Webscraping & API Project

Mikael Magnusson

Overview of project

- Website to scrape: börssnack.di.se

- Scrapes thread title and number of comments

- Takes an input how many pages it should scrape
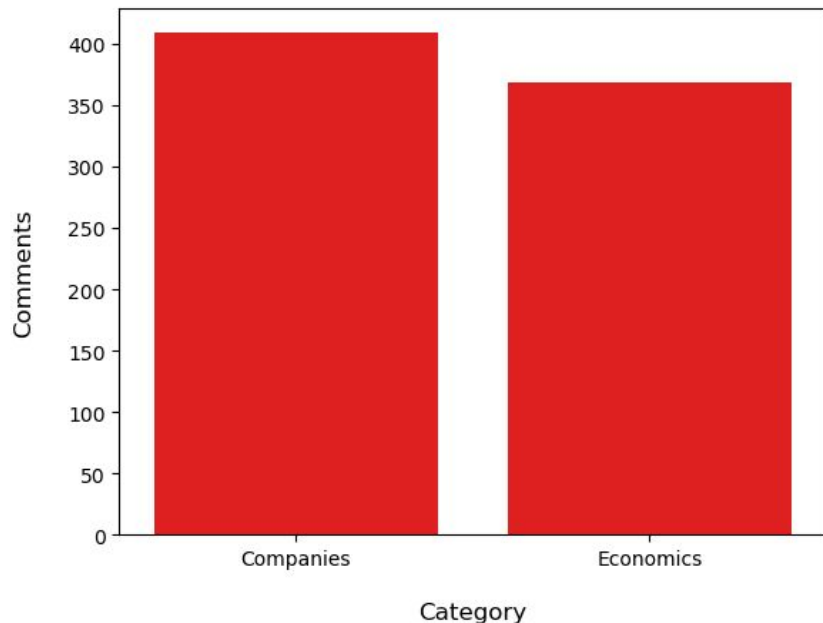
- Second scraper?

ChatGPT API

- Quickly summarizes current threads in the forum

  - Consolidates, categorizes and combines subjects

  - Counts each comment
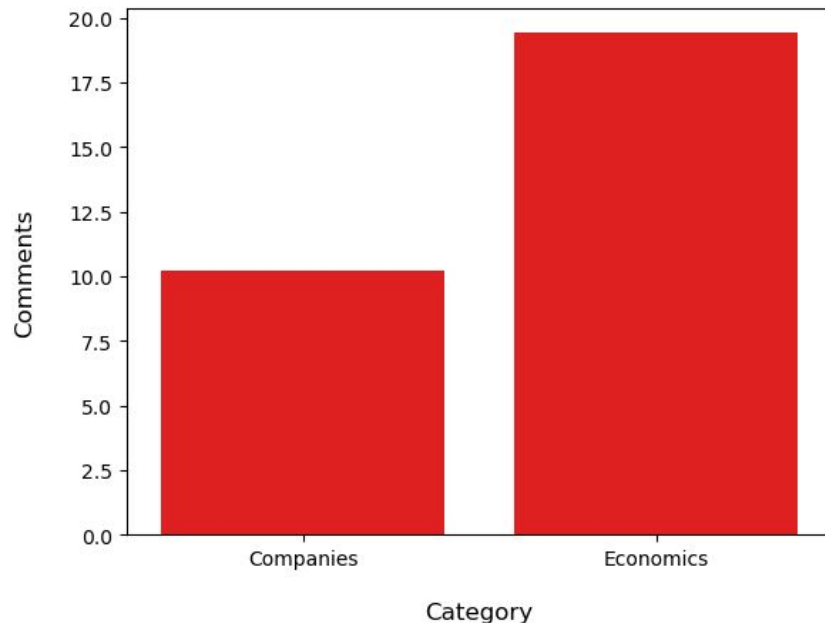
- Returns us a copy+paste data-frame

  ÄMNE | KOMMENTARER | TRÅDAR | M/B | TICKER | LISTA

- The gpt-4-turbo-preview model browses internet
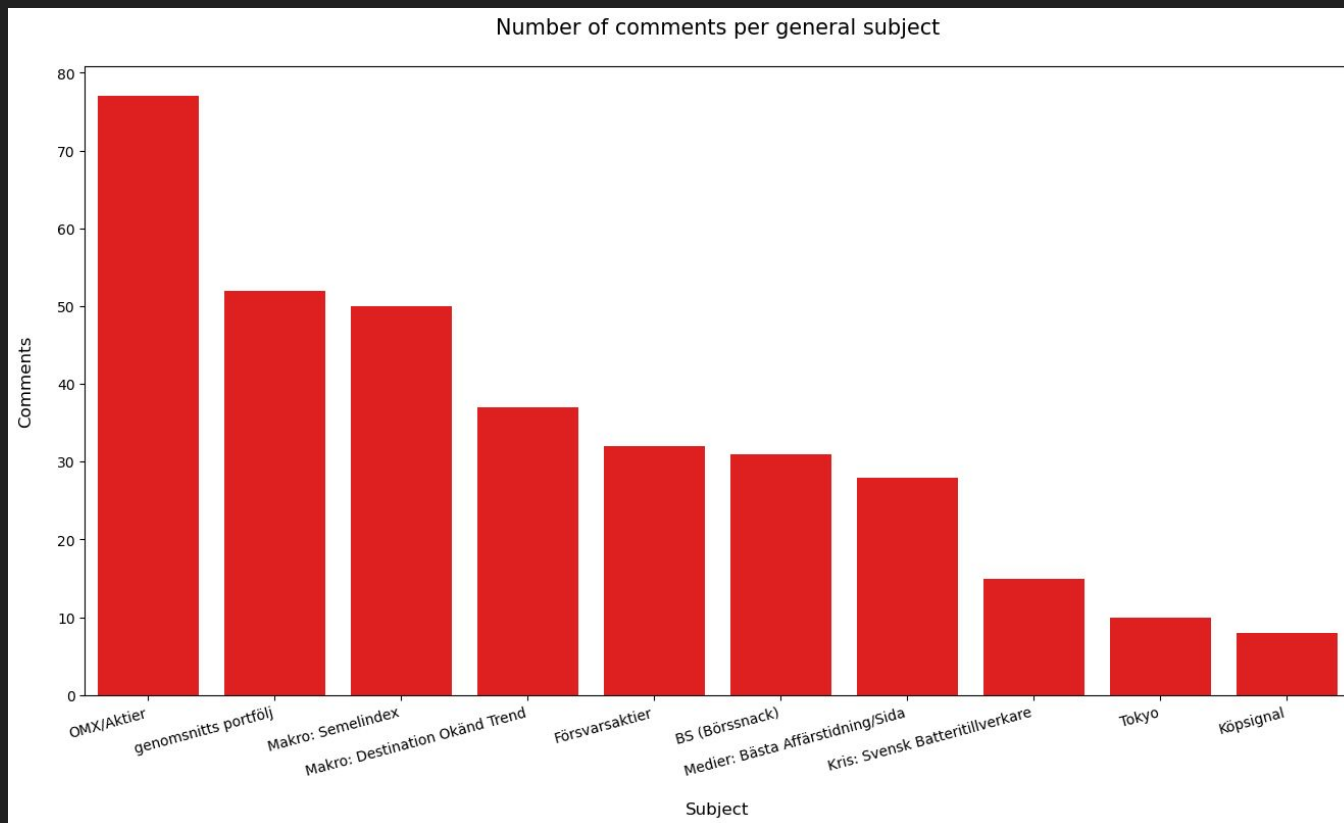
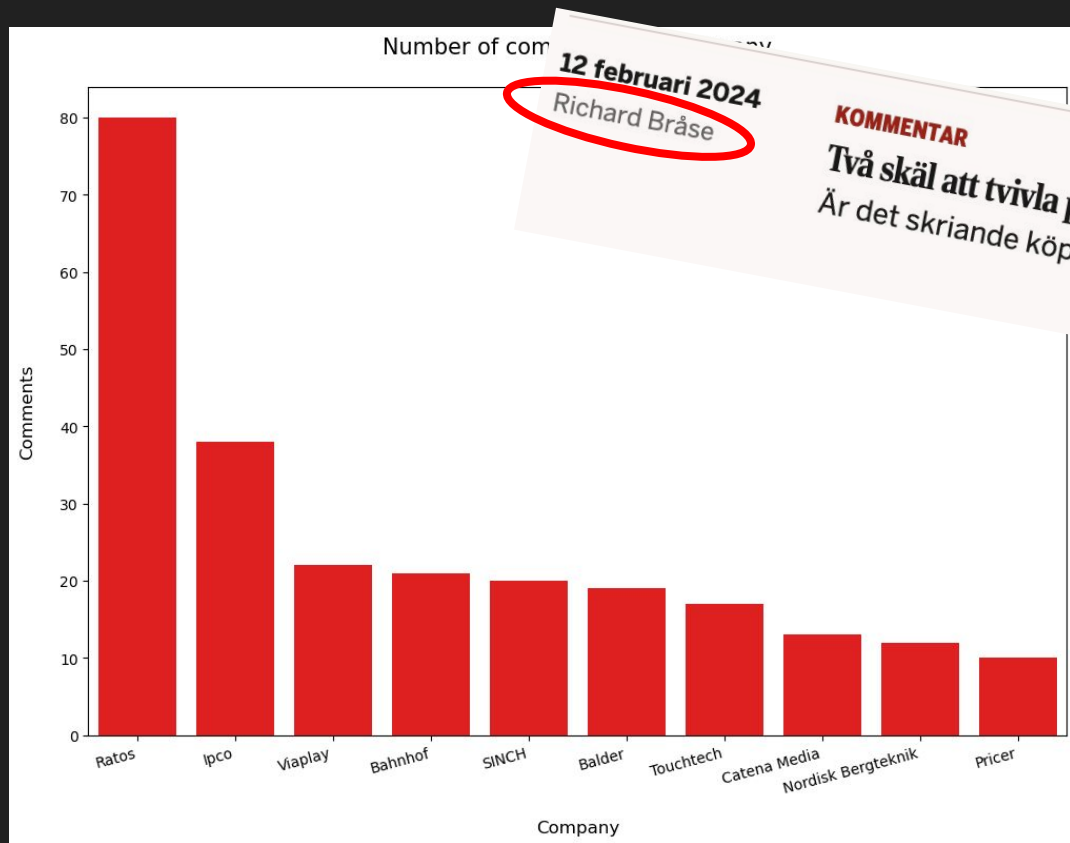# Data overview

Data overview

Number of comments per general subject

Data overview

Number of comments per company

12 februari 2024
Richard Bråse

KOMMENTAR

Två skäl att tvivla på Ratos potential
Är det skriande köpläge i Ratos? Det finns potential, men också skäl att...

(Bar chart: Comments vs Company)

- Ratos: 80
- Ipco: 38
- Viaplay: 22
- Bahnhof: 21
- SINCH: 20
- Balder: 19
- Touchtech: 17
- Catena Media: 13
- Nordisk Bergteknik: 12
- Pricer: 10

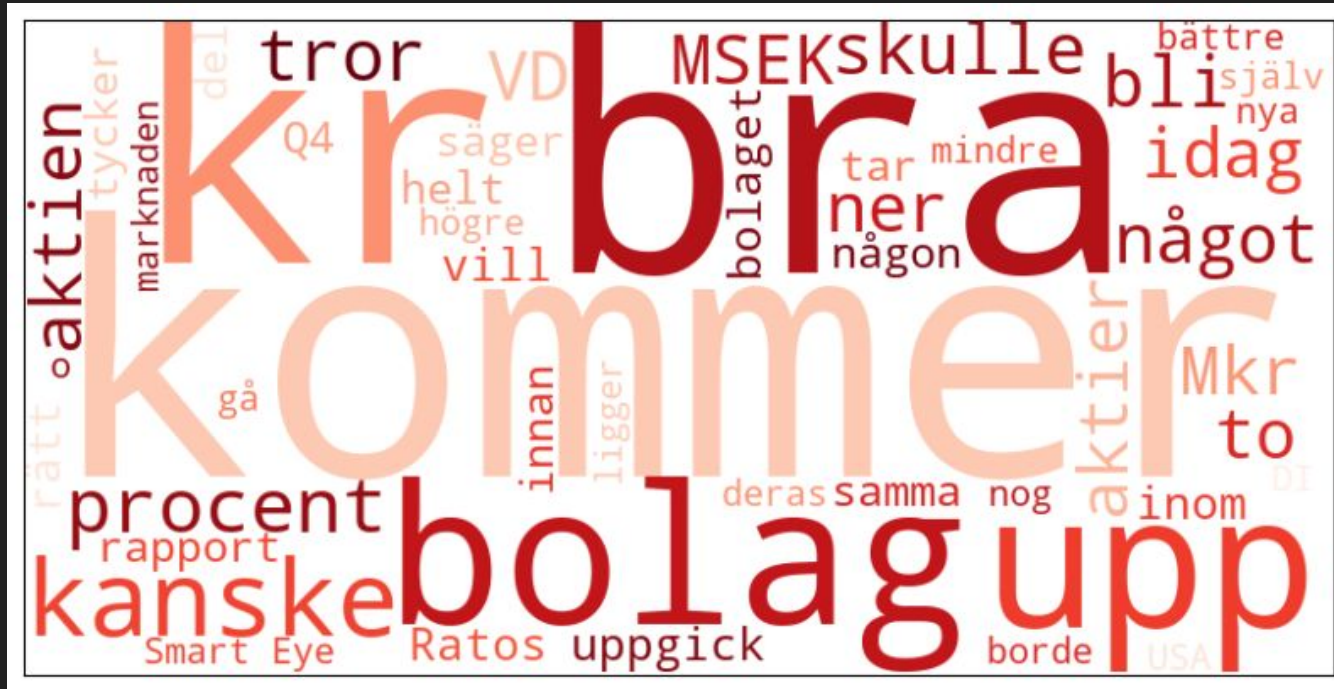Using InFront API with the tickers generated from ChatGPT-4

Second Scraper

- Enters threads and scrapes ALL text
        Thread to thread, page to page

- Allows us to do some more analysis

Second Scraper

Challenges & Highlights