Firstly, we unionid the 6 customer information datasets and show the situation of missing value with the other two datasets. Please refer to the plots in the slides.

The we merge all three data sets together together based on customer ID, and then handle those missing values. Since all the features in the data sets are assumed to be useful (relavent yo our target), we try to keep as many information as we can.

In campaign list: loan\_amount has NAs

In lending product info: msk\_mth\_since\_last\_limit\_review has NAs

In customer information nearly all numeric columns about the detailed information about clients have NAs. There are 11 columns with NAs.

Some columns just have fill NAs, we directly delete the rows of With NAs, these columns are: msk\_age, msk\_bank\_tenor, msk\_total\_saving\_bal. Process for others are as follows:

1. Loan\_amount is not NA only when clients borrow a loan, its reasonable. So, we can fill the NAs in this column with 0

2. msk\_mth\_since\_last\_limit\_review(Number of months since last credit card limit increase)

comparing to total number (720000), the NA values are not too great (42858). We will delect fill them with median

3. msk\_total\_credit\_trn\_amount : There are 73448 NAs. We fill them with average value of this column.

4. msk\_total\_credit\_trn\_atm\_amount: There are 279453 NAs. Since the min value of this column is not 0 but 10, we assume that the NAs represent for no transaction. So, we fill the NAs with 0.

5. msk\_total\_deposit\_trn\_count: Since the min value of this column is not 0 but 1, we assume that the NAs represent for no transaction. So, we fill the NAs with 0.

6. msk\_total\_credit\_trn\_count: There are 73448 NAs. Since it's too much, Since the min value of this column is not 0 but 1, we assume that the NAs represent for no transaction. So, we fill the NAs with 0.

7. msk\_total\_investment\_bal: There are 512912 NAs, which is more than 70%. Since it's too much, we just delete the column.

8. msk\_credit\_card\_revolving\_bal: There are 634861 NAs, which is more than 90%. Since it's too much, we just delete the column.

9. msk\_credit\_card\_bill\_bal: There are 25193 NAs. We fill them with average value of this column.

10. msk\_credit\_card\_limit: There are 25599 NAs. We fill them with average value of this column.