**Transactions worksheet**

* This worksheet contains 555 rows of data that contain missing values. There were 197 rows of data that were missing the same data which were:
  + Brand
  + Product line
  + Product class
  + Product size
  + Standard cost
  + Product first sold date

I have attached a csv file of the list of datapoints that contained the above mentioned missing values

* Under the “online order” feature, I noticed that 360 data points had missing values, I have also included the csv file that contains only these data points. Note that only two of the data points overlap with the 197 rows from earlier.
* The “product sold date” feature does not seem to have the correct information as the data under the category is not in the correct date format.

**NewCustomerList**

* Between the property valuation and Rank columns, column names are missing for the data. Therefore, we cannot figure out what the data represents.
* The oldest timestamp for the DOB is 1938-06-08 seems reasonable as a 82 year old could purchase a bike as well.
* This worksheet has 285 data points with null values. The features that contain null values include:
  + Last name
  + DOB
  + Job title
  + Job industry category

They do seem to be important features to have since Sprocket Central wants to see the demographics of their customers to narrow the focus of their marketing strategies.

* I included a csv file of all the data points with missing values

**customerDemographic worksheet**

* This worksheet contains a total of 1370 data points with missing values
* There is an error with the dataset where one person has a date of birth of 1843-12-21 which would make the person way over 100 years old. When sorted the dataset by the DOB, the person that have the lowest DOB is 1931-10-23 which still seems reasonable. It can be said that the datapoint with the 1843 DOB is an incorrect data entry.
* The “default” feature is filled with unreadable text symbols. Under this feature, there are datetime objects and numbers so it is unclear what “default” truly represents.
* Features within this worksheet that contain null values are:
  + Last name
  + DOB
  + Job title
  + Job industry category
  + Default
  + Tenure

**customerAddree\_df worksheet**

* The state column contains the state where the customer is located. However, this column used different abbreviations to describe the same state. For instance, New South Wales was described as “New South Wales” and “NSW”. These two groups of customers are from the same state but they may be segmented into different groups when we perform data analysis on this dataset.
  + Another state that had more than one name is Victoria. “Victoria” and “VIC” were used to describe customers from the Victoria state.
* There are no missing values in this worksheet.