

# Building a SQL editor around “fast”

Rill Developer’s no-run-button  
auto-EDA modeling tool

Hamilton Ulmer

Rill Data

@hamiltonulmer on Twitter





# Bids Sample Dash

Edit Metrics

Deploy

Last 24 Hours Mar 7-8, 2023 (1:00AM-1:00AM) Comparing to previous day Metric trends by hour

No filters selected

## All Measures

Media Spend  
\$447.9k  
+\$270.0k 151%



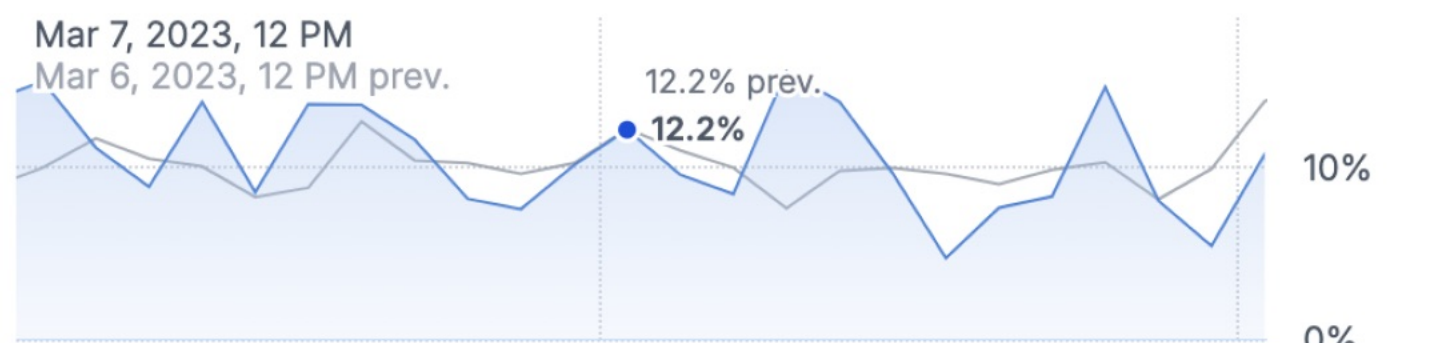
Bids  
876.2k  
+307.4k 54%



Impressions  
83.5k  
+26.4k 46%



Win Rate  
9.5%  
-0.5%



Clicks  
238  
-26 -9%



CTR  
0.3%  
-0.2%



## All Dimensions

showing top values by Media Spend

Adomain	Media Spend	Percentage
instacart.com	\$60.4k	525%
leafly.ca	\$55.1k	4%
hyundaiusa.com	\$31.3k	-27%
amgen.com	\$28.8k	455%
amfam.com	\$24.4k	4%
espaciourbano.cl	\$20.2k	632%
bacardi.com	\$14.2k	5%

Site Domain	Media Spend	Percentage
Not Available	\$253.9k	186%
today.com	\$22.0k	4%
soapoperaspy.com	\$17.3k	4%
eonline.com	\$16.6k	5%
history-a2z.com	\$12.8k	5%
msnbc.com	\$12.7k	2%
nbcnews.com	\$10.5k	5%

Bid Floor Bucket	Media Spend	Percentage
0.00-0.05	\$195.2k	326%
10.00-11.00	\$64.9k	570%
9.50-10.00	\$29.3k	7%
12.00-13.00	\$12.5k	4%
4.00-4.25	\$10.0k	881%
2.00-2.05	\$6.7k	607%
1.25-1.30	\$6.6k	140%

Device OS	Media Spend	Percentage
Android	\$166.6k	82%
iOS	\$88.6k	205%

Advertiser Name	Media Spend	Percentage
Instacart	\$60.4k	525%
Leafly	\$55.1k	4%
Hyundai	\$31.3k	-27%
Amgen Inc	\$28.8k	455%
American Family Insurance	\$24.4k	4%
GlaxoSmithKline	\$22.0k	453%
ESPACIO URBANO	\$20.2k	632%

App Name	Media Spend	Percentage
Not Available	\$223.9k	114%
Bally Sports	\$129.6k	413%
BallySports	\$23.4k	368%
ESPN	\$19.6k	1%
hulu	\$10.2k	5%
Healthline - FL - Mobile - Ne...	\$6.4k	429%
Tile Pair 3D - Tile Connect 3D	\$2.0k	7%

Campaign Name	Media Spend	Percentage
Leafly_MarketMight	\$55.1k	4%
Instacart_BrandBounce	\$49.4k	2%
Hyundai_Revitalize	\$31.3k	-27%
American Family Insurance...	\$24.4k	4%
Amgen Inc_BrandRamp	\$21.9k	1%
ESPACIO URBANO_PrimeP...	\$20.2k	632%
Bacardi_ExpandNow	\$14.2k	5%

Device OSV	Media Spend	Percentage
16	\$78.4k	223%
10	\$67.0k	120%

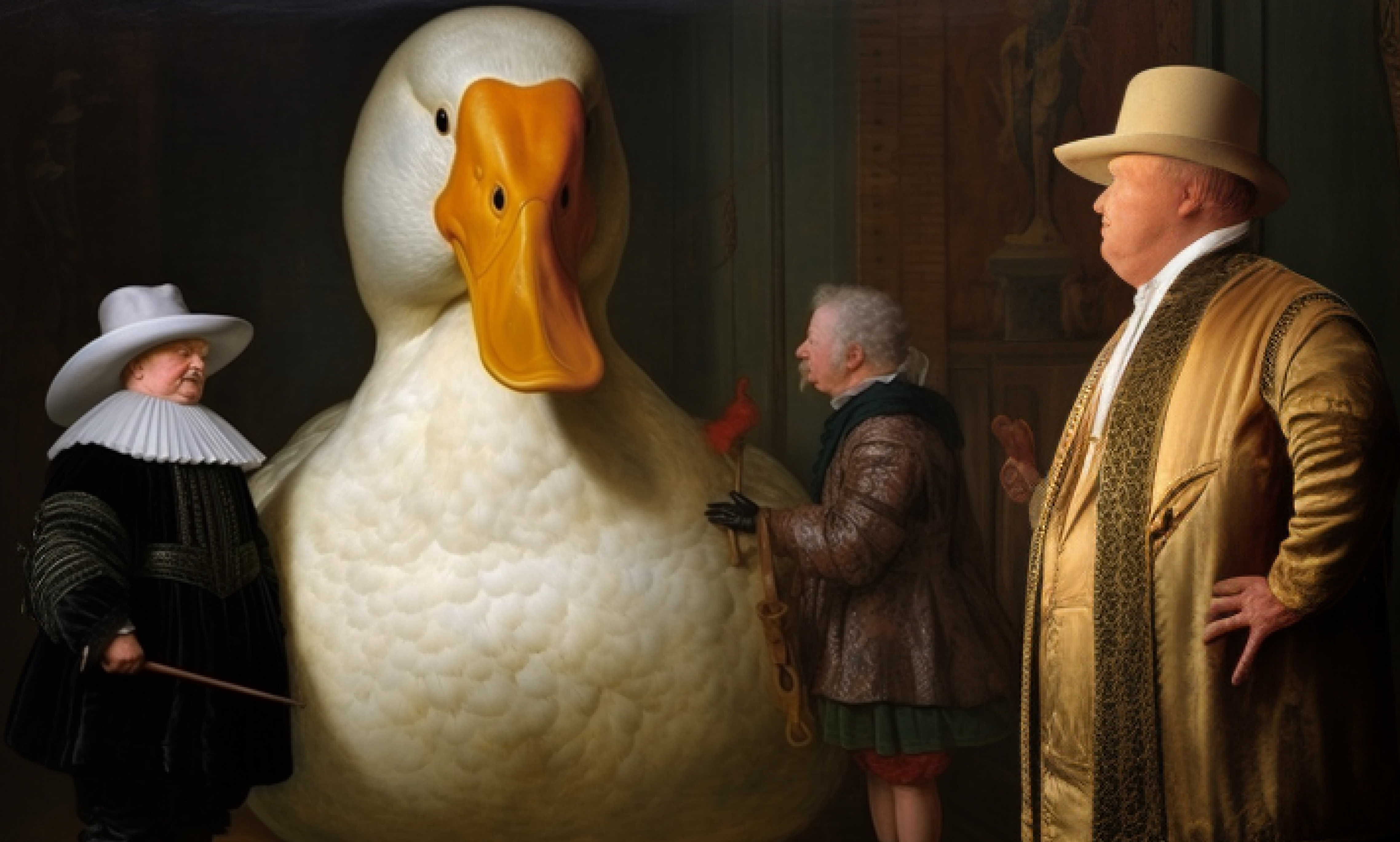
App or Site	Media Spend	Percentage
App	\$234.0k	208%
Site	\$214.0k	109%

Auction Type	Media Spend	Percentage
Second Price	\$261.3k	388%
First Price	\$186.6k	49%

Creative type	Media Spend	Percentage
VIDEO	\$371.8k	355%
BANNER	\$76.2k	-20%
NATIVE	\$7.99	no data

Device Region	Media Spend	Percentage
USA/OH	\$91.4k	3%
USA/CA	\$36.6k	36%













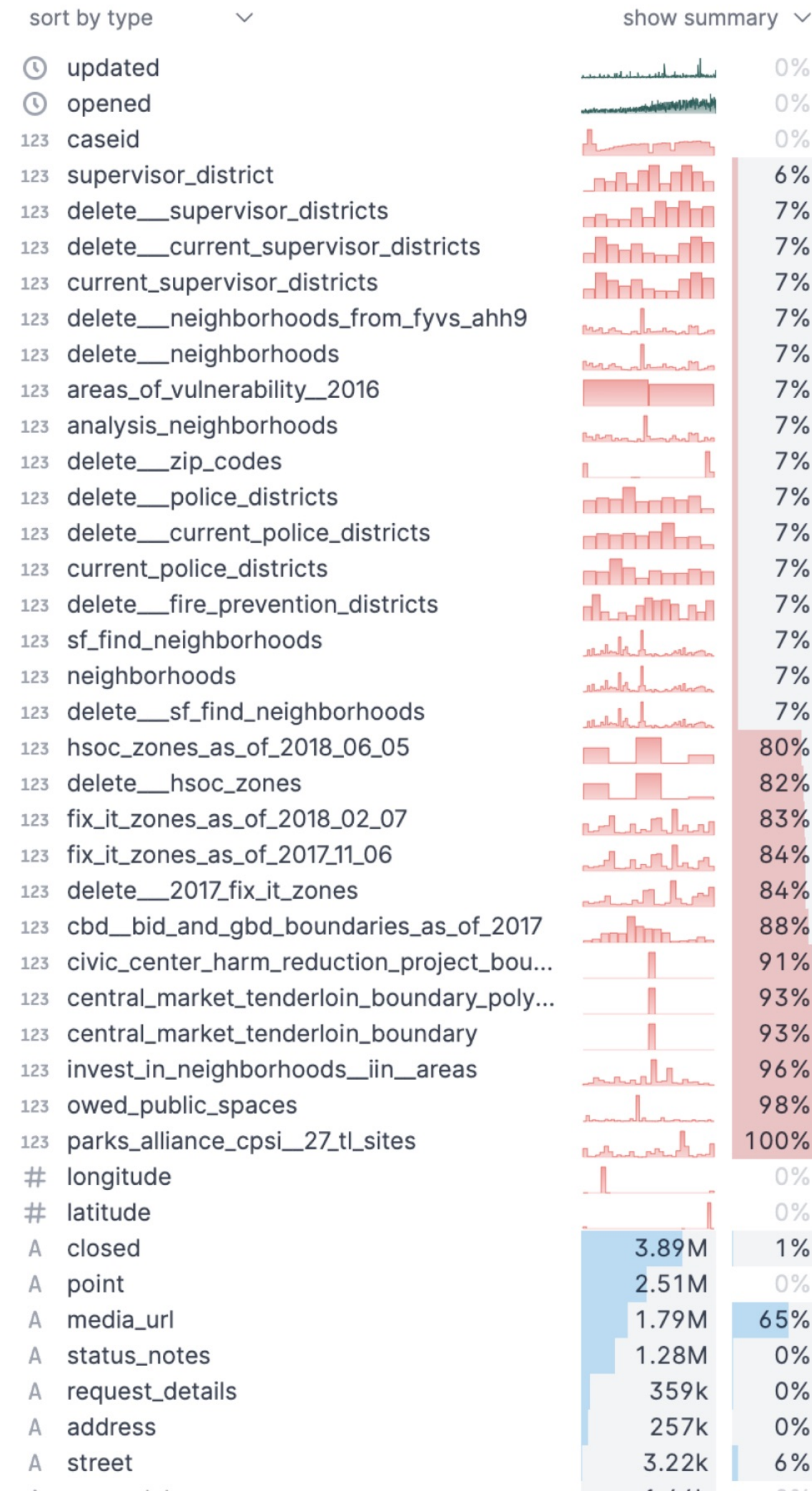
🤔 **what if your SQL editor  
didn't need this button?**



```
release_criteria_v1 X +
RUN SAVE SHARE SCHEDULE
1 WITH rc_included_tests AS (
2   -- Filter and correct table that defines the aggreg
3   SELECT
4     * REPLACE (
5       IFNULL(rc_tier, 5) AS rc_tier,
6       IFNULL(rc_test_aggregator, 'use_existing') AS
7       IFNULL(rc_subtest_aggregator, 'use_existing')
8       IFNULL(rc_replicate_aggregator, 'use_existing')
9       IFNULL(rc_ignore_first_replicates, 0) AS rc_ig
10      IFNULL(rc_target_type, 'none') AS rc_target_ty
11      IFNULL(test_extra_options, '') AS test_extra_o
12    )
```



# Rill's Auto-Profiler; spending the user's attention budget on EDA



## Compactness

x-pixel direction is most costly  
should feel like a simple receipt

## Easy scanning

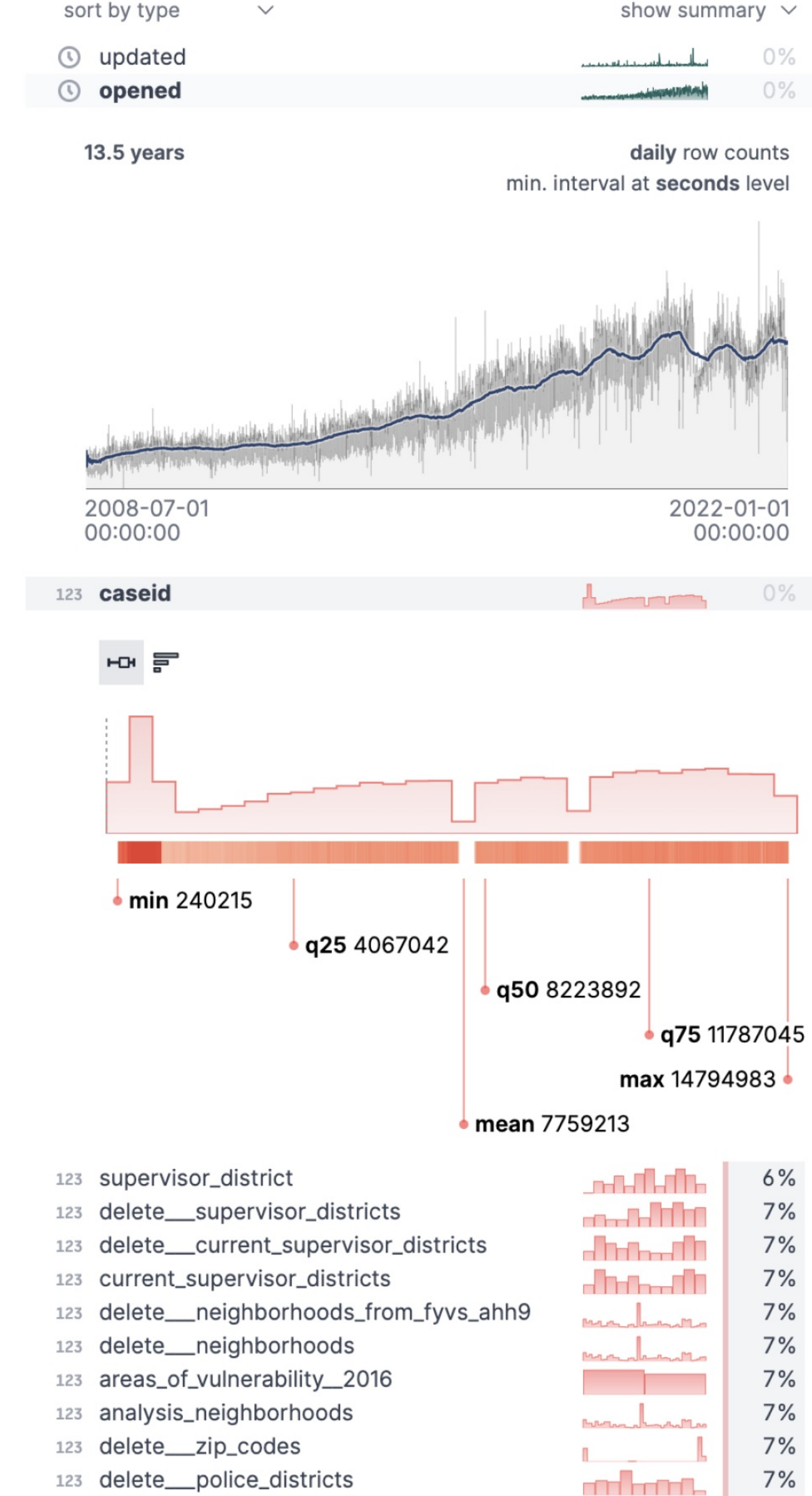
group by type, then nullity / uniques  
distribution plots enable gut checks to be explored further

## Provides details on demand

Richer diagnostics when you click on a profile

## Rendered iteratively

As DuckDB returns results, we update the interface  
Reduces the cognitive burden of waiting





- Sources
  - ISSUERS
  - OFFERINGS
  - RECIPIENTS
  - RELATEDPERSONS
  - SIGNATURES
  - + Add source
- Models
  - > offerings\_issuers\_model ...
  - + Add model
- Dashboards
  - offerings\_issuers
  - + Add dashboard

```

1
2 SELECT
3   SALE_DATE AS offering_saledate,
4   INDUSTRYGROUPTYPE AS offering_IndustryGroupType,
5   INVESTMENTFUNDTYPE AS offering_InvestmentFundType,
6   ISSUERS.ENTITYNAME AS issuer_EntityName,
7   RECIPIENTS.RECIPIENTNAME AS recipient_name,
8   CAST(
9     CASE
10      WHEN TOTALOFFERINGAMOUNT == 'Indefinite' THEN 0
11      ELSE TOTALOFFERINGAMOUNT
12    END
13    AS BIGINT) AS TotalOfferingAmount,
14  *
15 FROM
16   OFFERINGS
17   LEFT JOIN ISSUERS ON OFFERINGS.ACCESSIONNUMBER = ISSUERS.ACCESSIONNUMBER
18   LEFT JOIN RECIPIENTS ON OFFERINGS.ACCESSIONNUMBER = RECIPIENTS.ACCESSIONNUMBER
19 WHERE SALE_DATE BETWEEN DATE '2020-01-01' AND current_date()
20
    
```

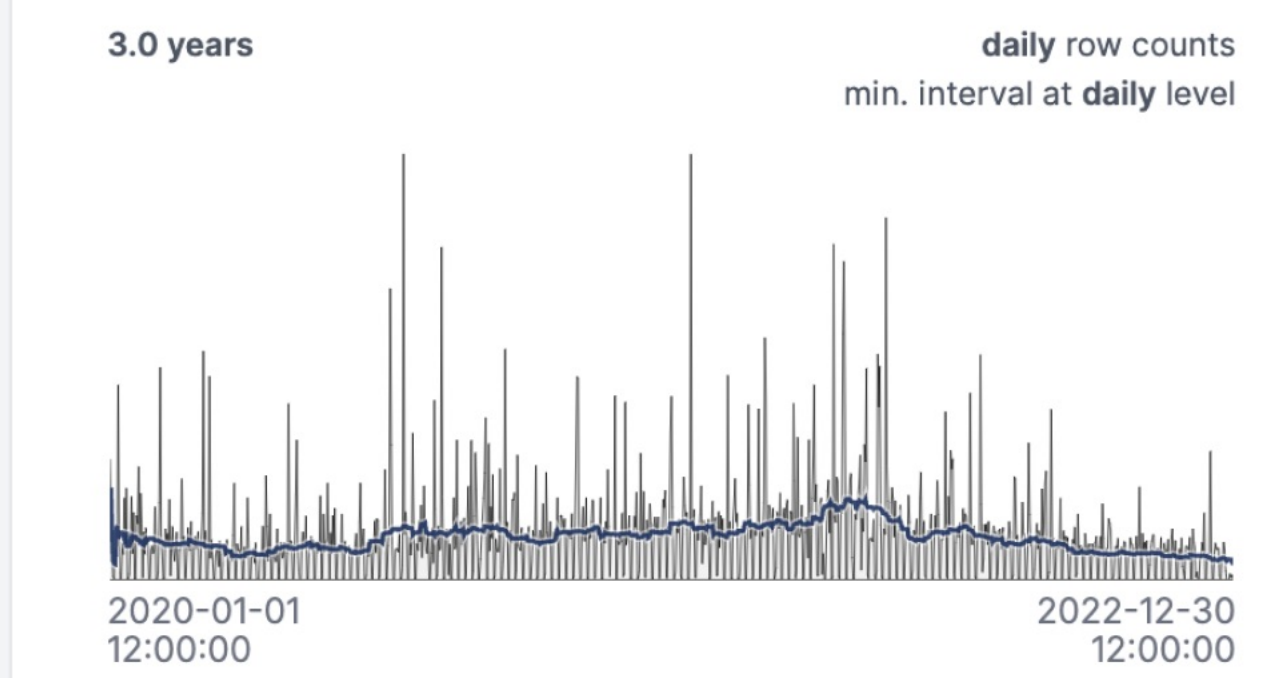
9.71% of source rows  
6 columns added  
**137,340 rows**  
**85 columns**

Referenced in this model

ISSUERS	552k rows
OFFERINGS	545k rows
RECIPIENTS	318k rows

Selected columns  
sort by type show summary

SALE\_DATE 0%



#	🕒 offering...	A offering_IndustryGroupType	A offering_InvestmentFundType	A issuer_EntityName
1	2022-01-...	Investing	null	3x5 Kaleidoscope Investors, LLC
2	2022-01-...	Investing	null	3x5 Kaleidoscope Investors, LLC
3	2020-08-...	Pooled Investment Fund	Hedge Fund	Faraday Capital Partners II, L.P.
4	2022-01-...	Other	null	EcoSoul Home Inc.
5	2021-03-...	Pooled Investment Fund	Hedge Fund	UBC Russell 3000 Index Fund, a series of Inve
6	2022-02-...	Other Technology	null	Biopharmaceutical Research Co
7	2022-02-...	Other	null	Filtravate, Inc.
8	2022-01-...	Other	null	TPE - NMC LeaseCo., LLC
9	2022-01-...	Manufacturing	null	NATIONAL MOLDING, LLC
10	2022-02-...	Pooled Investment Fund	Other Investment Fund	Lombard International Life Assurance Co Sep..
11	2021-09-	Investing	null	Selvarey Bum SPV LLC

🕒 offering_saledate	0%
123 TotalOfferingAmount	0%
123 TOTALNUMBERALREADYINVESTED	0%
123 TOTALAMOUNTSOLD	0%
123 SALESCOMM_DOLLARAMOUNT	0%
123 MINIMUMINVESTMENTACCEPTED	0%
123 ISSUER_SEQ_KEY	0%
123 GROSSPROCEEDSUSED_DOLLARAMOUNT	0%
123 FINDERSFEE_DOLLARAMOUNT	0%
123 YEAROFINC_VALUE_ENTERED	11%
123 RECIPIENT_SEQ_KEY	63%
123 NUMBERNONACCREDITEDINVESTORS	96%
A ACCESSIONNUMBER	103k 0%
A ACCESSIONNUMBER:1	103k 0%
A issuer_EntityName	77.9k 0%
A ENTITYNAME	77.9k 0%
A CIK	76.8k 0%
A STREET1	42.2k 0%
A ISSUERPHONENUMBER	39.3k 0%
A TOTALOFFERINGAMOUNT:1	33.3k 0%
A PREVIOUSACCESSIONNUMBER	18.1k 68%
A TOTALREMAINING	17.5k 0%
A ACCESSIONNUMBER:2	17.3k 63%





## Problems

**Profiling scales poorly in column space.**

And you can't sample columns.

Queries are multithreaded, but sequences of queries are serial.

**So queries need to be *fast*.**

**Rendering hundreds of profiles will weigh down the DOM** and make the app janky.

A single sparkline is cheap, but **hundreds or thousands of sparklines** could make the browser janky.

**Most profiles don't actually need to be recalculated** if the query DAG hasn't changed for that column.

## Solutions

**query priority queue** – only prioritize running queries the user needs to see.

**approximation methods** – humans can only perceive so much visual accuracy anyway, so don't be overly precise.

**shape-preserving dimensionality reduction methods** – M4, AM4, etc

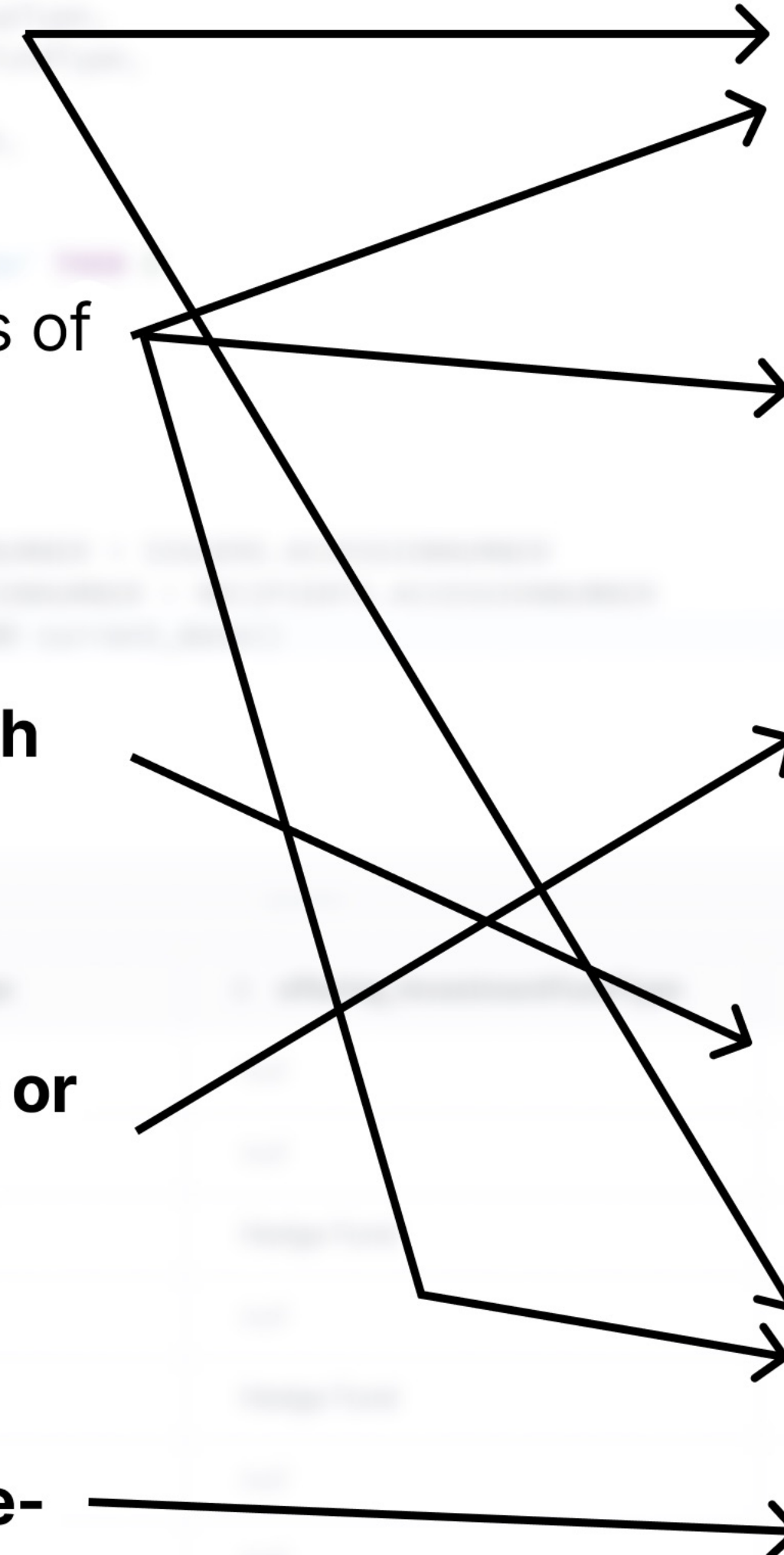
**virtualization & tanstack-style queries** – only query / render what's needed in the viewport.

**aggressive caching** – cache things the user has already seen.

**AST-aware profiling**

**\*COMING**

**SOON\***





A group of ducks dressed in formal, historical-style clothing are seated around a table. One duck in the center is wearing a black hat with a white plume and a dark dress. Another duck to the right is wearing a white dress with a large white ruff. The table is set with plates of food, including what looks like a turkey leg, and glasses. The background is dark and filled with more ducks in similar attire.

**Thanks for listening!**

1. follow along at [@hamiltonulmer](#) & [@rilldata](#) on twitter
2. talk to me @ the [Motherduck](#) afterparty
3. visit [rilldata.com/careers](https://rilldata.com/careers)

