Goodday 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

The data quality for Customer Addresses and Transactions data in the past 3 months had fairly acceptable data quality issues. The customer demographic however had more quality issues.

We encountered data issues such as

State problems along with solutions/approach

* Missing Data – several columns has missing data fields
* Inconsistency - Columns had the wrong the wrong input for its Fields. Eg in the gender column
* Invalid or suspicious data - A DoB stood out and look like and outlier. A customer over 150 years seems suspicious.
* There was missing customer IDs
* Duplicate and inconsistent address
* Wrong date Formats.

These data quality issues might skew our analysis and give a fairly unreliable analysis and conclusion.

In order to avoid this, These are the steeps we proposes 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