Know Your Customer:
How to Eliminate Bots and Predict Customer Lifetime Value

ASSURANCE

MIT MANAGEMENT SLOAN SCHOOL

Project team
Semi Hasaj
Ryme Kabak

Assurance team
Megan Dixon
Killian Farrell | Alison Borenstein

Faculty advisor
Jordan Levine

About Assurance IQ

Assurance IQ is a direct-to-consumer platform that transforms the buying experience for individuals seeking personalized health and financial wellness solutions by giving them recommendations for personalized insurance plans tailored to their needs
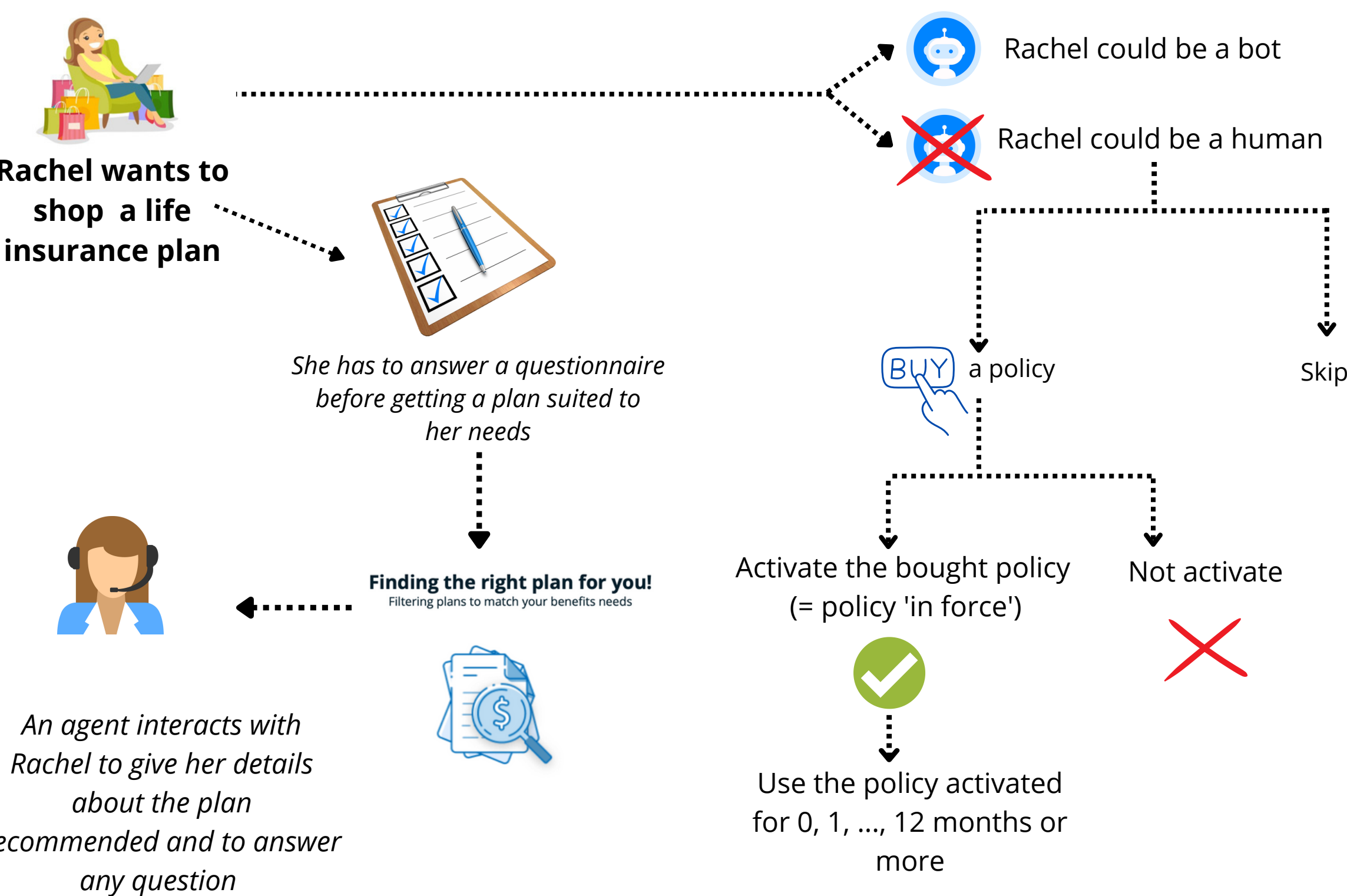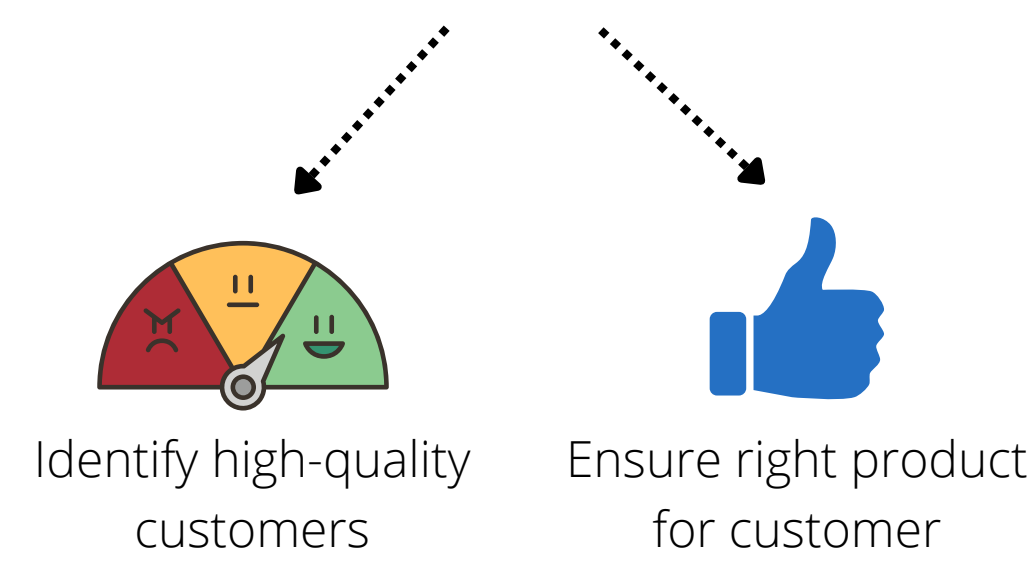
8 lines of business

Life Insurance | Health Insurance | Medicare | Money Center
Auto Insurance | Home Insurance | Home Loans | Personal Loans

500K visitors daily

150K contactable shoppers

How does it work?

Rachel wants to shop a life insurance plan

She has to answer a questionnaire before getting a plan suited to her needs

Finding the right plan for you! Filtering plans to match your benefits needs

An agent interacts with Rachel to give her details about the plan recommended and to answer any question

Rachel could be a bot

Rachel could be a human

BUY a policy | Skip

Activate the bought policy (= policy 'in force') | Not activate

Use the policy activated for 0, 1, …, 12 months or more

Problem statement

How can we better focus resources on the experience of the real customers

Identify high-quality customers | Ensure right product for customer

Secondary Goals:

COST — Minimize cost of fraud bot applications | Improve business decisions with customer LTV forecast
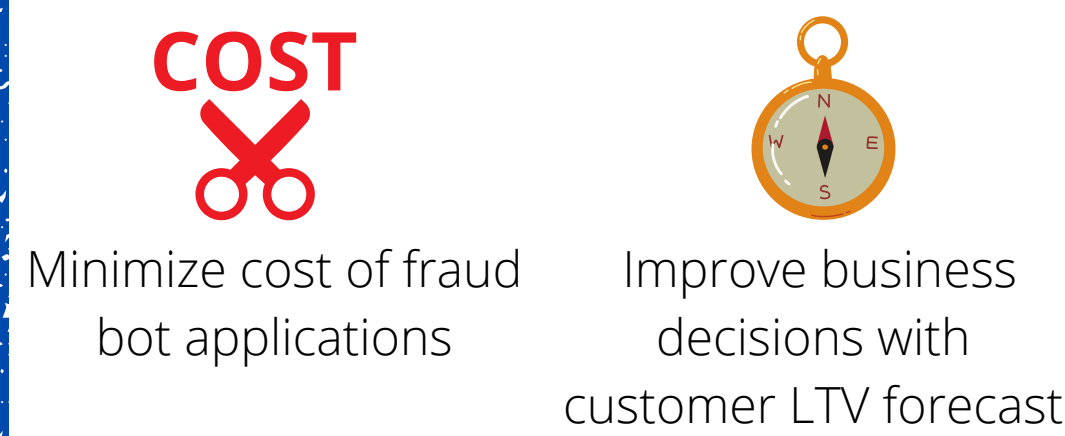
Dataset

Demographics: Location, Gender, marital status, BMI, medical conditions…

Website visitor data: IP address, questionnaire responses, device, browser…

Cost & Revenue: Marketing costs, agents and transfer costs, revenue from sales, revenue from ad clicks…

Calls summary: total inbound and outbound calls, duration of calls…

External data: Third party vendors data

A) Detect bots — bot or not

Goal: Build rule system to block bots without effecting true customers

Challenge: No Labels, Constantly changing bot behavior, trade-offs between various bot detection tools (**accuracy, cost, complexity**)

Strategy: Develop combinations of rules and analyze effectiveness using profit based analysis with/without rule

Results: Rule to block bots while improving profit

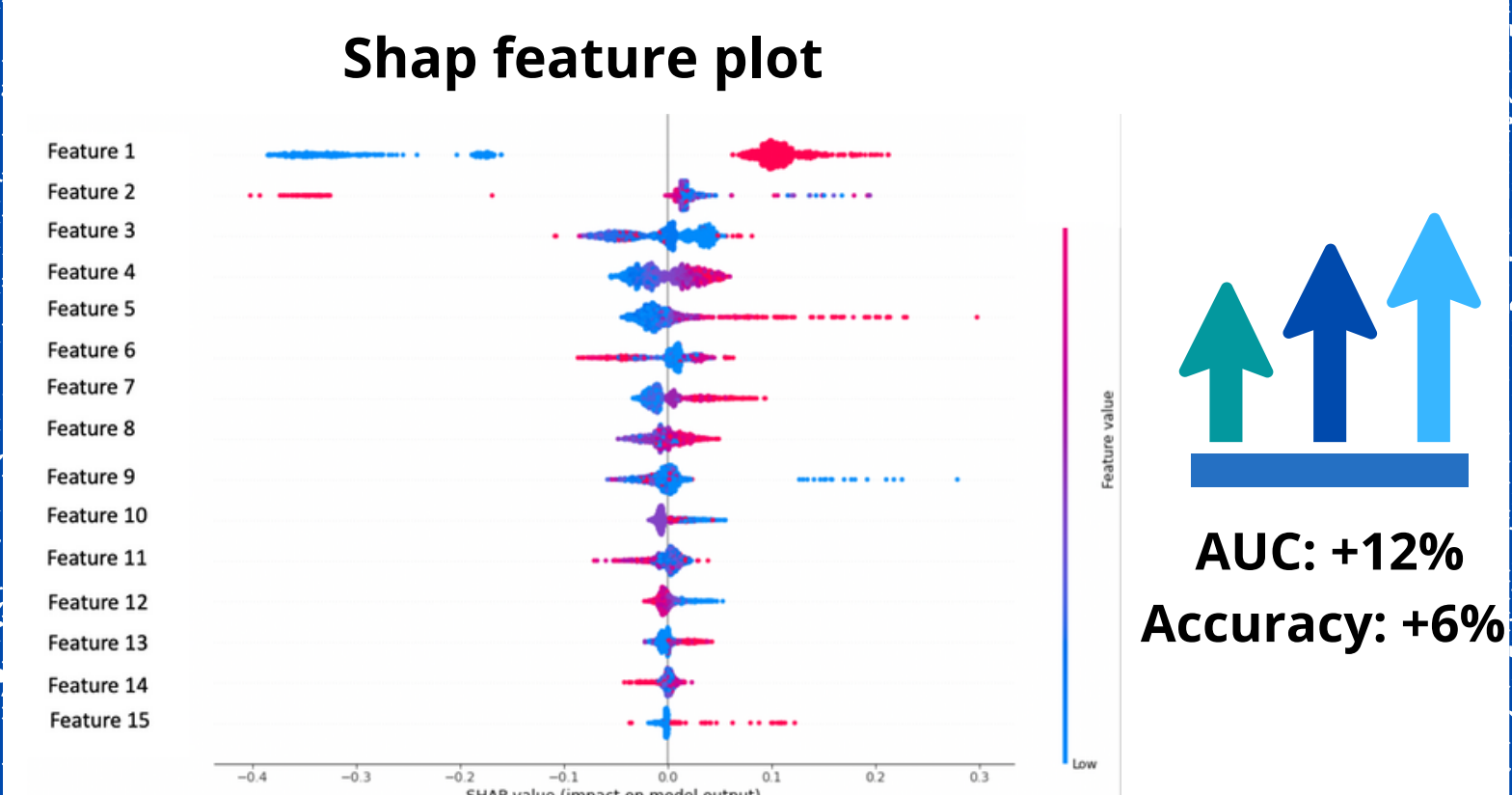| Rule	Visitors Blocked (%)	Total Profits (% change)	Profit per Visitor (% change)
5-features rule	1.3%	+0.01%	+1.48%
4-features rule	0.85%	+0.13%	+0.98%
2-features rule	0.53%	+0.09%	+0.64%

B) Predict policy activation

Goal: Improve accuracy when predicting if a policy will be activated

Strategy: Enrich predictions using more features

Understand business features driving effectuation

Validate model performance against baseline

**Results:** Production model, improved performance, understanding of feature impact

Shap feature plot

Feature 1
Feature 2
Feature 3
Feature 4
Feature 5
Feature 6
Feature 7
Feature 8
Feature 9
Feature 10
Feature 11
Feature 12
Feature 13
Feature 14
Feature 15

SHAP value (impact on model output)

AUC: +12%
Accuracy: +6%

C) Predict retention

Goal: Improve the current approach used to predict the retention

Strategy: We use a two-stage approach
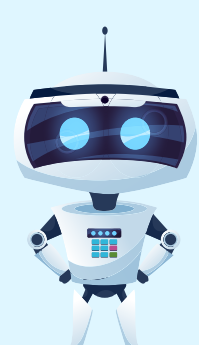
Classify-then-Regress

Results: We were able to achieve **strong out of sample** performance predicting the number of months during which a policy will be retained given it was submitted
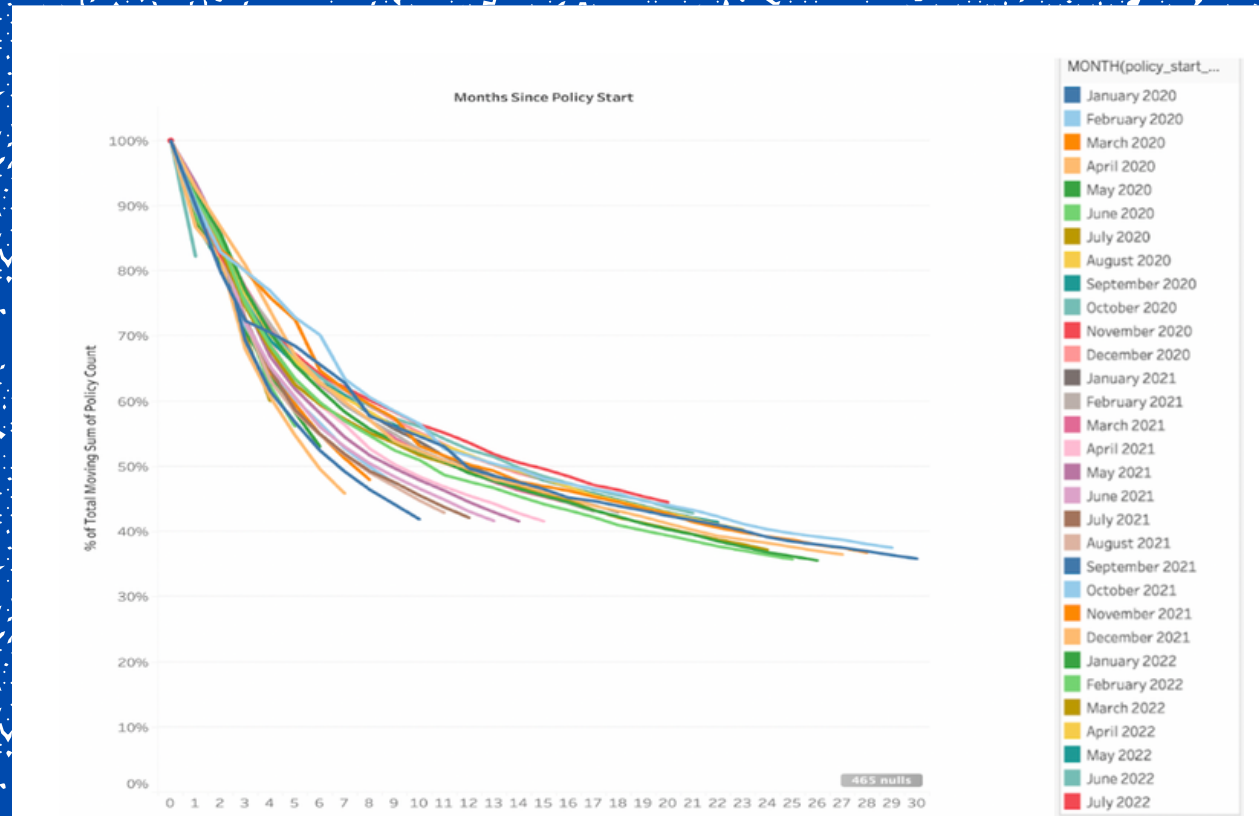
15% decrease in MAE

Business Impact

~1K likely bots detected

+$1.1M increased annual revenue forecast

+$10K increased monthly profit

Policy activation | valuable insights | Policy retention

Next steps

- Full production implementation and evaluation

- Testing of blocking vs. challenging users

- Continued improvement and deployment of LTV models