Retention Data and Customer Intelligence-Round 1

MINGLIANG WEI(11274244)

Starting Strategy

Our strategy is to find the people who has the most largetst possiblity to leave and invite them to prevent them from leaving.Firstly, I have checked the data and find out that there are so many missing values there. To avoid missing values which will cause a misleading to the regression, we will give the values of the NA obeying the characteristic of those variables who have missing values.(Average, Maximum,Medium based on the real business meaning environment) and changing those factors varibales into factor format.(e.g.)

```
streamraw$timeSinceLastTechProb[is.na(streamraw$timeSinceLastTechProb)]=100
streamraw$minutesVoice[is.na(streamraw$minutesVoice)]=200
```

Set the artificial variable 'Freq' which means the number of people who are using the plan in each single family to simulate the factor of conformity behavior, summary is shown as follows:

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.000   1.000   1.000   1.163   1.000   5.000
```

Seperate the data into the training set and the testing set and do the binary regression to get mod1, then predict the testing data based on the model we get.

| | p1 | nbAdultAvg | chrono | ... | gender | isWorkPhone | planType | data | dataAvgConsumption ▶ |
	<dbl>	<dbl>	<int>	<int>	<fctr>	<fctr>	<fctr>	<dbl>	<dbl>
624694	0.1241081	4	115	33	F	0	bring	9	4.455
666979	0.1240167	4	117	27	F	0	bring	7	2.185

2 rows | 1-10 of 30 columns

Adding predict_correction_p1 as our index showing the prediction correction rate of our model.

```
## [1] 0.0712
```

We can get the mod2 by using the stepwise of mod1 and by considering the correlations between variables, we can get mod3, after comparing the AUC between all the models, mod1 has the best performance so we choose it as our optimal model. But the prediction correction rate for our optimal model is too small as being a good binory model,after checking the previous dataset. We can get the leaving rate for all the clients.

```
leaving_rate
```

```
## [1] 0.02714581
```

Modifying Strategy

a.We find out that only 2.7% percentage of people will leave in three months, and the most largest possibiliy of mod1 is 12.4% combing the largest 7.12% prediction correction rate from all the models,which means that we don't have a big confidence to find out those who will leave in 3 months. b.Plus we are not sure if we invite them to come to our dinner event will help to change their mind from leaving, to remedy the weakness of our model, we modify our strategy from inviting those who has the largest possibilities to leave to the modified strategy that finding the expectation money we will lost for each person, it means that we will take the potential value of each customer into consideration. c.We will use the equations as follows to caculate the potential value of each customer:

$$PotentialValue = baseMonthlyRateForPlan + (baseMonthlyRateForPhone + cashDown + phonePrice + phoneBalance)^{1/2}$$
$$ExpectationLossingValue = PotentialValue * Probability$$

```
p1_score_cbind=p1_score_cbind%>%
  mutate(expectation_value_p1=(baseMonthlyRateForPlan+(baseMonthlyRateForPhone+cashDown+phonePrice+phoneBalanc
e)^0.5)*p1_score)
p1_score_cbind=p1_score_cbind[sort.list(p1_score_cbind$expectation_value_p1,decreasing=TRUE)[1:nrow(p1_score_cb
ind)],]%>%filter(Freq<=1)
```

Processing the score data and use the equation we mentioned above to predict the expectation money we will lost for each person and find the largest 8000 ones.
We will filter those people whose Freq is 1 because based on the rule of conformity behavior, the one who has the least degree pf conformity behavior will more likely to leave and then get our results.