

T.C.

FIRAT ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ



## **Literatüre Dayalı İnme Riski Veri Seti Üzerinde Keşifsel Veri Analizi Raporu**

**Zehra İlayda DEMİRCİ**

## 1.GİRİŞ

İnme, beyne giden kan akışının aniden kesilmesiyle birlikte beyin hücrelerinin hasar görmesi ya da ölmesiyle sonuçlanan ciddi bir sağlık problemidir. Genellikle aniden ortaya çıkar ve zamanında müdahale edilmediğinde kalıcı sağlık sorunlarına yol açabilir.

Bu çalışmada, inme riskini etkileyebilecek faktörleri anlamak amacıyla Kaggle üzerinden alınan bir veri seti kullanılmış ve Keşifsel Veri Analizi (Exploratory Data Analysis - EDA) yapılmıştır. EDA sayesinde, veri setindeki değişkenlerin birbirleriyle olan ilişkileri incelenmiş ve bazı çıkarımlar yapılmıştır.

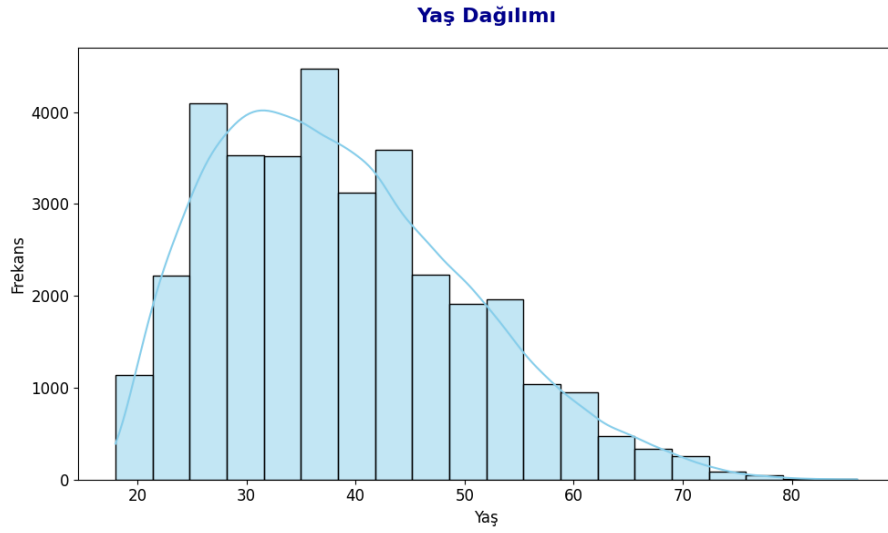
### 1.1 Veri Seti Hakkında Genel Bilgi

Veri seti, inme riskini tahmin etmek için semptomlar, demografik bilgiler ve tıbbi literatüre dayalı risk modellerini içermektedir. Çalışmanın doğruluğunu artırmak amacıyla, yaşa bağlı semptom olasılıkları, cinsiyete özgü risk faktörleri ve tıbbi olarak doğrulanmış özellik mühendisliği uygulanmıştır. Veri setinde 35.000 kayıt bulunmakta olup, her bir kayıta 17 özellik ve 2 hedef değişken bulunmaktadır ayrıca eksik veri bulunmamaktadır.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35000 entries, 0 to 34999
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   age                                   35000 non-null  int64
1   gender                               35000 non-null  object
2   chest_pain                           35000 non-null  int64
3   high_blood_pressure                 35000 non-null  int64
4   irregular_heartbeat                 35000 non-null  int64
5   shortness_of_breath                 35000 non-null  int64
6   fatigue_weakness                    35000 non-null  int64
7   dizziness                           35000 non-null  int64
8   swelling_edema                      35000 non-null  int64
9   neck_jaw_pain                       35000 non-null  int64
10  excessive_sweating                  35000 non-null  int64
11  persistent_cough                    35000 non-null  int64
12  nausea_vomiting                     35000 non-null  int64
13  chest_discomfort                    35000 non-null  int64
14  cold_hands_feet                     35000 non-null  int64
15  snoring_sleep_apnea                 35000 non-null  int64
16  anxiety_doom                        35000 non-null  int64
17  stroke_risk_percentage               35000 non-null  float64
18  at_risk                             35000 non-null  int64
dtypes: float64(1), int64(17), object(1)
memory usage: 5.1+ MB
None
```

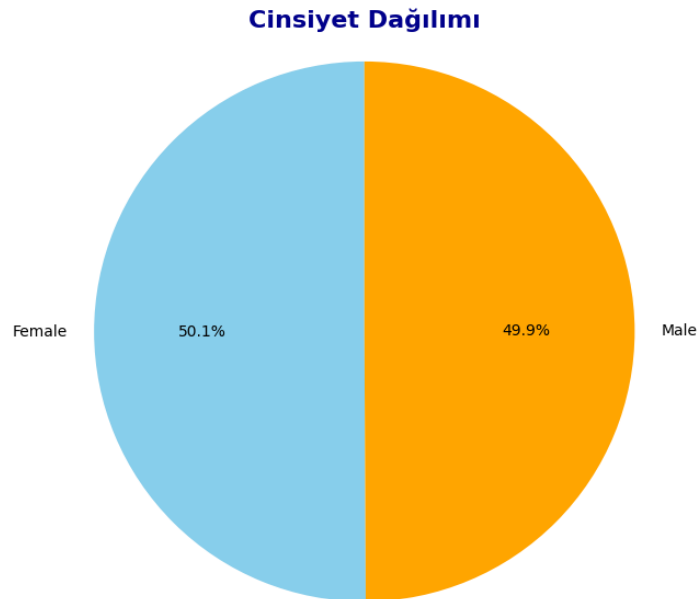
Şekil 1. Veri Setinin Özellikleri

## 2. Keşifsel Veri Analizi



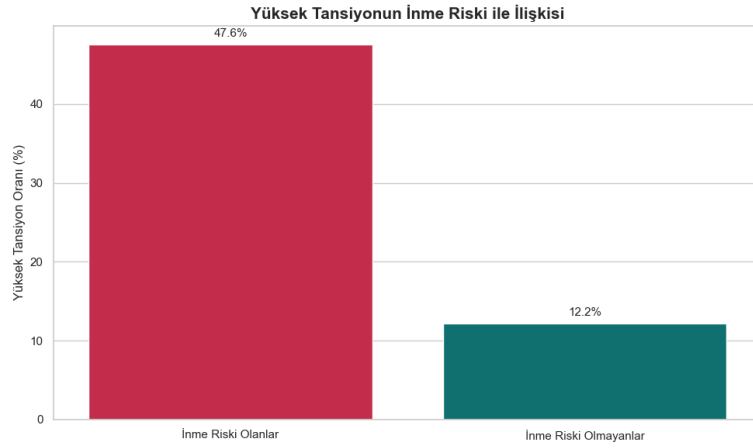
**Şekil 1.** Yaş Dağılımı Grafiği

Şekil 1'deki yaş dağılımı histogramı, veri setindeki katılımcıların yaşlarının belirgin bir şekilde 25 ile 50 yaş aralığında kümelendiğini göstermektedir. Bu aralıkta, özellikle 35-40 yaş bandında en yüksek sayıda birey bulunmaktadır. Dağılımın sağa doğru uzayan kuyruğu, ileri yaşlardaki katılımcıların sayısının azaldığını ancak yine de örneklemde temsil edildiklerini işaret etmektedir. Bu durum, inme riski analizinde yaşın önemli bir faktör olabileceği hipotezini destekleyebilir, zira bu yaş aralığı ve sonrası inme vakalarının görülme sıklığının arttığı bilinen bir dönemdir.



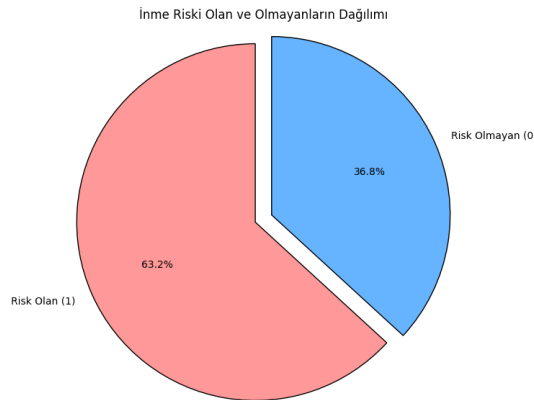
**Şekil 2.** Cinsiyet Dağılımı Grafiği

Veri setindeki cinsiyet dağılımı Şekil 2'de gösterilmektedir. Grafik incelendiğinde, kadın (%50.1) ve erkek (%49.9) katılımcı oranlarının birbirine oldukça yakın olduğu görülmektedir. Bu dengeli dağılım, cinsiyet faktörünün analizlerimizde belirgin bir önyargıya neden olma ihtimalini azaltmaktadır.



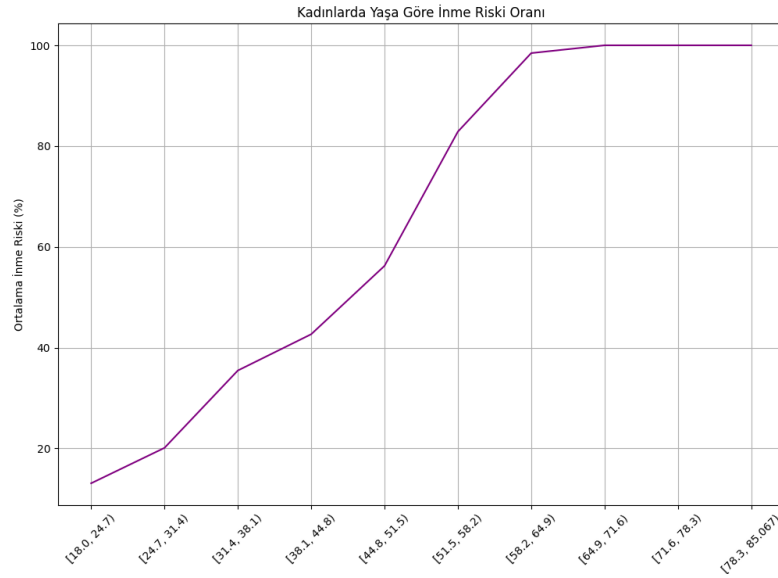
**Şekil 3.** Yüksek Tansiyonun İnme İle İlişkisi Grafiği

Şekil 3'e baktığımızda, yüksek tansiyonun inme riskiyle olan ilişkisini net bir şekilde görebiliyoruz. İnme geçirmiş bireylerin neredeyse yarısında (%47.6) yüksek tansiyon geçmiş bulunurken, inme geçirmeyenlerde bu oran sadece %12.2 civarında. Bu durum, yüksek tansiyonun inme için önemli bir risk faktörü olabileceğini düşündürüyor. Açıkçası, yüksek tansiyonu olan kişilerin inme geçirme olasılığı, olmayanlara göre çok daha yüksek gibi duruyor. Bu bulgu, analizimizin ilerleyen aşamalarında yüksek tansiyon değişkenine daha fazla odaklanmamız gerektiğini gösteriyor.

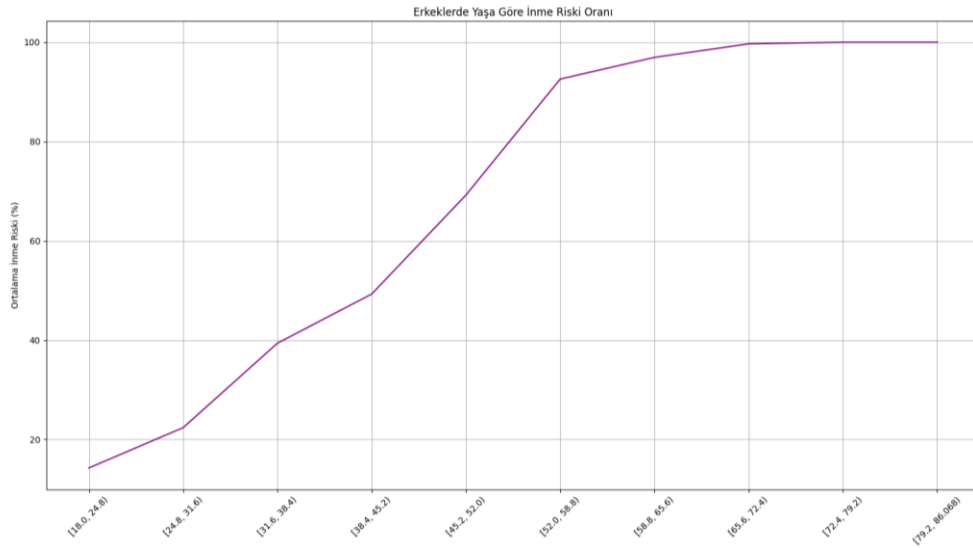


**Şekil 4.** İnme Riski Dağılımı Grafiği

Veri setindeki inme riski olan ve olmayan bireylerin dağılımına baktığımızda, örneklemin önemli bir kısmının inme riski taşıdığını görüyoruz. Pasta grafiğine göre, katılımcıların %63.2'si inme riski altındayken, %36.8'i herhangi bir risk taşımamaktadır. Bu durum, veri setinde inme riski olan bireylerin sayısının, olmayanlardan oldukça fazla olduğunu gösteriyor. Bu dengesiz dağılım, modeli eğitirken ve değerlendirirken dikkate alınması gereken önemli bir nokta olabilir.



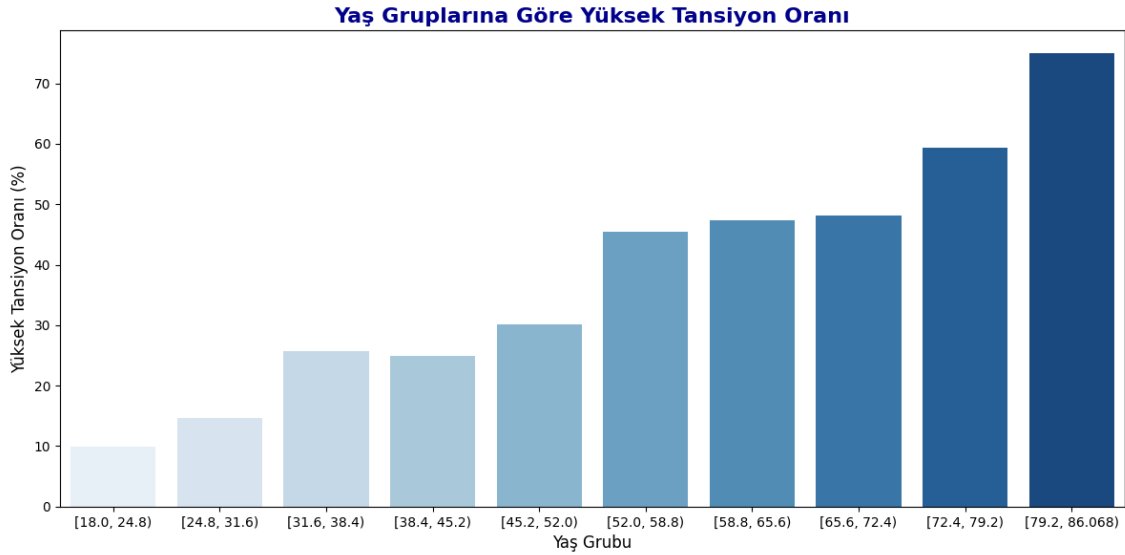
**Şekil 5.** Kadınlarda Yaşa Göre İnme Riski Oranı Grafiği



**Şekil 6.** Erkeklerde Yaşa Göre İnme Riski Oranı Grafiği

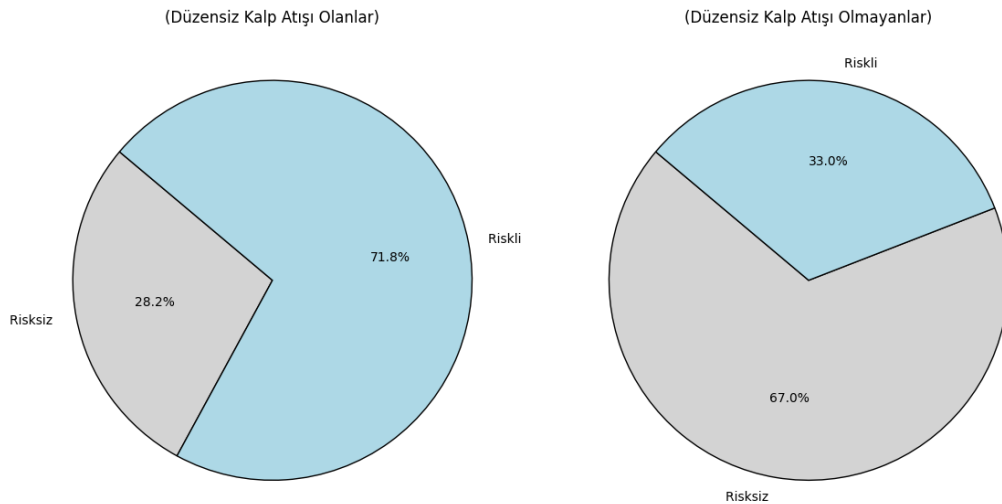
Şekil 5 ve Şekil 6, sırasıyla kadın ve erkek katılımcılarda yaş gruplarına göre ortalama inme riski oranlarını göstermektedir. Her iki grafikte de benzer bir eğilim göze çarpmaktadır:

yaşın ilerlemesiyle birlikte inme riski belirgin bir şekilde artmaktadır. Kadınlarda, genç yaş gruplarında (18-31 yaş aralığı) ortalama inme riski %20'nin altında seyrederken, bu oran 50 yaşından sonra hızla yükselerek 70 yaşın üzerindeki gruplarda %100'e yakın bir seviyeye ulaşmaktadır. Erkeklerde de benzer bir artış trendi izlenmektedir. Genç yaşlarda %20 civarında olan ortalama inme riski, 50 yaş sonrasında ivme kazanarak 70 yaş üzerindeki bireylerde yine %100'e yaklaşmaktadır.



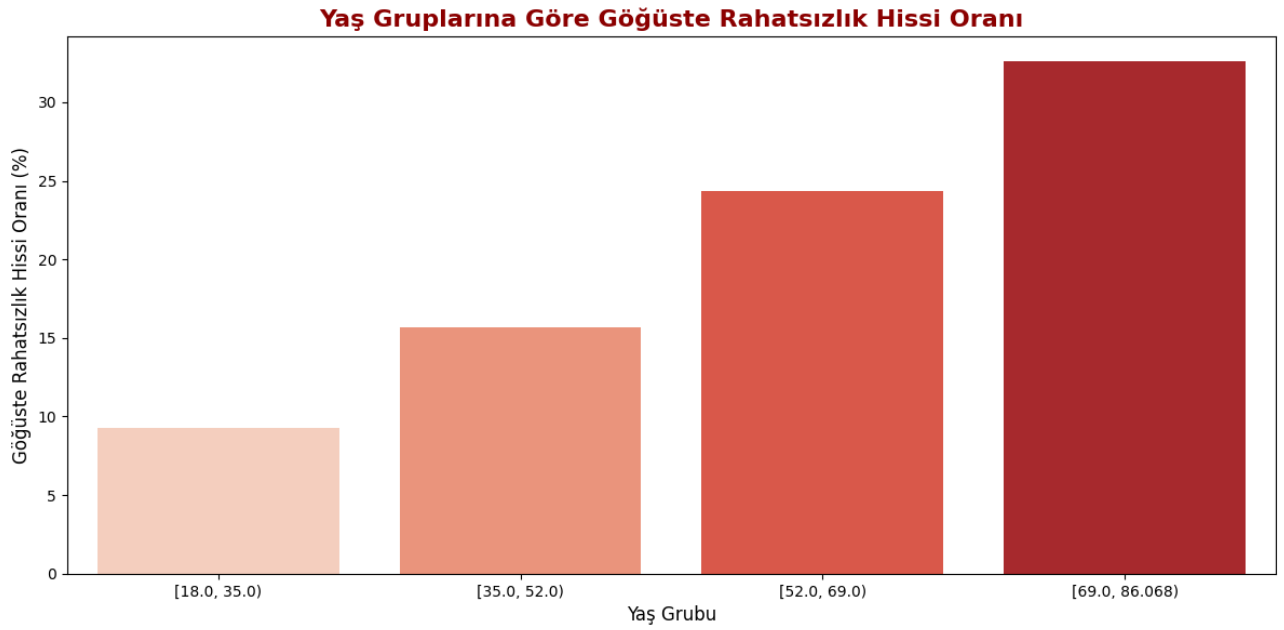
**Şekil 7. Yaş ve Hipertansiyon İlişkisi**

Bu grafikte, yaş gruplarına göre yüksek tansiyon oranları incelenmiştir. Genel olarak yaş ilerledikçe yüksek tansiyon oranının arttığı görülmektedir. Özellikle 52 yaş ve üzerindeki bireylerde oranlar belirgin şekilde yükselmekte, 79 yaş üstü grupta ise %75'e kadar ulaşmaktadır.



**Şekil 8. Düzensiz Kalp Atışının İnme Riskiyle İlişkisi**

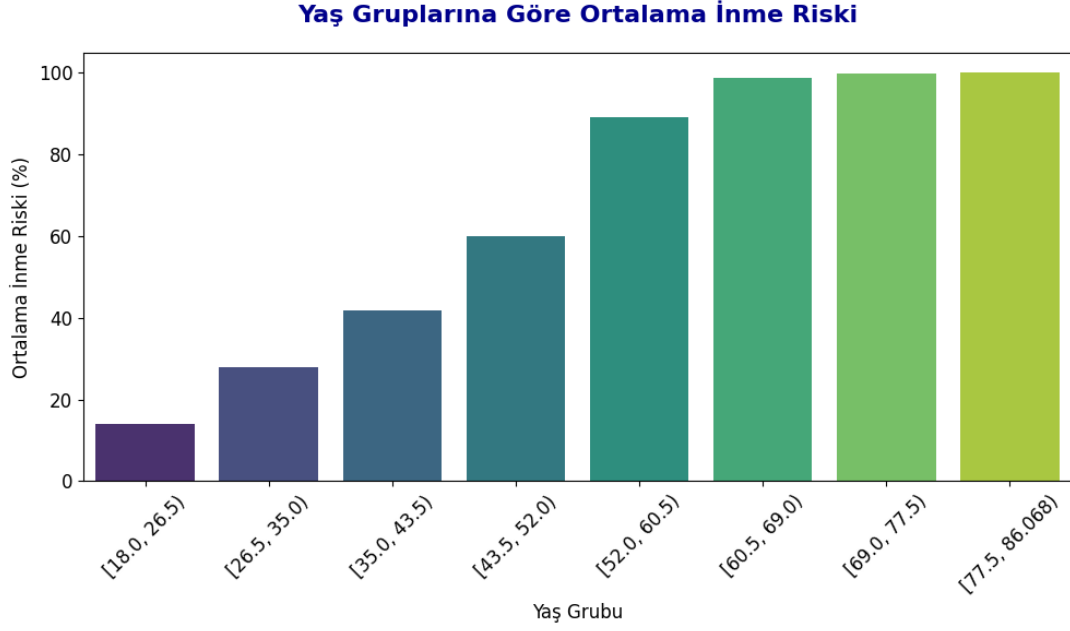
Grafikler, düzensiz kalp atışının inme riskiyle güçlü bir ilişkisi olduğunu göstermektedir. Düzensiz kalp atışı olan bireylerin yaklaşık %72'si inme riski taşırken, bu oran düzensiz kalp atışı olmayanlarda %33 civarındadır. Bu bulgu, düzensiz kalp atışının bir inme riski faktörü olabileceğini düşündürmektedir.



**Şekil 9.** Yaşa Göre Göğüste Rahatsızlık Hissi Oranı Grafiği

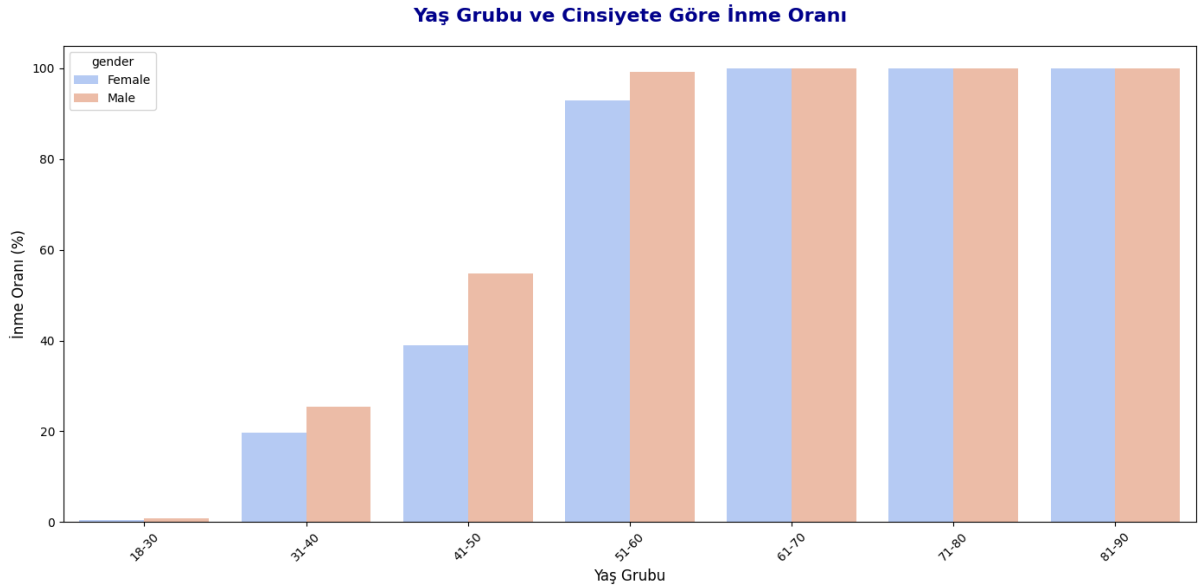
Şekil 9'daki grafik, yaş gruplarına göre göğüste rahatsızlık hissi oranını göstermektedir. Elde ettiğimiz verilere göre, yaşın ilerlemesiyle bu oranda belirgin bir artış gözlemlenmektedir. 18-35 yaş aralığında yaklaşık %9 olan göğüste rahatsızlık hissi, 35-52 yaş aralığında %16'ya, 52-69 yaş aralığında %24'e ve 69 yaş üzeri grupta ise %33'ün üzerine çıkmaktadır.

Bu bulgular, Cambridge Medicine'in "The Stroke Book" kaynağında belirtilen "göğüs ağrısının genç yetişkinlerde nadir olduğu ancak ateroskleroz ve koroner arter hastalığı nedeniyle yaşlı popülasyonlarda yaygınlaştığı" bilgisiyle paralellik göstermektedir. Veri setimizdeki göğüste rahatsızlık hissi oranının yaşla birlikte artması, bu tıbbi tespiti destekler niteliktedir ve inme riski değerlendirmesinde yaş faktörünün önemini bir kez daha vurgulamaktadır.



**Şekil 10.** Yaşa Göre Ortalama İnme Riski

Grafik, yaşın ilerlemesiyle inme riskinin belirgin şekilde arttığını göstermektedir. Risk, genç yaşlarda düşükken, özellikle 52 yaşından sonra hızlanarak 60 yaş üzerinde %90'ın üzerine çıkmaktadır. Bu doğrusal olmayan artış, "Harrison's Principles of Internal Medicine"da belirtilen inme insidansının 55 yaş sonrası her on yılda ikiye katlanması tespitiyle ve modeldeki sigmoid eğri ile uyumludur. Bu durum, yaşın inme riski tahmininde kritik bir faktör olduğunu güçlü bir şekilde desteklemektedir.



**Şekil 11.** Yaşa ve Cinsiyete Göre Ortalama İnme Riski



Önceki analizlerde yaşı inme riski üzerindeki belirleyici rolü vurgulanmıştı. Şekil 10, bu ilişkiyi cinsiyet faktörüyle birlikte ele almakta ve yaş gruplarına göre kadın ve erkeklerdeki inme oranlarını karşılaştırmaktadır.

Genel eğilim, yaş ilerledikçe her iki cinsiyette de inme oranlarının arttığını göstermektedir. Ancak, 18-50 yaş aralığında erkeklerde inme oranlarının kadınlara kıyasla daha yüksek olduğu dikkat çekmektedir. Örneğin, 41-50 yaş grubunda erkeklerde oran yaklaşık %55 iken, kadınlarda %40 civarındadır.

51 yaş ve üzerindeki gruplarda ise inme oranları her iki cinsiyette de artarak benzer seviyelere ulaşmakta, 61 yaş ve sonrasında %90'ın üzerine çıkmaktadır.

Bu durum, yaşı inme riski için baskın bir değişken olduğunu doğrulamakla birlikte, cinsiyetin özellikle genç ve orta yaş gruplarında önemli bir ayrıştırıcı faktör olabileceğini göstermektedir. Bu nedenle, risk analizlerinde yaş ve cinsiyet faktörlerinin birlikte değerlendirilmesi gerekmektedir.

### 3.SONUÇ

Bu keşifsel veri analizi (EDA) çalışması, inme riskini etkileyebilecek faktörleri anlamak amacıyla gerçekleştirilmiştir. Kaggle üzerinden sağlanan kapsamlı veri seti üzerinde uygulanan EDA teknikleri, inme riskiyle ilişkili önemli eğilimleri ve potansiyel biyolojik belirteçleri ortaya koymuştur.

Veri setindeki yaş dağılımı incelendiğinde, örneklemin büyük bir kısmının orta yaş aralığında yoğunlaştığı görülmüştür. Ancak, yaşı ilerlemesiyle inme riskindeki belirgin artış, biyolojik süreçlerle ilişkili yaşlanmanın inme patogenezi üzerindeki rolünü desteklemektedir. Cinsiyet dağılımının dengeli olması, cinsiyetin genel analizler üzerindeki potansiyel etkisini minimize ederken, yaşa göre risk analizinde genç ve orta yaşlardaki erkeklerdeki hafif yüksek risk oranı, cinsiyet hormonları veya genetik yatkınlık gibi biyolojik farklılıkların etkisini düşündürmektedir.

Yüksek tansiyon ve düzensiz kalp atışı gibi mevcut sağlık durumlarının inme riskiyle güçlü korelasyonu, kardiyovasküler sistem sağlığının beyin sağlığı üzerindeki doğrudan

etkisini vurgulamaktadır. Bu faktörlerin biyolojik mekanizmaları, inme riskini öngörmeye önemli biyolojik belirteçler olarak değerlendirilebilir. Yaşla birlikte artan göğüs rahatsızlık hissi oranı ise, ateroskleroz gibi yaşa bağlı biyolojik değişikliklerin inme riskini dolaylı olarak etkileyebileceğini düşündürmektedir.

Sonuç olarak, bu EDA çalışması, demografik faktörlerin yanı sıra temel sağlık göstergelerinin ve yaşa bağlı semptomların inme riskiyle anlamlı ilişkiler sergilediğini göstermiştir. Cinsiyetin de özellikle genç ve orta yaşlarda risk farklılıklarına katkıda bulunabileceği görülmüştür. Elde edilen bu ön bulgular, daha ileri biyoinformatik analizleri ve makine öğrenimi modelleri için değerli bir temel oluşturmaktadır. Elde edilen bu bulgular, inme riskini tahmin etmeye yönelik daha sonraki modelleme çalışmalarına önemli bir girdi sağlayacak ve risk faktörlerinin daha detaylı analiz edilmesinin önemini vurgulayacaktır.