Excel Step-by-Step Data Manipulation

Opportunity Atlas

Filtering and Manipulating Data

1. Downloaded Baltimore Median Household Income data from the Opportunity Atlas for All Income and All Race subgroups
2. Used Text-to-Columns tool to focus only on neighborhoods within Baltimore City

Creating Graphs

1. Created a Pivot Table in order to find the median household income for each neighborhood and inserted a bar graph (*Baltimore City Median Household Income by Neighborhood)* to visually interpret the results
2. Checked neighborhoods against the official Baltimore City neighborhood list to ensure that all neighborhoods included were a part of Baltimore City

Open Baltimore

Filtering and Manipulating Data

1. Downloaded the Vacant Buildings data from Open Baltimore, last updated on October 26, 2020
2. Added a new column, Number of Vacant Buildings, to each row as data was presented by address and neighborhood

Creating Graphs

1. Created a Pivot Table in order to sum the total number of vacant buildings per each neighborhood
2. Inserted a bar graph to represent this visually (*Total Number of Vacant Buildings in Baltimore City by Neighborhood)*
3. Used Pivot Table again to look at the vacant buildings total for each police district
4. Inserted a bar graph to represent this visually (*Total Number of Vacant Buildings in Baltimore City by Neighborhood)*

U.S. Census

Filtering and Manipulating Data

1. Downloaded the U.S. Census data for Baltimore City and focused on pct_black (percent Black residents per neighborhood), pct_white (percent White residents per neighborhood), racial_diversity_index (the diversity score for each neighborhood from 0-100), pct_households_belowpovertyline (percent of households below the poverty line)

Combining Data Sets

1. Used =VLOOKUP to match the median household income data and vacant buildings data to each neighborhood row in the U.S. Census datasheet
2. The U.S. Census data combined several neighborhoods together along with having individual neighborhoods for their rows, unlike the Opportunity Atlas and Open Baltimore data sets which had individual neighborhoods

3. Created a combined data set that encompassed vacant buildings data from Open Baltimore, median household income data from Opportunity Atlas, and demographic data from the U.S. census by:

1. Averaged the median household income data for the separate neighborhoods that were grouped together by the U.S. Census using =AVERAGE
2. Found the total sum of vacant buildings in the neighborhoods grouped together by the U.S. Census using = SUM

Creating Visuals: Combined Data Sets

Bar Graphs

1. Used the combined dataset to focus on the race breakdown for each neighborhood
2. Inserted a bar graph showing *Total Vacant Buildings by Percent of White Residents per Baltimore City Neighborhood*
3. Inserted a bar graph showing *Total Vacant Buildings by Percent of Black Residents per Baltimore City Neighborhood*

Scatter Graph

1. Created a simple linear regression for median household income using =SLOPE, =INTERCEPT, =RSQ, and =STDEV for the table
2. Calculated Predicted Vacant Buildings using the y = mx + b formula and the above values
3. Calculated the error using a subtraction formula and then created an Outliers column finding outliers greater than 2*STDEV to find percentage of total outliers for the neighborhoods
4. Inserted a scatter graph with linear trendline, *Vacant Buildings vs. Median Household Income in Baltimore City Neighborhoods*

Clustering: Combined Data Sets

Cluster Analysis

1. Numbered the neighborhood groups from 1 to 33
2. Calculated the Mean and Standard Deviation using AVERAGE and STDEV functions for the five variables (pct_black, pct_white, pct_householdsbelowpovertyline, median_income, vacant_buildings)
3. Calculated the z-eligible columns using the STANDARDIZE function for each of the three variables
4. Highlighted the whole data array to name it "Cluster" and numbered the z-scores columns in the table above
5. Used the VLOOKUP function to set up our anchors in finding the names for our random cluster numbers in the table above as well as respective z-values for each of the variables
6. Used the SUMXMY2 function to calculate the distance between the variable z-scores for each agency and the z-values of the anchors
7. Used the MIN function to find the smallest distance of the four distances in each agency row
8. Used the MATCH function to assign the smallest distance to an anchor
9. Calculated the sum of all the minimum distances by using the SUM function
10. Applied the Excel Solver operations for a cluster analysis on the dataset
    a. Set objective to our sum of the minimum distances cell
    b. Selected min

c. By changing variable cells to the cluster numbers in the table
d. Subject to the constraints with the cluster cells > = 1, < = 56, and = integer
e. Selected Evolutionary as the solving method

Table with Grouped Neighborhoods

1. Filtered the cluster analysis results to see which neighborhoods were grouped into cluster 1, 2, 3, or 4
2. Copied over the neighborhoods to create a table showing these cluster groupings