Clean Data Validation Steps

1.       64 fields must be there for all clean files.

2.       Make sure to follow data dictionary notes.

3.       Joining lunch and breakfast data? Make sure to join on claim date where needed with School id and District id.

4.       No Null values in School id and District id, Unique id if any investigate and escalate.

5.       Check Duplication steps in multiple location of the recipe, group by School id and District ID.

6.       Make sure to convert to numbers before formula or aggregation if not prep will concatenate.

7.       Check Data types at joins and validate missing values or non-matching result set.

8.       Pay attention to foreign key lengths and consider pre-padding data with zeros where required.

9.       Values in breakfast/enrollment/lunch paid, free, reduced should not be negative, further investigation needed if so.

10.   ID Fields should be String data types.

11.   Change School Year According to the file -- standardized format SYXX-YY.

12.   Clean file output names should match with data dictionary name with Clean- prepended to it.

13.   Check of School level standardization formats.

14.   Check on school type standardization formats.

15.   Breakfast model standardization formats.

16.   Operating days check -- never more than 20 or negative

17.   Check for no nulls in State-Reporting

18.   Look for outliers in provided number value fields

19.   School year September to May of next year - claim date.

Pain Points or lessons learned:

1.       Data being inconsistent for all states, which is expected but a little steam line in the process can help speed up the cleanup step.

2.       Some rookie mistakes (example misunderstanding some notes in Data dictionary, sometimes data its self, joins) I made and documented which can be avoided in future for someone new to pick (Documenting will help).

3.       Centralized documenting with version control.

4.       Version control on Tableau prep recipes (a systematic version control is needed as we are making multiple changes and overwriting excising one).

5.       Have to learn more on validation steps from Pam and Adam (subject matter experts) to avoid rework , in a way that we do maximum validation steps at clean data step or before releasing the data for validation from there Pam/Adam can have  a second validation to make sure .(this is avoid a full process load run by George )

6.       Wait time for validation: currently after clean data step, consolidation and enriching steps are run manually after hand full of data is cleaned and ready to be consolidated.

7.       One drive sync issues couple of times.

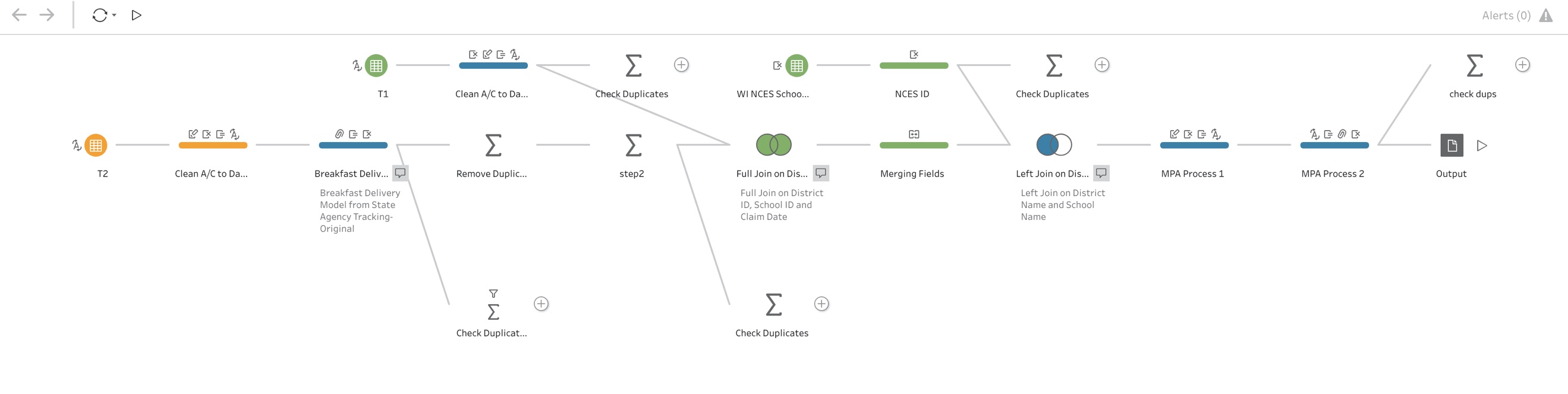
8.       Can learn what MPA team (subject matter experts) does once data is received (may be a scope for process improvement) .

9.       Tracking changes or feedback in Asana (a project management tool) and prioritizing.

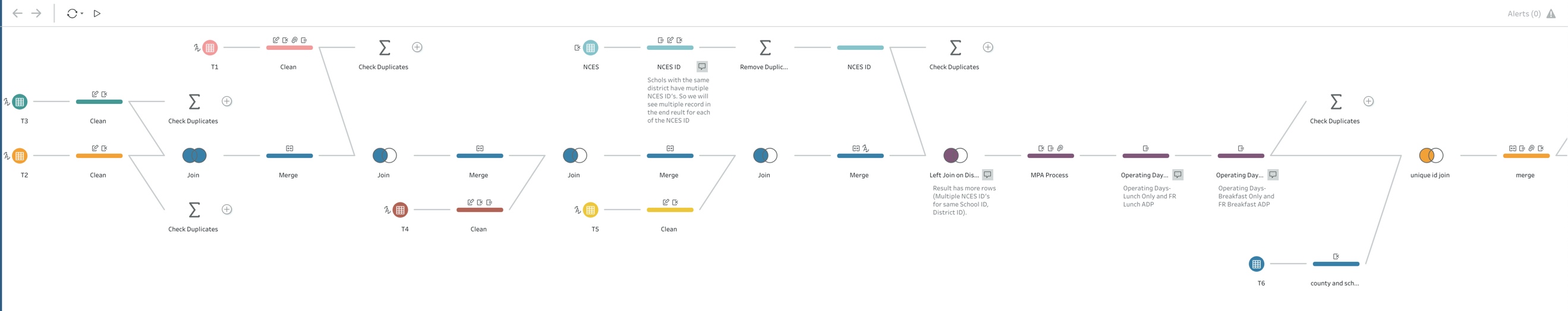
10.   Multiple validation failure points for same state which resulted in working on same state multiple times rerunning complete consolidation steps which is time consuming.

11.   Backfilling data for county and other required one's-- can this be a standard steps if data is missing for all states.

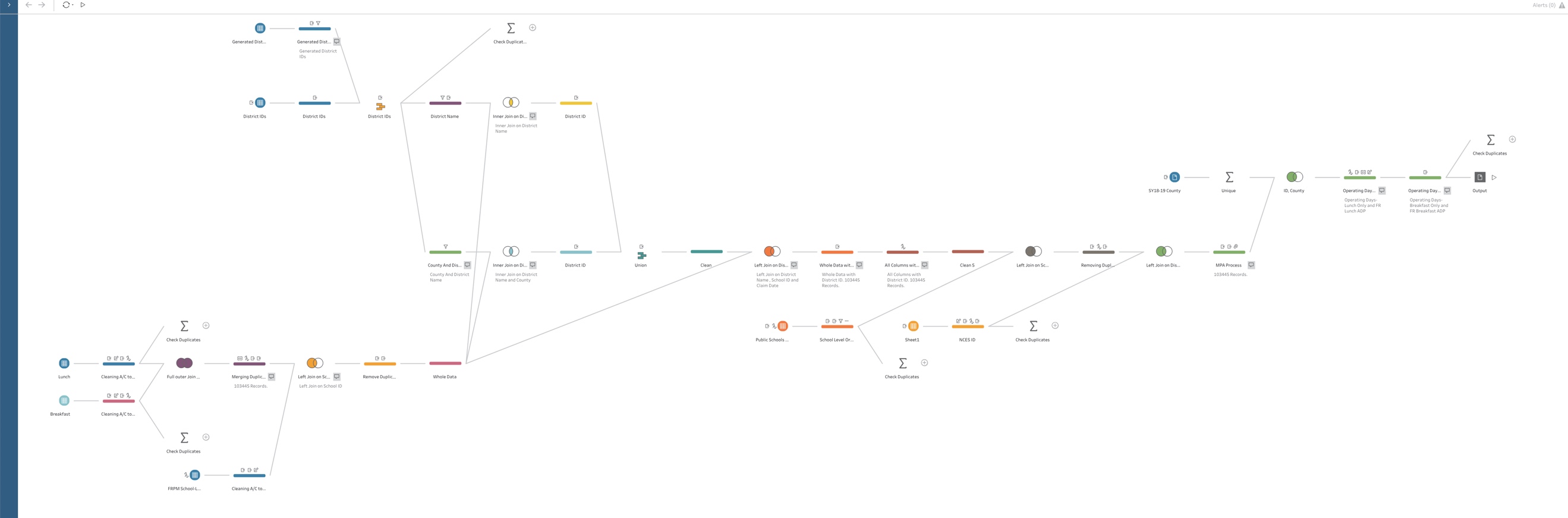
Easy State-School Year WI 18-19



Medium State-School Year WI 18-19



Hard State-School Year WI 18-19



Item fixed

Nebraska - SY17-18 , 18-19

·       Fixed NCES ID field

·       Change to number before adding fields

·       Operating days checked

·       New file created

INDIANA

·       Added missing fields

Breakfast Delivery Model from State Agency Tracking-Original

*And provision 2*

·       Changed Lower case Zip code to Zip Code

·       Changed Lower case Address line 1 to Line 1

·       Operating lunch and breakfast only name corrected

·       Call out operating days greater than 23 (11 Rows greater than 23 and less than 40)

·       New file created

MAINE

     -  Changed Lower case Address line 1 to Line 1

·       Call out operating days greater than 23 (9 Rows greater than 23 and less than 40)

·       New file created

TENNESSEE

·       Changed Lower case Zip code to Zip Code

·       Changed Lower case Address line 1 to Line 1

·       School year to 18-19

·       State reporting to TN for nulls

·       Unique id null fixed with above fix

·       Call out operating days greater than 23 (1 Rows greater than 23 days it has 31 days as operating)

·       New file created

KANSAS - 17-18 / 18-19

·       Changed Lower case Zip code to Zip Code

·       Replace Total enrolment 152,526 with 523 average of other months.

·       Replace Total enrolment 7603 with 833 average other two months.

·       School year 17-18 fixed

·       Call out operating days greater than 23 (253 Rows greater than 23 and less than 32)

·       New file created

FLORIDA

·       Change school year to SY 18 - 19

·       Claim date from 2017 - 2018  to 2018-2019

·       New file created

Louisiana

·       Breakfast Delivery Model from State Agency Tracking-Standardized fixed

·       Operating days above 23 and 31 days example greater than 100 . Escalated to Pam file not created yet . Waiting on pam's confirmation. (Confirmed that it is ok to have greater than 23 days for LA )

·       New file created

Wisconsin

    -   Negative values leave as Is per Adam

·       Fix State agency standardization metric

·       New file created

Florida

·       School level fixed (removed null field creation) and standardized

·       County field fixed where null used sponsor county field to fill in

·       Target Area  -- need fix  (work with George)

·       New file created

New York

·       Add Enrolment field from lunch template and merge with break fast

·       Fix FR enrolment Formula

·       New file created

California

·       Fixed operating days Lunch only column to fix FR lunch ADP numbers

(work with George)

Washington

·       Fixed CEP field

·       New file created

Louisiana

·       Target Area is null for some areas check  (work with George)

Montana

·       Fixed Zip code to have only 5 digits .

·       Check on Target Area

·       Formula change for FR enrolment %

·       Check Target Area (work with George)

(get more info from Adam for County Duplicate data)

Michigan

·       Unique ID  fix

·       Check on school id and district id

·       Recheck formulas

(waiting on some confirmation from Pam)

·       Keep District id and school id as is and move forward

·       New file created

South Caroline 17 - 18

·        Fix enrolment total and free from 6830 to 683

·       New file Created

Nebraska  18-19

·        School id length fix to constant 4 digits in unique id

·       New file created

Arizona

·       FR enrolment fixed

·       Breakfast and lunch numbers are showing up in clean data prep

·       New file created

New Hampshire

·        Duplicate issue is address data template 2

·       Combine or merge Meal type category for lunch and breakfast .

·       New file created