MATH 242 Midterm Project

Jiaxing and Gage

3/02/2024

```
library(dplyr)
library(readr)
library(ggplot2)
```

```
nyc_condos <- read.csv("data/nyc-condos_s24.csv")
```

```
# summary of dataset
str(nyc_condos)
```

```
## 'data.frame':    200 obs. of  16 variables:
##  $ Boro.Block.Lot         : chr  "1-01613-7501" "1-01171-7501" "3-02237-7519" "4-04955-7512" ...
##  $ Condo.Section          : chr  "0267-R1" "1058-R1" "3457-R1" "0278-R1" ...
##  $ Address                : chr  "1255 5 AVENUE" "200 RIVERSIDE BOULEVARD" "135 MIDDLETON STREET" "13
##  $ Neighborhood           : chr  "UPPER EAST SIDE (96-110)" "UPPER WEST SIDE (59-79)" "WILLIAMSBURG-C
##  $ Building.Classification: chr  "R4-CONDOMINIUM" "R4  -ELEVATOR" "R4-ELEVATOR" "R2-CONDOMINIUM" ...
##  $ Total.Units            : int  59 358 14 4 198 10 60 6 10 20 ...
##  $ Year.Built             : int  1925 1997 1942 1987 1963 1983 1928 1959 2005 2004 ...
##  $ Gross.SqFt             : int  63284 512280 26964 4010 206278 10962 61084 4497 9082 22295 ...
##  $ Estimated.Gross.Income : int  1613742 29871047 579187 60391 6266726 392220 742781 92683 242853 724
##  $ Gross.Income.per.SqFt  : num  25.5 58.3 21.5 15.1 30.4 ...
##  $ Estimated.Expense      : int  726500 5665817 205466 24782 2044215 162457 417204 37100 73837 21871
##  $ Expense.per.SqFt       : num  11.48 11.06 7.62 6.18 9.91 ...
##  $ Net.Operating.Income   : int  887242 24205230 373721 35609 4222511 229763 325577 55583 169016 5058
##  $ Full.Market.Value      : int  6857996 196582995 2914000 239000 32481000 1826000 2048000 437001 135
##  $ Market.Value.per.SqFt  : num  108.4 383.7 108.1 59.6 157.5 ...
##  $ Report.Year            : int  2015 2019 2016 2012 2012 2015 2014 2018 2019 2012 ...
```

```
summary(nyc_condos)
```

```
##  Boro.Block.Lot     Condo.Section        Address          Neighborhood
##  Length:200         Length:200         Length:200         Length:200
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##  Building.Classification  Total.Units       Year.Built     Gross.SqFt
##  Length:200              Min.   :  1.00   Min.   :1874   Min.   :   945
##  Class :character        1st Qu.: 12.75   1st Qu.:1925   1st Qu.: 17628
##  Mode  :character        Median : 28.00   Median :1987   Median : 36763
```

```
##                          Mean   : 64.17   Mean   :1970   Mean   : 77475
##                          3rd Qu.: 69.25   3rd Qu.:2005   3rd Qu.: 79066
##                          Max.   :546.00   Max.   :2016   Max.   :602694
##  Estimated.Gross.Income Gross.Income.per.SqFt Estimated.Expense
##  Min.   :   45152       Min.   :12.16         Min.   :   7089
##  1st Qu.:  440887       1st Qu.:21.51         1st Qu.: 171638
##  Median :  987986       Median :31.07         Median : 353376
##  Mean   : 2881794       Mean   :32.47         Mean   : 921141
##  3rd Qu.: 2843716       3rd Qu.:41.53         3rd Qu.: 850096
##  Max.   :30999403       Max.   :64.06         Max.   :8994024
##  Expense.per.SqFt Net.Operating.Income Full.Market.Value
##  Min.   : 4.030   Min.   :   33812     Min.   :   239000
##  1st Qu.: 8.102   1st Qu.:  280463     1st Qu.:  2161999
##  Median :10.405   Median :  572572     Median :  4496000
##  Mean   :10.879   Mean   : 1960653     Mean   : 15460160
##  3rd Qu.:13.150   3rd Qu.: 1967618     3rd Qu.: 15771504
##  Max.   :22.610   Max.   :24205230     Max.   :196582995
##  Market.Value.per.SqFt  Report.Year
##  Min.   : 27.97        Min.   :2012
##  1st Qu.:101.96        1st Qu.:2014
##  Median :156.13        Median :2017
##  Mean   :169.35        Mean   :2017
##  3rd Qu.:222.99        3rd Qu.:2020
##  Max.   :383.74        Max.   :2023
```

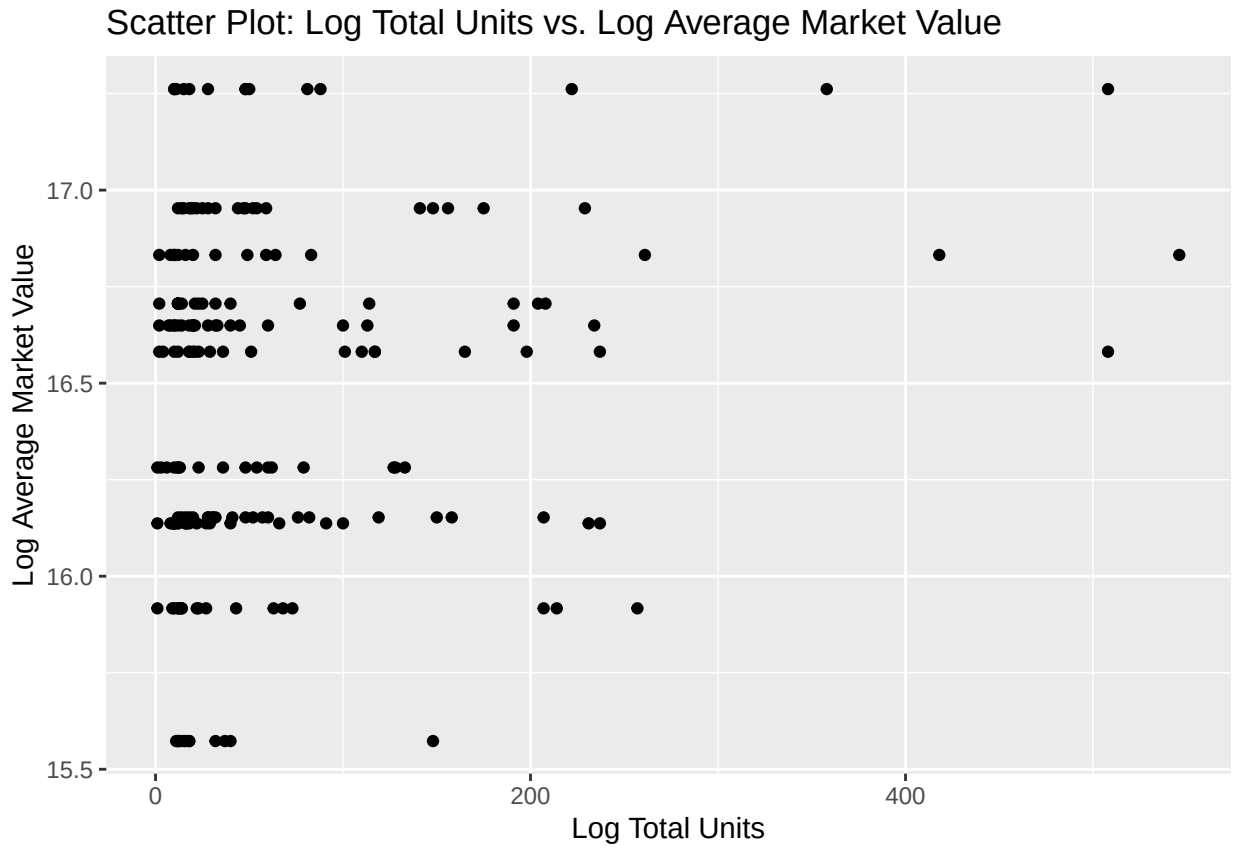
```r
# Check for missing values
colSums(is.na(nyc_condos))
```

```
##          Boro.Block.Lot           Condo.Section                 Address
##                       0                       0                       0
##            Neighborhood Building.Classification             Total.Units
##                       0                       0                       0
##              Year.Built             Gross.SqFt  Estimated.Gross.Income
##                       0                       0                       0
##   Gross.Income.per.SqFt       Estimated.Expense        Expense.per.SqFt
##                       0                       0                       0
##    Net.Operating.Income       Full.Market.Value   Market.Value.per.SqFt
##                       0                       0                       0
##             Report.Year
##                       0
```

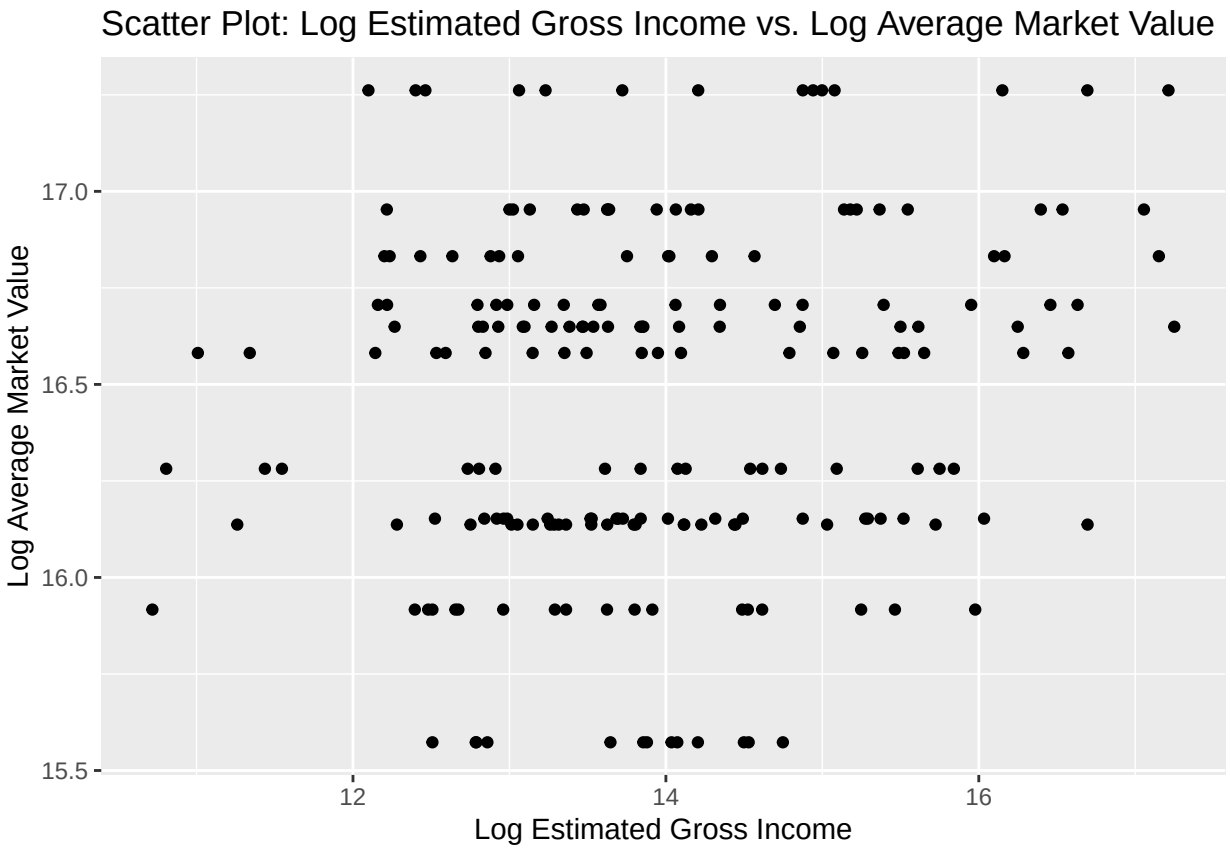
```r
# calculate average market value for each year
nyc_condos <- nyc_condos %>%
  group_by(Report.Year) %>%
  mutate(average_market_value = mean(Full.Market.Value, na.rm = TRUE))


# Log transform Gross SqFt
nyc_condos$log_GrossSqFt <- log(nyc_condos$Gross.SqFt)
# Log transform Estimated Gross Income
nyc_condos$log_EstimatedGrossIncome <- log(nyc_condos$Estimated.Gross.Income)
# Log transform Estimated Expense
nyc_condos$log_EstimatedExpense <- log(nyc_condos$Estimated.Expense)
```

```r
# Log transform Net Operating Income
nyc_condos$log_NetOperatingIncome <- log(nyc_condos$Net.Operating.Income)
# Log transform Full Market Value
nyc_condos$log_FullMarketValue <- log(nyc_condos$Full.Market.Value)

nyc_condos$log_average_market_value <- log(nyc_condos$average_market_value)
```

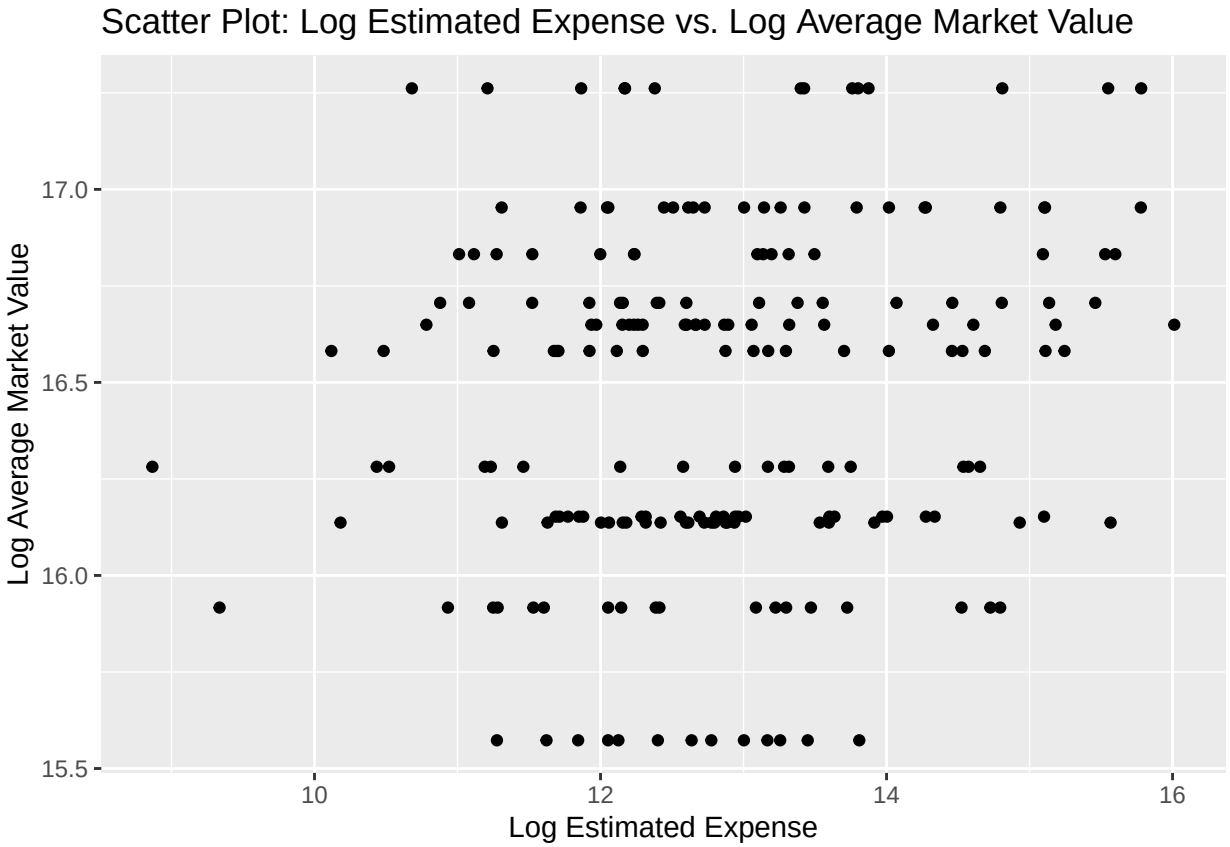
```r
# Load required library
library(ggplot2)

# Scatter plot: Total Units vs. Average Market Value (with log transformation)
ggplot(nyc_condos, aes(x = Total.Units, y = log_average_market_value)) +
  geom_point() +
  labs(x = "Log Total Units", y = "Log Average Market Value") +
  ggtitle("Scatter Plot: Log Total Units vs. Log Average Market Value")
```



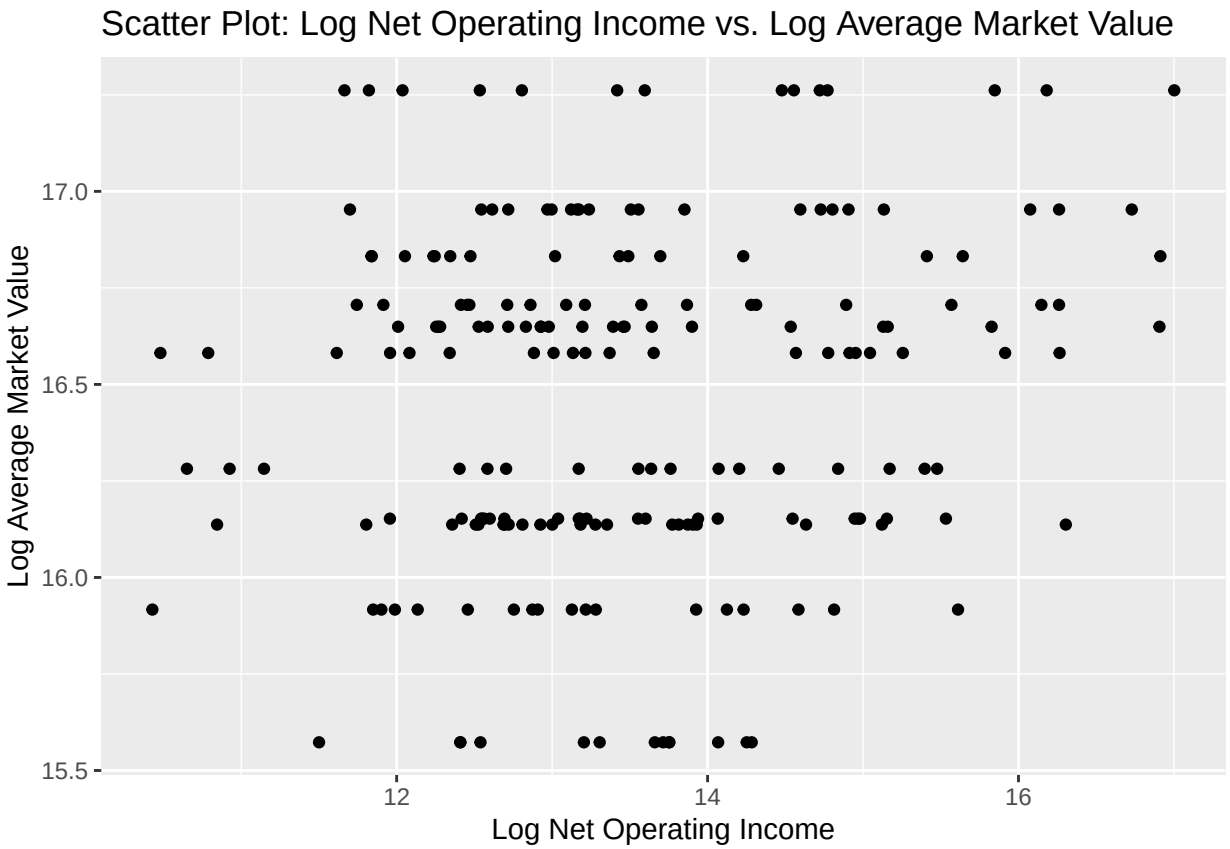

```r
# Scatter plot: log EstimatedGrossIncome vs. log Average Market Value
ggplot(nyc_condos, aes(x = log_EstimatedGrossIncome, y = log_average_market_value)) +
  geom_point() +
  labs(x = "Log Estimated Gross Income", y = "Log Average Market Value") +
  ggtitle("Scatter Plot: Log Estimated Gross Income vs. Log Average Market Value")
```



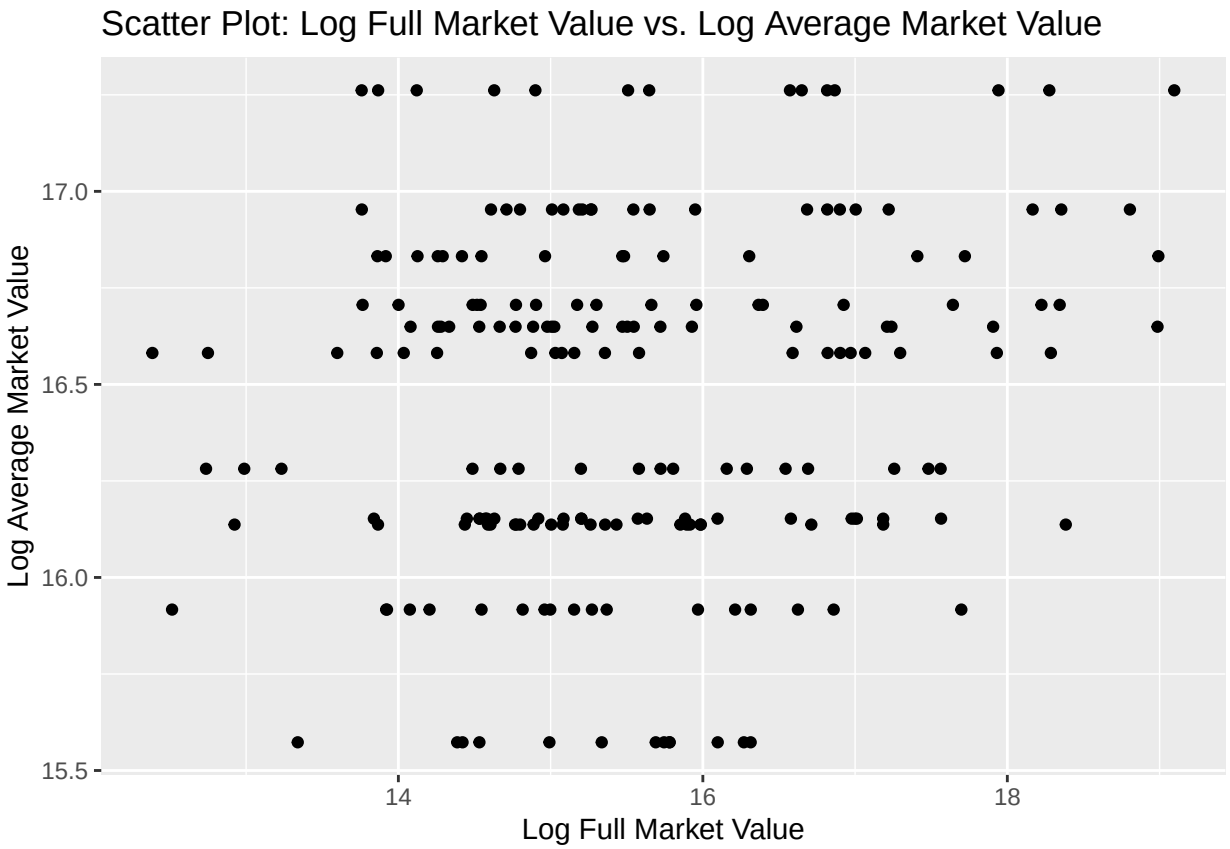

```
# Scatter plot: log EstimatedExpense vs. log Average Market Value
ggplot(nyc_condos, aes(x = log_EstimatedExpense, y = log_average_market_value)) +
  geom_point() +
  labs(x = "Log Estimated Expense", y = "Log Average Market Value") +
  ggtitle("Scatter Plot: Log Estimated Expense vs. Log Average Market Value")
```




```
# Scatter plot: log NetOperatingIncome vs. log Average Market Value
ggplot(nyc_condos, aes(x = log_NetOperatingIncome, y = log_average_market_value)) +
  geom_point() +
  labs(x = "Log Net Operating Income", y = "Log Average Market Value") +
  ggtitle("Scatter Plot: Log Net Operating Income vs. Log Average Market Value")
```




```
# Scatter plot: log FullMarketValue vs. log Average Market Value
ggplot(nyc_condos, aes(x = log_FullMarketValue, y = log_average_market_value)) +
  geom_point() +
  labs(x = "Log Full Market Value", y = "Log Average Market Value") +
  ggtitle("Scatter Plot: Log Full Market Value vs. Log Average Market Value")
```



Title:

Abstract

Introduction

Explanation of our variables:

1. **CondoSection**: Identification information for the condominium.
2. **Address**: Street address of the property.
3. **Neighborhood**: Name of the neighborhood where the property is located.
4. **BldgClassification**: Building classification code and description indicating the property's use.
5. **TotalUnits**: Total number of units in the building.

6. **YearBuilt**: Year the building was constructed.
7. **GrossSqFt**: Gross square footage of the building.
8. **EstGrossIncome**: Estimated gross income, calculated as income per square foot multiplied by gross square footage.
9. **GrossIncomePerSqFt**: Estimated gross income per square foot.
10. **EstimatedExpense**: Estimated expense, calculated as expense per square foot multiplied by gross square footage.
11. **ExpensePerSqFt**: Estimated expense per square foot.
12. **NetOperatingIncome**: Net operating income, calculated as estimated gross income minus estimated expense.
13. **FullMarketValue**: Current year's total market value of the property (land and building).
14. **MarketValuePerSqFt**: Market value per square foot, calculated as full market value divided by gross square footage.
15. **ReportYear**: Year of the report.
16. **Boro-Block-Lot**: Borough-Block-Lot location identifier for the property.

Methods

Results

Discussion

Conclusion

References