**İş Problemi:**

Dünyanın farklı yerlerinden müşterisi olan bir oyun şirketi, müşterilerinin bazı özelliklerini kullanarak seviye tabanlı müşteri tanımlamaları yapmak ve bu yeni müşteri tanımlarına göre segmentler oluşturup bu segmentlere göre yeni gelebilecek müşterilerin şirkete ortalama ne kadar kazandırabileceğini tahmin etmek istemektedir.

Veri setinde, kullanılan cihaz bilgisi, cinsiyet, ülke, yaş ve sağlanan kazanç bilgileri bulunmaktadır.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SOURCE | SEX | COUNTRY | AGE | PRICE |
| ios | male | bra | 17 | 39 |
| android | male | tur | 23 | 32 |
| android | female | fra | 16 | 43 |
| ios | male | bra | 32 | 34 |
| ios | female | ing | 47 | 45 |

*Veri setinin örnek bir bölümü*

**Çözüm:**

Veri setindeki “SOURCE”, “SEX”, “COUNTRY” ve “AGE” değişkenleri müşteriden müşteriye değişmektedir. Ancak tüm bu değişkenlerin aynı olduğu birçok müşteri de mevcuttur. Örneğin Türkiye’ de, erkek, 17-25 yaş arası Android kullanan birçok müşteri mevcuttur. Bu yaklaşımla yaş değişkeni de kategorilere ayrılarak her bir olasılığı bir karakter olarak tanımlayabiliriz. Böylece yeni gelen bir müşteri, tanımladığımız karakterlerden hangisi ile eşleşirse, yaklaşık o miktarda kazanç sağlayacaktır denilebilir.

**Adım-1:** Veri setinde boş ve yanlış değerler temizlenir. Yaş değişkeni belli sınıflara ayrılmalıdır.  *Pandas* kütüphanesindeki *cut* metodu kullanılarak istediğimiz aralıklara göre yaş değişkeni sınıflara ayrılır. Ardından her bir sınıfa isim verilir.

* 0-18 Yaş Aralığı : ‘0\_18’ ,
* 19-23 Yaş Aralığı : ’19\_23’ ,
* 24-30 Yaş aralığı : ‘24\_30’ ,
* 31-40 Yaş aralığı : ‘31\_40’ ,
* 40 Yaş Üstü : ‘40+’ ,

olarak isimlenidirlebilir. Bu bilgiler veri setine “AGE\_INTERVAL” olarak eklenir.

**Adım-2:** “SOURCE”, “SEX”, “COUNTRY” ve “AGE\_INTERVAL” değişkenlerinin tüm kombinasyonları için ortalama kazanç hesabı yapılır. Bunu yapmak için tüm bu değişkenleri birlikte ifade eden yeni bir değişken oluşturulur. Tüm bu değişkenler yan yana yazılarak da elde edilebilir. Elde edilen yeni değişken “CUSTOMER\_LEVEL ” olarak veri setine eklenir.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SOURCE | SEX | COUNTRY | AGE | AGE\_INTERVAL | CUSTOMER\_LEVEL | PRICE |
| ios | male | bra | 17 | 0\_18 | ios\_male\_bra\_0\_18 | 39 |
| android | male | tur | 23 | 19\_23 | android\_male\_tur\_19\_23 | 32 |
| android | female | fra | 16 | 0\_18 | android\_female\_fra\_0\_18 | 43 |
| ios | male | bra | 32 | 31\_40 | ios\_male\_bra\_31\_40 | 34 |
| ios | female | ing | 47 | 40+ | ios\_female\_ing\_40+ | 45 |

*Veri setinin son hali, CUSTOMER\_LEVEL tüm bilgilerin bulunduğu kolon*

**Adım-3:** Oluşturduğumuz yeni CUSTOMER\_LEVEL değişkeni tüm bilgileri içermekte ve tüm olasılıkları barındırmaktadır. Bu değişkenin her bir değeri için ortalama hesabı yapılıp yeni bir dataframe’ e kaydedilir. Oluşan yeni dataframe PRICE değerlerine göre segmentlere ayrılır.

|  |  |  |
| --- | --- | --- |
| CUSTOMERS\_LEVEL | PRICE | SEGMENT |
| bra\_androıd\_female\_0-18 | 35.645303 | B |
| bra\_androıd\_female\_19-23 | 34.077340 | C |
| usa\_ıos\_male\_0-18 | 33.983495 | C |
| usa\_ıos\_male\_41+ | 35.750000 | A |
| fra\_ıos\_female\_31-40 | 32.818182 | C |

*Tahmin yapılırken kullanılacak dataframe*

**Adım-4**: Yeni kayıt yapan bir kullanıcının bilgileri, tüm bilgileri alınarak en son oluşturduğumuz dataframe’ de sorgulanabilecek şekilde kaydedilir. Alınan kayıt bu dataframe’ de sorgulanarak eşleşen PRICE değeri getirilir. Örneğin Amerika’ lı 17 yaşında bayan bir IOS kullanıcısının getireceği kazanç şöyle tahmin edilir:

> df[df[“CUSTOMERS\_LEVEL”] == “usa\_ios\_female\_0\_18”] ;

customers\_level\_based PRICE SEGMENT

USA\_IOS\_FEMALE\_0-18 34.18 B

Oluşan çıktıya göre bu müşterinin yaklaşık 34.18 birim kazanç getirmesi beklenir.

**Sonuç:**

Uluslararası bir oyun şirketinin verileri, kural tabanlı sınıflandırma yapılarak analiz edildi. Bunun sonucunda gelen her bir müşteri için, eşleştiği karaktere göre getiri tahmini yapılabilir hale geldi. Müşteriler kazanç miktarlarına göre segmentlere ayrıldı. Böylece şirket kaynaklarını yönlendirirken müşteriler arasından daha kolay seçim yapabilir. Ayrıca müşterilerin ne kadar kazandıracağını yaklaşık olarak tahmin ederek ileriye yönelik bütçe planlaması yapabilir