***Step-by-Step Data Analysis***

1. Data Organization: pull “HireDate,” “AnnualSalary,” and “GrossPay” categorized by “AgencyName” only from FY2020
   1. assign “AgencyNumber” to individual data in “AgencyName” column
   2. isolate “Year Hired” from “HireDate” using YEAR function
   3. compute “Years Worked” with =2020-(Year Hired)
   4. compute z-scores for “Years Worked,” “AnnualSalary,” and “GrossPay” using STANDARDIZE function
   5. compute mean and standard deviation for “Years Worked,” “AnnualSalary,” and “GrossPay”
2. Cluster Analysis Preparation:
   1. create initial cluster table with three arbitrarily chosen “AgencyNumber”
   2. pull “AgencyName” and z-scores for each “AgencyNumber” using VLOOKUP function
   3. compute distance of data points to anchor numbers 1-3 using SUMXMY2 function with arbitrarily chosen agencies as reference data points
   4. compute minimum distance of data points to anchor number 1-3 using MINIMUM function
   5. compute the anchor number that each agency belongs to using MATCH function
   6. compute sum of minimum distances of data points to anchor numbers 1-3 using SUM function
3. Run Cluster Analysis: create constraints as integers equal to or between 1 and 17,225 (total number of data points = 17,225)
4. Data Visualizations:
   1. pull data for “Years Worked” and “GrossPay” and create scatter plot with trendline and R-squared value
   2. filter AnchorNumber to computer number of data points in each cluster, create bar graph comparing values