

**İST432-GRAFİKSEL VERİ ANALİZİ VİZE ÖDEVİ**

JITTER, RAINCLOUD, HEATMAPS VE CHOROPLETH GRAFİKLERİ İLE İSTATİSTİKSEL ANALİZ İNCELEMESİ

2025

***Dr. Öğr. Üyesi Onur Toka***

*ŞİMAL VİZE – 2210329027*

*İREM TARIM – 2210329047*

ÇANKAYA, ANKARA

İçindekiler Tablosu

[1. Analiz Yapılacak Veri 5](#_Toc198849776)

[1.1 Normallik Testi 6](#_Toc198849777)

[1.2. Korelasyon 7](#_Toc198849778)

[2. Çoklu Doğrusal Regresyon Modeli 8](#_Toc198849779)

[3. Raincloud Grafiği (Yağmur Bulutu) Nedir? 12](#_Toc198849780)

[4. Avantajları 13](#_Toc198849781)

[5. Dezavantajları 14](#_Toc198849782)

[6. Dikkat Edilmesi Gerekenler 14](#_Toc198849783)

[7. Raincloud Grafik 15](#_Toc198849784)

[7.1 Grup Karşılaştırma Testi 16](#_Toc198849785)

[7.2 Kurs Durumuna Göre Puanlar için Korelasyon 17](#_Toc198849786)

[8. Kümeleme Analizi ile Raincloud 18](#_Toc198849787)

[9. Raincloud için Sonuç ve Değerlendirme 20](#_Toc198849788)

[10. Jitter (Titreme) Grafiği Nedir? 21](#_Toc198849789)

[11. Avantajları 22](#_Toc198849790)

[12. Dezavantajları 23](#_Toc198849791)

[13. Dikkat Edilmesi Gerekenler 24](#_Toc198849792)

[14. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı İçin Jitter Grafik 25](#_Toc198849793)

[14.1. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için ANOVA Sonuçları 27](#_Toc198849794)

[14.2. Tukey Çoklu Karşılaştırma Testi 27](#_Toc198849795)

[15. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için Kümeleme 28](#_Toc198849796)

[15.1. Kümeleme ile Oluşturulmuş Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için ANOVA Sonuçları 29](#_Toc198849797)

[15.2 Tukey Çoklu Karşılaştırma Testi 30](#_Toc198849798)

[16. Jitter için Sonuç ve Değerlendirme 31](#_Toc198849799)

[17. Genel Sonuç ve Değerlendirme 31](#_Toc198849800)

[18. Isı Haritası (Heatmaps) Nedir? 33](#_Toc198849801)

[18.2. Isı Haritası Çeşitleri 34](#_Toc198849802)

[18.3. Avantajları 36](#_Toc198849803)

[18.4. Dezavantajları 37](#_Toc198849804)

[19. Analiz Yapılacak Veri 37](#_Toc198849805)

[19.1. Açıklayıcı Analizler 38](#_Toc198849806)

[19.2. Korelasyon 39](#_Toc198849807)

[19.2. Normallik Testi 40](#_Toc198849808)

[20. Grafik Yorumları 40](#_Toc198849809)

[21. Kümeleme Analizi 42](#_Toc198849810)

[21.1. Z Normalizasyonu ile K means Kümeleme 42](#_Toc198849811)

[21.1.1. Kümelere Göre Kruskal-Wallis ve Dunn’s Testi 43](#_Toc198849812)

[21.1.2. Isı Grafiği Gösterimi 44](#_Toc198849813)

[21.2. Z Normalizasyonu ile Hiyerarşik Kümeleme 45](#_Toc198849814)

[21.3. Z Normalizasyonu ile K Means++ 46](#_Toc198849815)

[21.3.1. Kruskal-Wallis ve Dunn’s Testi 47](#_Toc198849816)

[21.3.2. Isı Haritası 48](#_Toc198849817)

[22. Sonuç ve Değerlendirme 49](#_Toc198849818)

[23. Koroplet Haritası (Choropleth Map) Nedir? 50](#_Toc198849819)

[23.2. Avantajları ve Dezavantajları 51](#_Toc198849820)

[24. Isı Haritası ile Koroplet Farkı 52](#_Toc198849821)

[25. Analiz Edilecek Verinin Açıklanması 53](#_Toc198849822)

[25.2. Açıklayıcı Analizler 54](#_Toc198849823)

[25.3. Korelasyon 55](#_Toc198849824)

[25.4. Normallik Testi Sonuçları 56](#_Toc198849825)

[26. Kümeleme Analizi 56](#_Toc198849826)

[26.2. Z Normalizasyonu ile Yapılan K means ve K means++ Karşılaştırılması 56](#_Toc198849827)

[26.3. K means++ için Kruskal-Wallis Testi Sonuçları 58](#_Toc198849828)

[26.4. Dunn’s Testi Sonuçları 58](#_Toc198849829)

[27. Oranlaştırılmış Veri ile Analiz 60](#_Toc198849830)

[27.2. Açıklayıcı Analizler 60](#_Toc198849831)

[27.3. Korelasyon 61](#_Toc198849832)

[27.4. Normallik Testi 62](#_Toc198849833)

[28. Kümeleme Analizi 62](#_Toc198849834)

[28.2. Kruskal-Wallis Testi 64](#_Toc198849835)

[28.3. Dunn’s Testi 64](#_Toc198849836)

[29. Grafik Yorumlaması 66](#_Toc198849837)

[30. Sonuç ve Değerlendirme 70](#_Toc198849838)

[Kaynakça 71](#_Toc198849839)

# Analiz Yapılacak Veri

Veri seti, Kaggle platformunda yayımlanan Student Performance in Exams başlığıyla erişilebilen bir veri kümesidir. Amerika Birleşik Devletleri'ndeki lise öğrencilerinin çeşitli derslerden elde ettiği notlardan ve bu notları etkileyebilecek bazı sosyo-demografik faktörleri içermektedir. Veri seti toplam 1000 öğrenciye ait gözlem içermektedir ve 8 değişken bulunmaktadır. Bu değişkenler sürekli ve kategoriklerden oluşmaktadır.

**Cinsiyet:** Öğrencinin cinsiyetini belirtir.

**Irk / Etnik Köken:** Öğrencilerin etnik gruplarını temsil eder. Beş farklı kategori içerir: grup A, grup B, grup C, grup D, grup E. Ancak gruplar hangi etnik grupları kapsadığı hakkında bilgi sahibi olunmadığından analize eklenmeyecektir.

**Ebeveyn Eğitim Düzeyi:** Öğrencinin ebeveynlerinden birinin en yüksek eğitim düzeyini belirtir. Lise mezunu, üniversite deneyimi olan, Ön Lisans Diploması, Lisans Diploması, Yüksek Lisans Diploması olan olarak çeşitli kategoriler içerir.

**Öğle Yemeği Türü:** Öğrencinin okulda aldığı öğle yemeği türünü belirtir. Standart ve ücretsiz/indirimli olmak üzere iki gruba ayrılır.

**Sınav Hazırlık Kursu:** Öğrencinin sınav öncesinde herhangi bir hazırlık kursuna katılıp katılmadığını belirtir. Tamamladı ve katılmadı olmak üzere iki düzeylidir.

**Matematik Puanı:** Öğrencinin matematik sınavından aldığı sayısal puandır. Değer aralığı 0–100’dür.

**Okuma Puanı:** Öğrencinin okuma becerileri sınavından aldığı sayısal puandır. Aralık 0–100 aralığında yer alır.

**Yazma Puanı:** Öğrencinin yazma becerileri sınavından aldığı puanı gösterir. Diğer iki puan gibi 0–100 aralığındadır.

## Normallik Testi

Tablo 1.Shapiro-Wilk Sonuçları

|  |  |  |
| --- | --- | --- |
|  | **W** | **P Değeri** |
| **Matematik Puanı:** | 0.9890 | 0.0584 |
| **Okuma Puanı:** | 0.9900 | 0.1199 |
| **Yazma Puanı:** | 0.9890 | 0.0587 |

Değişkenlerin dağılımlarının normal olup olmadığını test etmek amacıyla uygulanan **Shapiro-Wilk normallik testi** sonuçlarına göre okuma puanları için p-değeri 0.1199 olup, 0.05 anlamlılık düzeyinin üzerindedir. Bu durumda yokluk hipotezi reddedilemez, yani okuma puanlarının dağılımı normal dağılıma uygun kabul edilir.

Matematik ve yazma puanlarının p-değerleri ise sırasıyla 0.0584 ve 0.0587 olup, 0.05 sınırına yakın olsa da sınırın biraz üzerindedir. Sonuçlar genel olarak üç değişkenin de dağılım yapısının normal dağılıma yakın olduğunu göstermektedir.

## Korelasyon

metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 1. Korelasyon Matrisi

Kategorik değişkenler ile puanlar arasındaki korelasyon analizi sonucunda; Sınav hazırlık kursuna katılım, öğrencilerin akademik performansları üzerinde farklı etkiler yarattığı görülmüştür. Hazırlık kursunu tamamlayan öğrencilerde okuma ve yazma becerileri arasındaki ilişki oldukça güçlüdür (r = 0.95). Bu yüksek korelasyon, iki becerinin birbiriyle yakın bir gelişim gösterdiğini ve öğrencilerin daha çok dilsel becerilerde benzer sonuçlar gösterdiğini ortaya koymaktadır.

Aynı grupta, matematik ile okuma (r = 0.82) ve yazma (r = 0.80) puanları arasındaki ilişkiler de güçlüdür; ancak, bu ilişkiler okuma-yazma arasındaki ilişkiye kıyasla bir miktar daha düşüktür. Bu bulgu, hazırlık kurslarının özellikle sözel beceriler üzerinde daha belirgin bir etkiye sahip olduğunu, matematik puanının ise daha çok bağımsız şekillenebileceğini düşündürmektedir. Hazırlık kursuna katılmayan öğrencilerde ise okuma ve yazma puanları arasındaki ilişki yine oldukça güçlüdür (r = 0.96). Bununla birlikte, bu grupta matematik puanı ile hem okuma (r = 0.86) hem de yazma (r = 0.86) puanları arasındaki korelasyonlar daha yüksektir. Bu durum, hazırlık kursuna katılmayan öğrencilerde becerilerin daha dengeli bir biçimde geliştiğini söyleyebilir. Öğle yemeği durumu matematik puanı üzerinde oldukça anlamlı çıkarken okuma puanı ve yazma puanı için benzer anlamlılık göstermiştir ve hazırlık kursu durumu ve ebeveyn eğitim durumu değişkenlerinin sınav puanları üzerinde farklı etkilere sahip olduğu görülmektedir.

# Çoklu Doğrusal Regresyon Modeli

Bu analizde, öğrencilerin ders puanlarını etkileyen faktörler çoklu doğrusal regresyon modeliyle incelenmiştir. Öğrencilerin matematik, okuma ve yazma alanındaki başarı puanlarını etkileyen faktörleri bir arada ele alabilmek amacıyla üç ayrı çoklu doğrusal regresyon modeli kurulmuştur.

Matematik Puanı için regresyon modeline göre,

Tablo 2. Matematik Puanı için Regresyon Sonuçları

|  |  |  |
| --- | --- | --- |
| **Değişken** | **Katsayı** | **Yorum** |
| **(Sabit Terim)** | 67.95 | Referans grup için beklenen matematik puanı. Bu grup: kursu tamamlayan, kız öğrenciler ve ebeveyn eğitimi lisans mezunu olanlar. |
| **Sınav Hazırlık Kursu: Katılmayanlar** | -7.17 | Kursa katılmayanların matematik puanı, kursa katılanlara göre ortalama 7,17 puan daha düşüktür (diğer değişkenler sabitken). |
| **Cinsiyet: Erkek** | +2.48 | Erkek öğrencilerin puanı, kızlara göre ortalama 2,48 puan daha yüksektir (diğer faktörler sabitken). |
| **Ebeveyn Eğitim Düzeyi:** |  |  |
| **Ön lisans mezunu** | +7.26 | Ebeveyni ön lisans mezunu olan öğrencilerin puanı, referansa göre 7,26 puan daha yüksektir. |
| |  | | --- | | **Lise mezunu** | | -2.51 | Ebeveyni yalnızca lise mezunu olan öğrenciler 2,51 puan daha düşük performans göstermektedir. |
| **Yüksek lisans mezunu** | +2.31 | Ebeveyni yüksek lisans mezunu olan öğrencilerin puanı 2,31 puan daha yüksektir. |
| |  | | --- | |  |  |  | | --- | | **Üniversiteyi tamamlanmamış** | | +0.26 | Etki çok küçüktür, fark anlamlı olmayabilir. |
| **Lise eğitimi tamamlanmamış** | -2.88 | Ebeveyni sadece ortaokul/lise düzeyinde olan öğrencilerin puanı 2,88 puan daha düşüktür. |
| **Öğle Yemeği Türü: Standart** | 13.3319 | Standart öğün yiyen öğrencilerin yazma puanı, ücretsiz/indirimli öğün yiyenlere göre ortalama 13,3 puan daha yüksek. |

Yazma Puanı için regresyon modeline göre,

Tablo 3. Yazma Puanı için Regresyon Sonuçları

|  |  |  |
| --- | --- | --- |
| **Değişken** | **Katsayı** | **Yorum** |
| **Sabit Terim** | 79.83 | Referans grup için beklenen Yazma puanı. Bu grup: kursu tamamlayan, kız öğrenciler ve ebeveyn eğitimi lisans mezunu olanlar |
| **Sınav Hazırlık Kursu: Katılmayanlar** | -11.58 | Kursa katılmayanların yazma puanı, kursa katılanlara göre ortalama 7,17 puan daha düşüktür (diğer değişkenler sabitken). |
| **Cinsiyet: Erkek** | |  | | --- | |  |  |  | | --- | | -10.19 | | Erkek öğrenciler, yazma puanında kızlara göre 10,19 puan daha düşük |
| **Ebeveyn Eğitim Düzeyi:** |  |  |
| **Ön lisans mezunu** | +4.23 | Ebeveyni ön lisans mezunu olan öğrencilerin puanı, referansa göre 4,23 puan daha yüksektir. |
| |  | | --- | |  |  |  | | --- | | **Lise mezunu** | | -3.02 | Ebeveyni yalnızca lise mezunu olan öğrenciler 3,02 puan daha düşük performans göstermektedir. |
| **Yüksek lisans mezunu** | +3.88 | Ebeveyni yüksek lisans mezunu olan öğrencilerin puanı 3,88 puan daha yüksektir. |
| |  | | --- | |  |  |  | | --- | | **Üniversiteyi tamamlanmamış** | | -0.73 | Etki çok küçüktür, fark anlamlı olmayabilir. |
| **Lise eğitimi tamamlanmamış** | -4.53 | Ebeveyni sadece ortaokul/lise düzeyinde olan öğrencilerin puanı 4,53 puan daha düşüktür. |
| **Öğle Yemeği Türü: Standart** | 10.506 | Standart öğün yiyen öğrencilerin yazma puanı, ücretsiz/indirimli öğün yiyenlere göre ortalama 10.50 puan daha yüksek. |

Okuma Puanı için regresyon modeline göre,

Tablo 4. Okuma Puanı için Regresyon Sonuçları

|  |  |  |
| --- | --- | --- |
| **Değişken** | **Katsayı** | **Yorum** |
| **Sabit Terim** | 79.10 | Referans grup için beklenen Okuma puanı. Bu grup: kursu tamamlayan, kız öğrenciler ve ebeveyn eğitimi lisans mezunu olanlar |
| **Sınav Hazırlık Kursu: Katılmayanlar** | -9.38 | Kursa katılmayanların Okuma puanı, kursa katılanlara göre 9,38 puan daha düşüktür (diğer değişkenler sabitken). |
| **Cinsiyet: Erkek** | |  | | --- | |  |  |  | | --- | | -8.79 | | Erkek öğrenciler, okuma puanında kızlara göre 8,79 puan daha düşük |
| **Ebeveyn Eğitim Düzeyi:** |  |  |
| **Ön lisans mezunu** | +3.12 | Ebeveyni ön lisans mezunu olan öğrencilerin puanı, referansa göre 3,12 puan daha yüksektir. |
| |  | | --- | |  |  |  | | --- | | **Lise mezunu** | | -2.52 | Ebeveyni yalnızca lise mezunu olan öğrenciler 2,52 puan daha düşük performans göstermektedir. |
| **Yüksek lisans mezunu** | +4.11 | Ebeveyni yüksek lisans mezunu olan öğrencilerin puanı 4,11 puan daha yüksektir. |
| |  | | --- | |  |  |  | | --- | | **Üniversiteyi tamamlanmamış** | | -1.39 | Etki küçüktür, Ebeveyni üniversiteyi tamamlanmamış olanların düzeyinde olan öğrencilerin puanı 1,39 puan daha düşüktür. Fark anlamlı olmayabilir. |
| **Lise eğitimi tamamlanmamış** | -3.62 | Ebeveyni sadece ortaokul/lise düzeyinde olan öğrencilerin puanı 3.62 puanı daha düşüktür. |
| **Öğle Yemeği Türü: Standart** | 10.458 | Standart öğün yiyen öğrencilerin yazma puanı, ücretsiz/indirimli öğün yiyenlere göre ortalama 10,45 puan daha yüksek. |

Sonuç olarak, hazırlık kursu, cinsiyet ve ebeveyn eğitimi öğle yemeği türü gibi faktörler, öğrencilerin akademik başarıları üzerinde anlamlı etkiler yaratmaktadır. Üç beceri türü arasında etki büyüklükleri farklılık göstermekle birlikte, özellikle sözel alanlarda (okuma ve yazma) hazırlık kursu ve cinsiyet farkları daha belirgin hale gelmiştir. Bu bulgulara göre analizlerle beraber yapılacak grafik görselleştirme için hangi faktörlerin daha çok etkilediği ile fikir sunmuş olur.

# Raincloud Grafiği (Yağmur Bulutu) Nedir?

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

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, diyagram, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 2. Raincloud Grafiğini oluşturan grafik türleri

Raincloud grafikleri, bir nicel değişkenin dağılımını, merkezi eğilim ölçülerini ve bireysel gözlemlerini aynı anda görselleştiren modern ve çok katmanlı veri gösterim araçlarıdır. Bu grafik türü, yoğunluk eğrisi (violin plot), kutu grafik (boxplot) ve ham veri noktalarını (jittered points) aynı düzlemde birleştirerek, geleneksel grafik türlerinin sunduğu bilgileri daha kapsamlı ve görsel olarak zengin bir biçimde sunar. Her bir grafik öğesi, verinin farklı yönlerini yansıtır: yoğunluk eğrisi verinin genel dağılım şeklini; kutu grafik medyan, çeyrekler ve uç değerler gibi merkezi özetleri; ham veri noktaları ise her bir gözlemi doğrudan temsil eder. Bu yapı sayesinde hem genelleştirilmiş eğilimler hem de örneklemin ayrıntılı bileşimi eş zamanlı olarak izlenebilir.

Raincloud grafiğinde her kategori için yoğunluk eğrisini, bu eğrinin altında kutu grafiği ve onun da altında veya yanında ham veri noktalarını barındıracak şekilde yapılandırılır. Bu yapı, kullanıcıya hem gruplar arası farkları hem de grup içi dağılım örüntülerini daha bütüncül biçimde analiz etme olanağı sunar. Özellikle örneklem sayısı düşük olduğunda veya verinin dağılımı simetrik olmadığında, geleneksel grafiklerin gizleyebileceği varyasyonları ve aykırılıkları raincloud grafikler ortaya çıkarabilir.

Bu grafik türü ilk kez 2019 yılında Allen ve çalışma arkadaşları tarafından bilimsel literatüre kazandırılmıştır. “Raincloud plots: a multi-platform tool for robust data visualization” başlıklı çalışmada tanıtılan bu grafik hem R hem Python hem de MATLAB gibi farklı yazılım ortamlarında kolaylıkla üretilebilmesi nedeniyle kısa sürede geniş bir kullanıcı kitlesi tarafından benimsenmiştir. Geliştiriciler, klasik boxplot grafiklerinin sınırlı bilgi sunması ve yoğunluk grafiklerinin ham veriyi gizlemesi gibi dezavantajlarını aşmak için bu çok bileşenli yapıyı önermişlerdir. Böylece hem nicel betimleme hem de görsel açıklık bir arada sağlanmıştır.

Raincloud grafikler, psikoloji, eğitim bilimleri, sağlık bilimleri ve davranışsal araştırmalar gibi bireysel farklılıkların yoğun olduğu disiplinlerde sıklıkla tercih edilmektedir. Özellikle grup karşılaştırmaları, deneysel araştırmalar ve skor dağılımları gibi alanlarda, grupların merkezi eğilimleri ile dağılımlarındaki çeşitlilikleri eş zamanlı olarak sunabilmesi bu grafik türünün öne çıkan özelliklerindendir.

# Avantajları

**Dağılım Yapısını Ayrıntılı Sunma:** Raincloud grafikler, verinin dağılımını yoğunluk eğrisiyle, merkezi eğilimlerini kutu grafikle ve her bir gözlemi noktasal olarak sunarak, kullanıcıya çok boyutlu bir analiz imkânı sağlar. Böylece hem genel eğilimler hem de istisnai değerler aynı grafik üzerinde gözlemlenebilir.

**Ham Veri Görünürlüğü Sağlama:** Klasik grafiklerin aksine, her bir gözlemi ayrı bir nokta olarak gösterdiğinden veri setinin tamamı doğrudan görselleştirilir. Bu sayede aykırı değerler veya kümelenme örüntüleri kolayca fark edilebilir.

**Küçük Örneklemler İçin Uygunluk:** Örneklem büyüklüğünün az olduğu durumlarda boxplot veya violin grafikler bilgi kaybına neden olabilirken, raincloud grafik bu sınırlamayı ortadan kaldırır. Her gözlemi temsil etmesi, analizin derinliğini artırır.

**Görsel ve Estetik Zenginlik:** Farklı renk seçenekleri, saydamlık düzeyleri ve grafik bileşenlerinin uyumlu birleşimi ile hem akademik hem sunumsal amaçlar için estetik açıdan etkileyici sonuçlar üretir.

**Gruplar Arası Karşılaştırmayı Kolaylaştırma:** Aynı grafik üzerinde birden fazla grubun verisi gösterilebildiği için gruplar arası farklar hem dağılım düzeyinde hem ortalama seviyesinde rahatlıkla kıyaslanabilir.

**Dağılım Şekline Göre Esneklik Sağlama:** Verinin normal dağılmadığı durumlarda bile etkili bir analiz sunar. Çünkü yoğunluk eğrileri, çarpıklık gibi dağılımsal özellikleri açıkça gösterir.

# 5. Dezavantajları

**Yorumlama Zorluğu:** Boxplot, yoğunluk eğrisi ve veri noktalarının birleşiminden oluşan bu grafik türü, grafik okuma alışkanlığı olmayan kullanıcılar için ilk etapta karmaşık ve anlaşılması güç olabilir.

**Grafik Alanının Yoğunluğu:** Çok sayıda grup içeren çalışmalarda veya büyük veri setlerinde grafik üzerindeki öğeler kalabalık hale gelebilir. Bu durum, görsel sadeleşmeyi zorlaştırabilir ve yorumlamayı güçleştirebilir.

**Nokta Çakışması Problemi:** Veri noktalarının birbirine yakın veya aynı değerde olması durumunda üst üste binmeler yaşanabilir. Bu da ham veri görünürlüğünü azaltabilir ve yanıltıcı bir yoğunluk algısı oluşturabilir.

**Teknik Bilgi Gerektirmesi:** Raincloud grafikler genellikle R, Python veya benzeri programlama dilleri ile oluşturulduğundan, bu araçlara hâkim olmayan kullanıcılar için üretimi teknik destek gerektirebilir.

**Aşırı Detayın Karmaşa Yaratması:** Her bileşen veri hakkında bilgi sunsa da bazı kullanıcılar için bu ayrıntı fazlalığı kafa karışıklığına neden olabilir. Özellikle sade görselleştirme bekleyen izleyiciler için aşırı yoğun bir grafik yapısı oluşturabilir.

# 6. Dikkat Edilmesi Gerekenler

Raincloud grafiklerini kullanırken dikkat edilmesi gereken bazı önemli noktalar vardır. Öncelikle, gruplarda yeterli sayıda gözlem olmalıdır. Çünkü çok az veri varsa hem yoğunluk eğrisi düzgün çizilmez hem de noktalar eksik ya da anlamsız görünebilir. Grupların gözlem sayısı çok farklıysa da grafik yanıltıcı olabilir; örneğin bir grubun dağılımı daha geniş ya da daha yoğun gibi görünebilir. Ayrıca grafik içinde kullanılan renkler, saydamlık oranı ve katmanların birbirine karışmaması da önemli bir detaydır. Yoğunluk eğrisinin ayarları dağılımı fazla dalgalı ya da düz göstermemeli, veriyi en doğru şekilde yansıtmalıdır. Bunun dışında, grafikte yer alan nokta katmanı yani bireysel gözlemler uç değerleri gösterebilir. Ama bu noktalar bazen boxplot ve yoğunlukla çakışmayabilir, bu yüzden dikkatli yorumlanmalıdır. Bir diğer önemli konu da Raincloud grafiği sadece görsel bir araçtır; yani grafik üzerinde fark var gibi görünse bile bu farkın gerçekten anlamlı olup olmadığını anlamak için mutlaka istatistiksel testler yapılmalıdır. Grafik sadece destekleyici bir araçtır, tek başına yeterli değildir. Son olarak, bu grafik türü herkesin çok alışık olduğu bir grafik tipi olmadığından, raporda grafik öğelerinin ne anlama geldiği kısaca açıklanmalı ve okuyucunun neyi nasıl okuyacağı belirtilmelidir. Böylece grafik hem daha anlaşılır olur hem de çalışmanın güvenilirliği artar.

# 7. Raincloud Grafik

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

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 3. Hazırlık Kursu Gruplarına Göre Skorların Raincloud Grafikleri

Bu çalışmada öğrencilerin matematik, okuma ve yazma becerilerindeki başarı düzeyleri, hazırlık kursuna katılıp katılmama durumuna göre karşılaştırmalı olarak incelenmiş ve bu karşılaştırma hem grafiksel hem istatistiksel yöntemlerle desteklenmiştir. Kullanılan Raincloud grafikleri, her bir başarı puanının yoğunluk dağılımını, bireysel gözlem noktalarını ve boxplot yapısını birleştirerek oldukça bütün ve zengin bir görselleştirme sunmuştur. Bu yönüyle hem betimleyici hem de sezgisel açıdan son derece işlevsel bir grafik türü olduğu görülmüştür.

Matematik puanına göre kursa katılan öğrencilerin katılmayanlara göre genel olarak daha yüksek düzeyde olduğu görülmüştür. Kursa katılan grupta dağılım daha simetrik ve yoğunluk merkezde toplanmışken, katılmayan grupta ise daha yaygın ve uç değerlere sahiptir ayrıca bu grupta 30 puanın altındaki değerlerin daha yoğun olduğu, bu grubun içinde düşük puana sahip öğrencilerin olduğu anlaşılmaktadır. Noktalar arasındaki boşlukların kursa katılmayan grupta başarıda daha fazla bulunduğundan notlar arasında daha fazla çeşitlilik olduğunu göstermektedir.

Okuma puanına göre Kursu tamamlayan grup hem medyan hem ortalama açısından belirgin şekilde daha yüksek skorlara sahiptir. Kursu tamamlayan grupta yoğunluk 75-85 aralığında çok belirgindir. Katılmayan grupta 50-65 aralığında ciddi bir yoğunluk bulunmaktadır. Kursa katılan grubun dağılımı daha dengeli ve simetrikken, diğer grup sola çarpık bir dağılım göstermektedir; yani düşük skorlar daha yaygındır.

Yazma puanına göre kursa katılanlar ve katılmayanlar arasındaki fark en belirgin olanıdır. Kursa katılan öğrencilerde yoğunluk eğrisi daha dik ve merkeziyken, katılmayan grupta dağılım hem daha yaygın hem de daha aşağıda konumlanmıştır. Kursa katılan grupta öğrencilerin skorları 70-85 arasında yoğunlaşırken, katılmayan grupta daha geniş bir alana yayılmış ve medyan düşmüştür. Katılmayan grupta medyan bariz şekilde daha aşağıda olup, yazılı anlatım becerilerinin daha düşük bir seviyede kaldığını göstermektedir. Kursa katılmayan öğrencilerde düşük skorlar (40-55 arası) belirgin biçimde daha fazla sayıda ve yoğunluktadır.

Özellikle yazma ve okuma skorlarında, kursun etkisi hem grafiksel hem de istatistiksel olarak net biçimde gözlemlenmektedir. Matematikteki etki ise daha sınırlı, ancak yine de anlamlı bir fark barındırmaktadır. Kursa katılan gruplarda dağılımların daha simetrik ve merkezi yoğunluğa sahip olması, bu öğrenciler arasında başarı düzeylerinin daha tutarlı olduğunu; buna karşılık katılmayan grupta ise başarı düzeylerinin daha heterojen bir yapı sergilediğini göstermektedir.

## 7.1 Grup Karşılaştırma Testi

Tablo 5. Grupları Karşılaşma Skorları

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Değişken** | **Ortalama (Tamamlayanlar)** | **Ortalama (Katılmayanlar)** | **t-değeri** | **p-değeri** |
| **Matematik** | 69.55 | 61.92 | 3.76 | 0.0002 |
| **Okuma** | 73.95 | 64.73 | 4.53 | <0.0001 |
| **Yazma** | 74.00 | 62.58 | 5.72 | <0.000001 |

Öğrencilerin başarı puanları, hazırlık kursunu tamamlayıp tamamlamadıklarına göre karşılaştırılmıştır. Gruplar arasında ortalamaların anlamlı şekilde farklı olup olmadığını test etmek için, her bir tür için elde edilen puanlara göre ayrı ayrı iki örneklem t-testi uygulanmıştır.

**Matematik puanları** için, hazırlık kursunu tamamlayan öğrencilerin ortalama puanı hazırlık kursuna katılmayanlara göre anlamlı derecede daha yüksektir. Bu, hazırlık kursunun matematik başarısı üzerinde olumlu bir etkisi olduğunu göstermektedir.

**Okuma puanları** için de benzer şekilde çok anlamlı bir fark bulunmuştur. Kursu tamamlayan öğrencilerin ortalama okuma puanı 73.95 iken, katılmayanlarınki 64.73’tür.

**Yazma puanları,** en yüksek farkı göstermiştir. Kursa katılan grubun ortalaması 74.00 iken, katılmayan grup yalnızca 62,58 puan almıştır. Bu fark son derece anlamlıdır ve yazma becerileri üzerinde kursun etkisinin çok belirgin olduğunu ortaya koyar.

Bu sonuçlar, **hazırlık kursunun hem sayısal hem de sözel alanlarda öğrencilerin başarısını anlamlı biçimde artırdığını** göstermektedir. Özellikle okuma ve yazma becerilerindeki yüksek farklar, sözel alanlarda daha güçlü bir etkiyi işaret etmektedir. Matematikte de anlamlı bir artış gözlemlenmiş olmakla birlikte, etkisi görece daha azdır.

## 7.2 Kurs Durumuna Göre Puanlar için Korelasyon

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

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.metin, ekran görüntüsü, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 4. Kurs için Korelasyon

Öğrencilerin sınav hazırlık kursuna katılıp katılmama durumlarına göre sınavdan aldıkları puanlar arasında yapılan korelasyon analizleri sonucunda;

Hazırlık kursunu tamamlayan öğrenciler grubunda, **okuma** ve **yazma** puanları arasındaki korelasyon katsayısı oldukça yüksektir. Bu sonuç, bu iki becerinin birbirini oldukça yakından takip ettiğini ve paralel geliştiğini göstermektedir. Matematik puanının okuma ve yazma ile olan korelasyonları da güçlü düzeydedir, ancak bu ilişkiler okuma-yazma ilişkisine kıyasla biraz daha zayıftır. Bu durum, hazırlık kursunun özellikle sözel beceriler üzerinde etkili olduğunu, matematik başarısının ise daha bağımsız bir yapıda gelişebildiğini düşündürmektedir.

Hazırlık kursuna katılmayan öğrenciler grubunda ise, yine okuma ve yazma puanları arasındaki korelasyon katsayısı oldukça yüksek olmakla birlikte, matematik puanı ile diğer iki beceri arasındaki korelasyonlar da daha yüksektir. Bu grupta bir öğrencinin matematikte gösterdiği başarı düzeyi, aynı zamanda onun okuma ve yazma becerileriyle de daha doğrudan ilişkilidir.

Her iki grupta da başarı puanları arasında yüksek düzeyde pozitif korelasyonlar gözlemlenmiştir. Ancak kursa katılmayan öğrenci grubunda tüm korelasyonların daha güçlü olması dikkat çekicidir. Bu durum, bu gruptaki öğrencilerin test başarılarının birbirine daha bağımlı olarak geliştiğini, yani test türleri arasında daha tutarlı bir başarı gösterdiklerini düşündürmektedir.

Öte yandan, hazırlık kursunu tamamlayan grupta özellikle okuma ve yazma becerileri arasında çok güçlü bir ilişki bulunmakla birlikte, matematik becerisi bu ilişki ağının dışında biraz daha bağımsız kalmaktadır. Bu durum, kursun sözel becerilere ağırlık vermesi veya öğrencilerin bu alanda daha fazla gelişim göstermesiyle açıklanabilir.

# 8. Kümeleme Analizi ile Raincloud

Bu çalışmada, öğrencilerin matematik, okuma ve yazma başarı puanlarına göre ortak başarı profillerine sahip gruplara ayrılabilmesi amacıyla **K-means kümeleme yöntemi** uygulanmıştır. Kümeleme analizi, benzer özelliklere sahip bireyleri aynı grup içinde toplayarak, veri içerisindeki bilgileri daha açık ve yorumlanabilir hale getirmeyi hedefleyen bir öğrenme yöntemidir.

Analiz kapsamında k=3 olacak şekilde modelleme yapılmış ve öğrenciler başarı düzeylerine göre **düşük, orta ve yüksek başarı profilleri** şeklinde üç kümeye ayrılmıştır.

sanat içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 5. Kümeleme Analizi ile Raincloud

Bu bölümde öğrencilerin matematik, okuma ve yazma alanlarındaki başarı düzeyleri sadece hazırlık kursuna katılıp katılmadıklarına göre değil, aynı zamanda başarı puanlarına göre yapılan kümeleme analizine göre incelenmiştir.

Matematik puanlarında, yeşil renkle gösterilen en yüksek başarı grubu hem kursa katılanlarda hem katılmayanlarda diğer gruplara göre daha ortada ve yoğun bir dağılım göstermiştir. Ancak sarı renkle gösterilen en düşük başarı grubunun, özellikle kursa katılmayan öğrenciler arasında çok daha fazla olduğu dikkat çekmektedir. Bu grup daha çok düşük puanlarda ve medyanın altında yoğunlaşarak, başarıda belirgin bir düşüklüğü ortaya koymuştur. Mor renkteki orta başarı grubu her iki grupta da yaygın olsa da kursa katılanlarda daha dengeli dağılmıştır.

Okuma puanlarına bakıldığında, gruplar arasındaki fark daha net görülmektedir. Yüksek başarı grubu (yeşil), kursa katılan öğrencilerde açıkça daha fazladır. Kursa katılmayanlarda ise daha çok orta (mor) ve düşük (sarı) başarı grupları yer almaktadır. Bu durum, okuma becerilerinde hem başarı düzeyinin hem de gruplar arası farkın daha belirgin olduğunu göstermektedir. Düşük başarı grubunun yoğunluk eğrisinin kısa ve alçak yapısı, bu öğrencilerin okuma becerilerinin oldukça sınırlı olduğunu göstermektedir.

Yazma puanlarında ise kümeler arasındaki fark en belirgin şekilde ortaya çıkmıştır. Özellikle düşük başarı grubu, kursa katılmayanlarda hem en geniş alana yayılmış hem de en düşük ortalamaya sahiptir. Orta başarı grubunun eğrisi kursa katılanlarda daha yoğunken, katılmayanlarda daha dağınık ve belirsiz bir yapıdadır. Yüksek başarı grubunun yazma puanlarında gösterdiği yoğunluk ve dengeli dağılım, bu beceride hazırlık kursunun etkisinin yanı sıra genel başarı düzeyinin de yüksek olduğunu düşündürmektedir.

Genel olarak bakıldığında, başarı puanlarına göre yapılan kümeleme analizi, öğrencilerin sadece ortalama puanlarına değil, aynı zamanda başarı dağılım yapılarına göre de anlamlı gruplar oluşturduğunu göstermektedir. Özellikle düşük başarı grubunun, kursa katılmayan öğrenciler arasında daha çok görülmesi dikkat çeken bir sonuçtur.

# 9. Raincloud için Sonuç ve Değerlendirme

Bu çalışmada öğrencilerin matematik, okuma ve yazma alanlarındaki başarı durumları, hem hazırlık kursuna katılıp katılmadıklarına göre karşılaştırılmış hem de başarı düzeylerine göre yapılan kümeleme analizi ile desteklenmiştir. Başarı dağılımları, Raincloud grafikleri yardımıyla görsel olarak detaylı şekilde incelenmiş, bireysel puanlar ile yoğunluk eğrileri birlikte değerlendirilmiştir. Hazırlık kursuna katılan öğrencilerin özellikle okuma ve yazma alanlarında daha yüksek ve dengeli dağılımlara sahip oldukları, matematikte ise bu farkın biraz daha az ama yine de anlamlı olduğu görülmüştür. Katılmayan öğrencilerde ise düşük başarıların daha çok toplandığı ve puanların daha dağınık olduğu dikkat çekmiştir.

K-means algoritmasıyla yapılan kümeleme sonucunda öğrenciler düşük, orta ve yüksek başarı gruplarına ayrılmıştır. Bu gruplar başarı dağılımı açısından anlamlı bir ayrım göstermiştir. Özellikle düşük başarı grubunun daha çok kursa katılmayan öğrencilerden oluştuğu, yüksek başarı grubunun ise genelde kursa katılanlardan meydana geldiği görülmüştür. Kümeleme sonuçlarının Raincloud grafikleriyle birlikte yorumlanması, öğrencilerin başarılarının sadece ortalama puanlarla değil, dağılım ve yoğunluk yapılarıyla da anlaşılabileceğini göstermektedir.

Elde edilen bulgular, hazırlık kurslarının özellikle sözel beceriler üzerinde olumlu etkileri olduğunu ortaya koymaktadır. Bunun yanında öğrencilerin başarılarını sadece bireysel puanlarla değil, benzer başarı gösteren gruplar üzerinden değerlendirmek de önemli bir bakış açısı sunmaktadır. Bu yönüyle çalışma, veri görselleştirme ve kümeleme gibi tekniklerin birlikte kullanılabildiği bir örnek sunmakta; özellikle düşük başarı grubundaki öğrencilere yönelik özel desteklerin gerekliliğini açıkça göstermektedir.

# 10. Jitter (Titreme) Grafiği Nedir?

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

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 6. Jitter Grafiği Örneği

Jitter Plot**,** çizilen noktaların konumlarına küçük miktarda rastgele kaydırma ekleyerek çakışan veri noktalarının daha iyi bir görünümünü sağlayan bir Şerit Plot, [Nokta Dağıtım Plot](https://datavizcatalogue.com/blog/chart-snapshot-dot-distribution-plot/) varyasyonudur. Jitter Plot, veri noktalarının konumlarını hafifçe rastgele hale getirerek çakışma nedeniyle oluşan veri noktalarının bulanıklığını azaltmaya yardımcı olur ve veri dağılımının daha net bir görünümünü sağlar.

Jitter Plot'un işlevi, bir değer ekseni boyunca noktalar çizerek birden fazla kategorideki veri dağılımını görselleştirmektir. Her nokta tek bir veri noktasını veya bir sayımı temsil edebilir. Bir Jitter Plot yatay veya dikey olarak görüntülenebilir. Bazen küçük bir çizgi biçiminde bir medyan işaretçisi görüntülenir.

Bir Jitter Plot'taki noktalar, kategorileri birbirinden ayırmak veya ek bir değişkeni görselleştirmek için renk kodlaması kullanabilir. Noktaların opaklığını azaltmak, örtüşme nedeniyle hala belirsiz olabilecek veri noktalarının görünürlüğüne yardımcı olabilir.

Bir [Beeswarm Plot](https://datavizcatalogue.com/blog/chart-snapshot-beeswarm-plot/" \t "_blank) görsel olarak bir Jitter Plot'a çok benzer, ancak rastgele titreşen veri noktaları yerine, bir Beeswarm Plot herhangi bir örtüşmenin oluşmasını önlemek için noktaların konumunu kaydıracaktır. Bir Beeswarm Plot, eksendeki çizilen konumu korumaya çalıştığı için dağılım şekline daha fazla vurgu yapar ve bu da değerlerin yoğunlaştığı 'sürü'de 'çıkıntılara' yol açar.

# 11. Avantajları

**Veri Yoğunluğunu Görselleştirme İmkânı:** Jitter plot, özellikle çok sayıda gözlemin aynı kategori altında toplandığı durumlarda, bu noktaların üst üste binmesini engelleyerek dağılım örüntülerini görünür kılar. Böylece hangi kategoride hangi değer aralıklarında yoğunluk oluştuğu grafik üzerinden doğrudan izlenebilir. Bu özellik, özellikle çakışan verilerin ayrıştırılmasında etkilidir.

**Ham Verinin Birebir Temsili:** Her bir gözlemi ayrı bir nokta ile ifade eden jitter plot, kullanıcıya veri setinin tamamını doğrudan sunar. Bu yönüyle özet istatistiklerin gizleyebileceği detaylar – örneğin aykırı değerler, kümelenmeler ya da boşluklar – daha kolay fark edilir. Bu durum, örneklem yapısının daha derinlemesine anlaşılmasına katkı sağlar.

**Küçük ve Orta Ölçekli Verilerde Etkinlik:** Örneklem büyüklüğü az olduğunda geleneksel kutu grafikler ya da histogramlar sınırlı bilgi sunabilir. Ancak jitter plot, her gözlemi açıkça gösterebildiği için bu kısıtı ortadan kaldırır. Küçük örneklemlerde dağılımın tüm ayrıntıları, gözlem bazında incelenebilir hâle gelir.

**Çoklu Değişken Kodlamaya Uygunluk:** Jitter plot, renk (color), şekil (shape), büyüklük (size) gibi estetik özellikler kullanılarak ikinci ya da üçüncü bir değişkenin de grafik içine entegre edilmesine olanak tanır. Bu sayede aynı görselde hem kategoriye göre dağılım hem de başka bir değişkene göre gruplandırma yapılabilir. Özellikle cinsiyet, grup veya sınıf karşılaştırmalarında bu özellik analiz derinliğini artırır.

**Görsel Sadelik ve Etkililik:** Grafik yapısı itibarıyla sade ancak bilgilendirici olan jitter plot, kullanıcıyı görsel karmaşadan uzak tutarak veriye odaklanmayı kolaylaştırır. Özellikle sunumlarda veya raporlamalarda veriyi hem yalın hem de etkili şekilde aktarmak isteyen kullanıcılar için ideal bir tercihtir.

**Geleneksel Grafiklerle Kombine Edilebilirlik:** Jitter plot’lar, violin plot veya boxplot gibi özetleyici grafiklerle birlikte kullanıldığında hem genel eğilimleri hem de bireysel gözlemleri aynı anda sunabilir. Bu kombinasyon, analizi zenginleştirir ve karar vericiye daha kapsamlı bir bakış açısı kazandırır.

# 12. Dezavantajları

**Aşırı Yayılma Nedeniyle Kategorik Belirsizlik:** Jitter plot’ta veri noktalarının üst üste binmesini önlemek amacıyla uygulanan yatay (veya dikey) kaydırma işlemi, verilerin ait oldukları kategorilerden görsel olarak uzaklaşmasına neden olabilir. Özellikle width değeri yüksek ayarlandığında, gözlemler başka kategori sınırlarına taşmış gibi algılanabilir. Bu durum, grafik yorumlamasında karışıklıklara yol açabilir.

**Merkezi Eğilim ve Yayılım Bilgisi Eksikliği:** Jitter plot, her ne kadar ham verileri ayrıntılı biçimde sunsa da medyan, çeyrekler, minimum ve maksimum gibi merkezi özet istatistikleri doğrudan göstermez. Bu eksiklik, özellikle gruplar arası karşılaştırmalar yapılırken verinin yapısını anlamada sınırlayıcı olabilir. Bu nedenle genellikle ek olarak boxplot gibi grafiklerle desteklenmesi gerekir.

**Büyük Örneklemlerde Görsel Yoğunluk Oluşturma:** Gözlem sayısı arttıkça jitter plot üzerinde veri noktaları birbirine çok yakın konumlanabilir ve görsel kalabalık oluşabilir. Bu durum, hem aykırı değerlerin ayırt edilmesini zorlaştırır hem de kategoriler arası farkların algılanmasını güçleştirir. Böyle durumlarda nokta boyutunun küçültülmesi ya da şeffaflığın artırılması gibi görsel önlemler alınmalıdır.

**Estetik Ayarların Hassasiyeti:** Jitter plot’ta geom\_jitter() fonksiyonuna ait parametrelerin (örneğin width, alpha, size) yanlış seçilmesi, grafik çıktısının anlamsız veya yanıltıcı olmasına neden olabilir. Özellikle görsel dengeyi koruyacak uygun parametrelerin belirlenmesi deneyim ve dikkat gerektirir. Bu da kullanıcı açısından ek bir çaba anlamına gelir.

**Karmaşık Değişken Yapılarında Yetersizlik:** Jitter plot, özellikle çok sayıda kategoriye ve çoklu boyuta sahip veri setlerinde okunabilirliğini kaybedebilir. Birden fazla kategorik değişkenin veya karmaşık etkileşimlerin yer aldığı durumlarda grafik yapısı sade kalabilir ve verinin tamamını yansıtmakta yetersiz kalabilir.

**Yalnızca Gözlemsel Veriye Odaklanma:** Bu grafik türü yalnızca ham verilerin görselleştirilmesine odaklandığından, istatistiksel anlamlılık, güven aralıkları ya da modelleme çıktıları gibi ileri düzey analiz bileşenlerini yansıtamaz. Bu yönüyle daha açıklayıcı veya öngörücü analizlerde destekleyici rol üstlenmesi gerekir.

# 13. Dikkat Edilmesi Gerekenler

Jitter plot kullanımında bazı önemli teknik ve yorumlayıcı unsurların göz önünde bulundurulması gerekir. Öncelikle, gözlem sayısının uygun düzeyde olması büyük önem taşır. Çok az sayıda veri noktası varsa jitter plot gereksiz görsel dağınıklık yaratabilir ve anlamlı bir desen sunamaz. Benzer şekilde, gruplar arası gözlem sayısı eşitsiz olduğunda, görsel yoğunluk farklılıkları yanlış yorumlara neden olabilir; örneğin büyük örnekleme sahip bir grup, diğer gruplara göre daha baskın ya da yaygın görünebilir.

Grafikte kullanılan yayılma (jitter) genişliği dikkatle belirlenmelidir. width değeri çok düşükse noktalar üst üste biner ve dağılım okunamaz hâle gelir. Tersine, bu değer çok yüksek ayarlanırsa noktalar kategori sınırlarının dışına taşabilir ve görsel olarak kategoriye aitliklerini yitirebilirler. Bu nedenle jitter ayarları verinin yoğunluğuna göre titizlikle seçilmelidir.

Renk ve saydamlık (alpha) değerleri, grafiğin okunabilirliği açısından kritik öneme sahiptir. Renk seçimi, özellikle grupların ya da kategorilerin kolay ayırt edilebilmesini sağlamalı; çok benzer tonlardan kaçınılmalıdır. Saydamlık ise noktaların üst üste geldiği bölgelerde yoğunluğu görsel olarak ifade etme açısından önemlidir. Saydamlığı fazla azaltmak noktaları silikleştirirken, çok azaltmamak da yoğun bölgeleri karanlıklaştırabilir.

Jitter plot, merkezi eğilim ve yayılım ölçülerini doğrudan göstermediği için yalnız başına yorum yapılmamalıdır. Gerekli durumlarda boxplot gibi özet grafiklerle birlikte kullanılması, daha güvenilir ve kapsamlı bir analiz sunar. Ayrıca grafik üzerinden gözle görülür bazı desenler ya da farklılıklar dikkat çekici olabilir; ancak bu görsel gözlemler, istatistiksel testlerle desteklenmediği sürece anlamlı sonuçlar olarak kabul edilmemelidir.

# 14. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı İçin Jitter Grafik

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 7. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanları için Jitter Grafiği

Bu grafik, öğrencilerin yazma puanlarının ebeveynlerinin eğitim düzeyine ve öğrencinin cinsiyetine göre nasıl değiştiğini göstermektedir. Jitter plot kullanılarak, her öğrencinin puanı üst üste binmeden gösterilmiş ve böylece dağılım daha net görülmüştür. Cinsiyet ayrımı renklerle gösterilmiştir

Grafiğe genel olarak bakıldığında, kız öğrencilerin yazma puanlarının erkek öğrencilere göre daha yüksek olduğu dikkat çekmektedir. Özellikle yüksek puanlarda kırmızı noktaların daha yoğun olduğu, mavi noktaların ise daha çok alt puanlarda toplandığı görülmektedir. Bu da yazma becerilerinde cinsiyetin etkili olabileceğini düşündürmektedir.

**Ebeveyn Eğitim Düzeyine Göre Dağılım**

Lise ve altı: Bu gruplarda yazma puanları çok geniş bir aralığa yayılmıştır. Hem düşük hem yüksek puanlar vardır. Ancak erkek öğrencilerde 50 puanın altında birçok örnek bulunması, düşük ebeveyn eğitim düzeyinin özellikle erkek öğrencileri daha çok etkilediğini gösterebilir.

Üniversiteyi tamamlamamış ve Ön lisans mezunu: Bu gruplarda yazma puanları daha çok orta seviyede toplanmıştır. Kız öğrenciler genellikle 70 ve üzeri puan alırken, erkek öğrenciler daha çok 50–70 arasında yer almaktadır. Cinsiyet farkı burada da kendini göstermektedir.

Lisans mezunu: Bu grupta yazma puanları daha yüksek ve sıkı bir şekilde toplanmıştır. Kız öğrenciler çoğunlukla 75–100 arasında yoğunlaşırken, erkek öğrenciler 60–85 arasında dağılmıştır. Bu da eğitim seviyesi arttıkça başarı düzeyinin de yükseldiğini gösterir.

Yüksek lisans mezunu: Bu grupta örnek sayısı az olsa da kız öğrencilerin yazma puanları büyük oranda 80–100 arasında yoğunlaşmıştır. Erkek öğrenciler ise daha seyrek dağılmış ve genelde daha düşük puanlarda kalmıştır. Bu grup aynı zamanda yazma başarısının en yüksek olduğu seviyeyi temsil etmektedir.

**Cinsiyete Dayalı Farklılık**

Grafikte genel olarak kız öğrencilerin yazma becerilerinde erkek öğrencilerden daha başarılı olduğu görülmektedir. Bu fark, özellikle ebeveynin eğitim düzeyi yüksek olduğunda daha belirgin hâle gelmektedir. Erkek öğrencilerin puanları daha geniş bir aralığa yayılmıştır, bu da onların başarı düzeylerinin daha değişken olduğunu göstermektedir. Bu farklılık hem öğretim hem de öğrenci desteği açısından önemlidir.

Ebeveynin eğitim düzeyi arttıkça öğrencilerin yazma puanlarının da genel olarak yükseldiği görülmektedir. Bu artış, hem kız hem erkek öğrencilerde mevcuttur; ancak kız öğrenciler bu yükselişten daha fazla fayda sağlamış görünmektedir. Özellikle lisans ve yüksek lisans gruplarında puanların daha yüksek ve dar bir aralıkta toplandığı görülmektedir. Bu da ebeveyn eğitiminin, özellikle yazma becerileri üzerinde olumlu etkisi olduğunu ortaya koymaktadır.

## 14.1. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için ANOVA Sonuçları

Tablo 6. ANOVA Sonuçları

| **Faktör** | **Anlamlı mı?** | **Yorum** |
| --- | --- | --- |
| **Ebeveyn Eğitim Düzeyi** | Evet (p = 0.00546) | Ebeveynlerin eğitim düzeyi, çocukların yazma puanları üzerinde anlamlı bir fark yaratıyor. |
| **Cinsiyet** | Çok güçlü anlamlı (p ≈ 0) | Cinsiyet, yazma puanlarını etkileyen en güçlü faktörlerden biri. Erkek ve kadın arasında anlamlı fark var. |
| **Ebeveyn Eğitim Düzeyi \* Cinsiyet Etkileşimi** | Hayır (p = 0.81955) | Eğitim düzeyi ve cinsiyetin birlikte etkisi anlamlı değil; yani cinsiyet farkı eğitim düzeyine göre değişmiyor. |

## 14.2. Tukey Çoklu Karşılaştırma Testi

Genel olarak ebeveyn eğitim düzeyleri arasında yazma puanları açısından çoğu grup arasında anlamlı fark yok. Hiçbir karşılaştırmanın p-değeri < 0.05 değil, yani istatistiksel olarak anlamlı bir fark çıkmamış. Ancak bazı gruplarda fark anlamlılığa oldukça yakın ve eğilimsel farklar gösteriyor.

Erkekler ve kadınlar arasında ortalama yazma puanlarında anlamlı ve güçlü bir fark var. Erkeklerin ortalama puanı kadınlardan yaklaşık 9 puan daha düşük (p ≈ 0.0000088).

Ebeveyn eğitim düzeyi ile cinsiyetin birlikte etkisi anlamlı bulunmadı (ANOVA sonuçlarıyla uyumlu). Tukey testi ile yapılan alt grup karşılaştırmalarında da çoğu etkileşim farkı anlamsız çıktı. Sadece bazı özel alt küçük anlamlılıklar görülmüş.

# 15. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için Kümeleme

metin, çizgi, diyagram, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 8. Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için Kümeleme

Bu grafikte öğrencilerin yazma puanları, ebeveyn eğitim düzeyine göre belirlenmiş üç eğitim grubuna göre kümelenmiş ve cinsiyete göre incelenmiştir. Bu öğrencilerin ebeveyn eğitim düzeylerine göre oluşturulan gruplar, daha anlamlı karşılaştırmalar yapılabilmesi için üç ana kümeye ayrılmıştır. İlk olarak, her bir ebeveyn eğitim düzeyi için öğrencilerin matematik, okuma ve yazma puanlarının ortalamaları hesaplanmıştır. Elde edilen ortalama değerler, eğitim düzeylerinin akademik başarıyla olan ilişkisini incelemek adına bir temel oluşturmuştur. Daha sonra bu ortalamalar dikkate alınarak benzer başarı seviyelerine sahip eğitim grupları birleştirilmiş ve toplamda üç ana eğitim kümesi (1: düşük, 2: orta, 3: yüksek) tanımlanmıştır.

Genel olarak grafik incelendiğinde, tüm gruplarda kız öğrencilerin yazma puanlarının erkek öğrencilere göre daha yüksek olduğu açıkça görülmektedir. Özellikle küme 3’te (ebeveyn eğitimi en yüksek grup: lisans ve yüksek lisans), kız öğrenciler hem sayıca daha fazla hem de puan olarak daha üst seviyelerde yer almaktadır. Bu grupta kız öğrenciler genellikle 75 ve üzeri puanlara sahipken, erkek öğrenciler daha düşük puanlara dağılmıştır. Bu durum, yüksek eğitim düzeyine sahip ailelerden gelen kız öğrencilerin yazma becerilerinde daha başarılı olduklarını göstermektedir.

Küme 2’de (orta düzey eğitim: ön lisans ve üniversite eğitimini tamamlamamış), benzer bir durum devam etmektedir. Kız öğrenciler genel olarak daha yüksek puanlar alırken, erkek öğrencilerin puanları daha geniş bir aralığa dağılmıştır. Ayrıca, ortalamanın altında kalan erkek öğrenci sayısı da daha fazladır. Küme 1’de (en düşük eğitim seviyesi: lise ve liseyi tamamlamamış), genel olarak her iki cinsiyette de puanlar daha düşüktür. Ancak burada da kız öğrenciler erkek öğrencilere göre daha yüksek puanlara sahip görünmektedir. Erkek öğrenciler ise hem daha kalabalık hem de puan olarak daha düşük seviyededir. Bu dağılım, yazma becerilerinde hem ebeveyn eğitim düzeyinin hem de cinsiyetin önemli etkileri olduğunu düşündürmektedir. Özellikle kız öğrencilerin, ebeveyn eğitimi yükseldikçe başarılarını daha çok artırdığı görülmektedir. Erkek öğrenciler ise her grupta daha düşük ortalama puanlara sahip oldukları için bu alanda daha fazla desteğe ihtiyaç duyabilir.

Sonuç olarak grafik, cinsiyet ile ebeveyn eğitimi etkileşiminin yazma başarısı üzerinde anlamlı farklar oluşturduğunu göstermektedir. Bu bulgular, özellikle düşük eğitim düzeyine sahip ailelerden gelen erkek öğrenciler için ek destek ve eğitim uygulamalarının önemli olabileceğini ortaya koymaktadır.

## 15.1. Kümeleme ile Oluşturulmuş Ebeveyn Eğitim Düzeyi ve Cinsiyete Göre Yazma Puanı için ANOVA Sonuçları

Tablo 7. ANOVA Sonuçları

| **Faktör** | **Anlamlı mı?** | **Yorum** |
| --- | --- | --- |
| **Ebeveyn Eğitim Düzeyi** | Evet (p = 0.000227) | Ebeveynlerin eğitim düzeyi, çocukların yazma puanları üzerinde anlamlı bir fark yaratıyor. |
| **Cinsiyet** | Çok güçlü anlamlı (p ≈ 0) | Cinsiyet, yazma puanlarını etkileyen en güçlü faktörlerden biri. Erkek ve kadın arasında anlamlı fark var. |
| **Ebeveyn Eğitim Düzeyi \* Cinsiyet Etkileşimi** | Hayır (p = 0.503) | Eğitim düzeyi ve cinsiyetin birlikte etkisi anlamlı değil; yani cinsiyet farkı eğitim düzeyine göre değişmiyor. |

Ebeveyn eğitim düzeyi ile kümeleme oluşturulmuş ebeveyn eğitim düzeyi arasındaki anova farkına göre anlam düzeylerinin daha kuvvetlendiği görülmüştür. Ebeveyn Eğitim Düzeyi \* Cinsiyet Etkileşimi ise red olsa bile öncesine göre p değerine diğerine göre daha çok yaklaştığı görülmektedir.

## 15.2 Tukey Çoklu Karşılaştırma Testi

Tablo 8. Tukey Testi Sonuçları

| **Karşılaştırma** | **Fark** | **Güven Aralığı** | **p-değeri** |
| --- | --- | --- | --- |
| Küme 2- Küme 1 | 5.84 | 0.60 – 11.08 | 0.025 |
| Küme 3- Küme 1 | 11.73 | 4.97 – 18.50 | 0.00018 |
| Küme 3- Küme 2 | 5.89 | -0.47 – 12.26 | 0.076 |

Gerçekleştirilen kümeleme analizi sonrasında yürütülen Tukey post hoc testi sonuçlarına göre, ebeveyn eğitim düzeyi ve cinsiyetin öğrencilerin yazma puanları üzerinde anlamlı etkiler yarattığı görülmüştür. İlk olarak, ebeveyn eğitim düzeyine göre oluşturulan üç kümeye ait karşılaştırmalarda; 1. küme (düşük eğitim düzeyi) ile 2. küme (orta düzey) ve özellikle 3. küme (yüksek eğitim düzeyi) arasında anlamlı farklar bulunmuştur. 1. ve 3. kümeler arasındaki fark (p < 0.001) oldukça belirgin olup, ebeveynlerinin daha yüksek eğitim düzeyine sahip olduğu öğrencilerin yazma puanlarının anlamlı şekilde daha yüksek olduğunu ortaya koymaktadır. Ayrıca, 2. ve 3. küme arasındaki fark anlamlılık sınırında yer almakta (p ≈ 0.075) ve bu da eğitimin belirli bir düzeyin üstüne çıktığında yazma başarısına olan katkının artabileceğini düşündürmektedir.

**Anlamlı Farklılıklar**

Tablo 9. Farklılık Açıklamaları

| **Karşılaştırma** | **p-değeri** | **Yorum** |
| --- | --- | --- |
| 1:male - 2:female | **0.0000917** | Küme 1’deki erkek öğrenciler, Küme 2’deki kızlara göre çok daha düşük puana sahip. |
| 2:male - 2:female | **0.0009650** | Küme 2’deki erkek öğrenciler de aynı kümedeki kızlardan anlamlı şekilde düşük. |
| 1:male - 3:female | **0.0000300** | En düşük grup: Küme 1 erkekleri. Küme 3’teki kızlarla fark çok büyük ve anlamlı. |
| 2:male - 3:female | **0.0002300** | Küme 2 erkekleri, Küme 3 kızlarından da anlamlı biçimde düşük puan almış. |

Cinsiyet değişkeni incelendiğinde ise, erkek öğrencilerin yazma puanlarının, kız öğrencilere göre anlamlı düzeyde daha düşük olduğu görülmektedir. Bu sonuç, cinsiyetin yazma becerisi üzerinde güçlü ve istatistiksel olarak anlamlı bir etkisi olduğunu ortaya koymaktadır. Etkileşim düzeyinde, yani hem cinsiyet hem de ebeveyn eğitim düzeyinin birlikte ele alındığı analizde ise en dikkat çekici farklar, düşük ve orta düzey eğitim kümesinde yer alan erkek öğrencilerin, aynı kümelerdeki kız öğrencilere göre anlamlı şekilde daha düşük yazma puanlarına sahip olmasıyla ortaya çıkmıştır. Özellikle 1. kümede yer alan erkek öğrenciler ile 2. ve 3. kümedeki kız öğrenciler arasında çok güçlü istatistiksel farklar bulunmuş; bu da düşük eğitim düzeyine sahip ebeveynlerden gelen erkek öğrencilerin yazma becerilerinde ciddi bir dezavantaj yaşadığını göstermektedir.

# 16. Jitter için Sonuç ve Değerlendirme

Kümeleme yapılmadan önce eğitim düzeyleri 6 kategori halinde yer almakta olup bu durum, gruplar arası farkların istatistiksel olarak zayıflamasına neden olabilmekteydi. Ancak kümelendirme ile eğitim düzeyleri anlamlı bir şekilde gruplanarak analiz gücü artırılmış, yazma puanları üzerindeki etkiler daha sade ve anlamlı biçimde ortaya konmuştur. Bu sayede özellikle belirli eğitim düzeylerinin öğrenciler üzerindeki etkisi net bir şekilde izlenebilmiş, cinsiyetle etkileşimi daha güçlü biçimde analiz edilebilmiştir. Grafikler incelendiğinde de bu bulgular desteklenmektedir: ebeveyn eğitim düzeylerinin tamamında kız öğrencilerin yazma puanlarının, erkek öğrencilere kıyasla daha yüksek olduğu dikkat çekmektedir. Özellikle orta ve yüksek başarı seviyelerinde kız öğrencilerin daha yoğun şekilde yer aldığı, buna karşılık erkek öğrencilerin daha çok düşük skor aralıklarında kümelendiği görülmektedir. Bu durum, yazma becerilerinde cinsiyetin önemli bir belirleyici olduğunu ve düşük eğitim düzeyindeki erkek öğrencilerin akademik anlamda daha kırılgan bir yapıya sahip olabileceğini düşündürmektedir.

# 17. Genel Sonuç ve Değerlendirme

Bu çalışmada, öğrencilerin matematik, okuma ve yazma alanlarındaki başarı düzeyleri hem hazırlık kursuna katılım durumuna göre karşılaştırıldı hem de başarılarına göre gruplandırılarak değerlendirildi. Hazırlık kursuna katılan öğrencilerin özellikle okuma ve yazma alanlarında daha yüksek puanlara sahip oldukları, matematik alanında ise farkın daha az ama yine de anlamlı olduğu görüldü. Kursa katılmayan öğrencilerin puanlarının daha dağınık olduğu ve düşük puanların daha çok bu grupta toplandığı gözlemlendi.

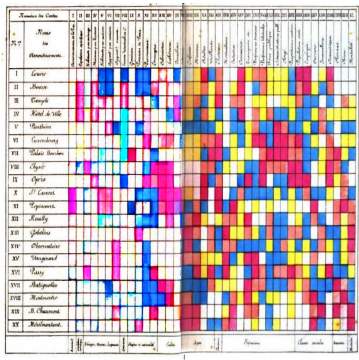
Raincloud grafiklerinin kullanımı ile her bir başarı puanının dağılımı ve öğrenci sayısı aynı anda görülebildi. Bu da verilerin daha iyi anlaşılmasını sağladı. Özellikle yazma puanlarında kursa katılan grubun daha dengeli, katılmayan grubun ise düşük puanlarda yoğunlaştığı görüldü.

K-means kümeleme yöntemiyle öğrenciler üç gruba ayrıldı: düşük, orta ve yüksek başarı grubu. Düşük başarı grubunun genelde kursa katılmayan öğrencilerden, yüksek başarı grubunun ise daha çok kursa katılan öğrencilerden oluştuğu tespit edildi. Böylece sadece ortalamalara bakmadan, öğrencilerin başarı seviyeleri daha net bir şekilde görülebildi.

Ayrıca cinsiyet ve ebeveyn eğitim düzeyinin özellikle yazma puanları üzerinde etkili olduğu anlaşıldı. Tüm eğitim düzeylerinde kız öğrencilerin erkek öğrencilere göre daha yüksek yazma puanları aldığı, özellikle düşük eğitim seviyesine sahip ailelerin erkek çocuklarında başarının daha düşük olduğu görüldü.

Sonuç olarak, bu çalışma hem kurs katılımının hem Ebeveyn eğitim durumuna göre öğrencilerin akademik durumlarıyla nasıl ilişkilendiğini ortaya koymuştur. Hazırlık kurslarının katkısı açıkça görülmüş, bazı öğrenci gruplarının daha fazla desteklenmesi gerektiği anlaşılmıştır.

# Isı Haritası (Heatmaps) Nedir?



Şekil 9. Toussaint Loua'dan (1873) Gölgeli Matris Gösterimi

Isı haritaları, ilgi duyulan temel bir değişkenin değerlerini, iki eksenli bir düzlemde renkli kareler (hücreler) aracılığıyla görselleştiren veri gösterim araçlarıdır. Her eksen, genellikle kategorik ya da sayısal olarak aralıklara ayrılmış iki değişkeni temsil eder; her hücrenin rengi ise bu eksenlerin kesişiminde kalan alandaki ana değişkenin değerine karşılık gelir. Isı haritaları farklı renk şemalarıyla oluşturulabilir. Gri tonlamalı haritalarda koyu renkler yüksek değerleri, açık renkler ise düşük değerleri ifade eder. Ancak insan gözü renkleri gri tonlarına göre daha kolay ayırt edebildiği için genellikle gökkuşağı renk skalası tercih edilir. Bu tür haritalarda kırmızı ve turuncu gibi sıcak renkler yüksek yoğunluklu bölgeleri, mavi ve mor gibi soğuk renkler ise düşük yoğunluklu bölgeleri temsil eder. Isı haritalarının kökeni, 19. yüzyılın sonlarında Toussaint Loua’nın Paris genelindeki sosyo-demografik değişimleri görselleştirmek için gölgelendirme haritaları kullanmasına dayanır. Modern anlamda bilgisayar destekli ısı haritalama teknolojisi ise 1990’lı yılların başında, yazılım tasarımcısı Cormac Kinney tarafından finansal piyasa verilerinin gerçek zamanlı görselleştirilmesi amacıyla geliştirilmiş ve ticari markaya dönüştürülmüştür. Günümüzde, ısı haritaları genellikle özel yazılımlar kullanılarak otomatik biçimde oluşturulmakta ve büyük veri kümelerinin analizinde yaygın olarak tercih edilmektedir. Her eksende yer alan değişkenler ister kategorik etiketler ister sayısal değerler olsun, herhangi bir türde olabilir. Sayısal değişkenler söz konusu olduğunda, bu değerler bir histogramdaki gibi aralıklara bölünerek, ilgi duyulan ana değişkenle ilişkili renklerin yerleştirileceği ızgara hücreleri oluşturulur. Hücrelerin renklendirilmesi, her bölmedeki gözlem sayısını (frekans), üçüncü bir değişkenin ortalama veya medyanı gibi özet istatistikleri temel alabilir. Bu bağlamda, bir ısı haritası; hücrelerin üzerinde renk kodlaması bulunan bir tablo veya matris gibi düşünülebilir. Bazı uygulamalarda ise sayısal olmayan niteliksel değerler (örneğin “düşük”, “orta”, “yüksek”) temel alınarak renklendirme yapılması da mümkündür. Bu tür grafik kullanımında kullanılan renk skalası önemlidir.

## Isı Haritası Çeşitleri

**Tıklama Isı Haritası:** Kullanıcıların bir web sayfasında en çok tıkladığı alanları gösterir. Hangi öğelerin ilgi çektiğini ve kullanıcı etkileşimini anlamaya yardımcı olur.

**Kaydırma Isı Haritası:** Ziyaretçilerin bir sayfayı ne kadar aşağı kaydırdığını gösterir. Sayfanın hangi bölümlerinin dikkat çekici olduğunu analiz eder.

**Fare Hareketi Isı Haritası:** Kullanıcıların imleci nasıl hareket ettirdiğini gösterir. Sayfa üzerindeki dikkat dağılımını ve etkileşim alanlarını görselleştirir.

**Göz İzleme Isı Haritası:** Kullanıcıların nereye baktığını göz hareketleriyle gösterir. Görsel dikkat ve bilgi işleme süreçlerini analiz eder.

**Dönüşüm Isı Haritası:** Kullanıcıların bir hedefe (ör. satın alma, form doldurma) ulaşırken izlediği adımları gösterir. Dönüşüm sürecindeki aksaklıkları belirlemek için kullanılır.

**Dikkat Isı Haritası:** Tıklama, gezinme ve diğer etkileşim verilerini birleştirerek, sayfada en çok dikkat çeken alanları bütüncül olarak gösterir.

**Mobil ve Masaüstü Isı Haritaları:** Mobilde dikey, masaüstünde hem dikey hem yatay kaydırma desenlerini gösterir. Farklı cihazlardaki kullanıcı davranışlarını karşılaştırmak için kullanılır.

Bunların dışında bazı çeşitleri Tablo 1’de kısaca açıklanmıştır.

Tablo 10. Isı Haritası Çeşitleri

|  |  |  |
| --- | --- | --- |
| ***TÜR*** | ***AÇIKLAMA*** | ***UYGULAMALAR*** |
| **Temel Matris Isı Haritası** | Değerleri renk kodlamasıyla 2 boyutlu bir ızgarada görüntüler. Kümeleme veya hiyerarşi olmadan basit yapıda oluşturur. | Gen ifadesi, korelasyon matrisleri, basit istatistikler gibi alanlarda kullanılır. |
| **Kümelenmiş Isı Haritası** | Hiyerarşik kümeleme ve dendrogramlar kullanarak benzer satırları/sütunları gruplayarak matris ısı haritalarını genişletir. | Genomik, müşteri segmentasyonu, biyolojik ağlar alanlarda kullanılır. |
| **Açıklamalı Isı Haritası** | Her hücre, renk gradyanıyla birlikte tam değeri göstermek için sayısal bir etiket içerir. | Pano raporlaması, iş zekâsı, sağlık verilerinde kullanılır. |
| **Coğrafi Isı Haritası** | Genellikle bölge veya konum koordinatlarına göre haritalara yoğunluk değerlerinin yerleştirilir. | Suç haritalama, seçim sonuçları, nüfus çalışmalarında kullanılır. |
| **Nokta Yoğunluk Isı Haritası** | Yoğunluğu nokta olaylarına göre görüntüler (bölgelere bağlı kalmadan). | Epidemiyoloji, trafik analizi, sıcak nokta tespitinde kullanılır. |
| **Etkileşimli Isı Haritası** | Yakınlaştırma, filtre, değerler için araç ipuçları gibi etkileşimli dijital haritalar. | Web kullanılabilirliği, pazarlama analizi, davranış takibinde kullanılır. |
| **3D/Hacimsel Isı Haritası** | Değerleri üç boyutlu veya hacim taramalarında (örneğin tıbbi görüntüleme) görselleştirir. | MRI analizi, nörobilim, BT tarama analizinde kullanılır. |
| **Zamansal (Takvim) Isı Haritası** | Zaman bloklarına göre düzenlenmiş ısı haritaları (takvim, haftalık, günlük desenler) | Aktivite görselleştirme, GitHub commit'leri, satış takibinde kullanılr. |
| **Çok Değişkenli Isı Haritası** | Her hücrede renk, boyut, şekil veya açıklama kullanarak birden fazla değişkeni kodlar. | Yapay zekâ yorumlanabilirliği, genomik, veri panolarında kullanılır. |
| **Korelogram** | Genellikle renk yoğunluğu ve işaretlerle çiftler halinde korelasyonları gösteren simetrik bir matris. | Ekonomi, psikoloji, eğitim araştırmasında kullanılır. |

## **Avantajları**

**Kullanıcı Davranışını Hızla Anlama:** Isı haritaları, kullanıcıların bir web sayfasında en çok ilgilendiği alanları renklerle gösterir. Özellikle dijital pazarlama ve web tasarım alanlarında kullanıcıların fare hareketleri, tıklamaları ve kaydırma alışkanlıkları analiz edilerek sayfa düzeni iyileştirilebilir.

**Veri Görselleştirme Kolaylığı:** Karmaşık veri setleri, ısı haritaları ile anlaşılır ve görsel olarak çekici bir hale getirilir. Böylece, büyük miktardaki veri kolayca analiz edilebilir hale gelir ve öne çıkan yoğunluklar hızla fark edilebilir.

**Pazarlama Stratejilerini Geliştirme:** Pazarlama ve satış stratejilerinde, kullanıcıların hangi ürünlere veya içeriklere daha fazla ilgi gösterdiğini anlamak için ısı haritaları kullanılır. Böylece, daha çok ilgi çeken ürünler ön plana çıkarılarak pazarlama kampanyalarının etkinliği artırılabilir.

**Daha Etkili İçerik Yerleşimi**: Web sitelerinde hangi içeriklerin daha çok dikkat çektiğini veya hangi bölümlerin daha az ilgi gördüğünü görmek, içerik yerleşimini optimize etmek için önemli bir veridir. Isı haritaları, önemli içeriklerin daha fazla ilgi çekecek şekilde konumlandırılmasına yardımcı olur.

**Karar Alma Süreçlerine Katkı:** Isı haritaları, kullanıcı davranışlarını anlamaya yönelik hızlı ve doğru bilgi sağlar. Bu sayede, veri analizine dayalı kararlar alarak işletmenin genel stratejileri daha etkili bir şekilde yönlendirilir.

**Kullanıcı Deneyimini İyileştirme:** Kullanıcıların web sitesindeki davranışları izlenerek sayfa düzeni, içerik akışı ve buton yerleşimleri geliştirilebilir. Böylece, ziyaretçiler için daha iyi bir deneyim sunulur ve sitenin genel performansı artar.

**Kaynakların Verimli Kullanımı:** Isı haritaları, web sitesi veya uygulama geliştirme süreçlerinde hangi alanlara odaklanılması gerektiğini belirler. Yoğun ilgi gören alanlar üzerinde iyileştirmeler yapılarak kaynaklar daha verimli bir şekilde kullanılabilir.

**Coğrafi Analizlerde Etkinlik Sağlama:** Isı haritaları, coğrafi veri analizlerinde de kullanılarak belirli bölgelerdeki yoğunlukları gösterir. Bu sayede, pazarlama veya lojistik stratejilerinde hangi bölgelere öncelik verilmesi gerektiği anlaşılır.

**Hızlı ve Kolay Analiz İmkânı:** Renkli görselleştirmeler sayesinde veriler hızlı bir şekilde analiz eder. Özellikle yoğun veri kümeleri için ısı haritaları, verileri görselleştirerek zaman kazandırır ve karar süreçlerini hızlandırır.

## Dezavantajları

**Veri Yorumlama Zorlukları**: Isı haritaları, kullanıcı davranışlarını görselleştirirken, bu verilerin doğru yorumlanması uzmanlık gerektirir. Yanlış yorumlamalar, hatalı kararlar alınmasına neden olabilir.

**Teknolojik Sınırlamalar**: Isı haritaları, kullanıcıların fare hareketleri ve tıklamaları gibi etkileşimlerini izlerken, mobil cihazlarda bu verilerin toplanması daha sınırlı olabilir.

**Renk Körlüğü ve Erişilebilirlik Sorunları**: Isı haritalarında kullanılan renk paletleri, renk körlüğü olan kullanıcılar için erişilebilirlik sorunları yaratabilir. Bu nedenle, renk seçiminde dikkatli olunmalıdır.

**Veri Gizliliği Endişeleri**: Kullanıcı davranışlarını izlemek, veri gizliliği ve etik konularında endişelere yol açabilir. Kullanıcıların izni olmadan veri toplamak, yasal sorunlara neden olabilir.

# Analiz Yapılacak Veri

Bu veri seti, Türkiye İstatistik Kurumu (TÜİK) tarafından yayımlanan 2024 yılı İşgücü İstatistikleri temel alınarak oluşturulmuştur. Veri setinde, 15-64 yaş aralığındaki nüfusun eğitim düzeyine göre işgücüne katılma sayıları, İBBS Düzey 2 (NUTS-2) 26 tane bölge bazında sunulmaktadır. Değişkenler sırasıyla,

F1: 15-64 yaş okuma yazma bilmeyenler, F3: 15-64 yaş lise ve dengi meslek okulu,

F2: 15-64 yaş lise altı eğitimliler, F4: 15-64 yaş yüksek öğretim olarak ayarlanmıştır.

Bu veri seti, eğitim düzeyinin bölgesel işgücüne katılım üzerindeki etkilerini analiz etmek amacıyla kullanılmış olup; ısı haritaları (heatmaps) aracılığıyla, farklı bölgelerdeki işgücüne katılım yoğunluklarının görsel olarak karşılaştırılmasına olanak tanımaktadır.

## 19.1. Açıklayıcı Analizler

Tablo 11. Açıklayıcı Analizler

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** | **Ortalama** | **SD** | **Medyan** | **Min** | **Maks** | **Aralık** | **Çarpıklık** | **Basıklık** | **Standart Hata** |
| **F1** | 26 | 24.42 | 18.83 | 20 | 5 | 85 | 80 | 1.57 | 2.06 | 3.69 |
| **F2** | 26 | 543 | 439.04 | 493 | 155 | 2476 | 2321 | 3.17 | 11.55 | 86.1 |
| **F3** | 26 | 359.88 | 375.03 | 260.5 | 69 | 2004 | 1935 | 3.21 | 11.31 | 73.55 |
| **F4** | 26 | 402.38 | 508.65 | 271 | 64 | 2639 | 2575 | 3.3 | 11.52 | 99.75 |

Eğitim düzeyi arttıkça, işgücüne katılımın hem ortalama düzeyinin hem de bölgesel farklılıklarının ciddi şekilde arttığı gözlemlenmektedir. Ortalamalara bakılarak okuma yazma bilmeyen (F1) kişi sayısının işgücüne katılımı diğerlerine göre az olduğu görülmektedir. Lise altı eğitimlilerin (F2) işgücüne katılma oranı oldukça yüksektir. Lise altı, lise dengi ve meslek okulu mezunu, yüksek öğretim mezunu için standart sapma, çarpıklık-basıklık katsayıları yüksek bu da bölgeler arasında eşitsizlik olduğunun göstergesidir. Özellikle yüksek öğretim mezunu için en büyük aralığa ve çarpıklık-basıklık katsayısına sahip olması çok uç değerler bulunduğu ve bölgesel uçurumlar olduğunu ortaya koyar. Türkiye’de eğitim düzeyi yükseldikçe işgücüne katılım sayısı da artmakta; ancak bu artış bölgesel eşitsizlikleri de beraberinde getirmektedir. Özellikle yükseköğrenim düzeyinde bazı bölgelerde büyük yoğunluklar, bazılarında ise çok düşük değerler görülmektedir. Bu durum, eğitim politikalarının ve bölgesel kalkınma stratejilerinin önemini vurgulamaktadır. Kümeleme yapılırken kesinlikle dengeli bir sonuç için Z normalizasyonu yapılmalıdır.

## Korelasyon

ekran görüntüsü, daire, renklilik, diyagram içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 10. Korelasyon Matrisi

Tablo 12. İlişki Tablosu

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **F1** | **F2** | **F3** | **F4** |
| **F1** | 1 | 0.29 | 0.172 | 0.157 |
| **F2** | 0.29 | 1 | 0.976 | 0.961 |
| **F3** | 0.172 | 0.976 | 1 | 0.993 |
| **F4** | 0.157 | 0.961 | 0.993 | 1 |

Okuma yazma bilmeyen gruba bakıldığında diğer gruplardan farklı bir dağılım yapısına sahip olduğu görülmektedir. Bu da eğitim düzeyi artıkça işgücüne katılanların sayılarının birbirinden bağımsız olduğunu gösterir. Diğer eğitim düzeyleri arasında yüksek korelasyon bulunmaktadır. Yani, bu düzeylerin işgücüne katılım sayıları birlikte artıp azalır. Bölgesel dağılımın çok benzer olduğunu gösterir.

## 19.2. Normallik Testi

Tablo 13. Shapiro-Wilk Testi Sonuçları

|  |  |  |
| --- | --- | --- |
|  | **W** | **P Değeri** |
| **F1** | 0.81 | 0.00027 |
| **F2** | 0.61 | 4.01e-07 |
| **F3** | 0.5993 | 2.902e-07 |
| **F4** | 0.56 | 1.118e-07 |

Yokluk hipotezi değişkenlerin dağılımı ile normal dağılım arasında fark yoktur olarak kurulduğunda p değerlerinin tümü 0.05’ten küçük olduğundan yokluk hipotezi reddedilir. Yani değişkenlerin dağılımı normal dağılıma uymamaktadır. En iyi uyum F1'dedir (W=0.81), ama bu da istatistiksel olarak normal değildir. Böylece kümeleme için Z normalizasyonu yapılması gerektiği bir kere daha kanıtlanmış oldu.

# Grafik Yorumları

metin, diyagram, harita içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

metin, diyagram, harita içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 11. Her Değişken için Türkiye Haritası Üzerinde Isı Haritaları

Genel olarak her choropleth grafiği dikkatlice incelendiğinde yükseköğretim mezunları (F4) büyük ölçüde İstanbul ve Ankara gibi metropollerde yoğunlaşmıştır. Bu durum, yüksek eğitimli bireylerin büyük şehirlerde istihdam olanaklarının daha fazla olduğunu gösterebilir. Lise altı ve dengi eğitim grupları ise daha dengeli bir yayılım göstermektedir. Okuma yazma bilmeyen grup ise Güneydoğu ve Doğu Anadolu’da sıklıkla bulunmaktadır. Bölgesel eğitim eşitsizliklerinin halen sürdüğüne işaret eder. İstanbul ise genel olarak her seviyede yüksek sayıda işgücü barındırdığı görülür. Sebep olarak nüfus büyüklüğü ve iş olanakları çeşitliliği gösterilebilir.

Grafikler RStudio’da ggplot2 paketi kullanılarak oluşturulmuş olup interaktif grafik oluşumu için plotly ve leaflet paketleri kullanılmıştır. Eğitim düzeyine göre bölgelerdeki işgücü katılımı, Şekil 3’teki gibi interaktif bir choropleth grafiği olarak bu bağlantı üzerinden görüntülenebilir. (<https://drive.google.com/drive/folders/1eaDqctXTnb5rKS8cbpFJySQU3Fs-JXhv?usp=drive_link>)

metin, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 12. Ana Veri için Isı Haritası Örneği (Kümeleme ve Z Normalizasyonu Olmadan)

Türkiye'nin İBBS-2 düzeyindeki 26 bölgesinde 15–64 yaş arası nüfusun dört farklı eğitim düzeyine göre (F1: Okuma yazma bilmeyen, F2: Lise altı, F3: Lise ve dengi, F4: Yükseköğretim) işgücüne katılım sayılarını ham değerlerle göstermektedir. Renk skalası, kişi sayılarına göre belirlenmiş olup, sarı tonlar yüksek değerleri, mor tonlar ise düşük değerleri temsil etmektedir. İstanbul F2, F3 ve F4 kategorilerinde en yüksek değerleri göstermektedir. Diğer bölgeler ise göreceli olarak işgücüne katılım azdır. Bu görselleştirme, işgücü katılım sayılarını olduğu gibi yansıtsa da z-normalizasyon ve kümeleme işlemleri uygulanmadığı için çeşitli metodolojik sınırlılıklar içermektedir. Ölçek uyumsuzluğunu, eğitim düzeyleri arasında çok büyük değer fark olmasından anlayabiliriz. Bu da F1’in kaybolmasına sebep olmuştur. Herhangi bir hiyerarşik bir kümeleme yapılmadığı için benzer özellik gösteren bölgeleri haritada yan yana getiremez. Bunlardan kaynaklı yorumlanabilirlik azalmış ve istatistiksel olarak sınıflama yapılamaz.

# Kümeleme Analizi

Kümeleme analizi, araştırmacıların bir dizi nesneyi, özellikleri bu tür diğer farklı kümelerden farklı olan küçük ama farklı kümelere ayırmasına veya gruplamasına olanak tanıyan istatistiksel bir araştırma yöntemidir. Analizde Z normalizasyonu kullanılarak K means, hiyerarşik kümeleme ve K means++ kullanılmıştır.

## 21.1. Z Normalizasyonu ile K means Kümeleme

Türkiye'deki bölgelerin 15–64 yaş grubundaki nüfusa ilişkin dört farklı eğitim düzeyine (F1: okuma yazma bilmeyen, F2: lise altı, F3: lise ve dengi, F4: yükseköğretim) göre işgücü katılım verilerini kullanarak k-ortalama (k-means) kümeleme analizi gerçekleştirmektedir. eri setine **3** kümeye ayırma işlemi uygulanmıştır. K-means algoritması, verideki gözlemleri (bölgeleri) eğitim düzeyi bazında birbirine benzer yapılar gösteren gruplar hâlinde toplar.

Tablo 14. Kümelere Dağılım

|  |  |  |
| --- | --- | --- |
| **Küme 1** | **Küme 2** | **Küme 3** |
| 1 | 3 | 22 |

Tablo 15. Kümelere Dağılan Ortalama Kişi Sayısı

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Küme** | **F1** | **F2** | **F3** | **F4** |
| **1** | 50 | 2476 | 2004 | 2639 |
| **2** | 65.66 | 420 | 188.33 | 146.33 |
| **3** | 17.63 | 471.9 | 308.54 | 335.6364 |

Küme 1, tüm eğitim düzeylerinde en yüksek işgücüne katılım ortalamalarına sahip bölge grubudur. Muhtemelen bu kümede İstanbul gibi büyük, gelişmiş metropol bölgeler yer almaktadır. Küme 2, diğerlerine göre F1’deki katılımın yüksek olması düşük eğitimli nüfusun yaygın olduğunu gösterir. Bu küme, genellikle Doğu ve Güneydoğu Anadolu’daki dezavantajlı bölgeleri içerebilir. Çünkü lise ve üniversite mezunu işgücünün çok az olmasından kaynaklanır. Küme 3; F1 seviyesi oldukça düşük, F2-F4 düzeylerinde yakın değerler olması orta seviyede işgücü katılımı olduğunu gösterir.

### 21.1.1. Kümelere Göre Kruskal-Wallis ve Dunn’s Testi

Tablo 16. Kruskal-Wallis Testi Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ki-kare** | **df** | **P değeri** |
| **F1** | 9.8797 | 2 | 0.007156 |
| **F2** | 3.0119 | 2 | 0.2218 |
| **F3** | 4.4354 | 2 | 0.1089 |
| **F4** | 5.3678 | 2 | 0.0683 |

Kruskal–Wallis testi, k-means kümeleme sonucu elde edilen 3 küme)arasında eğitim düzeyine göre işgücüne katılım dağılımlarının anlamlı şekilde farklılaşıp farklılaşmadığını test etmektedir. Sonuçlara göre F1’de kesin olarak bir fark olduğu söylenebilir.

Tablo 17. Dunn’s Testi Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z İstatistiği** | **Düzeltilmemiş p değeri** | **Düzeltilmiş p değeri** |
| **1-2** | -0.227 | 0.82 | 1 |
| **1-3** | 1.474 | 0.14 | 0.421 |
| **2-3** | 2.875 | 0.004 | 0.012 |

F1 grubu için yalnızca küme 2 ile küme 3 arasında anlamlı bir fark elde edilmiştir. Bu durum, düşük eğitimli bireylerin işgücü katılımı açısından özellikle Küme 2’nin (en dezavantajlı küme) belirgin şekilde ayrıştığını göstermektedir.

## 21.1.2. Isı Grafiği Gösterimi

metin, ekran görüntüsü, yazı tipi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 13. Kümeleme Analizi Uygulanmış Isı Haritası

Küme 1 (İstanbul), F2, F3, F4 kategorilerinde çok yüksek z-skorları (koyu kırmızı): yükseköğretim, lise ve alt düzey mezunlarda işgücü katılımı Türkiye ortalamasının oldukça üzerindedir. Yüksek eğitimli ve yoğun işgücü barındıran bir merkez yorumu yapılabilir. Küme 2 (Güneydoğu/Doğu Anadolu), okuma yazma bilmeyen ve lise altı eğitim seviyesi yüksek seviyededir. Daha az eğitimli ve düşük seviye formel eğitime sahip nüfusa sahiptir. Küme 3 (Orta seviye), bu bölgeler, eğitim seviyesi ve işgücü katılımı bakımından geçiş/orta düzeydeki yapıya sahiptir. Ne çok yüksek ne de çok düşük düzeydedir. Bu ısı haritası, bölgesel eğitim eşitsizliklerini ve bu eşitsizliğin işgücü piyasasına etkilerini görselleştirmek için yapılmıştır. Şekil 6’da bu açıklamalar choropleth grafi bazında çizilmiş Türkiye haritası üzerinde gösterilmiştir.

harita, sarı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.harita, sarı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 14. Türkiye Haritası Üzerinde Gösterimi

## Z Normalizasyonu ile Hiyerarşik Kümeleme

Bu kısımda Türkiye'deki bölgelerin eğitim düzeylerine göre hiyerarşik kümeleme analizi yapılmasını, ardından bu kümeleme sonuçlarının görsel olarak dendrogram (ağaç yapısı) ve ısı haritası (heatmap) ile sunulmasını sağlamak amaçlanmıştır. Küme dağılımında z normalizasyonu kullanıldığı için k means’te elde edilen sonuçlarla benzer sonuçlar elde edildiğinden ısı haritası ile dendrogram grafiği yorumlaması yapılacaktır.

metin, yazılım, web sitesi, web sayfası içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 15. Hiyerarşik Kümeleme Eklenmiş Isı Haritası

Şehirler hiyerarşik kümeleme algoritmasıyla benzerliğe göre gruplanmış, dendrogram bunu görselleştirmiştir. Şekil 5’te de yapılan grafiğe benzer bir yapı görülmektedir. İstanbul yüksek seviyedeki eğitim düzeylerini kapsarken büyük ölçüde bölgeler orta seviye eğitim düzeyinde işgücüne sahip olduğu görülmektedir. Buna ek olarak dendrogram grafiği de aynı şekli göstermektedir.

metin, ekran görüntüsü, paralel, doküman, belge içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 16. Oluşturulan Dendrogramın Ayrıntılı ve Kümelere Bölünmüş Hali

## Z Normalizasyonu ile K Means++

KMeans++, rastgele seçmek yerine daha iyi başlangıç ​​noktalarını otomatik olarak seçen KMeans algoritmasının geliştirilmiş bir sürümüdür. KMeans++'ın arkasındaki temel fikir, algoritmanın daha hızlı yakınsamasına yardımcı olan ve daha iyi kümeleme sonuçları veren yayılmalarını sağlamak için başlangıç ​​küme merkezlerini akıllı bir şekilde seçmesidir.

Tablo 18. K means++ Göre Dağılım

|  |  |  |
| --- | --- | --- |
| **Küme 1** | **Küme 2** | **Küme 3** |
| 21 | 1 | 4 |

Tablo 19. Kümelere Dağılan Ortalama Kişi Sayısı

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Küme** | **F1** | **F2** | **F3** | **F4** |
| **1** | 16.8 | 484.2857 | 319.95 | 348.5714 |
| **2** | 50 | 2476 | 2004 | 2639 |
| **3** | 58 | 368 | 158.5 | 125.75 |

Bir önceki kümeleme analizlerine benzer sonuçlar çıkmıştır. Küme 2’de yine İstanbul tek üyedir. Bu durum İstanbul’un istatistiksel uç nokta olduğunu doğrular. Diğer bölgelerle kıyaslandığında veri dağılımını bozacak kadar baskın bir profile sahiptir. Bu, bölgesel politika üretiminde İstanbul’a ayrı yaklaşılması gerektiğini gösterir. Küme 3 dezavantajlı yapıdadır. Eğitim ve işgücü açısından kırılgan, kırsal ve düşük gelişmişlik düzeyine sahip bölgelerdir. Küme 1 ise orta seviye bir yapıya sahiptir. Belirli bir denge bulunur.

### Kruskal-Wallis ve Dunn’s Testi

Tablo 20. Kruskal-Wallis Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ki-kare** | **df** | **P değeri** |
| **F1** | 11.522 | 2 | 0.003148 |
| **F2** | 3.9238 | 2 | 0.1406 |
| **F3** | 6.7664 | 2 | 0.03394 |
| **F4** | 7.9874 | 2 | 0.01843 |

Test sonucuna göre k means++ kümelerinin değişkenler içinde farklılık olan F1, F3 ve F4’tür. Bu 3 değişkene Dunn’s testi uygulanıp asıl farklılık yaratan bulunacaktır.

Tablo 21. F1 için Dunn’s Testi Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z İstatistiği** | **Düzeltilmemiş p değeri** | **Düzeltilmiş p değeri** |
| **1-2** | -1.53 | 0.125 | 0.375 |
| **1-3** | -3.14 | 0.001 | 0.004 |
| **2-3** | -0.13 | 0.895 | 1 |

En anlamlı fark küme 1 ile küme 3 arasında görülmektedir. Bu fark, okuryazarlık düzeyinin kümeleri ayırmada güçlü bir rol oynadığını gösterir.

Tablo 22. F3 için Dunn’s Testi Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z İstatistiği** | **Düzeltilmemiş p değeri** | **Düzeltilmiş p değeri** |
| **1-2** | -1.49 | 0.136 | 0.4 |
| **1-3** | 1.99 | 0.045 | 0.137 |
| **2-3** | 2.33 | 0.019 | 0.058 |

Küme 2-3 bakıldığında İstanbul’un F3 seviyesinde yoğun bir lise mezunu işgücüne sahip olduğu, buna karşılık dezavantajlı bölgelerde bu kitlenin işgücüne katılımda çok düşük kaldığı görülmektedir. Küme 1(Anadolu vb.) – 3 (Doğu vb.) için Bonferroni düzeltmesine bakıldığında anlamlı olmasa da üzerinde çalışılması gereken bir fark bulunabilir. Lise düzeyindeki bireylerin işgücü entegrasyonunda bölgesel eşitsizlik çok belirgindir.

Tablo 23. F4 için Dunn’s Testi Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z İstatistiği** | **Düzeltilmemiş p değeri** | **Düzeltilmiş p değeri** |
| **1-2** | -1.46 | 0.142 | 0.427 |
| **1-3** | 2.82 | 0.022 | 0.06 |
| **2-3** | 2.45 | 0.014 | 0.042 |

Küme 2 (İstanbul) ile küme 3 (Doğu) İstanbul ile doğudaki düşük eğitimli bölgeler arasında yüksek öğretimli işgücü açısından çok belirgin bir uçurum bulunmaktadır. Bu fark sosyoekonomik eşitsizliğin en açık göstergelerindendir. Orta seviyedekiler de incelendiğinde yüksek öğretim mezunu işgücü açısından belirgin şekilde geridedir.

### Isı Haritası

metin, ekran görüntüsü, diyagram, tasarım içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

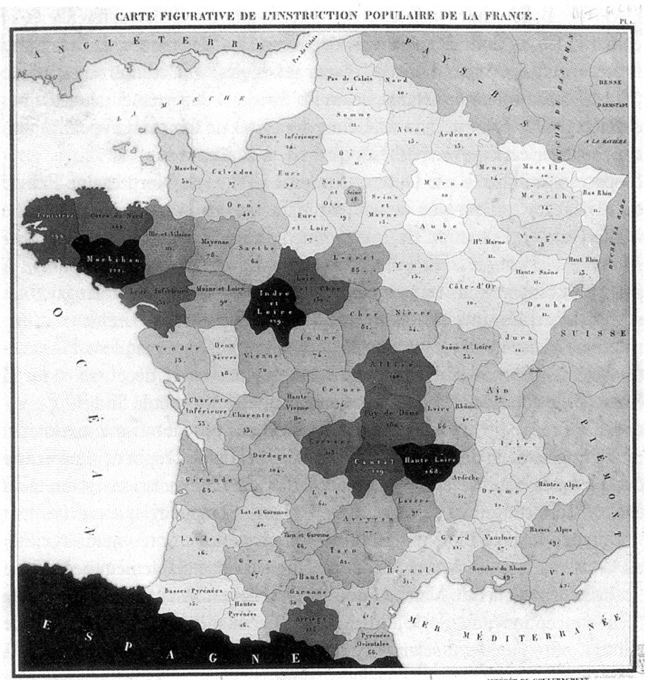
Şekil 17. Kümelere göre Isı Haritası

Şekil 9, k**-means++ algoritmasıyla oluşturulan 3 kümenin,** dört farklı eğitim düzeyine göre (F1–F4) **z-normalize edilmiş ortalama işgücü değerlerini** gösteren bir **ısı haritasıdır**.İstanbul’un (küme 2) eğitimli işgücü potansiyeli yüksek olması z-normalizasyon yapıldığında bile tüm sütunlarda kendini göstermesine sebep olmaktadır. Küme 1 ise ortalama illeri temsil ettiğinden Okuryazarlık sorunu çok yüksek olmamasına rağmen F3 ve F4 düzeylerinde İstanbul’un oldukça gerisindedir. Küme 3 ise dezavantajlı gruptur. Eğitim seviyesi düşük, okuryazarlık sorunları yüksek ve yükseköğretime erişim çok sınırlıdır.

# Sonuç ve Değerlendirme

Bu çalışmada, 15-64 yaş aralığındaki nüfusun 4 eğitim düzeyine göre işgücüne katılımının bölgesel dağılımını K means, hiyerarşik ve K means++ kümeleme yöntemlerini Z normalizasyonu kullanılarak pekiştirilmesiyle analiz edildi. Elde edilen kümeler hem ham verilerle hem de z-normalize edilmiş değerlerle farklı açılardan değerlendirilmiştir. Genel olarak bakıldığında İstanbul her algoritmada tek başına bir küme oluştururken yanında orta seviyeyi oluşturan karışık bir küme ve eğitim seviyesi çok düşük olan üçüncü küme ile birleşti. Yapılan analizler sonucunda Türkiye’de bölgesel işgücü yapısı eğitim düzeyine göre ciddi biçimde farklılaşırken bu farklılıklar yapılan test ve coğrafi kümelenme ile de kanıtlanmaktadır. Bu nedenle eğitim düzeyine göre farklılaştırılmış bölgesel politikalar geliştirilmelidir. Özellikle Doğu bölgelerine okuryazarlık seferberliği, meslek eğitimi ve istihdam teşvikleri üzerine çalışılmalıdır. İstanbul’da bulunan yüksek öğretim seviyesinden yararlanılarak AR-GE projeleri, teknoloji yatırımlı çalışmalar artırılmalıdır.

# Koroplet Haritası (Choropleth Map) Nedir?



Şekil 18. Dupin'in 1826 tarihli Fransa'daki okuryazarlık haritası

Koroplet haritası, adını Antik Yunanca *khôros* (alan, bölge) ve *plêthos* (çokluk) kelimelerinden alan, belirli bir coğrafi özelliğin mekânsal sayım birimleri (örneğin il, bölge, ülke) içindeki toplulaştırılmış istatistiksel değerlerini renk tonları (pseudocolor) aracılığıyla görselleştiren tematik harita türüdür. Bu haritalar, bir değişkenin coğrafi dağılımını analiz etmek veya bir bölge içindeki değişkenlik düzeyini ortaya koymak için etkili bir araçtır. Koroplet haritaları, genellikle hükümet veya resmî kurumlar tarafından yayınlanan istatistiksel verilerin il, eyalet veya ülke gibi idari birimlerde toplanması nedeniyle yaygın olarak kullanılır. Coğrafi Bilgi Sistemleri (GIS), elektronik tablolar ve diğer yazılım araçları ile kolayca üretilebilirler. Isı haritaları veya isaritmik haritalar ile benzerlik gösterse de koroplet haritaları sabit coğrafi sınırları temel alırken; diğer türler, değişkenin mekânsal örüntüsüne göre tanımlanan bölgeleri esas alır.

Bilinen en eski koroplet haritası, 1826 yılında Baron Pierre Charles Dupin tarafından hazırlanmış olup, Fransa'da temel eğitime erişimi iller bazında görselleştirmiştir. Bu çalışma, Fransa'da eğitim, sağlık, suç ve yaşam koşulları gibi alanlardaki diğer "ahlaki istatistiklerin" görselleştirilmesinde öncü olmuş ve kısa sürede benzer "cartes teintées" (renkli haritalar) üretilmeye başlanmıştır. 1850 sonrasında kromolitografinin yaygınlaşmasıyla birlikte, koroplet haritalarda daha zengin ve çeşitli renk kullanımları mümkün hale gelmiştir.

“Koroplet harita” terimi ilk olarak 1938 yılında coğrafyacı John Kirtland Wright tarafından literatüre kazandırılmış ve 1940’lı yıllarda kartograflar arasında yaygın bir kullanım kazanmıştır. Aynı yıl Glenn Trewartha, bu haritaları "oran haritaları" olarak yeniden tanıtmaya çalışsa da bu terim yaygınlık kazanmamıştır.

Bir koroplet haritası iki veri kümesini bir araya getirir: coğrafi alanın ayrı bölgelere bölünmesini temsil eden mekansal veriler ve her bölge içinde toplanan bir değişkeni temsil eden istatistiksel veriler.

Farklı yüzölçümüne veya nüfusa sahip bölgeler arasında ham verilerin (örneğin toplam vaka, boşanma sayısı) doğrudan karşılaştırılması yanıltıcı olabilir. Bu yüzden veri grafiğe çevrilmeden önce normalize edilmelidir. Ayrıca grafiğin başlığı okuyucunun anlayacabileceği seviyede olmalıdır. Eğer harita üzerindeki bölgeler çok farklı büyüklüklere sahipse, büyük alanlar görsel olarak daha baskın hale gelebilir ve verinin yanlış algılanmasına neden olabilir. Bu gibi durumlarda hexbin haritaları gibi alternatif görselleştirme yöntemleri değerlendirilerek bu tür hatalar engellenebilir.

## Avantajları ve Dezavantajları

Koroplet haritalar, bir haritada bir değişkeni temsil etmek için kullanılır. Bir değişkenin bir bölge genelindeki dağılımını göstermenin harika bir yoludur.

**Mekansal Desenleri Vurgulama:** Koroplet haritalar, coğrafi bölgeler arasında değişkenin dağılımını görselleştirerek genel eğilimleri ve aykırı değerleri kolaylıkla tanımlamayı mümkün kılar. Bu sayede mekânsal örüntüler açıkça ortaya konur.

**Karşılaştırma Kolaylığı:** Renk gradyanları aracılığıyla farklı bölgeler arasındaki veri değerlerinin karşılaştırılması kolaylaşır. Bu, özellikle bölgesel farklılıkların analizinde güçlü bir görsel araç sunar.

**Veri Entegrasyonu:** Koroplet haritalar, çeşitli kaynaklardan elde edilen istatistiksel verilerin coğrafi bilgi sistemleri (GIS) ile entegre edilmesine olanak tanır. Bu sayede daha kapsamlı ve çok katmanlı coğrafi analizler gerçekleştirilebilir.

**Ölçeklenebilirlik:** Bu haritalar, küresel düzeyden yerel idari birimlere kadar farklı coğrafi ölçeklerde uygulanabilirlik sunar. Böylece mikro ve makro düzeyde analizler yapılabilir.

**Özelleştirme İmkânı:** Kullanıcılar renk paletlerini, sınıflandırma yöntemlerini (doğal aralık, eşit aralık, kuantil vb.) ve interaktif özellikleri ihtiyaca göre özelleştirebilir. Bu, görselleştirmeyi amaca uygun hale getirir.

**Önyargı Azaltma:** Verilerin normalleştirilmesi (örneğin nüfusa göre oranlanması), bölge büyüklüğü veya nüfus yoğunluğu gibi etkenlerden kaynaklanabilecek algısal önyargıların önüne geçebilir. Ayrıca, koroplet haritalara alternatif olarak kartogram gibi yöntemlerin değerlendirilmesi, daha adil ve temsil gücü yüksek görselleştirmeler sağlayabilir.

Bunların yanı sıra, koroplet grafiğinde kesin olarak renkler üzerinden tüm verileri okumanın imkânı yoktur. Koroplet haritalarda bir başka sorun ise haritaların daha büyük alanlara odaklanması ve daha küçük bölgelerin gölgede kalarak izleyicinin algısını etkilemesidir. Koroplet haritalardaki yaygın bir hata haritalandırma aşamasında işlenmemiş verilerin kodlanmasıdır ve yanıltıcı sonuçlara neden olabilir. Koroplet haritalar yalnızca alanlara (bölgelere) bağlı verilerle kullanılabilir. Noktasal veriler ya da sürekli değişen yüzey verileri için uygun değildir.

# Isı Haritası ile Koroplet Farkı

Koroplet haritalar, genellikle idari birimler gibi ayrık (kesikli) coğrafi alanlara göre özetlenmiş kategorik veya oran-temelli verileri kullanırken ısı haritaları genellikle sürekli mekansal verilerle çalışır ve belirli bir alandaki olay yoğunluğunu yansıtır.

Koroplet haritaların amacı, verilerin coğrafi dağılımını ve bölgesel farklılıkları görselleştirmektir. Öte yandan, coğrafi ısı haritaları daha çok olayların mekânsal yoğunluğunu veya sıcak noktalarını vurgulamak amacıyla kullanılır.

Koroplet haritalar, iller, bölgeler, eyaletler gibi açık ve sabit coğrafi sınırlar temelinde oluşturulur. Coğrafi ısı haritaları ise sabit coğrafi sınırlara bağlı değildir; verilerin yoğunluk desenine göre şekillenen sürekli yüzeyler sunar.

Tablo 24. Grafik Türlerinin Farkları

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

# Analiz Edilecek Verinin Açıklanması

Bu çalışma kapsamında kullanılan veri seti, TÜİK’ten alınmış olup Türkiye'deki iller düzeyinde eşler arasındaki yaş farkına göre boşanma sayılarını içermektedir. Veri setinde yer alan 14 sütun, 81 ili ve her ilde yaş farkına göre sınıflandırılmış boşanma sayılarını göstermektedir. Değişkenler şekildeki gibi tanımlanmıştır.

**F1 – F10:** Kadın eşin erkek eşten 1 ila 10 yıl daha küçük olduğu durumlara karşılık gelen yaş farkı kategorileridir. Örneğin, F1 sütunu kadının eşinden 1 yaş küçük olduğu boşanma sayılarını, F5 ise 5 yaş küçük olduğu durumları göstermektedir.

**F11:** Kadın eşin erkek eşten 11 yaş ve daha fazla küçük olduğu boşanma sayılarını içerir.

**F12:** Erkek eşin kadın eşten küçük olduğu tüm boşanma vakalarını temsil eder.

**F13:** Eşlerin yaşlarının eşit olduğu durumu gösterir.

Bu veri, choropleth haritalar veya diğer görsel analiz araçlarıyla birlikte kullanılarak eşler arası yaş farkı boşanma oranlarının bölgesel dağılımını görmek, yüksek riskli bölgeleri tanımlamak ve kültürel yapının boşanma üzerindeki etkisini bölgesel olarak analiz edebilmek amaçlı kullanılacaktır.

Ayrıca grafik çiziminde önem arz eden normalleştirmenin neden önemli bir faktör olduğu da iki ayrı veri üzerinde yapılan kümeleme analizi ve grafik çizimi üzerinde gösterilecektir. İlk kısımda Z normalizasyonu ile elde edilen sonuçlar ve grafik çizimleri yapılırken ikinci kısımda yoğunluk ve oran hesaplaması üzerinden elde edilen veriyle aynı işlemler gerçekleştirilecektir.

## Açıklayıcı Analizler

Tablo 25. Açıklayıcı İstatistikler

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** | **Ortalama** | **SD** | **Medyan** | **Min** | **Maks** | **Aralık** | **Çarpıklık** | **Basıklık** | **Standart Hata** |
| **F1** | 81 | 209.57 | 431.24 | 82 | 7 | 3322 | 3315 | 5.19 | 31.86 | 47.62 |
| **F2** | 81 | 216.18 | 446.48 | 85 | 6 | 3478 | 3472 | 5.3 | 33.25 | 49.3 |
| **F3** | 81 | 206.84 | 418.18 | 83 | 7 | 3262 | 3255 | 5.29 | 33.24 | 46.18 |
| **F4** | 81 | 198.12 | 386.33 | 81 | 4 | 2980 | 2976 | 5.1 | 31.32 | 42.66 |
| **F5** | 81 | 179.04 | 345.78 | 77 | 6 | 2692 | 2686 | 5.16 | 32.29 | 38.19 |
| **F6** | 81 | 155.13 | 297.65 | 66 | 3 | 2277 | 2274 | 5 | 30.14 | 32.87 |
| **F7** | 81 | 132.7 | 253.94 | 58 | 4 | 1952 | 1948 | 5.05 | 30.72 | 28.04 |
| **F8** | 81 | 104.8 | 196.9 | 44 | 3 | 1500 | 1497 | 4.92 | 29.4 | 21.74 |
| **F9** | 81 | 81.48 | 156.03 | 37 | 2 | 1209 | 1207 | 5.11 | 31.65 | 17.23 |
| **F10** | 81 | 62 | 117.31 | 26 | 3 | 913 | 910 | 5.09 | 31.83 | 12.96 |
| **F11** | 81 | 197.82 | 355.18 | 78 | 12 | 2712 | 2700 | 4.81 | 28.84 | 39.22 |
| **F12** | 81 | 349.65 | 740.36 | 134 | 9 | 5823 | 5814 | 5.43 | 34.77 | 81.76 |
| **F13** | 81 | 178.76 | 375.79 | 67 | 6 | 2891 | 2885 | 5.22 | 31.96 | 41.5 |

Açıklayıcı analizler göz önüne alındığında ilk yorum olarak F1-F4 göze çarpmaktadır. Her ikisinin de ortalamasının ve medyanın yüksek olması boşanmanın yaşandığı en sık yaş farkı 1 yaş ile 4 yaş farkı olduğunu gösterir. F12’ye bakıldığında, erkeğin kadından genç olması durumu, standart sapmasının en yüksek olması bu kategorideki boşanmaların iller arasında çok değişken olduğunu belirtir. Buradan da bazı bölgelerin sosyokültürel olarak birbirinden farklılaştığını kanıtlar. Çarpıklık ve basıklık katsayılarına bakıldığında tümünde pozitif ve yüksektir. Bu da dağılımların sağa çarpık olduğunu gösterir. F12’nin çarpıklığı en yüksek olan olduğundan erkeğin kadından küçük yaşta olduğu evliliklerde bazı iller çok yüksek boşanma oranına sahip olduğunu göstermektedir.

## Korelasyon

ekran görüntüsü, kalıp, desen, düzen, metin, dikdörtgen içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 19. Ana Verideki Değişkenlerin Birbiriyle Olan İlişkisi

Şekil 2’de de görüleceği üzere her bir değişken için ilişki 1’e yakın olduğu analiz edilmiştir. Yani tüm değişkenler birbiriyle yüksek ve pozitif korelasyon içindedir. İller bazında birlikte artma eğiliminde olduğunu göstermektedir. Bu tür bir grafiğin elde edilmesinin en büyük sebebi verinin normalizasyon yapılmamış hali olmasıdır. Bu da değişkenler arasında yapay bir korelasyon oluşturur.

## Normallik Testi Sonuçları

Tablo 26. Shapiro-Wilk Testi Sonuçları

|  |  |  |
| --- | --- | --- |
|  | **W** | **P Değeri** |
| **F1** | 0.42578 | 2.615e-16 |
| **F2** | 0.4209 | 2.244e-16 |
| **F3** | 0.42764 | 2.773e-16 |
| **F4** | 0.44647 | 5.055e-16 |
| **F5** | 0.447 | 5.143e-16 |
| **F6** | 0.45398 | 6.45e-16 |
| **F7** | 0.45134 | 5.918e-16 |
| **F8** | 0.46267 | 8.578e-16 |
| **F9** | 0.45304 | 6.256e-16 |
| **F10** | 0.45751 | 7.239e-16 |
| **F11** | 0.47998 | 1.53e-15 |
| **F12** | 0.40708 | <2.2e-16 |
| **F13** | 0.41696 | <2.2e-16 |

Shapiro-Wilk testi sonuçlarına göre, tüm yaş farkı kategorilerinde boşanma sayılarına ilişkin veriler normal dağılım göstermemektedir (W < 0.5, p < 0.001). Bu test sonucunda veriye normalizasyon uygulanması gerektiği açıkça görülmektedir.

# Kümeleme Analizi

## Z Normalizasyonu ile Yapılan K means ve K means++ Karşılaştırılması

Tablo 27. K means Küme Dağılımı

|  |  |  |
| --- | --- | --- |
| **Küme 1** | **Küme 2** | **Küme 3** |
| 71 | 9 | 1 |

Tablo 28. K means++ Küme Dağılımı

|  |  |  |
| --- | --- | --- |
| **Küme 1** | **Küme 2** | **Küme 3** |
| 1 | 4 | 76 |

İki algoritma farklı başlangıç noktalarıyla çalışmasına rağmen, oluşan kümeler oldukça benzer sonuçlar vermektedir. Her iki yöntem de veri setinde baskın bir ana küme ve birkaç küçük küme olduğunu ortaya koymuştur. Bu yüzden Choropleth harita çiziminde, görsel denge ve anlamlı mekânsal desenlerin vurgulanabilmesi açısından **K-means++ sonucu** tercih edilmelidir.

metin, harita içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.metin, harita, atlas içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Bu iki grafikten görüleceği üzere her iki algoritma da benzer dağılımlar gösterse de sadece küme grubu değişmiş ve bazı iller K means algoritmasında diğerine göre daha yakın bulunduğu için aynı kümeye yerleştirilmiştir. Fakat yapısal olarak K means++ daha sağlıklı olduğu için analizin devamı ondan yapılacaktır.

Şekil 20. İki Algoritma Farkının Harita Üzerinde Gösterilmesi

## K means++ için Kruskal-Wallis Testi Sonuçları

Tablo 29. Kruskal-Wallis Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ki-kare** | **df** | **P değeri** |
| **F1** | 13.926 | 2 | 0.0009460 |
| **F2** | 13.926 | 2 | 0.0009463 |
| **F3** | 13.926 | 2 | 0.0009463 |
| **F4** | 13.927 | 2 | 0.0009459 |
| **F5** | 13.928 | 2 | 0.0009453 |
| **F6** | 13.928 | 2 | 0.0009454 |
| **F7** | 13.929 | 2 | 0.0009446 |
| **F8** | 13.929 | 2 | 0.0009448 |
| **F9** | 13.933 | 2 | 0.0009427 |
| **F10** | 13.931 | 2 | 0.0009440 |
| **F11** | 13.926 | 2 | 0.0009463 |
| **F12** | 13.926 | 2 | 0.0009460 |
| **F13** | 13.927 | 2 | 0.0009458 |

Kümeler, tüm yaş farkı kategorileri (F1–F13) için anlamlı biçimde farklılaşmaktadır (p < 0.001). Bu durum, boşanma profillerinin yaş farkına göre kümeler arasında sistematik farklılıklar gösterdiğini ortaya koymaktadır. Ki kare değerlerinin yakın olması değişkenlerin kümeler üzerinde ortak bir ayırt edici bir yapıya sahip olduğunu gösterir. Her değişken için hangi kümelerde farklılık olduğu Dunn’s testi ile analiz edilecektir.

## Dunn’s Testi Sonuçları

Tablo 30. F1 için Sonuçlar

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | 0.093 | 0.925 | 1.00 |
| **1 - 3** | 1.79 | 0.07 | 0.218 |
| **2 - 3** | 3.316 | 0.0009 | 0.002 |

Eşler arasında 1 yaş fark olması boşanma profilleri açısından kümeler arasında sadece 2-3 arasında gözlenmiştir. Küme 2’deki illerin dağınık olması yorum yapılmasını zorlaştırmaktadır.

Tablo 31. F2 için Sonuçlar

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | 0.093 | 0.93 | 1.00 |
| **1 - 3** | 1.79 | 0.07 | 0.22 |
| **2 - 3** | 3.32 | 0.00 | 0.002 |

Eşler arasında 2 yaş fark olması da aynı sonucu vermiştir. Diğer değişkenler için de p değerlerine bakıldığında hepsi aynı sonucu vermiştir. Küme 1’in tek üyeli olması farklılık açısından yorum yapılmasını imkânsız hale getirmiştir denebilir.

metin, ekran görüntüsü, diyagram, harita içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 21. ScatterPlot ile Gösterimi

Şekil 4’te F13 (aynı yaştaki eşlerin boşanma sayıları) değişkeni hem scatterplot hem de choropleth map ile ele alınmıştır. Bu ikili sayesinde hem istatistiksel olarak hem de coğrafi olarak **anormal sonuçlar** kolayca fark edilir.

# Oranlaştırılmış Veri ile Analiz

Koroplet haritalarda sıkça başvurulan bu yöntem, istatistikte kullanılan normalizasyon veya standardizasyon tekniklerine benzerlik gösterse de onlardan farklıdır. Genellikle, iki uzamsal olarak kapsamlı değişkenin oranı alınarak elde edilir. Her ne kadar bu tür oranlar yoğunluk temelli değişkenler üretse de yalnızca bazıları koroplet haritalarda anlamlı sonuçlar verir ve yaygın biçimde kullanılır.

Bu işlemin yapışma sebebi farklı nüfus büyüklüklerine sahip iller karşılaştırabilme ve mekânsal dağılımların daha anlamlı hale getirmektir.

## Açıklayıcı Analizler

Tablo 32. Oranlaştırılmış Veri için Açıklayıcı İstatistikler

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** | **Ortalama** | **SD** | **Medyan** | **Min** | **Maks** | **Aralık** | **Çarpıklık** | **Basıklık** | **Standart Hata** |
| **F1** | 81 | 0.09 | 0.01 | 0.09 | 0.05 | 0.13 | 0.08 | 0.25 | 1.03 | 0 |
| **F2** | 81 | 0.09 | 0.01 | 0.09 | 0.04 | 0.13 | 0.08 | -0.45 | 2.68 | 0 |
| **F3** | 81 | 0.09 | 0.01 | 0.09 | 0.06 | 0.11 | 0.06 | -0.4 | 0.14 | 0 |
| **F4** | 81 | 0.09 | 0.01 | 0.09 | 0.05 | 0.11 | 0.06 | -0.32 | 0.45 | 0 |
| **F5** | 81 | 0.08 | 0.01 | 0.08 | 0.04 | 0.1 | 0.06 | -1.06 | 2.39 | 0 |
| **F6** | 81 | 0.07 | 0.01 | 0.07 | 0.04 | 0.1 | 0.06 | 0.35 | 0.6 | 0 |
| **F7** | 81 | 0.06 | 0.01 | 0.06 | 0.04 | 0.08 | 0.04 | 0.13 | 0.06 | 0 |
| **F8** | 81 | 0.05 | 0.01 | 0.05 | 0.02 | 0.09 | 0.07 | 0.96 | 3.45 | 0 |
| **F9** | 81 | 0.04 | 0.01 | 0.04 | 0.01 | 0.06 | 0.05 | 0.05 | 0.99 | 0 |
| **F10** | 81 | 0.03 | 0.01 | 0.03 | 0.01 | 0.06 | 0.04 | 0.9 | 1.48 | 0 |
| **F11** | 81 | 0.1 | 0.02 | 0.09 | 0.05 | 0.16 | 0.11 | 0.77 | 0.47 | 0 |
| **F12** | 81 | 0.15 | 0.03 | 0.15 | 0.08 | 0.23 | 0.15 | 0.53 | 1.26 | 0 |
| **F13** | 81 | 0.08 | 0.01 | 0.08 | 0.05 | 0.11 | 0.06 | -0.13 | 0.42 | 0 |

Ortalamada öne çıkan değişkenler F12 ve F11 olmakla birlikte bu durum Türkiye’de boşanma oranlarının en çok kadının erkekten büyük olduğu ve 11 ve 11+ yaş farkı olduğu evlilikler üzerinde yoğunlaştığını gösterir. Genel olarak F9-F10 (9 ve 10 yaş fark) değişkenleri düşük orana sahip ve daha dengeli bir dağılıma sahip olduğu gözlenmiştir.

## Korelasyon

metin, ekran görüntüsü, sayı, numara, takvim içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 22. Oranlaştırılmış Veri Korelasyonu

Pozitif korelasyon grubuna bakıldığında F2-F4, F1-F13, F2-F13, F3-F13, F4-F13, F10-F11 gibi değişkenler için birlikte artış/azalış yaşadığı söylenebilir. Ayrıca başka bir deyişle bu yaş farkı oranları illerde benzer oranlarda bulunmaktadır.

Negatif korelasyonlar ise F3-F11, F7-F13, F1-F10, F8-F13, F11-F13 gibi kategoriler arasında güçlü bir korelasyon görülmektedir. Bazı bölgelerde kadın büyük erkek küçük olan boşanmaların çok yüksek, diğer yaş farkı kategorilerinin ise görece düşük olduğunu gösterir.

## Normallik Testi

Tablo 33. Shapiro-Wilk Sonuçları

|  |  |  |
| --- | --- | --- |
|  | **W** | **P Değeri** |
| **F1** | 0.97735 | 0.15570 |
| **F2** | 0.95923 | 0.01066 |
| **F3** | 0.98532 | 0.47540 |
| **F4** | 0.98050 | 0.24710 |
| **F5** | 0.93069 | 0.00026 |
| **F6** | 0.98455 | 0.43120 |
| **F7** | 0.98332 | 0.36640 |
| **F8** | 0.93590 | 0.00049 |
| **F9** | 0.97928 | 0.20680 |
| **F10** | 0.95527 | 0.00611 |
| **F11** | 0.95535 | 0.00617 |
| **F12** | 0.96419 | 0.02179 |
| **F13** | 0.97260 | 0.07627 |

Oranlama sonrası dağılımların tamamı normalliğe uymasa da yüzdelik olarak genelde uyan sayısı fazla olduğu için fazla sorun çıkarmayacağı üzerinde durulabilir. Kısmen normale yakınsaması beklenen bir durumdu çünkü yaş farkı türleri Türkiye’deki bölgeden bölgeye uç kategori olabilecek yapıdadır.

# Kümeleme Analizi

Bir önceki veri için K means++ tercih edildiği için ve sağlıklı bir karşılaştırma yapılmak istendiğinden bu veri setine de bu algoritma uygulanmıştır.

Tablo 34. Kümelere Dağılış

|  |  |  |
| --- | --- | --- |
| **Küme 1** | **Küme 2** | **Küme 3** |
| 24 | 38 | 19 |

Z normalizasyonu ile yapılan K means++ algoritmasında kümenin birinde tek üye bulurken bu sefer küme dağılımında daha sağlıklı bir dağılım olduğunu görmekteyiz. Değişkenlerin hangi kümeye daha yatkın olduğunu anlamak için her küme için ortalama boşanma oranları hesaplanmıştır.

Tablo 35. Her Küme için Dağılım Ortalaması

|  |  |  |  |
| --- | --- | --- | --- |
| **Değişken** | **Küme 1** | **Küme 2** | **Küme 3** |
| **F1** | 0.087971 | 0.095313 | 0.076470 |
| **F2** | 0.085434 | 0.096871 | 0.091517 |
| **F3** | 0.080755 | 0.094438 | 0.092061 |
| **F4** | 0.080160 | 0.092299 | 0.083200 |
| **F5** | 0.082326 | 0.081207 | 0.075535 |
| **F6** | 0.064533 | 0.066371 | 0.078703 |
| **F7** | 0.061196 | 0.056008 | 0.065180 |
| **F8** | 0.048927 | 0.043363 | 0.055551 |
| **F9** | 0.035537 | 0.034848 | 0.041746 |
| **F10** | 0.027775 | 0.026970 | 0.034946 |
| **F11** | 0.111814 | 0.081706 | 0.102002 |
| **F12** | 0.162183 | 0.148774 | 0.136432 |
| **F13** | 0.071388 | 0.081831 | 0.066657 |

Küme 1 için en yüksek oran F12 (erkeğin yaşının küçük olduğu durum) sahip olması bu kümenin geleneksel yapıda olmadığını gösterirken F11’in de büyük olması yaş farkının etkili olduğu söylenebilir. Küme 2 için yine F12 büyüktür fakat en büyük F12 değerine küme 1 sahip olduğundan etkisi buna daha az olduğu yorum yapılabilir. Diğer değişkenlerinin ortalama oran değerleri de birbirine yakın olduğu için bu kümede homojen bir dağılım olduğu görülmektedir. Küme 3, **F12** burada en düşük olduğundan kadının büyük yaşta olduğu evliliklerde bu kümede daha az boşanmayla sonuçlanmış olabilir. Buradaki bölgeler daha modern yapıya sahip olabilir.

## Kruskal-Wallis Testi

Tablo 36. Kruskal-Wallis Sonuçları

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Ki-kare** | **df** | **P değeri** |
| **F1** | 27.38800 | 2 | 0.00000113 |
| **F2** | 14.33000 | 2 | 0.00077320 |
| **F3** | 26.02900 | 2 | 0.00000223 |
| **F4** | 24.10500 | 2 | 0.00000583 |
| **F5** | 1.48790 | 2 | 0.47520000 |
| **F6** | 21.71400 | 2 | 0.00001927 |
| **F7** | 13.41100 | 2 | 0.00122400 |
| **F8** | 16.92300 | 2 | 0.00021140 |
| **F9** | 11.40700 | 2 | 0.00333400 |
| **F10** | 11.13200 | 2 | 0.00382500 |
| **F11** | 35.13800 | 2 | 0.00000002 |
| **F12** | 11.38800 | 2 | 0.00336600 |
| **F13** | 27.85000 | 2 | 0.00000090 |

3 farklı kümede boşanma oranlarının dağılımları arasında anlamlı fark olup olmadığını test etmek amaçlı yapılmıştır. **F5 hariç tüm kategorilerde** kümeler arasında anlamlı farklar olduğu görülmektedir. Kümeleme sonuçlarının sosyolojik farklılıklarla tutarlı olduğu görülmekte bu da kümelerin yorumlanabilir ve istatistiksel olarak geçerli olduğunu destekler.

## Dunn’s Testi

Tablo 37. F1 için Dunn’s Testi

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | -2.066191 | 0.03881049 | 0.11643150 |
| **1 - 3** | 3.017573 | 0.00254807 | 0.00764422 |
| **2 - 3** | 5.228120 | 0.00000017 | 0.00000051 |

Küme 3 diğer gruplardan farklı olduğu açıkça görülmektedir. Ortalama oran değerlerine bakıldığında (Tablo 11) en düşük değere sahip olan grup olarak göze çarpmaktadır.,

Tablo 38. F2 için Dunn’s Testi

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | -3.776278 | 0.00015919 | 0.00047757 |
| **1 - 3** | -1.750190 | 0.08008552 | 0.24025656 |
| **2 - 3** | 1.580780 | 0.11392841 | 0.34178522 |

Küme 1 ve 2’nin birbirinden farklı olduğu söylenebilir. Ortalamalara bakıldığında küme 1 değer olarak diğerlerinden düşük kalmaktadır.

Tablo 39. F3 için Dunn’s Testi

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | -5.023625 | 0.00000051 | 0.00000152 |
| **1 - 3** | -3.386130 | 0.00070886 | 0.00212658 |
| **2 - 3** | 0.941845 | 0.34627220 | 1.00000000 |

Küme 1 yine farklılık göstermektedir. Bazı bölgelerde 3 yaş farkı olan evliliklerde boşanma daha sık ya da nadir olabilir.

Tablo 40. F4 için Dunn’s Testi

|  |  |  |  |
| --- | --- | --- | --- |
| **Karşılaştırma** | **Z** | **P (Unadj)** | **P (Bonferroni)** |
| **1 - 2** | -4.551533 | 0.00000533 | 0.00001598 |
| **1 - 3** | -0.811443 | 0.41711120 | 1.00000000 |
| **2 - 3** | 3.330061 | 0.00086827 | 0.00260481 |

Burada da küme 2’nin ayırt edici bir farklılığı olduğu görülmektedir. Genel olarak diğer grup farklılıklarına da bakıldığında küme 2 ve 3’ün birçok yaş farkı kategorisinde farklılık yarattığı söylenebilir. F11 (11+ yaş farkı) ve F12 (kadının büyük olduğu) belirgin küme farkları görülmektedir. Bu farkların boşanma açısından hassas noktalar olduğu söylenebilir.

# Grafik Yorumlaması

metin, harita, atlas içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 23. Kümelerin Harita Üzerinde Görüntülenmesi

Kümelerin dağılımından da görüleceği üzere Küme 1 için yoğunlukla Akdeniz ve Doğu kısmında görülmektedir. F11 (11+ yaş farkı) ve F12 (kadının büyük olduğu) boşanma oranları yüksekti. Toplumsal normların dışına çıkan evlilik modelleri (kadın büyük, büyük yaş farkı) daha kırılgan olabilir.

Küme 2’nin daha çok Marmara, Ege ve Anadolu Bölgesi kısmında çoğunluğu sağlar. Bu küme F11 (11+ yaş farkı) en az olduğu gruptu. F1-F4 (1,2,3,4 yaş farkı olan) boşanma oranlarının fazlaydı. Bu durum bölgelerde küçük yaş farkının riskli olduğunun göstergesidir.

Küme 3 için ise dağınık bir yayılma olduğu söylenebilir. F6-F10 değerlerinin yüksek olması orta düzey yaş farkı riskli olabilir anlamına gelir.

diyagram, metin, harita içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 24. En Yaygın Boşanma Türü

F12’nin baskın olması Türkiye genelinde erkeğin kadından yaşça küçük olması evliliklerde boşanma riskini artırdığı söylenebilir. Bu da toplumun daha farklı bakış açılarına hazır olmadığı, gelenekselden vazgeçilemediğini gösterebilir. F11’in az yerde de olsa yaygın olması da büyük yaş farkı nesil farklılığına sebep olduğunu açıkça söylenebilir burada da iletişim konularında birtakım sıkıntılar doğurmasından kaynaklı boşanma riskini arttırdığı yorumu yapılabilir.

harita, ekran görüntüsü, metin içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 25. Oranların Grafikte Gösterimi (F11)

harita, ekran görüntüsü, metin içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 26. Oranların Grafikte Gösterimi (F12)

Şekil 7’ye ek olarak bu grafiklerin verilmesinin sebebi evlilik tipinin kırılganlığının yoğunluk derecesini vermek amaçlıdır. Burada F12’nin hangi şehirlerde oran olarak yüksek olduğu görülebilir. İstanbul, Aydın, Muğla gibi şehirler göze çarpmaktadır.

F11 için Tunceli, Şırnak gibi illerde F11 en yaygın kategori ama oran haritasında çok yüksek seviyede olmadıkları görülmektedir. Bu illerde 11+ yaş farkı kategorisinde diğer kategorilere göre daha çok boşanma var, ama toplam boşanma sayısı az veya dağılım daha dengeli yorumu yapılabilir.

diyagram, harita içeren bir resim

Yapay zeka tarafından oluşturulmuş içerik yanlış olabilir.

Şekil 27. Yoğunluk Grafiği

Bu harita, Türkiye’de illere göre boşanma yoğunluğunu (boşanma sayısı / km²) görselleştiriyor. Özellikle İstanbul’un yüksek seviyeye sahip olmasının sebepleri arasında metropol olmasına rağmen yüz ölçümünün küçük olması yani boşanma fazla ama alan da küçük olduğu için yoğunluk fazlaca göze çarpıyor. Genel olarak ise bakıldığında batı ve büyükşehirlerde boşanma sıklığı alan başına oldukça yoğundur. Modernleşme, yaşam biçimi veya kişisel yaşam tercihi olabilir. Doğu taraflarındaki düşük yoğunluk geleneksel yapıyı gösteriyor denebilir.

# Sonuç ve Değerlendirme

Bu çalışmada Türkiye’de 81 il üzerinden eşler arasındaki yaş farkına dayanarak yaşanan boşanma verilerini ele alarak koroplet haritaları ve istatistiksel analizler uygulanarak birtakım çıkarımlarda bulunulmaya çalışıldı. Açıklayıcı analizler, normallik testi, kümeleme analizleri ve görsellik ile bu çıkarsamalar desteklendi.

Boşanma oranları açısından en belirgin kategori F12 (kadının erkekten yaşça büyük olduğu) olarak bulunmuştur. Toplumun geleneksel normlarına aykırı olan bu evlilik biçimi daha aykırılık göstermiştir. Buna rağmen, F1-F4 (1,2,3,4 yaş farkı olan) en sık görülen boşanma türlerinden biri olmuştur.

Raporda başka kanıtlanmaya çalışılan normalleştirilmeyen verilerin sonuca etkisidir. İller arasında nüfus farkı olduğundan beklenen sonuçlar elde edilememiştir ve kümeleme de dengesiz bir dağılım görülmüştür.

# Kaynakça

Adams, A., W, L., Zhang, C., & Chen, X. (2020, Mayıs 7). *The disguised pandemic: the importance of data normalization in COVID-19 web mapping.* Mayıs 2025 tarihinde NIH: https://pmc.ncbi.nlm.nih.gov/articles/PMC7203028/ adresinden alındı

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., Jordy, v. L., & A Kievit, R. (2021, Ocak). *Raincloud plots: a multi-platform tool for robust data visualization.* Mayıs 2025 tarihinde NIH: https://pmc.ncbi.nlm.nih.gov/articles/PMC6480976/ adresinden alındı

Awati, R. (2023, Ağustos 22). *Definition of heat map (heatmap)*. Mayıs 2025 tarihinde TechTarget: https://www.techtarget.com/searchbusinessanalytics/definition/heat-map adresinden alındı

Belengeanu, A. (2022, Aralık 5). *Getting started with Raincloud plots in Python*. Mayıs 2025 tarihinde Medium: https://medium.com/@alexbelengeanu/getting-started-with-raincloud-plots-in-python-2ea5c2d01c11 adresinden alındı

*Chart Snapshot: Jitter Plots*. (2024, Temmuz 24). Mayıs 2025 tarihinde The Data Visualisation Catalogue Blog: https://datavizcatalogue.com/blog/chart-snapshot-jitter-plot/ adresinden alındı

*Choropleth map*. (tarih yok). Mayıs 2025 tarihinde from Data to Viz: https://www.data-to-viz.com/graph/choropleth.html adresinden alındı

*Choropleth map with R and ggplot2*. (tarih yok). Mayıs 2025 tarihinde https://r-graph-gallery.com/327-chloropleth-map-from-geojson-with-ggplot2.html adresinden alındı

Eisen, M., Spellman, P., Brown, P., & Botstein, D. (1998, Aralık 8). *Cluster analysis and display of genome-wide expression patterns.* Mayıs 2025 tarihinde NIH: https://pmc.ncbi.nlm.nih.gov/articles/PMC24541/ adresinden alındı

Fleetwood, D. (tarih yok). *Küme Analizi: Nedir ve Nasıl Kullanılır*. Mayıs 2025 tarihinde QuestionPro: https://www.questionpro.com/blog/tr/kume-analizi-nedir-ve-nasil-kullanilir/#what\_is\_cluster\_analysis adresinden alındı

Geoapify. (2024, Ağustos 1). *What is a Choropleth Map? Definition, Examples, and How to Create Custom Maps*. Mayıs 2025 tarihinde Geoapify: https://www.geoapify.com/what-is-a-choropleth-map-definition-examples-how-to-create/ adresinden alındı

*Heat map*. (tarih yok). Mayıs 2025 tarihinde Wikipedia: https://en.wikipedia.org/wiki/Heat\_map adresinden alındı

*Heat Map (Isı Haritası) Nedir ve Nasıl Çalışır?* (tarih yok). Mayıs 2025 tarihinde thro: https://thro.com.tr/heat-map-isi-haritasi-nedir-ve-nasil-calisir/ adresinden alındı

*Heatmap*. (tarih yok). Mayıs 2025 tarihinde Optimizely: https://www.optimizely.com/optimization-glossary/heatmap/ adresinden alındı

*Heatmap Definition*. (tarih yok). Mayıs 2025 tarihinde from Data to Viz: https://www.data-to-viz.com/graph/heatmap.html adresinden alındı

*İşgücü İstatistikleri, 2024*. (2025, Mart 20). Mayıs 2025 tarihinde TÜİK: https://data.tuik.gov.tr/Bulten/Index?p=Isgucu-Istatistikleri-2024-54059 adresinden alındı

*Jitter Plot*. (tarih yok). Mayıs 2025 tarihinde DVP: https://datavizproject.com/data-type/jitter-plot/#:~:text=A%20jitter%20plot%20is%20a,many%20individual%20one%2Ddimensional%20values adresinden alındı

*Koroplet Harita*. (tarih yok). Mayıs 2025 tarihinde Veri Görselleştirme Kataloğu: https://datavizcatalogue.com/TR/yontemleri/koroplet\_harita.html adresinden alındı

*Koroplet harita (renk tonlu harita) nedir? Tanım, Örnekler ve Araçlar*. (2023, Ekim 31). Mayıs 2025 tarihinde someka: https://www.someka.net/tr/blog/koroplet-harita-nedir/?srsltid=AfmBOoqJ3KfisP6KluON9A-tO8XPlJNAVImvuAJoZa9z15Ig5DNaMuJH#section2 adresinden alındı

*ML | K-means++ Algorithm*. (2025, Mart 11). Mayıs 2025 tarihinde GeeksforGeeks: https://www.geeksforgeeks.org/ml-k-means-algorithm/ adresinden alındı

Pedriquez, D. (2025, Şubat 17). *Choropleth map*. Mayıs 2025 tarihinde Venngage: https://venngage.com/blog/choropleth-map/ adresinden alındı

Seshapanpu, J. (2018). *Students Performance in Exams*. Mayıs 2025 tarihinde Kaggle: https://www.kaggle.com/datasets/spscientist/students-performance-in-exams adresinden alındı

Singh, A. (tarih yok). *How to Create Jitter Plot (Strip Plot) in Tableau*. Mayıs 2025 tarihinde DoingData: https://www.doingdata.org/blog/how-to-create-jitter-plot-strip-plot-in-tableau adresinden alındı

*Visualizing Distributions with Raincloud Plots (and How to Create Them with ggplot2)*. (2021, Haziran 6). Mayıs 2025 tarihinde CedricScherer: https://www.cedricscherer.com/2021/06/06/visualizing-distributions-with-raincloud-plots-and-how-to-create-them-with-ggplot2/ adresinden alındı

Yi, M. (tarih yok). *A complete guide to heatmaps*. Mayıs 2025 tarihinde Atlassian: https://www.atlassian.com/data/charts/heatmap-complete-guide adresinden alındı