



Finding the right table

Petros Likidis, Statistics Sweden



It is hard to find the right
table

The problem with the current
search

How many people live in Sweden?

Population per month by region, age and sex. Year 2025M01 -
2026M08

A naive solution

- Run all table title through an LLM to get the embedding vector
 - A numerical representation of the semantics of the title
- Take the search query run it through the same LLM to get the embedding vector
 - A numerical representation of the semantics of the query
- Do a distance search
 - Locate the titles that a semantical near the query

The result

- Not that impressive as I have hoped for
 - Got results that I did not expect
 - The distance was far
- About the same result when even if the query was in a different language
- The current titles are not enough
 - Need more information
- Title and queries are semantically different

Population per month by region, age and sex. Year 2025M01 - 2026M08

It is hard to find the right
table