NYC Property Sales

Mark Webster

29 January 2019

Introduction

Data Description

For my "Choose your own" project, I looked at sales data for properties in New York at this link: https://www.kaggle.com/new-york-city/nyc-property-sales

This data covered roperties sold in New York City over a 12-month period from September 2016 to September 2017.

The fields are as follows:

- BOROUGH: number from 1-4 indicating which borough the property is located at
- NEIGHBORHOOD: text - names of the neighbourhood the property is located at
- BUILDING CLASS CATEGORY: text - type of property (eg condo, single family residence, office building)
- TAX CLASS AT PRESENT: text - NYC tax category - differentiates between residential/other, and number of living spaces
- BLOCK: number - block number property is located at. Does not seem to be unique
- LOT: number - lot within the block. In most cases, the combination of block and lot is unique but not alwas (eg in the case of Condos)
- EASE-MENT: NA for all rows
- BUILDING CLASS AT PRESENT: as above, but for present day. NA in some cases
- ADDRESS: text - address. Includes apartment number at the end (in format ", 1A" for apartment 1A), but not always
- APARTMENT NUMBER: text - apartment number - not alwasy complete for apartments.
- ZIP CODE: number - zip code. Zero in some cases. Too granular for use in analysis - neighbourhood seems more useful
- RESIDENTIAL UNITS: number - how many residential units on the site. Note that this relates to the address as a whole, not the individual property sold (eg if a unit)
- COMMERCIAL UNITS: number - how many commercial units on the site
- TOTAL UNITS: number - usually equals RESIDENTIAL UNITS + COMMERCIAL UNITS
- LAND SQUARE FEET: text - how much land in the site. Note that this relates to the address as a whole, not the individual property sold (eg if a unit)
- GROSS SQUARE FEET: text - floor area of buildings on the site
- YEAR BUILT: number - year building was constructed
- TAX CLASS AT TIME OF SALE: NYC tax code - at time of sale
- BUILDING CLASS AT TIME OF SALE: NYC buliding class - more useful than "BUILDING CLASS AT PRESENT" as there are no NA values
- SALE PRICE: text - how much the property was sold for
- SALE DATE: date - when the property was sold

The data has 84548 rows. It is very difficiult to interpret as often key fields are NA or zero - for instance LAND SQUARE FEET and GROSS SQUARE FEET. It became clear that values such as these apply to the address as a whole - and not the apartment sold. For condos, the area columns are usually zero, giving us very little to go on apart from the location of the condo in terms of creating a model.

The data also has duplicate rows.

Task

The task is to create a model to predict sale price, given the data provided above.

I chose percentage error as measure when looking at the effectiveness of the model - ie the difference between the predicted and acual sale price, divided by the actual sale price. To get an overall estimate, I took the squares of the percentage errors, then the square root of the means of these errors.

Methods/Analysis

Cleaning Data

Rename Columns and Delete Irrelevant Ones

The raw data has all caps column names, with spaces in between words which makes them hard to use. There are also a few columns which have no values (such as EASE-MENT), and these were dropped. I selected the relevant cols and turn col names to lowecase, with no spaces

Convert Columns to Numeric

The key columns of 'land_area','gross_area' and 'sale_price' were in text format - I converted these to numeric and any NA values assigned a value of 0.

Some rows had 0 for the sale price - these rows were removed.

Remove Duplicate Rows

Duplicate rows - ie with the same sale price, address, sale date etc - were removed.

Incorporate Information on Building Class Codes and Segment into Datasets

Next, I found some information on building class codes from the NYC website: https://www1.nyc.gov/assets/finance/jump/hlpbldgcode.html

I converted the building class code details to a csv file and examined them. It was clear that some codes were of a very different nature to the rest of the data (eg CM MOBILE HOMES/TRAILER PARKS, Q2 PLAYGROUND). I created a "dataset" column in the csv file and classified these types of codes as "other".

Given the differences in data availability and pricing, I created the following categories in the dataset column:

- Condos - seem to be individual units, no land or gross area details
- Family Homes - houses with single or multiple families, generally have land and gross area details
- Rentals - seem to be largely whole buildings of rental apartments, generally have land and gross area details
- Commercial - industrial/warehouse/retail, generally have land and gross area details
- Offices - office buildings, generally have land and gross area details
- Vacant Land - land but no gross area details
- Other - parks, parking, and unclassified

You can see the differences in the dataset types by looking at the table below - many categories have no land or gross area information, others have 100% area information, with only a few categories which sometimes have area data and sometimes not:

Building Class	dataset	n	land_area_zero	gross_area_zero	avg_sale_price
A0 cape cod	Family Homes	285	0.0000000	0.0000000	664822.14
A1 two stories - detached sm or	Family Homes	4835	0.0000000	0.0028956	615568.59
A2 one story - permanent living	Family Homes	2007	0.0000000	0.0029895	542263.87
A3 large suburban residence	Family Homes	246	0.0000000	0.0081301	1453662.89
A4 city residence one family	Family Homes	143	0.0000000	0.0000000	4659219.66
A5 one family attached or semi-d	Family Homes	4133	0.0002420	0.0130656	539434.51
A6 summer cottage	Family Homes	78	0.0000000	0.0000000	217658.26
A7 mansion type or town house	Family Homes	8	0.0000000	0.0000000	9137187.50
A9 miscellaneous one family	Family Homes	1015	0.0000000	0.0000000	670661.97
B1 two family brick	Family Homes	2864	0.0000000	0.0034916	801036.20
B2 two family frame	Family Homes	3273	0.0000000	0.0122212	667596.63
B3 two family converted from one	Family Homes	2558	0.0000000	0.0003909	766130.26
B9 miscellaneous two family	Family Homes	993	0.0000000	0.0050352	899620.39
C0 three families	Rentals	2407	0.0000000	0.0041545	972239.89
C1 over six families without sto	Rentals	423	0.0000000	0.0023641	5044940.90
C2 five to six families	Rentals	394	0.0000000	0.0000000	1550123.87
C3 four families	Rentals	587	0.0000000	0.0000000	1268890.75
C4 old law tenement	Rentals	94	0.0000000	0.0000000	6670438.98
C5 converted dwellings or roomin	Rentals	93	0.0000000	0.0000000	3458733.91
C6 walk-up cooperative	Rentals	2505	0.9956088	0.9956088	489075.41
C7 walk-up apt. over six familie	Rentals	195	0.0000000	0.0000000	7843223.28
C8 walk-up co-op; conversion fro	Rentals	6	1.0000000	1.0000000	1910833.33
C9 garden apartments	Rentals	9	0.0000000	0.0000000	3514772.22
CM mobile homes/trailer parks	Other	1	0.0000000	0.0000000	90000.00
D0 elevator co-op; conversion fr	Rentals	207	1.0000000	1.0000000	2089478.30
D1 elevator apt; semi-fireproof	Rentals	102	0.0000000	0.0000000	15923137.96
D2 elevator apt; artists in resi	Rentals	5	0.0000000	0.0000000	13512579.20
D3 elevator apt; fireproof witho	Rentals	25	0.0000000	0.0000000	21095040.96
D4 elevator cooperative	Rentals	11325	0.9971744	0.9971744	776523.85
D5 elevator apt; converted	Rentals	9	0.0000000	0.0000000	56420726.67
D6 elevator apt; fireproof with	Rentals	14	0.0000000	0.0000000	61591999.43
D7 elevator apt; semi-fireproof	Rentals	43	0.0232558	0.0232558	27111863.86
D9 elevator apt; miscellaneous	Rentals	11	0.0000000	0.0000000	14163918.00
E1 fireproof warehouse	Commercial	105	0.0000000	0.0000000	5662530.28
E2 contractors warehouse	Commercial	14	0.0000000	0.0000000	1638635.93
E7 self-storage warehouses	Commercial	2	0.0000000	0.0000000	21300000.00
E9 miscellaneous warehouse	Commercial	47	0.0000000	0.0000000	5729701.34
F1 factory; heavy manufacturing	Commercial	16	0.0000000	0.0000000	9394750.06
F2 factory; special construction	Commercial	3	0.0000000	0.0000000	3491666.67
F4 factory; industrial semi-fire	Commercial	29	0.0000000	0.0344828	4092767.69
F5 factory; light manufacturing	Commercial	48	0.0000000	0.0208333	4712371.35
F9 factory; industrial-miscellan	Commercial	14	0.0000000	0.0000000	23038571.43
G0 garage; residential tax class	Commercial	44	0.0000000	0.7272727	422110.00
G1 all parking garages	Commercial	61	0.0000000	0.0000000	3860682.18
G2 auto body/collision or auto r	Commercial	59	0.0000000	0.0169492	2346364.02
G4 gas station with service/auto	Commercial	11	0.0000000	0.0000000	3012090.91
G5 gas station only with/without	Commercial	8	0.0000000	0.0000000	4679750.00
G6 licensed parking lot	Commercial	13	0.0000000	0.9230769	10859347.62
G7 unlicensed parking lot	Commercial	99	0.0101010	1.0000000	6309612.81
G8 car sales/rental with showroo	Commercial	3	0.0000000	0.3333333	1139284.67
G9 car wash or lubritorium facil	Commercial	22	0.0000000	0.0000000	2916772.45
G9 miscellaneous garage or gas s	Commercial	22	0.0000000	0.0000000	2916772.45

Building Class	dataset	n	land_area_zero	gross_area_zero	avg_sale_price
GU car sales/rental without show	Commercial	3	0.0000000	0.0000000	2283333.33
GW commercial garage other	Commercial	4	0.0000000	0.0000000	6790000.00
H1 luxury hotel	Commercial	3	0.0000000	0.0000000	154046666.67
H2 full service hotel	Commercial	5	0.0000000	0.0000000	135376040.00
H3 limited service; many affilia	Commercial	41	0.0000000	0.0000000	10902642.20
H4 motel	Commercial	4	0.0000000	0.2500000	3564972.75
H6 apartment hotel	Commercial	2	0.0000000	0.0000000	75350000.00
H8 dormitory	Commercial	6	0.0000000	0.0000000	75550000.00
H9 miscellaneous hotel	Commercial	1	0.0000000	0.0000000	30750000.00
HB boutique: 10-100 rooms, w/lux	Commercial	3	0.0000000	0.0000000	63749333.33
HH hostels- bed rentals in dormi	Commercial	1	0.0000000	0.0000000	42000000.00
HR sro- 1 or 2 people housed in	Commercial	2	0.0000000	0.0000000	2112500.00
HS extended stay/suite: amenitie	Commercial	1	0.0000000	0.0000000	35000000.00
I1 hospital, sanitarium, mental	Commercial	2	0.0000000	0.0000000	16762468.50
I3 dispensary	Commercial	1	0.0000000	0.0000000	4300000.00
I4 hospital; staff facility	Commercial	4	0.0000000	0.0000000	19468752.50
I5 health center, child center,	Commercial	9	0.0000000	0.0000000	5038909.33
I6 nursing home	Commercial	5	0.0000000	0.0000000	32262690.00
I7 adult care facility	Commercial	3	0.0000000	0.0000000	5713333.33
I9 miscellaneous hospital, healt	Commercial	5	0.0000000	0.0000000	6516500.00
J1 theatre; art type less than 4	Commercial	1	0.0000000	0.0000000	27000000.00
J4 legitimate theatre, sole use	Commercial	1	0.0000000	0.0000000	5000000.00
J8 multiplex picture theatre	Commercial	2	0.0000000	0.0000000	31875000.00
J9 miscellaneous theatre	Commercial	1	0.0000000	0.0000000	7211750.00
K1 one story retail building	Commercial	182	0.0000000	0.0000000	3307261.05
K2 multi-story retail building (	Commercial	91	0.0000000	0.0000000	5432030.45
K3 multi-story department store	Commercial	3	0.0000000	0.0000000	3265163.33
K4 predominant retail with other	Commercial	176	0.0000000	0.0000000	3990577.43
K5 stand-alone food establishmen	Commercial	13	0.0000000	0.0769231	3240307.69
K6 shopping center with or witho	Commercial	6	0.0000000	0.0000000	3925231.67
K7 banking facilities with or wi	Commercial	7	0.0000000	0.0000000	7750000.00
K8 big box retail: not affixed &	Commercial	1	0.0000000	0.0000000	225000.00
K9 miscellaneous store building	Commercial	10	0.0000000	0.1000000	4688077.00
L1 loft; over 8 stories (mid man	Rentals	4	0.0000000	0.0000000	21695530.25
L3 loft; semi-fireproof	Rentals	2	0.0000000	0.0000000	7625000.00
L8 loft; with retail stores othe	Rentals	9	0.0000000	0.0000000	8058239.67
L9 miscellaneous loft	Rentals	7	0.0000000	0.0000000	19407142.86
M1 church, synagogue, chapel	Commercial	40	0.0000000	0.0000000	3624380.48
M2 mission house (non-residentia	Commercial	1	0.0000000	0.0000000	86375000.00
M3 parsonage, rectory	Commercial	2	0.0000000	0.0000000	3172500.00
M4 convent	Commercial	4	0.0000000	0.0000000	6053750.00
M9 miscellaneous religious facil	Commercial	16	0.0000000	0.0000000	1898542.25
N2 home for indigent children, a	Commercial	8	0.0000000	0.0000000	18861207.62
N9 miscellaneous asylum, home	Commercial	8	0.0000000	0.0000000	4031053.62
O1 office only - 1 story	Office	21	0.0000000	0.0000000	5740399.14
O2 office only 2 - 6 stories	Office	53	0.0000000	0.0000000	8668464.11
O3 office only 7 - 19 stories	Office	4	0.0000000	0.0000000	174115450.25
O4 office only with or without c	Office	8	0.0000000	0.0000000	576937556.00
O5 office with comm - 1 to 6 sto	Office	54	0.0000000	0.0000000	20897568.31
O6 office with comm 7 - 19 stori	Office	18	0.0000000	0.0000000	82977398.28
O7 professional buildings/stand	Office	46	0.0000000	0.0000000	3216856.33
O8 office with apartments only (	Office	19	0.0000000	0.0000000	2436963.16

Building Class	dataset	n	land_area_zero	gross_area_zero	avg_sale_price
O9 miscellaneous and old style b	Office	1	0.0000000	0.0000000	5863222.00
P2 lodge room	Commercial	3	0.0000000	0.0000000	665000.00
P5 community center	Commercial	5	0.0000000	0.0000000	2045850.00
P6 amusement place, bath house,	Commercial	1	0.0000000	0.0000000	2550000.00
P8 library	Commercial	3	0.0000000	0.0000000	18850000.33
P9 miscellaneous indoor public a	Commercial	5	0.0000000	0.0000000	14558860.00
Q1 parks/recreation facilty	Other	3	0.0000000	1.0000000	17051933.33
Q8 marina, yacht club	Other	1	0.0000000	0.0000000	5000000.00
Q9 miscellaneous outdoor recreat	Other	2	0.0000000	1.0000000	395000.00
R0 special condominium billing l	Rentals	1	0.0000000	0.0000000	600000.00
R1 condo; residential unit in 2-	Condos	1041	1.0000000	1.0000000	1483606.32
R2 condo; residential unit in wa	Condos	667	1.0000000	1.0000000	639807.27
R3 condo; residential unit in 1-	Condos	1123	1.0000000	1.0000000	447738.24
R4 condo; residential unit in el	Condos	10386	1.0000000	1.0000000	2075128.46
R5 miscellaneous commercial	Commercial	14	1.0000000	1.0000000	11095766.86
R6 condo; resid.unit of 1-3 unit	Condos	150	1.0000000	1.0000000	1318610.87
R8 condo; comml.unit of 2-10 uni	Condos	38	1.0000000	1.0000000	4910371.84
R9 co-op within a condominium	Condos	1112	1.0000000	0.9991007	993925.23
RA cultural, medical, educationa	Commercial	5	1.0000000	1.0000000	19320643.40
RB office space	Office	247	1.0000000	1.0000000	4989003.07
RG indoor parking	Other	235	1.0000000	1.0000000	340298.03
RH hotel/boatel	Commercial	81	1.0000000	1.0000000	2015249.23
RK retail space	Commercial	77	1.0000000	1.0000000	7584901.04
RP outdoor parking	Other	52	1.0000000	1.0000000	67556.67
RR condo rentals	Condos	19	0.7368421	0.5789474	20362754.74
RR condominium rentals	Condos	19	0.7368421	0.5789474	20362754.74
RS non-business storage space	Condos	71	1.0000000	1.0000000	38449.06
RT terraces/gardens/cabanas	Condos	12	1.0000000	1.0000000	284467.75
RW warehouse/factory/industrial	Commercial	10	1.0000000	1.0000000	16405.00
S0 primarily 1 family with 2 sto	Family Homes	6	0.0000000	0.0000000	578472.33
S1 primarily 1 family with 1 sto	Family Homes	217	0.0000000	0.0000000	1030217.95
S2 primarily 2 family with 1 sto	Family Homes	447	0.0000000	0.0000000	1327699.06
S3 primarily 3 family with 1 sto	Family Homes	94	0.0000000	0.0000000	1839686.41
S4 primarily 4 family with 1 sto	Rentals	71	0.0000000	0.0281690	2025159.37
S5 primarily 5-6 family with 1 s	Rentals	57	0.0000000	0.0000000	3405981.35
S9 single or multiple dwelling w	Rentals	114	0.0000000	0.0000000	2646149.01
T2 pier, dock, bulkhead	Commercial	1	0.0000000	1.0000000	220000.00
V0 zoned residential; not manhat	Vacant Land	513	0.0038986	0.9766082	653288.17
V1 zoned commercial or manhattan	Vacant Land	192	0.0156250	0.9895833	6800232.26
V2 zoned commercial adjacent to	Vacant Land	1	0.0000000	1.0000000	225000.00
V3 zoned primarily residential;	Vacant Land	3	0.0000000	1.0000000	33666.67
V9 miscellaneous vacant land	Vacant Land	6	0.0000000	1.0000000	3307719.67
W1 public elementary, junior or	Commercial	4	0.0000000	0.0000000	5103916.00
W2 parochial school, yeshiva	Commercial	6	0.0000000	0.0000000	5395833.33
W3 school or academy	Commercial	4	0.0000000	0.0000000	6878750.00
W4 training school	Commercial	3	0.0000000	0.0000000	6286815.67
W8 other private school	Commercial	2	0.0000000	0.0000000	487505.00
W9 miscellaneous educational fac	Commercial	10	0.0000000	0.0000000	3730000.00
Y1 fire department	Commercial	1	0.0000000	0.0000000	1.00
Y3 prison, jail, house of detent	Commercial	1	0.0000000	0.0000000	16000000.00
Z0 tennis court, pool, shed, etc	Other	6	0.0000000	0.8333333	1090500.00
Z2 public parking area	Other	1	0.0000000	1.0000000	2.00

Building Class	dataset	n	land_area_zero	gross_area_zero	avg_sale_price
Z9 other miscellaneous	Other	56	0.0000000	0.7142857	7066476.46

Moving Records Between Datasets

Looking at the above table, we can see that for some condos, we have land and gross area - these were moved to the Rentals dataset.

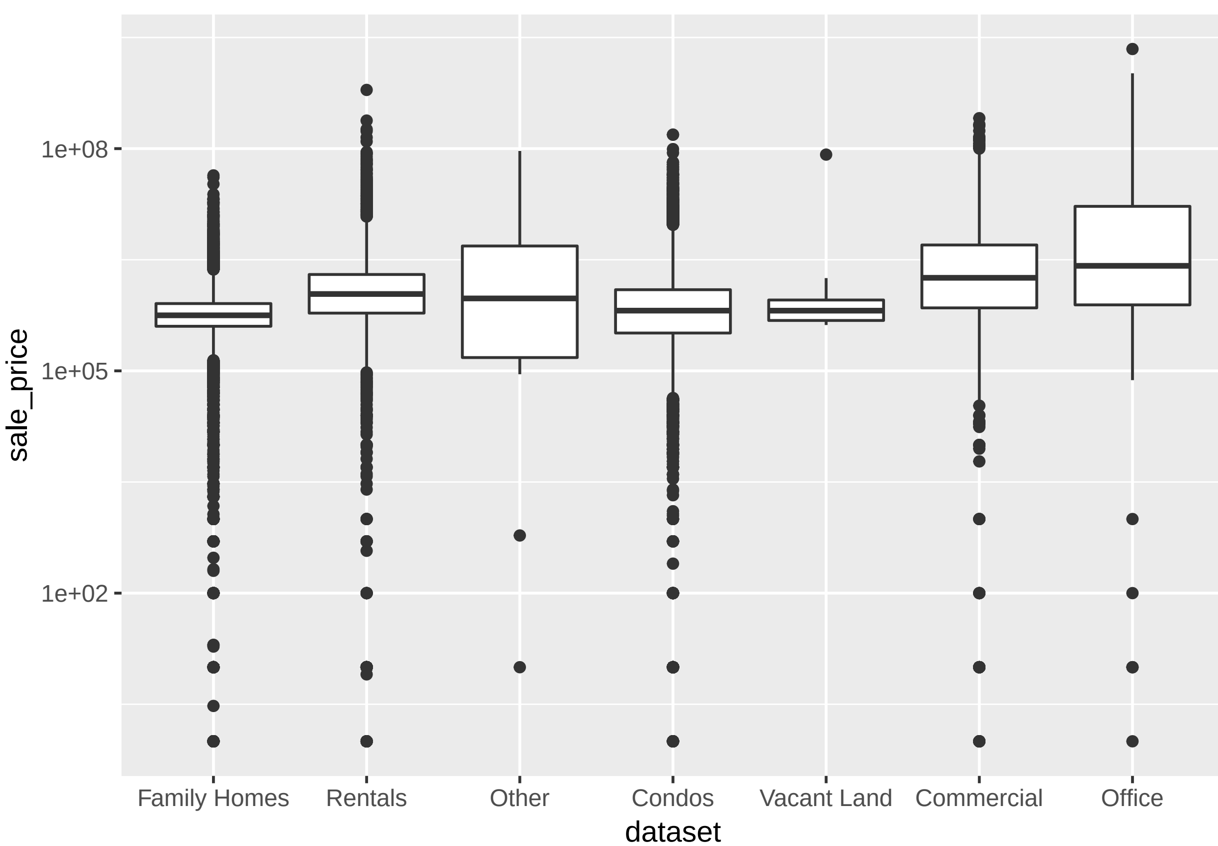
For some Rentals, we have no land or gross area - these were moved to the Condos dataset.

I also checked for situations were Commercial or Family Home records had no gross area - these would most likley be vacant lots (as there is no building onsite) so they were moved to the Vacant Land dataset.

Property Value Outliers

Examining the data, there were a lot of transactions for very low amounts - 0, USD 1, or USD 10. These are clearly transfers for no real consideration (eg gifts to family members) and hence not reflective of the commercial rate. It is hard to imagine buying any property in NYC for less than USD 50,000, so any rows with sale prices under this amount were excluded.

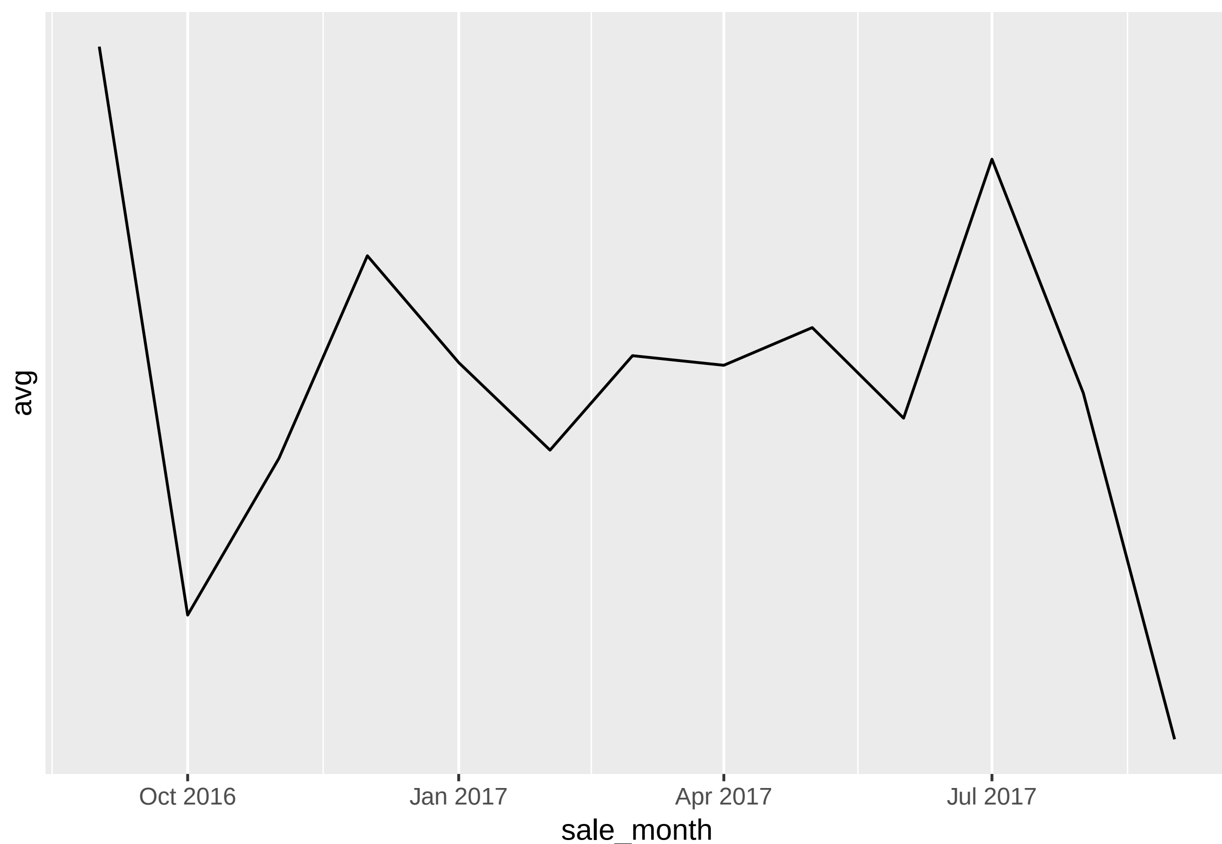
Next, the range of values is very large on the high end - with some properties valued at billions of dollars. Any property sold for over USD 10B was excluded.



Adding a Sale Month and Sale Week Column

New columns were created to reflect the sale month and sale week by rounding the sale date column. This was to see if there was a trend in property values over the year.

Looking at the average sale price per month for the Condos dataset, it looks like there is some month-to-month trend we could allow for in modelling.



Overall Trend by Dataset

Looking at the overall trend, we can see that most of the properties are Condos or Family Homes, with reasonable numbers in the Rentals and Commercial datasets. There are only small numbers in the Office, Vacant Land and Other datasets, so I did not attempt to fit these datasets.

dataset	count	avg	max	min	sd
Condos	28190	1301988.1	154250000	50661	2981096.7
Family Homes	22223	742168.1	43500000	51000	957849.7
Rentals	4474	3181292.9	620000000	52208	12541255.8
Commercial	1136	7102948.5	257500000	52500	19998557.0
Office	218	39973123.3	2210000000	75000	177777310.0
Other	17	8667608.5	93000000	90000	22256773.3
Vacant Land	14	6632445.9	83000000	415740	21983887.6

There is a large variation in average sale values and standard deviation by dataset, hence each dataset will be fitted separately.

Method

The data were segmented by dataset and models fitted for the most common categories - Condos, Family Homes, Rentals and Commercial.

I decided to try two different models in fitting the data.

Method 1 - Iterative Means

This method is a generalisation of the method used in the MovieLens project. I wanted to see if this method gave better results or ran more quickly than other regression methods.

The data has a number of different columns where the average price can be calculated for each column value. In particular these columns:

- neighbourhood
- units_res
- units_comm
- units_total
- year_built
- building_class
- bldg_category
- sale_month
- gross_area_category
- land_area_category

For example, an average sale price can be calculated for each neighbourhood in the neighbourhood column (similar to calculating an average rating for each user based on the userID column in Movielens).

Bins were used to separate gross area and land area into categories. A log scale was used to set the ranges for each bin given the large variation in areas.

One thing which was not clear to me in the MovieLens project was why we calculated the movie effect first, then calculated the user effect based on the residual. For this project, I wanted to create an algorithm to choose columns in an optimal order when calculating column effects.

The process was then as follows:

1. Calculate the overall mean sale price, and subtract this from the train_data sale price to create a residual
2. Create a list of columns we want to calculate effects for (column list)
3. Calculate the overall mean residuals by column value for each column remaining in the column list

- Y_hat is the mean residual for the column value
- Calculate the RMSE when comparing Y_hat to the actual residual value

5. Select the column x which gives the lowest RMSE
6. Save x into a ordered priority list of column effects, and the mean residuals by value of column x (analogous to a 'movies' table in MovieLens - the mean residual for each movie)
7. Delete x from the list of columns we want to calculate effects for
8. Update the residuals by subtracting the Y_hat for the column x
9. Return to item 2, and continue until we have generated a full priority list of column effects

Because we have saved the list of columns in order of importance, and the mean residuals by column, we can then apply the model which has been generated. This is analogous to the Movielens process except:

- We can potentially handle any number of columns; and
- We figure out which columns should be used first in calculating column effects

Here is some sample output for the "Family Homes" dataset:

```
## [1] "Overall Mean: RMSE =  0.262746073033956"
## [1] "neighbourhood : RMSE =  0.188802841182553"
## [1] "gross_area_category : RMSE =  0.17969555779873"
## [1] "land_area_category : RMSE =  0.176774437866295"
## [1] "year_built : RMSE =  0.17595393907815"
## [1] "sale_month : RMSE =  0.175565803751604"
## [1] "building_class : RMSE =  0.175199040209397"
## [1] "units_res : RMSE =  0.175197885110902"
## [1] "units_comm : RMSE =  0.175222609380503"
## [1] "bldg_category : RMSE =  0.175223730897435"
## [1] "units_total : RMSE =  0.175222817453603"
## [1] "Percentage Error: 0.354059757913845"

## [1] 0.3540598
```

The above printout applies the model trained on the training set, applied to the test set with RMSE effects when each column is added.

You can see that the neighbourhood, gross_area, land_area and year_built have the most significant column effects.

After the bulding_class column, there is very little benefit from adding additional column effects.

Method 2 - Binary Colums and LM

Because we have category columns (eg neighbourhood), it is difficult to perform traditional linear regression for this problem.

Accordingly, I created binary columns for key columns - in particular neighbourhood and property_class. This created one column for each value (eg for each neighbourhood), with a 1 if the property matched the value (ie was in the neighbourhood), and 0 otherwise.

I was able to use the other colums which had numerical data as is - particularly land_area, gross_area, year_built.

I then used LM to fit a model. KNN and RandomForest were also tried, but these did not work due to the size of the dataset.

Results

Below are the summarised percentage error results for both methods:

```
## [1] "Family Homes  lm fit : 0.38926879290017"
## [1] "Family Homes  iterative means fit : 0.354059757913845"
## [1] "Condos  lm fit : 25.9165384914405"
## [1] "Condos  iterative means fit : 1.33095915957527"
## [1] "Rentals  lm fit : 5.69959930254198"
## [1] "Rentals  iterative means fit : 5.28399827413907"
## [1] "Commercial  lm fit : 6.64454275323912"
## [1] "Commercial  iterative means fit : 2.98947876582526"
```

The Iterative Means method was considerably faster than Method 2 using LM.

Conclusions

The Iterative Means method seems like a useful approach - for instance:

- It allows you to quickly figure out which columns are the most significant
- Accuracy is better in many cases than using Method 2 and LM
- It runs a lot faster than Method 2