### 2023: Week 4 - New Customers

[January 25, 2023](https://preppindata.blogspot.com/2023/01/2023-week-4-new-customers.html)

Challenge by: Jenny Martin

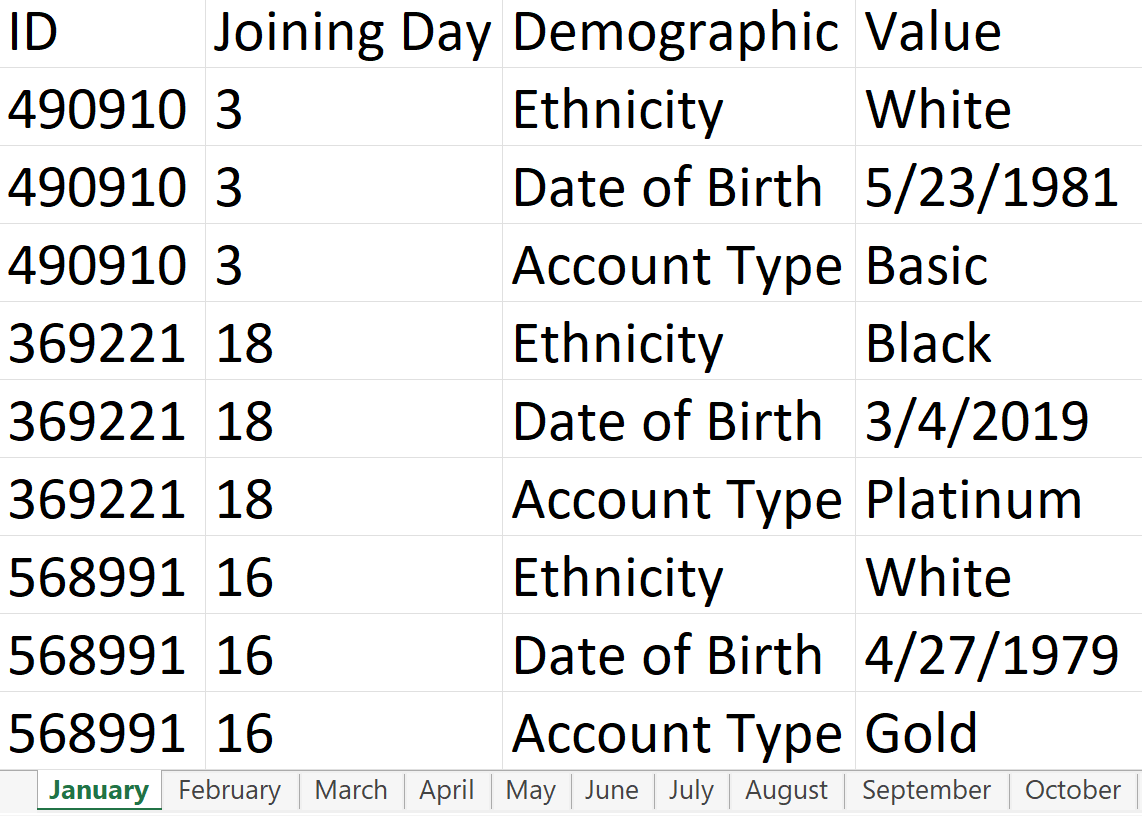
It's the final week of beginner month already! Hopefully you've picked up a lot of new skills and have enjoyed the challenges. We'll add another new technique and build on your existing skills with this challenge.

Data Source Bank acquires new customers every month. They are stored in separate tabs of an Excel workbook so it's "easy" to see which customers joined in which month. However, it's not so easy to do any comparisons between months. Therefore, we'd like to consolidate all the months into one dataset.

There's an extra twist as well. The customer demographics are stored as rows rather than columns, which doesn't make for very easy reading. So we'd also like to restructure the data.

### Inputs

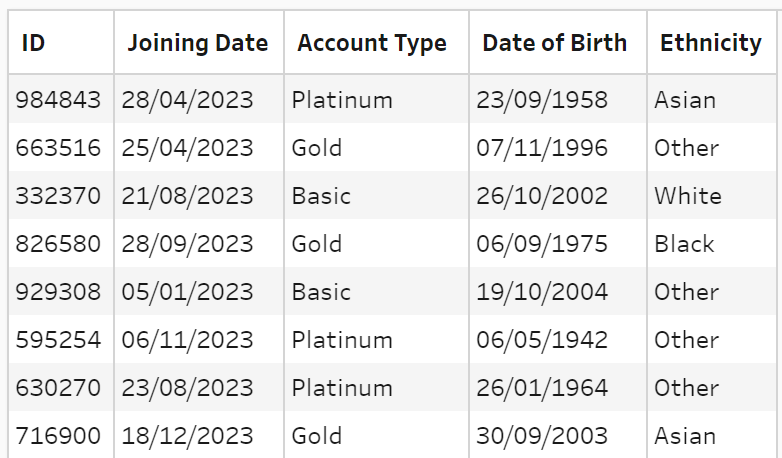
We have one excel file, with 12 different tabs, one for each month:

[](https://blogger.googleusercontent.com/img/a/AVvXsEiaT8Ijorc_YYWPq3od7zn-iyjOm2qzEdvgd0QHLYL7VTuJc5kiNM-SnT3da40PRBJYbNSK1pZs2mp13ieNm2gLXxg1QGI4QXYSYI0FepBFkEl-k25mBRALEfzecQkU16UcUnnWhjFPlg5ZJn0vUFzC9oXTAMN2MzDR6JC4Rcxjx2CNHnJ2kF4U0fQOEQ)

### Requirements

* [Input the data](https://docs.google.com/spreadsheets/d/1fwAMNqQZVMusRIlI2G6W0OHYjcH-DfwM/edit?usp=share_link&ouid=113142875990517226017&rtpof=true&sd=true)
* We want to stack the tables on top of one another, since they have the same fields in each sheet. We can do this one of 2 ways ([help](https://preppindata.blogspot.com/2019/10/how-tounion.html" \t "_blank)):
  + Drag each table into the canvas and use a union step to stack them on top of one another
  + Use a wildcard union in the input step of one of the tables
* Some of the fields aren't matching up as we'd expect, due to differences in spelling. Merge these fields together
* Make a Joining Date field based on the Joining Day, Table Names and the year 2023
* Now we want to reshape our data so we have a field for each demographic, for each new customer ([help](https://www.youtube.com/watch?v=1Pw1Ql6q0fQ&list=PL_t5OlLHbVGzjl3ygGsOnXbr9sxPOrogY&index=23" \t "_blank))
* Make sure all the data types are correct for each field
* Remove duplicates ([help](https://www.youtube.com/watch?v=-GW4EaOIKno&list=PL_t5OlLHbVGzjl3ygGsOnXbr9sxPOrogY&index=21" \t "_blank))
  + If a customer appears multiple times take their earliest joining date
* [Output the data](https://drive.google.com/file/d/1oyEoYwhZ2DOIXqiXPG2A3s_iHFDrfzAt/view?usp=share_link)

### Output

[](https://blogger.googleusercontent.com/img/a/AVvXsEi4ifWKQhn7mROh3Jawzrq0KjWKzRDHenbG1DoEEq1aA9_QyY8jU7ZKlIT1csWJCgbt8-Eg-P8WjRbXS9KXxjr1fDursLa2YTAu3RuL76AtRCSvOjI5ZGDRbV5JT-ItF-zJXcEPxWuFNL7t6izF0d2dIbiJBD1nuY_BAb5tqcv7bTjJz5fO-Bu1eM8XCg)

* 5 fields
  + ID
  + Joining Date
  + Account Type
  + Date of Birth
  + Ethnicity
* 989 rows (990 including headers)

You can download the [output from here](https://drive.google.com/file/d/1oyEoYwhZ2DOIXqiXPG2A3s_iHFDrfzAt/view?usp=share_link" \t "_blank)

Root link:

<https://preppindata.blogspot.com/2023/01/2023-week-4-new-customers.html>