



Two clicks to cluster and block thousands of spammers

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021





Story time



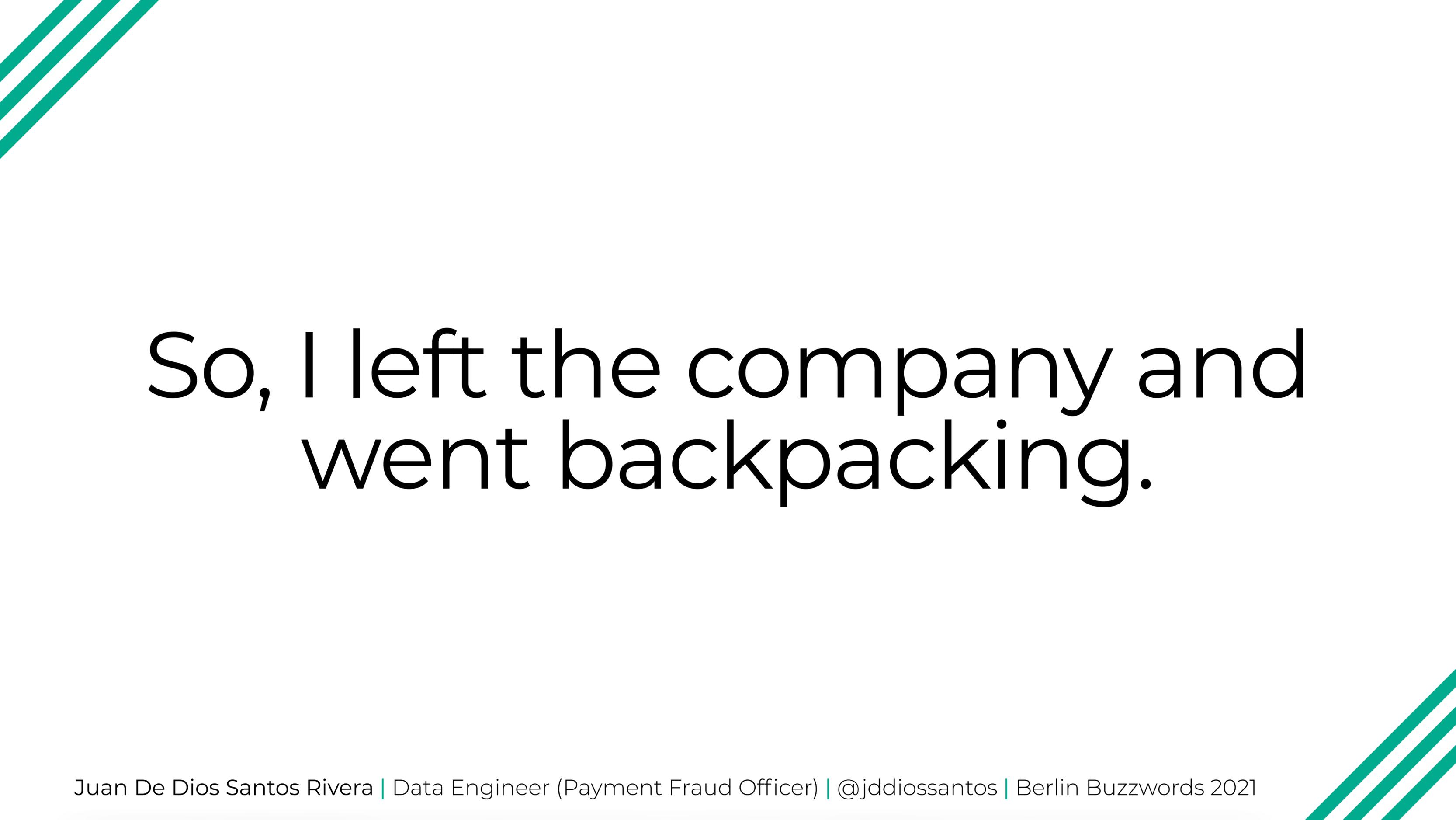
This is my second time
working at LOVVOO.



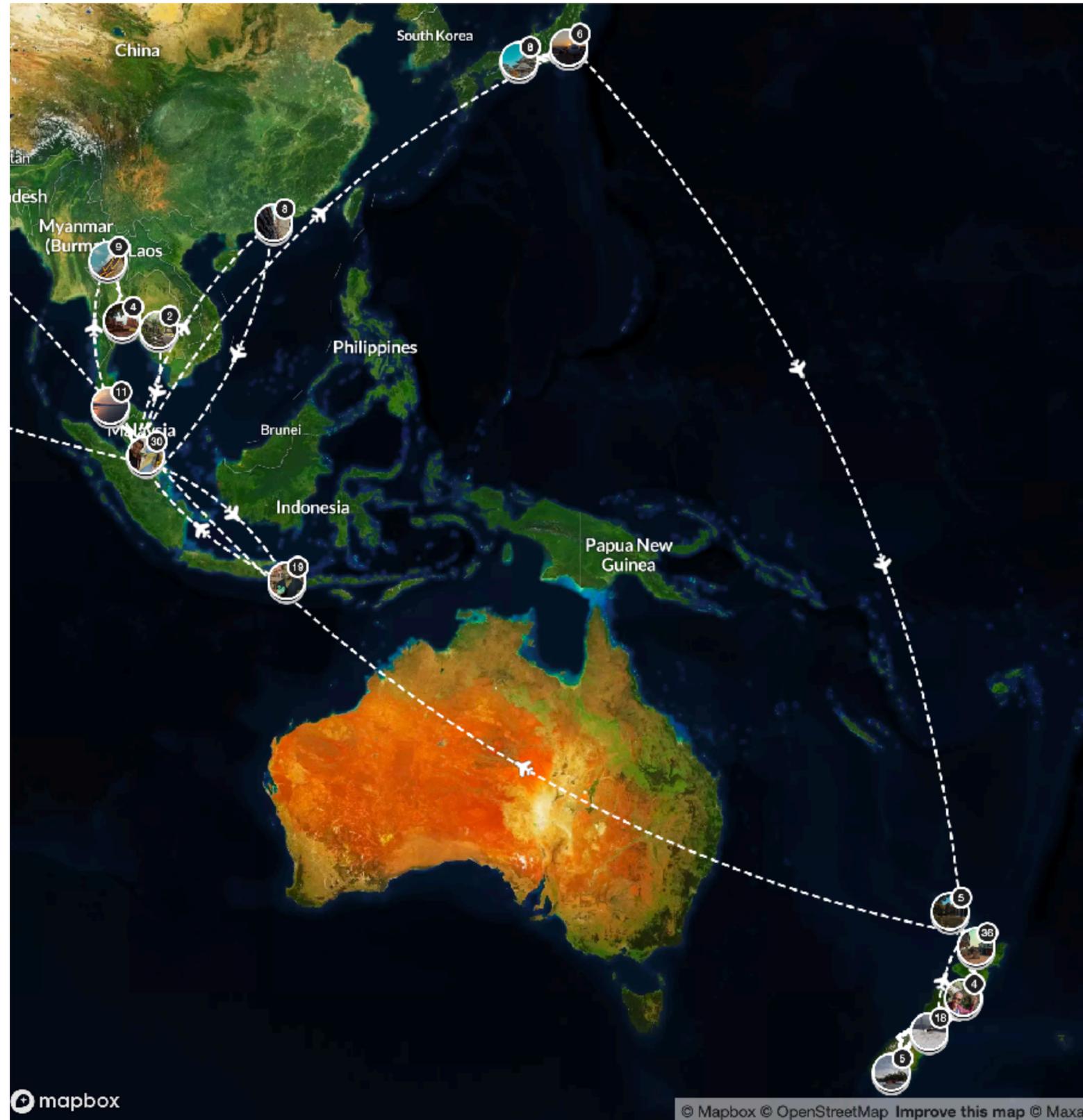
Previously, I worked from
2016 to 2019.



But I got the itch to replace the keyboard for a backpack, the IDE for a camera, and the trips to the coffee machine for trips to the beach.



So, I left the company and
went backpacking.





But then, COVID came.





“Meh. I miss my old job. I wish I would have done the **cluster** thing. If life give me a second chance there, I’ll do it.”

```
func whatToDo(c chance) {  
    switch c {  
    case first:  
        fmt.Println("This is the past.")  
    case second:  
        doCluster()  
    default:  
        fmt.Println("I don't know :/")  
    }  
}
```



I got the second chance.



And here I'll present my
clusters :)

Who am I?

Hi, I'm Juan :)

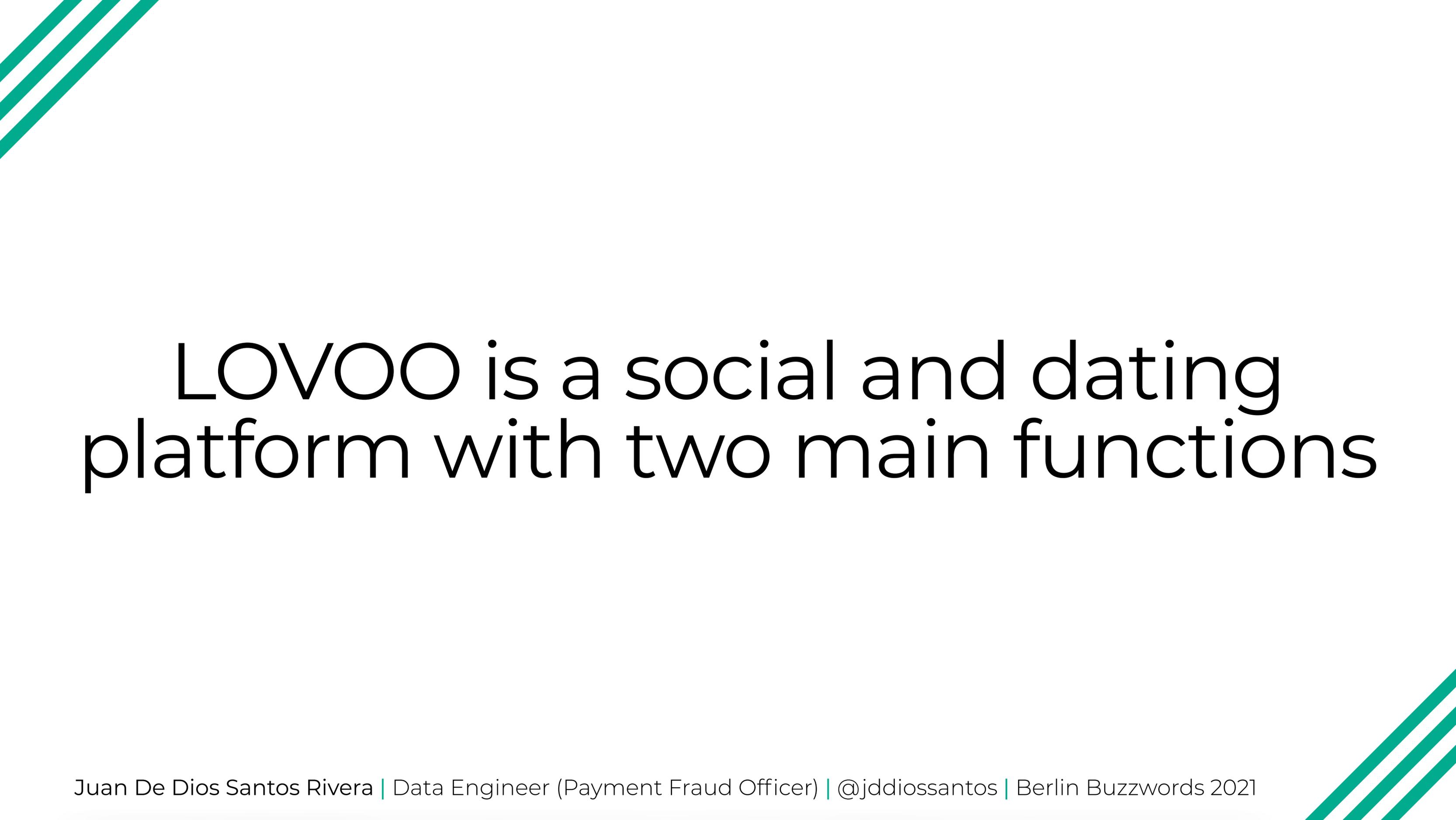
- Data Engineer at LOVOO
- Deal with fraud and spam
- Develop systems to moderate fraudulent and spam content
 - Using data, algorithms, machine learning, and common sense
- Person who likes clusters
- Former backpacker, photographer, and writer
- Author of *Practical TensorFlow.js* (Apress)

Agenda

- The context of the problem
- Clusters
- Architecture
- Features vector
- The clustering algorithm
- Outcome
- Takeaways

The context of the problem

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021



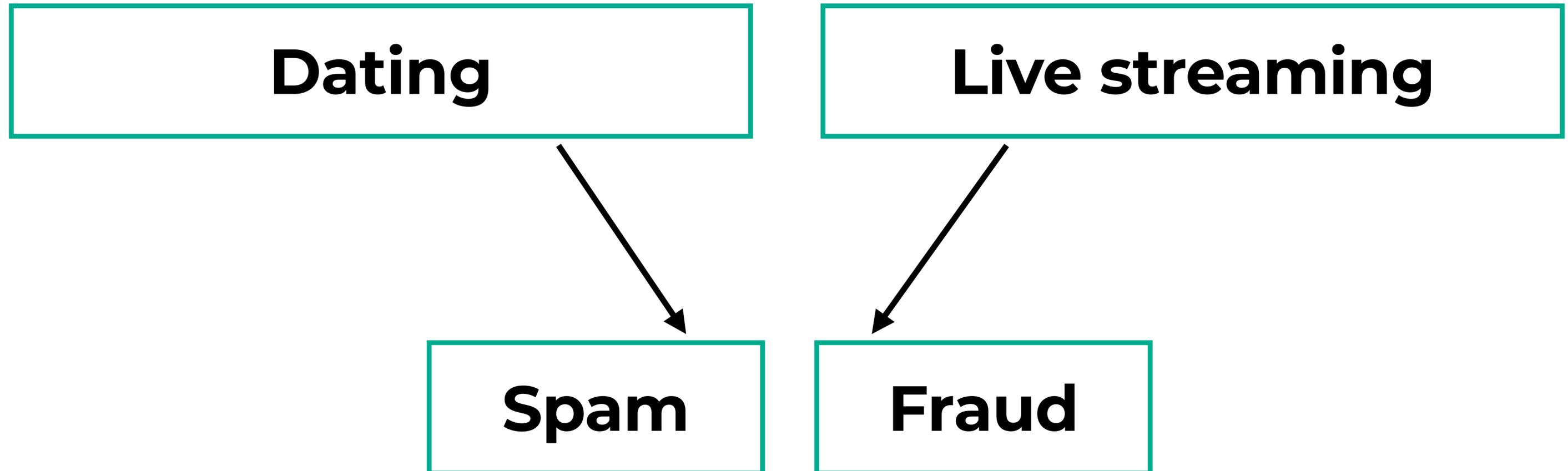
LOVVOO is a social and dating platform with two main functions

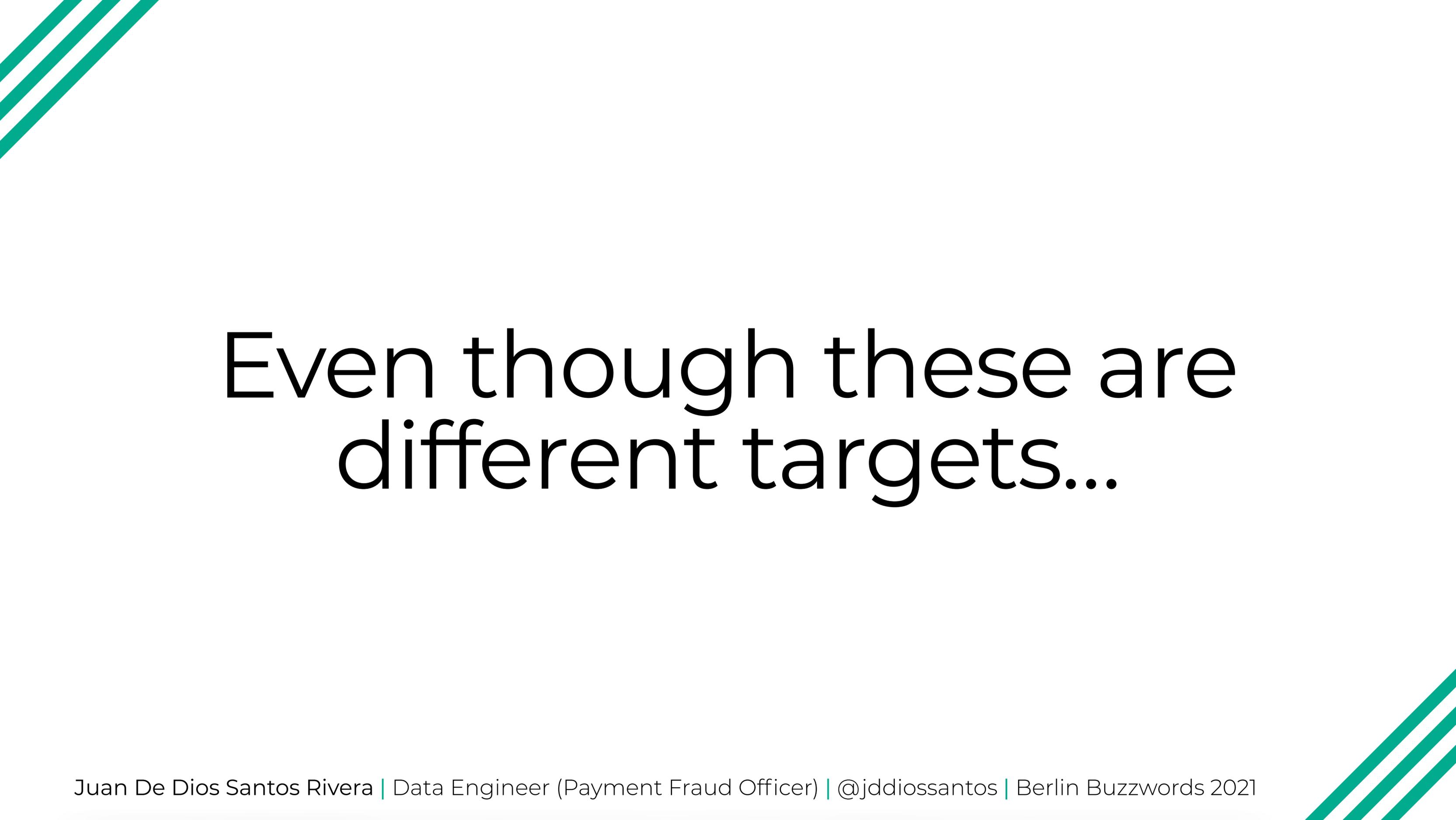
Context

Dating

Live streaming

Context





Even though these are
different targets...



...at the end, they are unwanted
content we want to remove.

Some of our techniques

Context

- Machine learning
 - Anomaly detectors
- Information retrieval algorithms
- Time series-based algorithms
- Heuristics
- Rule-based systems
- And one of the most important...



Conviction

Conviction

Context

- Our score collector
- It gathers scores from all the other systems
 - We call them **confidences** scores.
 - They are “grouped” by system, e.g, fraud-related scores and images-related scores.
- We create rules from them
 - e.g., if $C1$ and $C2 > N$ then, we block a user.

Conviction

But we have 200+ scores

Conviction

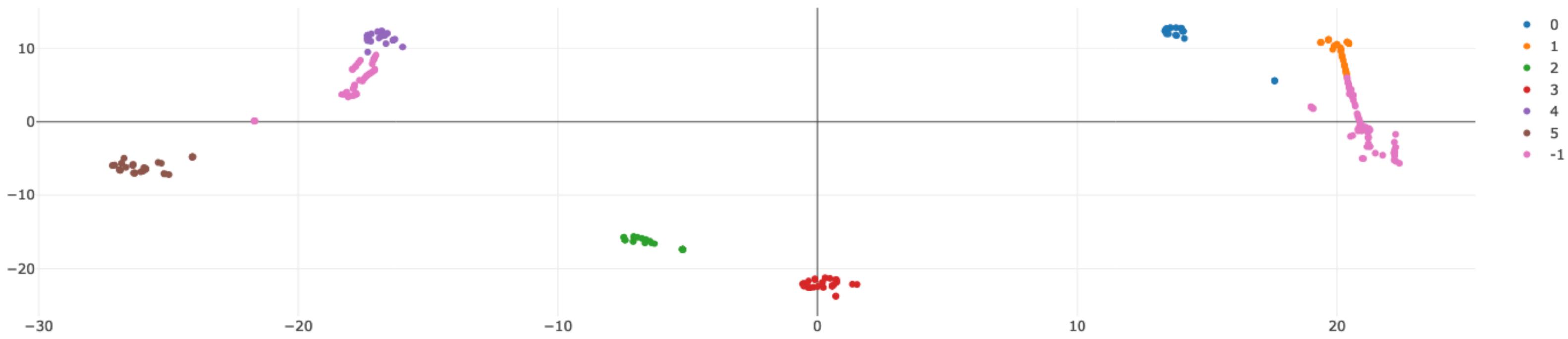
So it keeps getting harder to model
the rules by hand.

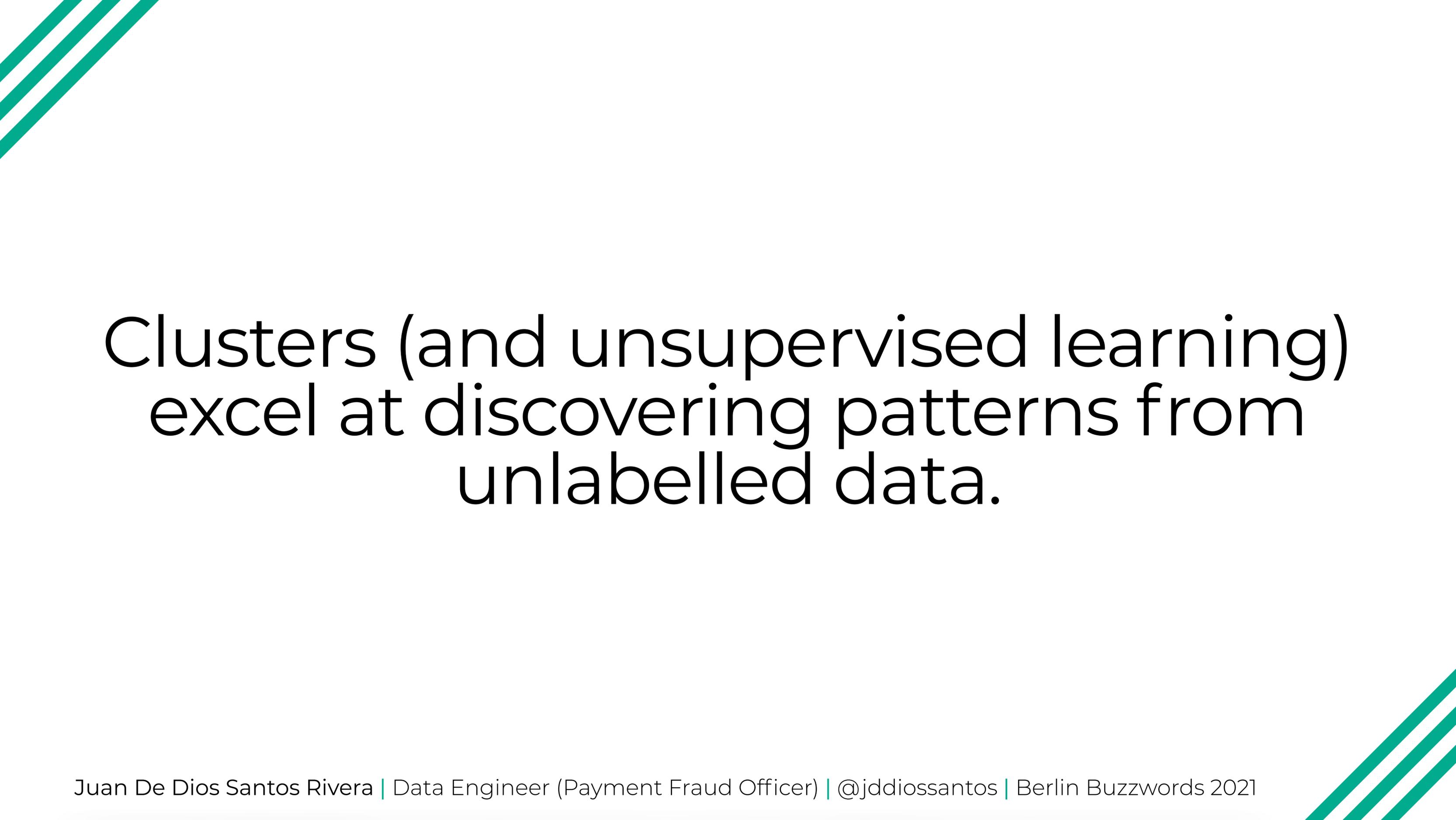
Conviction

Hence, the clusters :)

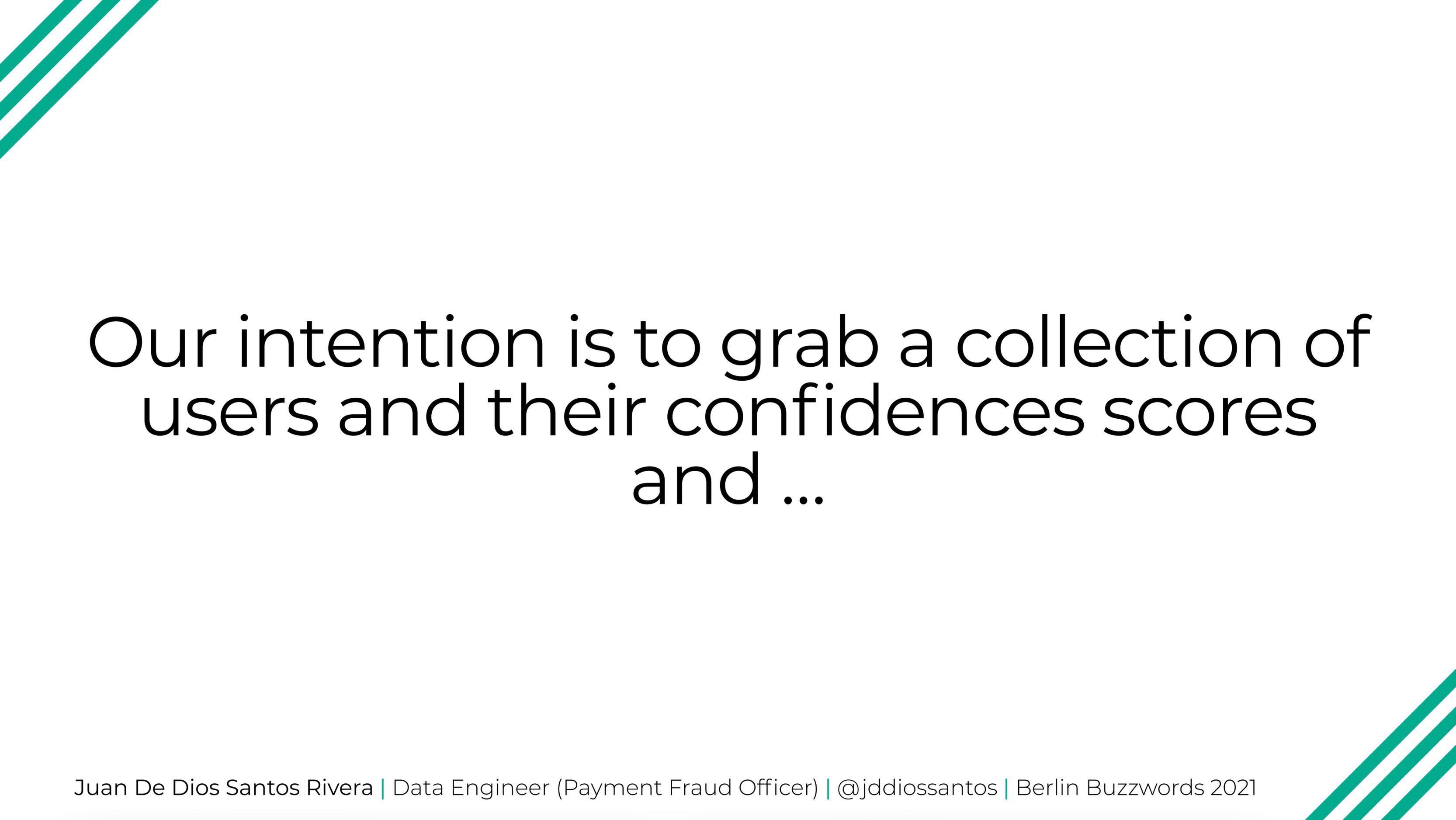
Clusters

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021





Clusters (and unsupervised learning)
excel at discovering patterns from
unlabelled data.



Our intention is to grab a collection of
users and their confidences scores
and ...

The intention

Clusters

- **Group and block them**
 - Depending on the percentage of already existing spammers within the cluster or on cluster analysis.
- **Create new rules for our rule-based system**
 - Work in progress :/
- **Visualising them**
 - To test if there's separation and because images are always pretty :)

Architecture

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021

Berlin Buzzwords Batch :)

Filter Batch Cluster Batch Create Action Delete Batch

Sections:

Batch Info:

- Added**
2021-05-09 09:51:25
- Created by**
Anonymous
- Origin**
manually created
- Total Items**
2

General LOVOO V4

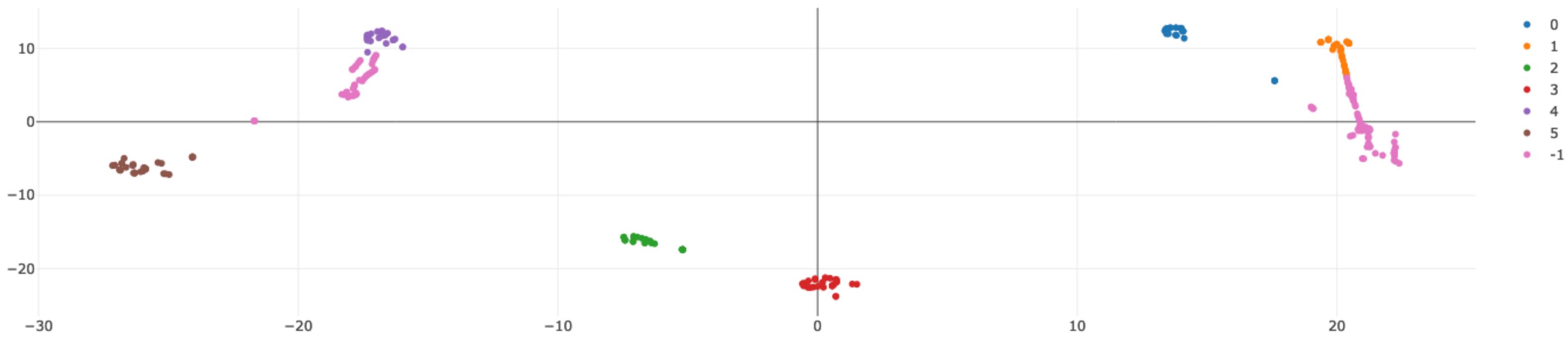
ID: user-1
Name: User 1
Reg Date: not set
Flags:
Other Flags:
Email: -
Super duper secret information
Freetext: -

General LOVOO V4

ID: user-2
Name: User 2
Reg Date: not set
Flags:
Other Flags:
Email: -
Super duper secret information
Freetext: -



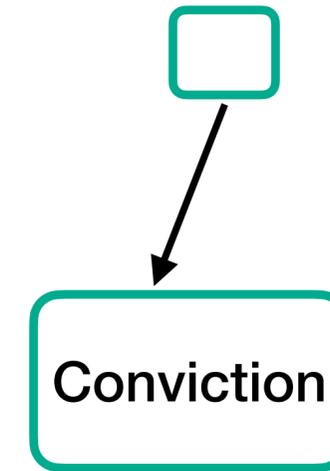
How do we go from that first
click to...

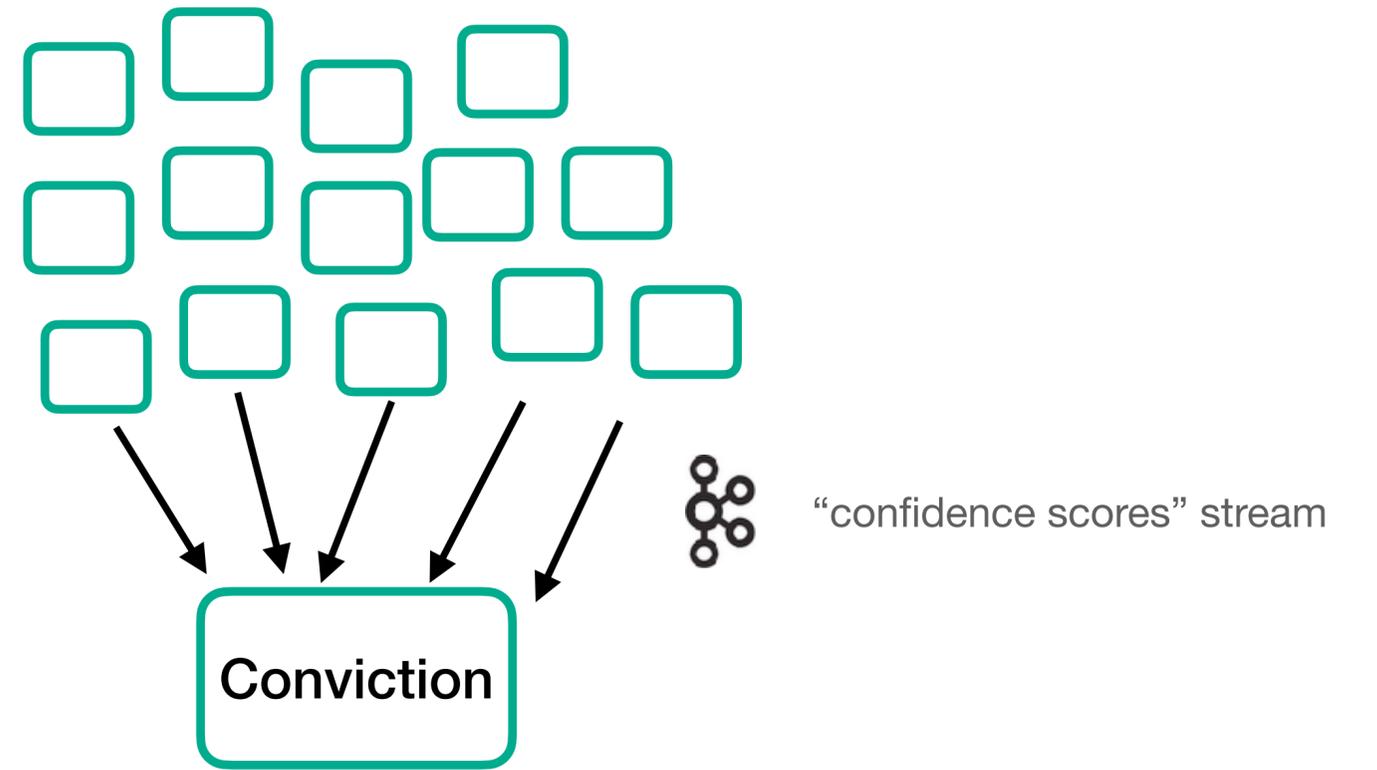


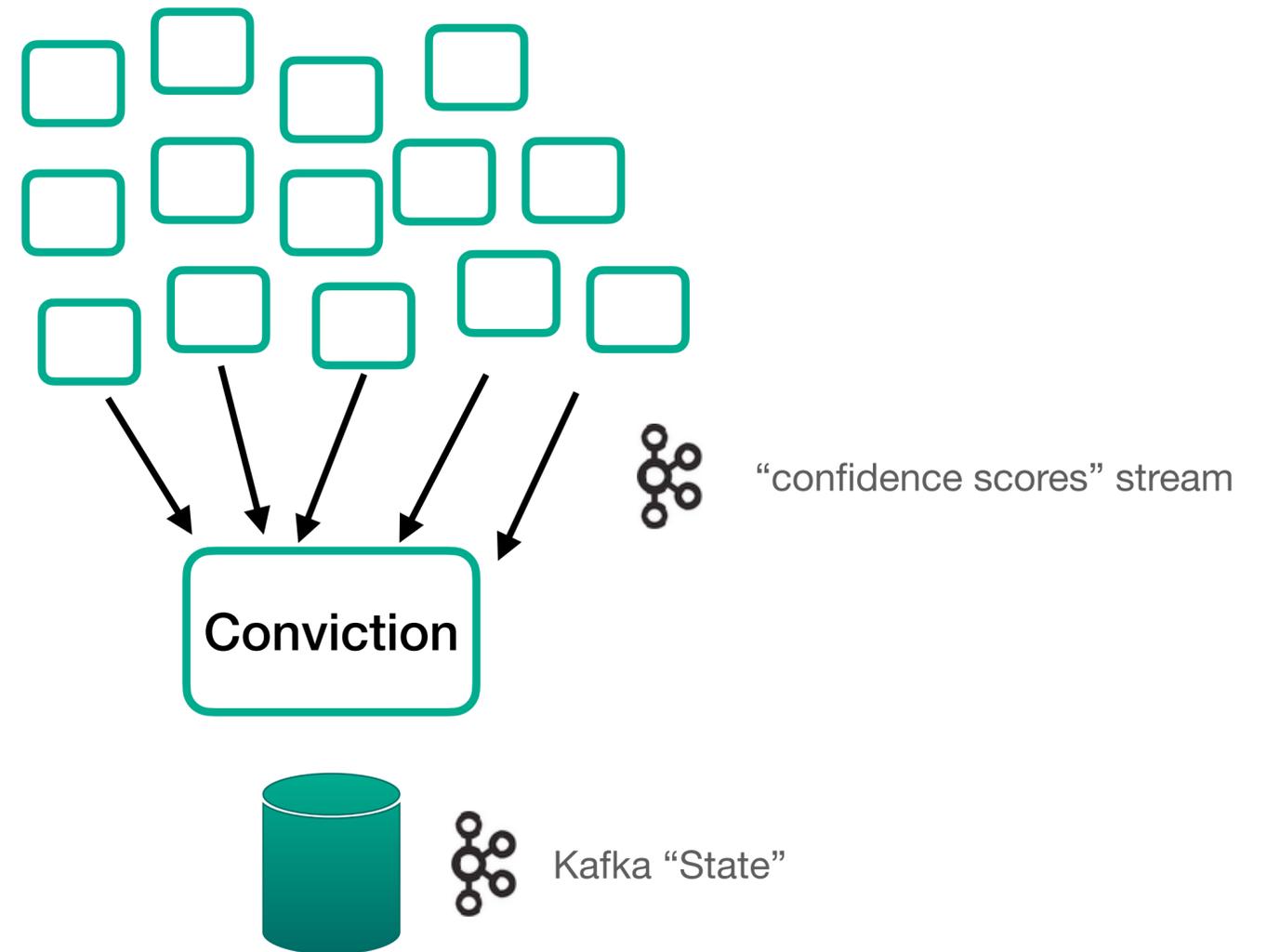


Like this:

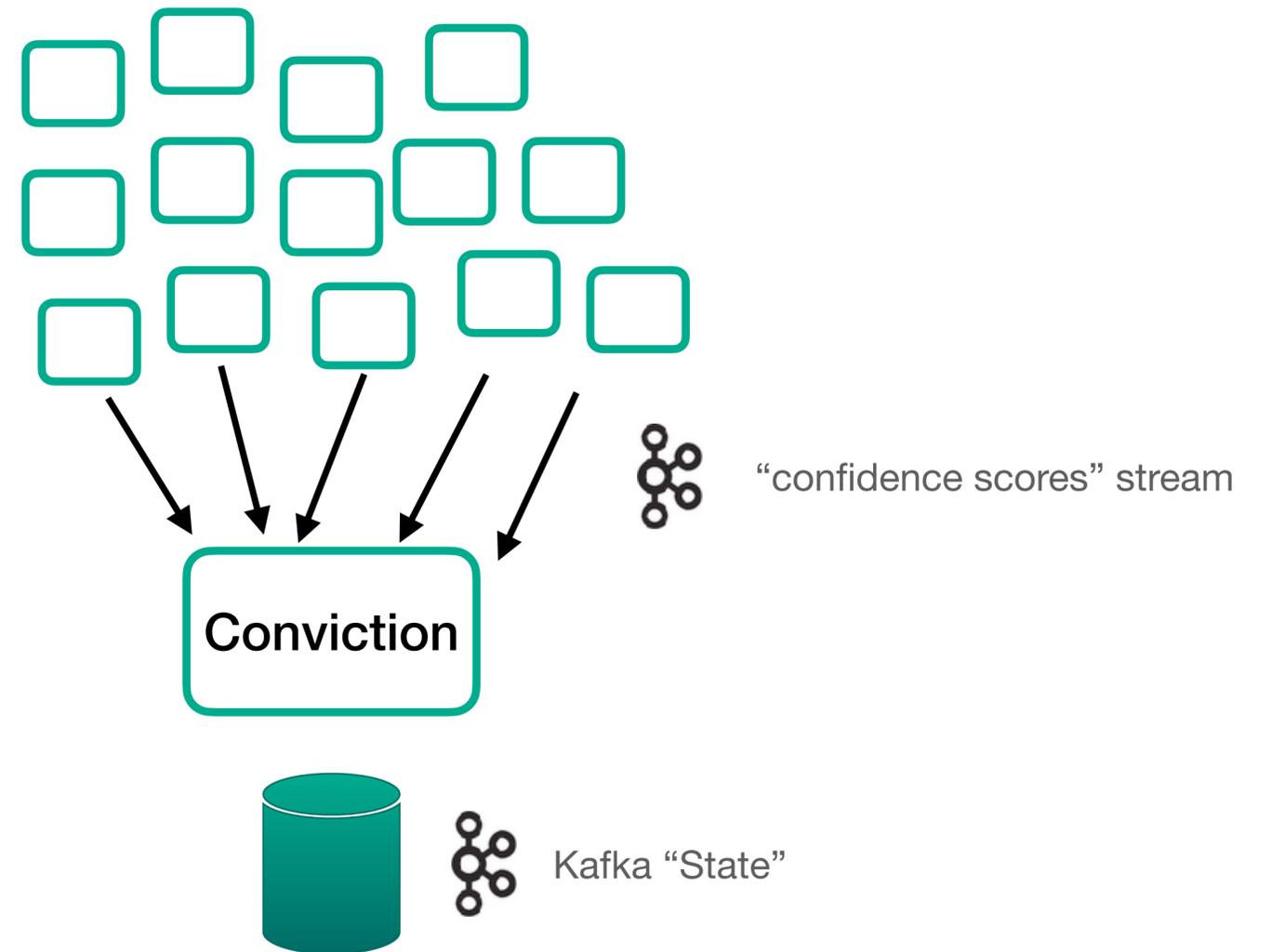
Conviction

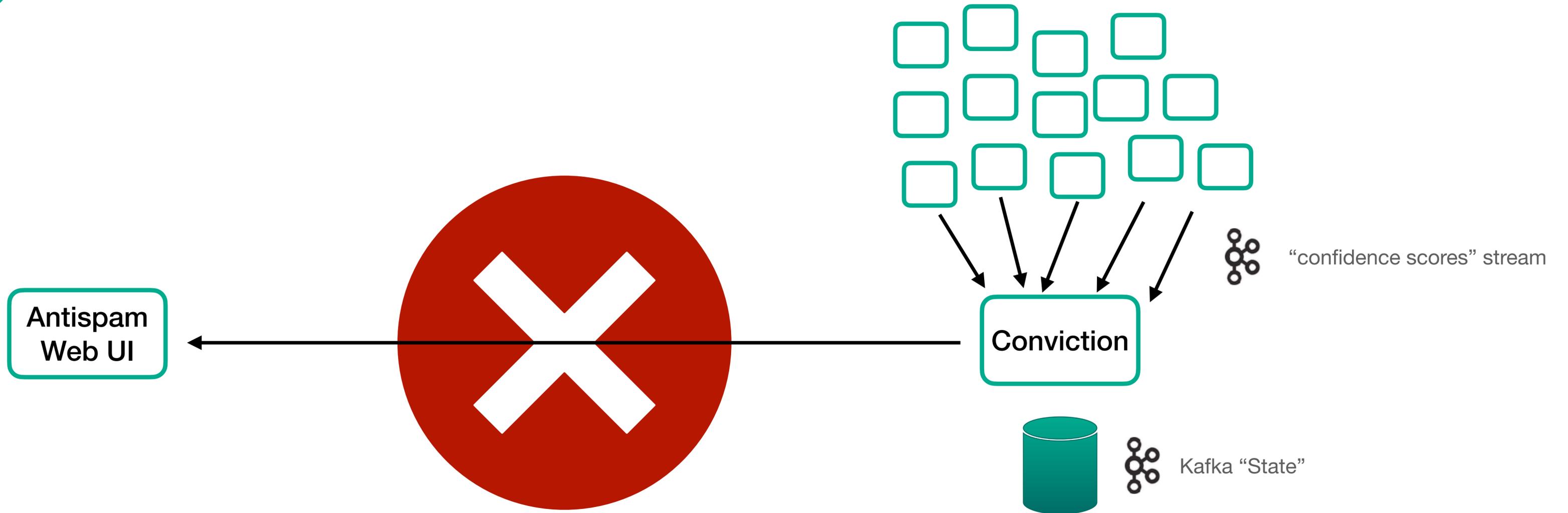


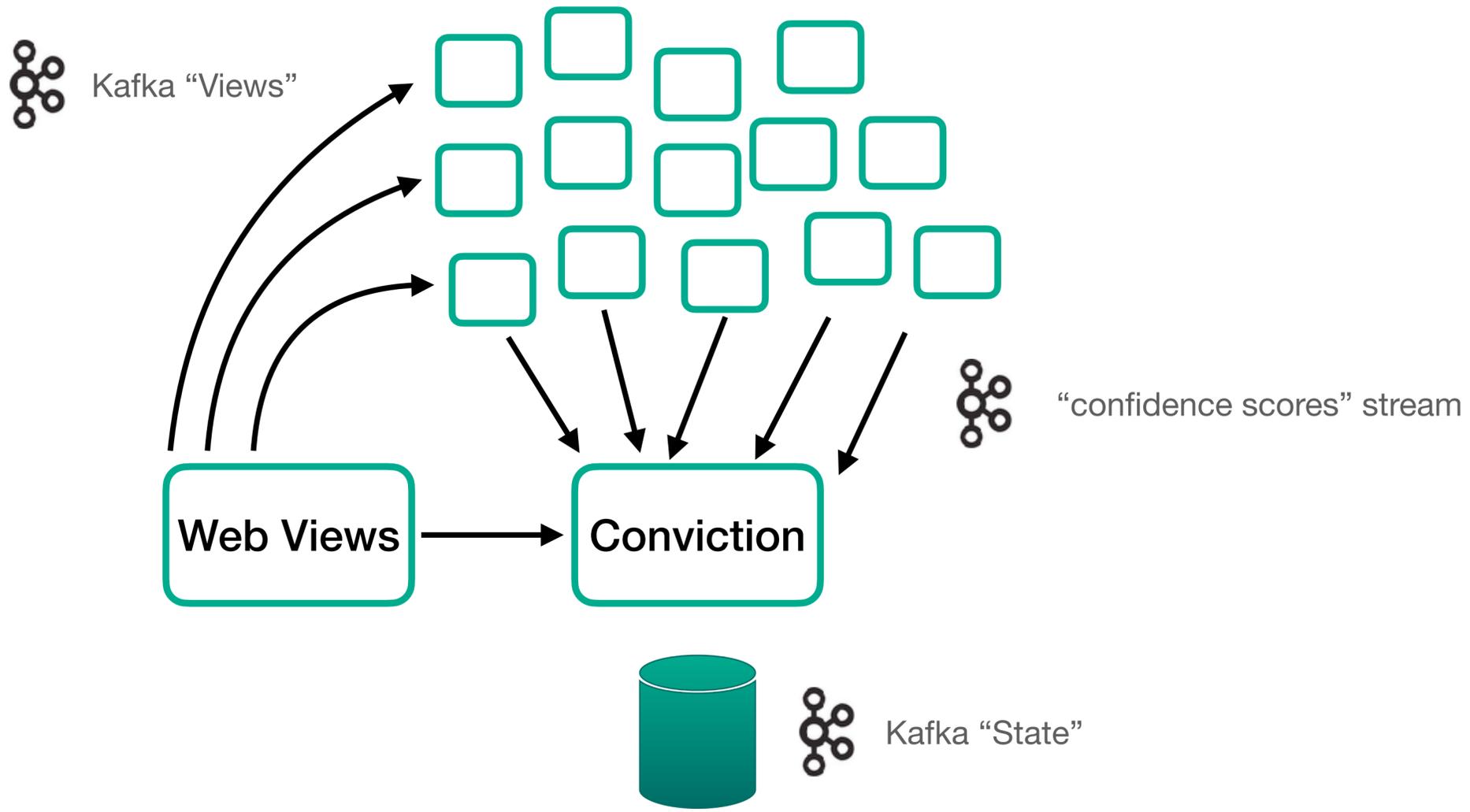


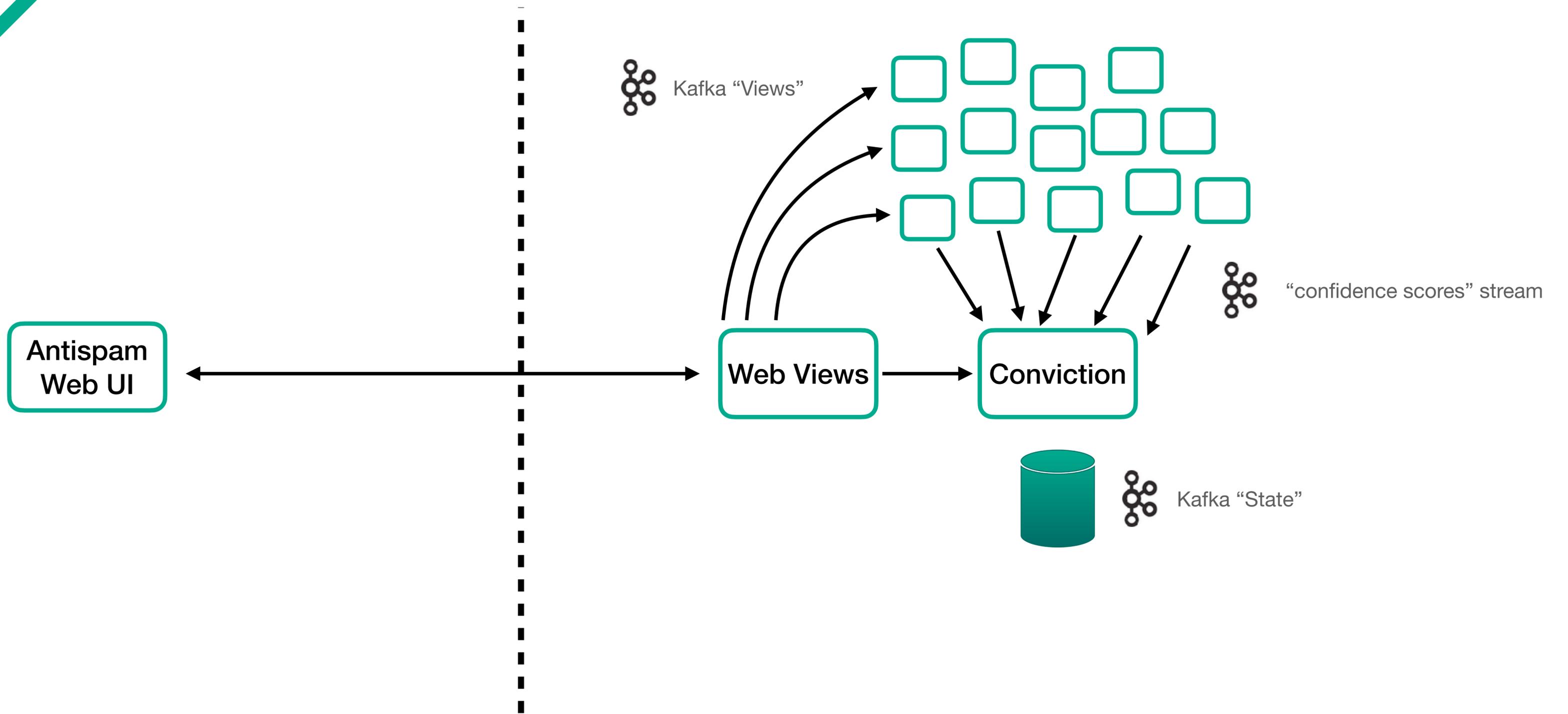


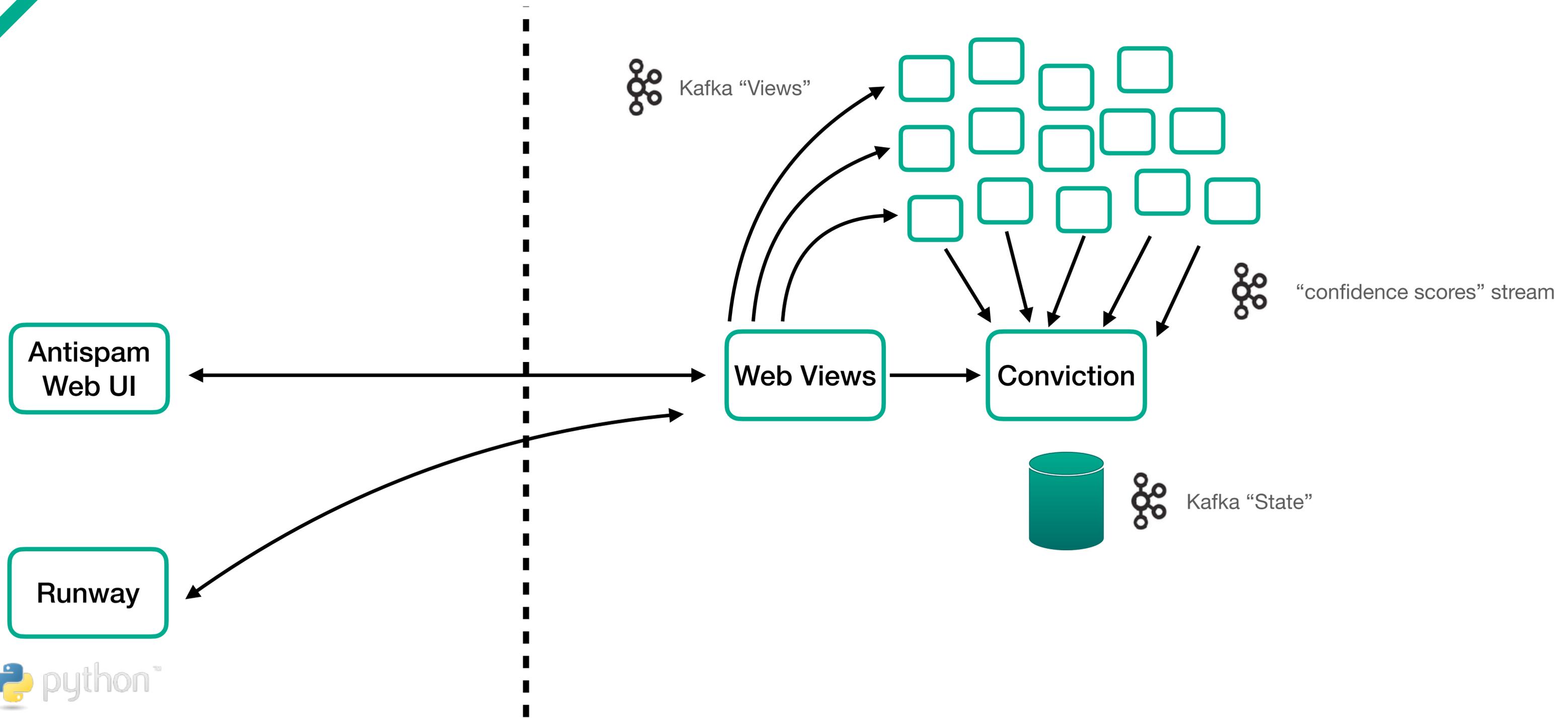
Antispam
Web UI

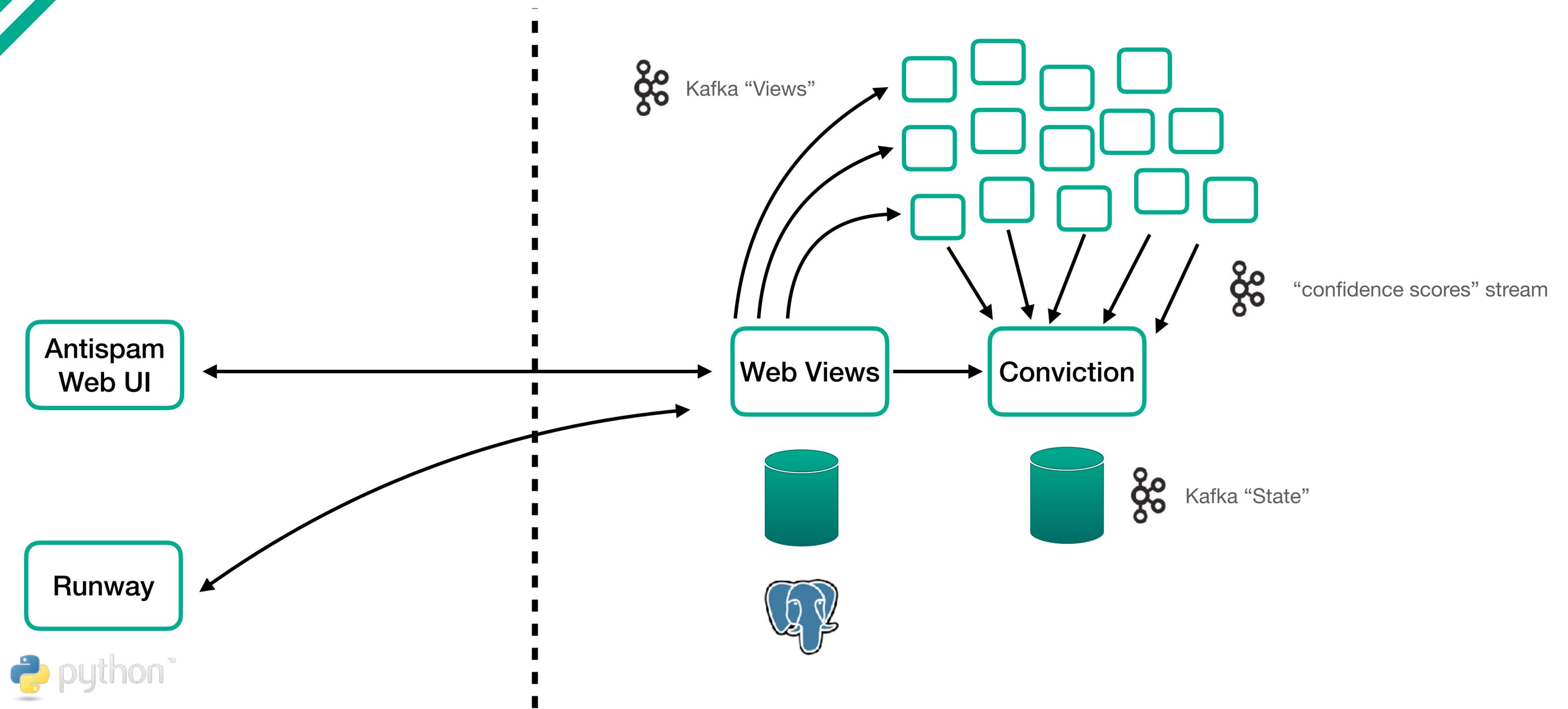


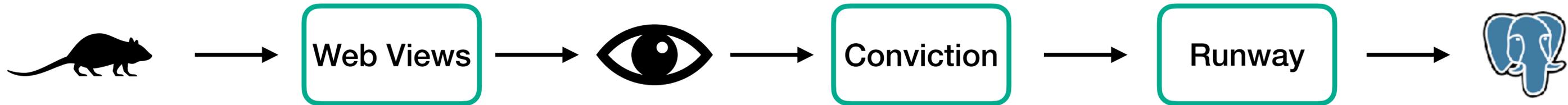


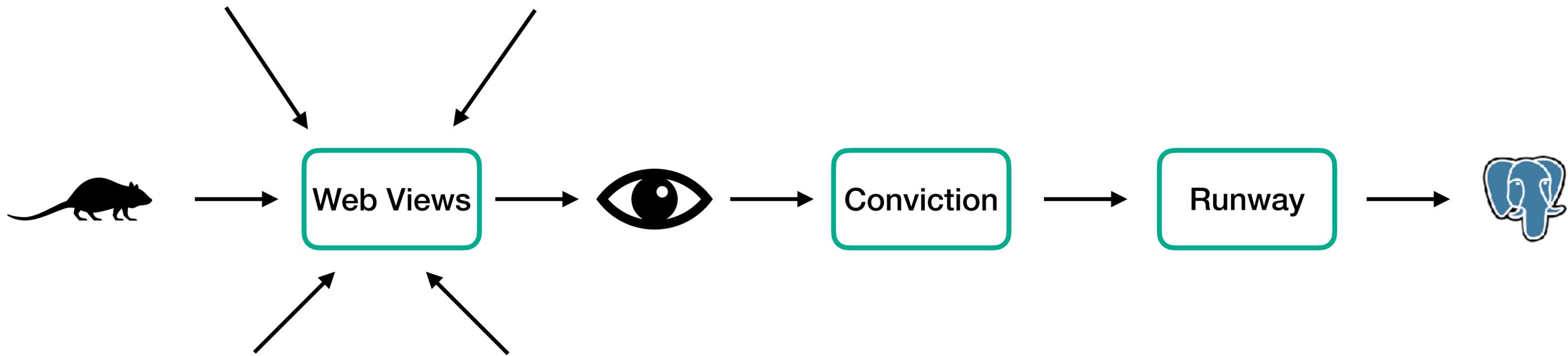










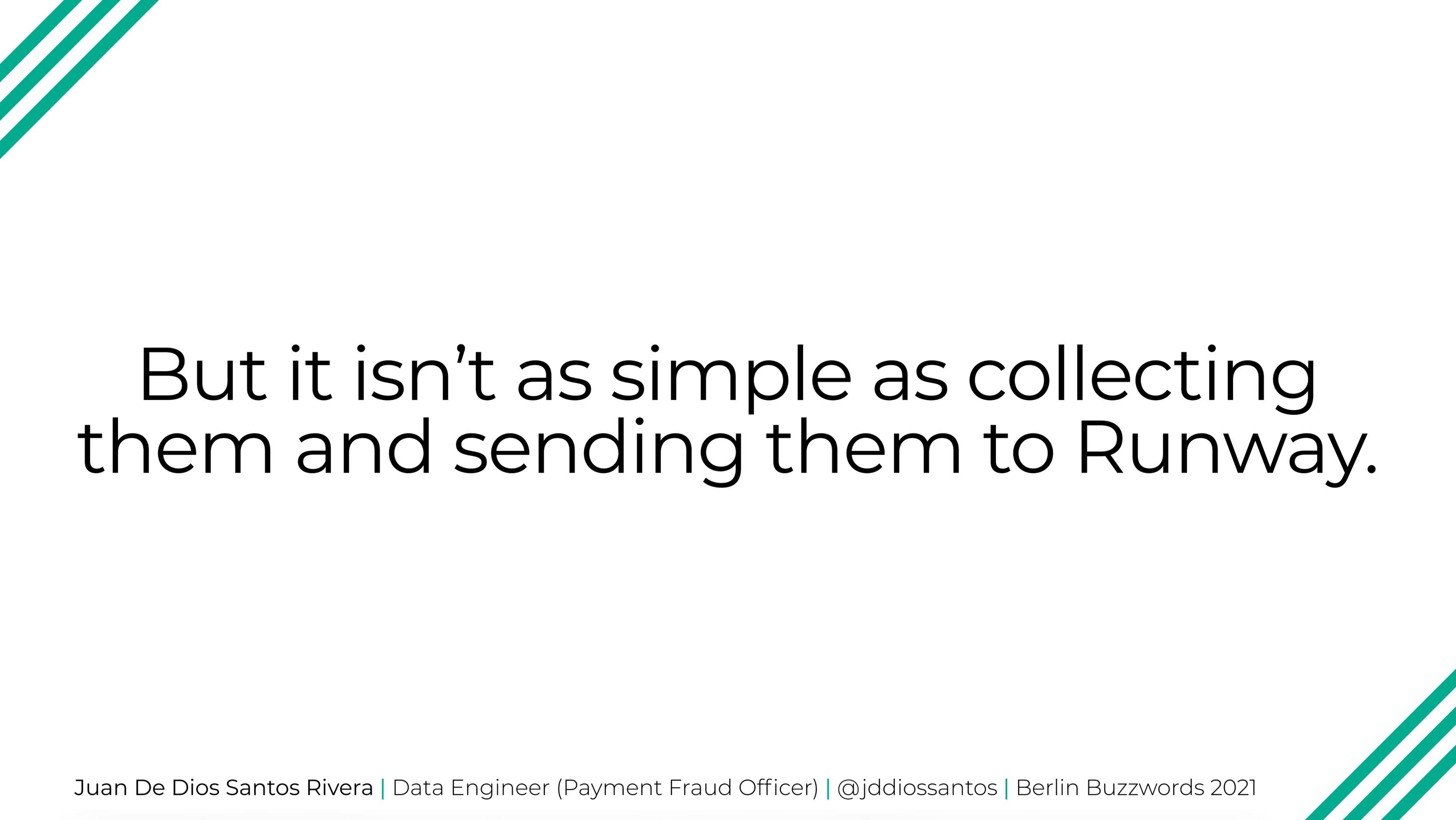


Feature vector

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021



The feature vector is the set of confidences we want to cluster



But it isn't as simple as collecting them and sending them to Runway.

Difficulties

Features vector

- Not every user has the the same confidences scores.
- For example,
 - user-1: c1, c2, c3,
 - user-2: c2, c3, c4
 - user-3: c5, c6

Features vector

- What we did is first create a union of all of our user's confidences scores
 - $\{c1, c2, c3\} \cup \{c2, c3, c4\} \cup \{c5, c6\} = \{c1, c2, c3, c4, c5, c6\}$
- Then, for each user, we get those confidences scores.
 - If the user doesn't have a score, we use the default value, e.g., 0.



Then, we cluster.

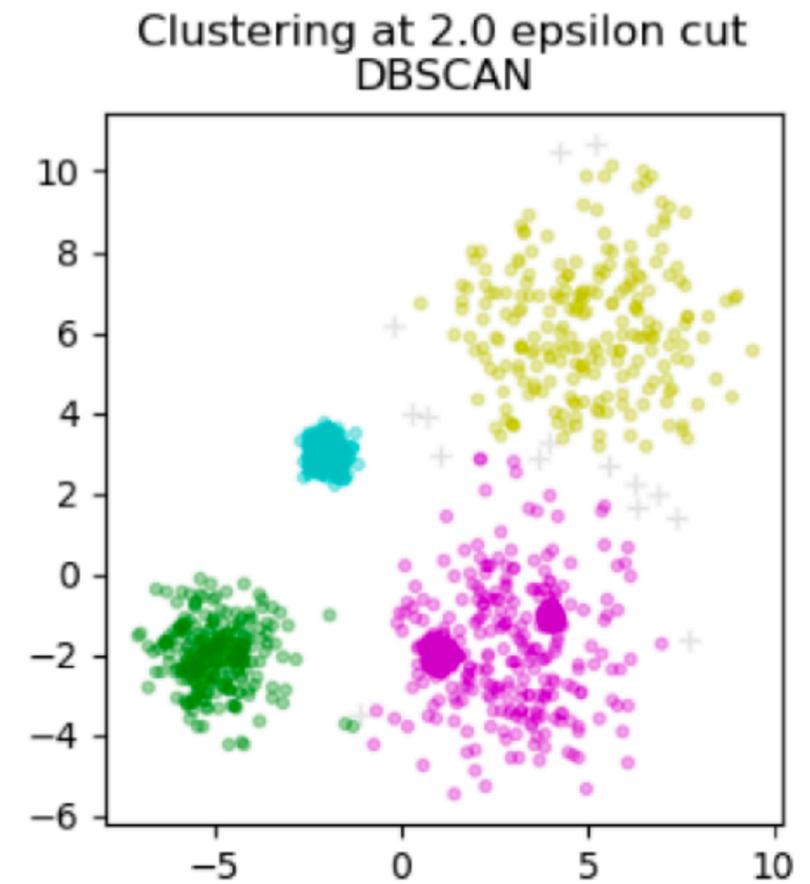
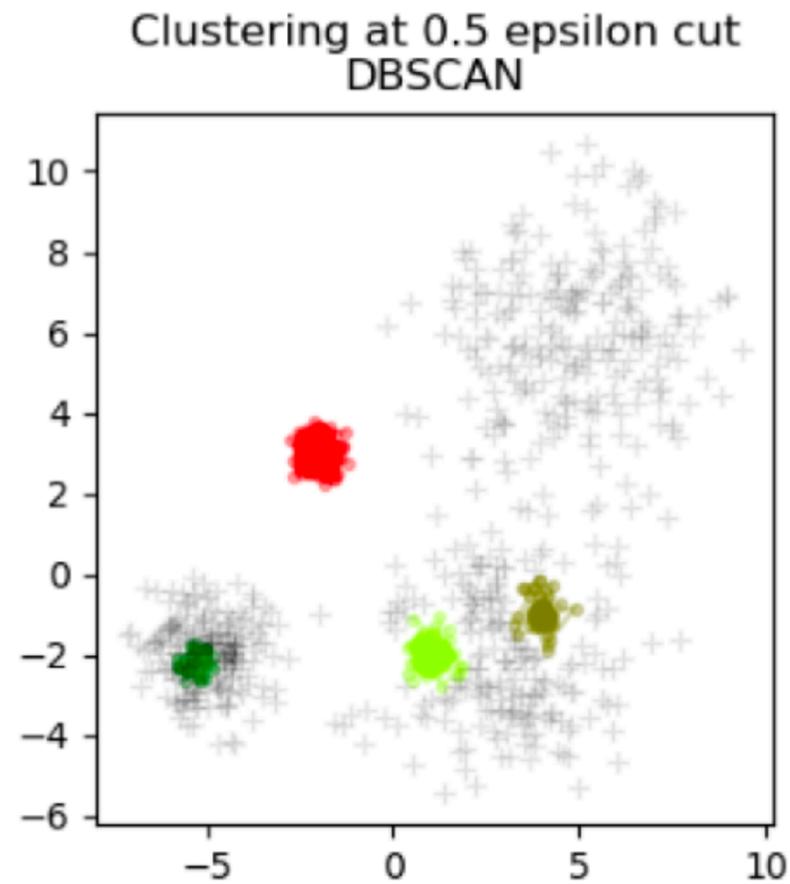
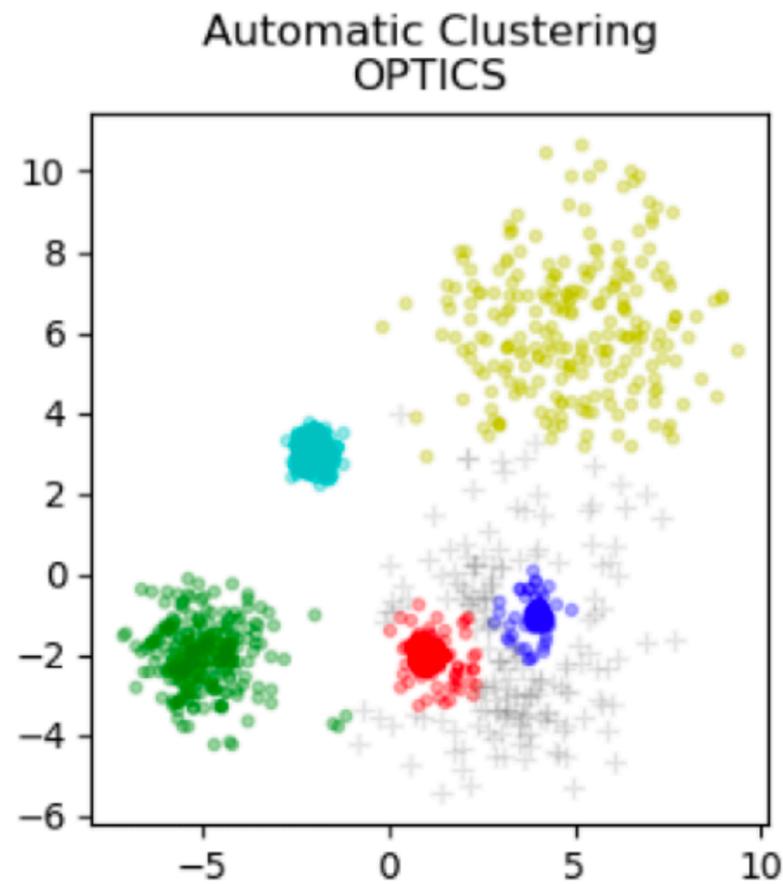
The clustering algorithm

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021

OPTICS

- **O**rdering **P**oints **T**o **I**dentify the **C**lustering **S**tructure
- Density-based algorithm
- Works by finding dense areas where a minimum number of points, *minPts* are within distance ϵ of each other.
- Controlling these two hyperparameters grants you flexibility but also means you have to find an optimal set of them.

OPTICS



Source:
https://scikit-learn.org/stable/auto_examples/cluster/plot_optics.html
by Shane Grigsby and Adrin Jalali

OPTICS

- We are using a default *minPts* of **20** that we can overwrite from the API call.

Outcome

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021

Berlin Buzzwords Batch :)

Filter Batch Cluster Batch Create Action Delete Batch

Sections:

Batch Info:

Added
2021-05-09 09:51:25

Created by
Anonymous

Origin
manually created

Total Items
2



General LOV00 V4

ID: user-1
Name: User 1
Reg Date: not set
Flags:
Other Flags:
Email: -
Super duper secret information
Freetext: -



General LOV00 V4

ID: user-2
Name: User 2
Reg Date: not set
Flags:
Other Flags:
Email: -
Super duper secret information
Freetext: -

Cluster batch ID

Clustering Settings:

- Fit a t-SNE model
- Apply MinMaxScaler to the data

Confidences groups to cluster:

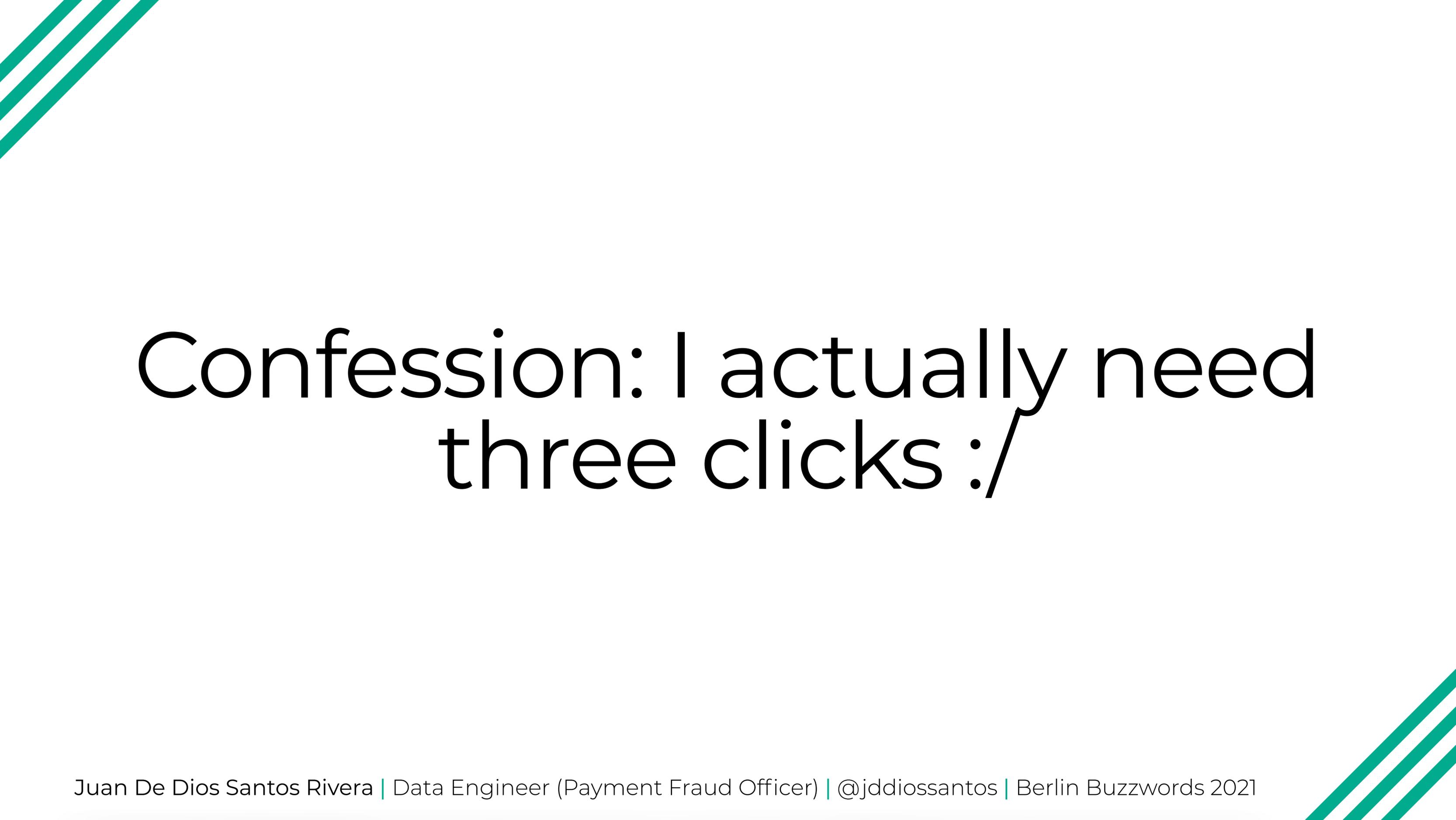
- Cluster all confidences

- A confidences
- B confidences
- C confidences
- D confidences
- E confidences
- F confidences
- G confidences
- H confidences
- I confidences
- J confidences
- K confidences
- L confidences

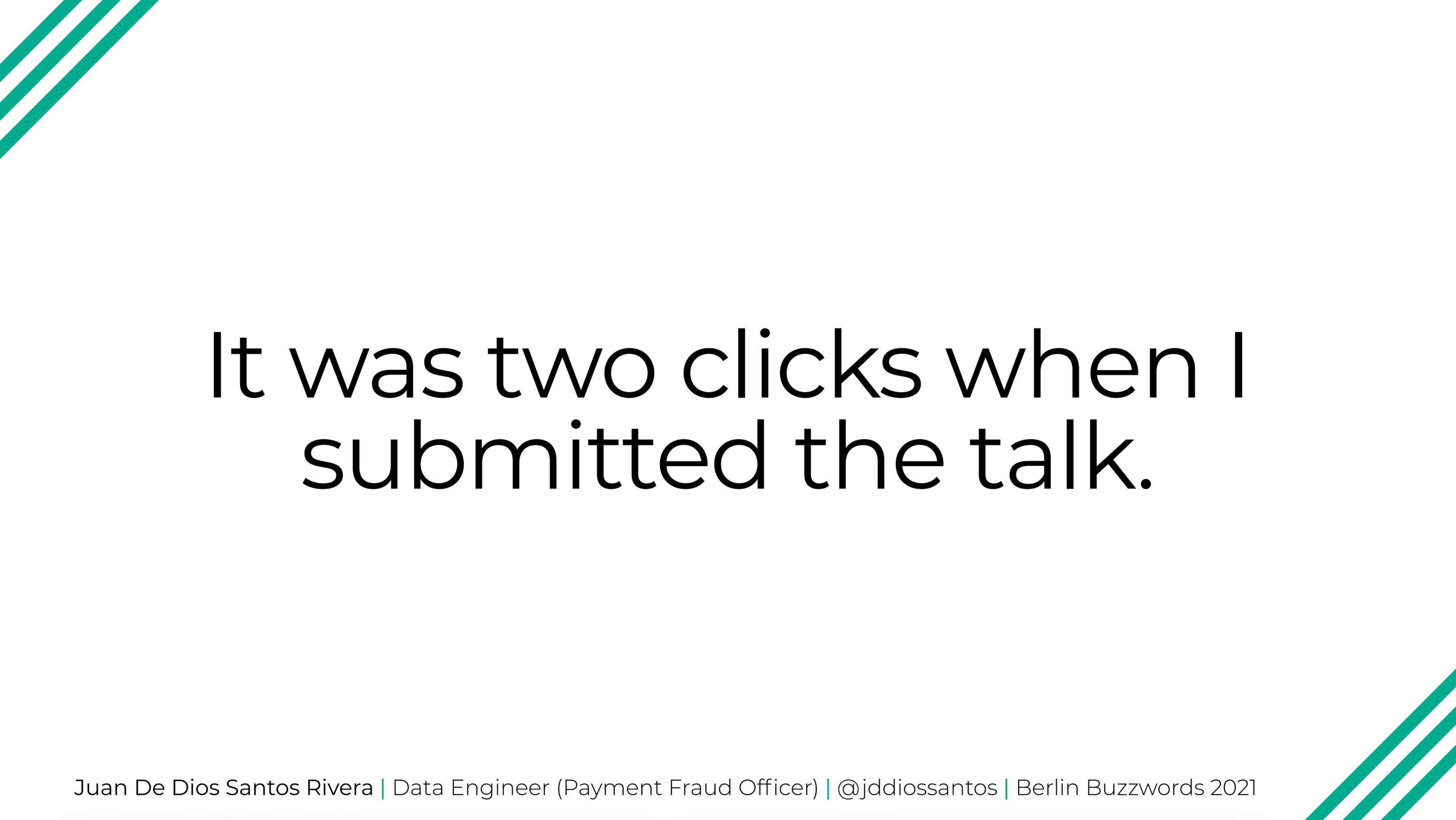
e.g. images-related confidences
or likes-related confidences

Cancel

Apply



Confession: I actually need
three clicks :/



It was two clicks when I
submitted the talk.



But I've made some
changes since then.



Still, three clicks are good.
Right?

✓ Done Clustering ID 123

Cluster Batch Info

Batch ID

abcse

Added

2021-05-15 13:21:06

Status

done

Items Clustered

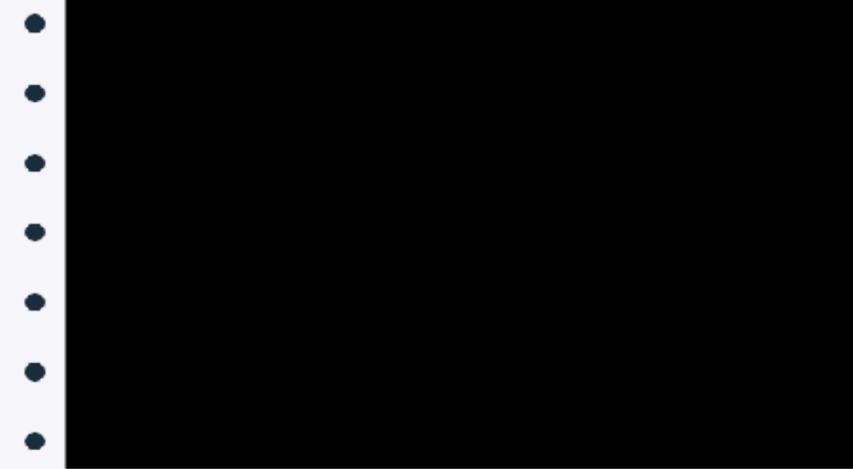
304

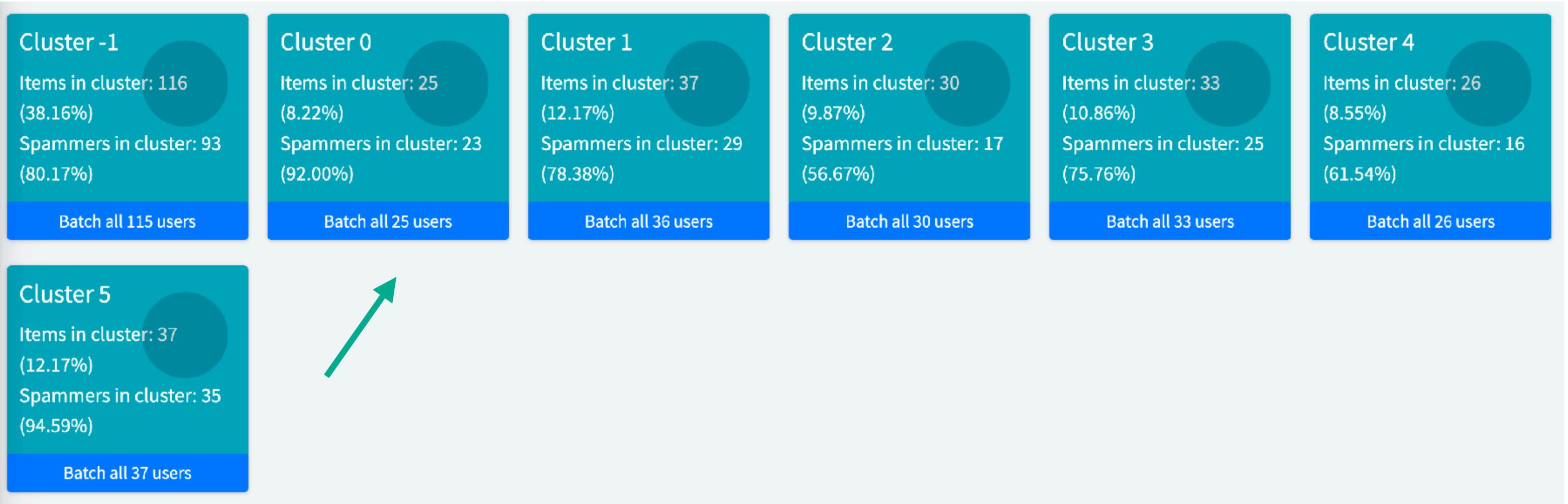
Features Clustered

161

Used Confidences

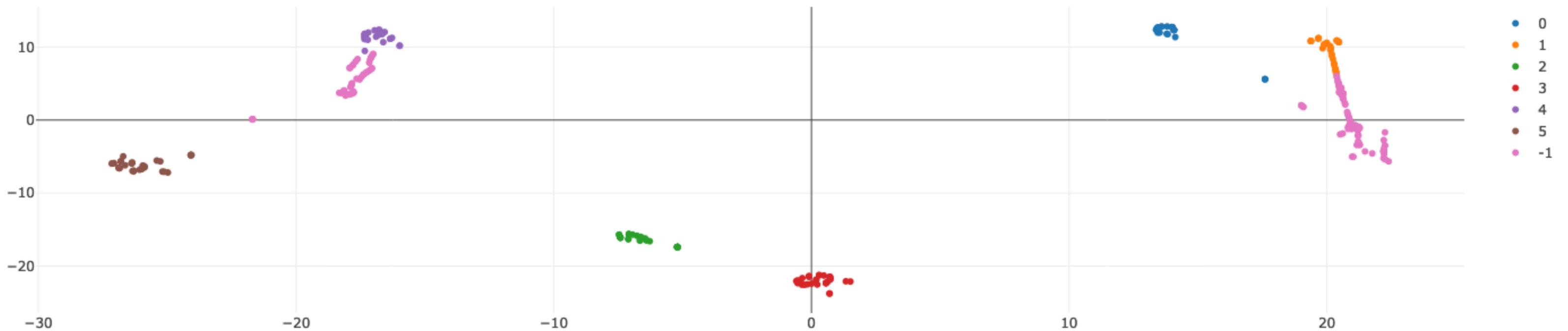
- reputation_male





That's the button that blocks
users ❤️

t-SNE visualisation



Clustered Users

Show entries

Se

User	Cluster	Is Spammer
	3	True

Clusters Descriptive Statistics

Show entries

Search:

Cluster	Feature	Mean	Std	Min	25%	50%	75%	Max
-1	photographer_num_pictures	1.53	4.42	0.00	1.00	1.00	1.00	45.00
0	photographer_num_pictures	1.08	0.40	0.00	1.00	1.00	1.00	2.00
1	photographer_num_pictures	1.76	5.80	0.00	1.00	1.00	1.00	36.00
2	photographer_num_pictures	0.93	0.37	0.00	1.00	1.00	1.00	2.00
3	photographer_num_pictures	0.58	0.56	0.00	0.00	1.00	1.00	2.00
4	photographer_num_pictures	0.92	0.27	0.00	1.00	1.00	1.00	1.00
5	photographer_num_pictures	1.22	1.18	0.00	1.00	1.00	1.00	6.00

Showing 1 to 7 of 7 entries (filtered from 1,127 total entries)

Previous **1** Next

Clusters Descriptive Statistics

Show entries

Search:

Cluster	Feature	Mean	Std	Min	25%	50%	75%	Max
-1	photographer_num_pictures	1.53	4.42	0.00	1.00	1.00	1.00	45.00
0	photographer_num_pictures	1.08	0.40	0.00	1.00	1.00	1.00	2.00
1	photographer_num_pictures	1.76	5.80	0.00	1.00	1.00	1.00	36.00
2	photographer_num_pictures	0.93	0.37	0.00	1.00	1.00	1.00	2.00
3	photographer_num_pictures	0.58	0.56	0.00	0.00	1.00	1.00	2.00
4	photographer_num_pictures	0.92	0.27	0.00	1.00	1.00	1.00	1.00
5	photographer_num_pictures	1.22	1.18	0.00	1.00	1.00	1.00	6.00

Showing 1 to 7 of 7 entries (filtered from 1,127 total entries)

Clusters Analysis

Overall Silhouette Score 0.7743101569200785

Davies Bouldin Score 0.38803664986451564

Cluster	Silhouette
0.0	0.81
1.0	0.84
2.0	0.60
3.0	0.90
4.0	0.83
5.0	0.73

Cluster Analysis

- Silhouette coefficient:
 - Measures the cohesion and separation of a cluster.
 - Ranges from -1 to 1, where -1 implies an incorrect clustering and 1, a dense and good clustering.
- Davies-Bouldin index:
 - Measures the separation among clusters.
 - Where similarity is the ratio between the intra-cluster distance and inter-cluster distance
 - Ranges from 0 to 1, where lower values indicate better clustering.

Clusters Analysis

Overall Silhouette Score 0.7743101569200785

Davies Bouldin Score 0.38803664986451564

Cluster	Silhouette
0.0	0.81
1.0	0.84
2.0	0.60
3.0	0.90
4.0	0.83
5.0	0.73

Takeaways

Juan De Dios Santos Rivera | @jddiossantos
Data Engineer (Payment Fraud Officer) @ LOVOO
Berlin Buzzwords 2021 | June 15, 2021

Takeaways

What I like

- The system works. It allows us to get cases we wouldn't catch with our current systems.
- Seeing a cluster's spam percentage.
 - Why didn't we catch this user?
 - What should we do to catch this user?
- Good way to create labelled datasets
- The statistics and cluster analysis

Takeaways

What I don't like

- OPTICS can be a bit slow :/

Takeaways

What could be better

- How to interpret the clusters' statistics and analysis
 - I'd like a functionality that shows which cluster has the lower or highest values for a particular feature

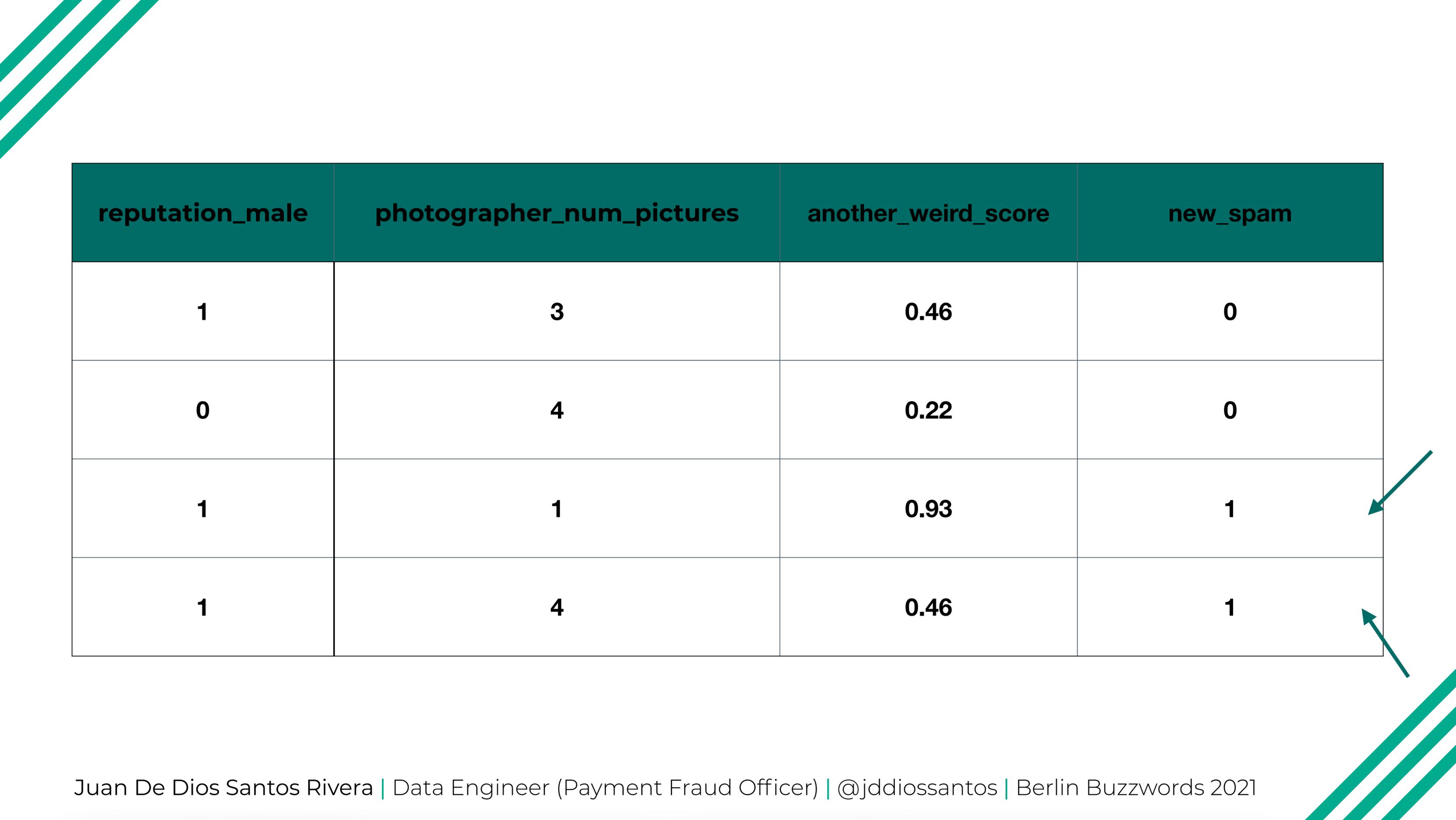
Takeaways

Future work

- We want to fit a supervised learning model using one or more clusters as the positive class and the rest as the negative class.
- Our first idea is to use a decision tree, and automatically build new Conviction rules from what the tree has learned.

reputation_male	photographer_num_pictures	another_weird_score	cluster	is_spam
1	3	0.46	-1	0
0	4	0.22	1	1
1	1	0.93	2	1
1	4	0.46	2	1

reputation_male	photographer_num_pictures	another_weird_score	cluster	is_spam
1	3	0.46	-1	0
0	4	0.22	1	1
1	1	0.93	2	1
1	4	0.46	2	1



reputation_male	photographer_num_pictures	another_weird_score	new_spam
1	3	0.46	0
0	4	0.22	0
1	1	0.93	1
1	4	0.46	1





And now, we fit the model
using this new dataset



✓ Done Clustering ID c2nug39gnfth8cdo6r50

[Download CSV](#) [Fit Tree](#)

Cluster Batch Info

Batch ID
c2mhb91gnftvc7ne3eg

Added
2021-05-27 20:42:21

Status
done

Items Clustered
360

Features Clustered
35

Used Confidences



Decision Tree Info

Status
-

Validation Average F1 Metric Score
-

Validation Standard Deviation Score
-

Fit Decision Tree

Choose positive labels:

- 1
- 0
- 1
- 2
- 3

Cancel

Fit

Antispam
Web UI

Runway

Web Views

Conviction



reputation_max	photographer_num_picture	another_weight_score	cluster	is_spam
1	3	0.66	-1	0
0	4	0.22	1	1
1	1	0.33	2	1
1	4	0.66	2	1

✓ Done Clustering ID c2nug39gnfth8cdo6r50

Download CSV

Fit Tree

Cluster Batch Info

Batch ID

c2mhb91gnftvc7ne3eg

Added

2021-05-27 20:42:21

Status

done

Items Clustered

360

Features Clustered

35

Used Confidences



Decision Tree Info

Status

done

Validation Average F1 Metric Score

0.3313736063133654

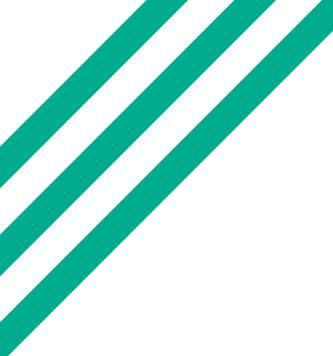
Validation Standard Deviation Score

0.17882785747870775

```

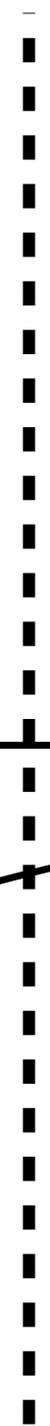
- [redacted] <= 1.07
  | [redacted] <= 2.88
  | | --- reputation_male <= 17.50
  | | | [redacted] <= 0.96
  | | | [redacted] <= 145983.00
  | | | [redacted] <= 130885.00
  | | | [redacted] <= 0.41
  | | | | [redacted] <= 48314.50
  | | | | | --- class: 0
  | | | | | [redacted] > 48314.50
  | | | | | [redacted] <= 1.02
  | | | | | | --- class: 0
  | | | | | [redacted] > 1.02
  | | | | | | --- tracer_confidence <= 0.03
  | | | | | | | --- class: 1
  | | | | | | --- tracer_confidence > 0.03
  | | | | | | | --- photographer_num_pictures <= 0.18
  | | | | | | | | --- class: 1
  | | | | | | | | --- photographer_num_pictures > 0.18
  | | | | | | | | --- class: 0

```



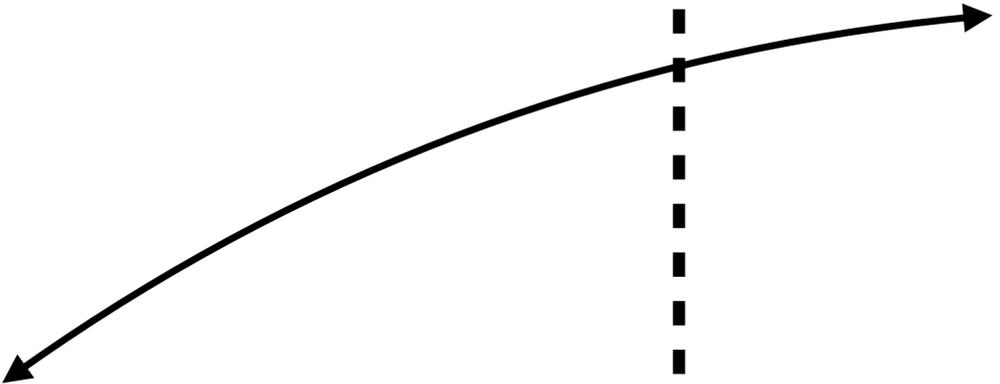
Antispam
Web UI

Runway



Web Views

Conviction



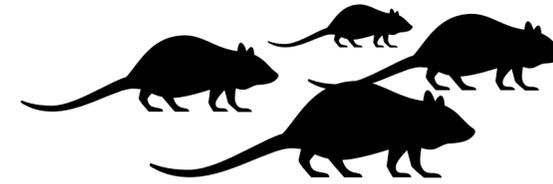
```
-----<= 1.07
-----<= 2.88
reputation_male <= 17.50
-----<= 0.96
-----<= 145983.00
-----entity_isp_users <= 130885.00
-----<= 0.41
-----<= 48314.50
-----class: 0
-----> 48314.50
-----<= 1.02
-----class: 0
-----> 1.02
-----tracer_confidence <= 0.03
-----class: 1
-----tracer_confidence > 0.03
-----photographer_num_pictures <= 0.18
-----class: 1
-----photographer_num_pictures > 0.18
-----class: 0
```





That's the future work.

Thanks :)



Say no to spam!

- Goka (our open-sourced Kafka library for Go): <https://github.com/lovoo/goka>
- *Fighting Spam in Dating Apps*: <https://dl.gi.de/handle/20.500.12116/21707>