TASK 1: DATA QUALITY ASSESEMENT

Dear [Client Name],

This is Jaynil from the KPMG Analytics Team. After reviewing the data for Sprocket Central Pty Ltd, I found several issues, which, if not addressed would relate to an unbalanced dataset. I have underlined the problems and ways to mitigate them in accordance with the data quality framework.

Data Quality Parameters	Issues	Mitigation
Accuracy	• Outliers in DOB column in multiple sheets. • Spelling error in 'job_category' column in *CustomerDemographic.*	• Create a new age column wherever DOB is mentioned for better analysis. • Find and replace the error with correct spelling.
Completeness	• Null values in many columns. • Default column in *CustomerDemographic* is filled with irrelevant values.	• Filter out and assign *NA* to null-values in columns like address, profession, etc… • Deleting the column for better analysis at further stages.
Consistency	• Gender attribute in multiple sheets is not consistent. • Some data fields have full state names instead of short forms.	• Edit values using find and replace. • For states, VLOOKUP can also be used.
Currency	• The 'Y' indicator in deceased_indicator is irrelevant.	• Discard the records with 'Y' as deceased_indicator.
Relevancy	• Redundant column of rank in NewCustomerList	• Keep only one rank column (Advisable to keep the one which formulates the rank).
Validity	• In *transactions,* product_first_sold_date is of wrong data type. • In *transactions,* list_price and standard_cost is not of currency type.	• Ensure all the values have relevant data types. This will be helpful while fetching data for model building.