**CURRENT PREDICTORS**

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| Sold.price: | the price in dollars that the house was sold for **will be our response variable (myresponse)- regression problem**  **Remove entries with no selling price (N/A)** |
| Zip code | the zip code (location) of the house (categorical)  **City instead** |
| Overview | Text description of inside and outside properties of the house (qualitative) |
| Type | whether each house is a single family home or condo/townhouse (categorical variable) |
| Year built | the year that the house was built (year-ie 2019) |
| Num.beds | number of beds in the house (integer) |
| Num.baths | number of bathrooms in the house (.5 is only bathroom no shower) |
| Living.area | size of living space within the house in sq. feet |
| Lot.area | size of land outside the house on the property (sq. feet); **not to be used, the majority of the houses do not include this info** |
| Masterbedroom.length | master bedroom’s length in feet, **NULL** |
| Masterbedroom.width | master bedroom’s width in feet, **NULL** |
| Kitchen.length | kitchen length in feet, **NULL** |
| Kitchen.width | kitchen width in feet, **NULL** |

**NEW VARIABLES:**

Age.of.home= how old the home is using 2019 as the base year, calculated by taking the year house was built and subtracting from 2019. (ex. Home built in 2019 equals 0)

Kitchen.area= Area of the kitchen (length x width)

Masterbedroom.area= Area of the master bedroom (length x width)

**DATA CLEANUP TASKS**

* 16 Missing Values in Sold.Price variables → Remove them from the model but keep them for later testing
* Kitchen.length x Kitchen.width → could be multiplied to create a NEW variable (**total kitchen area**)
* Masterbedroom.length x Masterbedroom.width → could be multiplied to create a NEW variable (**total master bedroom area**)
* Year.built → create variable for current age of the house by subtracting year built from 2019
* Zip Code → classify into different regions/towns within a 10 mile radius of Boston (zip codes = towns [**https://www.zipcodestogo.com/Massachusetts/**](https://www.zipcodestogo.com/Massachusetts/) **)**
  + Maybe determine how close it is from Boston?
    - Look up the towns and find how many miles away from Boston?
    - R Package - <https://cran.r-project.org/web/packages/zipcode/zipcode.pdf>
  + Determine population size of town → city, suburb, rural
  + Population growth
  + Average income in these towns
    - <https://www.bostonglobe.com/metro/2018/12/11/full-list-massachusetts-median-household-incomes-town/eZpgJkpB1uF2FVmpM4O8XO/story.html>
    - **Not all towns are included here**
  + Crime Rate
  + Education rating?
  + Distance from water
  + Number of stores/malls

Towns with multiple zip codes (based on codes from [**https://www.zipcodestogo.com/Massachusetts/**](https://www.zipcodestogo.com/Massachusetts/) **)**

|  |  |
| --- | --- |
| 2118 2127 2119 2128 2116 2125 2124 2122 2108 2111 2199 2114 2113 2210 2115 2215 2121 2109 2120 | Boston Zip Codes |
| 2451 2452 | Waltham Zip Codes |
| 2144 2145 | Somerville |
| 2169 2170 | Quincy |
| 1902 1904 1905 | Lynn |
| 2420 2421 | Lexington |

* + Towns with multiple zip codes

**DATA BACKGROUND INFO**

* Recently sold houses (February-August 2019) located within a ten-mile radius from Boston
* Realtor.com
  + The information included for house listings:
* 