**Subject: Addressing Data Quality Issues and Implementing Mitigation Strategies**

Dear Sprocket Central Team,

I hope this email finds you well. I wanted to bring to your attention some significant data quality issues we have identified within the datasets you provided. We understand the importance of accurate and reliable data for your business decisions, and we have developed strategies to mitigate these issues effectively.

**Transactions Data**:

1. Missing Mode of Delivery: We noticed that 360 records in the Transactions data are missing information regarding the mode of delivery (online or not). To address this, we recommend implementing a data validation process that ensures this field is completed for all transactions. This can be done through mandatory data entry fields or automated checks during data ingestion.
2. Missing Product Details: In 197 records, essential product details such as brand, product line, product class, product size, standard cost, and product first sold date are missing. To mitigate this issue, we propose conducting a thorough data cleansing exercise that involves cross-referencing the available information with trusted sources or contacting the respective departments to obtain the missing data. Establishing data collection protocols moving forward will also help maintain complete product information.

**New Customer List:**

1. Incomplete Customer Information: Within the New Customer List, we found 29 records with missing last names, 106 records without job titles, 165 records with "N/A" in the job industry category field, and 17 missing dates of birth (DOB). To address these issues, we suggest implementing data validation checks during the data entry process to ensure all mandatory fields are completed. Additionally, conducting a thorough data audit and contacting the customers directly to obtain the missing information will help improve the overall data quality.
2. Clarification on Gender 'U': We noticed that the dataset includes a gender code 'U,' and we would appreciate clarification regarding its meaning. Is it intended to represent "others" or does it indicate a lack of available data for gender? This clarification will assist us in accurately interpreting the data and maintaining consistency in our analysis.

**Customer Demographic:**

1. Incomplete Customer Information: The Customer Demographic dataset exhibits several data quality issues, including 125 missing last names, 87 missing DOBs, 506 missing job titles, 656 instances of "N/A" in the job industry category field, and 87 missing tenure information. Similar to the previous dataset, we recommend implementing data validation checks, conducting data audits, and contacting customers directly to obtain the missing information. This proactive approach will significantly improve the data completeness and accuracy.
2. Clarification on Gender 'U': As observed in the previous dataset, the Customer Demographic dataset also includes the gender code 'U.' We kindly request clarification regarding its meaning to ensure accurate interpretation and analysis of the data.
3. Incorrect DOB and Inappropriate Data in the Default Column: We identified one record in the Customer Demographic dataset with an incorrect DOB of 1843-12-21, which is not feasible based on the current year. Additionally, the Default column contains some inappropriate data, and we would appreciate clarification on its intended meaning to properly handle this column during analysis.Bottom of Form

Please let us know your thoughts on these strategies, and if you have any additional suggestions or concerns. We are eager to collaborate with you in resolving these data quality challenges and look forward to discussing this further.

Thank you for your attention to this matter. We value our partnership and remain dedicated to delivering accurate and actionable insights.

Best regards,

Jayant Yadav