



ALPINE  
INSIGHTS

# CUSTOMER CHURN PREDICTIONS IN TELECOMMUNICATION

12 MAY 2022

PRESENTATION BY



ALPINE  
INSIGHTS



# ABOUT US

---





# PRESENTATION TEAM



**MICHAEL BIGLER**  
/ Business Analyst /



**LIAM PHAN**  
/ Business Analyst /



**TANIA LOUREIRO**  
/ Statistician /



## REMAINING TEAM



**WILLIAM ELKIESS**  
/ Statistician /



**DAKOTA CUELLAR**  
/ Statistician /



**ILYANA EL MENDILI**  
/ Statistician /



# OUR AGENDA





# OUR AGENDA

1

PROBLEM  
PRESENTATION

2

EXPLORATORY  
DATA ANALYSIS

3

PREDICTIVE  
MODELS

4

MAIN DRIVERS OF  
CUSTOMER CHURN

5

CONCLUSIONS AND  
RECOMMENDATIONS



# PROBLEM PRESENTATION





# CUSTOMER CHURN





# DATA 5'000 customers



## 18 variables:

- duration of the customers' relationship with the company
- total number of day-time national calls
- total number of international calls
- ...

## The dependant variable:

- did the customer churn ?  
Yes / No



## HOW CAN CUSTOMER CHURN BE DETECTED AND WHAT ARE THE MOST IMPORTANT “DRIVERS”?

---

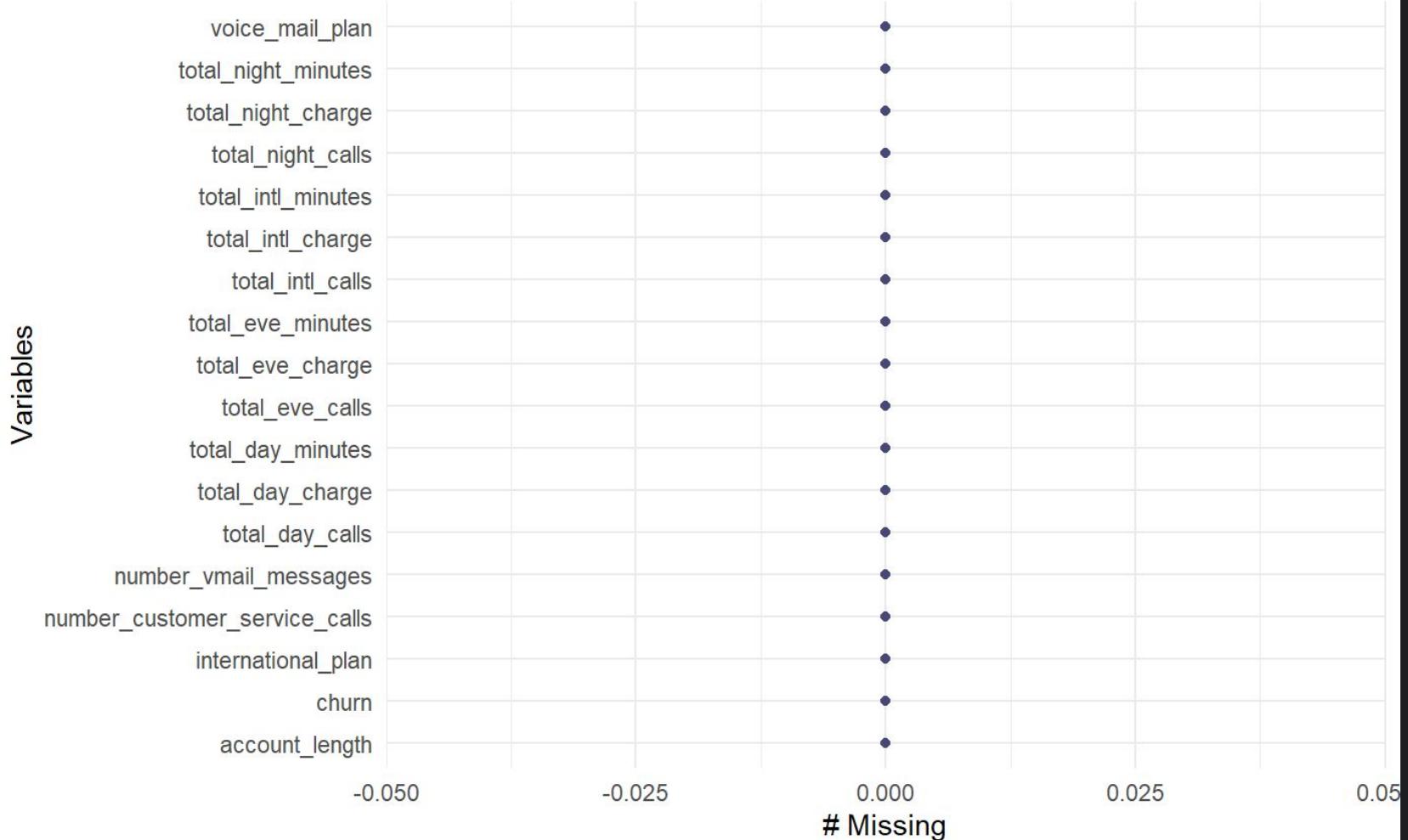


# EXPLORATORY DATA ANALYSIS AND FEATURE ENGINEERING





# MISSING VALUES



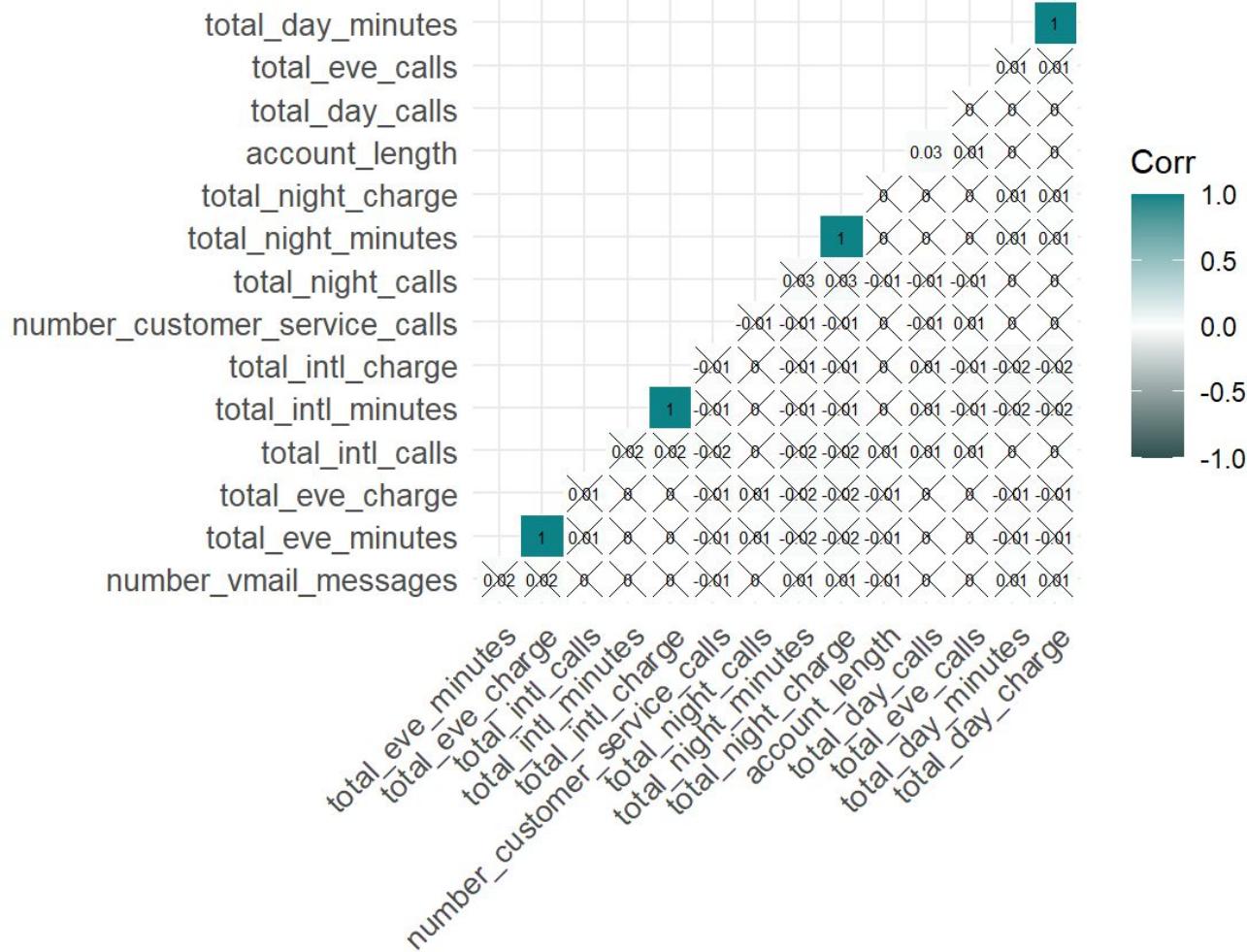


# OVERVIEW

	mean	sd	min	max
account_length	100.258600	39.6945595	1	243.00
international_plan	0.094600	0.2926909	0	1.00
voice_mail_plan	0.264600	0.4411641	0	1.00
number_vmail_messages	7.755200	13.5463934	0	52.00
total_day_minutes	180.288900	53.8946992	0	351.50
total_day_calls	100.029400	19.8311974	0	165.00
total_day_charge	30.649668	9.1620687	0	59.76
total_eve_minutes	200.636560	50.5513090	0	363.70
total_eve_calls	100.191000	19.8264958	0	170.00
total_eve_charge	17.054322	4.2968433	0	30.91
total_night_minutes	200.391620	50.5277893	0	395.00
total_night_calls	99.919200	19.9586859	0	175.00
total_night_charge	9.017732	2.2737627	0	17.77
total_intl_minutes	10.261780	2.7613957	0	20.00
total_intl_calls	4.435200	2.4567882	0	20.00
total_intl_charge	2.771196	0.7455137	0	5.40
number_customer_service_calls	1.570400	1.3063633	0	9.00
churn	0.141400	0.3484685	0	1.00



# CORRELATION



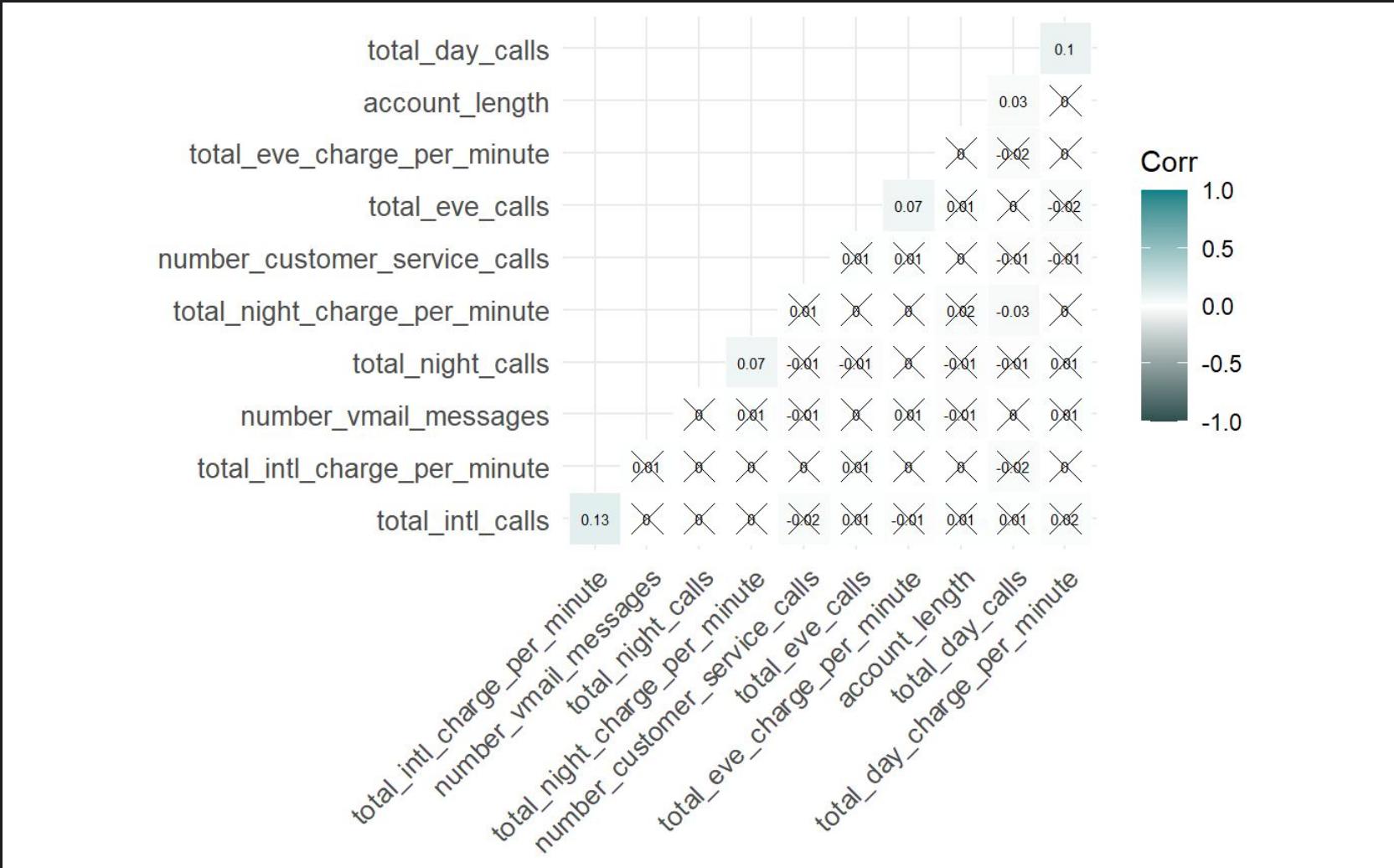


# SOLVING MULTICOLLINEARITY

**CHARGED MINUTES PER TOTAL MINUTES =  
RATIO OF CHARGED MINUTES OVER TOTAL  
MINUTES**

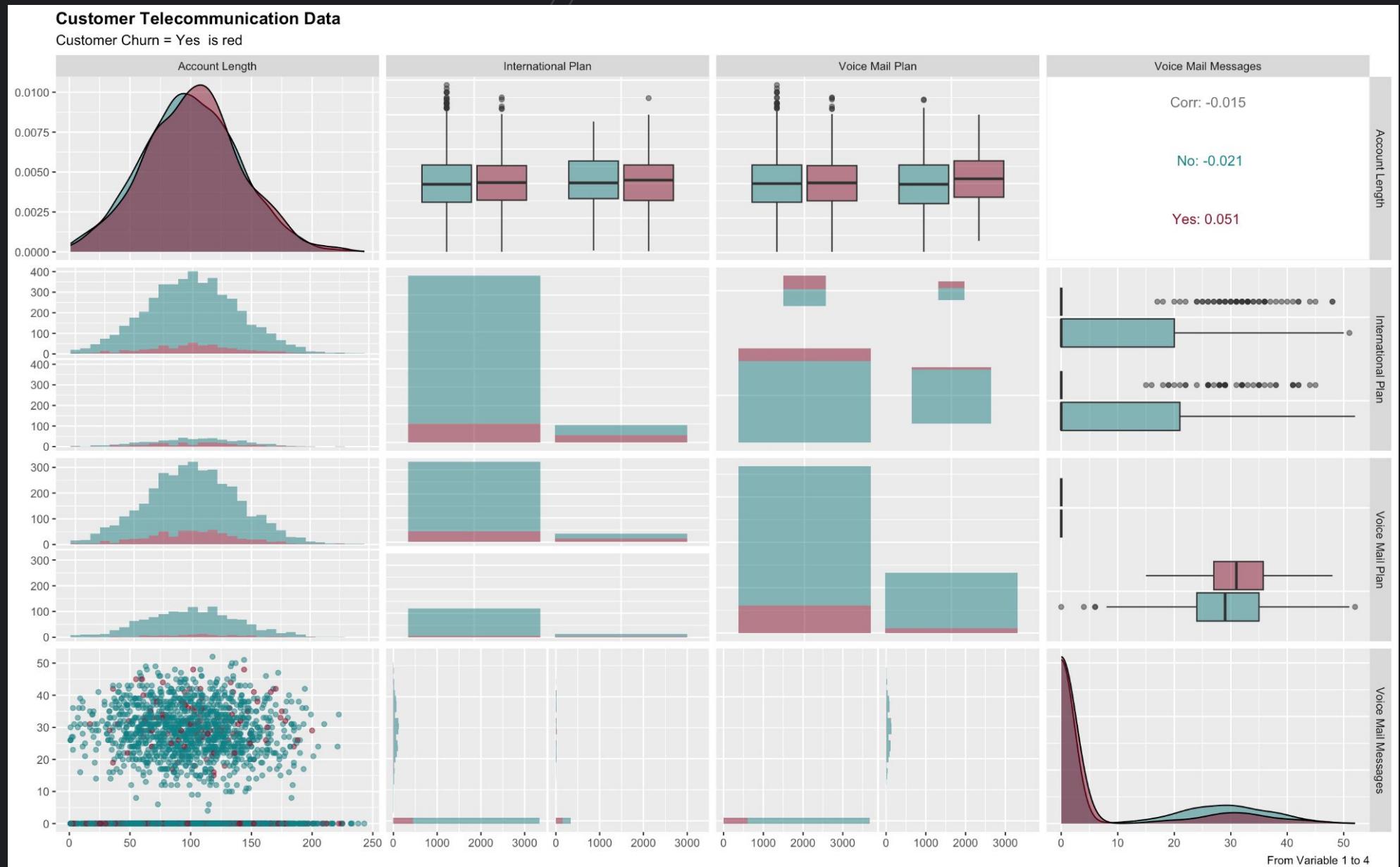


# CORRELATION AFTER COMBINING MINUTES AND CHARGE



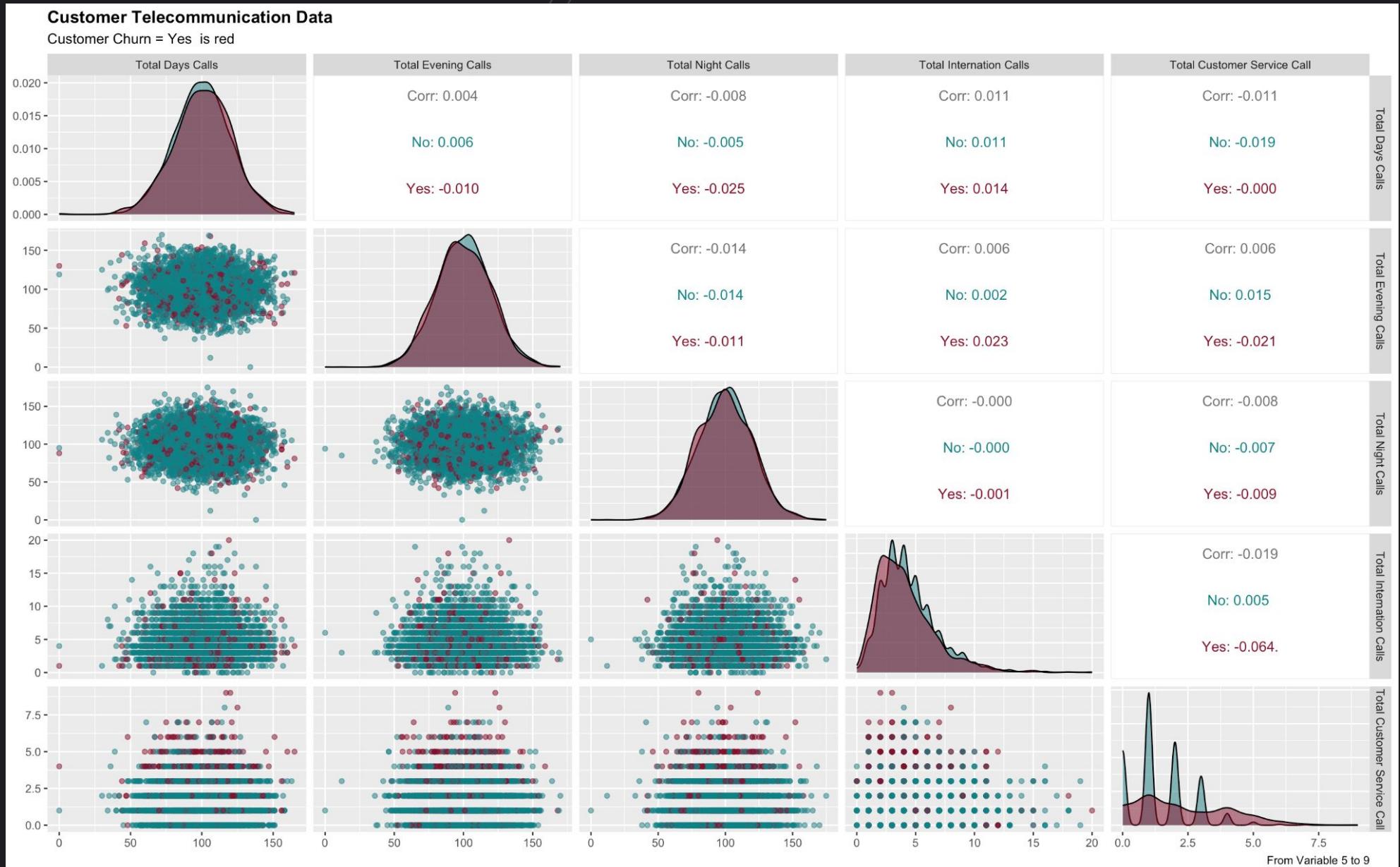


# PAIRPLOT





# PAIRPLOT



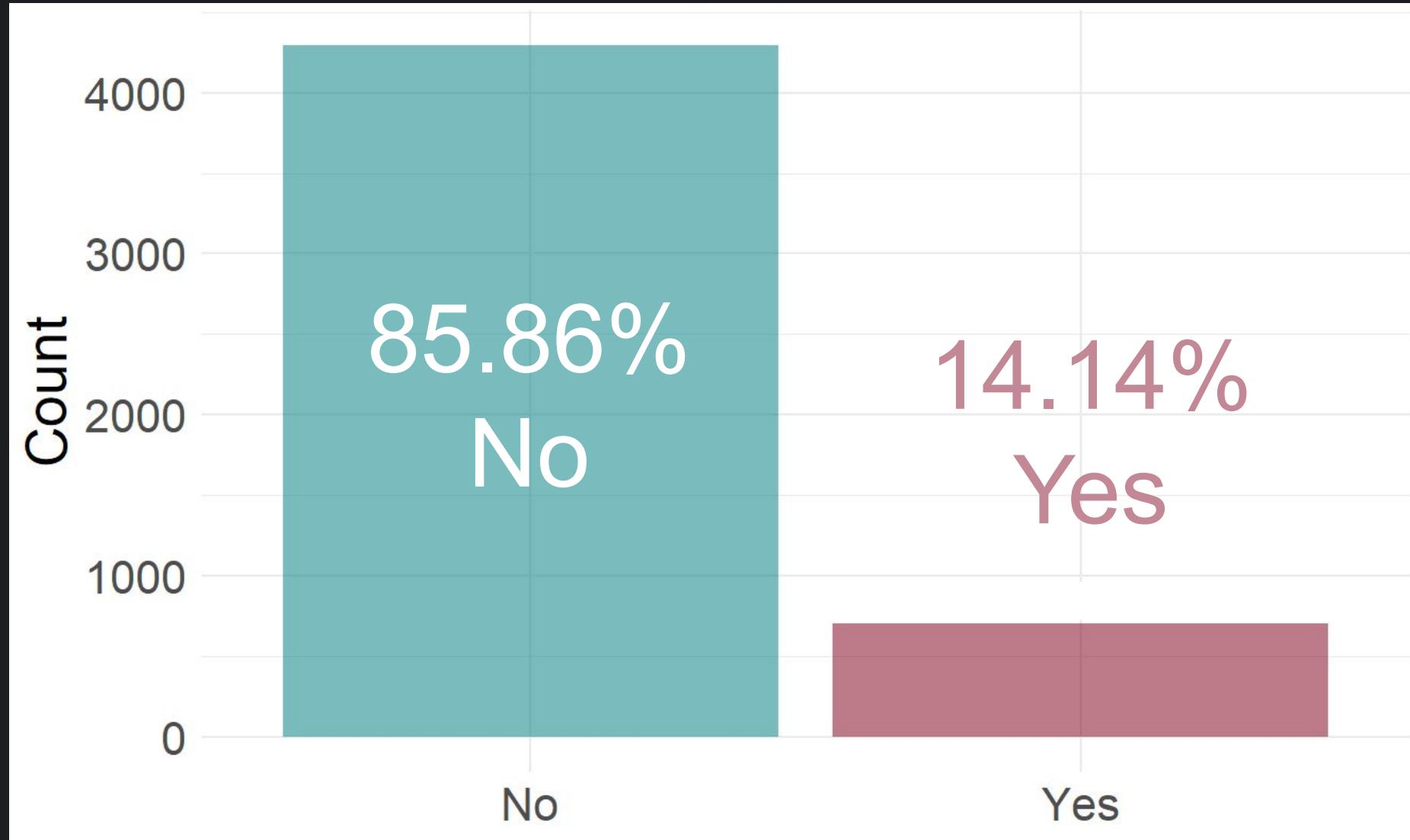


## ADDED FEATURES

$X \rightarrow X^2$

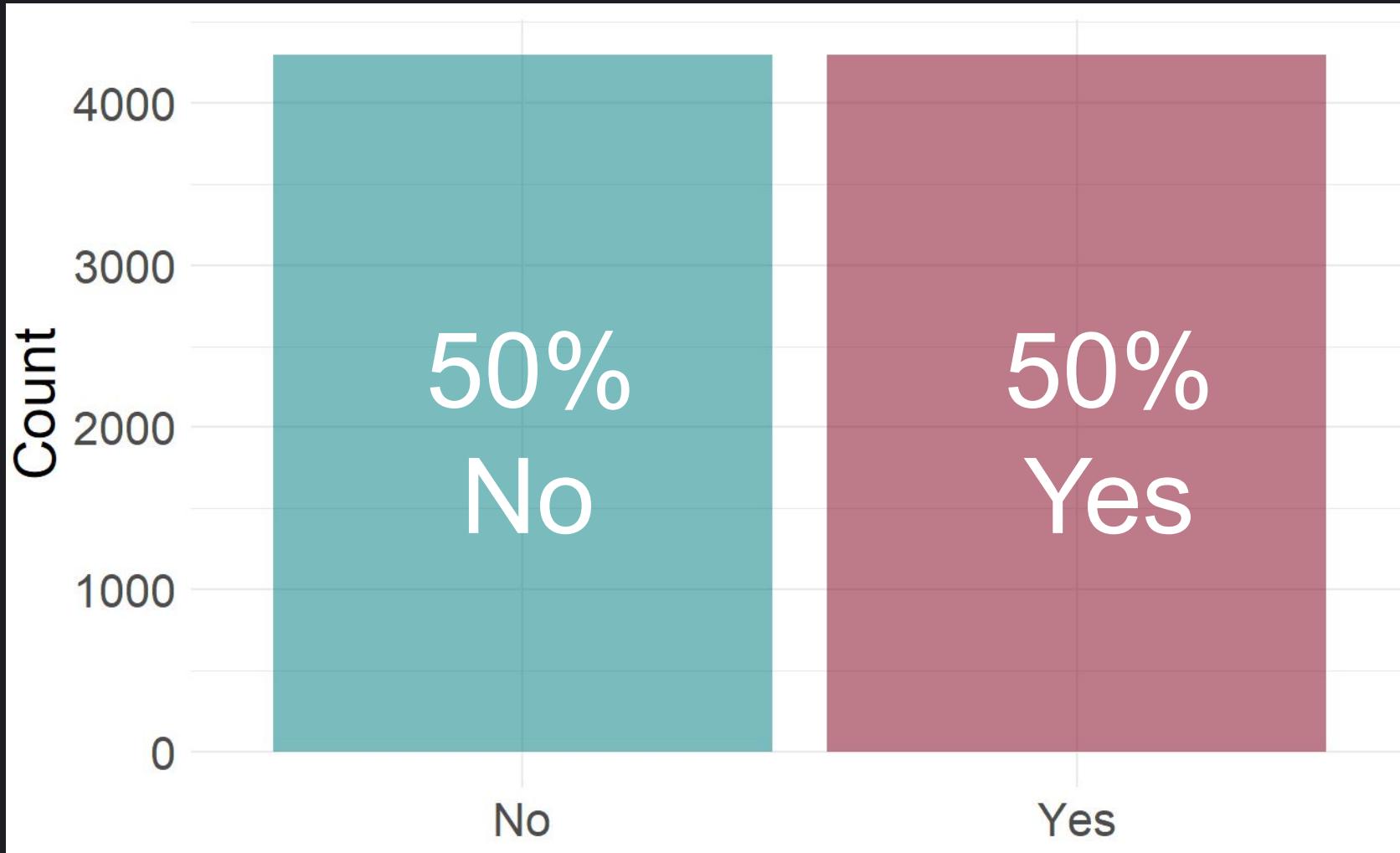


# BARPLOT OF CUSTOMER CHURN



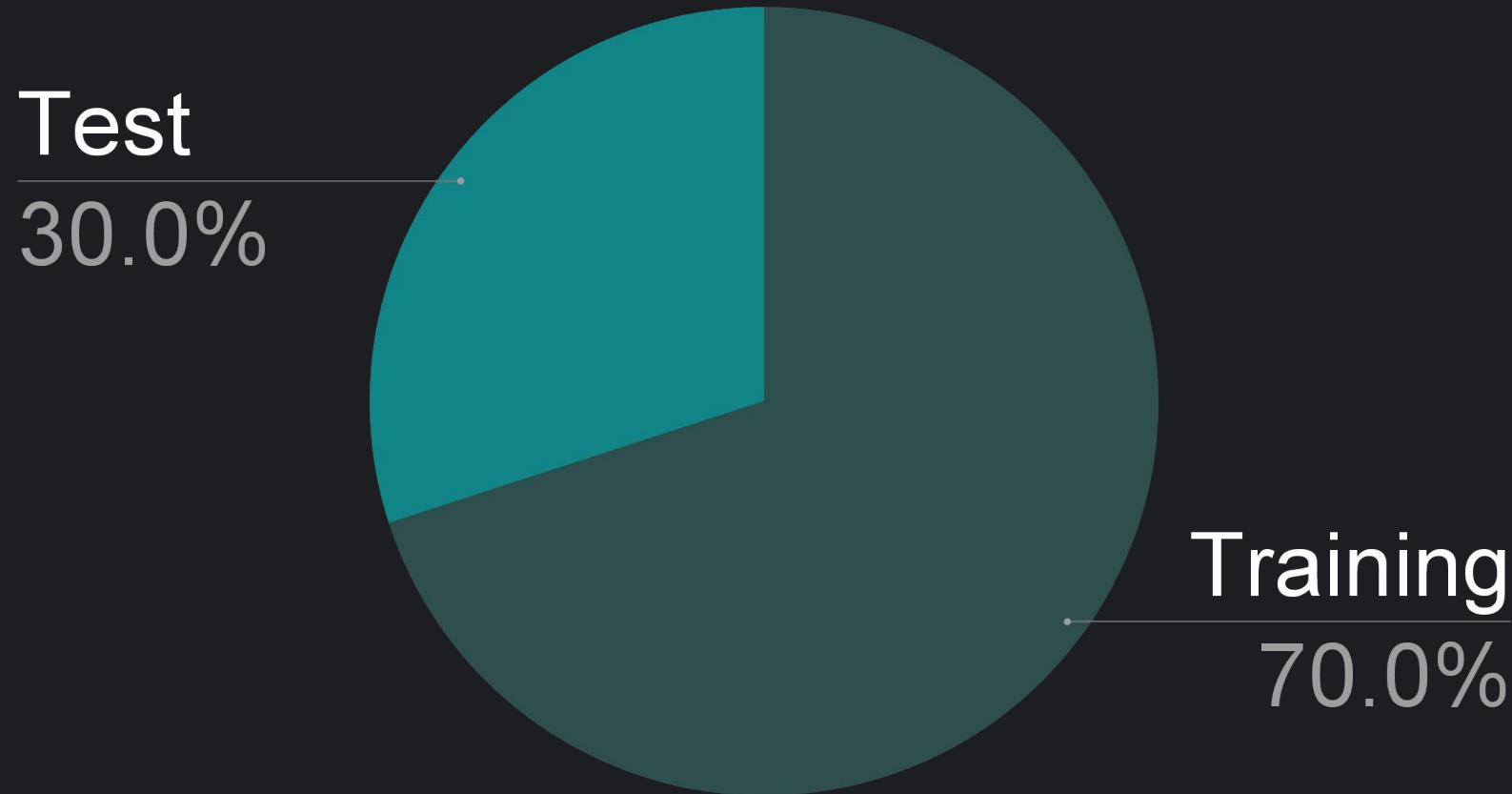


# RESAMPLED BARPLOT OF CUSTOMER CHURN





## FURTHER DATA PREPARATION



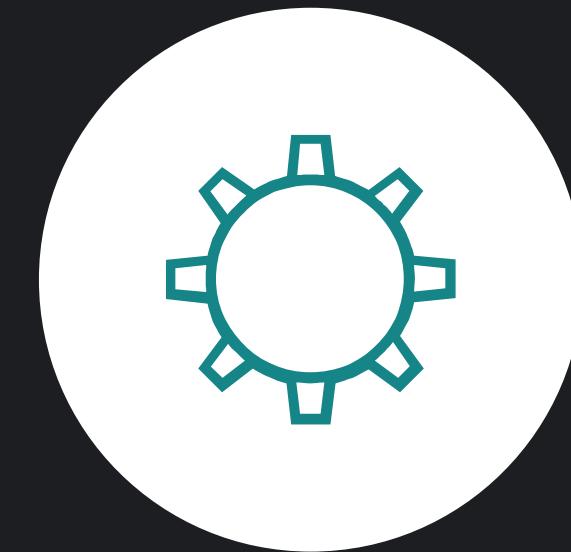


# FURTHER DATA PREPARATION

$$\sim(\mu, \sigma^2) \longrightarrow \sim(0, 1)$$



# PREDICTIVE MODELS





# Accuracy

The bias-variance tradeoff



# Accuracy

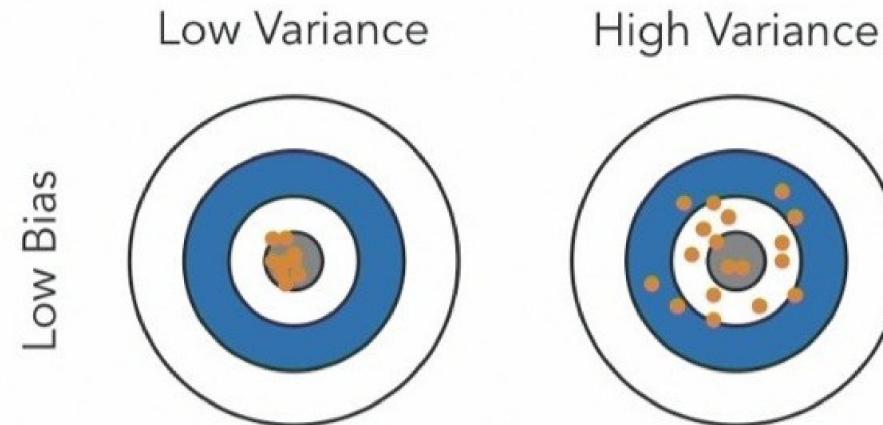
The bias-variance tradeoff





# Accuracy

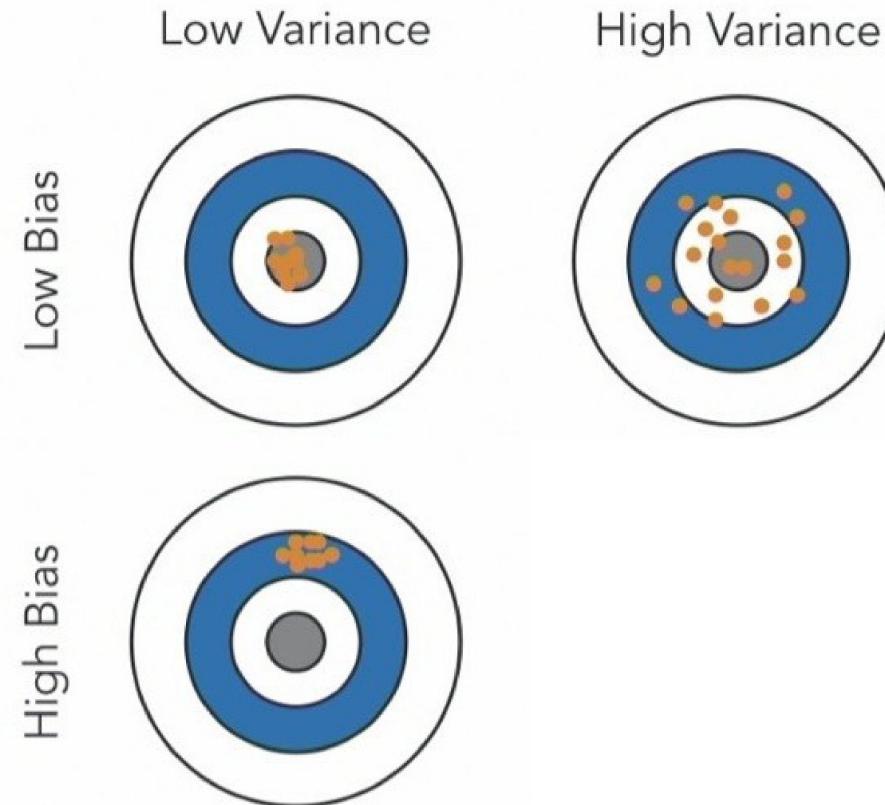
The bias-variance tradeoff





# Accuracy

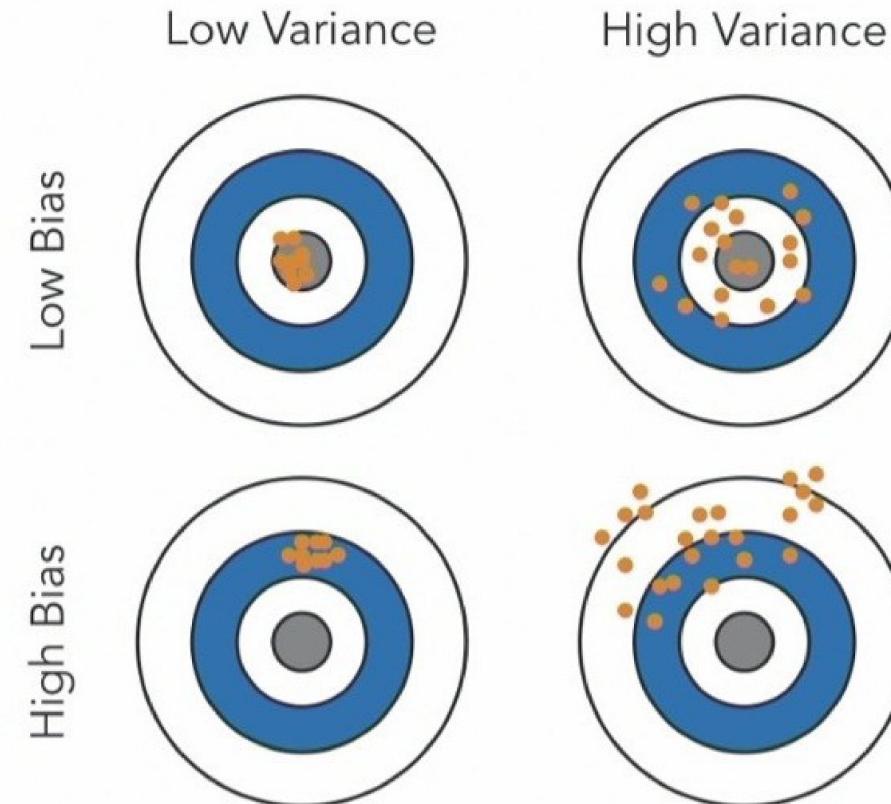
The bias-variance tradeoff





# Accuracy

The bias-variance tradeoff





# Confusion Matrix



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No		
	Yes		



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No		
	Yes		



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No	True Negative	
	Yes		



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No	True Negative	
	Yes		True Positive



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No	True Negative	
	Yes	False Positive	True Positive



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No	True Negative	False Negative
	Yes	False Positive	True Positive



# Confusion Matrix

		Actual	
		No	Yes
Predicted	No	True Negative	False Negative
	Yes	False Positive	True Positive

What we are interested in  
getting right



# Sensitivity VS Specificity

ALSO KNOWN AS TRUE POSITIVE RATE  
(TPR) AND TRUE NEGATIVE RATE (TNR)



# Sensitivity VS Specificity

ALSO KNOWN AS TRUE POSITIVE RATE (TPR) AND TRUE NEGATIVE RATE (TNR)

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$



# Sensitivity VS Specificity

ALSO KNOWN AS TRUE POSITIVE RATE (TPR) AND TRUE NEGATIVE RATE (TNR)

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$



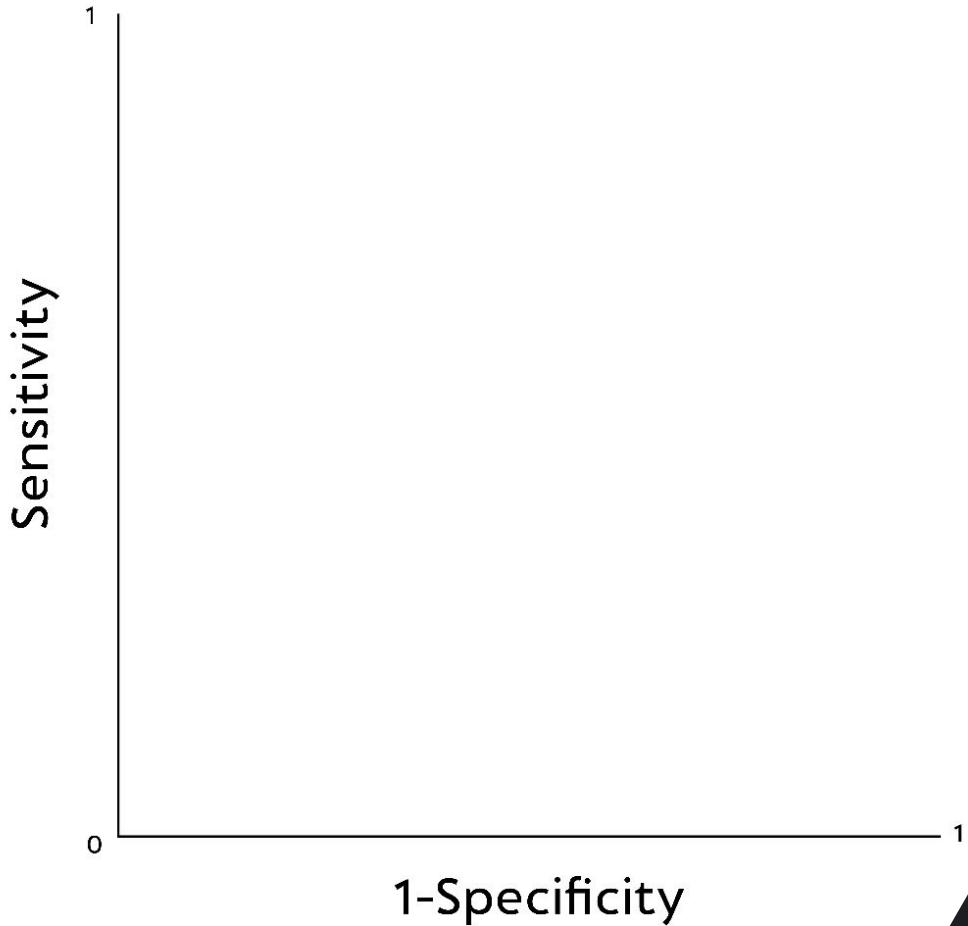
What we are interested in  
getting right

# Sensitivity VS Specificity

ALSO KNOWN AS TRUE POSITIVE RATE  
(TPR) AND TRUE NEGATIVE RATE (TNR)

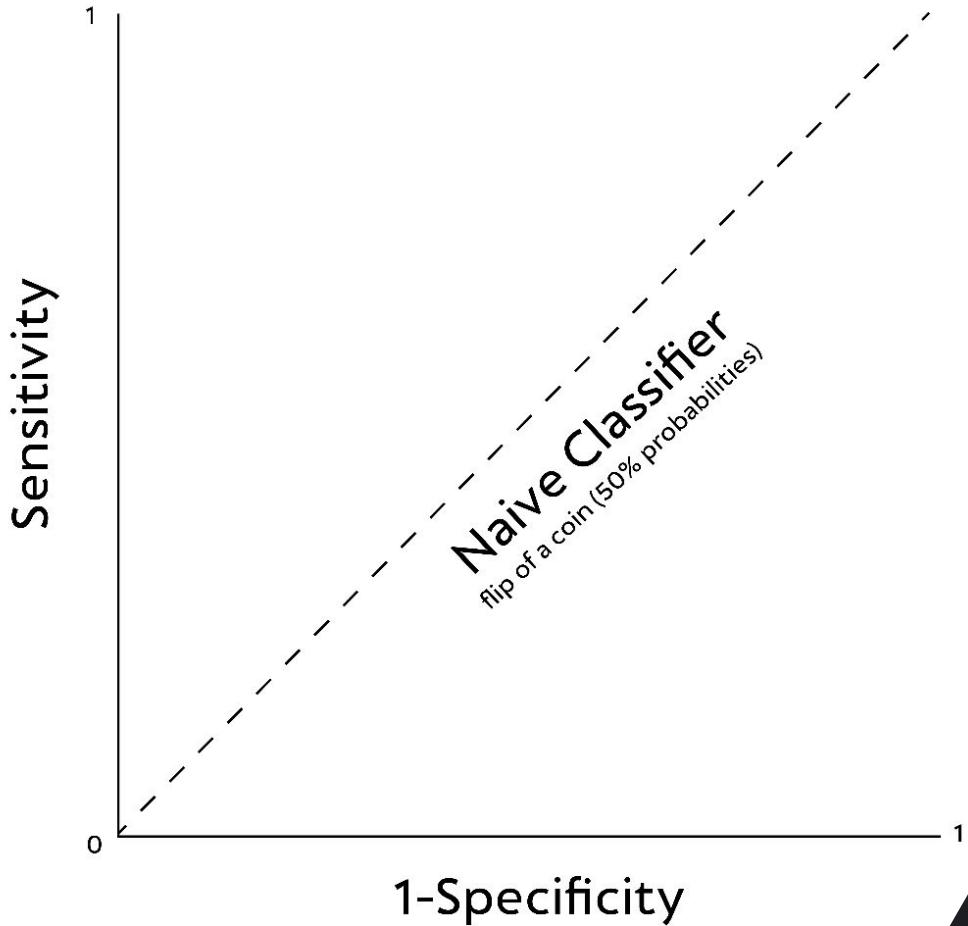
Sensitivity = 
$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Specificity = 
$$\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

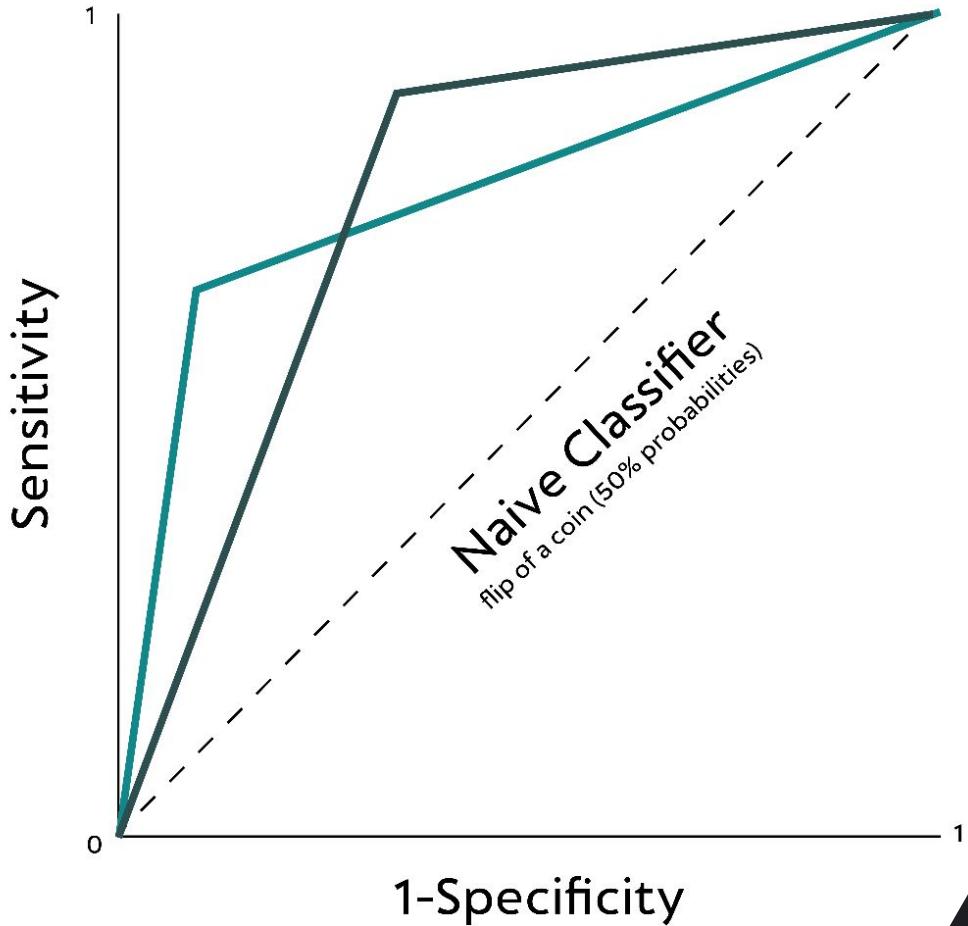


Also known as False  
Positive Rate

**ROC**  
Receiver operating  
characteristic curve



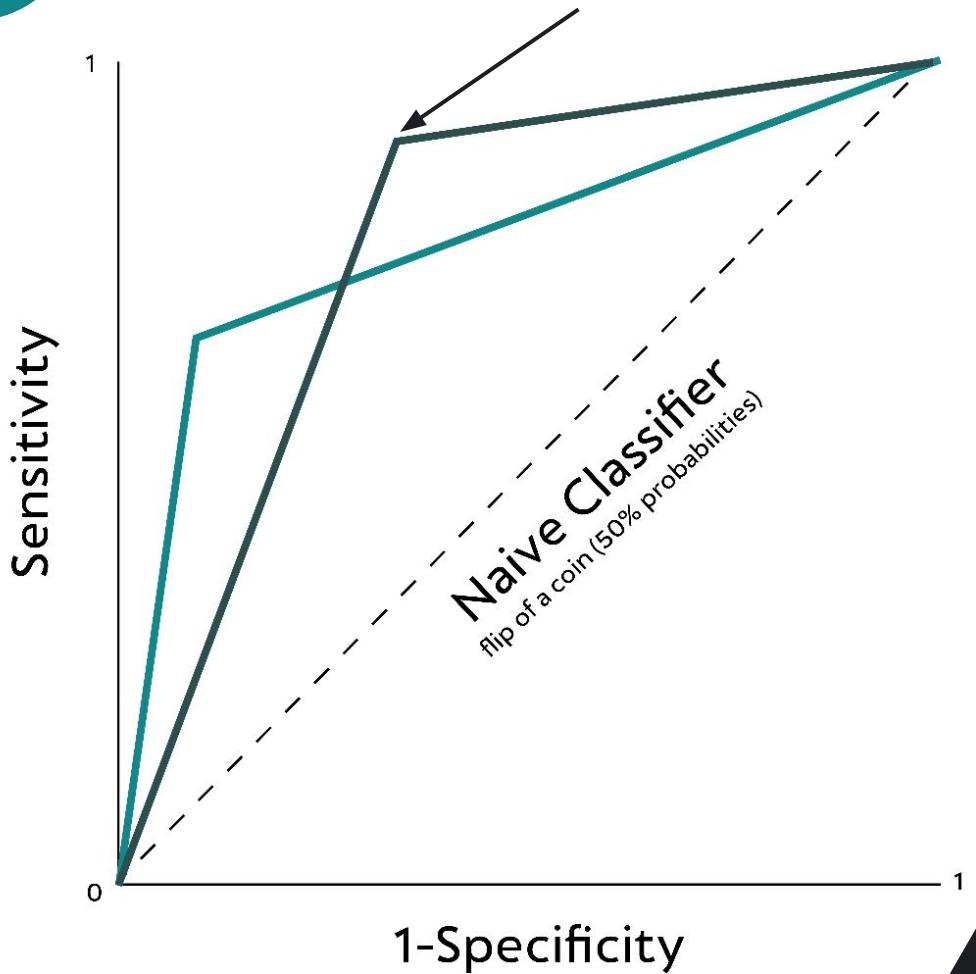
**ROC**  
Receiver operating  
characteristic curve



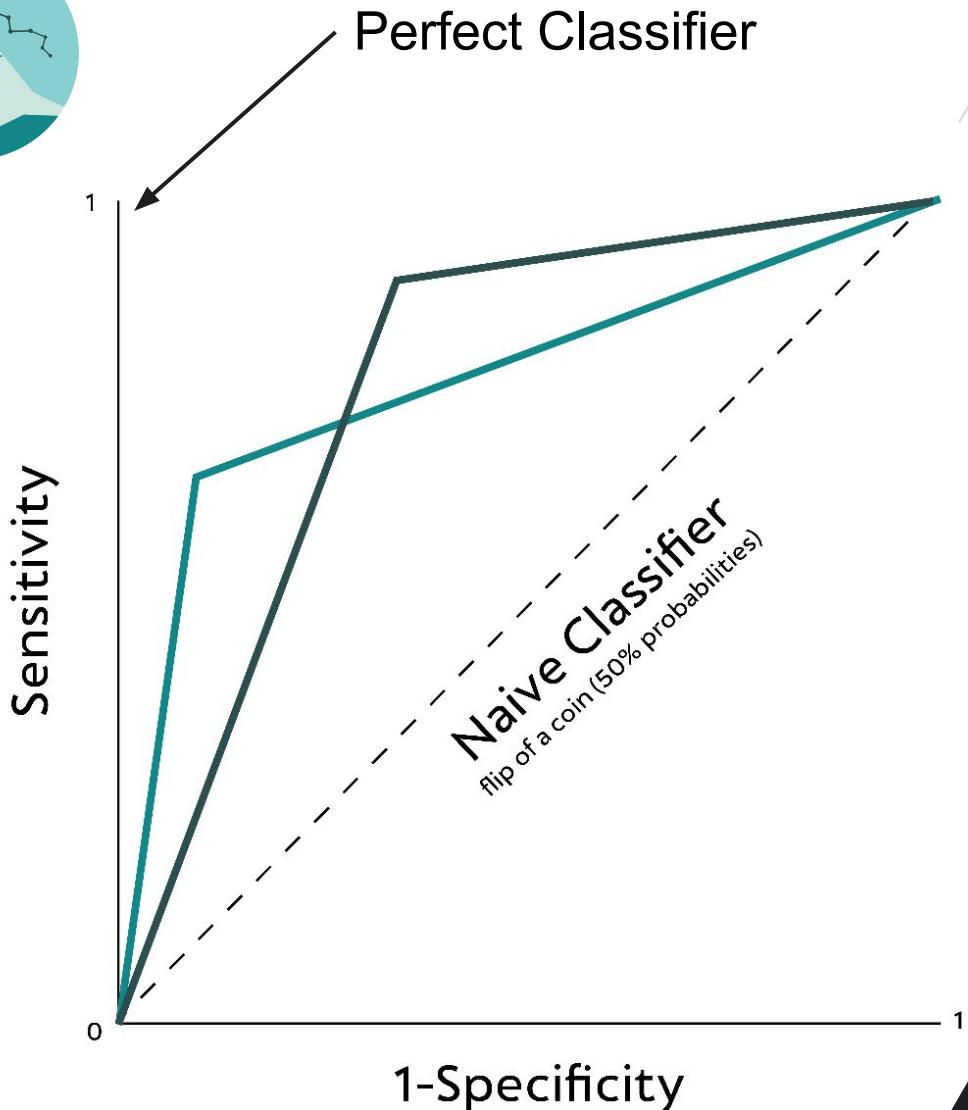
**ROC**  
Receiver operating  
characteristic curve



What we want for  
your model



**ROC**  
Receiver operating  
characteristic curve



**ROC**  
Receiver operating  
characteristic curve

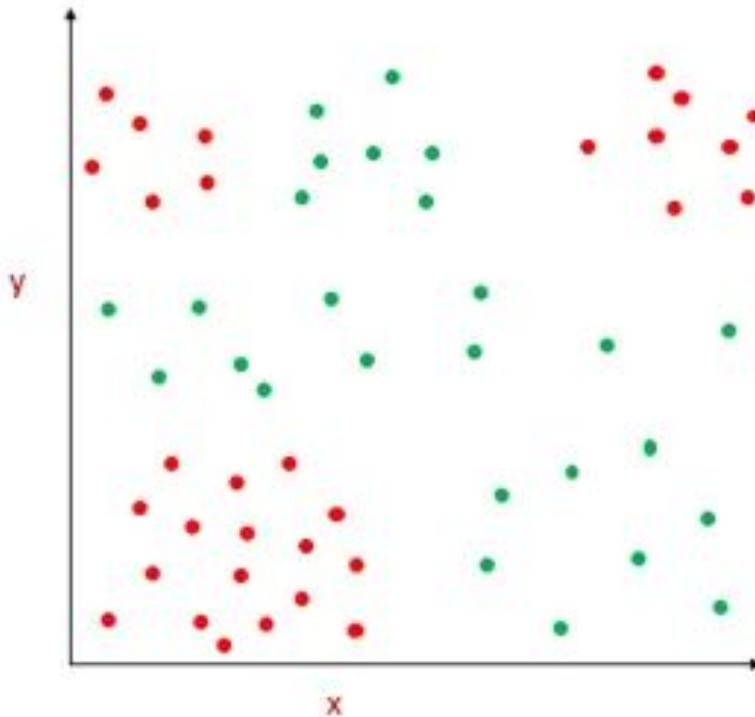


# BEST MODEL BAGGING TREES



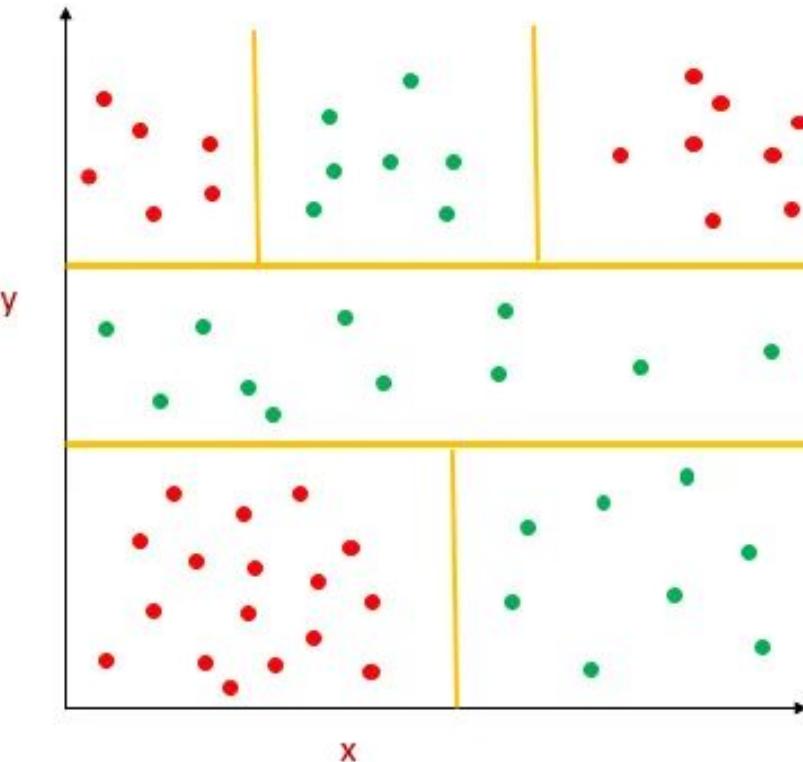


## How Trees work



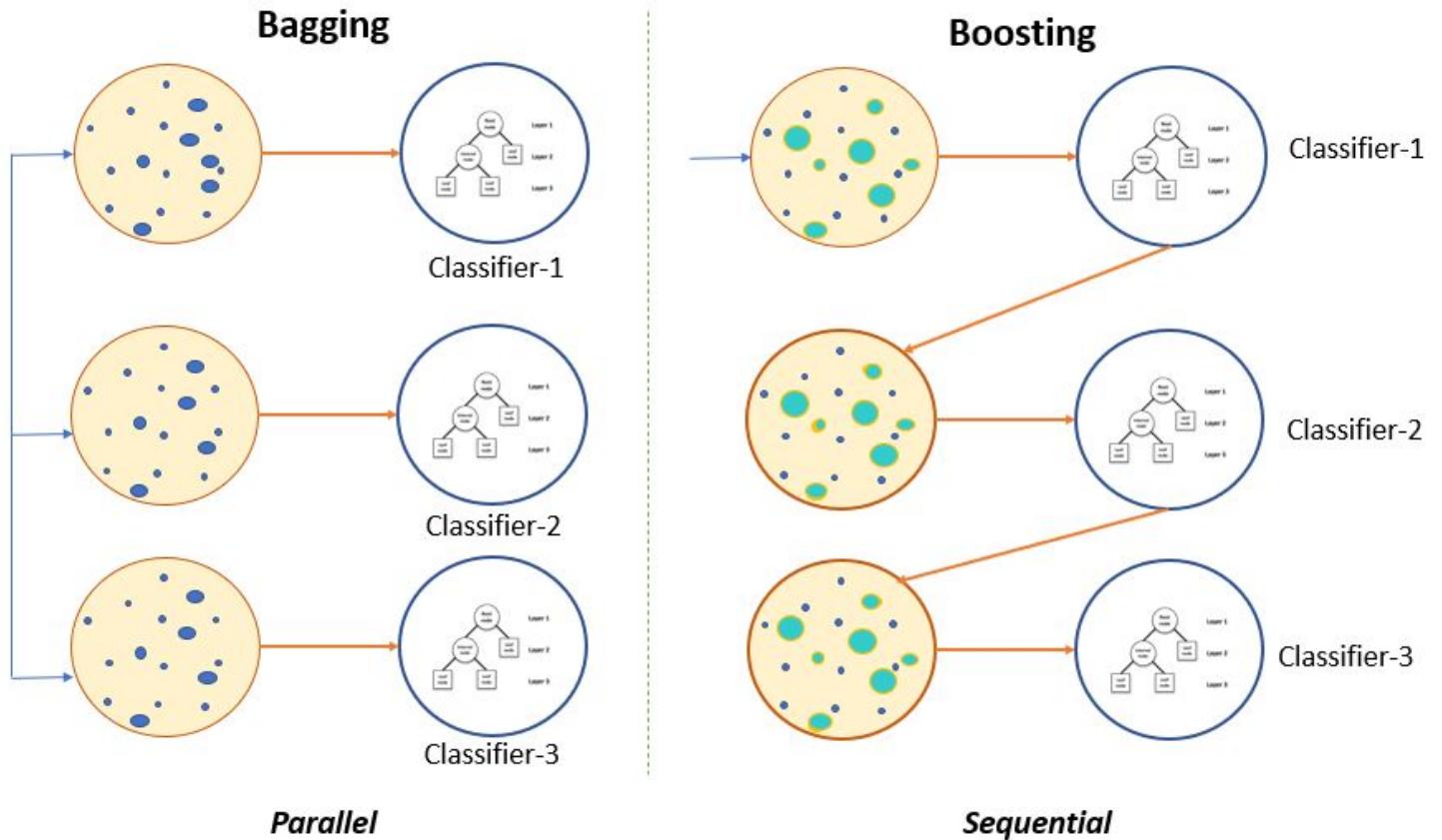


## How Trees work



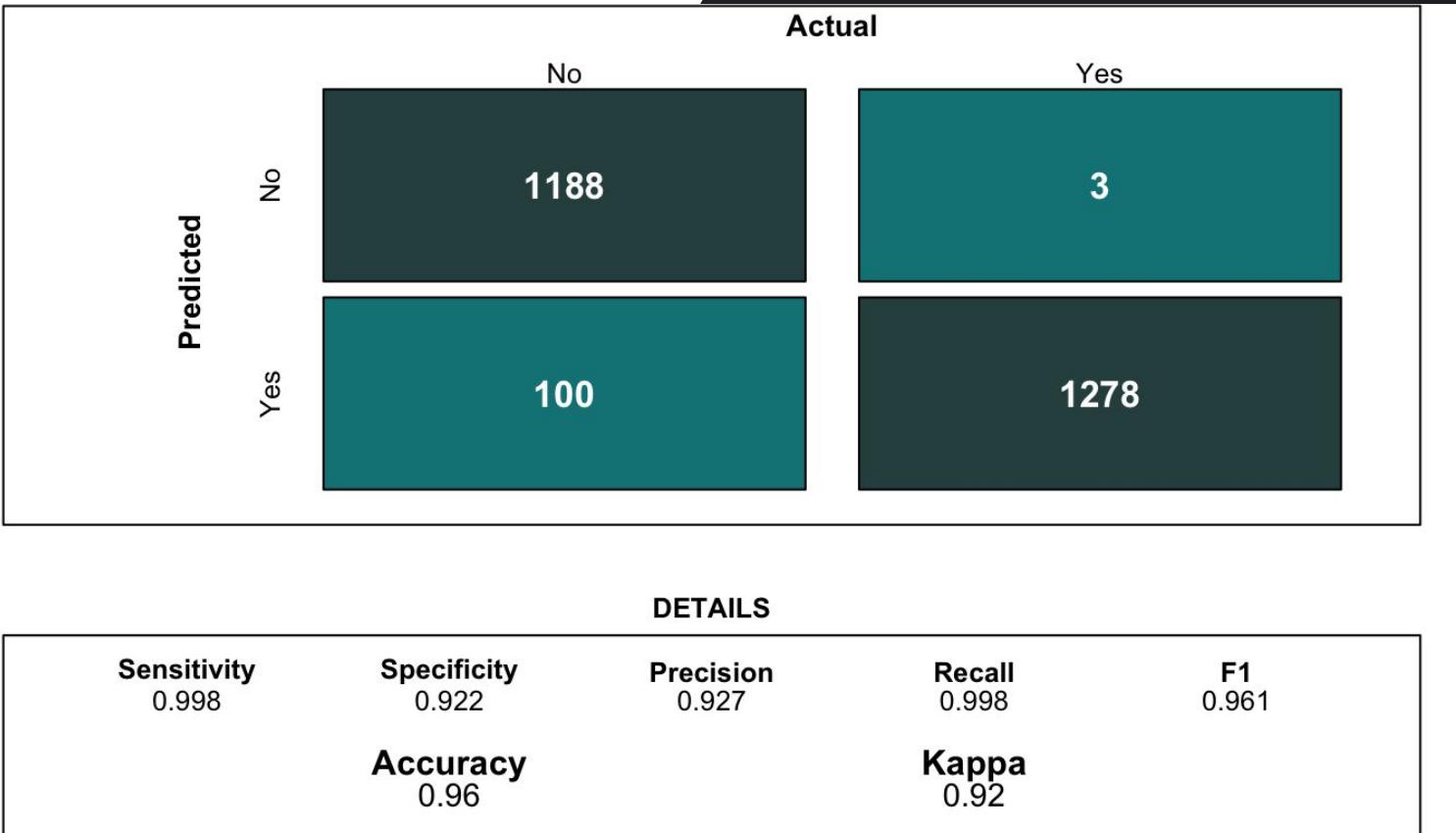


# Bagging vs. Boosting Trees



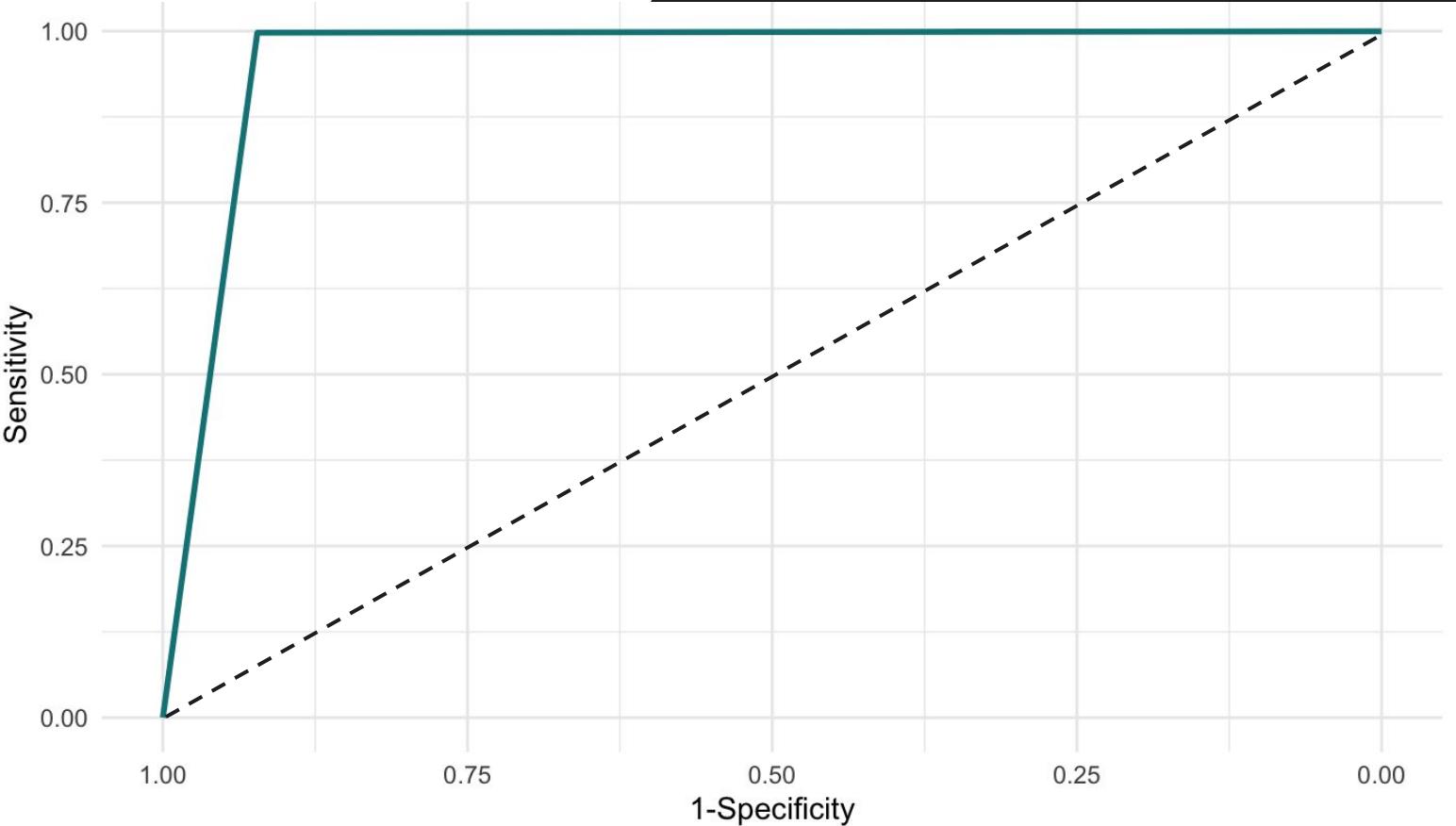


# RESULTS FOR OVERSAMPLED DATA



## Confusion Matrix

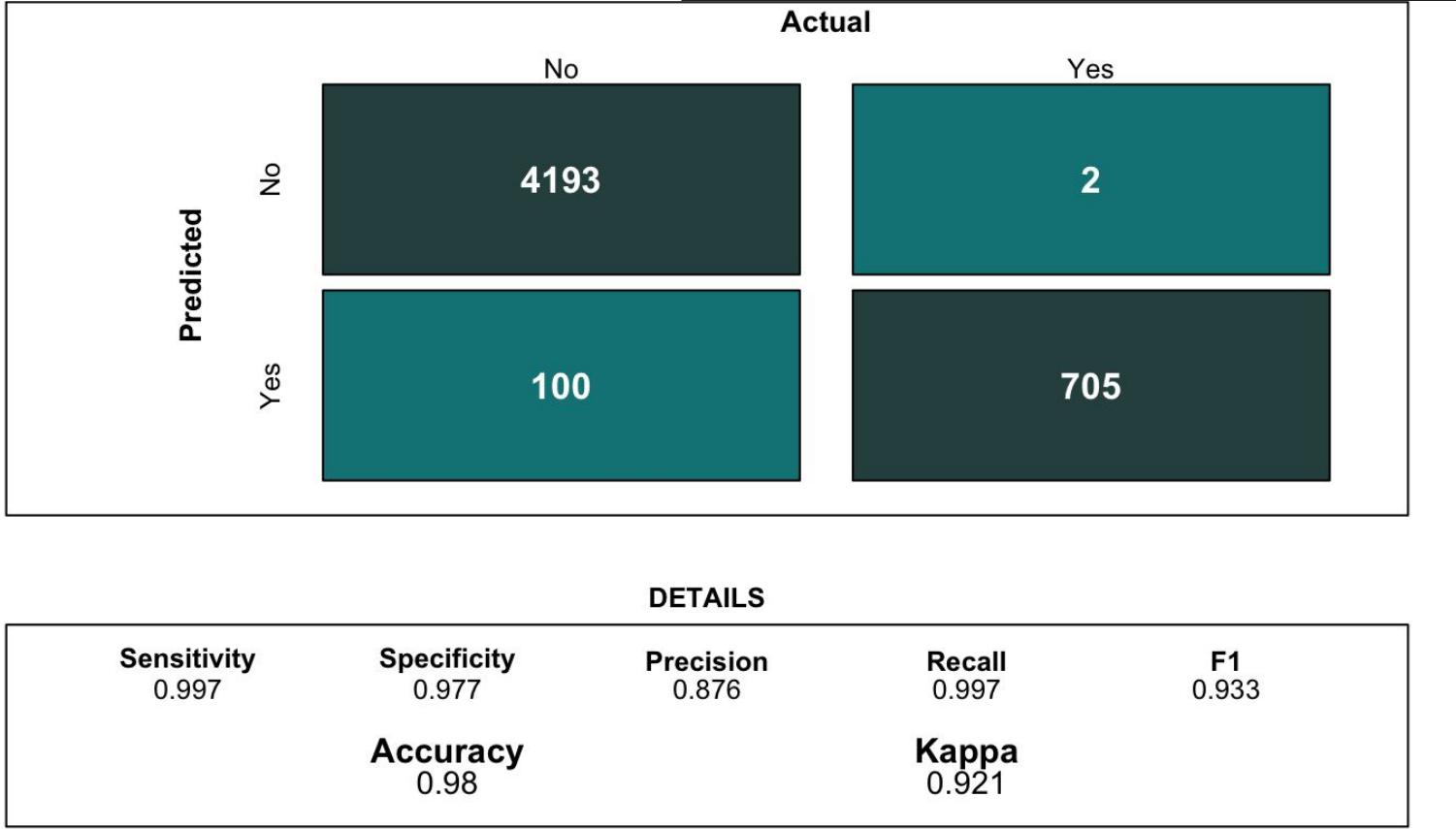
# RESULTS FOR OVERSAMPLED DATA



# ROC



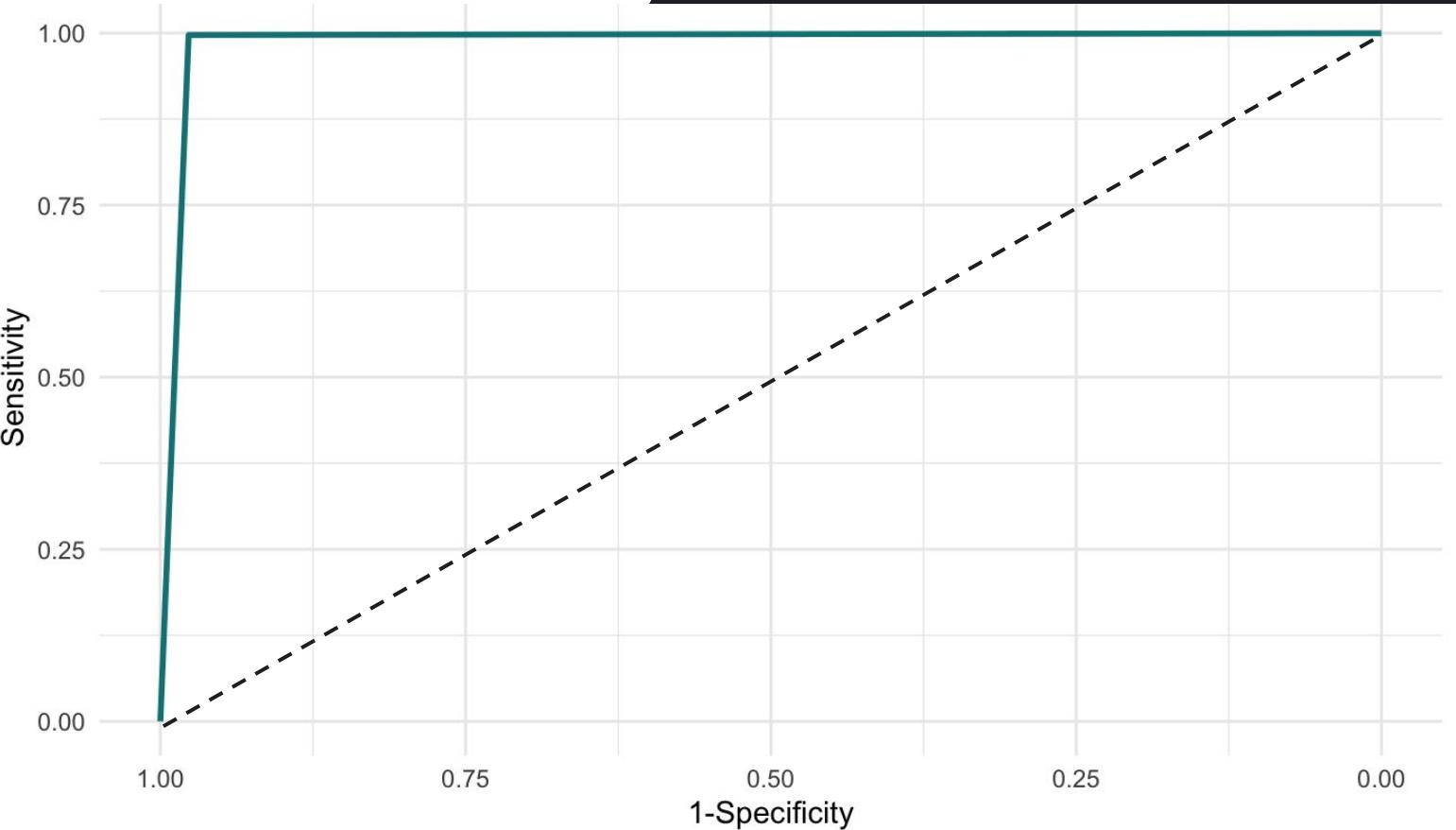
# RESULTS FOR ORIGINAL DATA



## Confusion Matrix



# RESULTS FOR ORIGINAL DATA



ROC



# BEST MODELS

## MAJORITY VOTING





# Voting System Amongst 3 Models

- Boosting Trees
- Bagging Trees
- KNN (K-nearest neighbor)





## Boosting Trees

VOTE “YES”





Boosting  
Trees

VOTE “YES”

Bagging  
Trees

VOTE “YES”

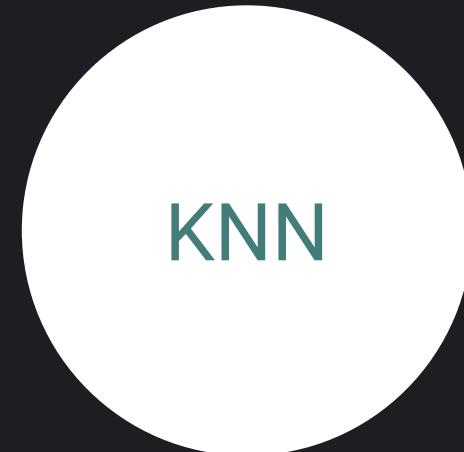




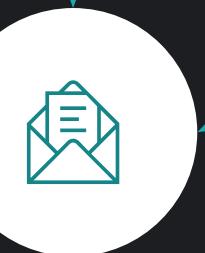
**VOTE “YES”**



**VOTE “YES”**



**VOTE “NO”**

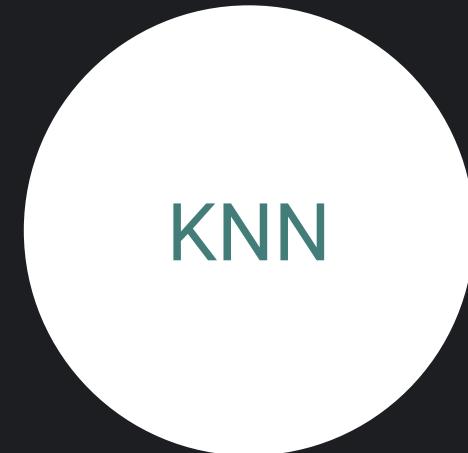




**VOTE “YES”**



**VOTE “YES”**



**VOTE “NO”**

A small white circle containing a teal icon of an envelope with a document inside. Below it is the text "FINAL VOTE “YES”" in a large, bold, white sans-serif font.

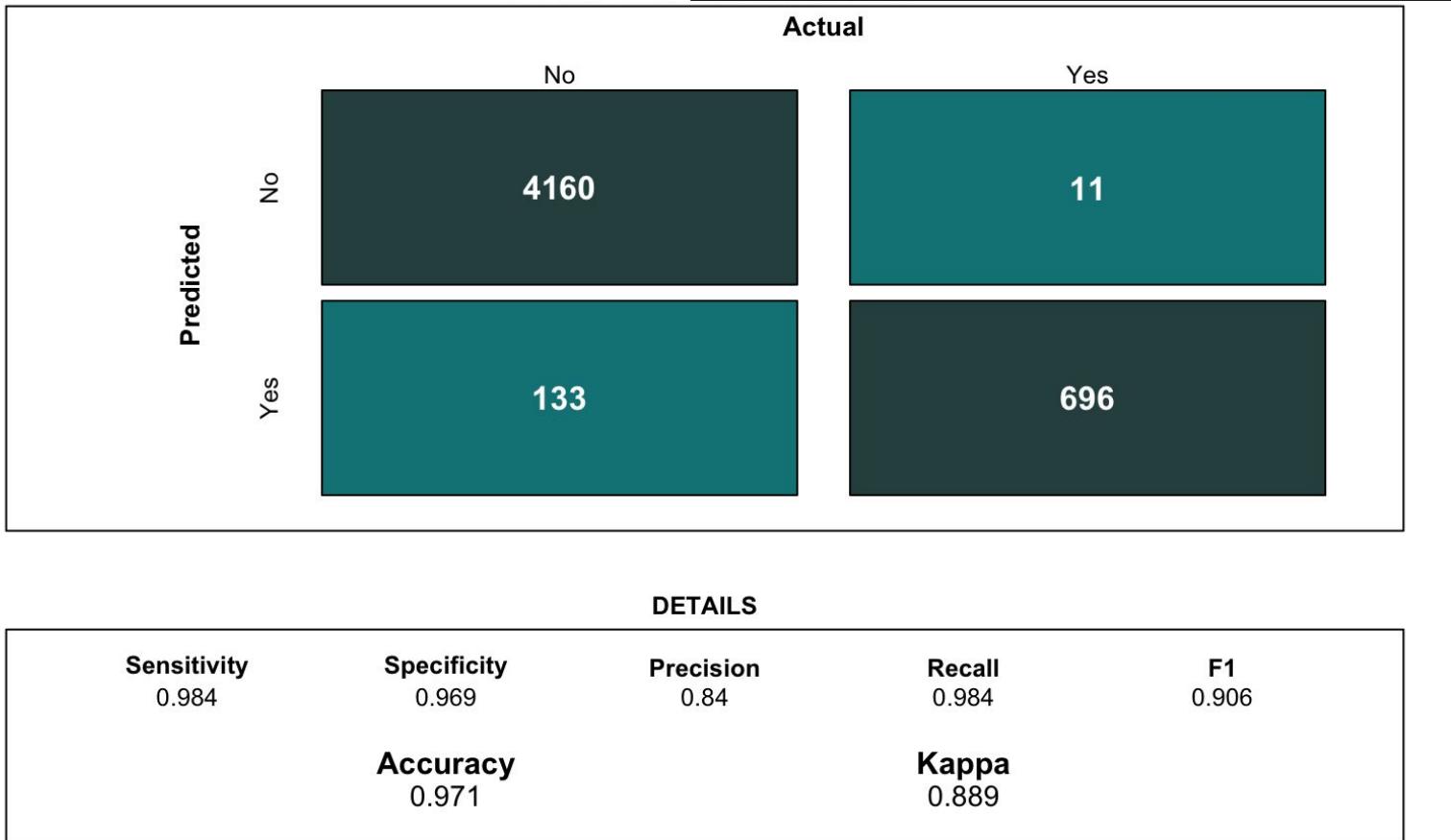


# Advantages

- Accuracy
- Robustness



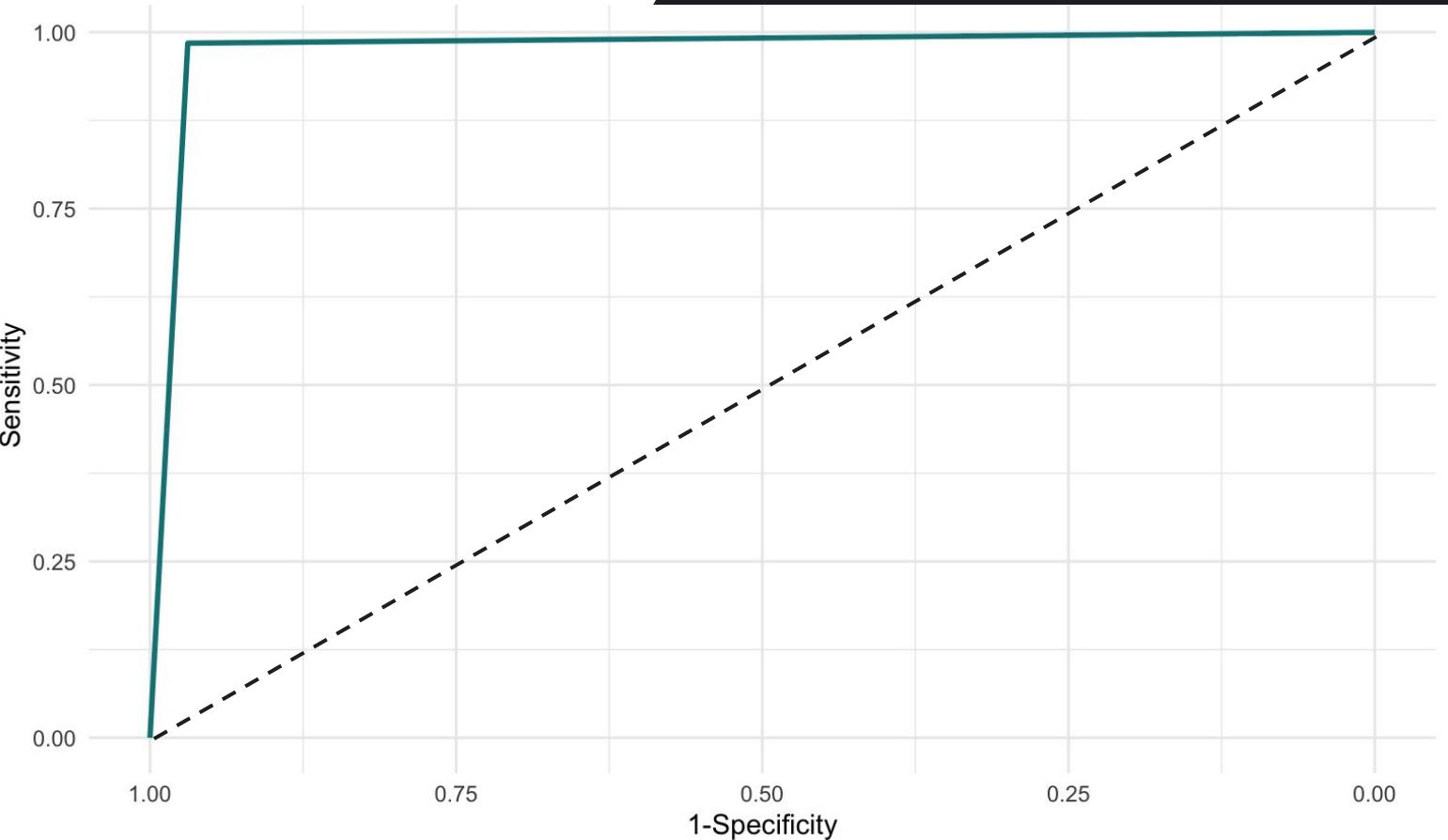
# RESULTS FOR ORIGINAL DATA



## Confusion Matrix



# RESULTS FOR ORIGINAL DATA



ROC



# MAIN DRIVERS OF CUSTOMER CHURN





## TOP 3 DRIVERS



DAY CHARGEABLE  
MINUTE / TOTAL DAY  
MINUTE

14%

IMPORTANCE PROPORTION OF THE VARIABLE FROM BAGGING TREES MODEL



## TOP 3 DRIVERS



DAY CHARGEABLE  
MINUTE / TOTAL DAY  
MINUTE

14%



ACCOUNT LENGTH (IN  
WEEKS)

5.5%

IMPORTANCE PROPORTION OF THE VARIABLE FROM BAGGING TREES MODEL



## TOP 3 DRIVERS



DAY CHARGEABLE  
MINUTE / TOTAL DAY  
MINUTE

14%



ACCOUNT LENGTH (IN  
WEEKS)

5.5%



EVENING CHARGEABLE  
MINUTE / TOTAL EVENING  
MINUTE

5.4%

IMPORTANCE PROPORTION OF THE VARIABLE FROM BAGGING TREES MODEL



# CONCLUSION & RECOMMENDATIONS



# BEST MODEL

# BAGGING TREES

Predicted churn in a customer who actually churned

SENSITIVITY





# BEST MODEL

# BAGGING TREES

Predicted churn in a customer who actually churned

SENSITIVITY

99.7%

Correct

0.3%

Incorrect



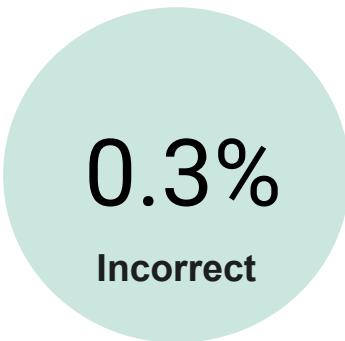


# BEST MODEL

## BAGGING TREES

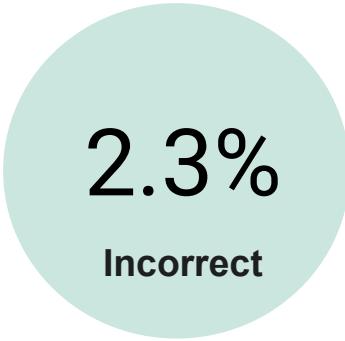
Predicted churn in a customer who actually churned

SENSITIVITY



Predicted no churn in a customer who actually didn't churn

SPECIFICITY

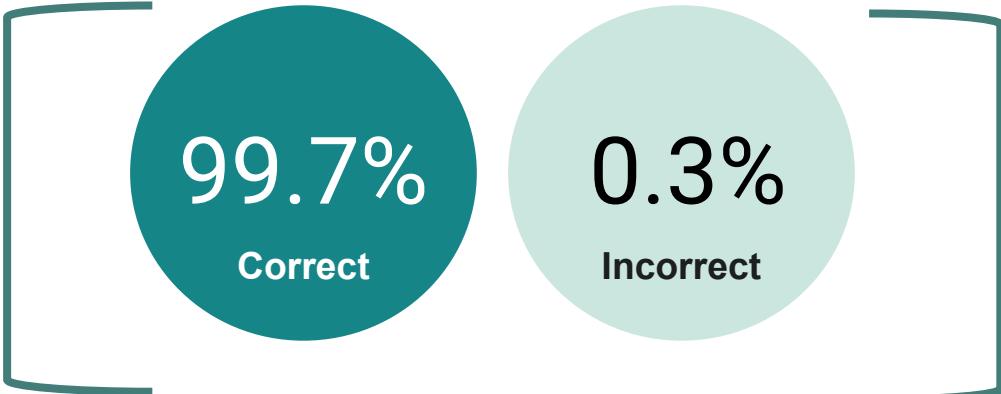




# BEST MODEL BAGGING TREES

Predicted churn in a customer who actually churned

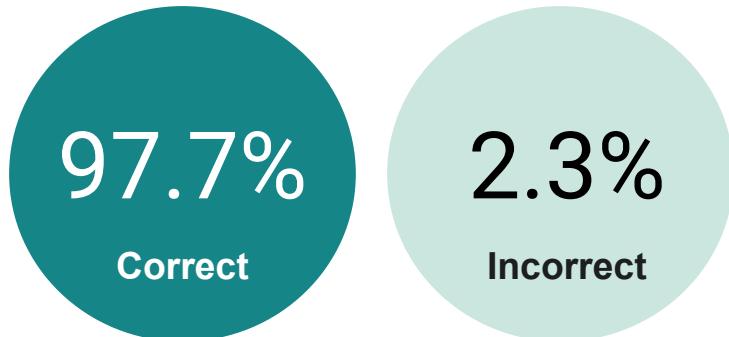
SENSITIVITY



YOUR MAIN  
TARGET

Predicted no churn in a customer who actually didn't churn

SPECIFICITY





# RECOMMENDATIONS

- Targeted offers based on the main drivers
- Improve the predictive model





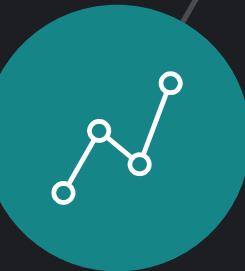
## STEPS TO A BETTER PREDICTIVE MODEL



COLLECT CUSTOMER  
SATISFACTION  
FEEDBACK



DATA EXPLORATION  
ANALYSIS AND FEATURE  
ENGINEERING



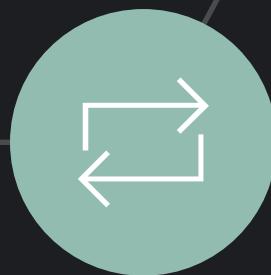
RE-EVALUATE MODEL  
SELECTION



## STEPS TO A BETTER PREDICTIVE MODEL



BEST MODEL TESTING  
ON NEW DATA AND  
FOLLOW-UP



AUTOMATE THE  
PROCESS



BETTER INFORMED  
DECISIONS BASED ON  
CUSTOMER FEEDBACK  
AND OTHER METRICS



**THANK YOU FOR YOUR  
ATTENTION**



ALPINE  
INSIGHTS

## CONTACT US

### MAILING ADDRESS

Av. Pictet-de-Rochemont 3, 1207 Genève

### EMAIL ADDRESS

[office@alpineinsights.com](mailto:office@alpineinsights.com)

### PHONE NUMBER

0041 77 482 57 12





# ADDITIONAL SLIDES

---



## OTHERS METRICS

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

$$\text{Cohen's kappa} = \frac{\text{Observed agreement} - \text{Expected agreement}}{1 - \text{Expected agreement}}$$