STEP BY STEP PROCESS FOR MINI PROJECT 3 ANALYSIS: **14** STEPS TOTAL

1. I downloaded data from Maryland State Board of Elections website for Democratic and Republican voting data from the 2016 general presidential election.
2. I cleaned up all of the data by:
   1. Separating Democratic and Republican data into two different sheets within the workbook called “Akhila Mini Proj 3 Analysis”
   2. Unmerging columns that were merged by default and deleting columns that weren’t necessary
   3. Highlighting the columns I was going to use with blue or red, depending on the party (Blue for Democrat, Red for Republican)
3. I added another column for each party called “total\_voters\_dem/rep” that summed the POLLS, EV, ABS, and PROV values for each county to get the total number of voters.
4. I consolidated all the data into another worksheet (named Consolidated Data BEFORE SOLVER) by using VLOOKUP with table arrays from the “Cleaned Up Democrat” and “Cleaned Up Republican” worksheets.
5. I calculated the mean and standard deviation for each of the four variables: turnout\_dem, total\_voters\_dem, turnout\_rep, and total\_voters\_rep.
6. I calculated the z-scores for each of the variables for each of the county to measure the distance of each of the values from the mean using the STANDARDIZE function.
7. I numbered each of the counties 1-24 by adding a column “anchor number” next to the county names.
8. I chose three random counties (1, 12, and 22) and wrote the corresponding number down. These would be anchors 1, 2, and 3.
9. Using VLOOKUP, I filled out all of the z-scores for each of the variables for the three selected counties.
10. I calculated the distance squared for each of the 24 counties’ z-scores from the z-scores of anchors 1, 2, and 3 using the SUMXMY2 function.
11. I calculated the lowest distance squared among each of the anchors for each county and matched that minimum number with the corresponding randomly chosen anchor numbers using the MIN and MATCH functions.
12. I calculated the sum of the minimum distance squared for all of the counties and used that as the objective for the Solver tool.
13. I used the Solver tool to determine what the actual anchors are for the data – anchors that correctly represent the three distinct clusters. I set constraints for the values to be >= 1, <= 24, and an integer.
14. I used the new anchors to analyze and interpret the data.