Dear Sprocket Central Pty Ltd team,

thank you for trusting in our service. I had a look in your worksheets. The quality of the data should be improved in the following area, so it can be used for the analysis:

In the “Customer Demographic” sheet I found missing values in 6 columns. For the “DOB” and “tenure” I would suggest replacing the empty values with the median value from their column. Also, in the “DOB” there is an accuracy error. The “DOB” of customer Jephthah Bachmann is “1843-12-11”. This is probably a typo or a very old entry.

The missing “job\_titles” and “job\_industry\_category” I would group in a new category and name it “others” or “miscellanies”. Then I would remove the “default” column, because the data have no relevance for me. The values are decoded and unreadable. I would help me, if you can deliver me more information about this column. The missing values in “last\_name” we don’t need to edit, because we can use instead the customer id.

Also, I noticed in the “gender” column some inconsistency. The database uses different abbreviation for the same category. For example, “F”, “Femal” and “Female”. So, I would edit the column, that only 3 categories exist: “F”, “U” and “M”.

In the Transactions table are also missing values, which could be filled with the upper suggestions. But the “product\_first\_sold\_date” column has a validly problem. The values should be dates and not numbers. It’s necessary to transform them to dates.

The Customer Address sheet has different notations in the “state” column similar to the gender column from the Customer Demographic sheet. Sometimes they used an abbreviation of the state and other times they used the full name.

I hoped this gave you an overview of my work.

I’m looking forward to hear from you.

Best regards

Daniel Schlon