1. Baltimore Employee Salary data was downloaded.
2. Data was filtered to only include object id, annual salary, gross pay and these columns were copied into a new sheet
3. These values were used to create a scatterplot of annual salary vs gross pay to determine how many clusters should be chosen (select both columns, insert scatter chart, add appropriate axis labels and insert shapes for lines to divide the data)
4. The mean and standard deviation of annual salary and gross pay were calculated using excel formulas (=AVG,=STDEV)
5. The formulas were applied to find standardized values for annual salary and gross pay
6. The distance between clusters was found using the excel formula: =SUMXMY2($D$3:$E$3,H12:I12) for dist1 and it was repeated and modified for clusters 1-4 and distances 1-4
7. The minimum squared distance was found using the =MIN() formula which chose the minimum squared distance out of the four options
8. Each data point was assigned to a cluster using the =MATCH() formula
9. Each cluster was given a anchor name so the anchor name for the anchor would show during the cluster analysis using =vlookup and the cluster data as the table array
10. The sum of the minimum squared distances was found using a =SUM() formula for the minimum squared distances column and this value was selected when the cluster analysis was run
11. The cluster analysis was run using solver
12. The solver add-in was enabled and the variable values were chosen as the four anchors, with the following parameters: Graphical user interface, text, application, email

    Description automatically generated
13. Under options, the ‘ignore integer constraints’ option was unchecked and under evolutionary, the mutation rate was changed to 0.5
14. The cluster analysis was run and the results from it were accepted
15. The final cluster nodes were analyzed to determine what they meant
16. A =COUNTIF formula was applied to determine the amount of job titles within each cluster, which allowed for the creation of a bar chart reflecting this data (same process as earlier chart but select bar chart type and disregard adding lines)
17. A final visualization was created by using the final anchor nodes and their data interpretations to reflect the characteristics found for each anchor node. Each cluster’s anchor node was included in a table and the corresponding interpretation of what their annual salary and gross pay compared to the average were added and arrows represent above or below average