Classification Pattern

* Individual listing (label = 1)
  + Check title field and description meta:
    - Example: https://boltonlandingfarmersmarket.com/
    - <title><business\_name>/<area/location> Farmer's/Farmers Market …</title>
    - <meta name="description" content="…">
    - https://www.manchesterct.gov/Government/Departments/Leisure-Family-and-Recreation-Department/Neighborhoods-and-Families-Division/Manchesters-Farmers-Market-Series
  + Most of them are update/event or have some links which leads to more info on the FM
    - <https://www.warwickadvertiser.com/news/local-news/locals-enjoy-the-lakeside-farmers-market-ac2602552>
  + Info for 1 specific FM but also have other FM info links?
    - <https://www.commongroundfarm.org/markets/newburgh-farmers-market>
  + Has Farmer MarketS in title but only list 1 FM info:
    - <https://411.info/business/Bellmore-Fresh-Farmers-Market-North-Bellmore-NY-6191497>
  + Farmer Market’s website
  + Farmer Market’s social media page
  + Farmer Market’s info/update with links to the Farmer Market
    - <a href="tel:5167850729"><span>(516) 785-0729</span></a>
  + Farmer Market’s info/update with text only
  + App map?
    - <https://moovitapp.com/index/en/public_transit-Eastgate_Plaza_Farmers_Market-Buffalo_NY-site_170784868-1402>
    - <https://vymaps.com/US/Newburgh-Farmers-Market-T9439019/>
* Multiple Listings (label = listings)
  + https://delaware.wicresources.org/farmers-market/farmers-market-locations/
  + An article listing multiple Farmer Markets with some of their info (name, time, location, …)
    - Example: <https://cremedelacreme.com/blog/farmers-markets-gilbert-az/>
    - Example: <https://www.newsbreak.com/minneapolis-mn/3015915576982-here-s-where-to-find-farmers-markets-in-minneapolis-in-2023>
  + An article with a link to FMs or FM database:
    - <https://www.adirondackalmanack.com/2023/06/new-farmers-markets-in-port-henry-and-schroon-lake-begin-this-weekend.html>
  + A database with Farmer Markets’ info
    - Example: <https://www.mda.state.mn.us/authorized-farmers-markets>
    - <https://www.nyc.gov/assets/doh/downloads/pdf/cdp/health-bucks-poster-mn.pdf>
  + KWs: Farmers Market Series, #x\_numbers farmer markets, Farmers Market Guide, … Area, gov site
* Irrelevant (label = other)
  + not Farmer Market listing:
    - Article with general information or updates about Farmer Markets without the Farmer Markets’ specific details or links to them
      * Example: <https://wtop.com/business-finance/2020/08/national-farmers-market-week-begins-today-dc-ranks-among-the-best-in-the-country/>, <https://sf.gov/step-by-step/apply-sell-goods-alemany-farmers-market>
      * <https://agriculture.ny.gov/consumer-benefits-farmers-markets>
    - Administrative procedures/regulations for Farmer Markets (sign up, regulations, laws), Nutrition Program
      * <https://sf.gov/step-by-step/apply-sell-goods-alemany-farmers-market>
    - Policy update:
      * <https://www.mytwintiers.com/news-cat/top-stories/new-program-coming-to-corning-farmers-market-to-help-snap-holders/>
    - KWs: Apply, Consumer Benefits. funding, grant
* Bad

Status Scan:



Initial scan for bad links (still need validating)

Bad link errors:

* 404 Client Error: -> true bad link
* 403 Client Error: Forbidden for url: …
  + Sever error -> true bad link
  + request get blocked (can use selenium to validate)
* HTTPSConnectionPool(host=’’, port=443): Max retries exceeded with url
  + (Caused by SSLError(SSLCertVerificationError(1, '[SSL: CERTIFICATE\_VERIFY\_FAILED] certificate verify failed: unable to get local issuer)/ SSLError(SSLError(1, '[SSL: UNSAFE\_LEGACY\_RENEGOTIATION\_DISABLED] unsafe legacy renegotiation
  + Caused by ConnectTimeoutError()
  + Read timed out. (read timeout= ) -> Set a higher timeout value
  + Caused by NameResolutionError("<urllib3.connection.HTTPSConnection object at >: Failed to resolve '’)