**Yanılgıma Övgü: Airbnb Kiralarının ve En Yakın Sağlık Turizmi ile İlintili Kuruluşa Olan Mesafenin İlişkisi**

*İstanbul’daki bir Airbnb’nin kirası ve o Airbnb’in en yakın sağlık turizmi ile ilintili kuruluşa olan mesafesi arasında herhangi bir korelasyon var mı? Cevabın “Evet!” olmasını çok istedim, fakat maalesef cevap hayır. Bu yazı yanılgıma bir övgüdür.*

A picture containing map

Description automatically generated

*Bu yazının konusu olan analizin sonucuna dair ufak bir ipuçu.*

Bir önceki analizimden (LINK) yola çıkarak İstanbul’daki sağlık kuruluşlarını biraz daha kurcalamak istedim. Aklımda üzerine düşmek istediğim iki tür sağlık kuruluşu vardı: estetik klinikleri ve saç ekim merkezleri.

Peki neden? Coronavirüs pandemisi öncesinde İstanbul’a uğradıysanız etrafta kafaları bandajlar ile sarılı halde dolaşan insan sayısının çokluğunu büyük ihtimalle fark etmişsinizdir.

Graphical user interface, text, application, email

Description automatically generated

*İstanbul’da sağlık turizmi (özellikle de saç ekimi) büyük bir piyasa.*

Google’a “İstanbul” ve “saç ekimi” kelimelerini yazarak yapacağınız ufak bir arama size bu garip görüntünün sebebini açıklayacaktır. İstanbul’da çok sayıda estetik merkezi ve saç ekim merkezi var. Bu merkezlerin müşterileri ise genelde yurt dışından gelen ve tatilleri sırasında birkaç kozmetik operasyonu aradan çıkarmak isteyen turistler. Bu kozmetik operasyonlar arasındaki en popülerlerinden biri de saç ekimi.

Ben böyle bir turizm aktivitesinin varlığından çok etkilendim ve kendime “Acaba sağlık turizmi, şehrin turizm ekonomisinde nasıl bir yere sahip?” diye sormaya başladım. Böyle bir ilişkiyi nasıl irdeleyebilirdim?

Airbnb kiralık daireleri, aynı diğer turizm merkezlerinde olduğu gibi, İstanbul’a tatile gelen ve konaklama ihtiyaçlarını karşılamak isteyen turistler için çok cazip bir seçenek. Kafamda bu iki olguyu birleştirdim ve bir soru oluşturdum, belki de Airbnb dairelerinin kiraları ve dairelerin en yakın sağlık turizmi ile ilintili kuruluşa olan mesafeleri arasında doğrusal (linear) veya monotonik bir ilişki vardır! Ne de olsa sağlık turizmi insanları İstanbul’a çeken büyük bir etken. Sağlık turizmi ile ilintili bir kuruluşa olan yakınlık, daha yüksek kiralarla korele olabilir.

**Sağlık Turizmi ile İlintili Kuruluşların Coğrafi Dağılımı**

Aklımdaki sorunun coğrafi bir tarafı da olduğundan dolayı, ilk iş olarak sağlık turizmi kuruluşlarının ve Airbnb dairelerinin İstanbul içinde nasıl dağıldığına bakmak mantıklı olabilir. Böylece bu iki olgunun arasındaki ilişkiyi daha iyi kavrayabiliriz. İBB’nin (İstanbul Büyükşehir Belediyesi) İstanbul’daki sağlık kuruluşları hakkında yayınladığı bir veritabanından ve kendi yazdığım web scraping kodundan yararlanarak İstanbul’daki sağlık turizmi ile ilintili kuruluşları haritaladım. Aşağıdaki iki harita, sağlık turizmi ile ilintili kuruluşları hem ilçe seviyesinde hem de tekli olarak gösteriyor.

Diagram

Description automatically generated

*İstanbul’daki sağlık turizmi ile ilintili kuruluşların dağılımını gösteren iki harita ve bir çubuk grafiği. Estetik merkezleri ve saç ekim merkezleri, diğer ilçelerle kıyasla daha merkezi ve erişimi daha kolay ilçelerde toplanmışa benziyorlar.*

En üstteki koroplet harita, sağlık turizmi ile ilintili kuruluşların en çok Şişli, Beşiktaş ve Kadıköy gibi güneyde bulunan ve merkezi ilçelerde kümelendiğini gösteriyor. Sırasıyla 33, 19 ve 17 sağlık kuruluşuna sahip bu üç ilçeye, İstanbul’un diğer ilçelerinden kolayca ulaşılabiliyor. Bu ilçeler, sıralamadaki 4. ve 5. İlçeler olan Ataşehir ve Üsküdar ile beraber günlük hayatta bir çok insanın uğradığı önemli merkezler. Ayrıca İstanbul’un güney kıyısındaki ilçelerin diğer ilçelere kıyasla daha kalabalık bir nüfusa sahip olduğuna da burada belirtmemiz gerekir.

Sağlık turizmi ile ilintili kuruluşların genel dağılımını gösteren haritaya bakınca yukarıdaki tespitimizin doğru olduğunu anlıyoruz. Güney ilçeleri içinde bile, çoğu sağlık turizmi kuruluşu güney kıyı hattına yakın bir şekilde konumlanmış. Şehrin tam güneyindeki ana kümelenmenin dışında çok az yerde sağlık turizmi ile ilintili bir kuruluş var.

Özellikle de turizm konusunda şehrin ağırlık merkezinin güneyde olduğunu göz önünde bulundurduğumuzda, yukarıda not ettiğimiz gözlemler aslında pek şaşırtıcı değiller. Turistler en çok güneydeki ilçeleri ziyaret ettiğinden dolayı, bu ilçeler sağlık turizmi ile ilintili bir klinik açmak isteyenler için ilk seçenek olmalı.

Gözlemlediğimiz coğrafi dağılımı başka faktörlerle ilişkilendirmek, Airbnb dairelerini incelemeden önce elimizdeki konu hakkında biraz daha bilgi sahibi olmamızı sağlayabilir. Mesela, bir ilçede bulunan sağlık turizmi ile ilintili kuruluş sayısıyla o ilçenin nüfusu ve yıllık ortalama hanehalkı geliri arasındaki ilişkiyi bir dağılım grafiği ile inceleyebiliriz.

Chart, scatter chart

Description automatically generated

*Bir ilçede bulunan sağlık turizmi ile ilintili kuruluşların sayısı (bağımlı değişken) ile iki farklı bağımsız değişken arasındaki ilişkiyi gösteren iki dağılım grafiği. Bağımlı değişken ve nüfus arasında herhangi bir doğrusal ilişki yok, fakat yıllık ortalama hanehalkı geliri ile bağımlı değişken arasında çok cılız bir doğrusal ilişki var.*

Yukarıdaki iki dağılım grafiği bize bir ilçenin nüfusu ile o ilçedeki sağlık turizmi ile ilintili kuruluşların sayısı arasında herhangi bir doğrusal ilişki olmadığını gösteriyor. Bu gerçeği göz önünde bulundurunca daha önce bahsettiğimiz şekilde güneydeki ilçelerin nüfus olarak diğer ilçelerden daha kalabalık olması, konumuz özelinde bir şey ifade etmiyor. Bunun yanında, bir ilçenin yıllık ortalama hanehalkı geliri ile o ilçedeki sağlık turizmi ile ilintili kuruluş sayısı arasında cılız bir doğrusal ilişki olduğunu görebiliyoruz.

Şimdi kiralık Airbnb dairelerin coğrafi dağılımını incelemeye geçebiliriz.

**Kiralık Airbnb dairelerinin coğrafi dağılımı**

Airbnb daireleri ile ilgili veriyi, *Inside Airbnb* adındaki bir proje aracılığıyla buldum. Yazıyı yazdığım vakitte halen neredeyse tüm dünyada etkili olan koronavirüs salgını turizme büyük bir darbe vurduğundan dolayı, analizimi Mayıs 2019 verilerine dayandırmayı seçtim.

Graphical user interface, map

Description automatically generated

*İstanbul’daki kiralık Airbnb dairelerinin dağılımını gösteren iki harita ve bir çubuk grafiği. En çok Airbnb dairesine sahip ilçelerin sıralaması bir, önceki sıralamayla ile tamamen aynı olmasa da benzerlik gösteriyor.*

Tepedeki koroplet haritaya bakınca, en çok kiralık Airbnb dairesinin bir önceki dağılıma benzer bir şekilde yine güneydeki boğaza yakın ilçelerde olduğunu görüyoruz. Önceki sıralamada ilk beşte yer alan Şişli, Beşiktaş ve Kadıköy ilçeleri bu sıralamada da ilk beşte. Fakat ilk sıralamanın aksine, bu sefer 1.lik ve 3.lük Beyoğlu ve Fatih’e ait.

Genel dağılıma baktığımızda kiralık Airbnb dairelerinin, sağlık turizmi ile ilintili kuruluşlara kıyasla daha az kümelenmiş bir şekilde dağıldığını görüyoruz. Airbnb dairelerinin fazlalığı ve haritanın ölçeği en büyük küme içerisindeki dağılımın detaylarını görmeyi engelliyor. Fakat yine de sağlık turizmi ile ilintili kuruluşların hiç olmadığı bazı yerlerde kiralık Airbnb dairelerinin olduğunu gözlemleyebiliyoruz. Kuzeydoğudaki Şile ilçesinde, Güneybatıdaki Silivri ilçesinde ve Adalar ilçesinde bulunan Airbnb daireleri, ana kümeden bağımsız kümeler oluşturuyor. Mesafe ve fiyat arasında herhangi bir analiz yapmadan bile, bu bağımsız kümelerin varlığının anlamlı bir doğrusal veya monotonik ilişki gözlemleme şansını azalttığını söyleyebiliriz.

Bir sonraki aşamaya geçmeden önce, bir ilçedeki kiralık Airbnb dairesi sayısının o ilçenin nüfusuyla ve yıllık ortalama hanehalkı geliriyle nasıl bir ilişkisinin olduğuna bakalım.

Chart, scatter chart

Description automatically generated

*Bir ilçede bulunan kiralık Airbnb dairesi sayısı (bağımlı değişken) ile iki farklı bağımsız değişken arasındaki ilişkiyi gösteren iki dağılım grafiği. Bağımlı değişken, ne nüfusla ne de yıllık ortalama hanehalkı geliriyle herhangi bir doğrusal ilişkiye sahip değil.*

Sonuçlar bir önceki korelasyon analizine çok benzer. Bir ilçede bulunan kiralık Airbnb dairelerinin sayısı ile, o ilçenin nüfusu veya yıllık ortalama hanehalkı geliri arasında anlamlı herhangi bir ilişki yok.

**Varlıktan Niteliğe: Airbnb Kiraları**

İlgilendiğimiz iki varlığın da coğrafi dağılımı hakkında bir anlayışa ulaştık. Fakat en nihayetinde aslında ilgilendiğimiz şey bu varlıklardan birine ait iki niteliğin (günlük kira ve en yakın sağlık kuruluşuna olan mesafe) arasındaki ilişki. Ana incelememizin bağımlı değişkeni Airbnb kiraları. Bu değişkenin başka bir değişkenle olan ilişkisini incelemeden önce kendisini incelemek mantıklı olabilir. Kiraların dağılımına bakmak için bir histogram grafiği kullanalım:

Chart, histogram

Description automatically generated

*Airbnb kiralarının dağılımını gösteren iki histogram. İlk histogram bize sadece dağılımın yüksek seviyede pozitif çarpıklığa (right skew) sahip olduğunu gösteriyor. İkinci histogram ise kira dağılımını daha farklı bir detay seviyesinde görmemizi sağlıyor.*

Yukarıdaki ilk histograma baktığımızda gördüğümüz tek şey, çok uzun bir sütun ve onu çevreleyen birkaç küçük sütün. Bu görüntü, ilgilendiğimiz değişkenler arasında anlamlı bir doğrusal veya monotonik ilişki bulma şansımızı bir hayli azaltıyor. Grafiğin bu yapısı, kira dağılımının yüksek seviyede pozitif çarpıklığa sahip olduğunu gösteriyor. Dağılımın çarpıklık değeri (skewness) 33.74. 0’dan büyük bir çarpıklık değeri, dağılımın sağına doğru aykırı değerler olduğunu gösterir. Bu aykırı değerler, dağılımın ortalamasını sağa doğru çeker, dağılımın şeklini normal dağılım şeklinden uzaklaştırır ve bizim doğrusal bir ilişki bulma şansımızı azaltır.

Kira dağılımının sağa doğru çarpık olması aynı zamanda histogram grafiğinin bize sunduğu bilgiyi kısıtlar. İlk histogram’ın şu anda göründüğü gibi görünmesinin sebebi, sayıca az ama etki olarak fazla olan bazı değerleri görselleştirmeye çalışmasıdır. Bu çabadan dolayı şeklin solunda olan değerlerin dağılımını detaylı bir şekilde inceleyemiyoruz. İkinci histogram, kira değerlerinin log10 dönüştürülmüş halini gösteriyor. Bu dönüşüm sağa doğru çarpık dağılımın solundaki değerleri daha iyi görmemizi sağlıyor. Ortalamaya daha yakın değerleri görebilmek yine de doğrusal bir ilişki keşfedebilme şansımızı pek arttırmıyor.

A picture containing chart

Description automatically generated

*Airbnb daire kiralarının dağılımını ilçe seviyesinde gösteren çoklu histogram grafiği. Gayet bonkör bir çarpıklık sınırı koysak bile çoğu dağılım halen yüksek seviyede sağa doğru çarpık.*

Airbnb kiralarının dağılımındaki sağa çarpıklığı, kiralara ilçe seviyesinde baktığımızda da görüyoruz. Çarpıklık değeri sınırını gayet bonkör ve tamamen keyfi bir şekilde 5 kabul etsek bile bazı ilçeler (sarı renkli) halen aşırı derecede sağa doğru çarpık. Çoğu ilçenin çarpıklık değeri ise en az 1.

Şu ana kadar yaptığımız incelemelerde, ilgilendiğimiz iki değişken arasında herhangi bir doğrusal veya monotonik ilişki bulmamızın çok zor olduğu sinyalini veren bir sürü işaretle karşılaştık. Yine de buraya kadar gelmişken en azından ne kadar hatalı bir varsayımda bulunduğumuza bakacağız. Fakat bundan önce “en yakın sağlık kuruluşuna olan mesafe” bilgisine nasıl ulaştığıma bakalım.

**Kısa bir ara: en yakın komşu analizini nasıl yaptım?**

**Map

Description automatically generated**

*Bu projede kullanılan kodun, en yakın komşu analizini doğru yaptığını gösteren soyut harita. Her bir Airbnb dairesi (mavi yuvarlak) en yakınındaki sağlık turizmi ile ilintili kuruluş (turuncu üçgen) ile eşleşmiş.*

Projenin kaynak kodunda (bkz. Referanslar), coğrafi verileri işlemek ve her Airbnb’ye en yakın sağlık turizmi ile ilintili kuruluşu (“en yakın komşu) bulmak için iki adet popüler Python GIS paketi (GeoPandas ve Shapely) kullandım. Bu analizden sonra, başka bir Python paketi yardımıyla her bir Airbnb ile eşleştiği kuruluşun arasındaki jeodezik mesafeyi hesapladım. Metre cinsinden jeodezik mesafe hesaplaması, iki nokta arasındaki en kısa yolu baz alır. Bu hesaplama yol yapısı ve yükseklik gibi faktörleri dikkate almaz. Hesabı düzgün yapabilmek adına, projede kullanılan coğrafi verileir WGS-84 koordinat referans sistemine (CRS) uygun hale getirdim.

Yukarıdaki soyut İstanbul haritası, en yakın komşu analizimin doğru çalıştığına dair bir kanıt. Bu haritada her mavi yuvarlak (Airbnb daireleri) en yakınındaki turuncu üçgene (sağlık turizmi ile ilintili kuruluş) beyaz bir çizgi ile bağlı. Haritanın çözünürlüğü her bir bağlantıyı takip etmek için uygun değil. Fakat yine de haritanın güneybatısındaki (Silivri) gibi alt kümelere bakınca eşleşmenin başarılı olduğunu görebiliyoruz.

**Hüsran: Airbnb kiraları ve en yakın sağlık turizmine olan mesafe arasındaki korelasyon**

Gereken tüm tanımları yaptığımıza ve neyi inceleyeceğimizi iyice anladığımıza göre asıl analizimize geçebiliriz.

Chart, scatter chart

Description automatically generated

*Airbnb kiraları ve en yakındaki sağlık kuruluşlarına olan mesafeleri gösteren dağılım grafiği. İki değişken arasında doğrusal veya monotonik bir ilişki yok.*

Yukarıdaki dağılım grafiği, yol boyunca gözlemlediğimiz sıkıntıları doğruluyor. İlgilendiğimiz iki değişken arasında doğrusal bir ilişki (r = 0.02) yok. Aynı şekilde anlamlı bir monotonik (rho = -0.01) ilişkiden söz etmemiz de mümkün değil.

Bu sonucun nedeni çok açık: Hem kira değeri verisinde hem de mesafe verisinde, diğer değerlere kıyasla aşırı bir sürü değer var. Sorumuzun cevabı şimdilik hayır.

**İkinci Deneme: Normalleştirilmiş Airbnb kiraları ve en yakın sağlık turizmine olan mesafe arasındaki korelasyon**

Doğrusal veya monotonik bir ilişki gözlemleyememizin sebeplerinden birinin, fiyat dağılımında bulunan çok sayıda uç değer olduğunu biliyoruz. Peki bu uç değerlerden kurtulsak ne olur?

Normalde işimize gelmeyen verileri göz ardı etmek, doğruculuk adına yapabileceğimiz en kötü şeylerden biri. Fakat bu analiz, az buzuk emek vermediğim ve eğlencesine yaptığım bir proje. Bu yüzden bir istatistik suçu işleyerek ilk başta düştüğüm yanılgıyı kanıtlamayı bir kez daha deneyebilirim.

Genelde dağılımın 1. çeyreğinden 1.5 IQR daha altta yer alan değerler ile aynı dağılımın 3. çeyreğinden 1.5 IQR daha üstte yer alan değerler, uç değer olarak tanımlanır. Biz de bu tanımı kullanalım ve dağılımımıza tekrar bakalım.

Chart, histogram

Description automatically generated

*Normalleştirilmiş Airbnb kiralarının dağılımını gösteren histogram grafiği. Kira dağılımı sonuç olarak normal dağılıma daha benzer bir hale gelmiş.*

Yukarıdaki metodu kullanınca kira dağılımının normal dağılıma daha benzer bir hale geldiğini görebiliyoruz. Çarpıklık değeri (skewness) 1.06 ve standart sapması 165.55 olan bu dağılım, öncekine kıyasla çok daha kısa bir aralığa dağılmış.

**Peki ilçe seviyesinde durum nasıl?**

**A picture containing chart

Description automatically generated**

*Normalleştirilmiş Airbnb daire kiralarının dağılımını ilçe seviyesinde gösteren çoklu histogram grafiği. Tüm dağılımların çarpıklık değeri 2.20’den küçük.*

**Peki şimdi elimizdeki normalleştirilmiş verileri kullanarak aradığımızı bulabilir miyiz?**

**Chart, scatter chart

Description automatically generated**

*Normalleştirilmiş Airbnb kiraları ve en yakındaki sağlık kuruluşlarına olan mesafeleri gösteren dağılım grafiği. İki değişken arasında doğrusal veya monotonik bir ilişki yok.*

Elimizdeki veri kümesini normalleştirsek de, iki değişken arasında ne doğrusal ne de monotonik bir ilişki bulamıyoruz.

**Üçüncü deneme: ilke seviyesinde korelasyon**

Pozitif bir cevap almak adına son bir deneme yapıp ölçeğimizi daraltabiliriz. Belki de bütün İstanbul’u değil de sadece bazı ilçeleri ele aldığımızda doğrusal veya monotonik bir ilişki bulabiliriz?

Dikkatimizi, sağlık turizmi ile ilintili kuruluş sayısının en çok olduğu Şişli, Beşiktaş, Kadıköy, Ataşehir ve Üsküdar ilçelerine kaydıralım. Önceki denemede işlediğimiz suçu benimseyelim ve bu ilçeler için de normalleştirme uygulayalım.

Chart

Description automatically generated

*Normalleştirilmiş Airbnb daire kiralarının dağılımını, hem ortaklaşa hem de ilçe seviyesinde gösteren çoklu histogram grafiği. Bu seçkinin çarpıklık değeri (1.31), normaleştirilmiş tüm ilçelerin çarpıklık değerinden (1.06) daha yüksek.*

Yukarıdaki büyük histogram, bu seçkinin çarpıklık değerinin (1.31) tüm ilçeleri içeren dağılımın çarpıklık değerinden (1.06) daha yüksek olduğunu gösteriyor. İlçe ilçe baktığımızda da pek farklı bir durum yok. Sonuç olarak analizimizi ilçe bazında yapmak pek bir fayda etmeyecekmiş gibi duruyor. Yine de başladığımızı bitirelim ve sonuca bakalım.

Aşağıdaki grafik en yakın komşu analizinin ilçe seviyesinde de çalıştığını gösteriyor.

A picture containing map

Description automatically generated

*Bu projede kullanılan kodun, ilçe bazında en yakın komşu analizini doğru yaptığını gösteren beş soyut harita. Her bir Airbnb dairesi (mavi yuvarlak) en yakınındaki sağlık turizmi ile ilintili kuruluş (turuncu üçgen) ile eşleşmiş.*

Bu dağılım grafikleri de analizimizin sonucunu gösteriyor:

Chart, scatter chart

Description automatically generated

*Normalleştirilmiş Airbnb kiraları ve en yakındaki sağlık kuruluşlarına olan mesafeleri gösteren beş dağılım grafiği. Seçilen beş ilçede de iki değişken arasında doğrusal veya monotonik bir ilişki yok.*

Bazı ilçelerde doğrusal ve monotonik ilişkilerin yönünü ve kuvvetini gösteren Pearson r ve Spearman rho katsayıları değişiklik gösterse de, yine de aradığımız türden bir ilişki bulduğumuzu kesinlikle söyleyemeyiz.

**Sonuç**

**Diagram, schematic

Description automatically generated** *Farklı şekillerde yanılmışım. Bu yazı, yanılgıma bir övgüdür. Çizim Nathan W. Pyle’a () ait.*

Neresinden bakarsak bakalım, Airbnb kiraları ve en yakın sağlık kuruluşuna olan mesafe arasında herhangi bir doğrusal veya monotonik ilişki bulunduğuna dair kanıt bulamadık. Bu yazı, en azından denediğimin göstergesidir.

**Referanslar**

Bu analiz, tekrarlanabilirlik ilkesi düşünülerek hazırlanmışdır. Tüm kaynak koduna, veri kümelerine ve referanslara bu GitHub repository’sinden (<https://github.com/ejgenc/Data-Analysis_Istanbul-Health-Services-Map>) ulaşabilirsiniz.