

Dear client

Thank you for the provided datasets. After the analysis I made on the datasets, here are the quality issues I have discovered.

There is incompleteness in *customer_id* between the Customer Demographic, Customer Address and Transactions columns. In the Customer Demographic there are 4000 entries, in the Customer Address 4003 and in the Transactions 3500 entries.

Suggestion: Update all the tables on regular basis, for example once per day or once per week.

Customer Demographic Table

- There are several missing values in *last_name*, *job_title*, *DOB*, *job_industry_category* and *tenure* columns.
- There are consistency issues in *gender* column. All entities in this column should be displayed as “M” for Male, “F” for Female or “U”.
- The *default* column is irrelevant and it must be dropped.
- There is a validity issue in *DOB* column. We have a date of birth that refers to the previous century.

Customer Address Table

- There is a consistency problem in *state* column. All states must be displayed as “New South Wales”, “Victoria” and “Quennsland”.

Transactions Table

- There are many missing values in *online_order*, *brand*, *product_line*, *product_class*, *product_size*, *standard_cost* and in *product_first_date* columns.
- The *list_price* column does not in right format. It should be in Currency format.
- The *product_first_sold_date* column is also not in the appropriate format. It should be Date.
- In *order_status* the Cancelled orders are not needed.
- There is missing a “Profit” column and it must be added.

During this Quality Analysis, I cleared the data so they will be ready for analysis.

Please let me know if you have any questions regarding mitigation.

Sincerely,

Lazaros Panagiotoulas