

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

2a. Wrangling Data

File: loyalty-cards-inc.csv and loyalty-cards-inc-visits.csv

Let’s introduce a few changes to the current dataset and merge it with the old loyalty program data. The goal is to assess spend, rewards obtained and redemption rates of different types of people. Start by importing process “Week2a”, and dataset loyalty-cards-inc-visits.csv.

1. Let’s introduce a new column calculating the percent rewards redeemed so far.
2. Let’s focus only on customers who spent at least 200 euros in the store.
3. GDPR is out there. Let’s get rid of customer names
4. Merge the dataset with file loyalty-cards-inc-visits.csv. This will enable us to add a column containing the number of customer visits to our store.
5. Create new unique ID variable and scrap the old one. Make sure the new variable is named “ID”, not “id”. (This is often useful when customer id’s are not consecutive, for example.)
6. Finally, create a new variable that contains the average spend for the person’s gender. This means all people from each gender will have the same value in this variable.