Did you learn anything of potential business value from this analysis?

In this module I learned about *data science frameworks:* When you don't know for sure the problem to be solved or the prediction to be made, and you need to understand the business process, identify the problem and the most suitable solution for the modeling, the use of those frameworks are excellent tools to organize not only the data but the ideas thorough next methodologies:

- Business understanding: Define business problem, goals and criteria for success.
- Data understanding: Categorize the model (Classification/Regression), analyze the attributes and their relationships.
- Data preparation: Handle missing values, outliers.
- Modeling: Apply multiple models and choose the most optimal one.
- Evaluation: Evaluate the models avoiding over-fitting.
- Deployment: Prepare PROD plan, monitoring and support.

Summarizing the data with specific facts, hypothesis, graphs and matrixes improves the capacity to understand the final product and make better and faster decisions.

What are the main lessons you've learned from this experience?

Python is an interpreted language, which can save you considerable time during program development:

- You could write a Unix shell script or Windows batch files.
- No compilation and linking is necessary.
- No variable or argument declarations are necessary.
- High-level data types allow you to express complex operations in a single statement.
- Easy to read: Statement grouping is done by indentation instead of beginning and ending brackets.

What recommendations would you give to the Guido regarding your findings?

Once the main population that is defaulted is identified, some extra requirements must be defined to assure that the customer would be able to pay the loan each month like: How many members of the family depend on the income of the potential customer? Is it the only support of the family? Work full-time or part-time? Any other debts? What are previous scores for loans canceled? How much money is spent by month? What's the month salary?

Keep in mind that higher scores mean fewer defaults and vice versa, so collect any additional data that can help to create better patterns to calculate more precisely the customer's score. If those scores are very close to reality, it would be easier to estimate the probability to get defaulted or not, for example using a lot of the available credit, may indicate that the person is overextended.

Pay attention to extreme values that don't reflect the reality in the data, this may cause that the predicting model generate false positive results, it is highly recommended to identify those extreme outliers and remove the dataset.

Group continuous data on categories to make easier the interpretation, such as setting up the ages by ranges.