1. Numerical data with high correlation to target
   1. ['transaction\_count', 'transaction\_count\_change', 'utilization\_ratio', 'transaction\_amount', 'revolving\_balance']
2. From high correlation visualization, the distribution of some data that have high correlation to target seem to be the same value as listed below but the pair between Transaction\_amount and transaction\_count we not sure how does this relation effect to model and we consider to test in training model session
   1. Utilization\_ratio and revolving\_balance
   2. Transaction\_amount and transaction\_count
3. Explore relation between numerical data
   1. Utilization\_ratio and revolving\_balance
   2. Transaction\_count and transaction\_amount
   3. Book\_period and age
   4. Open2buy and credit\_limit
4. Categorical data with high correlation to target
   1. ['contact\_num', 'month\_inactive', 'total\_product']

# New explore data

## 1 Dimension

### Hue = total\_product (group 0-2 and 2 - 10)

transaction\_amount

transaction\_count

### Hue = gender

open2buy show 2 normal distributions when the target is gender

credit\_limit (male seem to have more credit limit than female)

income (more than 60K is male, female have almost unknown income)

### Hue = Y (if we upsample this data)

transaction\_count

transaction\_count\_change

## 2 Dimensions

### Hue = total\_product (group 0-2 and 2 - 10)

Many features seem to be good after group

### Hue = Y

seem to can cluster many feature

transaction\_count & transaction\_count\_change

transaction\_count & transaction\_amount

age & transaction\_amount

age & transaction\_count

book\_period & transaction\_count

book\_period & transaction\_amount