STAC67 Case Study: Predictive Model For The Median Value Of Homes In Boston

Group 18 / Abbas Rai 1003957104 / Calvin Chan 999940364 / Huan Wang 1001707049/ Ruotong Zheng 1002586659/ Andrey Zhuravlev

# Abstract

Buying a home is one of the most important purchase decisions in oneâs life. There are many important factors for one to consider, especially the value of the home one is purchasing. In reality, the value of a home can be evaluated based on an in-depth analysis of the homeâs geographical location and several other environmental and socioeconomic factors. Our team has used a plethora of statistical tests and considered several influential variables associated with this important topic in order to develop a predictive model which best explains the value of homes in the city of Boston.

# Background and Significance

The debate surrounding purchasing the “ideal” home is one that is extremely prevalent within households across the globe. People work extremely hard to save up to buy a home, making it one of the most important investments in one’s lifetime. In fact, research shows that it takes the average American more than seven years to save up enough money just to deposit the down payment towards a home (Olsen, 2018). Moreover, future home buyers find themselves in an even more vulnerable position. Due to inflation and increasing price levels within the economy, there has been an increasing trend of the time needed to save up for a home in America, which has already risen by almost two years since 1988 (Olsen, 2018). The presence of these trends makes it even more important for individuals to make smart home purchases, in order to avoid facing severe economic repercussions. One of the fundamental factors which affects the value of a home is safety. A report published by the Center for American Progress concluded that âa 10% reduction in homicides would lead to a 0.83% increase in housing values the following year.â (Byloos, 2016). In addition, one must consider the presence of retail stores in the nearby area, which can actually lower the value of a house if located within close proximity (Matthews, 2006). We ultimately propose the following model to predict the median house price levels in Boston in order to provide home buyers and real estate agents with more holistic understanding of the factors which influence the value of homes and provide them with the necessary information to understand the housing market better and make better decisions overall.

# References

Olsen, S. (2018, October 22). Home Buyers Need 7.2 Years to Save Down Payments â 1.5 Years More Than in 1988. Retrieved from <https://www.zillow.com/research/how-many-years-down-payment-21734/>

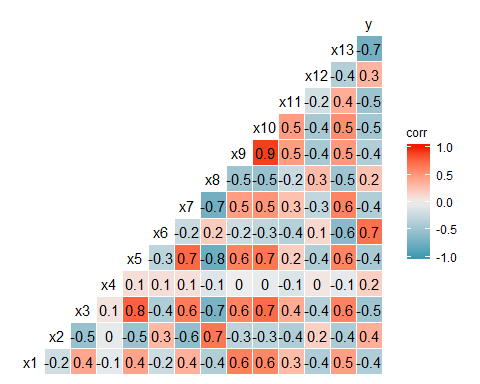
Byloos, M. (2016). Research Crime Rates and the Impact on Home Values | Homes.com. Retrieved from <https://www.homes.com/blog/2016/05/secure-new-home-research-crime-rates-impact-home-value/>

Matthews, J. (2006). Retail Proximity and Residential Values or Do Nearby Stores Really Run Down Property Values?. SSRN Electronic Journal. doi: 10.2139/ssrn.989049

## x1 x2 x3 x4 x5 x6 x7   
## 0.01501 0.00000 18.10000 0.00000 0.53800 5.71300 100.00000   
## x8 x9 x10 x11 x12 x13 y   
## 3.49520 24.00000 666.00000 20.20000 396.90000 8.05000 50.00000

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa



# Model

Half the data set is choosen to create the model and other half is for validation. Seed is set to 67 for consistent results.

# Model Creation

Stepwise Regression with function stepAIC with 12 variables (Akaike’s Information Criterion). X4 is ignored since only 6% of the data set has X4 at 1.

# Model validation

# Model diagnostics

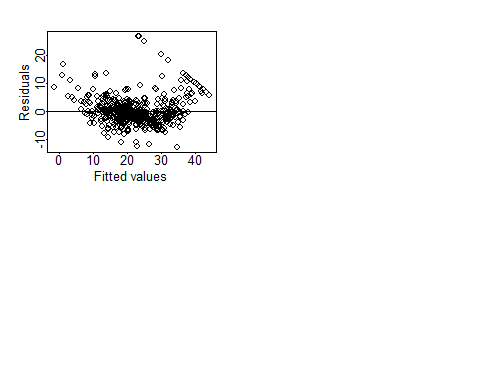
Residuals Vs. Fitted Values

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

 #Outlying Y observations We found the threshold(t\_crit) to be 3.925493. Using studentized deleted residuals we found that observations 369, 370, 372,and 373 above the threshold and therefore are the outlying Y observations.

# Outlying X observations (Leverage)

No leverage points is with guideline 2 which is Pii > 0.5. With the guideline Pii > 2\*p\_prime/n there are 46 leverage points. [9, 49, 55, 103, 142, 143, 144, 145, 146, 147, 148, 149 151, 152, 153, 154, 155, 156, 157, 160, 204, 205, 254, 258, 291, 292, 293, 352, 353, 354, 355, 356, 365, 366, 368, 369, 375, 411, 413, 415, 425, 489, 490, 491, 492, 493]

## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.05 0.02 0.03 0.02 0.01 0.01 0.01   
## 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30   
## 0.01 0.01 0.01 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01   
## 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45   
## 0.02 0.01 0.03 0.01 0.02 0.01 0.01 0.01 0.01 0.03 0.03 0.01 0.01 0.01 0.01   
## 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60   
## 0.01 0.01 0.02 0.05 0.02 0.01 0.01 0.01 0.01 0.04 0.03 0.03 0.03 0.02 0.01   
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75   
## 0.02 0.01 0.01 0.02 0.03 0.03 0.03 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01   
## 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01   
## 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.01 0.01 0.01 0.04 0.01 0.01   
## 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01   
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135   
## 0.02 0.02 0.02 0.03 0.02 0.02 0.03 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.02   
## 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150   
## 0.02 0.02 0.02 0.02 0.02 0.03 0.04 0.04 0.04 0.05 0.05 0.05 0.05 0.05 0.04   
## 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165   
## 0.04 0.05 0.05 0.04 0.04 0.06 0.07 0.02 0.02 0.05 0.02 0.02 0.03 0.03 0.02   
## 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180   
## 0.02 0.03 0.02 0.02 0.01 0.02 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01   
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195   
## 0.02 0.01 0.01 0.01 0.02 0.01 0.02 0.02 0.02 0.02 0.02 0.02 0.02 0.01 0.01   
## 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210   
## 0.03 0.03 0.03 0.03 0.04 0.04 0.03 0.03 0.05 0.05 0.01 0.00 0.01 0.01 0.02   
## 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225   
## 0.01 0.02 0.01 0.00 0.03 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.01 0.01 0.02   
## 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240   
## 0.03 0.02 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.01 0.01 0.01 0.01 0.01 0.01   
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255   
## 0.01 0.01 0.01 0.01 0.02 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.05 0.03   
## 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270   
## 0.03 0.03 0.04 0.02 0.02 0.02 0.02 0.03 0.02 0.02 0.03 0.02 0.03 0.02 0.01   
## 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.01 0.02 0.04 0.03   
## 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300   
## 0.01 0.03 0.01 0.01 0.01 0.04 0.04 0.04 0.02 0.02 0.02 0.02 0.02 0.02 0.02   
## 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315   
## 0.02 0.01 0.01 0.01 0.02 0.02 0.02 0.02 0.01 0.01 0.02 0.01 0.01 0.01 0.00   
## 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330   
## 0.01 0.01 0.01 0.00 0.00 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.02   
## 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345   
## 0.02 0.01 0.01 0.02 0.02 0.02 0.01 0.02 0.01 0.01 0.01 0.01 0.04 0.01 0.01   
## 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360   
## 0.03 0.03 0.03 0.02 0.03 0.02 0.04 0.04 0.06 0.05 0.06 0.02 0.02 0.02 0.02   
## 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375   
## 0.03 0.02 0.03 0.02 0.05 0.09 0.02 0.07 0.05 0.02 0.03 0.02 0.02 0.03 0.05   
## 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390   
## 0.02 0.02 0.02 0.02 0.01 0.02 0.02 0.01 0.01 0.02 0.02 0.02 0.03 0.02 0.01   
## 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405   
## 0.01 0.01 0.02 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.02 0.01 0.01 0.01 0.01   
## 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420   
## 0.01 0.03 0.02 0.02 0.02 0.05 0.04 0.05 0.02 0.04 0.04 0.04 0.02 0.03 0.03   
## 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435   
## 0.01 0.01 0.01 0.04 0.04 0.03 0.04 0.03 0.02 0.03 0.03 0.03 0.03 0.02 0.02   
## 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450   
## 0.02 0.03 0.03 0.04 0.01 0.01 0.02 0.01 0.02 0.01 0.03 0.01 0.01 0.01 0.01   
## 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465   
## 0.04 0.01 0.01 0.03 0.04 0.03 0.03 0.04 0.01 0.01 0.01 0.01 0.02 0.02 0.01   
## 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480   
## 0.01 0.03 0.01 0.01 0.02 0.01 0.02 0.01 0.01 0.01 0.02 0.01 0.02 0.01 0.01   
## 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495   
## 0.02 0.02 0.02 0.02 0.01 0.02 0.01 0.02 0.07 0.07 0.08 0.07 0.07 0.01 0.01   
## 496 497 498 499 500 501 502 503 504 505 506   
## 0.01 0.01 0.01 0.01 0.01 0.01 0.02 0.02 0.02 0.02 0.02

## 9 49 55 103 142 143 144 145 146 147 148 149 151 152 153 154 155 156   
## 9 49 55 103 142 143 144 145 146 147 148 149 151 152 153 154 155 156   
## 157 160 204 205 254 258 291 292 293 352 353 354 355 356 365 366 368 369   
## 157 160 204 205 254 258 291 292 293 352 353 354 355 356 365 366 368 369   
## 375 411 413 415 425 489 490 491 492 493   
## 375 411 413 415 425 489 490 491 492 493

## named integer(0)

# Influence

## Warning in plot.window(...): "id.method" is not a graphical parameter

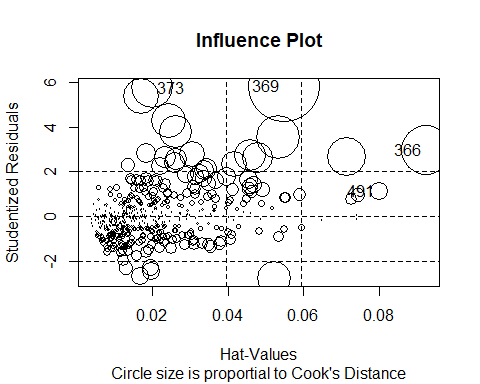
## Warning in plot.xy(xy, type, ...): "id.method" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is  
## not a graphical parameter  
  
## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is  
## not a graphical parameter

## Warning in box(...): "id.method" is not a graphical parameter

## Warning in title(...): "id.method" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.method" is not a  
## graphical parameter



## StudRes Hat CookD  
## 366 2.965550 0.09222629 0.08796623  
## 369 5.815164 0.05489718 0.18423487  
## 373 5.747572 0.02022483 0.06405406  
## 491 1.131307 0.07996003 0.01111687

There are no observations with a Cooks Distance greater than the 20th percentile of F(10,496).