KPMG Australia

Data Consulting Team

Sir/Madam XXX YYY

Dear Customer,

First, it is a pleasure to help you going forward with your activity at Sprocket Central Pty Ltd.

We well received the three datasets you provided to us for a thorough analysis.

As you may know, optimizing data quality and accuracy is a key component to drive company growth.

After the conduction of a data exploration, we identified multiple issues that may impact your marketing analysis & market understanding.

Based on the current situation of your datasets here are our observations, please let us know if something seems unclear for you.

|  |  |  |
| --- | --- | --- |
| **Date of reception** | **Concerned Sheet** | **Numbers of records** |
| 5th March, 2022 | Customer demographic | 4 000 |
| 5th March, 2022 | Customer address | 3 999 |
| 5th March, 2022 | Transaction data | 20 000 |

Noticeable data quality issues have been identified and here is the methodology we will conduct to mitigate the inconsistencies. You will also find recommendations to remove the causes later on.

**ACCURACY:**

* Dob column from CustomerDemographic and wrong values

*Mitigation: Filtering and extracting data. Calculate an “Age” column or a “age\_range” to make it more readable for human brain.*

* Typos in the “job\_industry\_category”

*Mitigation: Change the false values with the right one and clean it with an extracted list of errors and regular expressions use.*

**COMPLETENESS:**

* Null values

*Mitigation: in case of a small sample of missing value we can ignore them in the training set for the model. If it is a big part of the dataset and/or an important variable, we can make assumptions base one the mean values observed elsewhere in the dataset.*

* Additional Customer\_ids in transactions dataset compared to the others

*Mitigation: Make sure the datasets provided are from the same period. Missing data records could wrongly influence the results of the modelisation. In this situation we will only keep data that matches with CustomerAddress and CustomerDemographic.*

**CONSISTENCY:**

* Values as attributes should have a fixed syntax

*Mitigation: Change all the variations to one unified syntax. Data normalization helps with quality of analysis. Example: VIC, Victoria or F, Femal, Female*

* Numeric values should respect a decimal and currency syntax

*Mitigation: Convert characters to numeric, create a column with currency indication if the dataset contains multiple countries sells, fix the price to 2 decimals.*

**CURRENCY:**

* Deceased customers are irrelevant for a customer targeting analysis

*Mitigation: Remove all the records that contains “Y” in the deceased\_indicator of CustomerDemographic*

**RELEVANCY:**

* Cancelled orders could be irrelevant data for a customer to target recommendation

*Mitigation: Retain only values from sold products that are listed in Transactions.*

**VALIDITY:**

* Make sure that columns are correctly attributed to a data type

*Mitigation: Format columns to the right data type for the content.*

RECOMMENDATIONS:

Regarding the uniqueness, some of the data could be duplicated values in CustomerDemographic.

Based on the analysis of “First name”, “last name”, “gender” and “DOB”, it could be useful to check if there is not duplicated customers and in case of a doubt to make sure that there is not inactive account.

Furthermore, our proposal of improvement is to make the necessary fixtures to apply a Snowflake strategy of data to get it accurate, valid, consistent, avoid issues when there is a lack of uniqueness, etc.

Right now, you only get very partial data quality according to the market standards and it will be your main issue to generate meaningful reports. Unstandardized data always means more long-term costs to fix and make your databases working right away in your dashboards.

Following this in-depth analysis, we will pursue the data cleaning process. We will join all the datasets and make sure data quality is optimal, we will add new fields to the data to help the later analysis based on basic marketing questions.

We value your confidence in our advisory services and will be glad to help you bring Sprocket Central Pty Ltd as a brand with an active marketing strategy company.

It would be great to be able to exchange with your data team to ensure comprehension with generated data and your company understanding.

Best regards,

Valentin T.