Median Household Income (MHI) and Economic Status Determination Methodology

State Water Resources Control Board

Summary:

Median Household Income (MHI) is a measurement of the annual revenue a household earns within a given geographic area. Median Household Income divides the income distribution into two equal parts: one-half of the cases falling below the median income and one-half above the median[1]. Medians do not consider values at the upper and lower end of the distribution, but just focus on the middle value. Medians are often viewed as a better central measure than means due to the latter being distorted by a small number of extremely large values.

There are limitations however to using MHI. Data coverage is inconsistent; populated areas tend to be more precise than coverage over sparsely populated areas of California. The finest resolution of data also stops at the Block Group level. These block groups can cover a wide expanse of California's regions. Due to the random nature of sampling, comparisons over time are also not as accurate as one would assume. The same households are rarely being remeasured when the American Community Survey (ACS) is redone and changes in demographic characteristics are not taken into consideration. Despite these constraints though, MHI is still one of the most widely accepted and used metrics for estimating income for policy setting across the nation. The data availability is excellent and covers over multiple years and can be easily broken down into distinct socioeconomic bins.

MHI has been used historically to track economic status as well as economic trends over time. The Needs Assessment[2] for instance, utilizes MHI in assessing a system's affordability to its customer base through indicators such as %MHI[3]. The State Water Board uses MHI to determine Disadvantaged Community (DAC) and Severely Disadvantaged Community (SDAC) status. DAC is defined as an MHI that meets 80% of the statewide MHI where SDAC is set at 60% of the statewide MHI. MHI has an important impact on the available financial support for water systems. For instance, % MHI[4] is used in the Drinking Water State Revolving Fund (DWSRF) program as a metric

---

[1] Median Household Income https://www.census.gov/quickfacts/fact/note/US/INC110221

[2] State Water Board Drinking Water Needs Assessment Website: https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/needs.html

[3] % MHI: This indicator measures annual system-wide average residential customer charges for six Hundred Cubic Feet (HCF) per month relative to the annual Median Household Income (MHI) within a water system's service area. Six HCF indoor water usage per month is roughly equivalent to 50 gallons per person per day for a three-person household for 30 days.

[4] The Fund Expenditure Plan used an affordability threshold of 1.5% MHI to identify DAC water systems that may have customer charges that are unaffordable: FY 2020-21 Fund Expenditure Plan https://www.waterboards.ca.gov/water_issues/programs/grants_loans/sustainable_water_solutions/docs/sadwfep _2020_07_07.pdf

for determining whether a small DAC will receive repayable (loan) or non-repayable (e.g., grant or non-repayable) funding.

MHI is determined by joining ACS MHI data to Block group boundaries and intersecting them with the service area boundaries of all community water systems across the state of California. A weighted average MHI is then determined for each system, where the area of a block group's intersection within a water system boundary is used as the weight.

Data Sources:

- Service Area Boundary Layer (SABL) shapefile from the water49 database[5]
- Missing Service Area Boundary Layer locational data merged with SABL layer called "SABL+" (see Public Water Systems, Step 1)[6]
- 5 Year-Estimate ACS MHI Data (B191013)[7]
- Block Group Shapefiles[8]
- State Water Board Division of Financial Assistance (DFA) Income Surveys[9]

Step 1: Add Missing Systems to the Service Area Boundary Layer (SABL)

1. The service area boundary of water systems is used as the primary locational reference for all water systems included in the analysis. The SABL Shapefile includes most public water systems, however there are systems in the physical consolidation inventory missing from the dataset. Those missing systems[10] are given an estimated service area to incorporate them into the analysis.
2. The estimated service area starts from a single point of locational data tied to each system. These point locations can come from a variety of sources such as well locations, facility locations, or physical addresses.
3. In ArcPro, to merge two layers together, they need to be in the same layer format. The SABL layer is a polygon layer while the missing points are a point layer. To convert the missing points into polygons a 1-mile buffer is created around each point using the "Create Buffer" tool in ArcPro.

---

[5] System Area Boundary Layer (SABL) Look-up Tool (ca.gov) https://gispublic.waterboards.ca.gov/portal/apps/webappviewer/index.html?id=272351aa7db14435989647a86e6d3ad8

[6] SABL_Plus - Overview (ca.gov) https://www.waterboards.ca.gov/drinking_water/certlic/drinkingwater/docs/2023/20230714-final-cost-assessment-consolidation-white-paper.pdf

[7] Census Bureau Table B19013: Median Household Income in the Past 12 Months https://data.census.gov/table?t=Income+and+Poverty&g=040XX00US06$1500000&tid=ACSDT5Y2021.B19013

[8] 2022 TIGER/Line® Shapefiles (census.gov) https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2022&layergroup=Block+Groups

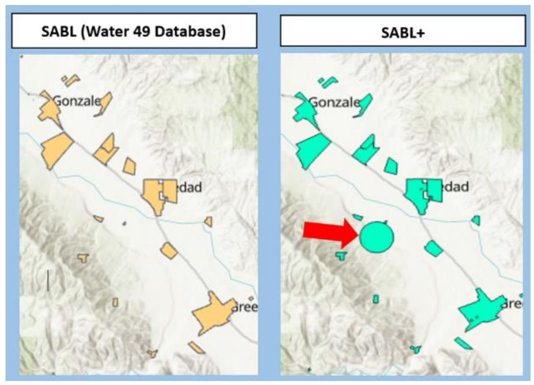
[9] The income surveys collected by the Department of Financial Assistance are not publicly available and must be requested. The income data used in this analysis is restricted to 2018 to present.

[10] The number of missing systems from the SABL layer changes over time as more systems are added. This list is specifically from a pull performed on January 1st, 2023.

a. The 1-mile distance was chosen as an approximation to the potential service area of each point since most of the missing systems are small water systems.

4. The SABL layer is then merged with these missing locations into a new layer that will be referred to as the “SABL+ layer” moving forward.

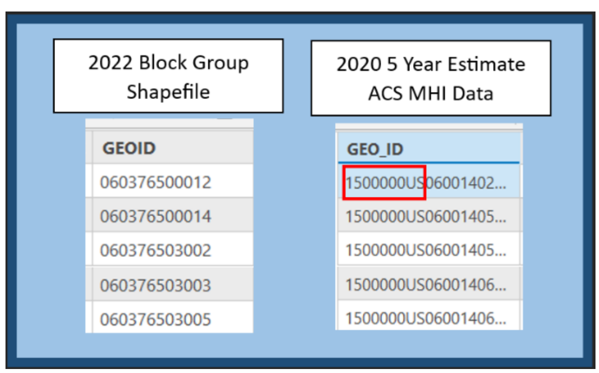
Figure 1: Visual comparison of the Standard SABL layer to the SABL+ layer. Note the addition of the estimated water system boundary for the missing system (see red arrow)



Step 2: GIS Analysis

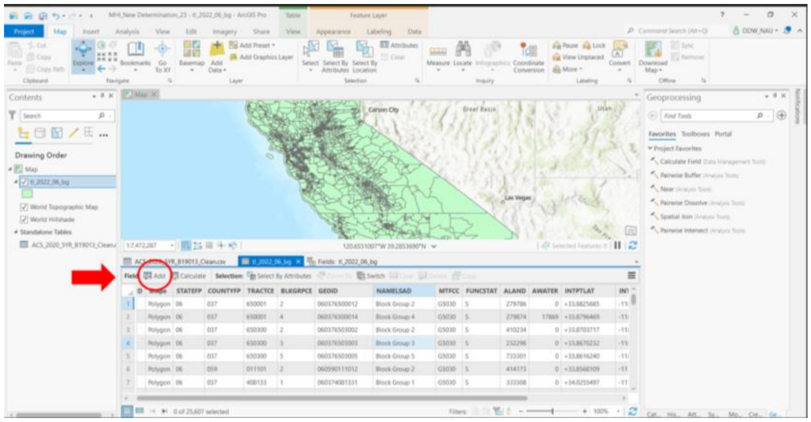
1. Import into the “Table of Contents” the following data layers:
    a. SABL+
    b. Block Group shapefiles
    c. ACS Data table with MHI.
2. For the Block Group Shapefile there is a field called “GEOID” which is a series of unique identifiers for each block group. However, this GEOID is not formatted properly for the analysis.

Figure 2: Highlight of the difference between the GEOID codes for the Block Group Shapefile and the 5 Year Estimate ACS MHI Data.



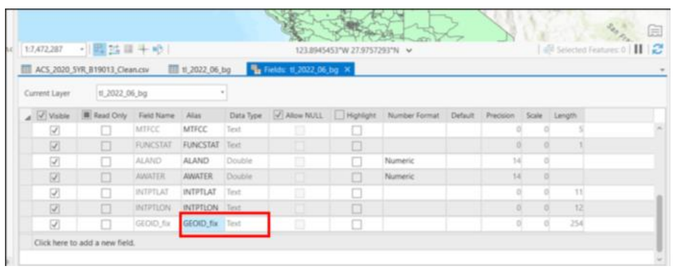
3. Use the “Add Field”.

Figure 3: Location of the “Add Field” tool, it sits on the top right corner of the “Attributes” Pane that typically opens on the bottom half of ArcPro’s User Interface.



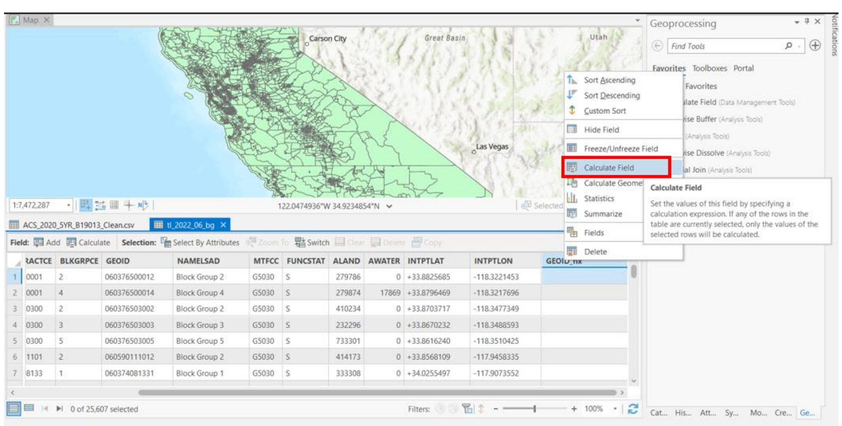
4. Label the “Alias” as “GEOID_Fix” and set the “Data Type” to “Text” and hit save when done.

Figure 4: Visual of the “Edit” window for new attributes.



5. Right click the new “GEOID_Fix” field and click on “Calculate field”.

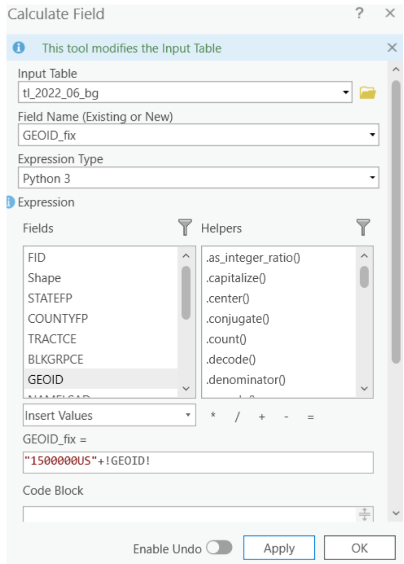
Figure 5: Visual of the “Calculate Field” tool, right click on the new field to open the drop-down menu.



6. Select the Block Group shapefile as the “Input”, “GEOID_Fix” for “Field Name” and “Python 3” for the “Expression type”.

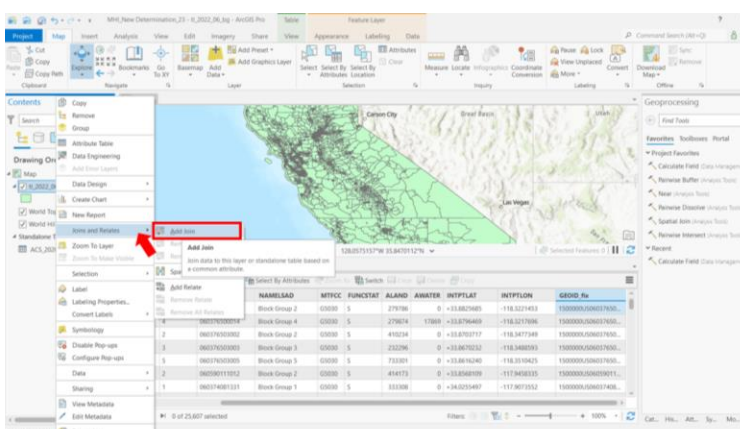
7. Input “GEODID_Fix” = “1500000US” + “!GEOID!. this expression adds the text “1500000US” Infront of the original code for “GEOID” to match the two different tables properly for the next step.

Figure 6: Calculate field window that will open after clicking the “Calculate Field” ribbon. This window can use various coding languages to specify how to manipulate the data but in this case Python 3 is being used.



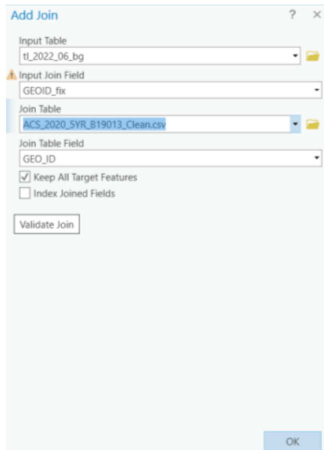
8. Right Click on the “Block Group” shapefile layer in the “Drawing Order” menu. Under the “Joins and Relates” menu, click the “Add Join” option.

Figure 7: Location of “Add Join” tool within ArcPro. Right click the layer that the join is being added to and highlight “Joins and Relates” to see the “Add Join” ribbon.



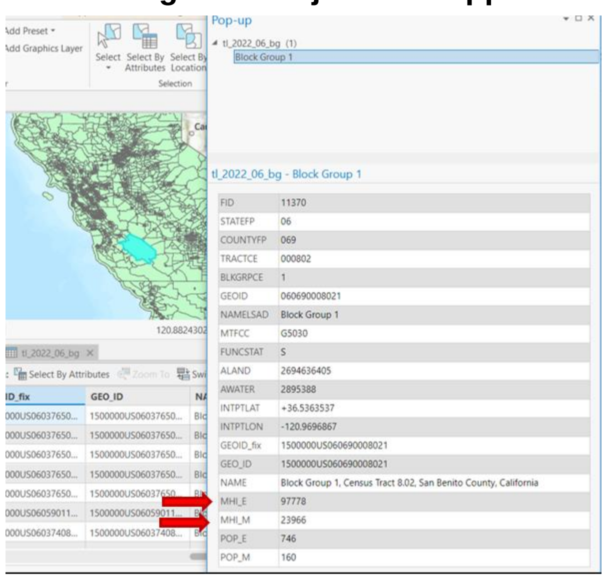
9. Select the “Block Group” shapefile for the “Input Table”, “GEOID_Fix” for the “Input Join Field”, the “2020 ACS MHI” data table for the “Join Table” and “GEO_ID” for the “Join Table Field”. Then hit okay.

Figure 8: Visual of the “Add Join” option window. Select the targeted layer to add the join to, select the table that is being joined, and finally select the field that allows records to be joined correctly.



10. Check the join by selecting any random Block Group shapefile on the map. If done successfully there will be a record for MHI (MHI_E) and MHI margin of error (MHI_M) in the attributes field.

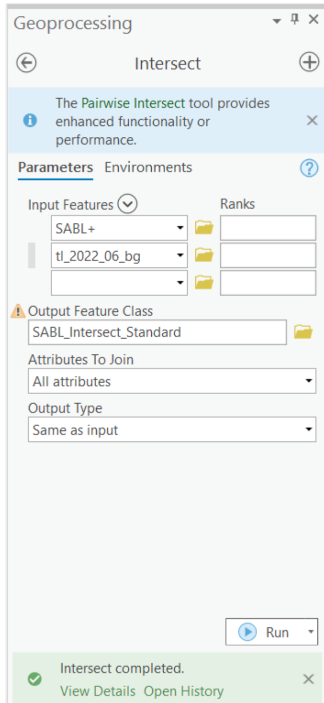
Figure 9: Use the “Select” tool located in the top ribbon under the “Map” pane to pick a random block group to check if the join was added successfully. The two arrows below show this block group has both an MHI and an MHI margin of error within its attribute table meaning that the join was applied correctly.



11. Select the “Intersect” tool and import the Block Group shapefile and the SABL layers into the tool.
12. Select “SABL+” and the “Block Group” shapefile as the “Input Features”, name the output as “Intersect” and specify where the intersect is being saved. Select “All Attributes” for “Attributes to Join” and “Same as Input” for “Output Type”.

Then hit “Run” a new layer representing the intersection of the two layers will be produced.

Figure 10: Visual of the “Intersect” tool window. Select the two layers that are going to be joined and specify where it will be saved on your project folder. Then hit run.



13. Select the “Table to Excel” tool to export the data table from the intersection layer to excel. Select the intersecting layer for “Input Table” and specify where the table will be saved on the computer. The hit “Run”.

Figure 11: Visual of the “Table to Excel” tool window. Select the layer that is going to have its data table exported and converted into an excel sheet for the

second half of the analysis. Specify where the new excel file will be saved and hit “Run”.

Step 3: Excel Analysis

1. Open the excel file that was just created from the last step. Filter the following data-points from the first tab into a “Clean_V1” new tab:
   i. PWSID
   ii. Population
   iii. Federal Classification
   iv. MHI_E
   v. MHI_MOE
   vi. Part Area

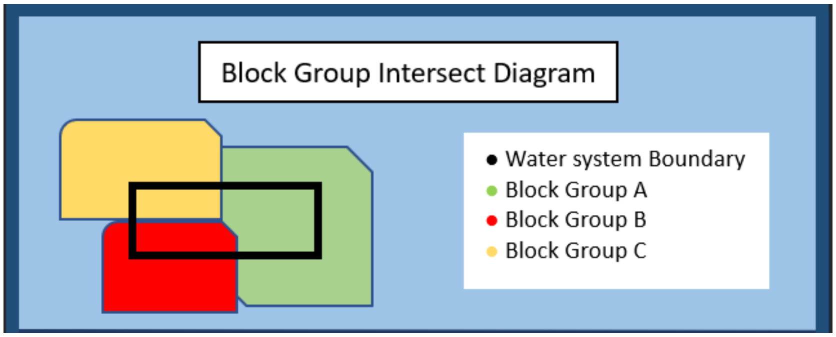
Figure 12: Snippet of the required data fields needed from the excel sheet produced by the intersection to be collected on a new tab.

PWSID	Population	Federal Classification	MHI_E	MHI_MOE	Part Area
CA3310001	270000	COMMUNITY	17269	8101	47397.81724
CA3310001	270000	COMMUNITY	17269	8101	24916784.66
CA3310007	45727	COMMUNITY	17269	8101	24916784.66
CA4910009	173648	COMMUNITY	17324	7949	232518.0242
CA1310001	26928	COMMUNITY	17328	17284	579557.6504

2. Calculating MHI
   a. Looking at Figure 13 the water system boundary is intersecting with 3 different block groups. Each of these intersections can be referred to as “Parts” and have their own MHI and area of intersection.
   b. Ideally, taking the average of all 3 block groups would work to give an average MHI estimate. However, the size of the different areas can vary and will skew the average unfairly as each Block Group does not have an equal area of coverage within the water system boundary.
   c. To overcome this MHI is calculated from a weighted average of all block group MHI’s that intersect within a water system’s boundary.

d. Where the weight is the area of the block group's intersection within the water system boundary over the total water system's area.

Figure 13: Illustration of the intersect between the block groups and the water system boundaries.

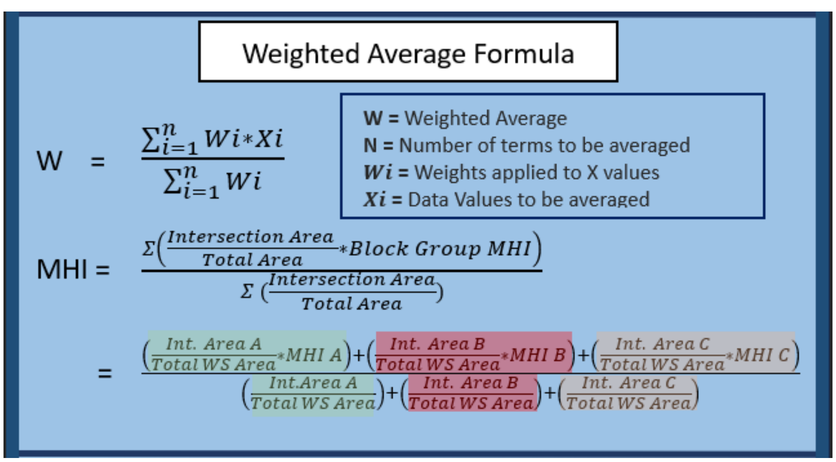


e. The weighted average is given by the following formula:
   i. $W = \frac{\sum_{i=1}^{n} Wi*Xi}{\sum_{i=1}^{n} Wi}$
   ii. Where:
      1. W = the weighted average
      2. n= the number of unique intersections within a water system
      3. Wi = weight (Area of Intersection/ Total area of Water system)
      4. Xi is the data value to be averaged (MHI by each block group)

f. This formula can be expanded to fit the data points we have access too here:
   i. $W = \frac{\Sigma\left(\frac{Intersection\ Area}{Total\ Area}*Block\ Group\ MHI\right)}{\Sigma\left(\frac{Intersection\ Area}{Total\ Area}\right)}$

g. The following diagram breaks down how this formula further:

Figure 14: Explanation of the weighted average formula used to determine MHI. The highlighted colors correspond to the Block Group intersecting areas seen above in Figure 13.



h. Before values can start being plugged into the formula, MHI for each intersection must have a Margin of Error[11] (MOE) applied.
i. In this analysis the MOE has been taken away or subtracted from the estimated MHI.
j. The method in which the MOE is applied is based on the approach the Division of Financial Assistance (DFA) utilizes for MHI determinations.

Figure 15: Diagram of how Margin of Error is applied based on the population of the water system.

Margin of Error

Water System Population	Maximum MOE that can be applied to total	Exception
500+	- $7500	If MOE < $7,500 but pop > 500 use original MOE
500-	- $15,000	If MOE < $15,000 but pop < 500 use original MOE

3. Add a new column in the "Clean_V1" tab called, "Adjusted MOE"
4. Apply the following formula:
    a. =IF(B2>500,IF(F2>7500,7500,F2),IF(F2>15000,15000,F2))

Figure 16: Snippet of the excel formula and its associated columns that is used to apply the Margin of Error formula described in figure 15.

SUM =IF(B2>500,IF(F2>7500,7500,F2),IF(F2>15000,15000,F2))

	A	B	C	D	E	F	G
1	PWSID	Population	Adjusted MOE	Federal Classification	MHI_E	MHI_MOE	Part Area
2	CA0110005	1430200	=IF(B2>500,IF(F2>7500,7500,F2),IF(F2>15000,15000,F2))	COMMUNITY	11114	122	105006.0916
3	CA1310800	4800	IF(logical_test, **[value_if_true]**, [value_if_false])	COMMUNITY	8913	375	2259883.251
4	CA1310003	7412	375	COMMUNITY	8913	375	255637.3096
5	CA3710020	1374790	417	COMMUNITY	11409	417	571008.179

    b. This formula applies the logic in figure 15 above to see what MOE will be applied to the MHI estimate.
5. Add a new Column labelled, "MHI w/ MOE".

[11] To account for the inaccuracy of the ACS MHI data, a Lower Bound MHI value is calculated by using a community's MHI Margin of Error (MOE) when determining DAC status When calculating a water system's rates as a percentage of MHI. The Lower Bound MHI is calculated by subtracting the Usable MOE from the MHI. The maximum Usable MOE applied to a community's MHI depends on the size of the community. The maximum Usable MOE for communities with 500 people or less is $15,000. The maximum Usable MOE for communities with greater than 500 people is $7,500. [Median Household Income Determinations [CWSRF/DWSRF Procedure Manual]](http://wiki/srf/doku.php?id=project:05prefa:03eligibility:01general:01mhi)
http://wiki/srf/doku.php?id=project:05prefa:03eligibility:01general:01mhi

   a. Subtract the values in “Adjusted MOE” from the values in the “MHI” column. This will be the set of MHI values that will be used in the final calculation.
6. Lastly remove any row without an MHI estimate.
   a. Some block groups do not have MHI data from the ACS data to account for missing data those records are removed from the final average.
   b. Select for all records with an MHI estimate and copy them over to a new Tab labelled, “Clean_V2”.
7. Add a new column and label as “Total Area”.
8. Select the entire sheet and insert into a pivot table, label the new tab created as “Pivot 1”.
   a. Filter the Pivot by “PWSID” for rows and “Sum of Part Area” for values.
9. Go back to “Clean_V2” and use the “vlookup” function to copy the values from “Pivot 1” into the column “Total Area” organized by PWSID.
10. Add new column labeled “Part/Total” and divide columns “Part Area” by “Total Area”.
11. Add new Column labeled “MHI * (Part/total)” and multiple “Part/Total” by the “MHI w/MOE”.
12. Insert the whole tab into a new pivot table labeled “Pivot 2”.
    a. Organize the pivot table by selecting “PWSID” for rows and both “Sum of MHI * Part/Total” and “Sum of Part/Total” for values.
13. Divide “Sum of MHI * Part/Total” by “Sum of Part/Total” for each PWSID
14. This will give you the weighted average for each system.
    a. Use the “Federal Classification” to filter for systems of interest, for the Risk Assessment it includes all community water systems and NTNC schools, for the Affordability Assessment it only includes community water systems.
15. The list of Determined MHI’s from this calculation must go through one last round of editing.
16. Due to the superior accuracy of the MHI results produced from the Division of Financial Assistance’s (DFA) income surveys, any system with data from that list will have its determined MHI replace with DFA MHI.
17. The result will be a list of water systems with their determined MHI from this methodology mixed with systems whose MHI comes from DFA where available.