Dear [Client point-of-contact],

Thank you for your decision to utilize KPMG's expertise for your organization. Our team has analyzed the quality of the raw data and identified several quality issues that need to be addressed. Additionally, we have provided recommendations to resolve these quality issues and improve the data's effectiveness.

	quality issue	Recomendation
1	**Missing values** • Transactions table has several attributes that contains missing value, such as online order, brand, product line, product class, product_size, standar_cost, product_first sold_date. • Customer Demographic table has missing value in several attributes: last_name, DOB, job_title, job_industry_category, default, and tenure.	Missing values can have an impact on the data, so it is recommended to remove rows containing missing values, As the percentage of missing values in the datasets is low as compared to the entire dataset.
2	**Inconsistent values within same column** In the "State" column of the Customer Address Table, several instances of inconsistent values were identified such as "VIC" and "Victoria", "NSW" and "New South Wales." Additionally, a similar problem exists in the "Gender" column of the Customer Demographic table.	Using the full names of the states instead of abbreviations for all records to ensure consistency across addresses. For the gender column, can use "female" and "male."
3	**Inappropriate datatype** The customer demographic column "dob" has an inappropriate datatype.	Converting the DOB into date or age column would be more relevant to the bussines data analytics
4	**Redundant Outliers.** Some data values stand out as outliers and can disrupt the entire dataset. For example, customer ID 34, belong to Jephthah Bachmann, whose recorded birth year is 1843, indicating an age of 179 years which is an error in the data in the Customer Demographic Table.	Remove the redundant data as it may skew the distribution of the dataset.

5	**Multiple DataTypes for a Single Column.** In the Customer Demographic table, there are certain records in the "default" attribute that include special character strings, leading to inconsistencies in the dataset.	Eliminate the special characters from the records and convert all the characters into numeric data to ensure consistent data type.
6	**Customer IDs are missing in certain tables.**	To keep one table as the master table containing relevant data for all other tables, customer ID values with missing data will be removed. This is because data for irrelevant customer IDs has been provided.

Please look into quality issues mentioned above along with the recommended changes to ensure the consistent quality of the dataset across all the tables. all the suggestions are matched we can proceed with deeper data analysis to find valuable insights for the company. Please let us know your thoughts on our first stage of cleaning.

Kind Regards,

Windi Pangesti