|  |  |
| --- | --- |
| **Logo, company name  Description automatically generated** | Graphical user interface, application  Description automatically generated |

Icon

Description automatically generated2b. Linear Regression

File: loyalty-cards-inc.csv and loyalty-cards-inc-new-users.csv

Great news. We have obtained a new dataset of customer leads for the loyalty program. It’s actually a sample that the vendor made available to see if we are interested in a larger dataset. Our goal is to predict the amount of rewards redeemed by each customer in this new dataset. Please see the specific requests below:

1. Let’s first find whether location and gender predict rewards redeemed. (We care about this variable since we obtain a commission from the retailer on each redemption). How does each variable affect predicted rewards redeemed?
2. Is the fit of the last regression good? What are the R-Squared, Root Mean Squared Error and the Mean Absolute Error?
3. Please also find whether location and gender predict rewards earned. What are the conclusions here?
4. Let’s forecast the rewards redeemed of the potential loyalty card subscribers in the new sample dataset. (Note: It may be useful to know that the LivesNearby variable in the original dataset is 1 when the customer lives within 5 Km of the supermarket.)
5. We earn 5 cents for each euro of rewards redeemed by customers. Assuming zero marketing costs and a 10% marketing effectiveness (i.e., new program subscriptions upon receiving promotional emails), what is our willingness to pay per dataset record?