

EDGAR Form D Screener

Problem: Manual screening of Form D filings is slow and error-prone.

Goal: Surface only CA LLCs in Residential claiming 506(b), one per company.

Sources: SEC EDGAR master index + primary_doc.xml (no API key; respectful rate limiting)

Approach: Index → resolve XML → verify 4 (or more) criteria → de-dupe by CIK → keep latest per company

Funnel: 2024-01-01 → 2025-01-01 (LLC, CA, Residential, Rule 506(b))

Forms: D & D/A Resolver: hybrid, Polite UA & throttling.

Stage	Count
Total filings (D + D/A)	55,737
"LLC" in company name	17,683
XML reachable	17,679
Unique CIKs in master	15,053
Unique companies kept (1 per CIK)	56

vs. unique issuers: kept 56 / 15,053 → 99.63% reduction

vs. D/D-A filings: kept 56 / 55,737 → ~99.90% reduction

Results:

- Kept 56 companies from 15,053 issuers (de-duped by CIK; latest only).
- Median fetch latency: 175.75 ms (avg 181.59 ms); 429/503 = 0, retries = 0 / 17,679.
- Final dataset: final_companies.csv

(https://github.com/eituranalp/SEC-EDGAR-Form-D-Screener/blob/main/docs/exampleRun/final_companies.csv)

Samples:

BLDG Pioneer LLC 2023-01-05 CIK 0001960560

XML: https://www.sec.gov/Archives/edgar/data/1960560/000168316823000028/primary_doc.xml

Stoakes Group LLC 2023-10-12 CIK 0001996708

XML: https://www.sec.gov/Archives/edgar/data/1996708/000199670823000001/primary_doc.xml

Repo:

<https://github.com/eituranalp/SEC-EDGAR-Form-D-Screener#>