STEP BY STEP PROCESS FOR MINI PROJECT 1 ANALYSIS: **12** STEPS TOTAL

1. I downloaded data from Opportunity Atlas as 8 separate Excel workbooks. Each workbook represented the teenage birth rates for a specific race in a specific location. Four races (Black, Asian, Hispanic, White) multiplied by two locations (Baltimore City, MD and Morris County, NJ) = 8 workbooks.
2. To consolidate all of the data into one workbook, I created another spreadsheet called “Akhila Mini Proj 1 Analysis” and copied and pasted the data in the Baltimore black workbook to serve as the starting point (I could’ve used data from any of the workbooks, I just chose that one.) Now I had each tract, the name of the tract, and the teenage birth rate (which I abbreviated to TBR) for the black race in the workbook.
3. I then used the VLOOKUP tool to transfer the TBRs for the other races (Asian, Hispanic, White) within Baltimore into the main workbook. I matched up the data by tract; since the tracts were the same among the sheets in Baltimore, I was able to use VLOOKUP to place the correct data by tract. An example formula is: =VLOOKUP(A2,the array from the worksheet for the individual races,3,FALSE). A2 represented the tract that I wanted it to match, and the 3 meant that the TBR was in the third column of the workbook I wanted to transfer data from. Then by double clicking the bottom right of the cell, I got the data for all of the tracts.
4. To make sense of the data, I made a Pivot Table and a Chart. I made the tracts the row and the average TBR for each of the races the values. This made the table show the corresponding TBRs to each tract very clearly. Then I produced a chart for the Pivot Table. I added chart elements like chart title and axis title.
5. I did the exact same VLOOKUP and Pivot Table & Chart process for Morris County, NJ.
6. In the next sheet, I copied and pasted the charts from Baltimore and Morris County side by side so I could analyze them. My comparison is written in the sheet itself as well in the README.
7. I was curious to see which tracts in each location had TBRs that were above the average rate for the black race. I transferred the tract numbers and the corresponding TBRs for the black race into another sheet (for each location). Then I calculated the average of each location’s TBRs with =AVERAGE(TBR value array). Then I created a column in each location that would track the tracts that were higher than the average. I used an if statement to do this. The formula format was: =IF(TBR value>average,1,0). I locked the average so that it would use the same average for each cell. So now I had a column for each location that had a bunch of 1s and 0s – 1 for above average and 0 for not above average.
8. I copied and pasted the tract numbers for only the 1s for each location into another sheet. Then I used VLOOKUP again to match the tract numbers with their name using the initial data sheet. The formula format was = VLOOKUP(TractNumber,data array from the original data worksheet,2,FALSE). The 2 represents that the name was in the second column of the array.
9. I counted how many tracts were in the new worksheet using =COUNT(all the tract numbers that corresponded to an above average rate) for each location.
10. I counted how many tracts each location had overall using =COUNT(all tract numbers from original data sheet).
11. I divided the results from Step 9 by Step 10 to get the percentage of tracts that had rates above average for the black race for each location.
12. I wrote down some questions that can be answered for future research based on what I did.