Good day Sir/Ma,

Regarding our recent project, after our quality assessment on the data received from you, we encountered some data quality issues which will be outlined in the email as well as how it can affect our analysis and way to mitigate these issues.

The summary of the data is as below

Table	Columns	Rows
CUSTOMER DEMOGRAPHICS	13	4000
CUSTOMER ADDRESSES	6	3999
TRANSACTIONS	13	20000

The following are data quality issues to be reviewed

MISSING VALUES

- CUSTOMER DEMOGRAPHICS - Missing data in Last name, DoB, job title, job industry category, default, tenure
- TRANSCATION has missing data in online order, brand, product line, product class, product size, product cost and product first sold date
- Missing ID

Solution – Review original master sheet for missing data. If missing data is less that 1%, drop else find a strategy with the existing data.

INCOHERENT & INCONSISTENT DATA

- Default column in CUSTOMER DEMOGRAPHICS have incoherent and meaningless data.
- Gender column in CUSTOMER DEMOGRAPHICS have a gender "U". Please, input for this should be validated.
- Date column in CUSTOMER DEMOGRAPHICS has an outlier. A DoB stood out and look like and outlier. A customer over 150 years seems suspicious. Please validate if its 1843 or 1943.

WRONG DATA TYPES

- Product first sold data column in Transactions have inappropriate data type.

To address these are the steps we propose in order to mitigate these data quality issues

- Review the original data if it was extracted rightly, if not we collect the right values,
- Else we fill the missing values and if they cannot be filled, we will drop these from our analysis.
- For inconsistent value we convert them to their appropriate groups
- Investigate outliers, correct or remove
- Verify the duplicate addresses

If you agree with this approach, please indicate so we can go along with it.

Thank You,

Gilbert Victor