

Hi,

I did data exploration and found some issues in Data that need to address to improve the quality of Data.

Transaction Data

Issues in Transaction Data are:

1. In Transaction Data, for some Transaction Id status, online order information whether it is True or False is not available whereas rest all information related to the same transaction Id is available.
2. Another issue in Transaction Data is, for approved order they don't have information regarding Brand name, Product line, Product class, Product size.
3. Also in Transaction Data, for some Transaction product Id mentioned is zero which is not possible, it should be some number starting from 1.

Customer Demographic Data

Issues in Customer Demographic Data:

1. The last name for so many customers is missing in data.
2. In field "gender", standard representation for gender where it is male or female is not followed in Datasheet.
Example, for customer ID 1, the "gender" field "F" is mentioned instead of writing "Female" as they wrote whole spelling for other customers.
3. They also kept Data of some customers who don't buy any item from their company.
For example, some customers do not make any purchases for bike from the last 3 years and their purchase count is zero still their record is maintained in Datasheet which is an increasing amount of Data.
4. The birth year of a few customers is before 1901, so they must be no more now. Hence we can omit data of such customers and clean our data. Also, the company doesn't have a record for the date of birth for a few customers.
5. The job title of some customers is also not mentioned in the company's data.

Current Address Data

Issues in Current Address Data:

In address field there are customers with house number/plot number- 0. This is error as house number/plot number cannot be zero.

All this error will make our data unclean. And it is suggested that company should collect all missing data so that at the time of analysis this missing value should not affect hypothesis.

Thanks.