CUSTOMER LIFECYCLE MANAGEMENT

PREDICTING CUSTOMER CHURN

for beverage machines

- machine learning solution

MAKING MORE POSSIBLE

INTRODUCTION

- The business
- The problem
- Importance of existing customers
- Beverage machine data
- Relationship amongst categorical data
- Choosing estimator
- Our machine learning solution
- Process in place
- How to diagnose
- First results
- Understanding predictions
- Conclusions

THE BUSINESS

- The main business is a full service for beverage machine including :
  - Beverage machines deployed at a customer (Free on Loan, Rental,…)
  - Beverage ingredients delivered (coffee beans, soluble coffee, powder milk and chocolate, ...)
  - Management of preventative maintenance and repairs

- The customers point of consumption (POC) is referring to the location where the machine is installed

- A 'churned' POC is a location where a machine has been definitively removed

THE PROBLEM

We have a high churn rate of the beverage machine rented/loaned in our business and the goal is to reduce the churn rate by predicting which customer is more likely to churn and try to retain these customers

A churned machine generates a one-time cost for removal and replacement and a variable cost for depreciation and storage whilst a new location is found

By giving the machines with the highest churn likelihood to Service and Sales managers, they can act on it

This will help to retain more customer's installation points and increase the company's deployed beverage machine park

20%

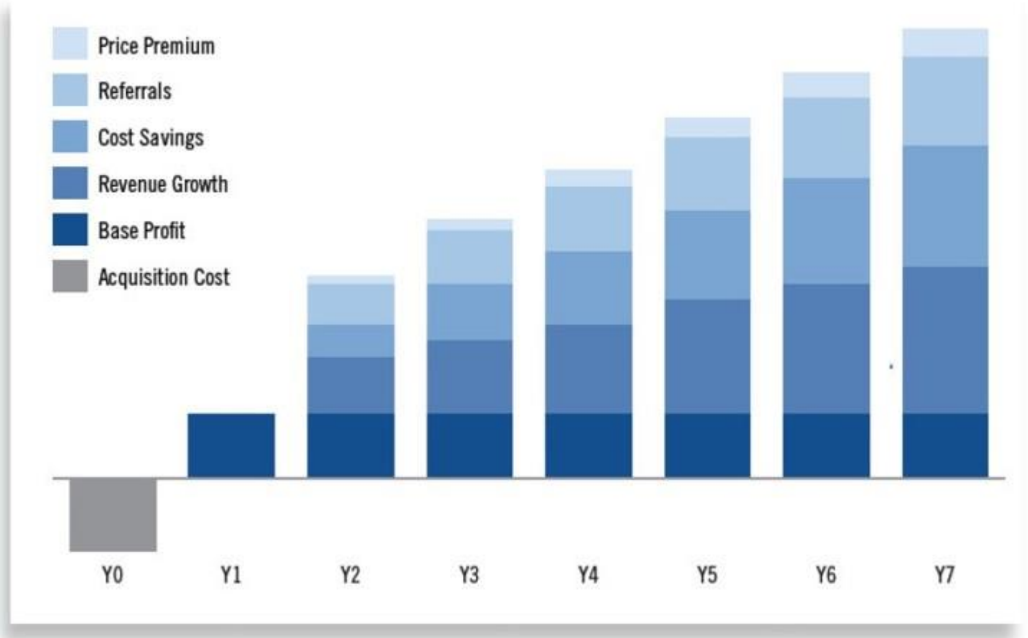
Average yearly churn

IMPORTANCE OF EXISTING CUSTOMERS

- When it comes to existing customers - they are our best ambassadors

- Loyal customers are the way to grow our business further :
    - Acquiring a new customer can cost 5 to 15 times more than retaining an existing customer
    - ROI : Increasing customer retention by 5% can increase profits from 25-95%
    - Loyalty : The success rate of selling to a customer you already have is 60-70%, while the success rate of selling to a new customer is 5-20%. Loyal customers spend more

Figure 1: Source of Profit from a Loyal Customer over Time



Adapted from F. Reichheld, *"The Loyalty Effect"*

Source : https://www.forbes.com/sites/jiawertz/2018/09/12/dont-spend-5-times-more-attracting-new-customers-nurture-the-existing-ones/?sh=7fbb5b675a8e

BEVERAGE MACHINE DATA

We keep monthly snapshot of the Beverage Machine Park and I saw two ways to manage the data:

- Proposal 1 : to keep all the data and not aggregate the monthly data of the machines
- Proposal 2 : to aggregate the data of the same Installation Point

Proposal 1

Inst. #	Month of snapshot	ID	Churn	Age in Month
Inst. 1	Jan	1	No	20
Inst. 2	Jan	2	No	48
Inst. 3	Jan	3	No	69
Inst. 4	Jan	4	Yes	45
Inst. 1	Feb	5	No	21
Inst. 2	Feb	6	No	49
Inst. 3	Feb	7	Yes	70
Inst. 5	Feb	8	No	25
Inst. 1	Mar	9	No	22
Inst. 2	Mar	10	No	50
Inst. 5	Mar	11	No	26
Inst. 6	Mar	12	No	30
Inst. 7	Mar	13	No	42
Inst. 8	Mar	14	No	7

Proposal 2

Inst. #	Latest month snapshot	ID	Churn	Age in Month	data available since (#month)
Inst. 1	Mar	1	No	22	3
Inst. 2	Mar	2	No	50	3
Inst. 3	Feb	3	Yes	70	1
Inst. 4	Jan	4	Yes	45	2
Inst. 5	Mar	5	No	26	2
Inst. 6	Mar	6	No	30	1
Inst. 7	Mar	7	No	42	1
Inst. 8	Mar	8	No	7	1

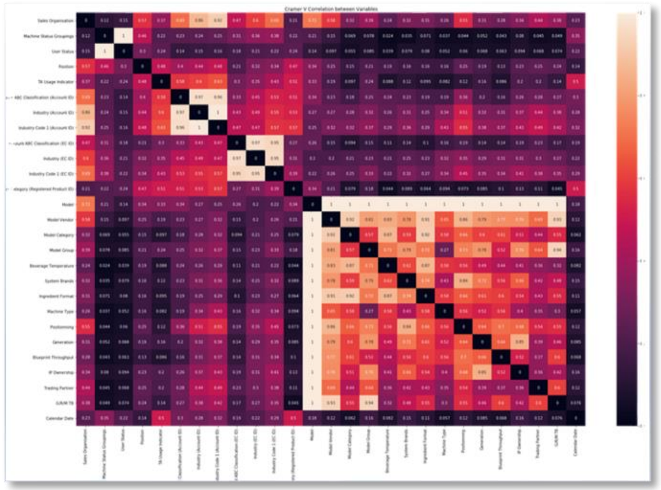
MISSING DATA

	Contract Installation Date	Contract Start Date	Contract End Date	Depreciation Start
data missing	89.0%	89.0%	89.0%	0.0%
min date	01.01.1980	01.01.1980	01.01.1980	01.01.1980
max date	snapshot date	snapshot date	in 10 years	snapshot date
replace missing and outliers by	median	median	mean	median
average	contract installation date was 871 days ago	contract started 869 days ago	contract ends in 910 days	depreciation started 1080 days ago

RELATIONSHIPS AMONGST CATEGORICAL FEATURES

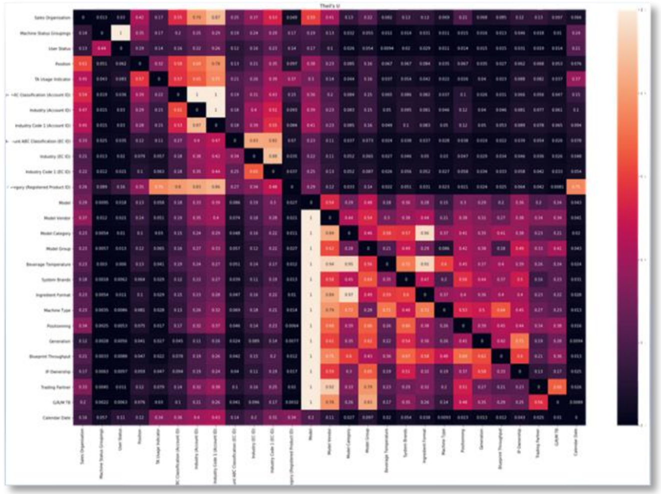
Cramér's V

- Cramér's V is a measure of association between two nominal variables, so we can get insights on the relationships amongst the features themselves
- Cramér’s V can be a heavily biased estimator of its population counterpart and will tend to overestimate the strength of association



Theil’s U

- Theil’s U measure, can determine the degree of association between two variables and by swapping the two variables, their values are different, so the heat map is not symmetrical
- e.g. we see a small association Positioning - Channel but we do not see it as much in the inverse



When features that are not highly dependent, they can add value to a model

FIVE ESTIMATORS

Logistic regression

makes predictions using a logistic function

KNeighbors

regression based on k-nearest neighbors

Decision Tree

learns a set of if-then-else rules that lead to a final decision

Random Forest

meta estimator that constructs a multitude of decision trees at training time

XGBoost

an ensemble of weak prediction models, typically decision trees

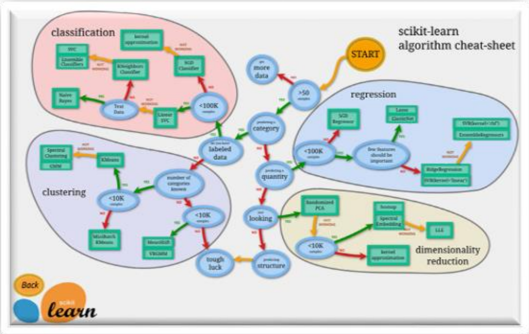
WHY USE THESE MODELS

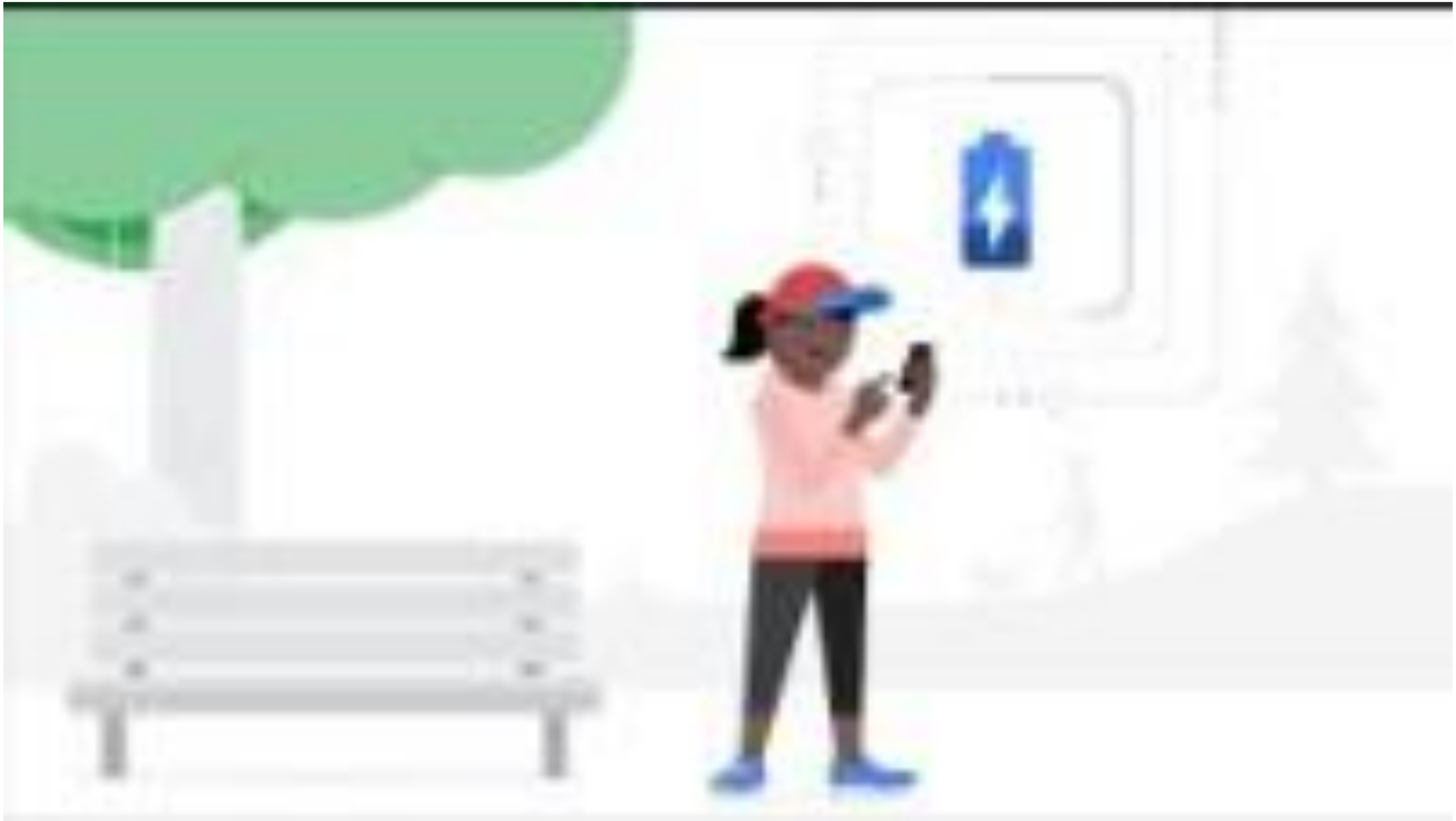
- I used the flowchart from Scikit Learn designed to help me to choose which estimators to try

- I preferred Logistic Regression to Linear SVM (Support Vector Machines):
  - generally SVM (or SVC) works well with unstructured and semi-structured data like text and images while logistic regression works with already identified independent variables

- Kneighbors Classifier was worth to be tried because it works differently
  - relies on distance for classification

- Random Forest and XGBoost are Ensemble classifier and why not try also decision tree which can help to understand the logic visually
  - random Forest is constructing a multitude of decision trees
  - it can handle categorical features very well and can work out of the box
  - XGBoost can perform even better but is more a black box





https://www.youtube.com/watch?v=qCdc3WPPGmM

OUR MACHINE LEARNING SOLUTION

Collect Data

- Trade assets data
- Telemetry data
- Commercial visits or calls
- Service data
- Others

Analyze

- Analyze data, clean and transform the data.

Algorithm

- Split data in two data sets (80% Train data, 20% Test data). Use Train data to feed several algorithms.

Evaluation

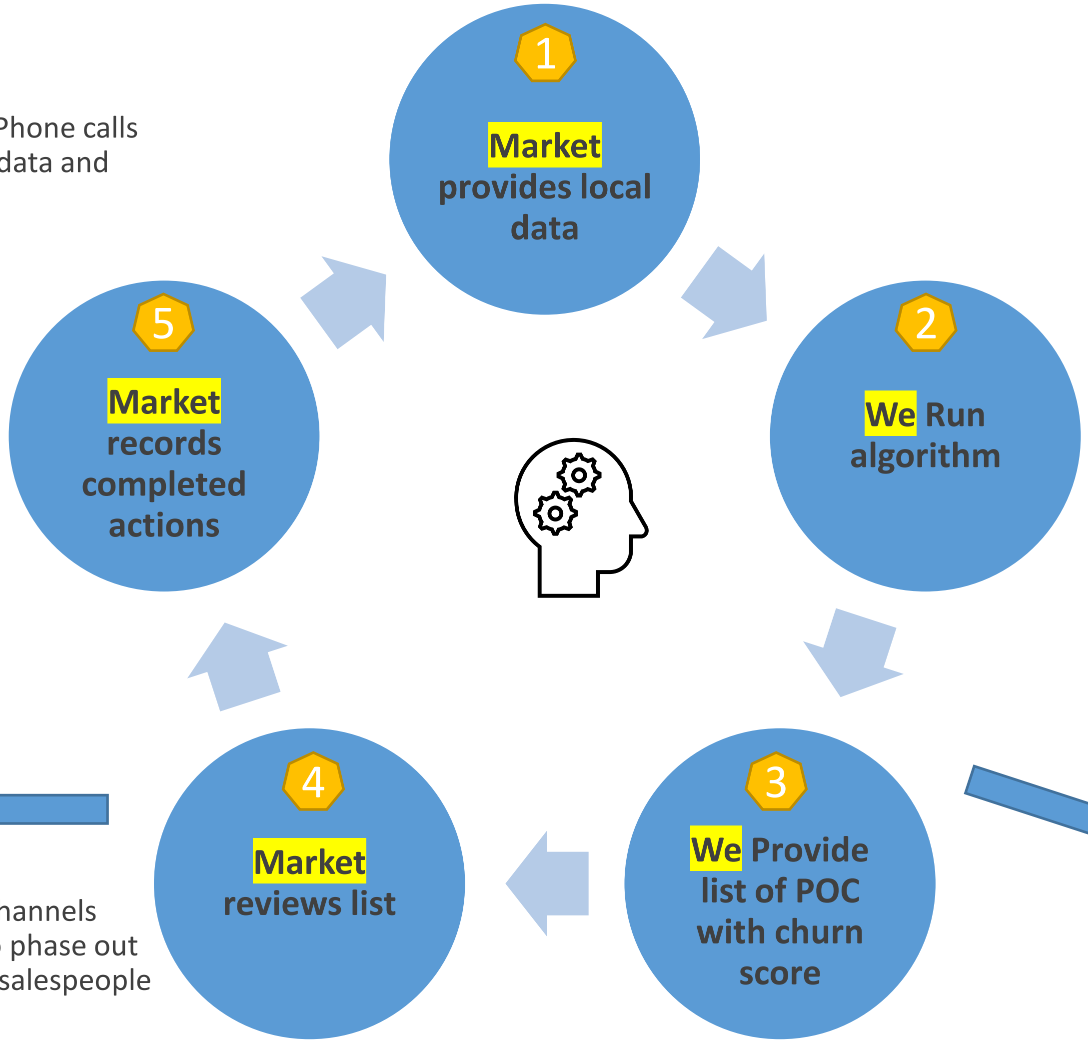
- Compare the predictions with actual results of Test data to evaluate the accuracy
- Benchmark from algorithms to identify the best algorithm

PROCESS

Improve with markets

- Corrected data relationship for Visits and Phone calls
- Markets understood the benefits of good data and started to correct their data

Three pilot Markets



Approach examples

- Focus on high volume locations
- Focus on top 20 overall
- Focus on top 20 in local market priority channels
- Focus on specific models that you plan to phase out
- Focus on specific sales territories and/or salespeople

Provided list

- Only half of the list is provided at the beginning for experiment purposes

HOW TO DIAGNOSE?

- Worked with colleagues from Markets and Regions to create a playbook for retention
- Checks examples:
  - How old is the machine? Is it an old model that should be changed?
  - When was the last commercial visit or phone call? Was it too long ago and customer engagement is needed?
  - Is this a good performing (throughput) location?
  - Does the channel have a higher-than-average churn rate?
  - Is the average sales decreasing?
  - How many incidents have occurred? When did the last incidents go back?
  - What kind of contract is it?
  - …

POC WITH HIGH RISK PREDICTIONS

Only half of the list is provided at the beginning for experiment purposes

Count of IP (high risk score shared with market)	**After 3 months (Dec 21 -Feb 22)**			
Sales Organisation	**not churned**	**churned**	**Total**	**% churned**
1	324	90	414	21,7%
2	39	5	44	11,4%
Grand Total	**363**	**95**	**458**	**20,7%**

Count of IP (high risk score hidden)	**After 3 months (Dec 21 -Feb 22)**			
Sales Organisation	**not churned**	**churned**	**Total**	**% churned**
1	298	108	406	26,6%
2	37	14	51	27,5%
Grand Total	**335**	**122**	**457**	**26,7%**

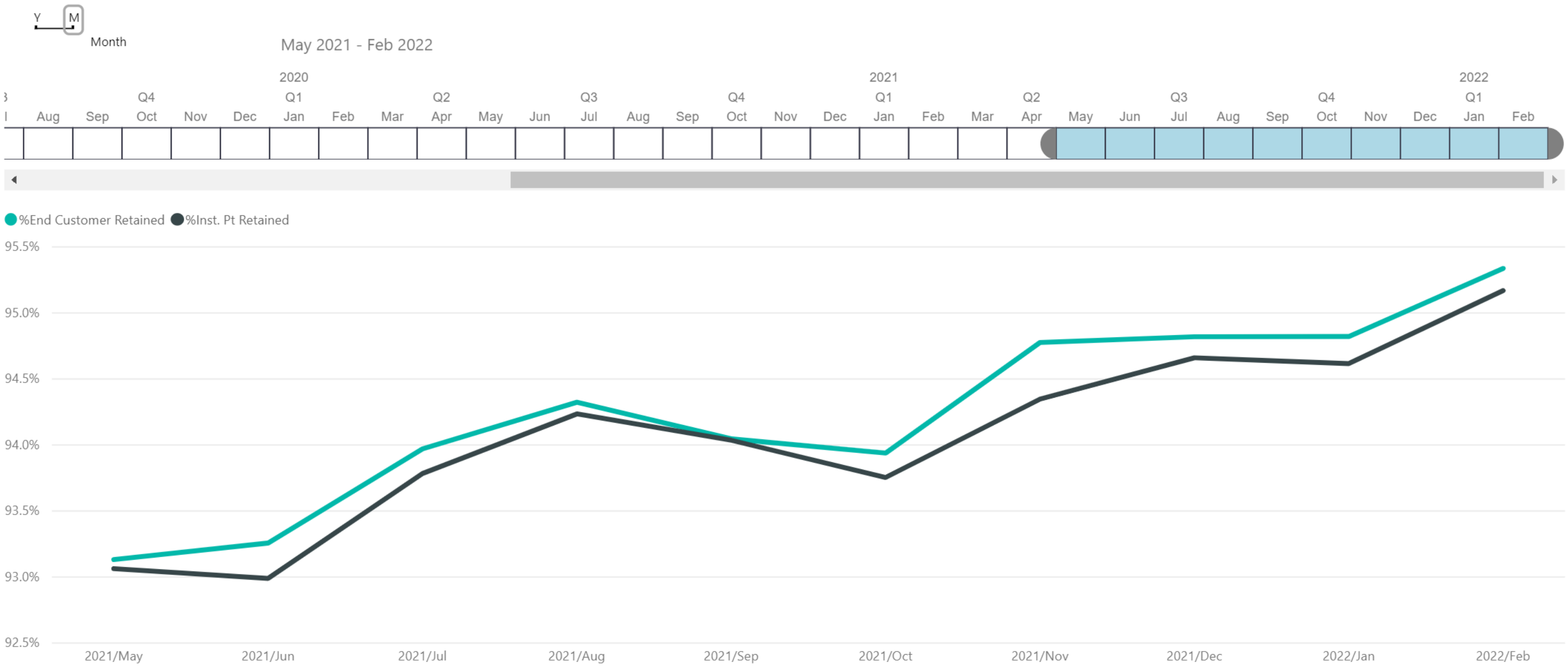
churn reduced by

6%

After three months, installation points with high risk score provided to markets had 6% less churn (26,7 to 20,7%) thanks to sales actions

PILOT MARKETS RETENTION RATE

We see an increase in the retention rate for pilot markets since we started the project



INTERPRETING PREDICTIONS

- A python library called treeinterpreter can help to interpret decision tree and random forest predictions

- We can also use SHAP library for interpreting black-box machine learning models using the Shapley values methodology

- SHAP and treeinterpreter library can decompose predictions

Enter the ID of a machine → Treeinterpreter library → Top N features contributing to the churn prediction

CONCLUSIONS

- Machine Learning can improve efficiency and reduce customer churn
- An important part of the project is also spending time explaining and convincing colleagues to use the predictions
- When you work with sales and service colleagues, you learn from their knowledge and ultimately improve the algorithm as well

LET’S RETAIN OUR CUSTOMER