

Implementation Guide — QCC Dutchie Scraper (Apify Actor)

Architecture

- PlaywrightCrawler with Apify proxies
- GraphQL intercept → normalized products
- DOM fallback → if GraphQL returns 0
- Screenshot+OCR (Tesseract+Sharp) → last resort
- AI parsing → normalized products[] and promotions[]
- Daily OCR dataset naming for audits
- Optional Google Chat run summary

Deploy

- `apify push` to Apify from local or link GitHub for auto builds
- Set env vars in Apify → Settings → Environment variables
- Create Schedules (02:00, 13:00 ET)

Extend

- Competitor URLs: add to `startUrls`
- Diff detection & alerts
- Sheet sync with Apps Script