**Data Quality Assessment**

Try to explain every issue find in the datasets.

**Transaction dataset**

1. In the column online\_order, there are 360 rows (1.8% total columns) with null values. Is it correct that these rows haven't a value? Try to find the why of these missing values (problem on data collection); in the meantime, we can probably choose between dropping these rows or fill these values for our analysis.
2. The columns brand, product\_line, product\_class, product\_size, standard\_cost and product\_first\_sold\_date have 197 null values. Every row with these null values has the product\_id = 0; this product\_id is present in other rows with these columns filled with values. The same as the last case, we can probably drop these rows and try to dig deeper to find the problem behind these missing values.
3. The column product\_first\_sold\_date is a number. It may probably be a DateTime column.
4. There are the same products that have different list\_price and standard\_cost.

**Customer demographic dataset**

1. The columns last\_name, DOB, job\_title, job\_industry\_category, default, tenure contain missing values. In this case, the percentage of rows with null values is relevant, and it isn't a good idea to drop these rows.
2. The gender column has six values (Female, Male, U, F, M, Femal). We need to understand what means "U" and correct the other three misspelled errors. Otherwise, it can be possible to force the input into four different categories: Male, Female, Other, Not Declared.
3. The default column contains values encoded incorrectly. Specify the correct encoding when ingesting or when export the data.
4. The max values of "past\_3\_years\_bike\_related\_purchases" is 99. Try to find out if it's a cap for the column of it's the real value.

**Customer address dataset**

1. The column state has two values written in an extended and abbreviated way. Try to have this value from a default list with the same format in the data collection phase.
2. Is the property column a rank from 1 to 12? What's the meaning of the value 1 and the value 12?