# **Project 11**

Data: ”employees”, ”travel\_expense”

Aim: create report in rtf.

Operations: Import the datasets in SAS. Create a dataset named “All” that contains only the employees for whom a travel expense was done and which contains the following information:

* JOB\_DEP: obtained concatenating Job\_Title and Department of each employee, separated by a blank space
* EMPLOYEE\_ID: identifier of the employee
* GENDER: employee gender
* GENDER\_CD: is equal to 0 for males and 1 for females
* SALARY: format this variable with the dollar sign $
* BIRTH\_DATE e HIRE\_DATE: format these dates as 14SEP2019.
* HIRE\_AGE: age in which the employee was hired; to be computed as the difference between HIRE\_DATE and BIRTH\_DATE and divide it by 365.25.
* JOB\_TITLE: job role for each employee
* TRIP\_ID: code identifier of the business trip
* EXP1 – EXP5: expenses for the business trip
* Using an array, create a character variable for each expense (name them EXP\_C1 EXP \_C2 EXP \_C3 EXP \_C4 EXP \_C5) that contains the character version of the expense, using the following categorization\*:

|  |  |
| --- | --- |
| Expense (numeric) | Expense (character) |
| <100 | Very low expense |
| >=100 and <300 | Low expense |
| >=300 and <500 | High expense |
| >=500 | Very high expense |
| . | No expenses |

\* use the following statement to assign a length of 25 letters to the new character variables:

length EXP\_C1 EXP \_C2 EXP \_C3 EXP \_C4 EXP \_C5 $ **20**;

Create a report excluding variables EXP1-EXP5. Consider all employees except the ones with Job\_Title ending with “II”. Assign a title of your choice.

Create a table that summarizes the mean, min, max, sd, and n related to the HIRE\_AGE grouped by SEX. Briefly discuss results, e.g. there seems to be a difference between males and females in the HIRE\_AGE?

Then, impute missing values of exp1 – exp5 by using the Last Observation Carried Forward (LOCF) method. Once imputed the missing values, compute the total expenses for each trip\_id. Finally, compute the contribution of each expense to the total amount, in percentage, creating appropriate variables (perc1, perc2, …). Create a report that shows EMPLOYEE\_ID, the imputed exp1 – exp5, total expenses, and perc1-perc5.