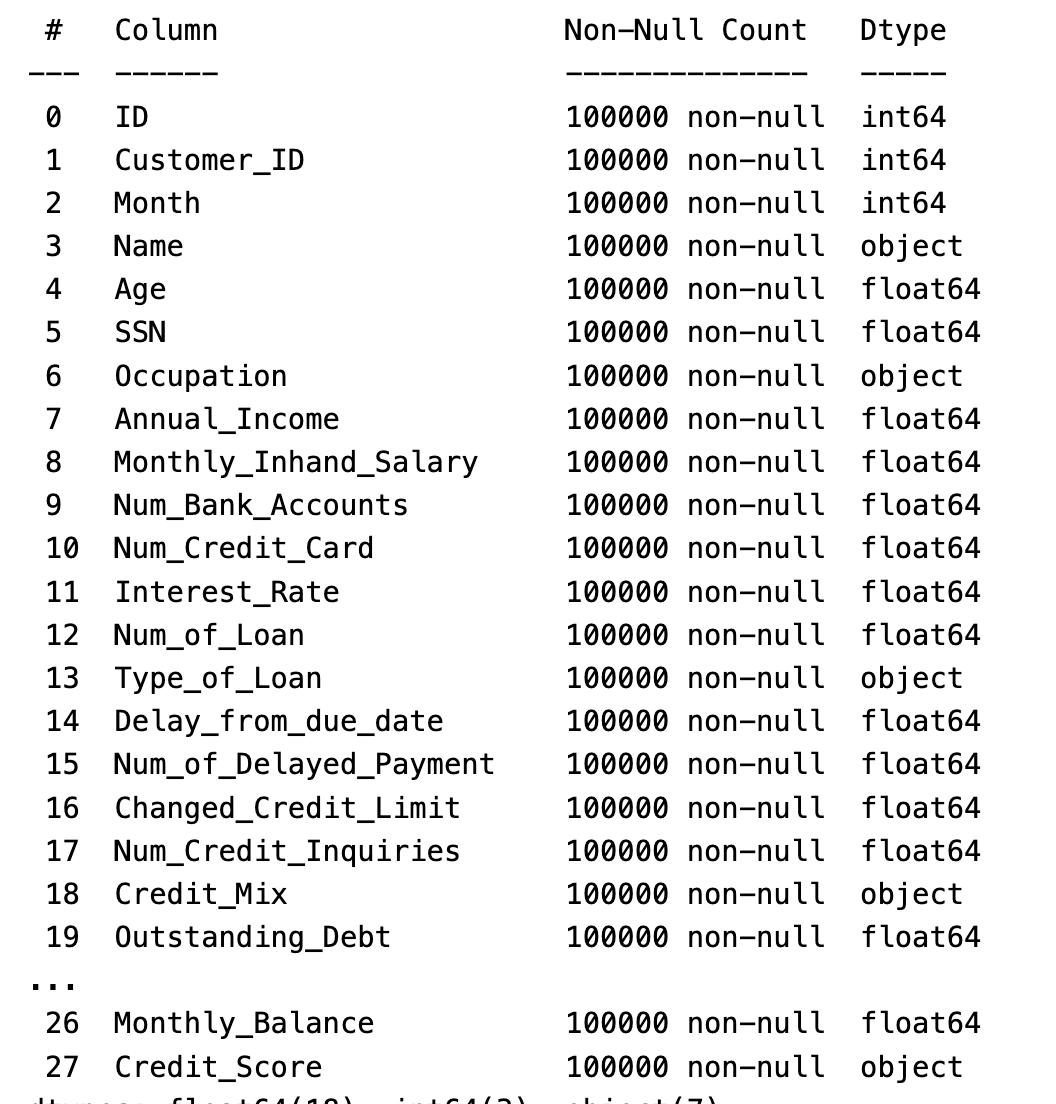
**Anexo. Traditional Credit Risk Model**

**Cleaning process and EDA**

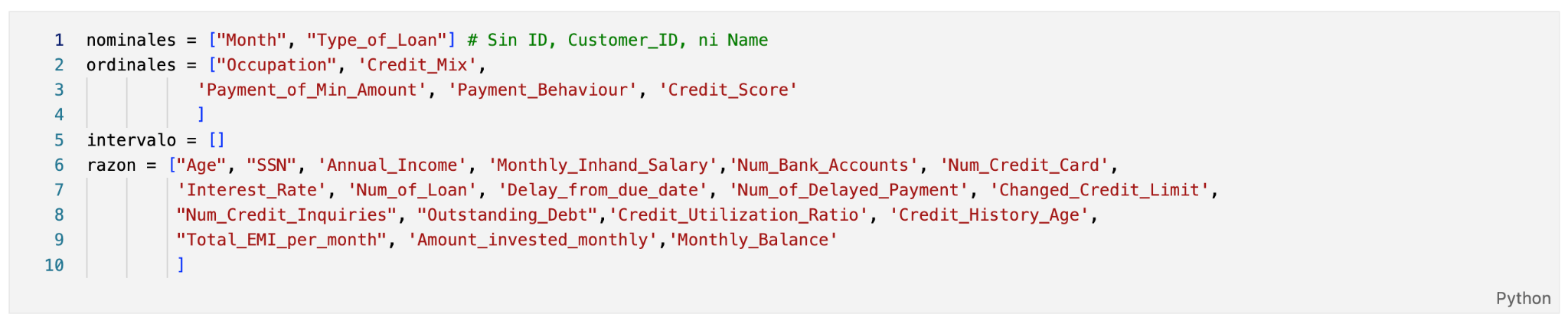
As part of the data cleaning process, we conducted the following steps. Some of this analysis was not even later used or taken into consideration for our models, but they were made to ensure good practice and to leave the bases for a possible project continuation and/or improvement.

1. **Handling Missing Values** – We checked for null values across all variables.



In this first result, apparently there were no null values, although each column was still analyzed to see if there were no “garbage” or incorrect values.

1. **Variable Classification** – We categorized the variables into nominal, ordinal, interval, and ratio scales.



1. **Cardinality Analysis** – For nominal and ordinal variables, we examined unique values to assess whether they exhibited high or low cardinality.
   1. **Low cardinality:** “Credit\_Mix”, “Payment\_of\_Min\_Amount”, “Payment\_Behaviour”, “Credit\_Score”
   2. **High cardinality:** ”Type\_of\_Loan”, “Occupation”
2. **Proportional Distribution** – We analyzed the proportion of each unique value within the nominal and ordinal variables.
3. **Duplicate Detection** – We checked for duplicate records to ensure data integrity. The result was 0 duplicates.
4. **Histogram Analysis –** To ensure we selected the best variables for our Traditional Credit Model, we plotted each of our ratio ratio variables in which each of the three Scorecard groups (bad, standard, good) was contrasted. This way, an analysis was made to determine the best possible predicting variables.

