



id: SI_001

model: openai/gpt-5.4-mini

seeds:

04_risiko: sårbarhet

03_relasjon: konkurranse

05_handling: implementer

01_perspektiv: etisk

02_tidhorisont: langsiktig

Instans SI_001

Seeds: etisk | langsiktig | konkurranse | sårbarhet | implementer

Å samle **komplett prisdata på alle varer i verden** er i praksis ikke et “scrape alt”-problem, men et **konkurranse- og infrastrukturproblem**. Mainstream-svaret er ofte “bruk nett-scraping og API-er”; mitt perspektiv er at det er **for sårbart, juridisk skjørt og strategisk kortsiktig**. Du trenger en modell som tåler at priser er dynamiske, lokale, lovregulerte og ofte skjult bak rabatter, pakningsstørrelser og medlemskap.

Hva “komplett” egentlig betyr

Pris er ikke bare én verdi. Du må definere:

- **Listepris**
- **Reell transaksjonspris**
- **Pris per enhet** (kg, liter, stk)
- **Geografisk pris**
- **Tidsstempel**
- **Tilgjengelighet/lager**
- **Betingelser** (lojalitetsprogram, kupong, abonnement)

Uten dette blir datasettet etisk og analytisk svakt: det sammenligner epler og pærer.

Etisk og sårbar tilnærming

Full global innsamling via aggressiv scraping skaper sårbarhet:

- juridisk risiko (GDPR, ToS, databaserett)
- blokkering fra plattformer
- skjevhet mot digitale markeder



- dårlig dekning for små butikker og fattige regioner

Et mer etisk og robust spor er **flernivå-innsamling**:

1. **Offentlige prisdatakilder** der de finnes
2. **Partneravtaler** med kjeder, grossister og markeds plasser
3. **Crowdsourcing/app-basert skanning** av kvitteringer og hyllepriser
4. **Syntetisk normalisering** av pakningsstørrelser og valuta

Konkurransperspektiv

Verdens komplette prisdata er i seg selv en strategisk moat. Den som bygger:

- høy datadekning
- lav forsinkelse
- god enhetsnormalisering
- kvalitetsscore per kilde

...får et kraftig konkurransefortrinn i retail analytics, innkjøp, inflasjonsovervåkning og eksport/import-intelligens.

Implementerbar arkitektur

Start smalt, ikke globalt:

- 5 land, 3 kategorier: dagligvarer, elektronikk, drivstoff
- normaliser til **pris per standardenhet**
- bygg en **master product graph** med GTIN/EAN, SKU, MPN
- bruk probabilistisk matching for å koble samme vare på tvers av kilder
- lag en "confidence score" per prisobservasjon

Langsiktig sannhet

Kompletthet er et mål, ikke et endepunkt. Den riktige strategien er å bygge et **levende prisnettverk** som kontinuerlig oppdateres, ikke en statisk database.

Neste steg:

1. Definer datamodellen: pris, enhet, geografi, tidspunkt, betingelser.
2. Velg 2-3 datakilder per land og lag en pilot for normalisering/matching.
3. Mål dekning og feilrate per kategori, ikke bare antall innsamlede priser.