Subject: Data Quality Analysis Report for SP Rocket Central Pty Limited

Dear [Name of the Client],

I hope this message finds you well. We have successfully received the three raw datasets from SP Rocket Central Pty Limited. As part of our preliminary analysis, we have thoroughly examined the data for quality assurance. Below, you will find a summary of our findings, along with corresponding recommendations to enhance the overall data effectiveness:

1. **Redundant Outliers:**

- **Issue:** We have identified certain data points that fall outside the expected range, potentially impacting the dataset integrity. For instance, customer ID "34" with the name of Jephthah Bachmann was recorded with a birth year of 1843, indicating an age of 175 years—an evident data discrepancy.

- **Recommendation:** We advise the removal of such redundant data to preserve the integrity and distribution of the dataset.

2. **Missing Values:**

- **Issue:** Several attributes, including "Online Order," "Brand Name," "Product Line," "Product Class," "Product Size," "Standard Cost," and "Product First Sold Date" in the Transactions table, contained empty values. Additionally, in the Customer Demographic table, attributes "Job Title," "Job Category," and "Tenure" exhibited instances of missing records.

- **Recommendation:** Given the relatively low percentage of missing values in relation to the dataset size, we propose the removal of these records.

3. **Inconsistent Entries Across Datasets:**

- **Issue:** A disparity in the number of customer IDs was observed between the Transactions table and the Customer Demographic and Customer Address tables. This incongruity renders the data unsuitable for analysis in cases of missing records.

- **Recommendation:** We suggest performing the analysis exclusively on synchronized data across all three customer tables, based on customer ID.

4. **Multiple Datatypes for a Single Column:**

- **Issue:** The "Standard Cost" attribute in the Transaction table exhibits varying data types, including special characters, leading to dataset inconsistency.

- **Recommendation:** We recommend the removal of special characters and standardization of data types to ensure consistency.

5. **Duplicate Values for the Same Column:**

- **Issue:** The "State" column in the Customer Address Table contains duplicate values, such as "VIC" & "Victoria," as well as "NSW" & "New South Wales." Similar duplication is observed in the "Gender" column of the Customer Demographic Dataset.

- **Recommendation:** To maintain uniformity, we advise utilizing state abbreviations instead of full names. For the Gender Column, we recommend imputing "U" with the distribution of the dataset.

We kindly request your attention to the quality issues and proposed changes to uphold consistent data quality across all tables. Once these adjustments are implemented, we can proceed with further analysis to derive valuable insights into the company.

Thank you for your cooperation.

Warm regards,

Shaghayegh Haghbin