



Kural Tabanlı Sınıflandırma ile  
Potansiyel Müşteri Getirisi Hesaplama

miuul

# İş Problemi

Bir oyun şirketi müşterilerinin bazı özelliklerini kullanarak seviye tabanlı (level based) yeni müşteri tanımları (persona) oluşturmak ve bu yeni müşteri tanımlarına göre segmentler oluşturup bu segmentlere göre yeni gelebilecek müşterilerin şirkete ortalama ne kadar kazandırabileceğini tahmin etmek istemektedir.

## Örneğin:

Türkiye'den IOS kullanıcısı olan 25 yaşındaki bir erkek kullanıcının ortalama ne kadar kazandırabileceği belirlenmek isteniyor.



İŞ PROBLEMİ  
VERİ SETİ

DEĞİŞKENLER  
GÖREVLER

# Veri Seti Hikayesi

**Persona.csv** veri seti uluslararası bir oyun şirketinin sattığı ürünlerin fiyatlarını ve bu ürünleri satın alan kullanıcıların bazı demografik bilgilerini barındırmaktadır. Veri seti her satış işleminde oluşan kayıtlardan meydana gelmektedir. Bunun anlamı tablo tekilleştirilmemiştir. Diğer bir ifade ile belirli demografik özelliklere sahip bir kullanıcı birden fazla alışveriş yapmış olabilir.

PRICE	SOURCE	SEX	COUNTRY	AGE
39	android	male	bra	17
39	android	male	bra	17
49	android	male	bra	17
29	android	male	tur	17
49	android	male	tur	17



# Değişkenler

## persona.csv

**PRICE** – Müşterinin harcama tutarı

**SOURCE** – Müşterinin bağlandığı cihaz türü

**SEX** – Müşterinin cinsiyeti

**COUNTRY** – Müşterinin ülkesi

**AGE** – Müşterinin yaşı

İŞ PROBLEMİ

VERİ SETİ

DEĞİŞKENLER

GÖREVLER

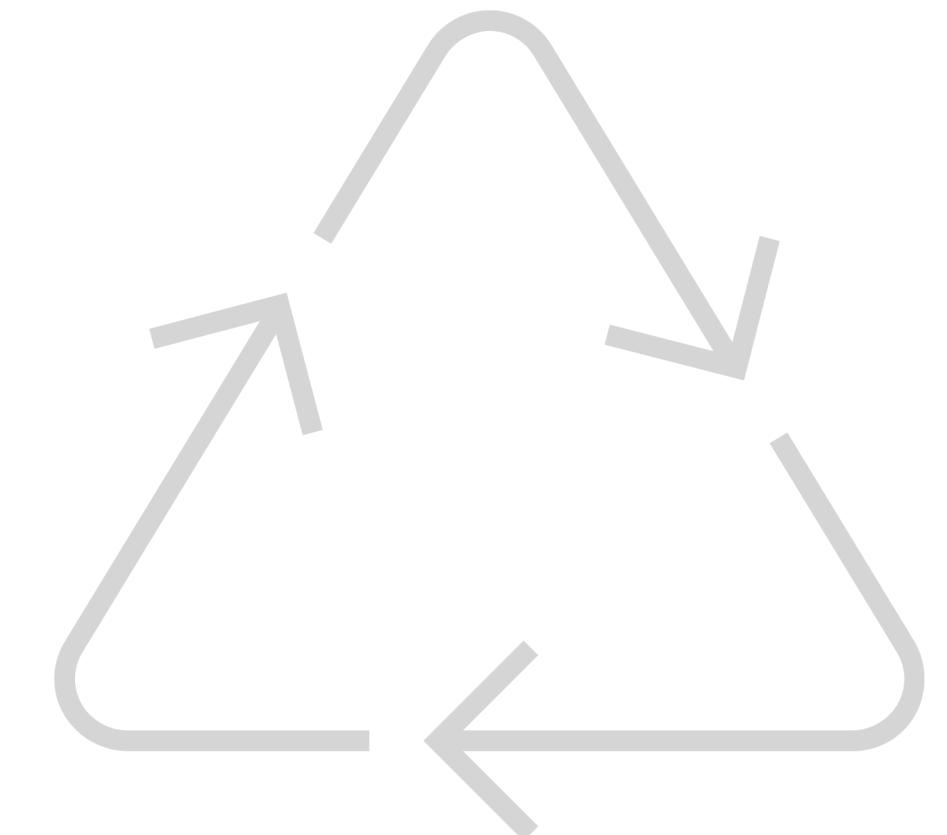


# Proje Görevleri

---

## Uygulama Öncesi Veri Seti :

PRICE	SOURCE	SEX	COUNTRY	AGE
39	android	male	bra	17
39	android	male	bra	17
49	android	male	bra	17
29	android	male	tur	17
49	android	male	tur	17



## Hedeflenen çıktı :

customers_level_based	PRICE	SEGMENT
BRA_ANDROID_FEMALE_0_18	35.6453	B
BRA_ANDROID_FEMALE_19_23	34.0773	C
BRA_ANDROID_FEMALE_24_30	33.8639	C
BRA_ANDROID_FEMALE_31_40	34.8983	B
BRA_ANDROID_FEMALE_41_66	36.7371	A

# Görev 1: Aşağıdaki Soruları Yanıtlayınız

**Soru 1:** persona.csv dosyasını okutunuz ve veri seti ile ilgili genel bilgileri gösteriniz.

**Soru 2:** Kaç unique SOURCE vardır? Frekansları nedir?

**Soru 3:** Kaç unique PRICE vardır?

**Soru 4:** Hangi PRICE'dan kaçar tane satış gerçekleşmiş?

**Soru 5:** Hangi ülkeden kaçar tane satış olmuş?

**Soru 6:** Ülkelere göre satışlardan toplam ne kadar kazanılmış?

**Soru 7:** SOURCE türlerine göre satış sayıları nedir?

**Soru 8:** Ülkelere göre PRICE ortalamaları nedir?

**Soru 9:** SOURCE'lara göre PRICE ortalamaları nedir?

**Soru 10:** COUNTRY-SOURCE kırılımında PRICE ortalamaları nedir?

**Görev 2:** COUNTRY, SOURCE, SEX, AGE kırılımında **ortalama kazançlar** nedir?**Elde edilmesi gereken çıktı:**

	COUNTRY	SOURCE	SEX	AGE	PRICE
0	bra	android	female	15	38.71
1				16	35.94
2				17	35.66
3				18	32.25
4				19	35.20

### Görev 3: Çıktıyı PRICE'a göre sıralayınız.

- Önceki sorudaki çıktıyı daha iyi görebilmek için sort\_values metodunu azalan olacak şekilde PRICE'a göre uygulayınız.
- Çıktıyı agg\_df olarak kaydediniz.

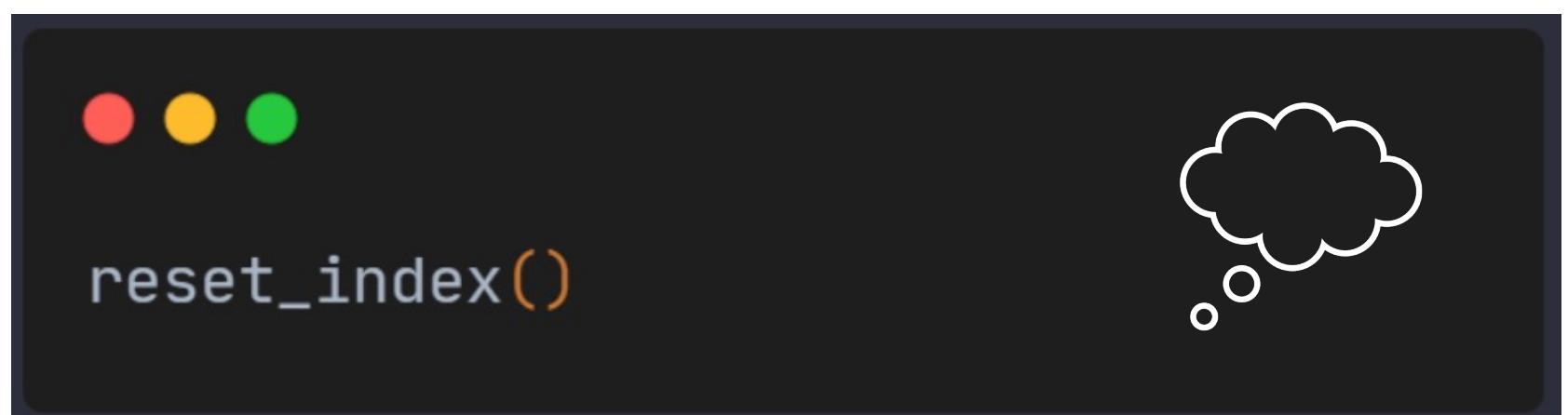
**Elde edilmesi gereken çıktı:**

	COUNTRY	SOURCE	SEX	AGE	PRICE
0	bra	android	male	46	59.0
1	usa	android	male	36	59.0
2	fra	android	female	24	59.0
3	usa	ios	male	32	54.0
4	deu	android	female	36	49.0

## Görev 4: Indekste yer alan isimleri değişken ismine çeviriniz.

- Üçüncü sorunun çıktısında yer alan PRICE dışındaki tüm değişkenler index isimleridir. Bu isimleri değişken isimlerine çeviriniz.

İpucu:



## Görev 5: Age değişkenini kategorik değişkene çeviriniz ve agg\_df'e ekleyiniz.

- Age sayısal değişkenini kategorik değişkene çeviriniz.
- Aralıkları ikna edici şekilde oluşturunuz.
- Örneğin: '0\_18', '19\_23', '24\_30', '31\_40', '41\_70'

**Elde edilmesi gereken örnek çıktı:**

COUNTRY	SOURCE	SEX	AGE	PRICE	AGE_CAT
bra	android	male	46	59.0	41_70
usa	android	male	36	59.0	31_40
fra	android	female	24	59.0	24_30
usa	ios	male	32	54.0	31_40
deu	android	female	36	49.0	31_40

## Görev 6: Yeni seviye tabanlı müşterileri (persona) tanımlayınız.

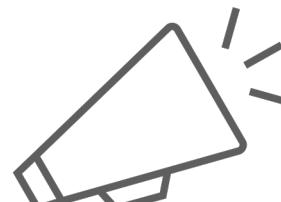
- Yeni **seviye tabanlı müşterileri (persona)** tanımlayınız ve veri setine değişken olarak ekleyiniz.
- Yeni eklenecek değişkenin adı: `customers_level_based`
- Önceki soruda elde edeceğiniz çıktıdaki gözlemleri bir araya getirerek **`customers_level_based`** değişkenini oluşturmanız gerekmektedir.

Bu tabloda bulunan gözlemler bir araya gelecek

COUNTRY	SOURCE	SEX	AGE	PRICE	AGE_CAT
bra	android	male	46	59.0	41_66
usa	android	male	36	59.0	31_40
fra	android	female	24	59.0	24_30
usa	ios	male	32	54.0	31_40
deu	ios	female	36	49.0	31_40

Elde edilmesi gereken çıktı

customers_level_based	PRICE
BRA_ANDROID_MALE_41_66	59.0
USA_ANDROID_MALE_31_40	59.0
FRA_ANDROID_FEMALE_24_30	59.0
USA_IOS_MALE_31_40	54.0
DEU_ANDROID_FEMALE_31_40	49.0

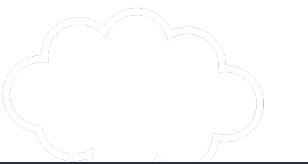


Dikkat! List comprehension ile `customers_level_based` değerleri oluşturulduktan sonra bu değerlerin tekillesştirilmesi gerekmektedir. Örneğin birden fazla şu ifadeden olabilir: USA\_ANDROID\_MALE\_0\_18. Bunları groupby'a alıp price ortalamalarını almak gerekmektedir.

## Görev 7: Yeni müşterileri (personaları) segmentlere ayırınız.

- Yeni müşterileri (Örnek: USA\_ANDROID\_MALE\_0\_18) **PRICE**'a göre 4 segmente ayırınız.
- Segmentleri **SEGMENT** isimlendirmesi ile değişken olarak **agg\_df**'e ekleyiniz.
- Segmentleri betimleyiniz (Segmentlere göre group by yapıp price mean, max, sum'larını alınız).

İpucu:



```
● ● ●  
pd.qcut(agg_df["PRICE"], 4, labels=["D", "C", "B", "A"])
```

## Görev 8: Yeni gelen müşterileri sınıflandırıp, ne kadar gelir getirebileceklerini tahmin ediniz.

- 33 yaşında ANDROID kullanan bir Türk kadını hangi segmente aittir ve ortalama ne kadar gelir kazandırması beklenir?
- 35 yaşında IOS kullanan bir Fransız kadını hangi segmente aittir ve ortalama ne kadar gelir kazandırması beklenir?

İpucu:



```
● ● ●  
new_user = "TUR_ANDROID_FEMALE_31_40"  
agg_df[agg_df["customers_level_based"] == new_user]
```

**miuul**

[miuul.com](http://miuul.com)