Hi there!

As directed, I have reviewed the provided datasets in accordance with the data quality framework and hereby list below the quality concerns along with suggestions to mitigate them.

Issues in dataset Customer Demographic:

* Accuracy: Gender column has both Female and Femal which is incorrect
* Completeness: Missing values: 125 last names, 87 DOBs, 497 job titles, 87 tenures
* Consistency: Gender column has M, Male, F, Female, etc. as separate categories, wealth segment has both affluent customer and high net worth as separate categories
* Currency: OK
* Relevancy: default column is unreadable and of no apparent value, tenure column ambiguous with apparent value, deceased indicator column can be removed by deleting records of dead customers and only keeping the rest, thus freeing up a column
* Validity: A date of ‘1843-12-21’ for customer id 34 is clearly not an allowable value, several mismatched job titles and industries (nurse practitioner and finance, nurse practitioner and manufacturing, etc.)
* Uniqueness: OK

Issues in dataset New Customer list:

* Accuracy: Gender column has both Female and Femal which is incorrect
* Completeness: Missing values: 29 last names, 17 DOBs, 106 job titles, 87 tenures, no customer id assigned, cannot link with addresses
* Consistency: wealth segment has both affluent customer and high net worth as separate categories
* Currency: OK
* Relevancy: property valuation and rank are not relevant to bike purchases; tenure, rank and value columns ambiguous with apparent value, Australia is the common country for all and can be omitted to free a column, deceased indicator column redundant as all new customers are alive
* Validity: Columns past\_3\_years\_bike\_related\_purchases, postcode and property valuation are numbers but formatted as text, several mismatched job titles and industries (speech pathologist and manufacturing, programmer and retail, nurse and property, etc.)
* Uniqueness: OK

Issues in dataset Customer Address:

* Accuracy: Has up to customer id 4003 but Customer Demographic has only till 4000, mismatch because customer ids 3, 10, 22 and 23 are missing and 3999 actual records present.
* Completeness: no customer id assigned for new customers, cannot link with addresses
* Consistency: state column has New South Wales, NSW, Victoria, VIC etc. as separate categories
* Currency: OK
* Relevancy: property valuation and rank are not relevant to bike purchases; Australia is the common country for all and can be omitted to free a column
* Validity:
* Uniqueness: OK

Issues in dataset Transactions:

* Accuracy: product\_first\_sold\_date column does not have date values
* Completeness: Missing values: 360 online\_order, 197 brand, 197 product line, 197 product class, 197 product size, 197 standard\_cost and 197 product\_first\_sold\_date
* Consistency: list price column has no dollar sign but standard\_cost has dollar sign, standard\_cost also has values with both 2 decimal places (e.g. $108.76) and 7 decimal places (e.g. 270.2999878) without the dollar sign.
* Currency: OK
* Relevancy: transaction date is already recorded, product\_first\_sold\_date irrelevant as a product can only be sold once, returns/exchanges etc. are separate transactions
* Validity: product\_first\_sold\_date column has values which are not in the date format, indeed they are not dates at all, probably mistitled column
* Uniqueness: OK

I would love your feedback on the same and if anything needs changing, please let me know.

Regards,

Debosmita