Airbnb Dataset Analysis

Abdul

2025-03-31

```r
# Load the dataset
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)
```

```r
## Task 1: Structure of the Dataset

# Load the dataset
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)

# Print structure of the dataset
str(airbnb_data)
```

```
## ’data.frame':    102599 obs. of  26 variables:
##  $ id                            : int  1001254 1002102 1002403 1002755 1003689 1004098 1004650 1005
##  $ NAME                          : chr  "Clean & quiet apt home by the park" "Skylit Midtown Castle"
##  $ host.id                       : num  8.00e+10 5.23e+10 7.88e+10 8.51e+10 9.20e+10 ...
##  $ host_identity_verified        : chr  "unconfirmed" "verified" "" "unconfirmed" ...
##  $ host.name                     : chr  "Madaline" "Jenna" "Elise" "Garry" ...
##  $ neighbourhood.group           : chr  "Brooklyn" "Manhattan" "Manhattan" "Brooklyn" ...
##  $ neighbourhood                 : chr  "Kensington" "Midtown" "Harlem" "Clinton Hill" ...
##  $ lat                           : num  40.6 40.8 40.8 40.7 40.8 ...
##  $ long                          : num  -74 -74 -73.9 -74 -73.9 ...
##  $ country                       : chr  "United States" "United States" "United States" "United State
##  $ country.code                  : chr  "US" "US" "US" "US" ...
##  $ instant_bookable              : logi  FALSE FALSE TRUE TRUE FALSE TRUE ...
##  $ cancellation_policy           : chr  "strict" "moderate" "flexible" "moderate" ...
##  $ room.type                     : chr  "Private room" "Entire home/apt" "Private room" "Entire home
##  $ Construction.year             : int  2020 2007 2005 2005 2009 2013 2015 2009 2005 2015 ...
##  $ price                         : chr  "$966 " "$142 " "$620 " "$368 " ...
##  $ service.fee                   : chr  "$193 " "$28 " "$124 " "$74 " ...
##  $ minimum.nights                : int  10 30 3 30 10 3 45 45 2 2 ...
##  $ number.of.reviews             : int  9 45 0 270 9 74 49 49 430 118 ...
##  $ last.review                   : chr  "10/19/2021" "5/21/2022" "" "7/5/2019" ...
##  $ reviews.per.month             : num  0.21 0.38 NA 4.64 0.1 0.59 0.4 0.4 3.47 0.99 ...
##  $ review.rate.number            : int  4 4 5 4 3 3 5 5 3 5 ...
##  $ calculated.host.listings.count: int  6 2 1 1 1 1 1 1 1 1 ...
##  $ availability.365              : int  286 228 352 322 289 374 224 219 180 375 ...
##  $ house_rules                   : chr  "Clean up and treat the home the way you’d like your home to
##  $ license                       : chr  "" "" "" "" ...
```

AirbnbData_Analysis

Trupti

2025-03-31

```r
# Set global options for code chunks

knitr::opts_chunk$set(echo = TRUE)


library(readr)

#Load the dataset using read_csv from readr
data <- read_csv("C:/Users/dtrup/OneDrive/Desktop/Airbnb_Open_Data.csv")
```

```
## Warning: One or more parsing issues, call ‘problems()‘ on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 102599 Columns: 26
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (13): NAME, host_identity_verified, host name, neighbourhood group, neig...
## dbl (11): id, host id, lat, long, Construction year, minimum nights, number ...
## lgl  (2): instant_bookable, license
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```r
# List the variables (columns) in the dataset
print(colnames(data))
```

```
##  [1] "id"                             "NAME"
##  [3] "host id"                        "host_identity_verified"
##  [5] "host name"                      "neighbourhood group"
##  [7] "neighbourhood"                  "lat"
##  [9] "long"                           "country"
## [11] "country code"                   "instant_bookable"
## [13] "cancellation_policy"            "room type"
## [15] "Construction year"              "price"
## [17] "service fee"                    "minimum nights"
## [19] "number of reviews"              "last review"
## [21] "reviews per month"              "review rate number"
## [23] "calculated host listings count" "availability 365"
## [25] "house_rules"                    "license"
```

```r
# Print the top 15 rows of the dataset
top_15_rows <- as.data.frame(head(data, 15))
print(top_15_rows)
```

```
##         id                                             NAME     host id
## 1  1001254               Clean & quiet apt home by the park 80014485718
## 2  1002102                            Skylit Midtown Castle 52335172823
## 3  1002403              THE VILLAGE OF HARLEM....NEW YORK ! 78829239556
## 4  1002755                                             <NA> 85098326012
## 5  1003689 Entire Apt: Spacious Studio/Loft by central park 92037596077
## 6  1004098        Large Cozy 1 BR Apartment In Midtown East 45498551794
## 7  1004650                                  BlissArtsSpace! 61300605564
## 8  1005202                                  BlissArtsSpace! 90821839709
## 9  1005754                  Large Furnished Room Near B’way 79384379533
## 10 1006307               Cozy Clean Guest Room - Family Apt 75527839483
## 11 1006859               Cute & Cozy Lower East Side 1 bdrm  1280143094
## 12 1007411                 Beautiful 1br on Upper West Side 18824631834
## 13 1007964                  Central Manhattan/near Broadway 88136055909
## 14 1008516   Lovely Room 1, Garden, Best Area, Legal rental 26802410424
## 15 1009068 Wonderful Guest Bedroom in Manhattan for SINGLES 88920244552
##    host_identity_verified host name neighbourhood group      neighbourhood
## 1             unconfirmed  Madaline            Brooklyn         Kensington
## 2                verified     Jenna           Manhattan            Midtown
## 3                    <NA>     Elise           Manhattan             Harlem
## 4             unconfirmed     Garry            Brooklyn       Clinton Hill
## 5                verified    Lyndon           Manhattan        East Harlem
## 6                verified  Michelle           Manhattan        Murray Hill
## 7                    <NA>   Alberta            Brooklyn Bedford-Stuyvesant
## 8             unconfirmed      Emma            Brooklyn Bedford-Stuyvesant
## 9                verified    Evelyn           Manhattan     Hell’s Kitchen
## 10            unconfirmed      Carl           Manhattan    Upper West Side
## 11               verified   Miranda           Manhattan          Chinatown
## 12               verified      Alan           Manhattan    Upper West Side
## 13               verified      <NA>           Manhattan     Hell’s Kitchen
## 14               verified     Darcy             brookln        South Slope
## 15               verified  Leonardo           Manhattan    Upper West Side
##         lat      long       country country code instant_bookable
## 1  40.64749 -73.97237 United States           US            FALSE
## 2  40.75362 -73.98377 United States           US            FALSE
## 3  40.80902 -73.94190 United States           US             TRUE
## 4  40.68514 -73.95976 United States           US             TRUE
## 5  40.79851 -73.94399 United States           US            FALSE
## 6  40.74767 -73.97500 United States           US             TRUE
## 7  40.68688 -73.95596 United States           US            FALSE
## 8  40.68688 -73.95596 United States           US            FALSE
## 9  40.76489 -73.98493 United States           US             TRUE
## 10 40.80178 -73.96723 United States           US            FALSE
## 11 40.71344 -73.99037 United States           US            FALSE
## 12 40.80316 -73.96545 United States           US             TRUE
## 13 40.76076 -73.98867 United States           US            FALSE
## 14 40.66829 -73.98779 United States           US             TRUE
## 15 40.79826 -73.96113 United States           US            FALSE
##    cancellation_policy       room type Construction year  price service fee
## 1               strict    Private room              2020   $966        $193
## 2             moderate Entire home/apt              2007   $142         $28
## 3             flexible    Private room              2005   $620        $124
## 4             moderate Entire home/apt              2005   $368         $74
## 5             moderate Entire home/apt              2009   $204         $41
## 6             flexible Entire home/apt              2013   $577        $115
## 7             moderate    Private room              2015    $71         $14
## 8             moderate    Private room              2009 $1,060        $212
## 9               strict    Private room              2005 $1,018        $204
## 10              strict    Private room              2015   $291         $58
## 11            flexible Entire home/apt              2004   $319         $64
## 12            flexible Entire home/apt              2008   $606        $121
## 13              strict    Private room              2008   $714        $143
## 14            moderate    Private room              2010   $580        $116
## 15            flexible    Private room              2019   $149         $30
##    minimum nights number of reviews last review reviews per month
## 1              10                 9  10/19/2021              0.21
## 2              30                45   5/21/2022              0.38
## 3               3                 0        <NA>                NA
## 4              30               270    7/5/2019              4.64
## 5              10                 9  11/19/2018              0.10
## 6               3                74   6/22/2019              0.59
## 7              45                49   10/5/2017              0.40
## 8              45                49   10/5/2017              0.40
## 9               2               430   6/24/2019              3.47
## 10              2               118   7/21/2017              0.99
## 11              1               160    6/9/2019              1.33
## 12              5                53   6/22/2019              0.43
## 13              2               188   6/23/2019              1.50
## 14              4               167   6/24/2019              1.34
## 15              2               113    7/5/2019              0.91
##    review rate number calculated host listings count availability 365
## 1                   4                              6              286
## 2                   4                              2              228
## 3                   5                              1              352
## 4                   4                              1              322
## 5                   3                              1              289
## 6                   3                              1              374
## 7                   5                              1              224
## 8                   5                              1              219
## 9                   3                              1              180
## 10                  5                              1              375
## 11                  3                              4                1
## 12                  4                              1              163
## 13                  4                              1              258
## 14                  4                              3               47
## 15                  3                              1               68
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8  House Guidelines for our BnB We are delighted to welcome you.  Check in Sun - Thurs by 8PM and Fri
## 9
## 10
## 11
## 12
## 13
## 14
## 15
##    license
## 1       NA
## 2       NA
## 3       NA
## 4       NA
## 5       NA
## 6       NA
## 7       NA
## 8       NA
## 9       NA
## 10      NA
## 11      NA
## 12      NA
## 13      NA
## 14      NA
## 15      NA
```

~~~
---
Title: "Task 4 - AirBnb Data Analysis"
Author: Akshat Singh
Date: "2025-03-31"
Output:
  pdf_document: default
    df_print: paged
output: pdf_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## R Markdown

This is Task 4 of the AirBnb Data Analysis our group has been conducting. In this task
we had to choose a variable of our choice from the dataset and define a user defined
function for the same variable.

We have used the Construction Year as our variable to print information of all the
propertied listed and constructed after 2020.

Below is the code for the same Task.

```{r cars}
getwd()
Airbnb_Open_Data <- read.csv("/cloud/project/Airbnb_Open_Data.csv", header=FALSE)
recent_properties <- as.numeric(Airbnb_Open_Data$V15)
recent_properties <- Airbnb_Open_Data[Airbnb_Open_Data$V15 > 2020, ]
print(recent_properties)

```


**Title: "Task 5 - AirBnb Data Analysis"**

In this task we were supposed to filter data regarding cancellation policy based on the
location of the property.

For this output we had set a condition for Brooklyn and hence it showed us the date of
the property as well as the cancellation policy of the particular BnB.

The code for the function is given below.

```{r cars2}
install.packages("dplyr")
library(dplyr)
Airbnb_Open_Data$V15 <- as.numeric(Airbnb_Open_Data$V15)
brooklyn_new_properties <- Airbnb_Open_Data %>%
  filter(Airbnb_Open_Data$V15 > 2010, Airbnb_Open_Data$V6 == "Brooklyn") %>%
  select(V15,V13)
print(brooklyn_new_properties)
```

## Including Plots
~~~

~~~~
```{r pressure, echo=FALSE}
plot(pressure)
```
~~~~

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Task 6

Abdul Ali

2025-04-01

```
## Task 2: Identify Variables and Reshape the Data

# Load the necessary libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Set working directory and load the dataset
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)

# Convert price to numeric
airbnb_data$price <- gsub("\\$", "", airbnb_data$price)  # Remove $
airbnb_data$price <- gsub(",", "", airbnb_data$price)    # Remove commas
airbnb_data$price <- as.numeric(airbnb_data$price)       # Convert to number

# Select variables and remove NA
selected_data <- airbnb_data %>%
  select(price, Construction.year, room.type, number.of.reviews, country) %>%
  na.omit()

# Reshape data - average price by room type and country
reshaped_data <- selected_data %>%
  group_by(room.type, country) %>%
  summarise(Average_Price = mean(price, na.rm = TRUE), .groups = "drop")

# Display the first few rows
head(reshaped_data)
```

```
## # A tibble: 6 x 3
##   room.type       country         Average_Price
##   <chr>           <chr>                     <dbl>
## 1 Entire home/apt ""                         604.
## 2 Entire home/apt "United States"            625.
## 3 Hotel room      ""                         859
## 4 Hotel room      "United States"            663.
## 5 Private room    ""                         621.
## 6 Private room    "United States"            625.
```

~~~
---
title: "AirbnbDataset_Analysis"
author: "Sohini Biswas"
date: "2025-03-30"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

## R Markdown


When you click the **Knit** button a document will be generated that includes both content
as well as the output of any embedded R code chunks within the document. You can embed an
R code chunk like this:

```{r cars}
summary(cars)
```

## Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```


# AUTHOR - SOHINI BISWAS

setwd("C:\\Users\\sohin\\Downloads")

getwd()

airbnb_data <- read.csv("airbnb_Data.csv", check.names = FALSE)

#   TASK 1 : Remove Missing Values in the Dataset
#   Check for missing values in each column
missing_values_per_column <- colSums(is.na(airbnb_data))
print("Missing values per column:")
print(missing_values_per_column)

#   Total missing values in the entire dataset
total_missing_values <- sum(is.na(airbnb_data))
print(paste("Total missing values:", total_missing_values))

#   Remove rows with any missing values
airbnb_cleaned_rows <- na.omit(airbnb_data)

#   Remove columns with any missing values
airbnb_cleaned_cols <- airbnb_data[, colSums(is.na(airbnb_data)) == 0]

#   TASK 2 : Identify and remove Duplicate Values in the Dataset
#   Identify duplicate rows
duplicate_rows <- duplicated(airbnb_data)

#   Count the number of duplicate rows
number_of_duplicate_rows <- sum(duplicate_rows)
print(paste("Number of duplicate rows:", number_of_duplicate_rows))

#   View the duplicate rows
~~~

```
print("Duplicate rows:")
print(airbnb_data[duplicate_rows, ])

#   Remove duplicate rows
airbnb_unique <- airbnb_data[!duplicate_rows, ]

#   Verify removal by counting rows before and after
print(paste("Number of rows before removing duplicates:", nrow(airbnb_data)))
print(paste("Number of rows after removing duplicates:", nrow(airbnb_unique))
```

Airbnb Dataset Analysis

Sonali

2025-03-30

##Introduction

In this assignment, we are performing a data analysis on an Airbnb dataset using R Programming.
The objective of this assignment is to apply basic data manipulation techniques on the dataset.
We will perform two tasks: 1. Reorder the dataset rows in descending order based on a numeric column. 2. Rename selected column names for better clarity.

##Load Libraries and Dataset

In this section, we will load the required library and the Airbnb dataset.
We will use the `dplyr` package for data manipulation and load the dataset from the local folder.

```
# Install dplyr package (only one time)
# install.packages("dplyr")

# Load library
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load dataset
airbnb_data <- read.csv("Airbnb_Sample.csv.csv", check.names = FALSE)

# View column names
colnames(airbnb_data)
```

```
##  [1] "id"                             "NAME"                          
##  [3] "host id"                        "host_identity_verified"        
##  [5] "host name"                      "neighbourhood group"           
##  [7] "neighbourhood"                  "lat"                           
##  [9] "long"                           "country"                       
## [11] "country code"                   "instant_bookable"              
## [13] "cancellation_policy"            "room type"                     
## [15] "Construction year"              "price"                         
## [17] "service fee"                    "minimum nights"                
## [19] "number of reviews"              "last review"                   
## [21] "reviews per month"              "review rate number"            
## [23] "calculated host listings count" "availability 365"              
## [25] "house_rules"                    "license"
```

Task 1: Reorder Multiple Rows in Descending Order

In this task, we will reorder the rows of the Airbnb dataset in descending order based on the `price` column. This will help us to see the listings with the highest prices at the top of the dataset.

```
# Reorder rows in descending order by price
airbnb_sorted <- airbnb_data %>%
  arrange(desc(price))

# View top 10 rows after sorting
head(airbnb_sorted, 10)
```

```
##         id                                             NAME     host id
## 1  1248766             Hancock Town House!-Stuyvesant Mews 17302896128
## 2  1068717                3 Story Town House in Park Slope  9853816495
## 3  1536515 Great room priv/bathrm Eastside location 70's ST 33469116543
## 4  1445938                 Lovely Hell's Kitchen Studio... 91331960557
## 5  1551427              Stay with a Jazz Singer in Harlem! 38298412868
## 6  1201821              Colorful Artistic Williamsburg Apt 59139947513
## 7  1173101             1BR: See Central Park from Terrace!   948625792
## 8  1363645             At a very nice area in the WestSide 27086952528
## 9  1526021               Eastern Parkway Brooklyn 1BR Flat 55368535774
## 10 1511661  Beautiful Room in a Beautiful New NYC Apartment 58156347387
##    host_identity_verified host name neighbourhood group      neighbourhood
## 1             unconfirmed     Lucia                     Bedford-Stuyvesant
## 2                             Darcy            Brooklyn        South Slope
## 3                verified     Perry           Manhattan    Upper East Side
## 4                verified     Scott           Manhattan     Hell's Kitchen
## 5                verified    Taylor           Manhattan             Harlem
## 6                verified    Parker            Brooklyn       Williamsburg
## 7                verified    Howard           Manhattan    Upper West Side
## 8             unconfirmed   Bennett           Manhattan Washington Heights
## 9                verified    Holmes            Brooklyn      Crown Heights
## 10            unconfirmed   Edwards           Manhattan             Harlem
##         lat      long       country country code instant_bookable
## 1  40.68669 -73.91989 United States           US            False
## 2  40.66499 -73.97925 United States           US             True
## 3  40.76850 -73.96034 United States           US            False
## 4  40.76147 -73.99152 United States           US             True
## 5  40.80192 -73.95827 United States           US             True
## 6  40.71125 -73.95613 United States           US             True
## 7  40.77728 -73.97818 United States           US            False
## 8  40.85099 -73.92822 United States           US            False
## 9  40.66847 -73.94875 United States           US            False
## 10 40.83177 -73.95000 United States           US             True
##    cancellation_policy       room type Construction year price service fee
## 1             moderate    Private room              2013 $999        $200
## 2               strict Entire home/apt              2021 $996        $199
## 3             moderate    Private room              2008 $996        $199
## 4               strict Entire home/apt              2016 $993        $199
## 5             flexible    Private room              2022 $993        $199
## 6               strict    Private room              2011  $99         $20
## 7             moderate Entire home/apt              2008 $989        $198
## 8             flexible    Private room              2021 $986        $197
## 9             flexible Entire home/apt              2013 $984        $197
## 10              strict    Private room              2003 $983        $197
##    minimum nights number of reviews last review reviews per month
## 1               1                27   10/8/2017              0.28
## 2               2                16  12/30/2018              0.24
## 3               7                93    6/7/2019              1.06
## 4               2                72   7/31/2018              0.82
## 5               1                 0                            NA
## 6               3                31    3/1/2019              0.31
## 7              13                38   2/15/2019              0.39
## 8              60                 4    7/3/2015              0.06
## 9               5                 1    9/4/2014              0.02
## 10              1                56   7/16/2018              0.64
##    review rate number calculated host listings count availability 365
## 1                   1                              4              181
## 2                  NA                             NA              146
## 3                   3                              1              270
## 4                   4                              1              125
## 5                   3                              1              160
## 6                   2                              1              419
## 7                   3                              1              105
## 8                  NA                              1              355
## 9                   3                              1              212
## 10                  5                              1              425
##
house_rules
## 1
## 2
## 3  House Rules 1. Check-in is 4 pm local time. If the unit is ready earlier, we’ll let you kn
ow. Check-out is normally 11 am local time, but we’d be happy to extend it as long as we don’t h
ave a cleaning scheduled. Just let us know.  2. All bookings require a security deposit of at le
ast $300, which will be refunded within 7 days of your check-out. 3. For security measures we re
quire all guests to provide proof of identification through ID verification on our own website. 
In order to check in we'll need a photo of your ID.  4. Our cancellation policy is as shown on o
ur ad and defined by the site you are booking through.  5. Don’t let $300 go up in smoke. Ther 
e's no smoking allowed in any Flatbook and a $300 fine for breaking this rule.  6. Unfortunately 
we don’t allow pets in any of our apartments. 7. Unless you’re staying in one of our specialty a
partments, we don’t allow parties or excessive noise. 8. If we find the place very messy, we hav
e to charge an extra $40 for every extra hour o
## 4
Respect our place please.
## 5
No smoking in the apartment, in the foyer or outside the building at any time. No pets. No parti
es.
## 6
Maximum of 2 guests staying in a bedroom.  No smoking anywhere in or outside the property.  No p
ets anywhere in or outside the property.  No parties anywhere in or outside the property.  No pa
rking on our property or in the lot.  No frying food.  No extreme temperature settings.   Any br
oken house rules, including evidence of cigarette butts, ashes, odors, stained linens, damaged f
urniture, or missing items will result in loss of deposit. All doors must be locked, and all lig
hts, fans, air conditioners, and heaters should be turned off each time you leave. We kindly ask
that you respect our property and our neighbors in this wonderful community.
## 7
No drugs, no smoking, no pets (our dog is enough pet for us). Make sure the front door is shut s
ecurely when you come and go, as our dog Lucy likes to sneak out and explore when we're not look
ing! No additional guests or visitors without prior permission. If you'd like to have guests ove
r please ask us first.
## 8  Check in/out: We can meet outside of my building at check-in between 4:00pm-6:00pm.  If yo
u can give me an estimated ETA, that would be helpful!  Please leave your keys in the studio whe
n you check-out Thursday afternoon.  I am very flexible with check-out between 12:00pm-3:00pm. G
o up 1 flight of stairs, and it will be the first door to the right.   The door locks automatica
lly, but there is a deadbolt and lock.  You do not have to deadbolt the door every time you leav
```

```
e.  TV – Press and hold the “All On” button to turn on TV Towels are on the top shelf in the bat
hroom, and the bed is all set!   I do have a portable A/C in the second window on the right.  Si
nce I live on the second floor, it does not get too warm, but if you do need it, please remember
to turn it off if you are leaving the building. Trash/recycling days are Sundays and Thursdays.
There’s trash bags on the third shelf in the bathroom closet to the left.  When you are checking
out on Thursday, please take the trash out r
## 9
## 10
NO Smoking
##    license
## 1       NA
## 2       NA
## 3       NA
## 4       NA
## 5       NA
## 6       NA
## 7       NA
## 8       NA
## 9       NA
## 10      NA
```

Task 2: Rename Some of the Column Names

In this task, we will rename some of the column names in the Airbnb dataset to make them clearer.
The following changes will be made: - `NAME` will be renamed to `listing_name` - `price` will be renamed to `listing_price` - `host id` will be renamed to `host_id_number`

```
# Rename selected column names
airbnb_renamed <- airbnb_data %>%
  rename(
    listing_name = NAME,
    listing_price = price,
    host_id_number = `host id`
  )

# View new column names
colnames(airbnb_renamed)
```

```
##  [1] "id"                             "listing_name"
##  [3] "host_id_number"                 "host_identity_verified"
##  [5] "host name"                      "neighbourhood group"
##  [7] "neighbourhood"                  "lat"
##  [9] "long"                           "country"
## [11] "country code"                   "instant_bookable"
## [13] "cancellation_policy"            "room type"
## [15] "Construction year"              "listing_price"
## [17] "service fee"                    "minimum nights"
## [19] "number of reviews"              "last review"
## [21] "reviews per month"              "review rate number"
## [23] "calculated host listings count" "availability 365"
## [25] "house_rules"                    "license"
```

Conclusion

In this assignment, we successfully performed basic data manipulation tasks using R Programming on the Airbnb dataset.
We reordered the dataset rows in descending order based on the `price` column and renamed selected column names to improve clarity.

New calculated variable

Ritvik Pande

2025-03-31

```
# Load the ds
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)
```

```
# Removing the dollar sign and converting to numeric for both columns
airbnb_data$price <- as.numeric(gsub("[$,]", "", airbnb_data$price))
airbnb_data$service.fee <- as.numeric(gsub("[$,]", "", airbnb_data$service.fee))
```

```
# Performing mathematical calculation and adding to new variable
airbnb_data$total_unit_price <- airbnb_data$price + airbnb_data$service.fee
```

```
# Printing the result
# Check the result
head(airbnb_data[, c("price", "service.fee", "total_unit_price")])
```

```
##   price service.fee total_unit_price
## 1   966         193             1159
## 2   142          28              170
## 3   620         124              744
## 4   368          74              442
## 5   204          41              245
## 6   577         115              692
```

Airbnb Data Analysis - Taining set using random number generator engine & Summary statistics

Aneesh

2025-03-31

Introduction:

This document demonstrates how to create a training set from the Airbnb dataset using a random number generator engine and provides summary statistics of the dataset.

Create Training and Test Sets:

```
library(dplyr) # Load dplyr if you haven't already
```

```
## Warning: package 'dplyr' was built under R version 4.4.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load dataset
Airbnb_Open_Data <- read.csv("Airbnb_Open_Data.csv", check.names = FALSE)

# Clean the 'price' column and create 'price_numeric'
Airbnb_Open_Data <- Airbnb_Open_Data %>%
  mutate(price_numeric = gsub("[$,]", "", price),
         price_numeric = as.numeric(price_numeric))

# Determine the training set size (e.g., 70% for training)
set.seed(123) # For reproducibility
train_size <- floor(0.70 * nrow(Airbnb_Open_Data))
train_indices <- sample(1:nrow(Airbnb_Open_Data), train_size)

# Create the training set
```

```
train_data <- Airbnb_Open_Data[train_indices, ]

# Create the test set
test_data <- Airbnb_Open_Data[-train_indices, ]

# Verify the number of rows in each set
cat("Number of rows in training set:", nrow(train_data), "\n")
```

```
## Number of rows in training set: 71819
```

```
cat("Number of rows in test set:", nrow(test_data), "\n")
```

```
## Number of rows in test set: 30780
```

```
cat("Number of rows in test set:", nrow(test_data), "\n")
```

```
## Number of rows in test set: 30780
```

Summary Statistics:

```
##        id                NAME              host id
##  Min.   : 1001254   Length:102599      Min.   :1.236e+08
##  1st Qu.:15085814   Class :character   1st Qu.:2.458e+10
##  Median :29136603   Mode  :character   Median :4.912e+10
##  Mean   :29146235                      Mean   :4.925e+10
##  3rd Qu.:43201198                      3rd Qu.:7.400e+10
##  Max.   :57367417                      Max.   :9.876e+10
##
##  host_identity_verified  host name          neighbourhood group
##  Length:102599          Length:102599      Length:102599
##  Class :character       Class :character   Class :character
##  Mode  :character       Mode  :character   Mode  :character
##
##
##
##
##  neighbourhood           lat             long          country
##  Length:102599      Min.   :40.50   Min.   :-74.25   Length:102599
##  Class :character   1st Qu.:40.69   1st Qu.:-73.98   Class :character
##  Mode  :character   Median :40.72   Median :-73.95   Mode  :character
##                     Mean   :40.73   Mean   :-73.95
##                     3rd Qu.:40.76   3rd Qu.:-73.93
##                     Max.   :40.92   Max.   :-73.71
##                     NA’s   :8       NA’s   :8
##  country code       instant_bookable cancellation_policy  room type
##  Length:102599      Mode :logical    Length:102599       Length:102599
##  Class :character   FALSE:51474      Class :character    Class :character
##  Mode  :character   TRUE :51020      Mode  :character    Mode  :character
##                     NA’s :105
##
##
```

```
##
##  Construction year    price           service fee        minimum nights
##  Min.   :2003      Length:102599      Length:102599      Min.   :-1223.000
##  1st Qu.:2007      Class :character   Class :character   1st Qu.:    2.000
##  Median :2012      Mode  :character   Mode  :character   Median :    3.000
##  Mean   :2012                                            Mean   :    8.136
##  3rd Qu.:2017                                            3rd Qu.:    5.000
##  Max.   :2022                                            Max.   : 5645.000
##  NA’s   :214                                             NA’s   :409
##  number of reviews last review        reviews per month review rate number
##  Min.   :   0.00   Length:102599      Min.   : 0.010    Min.   :1.000
##  1st Qu.:   1.00   Class :character   1st Qu.: 0.220    1st Qu.:2.000
##  Median :   7.00   Mode  :character   Median : 0.740    Median :3.000
##  Mean   :  27.48                      Mean   : 1.374    Mean   :3.279
##  3rd Qu.:  30.00                      3rd Qu.: 2.000    3rd Qu.:4.000
##  Max.   :1024.00                      Max.   :90.000    Max.   :5.000
##  NA’s   :183                          NA’s   :15879     NA’s   :326
##  calculated host listings count availability 365 house_rules
##  Min.   :  1.000                Min.   : -10.0   Length:102599
##  1st Qu.:  1.000                1st Qu.:   3.0   Class :character
##  Median :  1.000                Median :  96.0   Mode  :character
##  Mean   :  7.937                Mean   : 141.1
##  3rd Qu.:  2.000                3rd Qu.: 269.0
##  Max.   :332.000                Max.   :3677.0
##  NA’s   :319                    NA’s   :448
##    license          price_numeric
##  Length:102599      Min.   :  50.0
##  Class :character   1st Qu.: 340.0
##  Mode  :character   Median : 624.0
##                     Mean   : 625.3
##                     3rd Qu.: 913.0
##                     Max.   :1200.0
##                     NA’s   :247
```

Description of Create Training, Test Sets and Summary Statistics :

The code first loads the Airbnb dataset and prepares it for analysis by cleaning the price data. It then splits the data into training and test sets using a random selection process. The training set will be used to build a machine learning model, while the test set will be used to evaluate its performance. Finally, the summary() function is used to display descriptive statistics for all columns in the original dataset, providing an overview of its distribution and characteristics.

Conclusion:

In conclusion, the code performs essential data preparation steps on the Airbnb dataset. The price data is cleaned, and the dataset is split into training and testing sets to facilitate model development and evaluation. The summary statistics offer key insights into the dataset’s characteristics.

Statistical Functions

Ritvik Pande

2025-03-31

```r
# Load the ds
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)
```

```r
# Converting to numeric
minimum.nights_1 <- as.numeric(airbnb_data$minimum.nights)
head(minimum.nights_1)
```

```
## [1] 10 30  3 30 10  3
```

```r
# Mean of minimum nights
mean_1 <- mean(minimum.nights_1, na.rm = TRUE)
print(paste("Mean: ", mean_1))
```

```
## [1] "Mean:  8.13584499461787"
```

```r
# Median of minimum nights
median_1 <- median(minimum.nights_1, na.rm = TRUE)
print(paste("Median: ", median_1))
```

```
## [1] "Median:  3"
```

```r
# Mode of minimum nights
mode_1 <- function(x) {
  uniq_x <- unique(x)
  uniq_x[which.max(tabulate(match(x, uniq_x)))]
}

mode_1 <- mode_1(minimum.nights_1)
print(paste("Mode: ", mode_1))
```

```
## [1] "Mode:  1"
```

```r
# Range of minimum nights
range_1 <- range(minimum.nights_1, na.rm = TRUE)
range_1_value <- diff(range_1)
print(paste("Range: ", range_1_value))
```

```
## [1] "Range:  6868"
```

Airbnb Data Analysis - Scatter Plot & Bar Plot

Deepak Kumar

2025-03-31

Introduction:

This document explores two aspects of the Airbnb dataset: the relationship between the price of listings and the number of reviews they receive, and the comparison of average prices across different room types.

Scatter Plot: Price vs Reviews per Month:

```r
library(ggplot2)
```

```
## Warning: package ’ggplot2’ was built under R version 4.4.3
```

```r
library(dplyr)
```

```
## Warning: package ’dplyr’ was built under R version 4.4.3
```

```
##
## Attaching package: ’dplyr’
```

```
## The following objects are masked from ’package:stats’:
##
##     filter, lag
```

```
## The following objects are masked from ’package:base’:
##
##     intersect, setdiff, setequal, union
```

```r
library(scales)
```

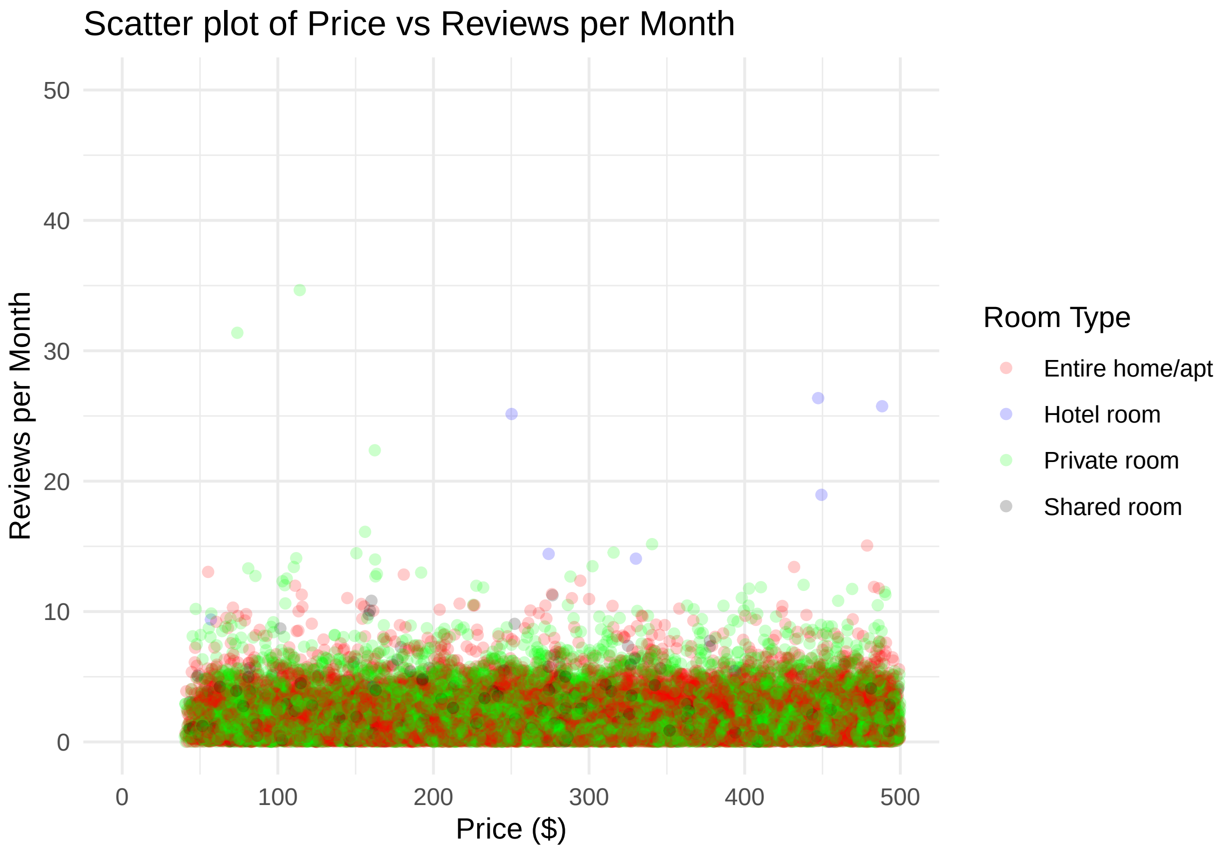
```
## Warning: package ’scales’ was built under R version 4.4.3
```

```r
# Load dataset
Airbnb_Open_Data <- read.csv("Airbnb_Open_Data.csv", check.names = FALSE)

# Clean the 'price' column and create 'price_numeric'
Airbnb_Open_Data <- Airbnb_Open_Data %>%
  mutate(price_numeric = gsub("[$,]", "", price),
         price_numeric = as.numeric(price_numeric))

# create the scatter plot
ggplot(Airbnb_Open_Data, aes(x = price_numeric, y = `reviews per month`, color = `room type`)) +
  geom_jitter(alpha = 0.2, width = 10, height = 1) +
  labs(title = "Scatter plot of Price vs Reviews per Month",
       x = "Price ($)",
       y = "Reviews per Month",
       color = "Room Type") +
  theme_minimal() +
  scale_color_manual(values = c("Entire home/apt" = "red",
                                "Hotel room" = "blue",
                                "Private room" = "green",
                                "Shared room" = "black")) +
  xlim(0, 500) +
  ylim(0, 50)
```

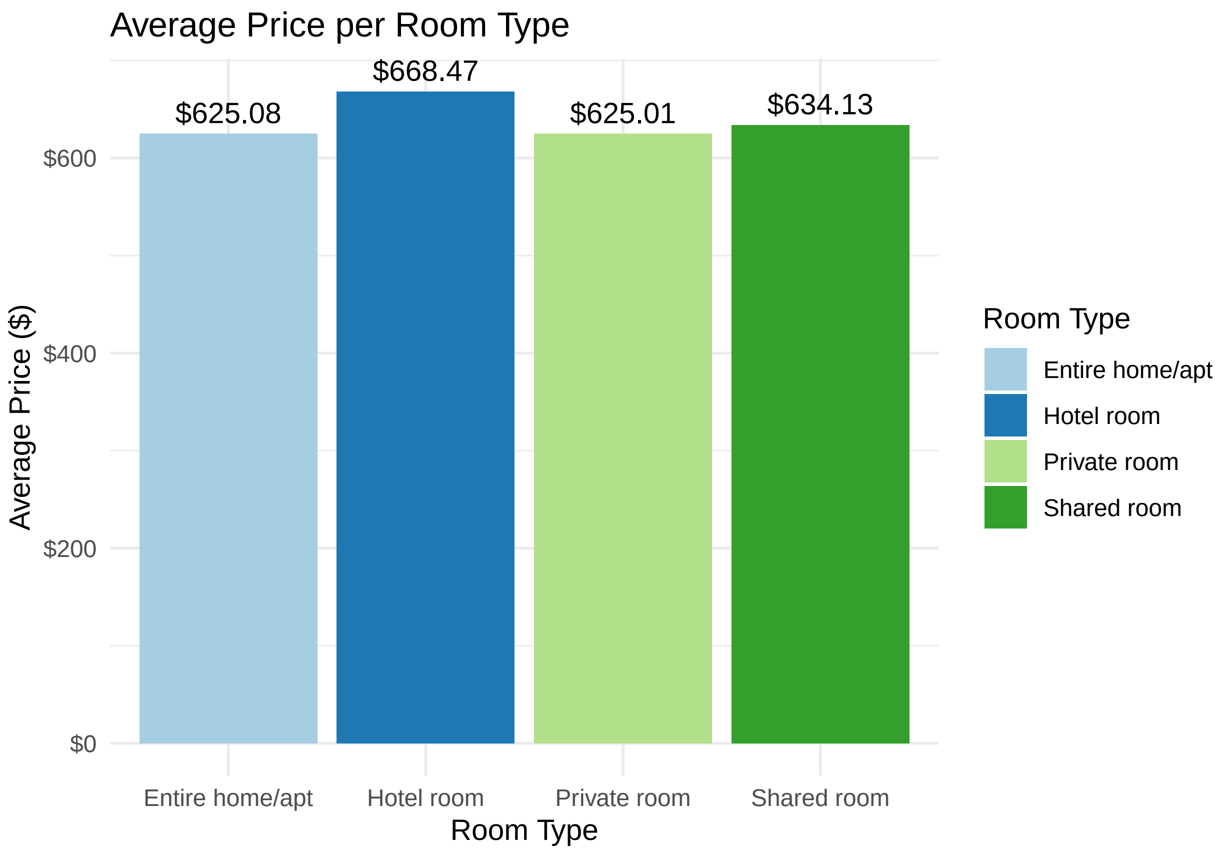
```
## Warning: Removed 75361 rows containing missing values or values outside the scale range
## (‘geom_point()‘).
```



Description of the Scatter Plot:

This scatter plot visualizes the relationship between the price of Airbnb listings and the number of reviews they've received, focusing on listings priced between $0 and $500 and with 0 to 50 reviews per month. Each point represents an individual Airbnb listing, and the color of the point indicates the type of room being offered: red for entire homes/apartments, blue for hotel rooms, green for private rooms, and black for shared rooms. The points are slightly jittered to help distinguish listings with similar price and review counts, especially in densely populated areas of the plot.

Bar Plot: Average Price per Room Type:



Description of the Bar Plot:

This bar plot displays the average price for different types of Airbnb listings: Entire home/apt, Hotel room, Private room, and Shared room. The height of each bar represents the average price in dollars, with the exact average price shown as a label above each bar. The colors differentiate the room types, as indicated by the legend on the right. This visualization allows for a quick comparison of the average listing price across different accommodation types.

Conclusion:

This analysis provides insights into the relationship between listing price and reviews, as well as the average prices for different room types on Airbnb.

Pearson Correlation

Ritvik Pande

2025-03-31

```
# Load the ds
setwd("C:/Users/ritvi/Downloads/Airbnb data")
airbnb_data <- read.csv("Airbnb_Open_Data.csv", stringsAsFactors = FALSE)
```

```
# Converting to numeric
Construction.year_1 <- as.numeric(airbnb_data$Construction.year)
head(Construction.year_1)
```

```
## [1] 2020 2007 2005 2005 2009 2013
```

```
review.rate.number_1 <- as.numeric(airbnb_data$review.rate.number)
head(review.rate.number_1)
```

```
## [1] 4 4 5 4 3 3
```

```
# Pearson correlation between property Construction year and Review ratings
correlation_value <- cor(Construction.year_1, review.rate.number_1, use = "complete.obs", method = "pea

# Print the correlation value
print(paste("Pearson Correlation: ", correlation_value))
```

```
## [1] "Pearson Correlation:  0.00475278600332765"
```