Hi there,

Upon Review of the 3 datasets, these are the issues which will affect the quality and accuracy of the model. The issues that will be faced while creating a model or conducting an analysis are mentioned below according to the dataset.

The birth date column in Customer demographic contains incorrect value such as 1843-12-21, which is around 170 years which will affect the accuracy of the model. The gender column has multiple values having the same meaning such as femal, female, f and such which basically adds up to inconsistency in the dataset. The default column consists gibberish values in them which basically makes that column useless for the analysis. From the customer demographic dataset columns named DOB, last name, job title, job industry, default and tenure consist missing values which affects the completeness of the data set eventually affecting the accuracy of the model.

In the CustomerAddress dataset the column n does not hold any values hence this column is not required, apart from that the entire dataset consist of 2 rows in the last that has missing values in every column and hence they can be ignored as well. Additionally, the column state has vic and Victoria pointing a single state, similarly nsw and new south wales points to a single place and hence creates data inconsistency.

In the Transactions dataset, we have similar issue with the previous dataset as it consists a row that does not hold any values. The online\_order, brand, product\_line, product\_class, standard cost, product first sold date and product\_size columns consists of some missing values.

These are the potential issues and they affect the quality of the dataset.

Best.