Subject: Addressing Data Quality Issues and Strategies for Mitigation

Dear Sir/Madam,

I hope this email finds you well. During our analysis of the tables, we have identified several data quality issues that require attention. To ensure the integrity and reliability of the data, we recommend implementing the following strategies to mitigate these issues:

Table: **TRANSACTIONS**

Data Quality Issues:

Date Format for Product First Sold Date: We noticed that the date format for the product first sold date was incorrect. We recommend updating the date format to ensure consistency and accuracy throughout the table.

Uniformity of Product First Sold Date: It appears that the same date is recorded for the product first sold in all transactions. This could indicate a data entry error or missing information. We recommend investigating and updating the product first sold date with the correct information.

Null Values in Multiple Columns: We observed the presence of null values in columns such as online_order, brand, product_line, product_class, product_size, standard_cost, and product_first_sold_date. To mitigate this issue, we suggest addressing these null values by either imputing missing data or considering alternative sources to complete the information.

Creation of a Profit Column: We have introduced a new column named "profit" to calculate and track the profit for each transaction. This addition will enhance the analysis and provide valuable insights into the profitability of products.

Table: **Customer Demographic**

Data Quality Issues:

Null Values in Multiple Columns: We noticed the presence of null values in the last_name, DOB, job_title, job_industry_category, and tenure columns. To address this, we recommend handling these null values by either imputing missing data or considering alternative data sources to complete the information accurately.

Correction of Gender Values: In the gender column, we observed unnecessary values such as "F" and "Femal." To ensure consistency, we recommend correcting these values to "Female." Additionally, we suggest replacing the value "U" with "Unrecognized" to indicate cases where the gender is unidentified.

Removal of Irrelevant Default Column: We suggest removing the "default" column as it contains irrelevant content with symbols that may not contribute to the analysis or insights.

Creation of an Age Column: We have introduced a new column named "age" based on the available information. This addition will provide valuable insights for demographic analysis and segmentation.

Table: **CustomerAddresses**

Data Quality Issues:

Consistency in State Abbreviations: To ensure uniformity, we recommend replacing the abbreviations "NSW," "VIC," and "QLD" with their expanded