# SAS Class 2: Exercises

\*\*\*\*\*\* PART 1 \*\*\*\*\*\*

1. Convert the celcius to farenheit in the excel file of ex1. Make sure it is rounded to 2 decimal points. The relation between Farenheit and Celcius is as follows:
2. Round all numeric variables in Ex2 to 3 decimals. Try to find the most efficient code.

\*\*\*\*\*\* PART 2 \*\*\*\*\*\*

**Information about the SAS dataset:**

The excel file “Sales.xls” is a subset from a database of a Belgian store. The goal is to get an idea about the total sales per month. The dataset contains the following variables:

|  |  |
| --- | --- |
| cust\_id | the customer id |
| date | date of purchase |
| cash | dummy variable indicating whether  the person paid cash (1) or not (0). |
| prod\_id | the product id |
| quantity | the amount that was bought |

* 1. Read in the excel file (use PROC IMPORT) and store it as a temporary SAS table named “Sales”.
  2. ID variables are generally seen as character variables and not as numeric ones. Convert it to the correct format. Second, sort the dataset on product id (low 🡪 high), date (chronological) and customer id (high 🡪 low). Make sure the rows with missing[[1]](#footnote-1) product ids come last (creating an additional variable could do the trick!) and make sure every row is unique! Save this new dataset as “Sales\_sorted”.
  3. Create three additional variables based on the dataset “sales\_sorted”. The first variable “price” should represent the price per unit for a given product id. You can find the corresponding prices in the table below.

|  |  |
| --- | --- |
| prod\_id | price |
| 56 | 1,55 |
| 42 | 6,99 |
| 86 | 5,45 |
| 91 | 6,77 |

The second variable “tot\_price” is calculated by multiplying the quantity with the price. We also want you to create a variable “month” that indicates the month of the year. Store the new dataset as “sales\_price”.

* 1. Now we will try to identify the month with the highest sales. In order to do so, we make an overview of the total sales per month.
  2. Show the relation between cash purchases and the month.
  3. Suppose you work for the quality control department. You want to contact 3 random customers to know whether they were satisfied with the company products. Select five random customer id’s from the table “sales\_sorted” and store them in a new dataset. In order to create a random sample, you have to use the UNIFORM-function. Use SAS HELP (or google) to learn how to use this function.
  4. Use the original imported sales file for this exercise. Calculate for each customer, the difference between a cash payment and a previous cash payment of that customer. If there is no previous cash payment, the difference should be set to zero.

\*\*\*\*\*\* PART 3 \*\*\*\*\*\*

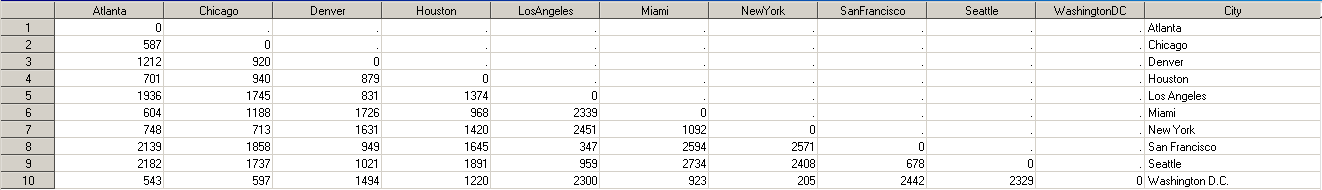
We have three datasets from a Belgian retailer. The ‘tickets.xlsx’ file contains ticket information of customers such as the amount they bought and what they paid. The ‘Customer\_info.xlsx’ file contains some information on the customers that have a customer loyalty card. Finally, the ‘External\_card.xlsx’ file contains information provided by an external company. The retailer has an arrangement with this company. If a customer scans its card of the external company during his visit to a store, he receives loyalty points at that external company. In addition, the company provides the retailer with some additional customer information on these customers. The retailer has to pay a fixed price (which is quite high) for this arrangement, as well as provide their customer information. The reason they decided to go the cooperation is that the external company has e-mail address information which they can use for marketing campaigns.

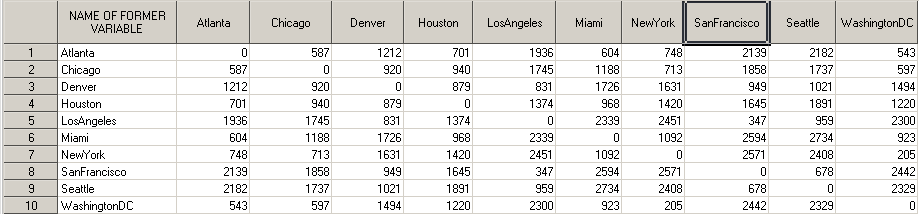
First summarize the tickets dataset so that it contains 1 line per customers (so it will be the average ticket amount per customer etc.). Also clean up the “Web\_Shop\_Client” variable to only contain the values “yes” and “no”. Merge the different excel files to create the following three datasets: 1) a dataset that contains those customers that have bought something according to the tickets file and have additional client information, but no external card information, 2) a dataset that contains those customers that have bought something according to the tickets file and have additional client information and external card information, 3) a dataset containing the remaining customers.

Decide based on these three created datasets and the information you have whether the cooperation should be continued or not. Try to be as detailed as possible exploiting the data you have.

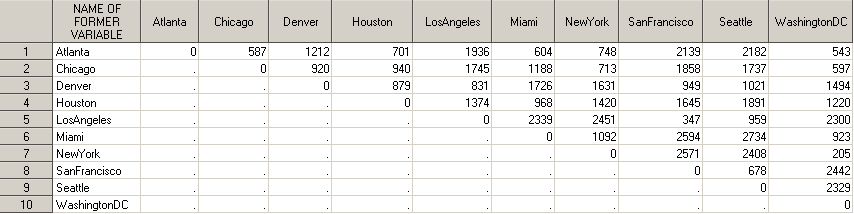
\*\*\*\*\*\*\* PAR T4 \*\*\*\*\*\*

Take a look at the “SASHELP.Mileages” dataset.



This is a lower triangular distance matrix. The goal is to make it a triangular matrix as below:

To do this, you first want to create the following upper triangular matrix (which will already give you part of the points for this question):



µ

Next, the original and new dataset can be set together to create the triangular matrix.

1. Note: Missing values in SAS are considered to be smaller than all numbers [↑](#footnote-ref-1)