Dear [name],

After reviewing the three datasets we received from Sprocket Central Pty Ltd, some notable data quality issues are encountered. Solutions and recommendations on how to clean the data and mitigate the issues are provided as follows. Please let us know if you have any questions.

Missing values in certain columns

We saw some empty values in some columns, such as brand, job_title, job_industry_category.

If the column is the key column and the number of missing values are small, we can refill the empty values. However, if the missing values take up a big amount of the entire column, we would drop the column if it is not a key column, or ask for further data for refilling.

Inconsistent spellings for the same attributes

We saw some inconsistent spellings in some columns, such as 'F', 'M', 'Female', 'Male' for genders. 'V' and 'Victoria' for 'Victoria'.

We will use the consistent expressions and replace the misleading ones.

Inconsistent Data type

We saw some the data type of product_forst_sold_date is float which should be datetime.

We will change the data type to datetime for further data exploration.

Abnormal data entry

We saw some abnormal data entries. For example, there is a DOB entry which is '1843-12-21 00:00:00'. And the values under column default are not readable.

We will drop the row or columns which contain abnormal data.

Best regards,
[Team name]