Hello,

Based on available dataset, there are 3 (three) sheet of data available, namely:

- 1. Transactions
- 2. NewCustomerList
- 3. CustomerDemographic

There are a few things to note:

1. There are some columns that don't have the correct data type.

```
transaction id
                                  int64
product_id
                                  int64
customer id
                                 int64
transaction date
                        datetime64[ns]
online order
                              float64
order status
                                object
                                object
brand
product line
                                object
product class
                                object
product size
                                object
list price
                              float64
standard cost
                              float64
product_first_sold_date
                                object
dtype: object
```

I think that column "product first sold date" should have a datetime data type.

2. There are quite a lot of missing values in each column.

```
0
transaction id
                           0
product id
customer id
                           0
                           0
transaction date
online order
                         360
order status
                           0
brand
                         197
product line
                         197
product class
                         197
product size
                         197
list price
                          0
standard cost
                         197
product_first_sold date 197
dtype: int64
```

There are 360 missing values in online order columns, 197 missing values in product line, product class, product size, standard cost,

and

product_first_sold_date

```
first name
                                          0.0
last name
                                          2.9
gender
                                          0.0
past_3_years_bike_related_purchases
                                          0.0
                                          1.7
                                         10.6
job_title
job_industry_category
                                         16.5
wealth_segment
                                          0.0
deceased_indicator
                                          0.0
owns car
                                          0.0
tenure
                                          0.0
address
                                          0.0
postcode
                                          0.0
state
                                          0.0
country
                                          0.0
property_valuation
                                          0.0
Rank
                                          0.0
Value
                                          0.0
dtype: float64
```

In the NewCustomerList sheet, the total percentage of missing values in column job_title reaches 10.6% which shows quite a lot of missing values.

- 3. In the given dataset, there is no duplication value.
- 4. There are inconsistent values, especially in field of gender

```
array(['F', 'Male', 'Female', 'U', 'Femal', 'M'], dtype=object)
```

There are 'F', 'Male', 'Female', 'Femal', and 'M'

5. There is one column that I don't know the meaning of the values.

Sheet CustomeDemographic, column 'default'