© 2019 SPLUNK INC.

Defense Against the Dark Arts: Splunk Edition

Melisa Napoles | Erika Strano

Wednesday October 23, 2019

© 2019 SPLUNK INC.

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or plans of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results may differ materially. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, it may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements made herein.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only, and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionalities described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Turn Data Into Doing, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2019 Splunk Inc. All rights reserved.

© 2019 SPLUNK INC.

Agenda

1. Why is the magic of Machine Learning (ML) essential to Defense Against the Dark Arts?
2. How are we seeing organizations detect the presence of dark wizards?
3. Let's journey into the dark forest and hunt down the death eaters!
4. Close your spell book and get ready to duel.
5. Concepts behind ML magic

****Muggle version of the agenda in the Appendix****

© 2019 SPLUNK INC.

Melisa Napoles
Sales Engineer | Splunk

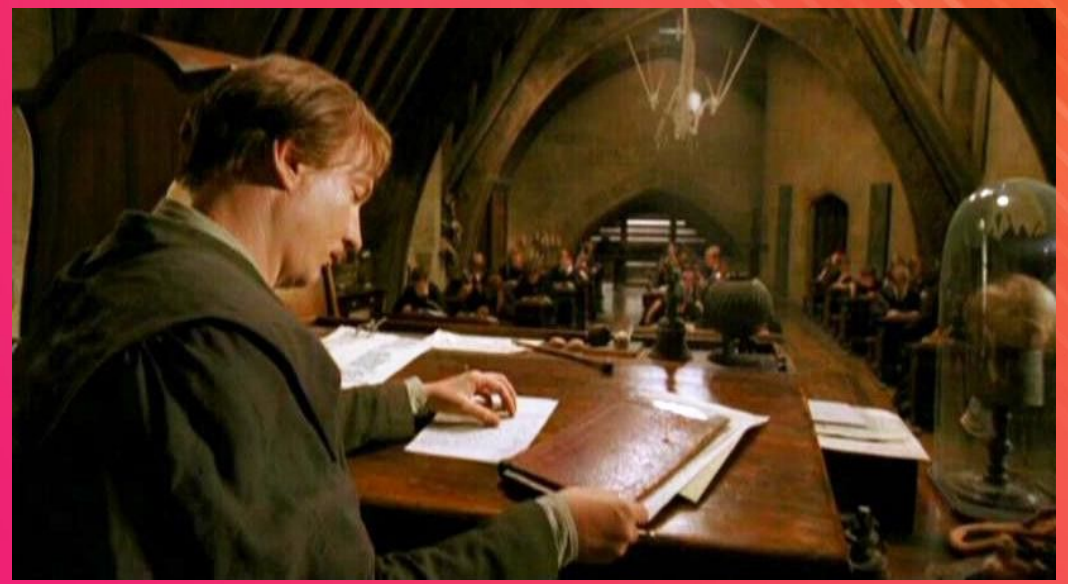
Erika Strano
Sales Engineer | Splunk

splunk> .conf19

© 2019 SPLUNK INC.

© 2019 SPLUNK INC.

Why is the magic of Machine Learning (ML) essential to Defense Against the Dark Arts?



splunk> .conf19

Machine Learning can help you ...

© 2019 SPLUNK INC.

© 2019 SPLUNK INC.

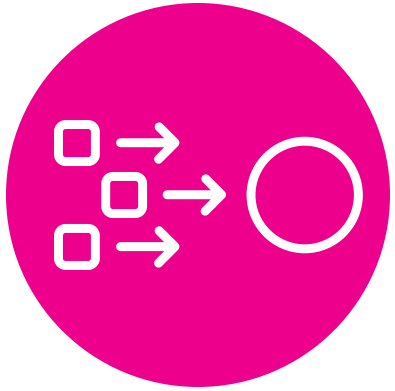
…so you can find your bears!




© 2019 SPLUNK INC.

How Does Machine Learning Work?

Do you believe in magic?



Use mathematical models to **learn** patterns in information

Catalog the patterns (and in some cases, iterate them as new data is received)

Use learned patterns to understand and interpret new data or make predictions


According to Gartner, a citizen data scientist uses analytics, but their title is not "data scientist" or similar. They complement data scientists and bring their OWN expertise and unique skills.
https://www.gartner.com/en/documents/3534848
https://blogs.gartner.com/carlie-idoine/2018/05/13/citizen-data-scientists-and-why-they-matter/

© 2019 SPLUNK INC.

Machine learning doesn’t solve cyber security, but it can help!

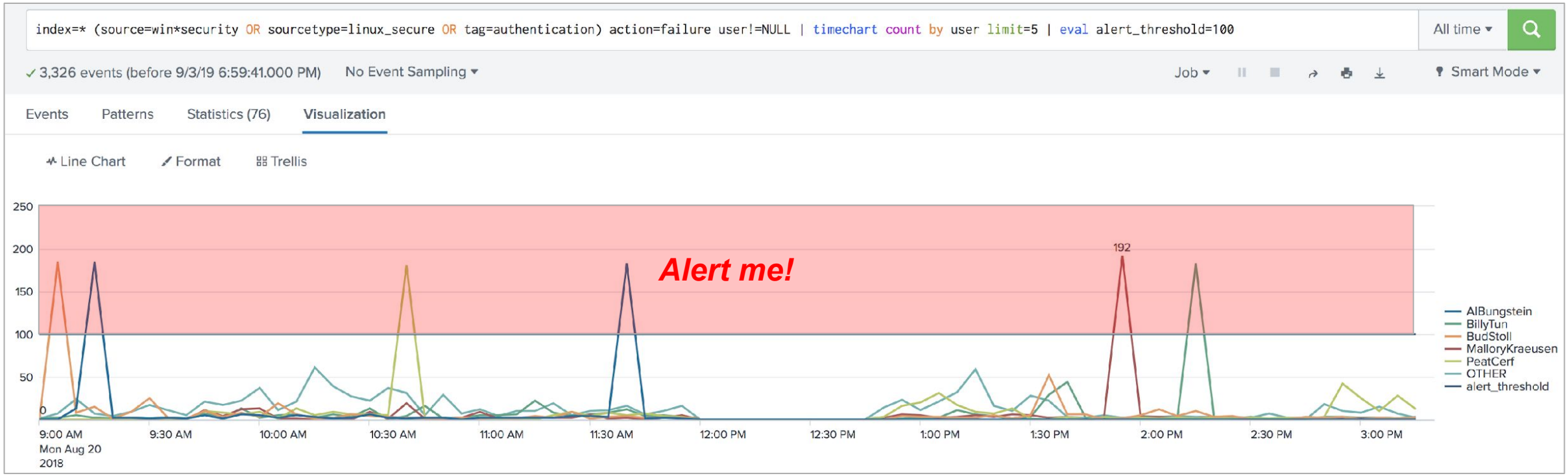
© 2019 SPLUNK INC.

How are we seeing organizations detect the presence of dark wizards?

© 2019 SPLUNK INC.

Who Finds Anomalies Like This Today?

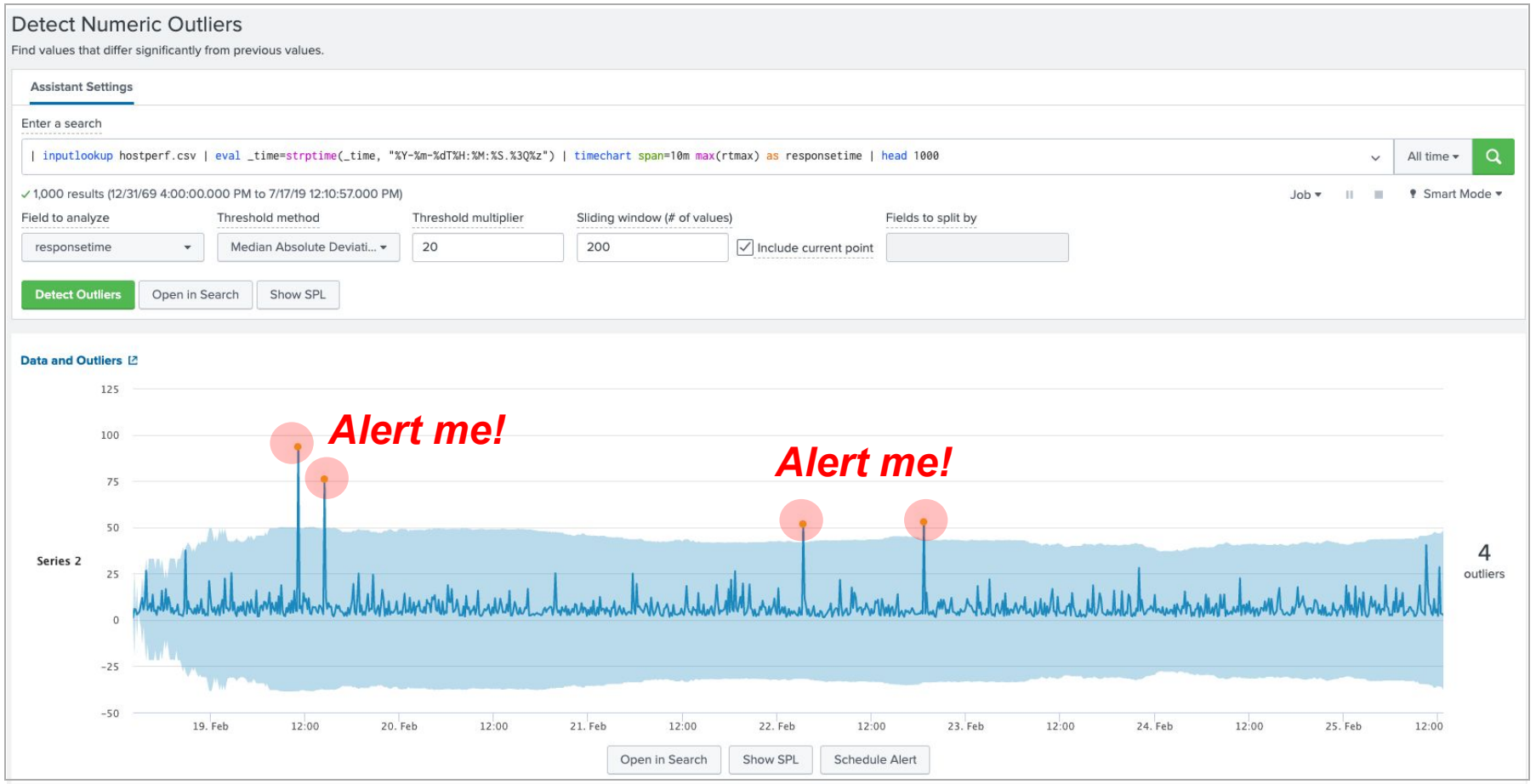
One search with a static threshold



© 2019 SPLUNK INC.

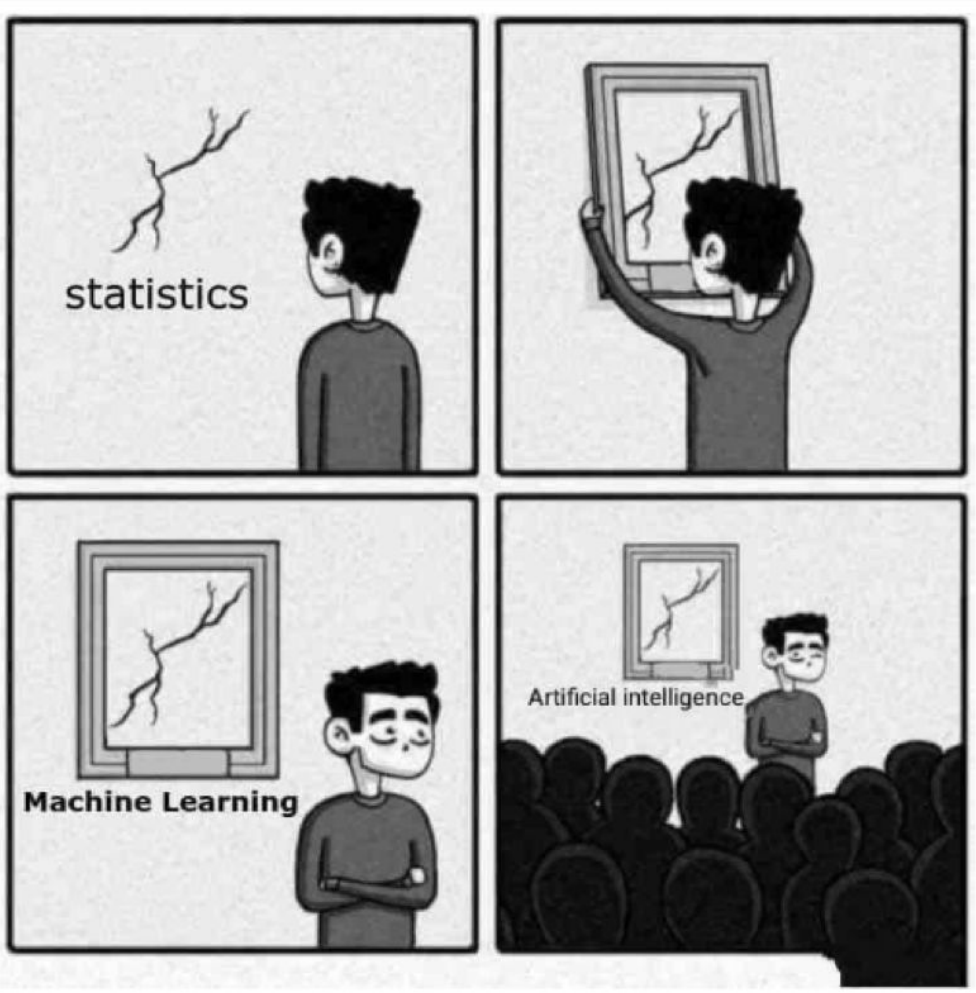
How About Like This?

One search with a dynamic threshold



© 2019 SPLUNK INC.

ML = Stats OR ML != Stats?



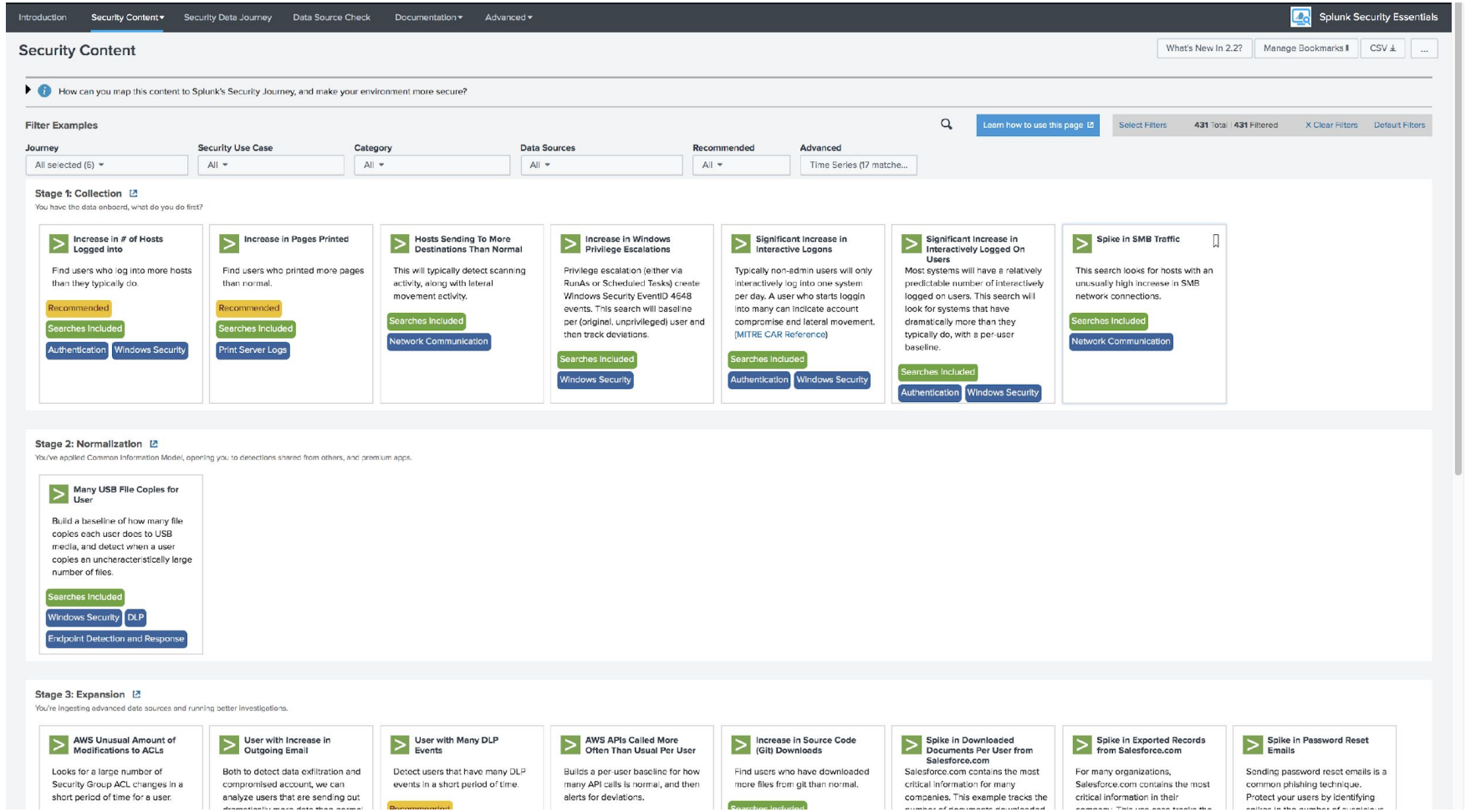
original comic by sandserif

splunk> .conf19

© 2019 SPLUNK INC.

Check Out SSE For a Jump Start with 17 Use-cases

Thanks Splunk Security Essentials (SSE)!



splunk> .conf19

© 2019 SPLUNK INC.

Is this scalable?

You're running your anomaly search on all your data every time...

© 2019 SPLUNK INC.

Instead, try...

fit first and then, *apply*

© 2019 SPLUNK INC.

…and make sure your data fits the model…

not the other way around.

© 2019 SPLUNK INC.

Enter Splunk Machine Learning Toolkit (MLTK)

Extends Splunk with new tools and guided modeling environment

Experiments and Assistants: Guided model building, testing, and deployment for common objectives

Showcases: Interactive examples for typical IT, security, business, and IoT use cases

Algorithms: 80+ standard algorithms (supervised & unsupervised)

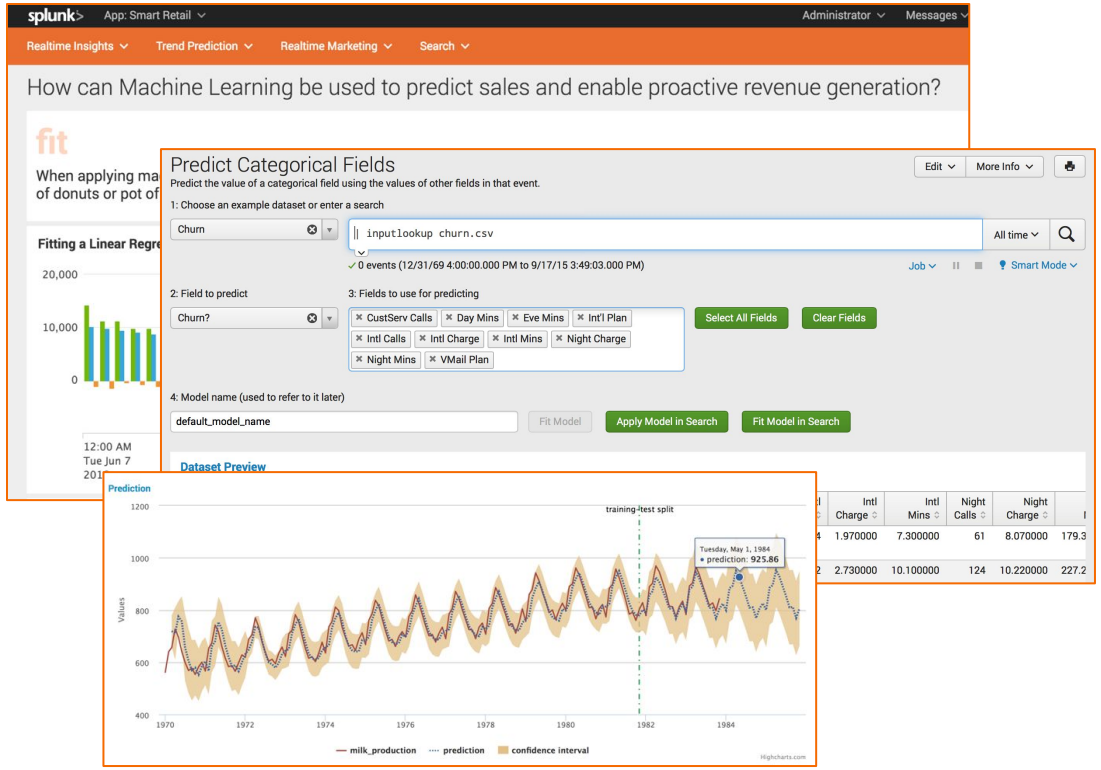
ML Commands: New SPL commands to fit, test, score and operationalize models

ML-SPL API: Extensibility to easily import any algorithm (proprietary / open source)

Python for Scientific Computing Library: Access to 300+ open source algorithms

Apache Spark MLLib: Support large scale model training via Spark Add-on for MLTK (LAR)

Tensorflow Container: Supports NN and GPU accelerated machine learning



splunk> .conf19

© 2019 SPLUNK INC.

In Fact, ESCU is Now Doing This with MLTK!

Check out the "New: Machine Learning in Splunk Enterprise Security Content Update" blog!

DNS Query Length Outliers

SMB Traffic Spike

Unusually Long Command Line

© 2019 SPLUNK INC.

Anomalies in the Context of MITRE

Initial Access	Execution	Persistence	Privilege Escalation	Defense Evasion	Credential Access	Discovery	Lateral Movement	Collection	Command And Control	Exfiltration	Impact
11 items	33 items	59 items	28 items	67 items	19 items	22 items	17 items	13 items	22 items	9 items	14 items
Drive-by Compromise	AppleScript	.bash_profile and .bashrc	Access Token Manipulation	Access Token Manipulation	Account Manipulation	Account Discovery	AppleScript	Audio Capture	Commonly Used Port	Automated Exfiltration	Data Destruction
Exploit Public-Facing Application	CMSTP	Accessibility Features	Accessibility Features	Binary Padding	Bash History	Application Window Discovery	Application Deployment Software	Automated Collection	Communication Through Removable Media	Data Compressed	Data Encrypted for Impact
External Remote Services	Command-Line Interface	Account Manipulation	AppCert DLLs	BITS Jobs	Brute Force	Browser Bookmark Discovery	Distributed Component Object Model	Clipboard Data	Connection Proxy	Data Encrypted	Defacement
Hardware Additions	Compiled HTML File	AppCert DLLs	AppInit DLLs	Bypass User Account Control	Credential Dumping	Domain Trust Discovery	Exploitation of Remote Services	Data from Information Repositories	Custom Command and Control Protocol	Data Transfer Size Limits	Disk Content Wipe
Replication Through Removable Media	Control Panel Items	AppInit DLLs	Application Shimming	Clear Command History	Credentials in Files	File and Directory Discovery	Logon Scripts	Data from Local System	Custom Cryptographic Protocol	Exfiltration Over Alternative Protocol	Disk Structure Wipe
Spearphishing Attachment	Dynamic Data Exchange	Application Shimming	Bypass User Account Control	CMSTP	Credentials in Registry	Network Service Scanning	Pass the Hash	Data from Network Shared Drive	Data Encoding	Exfiltration Over Command and Control Channel	Endpoint Denial of Service
Spearphishing Link	Execution through API	Authentication Package	DLL Search Order Hijacking	Code Signing	Exploitation for Credential Access	Network Share Discovery	Pass the Ticket	Data from Removable Media	Data Obfuscation	Exfiltration Over Other Network Medium	Firmware Corruption
Spearphishing via Service	Execution through Module Load	BITS Jobs	Dylib Hijacking	Compile After Delivery	Forced Authentication	Network Sniffing	Remote Desktop Protocol	Data Staged	Domain Fronting	Exfiltration Over Physical Medium	Inhibit System Recovery
Supply Chain Compromise	Exploitation for Client Execution	Bootkit	Exploitation for Privilege Escalation	Compiled HTML File	Hooking	Password Policy Discovery	Remote File Copy	Email Collection	Domain Generation Algorithms	Scheduled Transfer	Network Denial of Service
Trusted Relationship	Graphical User Interface	Browser Extensions	Extra Window Memory Injection	Component Firmware	Input Capture	Peripheral Device Discovery	Remote Services	Input Capture	Fallback Channels		Resource Hijacking
Valid Accounts	InstallUtil	Change Default File Association	File System Permissions Weakness	Component Object Model Hijacking	Input Prompt	Permission Groups Discovery	Replication Through Removable Media	Man in the Browser	Multi-hop Proxy		Runtime Data Manipulation
	Launchctl	Component Firmware	Hooking	Control Panel Items	Kerberoasting	Process Discovery	Shared Webroot	Screen Capture	Multi-Stage Channels		Service Stop
	Local Job Scheduling	Component Object Model Hijacking	Image File Execution Options Injection	DCShadow	Keychain	Query Registry	SSH Hijacking	Video Capture	Multiband Communication		Stored Data Manipulation
	LSASS Driver	Create Account	Launch Daemon	Deobfuscate/Decode Files or Information	LLMNR/NBT-NS Poisoning and Relay	Remote System Discovery	Taint Shared Content		Multilayer Encryption		Transmitted Data Manipulation
	Mshta	DLL Search Order Hijacking	New Service	Disabling Security Tools	Network Sniffing	Security Software Discovery	Third-party Software		Port Knocking		
	PowerShell	Dylib Hijacking	Path Interception	DLL Search Order Hijacking	Password Filter DLL	System Information Discovery	Windows Admin Shares		Remote Access Tools		
	Regsvcs/Regasm	External Remote Services	Plist Modification	DLL Side-Loading	Private Keys	System Network Configuration Discovery	Windows Remote Management		Remote File Copy		
	Regsvr32	File System Permissions Weakness	Port Monitors	Execution Guardrails	Securityd Memory	System Network Connections Discovery			Standard Application Layer Protocol		
	Rundll32	Hidden Files and Directories	Process Injection	Exploitation for Defense Evasion	Two-Factor Authentication Interception	System Owner/User Discovery			Standard Cryptographic Protocol		
	Scheduled Task	Hooking	Scheduled Task	Extra Window Memory Injection		System Service Discovery			Standard Non-Application Layer Protocol		
	Scripting	Hypervisor	Service Registry Permissions Weakness	File Deletion		System Time Discovery			Uncommonly Used Port		
	Service Execution	Image File Execution Options Injection	Setuid and Setgid	File Permissions Modification		Virtualization/Sandbox Evasion			Web Service		
	Signed Binary Proxy Execution	Kernel Modules and Extensions	SID-History Injection	File System Logical Offsets							
	Signed Script Proxy Execution	Launch Agent		Gatekeeper Bypass							
	Source	Launch Daemon		Group Policy Modification							
	Space after Filename			Hidden Files and Directories							
	Third-party Software			Hidden Users							
				Hidden Window							

splunk> .conf19

© 2019 SPLUNK INC.

Let’s journey into the dark forest and hunt down the death eaters!

splunk> .conf19

© 2019 SPLUNK INC.

Who In My Network is Executing Unusual DNS Query Lengths?

Which looks more unusually long to you?

ec2-13-193-103-139.us-west-1.compute.amazonaws.com

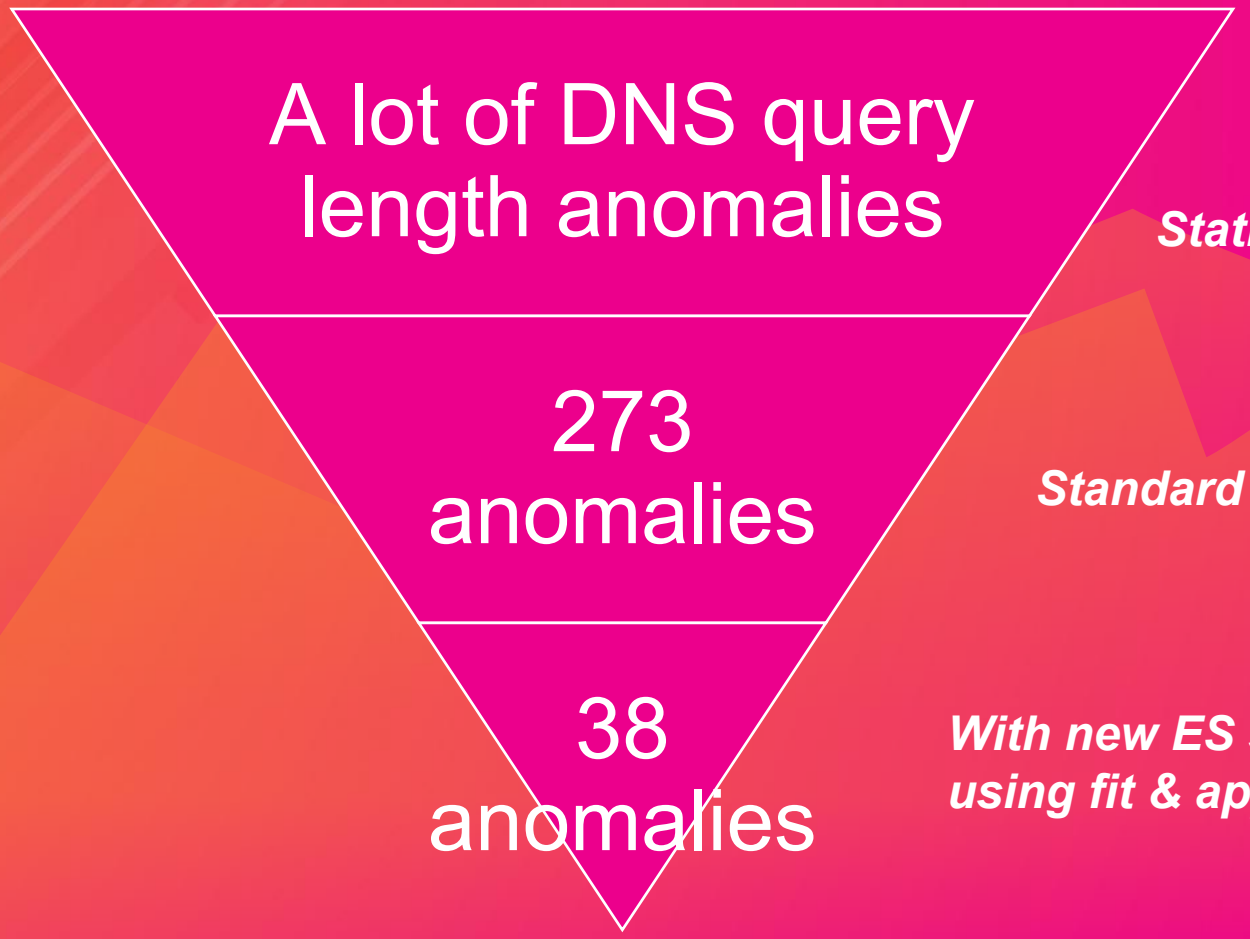
OR

p4-csddrc45xqiym-xmvtplcdqchjqs5r-322917-i2.anycast-stb.metric.gstatic.com

© 2019 SPLUNK INC.

Why Should You Care?

A look at our sample dataset



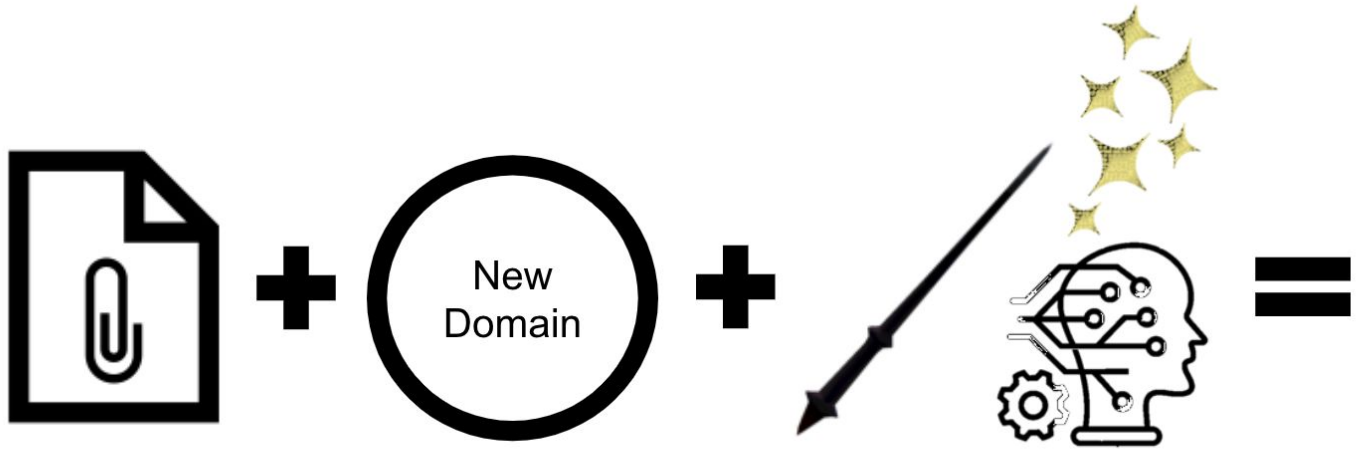
Static threshold

Standard Deviation

With new ES searches using fit & apply

© 2019 SPLUNK INC.

ML Can Help Add Context to Events



© 2019 SPLUNK INC.

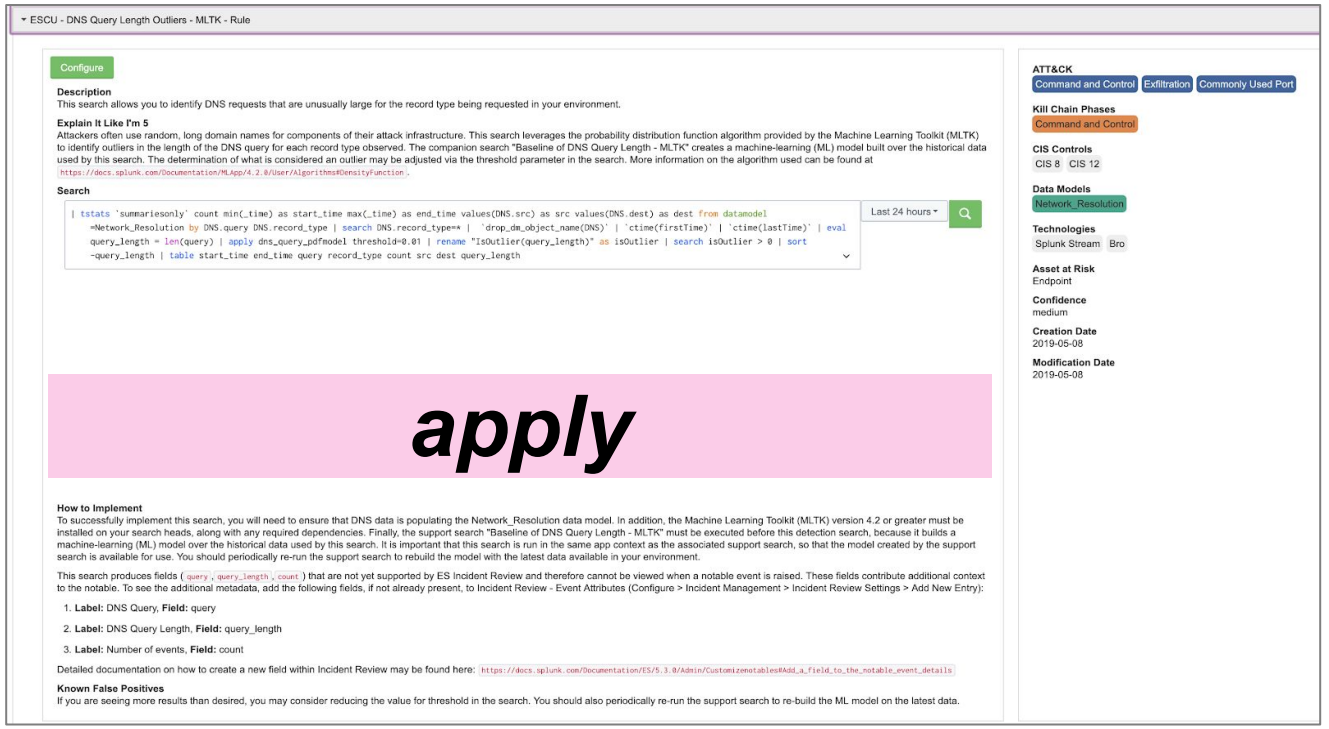
Close your spell book and get ready to duel.

splunk> .conf19

© 2019 SPLUNK INC.

Trusty ESCU Has Us Covered

Exfiltration Use Cases at our Fingertips

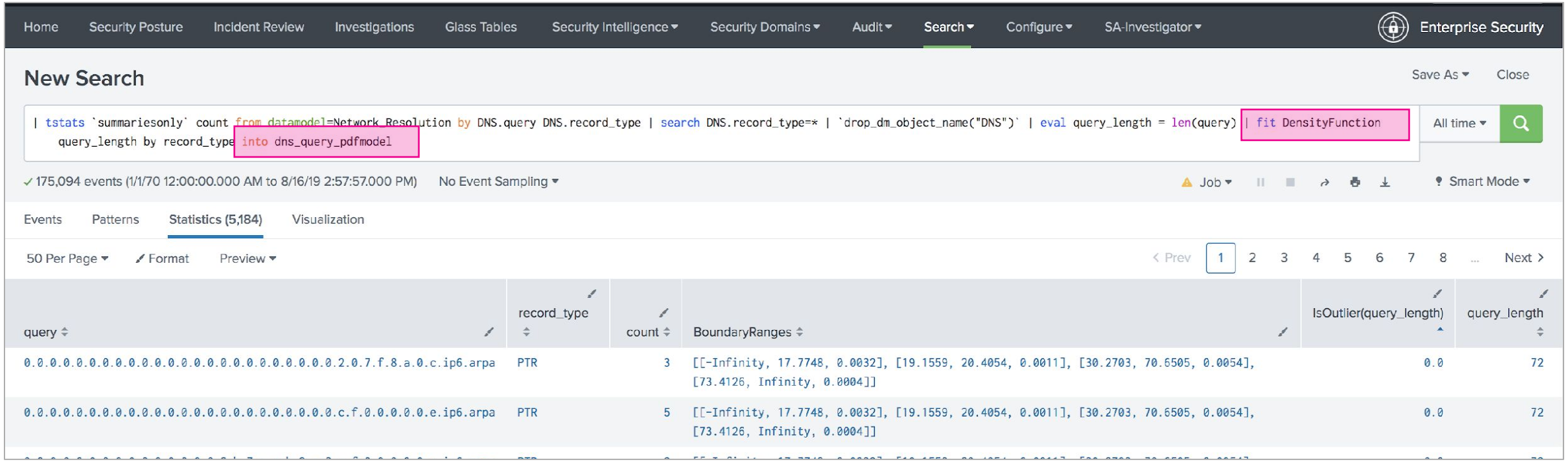


splunk> .conf19

© 2019 SPLUNK INC.

Creating a Model For Our Dataset

First the fit...



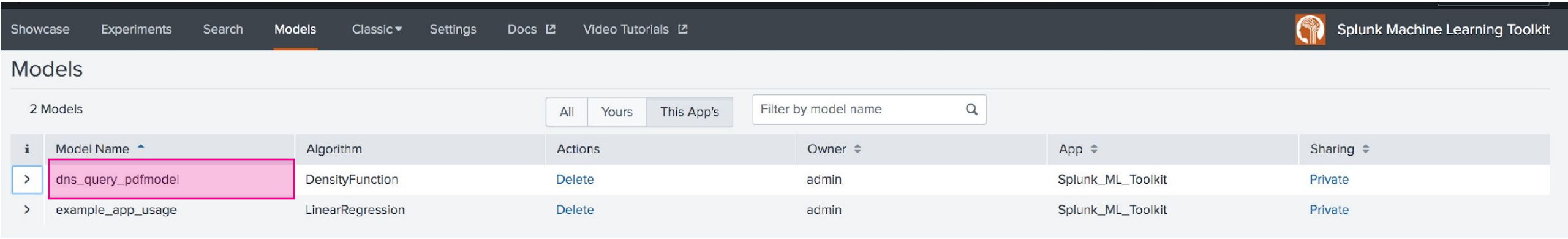
Notice we're building a model called "dns_query_pdfmodel"

splunk> .conf19

© 2019 SPLUNK INC.

Notice Your Model is Saved In MLTK!

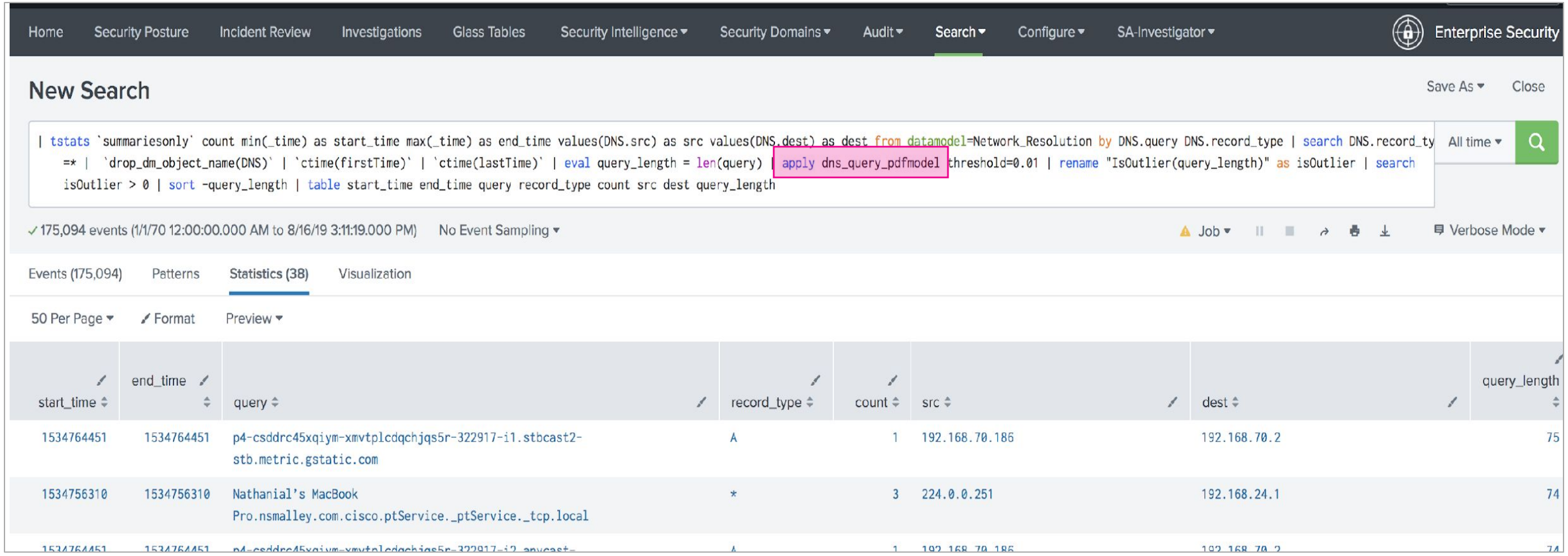
Bigger, stronger models to scale



© 2019 SPLUNK INC.

Applying the Model to Our Dataset

...then the apply!

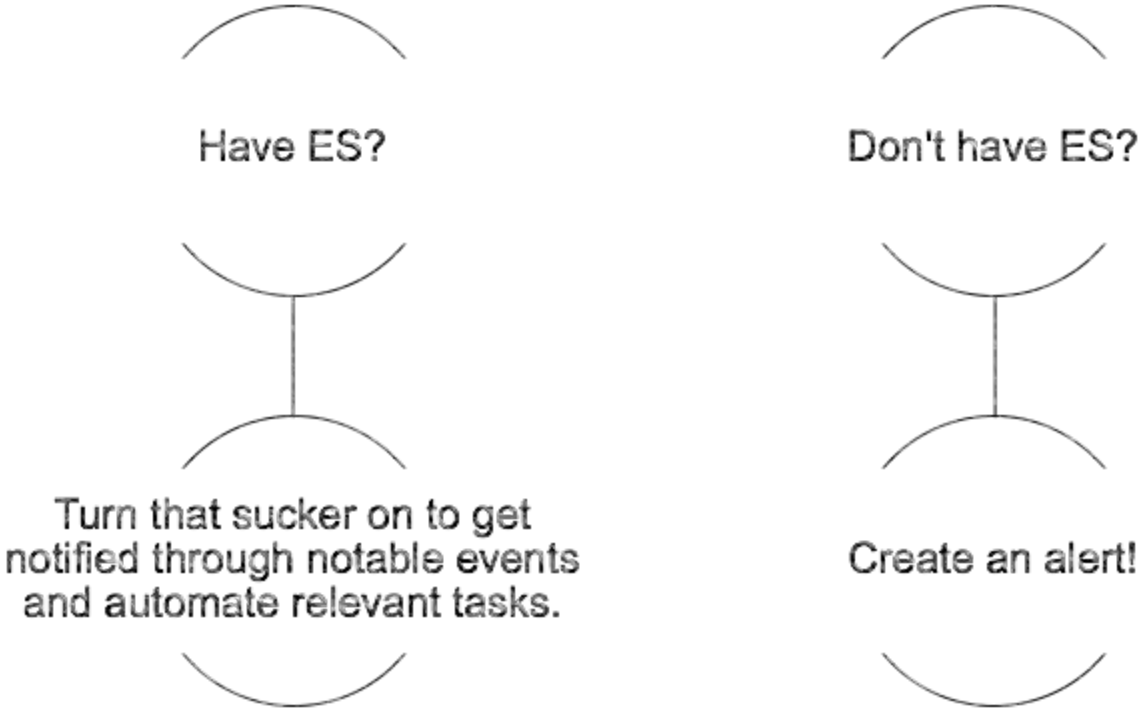


Notice our 38 anomalies

splunk> .conf19

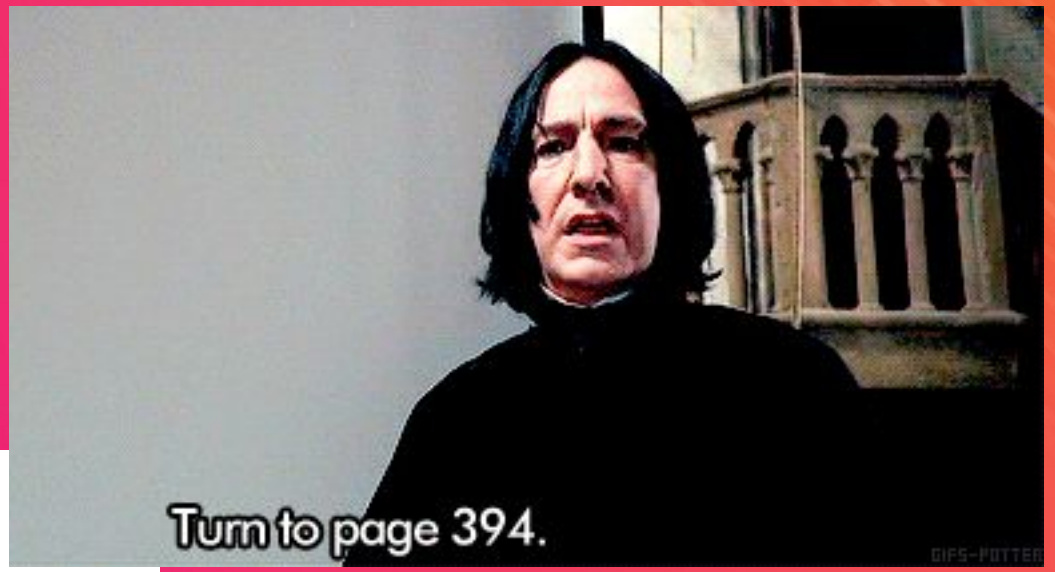
© 2019 SPLUNK INC.

Operationalize It!



© 2019 SPLUNK INC.

Concepts behind ML magic

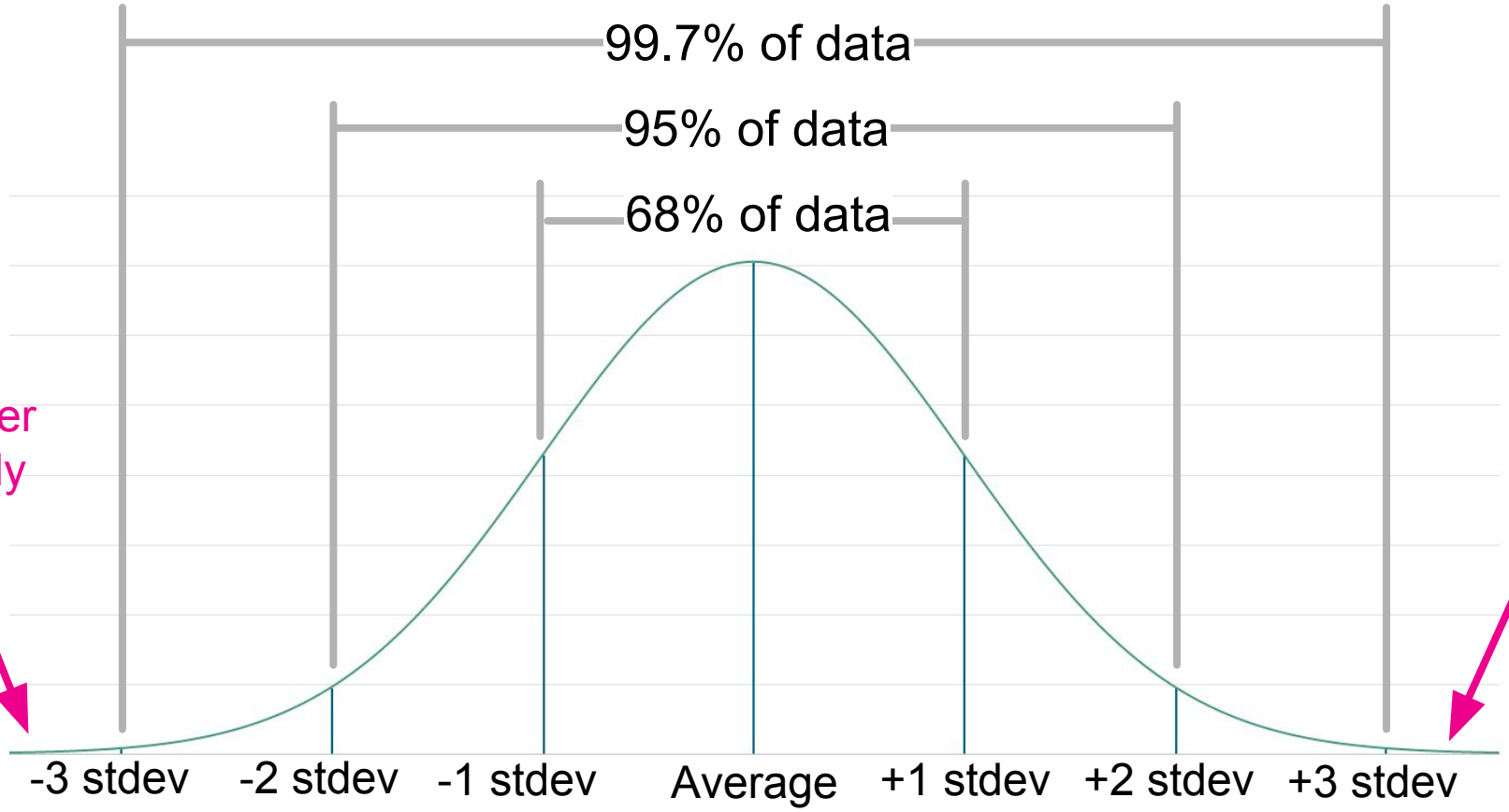


splunk> .conf19

© 2019 SPLUNK INC.

Normal Distribution

How likely is the data found within a population?



Values found over here are probably outliers

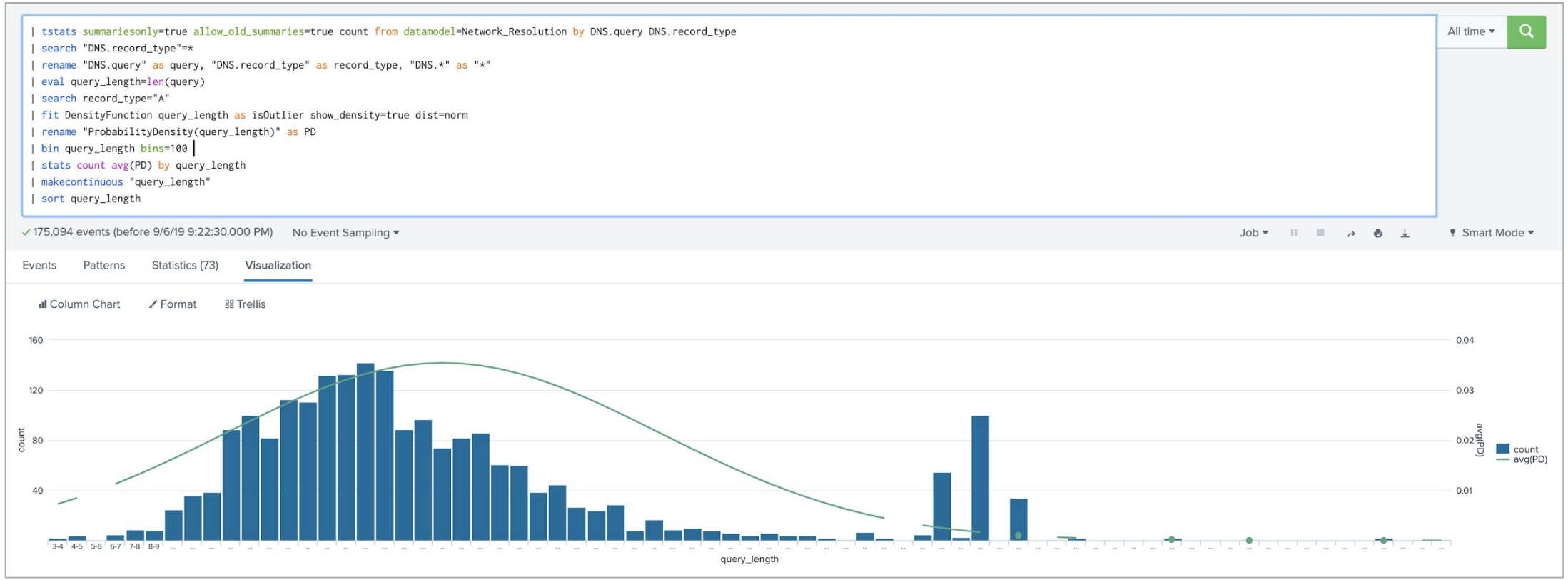
Values found over here are probably outliers

The percentages designate the area under the curve

splunk> .conf19

© 2019 SPLUNK INC.

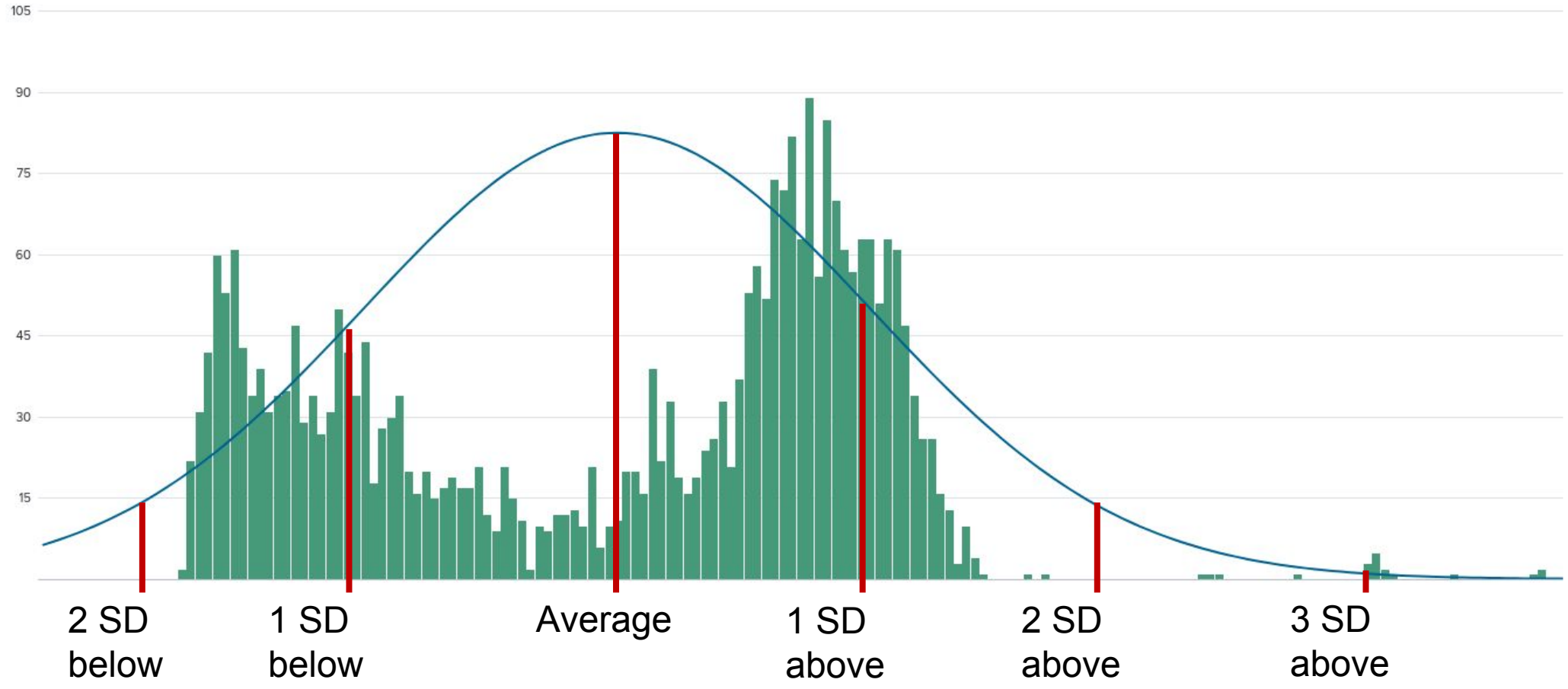
Look What Happens When We Assume Our Data is Normally Distributed...



splunk> .conf19

© 2019 SPLUNK INC.

Your Data May Not Always be So "Normal"

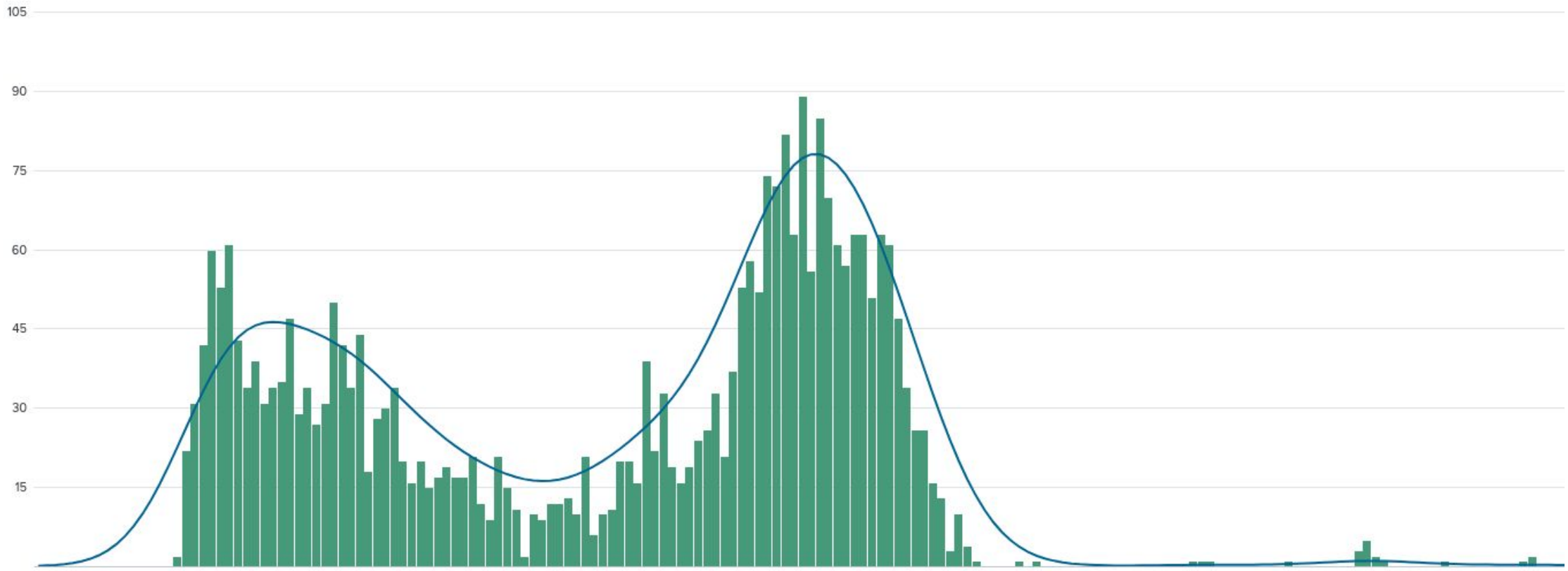


When viewing our data as a histogram, the average may not be so "average"

© 2019 SPLUNK INC.

What If We Could Follow the Shape of Our Data?

We can with the DensityFunction algorithm!



© 2019 SPLUNK INC.

Want to know more about the density function?

Go check out Eurus Kim's .conf19 talk:

The Two Most Common Machine Learning Solutions Everyone Needs to Know

© 2019 SPLUNK INC.

Which is Why ESCU is Upgrading Some Numeric Outlier Detection Use-cases!

The **DNS Query Length with High Standard Deviation** search

turned into

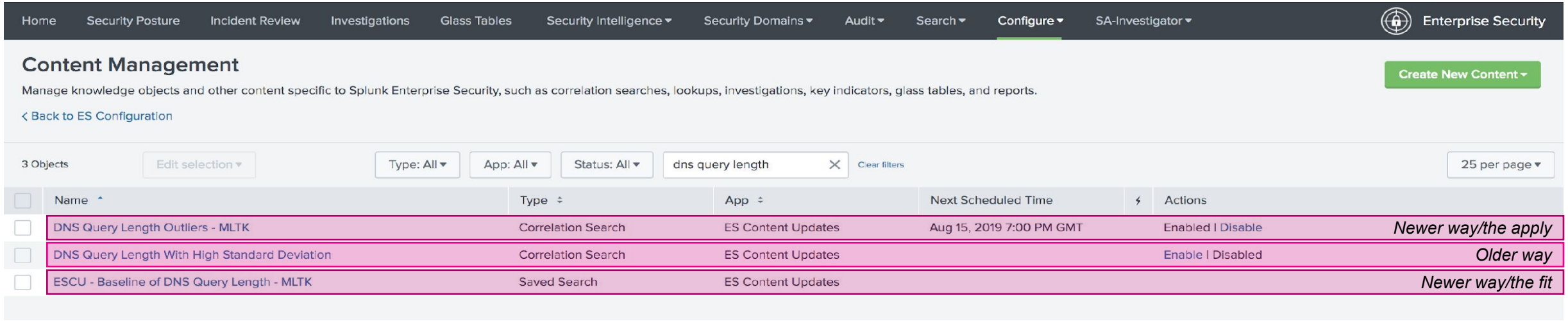
the fit **ESCU - Baseline of DNS Query Length -MLTK** search

+

the apply **DNS Query Length Outliers - MLTK** search

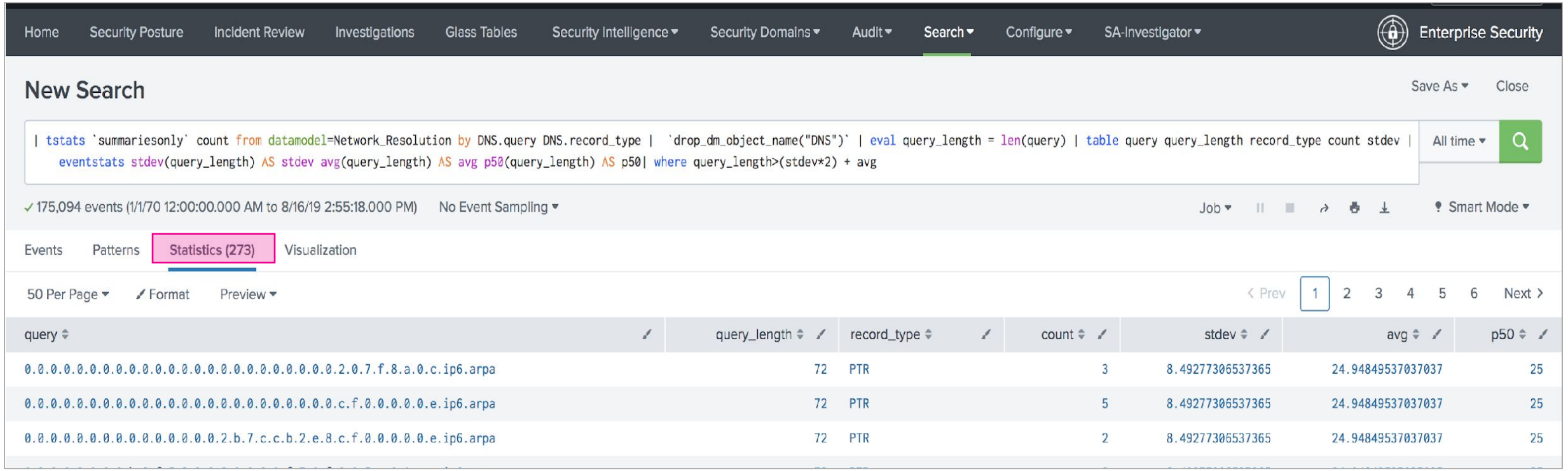
© 2019 SPLUNK INC.

If You Have ES, It Looks Like This



© 2019 SPLUNK INC.

Anomalies Using High Standard Deviation Search

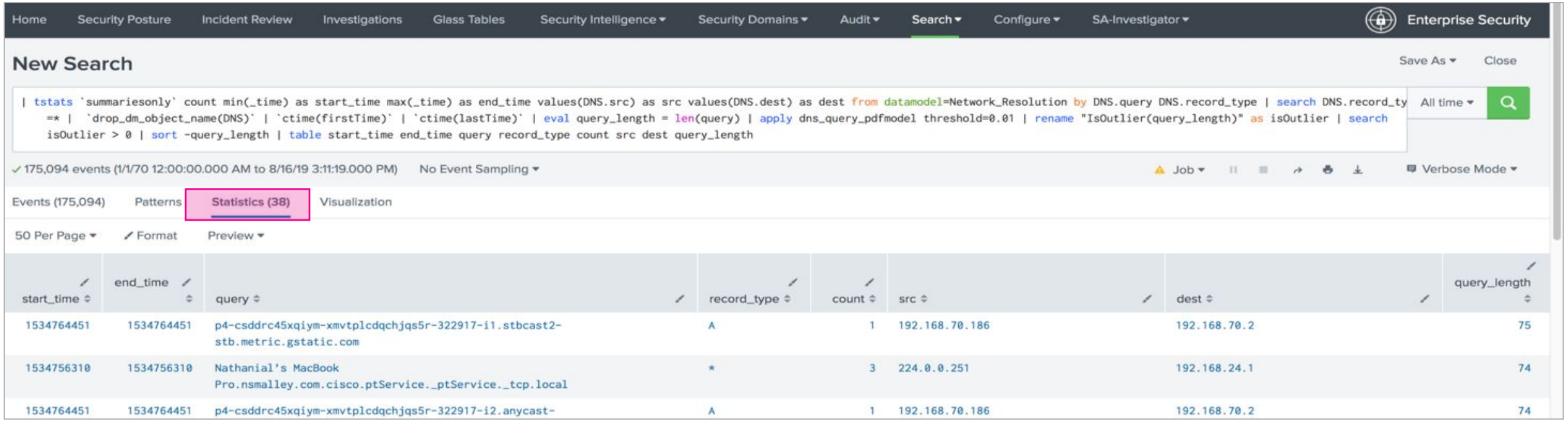


Notice our 273 anomalies

splunk> .conf19

© 2019 SPLUNK INC.

Anomalies Using Density Function



Notice our 38 anomalies

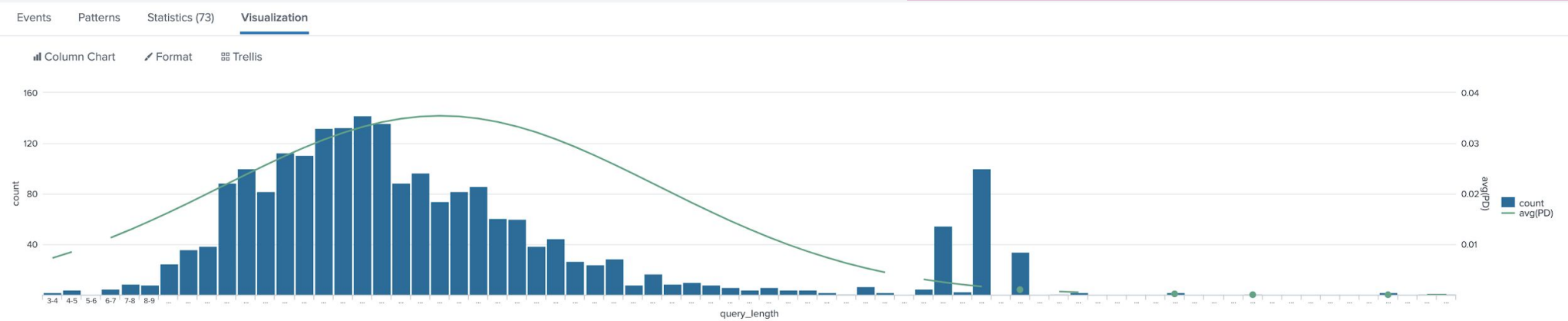
splunk> .conf19

© 2019 SPLUNK INC.

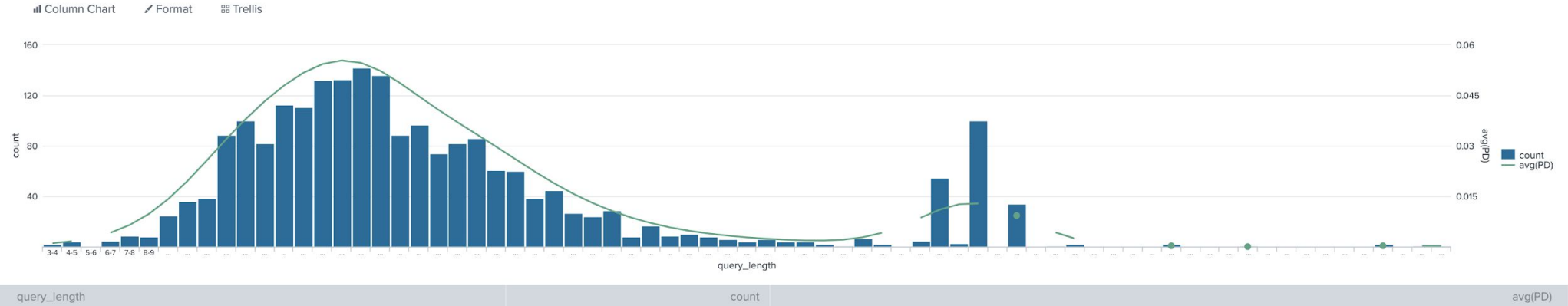
Oculus Reparo!

Fit the model to your data.

Normal curve



KDE curve



splunk> .conf19

© 2019 SPLUNK INC.

Key Takeaways

Hogwarts has served you well!

1. ML can positively impact your security practice!
2. You too can do data science; be a citizen data scientist.
3. Split one search into two with fit and apply concepts to make ML scale.
4. Download ESCU, MLTK and SSE for free!

© 2019 SPLUNK INC.

Special Thanks!

Ryan Kovar
Splunk Principal Security Strategist

David Veuve
Splunk Principal Security Strategist

Rico Valdez
Splunk Principal Security Research Engineer

Eurus Kim
Splunk ML Architect

“Datasciencery by the Splunk Field” DEFCON AI Village presentation.

“New: Machine Learning in Splunk Enterprise Security Content Update” blog.

“The Two Most Common Machine Learning Solutions Everyone Needs to Know”.conf19 session.

© 2019 SPLUNK INC.

Q&A

Melisa Napoles | Splunk Sales Engineer
Erika Strano | Splunk Sales Engineer

splunk> .conf19

© 2019 SPLUNK INC.

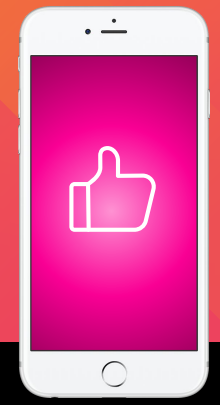
.conf19

splunk>

Thank You !

Go to the .conf19 mobile app to

RATE THIS SESSION



© 2019 SPLUNK INC.

Agenda

1. What is Machine Learning (ML) and why talk about it in Security?
2. How are organizations detecting anomalies today?
3. Example ML Security Detection in ESCU
4. How can you try this at home?
5. Concepts behind anomaly detection