SQL Queries

* output.csv
  + **Purpose** is to produce the final .csv that will be used for machine learning.
  + **Joins**: incr\_MRS1\_Setup to incr\_School1\_CountbyZipAvgRank
    - Brings in metrics from School Digger; appends to treated MRIS data. See “incr” queries for more information.
* incr\_MRIS\_FinalSetup
  + **Purpose** isto “stage” our MRIS data for the output.csv query. It contains the following calculated fields:
    - “PriceDiffFactor”
      * Calculation: (Close Price – List Price) / List Price
      * Positive values mean the property sold for more than the List Price; negative values mean the property sold for less.
    - “ZipMYrLookup”
      * Calculation: Zipcode + MonthYr of Closing Date
      * This field will be used later to join to data sources like CrimeDC and Zillow, both of which provide zip code level data by month, year.
    - “PostDateTrue”
      * Calculation: Close Date – DOMP
      * MRIS does not provide the date when listing was posted; this date is needed in order to evaluate seasonality.
    - “MostRecentSaleFlag”
      * Calculation: CASE WHEN Close Date = MaxCloseDt, “Y”
      * Flags the record for a property’s most recent sale. If the given property has only been sold one time, then this record will show “Y”.
  + **Joins:** 
    - Source\_MRIS to Lookup\_CSVDateVal
      * The Source\_MRIS table (generated from our raw listings data) was ingested with dates set in the “Excel integer date format”, ie the date 11/5/2016 appears on our source table as ‘42679’. This join transitions the date integers back to a recognizable [M]/[D]/[YYYY] format, stored as a string.
    - Source\_MRIS to incr\_MRIS1\_CountSales
      * Incr\_MRIS1\_CountSales provides a “SalesHistCount” that shows, per each address in our data, how many times it appears (in other words, how many times it’s been sold in the past).
      * It also gives us “MaxCloseDt”, which is used to generate the “MostRecentSaleFlag” calc field.
* incr\_MRIS1\_CountSales
  + **Purpose** is to roll up our MRIS data at an address level. It contains the following calculated fields:
    - “SalesHistCount”
      * Calculation: Count(Address)
      * Simply a count of how many times a given address occurs in the MRIS dataset.
    - “MaxCloseDt”
      * Calculation: Max(CloseDate)
      * Passes through the most recent Close Date for a given property address.
* incr\_School1\_CountbyZipAvgRank
  + **Purpose** is to roll up our SchoolDigger data at a zip-code level so that it can be linked to MRIS. It contains the following calculated fields:
    - “SchoolCount”
      * Calculation: Count(School)
      * Simply a count of schools by zip, regardless of whether the schools are Elementary, Middle, or High.
    - “SchoolAvgRank”
      * Calculation: Avg(SchoolDiggerStarRating)
      * Averages School Digger’s proprietary rating for all schools in a given zip code.