

## To Sprocket Central Pty Ltd:

Below mentioned are the data quality issues found in the dataset provided by your organization.

The dataset which you have provided had the following sheets:

1. Transactions
2. CustomerDemographics
3. CustomerAddress

Join key is Customerid

Issues faced in Transactions Sheet:

1. Online order column has 360 blank values: indicating there is no confirmation on the channel of order.
2. There are 197 entries which have Brand, Product Line, Product Class, Product Size, Standard cost as blank values which might indicate non branded or sales of preowned cycles.
3. List Price column is not in currency value i.e., in dollars which indicates as value not currency
4. Product\_First\_sold\_date column is not in a dd/mm/yyyy format which will make it difficult for a ml model to make predictions on.
5. List price is not of currency type. (dollars)

Issues faced in CustomerDemographic Sheet:

1. Last name column has 125 empty fields.
2. The gender column has multiple values for the indicating the same gender:
  - a. Male=M
  - b. Female=F
3. DOB column has 87 empty fields and one date of 1843 which is clearly a mistake
4. Job\_Title column has 506 empty fields.
5. Job\_industry\_type column has 656(n/a) values.
6. Decreased indicator shows that there are only two customers who show a decreased value hence the column can be omitted.
7. Default column is filled unknown values.

Issues faced in CustomerAddress Sheet:

1. The State column has multiple values indicating the same state:
  - NWS=New South Wales
  - Vic=Victoria
2. The country column is standard for all hence it can be dropped.