**Business Problem:**

A gaming company with customers from different parts of the world wants to make level-based customer definitions using some of the characteristics of its customers, create segments according to these new customer definitions, and estimate how much new customers can bring to the company on average based on these segments.

The data set includes information about the device used, gender, country, age and earnings.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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 |

*A sample of the dataset*

**Solution:**

The “SOURCE”, “SEX”, “COUNTRY” and “AGE” variables in the data set vary from customer to customer. However, there are many customers for whom all these variables are the same. For example, in Turkey, there are many male customers between the ages of 17-25 who use Android. With this approach, the age variable can be divided into categories and we can define each possibility as a character. Thus, it can be said that a new customer will earn approximately the same amount of profit if he/she matches one of the characters we have defined.

**Step-1:** Empty and incorrect values are cleared from the data set. The age variable should be divided into certain classes. Using the cut method in the Pandas library, the age variable is divided into classes according to the ranges we want. Then each class is named as follows

* Age Interval 0-18 : ‘0\_18’ ,
* Age Interval 19-23 : ’19\_23’ ,
* Age Interval 24-30 : ‘24\_30’ ,
* Age Interval 31-40 : ‘31\_40’ ,
* Age Interval More than 40 : ‘40+’ ,

This information is added to the data set as “AGE\_INTERVAL”.

**Step-2:** Average earnings are calculated for all combinations of the variables “SOURCE”, “SEX”, “COUNTRY” and “AGE\_INTERVAL”. To do this, a new variable is created that represents all these variables together. All these variables can also be obtained by writing them side by side. The resulting new variable is added to the data set as “CUSTOMER\_LEVEL”.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 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 |

*The final version of the data set, the CUSTOMER LEVEL containing all the information*

**Step-3:** The new CUSTOMER\_LEVEL variable we created contains all the information and all possibilities. The average is calculated for each value of this variable and saved in a new data frame. The resulting new data frame is divided into segments according to PRICE values.

|  |  |  |
| --- | --- | --- |
| 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 |

*Data frame to be used when making predictions*

**Step-4**: The information of a newly registered user is recorded in a way that can be queried in the last data frame we created by taking all the information. The received record is queried in this data frame and the matching PRICE value is returned. For example, the earnings of a 17-year-old female IOS user from America are estimated as follows:

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

customers\_level\_based PRICE SEGMENT

USA\_IOS\_FEMALE\_0-18 34.18 B

According to the resulting output, this customer is expected to bring approximately 34.18 units of profit.

**Conclusion:**

The data of an international gaming company was analyzed by rule-based classification. As a result, it became possible to estimate the return for each customer based on the character they matched. Customers were divided into segments based on their earnings amounts. Thus, the company can more easily choose among customers while directing its resources. It can also make future budget planning by approximately estimating how much customers will earn.