To,
Sprocket Central Pty Ltd

Subject - **Data Quality issues and strategies to mitigate these issues.**

Respected Sir/Madam,
 Thank you for providing us with the dataset from Sprocket, we have received the dataset and summarised the following data quality issue with the dataset. We have further given our comment about how we plan to overcome these data quality issue and laid out a plan to move forward with the data cleaning.

Sheet Name	Unique column	No. of Rows (Before cleaning)	No. of Rows (After cleaning)
Transactions	transaction_id	20000	19946
NewCustomerList	NA (*since it is the target sheet*)	1000	736
CustomerDemographic	customer_id	4000	2863
CustomerAddress	customer_id	4000	4000

Notable data quality issues that were encountered and the methods used to mitigate the identified data inconsistencies are as follows. Furthermore, recommendations have been provided to avoid the reoccurrence of data quality issues and improve the accuracy of the underlying data used to drive business decisions.

1. **Completeness / Missing Values**

   ● CustomerDemographic → DOB, Job Title, job_industry_category contain the null values.
   ● Transactions → Online Order, brand, Product Line, Product Class, Product Size, Standard Cost,.and Product first sold date contain the null values.
   ● NewCustomerList → *DOB, job title, job industry* contain the null values.

2. **Relevancy / Unrelated Data**

   ● CustomerDemographic → Gender contain the U value which does not indicate Male or Female. Deceased indicator → Y is deletd because we only want alive Customers. Default → column is deleted as it is not decodable.

3. **Consistency/ Up to date the Values**

   - CustomerDemographic → Gender changed values of F and Femal to Female and M to Male. Adding the New Column Age → To know the exact age till date.
   - CustomerAddress → Replacing the New South Wales to NSW and Victoria to VIC.

4. **Accuracy/ Column not in a proper format.**
   - Transactions → Chaged the column format from General to Currency of List Cost and Standard Cost. Also product first sold date changed type from number to date.
   - NewCustomerList → past_3_years_bike_related_purchases changed the type from text to mumber. Postcode and Property Valuation change type from text to number.

Moving forward, the team will continue with the data cleaning, standardisation and transformation process for the purpose of model analysis. Questions will be raised along the way and assumptions documented.After we have completed this, it would be great to spend some time with your data SME to ensure that all assumptions are aligned with Sprocket Central's understanding.

Kind regards,
Nikhil Zodape