The goal of the BDC-specific programs that import and process the “raw” extracted SOI data are to create the following set of consistently named and defined variables:

* Variables that must always be present or created:
  + asof\_date = the as of date for the investments
  + order\_within\_filing = an index value that keeps track of the order of the observations in the SEC filing.
    - This index needs to be manually created
    - This index does not need to be sequential but must achieve goal of reflecting SEC filing ordering if sorted by
  + issuer = the portfolio company name as reported in the given filing but cleaned up in the following ways
    - All letters capitalized
    - Extra spaces removed
    - Strange ASCII characters removed or replaced
    - Footnotes removed
  + cost = amortized cost in thousands of USD
  + fair\_value = fair value of debt in thousands of USD
  + instrument\_type\_raw\_upper = description of investment that must include key words such as “debt” and “equity” and also should include all available info to further characterize the type (e.g., “senior”). We will mine this field for keywords so the key thing is that the keywords are in there and also we should fix typos if noticed. This field may be a cleaned up version of investment\_descip\_raw (e.g., company names removed from the full description).
  + investment\_descrip\_raw = full description of investment as provided by the BDC. This field can be left “messy” meaning that it may include lots of extraneous information. The investment-level info will be extracted to individual variables (see below) and the key instrument info will be captured in “instrument\_type\_raw\_upper”. Therefore this field will serve primarily as a reference if needed.
  + industry = description of industry provided by the BDC but, if needed, the string is cleaned up to remove investment info such as numbers, percentage symbols, etc.
* Debt-specific investment-level variables for which at least one must be present
  + maturity\_date = reported maturity date. Almost always present.
  + basetype = base rate for variable-rate debt (LIBOR, PRIME, EURIBOR, OTHER)
  + Interest rates and spreads that are annualized (only if needed) and denoted in decimal (e.g., 10%=1)
    - allinrate = total interest rate reported. May represent sum of cash and PIK.
    - cash\_rate = interest rate that is to be paid in cash, not PIK
    - pik\_rate = interest rate that is to be paid as PIK (must be explicit)
    - spread = spread in percentage points (e.g., use value of 0.025 if listed as 250 basis points) above given base rate in basetype
    - spread\_pik = spread above given base rate to be paid as PIK (must be explicit).
* Investment-level variables that might not be present but should be created if available
  + principal = face value of debt in thousands of USD
  + principal\_curr = currency of principal because sometimes principal is in foreign currency so we can capture this in this field. Only put currency ticker (e.g., EUR or GBP) if not USD. If USD, leave missing as we will interpret missing values as USD.
  + floor\_rate= explicitly provided floor for the base rate (e.g., 0.01 if there is an explicit 1% libor floor).
  + acquisition\_date = date that BDC acquired or initiated investment
  + businessdescription = a more detailed description of the company’s business beyond an industry subheader. Example: see ARCC.

Additionally, Sam will create the following variables based on the above:

* Dummies to characterize an issuer that is not a true portfolio company
  + ind\_issuer\_clo = 1 if issuer represents a CLO security
  + ind\_issuer\_cdo = 1 if issuer represents a CDO security
  + ind\_issuer\_fund = 1 if issuer represents some sort of debt fund
  + ind\_issuer\_othernotcompany = 1 if issuer represents some other type of entity that should not be considered a true portfolio company. For example, a REIT or money market fund. Arguably, some of these could be “funds” but this category exists to make sure they are flagged regardless of specific reason for not being a portfolio company
* is\_debt = 1 if the investment is a debt investment (loan or security), 0 otherwise
* baserate = the base rate from which to add spread to compute total interest rate
  + this info is often provided in footnotes and needs to be manually collected
* allinrate\_final = the “total” interest rate as gleaned from interest rate variables
  + = allinrate if available
  + = baserate plus spread plus spread\_pik if allinrate is not available but baserate and spread are. Note this only applies to variable-rate debt
  + = cash\_rate plus pik\_rate if allinrate is not available but separate cash\_rates and pik\_rates are provided and the description implies they should be summed
    - As an example where this doesn’t work, it may be that the given PIK rate is indicated to be part of the overall rate, not in addition to. In this case, however, it seems like there is an allinrate provided so this step shouldn’t be needed
* Dummies to characterize where investment is being used as collateral
  + part\_of\_bdc\_clo = dummy to indicate given investment is being used as collateral in the BDC’s CLO. This information determined based on footnotes likely indicated in the raw portfolio company name.
  + If many secured facilities, there may be many dummies