

# Neden Airline Passenger Satisfaction Veri Setini Seçtim?

**Airline Passenger Satisfaction** veri setini tercih etmemin temel nedeni, **gerçek dünya verisine dayanması** ve analiz açısından yüksek düzeyde **istatistiksel çeşitlilik** sunmasıdır.

Bu veri seti, yolcu memnuniyetini etkileyen;

- Uçuş mesafesi,
- Hizmet kalitesi (Wi-Fi, yemek, eğlence),
- Kabin temizliği ve konfor,
- Uçuş öncesi ve sonrası süreçler

gibi birçok değişkeni içererek **çok boyutlu analiz** yapmaya olanak tanımaktadır.

Ayrıca, bu veri setinin hem **sayısal** hem de **kategorik** değişkenlerden oluşması; **eksik değer analizi**, **aykırı değer tespiti** ve **görselleştirme** gibi süreçleri etkin şekilde gerçekleştirmemi sağlamaktadır.

## Sayısal Değişkenlere Ait İstatistiksel Yorumlar

- Bu havayolu şirketini tercih eden yolcuların **ortalama yaşı 39.43** olup, bu durum **genç ve orta yaş arası** yolculardan oluşan bir müşteri kitlesine işaret etmektedir.
- **Ortalama uçuş mesafesi 1190 km** olup, genellikle **kısa ve orta mesafeli uçuşların** tercih edildiği gözlemlenmiştir.
- Hizmet kalitesine dair verilen puanlar **1 ile 5 arasında** değişmekte olup genel olarak **orta düzeyde memnuniyet** seviyesini göstermektedir.
  - En düşük puan: **Uçak içi Wi-Fi hizmeti** (2.73)
  - En yüksek puanlar: **Uçak içi hizmet (Inflight service)** (3.64) ve **Bagaj taşıma hizmeti (Baggage handling)** (3.63)
- **Kalkış gecikmeleri** ortalama **14.71 dakika**, **Varış gecikmeleri** ise ortalama **15.09 dakika** olarak hesaplanmıştır. Bu değerler, gecikmelerin **makul seviyede** olduğunu göstermektedir.

# Medyan ile Ortalama Farkı

Bazen veri setlerinde yer alan **aykırı değerler**, ortalamayı yanıltıcı hale getirebilir. Bu gibi durumlarda, daha güvenilir ve dengeli bir merkezi eğilim ölçüsü sunması açısından **medyan** tercih edilir.

Medyan ile ortalamanın **aynı ya da benzer** çıktığı durumlar genellikle **veri dağılımının simetrik olduğu** ve **uç değerlerin az** olduğu veri setlerinde gözlemlenir.

Bu veri seti özelinde dikkat çeken nokta, **kalkış ve varış gecikmesi** değişkenleridir. Her iki değişken için de medyan değerin **0.00 dakika** olması, uçuşların **yarısından fazlasında hiç gecikme yaşanmadığını** göstermektedir.

Ancak ortalama gecikme sürelerinin yaklaşık **15 dakika** olması, bazı uçuşlarda **çok yüksek gecikmeler** yaşandığını ve bu değerlerin ortalamayı yukarı çektiğini göstermektedir. Bu durum, ortalama ve medyan arasındaki farkı net şekilde ortaya koyar.

Bu nedenle, **tipik yolcu deneyimini yansıtmak** adına bu tür değişkenlerde **medyan değeri dikkate almak daha sağlıklı** bir analiz yaklaşımı olacaktır.

---

## Minimum Değer Analizi

- Veri setinde yer alan bazı sayısal değişkenlerin minimum değerleri dikkat çekicidir. Özellikle;

### Yolcu Profili:

- Minimum yaş değeri **7** olup, veri setinde çocuk yolcuların da bulunduğunu göstermektedir.

### ✈ Uçuş Özellikleri:

- En kısa uçuş mesafesi yalnızca **31 km** olup, bazı uçuşların çok kısa mesafeli olduğu anlaşılmaktadır.

### ⚠ Hizmet Puanlamaları:

- Inflight Wi-Fi, Food and Drink, Seat Comfort gibi birçok hizmet sütununda minimum değer **0**'dır.

Bu durum iki şekilde yorumlanabilir:

- Yolcular bu hizmetleri hiç almamış olabilir.
- Verilerde eksiklik veya yanlış girilmiş "0" değerleri olabilir.

## Gecikme Değerleri:

- Hem kalkış hem varış gecikmeleri için minimum değer **0**'dır. Bu da çok sayıda uçuşun zamanında gerçekleştiğini göstermektedir.

---

## Aykırı Değerler - Gecikme Süreleri

- Departure Delay in Minutes ve Arrival Delay in Minutes değişkenlerinde göze çarpan **çok büyük değerler** mevcuttur.
- Özellikle:
  - Maksimum kalkış gecikmesi: **1592 dakika**
  - Maksimum varış gecikmesi: **1584 dakika**
- Bu değerler, ortalama ciddi ölçüde yukarı çekmekte ve veri setinde **aykırı değer (outlier)** olarak değerlendirilmektedir.
- Bu nedenle, daha güvenilir analizler yapabilmek için bu uç değerler ya:
  - Veri ön işleme adımıyla filtrelenmeli,
  - Ya da medyan gibi daha sağlam istatistiklerle çalışılmalıdır.

---

## Standart Sapma Analizi

- Standart sapma**, verilerin ortalama etrafında ne kadar dağıldığını gösteren önemli bir ölçüdür.
- Flight Distance değişkeninde standart sapmanın **997.45 km** olması, uçuş mesafelerinin oldukça değişken olduğunu göstermektedir. Bu, bazı uçuşların çok kısa, bazılarının ise oldukça uzun olduğunu ortaya koymaktadır.
- Departure Delay ve Arrival Delay değişkenlerinde standart sapmanın sırasıyla **38.07 dakika** ve **38.47 dakika** olması, az sayıda uçuşta ciddi gecikmeler yaşandığını göstermekte; bu da ortalama gecikme sürelerinin güvenilirliğini azaltmaktadır.
- Hizmet kalitesine dair değişkenlerde (örneğin Inflight wifi service, Seat comfort) standart sapmanın **1.2-1.5** aralığında olması, yolcular arasında bu hizmetlerle ilgili **memnuniyet farklarının** bulunduğuna işaret etmektedir.
- Unnamed: 0 ve ID gibi değişkenlerin ise standart sapma açısından istatistiksel anlamı bulunmamaktadır. Bu değişkenler analiz kapsamı dışında tutulmalıdır.

---

## Eksik Değer Analizi

### Eksik Değerlerin Ele Alınması - Arrival Delay in Minutes

Veri setinde yalnızca Arrival Delay in Minutes değişkeninde eksik değer gözlemlenmiştir (toplam 393 satır). Bu durum, veri setinin genelinde eksik değer sorununun çok sınırlı olduğunu göstermektedir.

Bu eksik deęerleri ele almak için birkaç yaklaşım deęerlendirilmiřtir. Öncelikle, söz konusu deęişkenin ortalaması **15.09 dakika** iken, medyanı **0.00 dakika** olarak hesaplanmıştır. Bu durum, verinin sağa çarpık (positively skewed) dağıldığını ve yüksek uç deęerler içerdiğini göstermektedir. Bu nedenle, ortalama ile doldurma yöntemi veriyi bozabilir.

Bu sebeplerden dolayı, eksik deęerlerin **medyan ile doldurulması** tercih edilmiştir. Böylece verinin dağılım yapısı korunmuş ve analiz sonuçları daha güvenilir hale gelmiştir.

---

## Aykırı Deęer Analizi

---

Aykırı deęer analizinde sadece sayısal deęişkenler ele alınır.

---

## Aykırı Deęer Analizi: Gecikme Deęişkenleri

### Departure Delay in Minutes

- Boxplot grafięine göre, kalkış gecikmesi deęişkeninde çok sayıda aykırı deęer bulunmaktadır.
- Çoęu uçuřta gecikme 0-100 dakika arasında yoğunlařırken, **bazı uçuřlarda gecikme süresi 1000 dakikayı ařmaktadır**.
- Bu uç deęerler, veri setinin ortalamasını yukarı çekmekte ve dağılımı sağa çarpık hale getirmektedir.
- Medyan deęerin 0 olması, uçuřların yarısından fazlasının zamanında gerçekteřtiğini göstermektedir.

### Arrival Delay in Minutes

- Varış gecikmesi deęişkeninde de benzer şekilde çok sayıda uç deęer gözlemlenmiştir.
- Özellikle 1500 dakikaya varan gecikmeler, verinin geneline kıyasla istisnai durumlardır.
- Bu tür uç deęerler, ya gerçek istisnalar (örneęin teknik arıza) ya da olası veri giriş hataları olabilir.
- Aykırı deęerler nedeniyle ortalama ile medyan arasında ciddi fark oluşmakta, bu da medyanın analiz için daha güvenilir olduğunu göstermektedir.

---

**Sonuç olarak**, bu iki deęişkende aykırı deęerlerin etkisi büyüktür ve analiz sürecinde bu deęerler: - Görselleřtirme ve dağılım yorumlarında **ayrıca ele alınmalı**, - Gerekirse analizden **geçici olarak filtrelenmelidir** (örneęin üst sınır > 1000 olanlar).

---



## Flight Distance Değişkeni - Dağılım Analizi

Aşağıdaki histogram ve yoğunluk eğrisi (KDE), yolcuların uçuş mesafesi dağılımını göstermektedir.

- **Uçuş mesafeleri 0 ile ~5000 km** arasında değişmektedir.
- Dağılım **sağa çarpıktır** (right-skewed); bu da çoğu uçuşun kısa ve orta mesafeli olduğunu, az sayıda uzun mesafeli uçuş bulunduğunu göstermektedir.
- Özellikle **600-700 km civarındaki uçuşlar yoğunlukta** olup, bu değer aralığı veri setinde baskın bir rol oynamaktadır.
- **2500 km'nin üzerindeki uçuşlar daha seyrek** olup, veri setinde uç değer (outlier) olarak değerlendirilebilecek çok az sayıda uzun uçuş yer almaktadır.
- Bu durum, uçuş mesafesine dayalı analizlerde **medyan gibi sağlam istatistiklerin** tercih edilmesini gerekli kılmaktadır.
- Ayrıca, uçuş mesafesi ile hizmet kalitesi veya memnuniyet arasında ilişki kurulmak istenirse, bu dağılım yapısının göz önünde bulundurulması önemlidir.