Subject: Identification of Data Quality Issues in Provided Spreadsheets

Dear KPMG,

I hope this email finds you well.

As part of our ongoing efforts to ensure data integrity and accuracy, I have conducted a thorough review of the provided spreadsheets: "CustomerDemographic," "CustomerAddress," "Transactions," and "NewCustomerList." Below, I have detailed the data quality issues identified in each sheet:

1. **Transactions:**

*list\_price* has the wrong number format.

*product\_first\_sold\_date* was in number format instead of date format.

Missing values in *online\_order, brand, product\_line, product\_class,* and *product\_size.*

1. **NewCustomerList:**

Missing values for the next 4 rows after the *property\_evaluation column.*

Wrong number format.

Missing data on *last\_name*, *job\_title*, and *job\_industry\_category*.

1. **CustomerDemographic:**

Missing values on *last\_name*, *DOB*, and *job\_title.*

‘n/a’ values in *job\_industry\_category.*

*gender* column has 'Femal' instead of 'female' or 'F' and contains both 'Male' and 'M' values.

*default* column does not have human-readable datasets.

1. **CustomerAddress:**

*state* column has inconsistent data: ‘NSW’ and ‘New South Wales’, ‘VIC’ and ‘Victoria’, only ‘QLD’.

Addressing these data quality issues is crucial for improving our data reliability and ensuring more accurate analysis and reporting.

I recommend the following steps to resolve these issues:

**For Transactions:**

Correct the number format for *list\_price*.

Convert *product\_first\_sold\_date* to the appropriate date format.

Fill missing values in *online\_order, brand, product\_line, product\_class,* and *product\_size.*

**For NewCustomerList:**

Fill the missing values for the rows following the *property\_evaluation* column.

Correct the number format.

Update missing data on *last\_name, job\_title,* and *job\_industry\_category.*

**For CustomerDemographic:**

Address missing values in *last\_name, DOB,* and *job\_title.*

Replace n/a values in *job\_industry\_category.*

Standardize entries in the *gender* column.

Ensure *default* column entries are human-readable.

**For CustomerAddress:**

Standardize state names to ensure consistency (e.g., using either abbreviations or full names consistently).

I am available to discuss these findings and the next steps at your convenience. Please let me know a suitable time for a meeting, or if you have any questions or need further information.

Thank you for your attention to these matters.

Best regards,

Chigozie Uba

Junior Data Analyst