Mobile Service Churn Project

Group 1
Bai Shuang, Luo Yiling, Mao Xinming, Wang, Zichen,
Yang Wenjie, Zhao Ruimeng

Customer Churn

Who We Are?

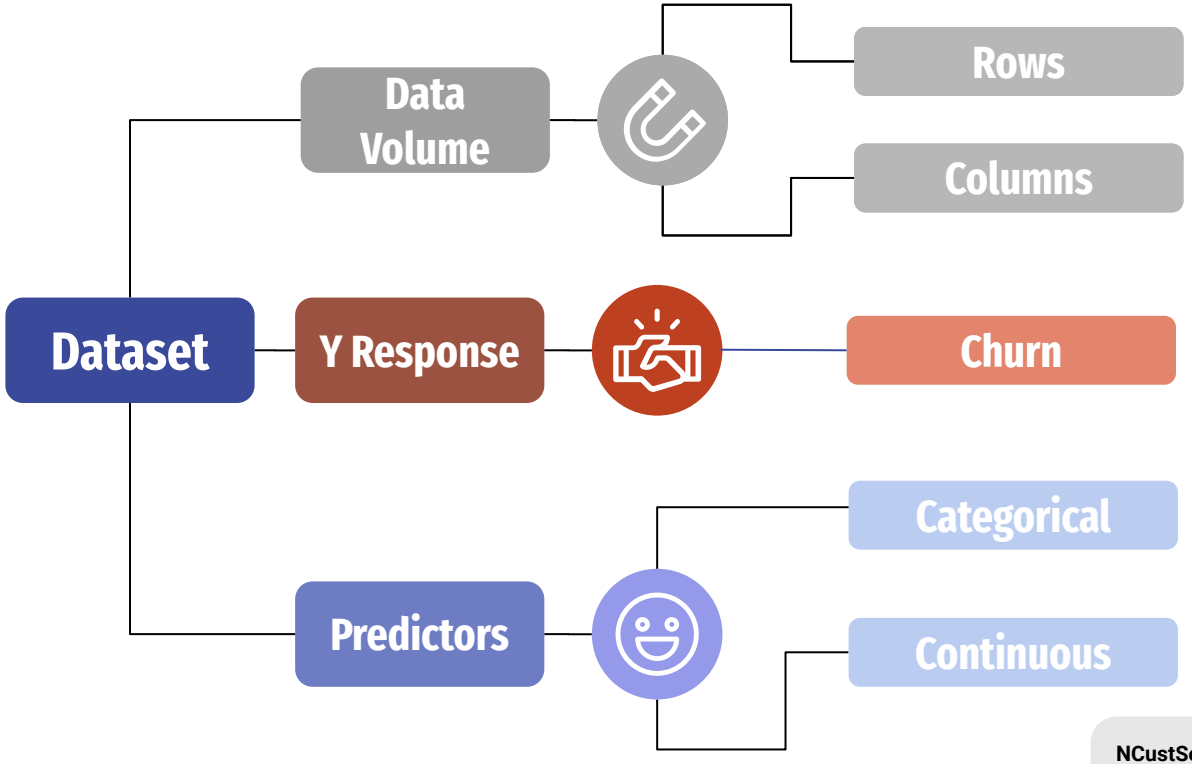
A mobile network company in an ultra-competitive industry

What We Are Trying To Do?

Identify **high churning customers** and study what predictors **impacting** on customer churn

Table of Contents

Task I: Dataset Exploration



- 3,332 rows in total
- 19 columns in total
- "Yes" or "No"
- Unbalanced data: 14.5% of customers churning (483 of Yes for "Churn")
- **"States", "IntlPlan" , "VMPlan"**

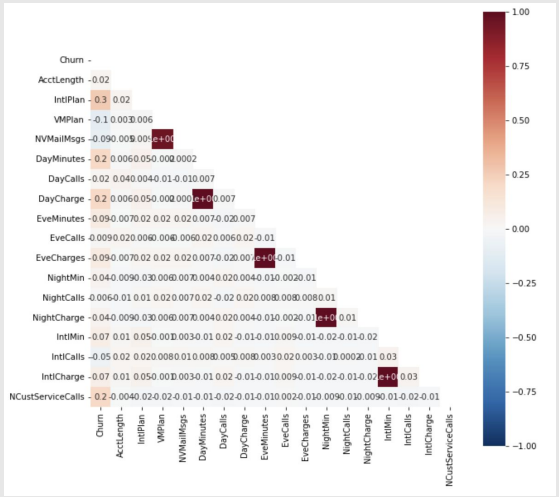
NCustServiceCalls
NVMailMsg
AcctLength
NCustServiceCalls

DayMinutes / DayCalls / DayCharge
EveMinutes / EveCalls/ EveCharges
NightMin / NightCalls / NightCharge
IntlMin / IntlCalls / IntlCharge

Task I: Data Descriptive Analysis

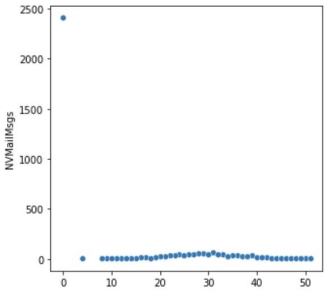
Correlation Matrix

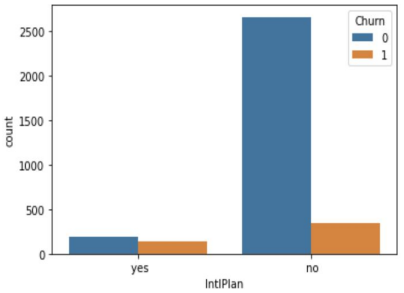
Generally not correlated, but certain variables corr = 1

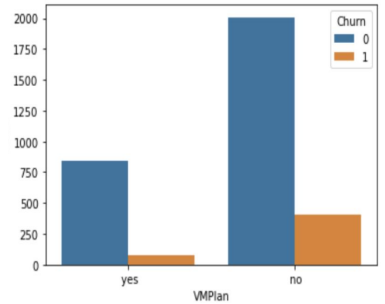


Explore on Specific predictors

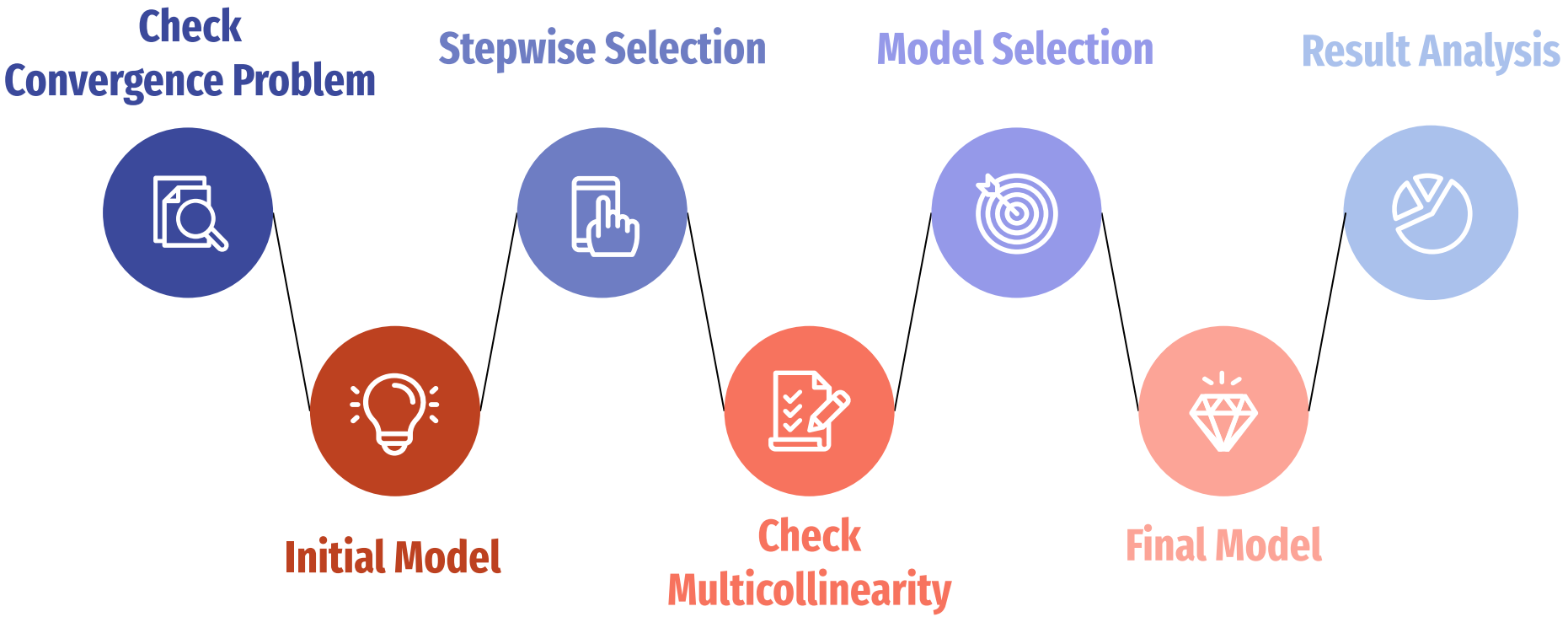
- the distribution of **NVMailMsgs** is skewed
- **With IntlPlan** group has higher churn rate and **without VoicePlan** group is also likely to churn







Task II: Model Building



Initial Model

Check Convergence Problem

```
library(detectseparation)
find_sep = glm(Churn~., family = binomial(link = "logit"), data = train_set,
               method = "detect_separation")
find_sep

## Implementation: ROI | Solver: lpsolve
## Separation: FALSE
```

- No occurrence of complete or quasi-complete separation
- Predictors of Initial Model: All features except DayMinutes, EveMinutes, NightMin, IntlMin

Reference: https://cran.r-project.org/web/packages/detectseparation/detectseparation.pdf

Variable Selection

All Variables — Initial Model

# of predictors: 14

# of variables: 64

AIC: 1770.2

Stepwise Selection — Reduced Model

# of predictors: 9

# of variables: 9

AIC: 1730.6

Variables	VIF
IntlPlan [yes]	1.062169
VMPlan [yes]	**15.338609**
NVMailMsgs	**15.300735**
DayCharge	1.050902
EveCharges	1.025418
NightCharge	1.017956
IntlCalls	1.012512
IntlCharge	1.017571
NCustServiceCalls	1.082091

Model Selection

Stepwise Model

Delete 'NVMailMsgs'

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
2,656	1,712.607			
2,601	1,642.161	55	70.44593	0.0783812

Model 3-1:
Churn ~ IntlPlan + VMPlan + DayCharge + EveCharges + NightCharge + IntlCalls + IntlCharge + NCustServiceCalls

Delete 'VMPlan'

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
2,656	1,719.467			
2,601	1,642.161	55	77.30579	0.02533413

Model 3-2:
Churn ~ IntlPlan + NVMailMsgs + DayCharge + EveCharges + NightCharge + IntlCalls + IntlCharge + NCustServiceCalls

Final Model

Model Result

Model: Churn ~ IntlPlan + VMPlan + DayCharge + EveCharges
+ NightCharge + IntlCalls + IntlCharge + NCustServiceCalls

Generalized Linear Model Regression Results

Dep. Variable:	Churn	No. Observations:	2665
Model:	GLM	Df Residuals:	2656
Model Family:	Binomial	Df Model:	8
Link Function:	logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-856.30
Date:	Fri, 09 Dec 2022	Deviance:	1712.6
Time:	11:33:58	Pearson chi2:	2.69e+03
No. Iterations:	6		
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-7.9496	0.582	-13.652	0.000	-9.091	-6.808
IntlPlan[T. yes]	2.0472	0.162	12.656	0.000	1.730	2.364
VMPlan[T. yes]	-0.9530	0.162	-5.865	0.000	-1.271	-0.634
DayCharge	0.0742	0.007	10.306	0.000	0.060	0.088
EveCharges	0.0808	0.015	5.424	0.000	0.052	0.110
NightCharge	0.0745	0.028	2.686	0.007	0.020	0.129
IntlCalls	-0.0930	0.028	-3.302	0.001	-0.148	-0.038
IntlCharge	0.3602	0.086	4.179	0.000	0.191	0.529
NCustServiceCalls	0.5002	0.045	11.229	0.000	0.413	0.587

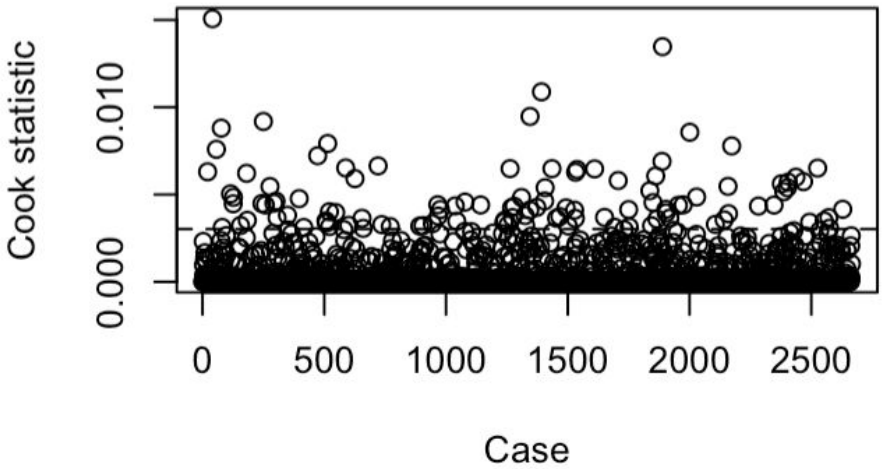
VIF

Variables	VIF
IntlPlan yes	1.062523
VMPlan yes	1.027730
DayCharge	1.049996
EveCharges	1.025334
NightCharge	1.017577
IntlCalls	1.011284
IntlCharge	1.017260
NCustServiceCalls	1.081643

No multicollinearity exists

Model Diagnosis

High Leverage Point



Outlier

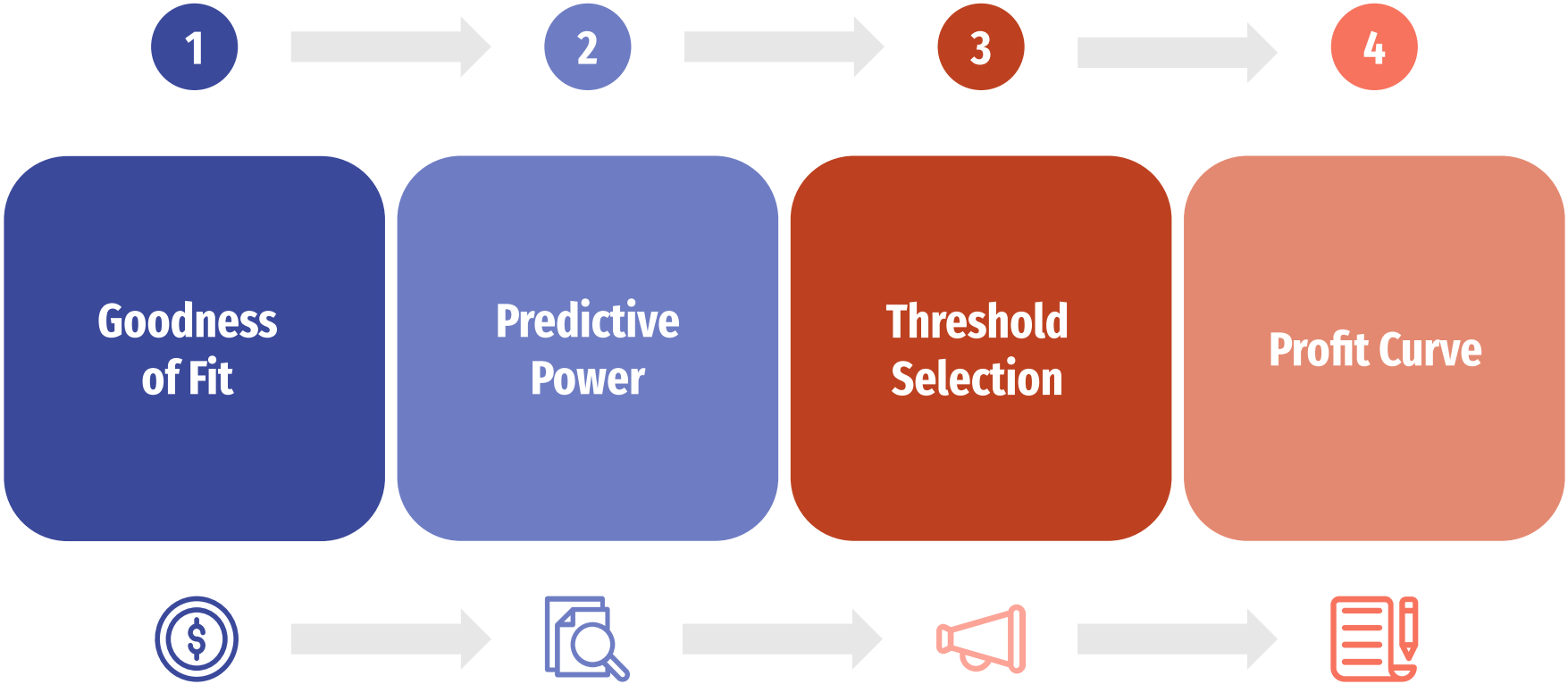
```
outlierTest(model3)
```

```
## No Studentized residuals with Bonferroni p < 0.05
```

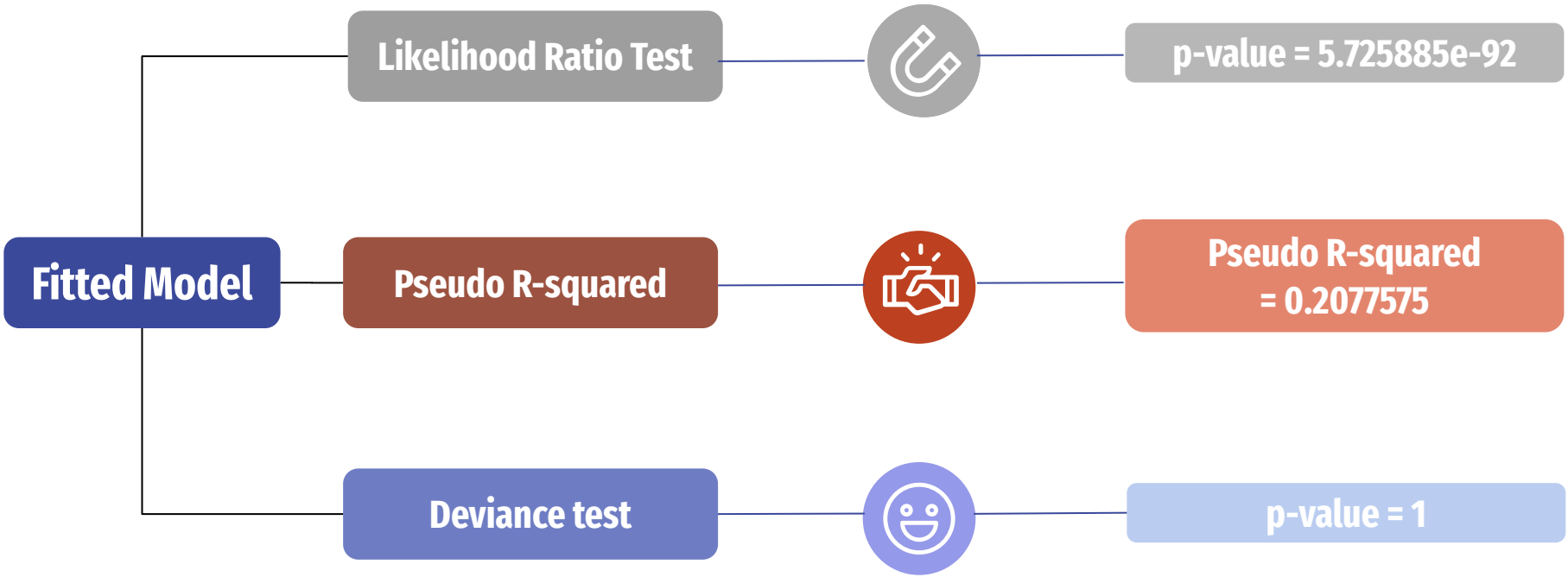
Some points have relatively extreme x values but ***can still be predicted well by the model***.

Reference: https://stat.ethz.ch/R-manual/R-devel/library/boot/html/glm.diag.plots.html

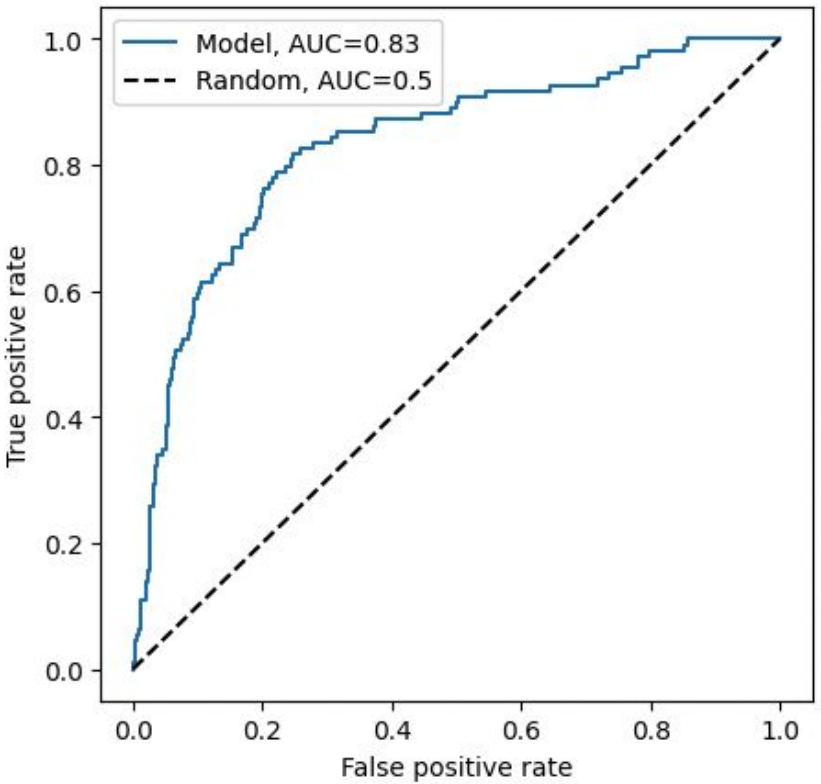
Task III: Evaluation



Goodness of Fit

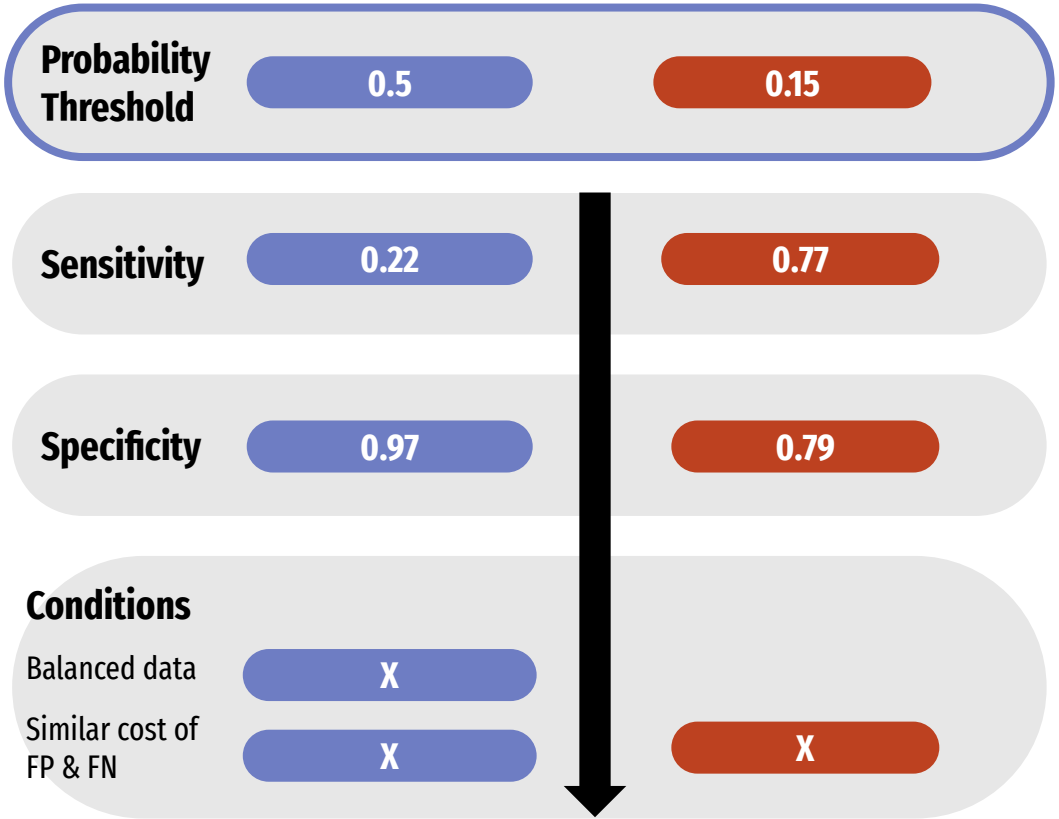


Predict Power: AUC



AUC = 0.83

Classification Threshold

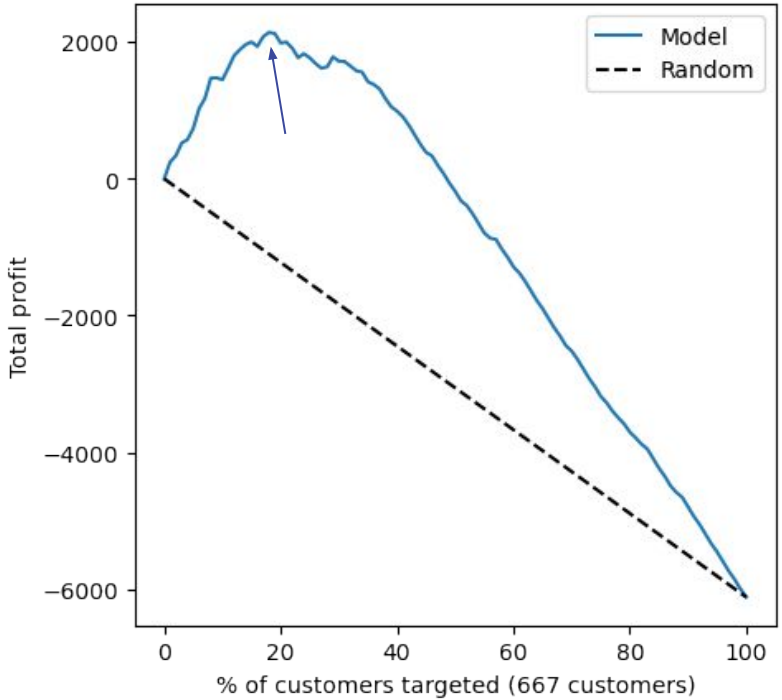


	cutoff	sensitivity	specificity	misClassError
0	0.00	1.000000	0.000000	0.836582
1	0.05	0.917431	0.419355	0.499250
2	0.10	0.853211	0.648746	0.317841
3	0.15	0.770642	0.788530	0.214393
4	0.20	0.660550	0.845878	0.184408
5	0.25	0.614679	0.892473	0.152924
6	0.30	0.522936	0.921147	0.143928
7	0.35	0.449541	0.942652	0.137931
8	0.40	0.348624	0.953405	0.145427
9	0.45	0.302752	0.965950	0.142429
10	0.50	0.220183	0.973118	0.149925

Profit Curve

- Assumption
  1. Incentive → Stay; No incentive → Churn
  2. Revenue:
     day + evening + night + international charges
  3. Cost: 20 dollars
- Top 26.44% with incentive (churn)
- Probability threshold = 0.18

	Churn	Not Churn
Y	revenue - cost	- cost
N	0	0



Task IV: Summary and Implications

Final Model

- Interpret the coefficients

  Percentage change of the odds of churn

- Calculate the standardized coefficients

  Measure the relative importance of the explanatory variables in a regression model

	coef	std	adj_coef	exp(coef)	exp(adj_coef)
IntlPlan	2.041353	0.295919	0.604075	7.701020	1.829558
VMPlan	−0.936289	0.447289	−0.418792	0.392080	0.657841
DayCharge	0.076539	9.257411	0.708555	1.079545	2.031054
EveCharges	0.084278	4.311311	0.363348	1.087931	1.438137
NightCharge	0.081568	2.275958	0.185646	1.084987	1.203995
IntlCalls	−0.091510	2.461450	−0.225248	0.912552	0.798318
IntlCharge	0.323891	0.753885	0.244177	1.382497	1.276570
NCustServiceCalls	0.512417	1.315652	0.674163	1.669321	1.962389

Business Implications

Indications

Suggestions

Charge

- Customers with higher day/evening/night/International charges are more likely to leave.
- Poor communication signals or expensive costs.

- Use a segmented charging strategy
- Provide discounts for excessive usage portion

Customer Service

- More customer service calls, more likely to leave
- Inefficient in solving customers' problem

- Retrieve history data to figure out the pain points
- Optimize the service process and efficiency

Voicemail Plan

- Customers with Voicemail Plan are less likely to leave
- Relatively well-designed and attractive product or good service in this plan

- Can be a selling point for the company
- Other products can refer to its strategy

International Plan

- A small percentage of users but high churn rate
- Maybe caused by poor international service

- Design different kinds of preferential packages to satisfy different types of customers' demands

Thanks !