedavi kaydı anomali olabilir.

145285 numaralı hastada diğer hastalardan farklı olarak “voltaren” alerjisi bulunmaktadır. Ayrıca tedavinin uygulama süresi de ortalamaya göre çok kısadır. Belki “voltaren” alerjisi bulunan bir hastaya bu tedavinin uygulanması riskli olabilir. Bu yüzden de her seansın dikkatli bir şekilde daha kısa sürelerle yapılması gerekebilir. Bu da tedavi süresinin artmasına sebep olabilir. Bu yüzden bu tedavi kaydı anomali olmayabilir. Daha iyi bir sonuca ulaşmak için bu alerjiye sahip olan başka hastaların da incelenmesi gerekmektedir. Ancak veri setinde bu alerjiye sahip bu tedaviyi görmüş tek hasta budur.

145530 numaralı hasta, veri setindeki en uzun tedavi süresine sahip hastadır. Özellikleri bu tedavi gören hastalara göre ortalama seviyede olmasına rağmen ortalama tedavi süresinden 4 standart sapma uzaklıktadır. Bu kaydın anomali olma ihtimali yüksektir.

145239 numaralı hasta, diğer hastalara göre daha yaşlıdır. Ancak ne bu hasta ile aynı kronik hastalıklara sahip hastalar ne de bu hastanın yaşına yakın hastalarda tedavi süresinde ortalamadan sapma görülmemektedir. Ayrıca bu hasta ile aynı kronik hastalıklara sahip ve aynı yaştaki hastalar da ortalama tedavi sürelerine sahiptir. Bu kayıt anomali olabilir.

145146 numaralı hasta aynı yaşta aynı tedaviyi 2 kere görmüştür. Birinde 15 seans tedavi görmüşken diğerinde 20 seans tedavi görmüştür. Tedavi süresindeki değişiklik çok büyük olmasa da bu tedavide çok az değişkenlik olduğu için standart sapması çok düşüktür. Bu yüzden de 5 seanslık bir değişim 3 standart sapmadan fazla değişikliğe karşılık gelmektedir.

145330 numaralı hasta ortalamaya göre daha yaşlıdır ve hiç kronik hastalığı yoktur. Ancak bu hastadan daha yaşlı ve kronik hastalığı olmayan hastaların tedavi sürelerinde buna benzer bir sapmaya rastlanmamıştır. Bu kayıt anomali olabilir.

145470 numaralı hasta ortalama özelliklere sahip bir hastadır. Bu tedavide ortalama 15 seans uygulanırken bu hastaya 10 seans uygulanmıştır. 5 seanslık fark çok büyük olmasa da bu tedavide farklı tedavi süresine sahip tek hasta budur. Bu yüzden 3 standart sapma uzaklıktadır.

Preprocessing

Analizde etkisiz olarak bulduğumuz "HastaNo", "Cinsiyet", "KanGrubu", "Uyruk" siliniyor. String tipinde olup float'a çevrilen "TedaviSuresi", "UygulamaSuresi" sütunları da siliniyor. "Alerji" ve "KronikHastalık" sütunlarında boş olan değerler "Bilinmiyor" olarak değiştiriliyor. "UygulamaYerleri", "Tanilar", "Bolum" sütunlarında boş olan değerler, aynı tedavide en çok geçen değerler olarak dolduruluyor. Böylece tüm boş değerler doldurulmuş oluyor. Tedavi süresini en çok etkileyen "TedaviAdi" ve "Tanilar"

sütunlarında sadece 1 kere geçen kategorilerde eğer tedavi süresi 15 seans ise bu kategoriler “Diğer” olarak değiştirildi. Böylece toplam 500’ün üzerinde olan eşsiz kategori sayısı 200 civarına indirildi. Böylece One-Hot encoding işlemi sonucunda daha az sütun olması sağlandı. "Yas", "KronikHastalıkCount", "UygulamaSuresi_num" sütunlarına StandardScaler uygulandı. "TedaviAdi", "Tanilar", "UygulamaYerleri", "Bolum", "Alerji", "KronikHastalik" sütunlarına One-Hot encoding uygulandı. Alternatif olarak çok fazla kategori içeren sütunlara Target Encoding yöntemlerle gibi her kategoriye özel, hedef özelliğe göre sayısal bir değer atanarak sütun sayısal formata dönüştürülebilir.