Dear client,

We like to thank you for choosing KPMG's expertise for your organisation. The dataset has been reviewed by our team and here is our initial review regarding the same.

The quality of the data will be summarised based on the Data Quality Framework we use as a guideline and it is as follows:

- Accuracy – correct values
- Completeness – data fields with values
- Consistency – values free from contradiction
- Currency – values up to date
- Relevancy – data item with value meta-data
- Uniqueness – records that are not duplicated
- Validity – data containing allowable values

Here is a brief summary table of our assessment:

	Customer Demographics	Customer Addresses	Transactions
Accuracy	DOB has deceased customers		
Completeness	Last_name, DOB, job_title, tenure	Customer_id's missing	Online_order, brand, product_line, product_class, product_size, standard_cost are all missing values
Consistency	Gender has 3 values		No data on all customer_id
Currency	Deceased customers in table		Cancelled orders not necessary
Relevancy	Default column	country column	product_first_sold_date not necessary
Uniqueness			
Validity	Age column needed		List_price and standard_cost not needed

We will now list out the data quality issues in detail along with our recommendations to improve the quality of data and the necessary action to improve the accuracy of your data to improve your business decisions.

It is as follows:

1. **The Customer Demographics has data from an extended timeline.**

Suggestion: Please ensure that all customers belong to the same timeline to get accurate results. The outliers can include customers above a certain age.

2. **The Transactions table has a lot of missing values in multiple columns listed in the summary table.**

Suggestion: Missing values highly affect your data so it is suggested to remove the rows containing the missing values as they are very few in number in comparison to the rows present in the table.

3. **The Customer Demographics column has DOB which has less relevancy.**

Suggestion: Converting the DOB into an age column would be more relevant to the business analytics.

4. **The Customer Addresses table has a country column with the same value across.**

Suggestion: The country can be added to the state column instead of adding a new column.

5. **The Transactions column has irrelevant columns pertaining to sold date. Profit column not included.**

Suggestion: We can remove the product sold date due to irrelevancy. The list price and standard cost can be removed and a profit column can be added to improve the validity.

6. **Customer id's are missing in certain tables.**

Suggestion: To keep one table as the master and have the data relevant to all those in the other tables and null values to be removed. As data for irrelevant customer_id's have been provided.

The above summary is concluded from the first stage of Data Quality checks. Our team will move forward with the next step, i.e. data cleaning, transformation, etc. This will be documented and sent at an agreed timeline.

Please let us know your thoughts on our first stage of cleaning. We will be happy to discuss your suggestions regarding this on our next meeting.

Thanks and Regards,

Abhijeet Patil