# What are we asking from **Tabular Data**?

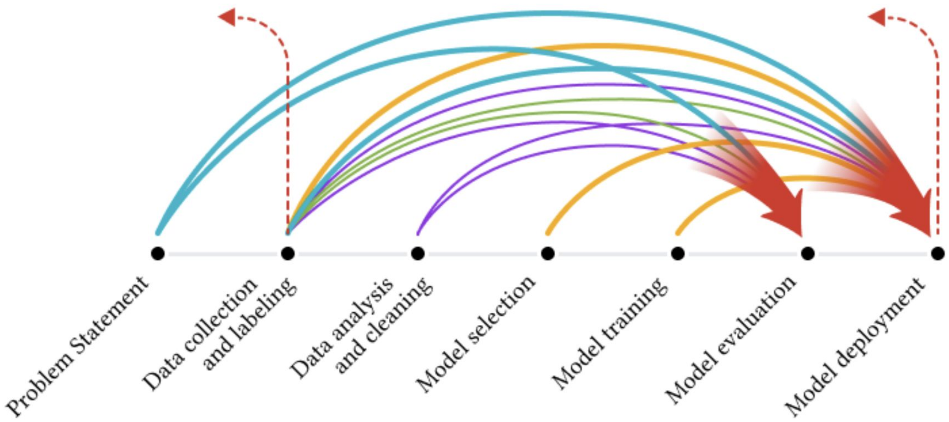
**Madelon Hulsebos**

6 December, Copenhagen

AI for Tabular Data workshop @ EurIPS 2025

As a data scientist,
in the “real world”,
I realized 3 things…

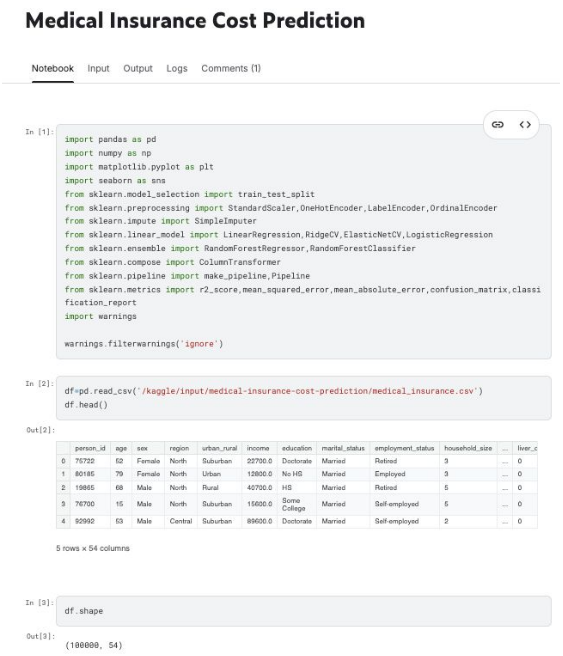
# **Realization 1:** most of my work was data work




Data science = “80% data work, 20% model work”

“Everyone wants to do the model work, not the data work”: Data Cascades in High-Stakes AI, Sambasivan et al., 2021

# **Realization 2**: everyone is doing the same…



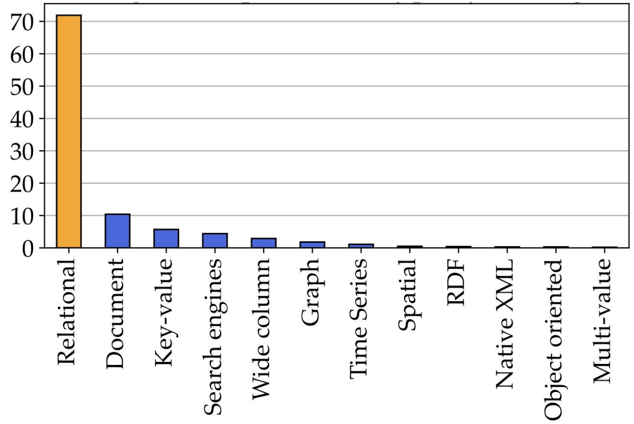


**over, and over, and over again…**

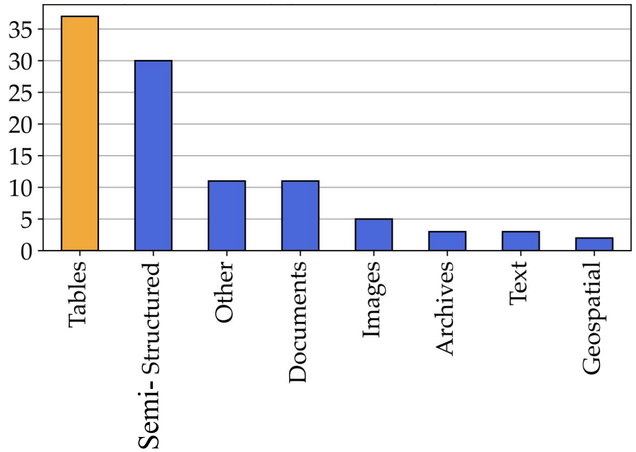

There must be latent patterns to *learn* (data, code, etc).

# **Realization 3:** tables prevail in the org data landscape

Popularity of database systems

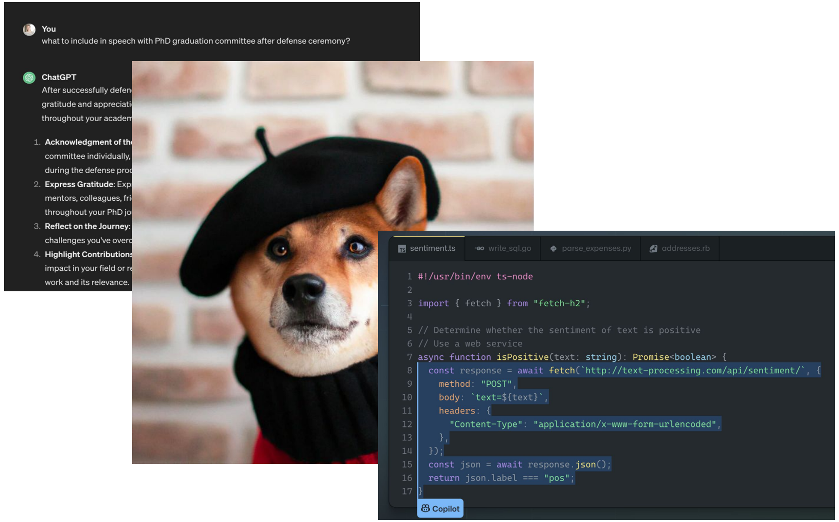


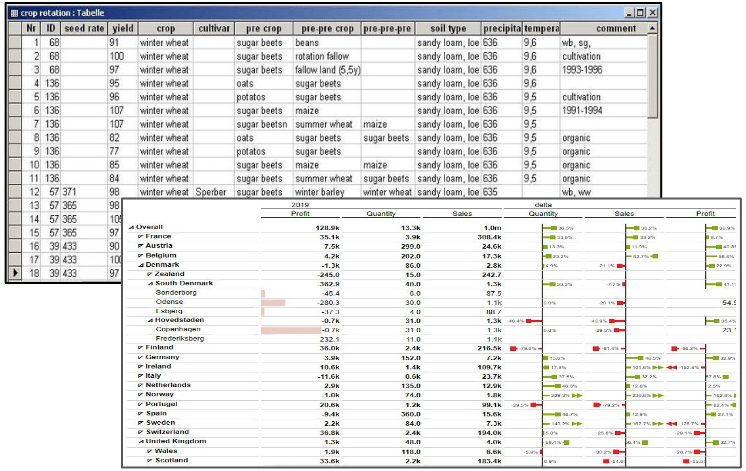

Distribution Google Dataset Search




For a reason: **tables serve high-value decisions**

# Surprisingly, “tables” ignored as modality in neural AI



crop rotation : Tabelle

| Nr | ID | seed rate | yield | crop | cultivar | pre crop | pre-pre crop | pre-pre-pre | soil type | precipita | tempera | comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 68 | | 91 | winter wheat | | sugar beets | beans | | sandy loam, loe | 636 | 9,6 | wb, sg, |
| 2 | 68 | | 100 | winter wheat | | sugar beets | rotation fallow | | sandy loam, loe | 636 | 9,6 | cultivation |
| 3 | 68 | | 97 | winter wheat | | sugar beets | fallow land (5,5y) | | sandy loam, loe | 636 | 9,6 | 1993-1996 |
| 4 | 136 | | 95 | winter wheat | | oats | sugar beets | | sandy loam, loe | 636 | 9,6 | |
| 5 | 136 | | 96 | winter wheat | | potatos | sugar beets | | sandy loam, loe | 636 | 9,5 | cultivation |
| 6 | 136 | | 107 | winter wheat | | sugar beets | maize | | sandy loam, loe | 636 | 9,5 | 1991-1994 |
| 7 | 136 | | 107 | winter wheat | | sugar beetsn | summer wheat | maize | sandy loam, loe | 636 | 9,5 | |
| 8 | 136 | | 82 | winter wheat | | oats | sugar beets | sugar beets | sandy loam, loe | 636 | 9,5 | organic |
| 9 | 136 | | 77 | winter wheat | | potatos | sugar beets | | sandy loam, loe | 636 | 9,5 | organic |
| 10 | 136 | | 85 | winter wheat | | sugar beets | maize | maize | sandy loam, loe | 636 | 9,5 | organic |
| 11 | 136 | | 84 | winter wheat | | sugar beets | summer wheat | sugar beets | sandy loam, loe | 636 | 9,5 | organic |
| 12 | 57 | 371 | 98 | winter wheat | Sperber | sugar beets | winter barley | winter wheat | sandy loam, loe | 635 | | wb, ww |
| 13 | 57 | 365 | 98 | | | | | | | | | |
| 14 | 57 | 365 | 10 | | | | | | | | | |
| 15 | 57 | 365 | 97 | | | | | | | | | |
| 16 | 39 | 433 | 90 | | | | | | | | | |
| 17 | 39 | 433 | 10 | | | | | | | | | |
| 18 | 39 | 433 | 97 | | | | | | | | | |

| | 2019 | | | delta | | |
|---|---|---|---|---|---|---|
| | Profit | Quantity | Sales | Quantity | Sales | Profit |
| Overall | 128.9k | 13.3k | 1.0m | 36.5% | 36.2% | 30.9% |
| France | 35.1k | 3.9k | 308.4k | 33.8% | 33.2% | 8.7% |
| Austria | 7.5k | 299.0 | 24.6k | 13.3% | 11.9% | 40.8 |
| Belgium | 4.2k | 202.0 | 17.3k | 23.2% | 82.7% | 98.6% |
| Denmark | -1.3k | 86.0 | 2.8k | 4.9% | -21.1% | 22.9% |
| Zealand | -245.0 | 15.0 | 242.7 | | | |
| South Denmark | -362.9 | 40.0 | 1.3k | 33.3% | -7.7% | 41.1 |
| Sonderborg | -45.4 | 6.0 | 87.5 | | | |
| Odense | -280.3 | 30.0 | 1.1k | 0.0% | -20.1% | 54.5 |
| Esbjerg | -37.3 | 4.0 | 88.7 | | | |
| Hovedstaden | -0.7k | 31.0 | 1.3k | -40.4% | -40.9% | 35.4% |
| Copenhagen | -0.7k | 31.0 | 1.3k | 0.0% | -29.5% | 23.1 |
| Frederiksberg | 232.1 | 11.0 | 1.1k | | | |
| Finland | 36.0k | 2.4k | 216.5k | -79.6% | -81.4% | -86.2% |
| Germany | -3.9k | 152.0 | 7.2k | 15.0% | 46.3% | 32.9% |
| Ireland | 10.6k | 1.4k | 109.7k | 17.8% | 101.6% | -152.5% |
| Italy | -11.6k | 0.6k | 23.7k | 37.5% | 37.2% | 57.6% |
| Netherlands | 2.9k | 135.0 | 12.9k | 46.3% | 12.6% | 2.5% |
| Norway | -1.0k | 74.0 | 1.8k | 229.3% | 230.8% | 162.8% |
| Portugal | 20.6k | 1.2k | 99.1k | -29.5% | -79.2% | 82.4% |
| Spain | -9.4k | 360.0 | 15.6k | 46.7% | 12.9% | 27.1% |
| Sweden | 2.2k | 84.0 | 7.3k | 143.2% | 167.7% | -128.7% |
| Switzerland | 36.8k | 2.4k | 194.0k | 5.0% | -25.6% | -29.1% |
| United Kingdom | 1.3k | 48.0 | 4.0k | 68.4% | 56.4% | 32.7% |
| Wales | 1.9k | 118.0 | 6.6k | -5.9% | -30.2% | -28.7% |
| Scotland | 33.6k | 2.2k | 183.4k | 0.9% | -54.6 | -55.5 |

**Text**, **Images**, **Code**…

**Tables**?!

# Did a PhD on table semantics, but had a larger vision.

**The Table Representation Learning workshop @ NeurIPS 2022 was born.**

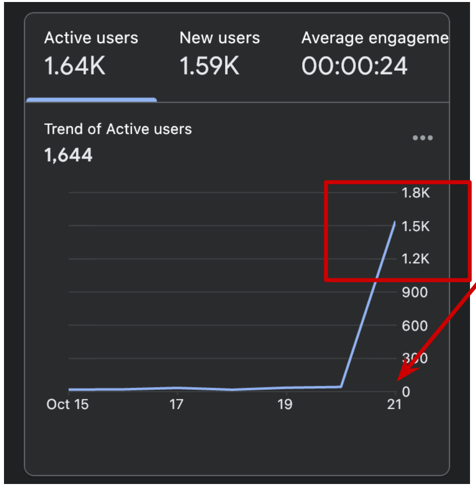
We received what we expected:

- Tabular QA / text-to-SQL,
- Synthetic data generation,
- Data preparation, etc.

And… neural models for predictive tabular ML.

# Anecdotes on neural models for tabular predictive ML

- Received neural predictive ML papers, rejected from main ML confs(?)
- Loud "pro-XGBoost" camp vs. small "pro-neural models" camp.
- Betted on pre-trained neural models for tables in '18: let's facilitate vision!
- And, 1 paper intro'd a pre-trained cross-table model: *TabPFN*. It went viral:




Frank's TabPFN tweet

Congrats team TabPFN, TabICL, ConTexTab, etc for pushing through and heading leaderboards!

# Tables weren't really cool in AI

But something happened in a tiny room in New Orleans at TRL @ NeurIPS 2022.

Great vibes, a wildly diverse community, trying to connect the dots.

**Tables were Back.**

# Fast-forward to 2025.

**Tabular AI is the “new hot topic” (quote CV researcher)**

**Tabular AI is Europe-led** 🇪🇺 (but let’s diversify).

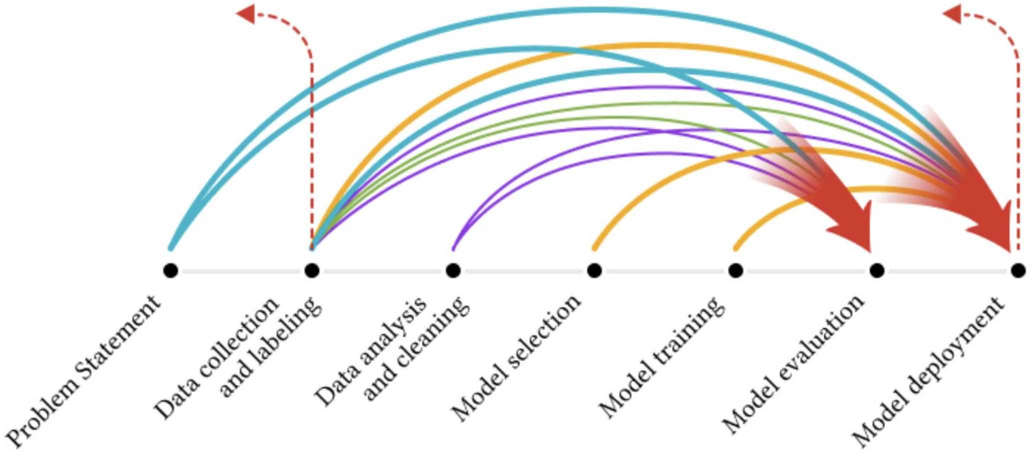
From an anonymous peer:

> *is there no tab workshop this year at main conf bc you all took to eurips?*
>
> *either way wish I could be at the workshop :)*

**And we’re only just beginning.**

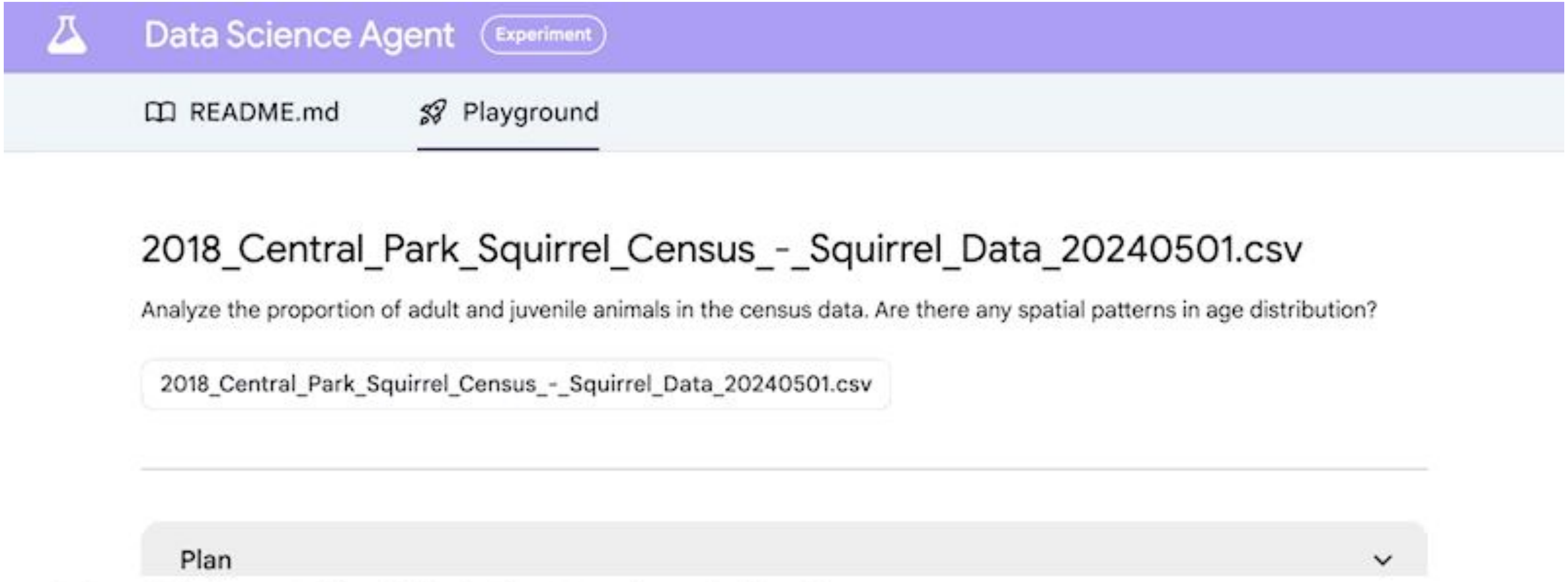
So, what *are we asking* from tabular data?

# Tabular pipelines are multi-faceted




*Replace “model” with “tool/dashboard” and it’s BI.

# LLM-enthusiasts make believe that DS is solved…



# It's not

| Model | *SDE* (69 tasks) | | *PM* (28 tasks) | | *DS* (14 tasks) | |
|---|---|---|---|---|---|---|
| | **Success** | **Score** | **Success** | **Score** | **Success** | **Score** |
| Closed model APIs | | | | | | |
| Claude-3.5-Sonnet | 30.43 | 38.02 | 35.71 | 51.31 | 14.29 | 21.70 |
| Gemini-2.0-Flash | 13.04 | 18.99 | 17.86 | 31.71 | 0.00 | 6.49 |
| GPT-4o | 13.04 | 19.18 | 17.86 | 32.27 | 0.00 | 4.70 |
| Gemini-1.5-Pro | 4.35 | 5.64 | 3.57 | 13.19 | 0.00 | 4.82 |
| Amazon-Nova-Pro-v1 | 2.90 | 6.07 | 3.57 | 12.54 | 0.00 | 3.27 |
| Open-weight models | | | | | | |
| Llama-3.1-405b | 5.80 | 11.33 | 21.43 | 35.62 | 0.00 | 5.42 |
| Llama-3.3-70b | 11.59 | 16.49 | 7.14 | 19.83 | 0.00 | 4.70 |
| Qwen-2.5-72b | 7.25 | 11.99 | 10.71 | 22.90 | 0.00 | 5.42 |
| Llama-3.1-70b | 1.45 | 4.77 | 3.57 | 15.16 | 0.00 | 5.42 |
| Qwen-2-72b | 2.90 | 3.68 | 0.00 | 7.44 | 0.00 | 4.70 |

(TheAgentCompany: Benchmarking LLM Agents on Consequential Real World Tasks, Xu et al., 2024)

# It all starts with retrieving the right data

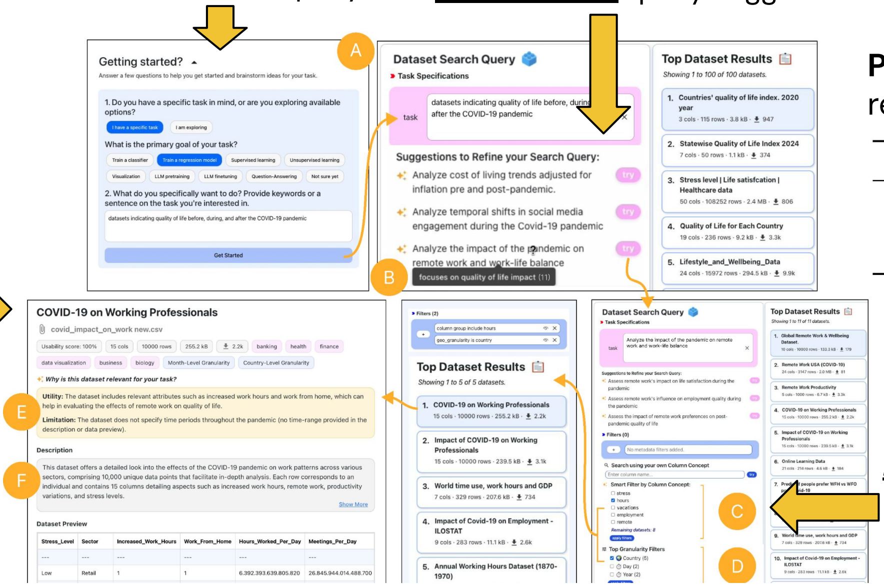
# What table to use for my task?

**In 2025**, we get "AGI" but *it still takes weeks to find the right dataset.*

<u>*Task-driven*</u> query

<u>*Result-driven*</u> query suggestions

<u>*Task-driven*</u> dataset relevance




<u>*Data-driven*</u> column suggestions

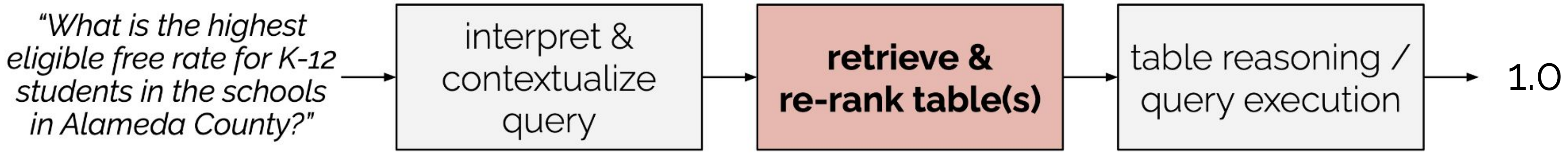
**Proactive search** makes it easier to find more relevant datasets, with higher success rate.

| Condition | Ease-of-use | Relevance | # Successes |
|---|---|---|---|
| (A) Kaggle | $\mu$=3.08; $\sigma$=0.51 | $\mu$=3.25; $\sigma$=1.05 | 7 *of* 12 |
| (B) Semantic Baseline | $\mu$=3.75; $\sigma$=0.45 | $\mu$=3.25; $\sigma$=0.86 | 6 *of* 12 |
| (C) DataScout | $\mu$=4.75; $\sigma$=0.45 | $\mu$=3.67; $\sigma$=0.78 | 10 *of* 12 |

*Rethinking Dataset Discovery with DataScout*
Lin, R., Chopra, B., Lin, W., Shankar, S., Hulsebos, M., Parameswaran, A., 2025.

UC Berkeley CWI

# How to query tables end-to-end? ~RAG over tables




How to best retrieve tables in end-to-end tabular QA? Some findings:

- BM25 **less effective than for text**, requires highly descriptive metadata.
- Retrieval over generated summaries helps (but tricky for enterprise-y tables)
- **Row-level retrieval most effective for relational DBs**, but infeasible (many, large tables).

There's a very long road ahead. Imagine predictive questions over data lakes?

*TARGET: benchmarking Table Retrieval for Generative Tasks*
Ji, X., Parker, G., Parameswaran, A., Hulsebos, M., 2024

*Metadata Matters in Dense Table Retrieval*
Gomm, D., Hulsebos, M., 2025

# Are We Asking the Right Questions?

In end-to-end tabular data analysis, we desire queries that perfectly express the ***insight need***: the data, the operation, and output.

Such queries are **platonic**, we won't see them.

But we need to understand **queries for e2d tabular data analysis** to build better systems and eval capabilities. (*join Daniel's talk for more*)

*Are We Asking the Right Questions? On Query Ambiguity in Tabular Data Analysis*
Gomm, D., Wolff, C., Hulsebos, M., 2025

CWI

# Tables are complex, they require context

# What is this table about?

crop rotation : Tabelle

| Nr | ID | seed rate | yield | crop | cultivar | pre crop | pre-pre crop | pre-pre-pre | soil type | precipita | tempera | comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 68 | | 91 | winter wheat | | sugar beets | beans | | sandy loam, loe | 636 | 9,6 | wb, sg, |
| 2 | 68 | | 100 | winter wheat | | sugar beets | rotation fallow | | sandy loam, loe | 636 | 9,6 | cultivation |
| 3 | 68 | | 97 | winter wheat | | sugar beets | fallow land (5,5y) | | sandy loam, loe | 636 | 9,6 | 1993-1996 |
| 4 | 136 | | 95 | winter wheat | | oats | sugar beets | | sandy loam, loe | 636 | 9,6 | |
| 5 | 136 | | 96 | winter wheat | | potatos | sugar beets | | sandy loam, loe | 636 | 9,5 | cultivation |
| 6 | 136 | | 107 | winter wheat | | sugar beets | maize | | sandy loam, loe | 636 | 9,5 | 1991-1994 |
| 7 | 136 | | 107 | winter wheat | | sugar beetsn | summer wheat | maize | sandy loam, loe | 636 | 9,5 | |
| 8 | 136 | | 82 | winter wheat | | oats | sugar beets | sugar beets | sandy loam, loe | 636 | 9,5 | organic |
| 9 | 136 | | 77 | winter wheat | | potatos | sugar beets | | sandy loam, loe | 636 | 9,5 | organic |
| 10 | 136 | | 85 | winter wheat | | sugar beets | maize | maize | sandy loam, loe | 636 | 9,5 | organic |
| 11 | 136 | | 84 | winter wheat | | sugar beets | summer wheat | sugar beets | sandy loam, loe | 636 | 9,5 | organic |
| 12 | 57 | 371 | 98 | winter wheat | Sperber | sugar beets | winter barley | winter wheat | sandy loam, loe | 635 | | wb, ww |
| 13 | 57 | 365 | 98 | winter wheat | Sperber | potatos | sugar beets | summer barl | sandy loam, loe | 635 | | cultivation, weed |
| 14 | 57 | 365 | 105 | winter wheat | Sperber | sugar beets | maize | maize | sandy loam, loe | 635 | | 1987-1992 |
| 15 | 57 | 365 | 97 | winter wheat | Sperber | sugar beets | winter wheat | sugar beets | sandy loam, loe | 635 | | |
| 16 | 39 | 433 | 90 | winter wheat | Okapi | summer barley | | | sandy loam, loe | 690 | 8,5 | oats, cultivation, weec |
| 17 | 39 | 433 | 100 | winter wheat | Okapi | oats | | | clay, silt | 690 | 8,5 | 1982-1986 |
| 18 | 39 | 433 | 97 | winter wheat | Okapi | winter wheat | | | clay, silt | 690 | 8,5 | |

Pre crop?
Pre-pre crop?
Pre-pre-pre?

A useful signal for data viz, data prep (integration, missing val mech.), ML, etc.

*Sherlock*: (tiny!) neural model surfaces **table semantics** by column embeddings.

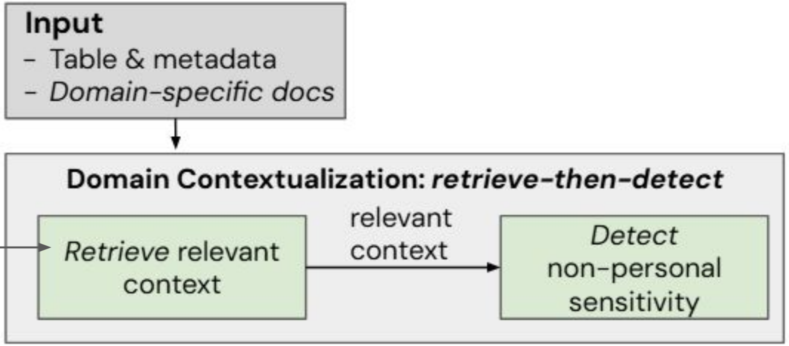
*Sherlock: A Deep Learning Approach to Semantic Data Type Detection,* Hulsebos et al., 2019.

mit media lab

# Is this data sensitive?

Understudied. Beyond fixed types (pii) sensitivity depends on context.
Requires ***context beyond table semantics***.

data governance
policy documents




Result:
- Better precision.
- Context-grounded (LLM) explanations.

Imagine in healthcare: contextualizing **EHR tables**, in **patient-doctor transcripts**, **medicine docs**, etc?

*Towards Contextual Sensitive Data, Telkamp and Hulsebos, 2025.*

But eventually,
we want that insight

# What is the answer to my question?

**Typical questions that drive decisions:**

- Analytical questions → *e.g.* what has happened?
- Predictive queries → *e.g.* what might happen?

Some stats from Gael's keynote at TRL @NeurIPS '24 resonated.

**Pypi #downloads last month** (updated):

- **Scikit-learn: 189,875,197**
- **Pandas: 502,255,215**

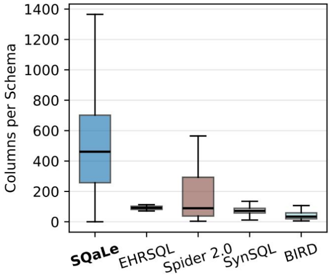
# Analytical questions: we can use SQL

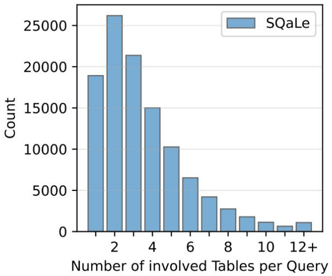
Why enter the *text*-to-*SQL* game?

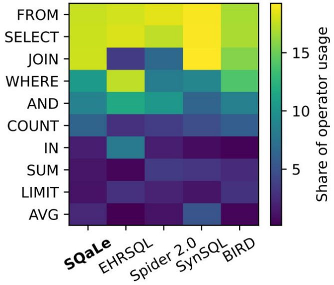
- Saturated (but not solved), everyone just throwing LLMs at it.
- SQL made for humans not machines: it's too flexible, different abstraction?

Then, Cornelius (PhD) convinced me that we need **SQaLe to specialize**.

- Large text-to-SQL dataset
- 500K+ (question, sql, schema)
- Grounded in *real* schemas




TBC

*SQaLe: A large text-to-SQL corpus grounded in real schemas, Wolff, Gomm, and Hulsebos, 2025.*

CWI

What *should we ask* from tabular data, next?

# Generalization versus specialization?

Deriving insights from tabular data to inform decisions is an inherently multi-faceted problem, **perfect for smaller specialized models**, but still looking for the right level of abstraction, which may vary.

Eventually tabular systems are a mix and match:

- Databases and other tools for analytical queries, data prep, etc.
- Some ML models for statistical reasoning
- Some human, document, and LLM contextualization
- Some agentic capabilities
- Some fluid interfaces that connect things together

# Still, many open questions

What is a table? What is a relation? How to deal with different representations?

What data do we encounter in practice?

What queries are asked? What can and should be asked? → check Daniel’s talk

How to contextualize tabular data w/ domain knowledge, e.g. from human input?

How do we want to enable interaction with tabular data?

What is the ideal scope for foundation models? When to generalize vs specialize?

How do we go full-cycle from data collection to decision? What can we “agentize”?

# Where does the community stand?

Excitement:

- **Tabular AI** has gained momentum – it's recognized widely!
- Academic / VC funding is flowing.
- There's much more to **explore and exploit**.

❤ **Tabular AI community**!

You, too? We're looking for committed members to foster the community.
*Reach out if you want to contribute, stay tuned for more initiatives :).*

# Reach out

TRL Lab

Madelon Hulsebos
Faculty

Xue (Effy) Li
Postdoctoral researcher

Daniel Gomm
PhD researcher

Cornelius Wolff
PhD researcher

Jan Henrik Bertrand
Research student (ELLIS MSc Honours Program)

Wojciech Kosiuk
Research student (ELLIS MSc Honours Program)

[madelon@cwi.nl](mailto:madelon@cwi.nl)

https://www.madelonhulsebos.com

https://trl-lab.github.io