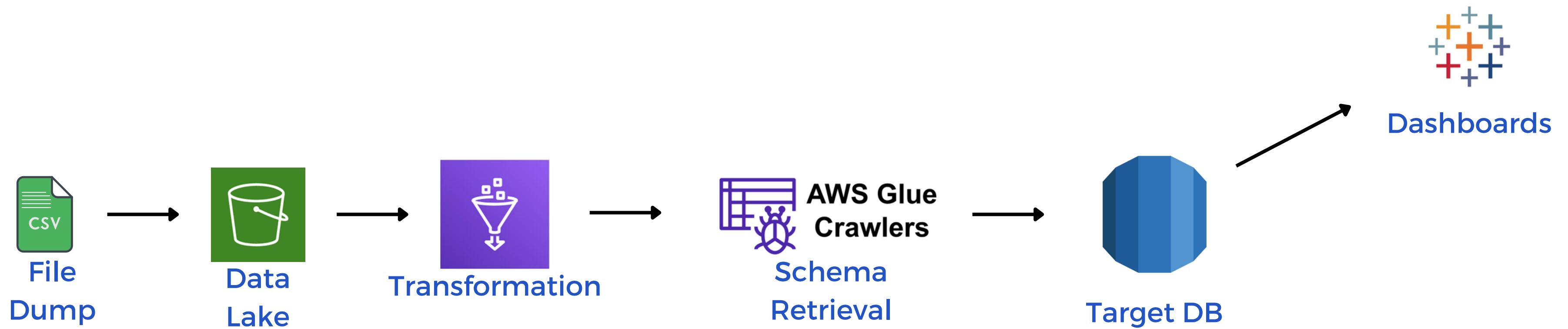


Unveiling Patterns in Telecom Churn

Automated ETL Pipeline



- Consolidated Storage
- ETL Health Monitoring
- Execution History
- Scheduled Weekly:
 - Job runs
 - Target DB Schema validation

Predictive Model

Feature Selection

Input Features	Description
bundleID	A unique ID combination of services (322)
contractID	A unique ID combination of contract, billing, and payment method (24)
numAdminTickets	Number of admin tickets
numTechTickets	Number of tech tickets
gender	Is customer is male or female
SeniorCitizen	If customer is a senior citizen or not (Yes or No)
Partner	If customer has a partner or not (Yes or No)
Dependents	If customer has dependents or not (Yes or No)
tenure	How long the customer has been with the company
MonthlyCharges	The amount charged to customer each month for services used

	Churn	bundleID	numAdminTickets	numTechTickets	contractID	gender	SeniorCitizen	Partner	Dependents	tenure	MonthlyCharges
0	No	17	0	0	7	Female	0	Yes	No	1	29.85
1	No	105	0	0	12	Male	0	No	No	34	56.95
2	Yes	113	0	0	8	Male	0	No	No	2	53.85
3	No	45	0	3	9	Male	0	No	No	45	42.30
4	Yes	129	0	0	7	Female	0	No	No	2	70.70
...
7038	No	241	0	0	16	Male	0	Yes	Yes	24	84.80
7039	No	285	0	5	14	Female	0	Yes	Yes	72	103.20
7040	No	33	0	0	7	Female	0	Yes	Yes	11	29.60
7041	Yes	258	0	0	8	Male	1	Yes	No	4	74.40
7042	No	176	2	0	21	Male	0	No	No	66	105.65

7043 rows × 11 columns

Originally 27 features

EDA

Continuous Features							
	Count	% Miss.	Card.	Min	1st Quart.	Mean	Median
bundleID	7043	0.0	322	1.00	129.0	188.846798	193.00
numAdminTickets	7043	0.0	6	0.00	0.0	0.515689	0.00
numTechTickets	7043	0.0	10	0.00	0.0	0.419566	0.00
contractID	7043	0.0	24	1.00	6.0	10.466988	8.00
SeniorCitizen	7043	0.0	2	0.00	0.0	0.162147	0.00
tenure	7043	0.0	73	0.00	9.0	32.371149	29.00
MonthlyCharges	7043	0.0	1585	18.25	35.5	64.761692	70.35

	3rd Quart.	Max	Std. Dev.
bundleID	267.00	322.00	89.535517
numAdminTickets	0.00	5.00	1.275299
numTechTickets	0.00	9.00	1.250117
contractID	16.00	24.00	6.430688
SeniorCitizen	0.00	1.00	0.368612
tenure	55.00	72.00	24.559481
MonthlyCharges	89.85	118.75	30.090047

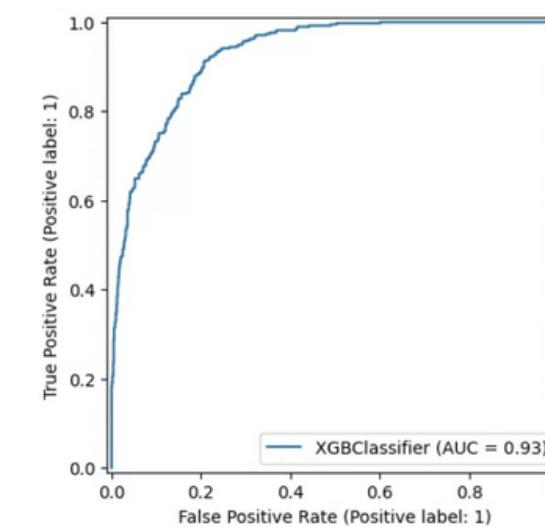
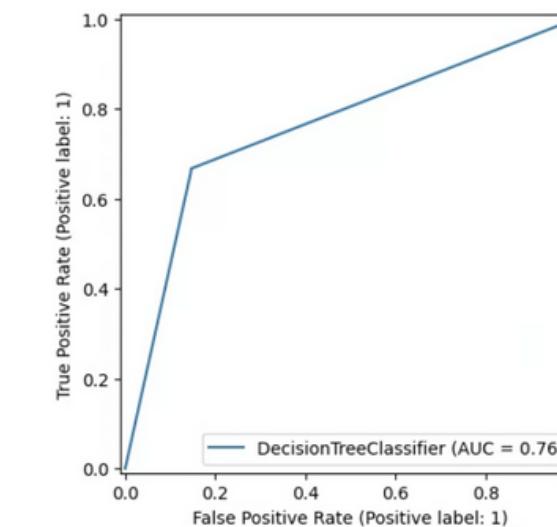
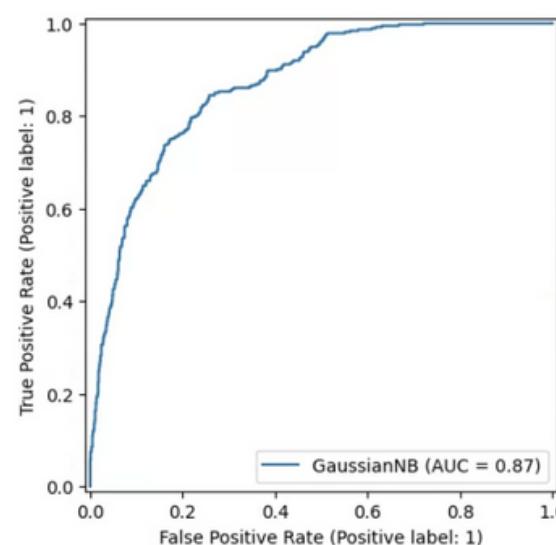
Categorical Features							
	Count	% Miss.	Card.	Mode	Mode Freq.	Mode %	2nd Mode
Churn	7043	0.0	2	No	5174	73.463013	Yes
gender	7043	0.0	2	Male	3555	50.475650	Female
...							
Churn					1869	26.536987	
gender					3488	49.524350	
Partner					3402	48.303280	
Dependents					2110	29.958824	

Predictive Model Cont.

Feature Engineering

```
# encoding dataset
label_encoder = LabelEncoder()
y_train = label_encoder.fit_transform(y_train)
y_test = label_encoder.transform(y_test)
x_train = pd.get_dummies(x_train)
x_test = pd.get_dummies(x_test)
```

Testing Different Models



Tuning the Best Model

```
# setting up lists of hyperparameters
loss1 = ['exponential', 'log_loss']
n_estimators1 = [50, 100, 150, 200, 250, 300]
learning_rate1 = [0.01, 0.1, 0.5, 1]
max_depth1 = [2, 4, 6, 8, 10, 12, 14, 16]
min_samples_leaf1 = [5, 10, 15, 20, 25, 30, 35, 40]
max_features1 = [None, 'sqrt', 'log2', 0.3, 0.4, 0.5, 0.6, 0.7, 0.8]
```

Results

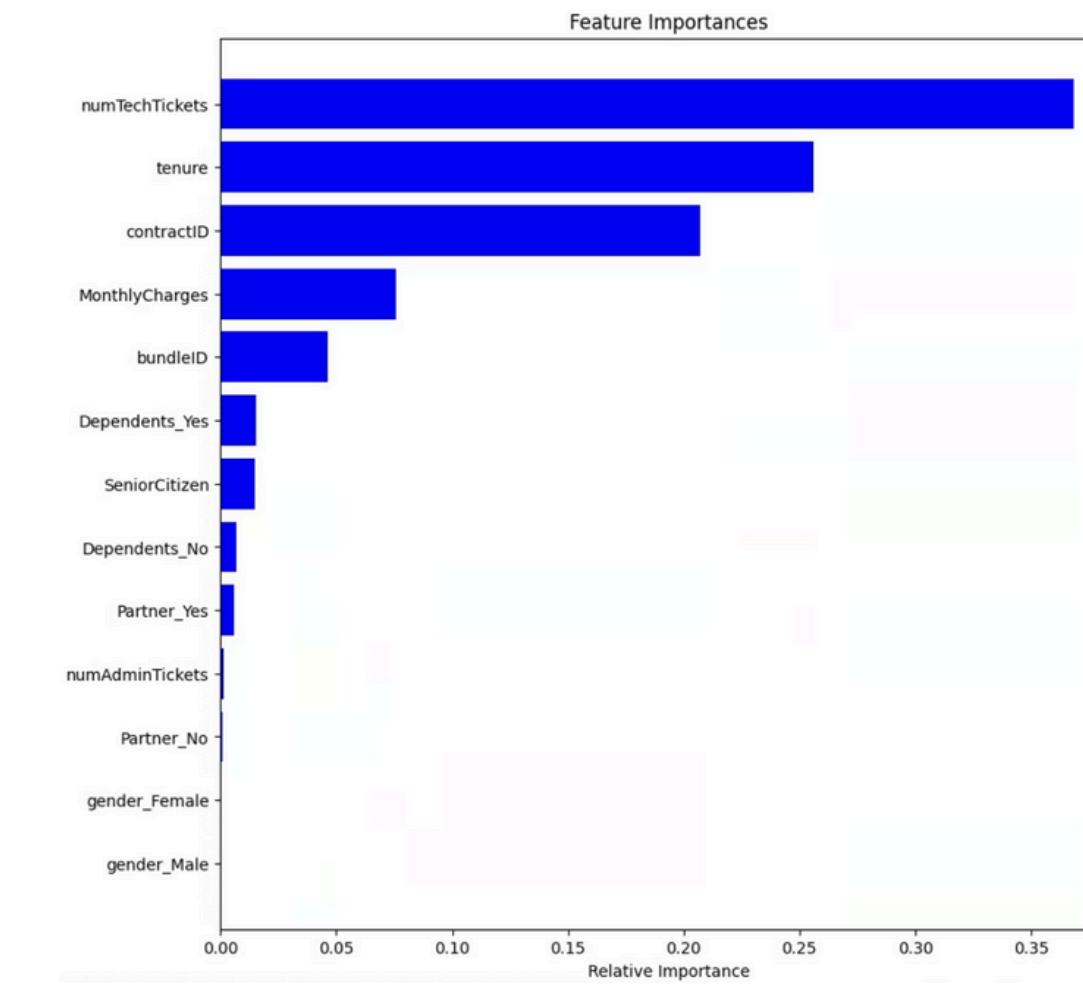
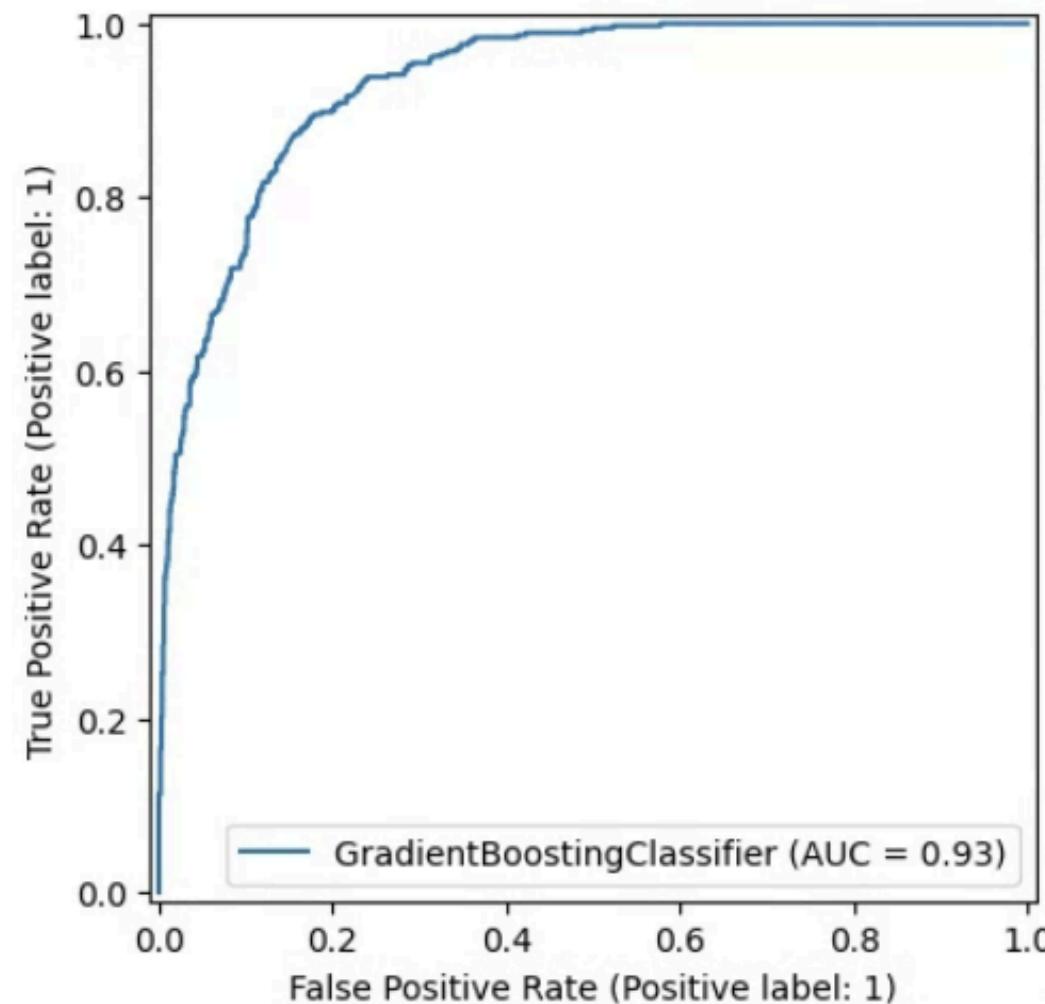
Gradient Boost

Accuracy: 86%

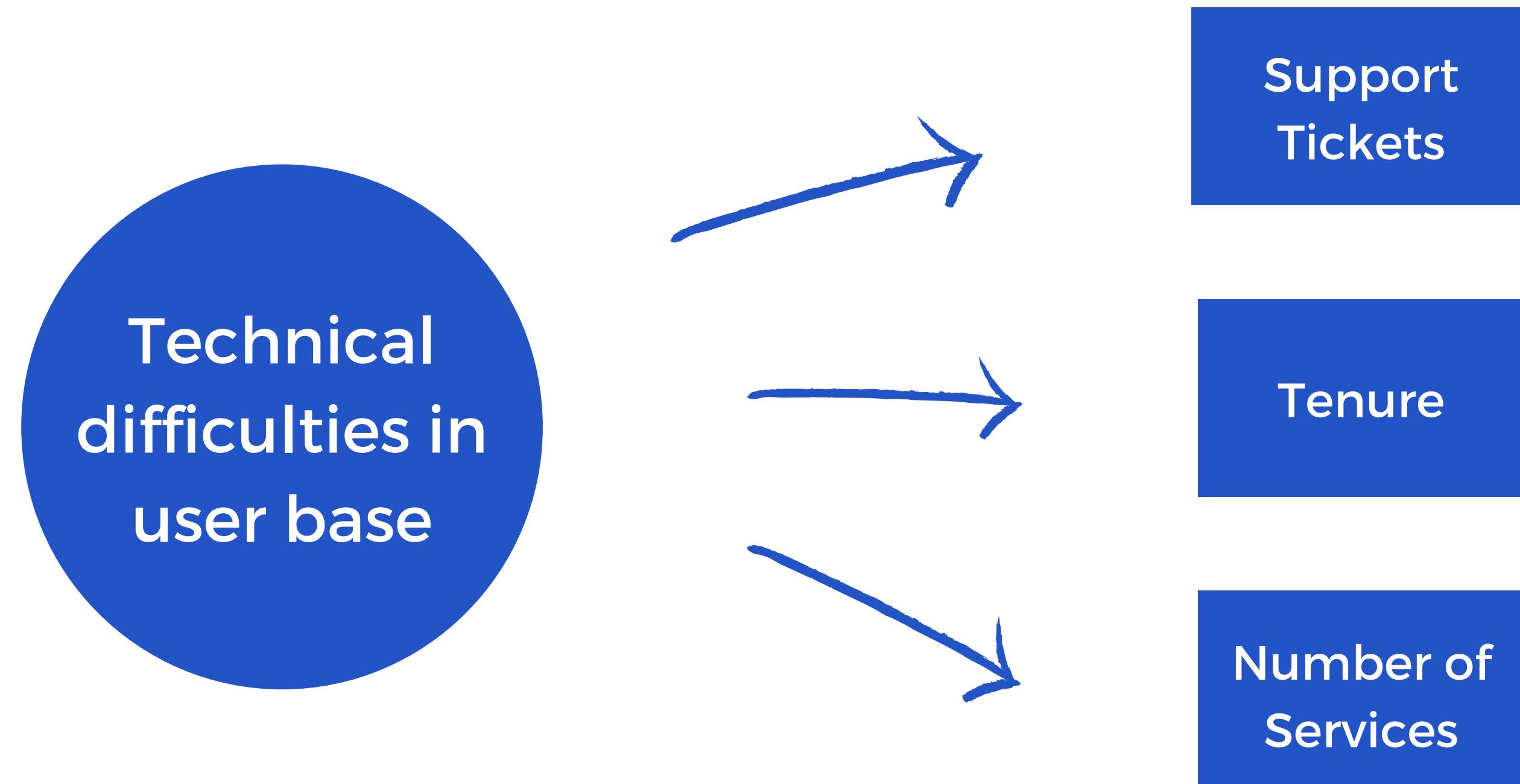
F1-Score: 73% for Churn

AUC = 0.93

[[947 89] [105 268]]		precision	recall	f1-score	support
	0	0.90	0.91	0.91	1036
	1	0.75	0.72	0.73	373
		accuracy		0.86	1409
		macro avg	0.83	0.82	1409
		weighted avg	0.86	0.86	1409
Accuracy: 0.8623136976579134					

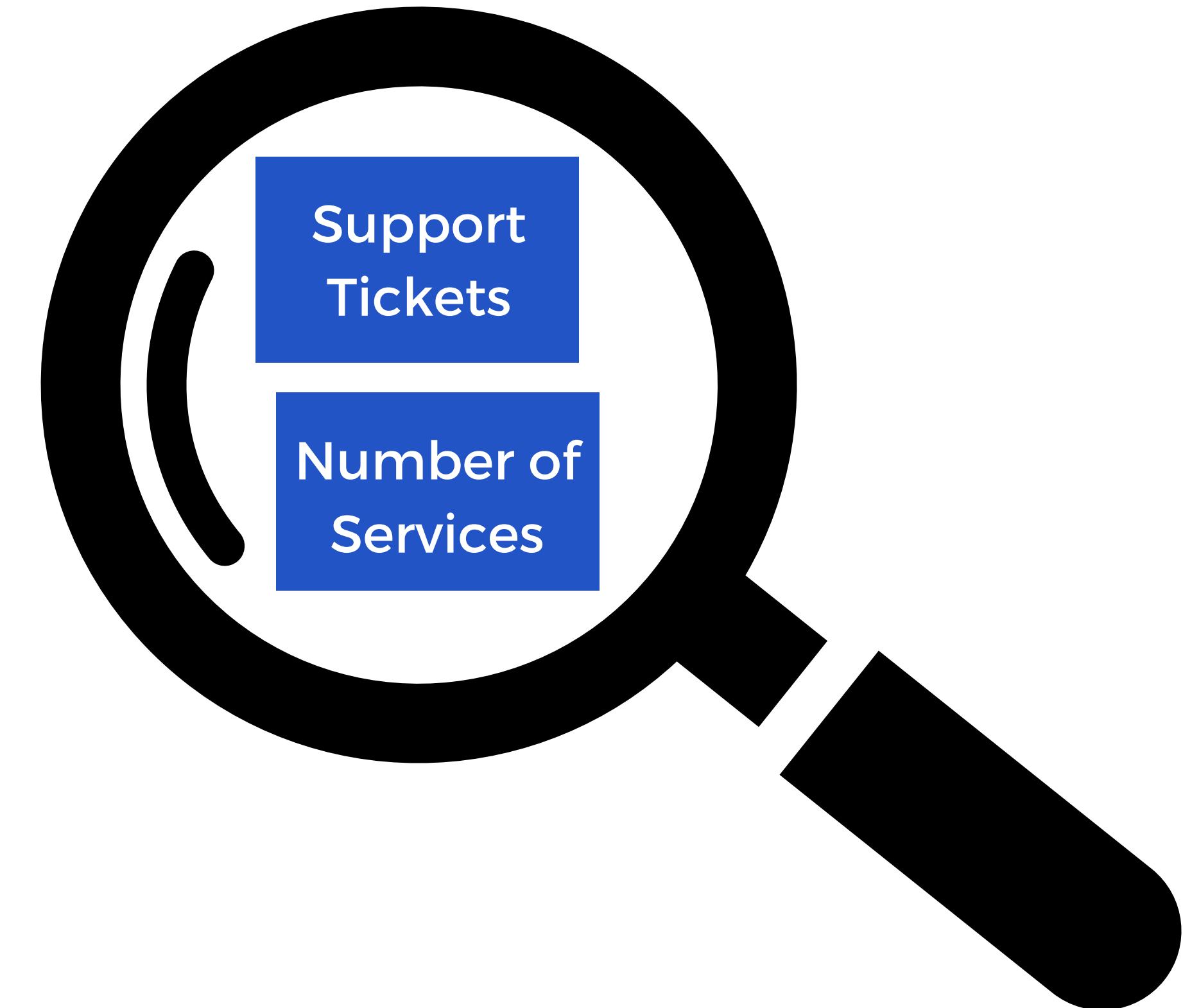


Descriptive Analytics

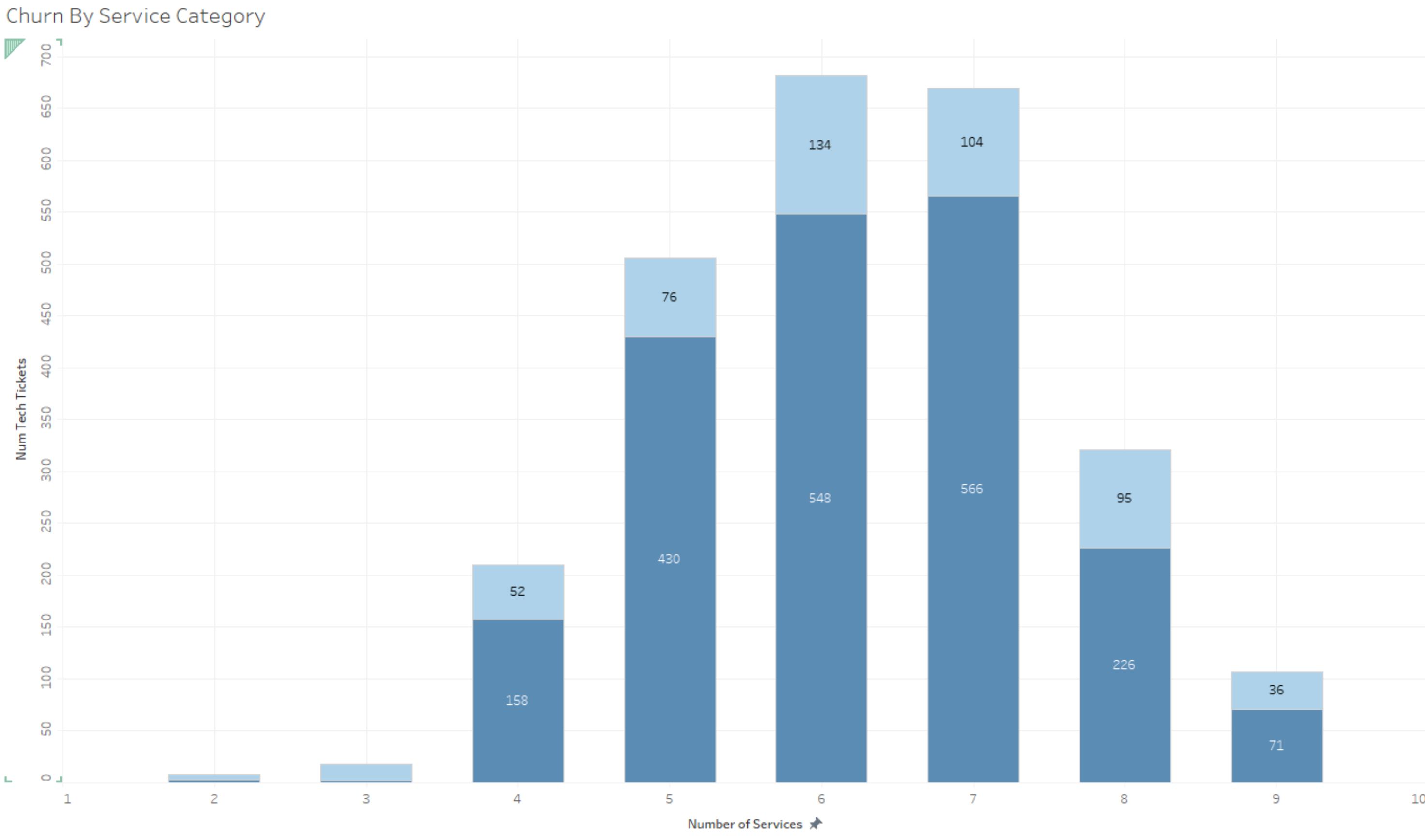


~~Descriptive Analytics~~

Diagnostic Analytics

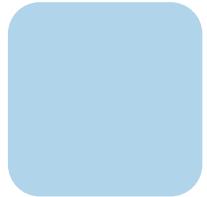
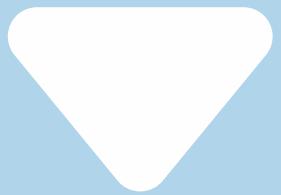


“Technical Support, how can I help?”

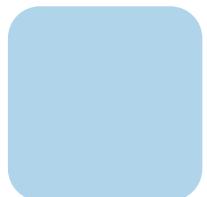


Customization

Phone Service



Single Line

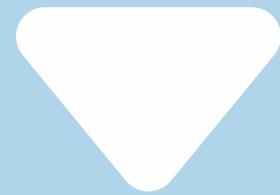


Multiple Lines



No Lines

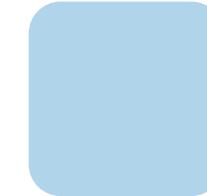
Internet Service



DSL

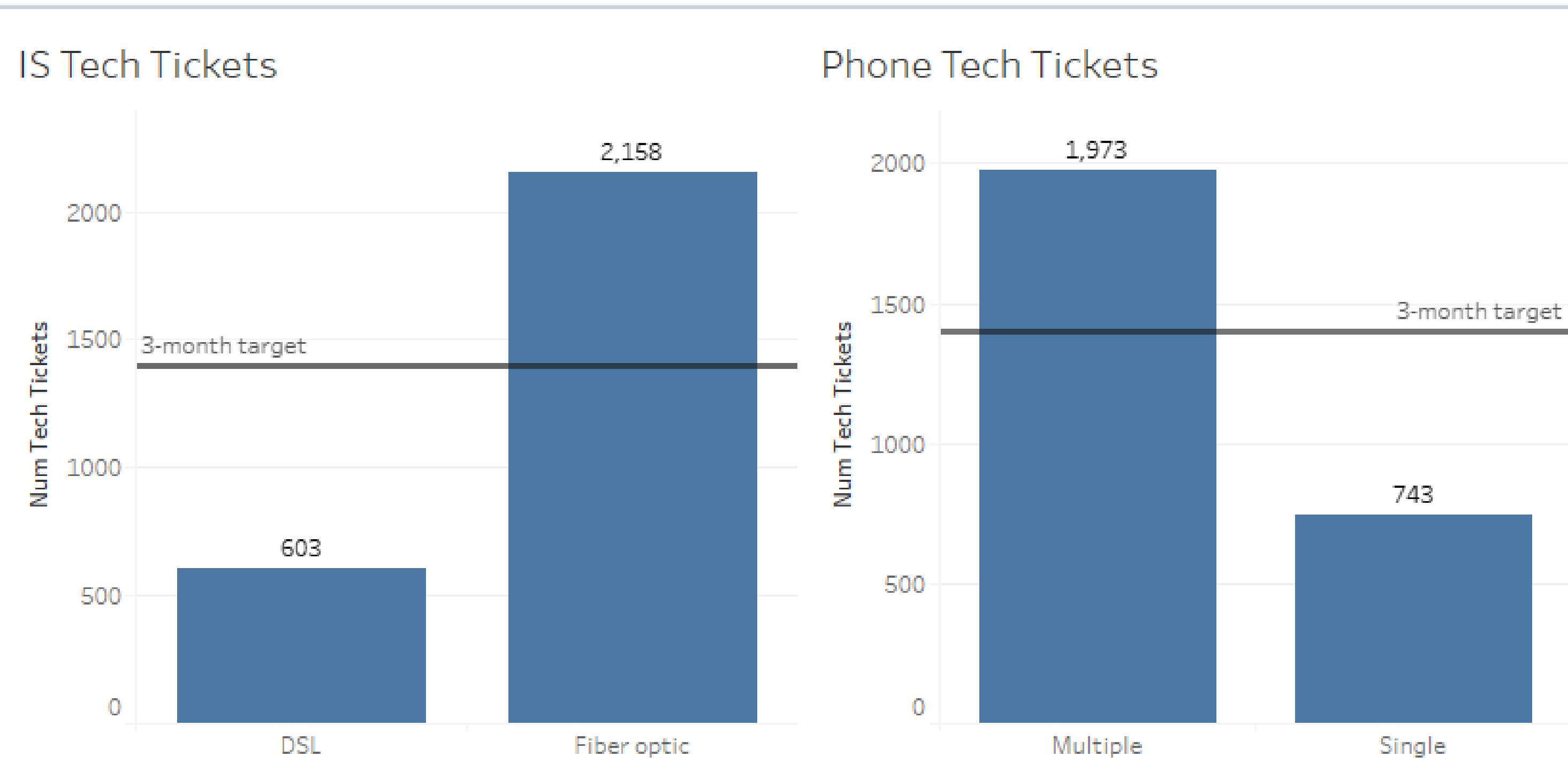


Fiber Optice



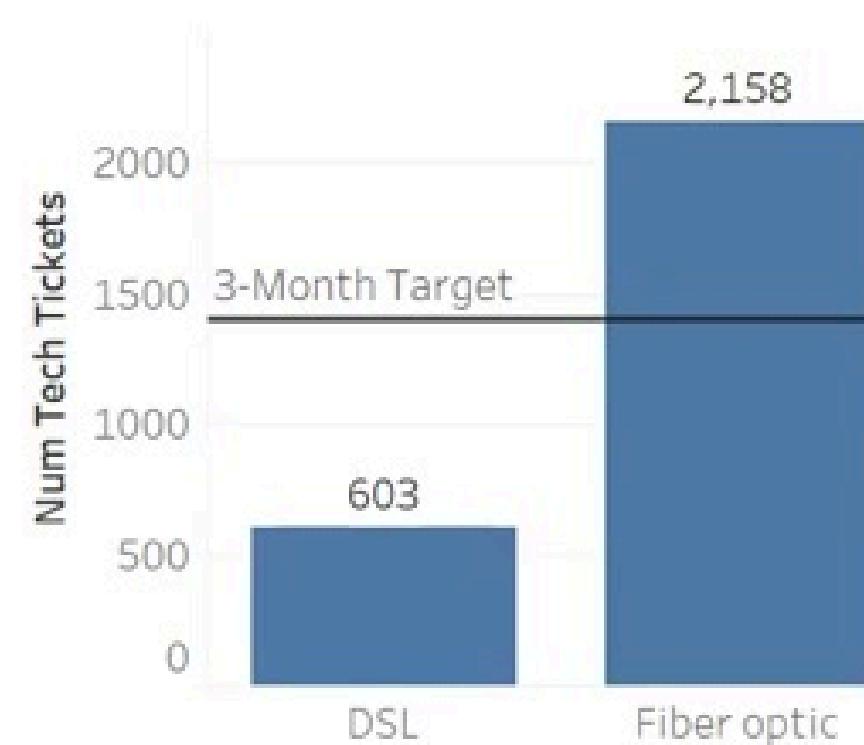
No Int Service

Volume as a target

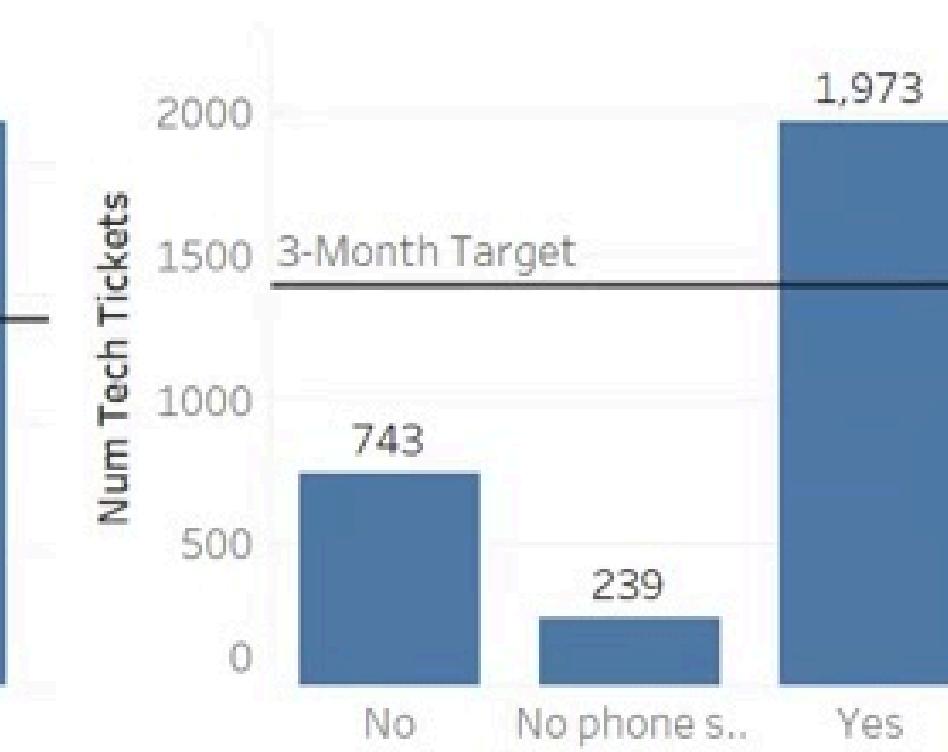


Dashboard

Internet Service Tickets



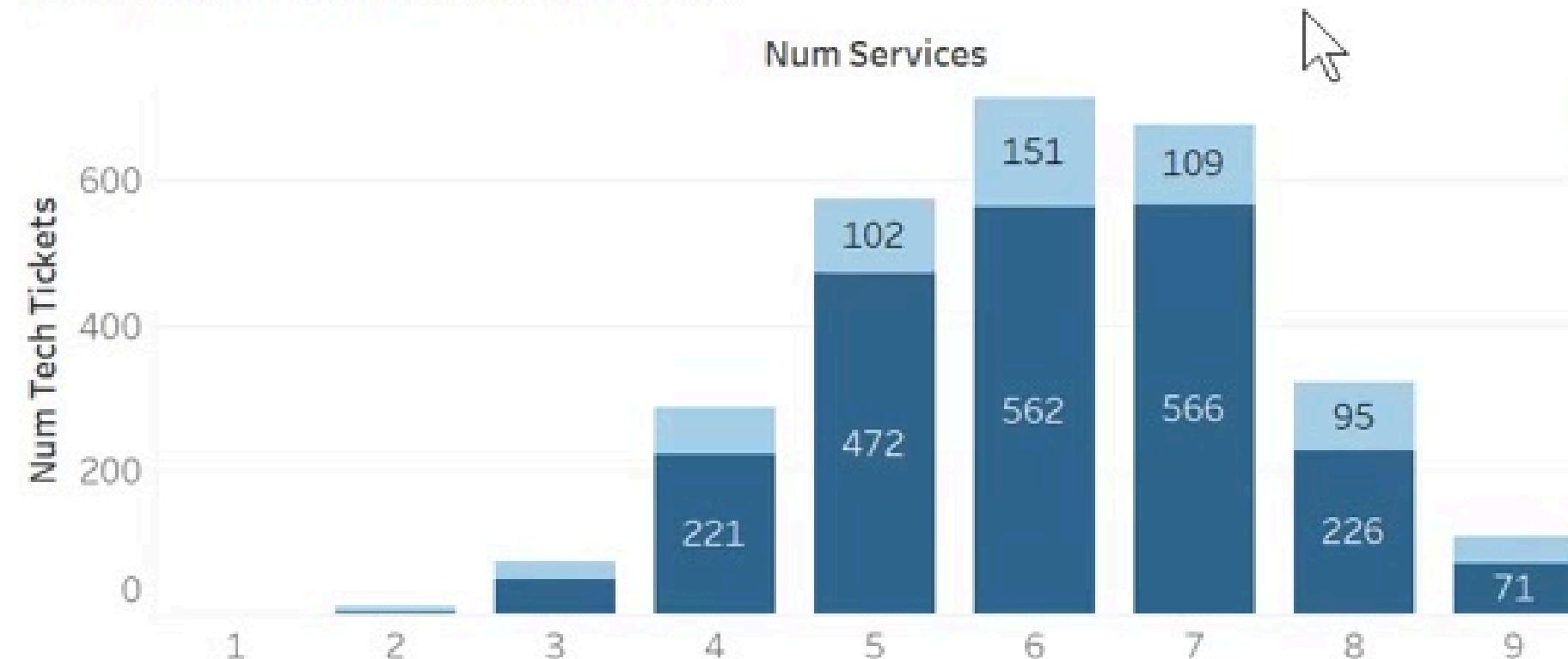
Phone Service Tickets



Churn Rate

26.54%

Technical Tickets & Service



Internet Service

(Multiple values) ▾

Multiple Lines

(All) ▾

Churn

No
Yes

A/B Test

**what to
A/B Test?**

**Technical
Support
Tickets**

**Consumer
Churn**

**Monthly
Revenue**

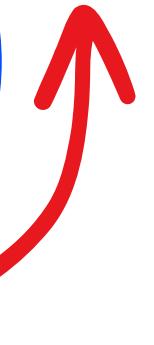
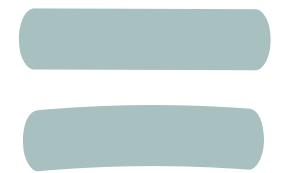
```
graph LR; A((Technical Support Tickets)) --> B((Consumer Churn)); B --> C((Monthly Revenue)); C --> A
```

**Technical
Support
Tickets**

**Consumer
Churn**

**Monthly
Revenue**

**Technical
Support
Tickets**



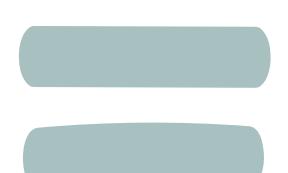
**Consumer
Churn**



**Monthly
Revenue**



**Technical
Support
Tickets**

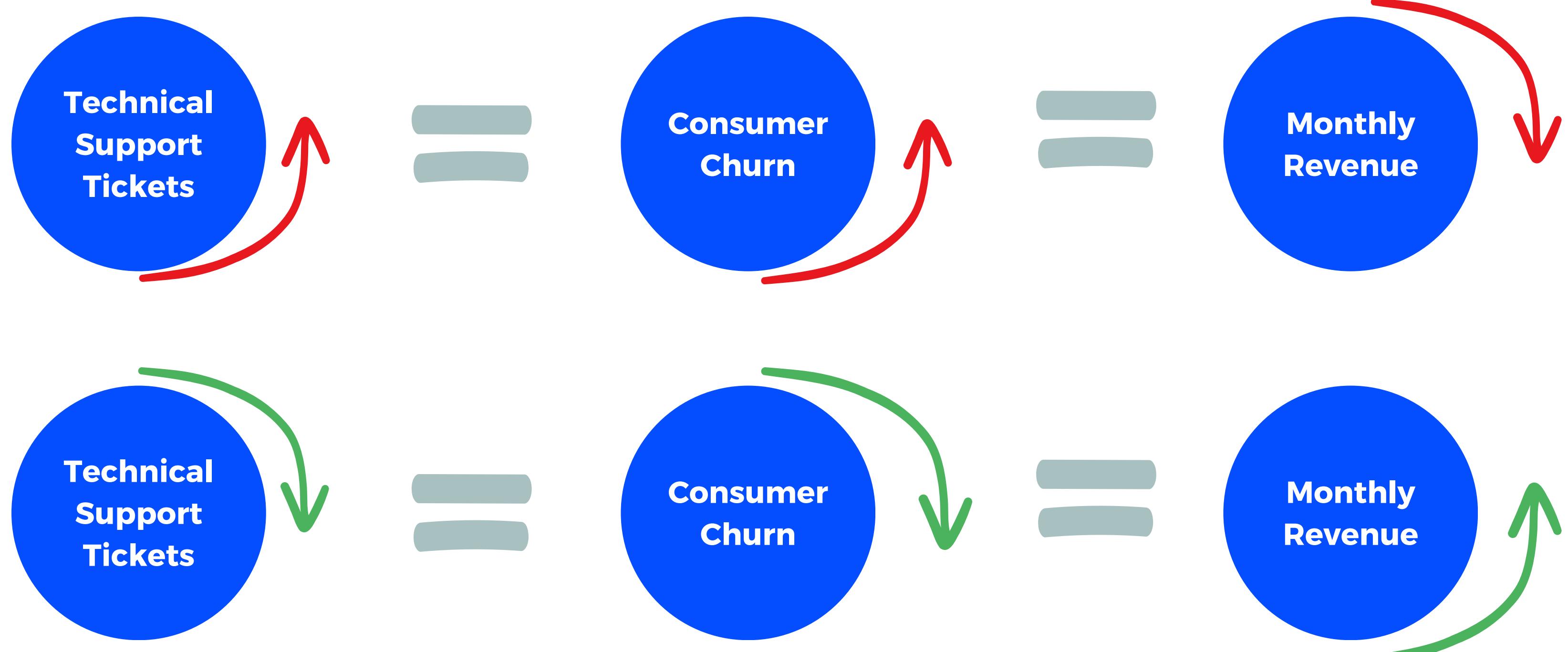


**Consumer
Churn**



**Monthly
Revenue**





Hypothesis: Reducing technical support **tickets** will lead to **fewer** customer **churn**.

Reduce technical friction
through **technical** development



Internet



Mobile



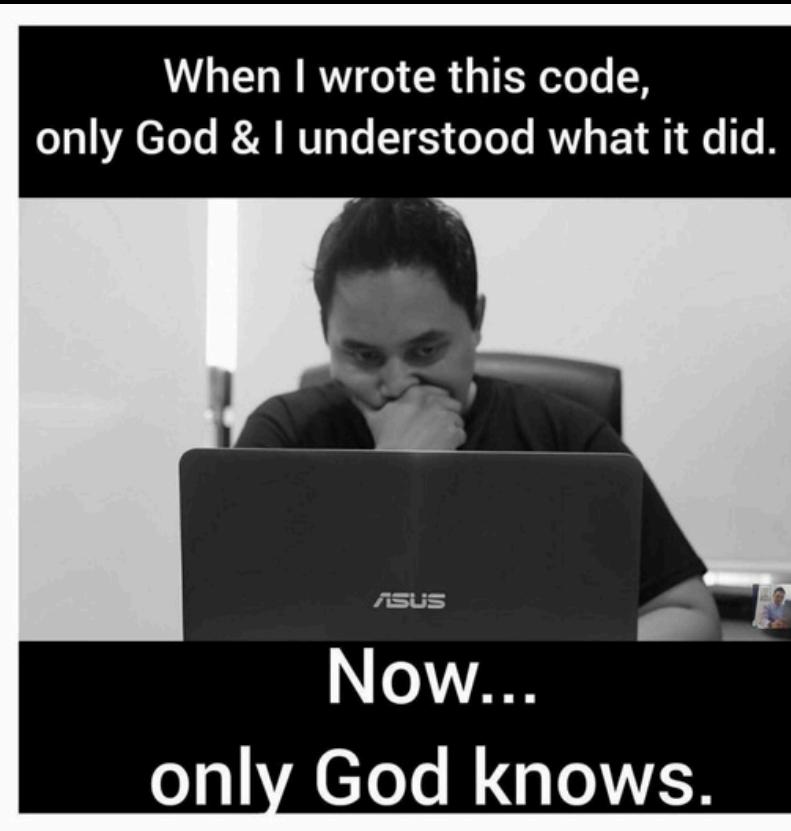
Streaming

Reduce technical friction
through **technical** development

Capital Intensive

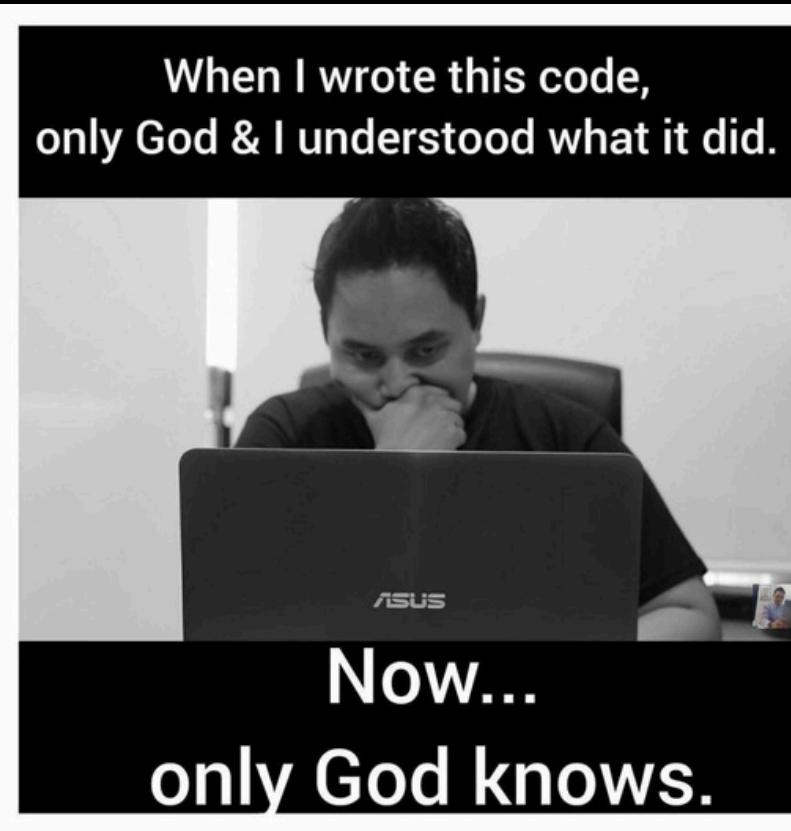


Reduce technical friction through **technical** development



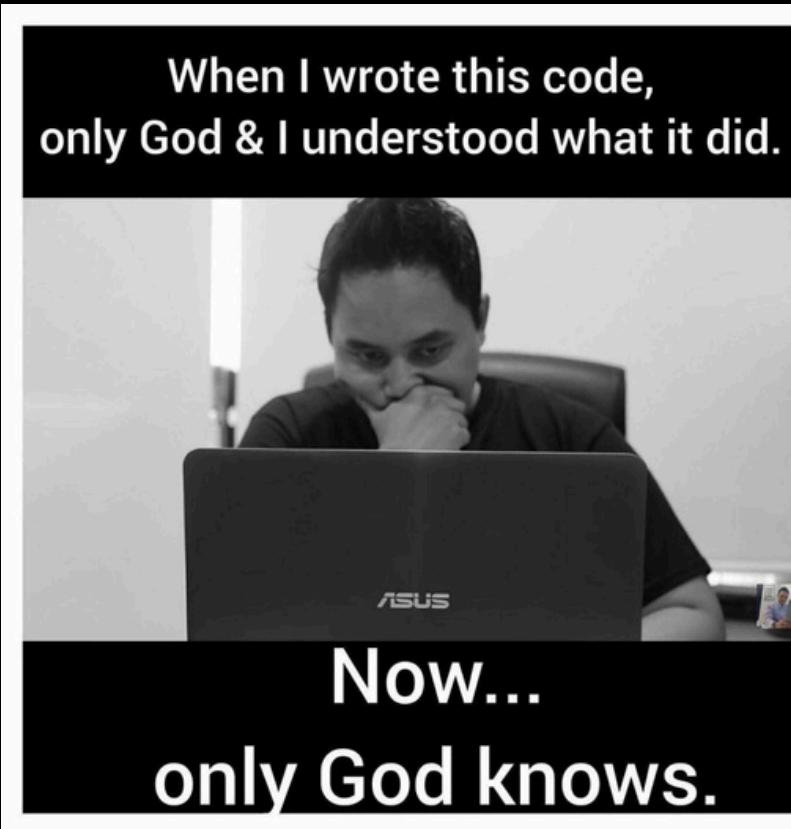
Reduce technical friction through **non-technical** development

Reduce technical friction through **technical** development



Reduce technical friction through **non-technical** development

Reduce technical friction
through **technical** development

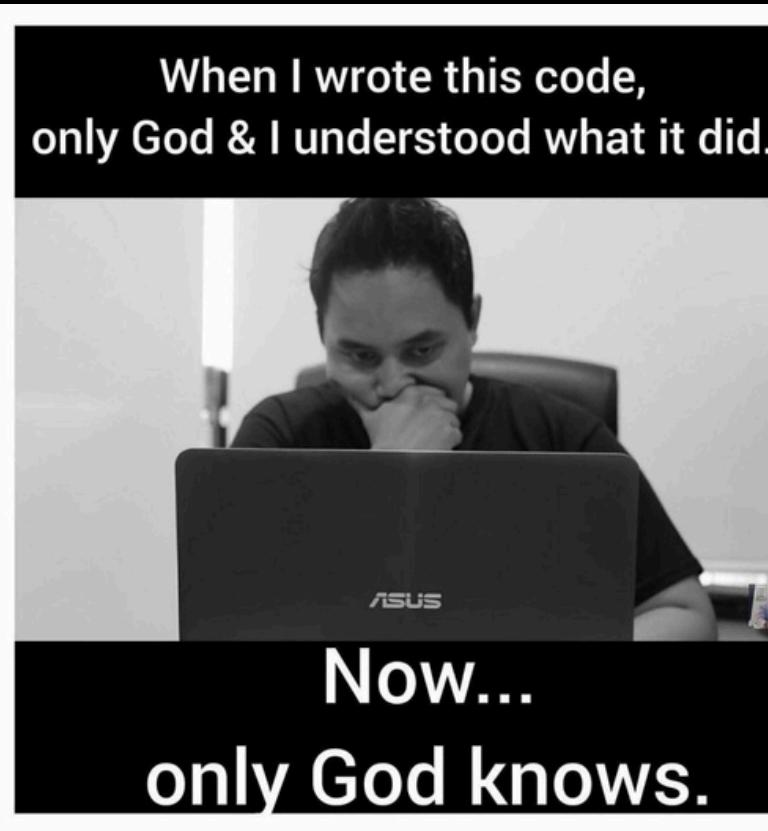


Reduce technical friction through
non-technical development

What?

Educate users

Reduce technical friction
through **technical** development



Reduce technical friction through
non-technical development

What?

Educate users



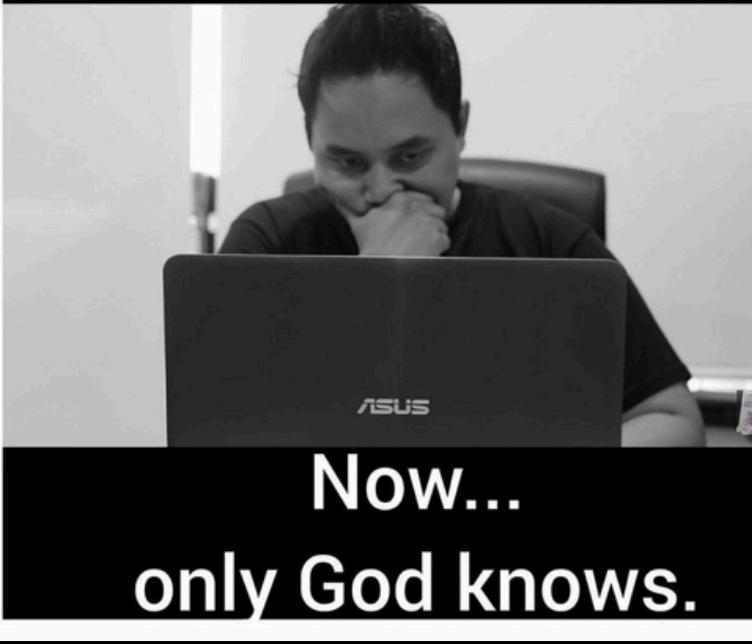
How?

Educational video

Reduce technical friction
through **technical** development



When I wrote this code,
only God & I understood what it did.



Reduce technical friction through
non-technical development

What?

Educate users

How?

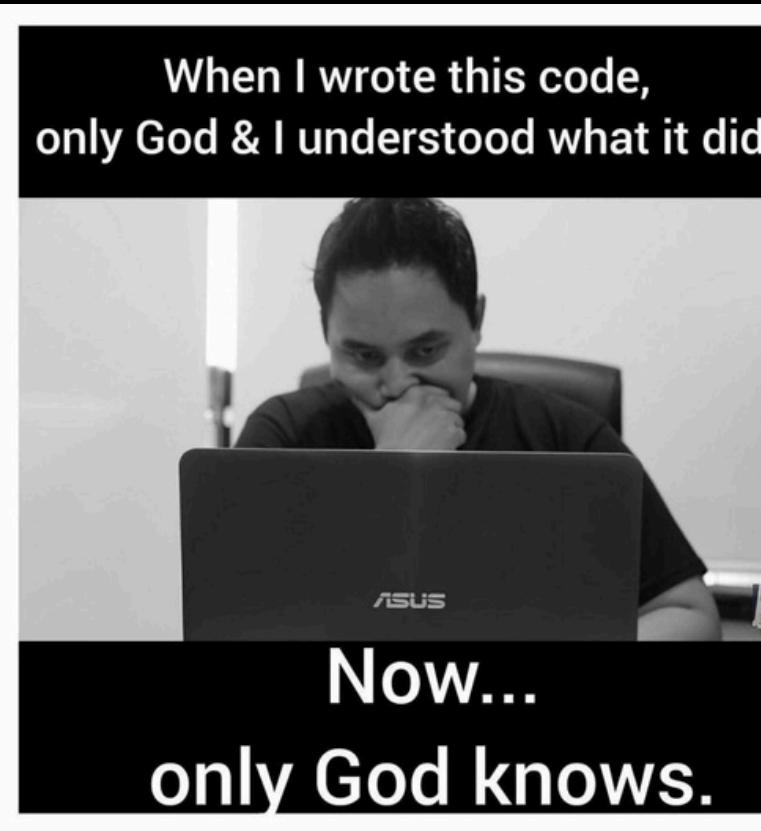
Educational video

Why?

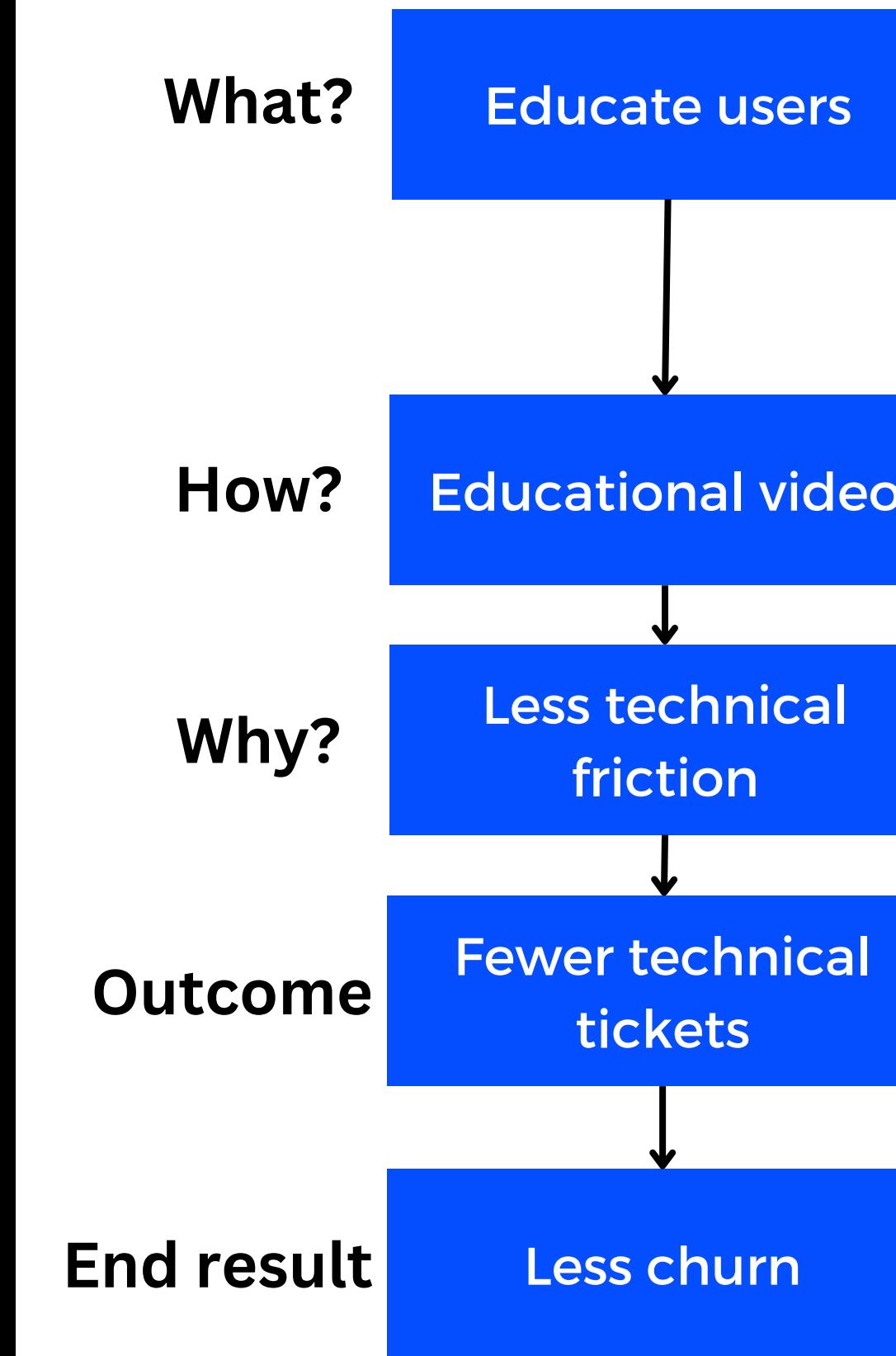
Less technical
friction

Engineering Power
Capital Intensive

Reduce technical friction
through **technical** development

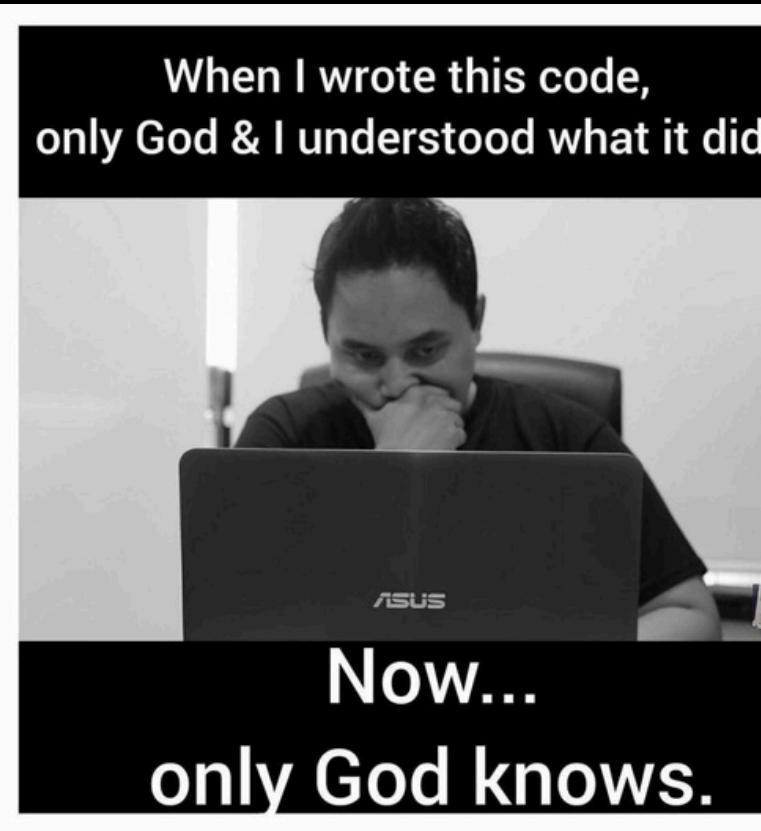


Reduce technical friction through
non-technical development



Engineering Power

Reduce technical friction
through **technical** development



Reduce technical friction through
non-technical development

What?

Educate users

How?

Educational video

Why?

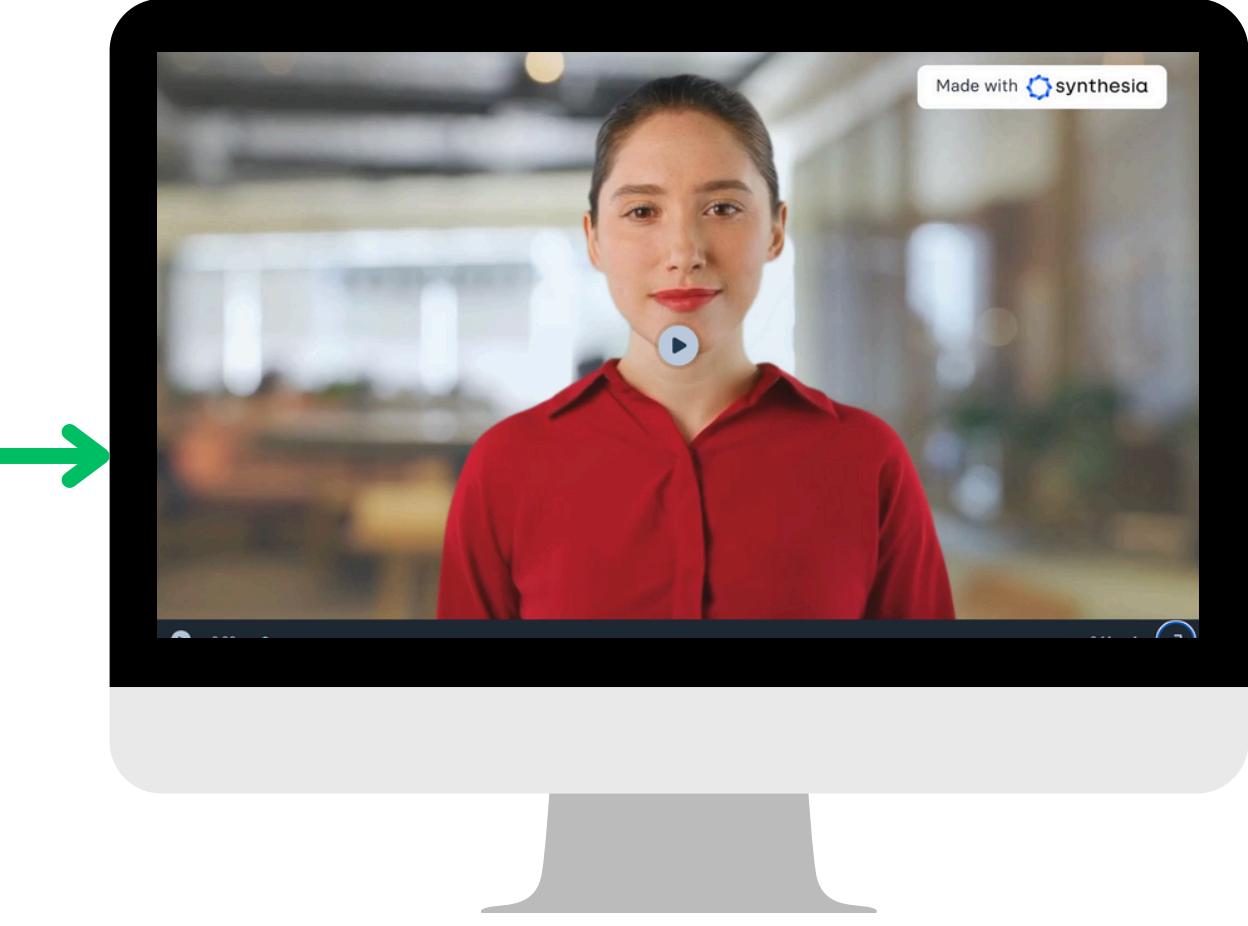
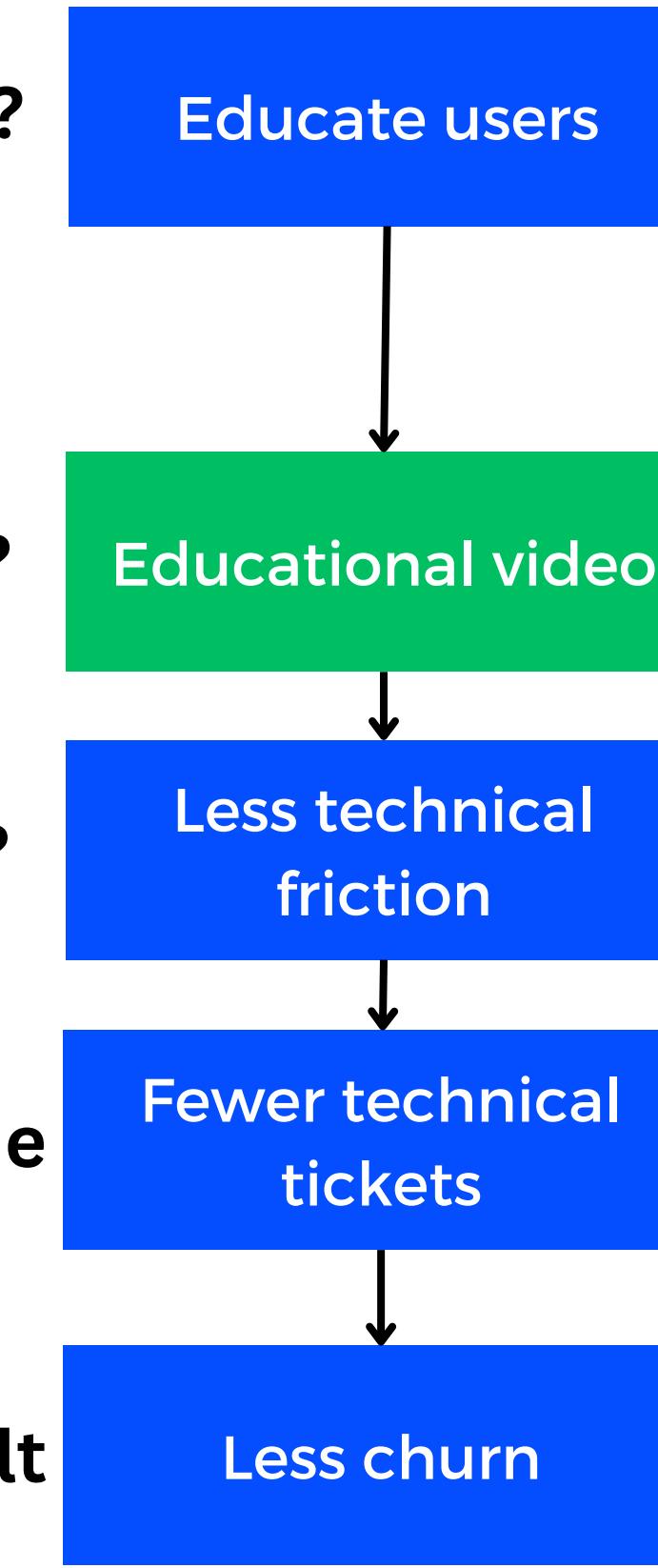
Less technical
friction

Outcome

Fewer technical
tickets

End result

Less churn

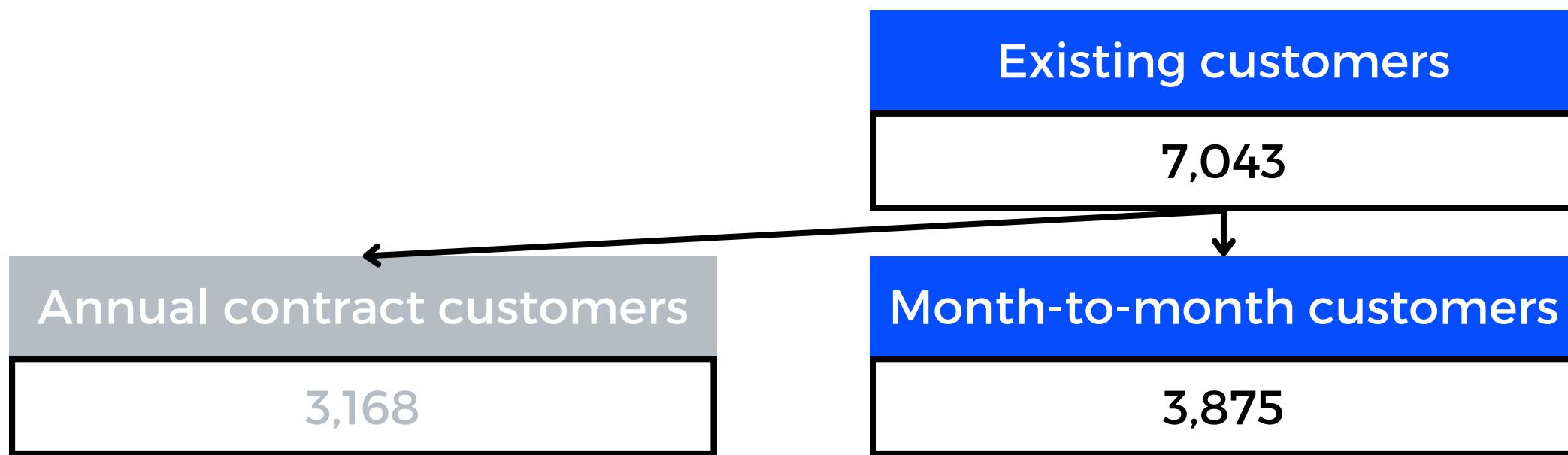


Who to A/B Test on?

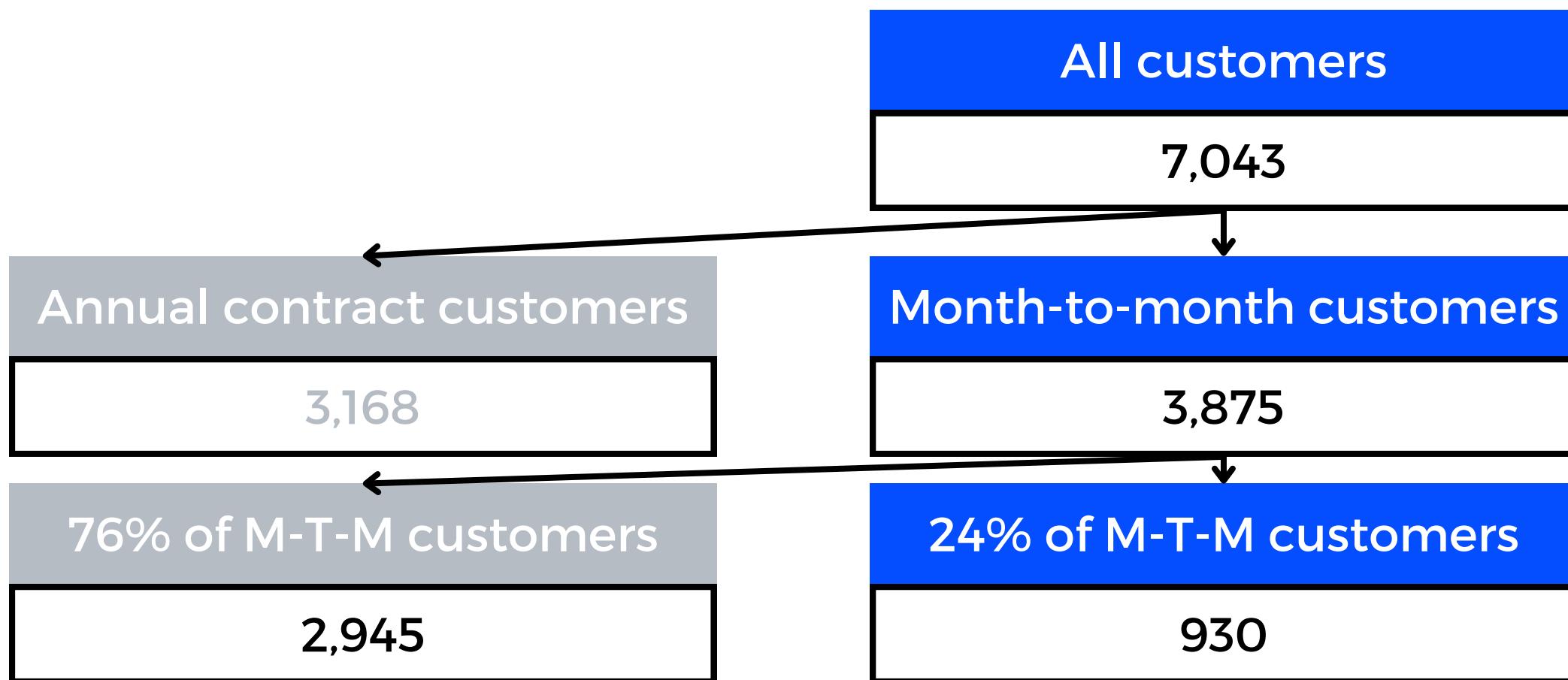
All customers

7,043

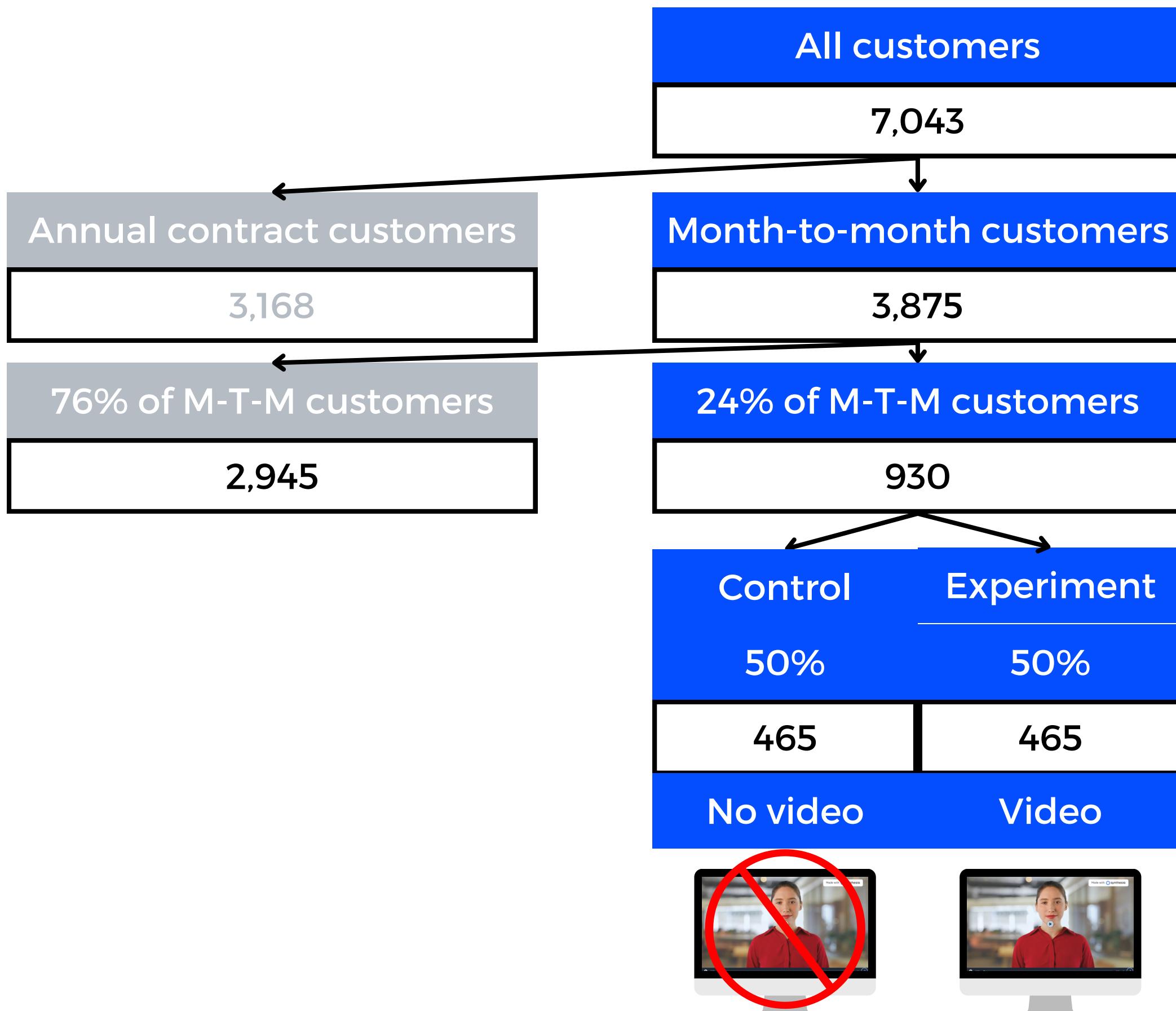
Who to A/B Test on?



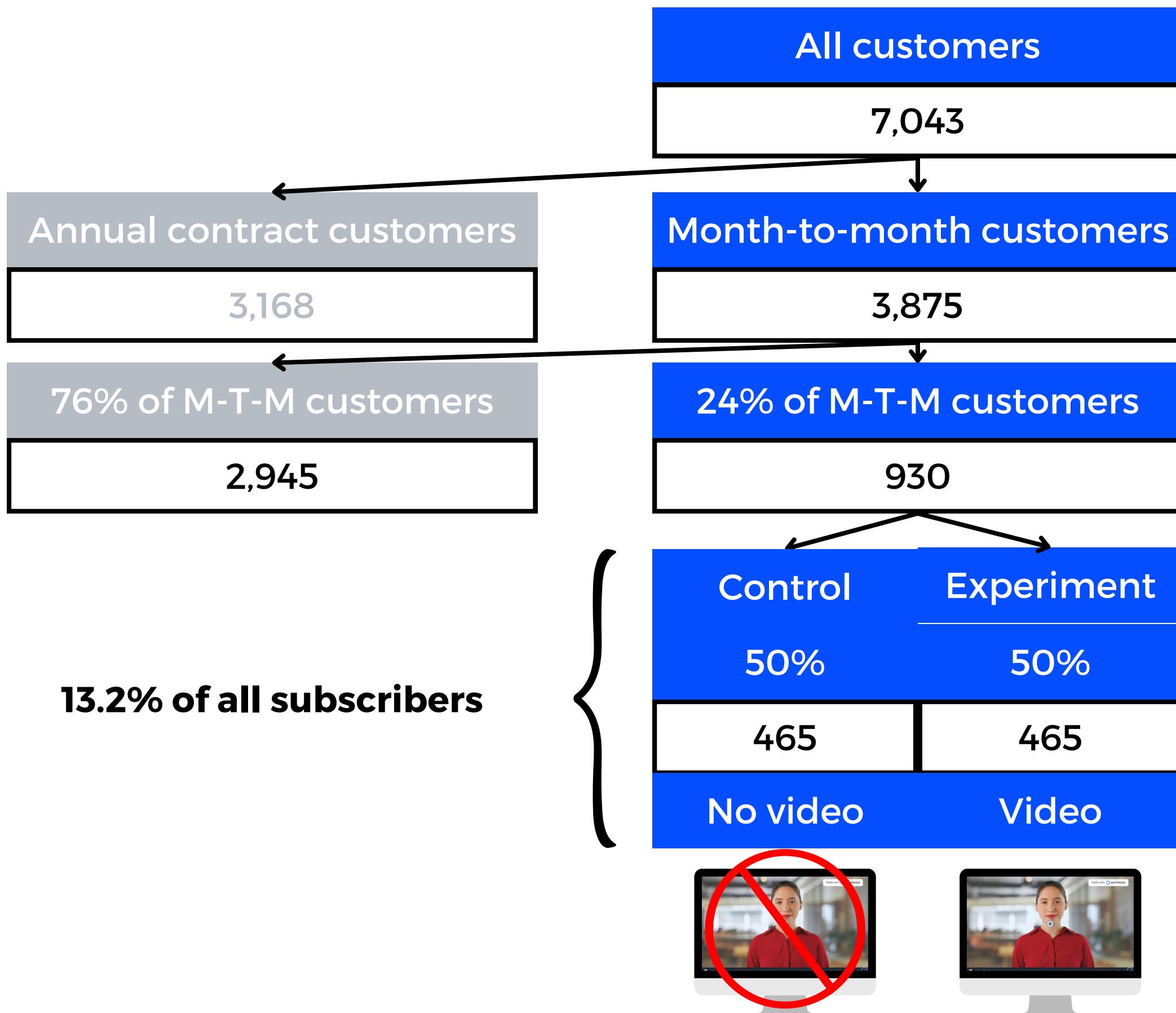
Who to A/B Test on?



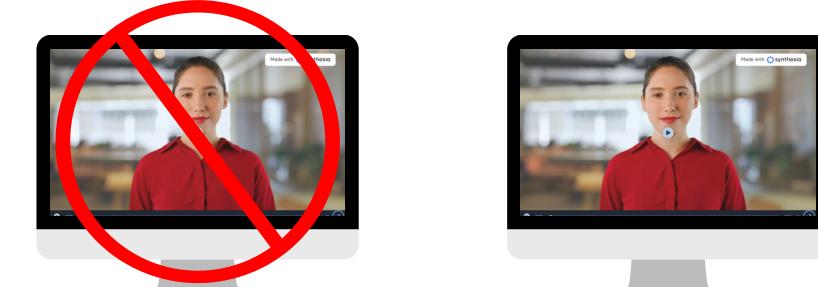
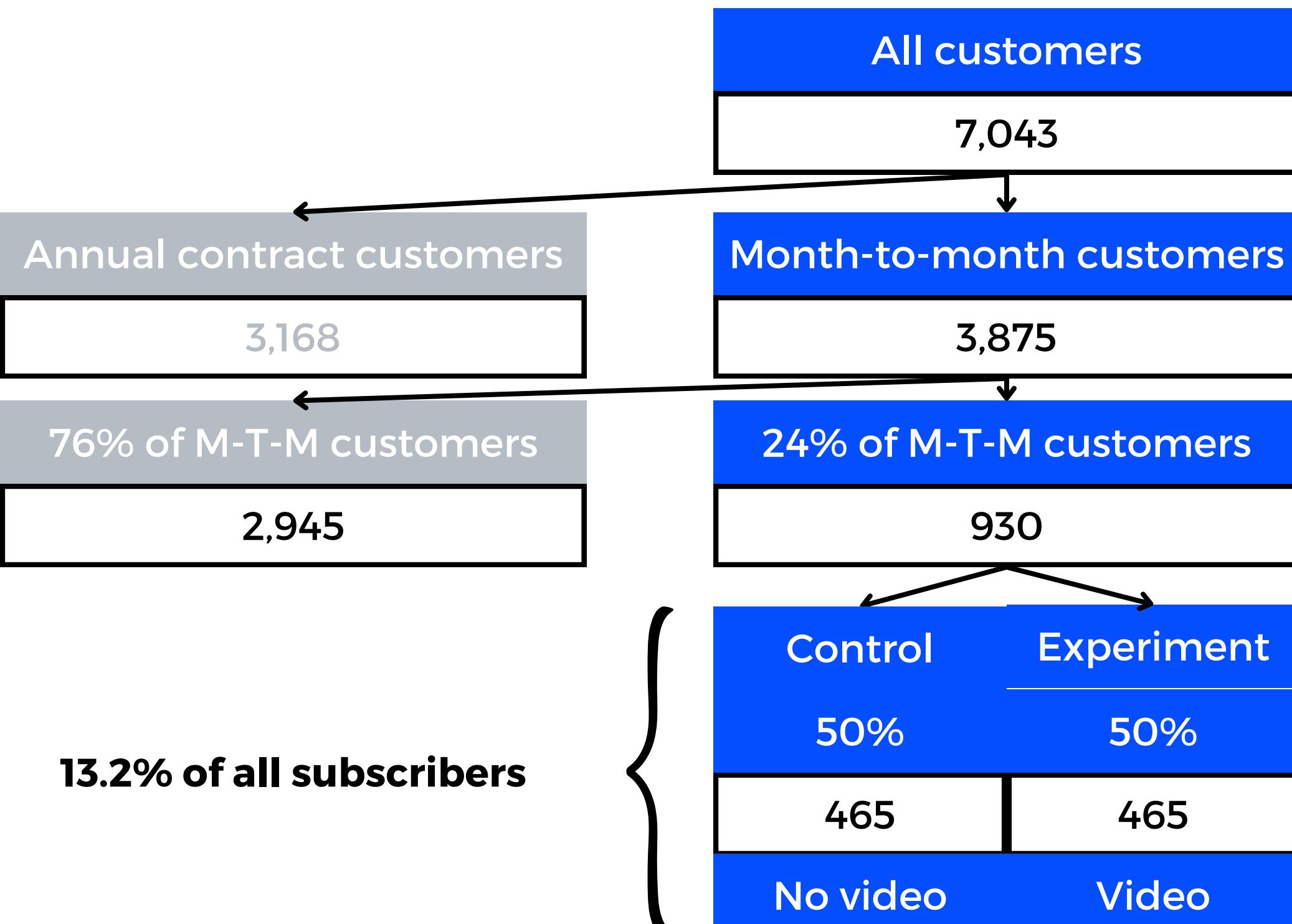
Who to A/B Test on?



Who to A/B Test on?

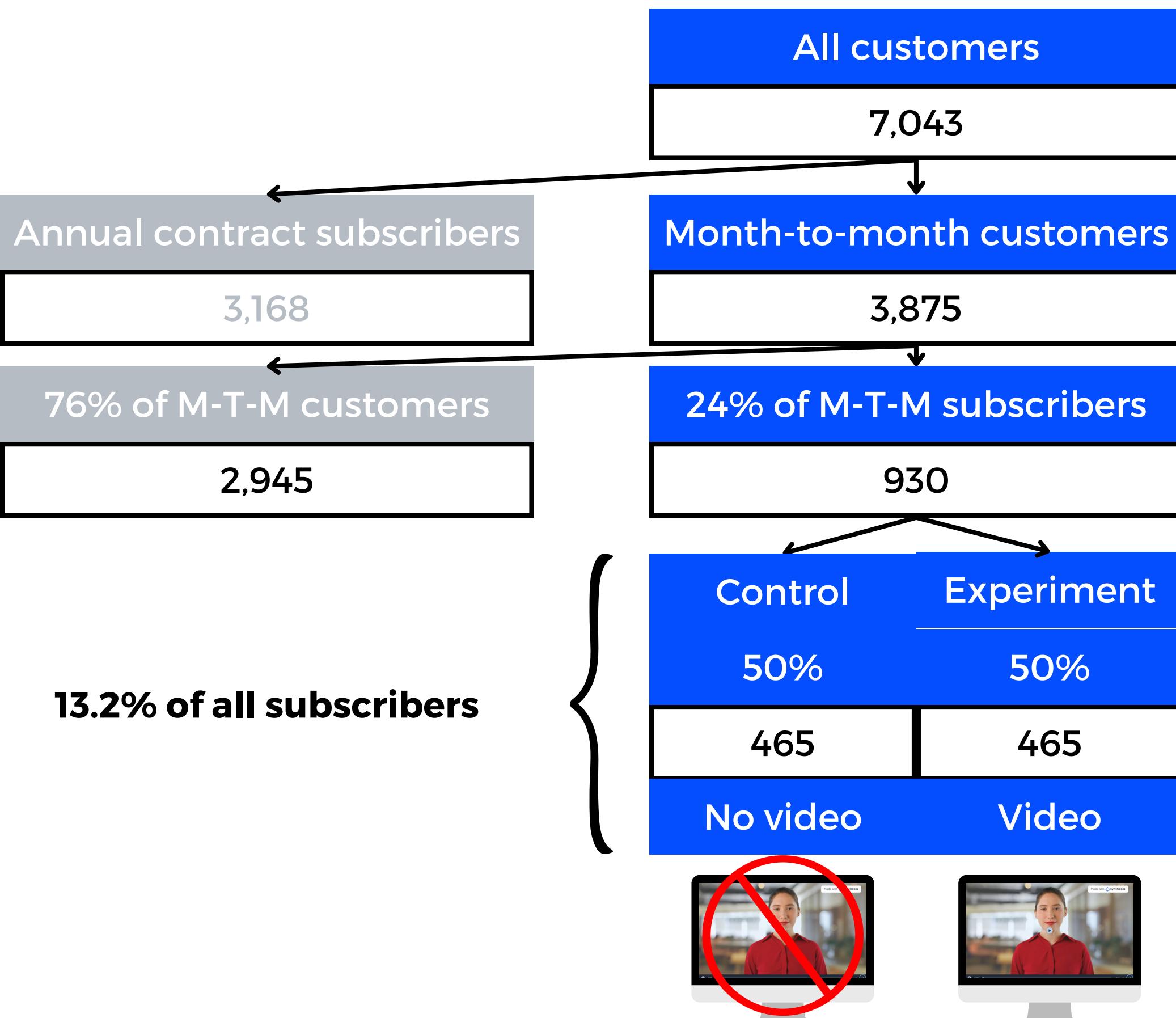


Who to A/B Test on?



H0	HA	
Subscribers that viewed the video will not exhibit a significant difference in monthly churn rate compared to subscribers that did not view the video.	Subscribers that viewed the video will exhibit a significant difference in monthly churn rate compared to subscribers that did not view the video.	
Characteristics	Control	Experimental
Month-to-month contract	Yes	Yes
Existing customer	Yes	Yes
Received the video	No	Yes

Who to A/B Test on?



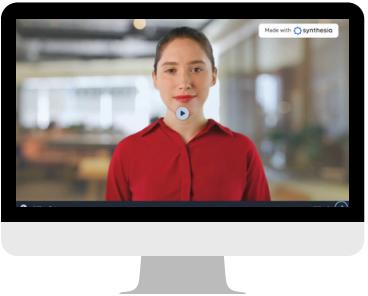
H0	HA	
Subscribers that viewed the video will not exhibit a significant difference in monthly churn rate compared to subscribers that did not view the video.	Subscribers that viewed the video will exhibit a significant difference in monthly churn rate compared to subscribers that did not view the video.	
Characteristics	Control	Experimental
Month-to-month contract	Yes	Yes
Existing customer	Yes	Yes
Received the video	No	Yes

Duration: 2 months

Implement A/B Testing

1

Hire ad agency to produce video



**4 minutes video
\$28,000**

6 weeks

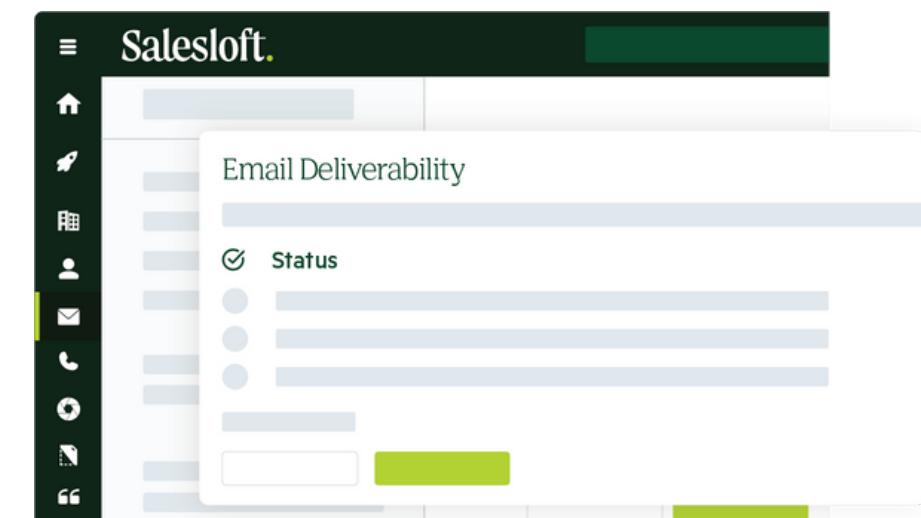
2

Identify control and experimental groups in CRM and backend

1 week

3

Set up an email sequence



2 weeks

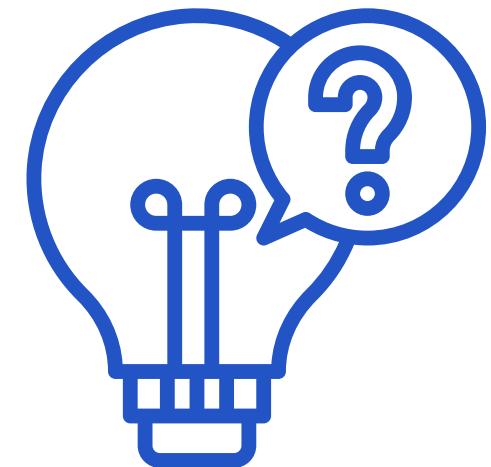
Experiment starts

Data Manifesto

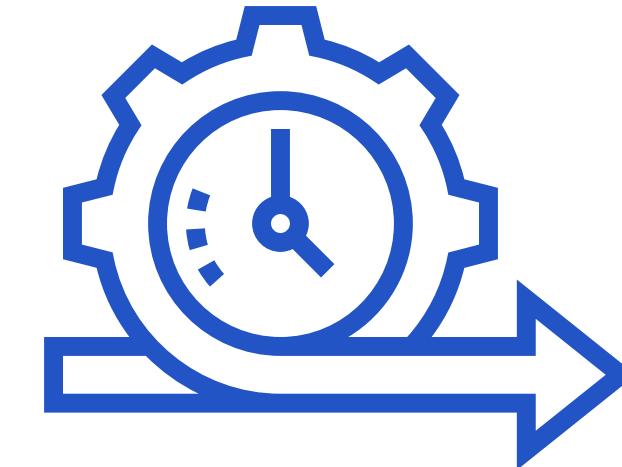
--- Purpose ---

“This manifesto represents our dedication to **infusing data throughout our operations**, recognizing its essential role in not only responding to current demands but also proactively shaping strategies and customer experiences for future trends.”

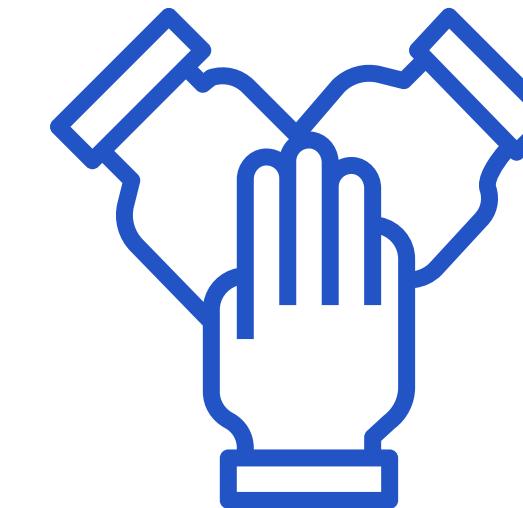
Values



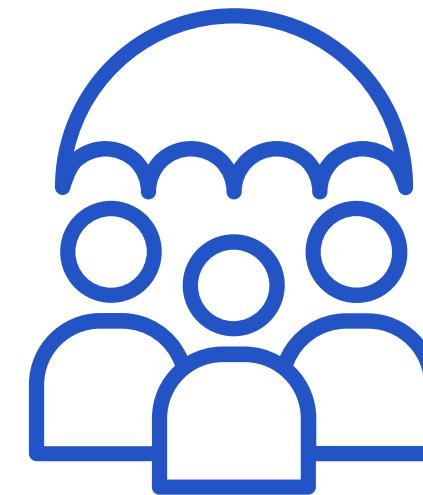
Curiosity



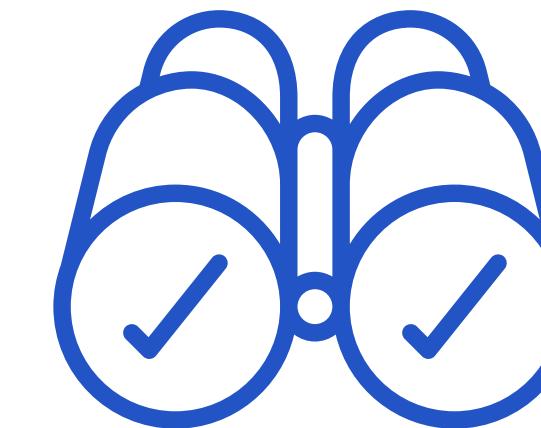
Agility



Collaboration



Inclusivity



Foresight

Core Principles

- **Data Literacy:** Foster an environment where every team member is fluent in data, ensuring informed decisions at all levels.
- **Analytics Maturity:** Progress through the analytics maturity scale, enhancing our predictive capabilities and strategic foresight.
- **Data-Informed Decision Making:** Ground every decision in data to stay aligned with our goals and customer needs.
- **Competitive Advantage:** Use data to identify and leverage our strengths in the market, staying ahead of industry curves.
- **Encouraging Continuous Discovery:** We will constantly delve into our data, unearthing opportunities for improvement, innovation, & customer engagement.