Dear client,

First of all, thank you for the dataset, I would like to give you an overview of the data quality assignment. We went through your data and the following is the assessment.

|  | **Transaction sheet** | **Customer Demographic sheet** | **Customer Address sheet** |
| --- | --- | --- | --- |
| No of records | 2000 | 4000 | 3999 |
| No of unique records | 2000 | 4000 | 3999 |
| No of blank records | \* | \* | none |
| Receiving date | 24 June 2023 | 24 June 2023 | 24 June 2023 |

\* - contains blank records, look below for details

Problem with the dataset

1. Blank records - the problem of data loss
2. Inconsistent datatype and attributes
3. Inconsistent values and attributes
4. Possible Data corruption and garbage values to be a part of the dataset

#### Problems with - Transaction sheet

1. 360 values are blanks in online order True / False (E)
2. 197 blank cells under brand (G)
3. 197 blank cells under product\_line (H)
4. 197 blank cells under product class (I)
5. 197 blank cells under product\_size (J)
6. 197 bank cells under standard\_cost (L)
7. 197 bank cells under product\_first\_sold (M)
8. The data type is inconsistent in product\_first\_sold (M) and is not in date format

### Problems with - Customer Demographic sheet

1. 125 blank cells under last name (C)
2. inconsistent data under gender (D)
3. 1- out of range data - under - past\_3\_years\_bike\_related\_purchases
4. 87 - blank records - under - past\_3\_years\_bike\_related\_purchases
5. 506 - blank records under - job title
6. 87 - blank records - under-tenure
7. default column has garbage value

We are moving forward with the process of data cleaning, wrangling etc. It will be great to spend some time with your data.

Regards