Dear Mr./Ms. [Client Name],

I would like to thank you for choosing KPMG Analytics, Information & Modelling team and providing us with the three datasets of your company. While working with the datasets, I noticed some data quality issues. Therefore, I am writing this email to identify those issues and my strategies to mitigate them.

Please take a look at the summary table below for a quick preview. More in-depth descriptions and recommendations for the future are also provided.

Summary Table

	Customer Demographic	**Customer Addresses**	**Transactions in the past 3 months**
Accuracy	- Inaccurate values in *DOB*. - Missing *age* data.		- Missing *profit* data.
Completeness	- Blanks in *job_title*. - Incomplete data in *customer_id*.	- Incomplete data in *customer_id*.	- Blanks in *online_order*. - Blanks in *brand*. - Incomplete data in *customer_id*.
Consistency	- Inconsistent values in *gender*.	- Inconsistent values in *state*.	
Currency	- Outdated values (deceased customers) in *deceased_indicator*.		
Relevancy	- Irrelevant *default* data.		- Irrelevant "Cancelled" data in *order_status*.
Validity			- Wrong data format in *list_price*. - Wrong data format in *product_first_sold_date*.
Uniqueness			

Detailed mitigations and recommendations for mentioned data quality issues:

1. Accuracy:
    a. Customer Demographic:
        Mitigation: Filter out the inaccurate values in *DOB*.
        Recommendation: Create an *age* column to calculate the age of customers. The *age* data is more comprehensible and easier to check for errors.

    b. Transactions:
        Recommendation: Create a *profit* column to calculate the profit earned. This data helps users in sales accuracy checking and future monetary analysis.

2. Completeness:
    a. **Customer Demographic:**

Mitigation:	Remove all the blank data fields.
Recommendation:	Simplify the *job_title* (such as *industry* or *department*) or provide drop down options for *job_title*.

    b. **Transactions:**

Mitigation:	Remove all the blank data fields in *online_order* and *brand.*
Recommendation:	Provide drop down options for *online_order* and *brand.*

    c. **Incomplete data in *customer_id:***

Mitigation:	Filter the data in *customer_id* from 1 to 3500.
Recommendation:	Ensure different spreadsheets are in sync and up-to-date to avoid errors in data analysis.

3. Consistency:
    a. **Customer Demographic:**

Mitigation:	Find & Replace all data values in *gender* into three values only: "Male", "Female" and "U".
Recommendation:	Identify the terminology first, then create drop down options for *gender* to avoid human errors due to manual entry.

    b. **Customer Addresses:**

Mitigation:	Find & Replace all data values in *state* into three values only: "NSW", "QLD" and "VIC".
Recommendation:	Identify the terminology first, then create drop down options for *state* to ensure the consistency of the data.

4. Currency:
Deceased customers data is not necessary for our data analysis since it does not impact the future estimated results.

Mitigation:	Remove all the data values with "Y" in *deceased_indicator.*
Recommendation:	Update the dataset frequently. Deceased customers are not current customers and would not affect the firm in the future, removing them will enhance the accuracy of future analysis.

5. Relevancy:
    a. **Customer Demographic:**

Mitigation:	Delete *default* column.
Recommendation:	Check the comprehensibility of all data. Ensure that the data is relevant and easy to understand.

    b. **Transactions:**

Mitigation:	Filter out the "Cancelled" data in *order_status.*
Recommendation:	Check the relevance of data in the analysis process. In this case, the "Cancelled" data can skew the dataset and affect future analysis.

6. Validity:
Wrong data format in *list_price* and *product_first_sold_date* would cause errors in calculation and interpretation.

Mitigation:	Change the format of *list_price* to currency and format of *product_first_sold_date* to short date.
Recommendation:	Ensure that all data are in valid format. Set the formats of all attributes to the commensurate ones before entering values.

All identified data quality issues with their corresponding mitigations and recommendations are summarized as above. Please feel free to contact us through email if you have any questions regarding these mentioned issues and their mitigations.

Sincerely,

Tran Van Loc