**Pipeline**

1. Daily Process

* Raw listings – daily
* Clean listings
* Dealer daily inv
  + aggregate individual listings to the day-level (description of daily inventory: price avg/sd, HHI, brands lisy, # new/used/offers)
  + function: gen\_dealer\_day\_inv(pattern = './listings/clean\_listings\_202?-??-??.zip')
  + output: f'./dealers/dealer\_daily\_inv\_{date}.csv'

1. Post Download Cleaning

* Run *caffeinate python Listings\_Cleaner.py sales*
  + filter\_VINs(): Clean listings => VINs only
  + find\_sold\_vins(): Sold.json (infer sales by comparing daily listings)
  + gen\_all\_sales():
    - inventory\_to\_sales(): Clean listings + Sold.json => ./sales/Sales\_{date}.csv
    - aggregated: './sales/allsales\_{min\_date}\_to\_{max\_date}.zip'
* Run *caffeinate python Listings\_Cleaner.py all\_dealer\_days*
  + Merge all dealer-day-level data & add all\_sales
* Basic Dealerbook – just keep dealer identifying information & address

1. Dealerbook Manager
   1. Basic\_dealer\_book => geocoded/demographics/dealerchains