

Defense Against the Dark Arts: Splunk Edition

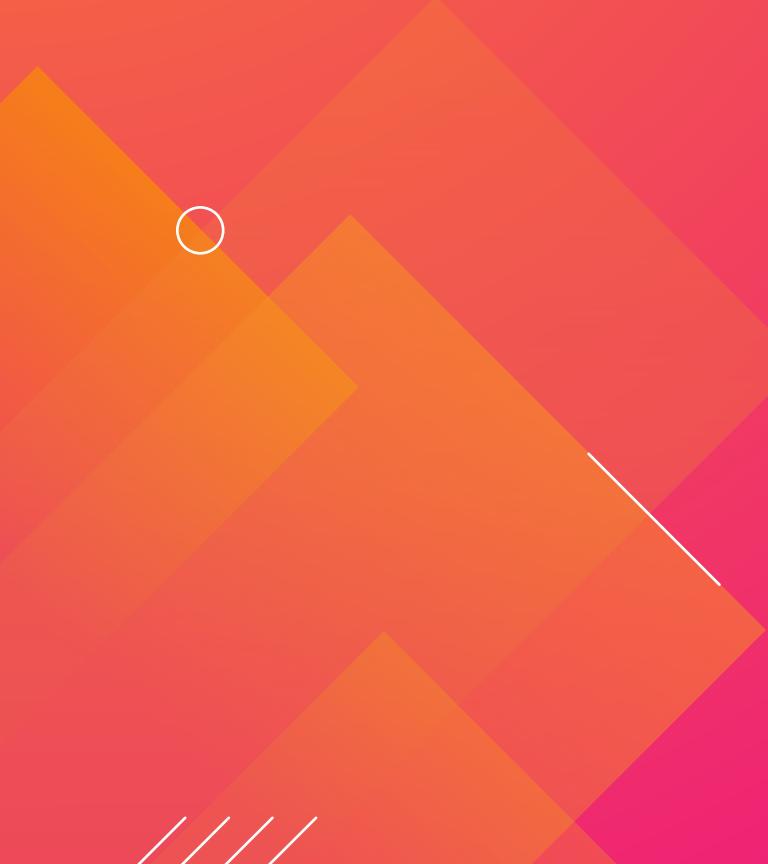
Melisa Napoles | Erika Strano

Wednesday October 23, 2019

.conf19
splunk®>



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 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.

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****

splunk> .conf19



Melisa Napolis
Sales Engineer | Splunk



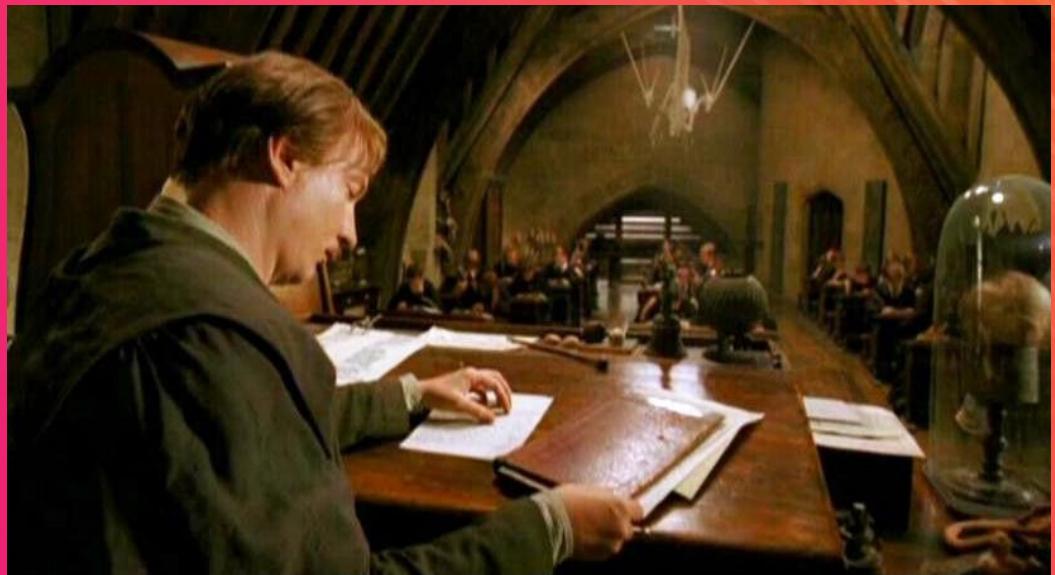
Erika Strano
Sales Engineer | Splunk



You're a wizard, Splunker!



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



Machine Learning can help you ...





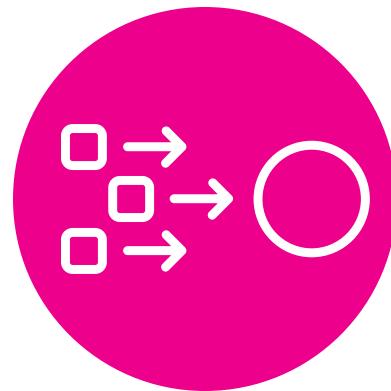


...so you can find your bears!

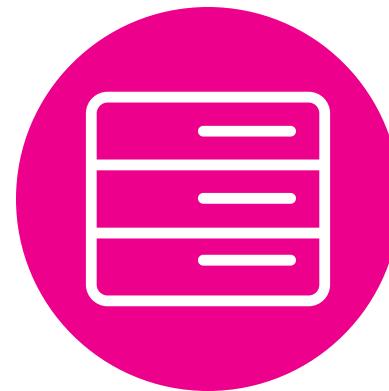


How Does Machine Learning Work?

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/>

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

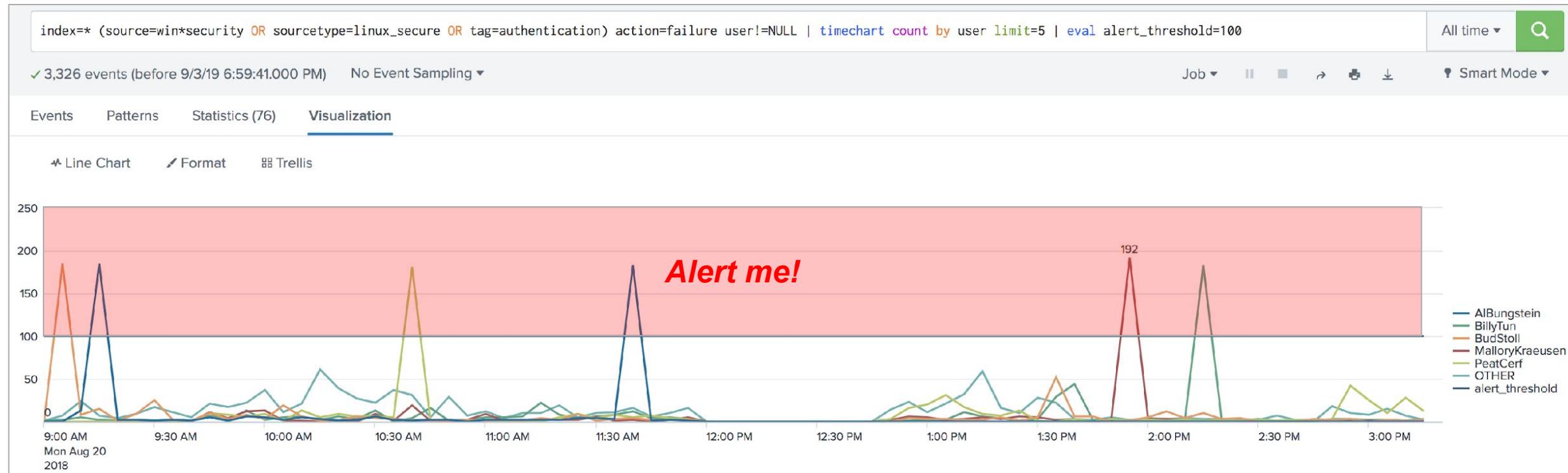


How are we seeing organizations detect the presence of dark wizards?



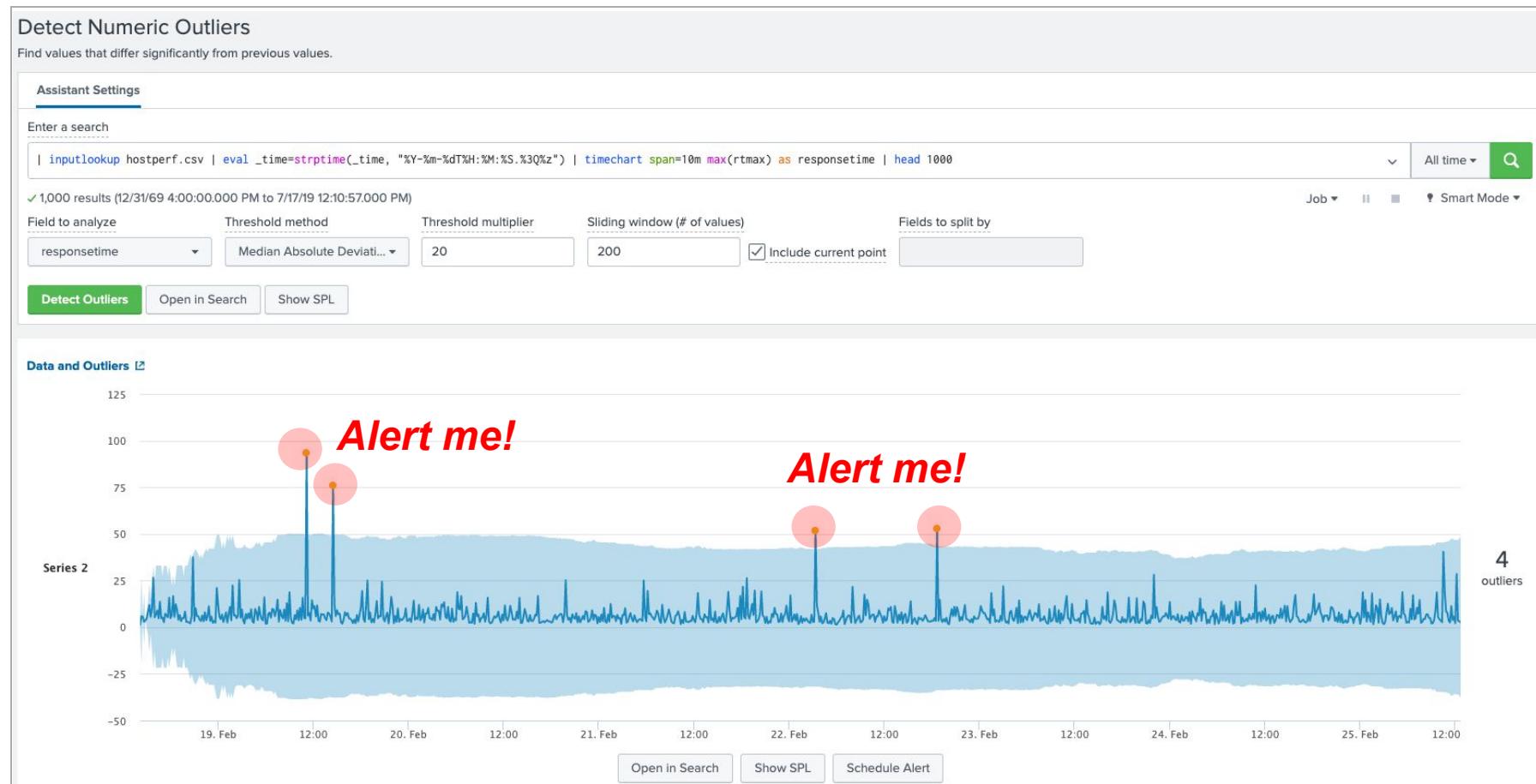
Who Finds Anomalies Like This Today?

One search with a static threshold

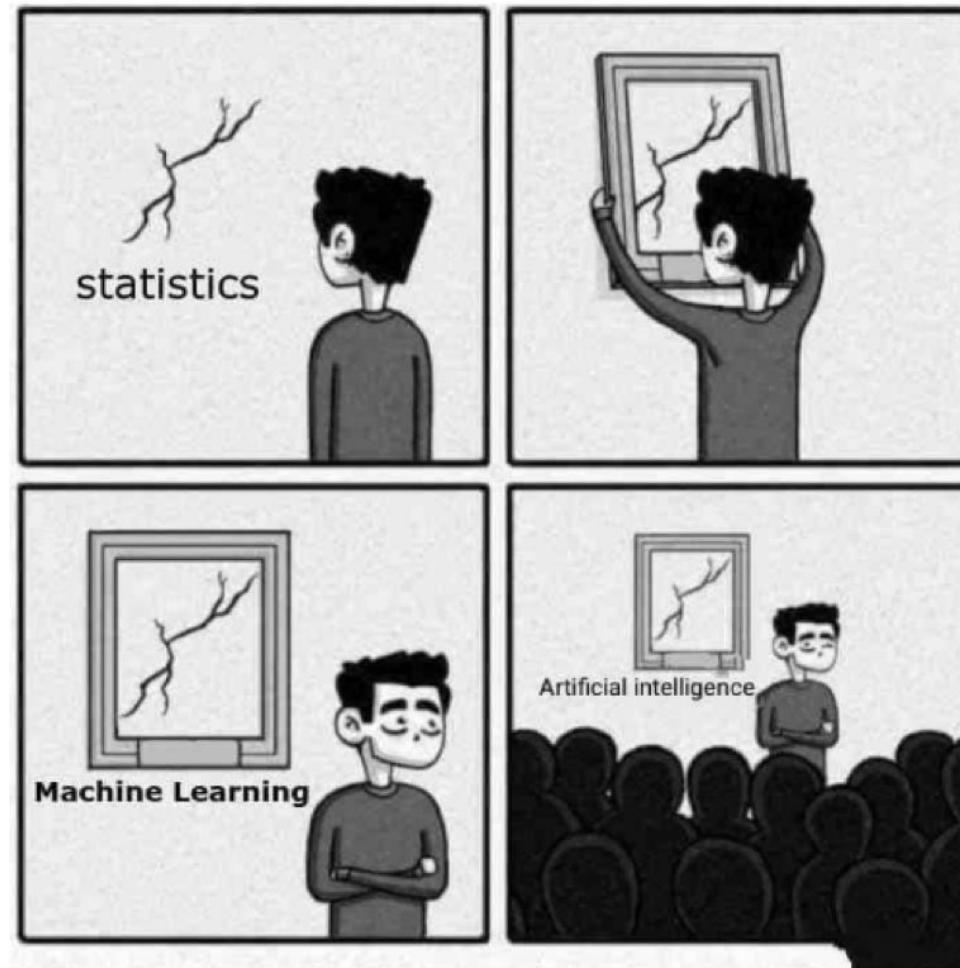


How About Like This?

One search with a dynamic threshold



ML = Stats OR ML != Stats?



original comic by sandserif

splunk> .conf19

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

Thanks Splunk Security Essentials (SSE)!

The screenshot shows the Splunk Security Essentials interface with the following sections:

- Header:** Introduction, Security Content (selected), Security Data Journey, Data Source Check, Documentation, Advanced.
- Top Right:** Splunk Security Essentials logo, What's New In 2.2?, Manage Bookmarks, CSV, ...
- Section Headers:**
 - Stage 1: Collection:** You have the data onboard, what do you do first?
 - Stage 2: Normalization:** You've applied Common Information Model, opening you to detections shared from others, and premium apps.
 - Stage 3: Expansion:** You're ingesting advanced data sources and running better investigations.
- Filter Examples:** Journey (All selected (6)), Security Use Case (All), Category (All), Data Sources (All), Recommended (All), Advanced (Time Series (17 matches...)).
- Search Bar:** Learn how to use this page, Select Filters, 431 Total, 431 Filtered, Clear Filters, Default Filters.
- Content Grid:** 17 use-cases are listed across three stages:
 - Stage 1:** Increase in # of Hosts Logged into, Increase in Pages Printed, Hosts Sending To More Destinations Than Normal, Increase in Windows Privilege Escalations, Significant Increase in Interactive Logons, Significant Increase in Interactively Logged On Users, Spike in SMB Traffic.
 - Stage 2:** Many USB File Copies for User.
 - Stage 3:** AWS Unusual Amount of Modifications to ACLs, User with Increase in Outgoing Email, User with Many DLP Events, AWS APIs Called More Often Than Usual Per User, Increase in Source Code (Git) Downloads, Spike in Downloaded Documents Per User from Salesforce.com, Spike in Exported Records from Salesforce.com, Spike in Password Reset Emails.

Is this scalable?

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

Instead, try...

fit first and then, **apply**

...and make sure your data fits
the model...

not the other way around.

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)

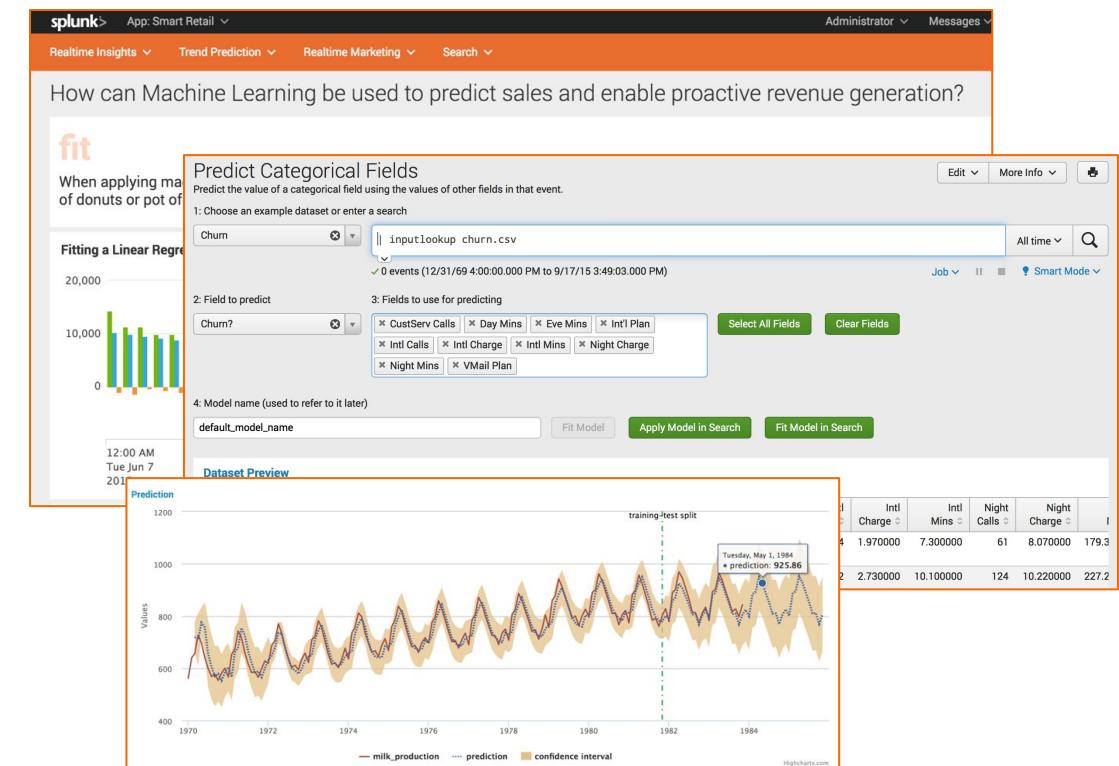
ML Commands: New SPL commands to fit, test, score and operationalize models

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



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

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



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



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

Why Should You Care?

A look at our sample dataset

A lot of DNS query length anomalies

273 anomalies

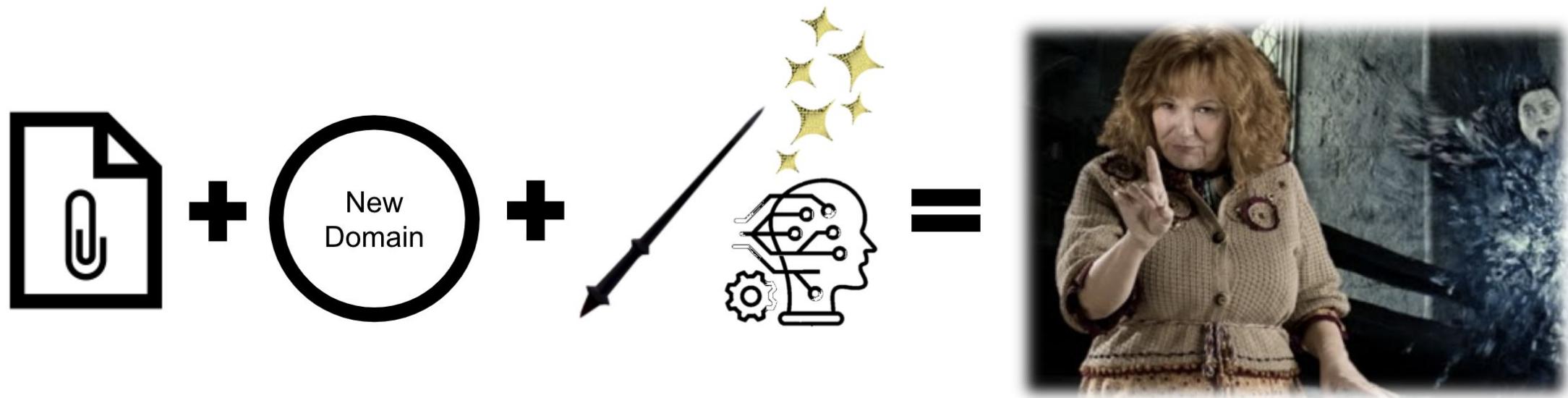
38 anomalies

Static threshold

Standard Deviation

With new ES searches using fit & apply

ML Can Help Add Context to Events





**Close your spell book and
get ready to duel.**



Trusty ESCU Has Us Covered

Exfiltration Use Cases at our Fingertips

ESCU - Baseline of DNS Query Length - MLTK

Configure

Description
This search is used to build a Machine Learning Toolkit (MLTK) model to characterize the length of the DNS queries for each DNS record type observed in the environment. By default, the search uses the last 30 days of data to build the model. The model created by this search is then used in the corresponding detection search, which uses it to identify outliers in the length of the DNS query.

Explain It Like I'm 5
Create a machine-learning (ML) model to characterize the length of DNS requests seen in your environment to help identify unusually long ones that may be indicative of attacker infrastructure or the use of DNS as a command-and-control channel in your environment.

Search

```
| tstats 'summariesonly' count from datamodel=Network_Resolution by DNS.query DNS.record_type | search DNS.record_type=* | drop_dm_object_name("DNS") | eval query_length = len(query) | fit DensityFunction query_length by record_type into dns_query_pdfmodel
```

fit

How to Implement
To successfully implement this search, you will need to ensure that DNS data is populating the Network_Resolution data model. In addition, you must have the Machine Learning Toolkit (MLTK) version >= 4.2 installed, along with any required dependencies. By default, the search builds the model using the past 30 days of data. You can modify the search window to build the model over a longer period of time, which may give you better results. You may also want to periodically re-run this search to rebuild the model with the latest data. More information on the algorithm used in the search can be found at <https://docs.splunk.com/Documentation/MLApp/4.2.0/User/Algorithms#DensityFunction>.

ESCU - DNS Query Length Outliers - MLTK - Rule

Configure

Description
This search allows you to identify DNS requests that are unusually large for the record type being requested in your environment.

Explain It Like I'm 5
Attackers often use random, long domain names for components of their attack infrastructure. This search leverages the probability distribution function algorithm provided by the Machine Learning Toolkit (MLTK) to identify outliers in the length of the DNS query for each record type observed. The companion search "Baseline of DNS Query Length - MLTK" creates a machine-learning (ML) model built over the historical data used by this search. The determination of what is considered an outlier may be adjusted via the threshold parameter in the search. More information on the algorithm used can be found at <https://docs.splunk.com/Documentation/MLApp/4.2.0/User/Algorithms#DensityFunction>.

Search

```
| tstats 'summariesonly' count min(_time) as start_time max(_time) as end_time values(DNS.src) as src values(DNS.dest) as dest from datamodel=Network_Resolution by DNS.query DNS.record_type | search DNS.record_type=* | 'drop_dm_object_name(DNS)' | 'ctime(firstTime)' | 'ctime(lastTime)' | eval query_length = len(query) | apply dns_query_pdfmodel threshold=0.01 | rename "IsOutlier(query_length)" as isOutlier | search isOutlier > 0 | sort -query_length | table start_time end_time query record_type count src dest query_length
```

apply

How to Implement
To successfully implement this search, you will need to ensure that DNS data is populating the Network_Resolution data model. In addition, the Machine Learning Toolkit (MLTK) version 4.2 or greater must be installed on your search heads, along with any required dependencies. Finally, the support search "Baseline of DNS Query Length - MLTK" must be executed before this detection search, because it builds a machine-learning (ML) model over the historical data used by this search. It is important that this search is run in the same app context as the associated support search, so that the model created by the support search is available for use. You should periodically re-run the support search to rebuild the model with the latest data available in your environment.

This search produces fields (`query`, `query_length`, `count`) that are not yet supported by ES Incident Review and therefore cannot be viewed when a notable event is raised. These fields contribute additional context to the notable. To see the additional metadata, add the following fields, if not already present, to Incident Review - Event Attributes (Configure > Incident Management > Incident Review Settings > Add New Entry):

1. Label: DNS Query, Field: query
2. Label: DNS Query Length, Field: query_length
3. Label: Number of events, Field: count

Detailed documentation on how to create a new field within Incident Review may be found here: https://docs.splunk.com/Documentation/ES/5.3.0/Admin/CustomizeTables#Add_a_field_to_the_notable_event_details

Known False Positives
If you are seeing more results than desired, you may consider reducing the value for threshold in the search. You should also periodically re-run the support search to re-build the ML model on the latest data.

ATT&CK
Command and Control | Exfiltration | Commonly Used Port

Kill Chain Phases
Command and Control

CIS Controls
CIS 8 CIS 12

Data Models
Network_Resolution

Technologies
Splunk Stream Bro

Asset at Risk
Endpoint
Confidence medium

Creation Date
2019-05-08

Modification Date
2019-05-08

Creating a Model For Our Dataset

First the fit...

The screenshot shows the Splunk Enterprise Security interface with a search bar containing a query:

```
| tstats `summariesonly` count from datamodel=Network_Resolution by DNS.query DNS.record_type | search DNS.record_type=* | `drop_dm_object_name("DNS")` | eval query_length = len(query) | fit DensityFunction query_length by record_type into dns_query_pdfmodel
```

The search results table displays two rows of data:

query	record_type	count	BoundaryRanges	IsOutlier(query_length)	query_length
0.2.0.7.f.8.a.0.c.ip6.arpa	PTR	3	<pre>[[--Infinity, 17.7748, 0.0032], [19.1559, 20.4054, 0.0011], [30.2703, 70.6505, 0.0054], [73.4126, Infinity, 0.0004]]</pre>	0.0	72
0.c.f.0.0.0.0.e.ip6.arpa	PTR	5	<pre>[[--Infinity, 17.7748, 0.0032], [19.1559, 20.4054, 0.0011], [30.2703, 70.6505, 0.0054], [73.4126, Infinity, 0.0004]]</pre>	0.0	72

Notice we're building a model called “dns_query_pdfmodel”

Notice Your Model is Saved In MLTK!

Bigger, stronger models to scale

The screenshot shows the Splunk Machine Learning Toolkit interface. At the top, there is a navigation bar with links: Showcase, Experiments, Search, Models (which is highlighted in orange), Classic ▾, Settings, Docs ▾, and Video Tutorials ▾. To the right of the navigation bar is the Splunk logo and the text "Splunk Machine Learning Toolkit". Below the navigation bar, the page title "Models" is displayed. A sub-header indicates "2 Models". There are filter buttons for "All", "Yours", and "This App's", and a search bar labeled "Filter by model name" with a magnifying glass icon. The main content area is a table with the following data:

i	Model Name	Algorithm	Actions	Owner	App	Sharing
>	dns_query_pdfmodel	DensityFunction	Delete	admin	Splunk_ML_Toolkit	Private
>	example_app_usage	LinearRegression	Delete	admin	Splunk_ML_Toolkit	Private

Applying the Model to Our Dataset

...then the apply!

New Search

Save As ▾ Close

```
| tstats `summariesonly` count min(_time) as start_time max(_time) as end_time values(DNS.src) as src values(DNS.dest) as dest from datamodel=Network_Resolution by DNS.query DNS.record_type | search DNS.record_ty
 *=* | `drop_dm_object_name(DNS)` | `ctime(firstTime)` | `ctime(lastTime)` | eval query_length = len(query) | apply dns_query_pdfmodel threshold=0.01 | rename "IsOutlier(query_length)" as isOutlier | search
 isOutlier > 0 | sort -query_length | table start_time end_time query record_type count src dest query_length
```

✓ 175,094 events (1/1/70 12:00:00.000 AM to 8/16/19 3:11:19.000 PM) No Event Sampling ▾

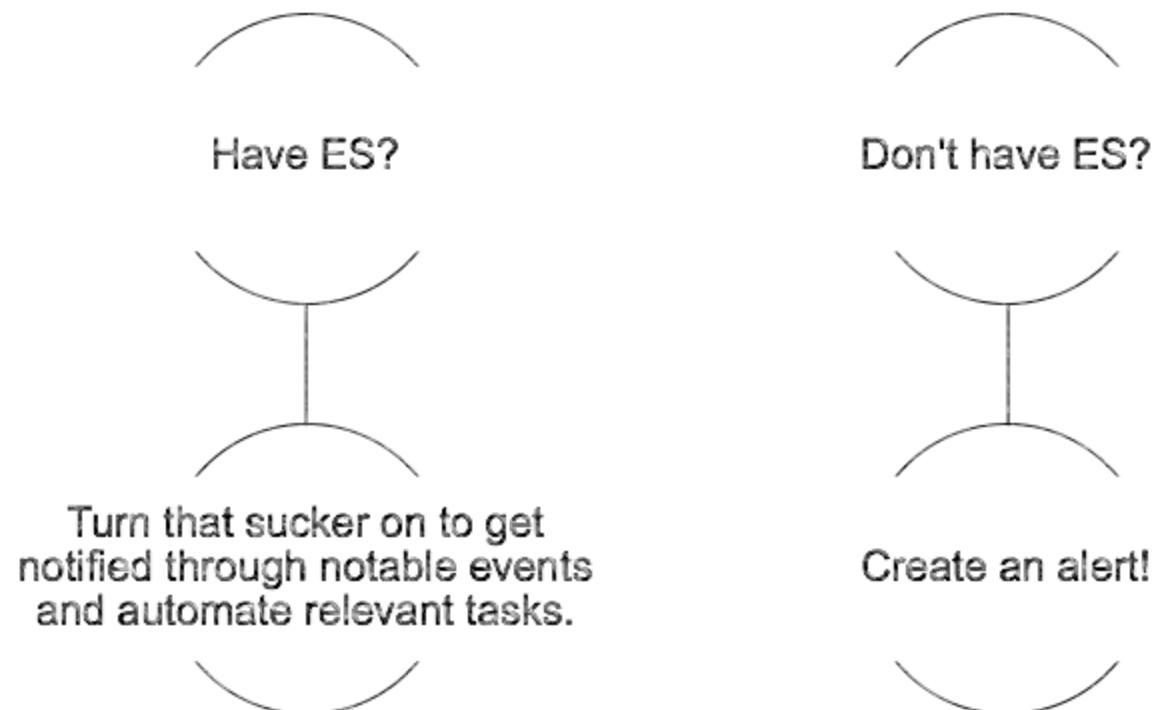
Events (175,094) Patterns Statistics (38) Visualization

50 Per Page ▾ Format Preview ▾

start_time	end_time	query	record_type	count	src	dest	query_length
1534764451	1534764451	p4-csddrc45xqiym-xmvtp lcdqchjqs5r-322917-i1.stbcast2-stb.metric.gstatic.com	A	1	192.168.70.186	192.168.70.2	75
1534756310	1534756310	Nathanial's MacBook Pro.nsmalley.com.cisco.ptService._ptService._tcp.local	*	3	224.0.0.251	192.168.24.1	74
1534764451	1534764451	p4-csddrc45xqiym-xmvtp lcdqchjqs5r-322917-i2.stbcast2-	A	1	192.168.70.186	192.168.70.2	74

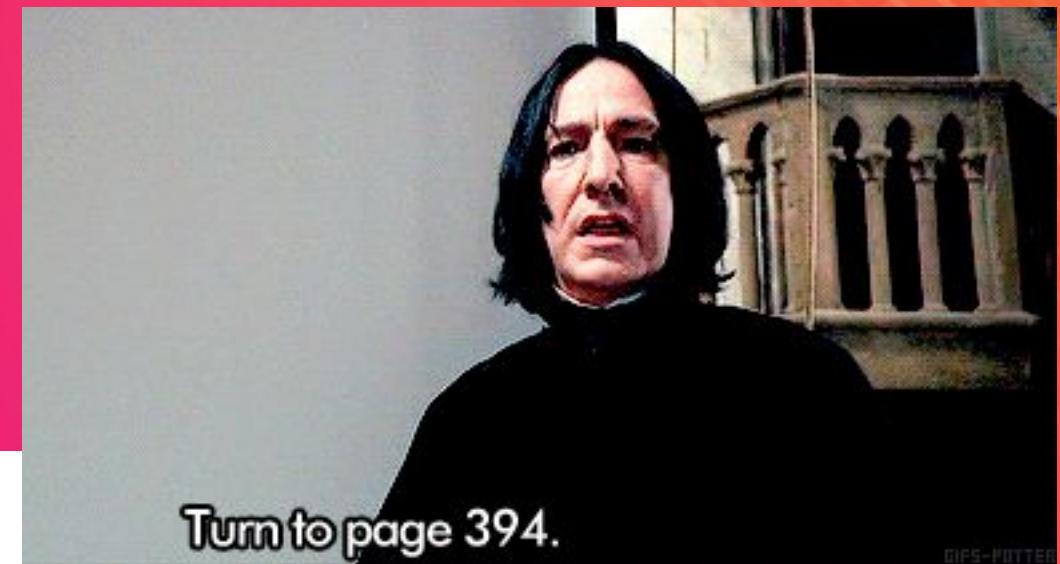
Notice our 38 anomalies

Operationalize It!



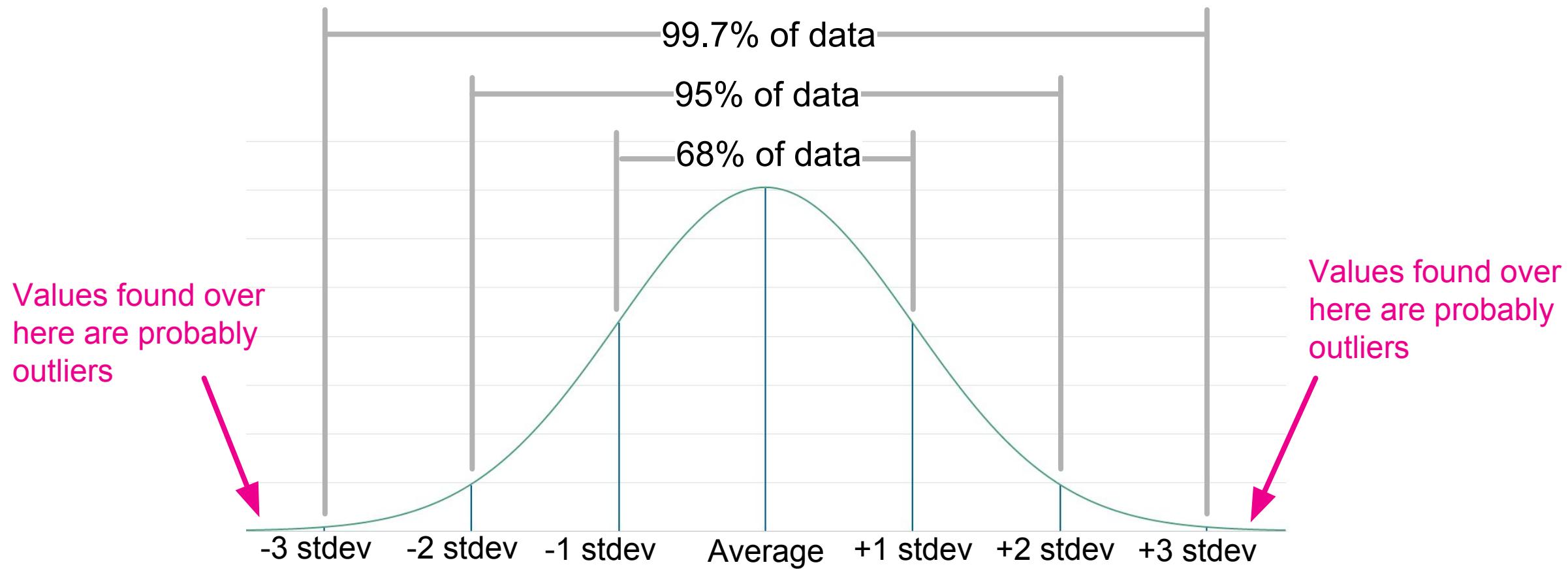


Concepts behind ML magic



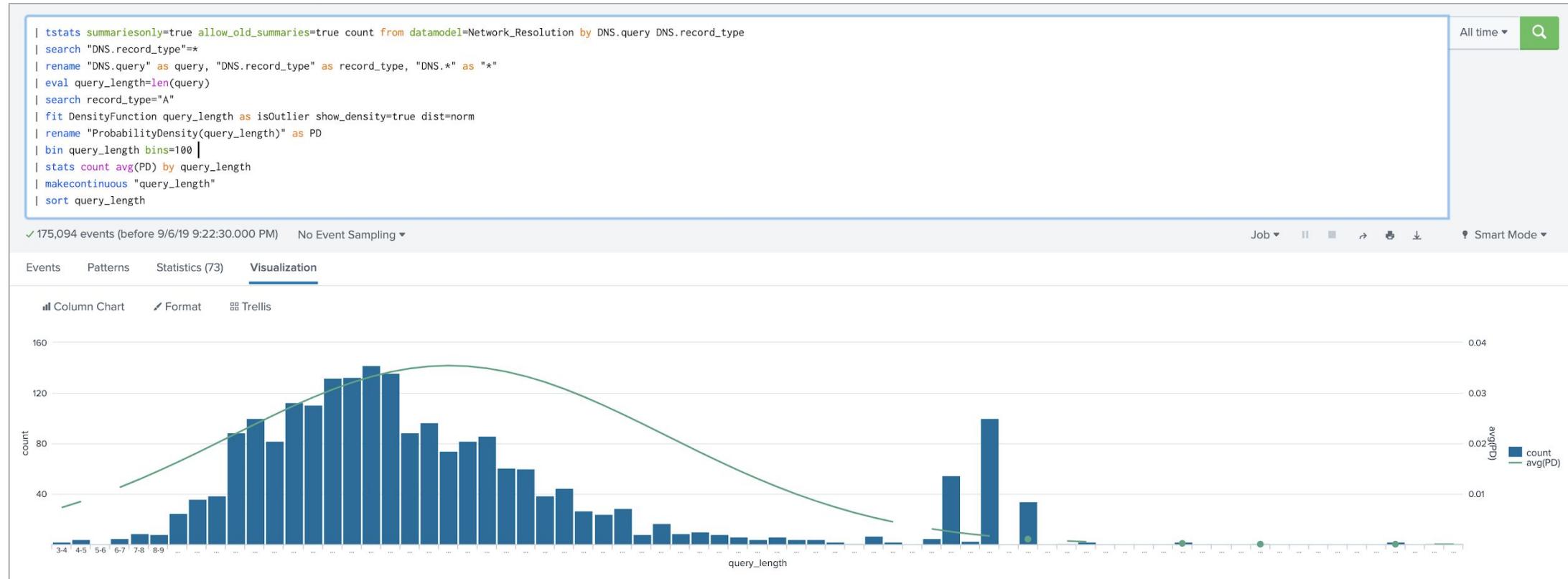
Normal Distribution

How likely is the data found within a population?

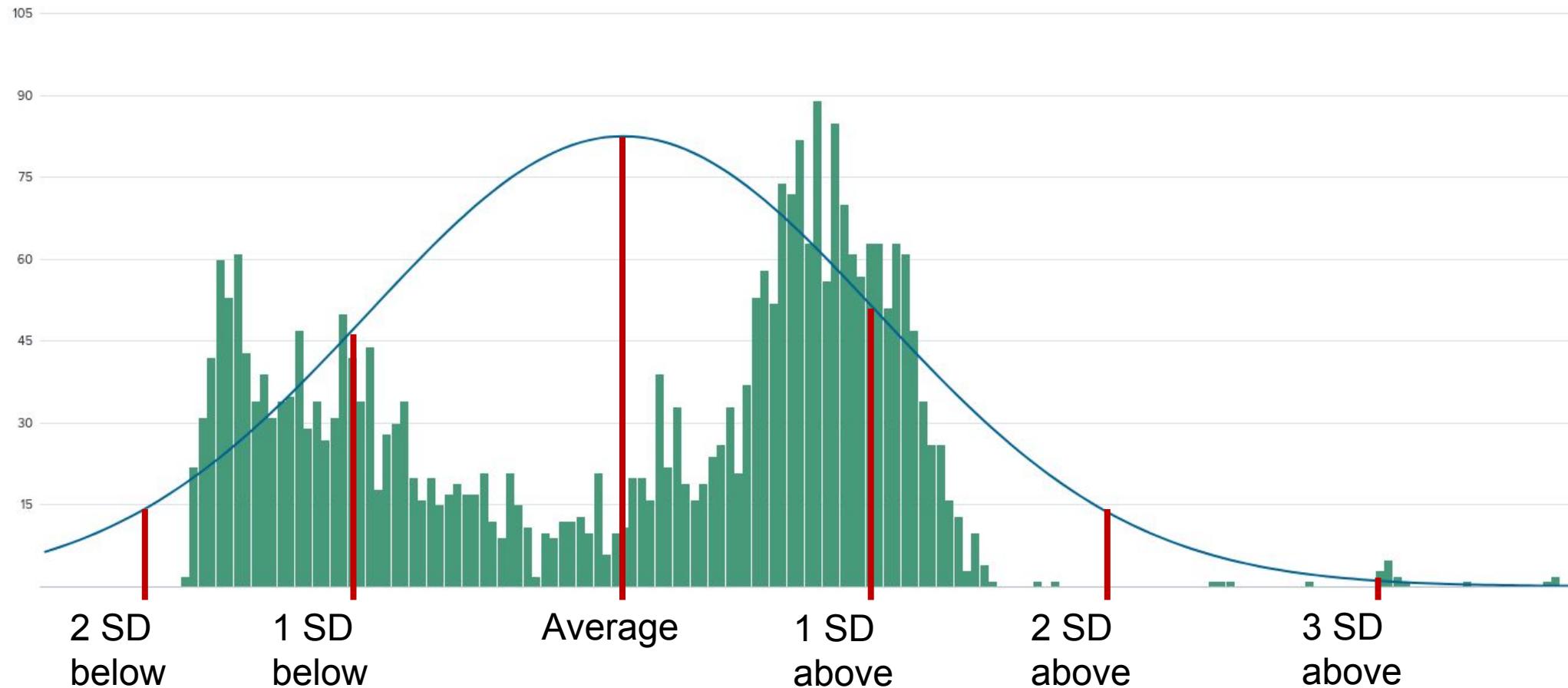


The percentages designate the area under the curve

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



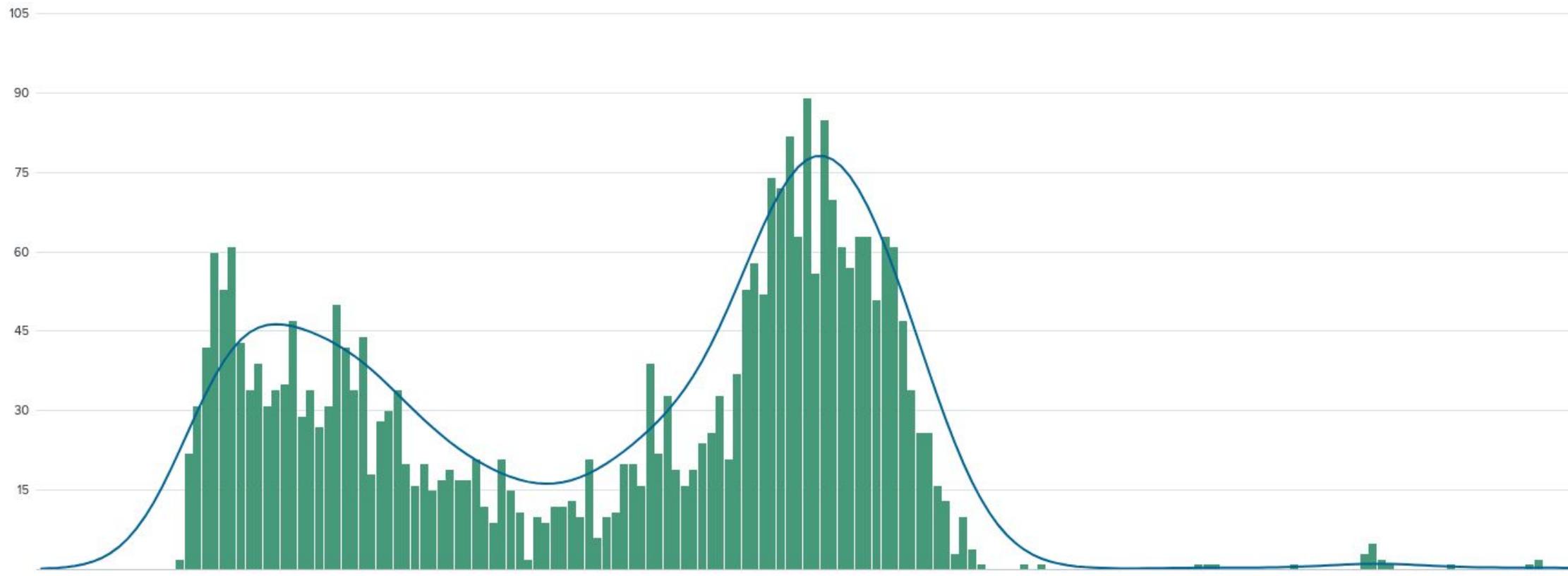
Your Data May Not Always be So “Normal”



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

What If We Could Follow the Shape of Our Data?

We can with the DensityFunction algorithm!



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

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

If You Have ES, It Looks Like This

Home Security Posture Incident Review Investigations Glass Tables Security Intelligence ▾ Security Domains ▾ Audit ▾ Search ▾ Configure ▾ SA-Investigator ▾  Enterprise Security

Content Management

Manage knowledge objects and other content specific to Splunk Enterprise Security, such as correlation searches, lookups, investigations, key indicators, glass tables, and reports.

[Create New Content ▾](#)

[◀ Back to ES Configuration](#)

3 Objects					Edit selection ▾	Type: All ▾	App: All ▾	Status: All ▾	dns query length 	Clear filters	25 per page ▾
<input type="checkbox"/>	Name ▾	Type ▾	App ▾	Next Scheduled Time	Actions						
<input type="checkbox"/>	DNS Query Length Outliers - MLTK	Correlation Search	ES Content Updates	Aug 15, 2019 7:00 PM GMT	Enabled Disable						Newer way/the apply
<input type="checkbox"/>	DNS Query Length With High Standard Deviation	Correlation Search	ES Content Updates		Enable Disabled						Older way
<input type="checkbox"/>	ESCU - Baseline of DNS Query Length - MLTK	Saved Search	ES Content Updates								Newer way/the fit

Anomalies Using High Standard Deviation Search

Screenshot of Splunk Enterprise Security interface showing a search results page for anomalies using high standard deviation search.

Search Bar:

```
| tstats `summariesonly` count from datamodel=Network_Resolution by DNS.query DNS.record_type | `drop_dm_object_name("DNS")` | eval query_length = len(query) | table query query_length record_type count stdev | eventstats stdev(query_length) AS stdev avg(query_length) AS avg p50(query_length) AS p50| where query_length > (stdev*2) + avg
```

Search Results:

- 175,094 events (1/1/70 12:00:00.000 AM to 8/16/19 2:55:18.000 PM)
- No Event Sampling
- Job ▾
- Smart Mode ▾

Statistics (273) Tab:

query	query_length	record_type	count	stdev	avg	p50
0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.2.0.7.f.8.a.0.c.ip6.arpa	72	PTR	3	8.49277306537365	24.94849537037037	25
0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.0.c.f.0.0.0.0.e.ip6.arpa	72	PTR	5	8.49277306537365	24.94849537037037	25
0.0.0.0.0.0.0.0.0.0.0.2.b.7.c.c.b.2.e.8.c.f.0.0.0.0.e.ip6.arpa	72	PTR	2	8.49277306537365	24.94849537037037	25

Notice our 273 anomalies

Anomalies Using Density Function

Screenshot of Splunk Enterprise Security interface showing a search results page for anomalies using the Density function.

Search Bar:

```
| tstats `summariesonly` count min(_time) as start_time max(_time) as end_time values(DNS.src) as src values(DNS.dest) as dest from datamodel=Network_Resolution by DNS.query DNS.record_type | search DNS.record_type !=* | `drop_dm_object_name(DNS)` | `ctime(firstTime)` | `ctime(lastTime)` | eval query_length = len(query) | apply dns_query_pdfmodel threshold=0.01 | rename "IsOutlier(query_length)" as isOutlier | search isOutlier > 0 | sort -query_length | table start_time end_time query record_type count src dest query_length
```

Statistics: 175,094 events (1/1/70 12:00:00.000 AM to 8/16/19 3:11:19.000 PM) No Event Sampling

Table: Statistics (38) rows

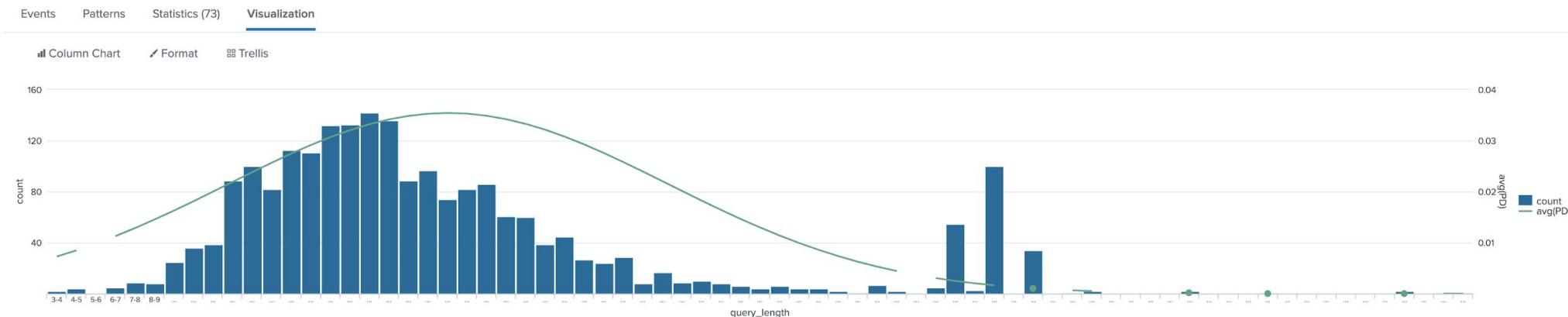
start_time	end_time	query	record_type	count	src	dest	query_length
1534764451	1534764451	p4-csddrc45xqiy-m-xmvtplcdqchjqs5r-322917-i1.stbcast2-stb.metric.gstatic.com	A	1	192.168.70.186	192.168.70.2	75
1534756310	1534756310	Nathaniel's MacBook Pro.nsmalley.com.cisco.ptService._ptService._tcp.local	*	3	224.0.0.251	192.168.24.1	74
1534764451	1534764451	p4-csddrc45xqiy-m-xmvtplcdqchjqs5r-322917-i2.anycast-	A	1	192.168.70.186	192.168.70.2	74

Notice our 38 anomalies

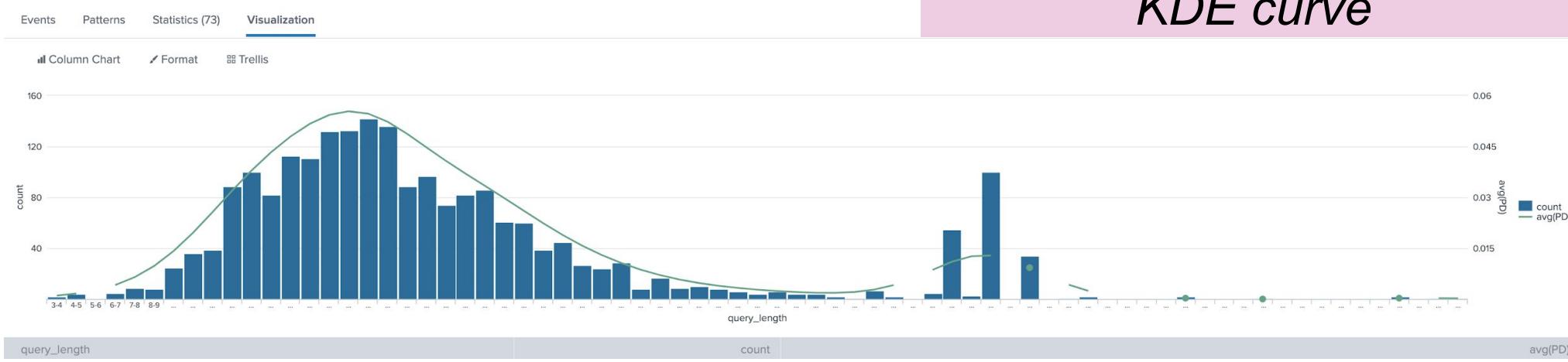
Oculus Reparo!

Fit the model to your data.

Normal curve



KDE curve



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!

Special Thanks!



Splunk Principal
Security Strategist



Splunk Principal
Security Strategist



Splunk Principal
Security Research
Engineer



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.



Q&A

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

.conf19

splunk>

Thank

You

!

Go to the .conf19 mobile app to

RATE THIS SESSION



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