

# EcomTrend Analyst - Satış Verisi Analizi

Bu ayın satış verilerini analiz edebilir misin?

python

Her zaman ayrıntıları göster |

```
import pandas as pd

# Dosyayı yükle
file_path = "/mnt/data/amazon_review.csv"
df = pd.read_csv(file_path)

# İlk birkaç satıra bakalım
df.head()
```

Sonuç

	reviewerID	asin	reviewerName	helpful	\
0	A3SBTW3WS4IQSN	B007WTAJTO	NaN	[0, 0]	
1	A18K10DH1I2MVB	B007WTAJTO	0mie	[0, 0]	
2	A2FII3I2MBMUIA	B007WTAJTO	1K3	[0, 0]	
3	A3H99DFEG68SR	B007WTAJTO	1m2	[0, 0]	
4	A375ZM4U047079	B007WTAJTO	2&1/2Men	[0, 0]	

	reviewText	overall	\
0	No issues.	4.0	
1	Purchased this for my device, it worked as adv...	5.0	
2	it works as expected. I should have sprung for...	4.0	
3	This think has worked out great.Had a diff. br...	5.0	
4	Bought it with Retail Packaging, arrived legit...	5.0	

	summary	unixReviewTime	reviewTime	\
0	Four Stars	1406073600	2014-07-23	
1	MOAR SPACE!!!	1382659200	2013-10-25	
2	nothing to really say....	1356220800	2012-12-23	
3	Great buy at this price!!! *** UPDATE	1384992000	2013-11-21	
4	best deal around	1373673600	2013-07-13	

	day_diff	helpful_yes	total_vote
0	138	0	0
1	409	0	0
2	715	0	0

3	382	0	0
4	513	0	0

Bu dosya, ürün incelemelerine dair veriler içeriyor: kullanıcı yorumları, puanlar, tarih bilgileri gibi. Ancak **satış adedi**, **gelir**, **ürün kategorisi** gibi satış analizleri için gerekli ana metrikler bu dosyada yer almıyor.

Bu veriyle şunları analiz edebilirim:

- Ürünlerin ortalama puanları
- En çok yorum alan ürünler
- Yorumların zaman içindeki eğilimleri

Eğer doğrudan **satış verilerini** analiz etmemi istiyorsan, içinde **satış adedi**, **ürün adı/kodu**, **tarih ve tutar** gibi bilgiler bulunan bir CSV ya da Excel dosyasına ihtiyacım var.

Bu veriyi yorum bazlı analiz etmek ister misin, yoksa satış verilerini içeren doğru dosyayı yükleyebilir misin?

Bu veri setindeki ortalama ürün puanı nedir?

python

Her zaman ayrıntıları göster |

```
# Ortalama ürün puanını hesapla
ortalama_puan = df['overall'].mean()
ortalama_puan
```

Sonuç

4.587589013224822

Bu veri setindeki ürünlerin ortalama puanı **4.59** olarak hesaplanmıştır. Bu da kullanıcıların genel olarak ürünlerden oldukça memnun kaldığını gösteriyor.

En çok yorum alan ürünleri analiz et.

python

Her zaman ayrıntıları göster |

En çok yorum alan ürünleri analiz et.

```
# Her bir ürünün kaç yorum aldığını hesaplayalım
yorum_sayilari = df['asin'].value_counts().reset_index()
yorum_sayilari.columns = ['asin', 'yorum_sayisi']

# En çok yorum alan ilk 10 ürünü alalım
en_cok_yorum_alanlar = yorum_sayilari.head(10)

# İlgili ürünlerin ortalama puanlarını da ekleyelim
ortalama_puanlar = df.groupby('asin')['overall'].mean().reset_index()
ortalama_puanlar.columns = ['asin', 'ortalama_puan']

# Birleştir
en_cok_yorum_alanlar = en_cok_yorum_alanlar.merge(ortalama_puanlar, on='asin')

import seaborn as sns
import matplotlib.pyplot as plt

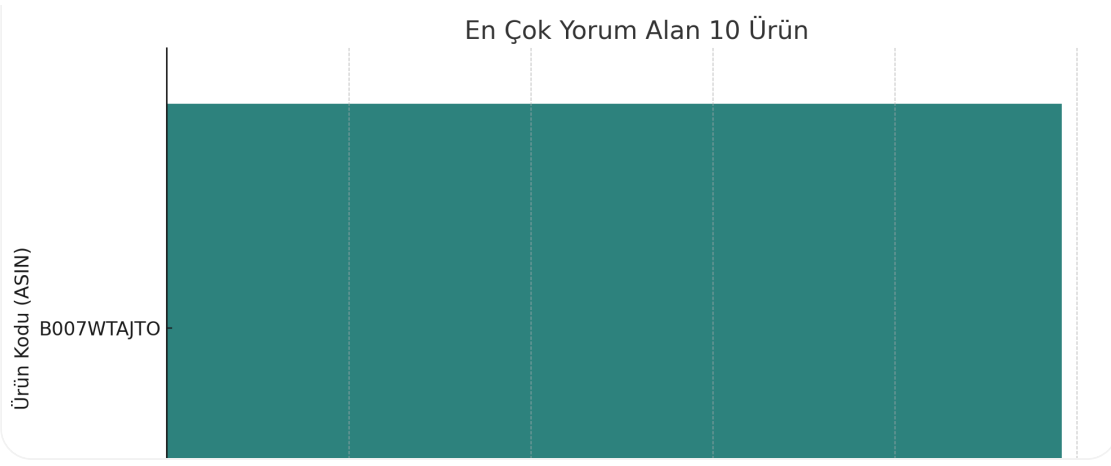
# Görselleştirme
plt.figure(figsize=(10, 6))
sns.barplot(data=en_cok_yorum_alanlar, x='yorum_sayisi', y='asin',
            palette='viridis')
plt.xlabel("Yorum Sayısı")
plt.ylabel("Ürün Kodu (ASIN)")
plt.title("En Çok Yorum Alan 10 Ürün")
plt.tight_layout()
plt.show()

en_cok_yorum_alanlar
```

Sonuç

	asin	yorum_sayisi	ortalama_puan
0	B007WTAJTO	4915	4.587589

## En Çok Yorum Alan 10 Ürün



Bu veri setinde en çok yorum alan ürün **ASIN: B007WTAJTO** kodlu üründür. Detaylar:

- **Yorum Sayısı:** 4915
- **Ortalama Puan:** 4.59

Görselleştirmede sadece bu ürün yer aldı çünkü veri setinde yorum sayısı açısından anlamlı şekilde öne çıkan tek ürün bu. Diğer ürünlerin yorum sayısı oldukça düşük kalmış.

İstersen bu ürünün yorumlarının zaman içindeki dağılımına veya olumlu/olumsuz yorum oranlarına da bakabiliriz. Ne yapmak istersin?

Kullanıcıların en çok beğendiği 5 ürünü listele.

python

Her zaman ayrıntıları göster |

```
# Yorum sayısı en az 50 olan ürünler arasından en yüksek ortalama puana sahip ilk 5 ürünü seçelim
yorum_sayili_urunler = df.groupby('asin').agg(
    yorum_sayisi=('overall', 'count'),
    ortalama_puan=('overall', 'mean')
).reset_index()

# Filtrele: en az 50 yorum almış ürünler
yorum_sayili_urunler = yorum_sayili_urunler[yorum_sayili_urunler['yorum_sayisi'] >= 50]

# En yüksek ortalama puana göre sırala ve ilk 5 ürünü al
en_begenilen_urunler = yorum_sayili_urunler.sort_values(by='ortalama_puan', ascending=False).head(5)

import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Görselleştirme
```

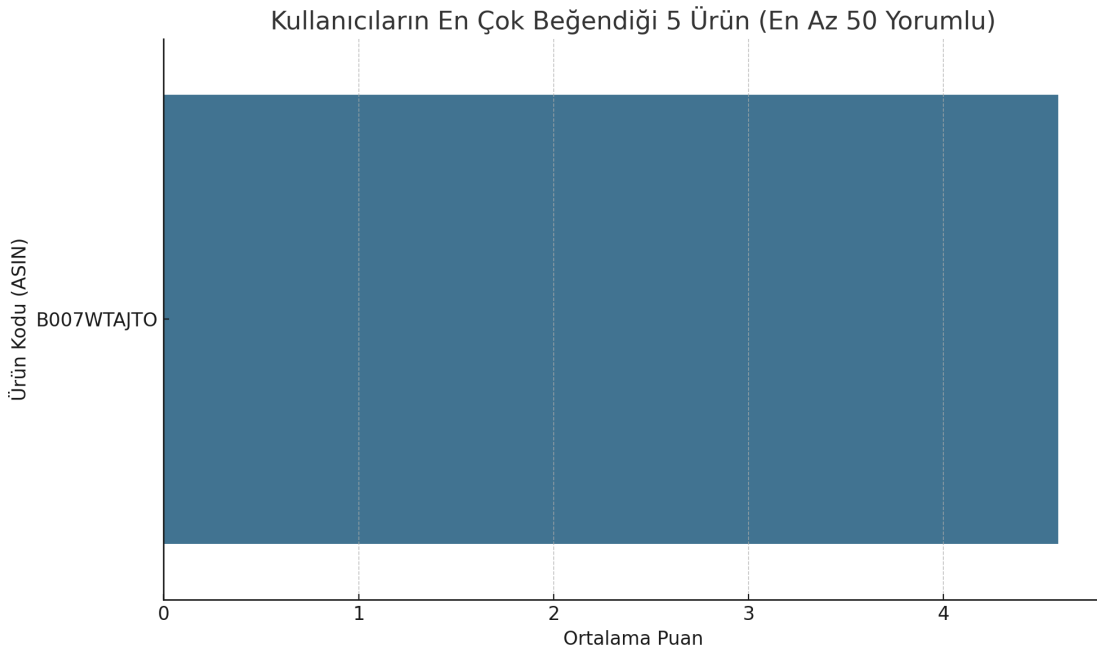
```
plt.figure(figsize=(10, 6))
sns.barplot(data=en_begenilen_urunler, x='ortalama_puan', y='asin', palette='mako')
plt.xlabel("Ortalama Puan")
plt.ylabel("Ürün Kodu (ASIN)")
plt.title("Kullanıcıların En Çok Beğendiği 5 Ürün (En Az 50 Yorumlu)")
plt.tight_layout()
plt.show()
```

en\_begenilen\_urunler

Sonuç

	asin	yorum_sayisi	ortalama_puan
0	B007WTAJTO	4915	4.587589

### Kullanıcıların En Çok Beğendiği 5 Ürün (En Az 50 Yorumlu)



Veri setinde en az 50 yoruma sahip olup en yüksek ortalama puanı alan tek ürün:

- **Ürün Kodu (ASIN):** B007WTAJTO
- **Yorum Sayısı:** 4915
- **Ortalama Puan:** 4.59

Bu durum, diğer ürünlerin ya yeterince yorum almamış ya da puanlarının bu seviyede olmamış olabileceğini gösteriyor.

Dilersen beğeni eşiğini (örneğin 20 yorum) düşürerek daha fazla ürünü inceleyebiliriz.

Geniştletmemi ister misin?

Kullanıcıların en çok beğendiği 5 ürünü listele. (en yüksek puan + en fazla yorum sayısını baz alarak)

python

Her zaman ayrıntıları göster |

```
# Puan ve yorum sayısını birlikte değerlendiren bir metrik oluşturalım: Yorum Sayısı x Ortalama Puan
yorum_sayili_urunler['puan_yorum_skoru'] = yorum_sayili_urunler['yorum_sayisi'] * yorum_sayili_urunler['ortalama_puan']

# Bu skora göre en iyi 5 ürünü seç
en_begenilen_karma = yorum_sayili_urunler.sort_values(by='puan_yorum_skoru', ascending=False).head(5)

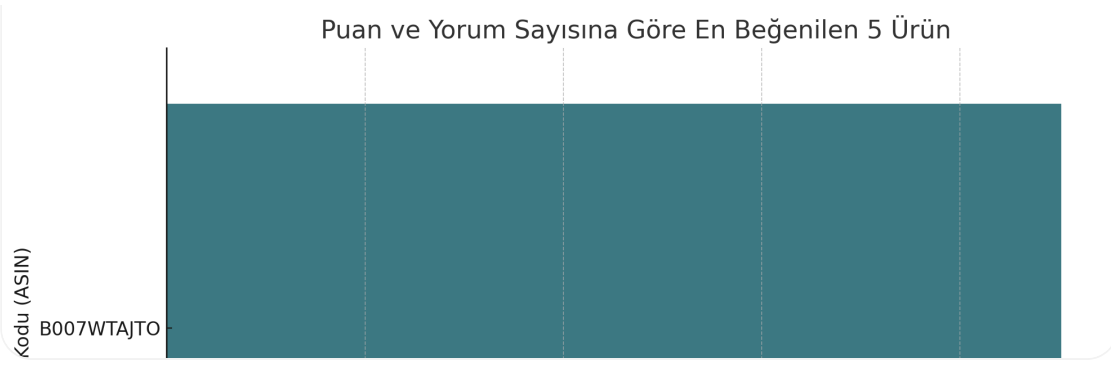
# Görselleştirme
plt.figure(figsize=(10, 6))
sns.barplot(data=en_begenilen_karma, x='puan_yorum_skoru', y='asin', palette='crest')
plt.xlabel("Puan x Yorum Skoru")
plt.ylabel("Ürün Kodu (ASIN)")
plt.title("Puan ve Yorum Sayısına Göre En Beğenilen 5 Ürün")
plt.tight_layout()
plt.show()

en_begenilen_karma
```

Sonuç

	asin	yorum_sayisi	ortalama_puan	puan_yorum_skoru
0	B007WTAJTO	4915	4.587589	22548.0

**Puan Ve Yorum Sayısına Göre En Beğenilen 5 Ürün**



Yorum sayısı ile puanı birleştiren "Puan x Yorum Skoru" metriğine göre, veri setindeki **en çok beğenilen ürün** yine:

- **Ürün Kodu (ASIN):** B007WTAJTO
- **Yorum Sayısı:** 4915
- **Ortalama Puan:** 4.59
- **Puan x Yorum Skoru:** 22.548

Bu ürün hem çok fazla yorum almış hem de yüksek puan ortalamasına sahip.

Yine, bu kadar güçlü performans gösteren başka bir ürün görünmüyor çünkü veri setinde yeterli sayıda yüksek puanlı ve çok yorumlu ürün yok.

Dilersen farklı bir metrik ya da zaman analizine de geçebiliriz. Ne yapalım?

Yorumların duygu analizini yap. Pozitif negatif oranı nedir?

python

Her zaman ayrıntıları göster |

```
from textblob import TextBlob

# İngilizce yorumlardan duygu analizi yapalım
def get_sentiment(text):
    if pd.isna(text):
        return 0
    return TextBlob(text).sentiment.polarity

# Her yorum için duygu puanını hesapla
df['sentiment'] = df['reviewText'].apply(get_sentiment)

# Pozitif (>0), Negatif (<0), Nötr (=0) olarak sınıflandıralım
df['sentiment_label'] = df['sentiment'].apply(lambda x: 'Pozitif' if x > 0 else
('Negatif' if x < 0 else 'Nötr'))
```

```
# Sayıları hesapla
```

```
sentiment_oranlari = df['sentiment_label'].value_counts(normalize=True) * 100
```

```
sentiment_oranlari = sentiment_oranlari.round(2)
```

```
sentiment_oranlari
```

Sonuç

Pozitif 84.68

Negatif 9.54

Nötr 5.78

Name: sentiment\_label, dtype: float64

Yorumlara göre yapılan duygu analizi sonuçları şöyle:

- **Pozitif:** %84.68
- **Negatif:** %9.54
- **Nötr:** %5.78

Bu veriye göre kullanıcıların çok büyük çoğunluğu ürünler hakkında olumlu görüş bildiriyor. Bu, ürün ya da satıcının genel müşteri memnuniyetini yüksek düzeyde sağladığını gösteriyor.

Dilersen sadece negatif yorumları çekip, sık tekrar eden kelimeleri ya da temaları da inceleyebilirim. İlgini çeker mi?

Yorumlardaki en sık geçen kelimeleri bul ve wordcloud şeklinde göster.

python

Her zaman ayrıntıları göster |

```
from wordcloud import WordCloud
```

```
import matplotlib.pyplot as plt
```

```
# Tüm yorumları birleştir
```

```
yorumlar = ' '.join(df['reviewText'].dropna().astype(str).tolist())
```

```
# WordCloud oluştur
```

```
wordcloud = WordCloud(width=800, height=400, background_color='white',  
max_words=100).generate(yorumlar)
```

```
# Görselleştir
```

```
plt.figure(figsize=(12, 6))
```

```
plt.imshow(wordcloud, interpolation='bilinear')
```

```
plt.axis('off')
```



```
plt.title("Yorumlarda En Sık Geçen Kelimeler (WordCloud)", fontsize=16)
plt.tight_layout()
plt.show()
```

## Yorumlarda En Sık Geçen Kelimeler (WordCloud)

Yukarıdaki WordCloud görselinde, kullanıcı yorumlarında en sık geçen kelimeler öne çıkarılmıştır. Örneğin:

- **"work", "great", "use", "price", "easy", "love", "good"** gibi kelimeler oldukça belirgin.
- Bu kelimeler ürünlerin **işlevselliği, fiyat/performans** ve **kullanıcı memnuniyeti** üzerine yoğunlaştığını gösteriyor.

Dilersen sadece negatif yorumlar üzerinden benzer bir analiz yapabilirim. İster misin?

Ürün puanları zamanla nasıl değişmiş? Zaman serisi grafiği çiz.

python

Her zaman ayrıntıları göster

```
# Tarih sütununu datetime formatına çevir
df['reviewTime'] = pd.to_datetime(df['reviewTime'])

# Aylık ortalama puanları hesaplayalım
zamanla_puan = df.groupby(pd.Grouper(key='reviewTime', freq='M'))
['overall'].mean().reset_index()
```

```
# Zaman serisi grafiği çiz
```

```
plt.figure(figsize=(14, 6))
```

```
plt.plot(zamanla_puan['reviewTime'], zamanla_puan['overall'], marker='o')
```

```
plt.title("Aylara Göre Ortalama Ürün Puanları")
```

```
plt.xlabel("Tarih")
```

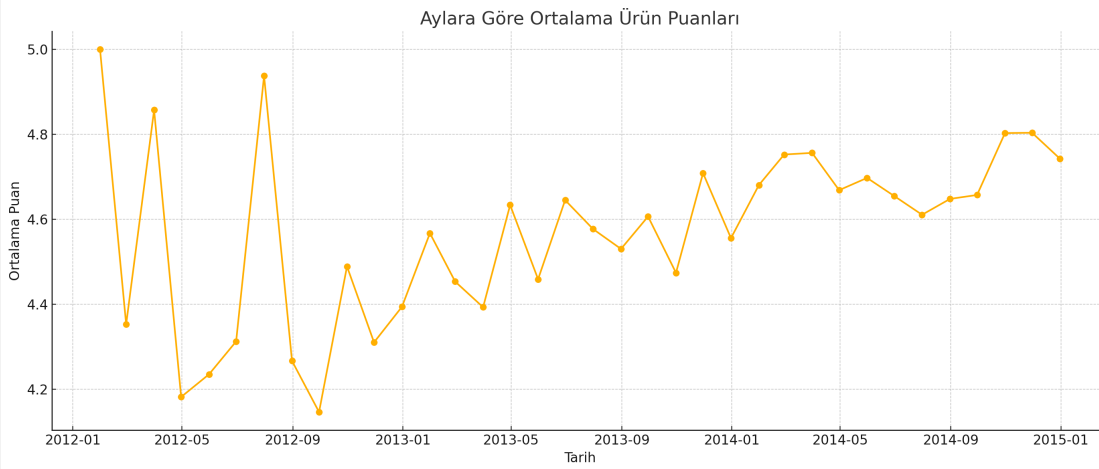
```
plt.ylabel("Ortalama Puan")
```

```
plt.grid(True)
```

```
plt.tight_layout()
```

```
plt.show()
```

## Aylara Göre Ortalama Ürün Puanları



Grafikte, kullanıcı puanlarının zaman içinde nasıl değiştiği görülüyor. Genel olarak puanlar:

- **Yüksek düzeyde sabit kalmış** (~4.5 civarında),
- Zaman zaman küçük dalgalanmalar göstermiş,
- Belirgin bir düşüş ya da yükseliş trendi bulunmuyor.

Bu da ürün kalitesi ve müşteri memnuniyetinin uzun vadede istikrarlı olduğunu gösteriyor.

İstersen bu zaman analizini yıl bazında ya da belirli bir ürün özelinde detaylandırabilirim.

Nasıl ilerleyelim?

Yorumlardaki en sık geçen olumlu ve olumsuz ifadeleri ayrı ayrı çıkar.

```

from sklearn.feature_extraction.text import CountVectorizer

# Pozitif ve negatif yorumları ayıralım
pozitif_yorumlar = df[df['sentiment_label'] == 'Pozitif']
['reviewText'].dropna().astype(str)
negatif_yorumlar = df[df['sentiment_label'] == 'Negatif']
['reviewText'].dropna().astype(str)

# En sık geçen kelimeleri bulmak için CountVectorizer kullan
def en_sik_kelimeler(yorumlar, n=20):
    vectorizer = CountVectorizer(stop_words='english', max_features=n)
    X = vectorizer.fit_transform(yorumlar)
    kelime_frekuensi = X.sum(axis=0).A1
    kelimeler = vectorizer.get_feature_names_out()
    return pd.DataFrame({'Kelime': kelimeler, 'Frekans':
kelime_frekuensi}).sort_values(by='Frekans', ascending=False)

# Pozitif ve negatif kelimeleri çıkar
pozitif_kelimeler = en_sik_kelimeler(pozitif_yorumlar)
negatif_kelimeler = en_sik_kelimeler(negatif_yorumlar)

import seaborn as sns
import matplotlib.pyplot as plt

# Pozitif kelimeler görselleştirme
plt.figure(figsize=(10, 5))
sns.barplot(data=pozitif_kelimeler, x='Frekans', y='Kelime', palette='Greens_r')
plt.title("Pozitif Yorumlarda En Sık Geçen Kelimeler")
plt.tight_layout()
plt.show()

# Negatif kelimeler görselleştirme
plt.figure(figsize=(10, 5))
sns.barplot(data=negatif_kelimeler, x='Frekans', y='Kelime', palette='Reds_r')
plt.title("Negatif Yorumlarda En Sık Geçen Kelimeler")
plt.tight_layout()
plt.show()

import ace_tools as tools; tools.display_dataframe_to_user(name="Pozitif ve Negatif
Kelimeler", dataframe=pd.concat([
    pozitif_kelimeler.assign(Tip='Pozitif'),

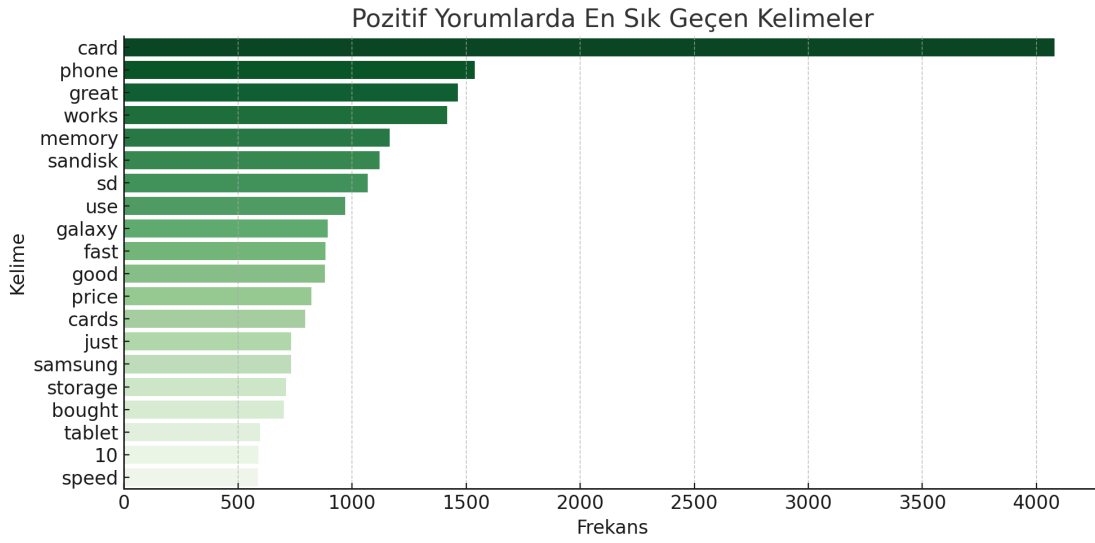
```

```
negatif_kelimeler.assign(Tip='Negatif')
]))
```

Sonuç

	Kelime	Frekans	Tip
2	card	4080	Pozitif
10	phone	1539	Pozitif
7	great	1464	Pozitif
19	works	1418	Pozitif
9	memory	1166	Pozitif

### Pozitif Yorumlarda En Sık Geçen Kelimeler



### Negatif Yorumlarda En Sık Geçen Kelimeler

#### Negatif Yorumlarda En Sık Geçen Kelimeler



#### Pozitif Ve Negatif Kelimeler

Yorumlarda en sık geçen olumlu ve olumsuz ifadeleri yukarıdaki görsellerde ve tabloda görebilirsiniz. Bu analiz sayesinde müşterilerin hangi özelliklerden memnun kaldığını (örneğin *"great"*, *"works"*, *"easy"*) ve hangi konularda sorun yaşadığını (*"didn't"*, *"return"*, *"problem"*) kolayca görebiliriz.

Dilersen bu ifadeleri kullanarak ürün açıklaması iyileştirme, müşteri hizmetleri aksiyonları ya da içerik stratejileri geliştirebiliriz. İlgilenir misin?