HDB\_Case\_Study

gabriel

2023-01-31

## Gabriel’s Capstone Project

In this project I will be taking a look at the pernicious effects of the HDB’s 99-year leasehold by quantifying the rate of intrinsic value deterioration, not accounting for inflation or market conditions.

### Contents

1. Data Preparation
2. Analysis
3. Key Findings
4. Conclusion

#### 1. Data Preperation

The data set used in this project can be found at <https://data.gov.sg/dataset/resale-flat-prices>  
This data is based on the date of registration for resale transactions between Jan 2012 and Jan 2023. Firstly since the data from the source splits the years up in 3 different files, we will have to append them together with power query.

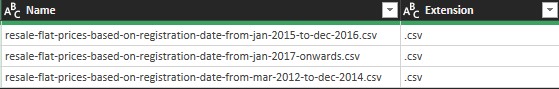


Fig.1 Appending all 3 files together

The next step is to calculate the Remaining Lease of each unit at the time of transaction. Which is given by the equation:

* Where;
  + RL = Remaining Lease
  + Y1 = Year transaction occurred
  + Y2 = Year of lease commenced

This is done with the following formula in Microsoft Excel:

=DATE(99-(YEAR([@month])-[@[lease\_commence\_date]]),MONTH([@month]),DAY([@month]))

Now that we have prepared our data, let’s get familiar with its contents before we begin analysis.

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.4.0 ✔ purrr 0.3.5   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

dataset <- read.csv("Book1.csv")  
glimpse(dataset)

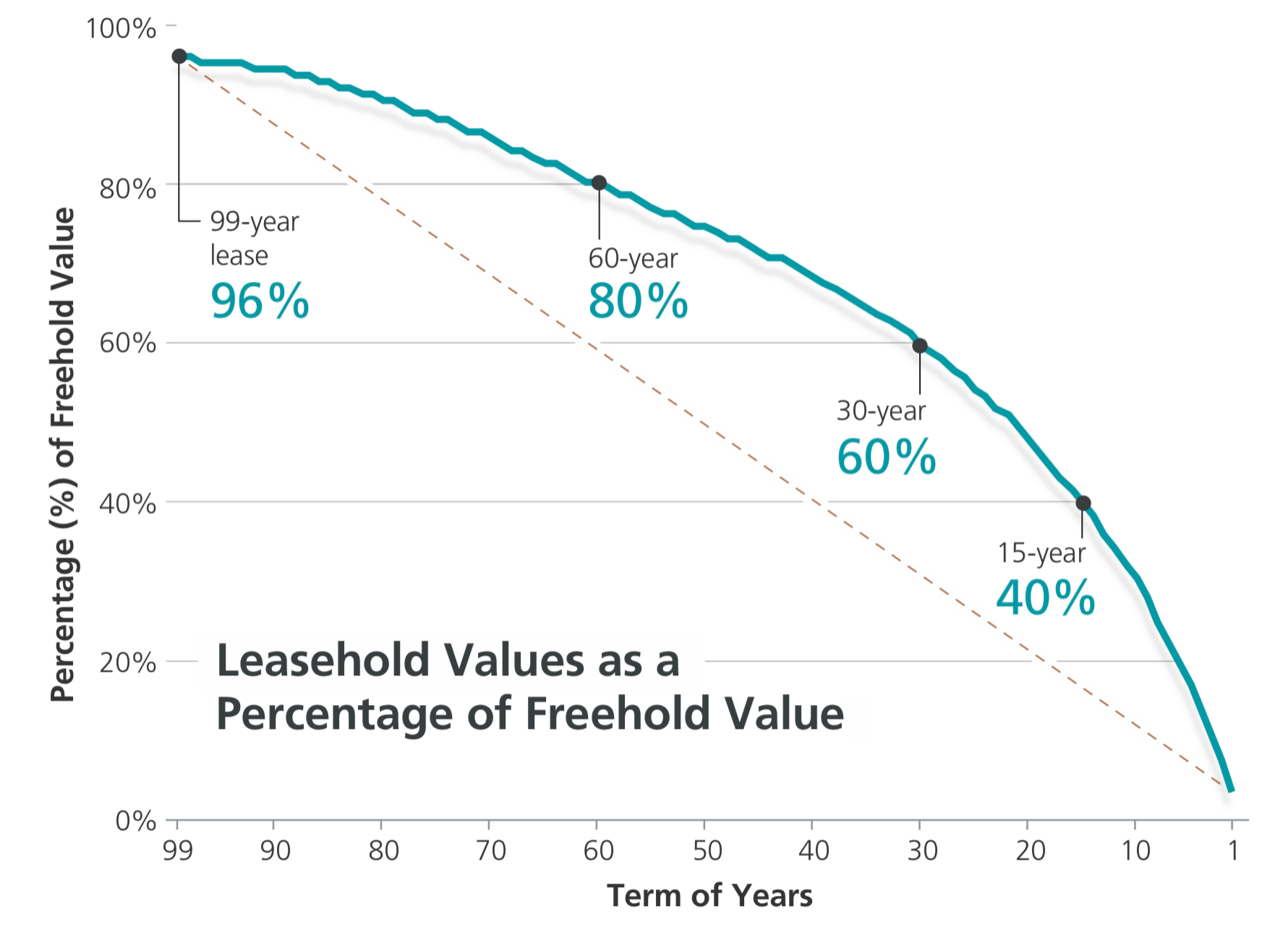
## Rows: 234,765  
## Columns: 11  
## $ month <chr> "03-2012", "03-2012", "03-2012", "03-2012", "03-20…  
## $ town <chr> "ANG MO KIO", "ANG MO KIO", "ANG MO KIO", "ANG MO …  
## $ flat\_type <chr> "2 ROOM", "2 ROOM", "3 ROOM", "3 ROOM", "3 ROOM", …  
## $ block <chr> "172", "510", "610", "474", "604", "154", "110", "…  
## $ street\_name <chr> "ANG MO KIO AVE 4", "ANG MO KIO AVE 8", "ANG MO KI…  
## $ storey\_range <chr> "06 TO 10", "01 TO 05", "06 TO 10", "01 TO 05", "0…  
## $ floor\_area\_sqm <int> 45, 44, 68, 67, 67, 68, 67, 67, 67, 67, 68, 67, 68…  
## $ flat\_model <chr> "Improved", "Improved", "New Generation", "New Gen…  
## $ lease\_commence\_date <int> 1986, 1980, 1980, 1984, 1980, 1981, 1978, 1979, 19…  
## $ remaining\_lease <int> 73, 67, 67, 71, 67, 68, 65, 66, 66, 72, 68, 67, 67…  
## $ resale\_price <int> 250000, 265000, 315000, 320000, 321000, 321000, 32…

unique(dataset$town)

## [1] "ANG MO KIO" "BEDOK" "BISHAN" "BUKIT BATOK"   
## [5] "BUKIT MERAH" "BUKIT PANJANG" "BUKIT TIMAH" "CENTRAL AREA"   
## [9] "CHOA CHU KANG" "CLEMENTI" "GEYLANG" "HOUGANG"   
## [13] "JURONG EAST" "JURONG WEST" "KALLANG/WHAMPOA" "MARINE PARADE"   
## [17] "PASIR RIS" "PUNGGOL" "QUEENSTOWN" "SEMBAWANG"   
## [21] "SENGKANG" "SERANGOON" "TAMPINES" "TOA PAYOH"   
## [25] "WOODLANDS" "YISHUN"

The output above shows the 11 columns that we’re working with and the 234,765 transactions that make up this data set. Each one of these transactions come from one of the 26 housing estates as seen above.

#### 2. Analysis

One might be tempted to assume a linear rate of deterioration in property value, trending toward zero. However that is not the case, the typical HDB’s value follows the proverbial Bala’s Curve (Fig.2). The property value is a concave downward graph that exponentially decreases it value the lower the tenure left. 

From our data set let’s find out the *average* deterioration of value by creating a linear model based on the covariance between resale price and the remaining lease. Next we import the “Broom” package to convert statistical analysis objects into tidy tibbles, we need this to save the estimated regression value into the slope variable.

price\_model = lm(resale\_price ~ remaining\_lease , data = dataset)  
summary(price\_model)

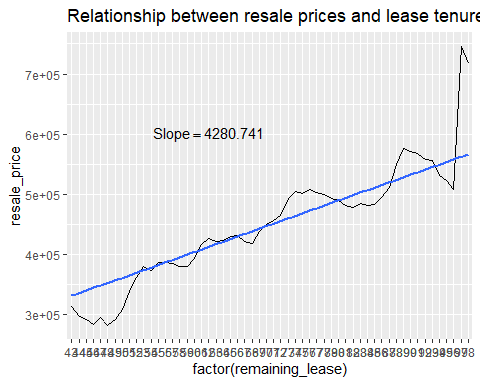
##   
## Call:  
## lm(formula = resale\_price ~ remaining\_lease, data = dataset)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -362166 -100340 -34674 66133 906748   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 147214.77 1791.11 82.19 <2e-16 \*\*\*  
## remaining\_lease 4280.74 23.61 181.28 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 143200 on 234763 degrees of freedom  
## Multiple R-squared: 0.1228, Adjusted R-squared: 0.1228   
## F-statistic: 3.286e+04 on 1 and 234763 DF, p-value: < 2.2e-16

library(broom)  
price\_model %>%   
 tidy() %>%   
 filter(term == "remaining\_lease") %>%   
 pull(estimate) -> slope

Now that we have the slope variable we can pass this into our line plot as an annotation.

dataset %>%   
 filter() %>%   
 ggplot(aes(x = factor(remaining\_lease), y = resale\_price, group = 1)) +  
 geom\_line(stat = 'summary', fun = 'mean') +  
 geom\_smooth(method = 'lm')+  
 labs(title = "Relationship between resale prices and lease tenure")+  
 annotate("text", x =20, y= 600000 ,label= (paste0("Slope ==", slope)), parse = TRUE )

## `geom\_smooth()` using formula = 'y ~ x'



## 3. Key Findingssd