Fair Share Dataset Logic

Notes: Keep in mind that at some point we will need to work in logic to either suppress VCS groups and neighborhoods with small sales numbers or do some grouping, based on feedback from our partners.

# Summary of Changes to Make (2/18/25)

* Remove race data from this aggregated table and create a separate race\_data table with the race variables pivoted. It should contain three variables:
  + Name
  + Race (Asian, Black, Hispanic, White)
  + Race\_percentage (the values from percent\_asian, percent\_black, etc)
* Union aggregated VCS and Neighborhood tables called financial\_data. “VCS” and “Neighborhood” variables are combined into a single “name” variable and each record is labeled as ‘vcs’ or ‘neighborhood’ in a “geog” variable.
* Add (probably 12) Durham County rows to financial\_data. Name = ’Durham’ and Geog = ‘county’. These rows should contain values for all variables – we will just treat all of Durham as a giant neighborhood.
* Add a second set of d\_over\_under and d\_count variables using the Durham median sales ratio to calculate d\_abs\_sales\_ratio.
* Create output files throughout the process for data validation.

# Filter out all but the most recent sales for properties with more than one sale

Connect to **qualified-residential-sales-info-2021-2024**, find the maximum sales date for each property and filter out any older sales.

# Create main dataset

Connect to **2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods** and the filtered table from above. Do a left join on REID (retain all records in **2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods**) and create a table with the following variables:

|  |  |  |
| --- | --- | --- |
| **Source Name** | **Variable (our name)** | **Source** |
| REID | REID | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| PIN | PIN | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| PIN\_EXT | PIN\_EXT | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| LOCATION\_ADDR | ADDRESS | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| LAND\_CLASS | PROPERTY\_TYPE | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| TOTAL\_PROP\_VALUE | ASSESSED\_VALUE | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| VCS | VCS | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| name\_2 | NEIGHBORHOOD | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| GID | NID | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| race\_asian\_per | PERCENT\_ASIAN | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| race\_black\_per | PERCENT\_BLACK | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| race\_hispanic\_per | PERCENT\_HISPANIC | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| race\_white\_per | PERCENT\_WHITE | 2024\_10\_All\_Durham\_BlockGroups\_with\_race\_neighborhoods |
| PRICE | PRICE | qualified-residential-sales-info-2021-2024 |
| SALE\_DATE | SALE\_DATE | qualified-residential-sales-info-2021-2024 |

# Create aggregated race dataset

* Find the mean of each race/ethnicity category by neighborhood
* Unpivot the data to create a table with three columns
  + Name (will contain values from NEIGHBORHOOD)
  + Race (Asian, Black, Hispanic, White)
  + Race\_percentage (will contain the percentage for each neighborhood and race)
* Find the mean of each race/ethnicity category by VCS
* Unpivot the data to create a table with three columns
  + Name (will contain values from VCS)
  + Race (Asian, Black, Hispanic, White)
  + Race\_percentage (will contain the percentage for each VCS group and race)
* Union this table with the table from the previous step.
* Find the mean of each race/ethnicity category for all of Durham County
* Unpivot the data to create a row with three columns
  + Name (‘Durham County’)
  + Race (Asian, Black, Hispanic, White)
  + Race\_percentage (will contain the percentage for each race)
* Add this row to the table from the previous steps and output as **race\_data**

# Calculate percent change for assessed value 2021-2024

* Connect to 2024 tax assessment file (available after March 1st) and join to main dataset on REID. Add 2024\_ASSESSED\_VALUE to dataset.
* Calculate the percent change between 2021 and 2024 assessed values.

# Calculate Adjusted Sales Price (ADJ\_PRICE) & Adjusted Sales Ratio (ADJ\_SALES\_RATIO)

* Use SALE\_DATE to find the sale year for each REID
* Find the median sales price for all of Durham for each year (2021, 2022, 2023, 2024)
* Calculate the difference between each year’s median sales price and come up with an adjustment factor for each year. This is how I did this in Tableau (MAX\_YEAR is 2024 in this case):

1 + ( ([MAX\_YEAR\_MED\_PRICE] - [MED\_PRICE\_BY\_YEAR]) / [MED\_PRICE\_BY\_YEAR])

These are the adjustment factors I got for each year – you should get something very similar:

* + 2021: 1.2292
  + 2022: 1.0172
  + 2023: 0.9970 (prices were higher in 2023 than in 2024)
  + 2024: 1.0000
* Multiply every sales price by the adjustment factor for the year in which it sold to create ADJ\_PRICE. This will adjust all sales to 2024 prices. Adjusted sales will go up for 2021 and 2022, go down for 2023, and stay the same for 2024.
* Find the ADJ\_SALES\_RATIO by dividing the assessed amount (ASSESSED\_VALUE) by ADJ\_PRICE

# Calculate Absolute Sales Ratio, categorize assessment, and create Sales Quartile

* Find the Median Adjusted Sales Ratio for each neighborhood (N\_MED\_ADJ\_SALES\_RATIO)
* Calculate N\_ABS\_ADJ\_SALES\_RATIO: ADJ\_SALES\_RATIO - N\_MED\_ADJ\_SALES\_RATIO
* Create N\_OVER\_UNDER:
  + If N\_ABS\_ADJ\_SALES\_RATIO is within 0.05 of N\_MED\_ADJ\_SALES\_RATIO, home is Accurate

Else if N\_ABS\_ADJ\_SALES\_RATIO is negative, home is Underassessed

* + Else if N\_ABS\_ADJ\_SALES\_RATIO is positive, home is Overassessed
* Create N\_SALES\_QUARTILE: Assign each REID to Q1, Q2, Q3, or Q4, based on ADJ\_PRICE and grouped by neighborhood
* Repeat all these steps by VCS group, using the VCS\_ prefix for all calculated variables
* Repeat all these steps for Durham County, using the D\_ prefix for all calculated variables

# Calculate D Absolute Sales Ratio, categorize assessment

* Calculate D\_N\_ABS\_ADJ\_SALES\_RATIO: ADJ\_SALES\_RATIO - D\_MED\_ADJ\_SALES\_RATIO (should have been calculated in the previous step)
* Create D\_N\_OVER\_UNDER:
  + If D\_N\_ABS\_ADJ\_SALES\_RATIO is within 0.05 of D\_MED\_ADJ\_SALES\_RATIO, home is Accurate

Else if D\_N\_ABS\_ADJ\_SALES\_RATIO is negative, home is Underassessed

* + Else if D\_N\_ABS\_ADJ\_SALES\_RATIO is positive, home is Overassessed

# Calculate COD and PRD by neighborhood, VCS group, and Durham County

**COD** = [(Σ|Sales Ratio - Median Sales Ratio| / n) / Median Sales Ratio] × 100%

**PRD** = (Average Sales Ratio / Weighted Average Sales Ratio) × 100%

**Weighted Average Sales Ratio** = Σ(Assessed Value × Sales Ratio )/ Σ (Assessed Value) × 100%

# Calculate Median Sales Ratio for dashboard display

This is the non-adjusted median sales ratio that will be displayed for each neighborhood/VCS. Right now, the plan is to not adjust the sales price for this one metric, but that may change after we meet with Hudson and L’Tanya.

* Calculate the sales ratio for each REID
* Find the mean for Durham County, each neighborhood and VCS group

# Create aggregated financial dataset

* Find the Median Adjusted Sales Price for each neighborhood (N\_MED\_ADJ\_SALES\_PRICE)
* Find the average change in assessment value for all of Durham (D\_ASSESSMENT\_CHANGE)
* Find the average change in assessment value for each neighborhood (N\_ASSESSMENT\_CHANGE)
* Find COUNTD(REID), aggregating on NEIGHBORHOOD, N\_SALES\_QUARTILE & N\_OVER\_UNDER
* Find D\_COUNTD(REID), aggregating on NEIGHBORHOOD, N\_SALES\_QUARTILE & D\_N\_OVER\_UNDER
* Create a table with the following variables:
  + NAME (will contain values from NEIGHBORHOOD)
  + GEOG (value for all rows is ‘neighborhood’)
  + NID
  + ASSESSMENT\_CHANGE
  + ADJ\_SALES\_QUARTILE
  + OVER\_UNDER
  + D\_OVER\_UNDER
  + COUNT (COUNTD of REID from step above)
  + D\_COUNT (D\_COUNTD of REID from step above)
  + MED\_ADJ\_SALES\_PROCE
  + MED\_SALES\_RATIO (from previous step)
  + COD
  + PRD
* Repeat these steps for VCS groups, create a table with the following variables, and union it with the table from the previous step
  + NAME (will contain values from VCS)
  + GEOG (value for all rows is ‘vcs’)
  + ASSESSMENT\_CHANGE
  + ADJ\_SALES\_QUARTILE
  + OVER\_UNDER
  + D\_OVER\_UNDER
  + COUNT
  + MED\_ADJ\_SALES\_PRICE
  + MED\_SALES \_RATIO
  + COD
  + PRD
* Repeat these steps for Durham County create a table with the following variables, union it with the table from the previous step, and output as **financial\_data**
  + NAME (‘Durham County’)
  + GEOG (value for all rows is ‘county)
  + ASSESSMENT\_CHANGE
  + ADJ\_SALES\_QUARTILE
  + OVER\_UNDER
  + D\_OVER\_UNDER
  + COUNT
  + D\_COUNT
  + MED\_ADJ\_SALES\_PRICE
  + MED\_SALES \_RATIO
  + COD
  + PRD