Titanic Veri Kümesi Analiz Raporu

Giriş

Bu analiz, 1912 yılında batmış olan Titanic gemisinin yolcu verilerini inceleyerek hayatta kalma oranlarını etkileyen faktörleri anlamayı amaçlamaktadır. Veri kümesi, yolcuların demografik bilgilerini, bilet bilgilerini ve hayatta kalıp kalmadıklarını içermektedir. Analiz, Pandas, Matplotlib ve Seaborn kütüphaneleri kullanılarak gerçekleştirilmiştir.

Veri Temizleme

Veri setinde bazı eksik ve gereksiz veriler bulunmaktaydı. Bu nedenle, aşağıdaki veri temizleme adımları uygulanmıştır:

Eksik Değerlerin Doldurulması: Age sütunundaki eksik değerler, sütunun medyanı ile doldurulmuştur. Embarked sütunundaki eksik değerler ise mod (en sık görülen değer) ile doldurulmuştur. Cabin sütunundaki eksik değerler çok fazla olduğu için bu sütun analizden çıkarılmıştır.

Veri Dönüşümleri: Sex sütunu, sayısal değerlere dönüştürülmüştür (male = 0, female = 1).

Keşifsel Veri Analizi (EDA)

Temel İstatistikler

Veri kümesindeki sayısal sütunların temel istatistikleri aşağıdaki gibidir:

Ortalama Yaş: 29.7 yıl

Ortalama Bilet Ücreti: 32.2

Hayatta Kalma Oranı: %38.4 (0.384)

Hayatta Kalma Oranı ve Cinsiyet

Cinsiyete göre hayatta kalma oranları incelendiğinde, kadınların hayatta kalma oranının erkeklere göre oldukça yüksek olduğu görülmüştür:

Erkeklerin Hayatta Kalma Oranı: %18.9

Kadınların Hayatta Kalma Oranı: %74.2

Hayatta Kalma Oranı ve Yolcu Sınıfı

Yolcu sınıfı ile hayatta kalma oranı arasındaki ilişki incelendiğinde, birinci sınıf yolcuların hayatta kalma olasılığının daha yüksek olduğu gözlemlenmiştir:

1. Sınıf: %62.96

2. Sınıf: %47.28

3. Sınıf: %24.24

Yaş Dağılımı ve Hayatta Kalma

Yaş dağılımı incelendiğinde, genç yolcuların (özellikle çocukların) hayatta kalma oranlarının daha yüksek olduğu tespit edilmiştir. 10 yaş altı çocukların hayatta kalma oranı %59.4'tür.

Ortalama Değeri En Yüksek Sütun

Veri setindeki sayısal sütunlar arasında ortalama değeri en yüksek olan sütun Fare sütunudur, yani bilet ücretleri diğer sayısal sütunlara göre en yüksek ortalamaya sahiptir.

Görselleştirme

Aşağıda, veri kümesinden elde edilen bazı önemli bulguları görselleştiren grafikler yer almaktadır:

Cinsiyete Göre Hayatta Kalma Oranı:

Kadınların hayatta kalma oranı erkeklere göre çok daha yüksek.

Yolcu Sınıfına Göre Hayatta Kalma Oranı:

Birinci sınıf yolcuların hayatta kalma oranı diğer sınıflara göre daha yüksek.

Yaş Dağılımı ve Hayatta Kalma:

Çocukların hayatta kalma oranı daha yüksek.

Sonuçlar

Bu analiz, Titanic felaketinde hayatta kalma oranını etkileyen başlıca faktörlerin cinsiyet, yolcu sınıfı ve yaş olduğunu ortaya koymuştur. Kadınlar, birinci sınıf yolcular ve çocuklar, hayatta kalma olasılığı daha yüksek olan gruplar olarak öne çıkmıştır. Bu bulgular, Titanic'teki "önce kadınlar ve çocuklar" kuralının gerçekten de uygulandığını göstermektedir.