Loading and Cleaning Sales Data

Kaushik Mohan

11/6/2018

Looking at NYC Annualized Real-Estate sales data from 2012 onwards. All the files downloaded from <https://www1.nyc.gov/site/finance/taxes/property-annualized-sales-update.page>

## Loading Data

## [1] "../data/annualized\_sales/2017\_queens.xls"  
## [1] "../data/annualized\_sales/2016\_brooklyn.xls"  
## [1] "../data/annualized\_sales/2016\_queens.xls"  
## [1] "../data/annualized\_sales/2015\_brooklyn.xls"  
## [1] "../data/annualized\_sales/2015\_queens.xls"  
## [1] "../data/annualized\_sales/2014\_brooklyn.xls"  
## [1] "../data/annualized\_sales/2014\_queens.xls"  
## [1] "../data/annualized\_sales/2013\_manhattan.xls"  
## [1] "../data/annualized\_sales/2013\_queens.xls"

We note 9 cases where there is an error reading in the file. Looks like it’s an issue with the size (correspondingly number of rows) and how .xls files deal with it. Saving them as .xlsx and reading them in seems to do the trick.

## Merging with MapPluto Data

MapPLUTO Data has more detailed information about each building including their exact geospatial location.

## [1] 28.5

##   
## A B C D R S   
## 0.26 0.51 0.91 0.42 99.99 1.54

##   
## A B C D R S   
## 0.05 0.09 0.12 0.07 28.12 0.04

We note that we have several cases not merged properly with the MapPluto data. Closer inspection reveals that majority of these are building class R which are Condos. We’ll deal with them separately.

We use the GeoClient API to get the locations for the Condos. Details in script 03\_geoclientAPI\_data.R.

## Merging with Property Assessment Roll Data

Lastly, we merge Property Assessment Roll archives data for more accurate information on apartment square footage. This is particularly essential as it corrects for condo unit square footage and also reduces the amount of missing data. We use the latest version of the assessment data FY18 for all the tax classes. The link to the data is below: <https://www1.nyc.gov/site/finance/taxes/property-assessment-roll-archives.page>

## 117 variables; Processing variable:1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117

## 117 variables; Processing variable:1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117

## [1] 0.29

##   
## A B C D R S   
## 0.01 0.01 0.00 0.00 1.00 0.02

##   
## A B C D R S   
## 0.00 0.00 0.00 0.00 0.28 0.00