



# BTK AKADEMİ DATATHON 2025

## Final Sunumu

**ROC Stars**

Hakkı Besim Bardakçı-Gamze Akkurt

# Ekip Üyeleri

---



Hakkı Besim Bardakçı

**Data Scientist**  
arabam.com



Gamze Akkurt

**Data Scientist**  
Freelance

# Sunum Akışı

---

- Problem Tanımı
- Literatür
- Keşifsel Veri Analizi
- Özellik Çıkarımı
- Modelleme
- Analiz Sonuçları
- Referanslar

# Problemin Tanımı

Kullanıcıların oturum bazlı etkileşimlerinden (görüntüleme, sepete ekleme, satın alma vb.) hareketle her oturumun değeri (session\_value) tahmin edilmektedir.

---

Toplam aktivite: 141,219

---

Kategori sayısı: 448

---

Toplam kullanıcı: 51,821

---

Ürün sayısı: 26,470

---

Toplam oturum: 70,736

---

Etkileşim türü: 4

Model performansı **MSE** (Mean Squared Error) metriği ile değerlendirilmektedir.

# Mevcut Yaklaşımlar

## Kullanılan Öznitelikler

- Kullanıcı davranışları
- Oturum özellikleri
- Ürün/ürün kategorisi bilgisi
- Zaman bilgisi
- Tarihsel kullanıcı verisi

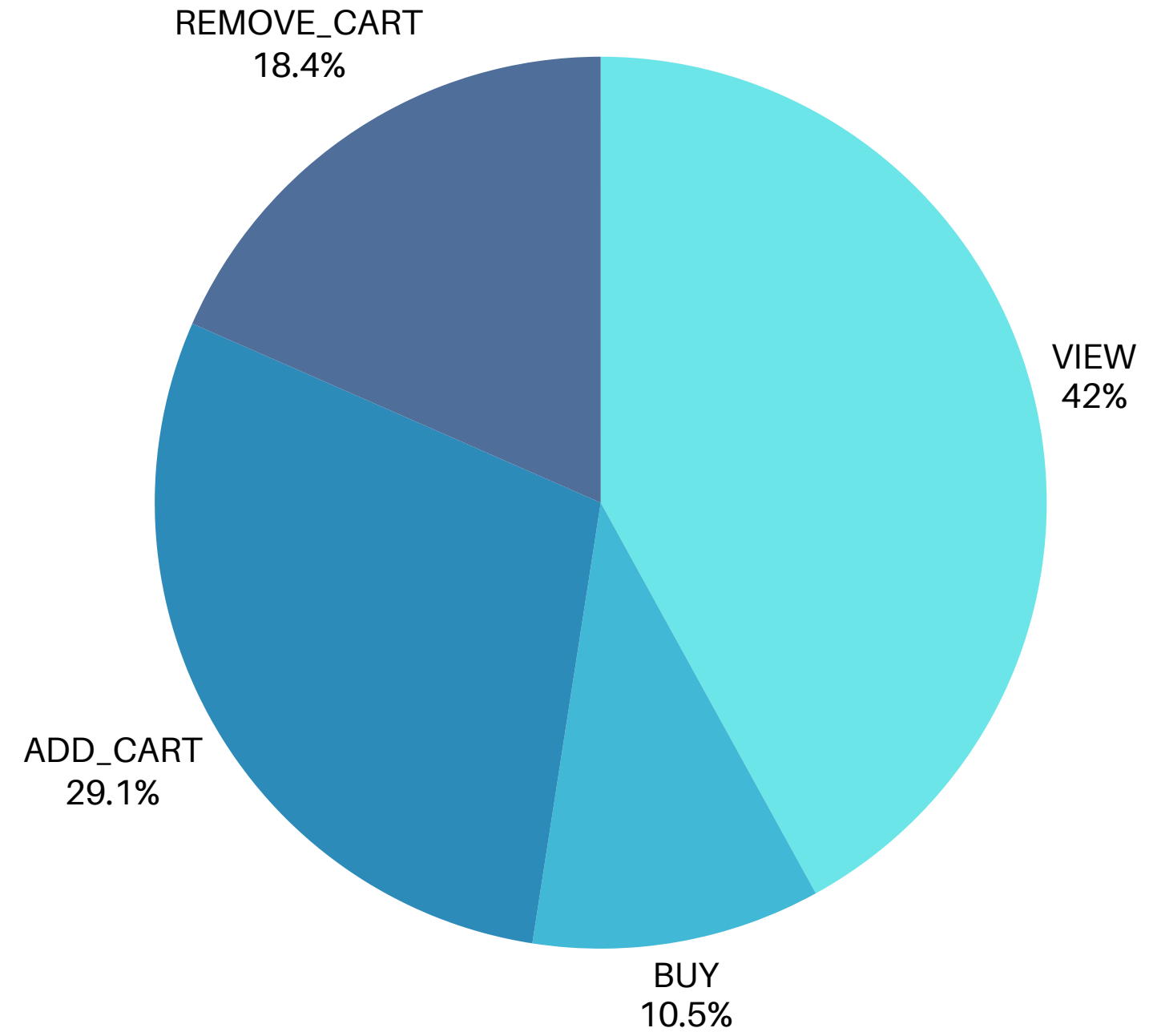
## Kullanılan Modeller

- GRU/LSTM
- Transformer tabanlı modeller (BERT4Rec, SASRec)
- GBDT / Random Forest

# LİTERATÜR

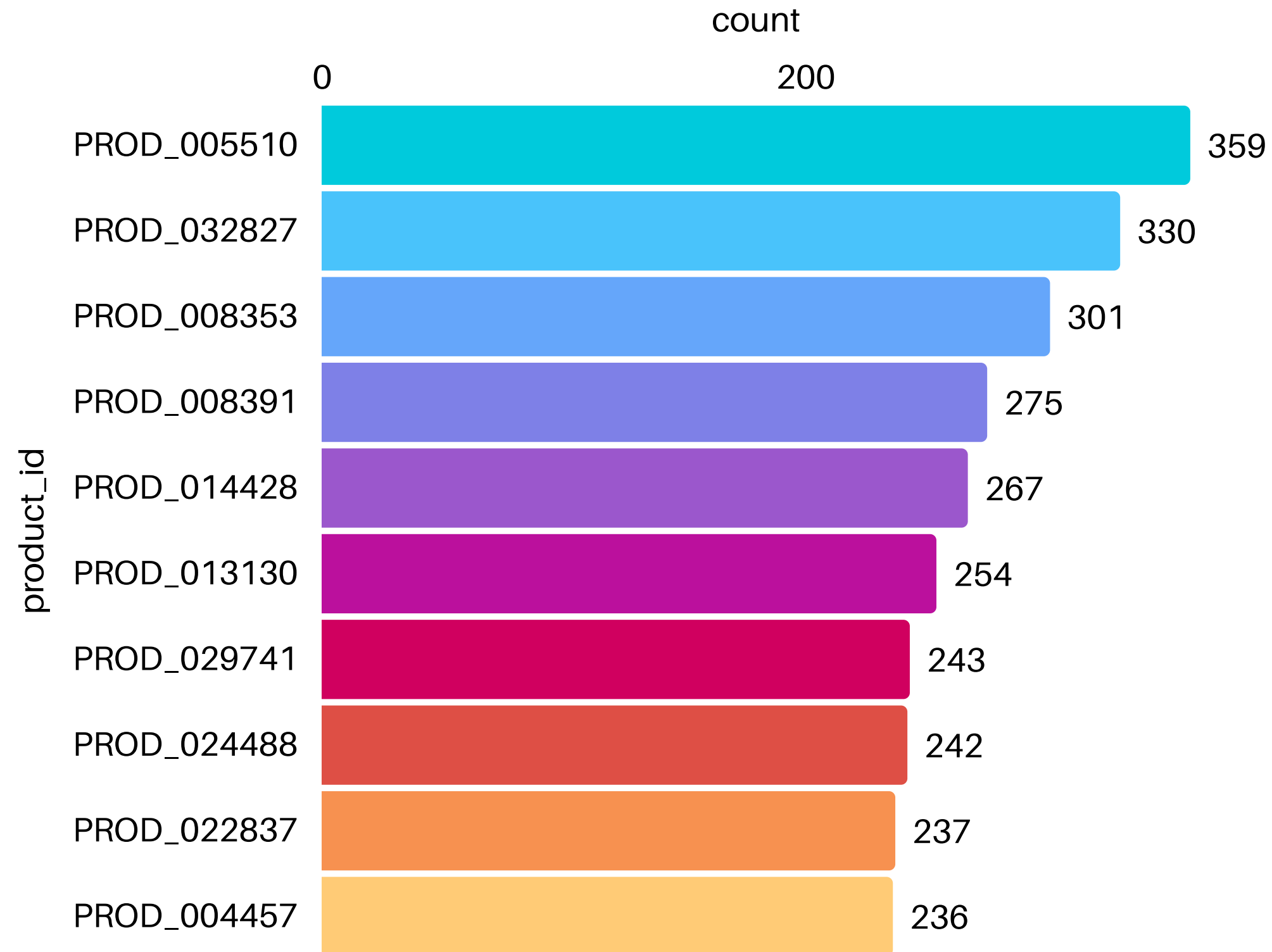
# VERİ ANALİZİ

## Etkinlik Türü



# VERİ ANALİZİ

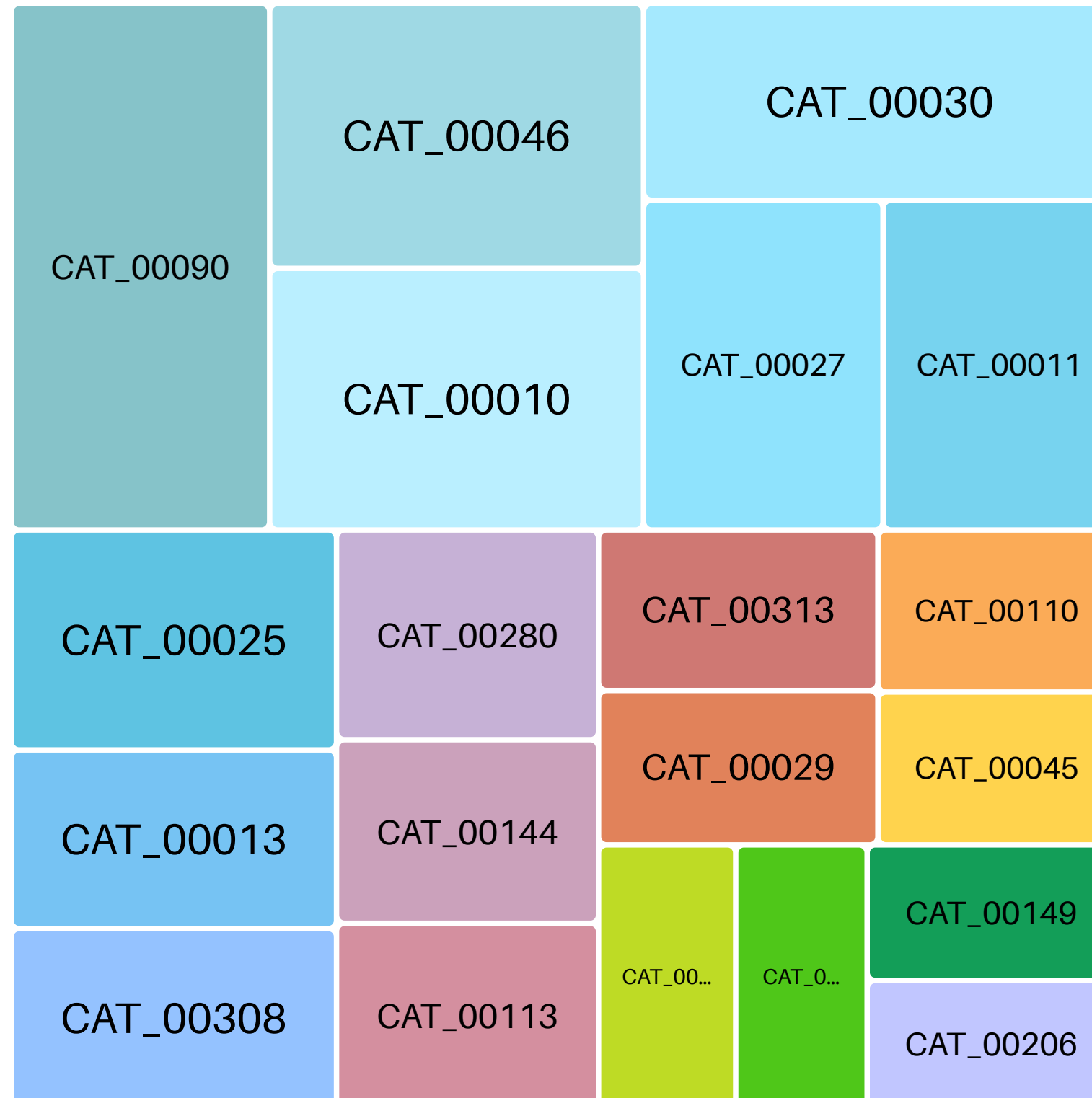
## Top 10 Ürün



# VERİ ANALİZİ

---

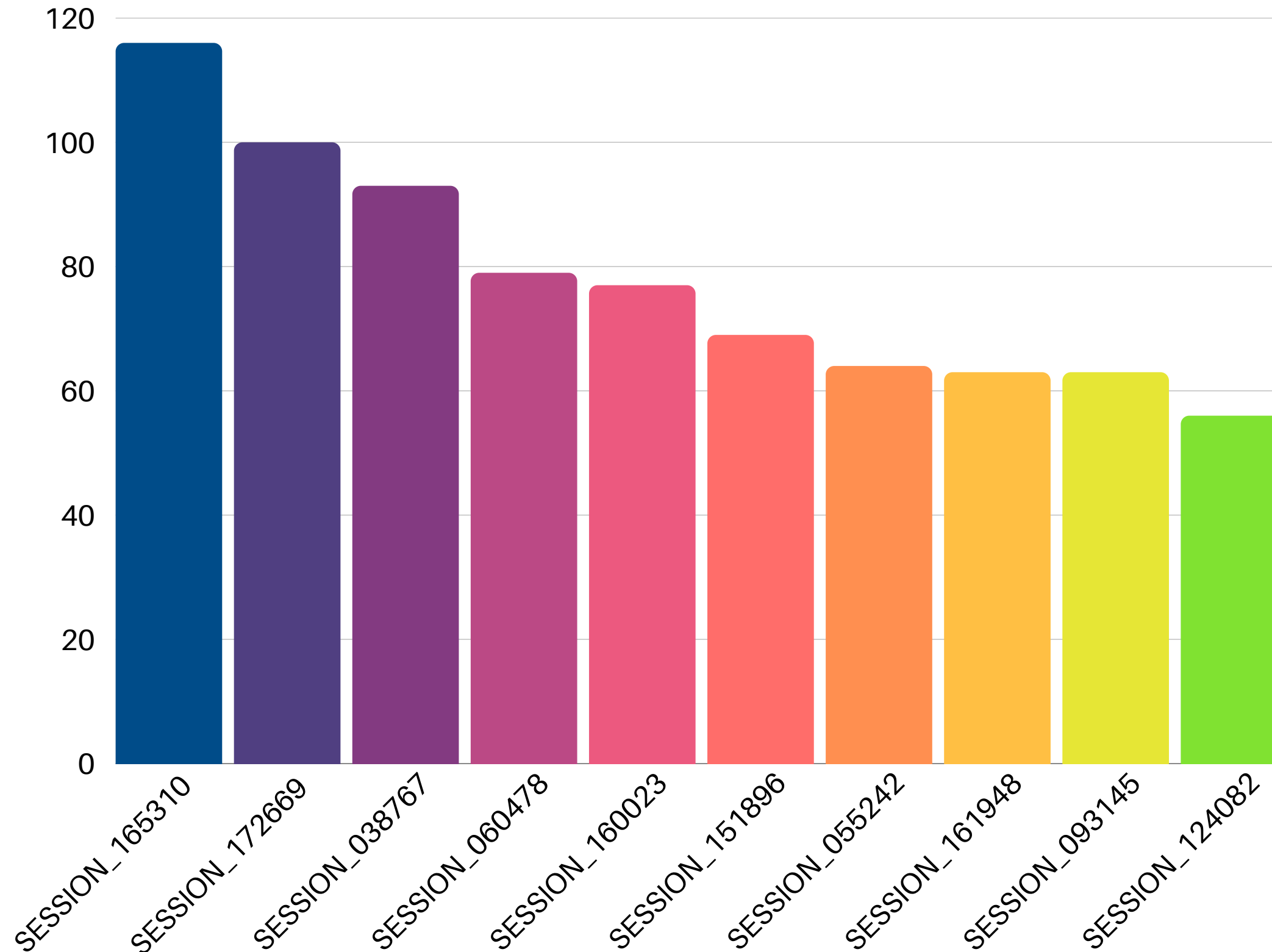
## Top 20 Kategori





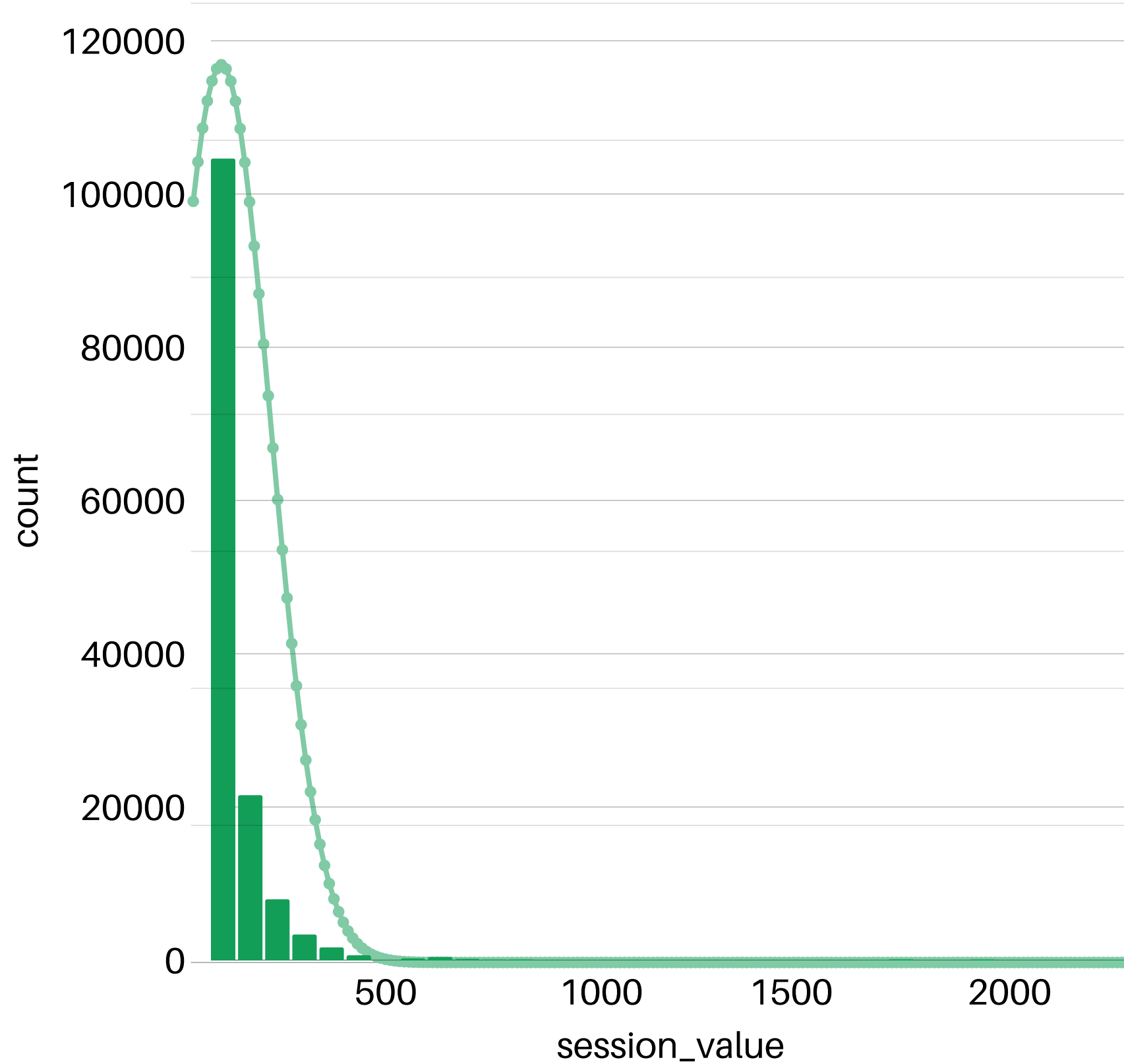
# VERİ ANALİZİ

## En Fazla 10 Kullanıcı Oturumu

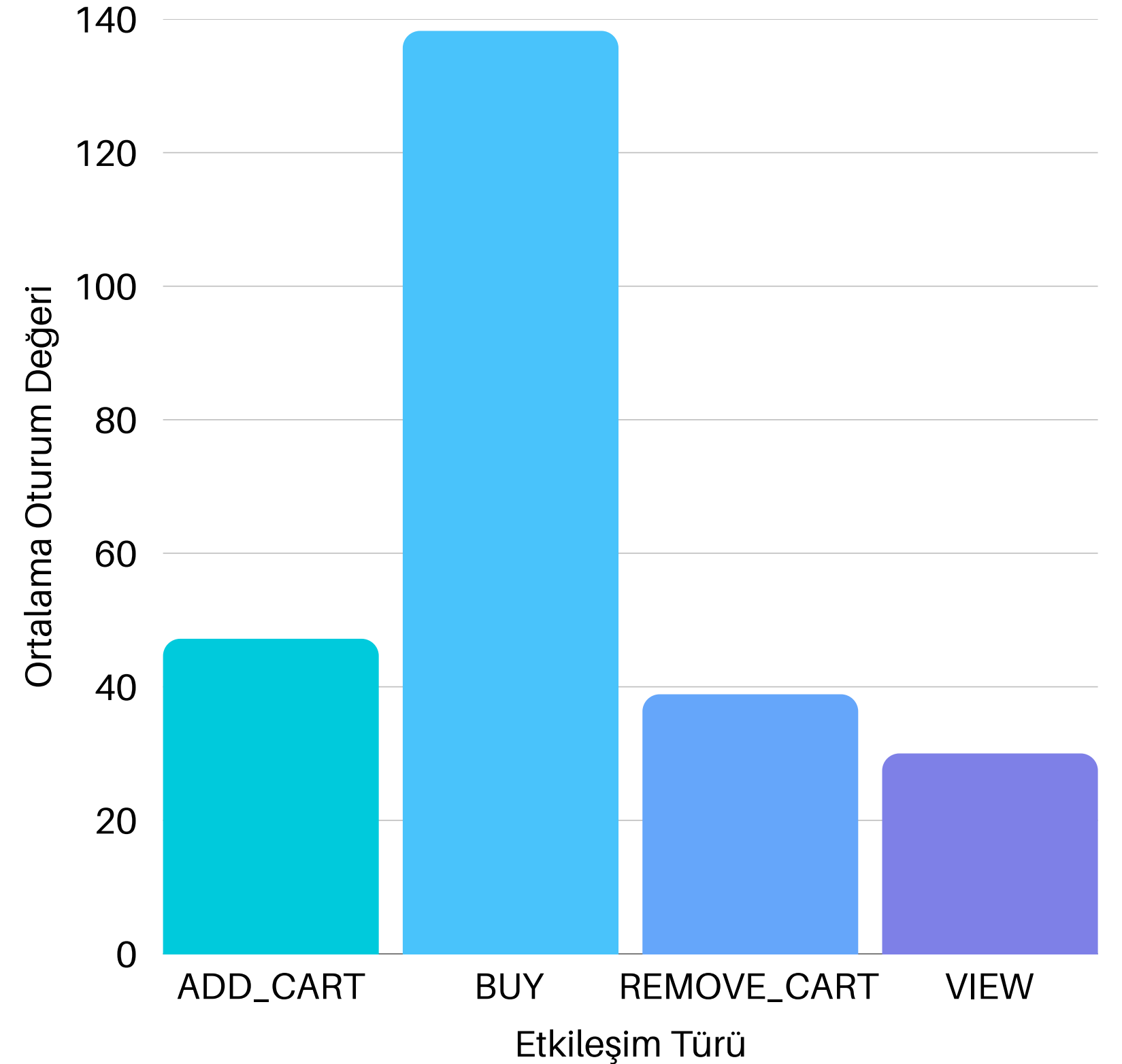


# VERİ ANALİZİ

## Oturum Değeri Dağılımı

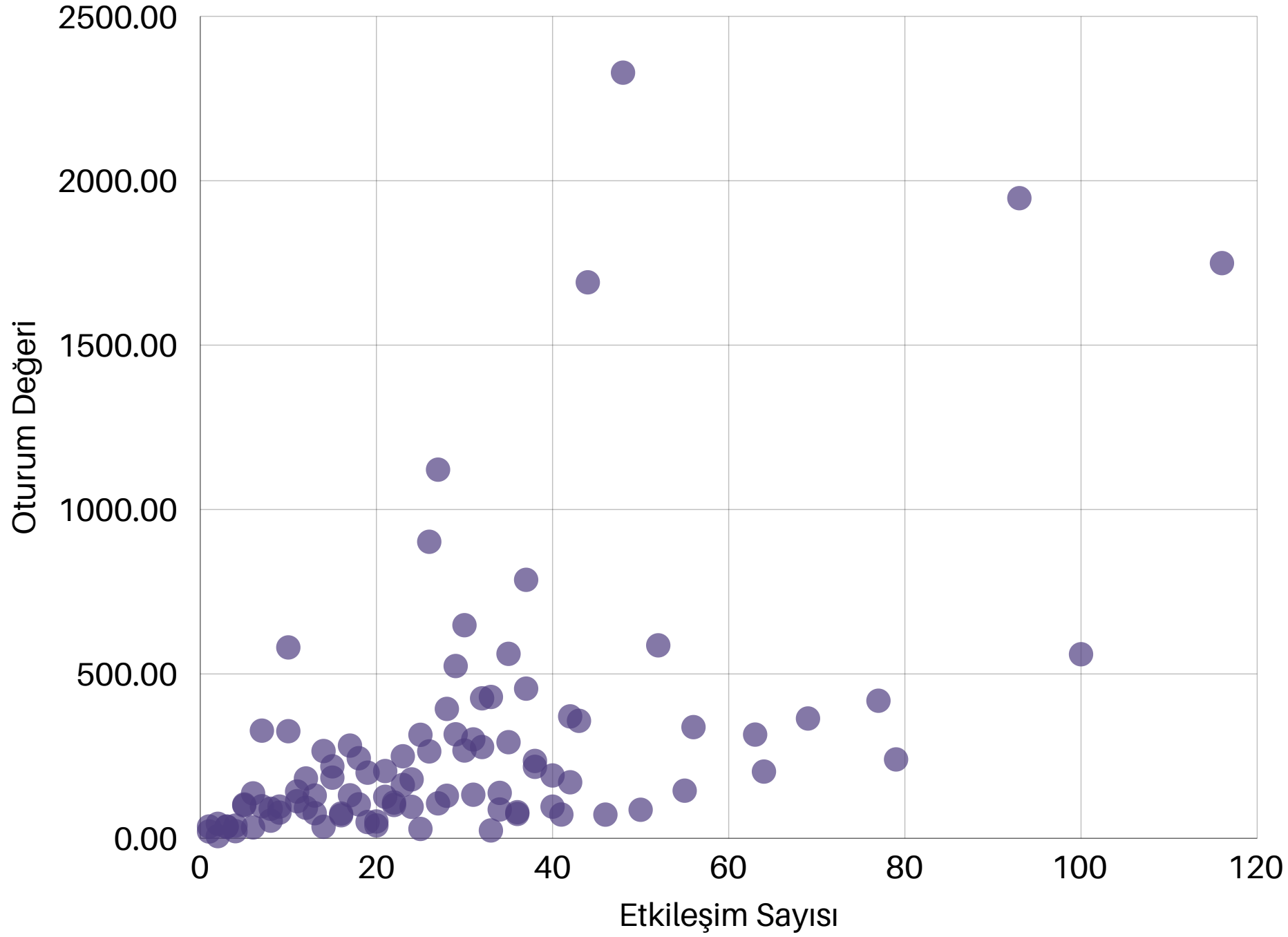


## Etkinlik Türlerine Göre Oturum Değeri

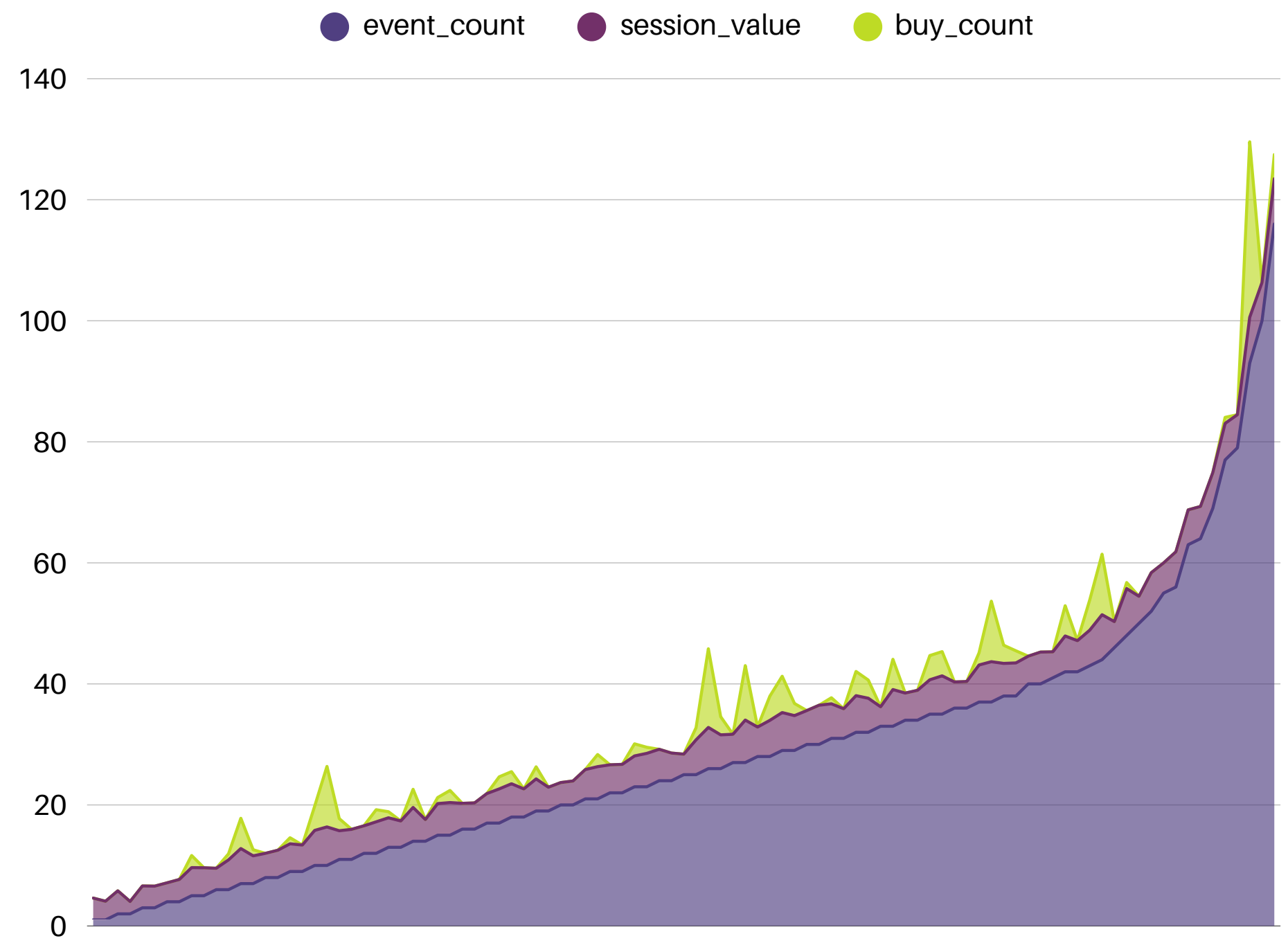


# VERİ ANALİZİ

## Etkileşim Sayısı vs Oturum Değeri Dağılımı

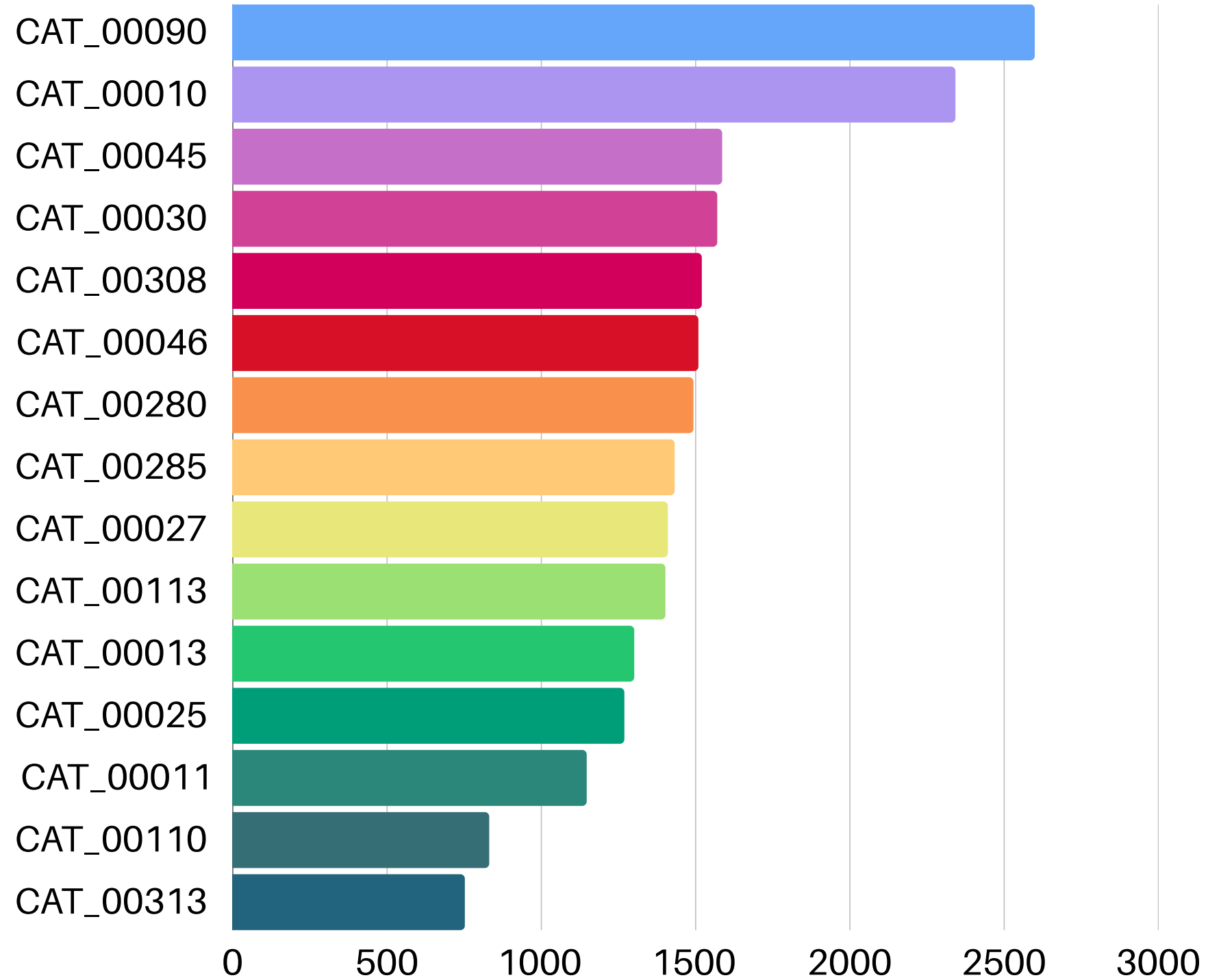


## Etkileşim Sayısı vs Buy Count vs log1p(Session Value) Grafiği

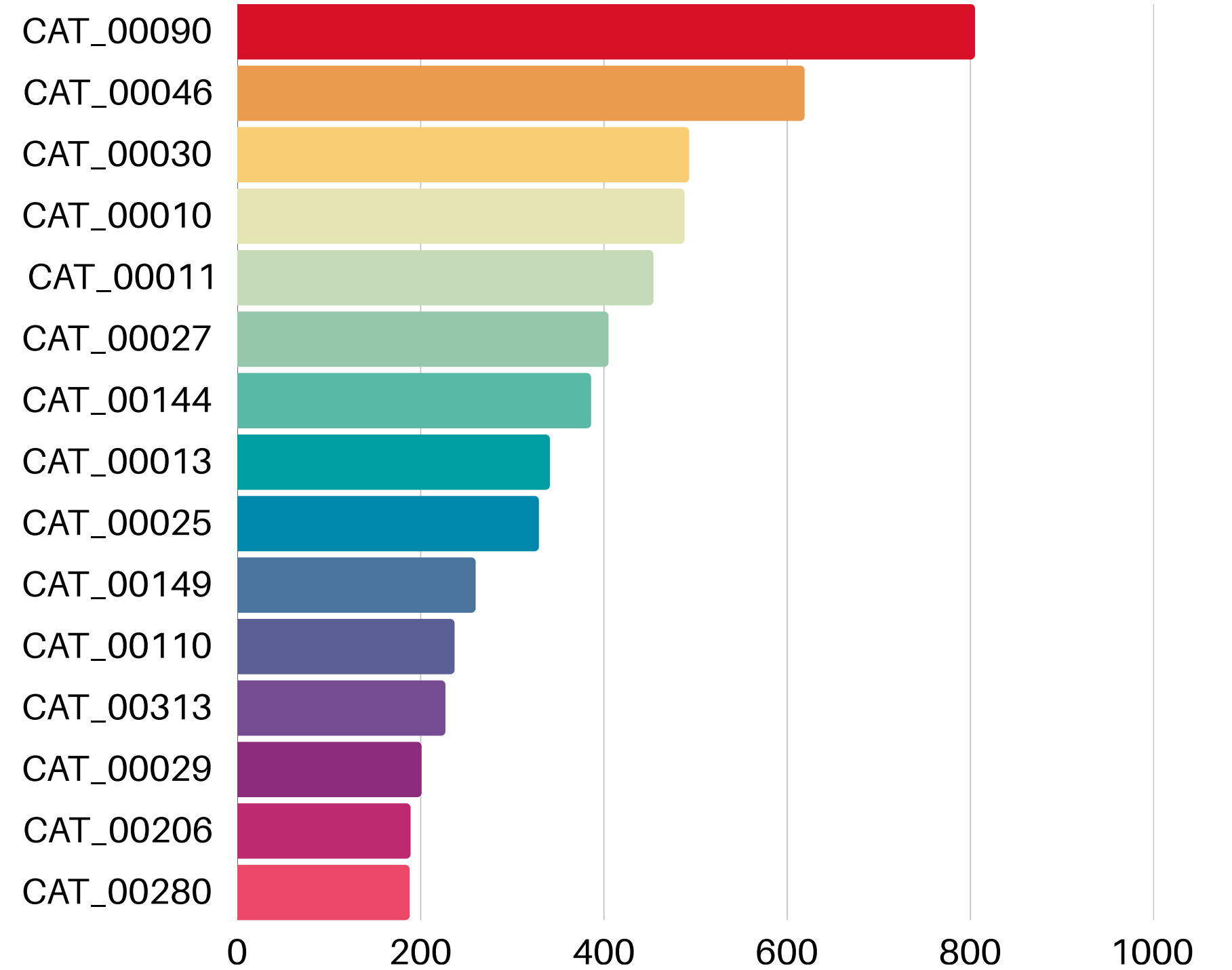


# VERİ ANALİZİ

## En Çok Görüntülenen 15 Kategori

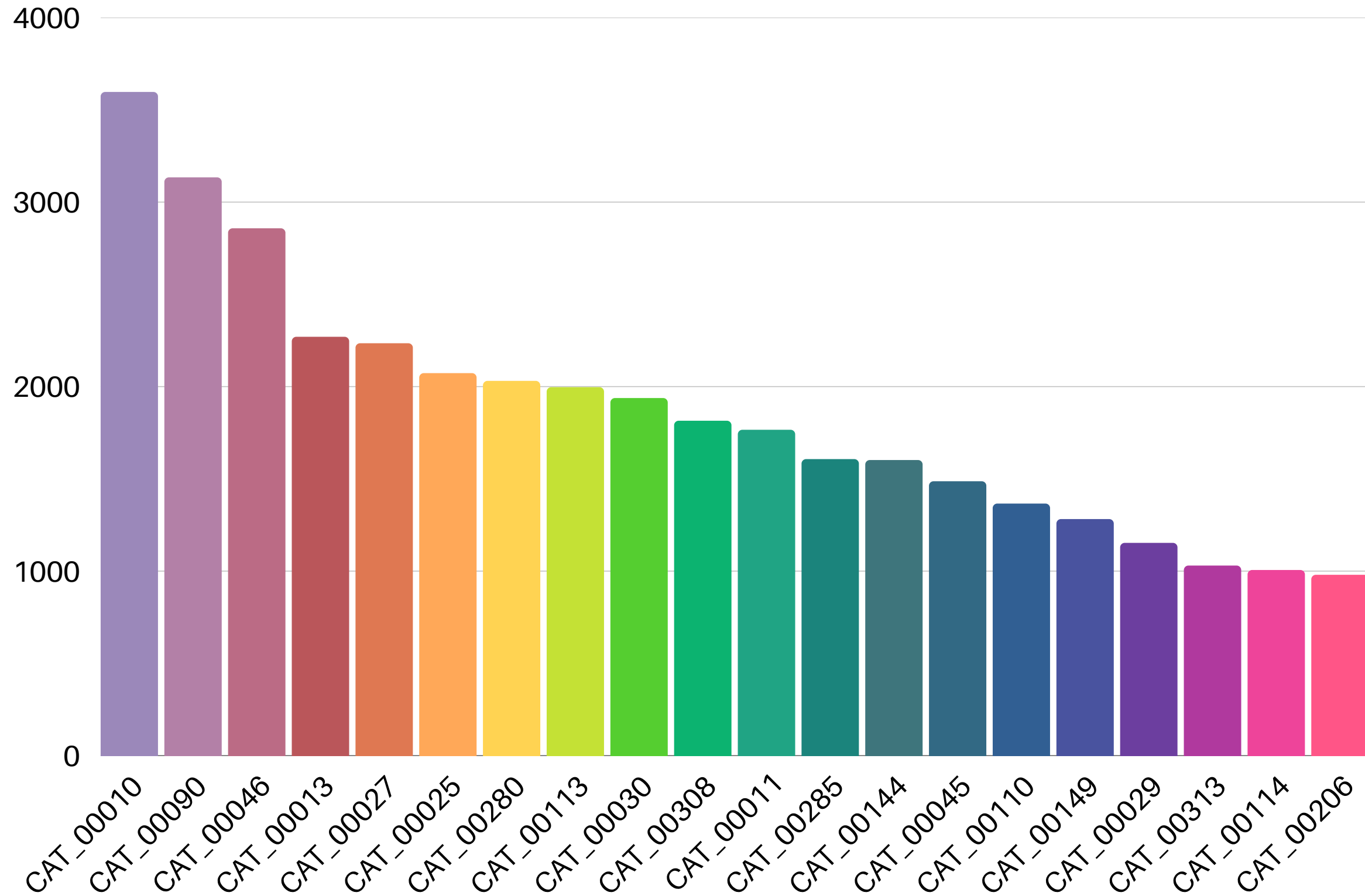


## En Çok Satın Alınan 15 Kategori



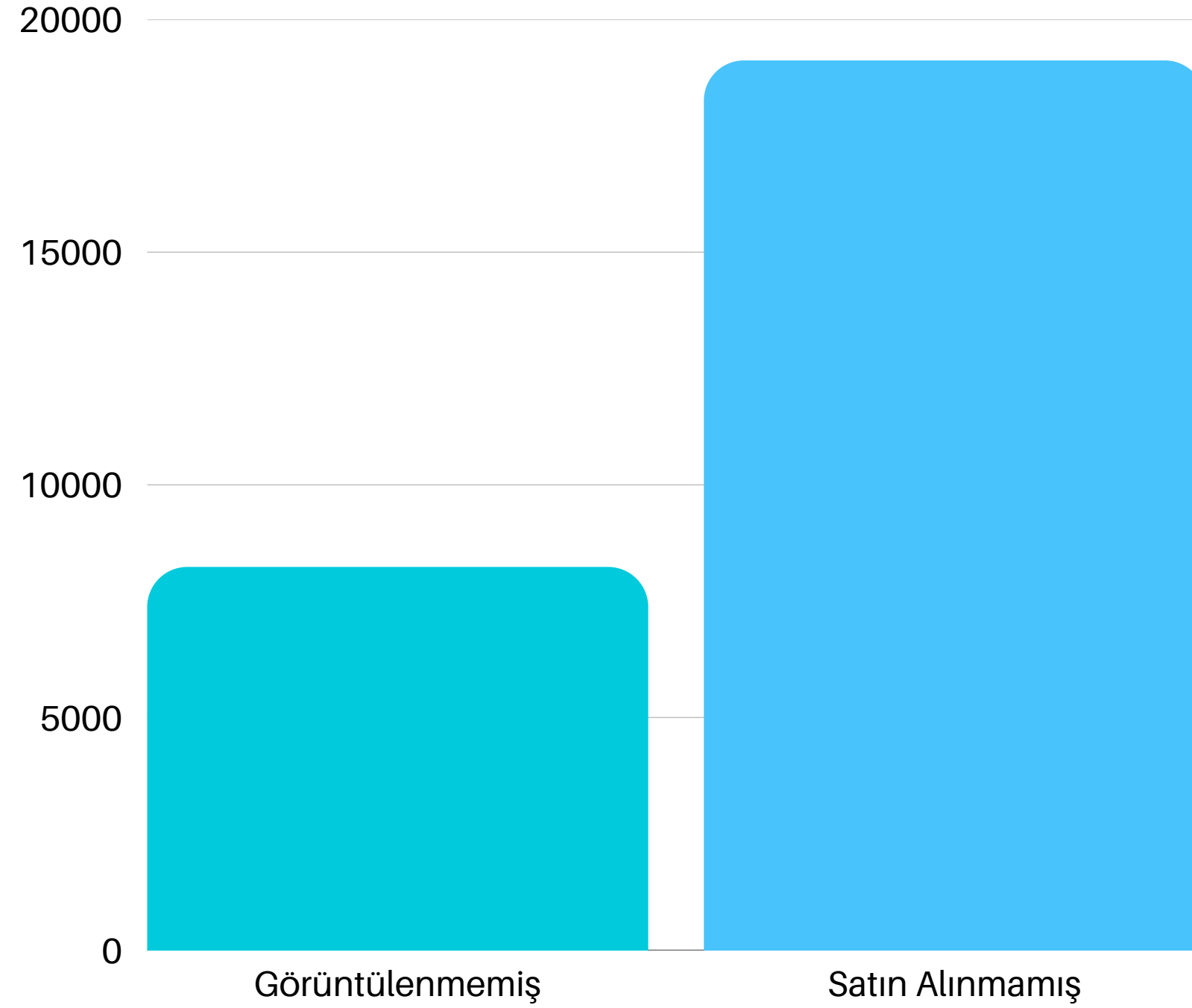
# VERİ ANALİZİ

En Çok Kullanıcısı Olan 20 Kategori



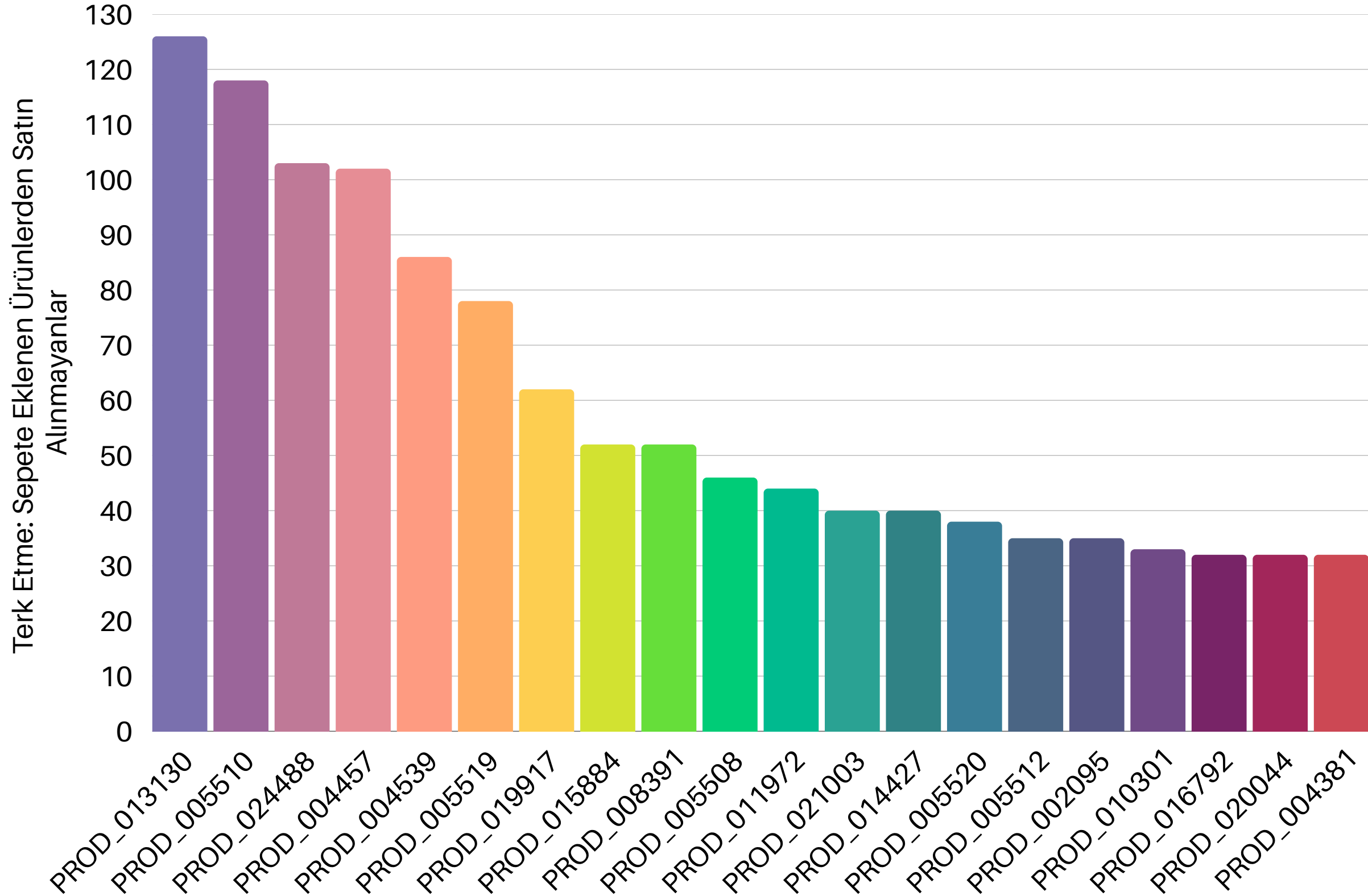
# VERİ ANALİZİ

Görüntülenmemiş veya Satın Alınmamış  
Ürün Sayısı



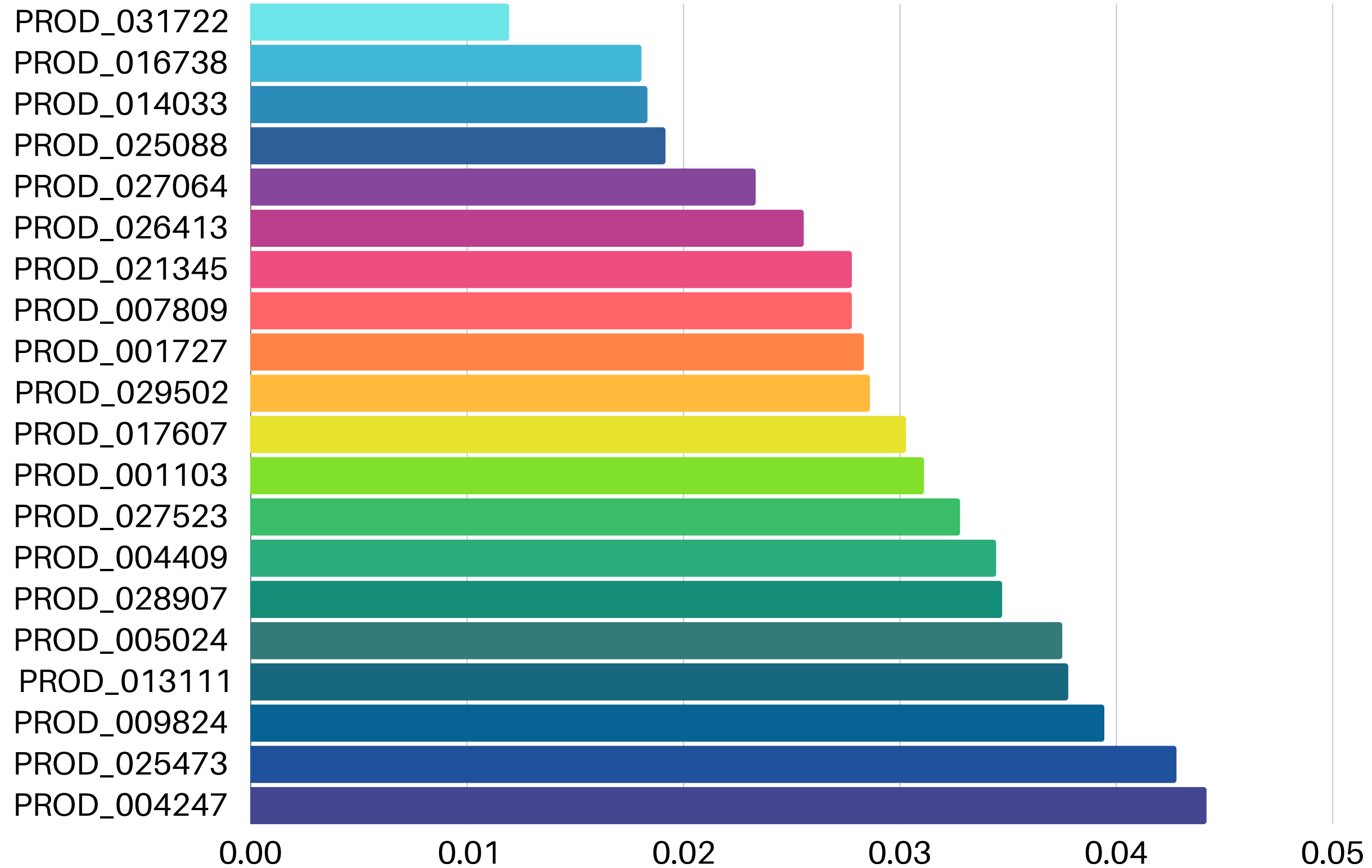
# VERİ ANALİZİ

## Sepette En Çok Bırakılan 20 Ürün



# VERİ ANALİZİ

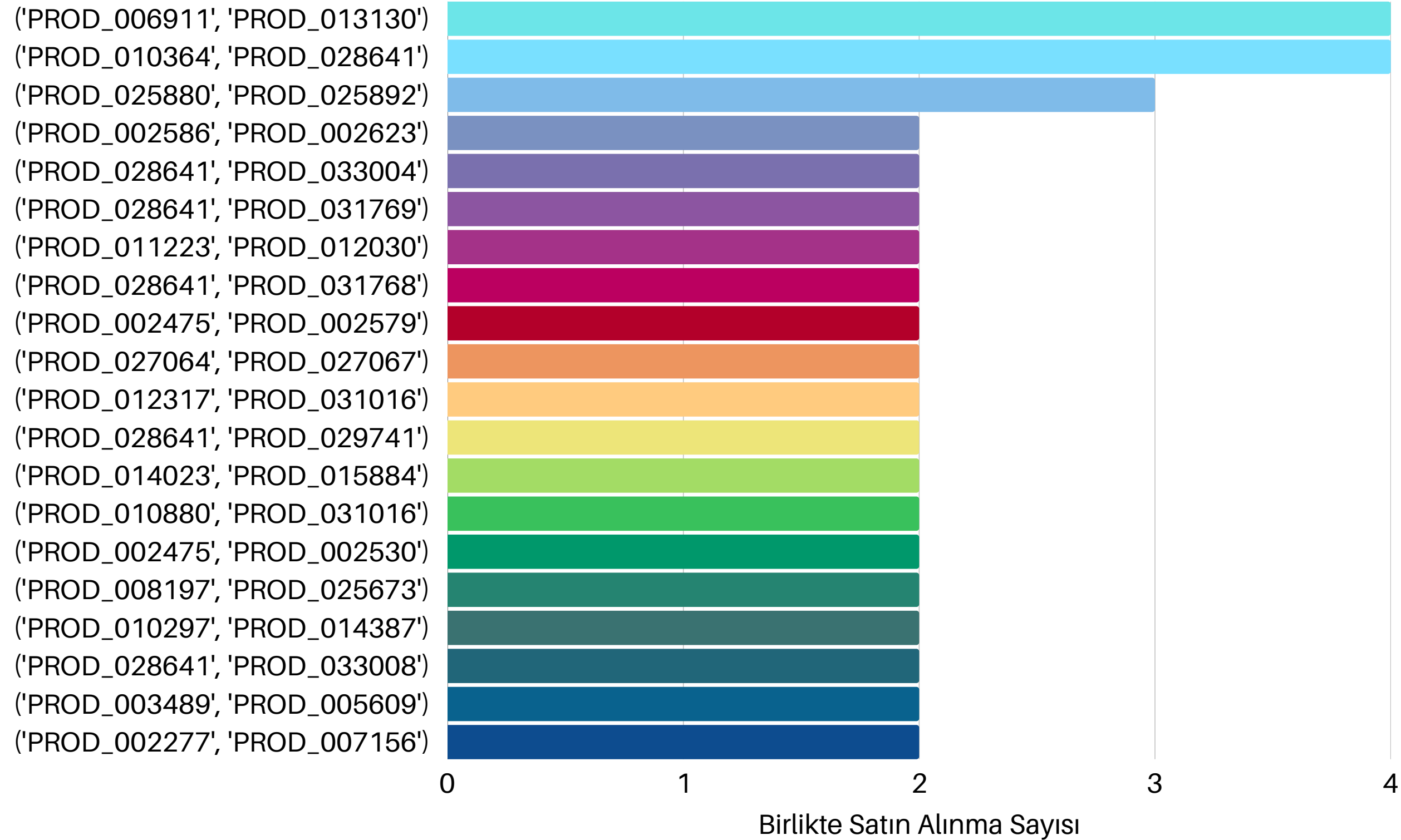
## Ortalama Satın Alma Süresi En Kısa 20 Ürün





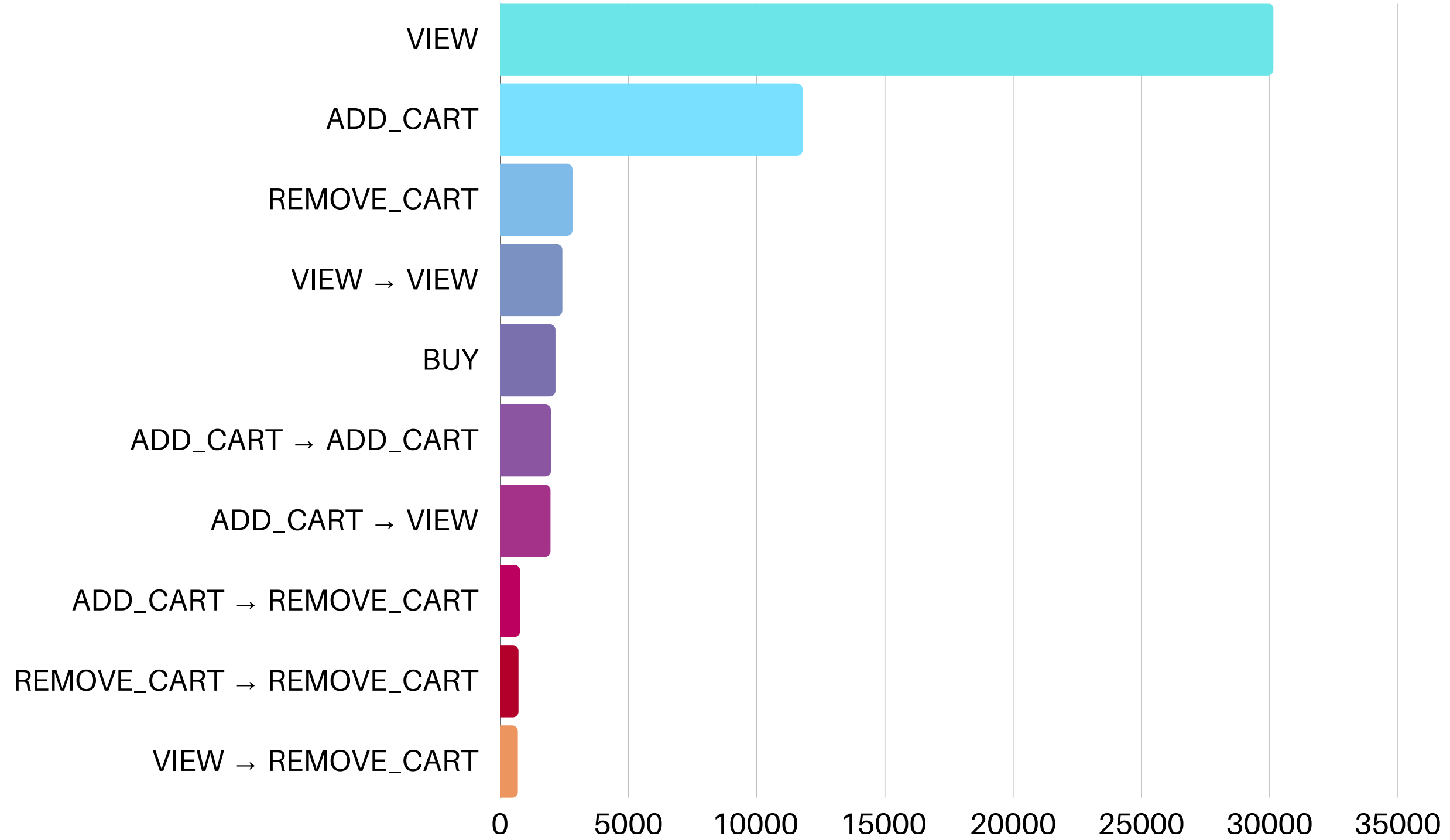
# VERİ ANALİZİ

## Birlikte En Çok Satın Alınan 20 Ürün



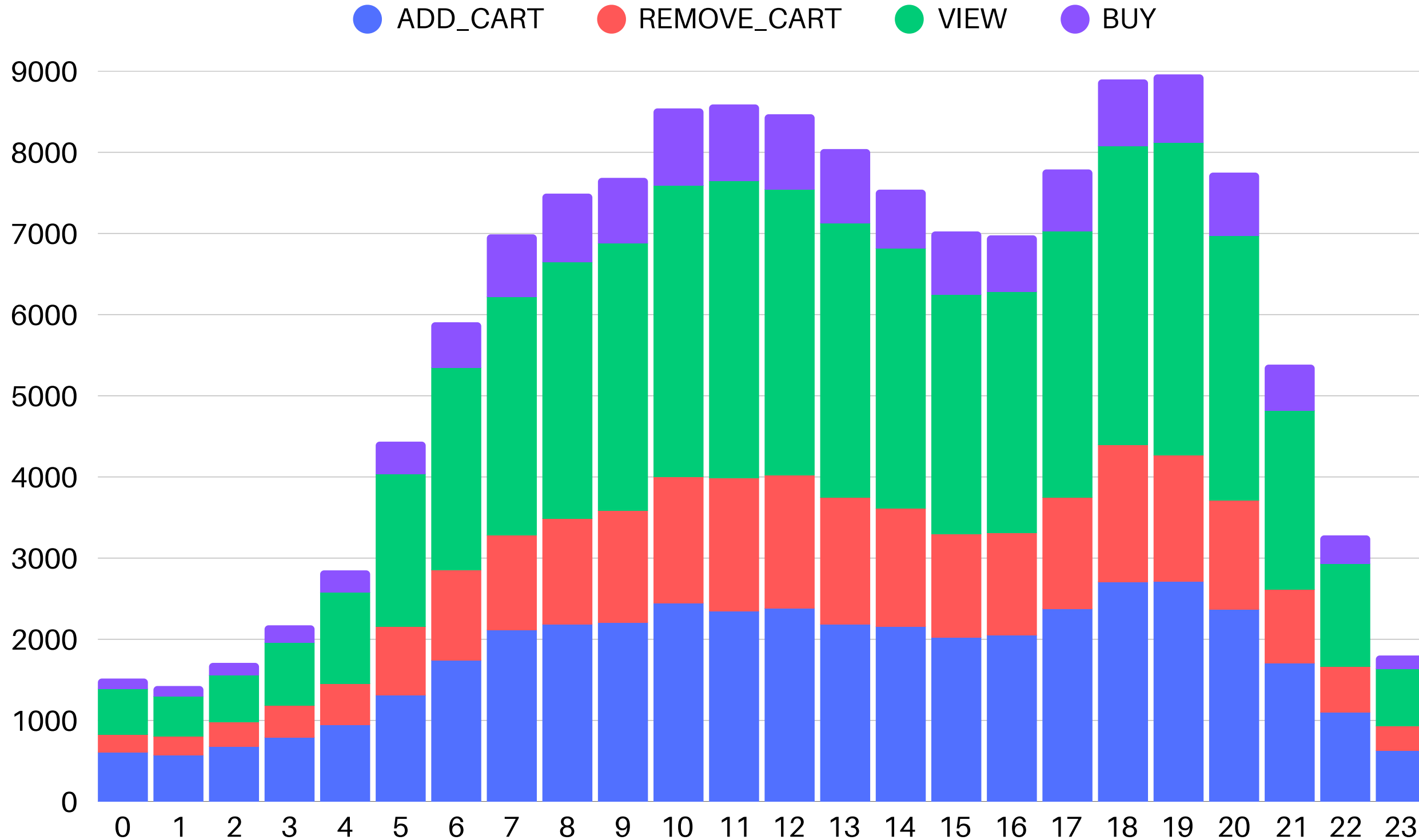
# VERİ ANALİZİ

En Popüler 10 Kullanıcı Etkinlik Dizisi



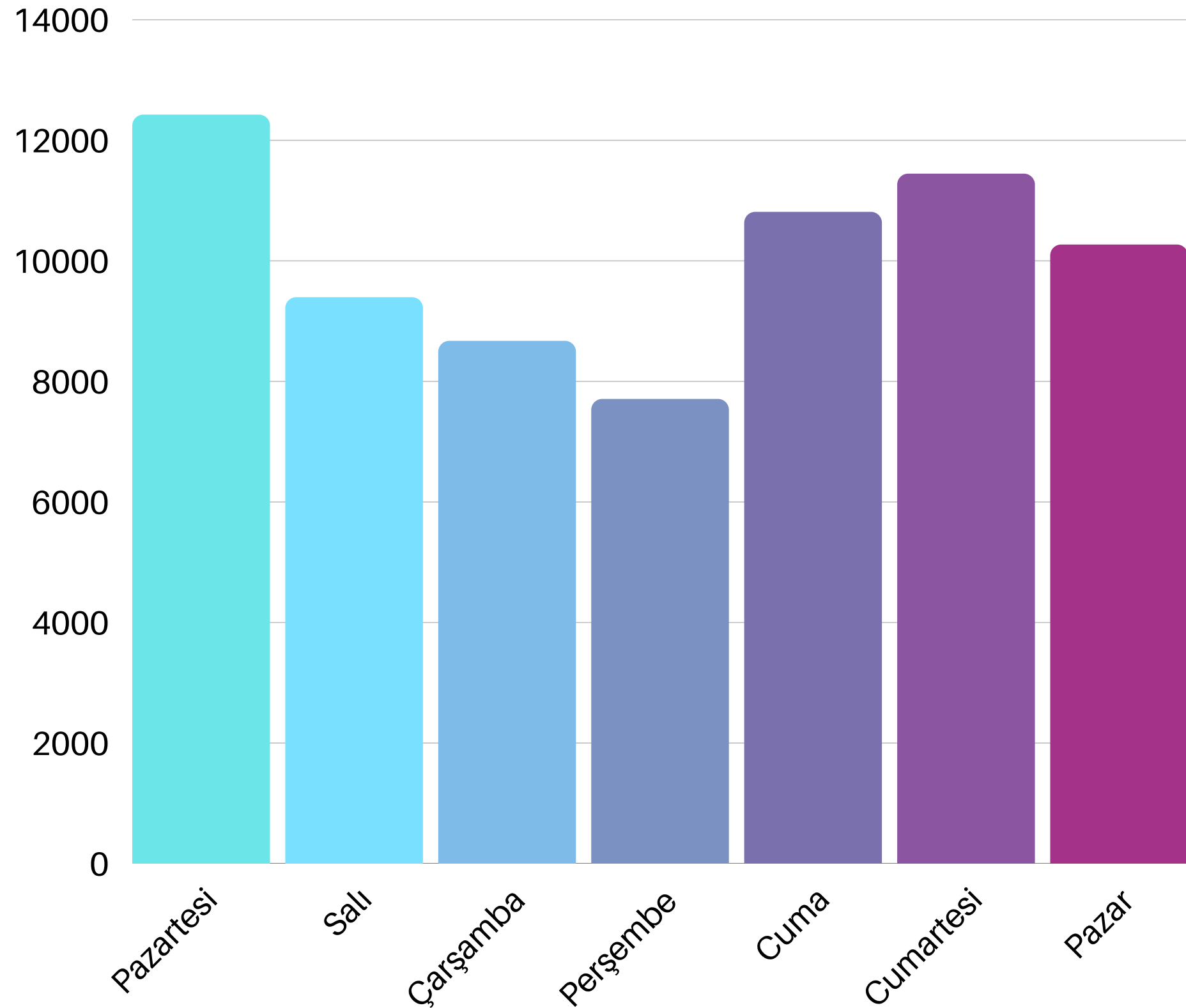
# VERİ ANALİZİ

Saat Bazında Etkinlik Dağılımı



# VERİ ANALİZİ

## Haftalık Oturum Sayısı Dağılımı



# Modelleme Stratejisi

## Event Count ve Session Value yüksek korelasyonu

---

- Session\_value değerini session bazlı olarak olay sayısı değeriyle normalize ettik. ( $\text{session\_value} / \text{olay\_sayisi}$ )
- Yapılan değişiklikten sonra tahmin anında her bir session için tahmin edilen session\_value değerlerini topladık. Fold skorlarını da aynı şekilde ölçtük.
- Train setindeki her bir veriye  $\log 1p(\text{session\_value})$  olacak şekilde ağırlık verdik. Bu sayede yüksek tahminlerdeki MSE değerlerini düşürmeyi amaçladık.

## Train verisetinde varolan bazı sessionların Testte de olması

---

- Preprocess işlemlerinden önce ortak sessionları train verisetine aktarıp ortak sessionlardan bir sözlük oluşturduk.
- Aktardığımız ortak sessionları da modeli eğittiğimiz verinin içine de dahil ettik.
- Oluşturduğumuz sözlüğü test verisetini tahmin ederken kullandık, ortak veriler için tekrar tahmin yapmak yerinde sözlükteki değerleri kullandık.

# Fold Stratejisi

## Fold Stratejisini belirlerken

---

- Session Value değerlerinin homojen olarak dağılmaması.
- Aynı Session'a ait değerlerinin aynı sette kalması gerekmesi.
- Modeli yeterince test edebilecek kadar fazla ancak çok fazla iterasyona sebep olmayacak sayıda az olması gerekmesi.

**StratifiedGroupKFold** kullanmakta karar kıldık.

## Stratification Label'ı belirleme

---

- Stratification label'ını belirlerken hem session toplamını hem de session ortalamasını denedik. En iyi sonucu session ortalaması verdiği için onunla devam ettik.
- Session ortalamasını 20 quantile'a böldük.

## Fold Grouping

---

User\_session kolonunu olduğu gibi kullandık.

## Fold Sayısı

---

Farklı n değerleri denedik n=5 fold değerinde leaderboard ile en yüksek korelasyonu sağladık

# Öznitelik Çıkarımı (Feature Engineering)

## STEP 1

### Anomalilerinin Tespiti

OOF tahminlerinde çok yüksek hata değerlerinin gelmesi üzerine toplamda 45 tane anomaly heuristici geliştirdik.

## STEP 2

### Oturum, Kullanıcı ve Zamansal Özellikler

Toplam oturum sayısı, ortalama olay aralığı, olayların saatlik dağılımı, toplam satın alma sayısı vb. özellikler çıkarılması

## STEP 3

### Kategorik Kodlama ve Embedding Özellikleri

Temel kategorik değişkenlerin sayısal forma dönüştürülmesi ardından **Truncated SVD** kullanarak düşük boyutlu embedding vektörleri oluşturulması

## STEP 4

### Oturum Düzeyi Embedding Kümeleme

Oturum embeddinglerinin contrastive learning kullanarak oluşturulup farklı **n\_cluster** değerleriyle **KMeans** kullanılarak kümelenmesiyle ve oluşan küme etiketlerinin, oturum davranışlarını temsil eden kategorik özellikler olarak çıkarılması.

# Anomali Tespiti

## Olağandışı Oturum Uzunlukları

- Toplam oturum: 70.736
- Uzun oturumlar : 14.701(%20)
- Kısa ve hızlı oturumlar: 30
- Süresi 0 olan oturumlar: 48.067(%67)

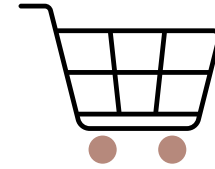
"very\_long\_duration"- "many\_duplicate\_timestamps"...



## Görüntüleme veya Sepete Ekleme Olmadan Satın Alma

- Öncesinde VIEW veya ADD\_CART olmayan satın almalar: 12.859(%18)

"purchase\_without\_add\_or\_view"



## Sepete Eklemeden Sepetten Çıkarma

- Öncesinde ADD\_CART olmayan REMOVE\_CART olay sayısı: 23.719(%33)

"remove\_without\_add"



## Olay Sırası Anomalileri

- Satın almalar veya sepetten çıkarma işlemleri beklenmedik şekildeki oturum sayısı: 3.822(%5)

"purchase\_without\_add\_or\_view"

## Gece Saatlerindeki Aktivite Artışları

- Toplam gece etkinliği: 14.105(%20)

"high\_frac\_events\_night"



# Model Seçimi

Farklı modeller test ettik:

- Catboost
- LGBM
- XGBoost
- Autogluon
- LSTM / Transformer

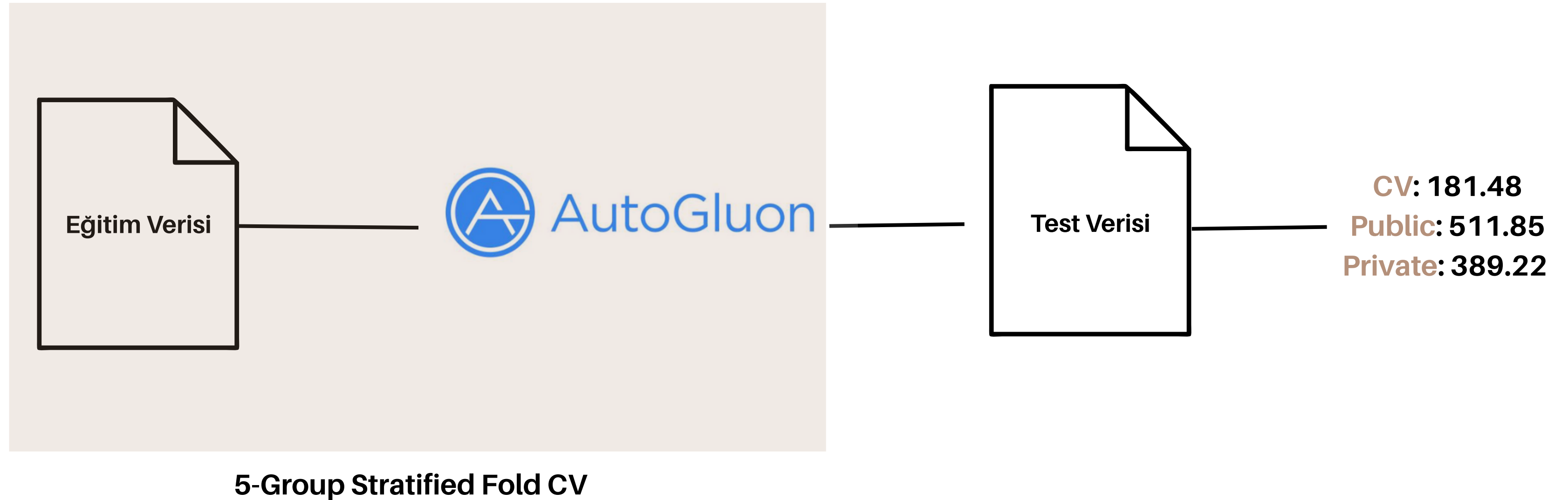
En iyi sonucu autogluon verdi.

Prototipleme kısmını en iyi 2. sonucu veren catboost ile yaptık.

Nihai model trainde ise autogluon kullandık.

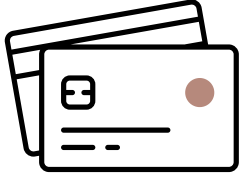


# Model Seçimi



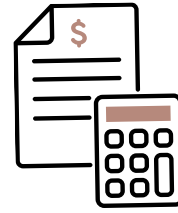
# Stratejik Öneriler | Analiz Bulguları

Doğru analiz, işinizi ileriye taşıyan güçlü değişimler yaratır.



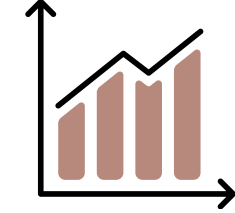
## Sepet Terk Etme & Dönüşüm

- Ödeme sürecini sadeleştirebiliriz daha az adım, şeffaf fiyatlandırma.
- Güven unsurları ekleyin (güvenli ödeme logoları, iade politikaları).
- Sepetten çıkış anında kampanya sunun (indirim pop-up'ı).



## Ürün & Kategori Optimizasyonu

- Fiyatlandırma stratejisini veya ürün sunumunu gözden geçirin (ör. düşük kaliteli görseller, eksik yorumlar).
- Çok görüntülenen ama düşük dönüşüm sağlayan ürünleri belirleyin
- Çapraz satış için paket/promosyonlar oluşturun (sıklıkla birlikte alınan ürünler).



## Veri Kalitesi & Takip İyileştirmeleri

- Olay kayıt sürecini denetleyin.
- Olay sırasının bütünlüğünü sağlayın (VIEW → ADD → BUY).
- Bu iyileştirme, hem analitik kalitesini hem de model performansını artıracaktır.

# Referanslar

---

## Session-Based Recommendation

Recurrent neural networks with top-k gains for session-based recommendations

Enhancing Collaborative Information with Contrastive Learning for Session-based Recommendation

STAMP: short-term attention/memory priority model for session-based recommendation

Behavior Sequence Transformer for E-commerce Recommendation in Alibaba

---

# Teşekkürler

---



ROC *Stars*