Lab 1: Data pre-processing

For the following exercises, work with the bank_marketing_training data set.

In this exercice we will focus on the following data preparation tasks:

1) Adding an index field
   - Create a new variable that assigns every record a unique integer.
2) Changing misleading field values
   - Find each instance of 999 in the days_since_previous variable and replaces it with the value NaN.
   - Create a histogram of the variable, use the hist() command.
3) Reexpressing categorical data as numeric data
   - Replicate the education variable, and name it education_numeric, next replacing its categorical values with numeric ones such as shown in the following table.
   - Plot a bar graph for the education variable

Categorical Value	Numeric Value
illiterate	0
basic.4y	4
basic.6y	6
basic.9y	9
high.school	12
professional.course	12[a]
university.degree	16
unknown	Missing

4) Standardizing the numeric fields
   - Standardize the age variable and save it as a new variable, age_z.
   - Plot a histogram graph for age and age_z
   - Check the skewness of the variable days_since_previous
5) Identifying outliers.
   - Find outliers in age variable by using the query() function, which identifies rows that meet a particular condition.
   - Report the age and marital status of the 15 people who have the largest age_z values.