Subject: Data Quality Assessment and Recommendations

Dear Sprocket Central Pty Ltd Team,

I hope this message finds you well. I would like to extend our gratitude for providing us with the datasets from Sprocket Central Pty Ltd. Our team has conducted a thorough data quality assessment, and I wanted to share our findings and recommendations with you.

Customer Demographic Table:

1. **Invalid DOB:** We noticed that the entry for "Jephthah Bachmann" with customer ID 34 contains an invalid Date of Birth (DOB).

2. **Gender Categorization:** The "gender" column contains mistyped entries such as "M" for Male and variations for Female (e.g., "F" and "Femal"). Standardizing these entries will ensure consistency.

3. **Job Title Categorization:** The "job_title" column can be transformed for better categorization, e.g., using "Web Developer" instead of "Web Developer I," "Web Developer II," etc.

4. **Missing Values:** Both the "job_title" and "job_industry_category" columns contain null values.

5. **Default Column:** The "default" column contains garbage values, which should be addressed.

6. **Data Types:** The "tenure" column should be of type Number, and the "DOB" column should be of type Short Date.

7. **Blanks in DOB:** The "DOB" column also contains blank values that require attention.

Customer Address Table:

1. **State Abbreviations:** The "state" column contains state abbreviations that should be replaced with full state names, such as "VIC" to "Victoria" and "NSW" to "New South Wales."

2. **Redundant "Country" Column:** The "country" column contains a single entry, "Australia," for all rows. We recommend dropping this column.

Transaction Table:

1. **Date Format:** The "transaction_date" column should be of type Short Date for consistency.

2. **Missing Values:** Several columns, including "online_order," "order_status," "brand," "product_line," "product_class," "product_size," "list_price," "standard_cost," and "product_first_sold_date," contain null values.

3. **Currency Data Types:** The "standard_cost" and "list_price" columns should be of the currency type. Additionally, the precision of these columns should not exceed two decimal places.

We believe that addressing these data quality issues is crucial for accurate and meaningful analysis. To mitigate these concerns and ensure the data is ready for phase two of our analysis, we recommend the following steps:

1. **Data Cleansing:** Perform data cleansing to correct invalid, inconsistent, or missing values.

2. **Standardization:** Standardize data entries, such as gender and state abbreviations, for consistency.

3. **Categorization:** Simplify job titles for better categorization and analysis.

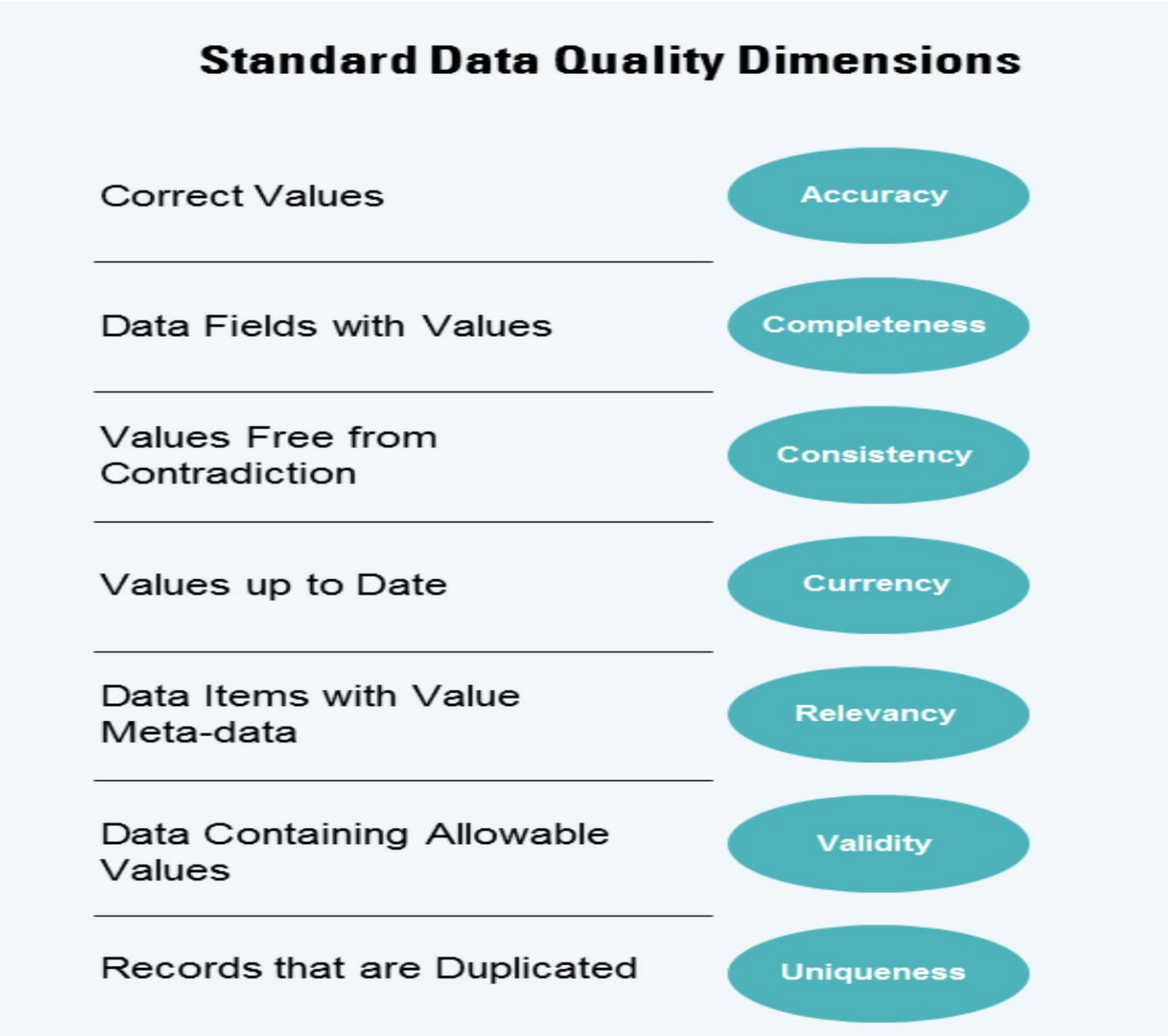
4. **Handling Missing Data:** Develop strategies for handling missing data in various columns.

5. **Data Type Adjustment:** Ensure that columns have the correct data types.

6. **Date Format:** Convert the "transaction_date" column to Short Date format.

7. **Remove Redundant Columns:** Consider dropping the "country" column from the Customer Address Table.

We have attached a Data Quality Framework Table that outlines the criteria and dimensions considered in our assessment. If you have any questions or require further clarification on any of these recommendations, please feel free to reach out.



Standard Data Quality Dimensions

Correct Values	Accuracy
Data Fields with Values	Completeness
Values Free from Contradiction	Consistency
Values up to Date	Currency
Data Items with Value Meta-data	Relevancy
Data Containing Allowable Values	Validity
Records that are Duplicated	Uniqueness

Thank you for entrusting us with this important task, and we look forward to working closely with you to optimize your data for enhanced business insights.

Best Regards,

Syed Yousuf Hasan

KPMG Data Analyst