Hello Sir,

As per our last meeting, I received the dataset regarding transactions and customer details of your company. While digging into the data, I found some issues that could interfere with the further analysis. I would like to draw your attention towards them before proceeding further with the analysis.

1. **Transactions Data**
2. **Inappropriate Data Types.**

|  |  |  |
| --- | --- | --- |
| **Column** | **Cuurent Data Type** | **Appropriate Data Type** |
| transaction\_date | text | date |
| product\_first\_sold\_date | float | date |
| standard\_cost | text | float |

To make use of such columns,they should be converted to their corresponding data types. product\_first\_sold\_date are actually pure numbers, so their understanding is a bit debatable.

1. **Missing Values**

1. For product\_id=0,there are 1181 transactions. Out of these records, 197 transactions do not have the these columns recorded: **brand, product\_line, product\_class,product\_size,list\_price,standard\_cost, product\_first\_sold\_date.**

These amount to 0.9% of the data. Because it is a very small amount, I belive such transactions should be removed from our final dataset.

2. For around 360 transactions, online\_order column is missing. Therefore it is difficult to know if it was an online or offline order. Until and unless, we would have to analyse your online transactions seperately for some agenda, we do not require to remove these records.

2**. New Customers Data**

**1. Missing Values**

1. job\_title data is missing for 106 new customers,even though their job industry is known for most of these records.

2. job\_industry\_catgeory is missing for 165 values

3. Date of birth(DOB) is not mentioned for gender ‘U’.(17 records)

**2. Ambiguous Data**

1. Some dates are quite ambiguous. For ex, 1843-12-21.

2. ‘tenure’ has different data type in this data from the corresponding column in demographics data.

3**. Customers Demographics Data**

**1. Missing Values**

1. DOB,default and tenure is missing for 87 out of 88 records for gender=’U’.

**2. Ambiguous Data**

1. Some dates are quite ambiguous. For ex, 1843-12-21.

2. default column has strange values.

3. Notations under gender column are inappropriate.

Female 2037 Male 1872 U 88 Femal 1 M 1

F 1

3**. Customers Addresses Data:** No problems until now.

**4. Joined Data**

**1. Joining Transactions and Customers Demographic**

For 507 customers , we have transactional data but no demographic data.

**2. joining Customers demographics and customer addresses**

For 4 customers, we have demographics data but no addresses data,while for 3 we have addresses data, but we don’t have demographics data.

**3. joining transactions and customer addresses**

Around 32 customers have transactional data but no addresses data.