

Registration Roll Churn Analysis for Dummies

EDA ELECTION DATA ANALYZER DEC 6, 2023

Heart, Comment, and Share icons

If you are a database administrator, developer, programmer, script writer or know how to use the Excel Data Model, this article is NOT.....for you. Take the EXIT ramp. LOL.

However, if you are a person of minimal to average technical skills and with the ability to use Excel in basic ways and, very importantly, only need to look at less than 1MM records (or less) for each snapshot period, the below method is a good introduction to the topic. You can use this method verbatim or some variation that works in your jurisdiction.

Thanks for reading Election Data Analyzer (EDA) Newsletter! Subscribe for free to receive new posts and support my work.

Type your email... Subscribe

Step 1

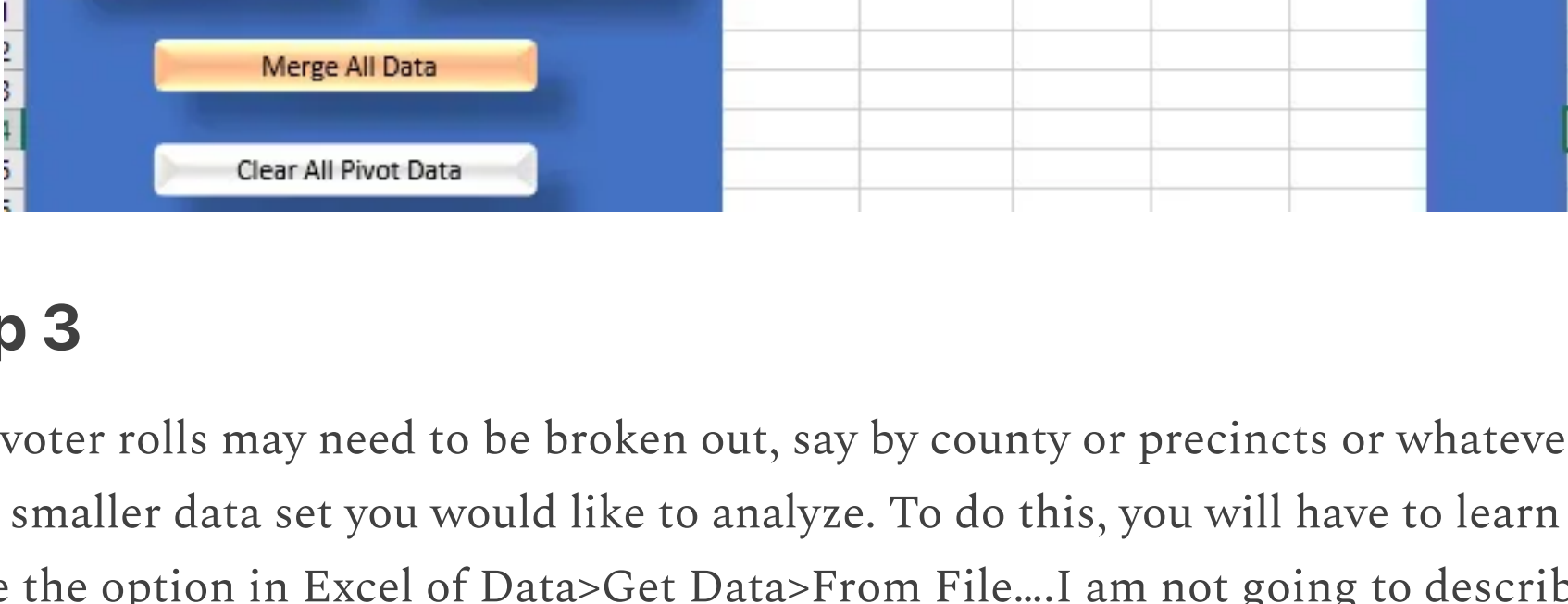
Obtain snapshots of your voter registration rolls. I typically use monthly snaps, but they can be at any periodicity, weekly, bi-monthly, quarterly etc...With this method, 3 snaps would be the minimum needed. 5 snaps is the maximum that can be analyzed in a single Excel tool that I built call the TPDA. If you have more than 5 snaps, which I typically do, you can combine the results of multiple TPDA files into a master data set.

Step 2

What is TPDA? This is an Excel.xlms file I developed for snapshot comparison. As stated above, if you want to look at smaller data sets, it works just fine. If you are looking to analyze ALL the counties in your state, you will need a different method using a script and a database. I am not going to cover database scripts. That is for the audience I cater too.

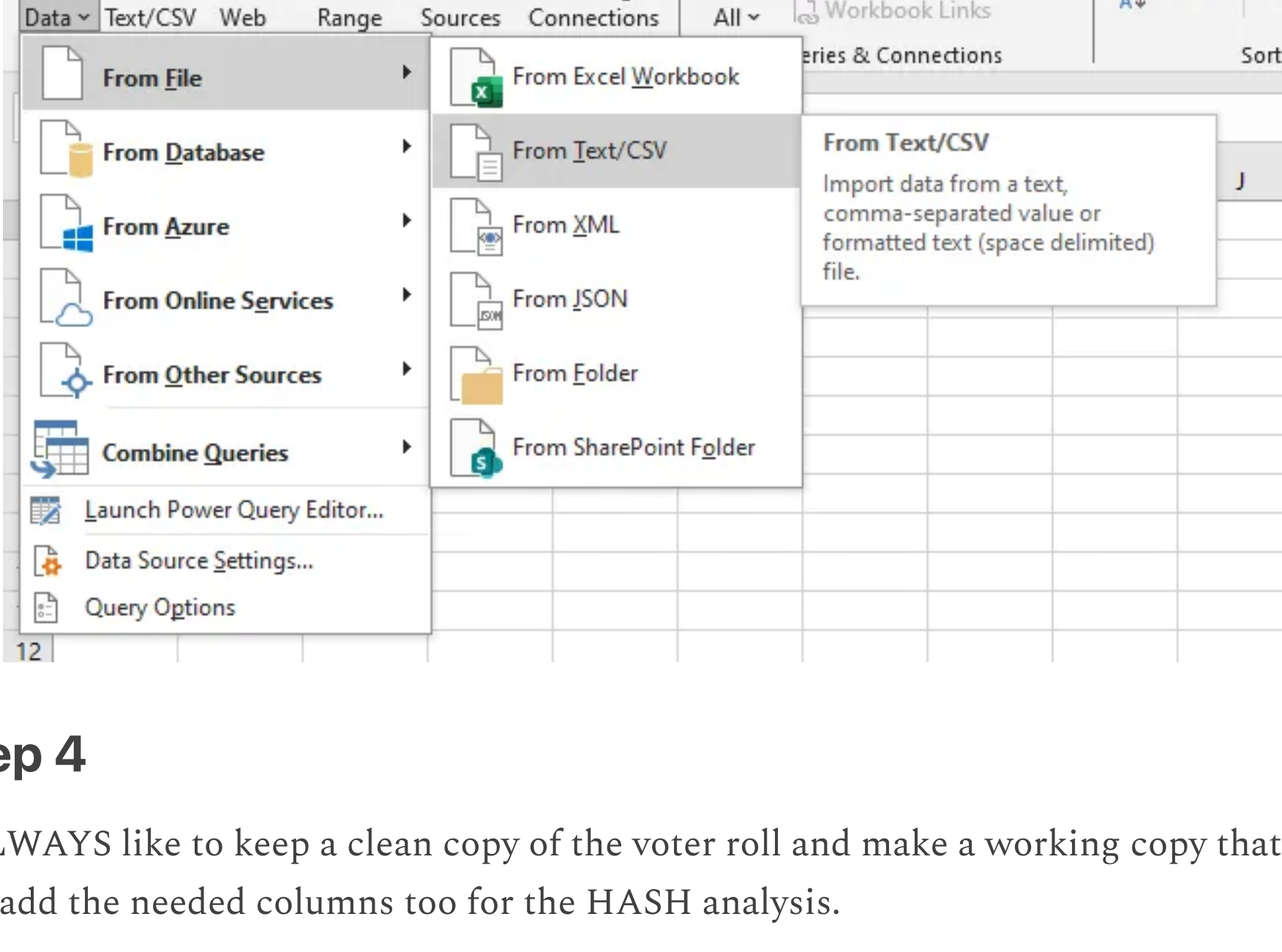
TPDA can be obtained on the Cause of America site only after joining the EDA team. Contact me there. I am looking to upload the TPDA file itself so it can be downloaded on demand. TPDA contains macros so if you are not comfortable downloading a file with macros, then proceed no further. Macros have to be enabled.

TPDA looks like this.



Step 3

Your voter rolls may need to be broken out, say by county or precincts or whatever other smaller data set you would like to analyze. To do this, you will have to learn how to use the option in Excel of Data>Get Data>From File....I am not going to describe this any further. You can seek out help articles on this topic. Your job here is to get your raw data organized and consistent. For example, extract out from the master file monthly snaps for one county over 6 months. Perhaps your data is already delivered to you by county in monthly snaps in which case, skip this step.



Step 4

I ALWAYS like to keep a clean copy of the voter roll and make a working copy that you can add the needed columns too for the HASH analysis.

Step 5

Using the working copy, take the .csv data and save the file as an .xls file. Then insert a Table on the data set (all columns all rows) using headers. This is standard Excel stuff so if you need to learn this part, look at a help article for Excel. After the table is created, add a column header called HASH on the far right of the table. Save it. Do that for all of your working copies of the rolls.

Step 6

Add the HASH formula to the new column. If you have ZERO experience with formulas in EXCEL then there may be some trial and error until you get it right. I will provide most of the exact syntax below. When I am building new formulas, I don't insert the formula in the table until I know it is correct. So pick out a single cell outside of the table and play around with a HASH like this.

= "1. "&[Voter ID]&"2. "&[Name]&"3. "&[Address]&"4. "&[Mailing Address]&"5. "&[Reg Date] etc.....

You need to use the syntax of the column headers in your voter roll and build out your HASH. Start with just a few properties until you get the hang of it and then build out the HASH with as many properties as you like. If it done correctly, the formula will yield this:

1. 00056789 2. Smith, Jim 3. 1234 Oak Street 4. PO Box 1800 5. 09/14/22 etc....

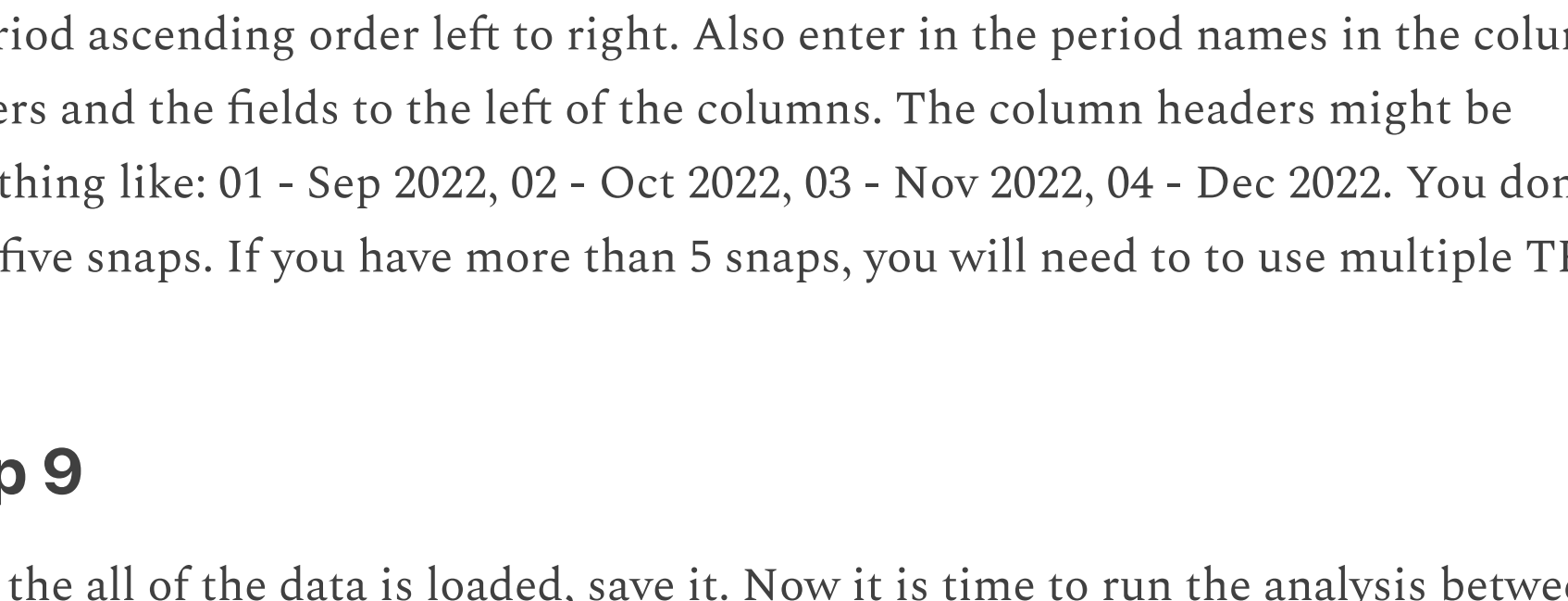
This is a long confusing string. I like to insert numeric delimiters in the HASH as markers so that later....we can parse out certain data, if needed. But more on that topic of PARSING later. You don't have to include the number if you don't want to.

Step 7

Very carefully, copy the formulas into the table first row in the HASH column and allow the table to calculate the HASH. This is going to take a few minutes and will substantially increase the size of the working file. After the formula is applied, you can copy the VALUES and paste back the VALUES only into the same column to make the file smaller. This writes over the formula. But before you do that, save the formula as you will need it for the other voter rolls. Repeat this HASH process for all roll files.

Step 8

Now we are ready, finally, to use the TPDA file itself show again below.



There are 5 blank white columns between the blue sections. You are going to paste in the VALUES only, not the formulas, into these columns of the HASH data. Paste them in period ascending order left to right. Also enter in the period names in the column headers and the fields to the left of the columns. The column headers might be something like: 01 - Sep 2022, 02 - Oct 2022, 03 - Nov 2022, 04 - Dec 2022. You don't need five snaps. If you have more than 5 snaps, you will need to use multiple TPDA files.

Step 9

After the all of the data is loaded, save it. Now it is time to run the analysis between snaps that will compare the HASH values from period to period. You will need to select the buttons in order and then wait and let it run. Depending on the size of your data set and computer, it will take some amount of minutes, perhaps even 10-15 minutes. Be patient and follow the prompts. The order you should select the buttons are

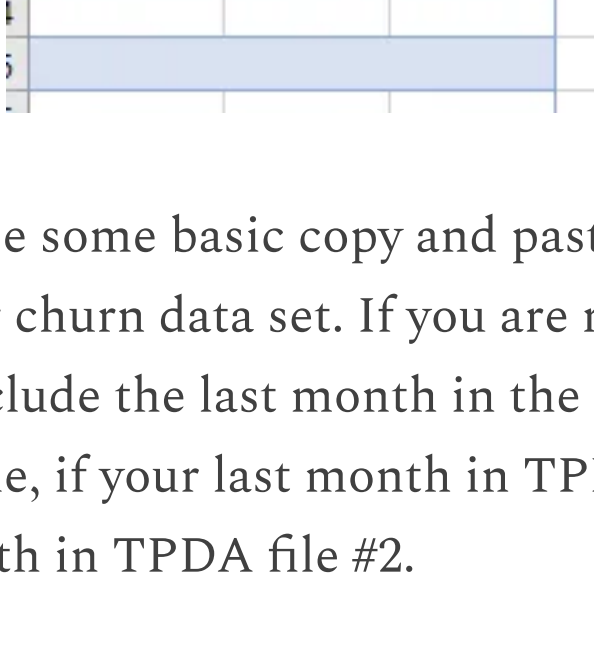
- Deletions: 1 to 2
- Additions: 2 to 1
- Deletions: 2 to 3
- Additions: 3 to 2
- etc....

Once all of those scripts have run on the data, create your merge data set by selecting "Merge All Data".

This is the holy grail....what I have referred to as the CHURN. Save it.

Step 10

In the TPDA file, locate the merged data set by going to the Dashboard tab.



From here, you will need to use some basic copy and paste skills. Copy out the results into a new file for your master churn data set. If you are running multiple TPDA analysis files, make sure to include the last month in the next file so that the data is linked across files. For example, if your last month in TPDA file #1 is Dec 2022 then make Dec 2022 your first month in TPDA file #2.

Combine all of the merge set data from all of your TPDA files into one master file and make a table out of that.

You have now after a bunch of time and very tedious work, extracted the HOLY GRAIL of roll registration analysis. Congratulations.

Step 11

From this point forward, it becomes difficult to proceduralize what to do with your churn data set. It will involve some creativity in how to use Excel or a database to analyze the data and come to some conclusions about it. You may need to parse out some fields from the HASH to gain some insights.

However, one standard convention I would always recommend is that you add a column to your master churn data set that permits analyzing the data in a time series. Apply the value to each period of churn data like this:

- 01 - Deleted July-Aug 2022
- 02 - Added Aug-July 2022
- 03 - Deleted Aug-Sep 2022
- 04 - Added Sep-Aug 2022
- etc....

The above should be applied to the appropriate rows in your combined churn data set.

Step 12

Parsing out data from the HASH is sometimes useful. There are many ways to do it. For example, I was interested in parsing out the party affiliation. To do this I chose to follow a two step process. I added three columns to my master churn data set called REP, DEM, UAF. I then used the following formula for each column.

- =IFERROR(IF(SEARCH("REP",[@[HASH]]),1),1),0)
- =IFERROR(IF(SEARCH("DEM",[@[HASH]]),1),1),0)
- =IFERROR(IF(SEARCH("UAF",[@[HASH]]),1),1),0)

This allowed me to calculate the party for each HASH and also gave me counts by Party.

I then added another column to the master churn set called Party and put this formula in it.

- =IF(@REP=1,"REP",IF(@DEM=1,"DEM",IF(@UAF=1,"UAF")))

These two steps effectively parse out the Party and Party counts from the HASH. There are may ways to do this and many other values that can be parsed out and analyzed.

Wrap Up

The goal here is to find the master Churn data set. What you do with it after that, requires some creativity for the analysis. Make no mistake, the Churn data set is gold, you just need to mine it.

References

El Paso County CO Voter Roll Wrap-up

ELECTION DATA ANALYZER - DEC 2

Read full story ->

Thanks for reading Election Data Analyzer (EDA) Newsletter! Subscribe for free to receive new posts and support my work.

Type your email... Subscribe

Comments

Write a comment...

Top, New, Community

Cast Vote Records for Non-Math-Minds

Visualizing CVR data

MAY 27, 2022 - ELECTION DATA ANALYZER

The NGO Project: CTCL 50 State Wrap Up Analysis

Voter Influence Peddling

JUN 5, 2022 - ELECTION DATA ANALYZER

Election Machines Are Immoral

An Op-Ed by the EDA

OCT 14 - ELECTION DATA ANALYZER

See all >

Ready for more?

Type your email... Subscribe