ADVANCED DATA ANALYSIS

Mrunali Vikas Patil

HOMEWORK 1 - 2024-01-28

PROBLEM 2 – MATRICES

Defining the matrices and vectors

```
Z <- matrix(c(1, 1, 1, 1,
              -1, 1, 0, 3), nrow=4, byrow=FALSE)

Y <- matrix(c(0, 8, 0, 6), byrow=TRUE)

M <- matrix(c(2, 11, 0,
              1, 3, 40,
              4, 28, 73), nrow=3, byrow=FALSE)

N <- matrix(c(-4, 7, 9,
              -3, 2, 7,
              0, 1, -8), nrow=3, byrow=FALSE)

v <-matrix(c(-3, 39, 15),byrow = TRUE)

w <- matrix(c(0, 10, 29),byrow = TRUE)
```

QUESTION A] v · w (dot product)

```
Dot_product <- sum(v * w)
print(Dot_product)

## [1] 825
```

QUESTION B] -3 * w

```
negative_three_times_w <- -3 * w
print(negative_three_times_w)

##      [,1]
## [1,]    0
## [2,]  -30
## [3,]  -87
```

QUESTION C] M * v

```
M_times_v <- M %*% v
print(M_times_v)
```

```
##      [,1]
## [1,]   93
## [2,]  504
## [3,] 2655
```

QUESTION D] M + N

```
M_plus_N <- M + N
print(M_plus_N)
```

```
##      [,1] [,2] [,3]
## [1,]   -2   -2    4
## [2,]   18    5   29
## [3,]    9   47   65
```

QUESTION E] M - N

```
M_minus_N <- M - N
print(M_minus_N)
```

```
##      [,1] [,2] [,3]
## [1,]    6    4    4
## [2,]    4    1   27
## [3,]   -9   33   81
```

QUESTION F] Z'Z

```
Z_transpose_Z <- t(Z) %*% Z
print(Z_transpose_Z)
```

```
##      [,1] [,2]
## [1,]    4    3
## [2,]    3   11
```

QUESTION G] (Z'Z)^-1

```
Z_transpose_Z_inv <- solve(Z_transpose_Z)
print(Z_transpose_Z_inv)
```

```
##             [,1]        [,2]
## [1,]  0.31428571 -0.08571429
## [2,] -0.08571429  0.11428571
```

QUESTION H] Z'Y

```
Z_transpose_Y <- t(Z) %*% Y
print(Z_transpose_Y)
```

```
##      [,1]
## [1,]   14
## [2,]   26
```

QUESTION I] β = (Z'Z)^-1 Z'Y

```r
beta <- Z_transpose_Z_inv %*% Z_transpose_Y
print(beta)
```

```
##          [,1]
## [1,] 2.171429
## [2,] 1.771429
```

QUESTION J] det(Z'Z)

```r
det_Z_transpose_Z <- det(Z_transpose_Z)
print(det_Z_transpose_Z)
```

```
## [1] 35
```

PROBLEM 3 - TYPES OD REGRESSION MODEL

LASSO REGRESSION (Least Absolute Shrinkage and Selection Operator)

Summary:

In the field of predictive modeling, standard regression methods often grapple with the risk of overfitting, especially when the set of candidate variables is large. This overfitting not only includes an excessive number of variables in the model but also tends to overestimate the model's performance, a phenomenon known as 'optimism bias'. This is particularly pronounced in predictions involving extreme risk outcomes. To mitigate these issues, penalized regression techniques like LASSO Regression are employed.

LASSO, which stands for Least Absolute Shrinkage and Selection Operator, is a technique that both shrinks coefficients and performs variable selection. It operates by imposing a constraint on the regression coefficients, essentially pulling them towards zero unless a certain threshold (denoted by lambda, λ) is met. This method is practical for model complexity control, as it automatically excludes variables whose coefficients shrink to zero. The process of selecting the optimal λ often involves k-fold cross-validation, a method where the dataset is split into equal parts for model training and validation, sequentially rotating through each subset.

While LASSO is effective and has computational advantages, it's not a cure-all for overfitting or optimism bias. It's also noted that LASSO may trade precise estimation of individual parameters for better overall prediction. Therefore, the interpretability of coefficients as independent risk factors can be compromised. The paper also touches upon related techniques like ridge regression and Elastic Net, indicating that the domain of penalized regression is rich with ongoing research.

References:

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58(1), 267–288.

Link:

https://academic.oup.com/bjs/article/105/10/1348/6122951

PROBLEM 4 - DATA ETHICS OR DATA INTEGRITY

Examining the role of virtue ethics and big data in enhancing viable, sustainable, and digital supply chain performance

Summary:

This article examines the intersection of virtue ethics, big data analytics, and their impact on enhancing sustainable and digital supply chain performance. The authors, Surajit Bag and colleagues, conducted a study utilizing the Ethical Theory of Organizing framework and Stakeholder theory to develop a theoretical model that explores the relationships between virtue ethics, big data analytics (BDA), and supply chain performance.

The study highlights the importance of virtue ethics in managing big data, particularly in the context of the fourth industrial revolution, where emerging technologies like big data analytics and AI significantly influence business operations. The authors argue that the ethical behavior of individuals handling big data is crucial, as it can lead to either positive outcomes, such as enhanced business performance and stakeholder trust, or negative consequences like poor decision-making and loss of reputation if data is mishandled.

The research utilized covariance-based structural equation modeling to test hypotheses drawn from the manufacturing industry, indicating that virtue ethics and ethical behavior of BDA personnel significantly contribute to adopting data-driven lean and green practices. These practices, in turn, enhance stakeholder trust and improve the performance of sustainable and digital supply chains.

The study's findings underscore the need for robust internal control mechanisms and ethical standards within organizations to leverage big data effectively for sustainable supply chain performance. It contributes to the literature by linking virtue ethics with big data management and providing empirical evidence on how these factors influence supply chain sustainability in the digital era.

Reference:

Bag, S., Rahman, M. S., Srivastava, G., Shore, A., & Ram, P. (2023). Examining the role of virtue ethics and big data in enhancing viable, sustainable, and digital supply chain performance. Technological Forecasting and Social Change, 186, 122154.

Link:

https://doi.org/10.1016/j.techfore.2022.122154

PROBLEM 5 - PREDICTING THE HOUSE PRICE

Load necessary libraries.

```
library(tidyverse)  # For data manipulation and visualization
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.3     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force
all conflicts to become errors
```

```
library(caret)      # For handling missing values and data preprocessing
```

```
## Loading required package: lattice
##
## Attaching package: 'caret'
##
## The following object is masked from 'package:purrr':
##
##     lift
```

```
library(car)        # For checking multicollinearity (VIF)
```

```
## 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
```

```
library(corrplot)   # For visualizing correlations
```

```
## corrplot 0.92 loaded
```

```
library(MASS)       # For stepwise regression using stepAIC
```

```
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
```

Read the dataset

```
housing_data <-
read.csv("/Users/mrunalipatil/Downloads/indian_housing_data.csv")
dim(housing_data)

## [1] 27900    91

head(housing_data)

##   exactPrice sqftPrice securityDeposit          propertyType    postedOn
## 1     240000       171              9 Multistorey Apartment Jun 20, '23
## 2      12000        12          12000 Multistorey Apartment Jun 19, '23
## 3      17000         7              9     Residential House Jun 21, '23
## 4       5000         9              9     Residential House Jun 23, '23
## 5      12000         9          24000 Multistorey Apartment Jun 24, '23
## 6      18000        16              9 Multistorey Apartment Jun 24, '23
##   noOfLifts maintenanceChargesFrequency maintenanceCharges
## 1         9                           9                 9
## 2         1                     Monthly              1500
## 3         9                           9                 9
## 4         9                           9                 9
## 5         1                     Monthly               500
## 6         9                           9                 9
##                locality     furnishing flrNum firstMonthCharges facing
## 1               Danapur Semi-Furnished      4                 9      9
## 2                     9 Semi-Furnished      4             25500      9
## 3 Phase 1 Ashiana Nagar Semi-Furnished Ground                 9      9
## 4               Kumhrar      Furnished      9                 9      9
## 5               Kumhrar    Unfurnished      1             36500   East
## 6            Lalji Tola    Unfurnished      1                 9  North
##   totalFlrNum  city carpetAreaUnit carpetArea brokerage bedrooms bathrooms
## 1           6 Patna              9          9         9        3         2
## 2           5 Patna          Sq-ft        900         9        2         2
## 3           2 Patna          Sq-ft       1300         9        3         3
## 4           3 Patna          Sq-ft        120         9        1         1
## 5           5 Patna          Sq-ft       1200         9        2         2
## 6           4 Patna          Sq-ft       1040         9        2         2
##   balconies Water_Storage Waste_Disposal Visitor_Parking Vaastu_Compliant
## 1         9             1              0               1                1
## 2         2             9              9               9                9
## 3         3             9              9               9                9
## 4         9             9              9               9                9
```

```
## 5         3             9             9              9                9
## 6         2             9             9              9                9
##
URLs
## 1           https://www.magicbricks.com/propertyDetails/3-BHK-1407-Sq-ft-
Multistorey-Apartment-FOR-Rent-Danapur-in-Patna&id=4d423636393433373437
## 2                    https://www.magicbricks.com/propertyDetails/2-BHK-
980-Sq-ft-Multistorey-Apartment-FOR-Rent-in-Patna&id=4d423637363030303937
## 3 https://www.magicbricks.com/propertyDetails/3-BHK-2500-Sq-ft-
Residential-House-FOR-Rent-Phase-1-Ashiana-Nagar-in-
Patna&id=4d423637363333393731
## 4               https://www.magicbricks.com/propertyDetails/1-BHK-120-Sq-
ft-Residential-House-FOR-Rent-Kumhrar-in-Patna&id=4d423637363638313337
## 5           https://www.magicbricks.com/propertyDetails/2-BHK-1200-Sq-ft-
Multistorey-Apartment-FOR-Rent-Kumhrar-in-Patna&id=4d423637363739323233
## 6        https://www.magicbricks.com/propertyDetails/2-BHK-1100-Sq-ft-
Multistorey-Apartment-FOR-Rent-Lalji-Tola-in-Patna&id=4d423631393339333635
##   Swimming_Pool Skydeck Service_Or_Goods_Lift Security
## 1             1       0                     0        1
## 2             9       9                     9        9
## 3             9       9                     9        9
## 4             9       9                     9        9
## 5             9       9                     9        9
## 6             9       9                     9        9
##   Retail_Boulevard___Retail_Shops__ Reserved_Parking
Rentable_Community_Space
## 1                                 0               1
0
## 2                                 9               9
9
## 3                                 9               9
9
## 4                                 9               9
9
## 5                                 9               9
9
## 6                                 9               9
9
##   RentOrSale Recreational_Pool Rain_Water_Harvesting RO_Water_System
## 1       Rent                 0                     1               0
## 2       Rent                 9                     9               9
## 3       Rent                 9                     9               9
## 4       Rent                 9                     9               9
## 5       Rent                 9                     9               9
## 6       Rent                 9                     9               9
##   Private_Terrace_Or_Garden Private_Garden Power_Back_Up Piped_Gas Park
## 1                         0              0             1         1    1
## 2                         9              9             9         9    9
## 3                         9              9             9         9    9
## 4                         9              9             9         9    9
```

```
## 5                          9              9              9         9    9
## 6                          9              9              9         9    9
##   Outdoor_Tennis_Courts Multipurpose_Hall Multipurpose_Courts
## 1                     1                 0                   0
## 2                     9                 9                   9
## 3                     9                 9                   9
## 4                     9                 9                   9
## 5                     9                 9                   9
## 6                     9                 9                   9
##   Mini_Cinema_Theatre Meditation_Area Maintenance_Staff     Long Lift
## 1                   0              0                 1 85.05633    0
## 2                   9              9                 9  9.00000    9
## 3                   9              9                 9 85.07996    9
## 4                   9              9                 9 85.18501    9
## 5                   9              9                 9 85.18501    9
## 6                   9              9                 9 85.14404    9
##   Library_And_Business_Centre Library Laundry_Service      Lat
## 1                           0       0               0 25.60590
## 2                           9       9               9  9.00000
## 3                           9       9               9 25.62143
## 4                           9       9               9 25.59309
## 5                           9       9               9 25.59309
## 6                           9       9               9 25.60508
##   Kids_Play_Pool_With_Water_Slides Kids_Play_Area Kids_Club
## 1                                0              0         0
## 2                                9              9         9
## 3                                9              9         9
## 4                                9              9         9
## 5                                9              9         9
## 6                                9              9         9
##   Jogging_and_Strolling_Track Internet_Or_Wi_Fi_Connectivity
Intercom_Facility
## 1                           1                               1
1
## 2                           9                               9
9
## 3                           9                               9
9
## 4                           9                               9
9
## 5                           9                               9
9
## 6                           9                               9
9
##   Indoor_Squash__And__Badminton_Courts Indoor_Games_Room
## 1                                    0                 0
## 2                                    9                 9
## 3                                    9                 9
## 4                                    9                 9
## 5                                    9                 9
```

```
## 6                                      9                 9
##   Health_club_with_Steam__Or__Jaccuzi Gymnasium Guest_Accommodation
## 1                                   0         1                    0
## 2                                   9         9                    9
## 3                                   9         9                    9
## 4                                   9         9                    9
## 5                                   9         9                    9
## 6                                   9         9                    9
##   Grand_Entrance_lobby Golf_Course Flower_Gardens Fire_Fighting_Equipment
## 1                    0           0              0                       0
## 2                    9           9              9                       9
## 3                    9           9              9                       9
## 4                    9           9              9                       9
## 5                    9           9              9                       9
## 6                    9           9              9                       9
##   Event_Space__And__Amphitheatre Earth_quake_resistant
Early_Learning_Centre
## 1                              0                    0
0
## 2                              9                    9
9
## 3                              9                    9
9
## 4                              9                    9
9
## 5                              9                    9
9
## 6                              9                    9
9
##   Dance_Studio DTH_Television_Facility Cycling__And__Jogging_Track
## 1            0                       1                          0
## 2            9                       9                          9
## 3            9                       9                          9
## 4            9                       9                          9
## 5            9                       9                          9
## 6            9                       9                          9
##   Cricket_net_practice Conference_Room Concierge_Services
## 1                    0               1                  0
## 2                    9               9                  9
## 3                    9               9                  9
## 4                    9               9                  9
## 5                    9               9                  9
## 6                    9               9                  9
##   Coffee_Lounge__And__Restaurants Club_House Canopy_Walk
## 1                               0          1           0
## 2                               9          9           9
## 3                               9          9           9
## 4                               9          9           9
## 5                               9          9           9
## 6                               9          9           9
```

```
##   Cafeteria_Or_Food_Court CCTV_Camera Barbeque_Pit Bar_Or_Lounge
Banquet_Hall
## 1                       1           0            0             1
1
## 2                       9           9            9             9
9
## 3                       9           9            9             9
9
## 4                       9           9            9             9
9
## 5                       9           9            9             9
9
## 6                       9           9            9             9
9
##   Bank__And__ATM Arts__And__Craft_Studio Air_Conditioned Activity_Deck4
## 1              0                       0               0              0
## 2              9                       9               9              9
## 3              9                       9               9              9
## 4              9                       9               9              9
## 5              9                       9               9              9
## 6              9                       9               9              9
##   AEROBICS_ROOM
## 1             0
## 2             9
## 3             9
## 4             9
## 5             9
## 6             9
```

```
str(housing_data)
```

```
## 'data.frame':    27900 obs. of  91 variables:
##  $ exactPrice                          : num  240000 12000 17000 5000
12000 18000 8500 10000 11000 7000 ...
##  $ sqftPrice                           : int  171 12 7 9 9 16 7 8 9 12 ...
##  $ securityDeposit                     : int  9 12000 9 9 24000 9 9 20000
11000 7000 ...
##  $ propertyType                        : chr  "Multistorey Apartment"
"Multistorey Apartment" "Residential House" "Residential House" ...
##  $ postedOn                            : chr  "Jun 20, '23" "Jun 19, '23"
"Jun 21, '23" "Jun 23, '23" ...
##  $ noOfLifts                           : chr  "9" "1" "9" "9" ...
##  $ maintenanceChargesFrequency         : chr  "9" "Monthly" "9" "9" ...
##  $ maintenanceCharges                  : num  9 1500 9 9 500 9 500 2000 9
9 ...
##  $ locality                            : chr  "Danapur" "9" "Phase 1
Ashiana Nagar" "Kumhrar" ...
##  $ furnishing                          : chr  "Semi-Furnished" "Semi-
Furnished" "Semi-Furnished" "Furnished" ...
##  $ flrNum                              : chr  "4" "4" "Ground" "9" ...
```

```
##  $ firstMonthCharges                    : num  9 25500 9 9 36500 9 9000
32000 22000 14000 ...
##  $ facing                               : chr  "9" "9" "9" "9" ...
##  $ totalFlrNum                          : int  6 5 2 3 5 4 3 5 2 2 ...
##  $ city                                 : chr  "Patna" "Patna" "Patna"
"Patna" ...
##  $ carpetAreaUnit                       : chr  "9" "Sq-ft" "Sq-ft" "Sq-ft"
...
##  $ carpetArea                           : int  9 900 1300 120 1200 1040
1000 930 1000 500 ...
##  $ brokerage                            : chr  "9" "9" "9" "9" ...
##  $ bedrooms                             : int  3 2 3 1 2 2 2 2 3 2 ...
##  $ bathrooms                            : int  2 2 3 1 2 2 1 2 1 1 ...
##  $ balconies                            : int  9 2 3 9 3 2 9 3 9 1 ...
##  $ Water_Storage                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Waste_Disposal                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Visitor_Parking                      : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Vaastu_Compliant                     : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ URLs                                 : chr
"https://www.magicbricks.com/propertyDetails/3-BHK-1407-Sq-ft-Multistorey-
Apartment-FOR-Rent-Danapur-in-Patna&id"| __truncated__
"https://www.magicbricks.com/propertyDetails/2-BHK-980-Sq-ft-Multistorey-
Apartment-FOR-Rent-in-Patna&id=4d423637363030303937"
"https://www.magicbricks.com/propertyDetails/3-BHK-2500-Sq-ft-Residential-
House-FOR-Rent-Phase-1-Ashiana-Nagar-i"| __truncated__
"https://www.magicbricks.com/propertyDetails/1-BHK-120-Sq-ft-Residential-
House-FOR-Rent-Kumhrar-in-Patna&id=4d42"| __truncated__ ...
##  $ Swimming_Pool                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Skydeck                              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Service_Or_Goods_Lift                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Security                             : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Retail_Boulevard___Retail_Shops__    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Reserved_Parking                     : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Rentable_Community_Space             : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ RentOrSale                           : chr  "Rent" "Rent" "Rent" "Rent"
...
##  $ Recreational_Pool                    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Rain_Water_Harvesting                : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ RO_Water_System                      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Private_Terrace_Or_Garden            : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Private_Garden                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Power_Back_Up                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Piped_Gas                            : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Park                                 : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Outdoor_Tennis_Courts                : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Multipurpose_Hall                    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Multipurpose_Courts                  : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Mini_Cinema_Theatre                  : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Meditation_Area                      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Maintenance_Staff                    : int  1 9 9 9 9 9 9 9 9 9 ...
```

```
##  $ Long                                 : num  85.1 9 85.1 85.2 85.2 ...
##  $ Lift                                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Library_And_Business_Centre          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Library                              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Laundry_Service                      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Lat                                  : num  25.6 9 25.6 25.6 25.6 ...
##  $ Kids_Play_Pool_With_Water_Slides     : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Kids_Play_Area                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Kids_Club                            : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Jogging_and_Strolling_Track          : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Internet_Or_Wi_Fi_Connectivity       : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Intercom_Facility                    : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Indoor_Squash__And__Badminton_Courts: int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Indoor_Games_Room                    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Health_club_with_Steam__Or__Jaccuzi : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Gymnasium                            : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Guest_Accommodation                  : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Grand_Entrance_lobby                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Golf_Course                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Flower_Gardens                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Fire_Fighting_Equipment              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Event_Space__And__Amphitheatre       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Earth_quake_resistant                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Early_Learning_Centre                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Dance_Studio                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ DTH_Television_Facility              : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Cycling__And__Jogging_Track          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Cricket_net_practice                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Conference_Room                      : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Concierge_Services                   : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Coffee_Lounge__And__Restaurants      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Club_House                           : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Canopy_Walk                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Cafeteria_Or_Food_Court              : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ CCTV_Camera                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Barbeque_Pit                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Bar_Or_Lounge                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Banquet_Hall                         : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Bank__And__ATM                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Arts__And__Craft_Studio              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Air_Conditioned                      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Activity_Deck4                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ AEROBICS_ROOM                        : int  0 9 9 9 9 9 9 9 9 9 ...
```

Displaying summary statistics

```
summary(housing_data)
```

```
##    exactPrice          sqftPrice          securityDeposit   propertyType
##  Min.   :9.000e+00   Min.   :        0   Min.   :      1   Length:27900
```

```
##  1st Qu.:1.300e+04   1st Qu.:       11   1st Qu.:      9   Class
:character
##  Median :3.000e+04   Median :       21   Median :      9   Mode
:character
##  Mean   :5.428e+06   Mean   :    42933   Mean   :  24079
##  3rd Qu.:5.270e+06   3rd Qu.:     3864   3rd Qu.:  14000
##  Max.   :3.250e+09   Max.   :200000000   Max.   :5000000
##    postedOn           noOfLifts         maintenanceChargesFrequency
##  Length:27900       Length:27900       Length:27900
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##  maintenanceCharges    locality          furnishing           flrNum
##  Min.   :0.000e+00   Length:27900       Length:27900       Length:27900
##  1st Qu.:9.000e+00   Class :character   Class :character   Class
:character
##  Median :9.000e+00   Mode  :character   Mode  :character   Mode
:character
##  Mean   :2.902e+05
##  3rd Qu.:9.000e+00
##  Max.   :8.076e+09
##  firstMonthCharges      facing           totalFlrNum          city
##  Min.   :9.000e+00   Length:27900       Min.   :  1.000   Length:27900
##  1st Qu.:9.000e+00   Class :character   1st Qu.:  2.000   Class :character
##  Median :9.000e+00   Mode  :character   Median :  4.000   Mode  :character
##  Mean   :3.328e+05                      Mean   :  5.666
##  3rd Qu.:3.000e+04                      3rd Qu.:  7.000
##  Max.   :8.077e+09                      Max.   :200.000
##  carpetAreaUnit       carpetArea     brokerage            bedrooms
##  Length:27900       Min.   :    1   Length:27900       Min.   : 1.000
##  Class :character   1st Qu.:    9   Class :character   1st Qu.: 2.000
##  Mode  :character   Median :  125   Mode  :character   Median : 2.000
##                     Mean   :  610                      Mean   : 2.673
##                     3rd Qu.: 1050                      3rd Qu.: 3.000
##                     Max.   :13000                      Max.   :10.000
##    bathrooms        balconies      Water_Storage   Waste_Disposal
##  Min.   : 1.000   Min.   : 1.000   Min.   :0.000   Min.   :0.000
##  1st Qu.: 2.000   1st Qu.: 2.000   1st Qu.:9.000   1st Qu.:9.000
##  Median : 2.000   Median : 3.000   Median :9.000   Median :9.000
##  Mean   : 2.483   Mean   : 4.677   Mean   :7.198   Mean   :7.184
##  3rd Qu.: 3.000   3rd Qu.: 9.000   3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :10.000   Max.   :10.000   Max.   :9.000   Max.   :9.000
##  Visitor_Parking Vaastu_Compliant     URLs           Swimming_Pool
##  Min.   :0.00    Min.   :0.000    Length:27900       Min.   :0.000
##  1st Qu.:9.00    1st Qu.:9.000    Class :character   1st Qu.:9.000
##  Median :9.00    Median :9.000    Mode  :character   Median :9.000
##  Mean   :7.21    Mean   :7.191                       Mean   :7.253
##  3rd Qu.:9.00    3rd Qu.:9.000                       3rd Qu.:9.000
```

```
##  Max.   :9.00    Max.   :9.000                       Max.   :9.000
##     Skydeck      Service_Or_Goods_Lift    Security
##  Min.   :0.000   Min.   :0.000         Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000         1st Qu.:9.000
##  Median :9.000   Median :9.000         Median :9.000
##  Mean   :7.122   Mean   :7.151         Mean   :7.287
##  3rd Qu.:9.000   3rd Qu.:9.000         3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000         Max.   :9.000
##  Retail_Boulevard___Retail_Shops__ Reserved_Parking
Rentable_Community_Space
##  Min.   :0.00                      Min.   :0.000    Min.   :0.00
##  1st Qu.:9.00                      1st Qu.:9.000    1st Qu.:9.00
##  Median :9.00                      Median :9.000    Median :9.00
##  Mean   :7.13                      Mean   :7.245    Mean   :7.13
##  3rd Qu.:9.00                      3rd Qu.:9.000    3rd Qu.:9.00
##  Max.   :9.00                      Max.   :9.000    Max.   :9.00
##   RentOrSale        Recreational_Pool Rain_Water_Harvesting
RO_Water_System
##  Length:27900       Min.   :0.000     Min.   :0.000         Min.   :0.000
##  Class :character   1st Qu.:9.000     1st Qu.:9.000         1st Qu.:9.000
##  Mode  :character   Median :9.000     Median :9.000         Median :9.000
##                     Mean   :7.128     Mean   :7.225         Mean   :7.146
##                     3rd Qu.:9.000     3rd Qu.:9.000         3rd Qu.:9.000
##                     Max.   :9.000     Max.   :9.000         Max.   :9.000
##  Private_Terrace_Or_Garden Private_Garden  Power_Back_Up     Piped_Gas
##  Min.   :0.000             Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000             1st Qu.:9.000   1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000             Median :9.000   Median :9.000   Median :9.000
##  Mean   :7.151             Mean   :7.116   Mean   :7.258   Mean   :7.155
##  3rd Qu.:9.000             3rd Qu.:9.000   3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000             Max.   :9.000   Max.   :9.000   Max.   :9.000
##       Park      Outdoor_Tennis_Courts Multipurpose_Hall
Multipurpose_Courts
##  Min.   :0.00   Min.   :0.000         Min.   :0.000     Min.   :0.000
##  1st Qu.:9.00   1st Qu.:9.000         1st Qu.:9.000     1st Qu.:9.000
##  Median :9.00   Median :9.000         Median :9.000     Median :9.000
##  Mean   :7.23   Mean   :7.166         Mean   :7.147     Mean   :7.151
##  3rd Qu.:9.00   3rd Qu.:9.000         3rd Qu.:9.000     3rd Qu.:9.000
##  Max.   :9.00   Max.   :9.000         Max.   :9.000     Max.   :9.000
##  Mini_Cinema_Theatre Meditation_Area Maintenance_Staff      Long
##  Min.   :0.000       Min.   :0.00    Min.   :0.000     Min.   : 0.00
##  1st Qu.:9.000       1st Qu.:9.00    1st Qu.:9.000     1st Qu.:75.69
##  Median :9.000       Median :9.00    Median :9.000     Median :77.44
##  Mean   :7.128       Mean   :7.17    Mean   :7.183     Mean   :72.86
##  3rd Qu.:9.000       3rd Qu.:9.00    3rd Qu.:9.000     3rd Qu.:80.85
##  Max.   :9.000       Max.   :9.00    Max.   :9.000     Max.   :91.29
##       Lift       Library_And_Business_Centre    Library
Laundry_Service
##  Min.   :0.000   Min.   :0.000               Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000               1st Qu.:9.000   1st Qu.:9.000
```

```
##  Median :9.000   Median :9.000               Median :9.000   Median :9.000
##  Mean   :7.258   Mean   :7.139               Mean   :7.121   Mean   :7.145
##  3rd Qu.:9.000   3rd Qu.:9.000               3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000               Max.   :9.000   Max.   :9.000
##       Lat        Kids_Play_Pool_With_Water_Slides Kids_Play_Area
##  Min.   : 0.00   Min.   :0.000                    Min.   :0.000
##  1st Qu.:17.30   1st Qu.:9.000                    1st Qu.:9.000
##  Median :23.19   Median :9.000                    Median :9.000
##  Mean   :22.24   Mean   :7.139                    Mean   :7.224
##  3rd Qu.:26.91   3rd Qu.:9.000                    3rd Qu.:9.000
##  Max.   :85.06   Max.   :9.000                    Max.   :9.000
##    Kids_Club     Jogging_and_Strolling_Track
Internet_Or_Wi_Fi_Connectivity
##  Min.   :0.000   Min.   :0.000               Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000               1st Qu.:9.000
##  Median :9.000   Median :9.000               Median :9.000
##  Mean   :7.141   Mean   :7.186               Mean   :7.161
##  3rd Qu.:9.000   3rd Qu.:9.000               3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000               Max.   :9.000
##  Intercom_Facility Indoor_Squash__And__Badminton_Courts Indoor_Games_Room
##  Min.   :0.000     Min.   :0.000                        Min.   :0.000
##  1st Qu.:9.000     1st Qu.:9.000                        1st Qu.:9.000
##  Median :9.000     Median :9.000                        Median :9.000
##  Mean   :7.201     Mean   :7.167                        Mean   :7.205
##  3rd Qu.:9.000     3rd Qu.:9.000                        3rd Qu.:9.000
##  Max.   :9.000     Max.   :9.000                        Max.   :9.000
##  Health_club_with_Steam__Or__Jaccuzi   Gymnasium     Guest_Accommodation
##  Min.   :0.000                       Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000                       1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000                       Median :9.000   Median :9.000
##  Mean   :7.129                       Mean   :7.269   Mean   :7.128
##  3rd Qu.:9.000                       3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000                       Max.   :9.000   Max.   :9.000
##  Grand_Entrance_lobby  Golf_Course    Flower_Gardens
Fire_Fighting_Equipment
##  Min.   :0.000        Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000        1st Qu.:9.000   1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000        Median :9.000   Median :9.000   Median :9.000
##  Mean   :7.129        Mean   :7.122   Mean   :7.176   Mean   :7.217
##  3rd Qu.:9.000        3rd Qu.:9.000   3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000        Max.   :9.000   Max.   :9.000   Max.   :9.000
##  Event_Space__And__Amphitheatre Earth_quake_resistant
Early_Learning_Centre
##  Min.   :0.000                  Min.   :0.000         Min.   :0.000
##  1st Qu.:9.000                  1st Qu.:9.000         1st Qu.:9.000
##  Median :9.000                  Median :9.000         Median :9.000
##  Mean   :7.144                  Mean   :7.156         Mean   :7.126
##  3rd Qu.:9.000                  3rd Qu.:9.000         3rd Qu.:9.000
##  Max.   :9.000                  Max.   :9.000         Max.   :9.000
##   Dance_Studio   DTH_Television_Facility Cycling__And__Jogging_Track
```

```
##  Min.   :0.000   Min.   :0.000           Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000           1st Qu.:9.000
##  Median :9.000   Median :9.000           Median :9.000
##  Mean   :7.121   Mean   :7.157           Mean   :7.165
##  3rd Qu.:9.000   3rd Qu.:9.000           3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000           Max.   :9.000
##  Cricket_net_practice Conference_Room Concierge_Services
##  Min.   :0.000        Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000        1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000        Median :9.000   Median :9.000
##  Mean   :7.121        Mean   :7.144   Mean   :7.126
##  3rd Qu.:9.000        3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000        Max.   :9.000   Max.   :9.000
##  Coffee_Lounge__And__Restaurants   Club_House     Canopy_Walk
##  Min.   :0.000                   Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000                   1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000                   Median :9.000   Median :9.000
##  Mean   :7.127                   Mean   :7.243   Mean   :7.123
##  3rd Qu.:9.000                   3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000                   Max.   :9.000   Max.   :9.000
##  Cafeteria_Or_Food_Court  CCTV_Camera     Barbeque_Pit   Bar_Or_Lounge
##  Min.   :0.000           Min.   :0.000   Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000           1st Qu.:9.000   1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000           Median :9.000   Median :9.000   Median :9.000
##  Mean   :7.153           Mean   :7.155   Mean   :7.123   Mean   :7.134
##  3rd Qu.:9.000           3rd Qu.:9.000   3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000           Max.   :9.000   Max.   :9.000   Max.   :9.000
##   Banquet_Hall   Bank__And__ATM  Arts__And__Craft_Studio Air_Conditioned
##  Min.   :0.000   Min.   :0.000   Min.   :0.000           Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000   1st Qu.:9.000           1st Qu.:9.000
##  Median :9.000   Median :9.000   Median :9.000           Median :9.000
##  Mean   :7.176   Mean   :7.135   Mean   :7.122           Mean   :7.142
##  3rd Qu.:9.000   3rd Qu.:9.000   3rd Qu.:9.000           3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000   Max.   :9.000           Max.   :9.000
##  Activity_Deck4  AEROBICS_ROOM
##  Min.   :0.000   Min.   :0.000
##  1st Qu.:9.000   1st Qu.:9.000
##  Median :9.000   Median :9.000
##  Mean   :7.127   Mean   :7.147
##  3rd Qu.:9.000   3rd Qu.:9.000
##  Max.   :9.000   Max.   :9.000
```

DATA PREPROCESSING

Checking unique values

```
sapply(housing_data, function(x) length(unique(x)))
```

```
##                            exactPrice                             sqftPrice
##                                  2018                                  4916
##                         securityDeposit                          propertyType
##                                   314                                     7
##                               postedOn                           noOfLifts
##                                   174                                    11
##           maintenanceChargesFrequency                     maintenanceCharges
##                                     6                                   257
##                              locality                            furnishing
##                                  3831                                     4
##                                 flrNum                      firstMonthCharges
##                                    62                                  1947
##                                 facing                            totalFlrNum
##                                     9                                    81
##                                   city                       carpetAreaUnit
##                                   20                                    10
##                             carpetArea                             brokerage
##                                  1596                                    19
##                               bedrooms                             bathrooms
##                                   10                                    10
##                               balconies                         Water_Storage
##                                   10                                     3
##                          Waste_Disposal                       Visitor_Parking
##                                     3                                     3
##                        Vaastu_Compliant                                URLs
##                                     3                                 27870
##                           Swimming_Pool                              Skydeck
##                                     3                                     3
##                  Service_Or_Goods_Lift                             Security
##                                     3                                     3
##    Retail_Boulevard___Retail_Shops__                      Reserved_Parking
##                                     3                                     3
##               Rentable_Community_Space                            RentOrSale
##                                     3                                     3
##                       Recreational_Pool                   Rain_Water_Harvesting
##                                     3                                     3
##                         RO_Water_System               Private_Terrace_Or_Garden
##                                     3                                     3
##                          Private_Garden                         Power_Back_Up
##                                     3                                     3
##                               Piped_Gas                                  Park
##                                     3                                     3
##                  Outdoor_Tennis_Courts                      Multipurpose_Hall
##                                     3                                     3
```

```
##                   Multipurpose_Courts                     Mini_Cinema_Theatre
##                                     3                                       3
##                        Meditation_Area                       Maintenance_Staff
##                                     3                                       3
##                                  Long                                    Lift
##                                  7059                                       3
##           Library_And_Business_Centre                                 Library
##                                     3                                       3
##                        Laundry_Service                                     Lat
##                                     3                                    7099
##     Kids_Play_Pool_With_Water_Slides                       Kids_Play_Area
##                                     3                                       3
##                               Kids_Club            Jogging_and_Strolling_Track
##                                     3                                       3
##       Internet_Or_Wi_Fi_Connectivity                     Intercom_Facility
##                                     3                                       3
## Indoor_Squash__And__Badminton_Courts                     Indoor_Games_Room
##                                     3                                       3
##  Health_club_with_Steam__Or__Jaccuzi                               Gymnasium
##                                     3                                       3
##                   Guest_Accommodation                  Grand_Entrance_lobby
##                                     3                                       3
##                            Golf_Course                       Flower_Gardens
##                                     3                                       3
##               Fire_Fighting_Equipment         Event_Space__And__Amphitheatre
##                                     3                                       3
##                 Earth_quake_resistant                  Early_Learning_Centre
##                                     3                                       3
##                           Dance_Studio                DTH_Television_Facility
##                                     3                                       3
##           Cycling__And__Jogging_Track                   Cricket_net_practice
##                                     3                                       3
##                        Conference_Room                     Concierge_Services
##                                     3                                       3
##      Coffee_Lounge__And__Restaurants                             Club_House
##                                     3                                       3
##                            Canopy_Walk                Cafeteria_Or_Food_Court
##                                     3                                       3
##                            CCTV_Camera                           Barbeque_Pit
##                                     3                                       3
##                          Bar_Or_Lounge                           Banquet_Hall
##                                     3                                       3
##                         Bank__And__ATM                Arts__And__Craft_Studio
##                                     3                                       3
##                        Air_Conditioned                           Activity_Deck4
##                                     3                                       3
##                          AEROBICS_ROOM
##                                     3
```

```
library(dplyr)
```

Remove the 'URLs' column

```
housing_data <- dplyr::select(housing_data, -URLs) # Replace 'URLs' with the exact column name
```

Remove the 'postedOn' column

```
housing_data <- dplyr::select(housing_data, -postedOn) # Replace 'postedOn' with the exact column name
```

Remove the 'locality' column

```
housing_data <- dplyr::select(housing_data, -locality) # Replace 'postedOn' with the exact column name
dim(housing_data)
```

```
## [1] 27900    88
```

Checking the structure of the data frame to confirm the columns have been removed

```
str(housing_data)
```

```
## 'data.frame':    27900 obs. of  88 variables:
##  $ exactPrice                          : num  240000 12000 17000 5000
12000 18000 8500 10000 11000 7000 ...
##  $ sqftPrice                           : int  171 12 7 9 9 16 7 8 9 12 ...
##  $ securityDeposit                     : int  9 12000 9 9 24000 9 9 20000
11000 7000 ...
##  $ propertyType                        : chr  "Multistorey Apartment"
"Multistorey Apartment" "Residential House" "Residential House" ...
##  $ noOfLifts                           : chr  "9" "1" "9" "9" ...
##  $ maintenanceChargesFrequency         : chr  "9" "Monthly" "9" "9" ...
##  $ maintenanceCharges                  : num  9 1500 9 9 500 9 500 2000 9
9 ...
##  $ furnishing                          : chr  "Semi-Furnished" "Semi-
Furnished" "Semi-Furnished" "Furnished" ...
##  $ flrNum                              : chr  "4" "4" "Ground" "9" ...
##  $ firstMonthCharges                   : num  9 25500 9 9 36500 9 9000
32000 22000 14000 ...
##  $ facing                              : chr  "9" "9" "9" "9" ...
##  $ totalFlrNum                         : int  6 5 2 3 5 4 3 5 2 2 ...
##  $ city                                : chr  "Patna" "Patna" "Patna"
"Patna" ...
##  $ carpetAreaUnit                      : chr  "9" "Sq-ft" "Sq-ft" "Sq-ft"
...
##  $ carpetArea                          : int  9 900 1300 120 1200 1040
1000 930 1000 500 ...
##  $ brokerage                           : chr  "9" "9" "9" "9" ...
```

```
##  $ bedrooms                                : int  3 2 3 1 2 2 2 2 3 2 ...
##  $ bathrooms                               : int  2 2 3 1 2 2 1 2 1 1 ...
##  $ balconies                               : int  9 2 3 9 3 2 9 3 9 1 ...
##  $ Water_Storage                           : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Waste_Disposal                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Visitor_Parking                         : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Vaastu_Compliant                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Swimming_Pool                           : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Skydeck                                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Service_Or_Goods_Lift                   : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Security                                : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Retail_Boulevard___Retail_Shops__       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Reserved_Parking                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Rentable_Community_Space                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ RentOrSale                              : chr  "Rent" "Rent" "Rent" "Rent"
...
##  $ Recreational_Pool                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Rain_Water_Harvesting                   : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ RO_Water_System                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Private_Terrace_Or_Garden               : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Private_Garden                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Power_Back_Up                           : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Piped_Gas                               : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Park                                    : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Outdoor_Tennis_Courts                   : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Multipurpose_Hall                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Multipurpose_Courts                     : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Mini_Cinema_Theatre                     : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Meditation_Area                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Maintenance_Staff                       : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Long                                    : num  85.1 9 85.1 85.2 85.2 ...
##  $ Lift                                    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Library_And_Business_Centre             : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Library                                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Laundry_Service                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Lat                                     : num  25.6 9 25.6 25.6 25.6 ...
##  $ Kids_Play_Pool_With_Water_Slides        : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Kids_Play_Area                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Kids_Club                               : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Jogging_and_Strolling_Track             : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Internet_Or_Wi_Fi_Connectivity          : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Intercom_Facility                       : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Indoor_Squash__And__Badminton_Courts: int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Indoor_Games_Room                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Health_club_with_Steam__Or__Jaccuzi : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Gymnasium                               : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Guest_Accommodation                     : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Grand_Entrance_lobby                    : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Golf_Course                             : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Flower_Gardens                          : int  0 9 9 9 9 9 9 9 9 9 ...
```

```
##  $ Fire_Fighting_Equipment              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Event_Space__And__Amphitheatre       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Earth_quake_resistant                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Early_Learning_Centre                : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Dance_Studio                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ DTH_Television_Facility              : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Cycling__And__Jogging_Track          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Cricket_net_practice                 : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Conference_Room                      : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Concierge_Services                   : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Coffee_Lounge__And__Restaurants      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Club_House                           : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Canopy_Walk                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Cafeteria_Or_Food_Court              : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ CCTV_Camera                          : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Barbeque_Pit                         : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Bar_Or_Lounge                        : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Banquet_Hall                         : int  1 9 9 9 9 9 9 9 9 9 ...
##  $ Bank__And__ATM                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Arts__And__Craft_Studio              : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Air_Conditioned                      : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ Activity_Deck4                       : int  0 9 9 9 9 9 9 9 9 9 ...
##  $ AEROBICS_ROOM                        : int  0 9 9 9 9 9 9 9 9 9 ...
```

Converting the bedrooms, bathrooms, and balconies columns to factors

```
housing_data <- housing_data %>%
  mutate(
    bedrooms = as.factor(bedrooms),
    bathrooms = as.factor(bathrooms),
    balconies = as.factor(balconies)
  )
```

Handling NA values (replace '9' with mean/mode)

```
housing_data[housing_data == 9] <- NA
```

Calculating the percentage of NA values in each column

```
na_percentage <- sapply(housing_data, function(x) sum(is.na(x)) / length(x) *
100)
print(na_percentage)
```

```
##                            exactPrice                             sqftPrice
##                            4.32258065                           13.18637993
##                         securityDeposit                           propertyType
##                            69.72043011                           0.04301075
```

```
##                           noOfLifts          maintenanceChargesFrequency
##                          82.34408602                          76.36559140
##                  maintenanceCharges                           furnishing
##                          83.82078853                           2.43369176
##                               flrNum                 firstMonthCharges
##                          24.03584229                          68.19354839
##                               facing                          totalFlrNum
##                          52.44802867                           7.99641577
##                                 city                       carpetAreaUnit
##                           0.04301075                          45.86021505
##                           carpetArea                            brokerage
##                          46.33333333                          79.26523297
##                             bedrooms                            bathrooms
##                           2.11469534                           2.61290323
##                            balconies                         Water_Storage
##                          38.83154122                          79.06810036
##                       Waste_Disposal                     Visitor_Parking
##                          79.06810036                          79.06810036
##                     Vaastu_Compliant                      Swimming_Pool
##                          79.06810036                          79.06810036
##                              Skydeck               Service_Or_Goods_Lift
##                          79.06810036                          79.06810036
##                             Security    Retail_Boulevard___Retail_Shops__
##                          79.06810036                          79.06810036
##                     Reserved_Parking            Rentable_Community_Space
##                          79.06810036                          79.06810036
##                           RentOrSale                   Recreational_Pool
##                           0.11111111                          79.06810036
##                 Rain_Water_Harvesting                     RO_Water_System
##                          79.06810036                          79.06810036
##             Private_Terrace_Or_Garden                      Private_Garden
##                          79.06810036                          79.06810036
##                        Power_Back_Up                            Piped_Gas
##                          79.06810036                          79.06810036
##                                 Park                Outdoor_Tennis_Courts
##                          79.06810036                          79.06810036
##                    Multipurpose_Hall                  Multipurpose_Courts
##                          79.06810036                          79.06810036
##                  Mini_Cinema_Theatre                     Meditation_Area
##                          79.06810036                          79.06810036
##                    Maintenance_Staff                                 Long
##                          79.06810036                           7.77419355
##                                 Lift          Library_And_Business_Centre
##                          79.06810036                          79.06810036
##                              Library                      Laundry_Service
##                          79.06810036                          79.06810036
##                                  Lat     Kids_Play_Pool_With_Water_Slides
##                           7.77419355                          79.06810036
##                       Kids_Play_Area                            Kids_Club
##                          79.06810036                          79.06810036
```

```
##           Jogging_and_Strolling_Track       Internet_Or_Wi_Fi_Connectivity
##                             79.06810036                          79.06810036
##                      Intercom_Facility Indoor_Squash__And__Badminton_Courts
##                             79.06810036                          79.06810036
##                      Indoor_Games_Room  Health_club_with_Steam__Or__Jaccuzi
##                             79.06810036                          79.06810036
##                               Gymnasium                   Guest_Accommodation
##                             79.06810036                          79.06810036
##                  Grand_Entrance_lobby                         Golf_Course
##                             79.06810036                          79.06810036
##                        Flower_Gardens              Fire_Fighting_Equipment
##                             79.06810036                          79.06810036
##       Event_Space__And__Amphitheatre                Earth_quake_resistant
##                             79.06810036                          79.06810036
##                 Early_Learning_Centre                          Dance_Studio
##                             79.06810036                          79.06810036
##               DTH_Television_Facility          Cycling__And__Jogging_Track
##                             79.06810036                          79.06810036
##                  Cricket_net_practice                      Conference_Room
##                             79.06810036                          79.06810036
##                    Concierge_Services      Coffee_Lounge__And__Restaurants
##                             79.06810036                          79.06810036
##                             Club_House                          Canopy_Walk
##                             79.06810036                          79.06810036
##               Cafeteria_Or_Food_Court                          CCTV_Camera
##                             79.06810036                          79.06810036
##                           Barbeque_Pit                       Bar_Or_Lounge
##                             79.06810036                          79.06810036
##                           Banquet_Hall                     Bank__And__ATM
##                             79.06810036                          79.06810036
##               Arts__And__Craft_Studio                    Air_Conditioned
##                             79.06810036                          79.06810036
##                         Activity_Deck4                      AEROBICS_ROOM
##                             79.06810036                          79.06810036
```

Identifying columns with more than 70% NA values

```
columns_to_remove <- names(na_percentage[na_percentage > 70])
```

Removing the identified columns from the dataframe

```
housing_data <- housing_data[, !(names(housing_data) %in% columns_to_remove)]
```

Checking the structure of the updated dataframe

```
dim(housing_data)
```

```
## [1] 27900    18
```

Function to calculate the mode, returning NA if there are no non-NA values

```
getMode <- function(v) {
  if (all(is.na(v))) {
    return(NA)
  } else {
    uniqv <- unique(na.omit(v))
    return(uniqv[which.max(tabulate(match(v, uniqv)))])
  }
}
```

Updated function to replace NA values

```
replaceNA <- function(df) {
  for (col in names(df)) {
    if (is.numeric(df[[col]])) {
      df[[col]][is.na(df[[col]])] <- mean(df[[col]], na.rm = TRUE)
    } else if (is.factor(df[[col]]) || is.character(df[[col]])) {
      mode_value <- getMode(as.character(df[[col]]))
      if (!is.na(mode_value)) {
        df[[col]][is.na(df[[col]])] <- mode_value
        if (is.factor(df[[col]])) {
          df[[col]] <- as.factor(df[[col]])
        }
      }
    }
  }
  return(df)
}
```

Applying the function to your dataframe

```
housing_data <- replaceNA(housing_data)
```

Checking NA percentages again

```
sapply(housing_data, function(x) sum(is.na(x)) / length(x) * 100)
```

```
##        exactPrice         sqftPrice   securityDeposit      propertyType
##                 0                 0                 0                 0
##        furnishing            flrNum firstMonthCharges            facing
##                 0                 0                 0                 0
##       totalFlrNum              city    carpetAreaUnit        carpetArea
##                 0                 0                 0                 0
##          bedrooms         bathrooms         balconies        RentOrSale
##                 0                 0                 0                 0
##              Long               Lat
##                 0                 0
```

```r
dim(housing_data)
```

```
## [1] 27900    18
```

```r
library(dplyr)
```

#Converting area units # Defining conversion factors from each unit to sq-ft

```r
conversion_factors <- c(
  "Sq-ft" = 1,
  "Kanal" = 5445,   # Assuming 1 Kanal = 5445 sq-ft
  "Marla" = 272.25, # Assuming 1 Marla = 272.25 sq-ft
  "Sq-yrd" = 9,     # Assuming 1 Sq-yrd = 9 sq-ft
  "Biswa1" = 1519.994, # Example conversion, adjust as necessary
  "Sq-m" = 10.7639, # Assuming 1 Sq-m = 10.7639 sq-ft
  "Rood" = 10890,   # Assuming 1 Rood = 10890 sq-ft
  "Biswa2" = 1519.994, # Example conversion, adjust as necessary
  "Acre" = 43560    # Assuming 1 Acre = 43560 sq-ft
)
```

Function to convert areas to sq-ft based on unit

```r
convert_to_sqft <- function(area, unit) {
  if (!is.na(unit) && (unit %in% names(conversion_factors))) {
    return(area * conversion_factors[unit])
  } else {
    return(NA) # Return NA if unit is not recognized or NA
  }
}
```

Applying the conversion function to the carpetArea column

```r
housing_data$carpetArea_in_sqft <- mapply(convert_to_sqft,
housing_data$carpetArea, housing_data$carpetAreaUnit)
```

Optionally, remove the original carpetArea and carpetAreaUnit columns

```r
housing_data$carpetArea <- NULL
housing_data$carpetAreaUnit <- NULL
```

Converting factors to dummy variables using model.matrix '-1' removes the intercept term to get a full set of dummy variables

```r
dummy_vars <- model.matrix(~ . - 1, data = housing_data)
```

Converting to a dataframe

```
dummy_data <- as.data.frame(dummy_vars)
dim(dummy_data)

## [1] 27900   129
```

Fit a linear regression model

```
model <- lm(exactPrice ~ ., data = dummy_data)
```

Summary of the model to check coefficients and their significance

```
summary(model)

##
## Call:
## lm(formula = exactPrice ~ ., data = dummy_data)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -196122529   -2544024     -80661    1936539 3184473219
##
## Coefficients: (4 not defined because of singularities)
##                                        Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                            1.130e+07  2.428e+07   0.465
0.641701
## sqftPrice                              1.302e+00  6.379e-02  20.414  < 2e-
16
## securityDeposit                       -1.285e+01  1.529e+00  -8.406  < 2e-
16
## `propertyTypeBuilder Floor Apartment` -4.976e+06  7.886e+05  -6.310 2.83e-
10
## `propertyTypeMultistorey Apartment`   -4.947e+06  6.941e+05  -7.127 1.05e-
12
## propertyTypePenthouse                 -3.403e+06  2.225e+06  -1.530
0.126059
## `propertyTypeResidential House`       -2.955e+06  6.616e+05  -4.467 7.97e-
06
## `propertyTypeStudio Apartment`        -5.763e+06  1.698e+06  -3.394
0.000690
## propertyTypeVilla                             NA         NA      NA
NA
## `furnishingSemi-Furnished`            -5.024e+05  4.102e+05  -1.225
0.220659
## furnishingUnfurnished                 -3.895e+05  4.185e+05  -0.931
0.351989
## flrNum10                               1.058e+05  1.652e+06   0.064
0.948936
```

```
## flrNum11                          -3.580e+06  2.146e+06  -1.668
0.095340
## flrNum12                          -2.989e+06  2.185e+06  -1.368
0.171427
## flrNum13                          -4.933e+05  3.292e+06  -0.150
0.880877
## flrNum14                           4.701e+06  3.023e+06   1.555
0.119906
## flrNum15                          -2.359e+06  3.009e+06  -0.784
0.433161
## flrNum16                          -6.701e+06  4.025e+06  -1.665
0.096002
## flrNum17                          -2.101e+06  4.126e+06  -0.509
0.610625
## flrNum18                           5.388e+06  3.717e+06   1.450
0.147160
## flrNum19                           4.477e+06  4.203e+06   1.065
0.286782
## flrNum2                           -1.403e+05  4.584e+05  -0.306
0.759650
## flrNum20                           2.400e+06  3.614e+06   0.664
0.506641
## flrNum21                          -4.114e+06  4.549e+06  -0.904
0.365838
## flrNum22                          -1.250e+07  6.077e+06  -2.057
0.039678
## flrNum23                           5.773e+06  9.154e+06   0.631
0.528245
## flrNum24                           6.408e+06  6.161e+06   1.040
0.298323
## flrNum25                          -2.251e+06  6.742e+06  -0.334
0.738492
## flrNum26                           1.650e+06  8.582e+06   0.192
0.847558
## flrNum27                          -2.206e+06  8.565e+06  -0.258
0.796766
## flrNum28                          -2.652e+05  7.352e+06  -0.036
0.971223
## flrNum29                          -1.086e+06  1.209e+07  -0.090
0.928382
## flrNum3                           -1.127e+05  5.674e+05  -0.199
0.842579
## flrNum30                           9.774e+06  5.665e+06   1.725
0.084463
## flrNum31                          -9.617e+06  1.391e+07  -0.691
0.489475
## flrNum32                           3.448e+07  9.916e+06   3.477
0.000508
## flrNum33                          -2.982e+06  9.881e+06  -0.302
0.762823
```

```
## flrNum34                              -1.456e+06  9.902e+06  -0.147
0.883138
## flrNum35                               1.031e+07  1.208e+07   0.854
0.393353
## flrNum36                               1.374e+07  1.085e+07   1.266
0.205590
## flrNum37                              -9.617e+06  1.398e+07  -0.688
0.491404
## flrNum38                               3.811e+06  9.945e+06   0.383
0.701541
## flrNum39                               4.701e+07  1.399e+07   3.361
0.000777
## flrNum4                               -3.841e+05  6.775e+05  -0.567
0.570767
## flrNum40                              -9.582e+06  1.703e+07  -0.563
0.573709
## flrNum41                               1.295e+07  1.714e+07   0.756
0.449883
## flrNum42                               2.355e+07  9.928e+06   2.372
0.017675
## flrNum43                               4.520e+07  2.413e+07   1.874
0.060986
## flrNum44                               3.104e+07  1.707e+07   1.819
0.068979
## flrNum45                               4.213e+07  1.394e+07   3.022
0.002515
## flrNum46                               1.126e+07  2.409e+07   0.468
0.640120
## flrNum47                               1.814e+07  2.408e+07   0.754
0.451144
## flrNum5                               -6.268e+05  7.852e+05  -0.798
0.424665
## flrNum50                               7.105e+07  2.409e+07   2.950
0.003185
## flrNum53                               7.188e+06  2.410e+07   0.298
0.765519
## flrNum54                               1.719e+08  2.421e+07   7.102 1.26e-
12
## flrNum55                               3.544e+07  2.408e+07   1.472
0.141108
## flrNum56                               4.087e+07  1.717e+07   2.381
0.017285
## flrNum58                               1.510e+07  1.713e+07   0.882
0.378032
## flrNum6                               -2.980e+05  1.027e+06  -0.290
0.771675
## flrNum60                              -2.117e+07  2.409e+07  -0.879
0.379461
## flrNum61                               1.023e+08  2.416e+07   4.233 2.31e-
05
```

```
## flrNum63                                7.159e+07  2.416e+07   2.964
0.003040
## flrNum65                                6.716e+06  2.412e+07   0.278
0.780646
## flrNum66                                2.594e+08  2.440e+07  10.631  < 2e-
16
## flrNum7                                -1.656e+06  1.225e+06  -1.352
0.176454
## flrNum70                                1.121e+08  2.455e+07   4.564 5.04e-
06
## flrNum8                                -1.422e+06  1.372e+06  -1.036
0.300088
## flrNumGround                            1.552e+06  4.746e+05   3.270
0.001077
## `flrNumLower Basement`                 -2.700e+06  2.908e+06  -0.929
0.353074
## `flrNumUpper Basement`                 -4.013e+05  2.789e+06  -0.144
0.885571
## firstMonthCharges                       1.068e-04  2.974e-03   0.036
0.971350
## facingNorth                             5.978e+05  6.013e+05   0.994
0.320111
## `facingNorth - East`                   -6.316e+05  5.537e+05  -1.141
0.253987
## `facingNorth - West`                    5.578e+05  1.113e+06   0.501
0.616282
## facingSouth                             1.796e+04  1.102e+06   0.016
0.987001
## `facingSouth - East`                    3.131e+05  1.098e+06   0.285
0.775592
## `facingSouth -West`                     4.726e+06  1.495e+06   3.162
0.001570
## facingWest                              5.788e+05  7.128e+05   0.812
0.416775
## totalFlrNum                             1.667e+05  3.464e+04   4.813 1.50e-
06
## cityAgartala                           -6.992e+06  2.457e+07  -0.285
0.775931
## cityBangalore                          -7.239e+06  2.402e+07  -0.301
0.763106
## cityBhopal                             -1.171e+07  2.401e+07  -0.488
0.625777
## cityChandigarh                         -5.495e+06  2.401e+07  -0.229
0.818982
## cityChennai                            -7.521e+06  2.402e+07  -0.313
0.754166
## cityDehradun                           -9.557e+06  2.401e+07  -0.398
0.690599
## cityGandhinagar                        -8.515e+06  2.401e+07  -0.355
0.722859
```

```
## cityGangtok                          -1.267e+07  2.630e+07  -0.482
0.629994
## cityGoa                              -4.979e+06  2.401e+07  -0.207
0.835720
## cityHyderabad                        -9.056e+06  2.401e+07  -0.377
0.706049
## cityJaipur                           -1.003e+07  2.400e+07  -0.418
0.676116
## cityKolkata                          -7.600e+06  2.401e+07  -0.317
0.751607
## cityLucknow                          -9.714e+06  2.401e+07  -0.405
0.685767
## cityMumbai                            1.354e+07  2.402e+07   0.564
0.573036
## `cityNew Delhi`                      -6.595e+06  2.442e+07  -0.270
0.787135
## `cityNew-Delhi`                      -8.128e+06  2.402e+07  -0.338
0.735048
## cityPatna                            -8.095e+06  2.401e+07  -0.337
0.735958
## cityRaipur                           -1.200e+07  2.401e+07  -0.500
0.617240
## bedrooms2                             2.478e+06  6.027e+05   4.111 3.95e-
05
## bedrooms3                             3.705e+06  7.210e+05   5.139 2.78e-
07
## bedrooms4                             5.206e+06  9.621e+05   5.411 6.33e-
08
## bedrooms5                             7.610e+06  1.398e+06   5.443 5.28e-
08
## bedrooms6                             2.962e+07  1.714e+06  17.283  < 2e-
16
## bedrooms7                            -2.925e+05  2.629e+06  -0.111
0.911391
## bedrooms8                            -2.685e+06  3.321e+06  -0.809
0.418809
## bedrooms9                                   NA         NA      NA
NA
## bedrooms10                            1.179e+07  4.199e+06   2.808
0.004984
## bathrooms2                           -1.444e+06  5.449e+05  -2.649
0.008074
## bathrooms3                           -8.124e+05  7.202e+05  -1.128
0.259316
## bathrooms4                            2.788e+06  9.778e+05   2.851
0.004363
## bathrooms5                            1.695e+07  1.369e+06  12.381  < 2e-
16
## bathrooms6                            5.525e+06  1.944e+06   2.841
0.004496
```

```
## bathrooms7                                2.022e+07  2.889e+06   6.999 2.63e-
12
## bathrooms8                                2.609e+07  4.294e+06   6.075 1.26e-
09
## bathrooms9                                       NA         NA      NA
NA
## bathrooms10                               2.958e+07  6.845e+06   4.321 1.56e-
05
## balconies2                                3.509e+05  3.875e+05   0.906
0.365188
## balconies3                                6.260e+05  5.409e+05   1.157
0.247163
## balconies4                                5.665e+06  8.791e+05   6.444 1.18e-
10
## balconies5                                6.072e+06  1.917e+06   3.168
0.001537
## balconies6                                2.766e+07  3.387e+06   8.165 3.34e-
16
## balconies7                                1.606e+07  7.011e+06   2.290
0.022024
## balconies8                                1.679e+08  6.296e+06  26.663  < 2e-
16
## balconies9                                       NA         NA      NA
NA
## balconies10                               4.077e+06  1.202e+07   0.339
0.734584
## RentOrSaleSale                            8.894e+06  3.697e+05  24.061  < 2e-
16
## Long                                     -1.147e+04  4.028e+04  -0.285
0.775794
## Lat                                      -1.108e+04  7.179e+04  -0.154
0.877357
## carpetArea_in_sqft                        4.874e+00  7.721e-01   6.312 2.80e-
10
##
## (Intercept)
## sqftPrice                                ***
## securityDeposit                          ***
## `propertyTypeBuilder Floor Apartment`    ***
## `propertyTypeMultistorey Apartment`      ***
## propertyTypePenthouse
## `propertyTypeResidential House`          ***
## `propertyTypeStudio Apartment`           ***
## propertyTypeVilla
## `furnishingSemi-Furnished`
## furnishingUnfurnished
## flrNum10
## flrNum11                                 .
## flrNum12
## flrNum13
```

```
## flrNum14
## flrNum15
## flrNum16                          .
## flrNum17
## flrNum18
## flrNum19
## flrNum2
## flrNum20
## flrNum21
## flrNum22                          *
## flrNum23
## flrNum24
## flrNum25
## flrNum26
## flrNum27
## flrNum28
## flrNum29
## flrNum3
## flrNum30                          .
## flrNum31
## flrNum32                          ***
## flrNum33
## flrNum34
## flrNum35
## flrNum36
## flrNum37
## flrNum38
## flrNum39                          ***
## flrNum4
## flrNum40
## flrNum41
## flrNum42                          *
## flrNum43                          .
## flrNum44                          .
## flrNum45                          **
## flrNum46
## flrNum47
## flrNum5
## flrNum50                          **
## flrNum53
## flrNum54                          ***
## flrNum55
## flrNum56                          *
## flrNum58
## flrNum6
## flrNum60
## flrNum61                          ***
## flrNum63                          **
## flrNum65
## flrNum66                          ***
```

```
## flrNum7
## flrNum70                              ***
## flrNum8
## flrNumGround                          **
## `flrNumLower Basement`
## `flrNumUpper Basement`
## firstMonthCharges
## facingNorth
## `facingNorth - East`
## `facingNorth - West`
## facingSouth
## `facingSouth - East`
## `facingSouth -West`                   **
## facingWest
## totalFlrNum                           ***
## cityAgartala
## cityBangalore
## cityBhopal
## cityChandigarh
## cityChennai
## cityDehradun
## cityGandhinagar
## cityGangtok
## cityGoa
## cityHyderabad
## cityJaipur
## cityKolkata
## cityLucknow
## cityMumbai
## `cityNew Delhi`
## `cityNew-Delhi`
## cityPatna
## cityRaipur
## bedrooms2                             ***
## bedrooms3                             ***
## bedrooms4                             ***
## bedrooms5                             ***
## bedrooms6                             ***
## bedrooms7
## bedrooms8
## bedrooms9
## bedrooms10                            **
## bathrooms2                            **
## bathrooms3
## bathrooms4                            **
## bathrooms5                            ***
## bathrooms6                            **
## bathrooms7                            ***
## bathrooms8                            ***
## bathrooms9
```

```
## bathrooms10                              ***
## balconies2
## balconies3
## balconies4                               ***
## balconies5                               **
## balconies6                               ***
## balconies7                               *
## balconies8                               ***
## balconies9
## balconies10
## RentOrSaleSale                           ***
## Long
## Lat
## carpetArea_in_sqft                       ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.4e+07 on 27775 degrees of freedom
## Multiple R-squared:  0.212,  Adjusted R-squared:  0.2085
## F-statistic: 60.27 on 124 and 27775 DF,  p-value: < 2.2e-16
```

To improve the Adjusted R-squared we will convert the flrNum column from character to numeric, where specific non-numeric values like “Ground”, “Lower Basement”, and “Upper Basement” need to be assigned specific numeric values.

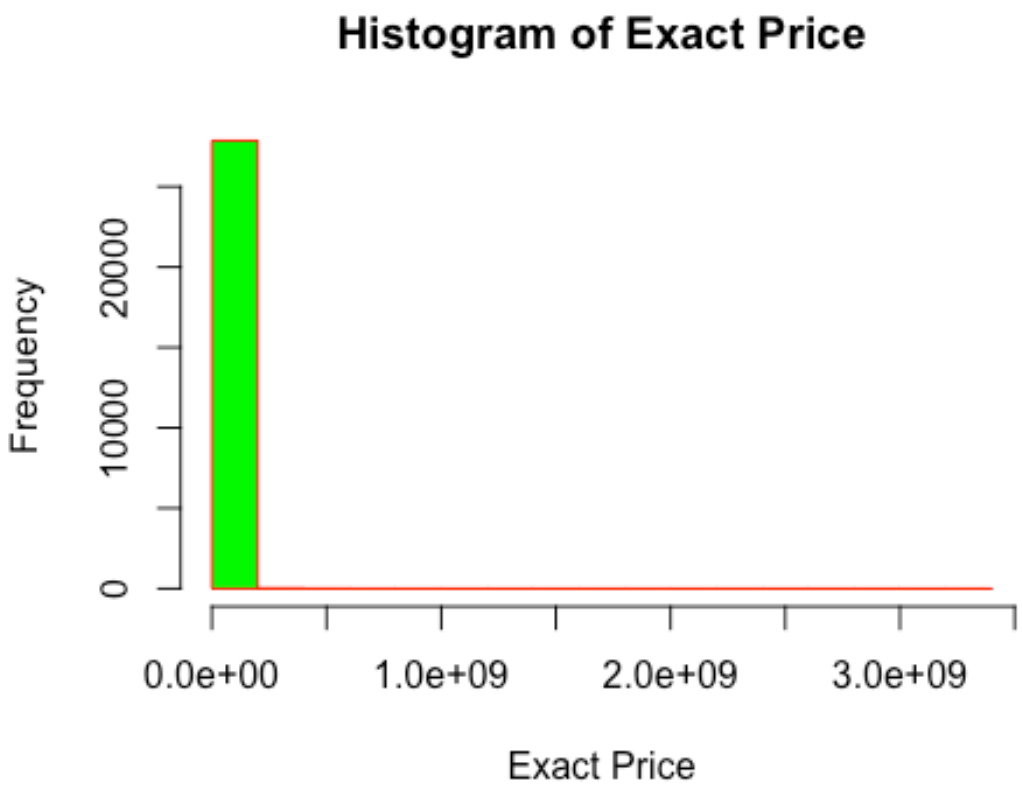
Let’s assign 0 to “Ground”, -1 to “Lower Basement”, and -2 to “Upper Basement”.

Converting the flrNum column to numeric, with specific assignments for non-numeric values

```
housing_data <- housing_data %>%
  mutate(flrNum = case_when(
    flrNum == "Ground" ~ "0",
    flrNum == "Lower Basement" ~ "-1",
    flrNum == "Upper Basement" ~ "-2",
    TRUE ~ as.character(flrNum)  # Keep other values as they are
  )) %>%
  mutate(flrNum = as.numeric(flrNum))  # Convert the modified column to
numeric
```

Plotting histogram of target variable

```
library(e1071)
hist(housing_data$exactPrice, main = "Histogram of Exact Price", xlab =
"Exact Price", col = "green", border = "red")
```



Calculating skewness of Exact Price

```
exactPrice_skewness <- skewness(housing_data$exactPrice)
```

Skewness of Exact Price

```
cat("Skewness of Exact Price:", exactPrice_skewness, "\n")

## Skewness of Exact Price: 68.05725
```

Checking The Outliers in Target Variable

Calculating the quartiles and IQR for exactPrice

```
Q1 <- quantile(housing_data$exactPrice, 0.25)
Q3 <- quantile(housing_data$exactPrice, 0.75)
IQR <- Q3 - Q1
```

Defining the lower and upper bounds for outliers

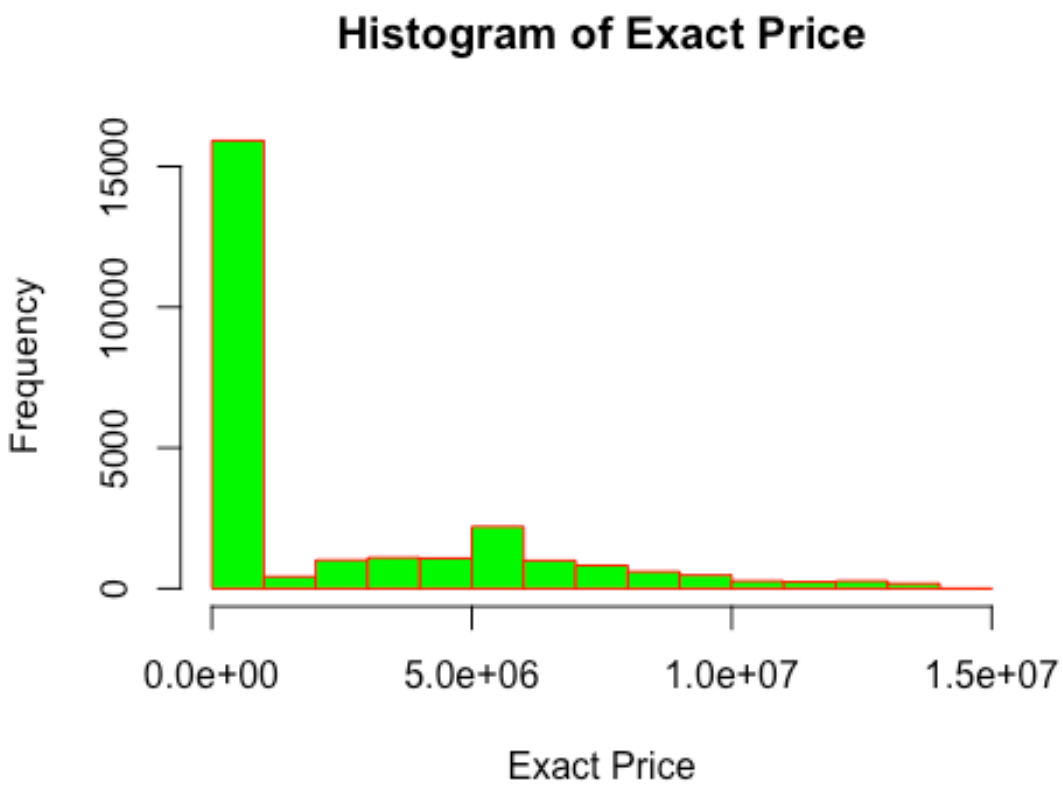
```
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR
```

Filtering the dataframe to exclude outliers in exactPrice and overwrite the original dataframe

```
housing_data <- housing_data %>%
  filter(exactPrice >= lower_bound & exactPrice <= upper_bound)

skewness(housing_data$exactPrice)

## [1] 1.340153
```

Checking the dimensions of the updated dataframe to confirm the removal of outliers

```
dim(housing_data)

## [1] 25601    17

hist(housing_data$exactPrice, main = "Histogram of Exact Price", xlab =
"Exact Price", col = "green", border = "red")
```



The histogram shows that the distribution of Exact Price is right-skewed, with a majority of properties clustered at the lower price range and fewer properties with higher prices. This suggests that affordable properties are more common in the dataset than high-priced ones.

The long tail to the right indicates that there are properties with significantly higher prices compared to the rest, which could be considered outliers.

Calculating Correlation Matrix

Calculating the correlation matrix for numeric variables in the dataframe

```
cor_matrix <- cor(housing_data %>% select_if(is.numeric))
```

Print the correlation matrix

```
print(cor_matrix)
```

```
##                       exactPrice     sqftPrice securityDeposit
flrNum
## exactPrice          1.000000e+00  0.0724850043    0.0058829747
0.068185228
## sqftPrice           7.248500e-02  1.0000000000   -0.0085262644 -
0.022509558
```

```
## securityDeposit     5.882975e-03 -0.0085262644    1.0000000000
0.055069463
## flrNum              6.818523e-02 -0.0225095575    0.0550694626
1.000000000
## firstMonthCharges   7.907576e-05 -0.0007283299   -0.0007531494 -
0.003318893
## totalFlrNum         1.451647e-01 -0.0065796222    0.0579298672
0.645108120
## Long               -7.477199e-03  0.0014070940   -0.0122381439 -
0.037655324
## Lat                 1.948436e-01  0.0079178720   -0.1615933535 -
0.029730390
## carpetArea_in_sqft  1.024215e-03 -0.0006765209    0.0008926242 -
0.002798513
##                    firstMonthCharges  totalFlrNum         Long
Lat
## exactPrice              7.907576e-05  0.145164677 -0.007477199
0.194843639
## sqftPrice              -7.283299e-04 -0.006579622  0.001407094
0.007917872
## securityDeposit        -7.531494e-04  0.057929867 -0.012238144 -
0.161593354
## flrNum                 -3.318893e-03  0.645108120 -0.037655324 -
0.029730390
## firstMonthCharges       1.000000e+00 -0.004753938 -0.001375345
0.005910621
## totalFlrNum            -4.753938e-03  1.000000000 -0.028796570 -
0.017014573
## Long                   -1.375345e-03 -0.028796570  1.000000000
0.054603927
## Lat                     5.910621e-03 -0.017014573  0.054603927
1.000000000
## carpetArea_in_sqft     -1.011596e-05 -0.002275179 -0.003536640
0.004242935
##                    carpetArea_in_sqft
## exactPrice              1.024215e-03
## sqftPrice              -6.765209e-04
## securityDeposit         8.926242e-04
## flrNum                 -2.798513e-03
## firstMonthCharges      -1.011596e-05
## totalFlrNum            -2.275179e-03
## Long                   -3.536640e-03
## Lat                     4.242935e-03
## carpetArea_in_sqft      1.000000e+00
```

Calculate eigenvalues of the correlation matrix

```
eigenvalues <- eigen(cor_matrix)$values
print(eigenvalues)
```

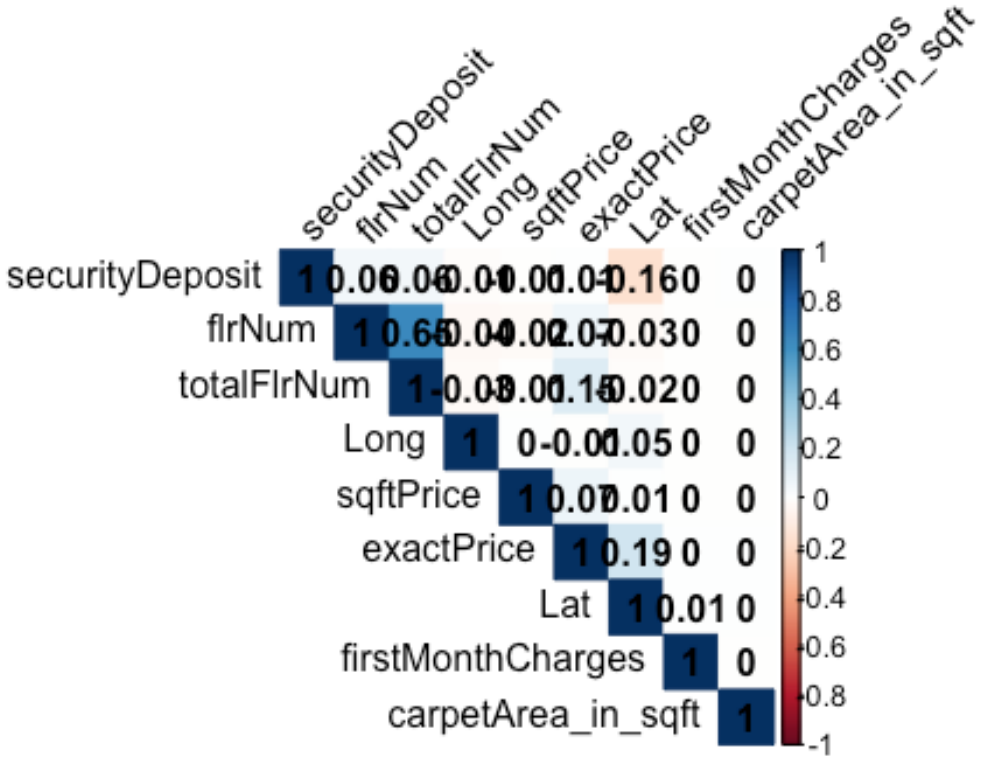
```
## [1] 1.6916806 1.2666851 1.0301288 1.0017415 0.9999468 0.9902040 0.9407855
## [8] 0.7288473 0.3499804
```

Inspecting the eigenvalues for very small values close to zero, which would indicate multicollinearity

```
very_small_eigenvalues <- eigenvalues[abs(eigenvalues) < 1e-10]
print(very_small_eigenvalues)

## numeric(0)
```

Visualizing the Correlation Matrix

```
library(corrplot)
corrplot(cor_matrix, method = "color", type = "upper", order = "hclust",
         tl.col = "black", tl.srt = 45, # Text label color and rotation
         addCoef.col = "black") # Add correlation coefficients to plot
```



Getting column names for numeric columns

```
numeric_cols <- names(housing_data)[sapply(housing_data, is.numeric)]
```

Subset the dataframe to only include numeric columns

```
numeric_data <- housing_data[numeric_cols]
```

Getting column names for categorical columns (factors and characters)

```
categorical_cols <- names(housing_data)[sapply(housing_data, function(x)
is.factor(x) | is.character(x))]
```

Print the column names

```
print(numeric_cols)
```

```
## [1] "exactPrice"         "sqftPrice"          "securityDeposit"
## [4] "flrNum"             "firstMonthCharges"  "totalFlrNum"
## [7] "Long"               "Lat"                "carpetArea_in_sqft"
```

```
print(categorical_cols)
```

```
## [1] "propertyType" "furnishing"   "facing"       "city"         "bedrooms"
## [6] "bathrooms"    "balconies"    "RentOrSale"
```

#Applying linear regression model to check the vif values for numerical columns.

```
model_vif <- lm(exactPrice ~ ., data = numeric_data)
library(car)
```

Calculating VIF

```
vif_values <- vif(model_vif)
print(vif_values)
```

```
##          sqftPrice    securityDeposit             flrNum
firstMonthCharges
##           1.000711           1.030476           1.716417
1.000061
##        totalFlrNum               Long                Lat
carpetArea_in_sqft
##           1.714664           1.004357           1.030282
1.000044
```

All VIF values are close to 1, with the highest observed value being approximately 1.716 for flrNum and totalFlrNum, which is well below the commonly used threshold of 5 or 10 that might indicate problematic multicollinearity. Therefore, we can conclude that the predictors in the model exhibit low multicollinearity.

Question A)

Multicollinearity is measured in the correlation plot by looking at the correlation coefficients between independent variable pairs.The coefficient values range from -1 to 1, where values close to 1 or -1 indicate a strong positive or negative linear relationship, respectively, and values close to 0 indicate no linear relationship.

From the plot, it's observed that:

The variables flrNum (which could represent the floor number of the property) and totalFlrNum (likely representing the total number of floors in the building) have a correlation coefficient of 0.65. This suggests a moderate positive correlation, which is expected since the floor number will not exceed the total number of floors. All other variables shown in the plot have correlation coefficients lower than 0.2 with respect to each other, indicating weak linear relationships.

To address the issue of multicollinearity:

I checked for multicollinearity by calculating the correlation matrix for numeric variables in the dataframe and visualizing it using the corrplot function from the corrplot library in R. This method is effective for identifying pairs of variables that may have high multicollinearity. After looking at the plot , there is no multicollinearity present between the independent Variables as no variables has values greater than 0.8

Before running the regression, other issues with the dataset which we considered are: Ensured there are no missing values within the predictors, or appropriately handling them if there are.

Checking for outliers that could skew the results and addressing them. Verifying that categorical variables are correctly encoded as dummy variables.

EXPLORATORY DATA ANALYSIS

Question B)

1) Histogram of Exact Price

```
library(ggplot2)
ggplot(housing_data, aes(x = exactPrice)) +
  geom_histogram(bins = 30, fill = "blue", alpha = 0.7) +
  labs(title = "Distribution of Exact Price", x = "Exact Price", y = "Count")
```

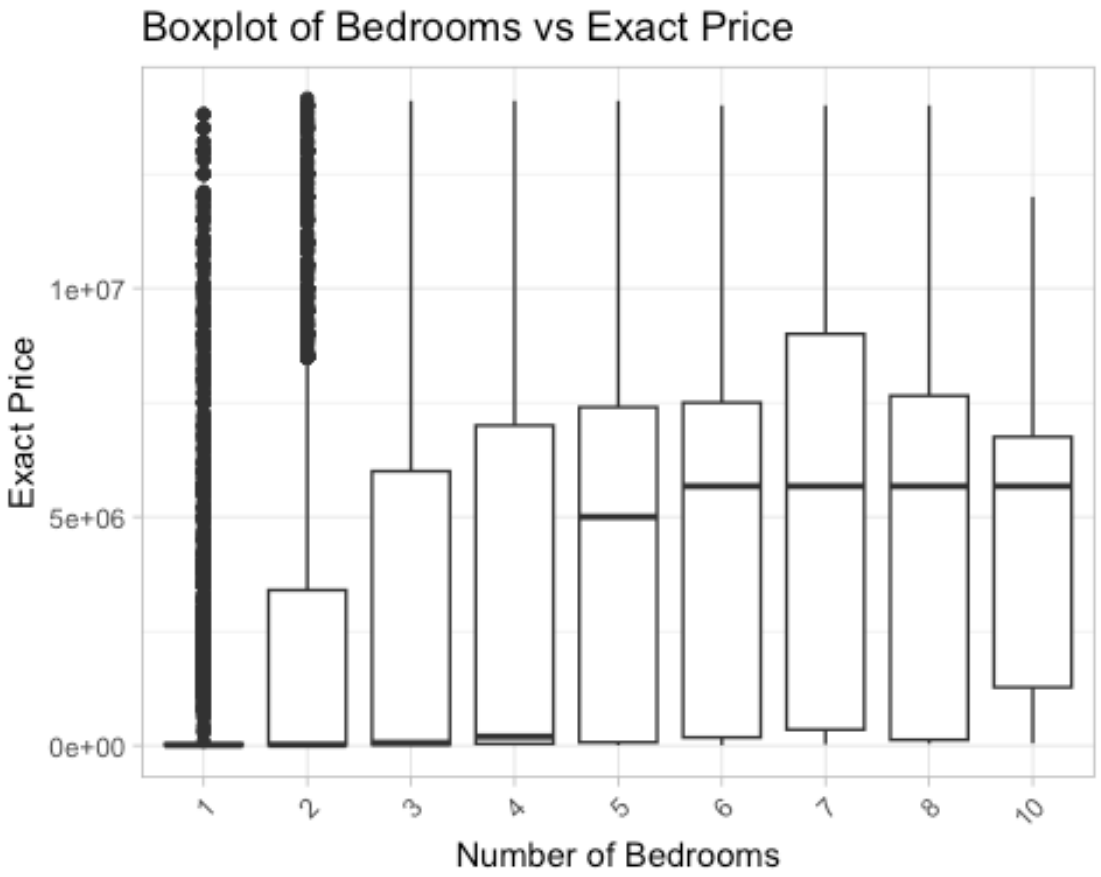
```
+
  theme_minimal()
```

The distribution of Exact Price is highly right skewed, indicating that a large number of properties have prices clustered at the lower end of the spectrum, while a few properties have very high prices.

The majority of properties fall within the lower price range, as evidenced by the tall bar on the far left, suggesting that affordable properties dominate this market.

2) Boxplot for 'bedrooms' vs 'exactPrice'

```
ggplot(housing_data, aes(x = bedrooms, y = exactPrice)) +
  geom_boxplot() +
  labs(title = "Boxplot of Bedrooms vs Exact Price",
       x = "Number of Bedrooms",
       y = "Exact Price") +
  theme_light() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  # Rotate x axis
labels for better readability
```



The plot indicates a general trend where properties with more bedrooms tend to have higher prices. This suggests that bedroom count is a significant factor in property valuation.

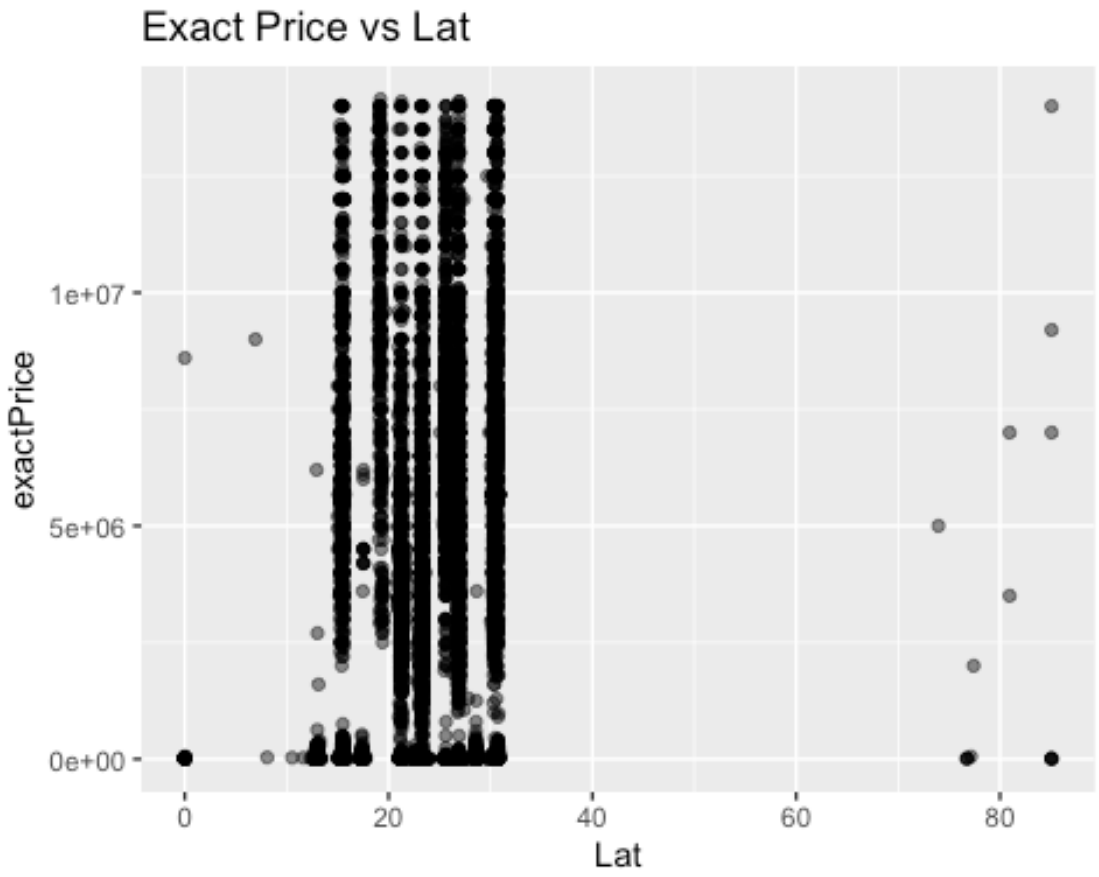
A substantial number of outliers particularly in properties with fewer bedrooms, points to some properties being priced significantly higher than the median.

For properties with 6 or more bedrooms, the wide range in prices could mean a varied market for such properties.

The data shows a ceiling effect with 8 to 10 bedroom properties, suggesting that at a certain point, additional bedrooms do not significantly increase the property's value.

3) Scatter plot for geographical distribution with color representing flrNum

```
ggplot(housing_data, aes(x = Lat, y = exactPrice)) +
  geom_point(alpha = 0.5) +
  labs(title = "Exact Price vs Lat", x = "Lat", y = "exactPrice")
```



The scatter plot shows the relationship between property prices and their latitude, revealing clusters at specific latitudes which likely correspond to particular cities or regions.

The wide range of prices at these latitudes suggests that within each region, properties vary significantly in cost, from budget-friendly to highly expensive.

4) Scatterplot for exactPrice vs securityDeposit

```
ggplot(housing_data, aes(x = securityDeposit, y = exactPrice)) +
  geom_point(aes(color = securityDeposit), alpha = 0.6) +
  scale_color_gradient(low = "blue", high = "red") +
  theme_minimal() +
  labs(title = "Scatterplot of Exact Price vs Security Deposit",
       x = "Security Deposit",
       y = "Exact Price")
```

The scatterplot visualizes the relationship between the exact price of properties and the security deposit required.

It indicates that while most properties have a relatively low security deposit, the exact price varies significantly, with a concentration of properties at the lower price end and some outliers with high prices.

5) Jitter plot for RentOrSale vs exactPrice

```
ggplot(housing_data, aes(x = RentOrSale, y = exactPrice)) +
  geom_jitter(aes(color = RentOrSale), width = 0.2, alpha = 0.6) +
  theme_minimal() +
  labs(title = "Jitter Plot of Exact Price by Rent or Sale",
       x = "Rent or Sale",
       y = "Exact Price")
```

The jitter plot reveals a stark contrast in Exact Price between properties available for rent and those for sale, with sale prices markedly higher than rental prices.

It also shows a dense concentration of rental prices at the lower end, indicating a more uniform and affordable rental market compared to the wide range and higher values seen in the sales market.

6) Barplot for bedrooms vs exactPrice

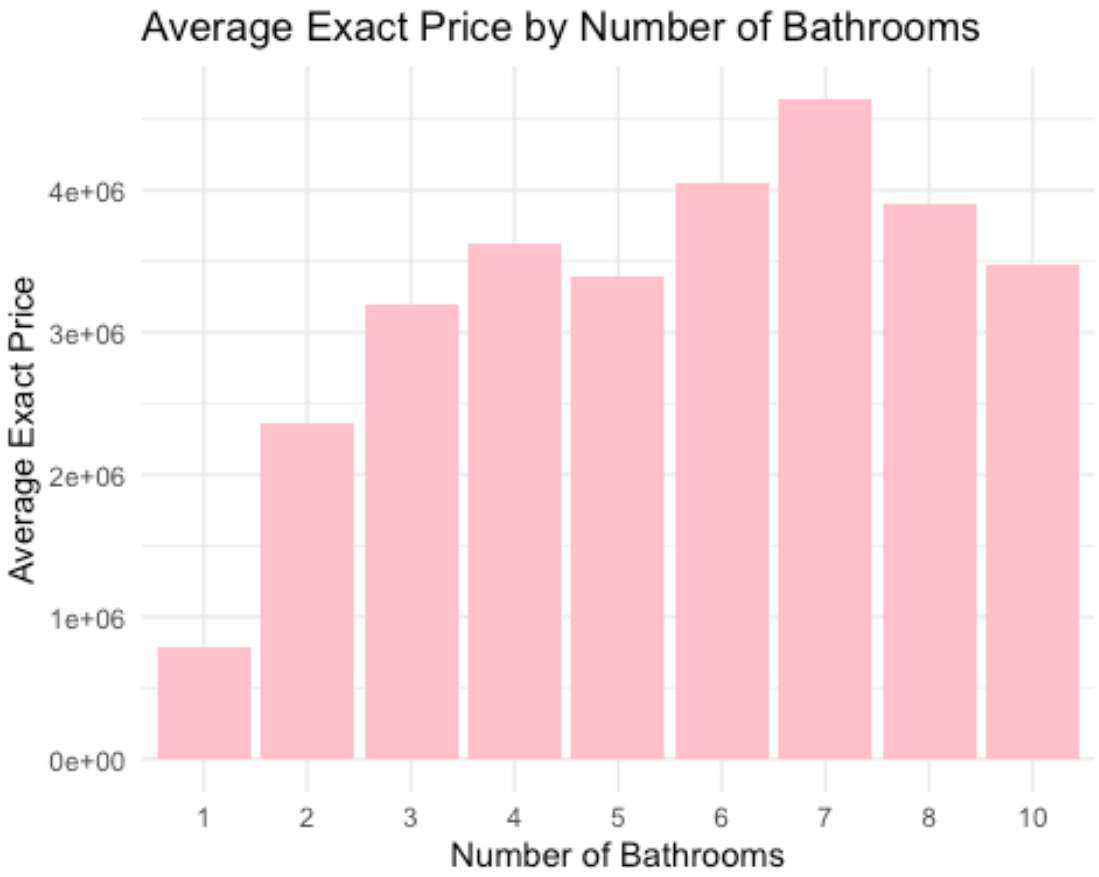
```
ggplot(housing_data, aes(x = factor(bedrooms), y = exactPrice)) +
  geom_bar(stat = "summary", fun = "mean", fill = "green",) +
  labs(title = "Average Exact Price by Number of Bedrooms",
       x = "Number of Bedrooms",
       y = "Average Exact Price") +
  theme_minimal()
```

The bar chart illustrates a general increase in the average exact price of properties as the number of bedrooms grows, indicating that larger properties with more bedrooms are typically more expensive.

There's a noticeable peak at 7 bedrooms, after which the average price slightly dips for properties with 8 bedrooms.

7) Barplot for bathrooms vs exactPrice

```
ggplot(housing_data, aes(x = factor(bathrooms), y = exactPrice)) +
  geom_bar(stat = "summary", fun = "mean", fill = "pink") +
  labs(title = "Average Exact Price by Number of Bathrooms",
       x = "Number of Bathrooms",
       y = "Average Exact Price") +
  theme_minimal()
```

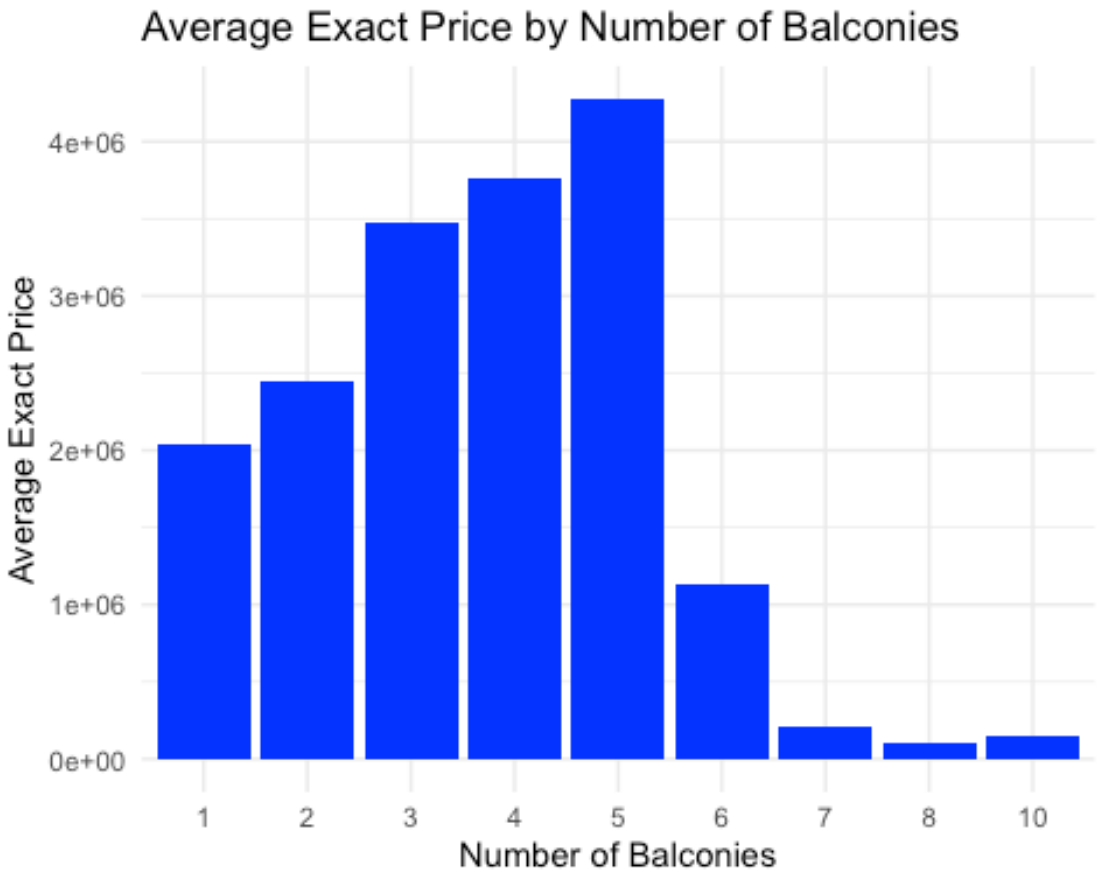


The bar chart shows that the average exact price of properties tends to increase with the number of bathrooms, indicating that properties with more bathrooms are typically priced higher.

The chart peaks at 7 bathrooms, suggesting that properties at this level of amenity command the highest average prices within the dataset.

8) Barplot for balconies vs exactPrice

```
ggplot(housing_data, aes(x = factor(balconies), y = exactPrice)) +
  geom_bar(stat = "summary", fun = "mean", fill = "blue") +
  labs(title = "Average Exact Price by Number of Balconies",
       x = "Number of Balconies",
       y = "Average Exact Price") +
  theme_minimal()
```



The bar chart suggests that properties with 4 to 5 balconies command the highest average prices, hinting at a premium for additional outdoor space.

However, there's a notable decrease for properties with more than 5 balconies, which could suggest a point of diminishing returns or a niche market for such features.

Converting to Dummies

Now Let's Convert factors to dummy variables using model.matrix again

```
dummy_vars <- model.matrix(~ . - 1, data = housing_data)
```

Converting to a dataframe

```
dummy_data <- as.data.frame(dummy_vars)
```

Applying Model

Fit a linear regression model.

```
model <- lm(exactPrice ~ ., data = dummy_data)
```

Summary of the model to check coefficients and their significance

```
summary(model)
```

```
##
## Call:
## lm(formula = exactPrice ~ ., data = dummy_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7145984  -789768   -74071   454153  8852800
##
## Coefficients: (4 not defined because of singularities)
##                                          Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                             5.669e+06  1.648e+06   3.439
0.000584
## sqftPrice                               1.081e+00  1.643e-01   6.583 4.71e-
11
## securityDeposit                        -1.258e+00  1.061e-01 -11.855  < 2e-
16
## `propertyTypeBuilder Floor Apartment` -5.773e+05  5.518e+04 -10.464  < 2e-
16
## `propertyTypeMultistorey Apartment`   -2.807e+05  4.880e+04  -5.753 8.86e-
09
## propertyTypePenthouse                   3.038e+05  1.760e+05   1.726
0.084285
## `propertyTypeResidential House`       -3.130e+04  4.797e+04  -0.653
0.514074
## `propertyTypeStudio Apartment`        -2.602e+05  1.161e+05  -2.241
0.025034
## propertyTypeVilla                             NA         NA      NA
NA
## `furnishingSemi-Furnished`             6.014e+04  2.903e+04   2.072
0.038291
## furnishingUnfurnished                 -6.757e+04  2.977e+04  -2.270
0.023208
## flrNum                                -1.232e+04  5.116e+03  -2.408
0.016050
## firstMonthCharges                      8.050e-06  2.018e-04   0.040
0.968181
## facingNorth                            1.267e+05  4.320e+04   2.933
0.003364
## `facingNorth - East`                   1.968e+05  3.978e+04   4.948 7.54e-
07
## `facingNorth - West`                   1.038e+05  8.154e+04   1.273
0.203024
## facingSouth                            1.532e+05  7.796e+04   1.965
0.049460
## `facingSouth - East`                   2.737e+04  7.890e+04   0.347
0.728668
```

```
## `facingSouth -West`                  -2.860e+04  1.105e+05  -0.259
0.795712
## facingWest                            9.529e+04  5.146e+04   1.852
0.064059
## totalFlrNum                           1.735e+04  2.798e+03   6.202 5.68e-
10
## cityAgartala                         -5.577e+06  1.667e+06  -3.346
0.000821
## cityBangalore                        -5.672e+06  1.630e+06  -3.481
0.000501
## cityBhopal                           -6.968e+06  1.629e+06  -4.279 1.89e-
05
## cityChandigarh                       -5.638e+06  1.629e+06  -3.461
0.000539
## cityChennai                          -5.686e+06  1.630e+06  -3.490
0.000484
## cityDehradun                         -5.574e+06  1.629e+06  -3.422
0.000623
## cityGandhinagar                      -6.123e+06  1.629e+06  -3.759
0.000171
## cityGangtok                          -6.198e+06  1.785e+06  -3.473
0.000516
## cityGoa                              -5.169e+06  1.629e+06  -3.173
0.001511
## cityHyderabad                        -5.932e+06  1.629e+06  -3.642
0.000271
## cityJaipur                           -6.198e+06  1.629e+06  -3.805
0.000142
## cityKolkata                          -5.730e+06  1.629e+06  -3.518
0.000436
## cityLucknow                          -5.876e+06  1.629e+06  -3.608
0.000309
## cityMumbai                           -2.326e+06  1.632e+06  -1.426
0.153980
## `cityNew Delhi`                      -5.573e+06  1.657e+06  -3.364
0.000771
## `cityNew-Delhi`                      -5.876e+06  1.629e+06  -3.606
0.000311
## cityPatna                            -5.172e+06  1.629e+06  -3.176
0.001495
## cityRaipur                           -6.836e+06  1.629e+06  -4.197 2.72e-
05
## bedrooms2                             7.252e+04  4.207e+04   1.724
0.084759
## bedrooms3                             4.876e+05  5.051e+04   9.652  < 2e-
16
## bedrooms4                             7.748e+05  7.001e+04  11.068  < 2e-
16
## bedrooms5                             9.892e+05  1.138e+05   8.694  < 2e-
16
```

```
## bedrooms6                          1.022e+06  1.493e+05   6.843 7.93e-
12
## bedrooms7                          1.659e+06  2.409e+05   6.886 5.86e-
12
## bedrooms8                          9.096e+05  2.968e+05   3.064
0.002185
## bedrooms9                                 NA         NA      NA
NA
## bedrooms10                         8.109e+05  4.040e+05   2.007
0.044762
## bathrooms2                         2.143e+05  3.753e+04   5.710 1.14e-
08
## bathrooms3                         6.564e+05  5.006e+04  13.110  < 2e-
16
## bathrooms4                         9.431e+05  7.192e+04  13.113  < 2e-
16
## bathrooms5                         7.351e+05  1.134e+05   6.480 9.36e-
11
## bathrooms6                         1.023e+06  1.806e+05   5.664 1.50e-
08
## bathrooms7                         8.629e+05  2.766e+05   3.120
0.001809
## bathrooms8                         1.055e+05  5.107e+05   0.207
0.836349
## bathrooms9                                NA         NA      NA
NA
## bathrooms10                       -9.394e+04  6.892e+05  -0.136
0.891585
## balconies2                         1.321e+04  2.722e+04   0.485
0.627548
## balconies3                         2.852e+05  3.888e+04   7.335 2.29e-
13
## balconies4                         3.880e+05  6.760e+04   5.740 9.60e-
09
## balconies5                         7.662e+05  1.584e+05   4.837 1.33e-
06
## balconies6                        -3.585e+05  3.474e+05  -1.032
0.302176
## balconies7                        -1.008e+06  7.462e+05  -1.350
0.176877
## balconies8                        -1.202e+06  1.164e+06  -1.033
0.301830
## balconies9                                NA         NA      NA
NA
## balconies10                       -6.202e+05  9.436e+05  -0.657
0.510984
## RentOrSaleSale                     5.787e+06  2.590e+04 223.456  < 2e-
16
## Long                              -2.812e+03  2.768e+03  -1.016
0.309551
```

```
## Lat                                      5.290e+03  5.019e+03   1.054
0.291963
## carpetArea_in_sqft                      -8.049e-02  5.342e-02  -1.507
0.131872
##
## (Intercept)                             ***
## sqftPrice                               ***
## securityDeposit                         ***
## `propertyTypeBuilder Floor Apartment`   ***
## `propertyTypeMultistorey Apartment`     ***
## propertyTypePenthouse                   .
## `propertyTypeResidential House`
## `propertyTypeStudio Apartment`          *
## propertyTypeVilla
## `furnishingSemi-Furnished`              *
## furnishingUnfurnished                   *
## flrNum                                  *
## firstMonthCharges
## facingNorth                             **
## `facingNorth - East`                    ***
## `facingNorth - West`
## facingSouth                             *
## `facingSouth - East`
## `facingSouth -West`
## facingWest                              .
## totalFlrNum                             ***
## cityAgartala                            ***
## cityBangalore                           ***
## cityBhopal                              ***
## cityChandigarh                          ***
## cityChennai                             ***
## cityDehradun                            ***
## cityGandhinagar                         ***
## cityGangtok                             ***
## cityGoa                                 **
## cityHyderabad                           ***
## cityJaipur                              ***
## cityKolkata                             ***
## cityLucknow                             ***
## cityMumbai
## `cityNew Delhi`                         ***
## `cityNew-Delhi`                         ***
## cityPatna                               **
## cityRaipur                              ***
## bedrooms2                               .
## bedrooms3                               ***
## bedrooms4                               ***
## bedrooms5                               ***
## bedrooms6                               ***
## bedrooms7                               ***
```

```
## bedrooms8                                **
## bedrooms9
## bedrooms10                               *
## bathrooms2                               ***
## bathrooms3                               ***
## bathrooms4                               ***
## bathrooms5                               ***
## bathrooms6                               ***
## bathrooms7                               **
## bathrooms8
## bathrooms9
## bathrooms10
## balconies2
## balconies3                               ***
## balconies4                               ***
## balconies5                               ***
## balconies6
## balconies7
## balconies8
## balconies9
## balconies10
## RentOrSaleSale                           ***
## Long
## Lat
## carpetArea_in_sqft
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1628000 on 25535 degrees of freedom
## Multiple R-squared:  0.7748, Adjusted R-squared:  0.7742
## F-statistic:  1351 on 65 and 25535 DF,  p-value: < 2.2e-16
```

Creating a train/test partition

```
set.seed(123)
splitIndex <- createDataPartition(dummy_data$exactPrice, p = 0.8, list =
FALSE)
train_set <- dummy_data[splitIndex, ]
test_set <- dummy_data[-splitIndex, ]
```

Checking the dimensions of the train and test sets

```
dim(train_set)

## [1] 20483    70

dim(test_set)

## [1] 5118   70
```

Fitting a linear regression model on the training set

```
initial_model <- lm(exactPrice ~ ., data = train_set)
```

Summary of the initial model to check coefficients and their significance

```
summary(initial_model)
```

```
##
## Call:
## lm(formula = exactPrice ~ ., data = train_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7475432  -801859   -67154   455627  8591137
##
## Coefficients: (4 not defined because of singularities)
##                                          Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                             5.723e+06  1.654e+06   3.461
0.000539
## sqftPrice                               1.000e+00  1.658e-01   6.034 1.63e-
09
## securityDeposit                        -1.296e+00  1.169e-01 -11.087  < 2e-
16
## `propertyTypeBuilder Floor Apartment` -5.873e+05  6.200e+04  -9.473  < 2e-
16
## `propertyTypeMultistorey Apartment`   -2.818e+05  5.500e+04  -5.123 3.03e-
07
## propertyTypePenthouse                  2.823e+05  1.908e+05   1.479
0.139112
## `propertyTypeResidential House`       -3.782e+04  5.412e+04  -0.699
0.484655
## `propertyTypeStudio Apartment`        -3.105e+05  1.286e+05  -2.414
0.015796
## propertyTypeVilla                             NA         NA      NA
NA
## `furnishingSemi-Furnished`             6.614e+04  3.251e+04   2.034
0.041932
## furnishingUnfurnished                 -6.169e+04  3.330e+04  -1.853
0.063936
## flrNum                                -1.713e+04  5.800e+03  -2.954
0.003145
## firstMonthCharges                      1.216e-05  2.021e-04   0.060
0.952043
## facingNorth                            1.190e+05  4.844e+04   2.456
0.014053
## `facingNorth - East`                   2.138e+05  4.467e+04   4.786 1.72e-
06
```

```
## `facingNorth - West`                  1.225e+05  9.115e+04   1.344
0.179091
## facingSouth                           1.525e+05  8.591e+04   1.775
0.075979
## `facingSouth - East`                 -2.829e+04  8.817e+04  -0.321
0.748313
## `facingSouth -West`                   2.365e+04  1.222e+05   0.194
0.846514
## facingWest                            7.648e+04  5.730e+04   1.335
0.181980
## totalFlrNum                           2.456e+04  3.416e+03   7.190 6.70e-
13
## cityAgartala                         -5.541e+06  1.678e+06  -3.303
0.000960
## cityBangalore                        -5.693e+06  1.632e+06  -3.488
0.000487
## cityBhopal                           -6.948e+06  1.631e+06  -4.260 2.05e-
05
## cityChandigarh                       -5.646e+06  1.631e+06  -3.461
0.000540
## cityChennai                          -5.704e+06  1.632e+06  -3.495
0.000475
## cityDehradun                         -5.563e+06  1.631e+06  -3.410
0.000650
## cityGandhinagar                      -6.130e+06  1.631e+06  -3.758
0.000172
## cityGangtok                          -6.330e+06  1.887e+06  -3.355
0.000795
## cityGoa                              -5.188e+06  1.631e+06  -3.180
0.001475
## cityHyderabad                        -5.936e+06  1.631e+06  -3.639
0.000275
## cityJaipur                           -6.196e+06  1.631e+06  -3.799
0.000146
## cityKolkata                          -5.727e+06  1.631e+06  -3.511
0.000448
## cityLucknow                          -5.883e+06  1.631e+06  -3.607
0.000311
## cityMumbai                           -2.457e+06  1.635e+06  -1.503
0.132835
## `cityNew Delhi`                      -5.534e+06  1.663e+06  -3.328
0.000875
## `cityNew-Delhi`                      -5.875e+06  1.632e+06  -3.600
0.000319
## cityPatna                            -5.143e+06  1.631e+06  -3.154
0.001614
## cityRaipur                           -6.831e+06  1.631e+06  -4.188 2.83e-
05
## bedrooms2                             7.928e+04  4.713e+04   1.682
0.092557
```

```
## bedrooms3                          4.963e+05  5.671e+04   8.750  < 2e-
16
## bedrooms4                          7.662e+05  7.881e+04   9.722  < 2e-
16
## bedrooms5                          9.854e+05  1.284e+05   7.672 1.77e-
14
## bedrooms6                          9.467e+05  1.674e+05   5.656 1.57e-
08
## bedrooms7                          1.587e+06  2.648e+05   5.996 2.06e-
09
## bedrooms8                          9.425e+05  3.308e+05   2.849
0.004387
## bedrooms9                                 NA         NA      NA
NA
## bedrooms10                         1.156e+06  4.598e+05   2.514
0.011940
## bathrooms2                         1.919e+05  4.203e+04   4.567 4.99e-
06
## bathrooms3                         6.520e+05  5.621e+04  11.600  < 2e-
16
## bathrooms4                         9.946e+05  8.090e+04  12.294  < 2e-
16
## bathrooms5                         7.907e+05  1.287e+05   6.144 8.18e-
10
## bathrooms6                         9.518e+05  2.033e+05   4.683 2.84e-
06
## bathrooms7                         1.060e+06  3.045e+05   3.482
0.000498
## bathrooms8                         6.425e+05  5.887e+05   1.091
0.275102
## bathrooms9                                NA         NA      NA
NA
## bathrooms10                       -9.770e+04  6.975e+05  -0.140
0.888611
## balconies2                         1.130e+04  3.042e+04   0.371
0.710331
## balconies3                         3.001e+05  4.346e+04   6.907 5.11e-
12
## balconies4                         4.299e+05  7.559e+04   5.687 1.31e-
08
## balconies5                         7.069e+05  1.735e+05   4.074 4.65e-
05
## balconies6                        -8.177e+05  4.023e+05  -2.033
0.042104
## balconies7                        -1.110e+06  8.434e+05  -1.316
0.188281
## balconies8                        -1.760e+06  1.636e+06  -1.076
0.282072
## balconies9                                NA         NA      NA
NA
```

```
## balconies10                               -4.902e+05  1.159e+06  -0.423
0.672459
## RentOrSaleSale                             5.771e+06  2.905e+04 198.646  < 2e-
16
## Long                                      -2.922e+03  2.991e+03  -0.977
0.328628
## Lat                                        2.701e+03  5.739e+03   0.471
0.637833
## carpetArea_in_sqft                        -1.668e+00  2.821e+00  -0.591
0.554323
##
## (Intercept)                               ***
## sqftPrice                                 ***
## securityDeposit                           ***
## `propertyTypeBuilder Floor Apartment`     ***
## `propertyTypeMultistorey Apartment`       ***
## propertyTypePenthouse
## `propertyTypeResidential House`
## `propertyTypeStudio Apartment`            *
## propertyTypeVilla
## `furnishingSemi-Furnished`                *
## furnishingUnfurnished                     .
## flrNum                                    **
## firstMonthCharges
## facingNorth                               *
## `facingNorth - East`                      ***
## `facingNorth - West`
## facingSouth                               .
## `facingSouth - East`
## `facingSouth -West`
## facingWest
## totalFlrNum                               ***
## cityAgartala                              ***
## cityBangalore                             ***
## cityBhopal                                ***
## cityChandigarh                            ***
## cityChennai                               ***
## cityDehradun                              ***
## cityGandhinagar                           ***
## cityGangtok                               ***
## cityGoa                                   **
## cityHyderabad                             ***
## cityJaipur                                ***
## cityKolkata                               ***
## cityLucknow                               ***
## cityMumbai
## `cityNew Delhi`                           ***
## `cityNew-Delhi`                           ***
## cityPatna                                 **
## cityRaipur                                ***
```

```
## bedrooms2                                .
## bedrooms3                                ***
## bedrooms4                                ***
## bedrooms5                                ***
## bedrooms6                                ***
## bedrooms7                                ***
## bedrooms8                                **
## bedrooms9
## bedrooms10                               *
## bathrooms2                               ***
## bathrooms3                               ***
## bathrooms4                               ***
## bathrooms5                               ***
## bathrooms6                               ***
## bathrooms7                               ***
## bathrooms8
## bathrooms9
## bathrooms10
## balconies2
## balconies3                               ***
## balconies4                               ***
## balconies5                               ***
## balconies6                               *
## balconies7
## balconies8
## balconies9
## balconies10
## RentOrSaleSale                           ***
## Long
## Lat
## carpetArea_in_sqft
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1630000 on 20417 degrees of freedom
## Multiple R-squared:  0.7744, Adjusted R-squared:  0.7737
## F-statistic:  1078 on 65 and 20417 DF,  p-value: < 2.2e-16
```

Making predictions on test data

```
predictions <- predict(initial_model, newdata = test_set)

## Warning in predict.lm(initial_model, newdata = test_set): prediction from
## rank-deficient fit; attr(*, "non-estim") has doubtful cases

dim(test_set)

## [1] 5118   70
```

Backward Elimination

Performing backward elimination on the training set. Commenting the step function as it takes time to knit the file.

Final_model <- step(initial_model, direction = "backward")

```
Final_model_with_selected_variables <- lm(exactPrice ~ sqftPrice +
securityDeposit + `propertyTypeBuilder Floor Apartment` +
                      `propertyTypeMultistorey Apartment` +
propertyTypePenthouse +
                      `propertyTypeStudio Apartment` + `furnishingSemi-
Furnished` +
                      furnishingUnfurnished + flrNum + facingNorth +
`facingNorth - East` +
                      facingSouth + totalFlrNum + cityAgartala +
cityBangalore +
                      cityBhopal + cityChandigarh + cityChennai +
cityDehradun +
                      cityGandhinagar + cityGangtok + cityGoa + cityHyderabad
+
                      cityJaipur + cityKolkata + cityLucknow + cityMumbai +
`cityNew Delhi` +
                      `cityNew-Delhi` + cityPatna + cityRaipur + bedrooms2 +
bedrooms3 +
                      bedrooms4 + bedrooms5 + bedrooms6 + bedrooms7 +
bedrooms8 +
                      bedrooms10 + bathrooms2 + bathrooms3 + bathrooms4 +
bathrooms5 +
                      bathrooms6 + bathrooms7 + balconies3 + balconies4 +
balconies5 +
                      balconies6 + RentOrSaleSale, data = train_set)

summary(Final_model_with_selected_variables)

##
## Call:
## lm(formula = exactPrice ~ sqftPrice + securityDeposit +
`propertyTypeBuilder Floor Apartment` +
##     `propertyTypeMultistorey Apartment` + propertyTypePenthouse +
##     `propertyTypeStudio Apartment` + `furnishingSemi-Furnished` +
##     furnishingUnfurnished + flrNum + facingNorth + `facingNorth - East` +
##     facingSouth + totalFlrNum + cityAgartala + cityBangalore +
##     cityBhopal + cityChandigarh + cityChennai + cityDehradun +
##     cityGandhinagar + cityGangtok + cityGoa + cityHyderabad +
##     cityJaipur + cityKolkata + cityLucknow + cityMumbai + `cityNew Delhi`
+
##     `cityNew-Delhi` + cityPatna + cityRaipur + bedrooms2 + bedrooms3 +
##     bedrooms4 + bedrooms5 + bedrooms6 + bedrooms7 + bedrooms8 +
##     bedrooms10 + bathrooms2 + bathrooms3 + bathrooms4 + bathrooms5 +
```

```
##     bathrooms6 + bathrooms7 + balconies3 + balconies4 + balconies5 +
##     balconies6 + RentOrSaleSale, data = train_set)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7420132  -800909   -70866   458548  8571449
##
## Coefficients:
##                                     Estimate Std. Error t value
Pr(>|t|)
## (Intercept)                        5.523e+06  1.631e+06   3.387
0.000709
## sqftPrice                          9.952e-01  1.657e-01   6.008 1.91e-
09
## securityDeposit                   -1.292e+00  1.160e-01 -11.138  < 2e-
16
## `propertyTypeBuilder Floor Apartment` -5.546e+05  4.114e+04 -13.482  < 2e-
16
## `propertyTypeMultistorey Apartment`   -2.495e+05  3.154e+04  -7.911 2.68e-
15
## propertyTypePenthouse               3.088e+05  1.859e+05   1.660
0.096841
## `propertyTypeStudio Apartment`     -2.799e+05  1.194e+05  -2.345
0.019021
## `furnishingSemi-Furnished`          6.572e+04  3.249e+04   2.023
0.043074
## furnishingUnfurnished              -6.145e+04  3.325e+04  -1.848
0.064622
## flrNum                             -1.676e+04  5.787e+03  -2.897
0.003775
## facingNorth                         1.134e+05  4.795e+04   2.364
0.018067
## `facingNorth - East`                2.086e+05  4.386e+04   4.756 1.99e-
06
## facingSouth                         1.490e+05  8.564e+04   1.740
0.081945
## totalFlrNum                         2.422e+04  3.409e+03   7.104 1.25e-
12
## cityAgartala                       -5.560e+06  1.677e+06  -3.315
0.000919
## cityBangalore                      -5.712e+06  1.631e+06  -3.502
0.000462
## cityBhopal                         -6.938e+06  1.631e+06  -4.255 2.10e-
05
## cityChandigarh                     -5.618e+06  1.631e+06  -3.445
0.000572
## cityChennai                        -5.734e+06  1.631e+06  -3.516
0.000439
## cityDehradun                       -5.541e+06  1.631e+06  -3.398
0.000681
```

```
## cityGandhinagar                    -6.107e+06  1.631e+06  -3.744
0.000181
## cityGangtok                        -6.349e+06  1.886e+06  -3.366
0.000764
## cityGoa                            -5.190e+06  1.631e+06  -3.183
0.001462
## cityHyderabad                      -5.947e+06  1.631e+06  -3.646
0.000267
## cityJaipur                         -6.173e+06  1.631e+06  -3.786
0.000154
## cityKolkata                        -5.755e+06  1.631e+06  -3.529
0.000418
## cityLucknow                        -5.875e+06  1.631e+06  -3.603
0.000316
## cityMumbai                         -2.435e+06  1.634e+06  -1.490
0.136226
## `cityNew Delhi`                    -5.536e+06  1.662e+06  -3.330
0.000870
## `cityNew-Delhi`                    -5.849e+06  1.631e+06  -3.585
0.000337
## cityPatna                          -5.151e+06  1.631e+06  -3.159
0.001584
## cityRaipur                         -6.839e+06  1.631e+06  -4.194 2.76e-
05
## bedrooms2                           8.182e+04  4.701e+04   1.741
0.081763
## bedrooms3                           5.023e+05  5.638e+04   8.910  < 2e-
16
## bedrooms4                           7.750e+05  7.834e+04   9.893  < 2e-
16
## bedrooms5                           9.843e+05  1.278e+05   7.705 1.37e-
14
## bedrooms6                           9.419e+05  1.657e+05   5.684 1.33e-
08
## bedrooms7                           1.626e+06  2.571e+05   6.324 2.61e-
10
## bedrooms8                           1.075e+06  3.086e+05   3.483
0.000496
## bedrooms10                          1.153e+06  4.459e+05   2.585
0.009740
## bathrooms2                          1.916e+05  4.173e+04   4.592 4.41e-
06
## bathrooms3                          6.534e+05  5.555e+04  11.763  < 2e-
16
## bathrooms4                          1.000e+06  7.977e+04  12.538  < 2e-
16
## bathrooms5                          7.940e+05  1.267e+05   6.265 3.80e-
10
## bathrooms6                          9.488e+05  2.011e+05   4.717 2.41e-
06
```

```
## bathrooms7                                9.552e+05  2.927e+05   3.263
0.001104
## balconies3                                2.998e+05  4.120e+04   7.276 3.56e-
13
## balconies4                                4.291e+05  7.389e+04   5.807 6.47e-
09
## balconies5                                7.218e+05  1.722e+05   4.191 2.79e-
05
## balconies6                               -7.962e+05  4.002e+05  -1.989
0.046665
## RentOrSaleSale                            5.772e+06  2.891e+04 199.605  < 2e-
16
##
## (Intercept)                              ***
## sqftPrice                                ***
## securityDeposit                          ***
## `propertyTypeBuilder Floor Apartment`    ***
## `propertyTypeMultistorey Apartment`      ***
## propertyTypePenthouse                    .
## `propertyTypeStudio Apartment`           *
## `furnishingSemi-Furnished`               *
## furnishingUnfurnished                    .
## flrNum                                   **
## facingNorth                              *
## `facingNorth - East`                     ***
## facingSouth                              .
## totalFlrNum                              ***
## cityAgartala                             ***
## cityBangalore                            ***
## cityBhopal                               ***
## cityChandigarh                           ***
## cityChennai                              ***
## cityDehradun                             ***
## cityGandhinagar                          ***
## cityGangtok                              ***
## cityGoa                                  **
## cityHyderabad                            ***
## cityJaipur                               ***
## cityKolkata                              ***
## cityLucknow                              ***
## cityMumbai
## `cityNew Delhi`                          ***
## `cityNew-Delhi`                          ***
## cityPatna                                **
## cityRaipur                               ***
## bedrooms2                                .
## bedrooms3                                ***
## bedrooms4                                ***
## bedrooms5                                ***
## bedrooms6                                ***
```

```
## bedrooms7                                  ***
## bedrooms8                                  ***
## bedrooms10                                 **
## bathrooms2                                 ***
## bathrooms3                                 ***
## bathrooms4                                 ***
## bathrooms5                                 ***
## bathrooms6                                 ***
## bathrooms7                                 **
## balconies3                                 ***
## balconies4                                 ***
## balconies5                                 ***
## balconies6                                 *
## RentOrSaleSale                             ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1630000 on 20432 degrees of freedom
## Multiple R-squared:  0.7743, Adjusted R-squared:  0.7737
## F-statistic:  1402 on 50 and 20432 DF,  p-value: < 2.2e-16
```

Making predictions on the test set

```
predictions <- predict(Final_model_with_selected_variables, newdata =
test_set)
```

Calculating Mean Squared Error (MSE)

```
mse <- mean((test_set$exactPrice - predictions)^2)
cat("Mean Squared Error (MSE):", mse, "\n")

## Mean Squared Error (MSE): 2.63445e+12
```

Calculating Mean Absolute Error (MAE)

```
mae <- mean(abs(test_set$exactPrice - predictions))
cat("Mean Absolute Error (MAE):", mae, "\n")

## Mean Absolute Error (MAE): 1058098
```

Question B)

Significance of Coefficients:

Variables like sqftPrice, securityDeposit, propertyTypeBuilder Floor Apartment, propertyTypeMultistorey Apartment, and many city-specific dummies (e.g., cityAgartala, cityBangalore, cityBhopal) have p-values well below the 0.05 threshold, indicating strong evidence against the null hypothesis. This suggests these predictors have a significant impact on exactPrice.

Predictors with Marginal Significance:

Some variables like propertyTypePenthouse, furnishingUnfurnished, and facingSouth have p-values close to 0.05, indicating weaker evidence of their effect on exactPrice.

Insignificant Predictors:

cityMumbai has a p-value greater than 0.05, suggesting that its coefficient is not significantly different from zero at the conventional 5% significance level.

Choice of Backward Elimination:

The backward elimination method was selected because of its iterative approach to model simplification. Starting with a complete model that contains all potential predictors, it gradually removes the least significant variable (the one with the greatest p-value over the selected significance level) until all remaining variables contribute significantly to the model. This approach helps in the reduction of model complexity while maintaining the model's explanatory power.

Overall Significance of the Regression Fit:

The F-statistic is extremely high, and the associated p-value is less than 2.2e-16, indicating that the regression model is statistically significant. This means that there is strong evidence against the null hypothesis that all coefficients are zero.

The Multiple R-squared value of 0.7743 suggests that approximately 77.43% of the variability in exactPrice is explained by the model. The Adjusted R-squared, which accounts for the number of predictors in the model, is also high at 0.7737, indicating a good fit.

Significant Predictors at the 0.05 Level:

Numerous predictors are significantly different from zero at the 0.05 level, including sqftPrice, securityDeposit, various propertyType categories, flrNum, totalFlrNum, multiple city categories, and specific counts of bedrooms, bathrooms, and balconies. The RentOrSaleSale variable, indicating properties for sale, shows an extremely significant positive coefficient, suggesting that properties for sale are priced significantly higher than properties for rent, holding all else constant.

Conclusion:

The model provides valuable insights into factors affecting property prices. The significant predictors highlight the importance of property size (sqftPrice, bedrooms, bathrooms), location (city categories), property type, and additional features like balconies in determining the exact price. The choice of backward elimination has ensured that the final model is simplified yet retains variables that are most influential in predicting exactPrice. The high R-squared values indicate a robust model fit, making this model a reliable tool for understanding and predicting property prices based on the given predictors.