Weekly Challenge

Solve the challenge, share your solution and summit the ranks of our Community!

Also available in | Français | Português | Español | 🇮🇳

IDEAS WANTED

We're actively looking for ideas on how to improve Weekly Challenges and would love to hear what you think!

SUBMIT FEEDBACK

Weekly Challenge

Challenge #54: Data Prep Address Parsing

GeneR
Alteryx Alumni (Retired)

The link to last week's challenge (exercise #53) is HERE.

This week's challenge is to parse out City, State and ZIP code from some unformatted input data.

The data is in a nonstandard format - it is missing commas and some city names are two words and some city name are only one word, making parsing a challenge. You need to be able to parse out the city name, state, and zip code if available.

Your goal is to create a process that will transform the data into a data table with separated columns for City, State, and ZIP.

Enjoy and as always I look forward to seeing some creative solutions.

challenge_54_start_file.yxmd

challenge_54_solution.yxmd

Basic | Data Preparation | Join | Parse | Preparation

Share | 17 LIKES | Reply

mceleavey
17 - Castor

Seeing as we are Alteryxing here, I decided to go a little further and "repair" the gaps in the data, basically because it's easy in Alteryx, so why not?

▷ Spoiler

Challenge 54 Solution.yxmd

Share | 18 LIKES | Reply

GeneR
Alteryx Alumni (Retired)

Nice!

Share | 0 LIKES | Reply

Hi,

So my license doesn't have access to the address parsing tools, so I took a more manual approach.

▷ Spoiler

challenge_54_start_file_complete.yxmd

Share 6 LIKES Reply

TomWelgemoed
12 - Quasar

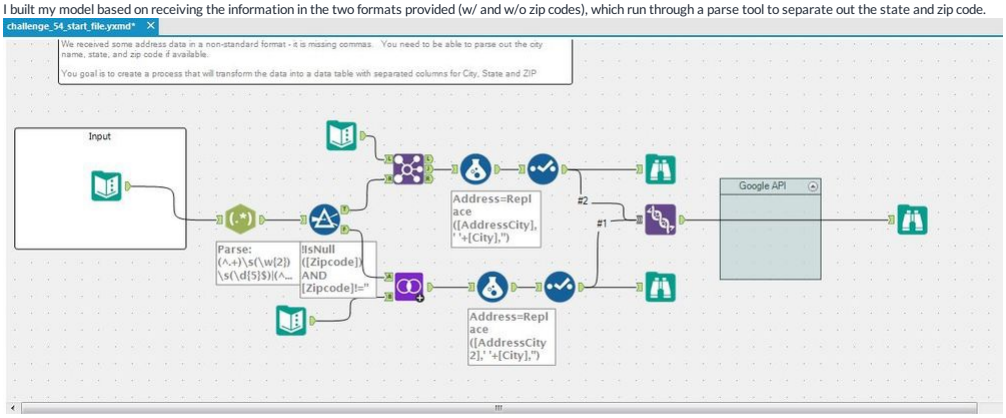
Great stuff Phil for doing this without address inputs.

Share 0 LIKES Reply

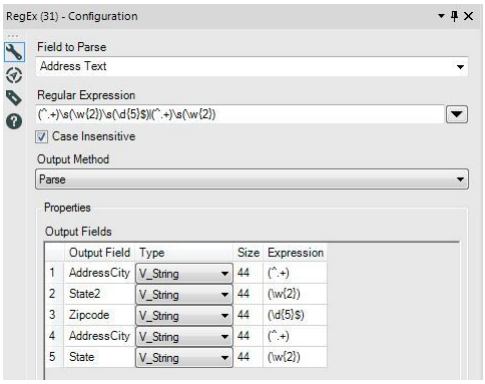
andrewdatakim
12 - Quasar

Hi Everyone,

I built my model based on receiving the information in the two formats provided (w/ and w/o zip codes), which run through a parse tool to separate out the state and zip code.

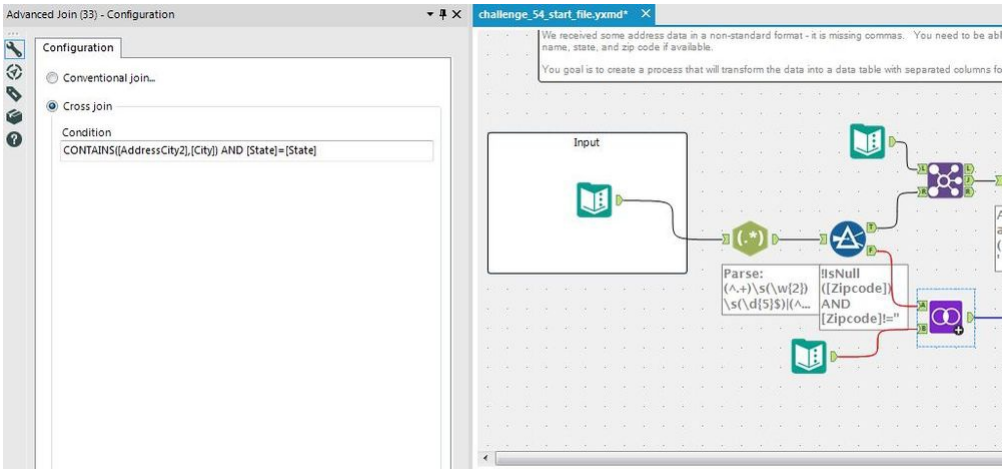


Address Parse Workflow



RegEx Address Parse

The next step was to divide the two into those with and without zip codes. The ones with zip codes can be referenced to a zip code repository and joined giving you all of the information you are looking for with the correct formatting. Those without zip codes need to be cross referenced using an Advanced Join (this tool allows you to specify multiple criteria for a cross join https://gallery.alteryx.com/?_ga=1.45648100.198713632.1460495285#!app/Advanced-Join/547f8df96ac90f0f...)



Advance Join Logic

The next set of formula tools removes the cities from the addresses and leaves the Selects and the Union for cleanup. The addition that I would make for further improvement is to add the Google API to run the address records where zip code was not given to add the zip code, at which point any zip codes that you were given that are either not in your repository or are invalid could be corrected/ stored.

The major weakness of this workflow is the initial parse which can currently only process the two options. I would want to create a repository for the most common submissions and insure that it stays case insensitive.

Address Parse from Raw String.yxzp

Share 5 LIKES Reply

NicoleJohnson
15 - Aurora

My solution. I got a bit stuck (clearly, i need a crash course in RegEx, and don't have access to the address parse tools either), so I ended up deciding to create a text input of common street names (Street, St, Drive, Dr, etc.), which allowed me to split out the number & street after using the Find & Replace tool. From there, string formulas let me do the rest. For this to work for any address, I'd likely need to expand the common street names text input to include additional ones (Place, Close, Court, etc.) It's not elegant, and I'm not convinced yet that it would work 100% of the time... but it worked.

Spoiler

challenge_54_NicoleJohnson.yxmd

Share 2 LIKES Reply

SeanAdams
17 - Castor

solution attached. Bit bashful about this one 'cause it's just not pretty (in fact - went back and re-did it once I checked the solution afterwards :-))
Love Mark's idea (@mceleavey to use the Google API to do some cleanup)

Spoiler

challenge_54_SeanSolution.yxmd

Share 2 LIKES Reply

estherb47
15 - Aurora

Had some challenges with the RegEx when I was trying to do it with one RegEx parse tool (could not get it to select from the numbers through the Street/Road/Ave/etc. Still managed to clean this with 3 tools!

Spoiler

challenge_54_EHB_solution.yxmd

Share 2 LIKES Reply

LordNeilLord
15 - Aurora

This was most definitely not a beginner exercise!

Spoiler

challenge_54_LNL.yxmd

Share 2 LIKES Reply
