

# Amazon Ürün Yorumları Veri Analizi ve Görselleştirme Raporu

## 1. Kullanılan Araçlar ve Yöntemler

- Pandas: Veri manipülasyonu ve analiz işlemleri için
- Matplotlib & Seaborn: Grafik oluşturmak ve görselleştirme için
- Python: Veri analizi ortamı
- Temel NLP ön işlemleri: Metin temizliği ve ön işleme

## 2. Veri Kümesi Tanıtımı

Veri seti, Amazon ürünlerine ait yaklaşık 3.6 milyon etiketlenmiş yorumu içermektedir. Bu veri setinde eksik değerler bulunmamaktadır.

## 3. Veri Ön İşleme

Veri ön işleme adımlarında aşağıdaki işlemler uygulanmıştır:

- Fazla boşlukların kaldırılması
- Tüm metinlerin küçük harfe çevrilmesi
- Noktalama işaretlerinin silinmesi
- İngilizce stop-words temizliği

## 4. Temel İstatistikler

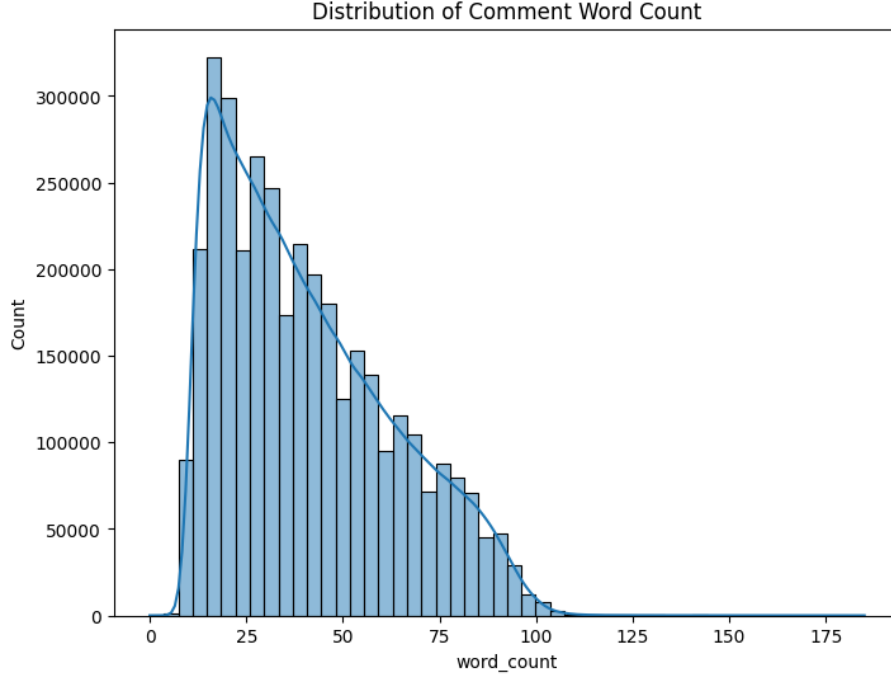
Temizlenen yorum metinlerinden temel özellikler çıkarılmıştır. Kullanılan yöntemler:

- Kelime sayısı:** Her yorumdaki toplam kelime sayısı hesaplandı
- Karakter sayısı:** Yorumların uzunluklarını anlamak için karakter sayısı alındı
- Cümle sayısı:** Basit şekilde nokta sayısına (+1) bakarak cümle sayısı tahmin edildi

## 5. Görselleştirme ve Bulgular

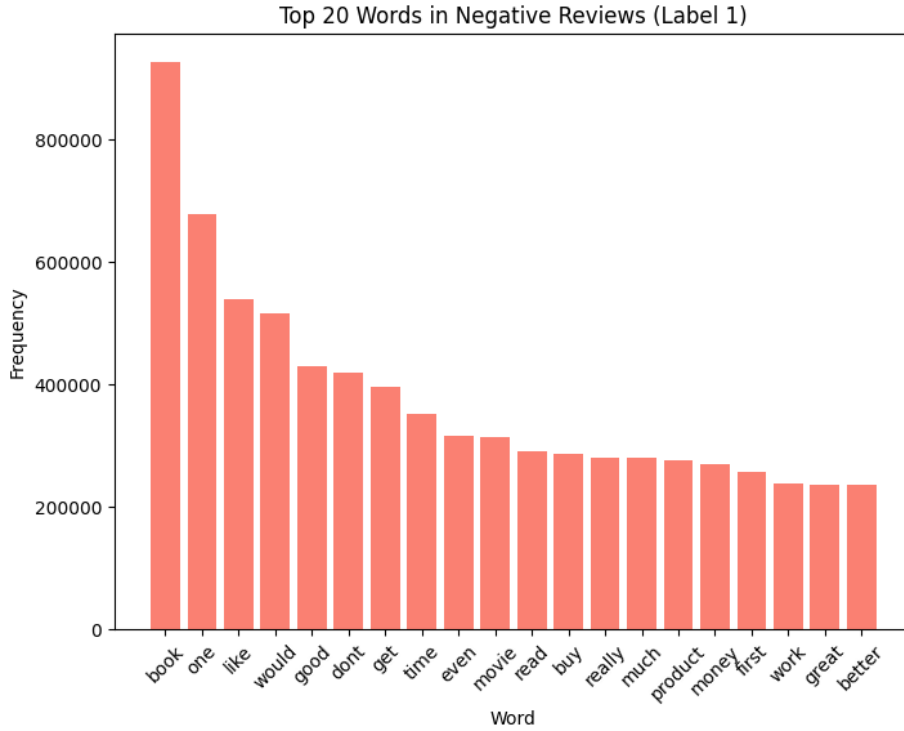
## 5.1. Yorum Uzunluğu Dağılımı

Aşağıdaki histogram, yorumların kelime sayılarına göre dağılımını göstermektedir. Çoğu yorumun 20–30 kelime civarında yoğunlaştığı, uzun yorumların ise daha seyrek olduğu görülmektedir:



## 5.2. Negatif Yorumlarda En Çok Kullanılan 20 Kelime

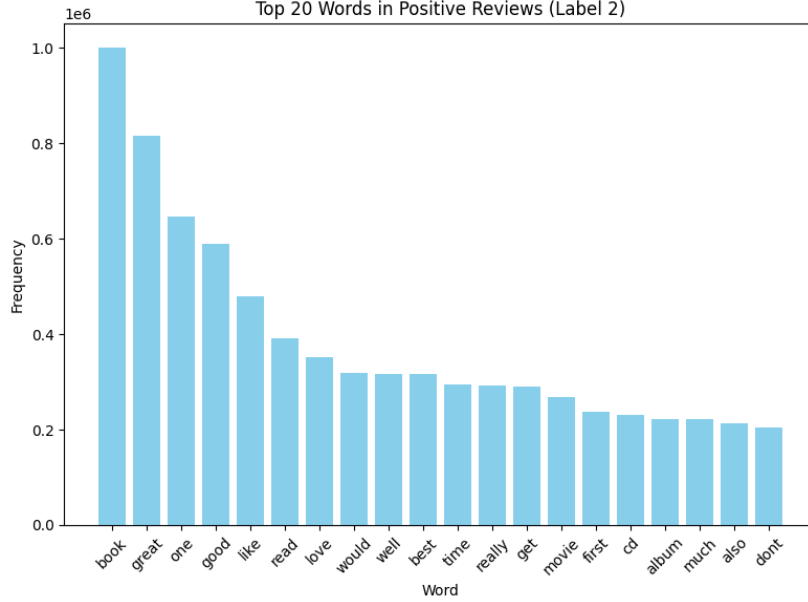
Negatif yorumlarda öne çıkan 20 kelime ise şu şekildedir



Negatif yorumlarda da “book” gibi kelimeler sıkça yer almakla birlikte, “dont”, “money”, “better”, “product” gibi problem vurgusu taşıyan sözcükler dikkat çekmektedir. Bu da kullanıcıların şikâyet odaklı dil kullandığını göstermektedir.

### 5.3. Pozitif Yorumlarda En Çok Kullanılan 20 Kelime

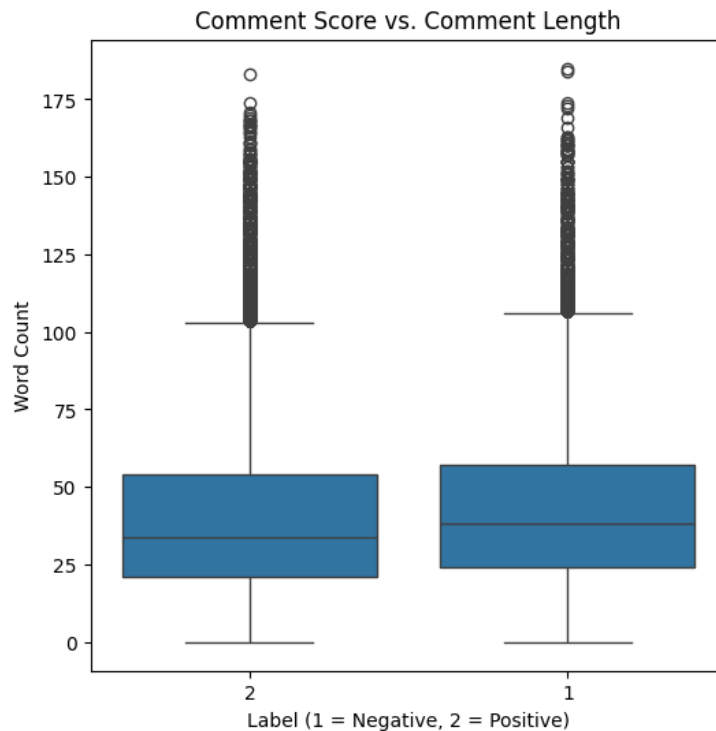
Pozitif yorumlarda öne çıkan 20 kelime ise şu şekildedir



“book”, “great”, “good”, “love” gibi olumlu çağrışımlar yapan kelimeler pozitif yorumlarda sıkça yer almıştır. Bu durum, veri setinin tutarlılığını desteklemektedir.

### 5.4. Yorum Puanı ve Yorum Uzunluğu İlişkisi

Aşağıdaki boxplot, pozitif ve negatif yorumlarda kelime sayılarını karşılaştırmaktadır



Hem negatif hem de pozitif yorumlarda ortalama kelime sayılarının benzer dağıldığı, negatif yorumlarda hafifçe daha düşük bir medyan değeri olduğu gözlenmektedir.