Dear Client,

I hope this email finds you well.

I have conducted a preliminary analysis of the data and have identified some areas where the quality of the data could be improved. Improving the quality of your data is essential to gain accurate insights about your customers and make informed decisions based on that information.

Using the Reference Data Quality Framework Table, I have assessed the data based on criteria and dimensions that are important for data quality. Based on this assessment, I have identified the following data quality issues:

|  |  |  |
| --- | --- | --- |
| Table | Inaccurate Data | Inconsistent Data |
| Customer Addresses |  | State, address |
| Customer Demographics | Job\_titles, DOB, deceased\_indicator | Gender, DOB, tenure |
| Transactions | product\_id, brand, product\_line, product\_class , product\_size, standard\_cost ,online\_order | product\_id, brand, product\_line, product\_class , product\_size, standard\_cost,product\_first\_sold\_date |

**Inaccurate data:**

There are instances where the data provided is incomplete. For instance, there are missing values for some of the fields in the customer demographic data and Transactions data. This can lead to inaccurate insights and affect the reliability of your marketing strategy.

Table

Description automatically generated

**Inconsistent data:**

The customer addresses dataset has some duplicate entries, which can cause confusion when analyzing the data. This can lead to over-representation of certain customers, leading to incorrect insights.

* State codes in the table are not consistent with standard coding; for example, NSW and New South Wales are same yet following different naming convention.

Table

Description automatically generated

* Duplicate address

Table

Description automatically generated

**Customer Demographics:**

* Gender data is inconsistent across the column.

Table

Description automatically generated

* Default column has no interpretable information so shall be dropped from table.
* 506 customers don’t have job titles mentioned in table.
* DOB column has one inaccurately recorded and has 87 customers have no records in DOB.
* DOB column has inaccuracy which makes data from deceased\_indicator column also unreliable.
* The same records in Gender columns also needs to be verified for accuracy.
* Similar inconsistencies are there in tenure column.

The transaction data has inconsistencies in the format of the data. This can affect the accuracy of data analysis and may lead to wrong inferences.

Table

Description automatically generated

**To mitigate these issues, I recommend the following strategies:**

**Data cleaning:**

This involves identifying and correcting errors, inconsistencies, and incompleteness in the data. This can be done by using data profiling, data validation, and data standardization techniques.

**Data integration:**

Combining the different datasets into a single dataset and creating unique identifiers for each customer to ensure that all data relating to a specific customer is linked.

**Data enrichment:**

This involves adding more data to improve the quality of the data. This can include external data such as social media data, web browsing behavior and other external data sources to gain insights into customer behavior and preferences.

**Data analysis:**

Applying statistical techniques such as regression analysis, clustering and segmentation to gain insights into customer behavior, preferences and purchasing habits.

**Data visualization:**

Presenting the insights gained from data analysis in the form of interactive dashboards, charts and graphs to make the insights more accessible and understandable.

By adopting these strategies, we can improve the quality of your customer datasets and provide accurate insights that will help you optimize your marketing strategy and drive business growth.

Please let me know if you have any questions or concerns. I look forward to discussing this with you further.

Best regards,

Vishita