

Dear [Client point-of-contact],

I hope this message finds you well. Your assistance in providing the four datasets from Sprocket Central Pty Ltd is greatly appreciated. I've compiled a summary of statistics from these datasets, highlighted in the table below. Should there be any discrepancies between these figures and your expectations, please let me know.

Customer Demographic (Total records 4000)

FIELD NAME	ERRORS
<u>DOB</u>	01 record 1843 87 records Blanks
<u>last_name</u>	125 records Blanks
<u>Gender</u>	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
<u>job_title</u>	506 records Blanks
<u>job_industry</u>	656 records mention 'N/A'
<u>Default</u>	3317 records value 'special characters' includes null and Blanks
<u>Tenure</u>	87 records Blanks

Transactions (Total records 20000 -past 3months)

FIELD NAME	ERRORS
<u>Online_order</u>	94 records Blanks
<u>brand</u>	48 records Blanks
<u>product_line</u>	48 records Blanks
<u>product_class</u>	48 records Blanks
<u>product_size</u>	48 records Blanks
<u>standard_cost</u>	48 records Blanks
<u>product_first_sold_date</u>	48 records Blanks

Throughout the review process, several data quality issues surfaced, accompanied by methods employed to address these inconsistencies. Additionally, recommendations have been outlined to prevent future occurrences and elevate the precision of the data pivotal to driving business decisions.

Here's an overview of the identified issues and the respective actions taken:

1. Empty Values in Certain Records:

- For fields like the brand of purchase or job title with empty values, steps were taken based on the extent of missing data.

2. Missing Fields in the Transactions Dataset:

- Records with missing fields (less than 3%) were removed from the training dataset to preserve data integrity.

3. Inconsistent Representations:

- Values across datasets were standardized to ensure consistency, such as representing "Victoria" consistently as "VIC" and resolving multiple gender representations into distinct categories.

4. Inaccurate Values:

- Instances of inaccurate data were flagged for further investigation and potential corrective actions.

5. Additional Customer IDs in Specific Datasets:

- Alignment issues were identified across datasets, prompting the use of the Customer Master list for model training to maintain completeness.

Ensuring data quality remains a priority to yield accurate results. We're committed to collaborating closely with you to rectify these issues and implement recommended strategies for enhanced data reliability.

Our team will continue with the data refinement process, documenting assumptions and raising queries as needed. Post-cleaning, aligning assumptions with Sprocket Central's understanding alongside your data SME is crucial.

Your feedback on these matters is invaluable. Please feel free to reach out if you have any questions or concerns.

Best regards,

Joe Junior Data Analyst at KPMG