Measuring What Matters to Streamline Security Operations with Splunk

Keshia Levan
Detection Engineering Lead | Red Canary

splunk> .conf19

Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance 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 could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

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, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward-looking statements we may make. 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 functionality described or to include any such feature or functionality in a future release.

A Few Words About Me

- Detection Engineer / Site Reliability Engineer @ Red Canary
- Red Canary Does Managed Endpoint Detection & Response (EDR)
- Tries to make useful content from logs and metrics
- Messes with Code; Interested in Data Science; Formally Trained in Neither
- Is Convinced You Can Use Splunk to Answer Any Questions if Creative Enough (and you actually have the data…)

A Brief Word On Common Terminology

- Detector / Signature / Alert – The Logic You Use To Detect Behavior

ruby_code

```
not_trusted?
&&

(process_path_downcased.include?('appdata')
||

process_path_downcased.include?('application data'))
&&

childproc_name_downcased == 'sh.exe'
```

- We focus mostly on Endpoint Data (Process Starts, Filemods etc.)

Noise – It's A Common Problem

- Humans Can Only Do So Much
- Repetitive Tasks Cause Fatigue and Mistakes
- Low Fidelity High Volume Alerts Contribute
- Alerts that provide no value and take a long time – Worst
- Short term, high value aka ("evil") – Best
- Reality - much more in the middle

Lack of Data & Context:

- Point in Time View of Alerts
  - Csv
  - Dashboard
- We tracked how effective a detector was, but only as a point in time
- No way to drill down into
  - Which host
  - Which process
  - FP vs TP characteristics

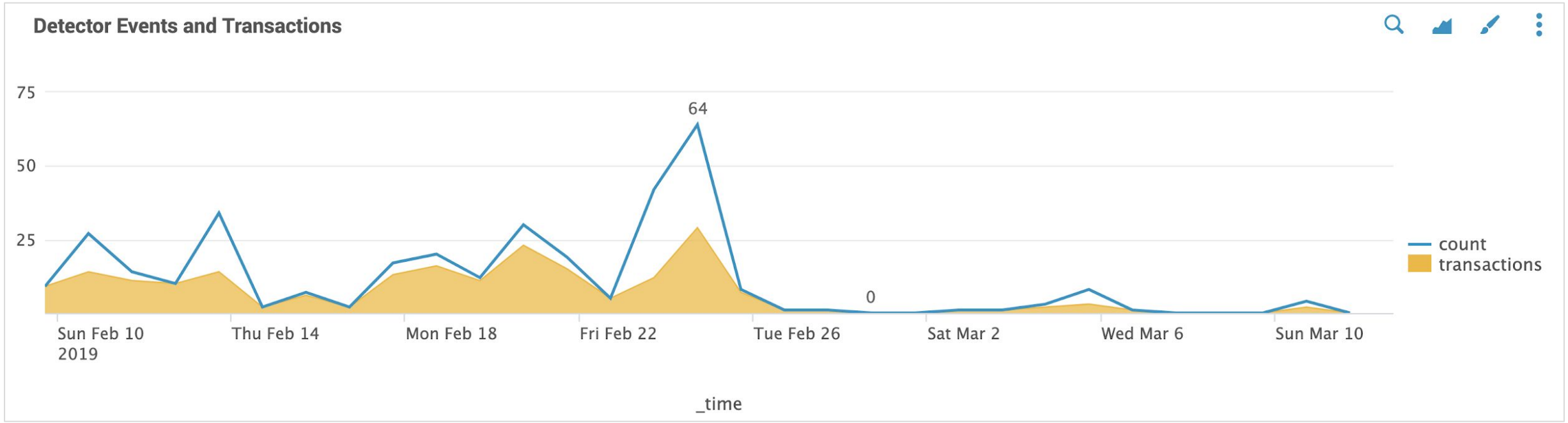
No concept of effort and time, no counts

Expected Impact from 2017-12-15 — Reset

Detection Classification	Conversion Rate	Classification Impact	Expected Impact
Malicious Software	5.556%	15.0	0.833
Malicious Software / Credential Theft	2.381%		0.000
Malicious Software / Hacking Tool	3.175%		0.000
Suspicious Activity	7.143%	10.0	0.714
Suspicious Activity / Network	6.349%		0.000
Suspicious Activity / Process	7.143%		0.000
		Total:	**1.548**

Data Was Needed:

- To Provide the ability to do Historical Analysis
- To Identify impact from changes



Content Overview:

This is a short talk, as such instead of focusing on the entire project will be focusing on small pieces of putting this together that were challenging and might help someone else.

- Summarizing Data
- Storing The Calculated Data
- Example Analytics on the Data

High Level Questions

Where is my team spending the most effort?

- Am I spending more time on detected events or false positives?
- Of the False Positives – Which one's take the most effort and time?
- Which set of my high fidelity signatures are close enough to perfect that can maybe automate them away?
- Are spikes skewing my analysis, methods to normalize data? (Why counts and averages can lie)

Per Alert Questions:

Which of my individual rules are terrible?

- How much time did I spend on Rule 1077 last month?
- How many of Rule 1077's events converted to actual incidents?

datestamp	detector_id	event_count	detected_count	not_detected_count	human_transaction_time	human_transaction_time_detected
2019-03-23T12:00Z	790	965	0	965	3758.59343771	0
2019-03-22T12:00Z	790	8087	0	8087	99078.743190333	0
2019-03-19T12:00Z	790	106	13	93	8712.67749321	2202.1169041
2019-03-16T12:00Z	790	33	0	33	953.83439033	0
2019-03-20T12:00Z	790	45	7	38	3933.0879424	1835.7224285
2019-03-15T12:00Z	790	54	1	53	1501.61993620	27.924000
2019-03-21T12:00Z	790	76	9	67	6731.11095707	2486.6014997
2019-03-24T12:00Z	790	2633	1	2632	10297.28082905	145.256000
2019-03-18T12:00Z	790	40	4	36	2847.63147319	287.42864019
2019-03-17T12:00Z	790	6	0	6	608.173500	0

Reducing Data

- Logs are voluminous but revealing
- Metrics are essentially just numbers
  - Great for trends
  - Bad for context
- Data reduction expedites searches and queries and helps with retention

What data do you want to keep?

- You may want to keep data not strictly needed for metrics
- What questions might you have after identifying a seemingly problematic rule or group of rules
- Metrics are essentially just numbers
  - Severity of incident related to alert
  - Environments that tripped alert
  - Hosts online and active during the timeframe pulled
  - Unique incidents per rule or day

Our Metrics

- True positive/false positive percentages per detector
- Time expenditure on transactions by detection rule
  - Confirmed detected events
  - Confirmed false positive events
- Conversion percentage per detection rule

Note: *Transaction= Grouping of events on a given endpoint at a point in time. We record the action taken on that grouping.*

Summarizing Data

.conf19
splunk>

Example: Calculating Time Metric

- For each event calculate the duration that it was worked:

```
| eval t_claim=strptime(claimed_at, "%Y-%m-%dT%H:%M:%S.%3NZ")
| eval t_term=strptime(terminal_at, "%Y-%m-%dT%H:%M:%S.%3NZ")
| eval t_duration=t_term-t_claim
```

- For each event add a field, so we can separate humans from automation
  - Analyst: analyst10@redcanary.com = human
  - Analyst: bot50@redcanary = bot

```
| eval freewill=if(like(analyst, "%noreply@redcanary%"), "bot", "human")
```

Example: Creating a Daily Sum

- Decide the smallest unit you might want to chart – Hours? Days?

```
| eval datestamp=(_time)
| convert timeformat="%Y-%m-%dT12:00Z" ctime(datestamp)
| eventstats count as total_events_all_detectors by date
```

- Sum time spent for each signature

```
| eventstats sum(eval(if(freewill="human", fractional_duration, 0))) AS human_transaction_time by detector_id,datestamp
```

- We will end up with one row per signature per day

datestamp ^	detector_id	total_time	human_per_signature	total_true_positive_human
2019-08-27T12:00Z	1095	216201.301743205	1846.177000	107526.54995846
2019-08-27T12:00Z	1089	216201.301743205	17.465000	107526.54995846

Answers: How much time did I spend on Rule XX last month?

New Search

Save As | New Table | Close

```
index=cirt_summary_index | stats sum(human_transaction_time) AS total_time by detector_id | eval
    time_in_hours=(total_time/3600) | table detector_id time_in_hours
```

Last 30 days

✓ 6,966 events (8/5/19 12:00:00.000 AM to 9/4/19 1:12:42.000 PM) No Event Sampling | Job | Fast Mode

Events | Patterns | Statistics (522) | Visualization

100 Per Page | Format | Preview

< Prev 1 2 3 4 5 6 Next >

detector_id	time_in_hours
1938	2.98079165294
1939	0.0251108333
1943	7.29522484461
1944	0.87152268519
1945	1.63469717476
1951	0.04423388889
1952	0.1830170835
1956	2.79656737100

You'll repeat this for as many data points as you care about - essentially summarize or enrich the things you want to keep and ask questions on.

```
    human_hidden_transactions by detector_id, datestamp
| eventstats dc(eval(if(freewill="human" AND terminal_stage="detected", transaction_id, NULL))) AS
    human_detected_transactions,dc(eval(if(freewill="human" AND terminal_stage="detected" AND
    'associated_detections{}.classification'="Malicious Software", transaction_id, NULL))) AS
    human_malicious_transactions, dc(eval(if(freewill="human" AND terminal_stage="detected" AND
    'associated_detections{}.classification'="Suspicious Activity", transaction_id, NULL))) AS
    human_suspicious_transactions, dc(eval(if(freewill="human" AND terminal_stage="detected" AND
    'associated_detections{}.classification'="Unwanted Software", transaction_id, NULL))) AS
    human_unwanted_transactions by detector_id,datestamp
```

Storing the Calculated Data

.conf19
splunk>

How do I keep data past normal retention?

- I can run the previous calculations every time I want to ask a question.
  - It will take a really long time each time I want to change something
  - I will only be able to run across maybe 30 days, data is constantly rolling off
- Simple Answer - Use a lookup table:

```
    unique_detections_suspicious, unique_detections_unwanted, unique_endpoints, uni
| dedup datestamp, detector_id | outputlookup append=t transaction_aggregate.csv
```

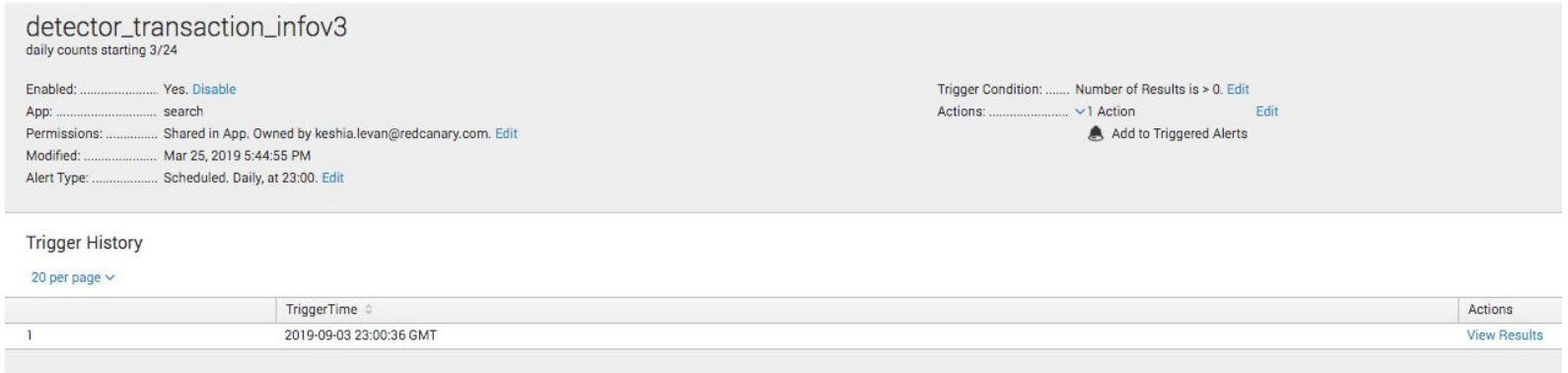
Lookup Limitations:

- Row Limitations
- Time in Dashboards Can Be “Difficult”

```
    <panel>
      <table>
        <search>
          <query>| inputlookup event_aggregate.csv
| eval _time=strptime(_time,"%Y-%m-%dT%H:%MZ")
| sort - _time
| addinfo
| where _time&gt;=info_min_time AND (_time&lt;=info_max_time OR info_max_time="+Infinity")
| eventstats sum(daily_human_detected_count) as human_total_detected  sum(eval(daily_human_det
```

Lookup Table: Alerts Work Well

- Automate Collection of Data – no one will remember to run a search daily, monthly (or hourly…)
  - List in triggered alerts is a good option

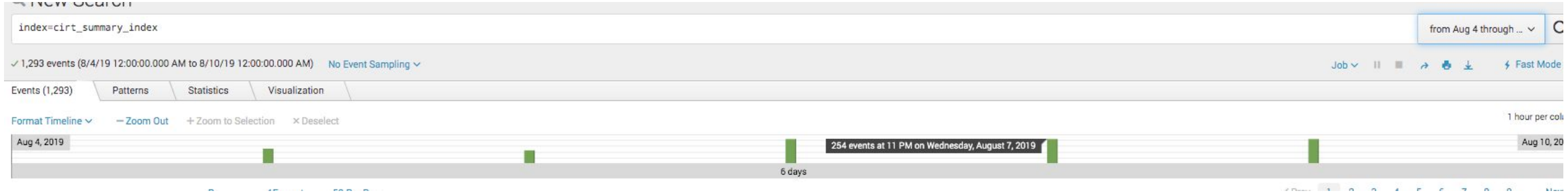


- Backup your data – super easy to blow everything away with one bad command
  - Monthly alert that sends CSV to e-mail was our hack at out of Splunk Storage

Another Option: Summary Index

```
human_transaction_time_malicious, human_transaction_time_suspicious, human_transaction_time_unwanted, human_transaction_time_not_detected,
unique_detections_suspicious, unique_detections_unwanted, unique_endpoints, unique_subdomains
| dedup datestamp, detector_id | addinfo  | collect index=cirt_summary_index addtime=false marker="report_name=\"transaction_data_summary\""
```

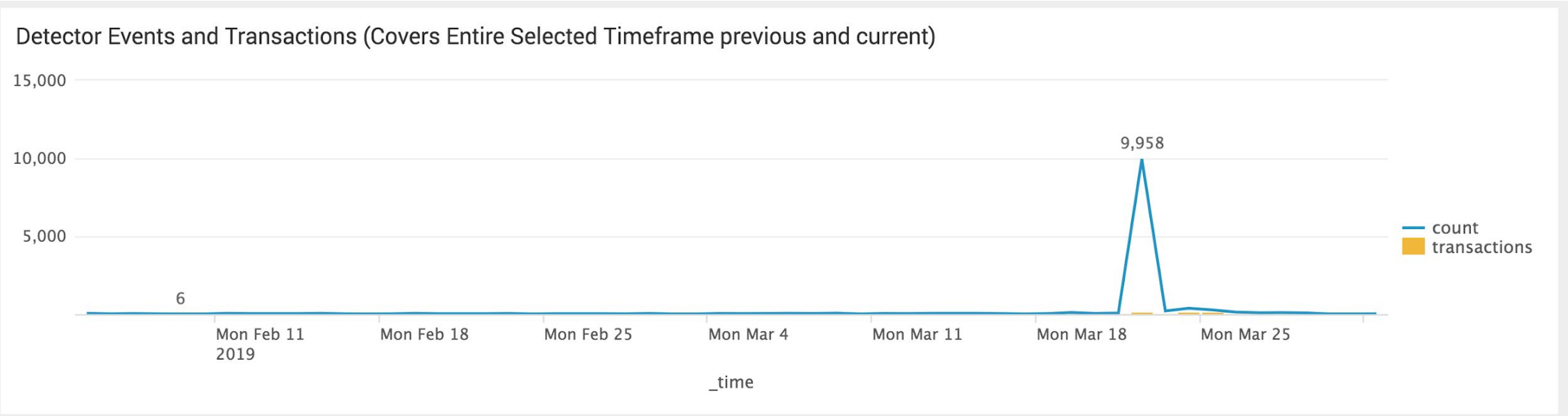
- Simple command took a good amount of time to figure
- You can search this index almost like normal



Example Analytics on the Data

.conf19
splunk>

Outliers: Are Spikes skewing my analysis?

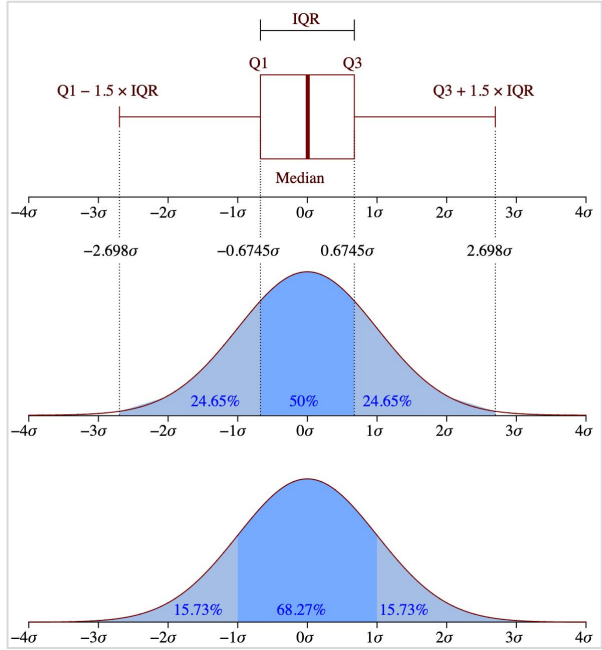


Detector Conversion

human_conversion_rate_period	human_conversion_rate_previous_period	diff
0.89	8.83	-7.94

Calculating IQR

- Splunk did all the hard work here
  - https://docs.splunk.com/Documentation/Splunk/7.3.1/Search/Findingandremovingoutliers



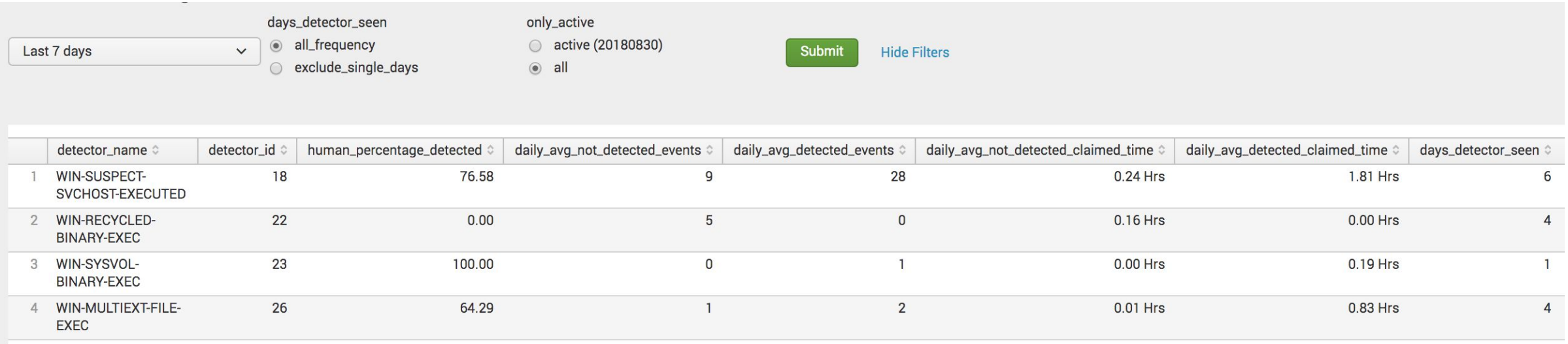

```
eventstats p25(human_transaction_time_not_detected) as not_detected_time_p25 p75(human_transaction_time_not_detected)
  as not_detected_time_p75 by detector_id

eval not_detected_time_IQR=(not_detected_time_p75-not_detected_time_p25)

eval not_detected_time_upperBound=(not_detected_time_p25+1.5*not_detected_time_IQR)

eval not_detected_time_isOutlier=if(human_transaction_time_not_detected > not_detected_time_upperBound, 1, 0)

eventstats sum(eval(if(not_detected_time_isOutlier="0", human_transaction_time_not_detected, 0))) as sum_transaction_time_not_detected by detector_id

eval adj_transaction_time_not_detected_hrs=round(sum_transaction_time_not_detected/3600,2)
```

Data Is Now More Representative of Signature

- Average is 44, versus 800
- Help us stack signatures for comparison

detector_id	adjusted_not_detected_count_avg	not_detected_count_avg	human_not_detected_count	not_detected_count_p25	not_detected_count_p75	not_detected_count_isOutlier	remove_outliers
790		803.7142857142857	213	35	183	1	
790		803.7142857142857	9957	35	183	1	
790	44.9	803.7142857142857	67	35	183	0	NotOutlier
790	44.9	803.7142857142857	36	35	183	0	NotOutlier
790	44.9	803.7142857142857	93	35	183	0	NotOutlier
790	44.9	803.7142857142857	35	35	183	0	NotOutlier
790	44.9	803.7142857142857	6	35	183	0	NotOutlier
790	44.9	803.7142857142857	33	35	183	0	NotOutlier
790	44.9	803.7142857142857	53	35	183	0	NotOutlier
790	44.9	803.7142857142857	48	35	183	0	NotOutlier
790	44.9	803.7142857142857	54	35	183	0	NotOutlier
790	44.9	803.7142857142857	24	35	183	0	NotOutlier

Using Dataset: How many of Rule XX's events converted to actual incidents?

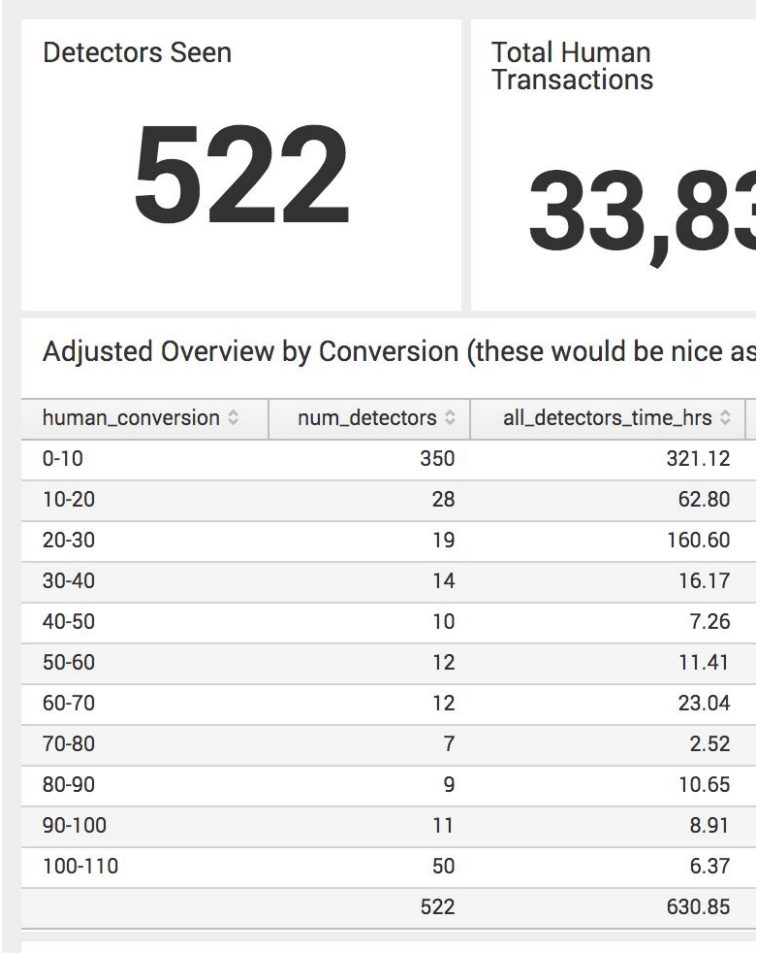


	detector_name	detector_id	human_percentage_detected	daily_avg_not_detected_events	daily_avg_detected_events	daily_avg_not_detected_claimed_time	daily_avg_detected_claimed_time	days_detector_seen
1	WIN-SUSPECT-SVCHOST-EXECUTED	18	76.58	9	28	0.24 Hrs	1.81 Hrs	6
2	WIN-RECYCLED-BINARY-EXEC	22	0.00	5	0	0.16 Hrs	0.00 Hrs	4
3	WIN-SYSVOL-BINARY-EXEC	23	100.00	0	1	0.00 Hrs	0.19 Hrs	1
4	WIN-MULTIEXT-FILE-EXEC	26	64.29	1	2	0.01 Hrs	0.83 Hrs	4

```
| eval
human_conversion=round(human_total_detected/human_total_detector_events*1
00,2)
```

Am I spending more time on detected events or false positives?

- For this time period 50% of human time was spent on things that convert less than 10% of the time.
- Maybe look there to improve that signature set first.
- We can drill into which specific one's they are with the data we saved.



Detectors Seen

522

Total Human Transactions

33,83

Adjusted Overview by Conversion (these would be nice as

human_conversion	num_detectors	all_detectors_time_hrs
0-10	350	321.12
10-20	28	62.80
20-30	19	160.60
30-40	14	16.17
40-50	10	7.26
50-60	12	11.41
60-70	12	23.04
70-80	7	2.52
80-90	9	10.65
90-100	11	8.91
100-110	50	6.37
	522	630.85

You'll Still Need Logs

- The previous dataset is meant to show a starting place or a trend.
- There's a tradeoff between metrics and logs.
- In order to adequately tune a signature you need to know:
  - What specific things are the signature hitting on we don't care about?
  - Is there anything specific I can add or remove that will cause it to be more targeted?
  - Is it even looking for the correct behavior?
  - Log data may not go back as long, but it will have more detail.

Conclusion:

- Summary indexes and lookup tables are good tools for summarizing large data sets that are difficult to search and retain
- Lessons Learned – building out the backend search is important, probably content to use it is more so and ensuring the team has the right tools
- There's way more than one way to do all of this, you tend to start with what you know.