**Data Cleansing Logic and UI/UX**

* + Data Cleansing add rules to a list of cleansing options to include (recipe ingredients). Each rule will have an input columns, cleaning options and output column or replace existing values. Each rule may be used “n” times if the user wants to invoke the same option multiple times.
    - Removal
      * Punctuation, spaces, numbers, letters, leading spaces, trailing spaces, non-printable chars, free field for user input
    - Replace
      * string with another string
        + Case sensitive yes/no, default yes
    - Copy Column
    - Case options
      * Upper, Lower, Capitalize
    - Fill Blanks with static test
    - Statement Generator this will allow the user to create 1 or more new columns based with data based on the following statement options .
      * Conditional If <COLUMN1> <OPERATOR (mathematical and string> COLUMN 2 (AND/OR/NOT…..) then output (Output can be static or column(s) value> into a new column
      * SUM(COLUMN1,COLUMN2,COLUMN(n))
        + Average, Difference, Divided, Multiply
      * into a new column
      * Range(FromDateONumber,ToDateOrNumber,Output If In Range, Output if our of Range)
        + Create a flag column

Define value for valid (static or original)

Define value for invalid (static or original)

* + - * Merge(delimiter,1st not empty (boolean default=0),COLUMN1,COLUMN1,COLUMN(n)
        + In a new column
      * Split(COLUMN1, delimiter)
        + Into multiple columns
      * Left/Right(COLUMN1, n digits)
        + Into a new column
      * Search(COLUMN1,char)
        + new column with True/False
      * Dynamic string (“Text”, (COLUMN1,(COLUMN2, (COLUMNn)
        + Create new columns with True/False
    - Date Conversion
      * Allow user to define a desired output format for dates and then format all dates to a column in this format
        + Allow overwrite or create a new column
    - Patterns- apply a regular expression and allow user to extract detected patterns into 1 or more columns
      * We will supply a global or private library of regular expressions for the user to apply
      * user can copy and paste his own RegEx
      * If possible create RegEx Gui designer, screenshot below
    - Dictionaries - allow users to upload dictionaries in order to normalize/standardize values
      * Allow to replace existing value, extract to a new column (define column header), delete original value or create a flag column based on a value
      * tokenize the selected column and allow user to define delimiters for tokenizing
      * Allow users to upload created dictionaries for use
    - Validation
      * Conditions
        + Dates - allow user to enter a date range
        + email - check box to check if email is valid
        + phone number - check box if email is valid
        + List - to check if a value/string contains the user input

1 per line

* + - * Output Options
        + Delete values

Valid

Invalid

* + - * + Use original

Valid

Invalid

* + - * + Change to static value (user input)

Valid

Invalid

* + - * Output result for Invalid Data
        + Options

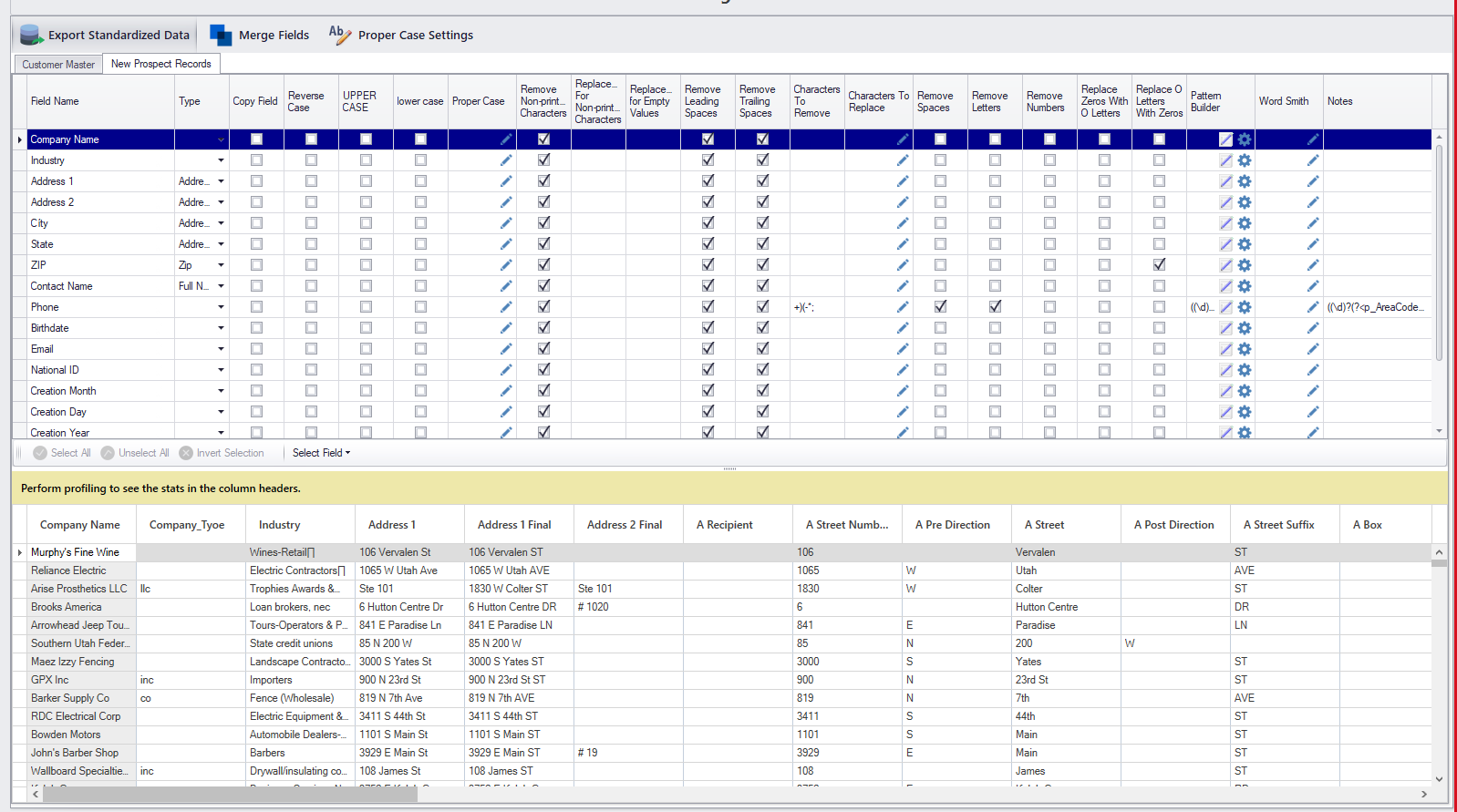
Delete values

Using original

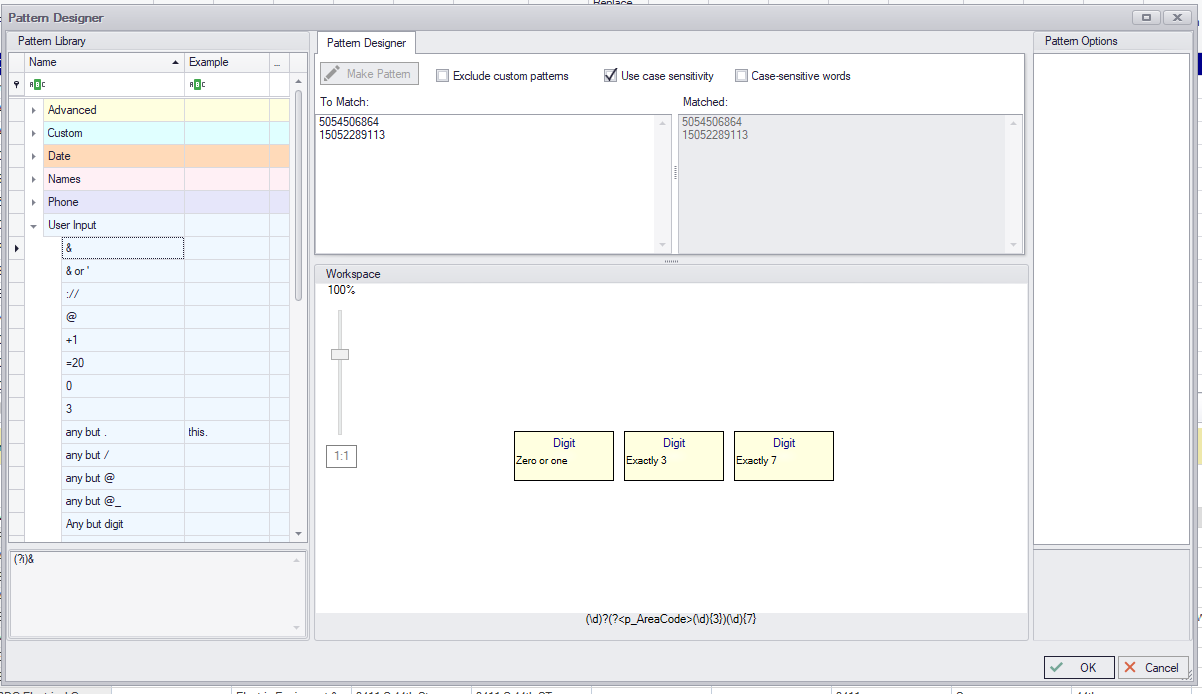
Change to static value (user input)

* + - * + Create a flag column for invalid values with custom flag value (user input, default=1)
    - Conversions
      * Allow user to convert measurement and units from 1 concept to another
        + Ex: 1 meter → 39.3701 inches
      * Allow user to add multiple columns with measurement info
      * Create 1 column with destination measurement value and 1 column with destination measurement name (2 columns in total)

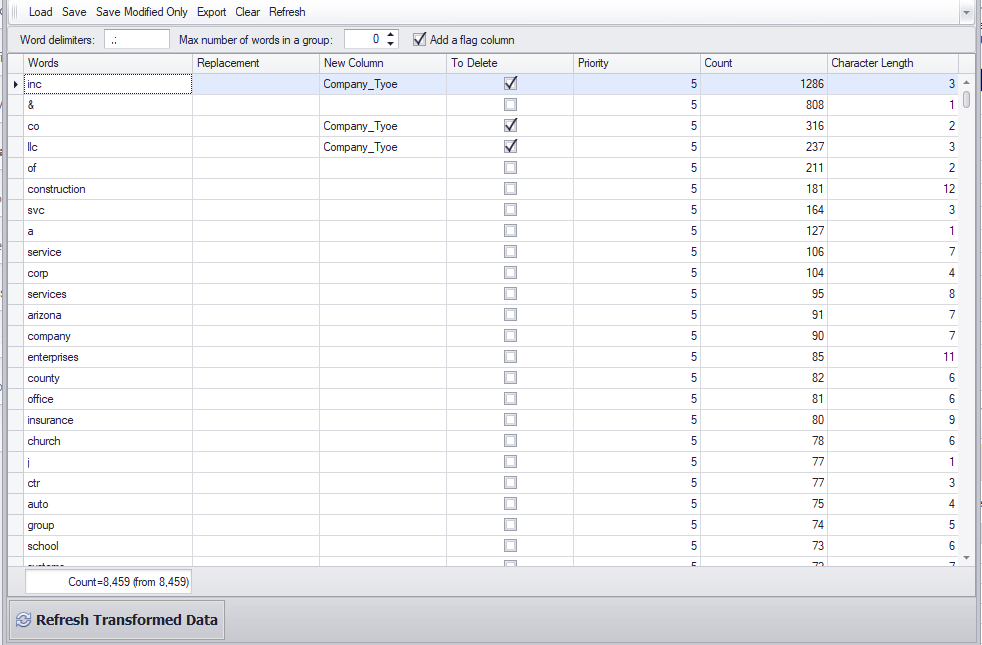
Sample Cleansing



**Sample Regex Designer**

****

**Sample Dictionaries**

****