Hi <Client Name>,

Thank you for sending the datasets. I was reviewing the datasets and found some issues. PFB the issues associated with each dataset along with recommendations to mitigate the issues.

**Transactions**

1. Issue – Data is from Jan 2017 to Dec 2017. Therefore, not current.

Solution – Upload data till Dec 2020.

1. Issue - There are rows where transaction\_id is given but none of the product details (product\_line, brand, product\_class and product\_size) are given. Without product details, rows do not make sense. Therefore, data is not complete.

Solution – Give necessary product details corresponding to each transaction\_id.

1. Issue - product\_first\_sold\_date has Data Type float.

Solution - Should have datetime Data Type.

1. Issue - product\_id is not unique for a single/multiple product detail column(s). Therefore, it is insignificant.

Solution – Make product\_id unique for a single/multiple product detail column(s).

1. Issue - On the same transaction\_date, for the same product\_line, brand, product\_class, product\_size there are two different standard\_cost. This is not possible. There is cleary some mistake here.

Solution – Check standard cost for such combinations.

1. Issue – For the same combination of product details (product\_line, brand, product\_class and product\_size) there is high fluctuation in standard\_cost over a period of just 1 year.

Solution – Filter data based on every combination of product details. Check and rectify standard\_cost fluctuations.

**CustomerDemographic and CustomerAddress**

I performed inner join on both the datasets and then analysed the resultant dataset.

1. Issue - Jephthah was born in 1843 and is not deceased. This means he is 175 yrs. old. This is clearly an error.

Solution – Either deceased\_indicator is wrong or DOB. Check both.

1. Issue - Approximately 50% of the Affluent and High Net Worth Customers do not even own a car. This seems strange.

Solution – Check owns\_car corresponding to each wealth\_segment.

1. Issue - There is one entry of M for Male and one of F for Female in gender column.

Solution – Change M to Male and F to Female.

1. Issue - There are entries which have same address, state, and country but different postcode. This is incorrect.

Solution – Check postcode and state corresponding to same address.

1. Issue - Same state is mentioned in two formats. New South Wales is the same as NSW. Victoria is the same as VIC.

Solution – Change New South Wales to NSW and Victoria to VIC.

**NewCustomerList**

1. Issue - wealth\_segment vs property\_valuation stats seems to be incorrect.

|  |  |  |
| --- | --- | --- |
| property\_valuation  %age of ppl in  wealth\_segment | 1 | 12 |
| Affluent Customer | High (3.7) | Less (3.3) |
| Mass Customer | Less (2.7) | High (4.3) |

Solution – Recheck entries in property\_valuation and wealth\_segment columns if possible.

1. Issue - Approximately 50% of the Affluent and High Net Worth Customers do not even own a car.

Solution - Recheck entries in owns\_car and wealth\_segment columns if possible.

Thanks and Regards,

Apar Garg