|  |
| --- |
| To: KPMG companies |
| From: Nick Nguyen |
| Subject: Data quality assessment |
| Dear KPMG, This email include my part to identify the quality issues and strategies to mitigate these problem.  I look forward to your responses.  Sincerely,  Nick Nguyen |

# Transaction

* Problems 1: Missing values

A picture containing table

Description automatically generated

Table

Description automatically generated

There are missing values in online\_order (360 missing values), brand (197 missing values), product\_line (197 missing values), product\_class (197 missing values), product\_size (197 missing values), standard\_cost (197 missing values), product\_first \_sold\_date (197 missing values).

* Solution 1: Drop missing values

Text

Description automatically generated with low confidence

# New Customer List

* Problem 1: Missing values

Table

Description automatically generated

There are some missing values last\_name (29 missing values), DOB(17 missing values), job\_title(106 missing values), job\_industry\_category (165 missing values)

* Solution 1: Drop missing values

Table

Description automatically generated with medium confidence

* Problem 2: The unclear data

Table

Description automatically generated with medium confidence

When check values of genders, U means the input error, or unknown gender or the customer does not want to mention.

* Solution 2: Fix the values

Because of “U” value containing missing values in other columns. When we remove missing values, we also remove “U” value.

* Problem 3: Loss of column name

Table

Description automatically generated

There are some columns missing the column\_name

* Solution 3: Remove these columns

Table

Description automatically generated

# Customer Demographic

* Problem 1: Missing values

Graphical user interface, text

Description automatically generated

Table

Description automatically generated

There are some missing values last\_name(125 missing values), DOB(87 missing values), job\_title(506 missing values), job\_industry\_cateogry(656 missing values), default(302 missing values), tenure(87 missing values)

* Solution 1: Drop missing values

Graphical user interface, text, application

Description automatically generated

* Problem 2: Unclear values

Table

Description automatically generated

There are the informality of values Females and F, Males and M. U means the input error, or unknown gender or the customer does not want to mention

* Solution 2: Fix values

We will replace “M”, ” F”, ”Femal” value bye “Male” and “Female”. When we remove missing values, we also remove “U” and “M value.

Graphical user interface, text, application, email

Description automatically generated

* Problem 3: Irrelevant column

Text

Description automatically generated

* Solution 3: Drop the “Default” column

Graphical user interface, text, application, email

Description automatically generated

* Problem 4: Strange values

Graphical user interface, text, application, email

Description automatically generated

The customer (Jephthah Bachmann, you'll notice that they were born in 1843, meaning that they are 175 years old!

* Solution 4: Fix the values

After removing values, the value of Jephthah Bachmann is removed too

# Customer address

* Problem 1: Unclear values

Text

Description automatically generated with low confidence

* Solution 1: Fix the values

Graphical user interface, text

Description automatically generated