

# WDI Veri Seti Kapsamlı Keşifçi Veri Analizi (EDA) Raporu

## 1. Veri Yapısı ve Setlerin Bütünlüğü

Veri Seti Adı	Dosya Adı	Boyut (Satır)	Ana İçerik	Ana Veri Tipi	Eksik Veri Durumu
Ana Veri (EDA 3)	WDICSV.csv	403.256	Yıllara Göre Sayısal Gösterge Değerleri	Sayısal/Kategorik	Yüksek (Kronik ve Rastgele)
Dipnotlar (EDA 4)	WDIfootnote.csv	863.158	Metodolojik Ek Açıklamalar	Kategorik	Mükemmel (Tam Dolu)
Ülke Meta (EDA 1)	WDICountry.csv	265	Ülke Yapısal Bilgileri	Kategorik/Sayısal	Yüksek (Özellikle Sayısalda)
Gösterge Meta (EDA 5)	WDISeries.csv	1516	Gösterge Tanımları/Metodolojisi	Kategorik	Yüksek (Kritik Alanlarda)
Ülke-Seri (EDA 2)	WDIcountry-series.csv	8086	Ülke-Seri Eşleşmeleri	Kategorik	Mükemmel (Tam Dolu)
Seri-Zaman (EDA 6)	WDIservices-time.csv	142	Serilerin Zamana Bağlı Dipnotları	Kategorik	Mükemmel (Tam Dolu)

## 2. KRİTİK BULGU: ANA VERİ KALİTESİ VE ETİK KISITLAMALAR (ASA B.1, B.2)

### 2.1. Yüksek Eksik Veri Oranı ve Zaman Serisi Kısıtlaması (EDA 3)

#### Bulgu:

Ana veri seti (WDICSV.csv) yüksek oranda eksik veri içermektedir. Özellikle 1960–1969 döneminde eksiklik oranı çok yüksektir ve non-null gözlem sayısı 40.000’in altındadır. Ayrıca 2022–2024 gibi güncel yıllarda dahi veri raporlama eksiklikleri gözlemlenmektedir.

**Etik Not:**

Analiz öncesinde uygun bir yıl aralığı belirlenmeli ve bu aralık dışında kalan büyük ölçüde boş sütunlar çıkarılmalıdır. Eksik verilerin silinmesi veya doldurulması durumunda, sonuçların genelleştirilebilirliğinin sınırlı olduğu açıkça raporlanmalıdır.

**2.2. Uç Değerler ve Ölçeklendirme Sorunları (EDA 3)****Bulgu:**

Veri setinde  $3.08 \times 10^{21}$  ve  $1.09 \times 10^{22}$  gibi aşırı büyük pozitif değerler ile  $2.59 \times 10^{14}$  gibi negatif minimum değerler tespit edilmiştir. Bu durum, veri tutarlılığı ve ölçeklendirme açısından ciddi sorunlara işaret etmektedir.

**Etik Not:**

Bu uç değerler, veri girişi hataları veya ölçü birimi farklılıklarından kaynaklanıyor olabilir. Analiz öncesinde doğrulanmaları, dönüştürülmeleri veya dışlanmaları gerekmektedir.

**3. META VERİ KALİTESİ VE BÖLGESEL / KONUSAL DENGELER (ASA B.7)****3.1. Metodolojik Bilgi Eksikliği (EDA 5)****Bulgu:**

Meta veri setinde (WDISeries.csv) Short definition sütunu tamamen, Base Period sütunu %94,8 ve Notes from original source sütunu %92,6 oranında eksiktir. Bu sütunlar göstergelerin metodolojik arka planını tanımlamaktadır.

**Etik Not:**

Metodolojik bilgilerin büyük ölçüde eksik olması, göstergelerin sınırlamalarının bilinmediği anlamına gelmekte ve karşılaştırmalı analizlerin güvenilirliğini azaltmaktadır.

**3.2. Bölgesel ve Konusal Temsil Dengesizliği (EDA 1, EDA 5)****Bulgu:**

Gözlemler ağırlıklı olarak Europe & Central Asia bölgesinde yoğunlaşırken, North America ve South Asia gibi bölgelerde gözlem sayısı düşüktür. Ayrıca gösterge konularında ekonomik göstergelerin belirgin biçimde baskın olduğu görülmektedir.

**Etik Not:**

Bu bölgesel ve konusal dengesizlikler, sonuçların genellelenebilirliğini sınırlandırmakta ve bulguların bağlamsal olarak yorumlanmasını gerekli kılmaktadır.

## 2. Kategorik ve İlişkisel Analiz Özetleri

EDA Raporu	Analiz Edilen Set	Önemli Kategorik Bulgular	İlişkisel Bulgular (Korelasyon/Çapraz Tablo)
EDA 1	WDI Ülke Meta	2015 sonrası veri yoğunluğu ve bölgesel eşitsizlikler.	<b>Latest industrial data – Latest trade data</b> arasında zayıf-orta düzeyde pozitif ilişki (0.37).
EDA 2	Ülke-Seri Kodları	Ülke kodları ve seri kodları arasında varyasyon/yoğunluk gözlemlendi.	Hangi ülke kodlarının hangi seri kodları ile en sık eşleştiği belirlendi.
EDA 4	Dipnotlar	Bazı ülkelerin/serilerin/yılların dipnotlarda baskın olması.	Hangi ülkelerin hangi göstergeler için daha sık dipnot gerektirdiğini gösteren haritalama yapıldı.
EDA 5	Gösterge Metodolojisi	Metodolojik standartlar ve yasal kullanım kısıtlamaları hakkında bilgi edinildi.	Hangi ana konuların (Topic) hangi gösterge kodları ile en sık eşleştiği analiz edildi.

## 3. Nihai Değerlendirme ve İleri Adımlar (ASA C.4)

Tüm EDA'lar, WDI veri setinin büyük, zengin, ancak **yüksek oranda dağınık ve eksik veri içeren** bir yapıya sahip olduğunu ortaya koymuştur. Meta veri katmanları eksiksiz olsa da, kritik metodolojik sütunlar ve asıl zaman serisi verileri kronik eksiklikler içermektedir.

Analizin bir sonraki aşamasına geçmeden önce:

1. **Zaman Aralığı Belirleme:** Analiz hedefine göre bir yıl aralığı seçilmeli ve bu aralık dışındaki boş yıllar veri setinden çıkarılmalıdır.
2. **Uç Değer Yönetimi:** Ana veri setindeki aşırı uç değerler incelenmeli ve normalleştirme/dönüştürme stratejileri belirlenmelidir.
3. **Etik İfşaat:** Tüm analiz sonuçlarının, özellikle eksik verilerden ve bölgesel/konusal dengesizliklerden dolayı **filtrelenmiş alt küme için geçerli olduğu ve genel tüm veri setine genellenemeyeceği** şeffaflıkla belirtilmelidir.