Dear Client,

We would like to thank you for giving these datasets. After exploratory data analysis from the first stage data preprocessing, we got the following over view.

Transactions

The main problem of transactions table is missing values

(Columns Name – Num of Missing Values)

```
online_order             360
brand                    197
product_line             197
product_class            197
product_size             197
standard_cost            197
product_first_sold_date  197
```

NewCustomerList

There are 7 features that are missing values

(Columns Name – Num of Missing Values)

```
online_order             360
brand                    197
product_line             197
product_class            197
product_size             197
standard_cost            197
product_first_sold_date  197
```

CustomerDeographic

There are 6 features that are missing values

(Columns Name – Num of Missing Values)

```
last_name                    125
DOB                           87
job_title                    506
job_industry_category        656
default                      302
tenure                        87
```

In addition to missing values, there is a strange variable column called "default", the content of it is all about messy code.

CustomerAddress

There are 6 features that are missing values

(Columns Name – Num of Missing Values)

```
last_name                    125
DOB                           87
job_title                    506
job_industry_category        656
default                      302
tenure                        87
```

The above is the problem we found, if you have any questions, please contact us as soon as possible.

Sincerely,

ZheHan