Hello Sir,

As per the given dataset of Sprocket Central Pty Ltd, on analysing the data using Python(Jupyter notebook) there were certain findings observed in respect of the data quality issues which majorly include inconsistency and a timeliness parameter cannot be analyzed.

Below are few finding listed one by one.

1.Transactions (Sheet 2)

* Everything seems to be alright with the type of data of the variables, minimum values, maximum values, and format.
* Therefore, we assume the accuracy of the datasets given is high.
* However,there are missing values in the overall dataset, in this sheet amounting to almost 1.8% of missing values.

2.NewCustomerList (Sheet 3)

* There is a problem in these columns -purchases on bike related purchases, DOB, postcode, and property valuation for having a wrong format.
* The job\_title variable shows a significant missing percentage with 10.6%

3.CustomerDemographic (Sheet 4)

* The column ‘Default’ has no value as it does not add any value to the dataset and it is a group of random characters.
* Here again, the ‘DOB’ column has a wrong format.
* The job\_title variable shows a significant missing percentage with 12.65%;

#### 4. Customer Address (Sheet 5)

#### There are no missing values in the dataset and therefore, it can be assumed that the accuracy is high.

## Mitigation measures:

* Some cleanliness in the data and standardization will increase the accuracy of the dataset.
* To take care of the missing data that if they are missing at random,it can be ignored for future analysis. If not, they can be predicted with the mean or median of the categories missing.
* It is recommended to add a Customer ID for this matter in the customer list.Because of this,a lot of data is inconsistent across the datasets.
* Besides the variable default in the CustomerDemographics dataset, all the variables seem to be relevant.
* A standardized format must be implemented on the DOB column especially.

Please review this and let me know if anymore clarity needed.

Kind Regards,

Nikita Gupta.