Dear Sir/Ma'am,

We have reviewed the datasets provided by your company, Sprocket Central Pty Ltd.

The datasets provided to us were:

1. Customer Demographic
2. Customer Addresses
3. Transaction data in the past three months

There were multiple issues with the datasets, and we have listed all of them down below, along with ways to mitigate these issues.

- **Accuracy Issues:**

Issues: 'DOB' field in the CustomerDemographic dataset contains inaccurate values; the 'job_industry_category' field contains misspelled category values.

Mitigations: Filtered outliers in the dataset. Fixed misspelled categories.

- **Completeness Issues:**

Issues: 'DOB', 'job_title', 'job_industry_category', and 'tenure' fields in the CustomerDemographic dataset contain missing values; 'standard_cost', 'brand', 'product_line', 'product_class', 'product_size' and 'product_first_sold_date' fields in the Transactions dataset contain missing values for several transactions. Datasets were not in synchronization with join keys, 'customer_id' found to be inconsistent among datasets.

Mitigations: Filtered out records containing incomplete data.

- **Consistency Issues:**

Issues: 'gender' field in the CustomerDemographics dataset contains acronyms and misspellings of the major categorical values; the 'state' field in the CustomerAddresses contains multiple versions to identify the same categories; Additional Customer IDs in the Transactions and CustomerAddresses datasets indicate missing records in the CustomerDemographics dataset.

Mitigations: Replaced extended values and acronyms in the necessary cases using regular expressions. Customers with complete data shall be considered.

- **Currency Issues:**

Issues: The CustomerDemographics dataset contains records pertaining to deceased customers.

Mitigations: Filtered out demographic data for deceased customers by checking the 'deceased_indicator' field.

- **Relevancy Issues:**

Issues: Incomprehensible 'default' field in the CustomerDemographics dataset; Cancelled orders present in the Transactions dataset which may not be relevant and lead to incorrect predictions which could lead to business decisions based on inaccurate analysis. Moreover, all the field names in all three datasets viz. CustomerDemographic, CustomerAddress and Transactions, need to have meta-data regarding what they are supposed to contain.

Mitigations: Removed the ‘default’ field and the records pertaining to cancelled orders in the respective datasets.

• **Validity Issues:**

Issue: ‘product_first_sold_date’ incorrectly formatted as numeric data in Transactions dataset.

Mitigation: Field converted to Date type. Apart from this, there are some fields in the datasets about which we would like to have some clarifications:

1. ‘default’ field in CustomerDemographic dataset contains garbage values. We would like to know what it was originally supposed to contain.
2. The purpose of ‘tenure’ in CustomerDemographic dataset and ‘property_valuation’ in CustomerAddress dataset needs to be stated.

That is all from our side. Let us know if you have any queries regarding our review.

Thank You.

Best Regards,

Purva