Stuck between a ROC and a hard place

Holly Stewart
Principal Research Lead
Windows Defender Antivirus

Microsoft

Overview
Stuck between a ROC and a hard place

Why do we need machine learning, anyway? What's the value, and how does it help protect people?

What are the inherent problems with machine learning, and why are these an issue for security researchers?

How can we resolve these inherent issues? How can we listen to our customers to make better decisions?

WHY MACHINE LEARNING? **BILLIONS OF SIGNALS**

Windows Defender ATP

Signals from hundreds of millions of customers

Microsoft Edge and Internet Explorer

8B internet downloads

Bing

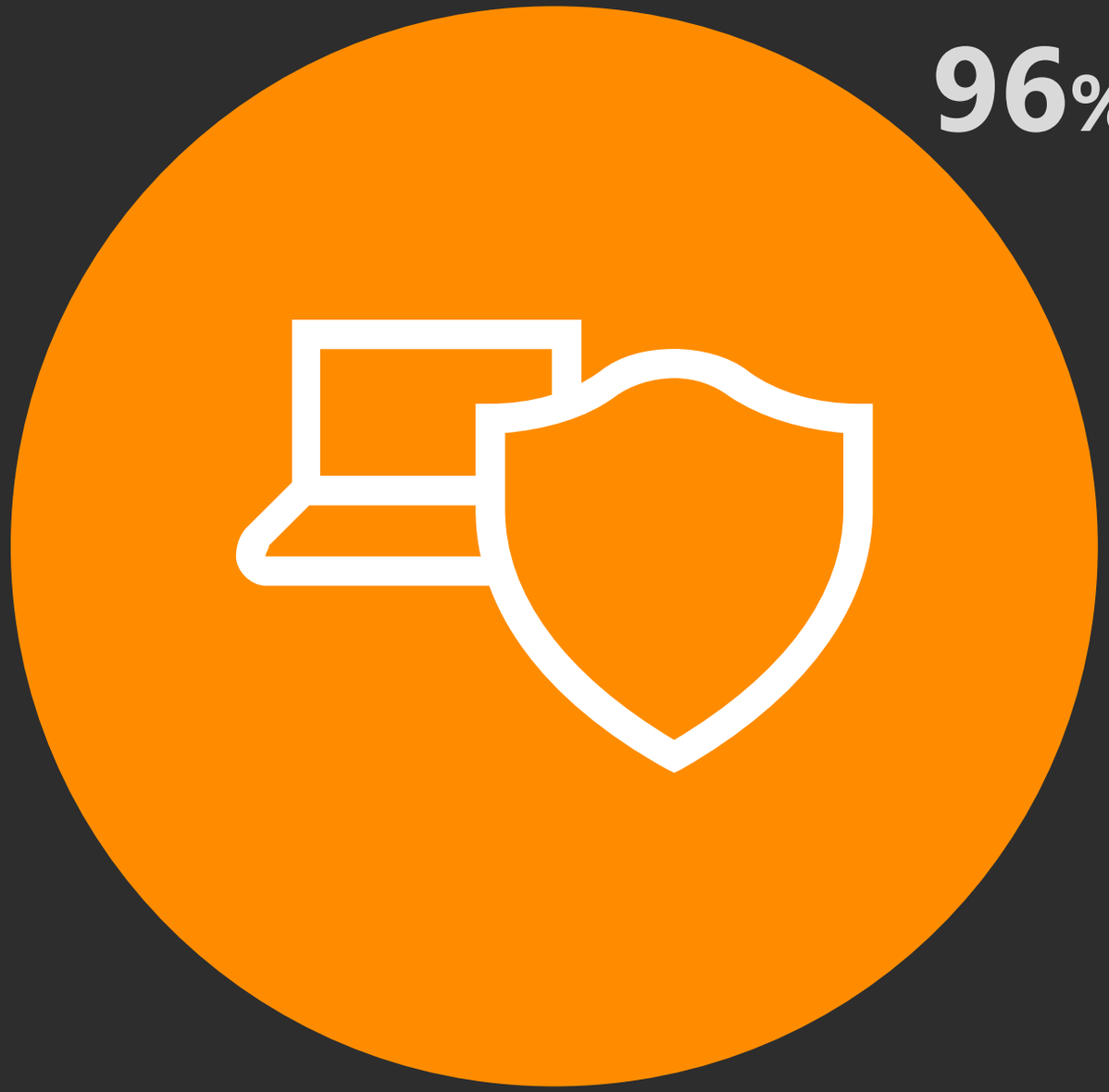
18B web pages scanned

Office 365

400B emails analyzed

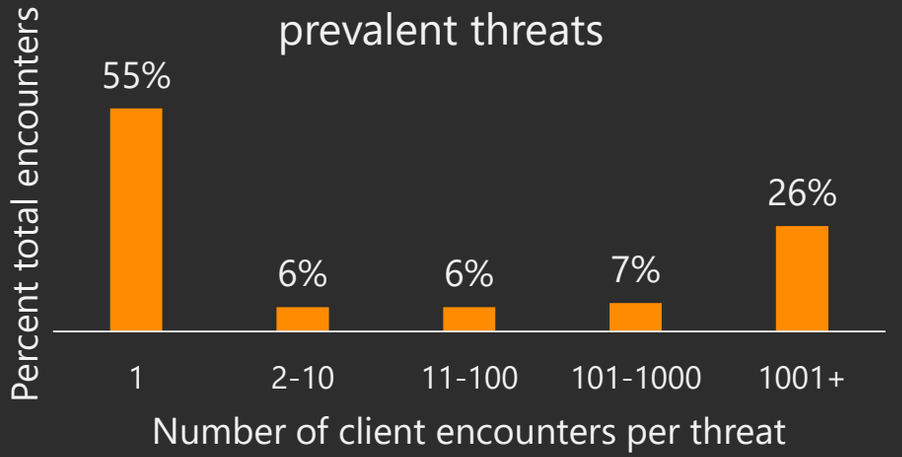
Microsoft

WHY MACHINE LEARNING? **THREAT LANDSCAPE**



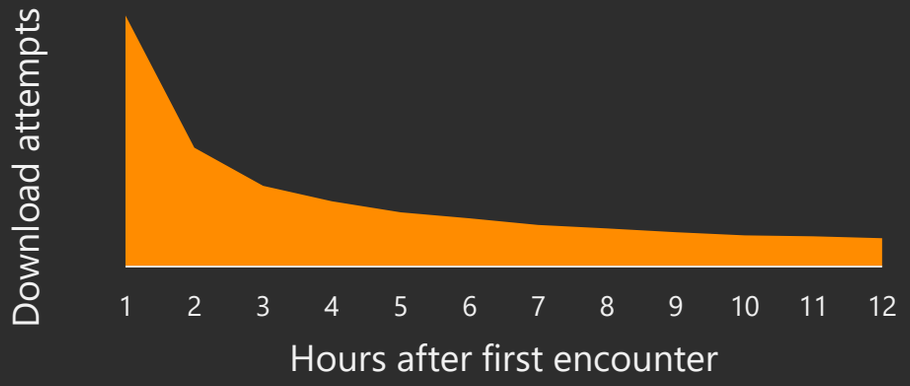
96% of malware are seen only once

Customer impact of unique and prevalent threats



SOURCE: WINDOWS DEFENDER ANTIVIRUS, Q1 2017

Malicious downloads encountered more than once

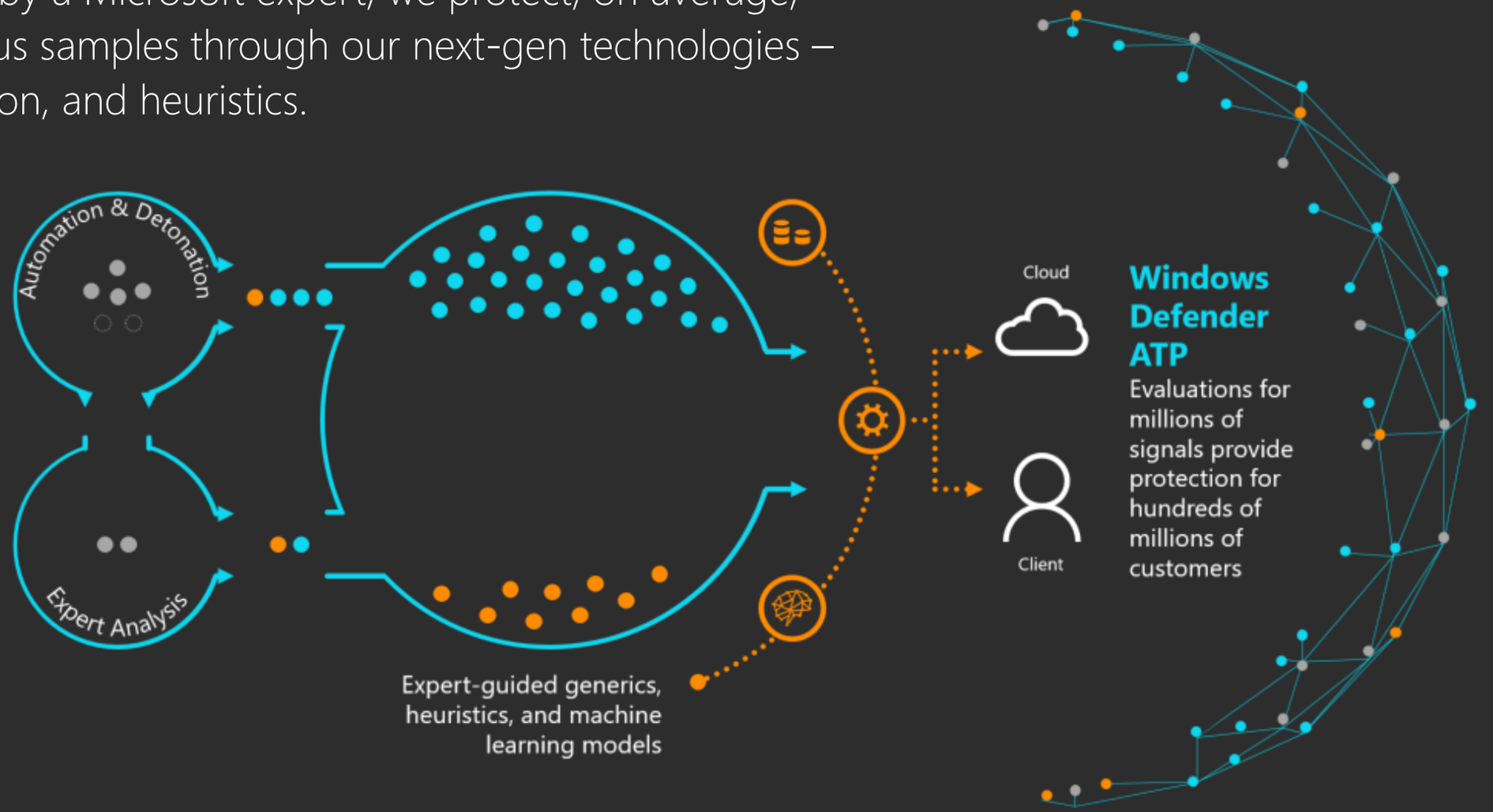


SOURCE: WINDOWS DEFENDER ANTIVIRUS, AUGUST 2017

WHY MACHINE LEARNING? **SCALE**

Supervised machine learning scales human expertise

For every sample analyzed by a Microsoft expert, we protect, on average, against 4,500 other malicious samples through our next-gen technologies – machine learning, automation, and heuristics.

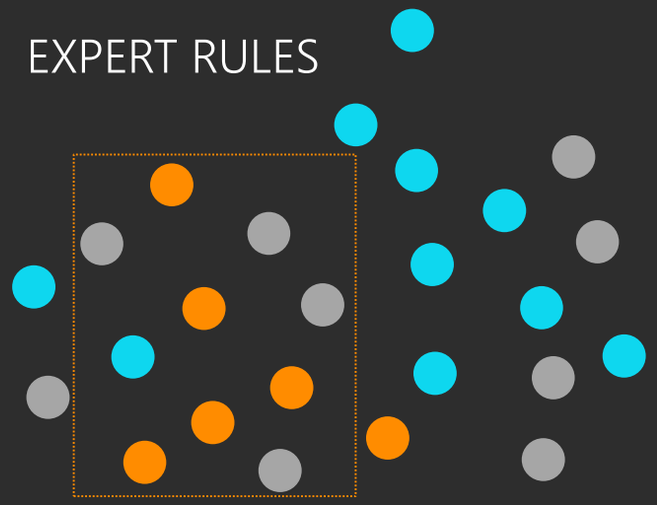


WHY MACHINE LEARNING? **PRECISION**

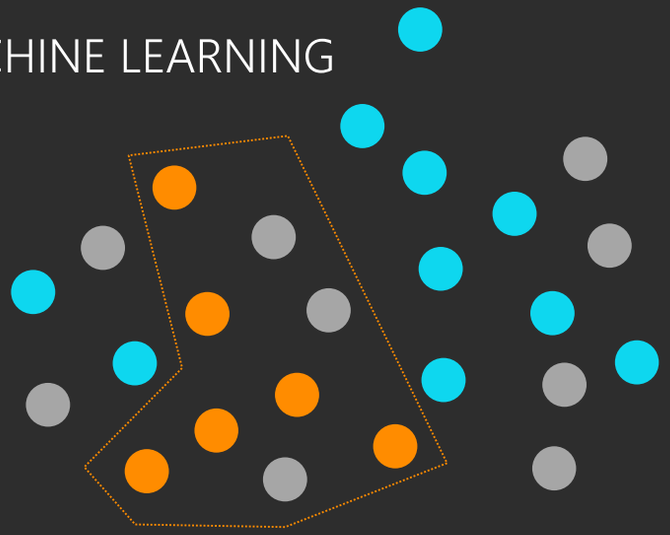
Supervised machine learning computes hundreds of thousands of variables into precise categories

Humans can create expert rules that combine tens or maybe hundreds of signal data, but machines compute highly dimensional data

EXPERT RULES



MACHINE LEARNING

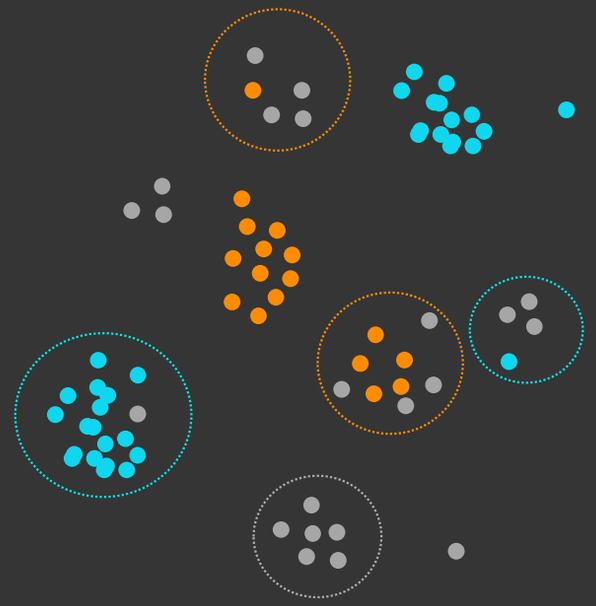


WHY MACHINE LEARNING? **HUMAN BIAS**

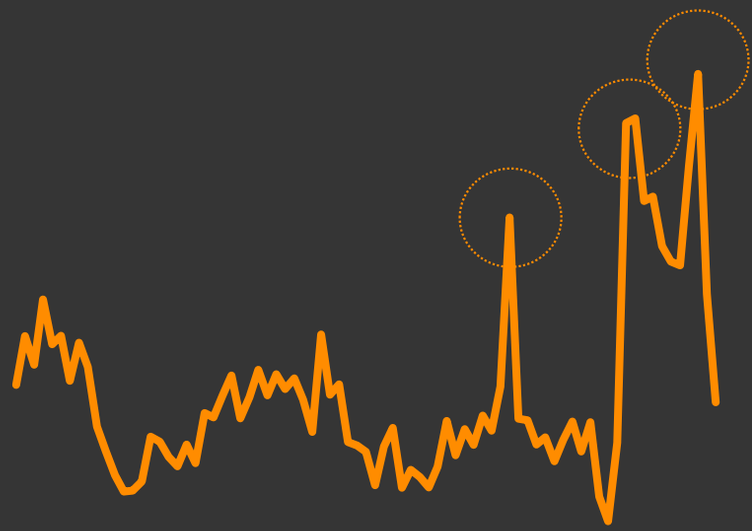
Unsupervised learning helps remove human bias

Machines can remove the human bias that come with expertise to reveal unexpected insights

UNKNOWN UNKNOWNS



ANOMALY DETECTION

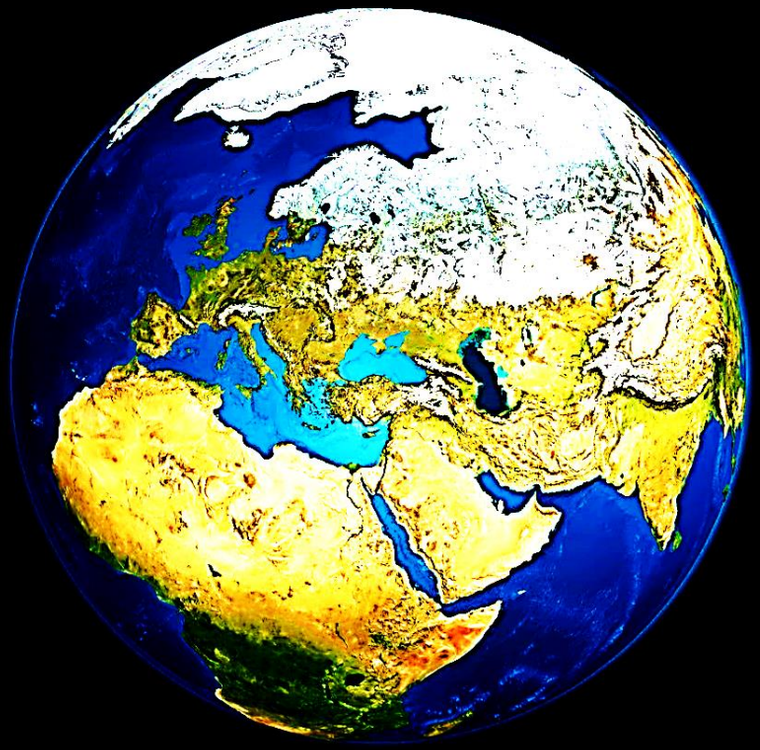


Wait. What's the problem?

MODELS ARE AN ABSTRACT REPRESENTATION OF REALITY

The one true earth…

Multidimensional model…



Two-dimensional model

By definition* machine learning models are imperfect

*else they would not be a “model”

How we measure machine learning models

CONFUSION TABLE

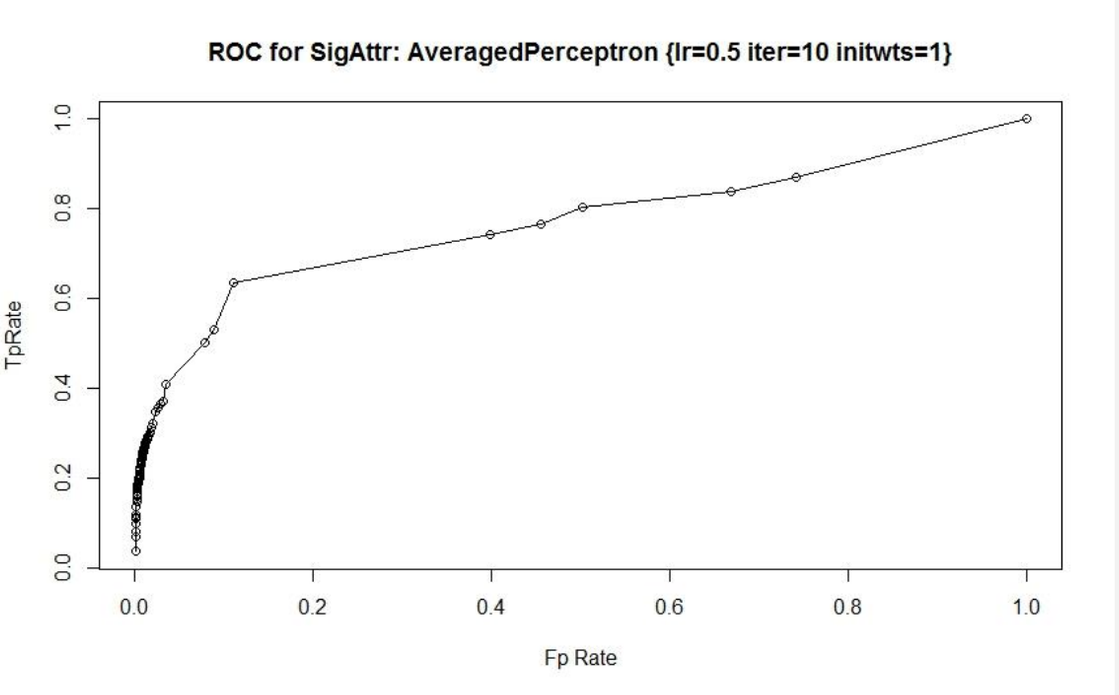
```
                                    ||================================||
PREDICTED                           || positive | negative   | Recall||
TRUTH                               ||================================||
positive                            ||   65,975 |  277,058   | 0.1923||
negative                            ||   48,608 | 11,179,862 | 0.9957||
                                    ||================================||
Precision                           ||   0.5758 |   0.9758   |       ||
                                    ||================================||

OVERALL 0/1 ACCURACY: 0.971856
```

ACCURACY, PRECISION, AREA UNDER THE CURVE

AUC:	0.828116 (0.0000)
Accuracy:	0.971856 (0.0000)
Positive precision:	0.575783 (0.0000)
Positive recall:	0.192328 (0.0000)
Negative precision:	0.975817 (0.0000)
Negative recall:	0.995671 (0.0000)
Log-loss:	0.155248 (0.0000)
Log-loss reduction:	19.396277 (0.0000)
F1 Score:	0.288342 (0.0000)
AUPRC:	0.323377 (0.0000)

Ubiquitous Receiver Operating Characteristic (ROC) Curve



How do we strike the right balance?

Our Approach

Retrospectively measure…

FNs - false negatives (misses)
FPs - false positives (incorrect detections)

Impact to consumers

Are people more likely to switch from Windows Defender Antivirus to another product after an FN or FP event?
(We call this switch ***customer churn***.)

Source: Consumer Windows Defender Antivirus customers on Windows 10 who used the Microsoft Malicious Software Removal Tool, Jan.-Apr. 2017

Measuring FNs

Threat active upon detection

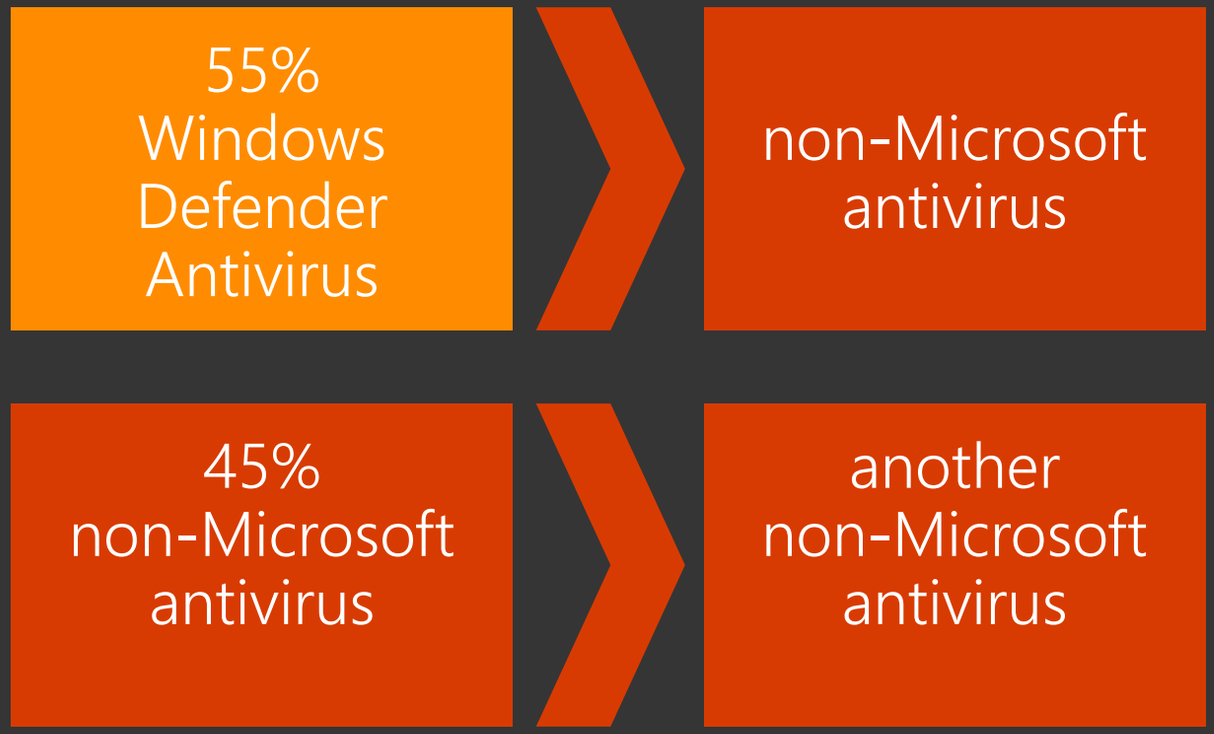
Classifier, threat report or researcher later marked file or behavior as malicious and client sent telemetry-only report (did not block)

Measuring FPs

Classifier or researcher later marked file or certificate as clean and reported as threat

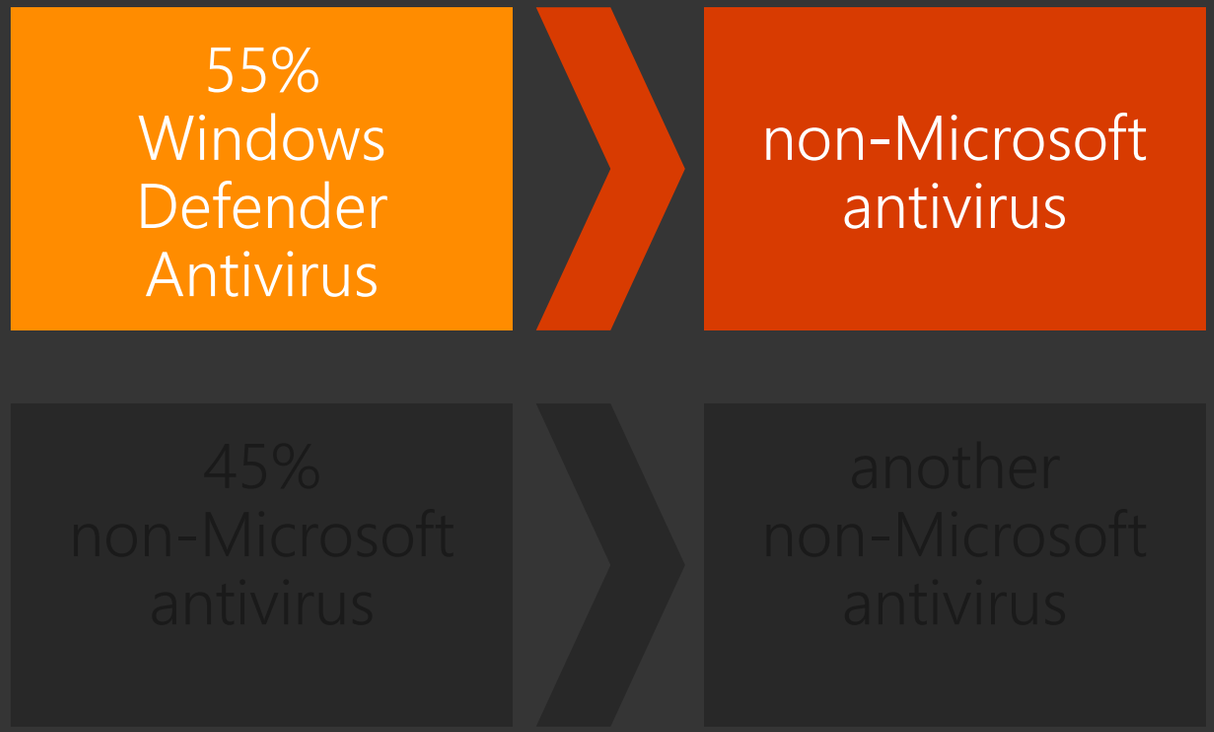
Windows 10 Antivirus Customer Churn

Insight: Lots of people change their antivirus vendor on Windows 10



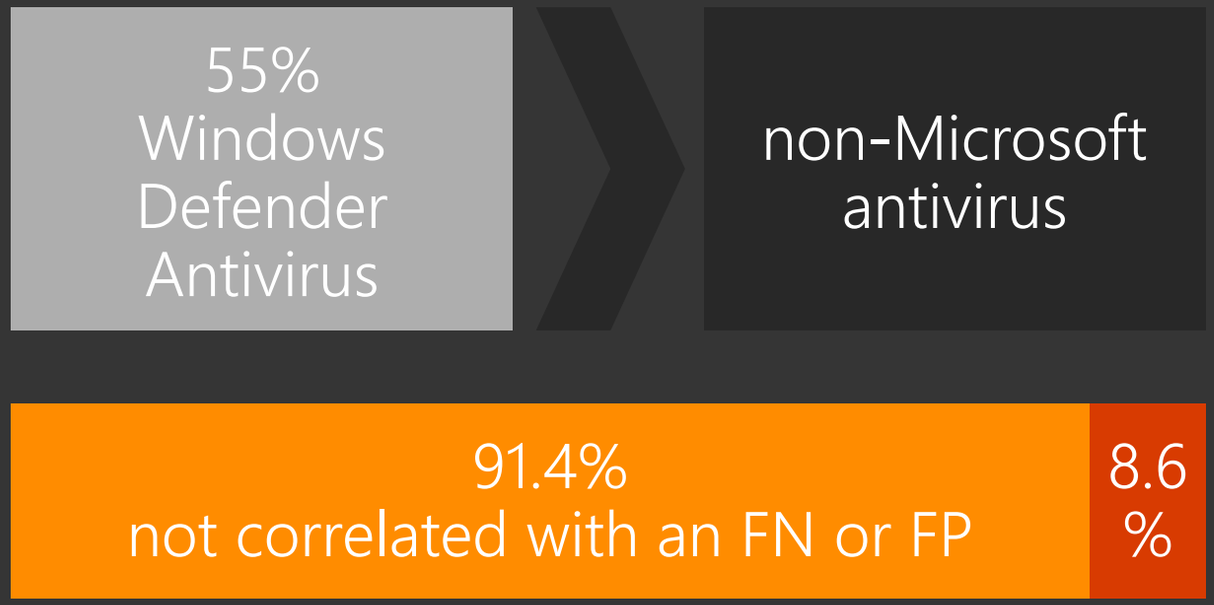
Data: 18 million computers switched to a non-Microsoft antivirus

Windows 10 Antivirus Customer Churn



Did an FP or FN even correlate with churn?

Insight: Most churn appears to be unrelated to FNs and FPs



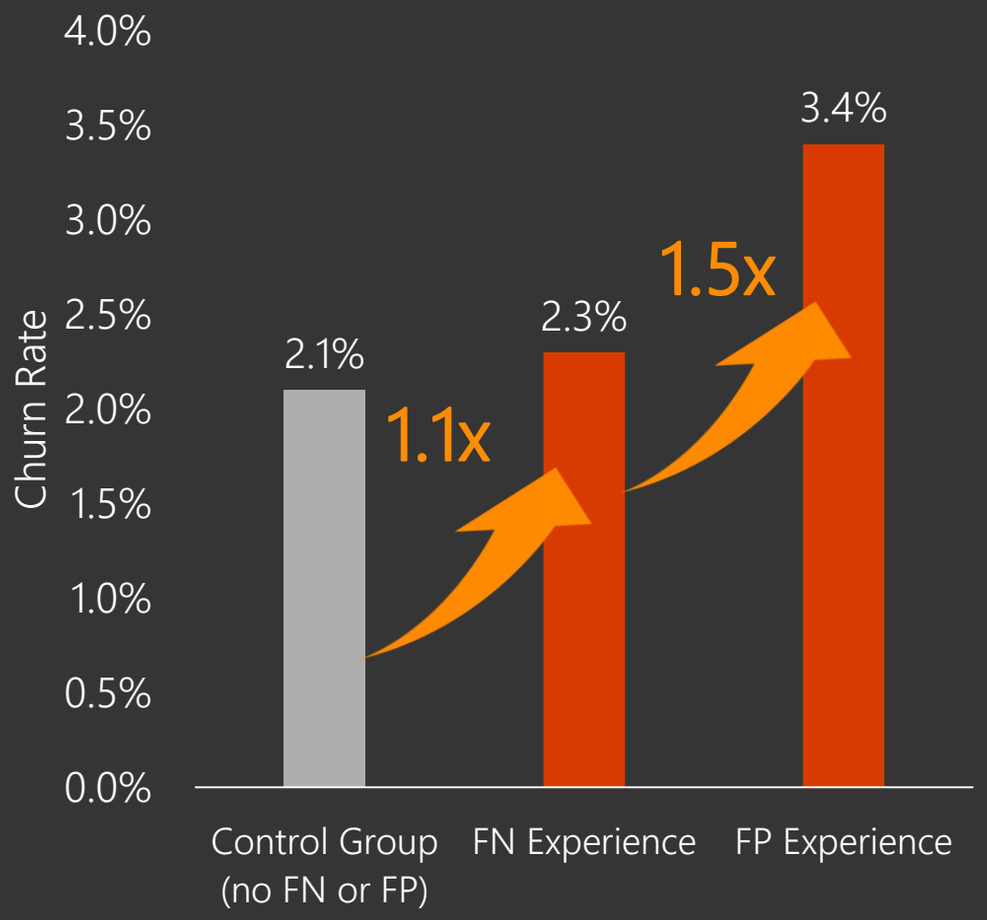
Data: Of the computers that switched to a non-Microsoft antivirus, only 8.6% were correlated with a false positive or a false negative.

Which is most highly correlated with churn?

Insights:

People are 1.1 times more likely to churn after an FN, in comparison to the control group

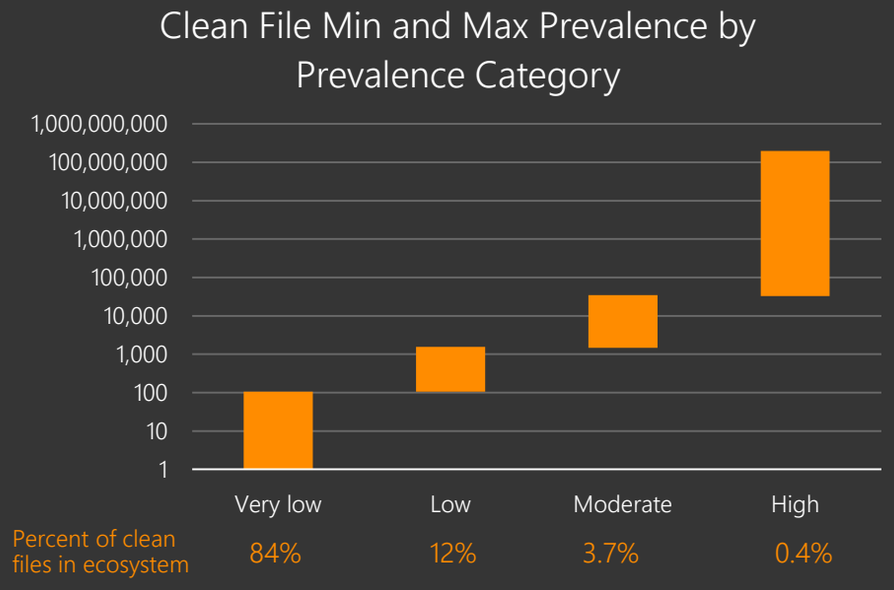
People are 1.5 times more likely to churn after an FP, in comparison to an FN

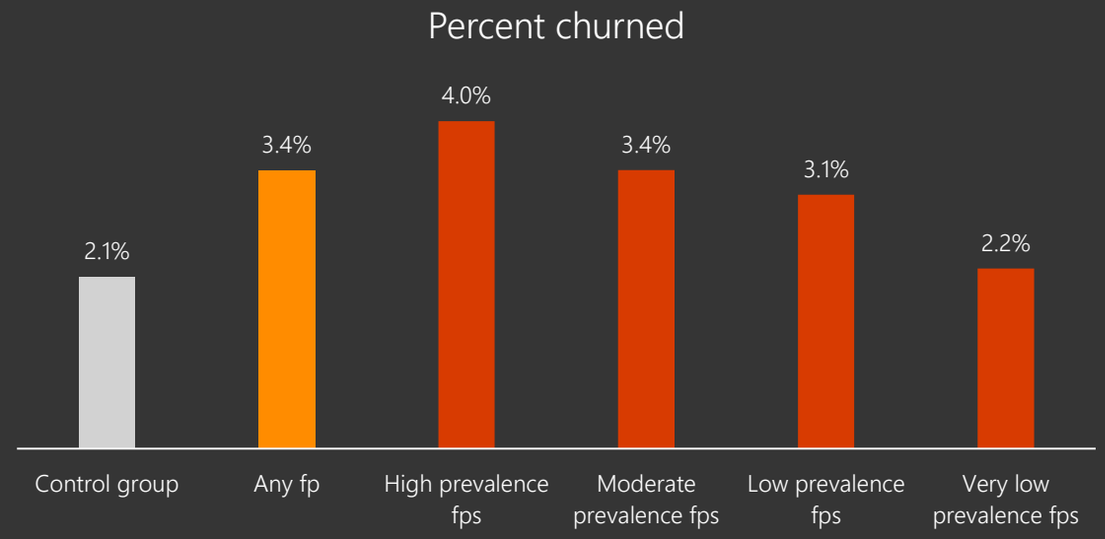


False positive impact

Does the severity of the FP matter?

Insight: Highly prevalent files were much more correlated with churn, while low prevalence files had little impact.

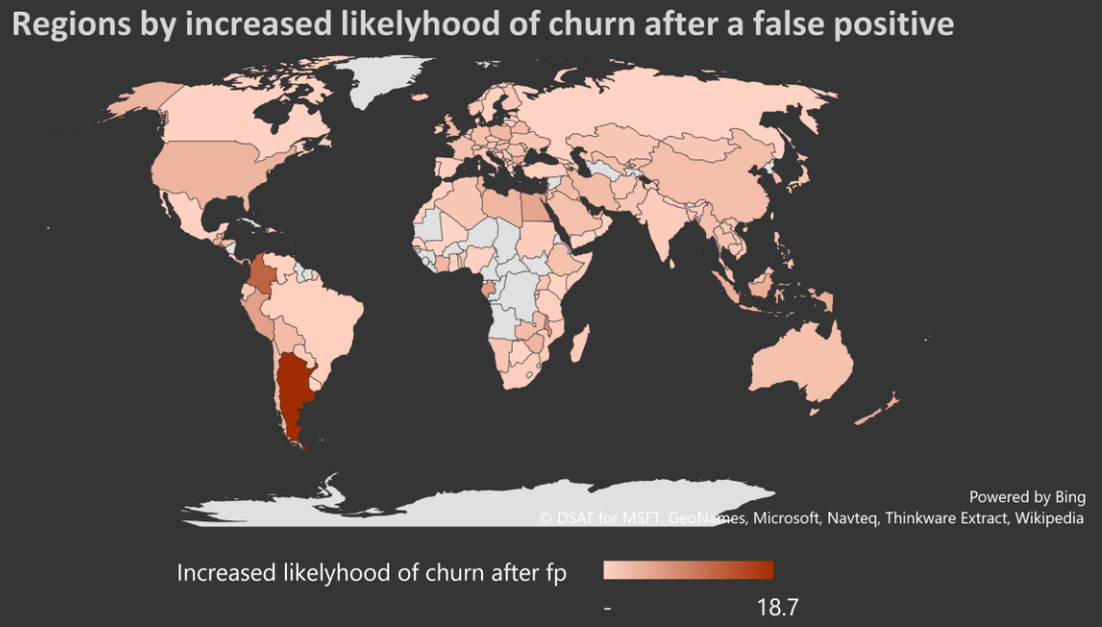




Data: People experiencing a highly prevalent FP were 1.9 times as likely to switch.

Are some populations more sensitive to FPs?

Insight: Some appear to be incredibly sensitive to FPs.



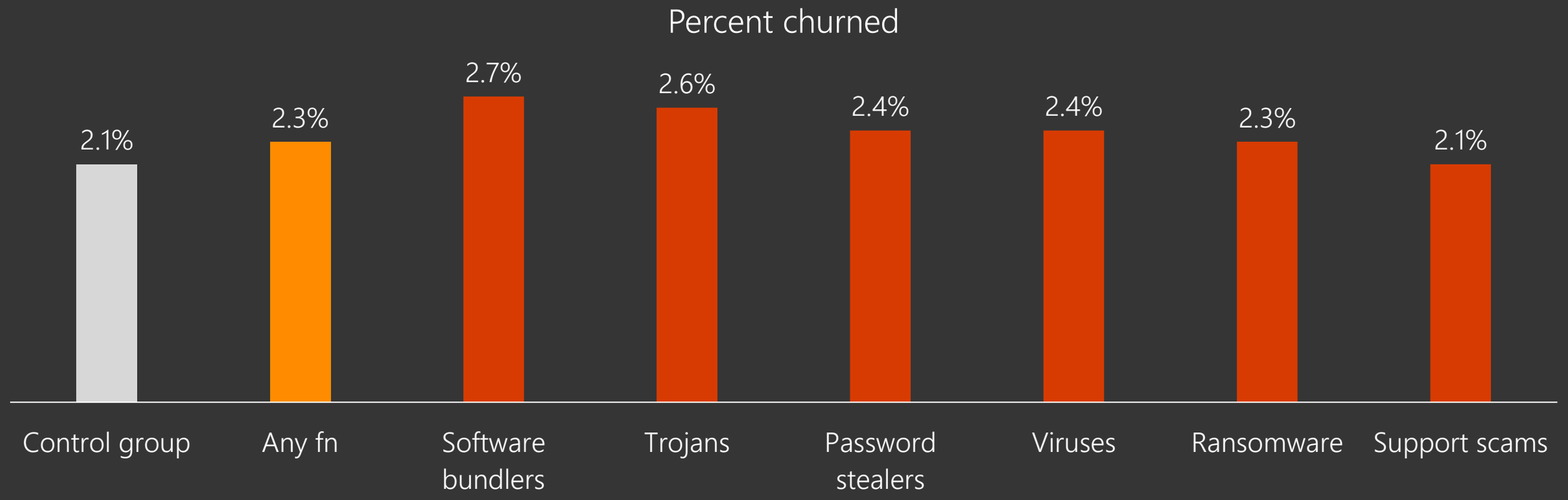
Region	Control churn	Fp churn	Increased likelihood of churn after FP
Argentina	0.2%	3.2%	18.7
Colombia	0.2%	3.1%	12.6
Indonesia	1.2%	4.7%	4.1
United States	2.8%	10.1%	3.6
United Arab Emirates	1.0%	3.3%	3.5
Poland	3.4%	11.1%	3.3

Data: People in Argentina and Colombia were respectively 18.7 and 12.6 times more likely to switch to another antivirus, while 4 other countries are more than double the average.

False negative impact

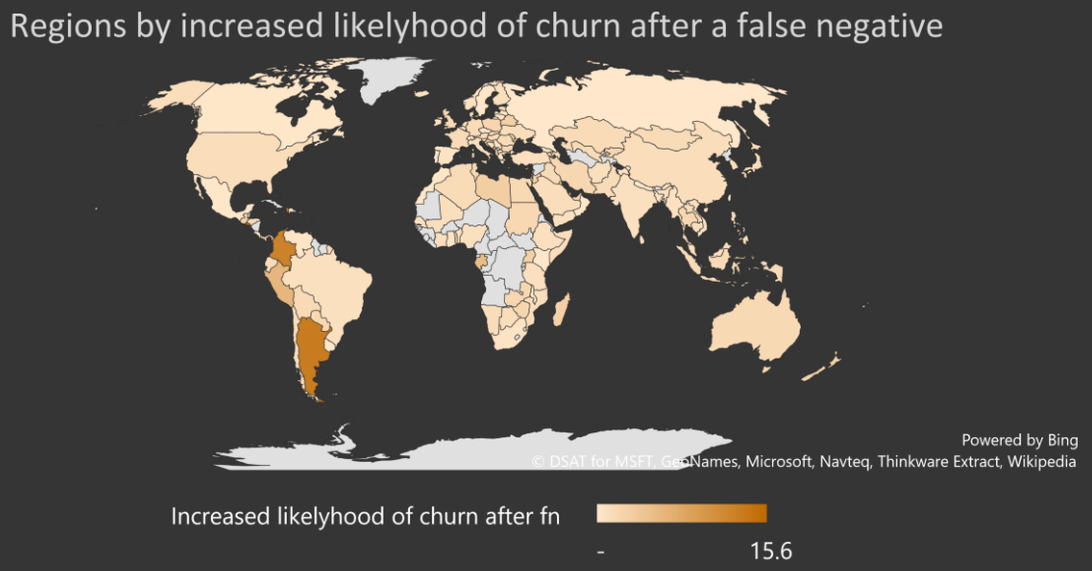
Does the category of the FN matter?

Insight: Bundlers surprisingly topped the list of FN categories, whereas highly visible threats like ransomware and support scams were closer to the FN average.



Are some populations more sensitive to FNs?

Insight: Some appear to be incredibly sensitive to FNs.



Region	Control churn	Fn churn	Increased likelihood of churn after FN
Argentina	0.2%	2.2%	13.2
Colombia	0.2%	3.1%	12.5
Israel	0.7%	2.1%	3.3
United Arab Emirates	1.0%	2.5%	2.6
Poland	3.4%	7.3%	2.2

Data: People in Argentina and Colombia were respectively 13.2 and 12.5 times more likely to switch to another antivirus, while 3 other countries are more than twice the average.

Conclusions

Key Insights and Questions

Some expected results

High prevalence FPs are more impactful than low prevalence FPs, and very low prevalence FPs don't have much impact at all.

People are 1.5 times as likely to churn because of an FP in comparison to an FN

Which do your customers experience more? Have you taken a balanced approach?

Geographical sensitivities

Some regions are especially sensitive to FPs and FNs. What can you do to better understand the applications they use to prevent FPs and threats that are specific to their geography?

Questions?