IMPORTING PROCESS:

In the importing process I didn't carry out any special code. I imported all the tables manually, All the tables have been imported, replacing the "." for "_" both in the tables 'names and the columns´ names.

I can highlight that for each specific table I created a table copy to work through it instead of using the original one.

DISCOVER INSIGHTS PROCESS:

In this analysis I carried out several steps.

The first one, I tried to organise all the data to categorize each call with their kind of call_type, I did this manipulating the calls_copy table and creating different tables for all the different types (international, roaming, etc.). for those calls that were neither international nor roaming I created a category called Fucto. Then, for all those calls with fucto category I created another table to store them. The next step was to calculate the weekday of the call to be able to classify them into peak/off-peak categories.

I calculated the day of the week and this is a really important point to consider: I managed to convert it into days of the week "names" but unfortunately I haven't been able to translate it into English with another command nor changed it through "tools" in SQL developer. So, I have to clarify that, starting from Monday, days of the week in my language are (**Lunes, Martes, Miercoles, Jueves, Viernes, Sábado, Domingo**).

So, then I create a table called fucto_days_1 combining the calls classified as fucto(no international no roaming) and I fusion them with the days_table(with the weekdays already) and classified the calls into(peak/nopeak) and as you can se in the code that in the functions **LIKE 'D%'** / **NO LIKE 'S%'** the letters correspond to Dominso(Sunday) and Sabado(Satuday).

Then I have been combining all the tables to get the specific data required, and then I calculated the main points for the analysis:

NUMBEROFCALLS: number of total calls did for each client.

MAX_CALL_COST: maximum cost of all the calls did for each client.

AVR_CALL_DURATION: average of minutes of all the calls done.

AVR_CALL_COST: average of cost of all the calls done.

SERVICE_CALLS_NUM: number of calls to the customer service.

AVG_CS_DURATION: average of minutes of all the calls done to the customer service.

VOICEMAILS_NUM: number of voicemails left.

AVG_VOICE_DURATION: average of minutes of all voicemails left.

Once I calculate all these variables for each customer, I put together all of them with the table customers (containing start_date_contract, end_day_contract, Plan, etc.) and having all in a table called CUSTOMERS_TOTAL_TABLE. In his table, I assume that the ID of each client is their PHONE_NUMBER.

In order to capture the output for 3 separate users, I added some code to randomly show all the insights of 3 random clients.

BUILD A MODEL

First of all, I had to modify a bit the main table for the model which is the CUSTOMERS_TOTAL_TABLE. I had to modify some columns in order to make it working, and specially amend the date type columns, converting them into moth_year.

Then I create the test and apply tables based on the target variable (end of the contract) considering those nulls and those no nulls.

I managed to build a model following the steps from the lectures. I tested and the model is running fine even if getting the results, I just figured out that the data was wrong and some predictions doesn't make sense. I tried to amend the data but it was giving me some errors, so I discarded it and I assuming is more important the process and control of commands that the content itself for the assignment.