Alexis User Guide

Team Data Wranglers

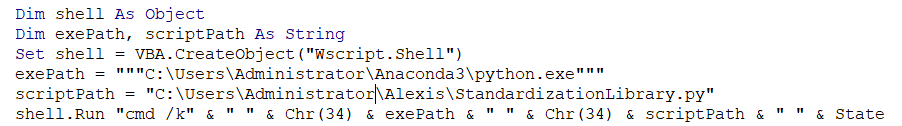
GMU DAEN 690 Section 4

Professor Robert E Kraig

**How to Use Alexis**

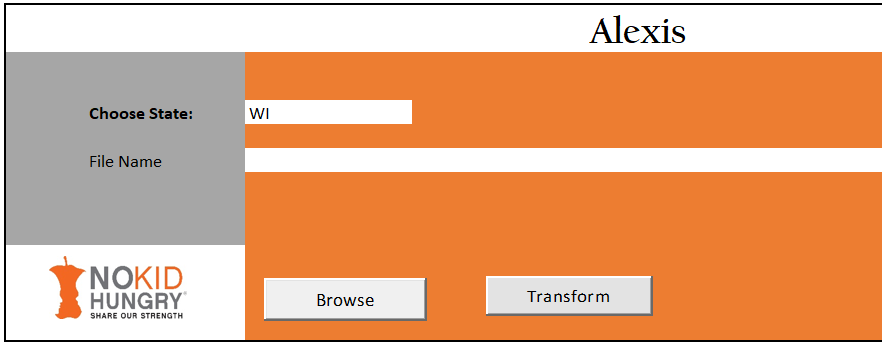
Alexis requires the below components in order to run:

1. Alexis UI which is macro enabled to trigger the python code (macro can be found in the transform button). Note the macro needs the path to the python executable and the standardization library in order to work.



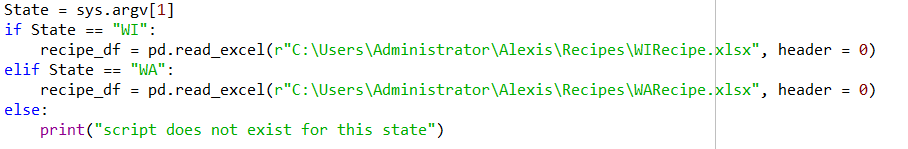
1. Recipes for each state
2. Standardization Library code

The user should enter the state shorthand in Choose State field in order to trigger the correct recipe for that state. Then click the transform button to start the code.



Note: The UI does have an additional functionality which brings in raw data files through the browse button as sheets in the UI. This has not yet been integrated to the code and requires design to work for Alexis.

The standardization library code contains all function definitions. Then the main section lists the recipe locations as well as where the clean csv should be saved to.





The remaining section of the main is reading each line of the recipe and applying the transformations.

**How to Use Actions in a Recipe**

An ACTION in Alexis is the name of a supported transformation. Following is the list of supporting actions and how to use them in an Alexis recipe.

GENERATEKEY – this function will generate a key column which will be used in a full outer join. All raw data files for which a full outer join is required should be listed one after another in the beginning of the recipe (under Raw\_File\_Name\_1). The sheet name should be mentioned under Worksheet. This function will read all of them and generate a key column for each.

Usage:

Raw\_File\_Name\_1: Full path to the location of the raw file.

Worksheet: Sheet name of the data required for transformation.

ACTION: GENERATEKEY

CalculationLogic1: This should be a placeholder name for the key column. Ex: key

CalculationLogic2 +: The remaining parameters are the names of the columns that will make up the key.

GENERATESCHOOLKEY – This function will generate the keys required for left join. This is usually the case for the nces school data.

Usage:

ACTION: GENERATESCHOOLKEY

CalculationLogic1: The placeholder value for key column should be Key.

CalculationLogic2 +: The remaining parameters are the names of the columns that will make up the key.

FullOuterJoin – performs full outer join of all raw data files mentioned with generatekey action.

Usage:

ACTION: FullOuterJoin

CalculationLogic1: name of key column used for generatekey action.

LeftOuterJoin- performs left join of raw data files mentioned

This function requires the path to the raw data file being introduced as well as the sheet name.

Usage:

Raw\_File\_Name\_1: Full path to the location of the raw file.

Worksheet: Sheet name of the data required for transformation.

ACTION: LeftOuterJoin

CalculationLogic1: Key

CalculationLogic2+: The names of the columns used in the new raw file that match Key (used to perform the merge).

SchoolTypeOriginalFormula: Formula used to generate the Wisconsin school type original column.

Usage:

ACTION: SchoolTypeOriginalFormula

CalculationLogic1: School Type-Original (clean name)

CalculationLogic2: Public (Y/N)

CalculationLogic3: School Level-Original

RENAME: Rename a column

Usage:

ACTION: RENAME

CalculationLogic1: Clean field name

CalculationLogic2: Raw field name

PREFIX\_ZEROES: Pad zeroes in front of string to make the equal length.

Usage:

ACTION: RENAME

CalculationLogic1: Desired string length

CalculationLogic2: Column name to apply padding zeroes on.

DefaultValue: Create a column where all records have the same value.

Usage:

ACTION: DefaultValue

CalculationLogic1: Clean field name

CalculationLogic2: Constant Value each record should have

Concatenate: Concatenate given fields separated by “@” symbol. This function can be extended to be any symbol if the @ (or any other symbol) becomes the values for CalculationLogic5 and function is updated accordingly. Order in which the raw fields are entered is the order they will appear in the final concatenated column.

Usage:

ACTION: Concatenate

CalculationLogic1: Raw field to be concatenated

CalculationLogic2: Raw field to be concatenated

CalculationLogic3: Raw field to be concatenated

SETNULL: Create an empty column

Usage:

ACTION: SETNULL

CalculationLogic1: Clean column name

REMOVECOLUMN: Drop column from dataframe

Usage:

ACTION: REMOVECOLUMN

CalculationLogic1: Column name to be removed.

MergeNullColumn- For same name columns that come from different raw data files. The columns are merged by checking if the right has any NULL records. If there are, it will check of the left has the value for that record. If it does, it will pull the value in, otherwise it will remain NULL. This function can only be used 2 columns at a time.

Usage:

ACTION: MergeNullColumn

CalculationLogic1: Clean column name

CalculationLogic2: Column to be merged

CalculationLogic3: Column to be merged

DATEMMDDYYYY- switches date format to mm/dd/yyyy.

Usage:

ACTION: DATEMMDDYYYY

CalculationLogic1: Clean column name

CalculationLogic2: Raw date column name

CLAIMYEAR- get the claim year from the claim date

Usage:

ACTION: CLAIMYEAR

CalculationLogic1: Clean column name

CalculationLogic2: Raw date column name

CLAIMMONTH- get the claim month from the claim date

Usage:

ACTION: CLAIMMONTH

CalculationLogic1: Clean column name

CalculationLogic2: Raw date column name

CONCATMODELS- This is used for Breakfast Delivery Model from State Agency Tracking-Original. This function will concatenate all of the model column names. This function currently supports 7 models.

Usage:

ACTION: CONCATMODELS

CalculationLogic1: Clean column name

CalculationLogic2: Column name of model (Ex: TRADITIONAL\_MODEL)

CalculationLogic3: Column name of model

CalculationLogic4: Column name of model

CalculationLogic5: Column name of model

CalculationLogic6: Column name of model

CalculationLogic7: Column name of model

CONCATMODELSSTRING- In the case that the records for the model columns are not “Y” and instead filled with the name of the model. This function will skip replacing “Y” records with the name of the model. This function is a shorter version of CONCATMODELS.

Usage:

ACTION: CONCATMODELSSTRING

CalculationLogic1: Clean column name

CalculationLogic2: Column name of model (Ex: TRADITIONAL\_MODEL)

CalculationLogic3: Column name of model

CalculationLogic4: Column name of model

CalculationLogic5: Column name of model

CalculationLogic6: Column name of model

CalculationLogic7: Column name of model

CONCATSCHOOLS- This is used for school level original column for Washington.

Usage:

ACTION: CONCATMODELS

CalculationLogic1: Clean column name

CalculationLogic2+ - CalculationLogic2-14 will be the grades Pre-K-12.

BreakfastOpDaysFormula- This is the formula for the breakfast operating days.

Usage:

ACTION: BreakfastOpDaysFormula

CalculationLogic1: Clean column name

CalculationLogic2: Operating Days

LunchOpDaysFormula- This is the formula for lunch operating days.

Usage:

ACTION: LunchOpDaysFormula

CalculationLogic1: Clean column name

CalculationLogic2: Operating Days

FREnrollmentFormula- This is the FR Enrollment formula

Usage:

ACTION: FREnrollmentFormula

CalculationLogic1: FR Enrollment

CalculationLogic2: Enrollment-Free

CaculationLogic3: Enrollment-Reduced

CalculationLogic4: Enrollment-Free and Reduced

FREnrollmentPercentageFormula- This is the FR Enrollment Percentage formula

Usage:

ACTION: FREnrollmentPercentageFormula

CalculationLogic1: FR Enrollment Percentage

CalculationLogic2: FR Enrollment

CaculationLogic3: Enrollment-Total

CalculationLogic4: Lunch Meals-Free

CalculationLogic5: Lunch Meals-Paid

FRBreakfastMealsFormula- This is the FR Breakfast Meals formula

Usage:

ACTION: FRBreakfastMealsFormula

CalculationLogic1: FR Breakfast Meals

CalculationLogic2: Breakfast Meals-Free

CaculationLogic3: Breakfast Meals-Reduced

FRBreakfastADPFormula- This is the FR Breakfast ADP formula

Usage:

ACTION: FRBreakfastADPFormula

CalculationLogic1: FR Breakfast ADP

CalculationLogic2: FR Breakfast Meals

CaculationLogic3: Operating Days-Breakfast Only

CaculationLogic4: Operating Days

FRLunchMealsFormula- This is the FR Lunch Meals formula

Usage:

ACTION: FRLunchMealsFormula

CalculationLogic1: FR Lunch Meals

CalculationLogic2: Lunch Meals-Free

CalculationLogic3: Lunch Meals-Reduced

FRLunchADPFormula- This is the FR Lunch ADP formula

Usage:

ACTION: FRLunchADPFormula

CalculationLogic1: FR Lunch ADP

CalculationLogic2: FR Lunch Meals

CaculationLogic3: Operating Days-Lunch Only

CaculationLogic4: Operating Days

CEPYN- This is the formula from Washington to generate the CEP(Y/N) column

Usage:

ACTION: CEPYN

CalculationLogic1: CEP (Y/N)

CalculationLogic2: Site Claiming Option

SCHOOLLevelSTD- Generates the school level standardized column. The function contains the data dictionary for the value mapping.

Usage:

ACTION: SCHOOLLevelSTD

CalculationLogic1: School Level-Standardized

CalculationLogic2: School Level-Original

SCHOOLTypeSTD- Generates the school type standardized column. The function contains the data dictionary for the value mapping.

Usage:

ACTION: SCHOOLTypeSTD

CalculationLogic1: School Type-Standardized

CalculationLogic2: School Type-Original

BreakDelModelSTD- This will generate the breakfast delivery model standardized column.

Usage:

ACTION: BreakDelModelSTD

CalculationLogic1: Breakfast Delivery Model from State Agency Tracking-Standardized

CalculationLogic2: Breakfast Delivery Model from State Agency Tracking-Original