Dear Client,

Thank you for providing us with the Sales information for your business. Upon my first review of the dataset, I have found a list of issues that will affect our ability to accurately analyse the data and provide useful insights. Possible solutions for these issues are also included. The list is as follows:

Inconsistent formatting e.g. the gender column in the Customer Demographic worksheet has 3 representations of Female i.e. F, Femal, Female. This can be mitigated by using drop down boxes or option boxes in forms to prevent the possibility of incorrect spelling and unique formats

Other examples include:

Transactions worksheet

- standard cost column has non number values

- product\_first\_sold\_date column is formatted as a number

- list price column is formatted as a number with varying levels of precision

NewCustomerList worksheet

- past\_3\_years\_bike\_related\_purchases column is not a number

CustomerAddress worksheet

- State has inconsistent formatting eg New South Wales and NSW

Incomplete values e.g. in the Transactions worksheet the standard cost column has blank values. There are 197 transactions where nothing is bought, which causes brand, product\_line, product\_class and product\_size to not have any values. There are also many blanks in - job title column in the NewCustomerList worksheet as well as job title last name and date of birth in the CustomerDemographic worksheet. This can be mitigated by using compulsory fields or using default values where appropriate.

Scale is ambiguous for several columns including property\_valuation, rank and value columns in the NewCustomerList worksheet as well as wealth segment in the Customer Demographic worksheet.

This can be mitigated by including a meta data worksheet for all the columns with predetermined value ranged.

Incorrect values e.g. The default column in the Customer Demographic worksheet doesn’t show any useful information or has been encoded incorrectly and Jephthah Bachmann date of birth is wrong and shows his year of birth as 1843. It might be worth removing these values or checking how these values were obtained and changing how the records are recorded so that they can be used.

If you have any further questions regarding the issues, feel free to contact me.

Kind Regards,

Owusu