R PROGRAMMIN PROJECT

EXPLORATORY DATA ANALYSIS

REPORT (INFERENCES)

RAASHI BAFNA

PES2UG22CS422

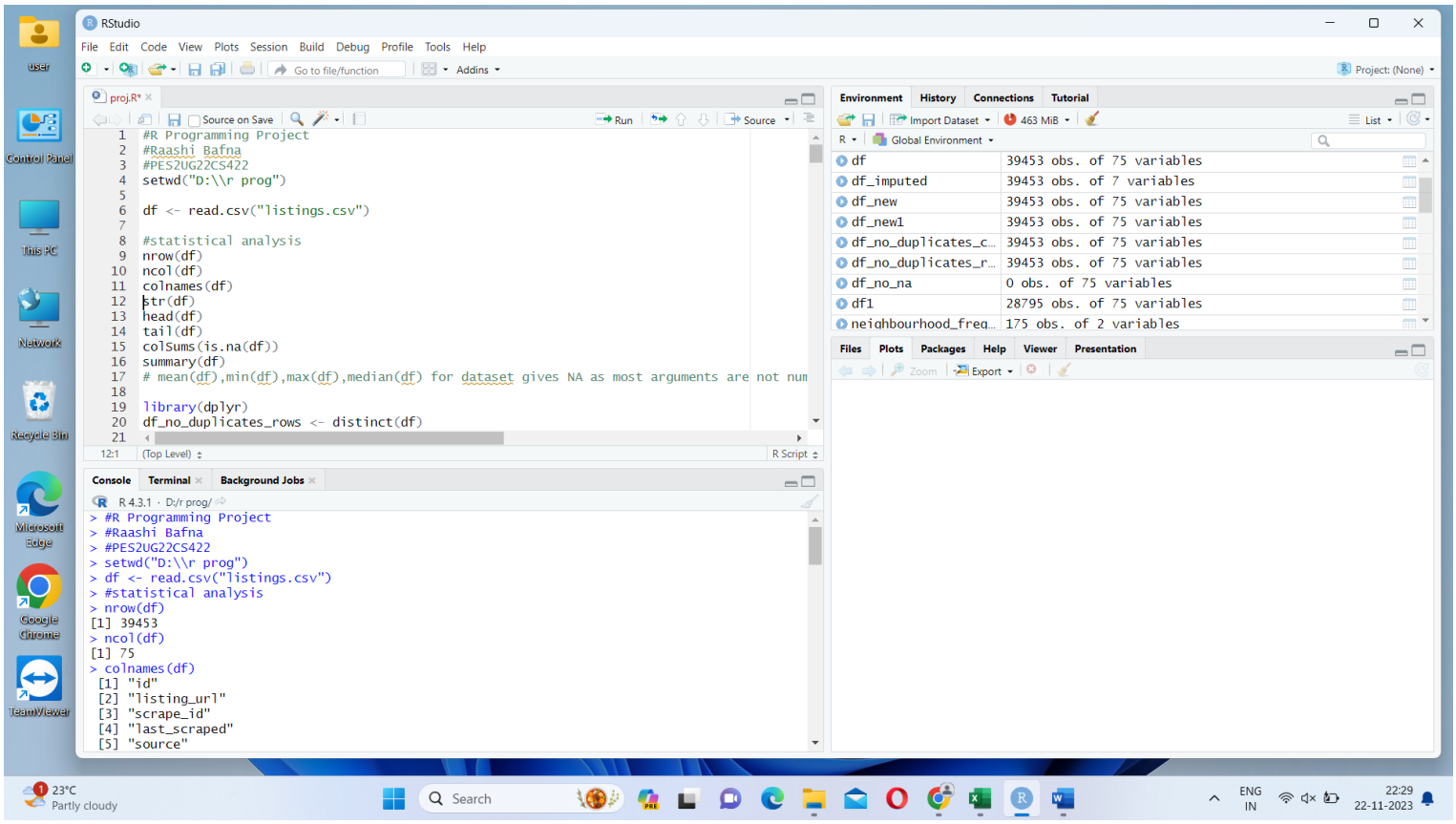
PES2202201371

SEM :3 SEC: G

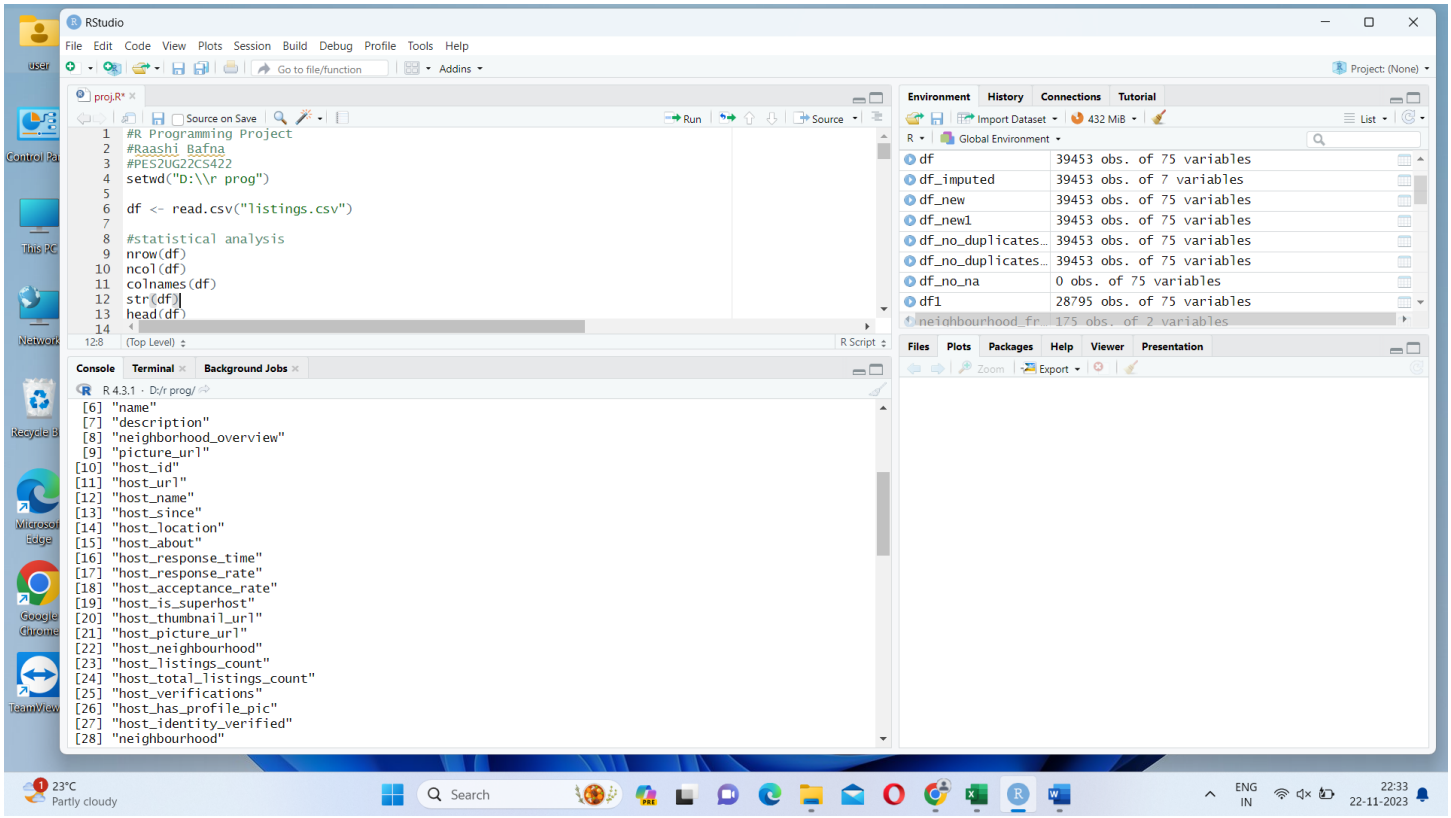
FINAL INSIGHTS ARE MENTIONED AT THE END OF THE FILE(pg.35)

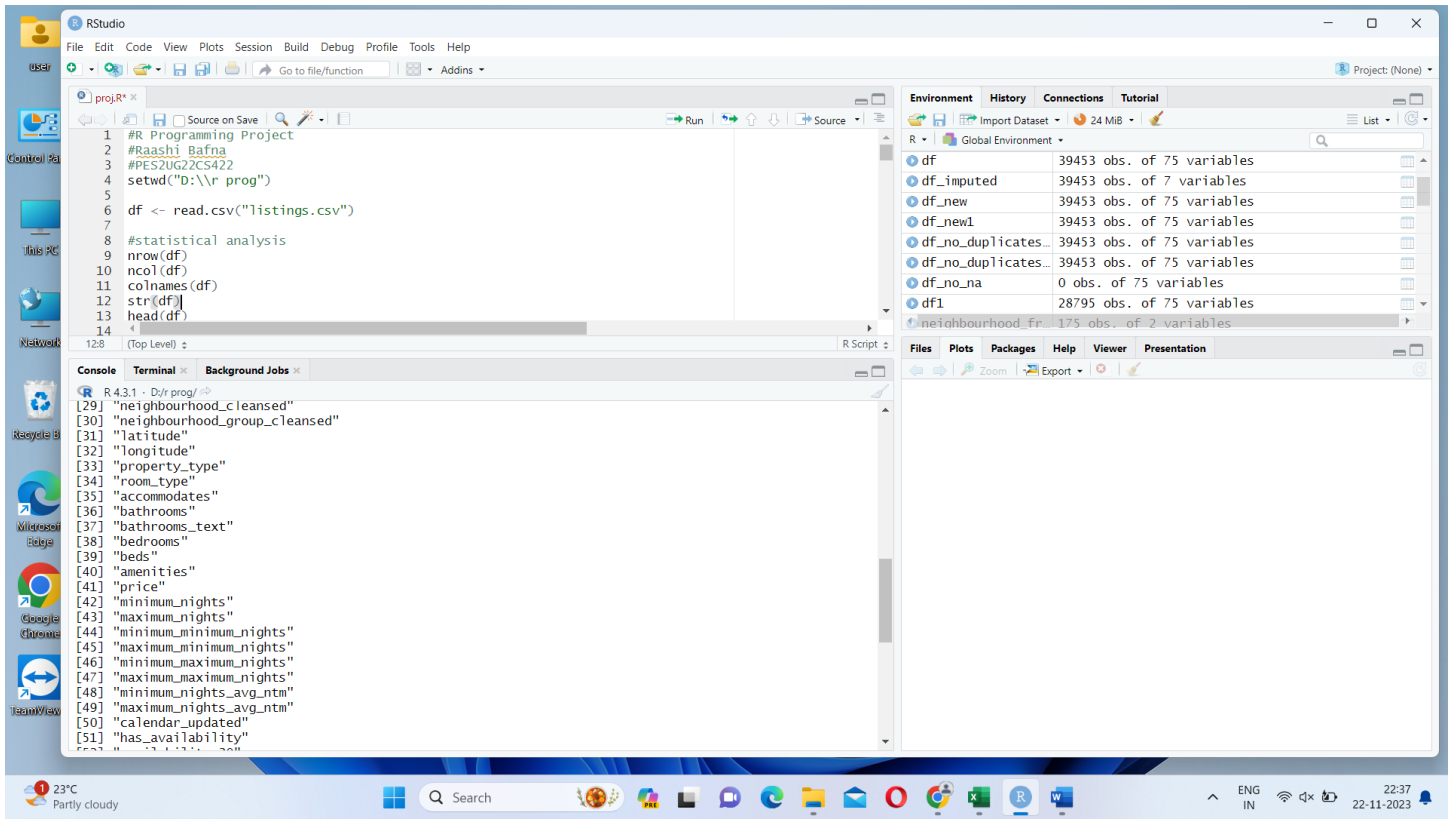
CODE WITH OUTPUT SCREENSHOTS:

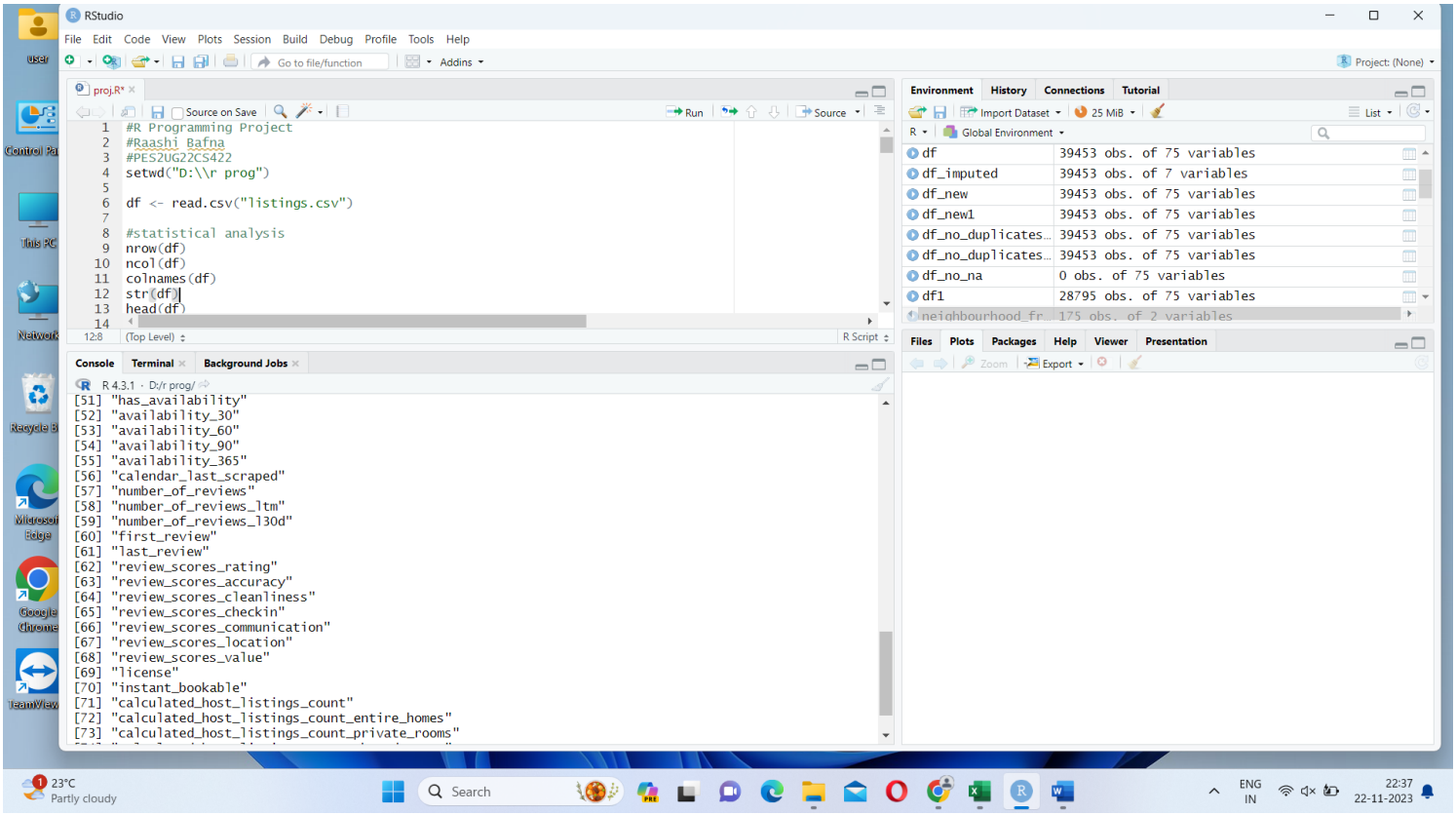
Statistical Analysis

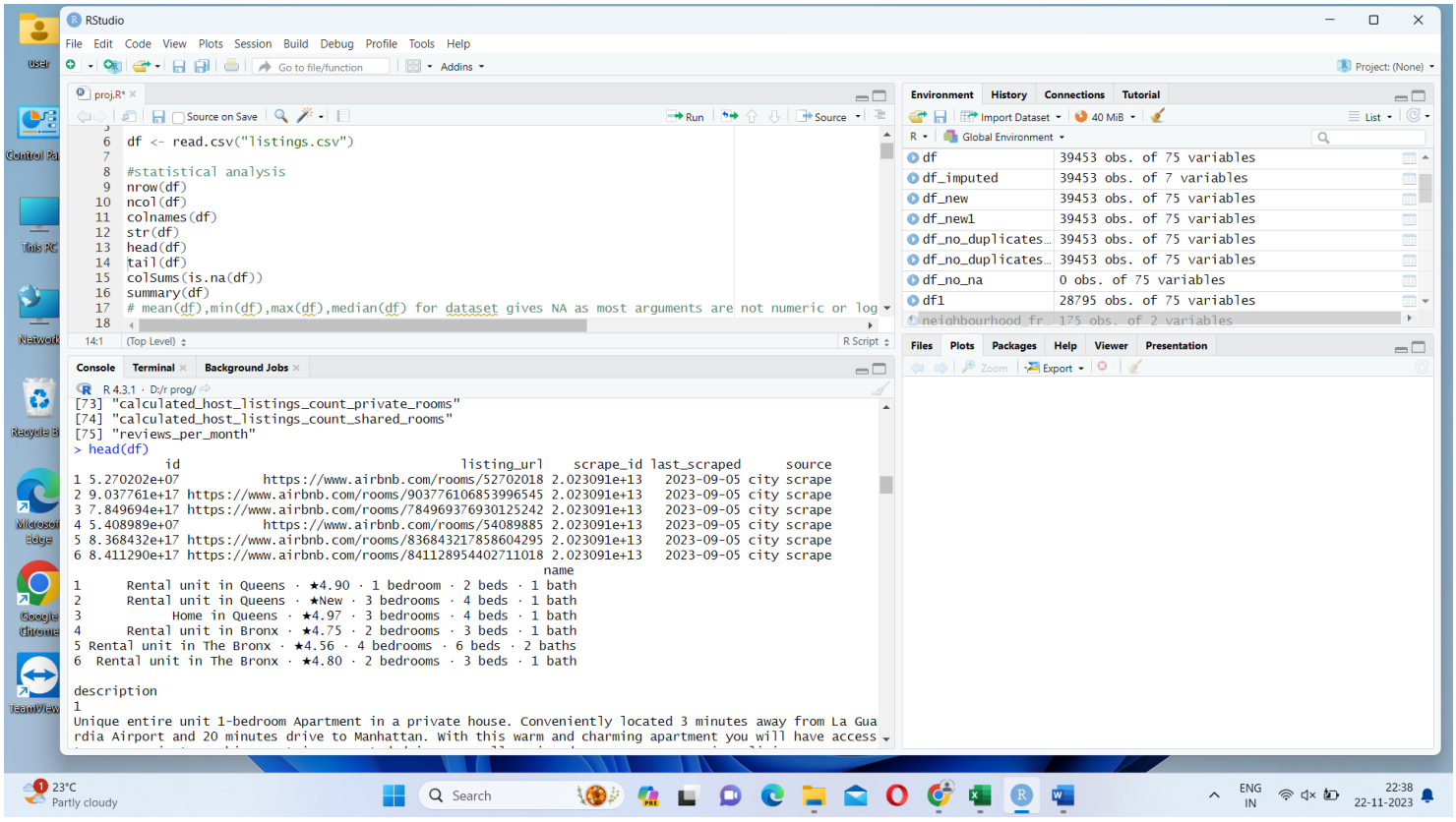


The dataset has 39453 rows and 75 columns

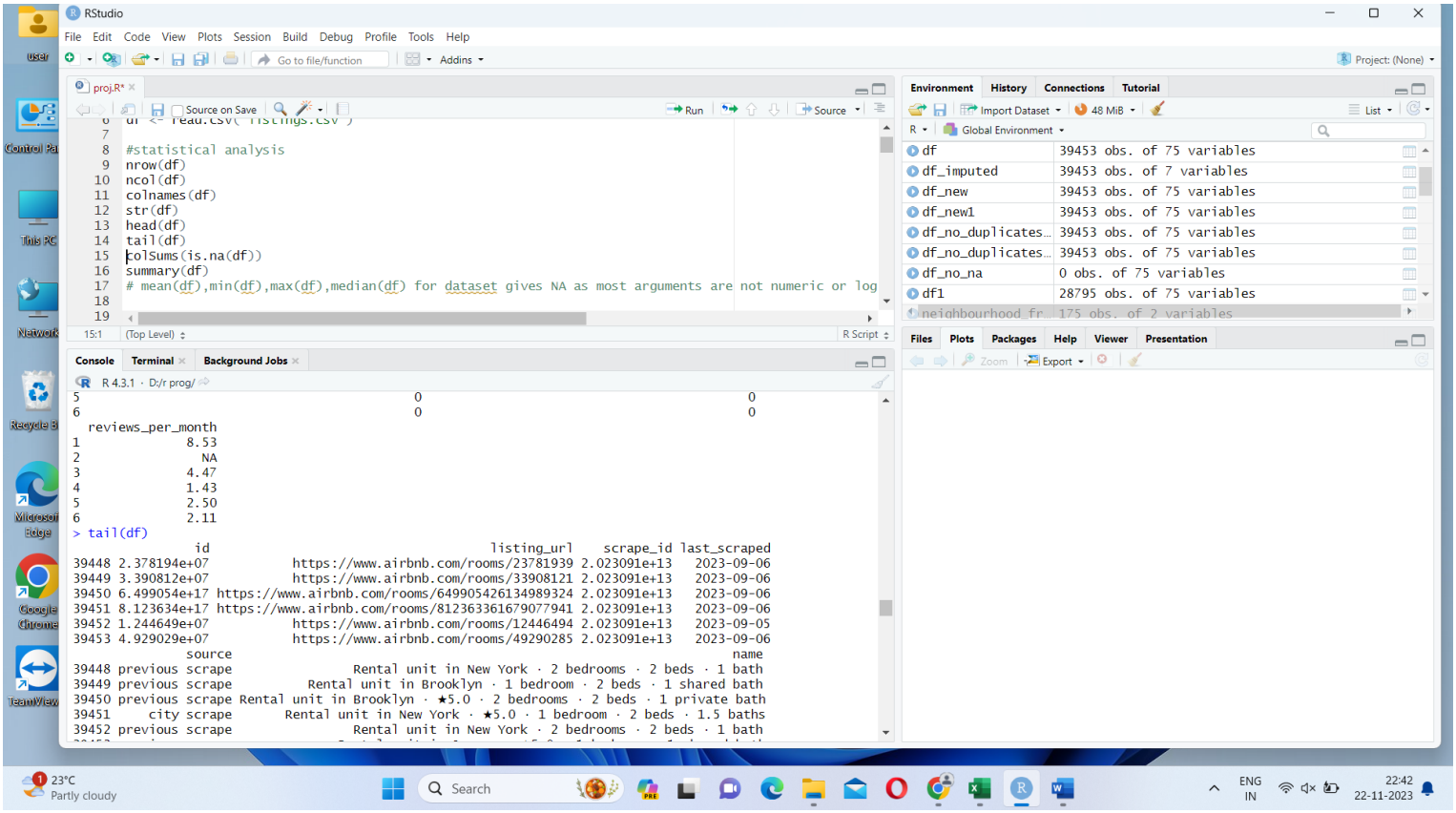

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
1 #R Programming Project
2 #Raashi Bafna
3 #PES2UG22CS422
4 setwd("D:\\r prog")
5
6 df <- read.csv("listings.csv")
7
8 #statistical analysis
9 nrow(df)
10 ncol(df)
11 colnames(df)
12 str(df)
13 head(df)
14
Environment History Connections Tutorial
Import Dataset 432 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[6] "name"
[7] "description"
[8] "neighborhood_overview"
[9] "picture_url"
[10] "host_id"
[11] "host_url"
[12] "host_name"
[13] "host_since"
[14] "host_location"
[15] "host_about"
[16] "host_response_time"
[17] "host_response_rate"
[18] "host_acceptance_rate"
[19] "host_is_superhost"
[20] "host_thumbnail_url"
[21] "host_picture_url"
[22] "host_neighbourhood"
[23] "host_listings_count"
[24] "host_total_listings_count"
[25] "host_verifications"
[26] "host_has_profile_pic"
[27] "host_identity_verified"
[28] "neighbourhood"
23°C Partly cloudy
Search
ENG IN
22:33 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
1 #R Programming Project
2 #Raashi Bafna
3 #PES2UG22CS422
4 setwd("D:\\r prog")
5
6 df <- read.csv("listings.csv")
7
8 #statistical analysis
9 nrow(df)
10 ncol(df)
11 colnames(df)
12 str(df)
13 head(df)
14
Environment History Connections Tutorial
Import Dataset 24 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[29] "neighbourhood_cleansed"
[30] "neighbourhood_group_cleansed"
[31] "latitude"
[32] "longitude"
[33] "property_type"
[34] "room_type"
[35] "accommodates"
[36] "bathrooms"
[37] "bathrooms_text"
[38] "bedrooms"
[39] "beds"
[40] "amenities"
[41] "price"
[42] "minimum_nights"
[43] "maximum_nights"
[44] "minimum_minimum_nights"
[45] "maximum_minimum_nights"
[46] "minimum_maximum_nights"
[47] "maximum_maximum_nights"
[48] "minimum_nights_avg_ntm"
[49] "maximum_nights_avg_ntm"
[50] "calendar_updated"
[51] "has_availability"
23°C Partly cloudy
Search
ENG IN
22:37 22-11-2023



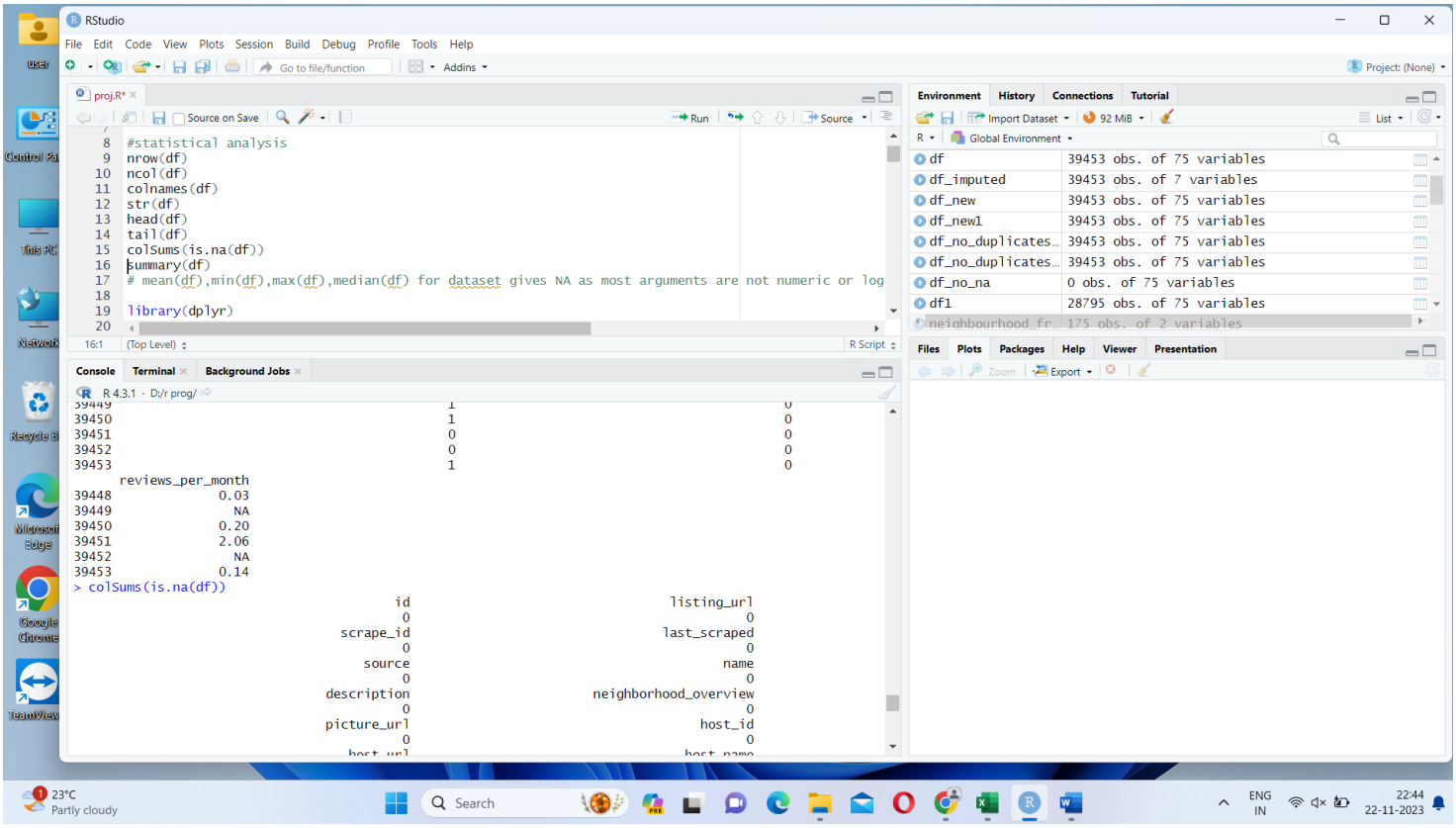


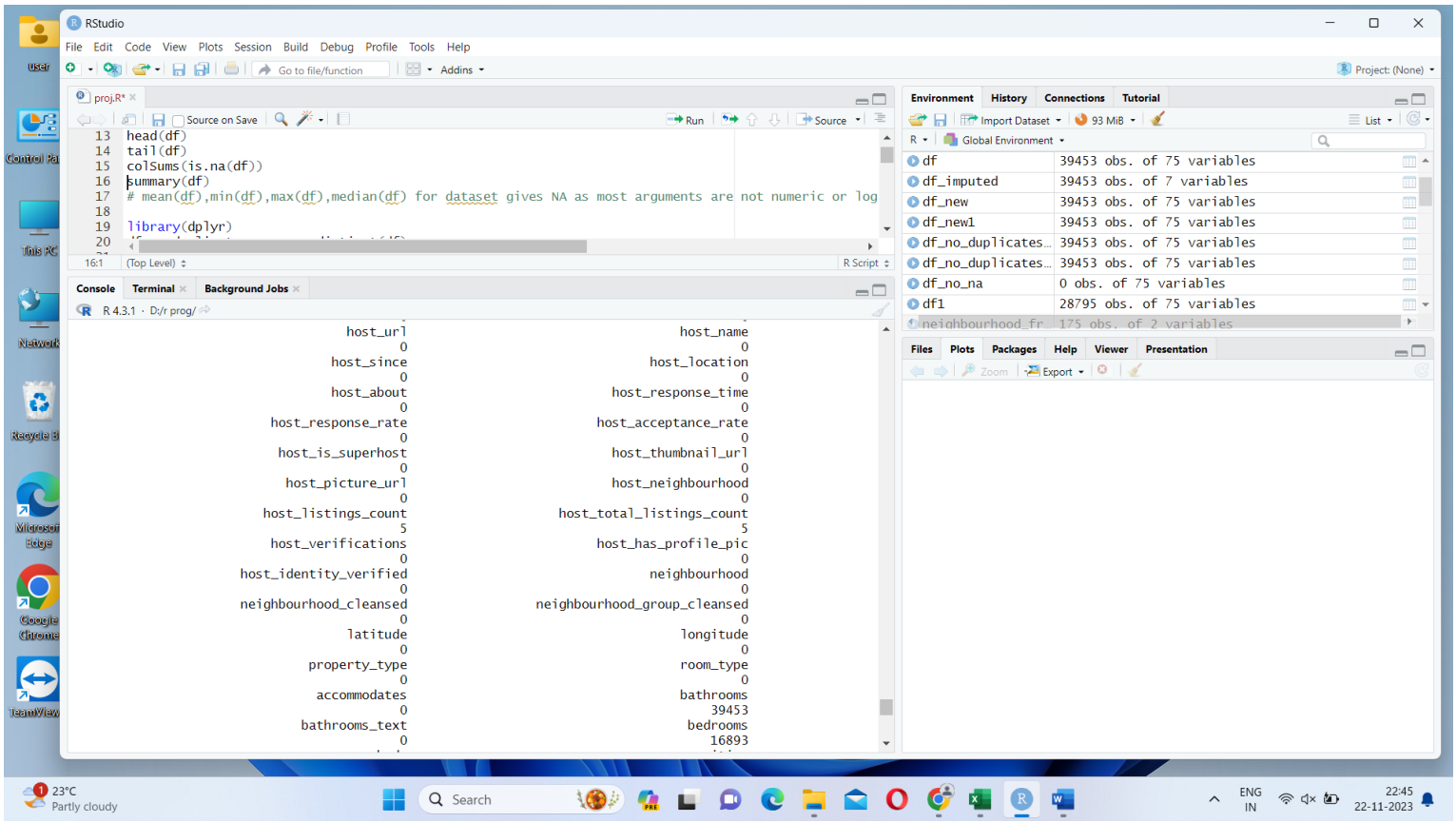
When head(df) is executed the first few lines of the dataset are displayed all screenshots have not been included as it would be very long

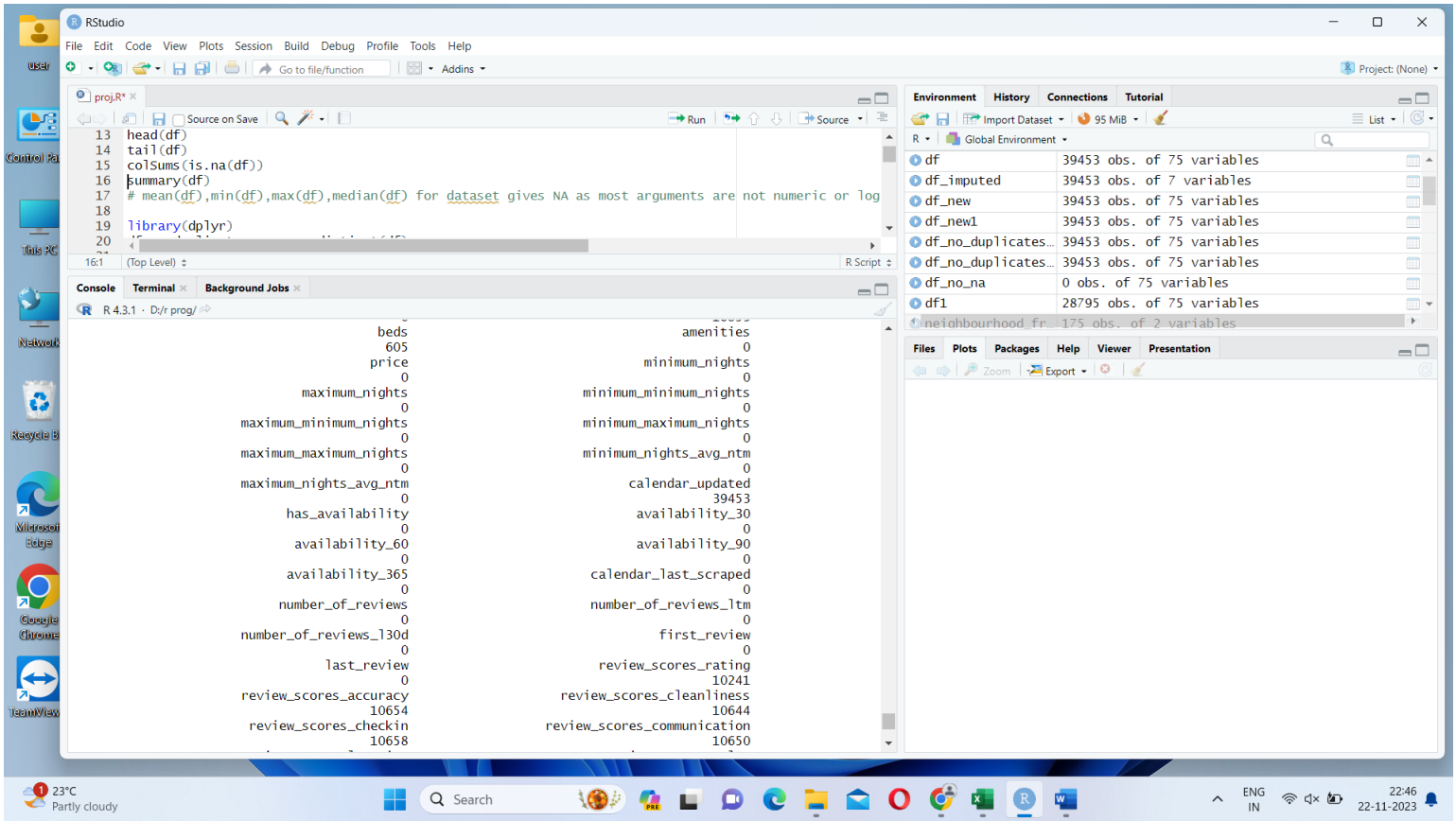


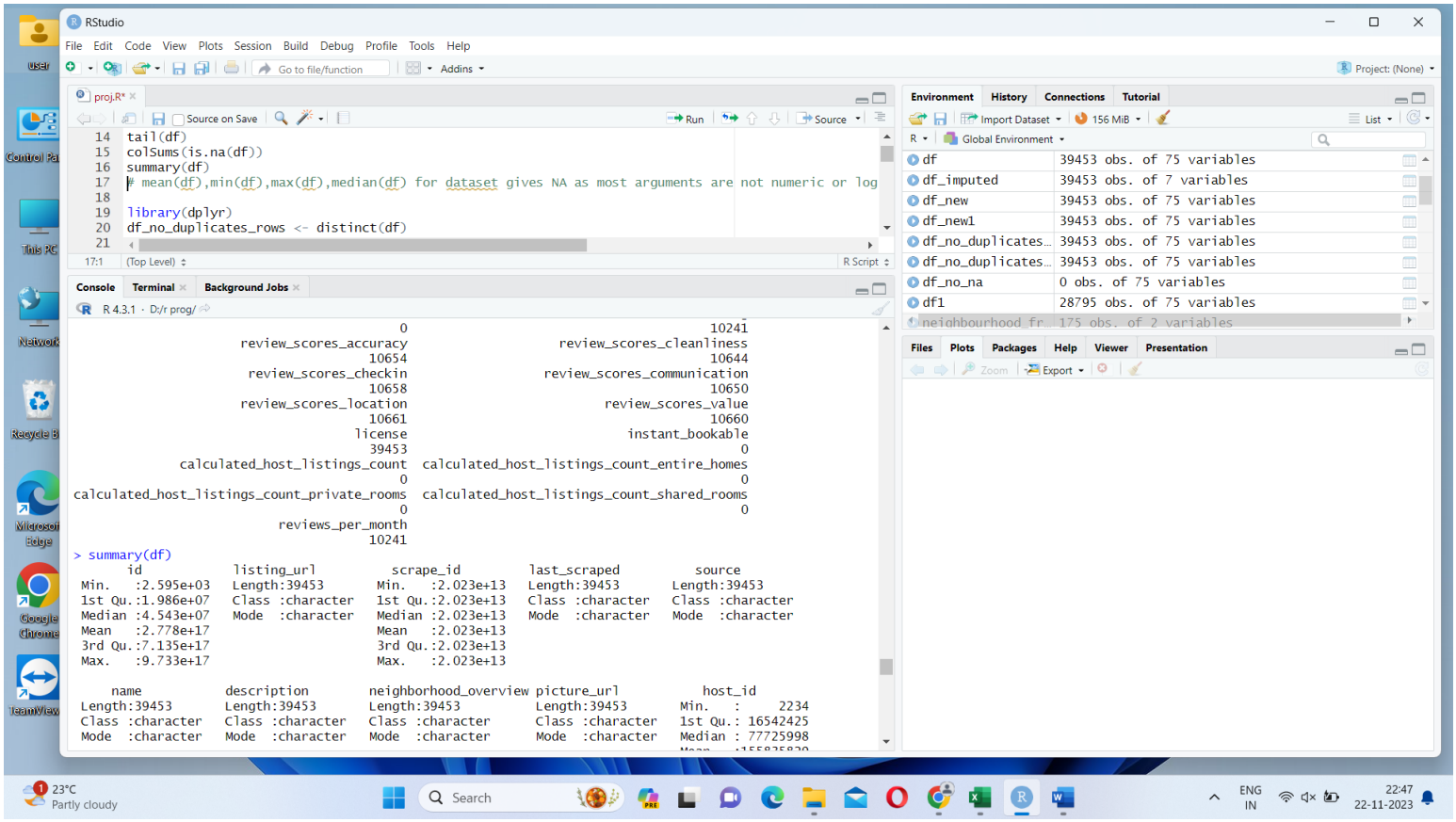
when tail(df) is executed the last few lines of the dataset are displayed

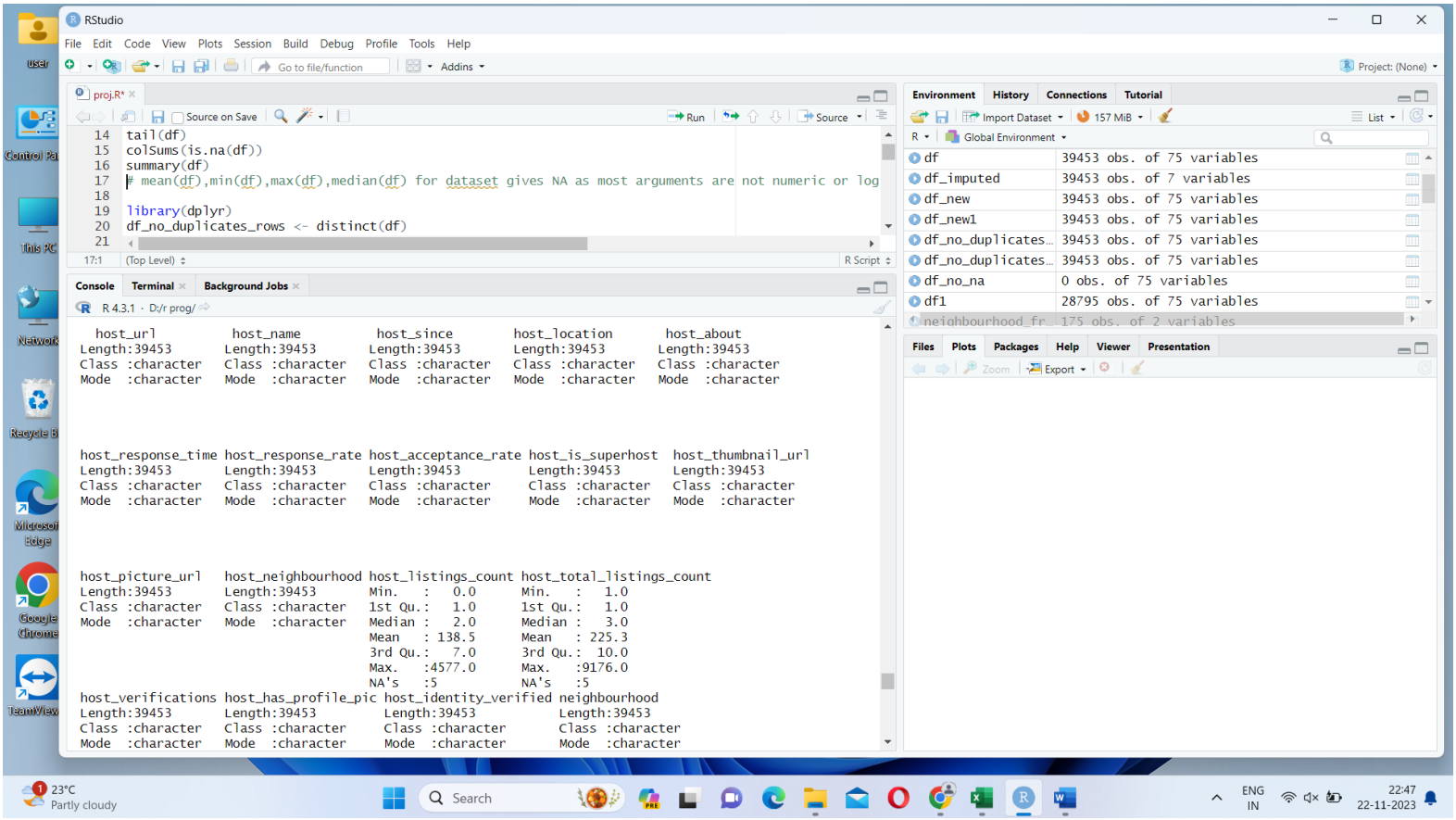
all screenshots have not been included as it would be very long

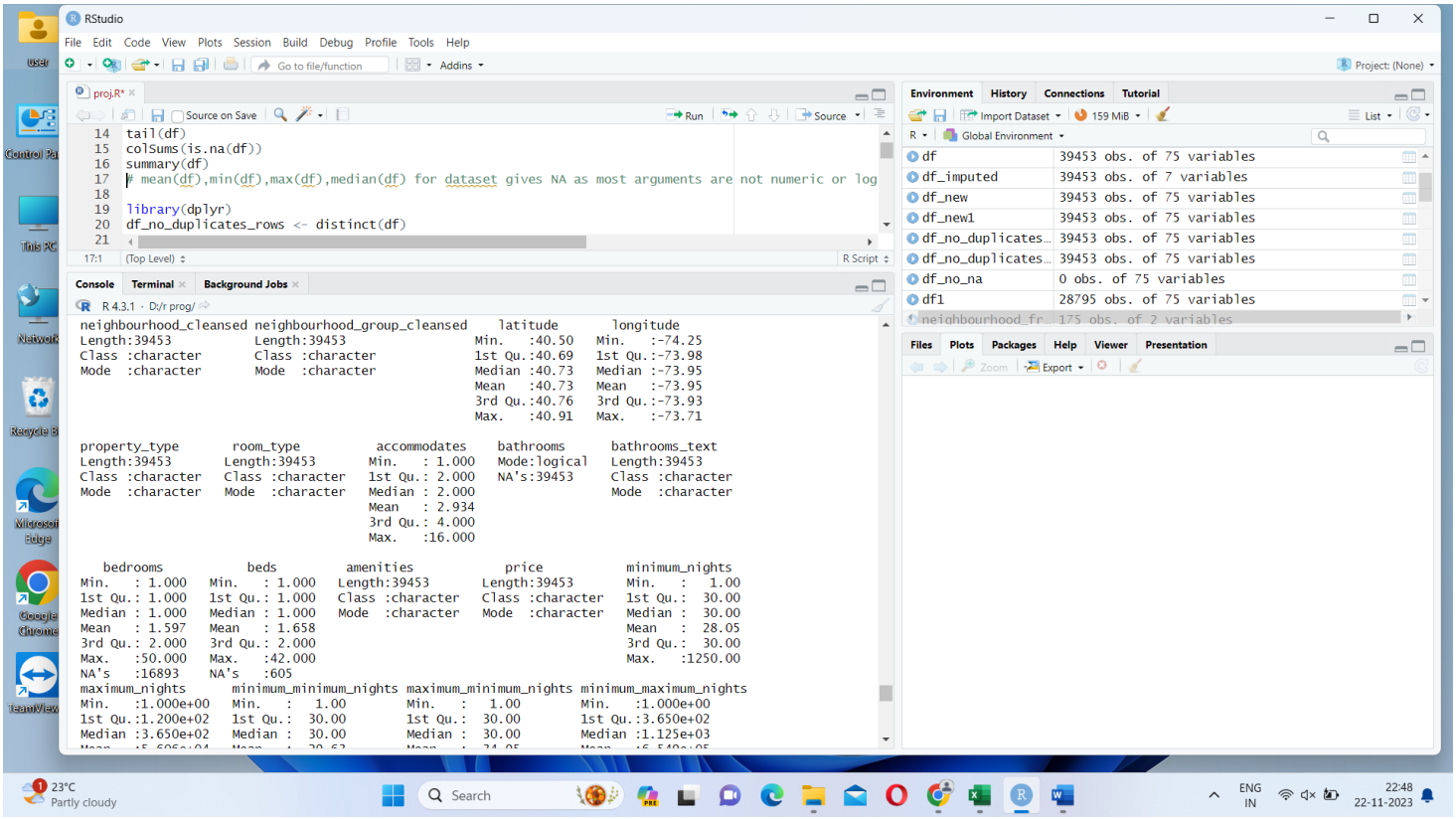


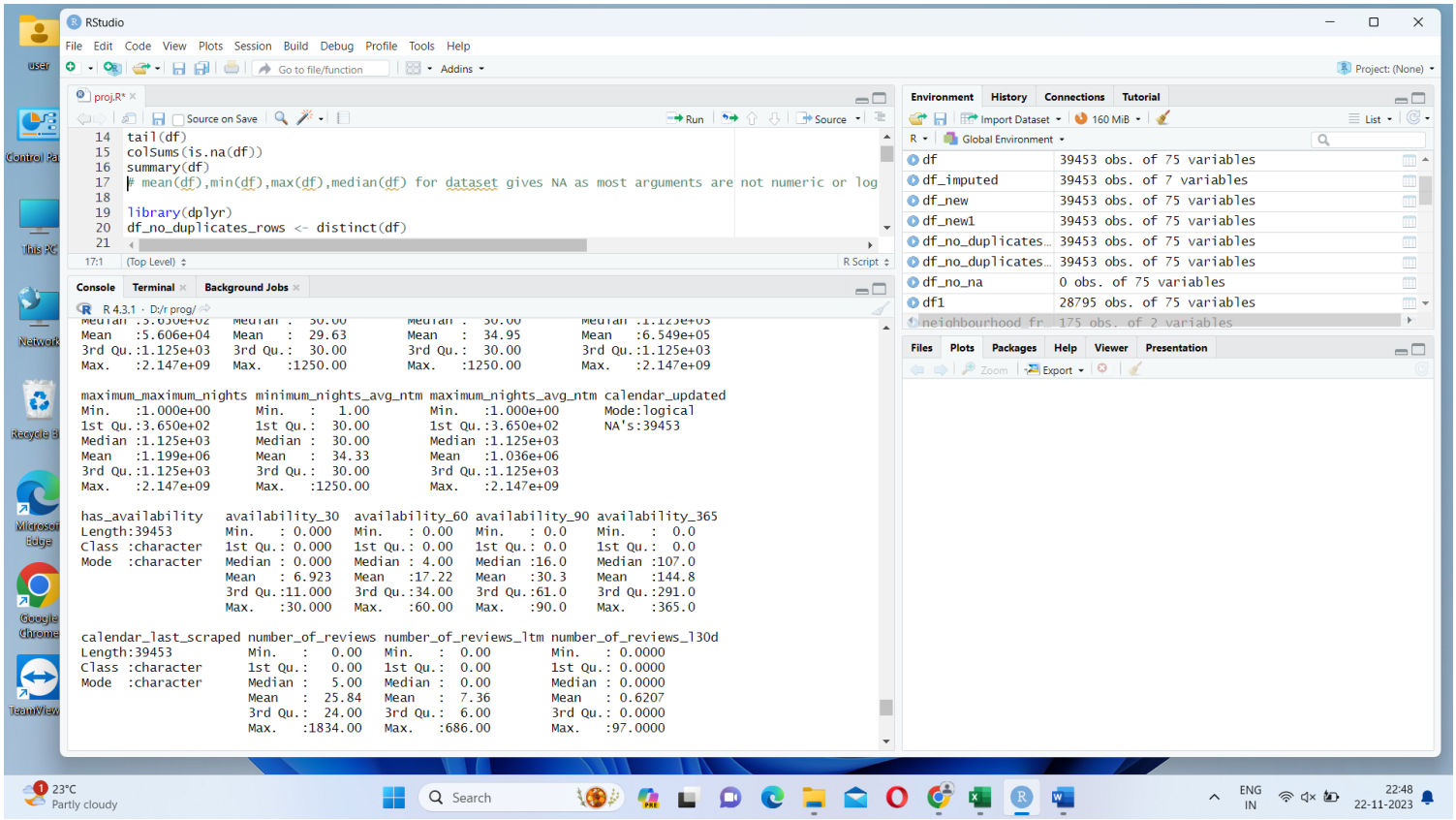

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
13 head(df)
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
host_url 0
host_name 0
host_since 0
host_location 0
host_about 0
host_response_time 0
host_response_rate 0
host_acceptance_rate 0
host_is_superhost 0
host_thumbnail_url 0
host_picture_url 0
host_neighbourhood 0
host_listings_count 5
host_total_listings_count 5
host_verifications 0
host_has_profile_pic 0
host_identity_verified 0
neighbourhood 0
neighbourhood_cleansed 0
neighbourhood_group_cleansed 0
latitude 0
longitude 0
property_type 0
room_type 0
accommodates 0
bathrooms 39453
bathrooms_text 0
bedrooms 16893
Environment History Connections Tutorial
Import Dataset 93 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:45 22-11-2023

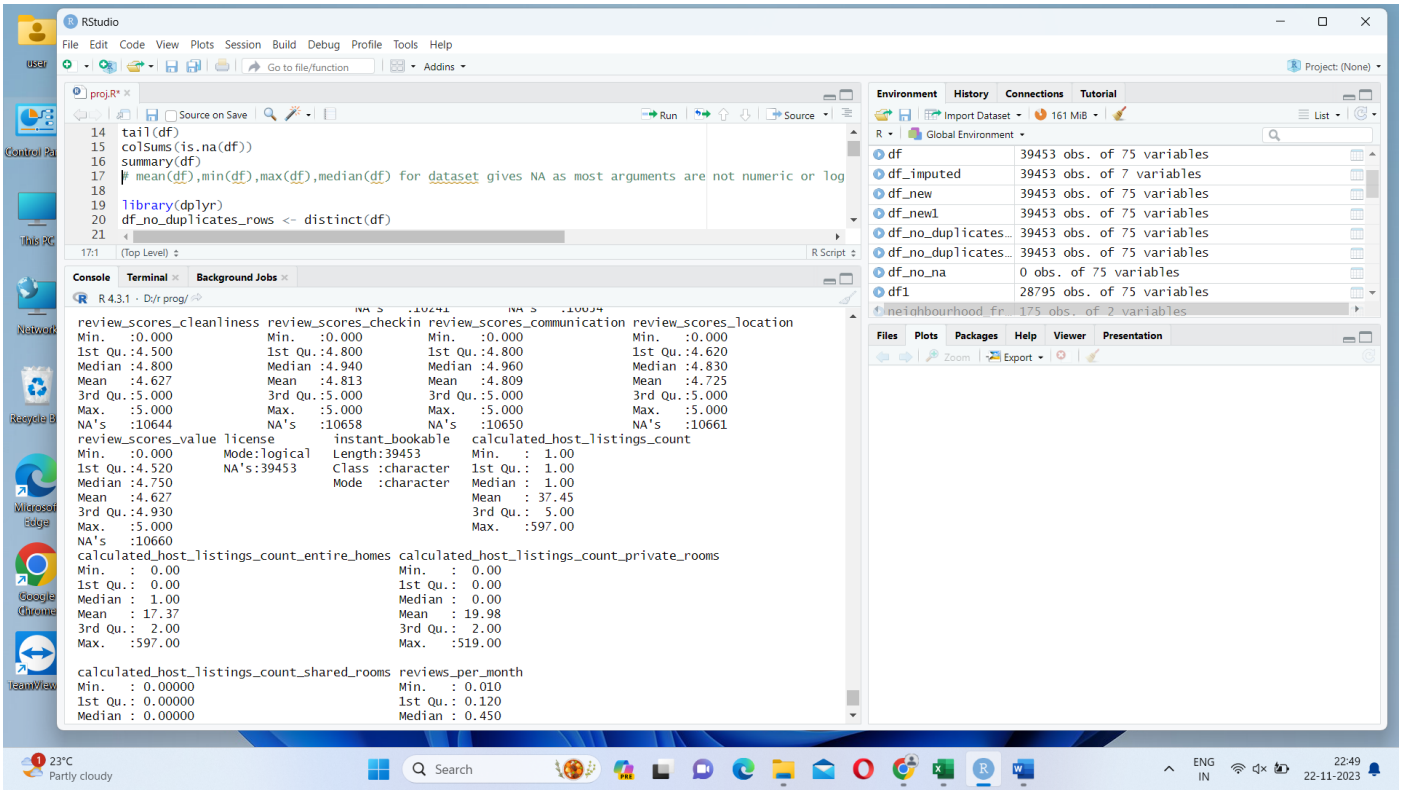

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
13 head(df)
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
beds 605
amenities 0
price 0
minimum_nights 0
maximum_nights 0
minimum_minimum_nights 0
maximum_minimum_nights 0
minimum_maximum_nights 0
maximum_maximum_nights 0
minimum_nights_avg_ntm 0
maximum_nights_avg_ntm 0
calendar_updated 39453
has_availability 0
availability_30 0
availability_60 0
availability_90 0
availability_365 0
calendar_last_scraped 0
number_of_reviews 0
number_of_reviews_ltm 0
number_of_reviews_l30d 0
first_review 0
last_review 0
review_scores_rating 10241
review_scores_accuracy 10654
review_scores_cleanliness 10644
review_scores_checkin 10658
review_scores_communication 10650
Environment History Connections Tutorial
Import Dataset 95 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:46 22-11-2023

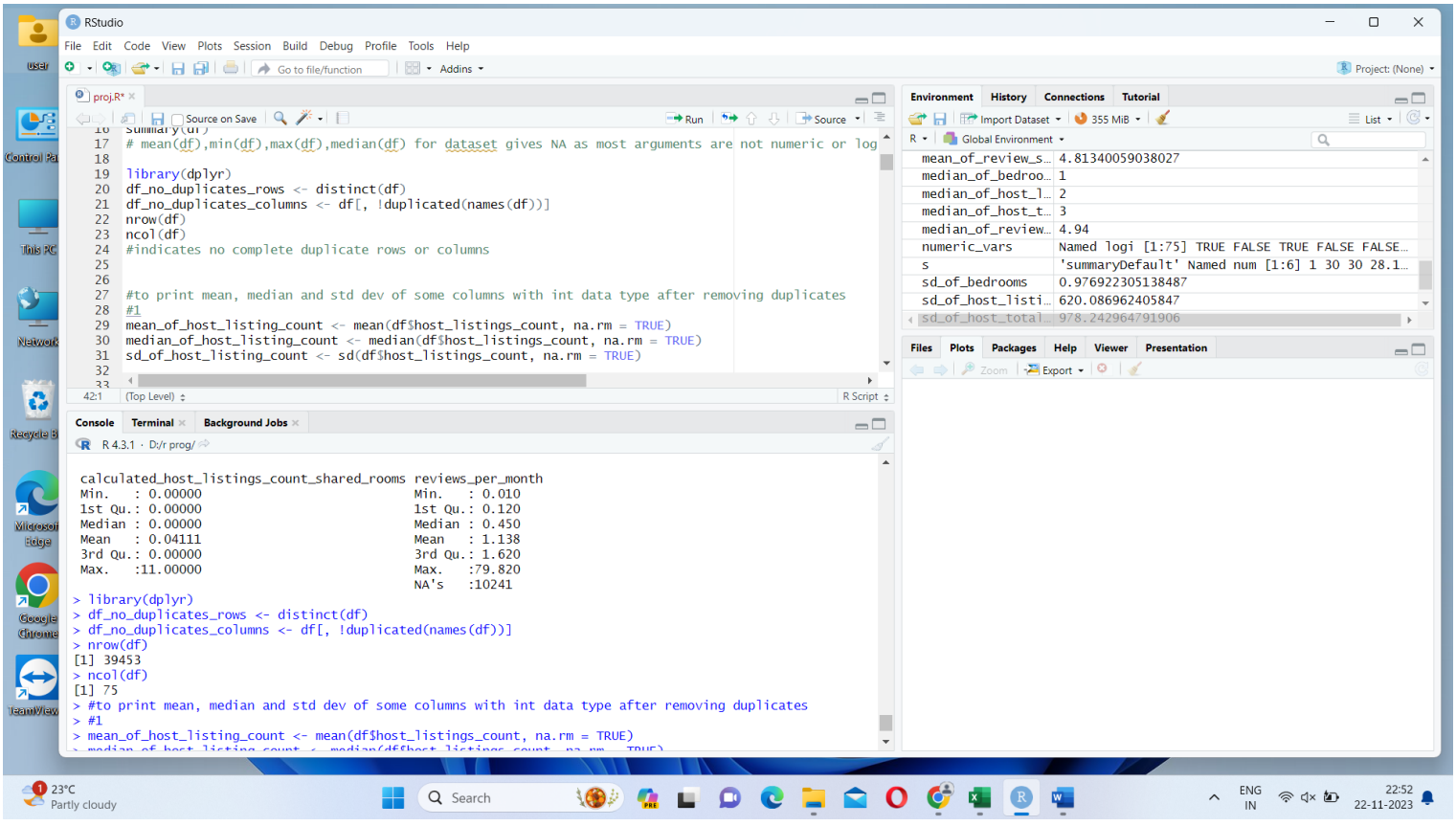

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
0 10241
review_scores_accuracy review_scores_cleanliness
10654 10644
review_scores_checkin review_scores_communication
10658 10650
review_scores_location review_scores_value
10661 10660
license instant_bookable
39453 0
calculated_host_listings_count calculated_host_listings_count_entire_homes
0 0
calculated_host_listings_count_private_rooms calculated_host_listings_count_shared_rooms
0 0
reviews_per_month
10241
> summary(df)
id listing_url scrape_id last_scraped source
Min. :2.595e+03 Length:39453 Min. :2.023e+13 Length:39453 Length:39453
1st Qu.:1.986e+07 Class :character 1st Qu.:2.023e+13 Class :character Class :character
Median :4.543e+07 Mode :character Median :2.023e+13 Mode :character Mode :character
Mean :2.778e+17 Mean :2.023e+13
3rd Qu.:7.135e+17 3rd Qu.:2.023e+13
Max. :9.733e+17 Max. :2.023e+13
name description neighborhood_overview picture_url host_id
Length:39453 Length:39453 Length:39453 Length:39453 Min. : 2234
Class :character Class :character Class :character Class :character 1st Qu.: 16542425
Mode :character Mode :character Mode :character Mode :character Median : 77725998
Environment History Connections Tutorial
Import Dataset 156 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
23°C Partly cloudy
Search
ENG IN
22:47 22-11-2023

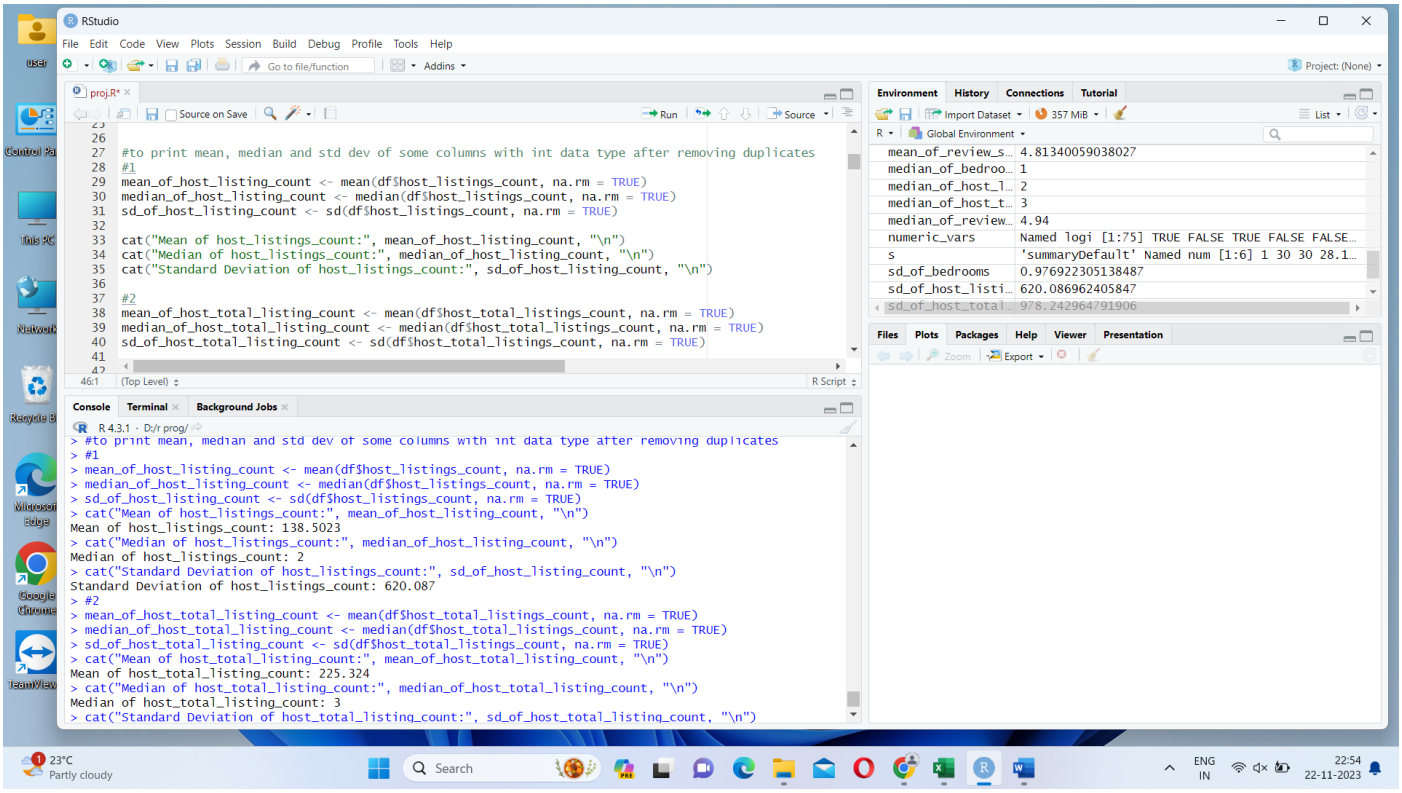

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
host_url host_name host_since host_location host_about
Length:39453 Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character Mode :character
host_response_time host_response_rate host_acceptance_rate host_is_superhost host_thumbnail_url
Length:39453 Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character Mode :character
host_picture_url host_neighbourhood host_listings_count host_total_listings_count
Length:39453 Length:39453 Min. : 0.0 Min. : 1.0
Class :character Class :character 1st Qu.: 1.0 1st Qu.: 1.0
Mode :character Mode :character Median : 2.0 Median : 3.0
Mean : 138.5 Mean : 225.3
3rd Qu.: 7.0 3rd Qu.: 10.0
Max. :4577.0 Max. :9176.0
NA's :5 NA's :5
host_verifications host_has_profile_pic host_identity_verified neighbourhood
Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character
Environment History Connections Tutorial
Import Dataset 157 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
23°C Partly cloudy
Search
ENG IN
22:47 22-11-2023

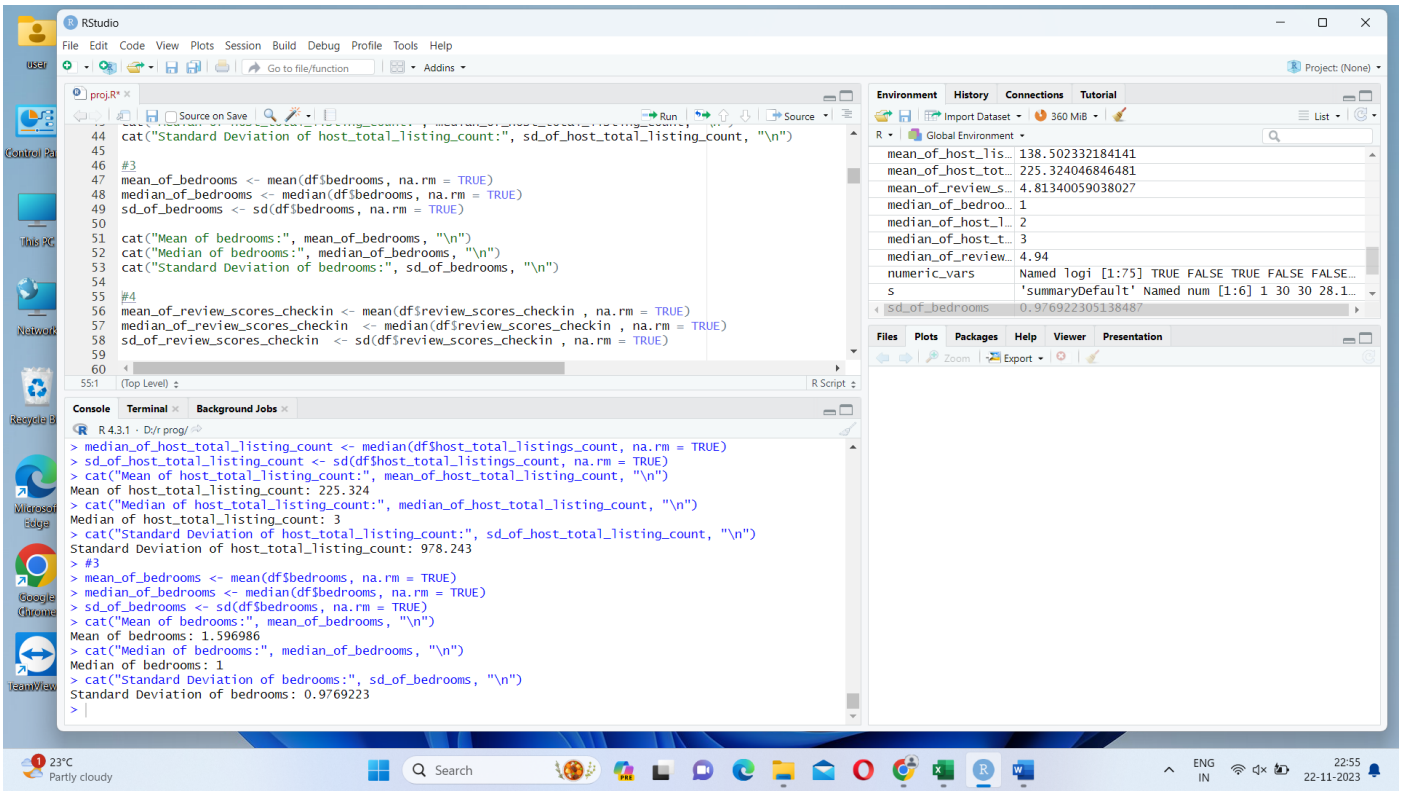

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
neighbourhood_cleansed neighbourhood_group_cleansed latitude longitude
Length:39453 Length:39453 Min. :40.50 Min. :-74.25
Class :character Class :character 1st Qu.:40.69 1st Qu.:-73.98
Mode :character Mode :character Median :40.73 Median :-73.95
Mean :40.73 Mean :-73.95
3rd Qu.:40.76 3rd Qu.:-73.93
Max. :40.91 Max. :-73.71
property_type room_type accommodates bathrooms bathrooms_text
Length:39453 Length:39453 Min. : 1.000 Mode:logical Length:39453
Class :character Class :character 1st Qu.: 2.000 NA's:39453 Class :character
Mode :character Mode :character Median : 2.000 Mode :character
Mean : 2.934
3rd Qu.: 4.000
Max. :16.000
bedrooms beds amenities price minimum_nights
Min. : 1.000 Min. : 1.000 Length:39453 Length:39453 Min. : 1.00
1st Qu.: 1.000 1st Qu.: 1.000 Class :character Class :character 1st Qu.: 30.00
Median : 1.000 Median : 1.000 Mode :character Mode :character Median : 30.00
Mean : 1.597 Mean : 1.658 Mean : 28.05
3rd Qu.: 2.000 3rd Qu.: 2.000 3rd Qu.: 30.00
Max. :50.000 Max. :42.000 Max. :1250.00
NA's :16893 NA's :605
maximum_nights minimum_minimum_nights maximum_minimum_nights minimum_maximum_nights
Min. :1.000e+00 Min. : 1.00 Min. : 1.00 Min. :1.000e+00
1st Qu.:1.200e+02 1st Qu.: 30.00 1st Qu.: 30.00 1st Qu.:3.650e+02
Median :3.650e+02 Median : 30.00 Median : 30.00 Median :1.125e+03
Environment History Connections Tutorial
Import Dataset 159 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C
Partly cloudy
Search
ENG
IN
22:48
22-11-2023

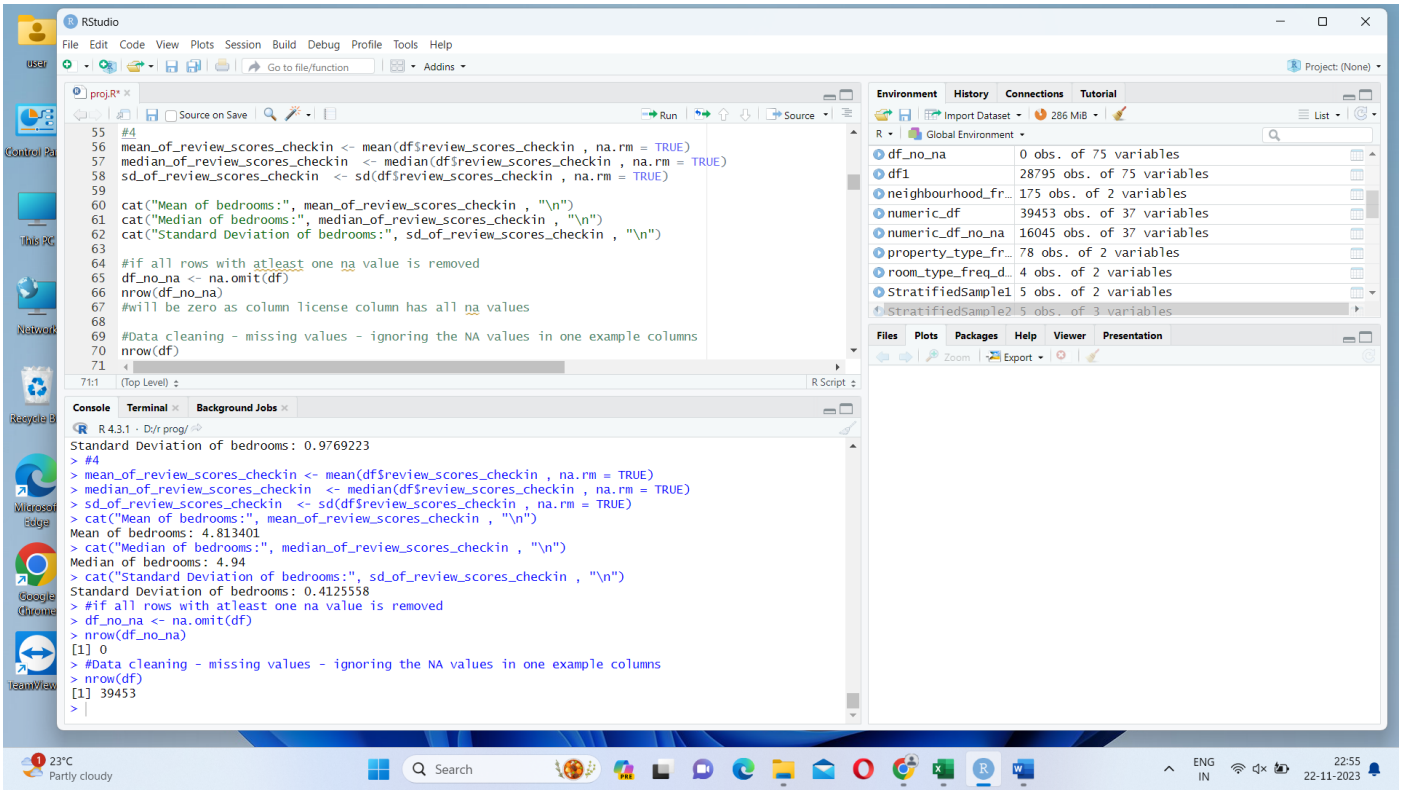

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Mean :5.606e+04 Mean : 29.63 Mean : 34.95 Mean :6.549e+05
3rd Qu.:1.125e+03 3rd Qu.: 30.00 3rd Qu.: 30.00 3rd Qu.:1.125e+03
Max. :2.147e+09 Max. :1250.00 Max. :1250.00 Max. :2.147e+09
maximum_maximum_nights minimum_nights_avg_ntm maximum_nights_avg_ntm calendar_updated
Min. :1.000e+00 Min. : 1.00 Min. :1.000e+00 Mode:logical
1st Qu.:3.650e+02 1st Qu.: 30.00 1st Qu.:3.650e+02 NA's:39453
Median :1.125e+03 Median : 30.00 Median :1.125e+03
Mean :1.199e+06 Mean : 34.33 Mean :1.036e+06
3rd Qu.:1.125e+03 3rd Qu.: 30.00 3rd Qu.:1.125e+03
Max. :2.147e+09 Max. :1250.00 Max. :2.147e+09
has_availability availability_30 availability_60 availability_90 availability_365
Length:39453 Min. : 0.000 Min. : 0.00 Min. : 0.0 Min. : 0.0
Class :character 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0
Mode :character Median : 0.000 Median : 4.00 Median :16.0 Median :107.0
Mean : 6.923 Mean :17.22 Mean :30.3 Mean :144.8
3rd Qu.:11.000 3rd Qu.:34.00 3rd Qu.:61.0 3rd Qu.:291.0
Max. :30.000 Max. :60.00 Max. :90.0 Max. :365.0
calendar_last_scraped number_of_reviews number_of_reviews_ltm number_of_reviews_l30d
Length:39453 Min. : 0.00 Min. : 0.00 Min. : 0.0000
Class :character 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.0000
Mode :character Median : 5.00 Median : 0.00 Median : 0.0000
Mean : 25.84 Mean : 7.36 Mean : 0.6207
3rd Qu.: 24.00 3rd Qu.: 6.00 3rd Qu.: 0.0000
Max. :1834.00 Max. :686.00 Max. :97.0000
Environment History Connections Tutorial
Import Dataset 160 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C
Partly cloudy
Search
ENG
IN
22:48
22-11-2023

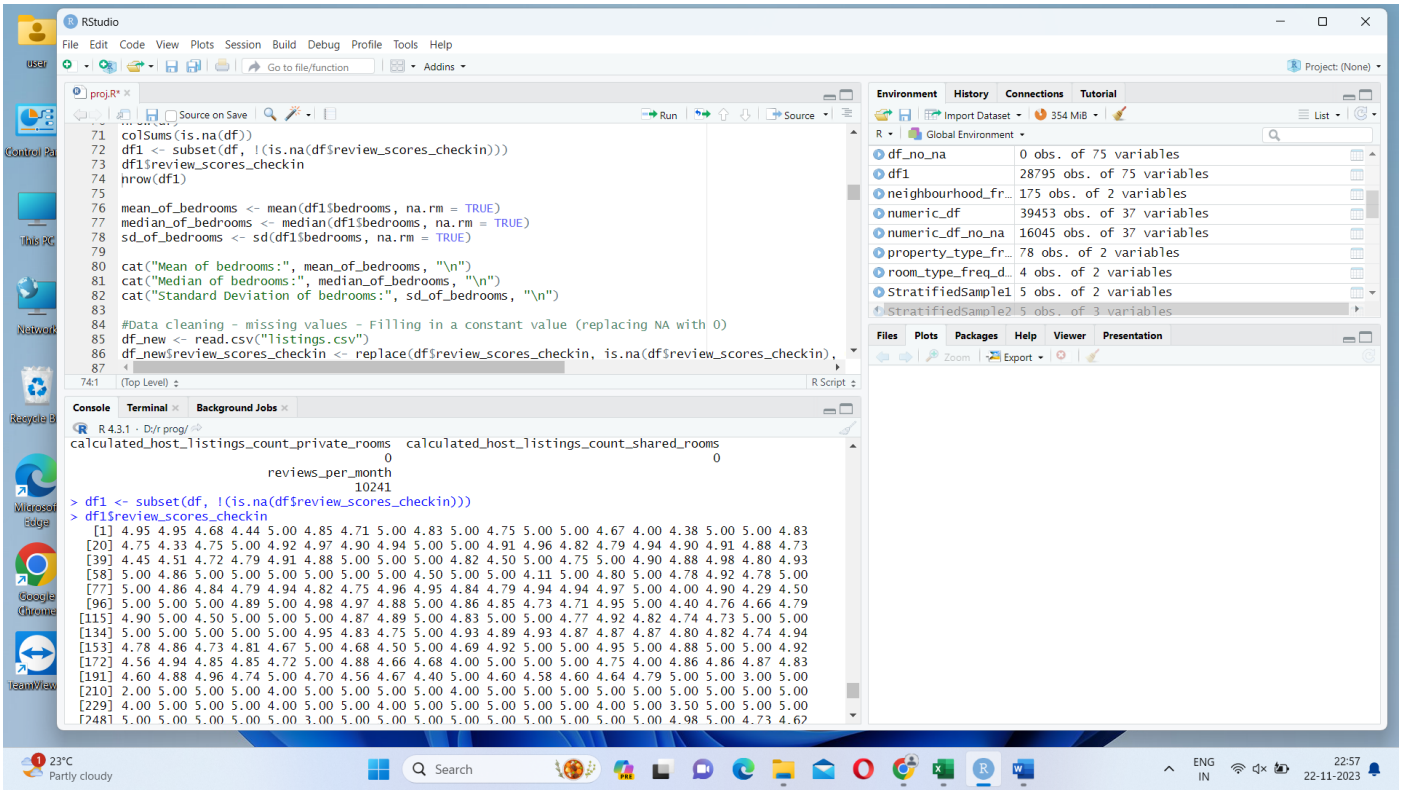

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
review_scores_cleanliness review_scores_checkin review_scores_communication review_scores_location
Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000
1st Qu.:4.500 1st Qu.:4.800 1st Qu.:4.800 1st Qu.:4.620
Median :4.800 Median :4.940 Median :4.960 Median :4.830
Mean :4.627 Mean :4.813 Mean :4.809 Mean :4.725
3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.:5.000
Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
NA's :10644 NA's :10658 NA's :10650 NA's :10661
review_scores_value license instant_bookable calculated_host_listings_count
Min. :0.000 Mode:logical Length:39453 Min. : 1.00
1st Qu.:4.520 NA's:39453 Class :character 1st Qu.: 1.00
Median :4.750 Mode :character Median : 1.00
Mean :4.627 Mean : 37.45
3rd Qu.:4.930 3rd Qu.: 5.00
Max. :5.000 Max. :597.00
NA's :10660
calculated_host_listings_count_entire_homes calculated_host_listings_count_private_rooms
Min. : 0.00 Min. : 0.00
1st Qu.: 0.00 1st Qu.: 0.00
Median : 1.00 Median : 0.00
Mean : 17.37 Mean : 19.98
3rd Qu.: 2.00 3rd Qu.: 2.00
Max. :597.00 Max. :519.00
calculated_host_listings_count_shared_rooms reviews_per_month
Min. : 0.00000 Min. : 0.010
1st Qu.: 0.00000 1st Qu.: 0.120
Median : 0.00000 Median : 0.450
Environment History Connections Tutorial
Import Dataset 161 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:49 22-11-2023

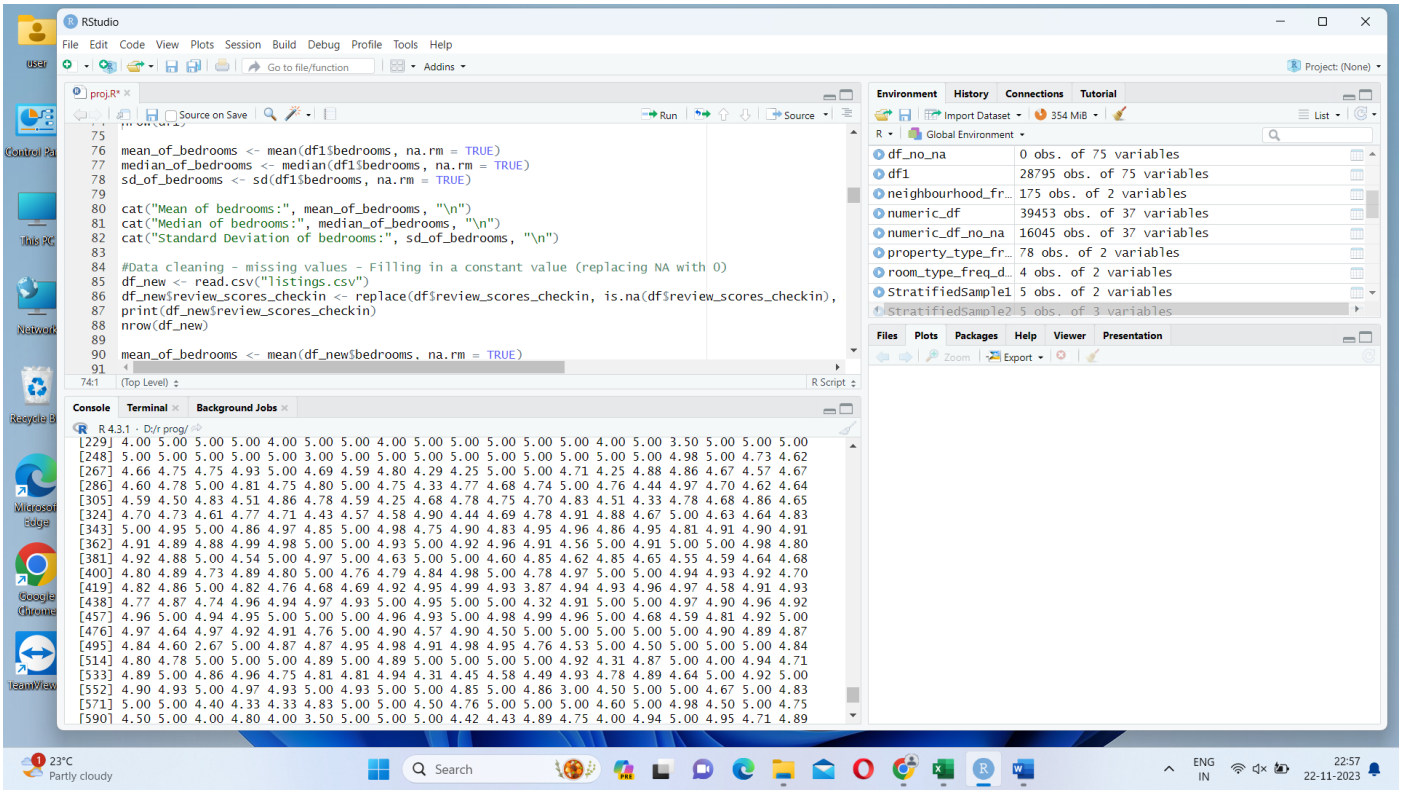

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21 df_no_duplicates_columns <- df[, !duplicated(names(df))]
22 nrow(df)
23 ncol(df)
24 #indicates no complete duplicate rows or columns
25
26
27 #to print mean, median and std dev of some columns with int data type after removing duplicates
28 #1
29 mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
30 median_of_host_listing_count <- median(df$host_listings_count, na.rm = TRUE)
31 sd_of_host_listing_count <- sd(df$host_listings_count, na.rm = TRUE)
32
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
calculated_host_listings_count_shared_rooms reviews_per_month
Min. : 0.00000 Min. : 0.010
1st Qu.: 0.00000 1st Qu.: 0.120
Median : 0.00000 Median : 0.450
Mean : 0.04111 Mean : 1.138
3rd Qu.: 0.00000 3rd Qu.: 1.620
Max. :11.00000 Max. :79.820
NA's :10241
> library(dplyr)
> df_no_duplicates_rows <- distinct(df)
> df_no_duplicates_columns <- df[, !duplicated(names(df))]
> nrow(df)
[1] 39453
> ncol(df)
[1] 75
> #to print mean, median and std dev of some columns with int data type after removing duplicates
> #1
> mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
Environment History Connections Tutorial
Import Dataset 355 MiB
Global Environment
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.976922305138487
sd_of_host_listi... 620.086962405847
sd_of_host_total... 978.242964791906
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:52 22-11-2023

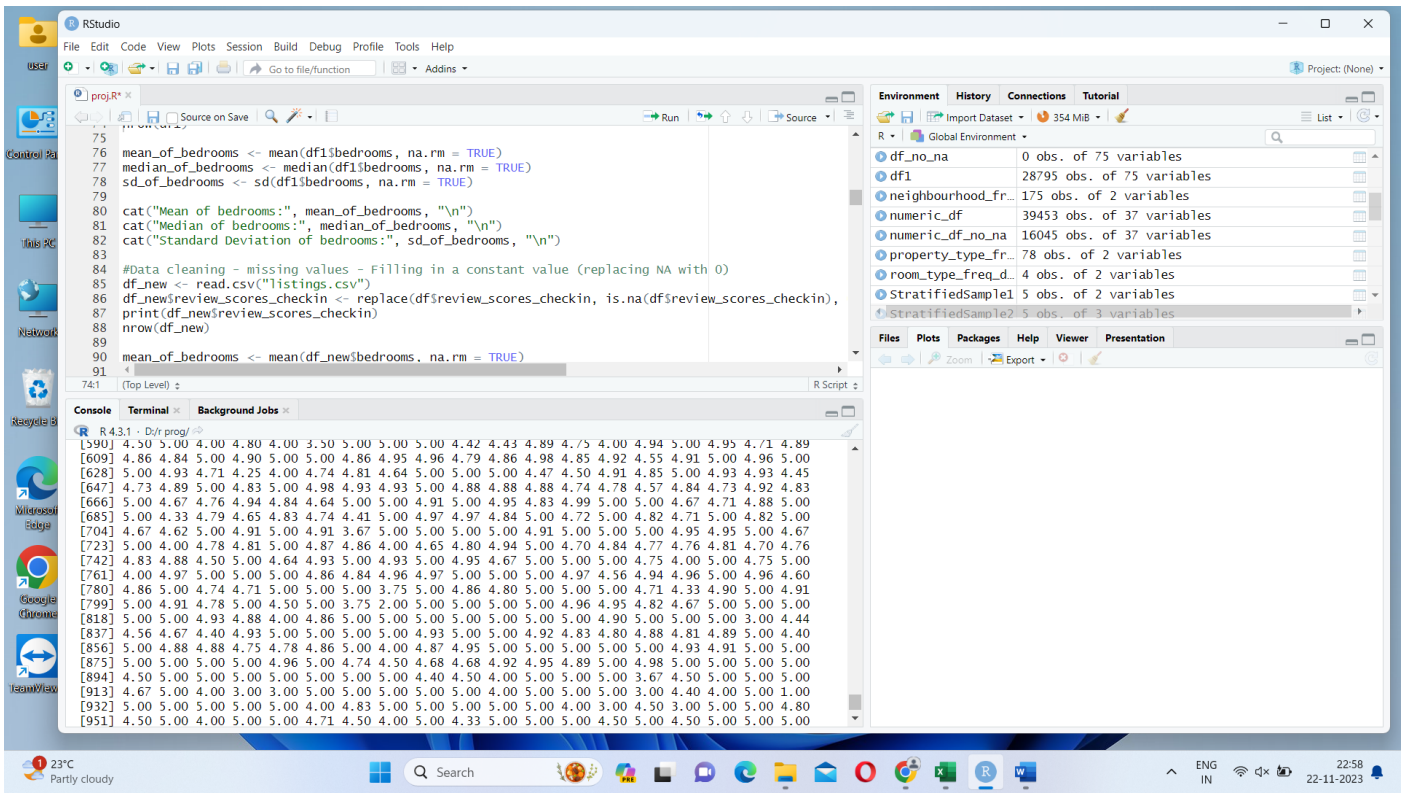

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
#to print mean, median and std dev of some columns with int data type after removing duplicates
#1
mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
median_of_host_listing_count <- median(df$host_listings_count, na.rm = TRUE)
sd_of_host_listing_count <- sd(df$host_listings_count, na.rm = TRUE)
cat("Mean of host_listings_count:", mean_of_host_listing_count, "\n")
cat("Median of host_listings_count:", median_of_host_listing_count, "\n")
cat("Standard Deviation of host_listings_count:", sd_of_host_listing_count, "\n")
#2
mean_of_host_total_listing_count <- mean(df$host_total_listings_count, na.rm = TRUE)
median_of_host_total_listing_count <- median(df$host_total_listings_count, na.rm = TRUE)
sd_of_host_total_listing_count <- sd(df$host_total_listings_count, na.rm = TRUE)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> #to print mean, median and std dev of some columns with int data type after removing duplicates
> #1
> mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
> median_of_host_listing_count <- median(df$host_listings_count, na.rm = TRUE)
> sd_of_host_listing_count <- sd(df$host_listings_count, na.rm = TRUE)
> cat("Mean of host_listings_count:", mean_of_host_listing_count, "\n")
Mean of host_listings_count: 138.5023
> cat("Median of host_listings_count:", median_of_host_listing_count, "\n")
Median of host_listings_count: 2
> cat("Standard Deviation of host_listings_count:", sd_of_host_listing_count, "\n")
Standard Deviation of host_listings_count: 620.087
> #2
> mean_of_host_total_listing_count <- mean(df$host_total_listings_count, na.rm = TRUE)
> median_of_host_total_listing_count <- median(df$host_total_listings_count, na.rm = TRUE)
> sd_of_host_total_listing_count <- sd(df$host_total_listings_count, na.rm = TRUE)
> cat("Mean of host_total_listing_count:", mean_of_host_total_listing_count, "\n")
Mean of host_total_listing_count: 225.324
> cat("Median of host_total_listing_count:", median_of_host_total_listing_count, "\n")
Median of host_total_listing_count: 3
> cat("Standard Deviation of host_total_listing_count:", sd_of_host_total_listing_count, "\n")
Environment History Connections Tutorial
Import Dataset 357 MiB
Global Environment
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.976922305138487
sd_of_host_listi... 620.086962405847
sd_of_host_total... 978.242964791906
Files Plots Packages Help Viewer Presentation
23°C Partly cloudy
Search
ENG IN
22:54 22-11-2023

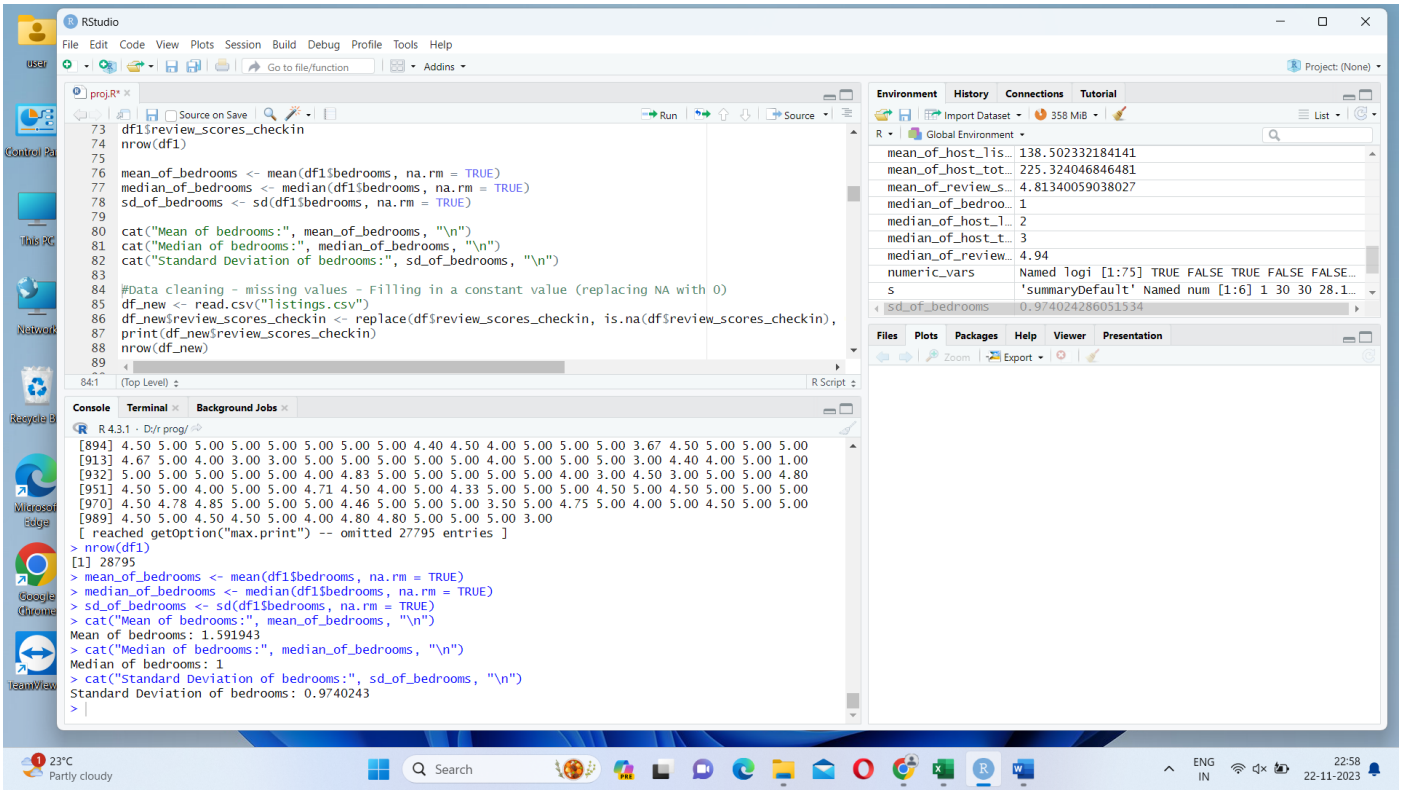

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
cat("Standard Deviation of host_total_listing_count:", sd_of_host_total_listing_count, "\n")
#3
mean_of_bedrooms <- mean(df$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#4
mean_of_review_scores_checkin <- mean(df$review_scores_checkin , na.rm = TRUE)
median_of_review_scores_checkin <- median(df$review_scores_checkin , na.rm = TRUE)
sd_of_review_scores_checkin <- sd(df$review_scores_checkin , na.rm = TRUE)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> median_of_host_total_listing_count <- median(df$host_total_listings_count, na.rm = TRUE)
> sd_of_host_total_listing_count <- sd(df$host_total_listings_count, na.rm = TRUE)
> cat("Mean of host_total_listing_count:", mean_of_host_total_listing_count, "\n")
Mean of host_total_listing_count: 225.324
> cat("Median of host_total_listing_count:", median_of_host_total_listing_count, "\n")
Median of host_total_listing_count: 3
> cat("Standard Deviation of host_total_listing_count:", sd_of_host_total_listing_count, "\n")
Standard Deviation of host_total_listing_count: 978.243
> #3
> mean_of_bedrooms <- mean(df$bedrooms, na.rm = TRUE)
> median_of_bedrooms <- median(df$bedrooms, na.rm = TRUE)
> sd_of_bedrooms <- sd(df$bedrooms, na.rm = TRUE)
> cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
Mean of bedrooms: 1.596986
> cat("Median of bedrooms:", median_of_bedrooms, "\n")
Median of bedrooms: 1
> cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
Standard Deviation of bedrooms: 0.9769223
Environment History Connections Tutorial
Import Dataset 360 MiB
Global Environment
mean_of_host_lis... 138.502332184141
mean_of_host_tot... 225.324046846481
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.976922305138487
Files Plots Packages Help Viewer Presentation
23°C Partly cloudy
Search
ENG IN
22:55 22-11-2023

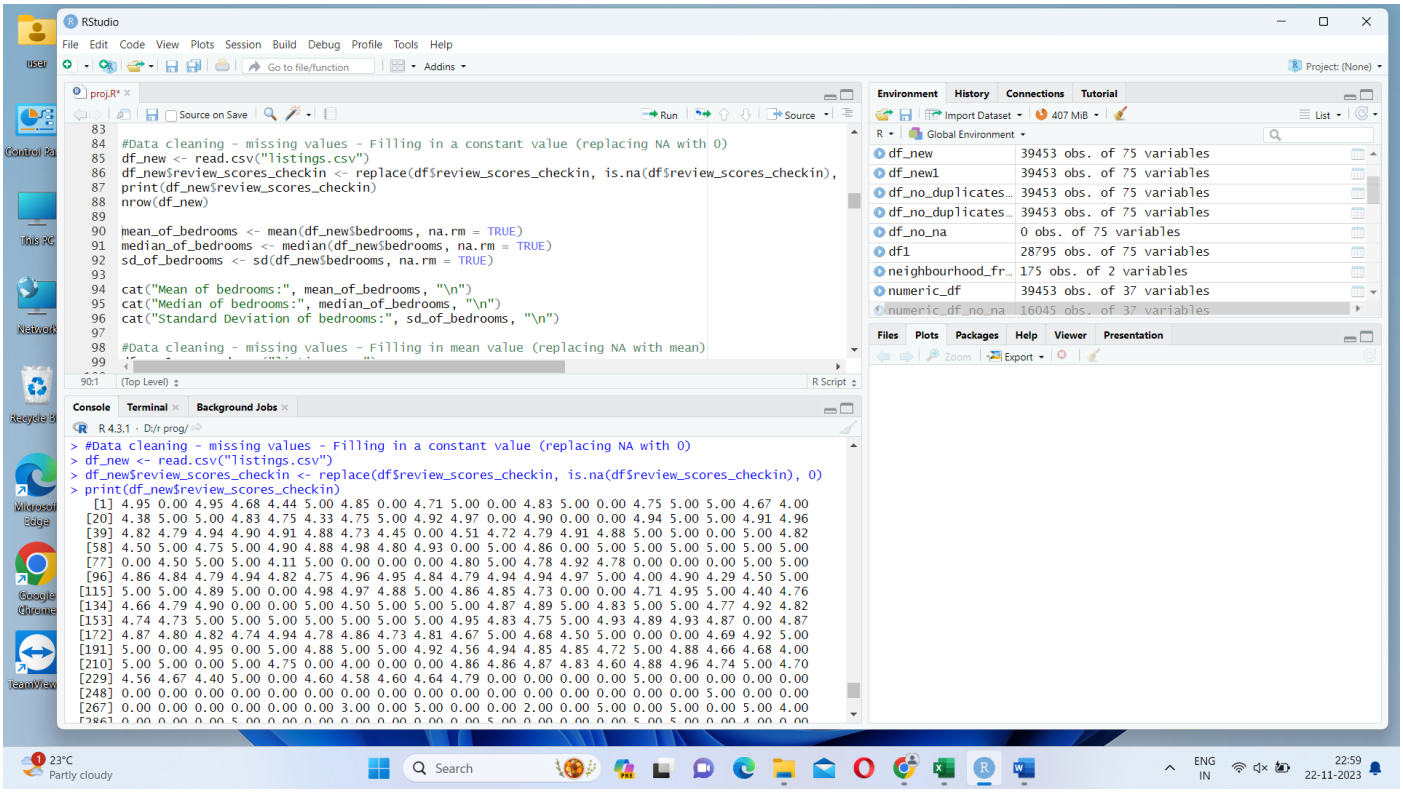

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
#4
mean_of_review_scores_checkin <- mean(df$review_scores_checkin , na.rm = TRUE)
median_of_review_scores_checkin <- median(df$review_scores_checkin , na.rm = TRUE)
sd_of_review_scores_checkin <- sd(df$review_scores_checkin , na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_review_scores_checkin , "\n")
cat("Median of bedrooms:", median_of_review_scores_checkin , "\n")
cat("Standard Deviation of bedrooms:", sd_of_review_scores_checkin , "\n")
#if all rows with atleast one na value is removed
df_no_na <- na.omit(df)
nrow(df_no_na)
#will be zero as column license column has all na values
#Data cleaning - missing values - ignoring the NA values in one example columns
nrow(df)
71:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Standard Deviation of bedrooms: 0.9769223
> #4
> mean_of_review_scores_checkin <- mean(df$review_scores_checkin , na.rm = TRUE)
> median_of_review_scores_checkin <- median(df$review_scores_checkin , na.rm = TRUE)
> sd_of_review_scores_checkin <- sd(df$review_scores_checkin , na.rm = TRUE)
> cat("Mean of bedrooms:", mean_of_review_scores_checkin , "\n")
Mean of bedrooms: 4.813401
> cat("Median of bedrooms:", median_of_review_scores_checkin , "\n")
Median of bedrooms: 4.94
> cat("Standard Deviation of bedrooms:", sd_of_review_scores_checkin , "\n")
Standard Deviation of bedrooms: 0.4125558
> #if all rows with atleast one na value is removed
> df_no_na <- na.omit(df)
> nrow(df_no_na)
[1] 0
> #Data cleaning - missing values - ignoring the NA values in one example columns
> nrow(df)
[1] 39453
Environment History Connections Tutorial
Import Dataset 286 MiB
Global Environment
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
room_type_freq_d... 4 obs. of 2 variables
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:55 22-11-2023

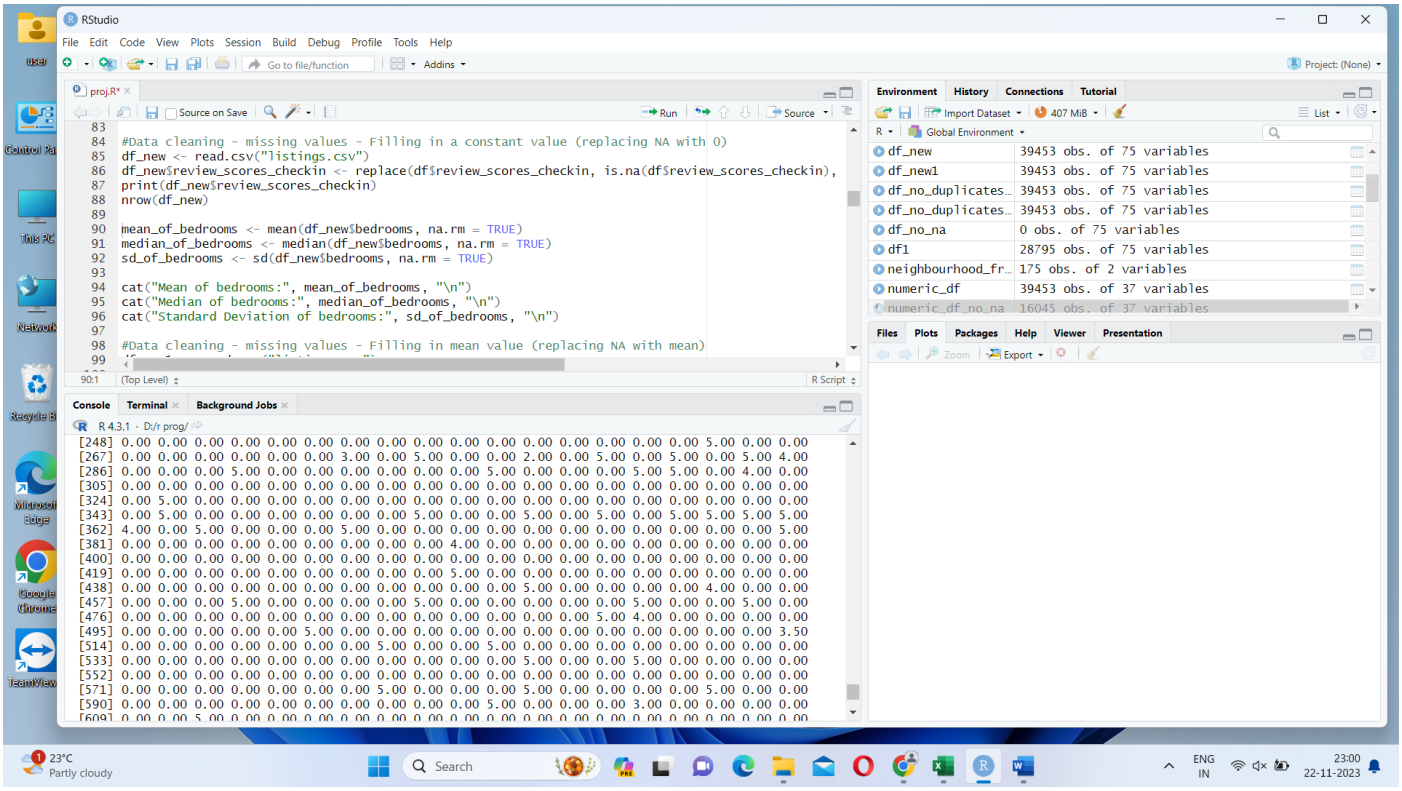

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
colSums(is.na(df))
df1 <- subset(df, !(is.na(df$review_scores_checkin)))
df1$review_scores_checkin
nrow(df1)
mean_of_bedrooms <- mean(df1$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df1$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df1$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
74:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
calculated_host_listings_count_private_rooms calculated_host_listings_count_shared_rooms
0 0
reviews_per_month
10241
> df1 <- subset(df, !(is.na(df$review_scores_checkin)))
> df1$review_scores_checkin
[1] 4.95 4.95 4.68 4.44 5.00 4.85 4.71 5.00 4.83 5.00 4.75 5.00 5.00 4.67 4.00 4.38 5.00 5.00 4.83
[20] 4.75 4.33 4.75 5.00 4.92 4.97 4.90 4.94 5.00 5.00 4.91 4.96 4.82 4.79 4.94 4.90 4.91 4.88 4.73
[39] 4.45 4.51 4.72 4.79 4.91 4.88 5.00 5.00 5.00 4.82 4.50 5.00 4.75 5.00 4.90 4.88 4.98 4.80 4.93
[58] 5.00 4.86 5.00 5.00 5.00 5.00 5.00 5.00 4.50 5.00 5.00 4.11 5.00 4.80 5.00 4.78 4.92 4.78 5.00
[77] 5.00 4.86 4.84 4.79 4.94 4.82 4.75 4.96 4.95 4.84 4.79 4.94 4.94 4.97 5.00 4.00 4.90 4.29 4.50
[96] 5.00 5.00 5.00 4.89 5.00 4.98 4.97 4.88 5.00 4.86 4.85 4.73 4.71 4.95 5.00 4.40 4.76 4.66 4.79
[115] 4.90 5.00 4.50 5.00 5.00 5.00 4.87 4.89 5.00 4.83 5.00 5.00 4.77 4.92 4.82 4.74 4.73 5.00 5.00
[134] 5.00 5.00 5.00 5.00 5.00 4.95 4.83 4.75 5.00 4.93 4.89 4.93 4.87 4.87 4.87 4.80 4.82 4.74 4.94
[153] 4.78 4.86 4.73 4.81 4.67 5.00 4.68 4.50 5.00 4.69 4.92 5.00 5.00 4.95 5.00 4.88 5.00 5.00 4.92
[172] 4.56 4.94 4.85 4.85 4.72 5.00 4.88 4.66 4.68 4.00 5.00 5.00 5.00 4.75 4.00 4.86 4.86 4.87 4.83
[191] 4.60 4.88 4.96 4.74 5.00 4.70 4.56 4.67 4.40 5.00 4.60 4.58 4.60 4.64 4.79 5.00 5.00 3.00 5.00
[210] 2.00 5.00 5.00 5.00 4.00 5.00 5.00 5.00 5.00 4.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00
[229] 4.00 5.00 5.00 5.00 4.00 5.00 5.00 4.00 5.00 5.00 5.00 5.00 4.00 5.00 3.50 5.00 5.00 5.00
[248] 5.00 5.00 5.00 5.00 5.00 3.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 4.98 5.00 4.73 4.62
Environment History Connections Tutorial
Import Dataset 354 MiB
Global Environment
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
room_type_freq_d... 4 obs. of 2 variables
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:57 22-11-2023

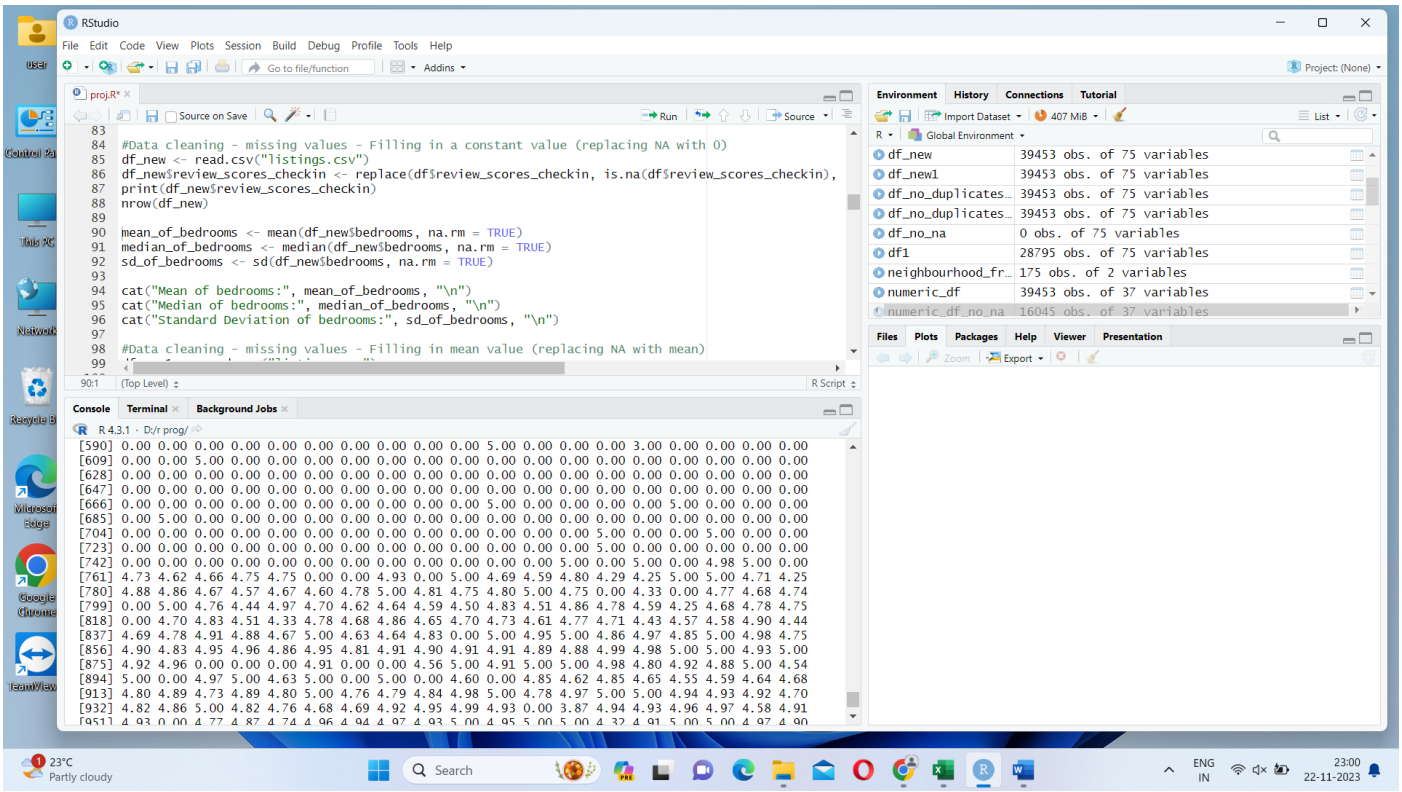

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
mean_of_bedrooms <- mean(df1$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df1$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df1$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
print(df_new$review_scores_checkin)
nrow(df_new)
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
Environment History Connections Tutorial
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
room_type_freq_d... 4 obs. of 2 variables
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
[229] 4.00 5.00 5.00 5.00 4.00 5.00 5.00 4.00 5.00 5.00 5.00 5.00 5.00 4.00 5.00 3.50 5.00 5.00 5.00
[248] 5.00 5.00 5.00 5.00 5.00 3.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 5.00 4.98 5.00 4.73 4.62
[267] 4.66 4.75 4.75 4.93 5.00 4.69 4.59 4.80 4.29 4.25 5.00 5.00 4.71 4.25 4.88 4.86 4.67 4.57 4.67
[286] 4.60 4.78 5.00 4.81 4.75 4.80 5.00 4.75 4.33 4.77 4.68 4.74 5.00 4.76 4.44 4.97 4.70 4.62 4.64
[305] 4.59 4.50 4.83 4.51 4.86 4.78 4.59 4.25 4.68 4.78 4.75 4.70 4.83 4.51 4.33 4.78 4.68 4.86 4.65
[324] 4.70 4.73 4.61 4.77 4.71 4.43 4.57 4.58 4.90 4.44 4.69 4.78 4.91 4.88 4.67 5.00 4.63 4.64 4.83
[343] 5.00 4.95 5.00 4.86 4.97 4.85 5.00 4.98 4.75 4.90 4.83 4.95 4.96 4.86 4.95 4.81 4.91 4.90 4.91
[362] 4.91 4.89 4.88 4.99 4.98 5.00 5.00 4.93 5.00 4.92 4.96 4.91 4.56 5.00 4.91 5.00 5.00 4.98 4.80
[381] 4.92 4.88 5.00 4.54 5.00 4.97 5.00 4.63 5.00 5.00 4.60 4.85 4.62 4.85 4.65 4.55 4.59 4.64 4.68
[400] 4.80 4.89 4.73 4.89 4.80 5.00 4.76 4.79 4.84 4.98 5.00 4.78 4.97 5.00 5.00 4.94 4.93 4.92 4.70
[419] 4.82 4.86 5.00 4.82 4.76 4.68 4.69 4.92 4.95 4.99 4.93 3.87 4.94 4.93 4.96 4.97 4.58 4.91 4.93
[438] 4.77 4.87 4.74 4.96 4.94 4.97 4.93 5.00 4.95 5.00 5.00 4.32 4.91 5.00 5.00 4.97 4.90 4.96 4.92
[457] 4.96 5.00 4.94 4.95 5.00 5.00 5.00 4.96 4.93 5.00 4.98 4.99 4.96 5.00 4.68 4.59 4.81 4.92 5.00
[476] 4.97 4.64 4.97 4.92 4.91 4.76 5.00 4.90 4.57 4.90 4.50 5.00 5.00 5.00 5.00 5.00 4.90 4.89 4.87
[495] 4.84 4.60 2.67 5.00 4.87 4.87 4.95 4.98 4.91 4.98 4.95 4.76 4.53 5.00 4.50 5.00 5.00 5.00 4.84
[514] 4.80 4.78 5.00 5.00 5.00 4.89 5.00 4.89 5.00 5.00 5.00 5.00 4.92 4.31 4.87 5.00 4.00 4.94 4.71
[533] 4.89 5.00 4.86 4.96 4.75 4.81 4.81 4.94 4.31 4.45 4.58 4.49 4.93 4.78 4.89 4.64 5.00 4.92 5.00
[552] 4.90 4.93 5.00 4.97 4.93 5.00 4.93 5.00 5.00 4.85 5.00 4.86 3.00 4.50 5.00 5.00 4.67 5.00 4.83
[571] 5.00 5.00 4.40 4.33 4.33 4.83 5.00 5.00 4.50 4.76 5.00 5.00 5.00 4.60 5.00 4.98 4.50 5.00 4.75
[590] 4.50 5.00 4.00 4.80 4.00 3.50 5.00 5.00 5.00 4.42 4.43 4.89 4.75 4.00 4.94 5.00 4.95 4.71 4.89
23°C Partly cloudy
Search
ENG IN
22:57 22-11-2023

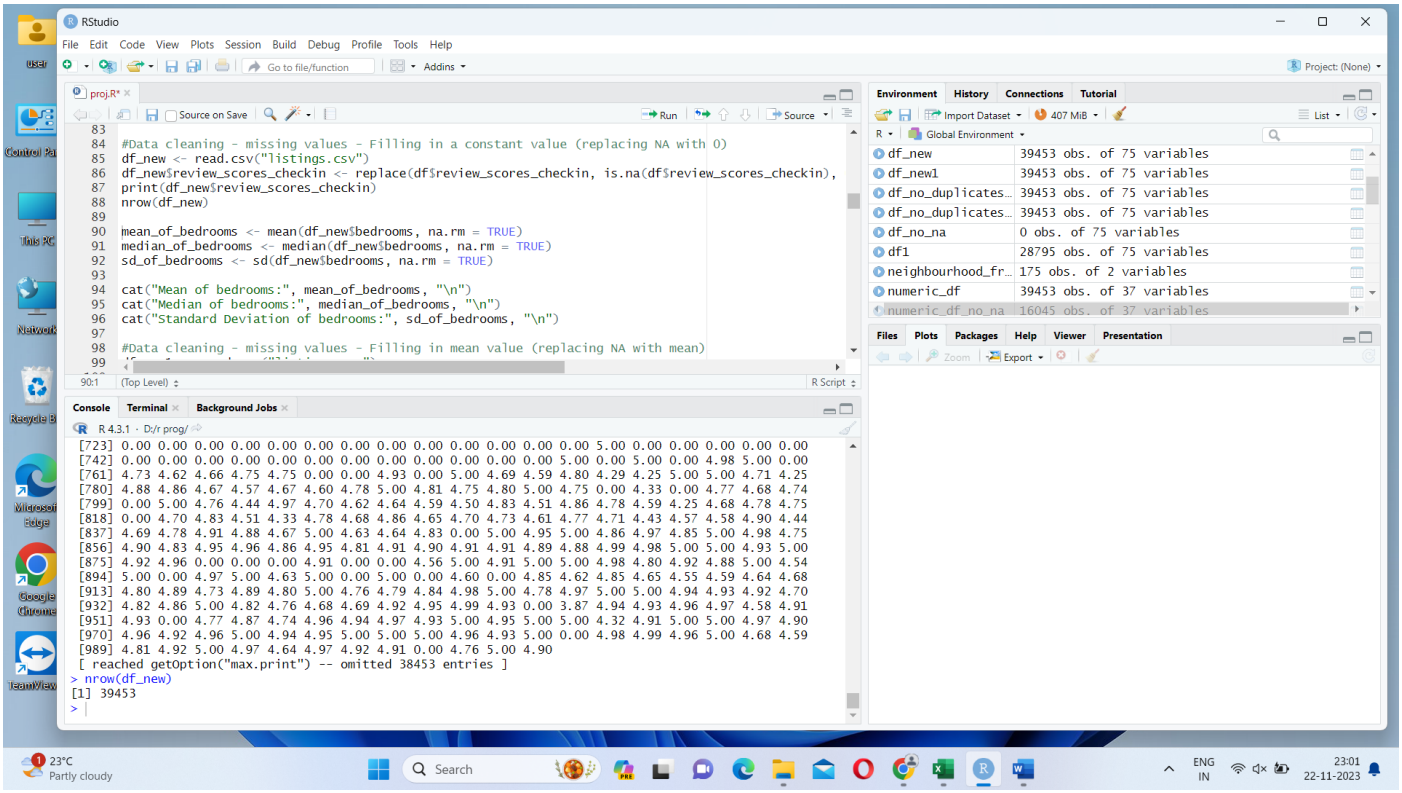

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
mean_of_bedrooms <- mean(df1$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df1$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df1$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
print(df_new$review_scores_checkin)
nrow(df_new)
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
Environment History Connections Tutorial
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
room_type_freq_d... 4 obs. of 2 variables
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
[590] 4.50 5.00 4.00 4.80 4.00 3.50 5.00 5.00 5.00 4.42 4.43 4.89 4.75 4.00 4.94 5.00 4.95 4.71 4.89
[609] 4.86 4.84 5.00 4.90 5.00 5.00 4.86 4.95 4.96 4.79 4.86 4.98 4.85 4.92 4.55 4.91 5.00 4.96 5.00
[628] 5.00 4.93 4.71 4.25 4.00 4.74 4.81 4.64 5.00 5.00 5.00 4.47 4.50 4.91 4.85 5.00 4.93 4.93 4.45
[647] 4.73 4.89 5.00 4.83 5.00 4.98 4.93 4.93 5.00 4.88 4.88 4.88 4.74 4.78 4.57 4.84 4.73 4.92 4.83
[666] 5.00 4.67 4.76 4.94 4.84 4.64 5.00 5.00 4.91 5.00 4.95 4.83 4.99 5.00 5.00 4.67 4.71 4.88 5.00
[685] 5.00 4.33 4.79 4.65 4.83 4.74 4.41 5.00 4.97 4.97 4.84 5.00 4.72 5.00 4.82 4.71 5.00 4.82 5.00
[704] 4.67 4.62 5.00 4.91 5.00 4.91 3.67 5.00 5.00 5.00 5.00 4.91 5.00 5.00 5.00 4.95 4.95 5.00 4.67
[723] 5.00 4.00 4.78 4.81 5.00 4.87 4.86 4.00 4.65 4.80 4.94 5.00 4.70 4.84 4.77 4.76 4.81 4.70 4.76
[742] 4.83 4.88 4.50 5.00 4.64 4.93 5.00 4.93 5.00 4.95 4.67 5.00 5.00 5.00 4.75 4.00 5.00 4.75 5.00
[761] 4.00 4.97 5.00 5.00 5.00 4.86 4.84 4.96 4.97 5.00 5.00 5.00 4.97 4.56 4.94 4.96 5.00 4.96 4.60
[780] 4.86 5.00 4.74 4.71 5.00 5.00 5.00 3.75 5.00 4.86 4.80 5.00 5.00 5.00 4.71 4.33 4.90 5.00 4.91
[799] 5.00 4.91 4.78 5.00 4.50 5.00 3.75 2.00 5.00 5.00 5.00 5.00 4.96 4.95 4.82 4.67 5.00 5.00 5.00
[818] 5.00 5.00 4.93 4.88 4.00 4.86 5.00 5.00 5.00 5.00 5.00 4.90 5.00 5.00 5.00 5.00 3.00 4.44
[837] 4.56 4.67 4.40 4.93 5.00 5.00 5.00 5.00 4.93 5.00 5.00 4.92 4.83 4.80 4.88 4.81 4.89 5.00 4.40
[856] 5.00 4.88 4.88 4.75 4.78 4.86 5.00 4.00 4.87 4.95 5.00 5.00 5.00 5.00 4.93 4.91 5.00 5.00
[875] 5.00 5.00 5.00 5.00 4.96 5.00 4.74 4.50 4.68 4.68 4.92 4.95 4.89 5.00 4.98 5.00 5.00 5.00 5.00
[894] 4.50 5.00 5.00 5.00 5.00 5.00 5.00 5.00 4.40 4.50 4.00 5.00 5.00 5.00 3.67 4.50 5.00 5.00 5.00
[913] 4.67 5.00 4.00 3.00 3.00 5.00 5.00 5.00 5.00 5.00 4.00 5.00 5.00 5.00 3.00 4.40 4.00 5.00 1.00
[932] 5.00 5.00 5.00 5.00 5.00 4.00 4.83 5.00 5.00 5.00 5.00 4.00 3.00 4.50 3.00 5.00 5.00 4.80
[951] 4.50 5.00 4.00 5.00 5.00 4.71 4.50 4.00 5.00 4.33 5.00 5.00 5.00 4.50 5.00 4.50 5.00 5.00 5.00
23°C Partly cloudy
Search
ENG IN
22:58 22-11-2023

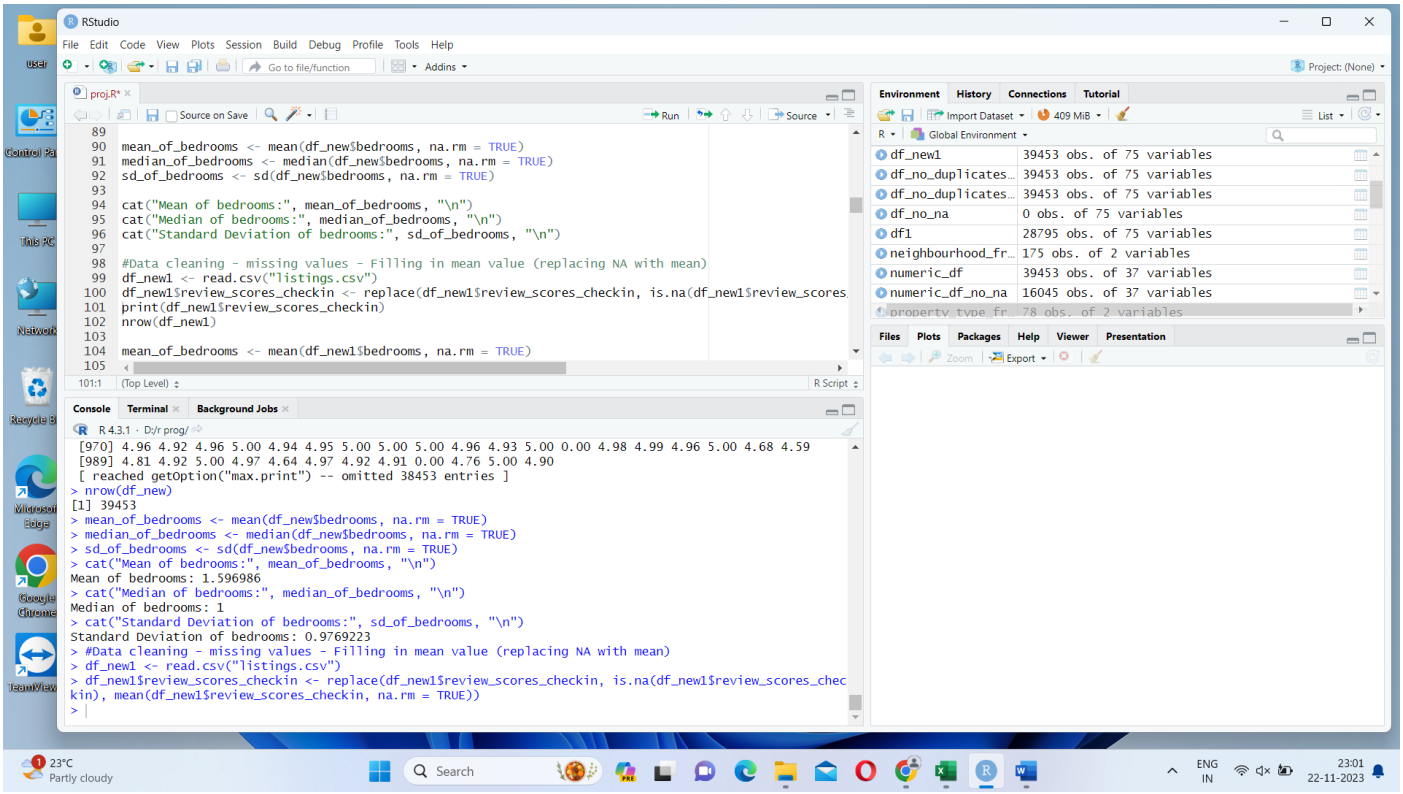

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
73 df1$review_scores_checkin
74 nrow(df1)
75
76 mean_of_bedrooms <- mean(df1$bedrooms, na.rm = TRUE)
77 median_of_bedrooms <- median(df1$bedrooms, na.rm = TRUE)
78 sd_of_bedrooms <- sd(df1$bedrooms, na.rm = TRUE)
79
80 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
81 cat("Median of bedrooms:", median_of_bedrooms, "\n")
82 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
83
84 #Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
85 df_new <- read.csv("listings.csv")
86 df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
87 print(df_new$review_scores_checkin)
88 nrow(df_new)
89
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[894] 4.50 5.00 5.00 5.00 5.00 5.00 5.00 4.40 4.50 4.00 5.00 5.00 5.00 3.67 4.50 5.00 5.00 5.00
[913] 4.67 5.00 4.00 3.00 3.00 5.00 5.00 5.00 5.00 4.00 5.00 5.00 5.00 3.00 4.40 4.00 5.00 1.00
[932] 5.00 5.00 5.00 5.00 5.00 4.00 4.83 5.00 5.00 5.00 5.00 4.00 3.00 4.50 3.00 5.00 5.00 4.80
[951] 4.50 5.00 4.00 5.00 5.00 4.71 4.50 4.00 5.00 4.33 5.00 5.00 5.00 4.50 5.00 4.50 5.00 5.00 5.00
[970] 4.50 4.78 4.85 5.00 5.00 5.00 4.46 5.00 5.00 5.00 3.50 5.00 4.75 5.00 4.00 5.00 4.50 5.00 5.00
[989] 4.50 5.00 4.50 4.50 5.00 4.00 4.80 4.80 5.00 5.00 5.00 3.00
[reached getOption("max.print") -- omitted 27795 entries]
> nrow(df1)
[1] 28795
> mean_of_bedrooms <- mean(df1$bedrooms, na.rm = TRUE)
> median_of_bedrooms <- median(df1$bedrooms, na.rm = TRUE)
> sd_of_bedrooms <- sd(df1$bedrooms, na.rm = TRUE)
> cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
Mean of bedrooms: 1.591943
> cat("Median of bedrooms:", median_of_bedrooms, "\n")
Median of bedrooms: 1
> cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
Standard Deviation of bedrooms: 0.9740243
>
Environment History Connections Tutorial
Import Dataset 358 MiB
Global Environment
mean_of_host_lis... 138.502332184141
mean_of_host_tot... 225.324046846481
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.974024286051534
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:58 22-11-2023

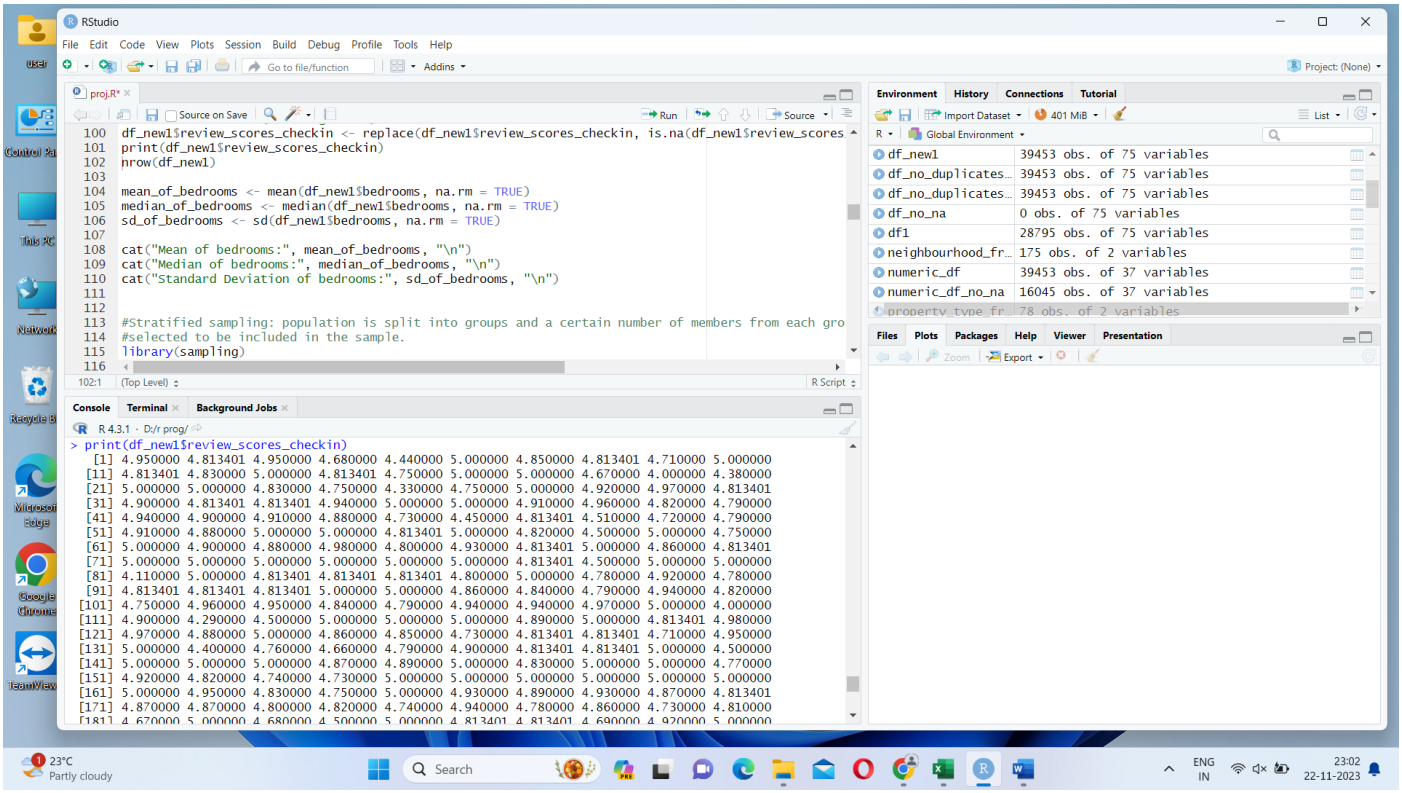

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
83
84 #Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
85 df_new <- read.csv("listings.csv")
86 df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
87 print(df_new$review_scores_checkin)
88 nrow(df_new)
89
90 mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
91 median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
92 sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
93
94 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
95 cat("Median of bedrooms:", median_of_bedrooms, "\n")
96 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
97
98 #Data cleaning - missing values - Filling in mean value (replacing NA with mean)
99
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> #Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
> df_new <- read.csv("listings.csv")
> df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin), 0)
> print(df_new$review_scores_checkin)
[1] 4.95 0.00 4.95 4.68 4.44 5.00 4.85 0.00 4.71 5.00 0.00 4.83 5.00 0.00 4.75 5.00 5.00 4.67 4.00
[20] 4.38 5.00 5.00 4.83 4.75 4.33 4.75 5.00 4.92 4.97 0.00 4.90 0.00 0.00 4.94 5.00 5.00 4.91 4.96
[39] 4.82 4.79 4.94 4.90 4.91 4.88 4.73 4.45 0.00 4.51 4.72 4.79 4.91 4.88 5.00 5.00 0.00 5.00 4.82
[58] 4.50 5.00 4.75 5.00 4.90 4.88 4.98 4.80 4.93 0.00 5.00 4.86 0.00 5.00 5.00 5.00 5.00 5.00 5.00
[77] 0.00 4.50 5.00 5.00 4.11 5.00 0.00 0.00 0.00 4.80 5.00 4.78 4.92 4.78 0.00 0.00 0.00 5.00 5.00
[96] 4.86 4.84 4.79 4.94 4.82 4.75 4.96 4.95 4.84 4.79 4.94 4.94 4.97 5.00 4.00 4.90 4.29 4.50 5.00
[115] 5.00 5.00 4.89 5.00 0.00 4.98 4.97 4.88 5.00 4.86 4.85 4.73 0.00 0.00 4.71 4.95 5.00 4.40 4.76
[134] 4.66 4.79 4.90 0.00 0.00 5.00 4.50 5.00 5.00 5.00 4.87 4.89 5.00 4.83 5.00 5.00 4.77 4.92 4.82
[153] 4.74 4.73 5.00 5.00 5.00 5.00 5.00 5.00 4.95 4.83 4.75 5.00 4.93 4.89 4.93 4.87 0.00 4.87
[172] 4.87 4.80 4.82 4.74 4.94 4.78 4.86 4.73 4.81 4.67 5.00 4.68 4.50 5.00 0.00 0.00 4.69 4.92 5.00
[191] 5.00 0.00 4.95 0.00 5.00 4.88 5.00 5.00 4.92 4.56 4.94 4.85 4.85 4.72 5.00 4.88 4.66 4.68 4.00
[210] 5.00 5.00 0.00 5.00 4.75 0.00 4.00 0.00 0.00 4.86 4.86 4.87 4.83 4.60 4.88 4.96 4.74 5.00 4.70
[229] 4.56 4.67 4.40 5.00 0.00 4.60 4.58 4.60 4.64 4.79 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00
[248] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00
[267] 0.00 0.00 0.00 0.00 0.00 0.00 3.00 0.00 5.00 0.00 0.00 2.00 0.00 5.00 0.00 5.00 5.00 5.00 4.00
Environment History Connections Tutorial
Import Dataset 407 MiB
Global Environment
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:59 22-11-2023

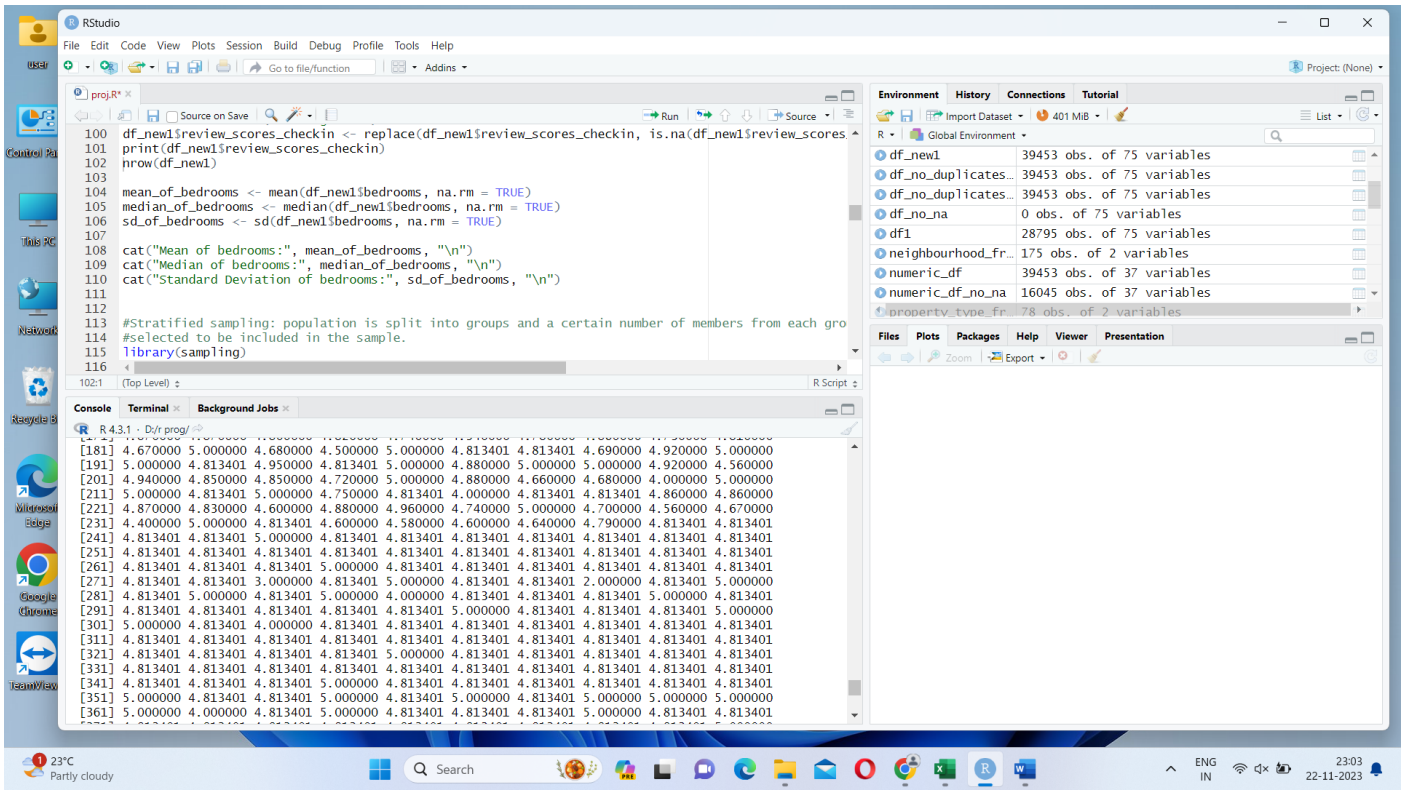

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
print(df_new$review_scores_checkin)
nrow(df_new)
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in mean value (replacing NA with mean)
Environment History Connections Tutorial
Import Dataset
Global Environment
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates 39453 obs. of 75 variables
df_no_duplicates 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
Files Plots Packages Help Viewer Presentation
Export
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[248] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00
[267] 0.00 0.00 0.00 0.00 0.00 0.00 3.00 0.00 5.00 0.00 0.00 2.00 0.00 5.00 0.00 5.00 0.00 5.00 4.00
[286] 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 5.00 5.00 0.00 4.00 0.00
[305] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[324] 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[343] 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00 5.00 0.00 5.00 5.00 5.00 5.00
[362] 4.00 0.00 5.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00
[381] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 4.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[400] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[419] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[438] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 4.00 0.00 0.00
[457] 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00
[476] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 4.00 0.00 0.00 0.00 0.00
[495] 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 3.50
[514] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[533] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00
[552] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[571] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00
[590] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 3.00 0.00 0.00 0.00 0.00
[609] 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[628] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[647] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[666] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00
[685] 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[704] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00
[723] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00
[742] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 5.00 0.00 4.98 5.00 0.00
[761] 4.73 4.62 4.66 4.75 4.75 0.00 0.00 4.93 0.00 5.00 4.69 4.59 4.80 4.29 4.25 5.00 5.00 4.71 4.25
[780] 4.88 4.86 4.67 4.57 4.67 4.60 4.78 5.00 4.81 4.75 4.80 5.00 4.75 0.00 4.33 0.00 4.77 4.68 4.74
[799] 0.00 5.00 4.76 4.44 4.97 4.70 4.62 4.64 4.59 4.50 4.83 4.51 4.86 4.78 4.59 4.25 4.68 4.78 4.75
[818] 0.00 4.70 4.83 4.51 4.33 4.78 4.68 4.86 4.65 4.70 4.73 4.61 4.77 4.71 4.43 4.57 4.58 4.90 4.44
[837] 4.69 4.78 4.91 4.88 4.67 5.00 4.63 4.64 4.83 0.00 5.00 4.95 5.00 4.86 4.97 4.85 5.00 4.98 4.75
[856] 4.90 4.83 4.95 4.96 4.86 4.95 4.81 4.91 4.90 4.91 4.91 4.89 4.88 4.99 4.98 5.00 5.00 4.93 5.00
[875] 4.92 4.96 0.00 0.00 0.00 4.91 0.00 0.00 4.56 5.00 4.91 5.00 5.00 4.98 4.80 4.92 4.88 5.00 4.54
[894] 5.00 0.00 4.97 5.00 4.63 5.00 0.00 5.00 0.00 4.60 0.00 4.85 4.62 4.85 4.65 4.55 4.59 4.64 4.68
[913] 4.80 4.89 4.73 4.89 4.80 5.00 4.76 4.79 4.84 4.98 5.00 4.78 4.97 5.00 5.00 4.94 4.93 4.92 4.70
[932] 4.82 4.86 5.00 4.82 4.76 4.68 4.69 4.92 4.95 4.99 4.93 0.00 3.87 4.94 4.93 4.96 4.97 4.58 4.91
[951] 4.93 0.00 4.77 4.87 4.74 4.96 4.94 4.97 4.93 5.00 4.95 5.00 5.00 4.32 4.91 5.00 5.00 4.97 4.90
23°C
Partly cloudy
Search
ENG
IN
23:00
22-11-2023

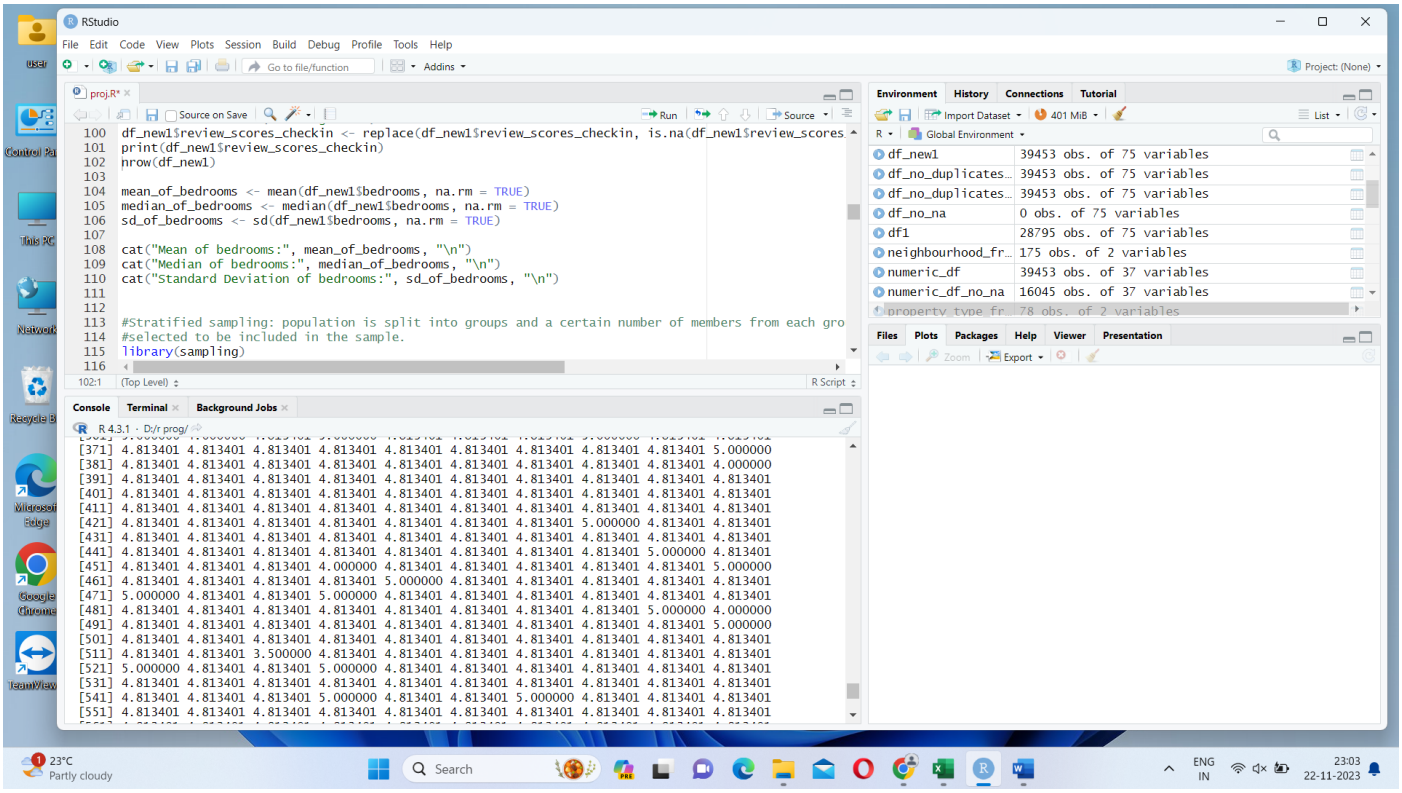

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
print(df_new$review_scores_checkin)
nrow(df_new)
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in mean value (replacing NA with mean)
Environment History Connections Tutorial
Import Dataset
Global Environment
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates 39453 obs. of 75 variables
df_no_duplicates 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
Files Plots Packages Help Viewer Presentation
Export
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[590] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 3.00 0.00 0.00 0.00 0.00
[609] 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[628] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[647] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[666] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00
[685] 0.00 5.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
[704] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 5.00 0.00
[723] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00
[742] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 5.00 0.00 4.98 5.00 0.00
[761] 4.73 4.62 4.66 4.75 4.75 0.00 0.00 4.93 0.00 5.00 4.69 4.59 4.80 4.29 4.25 5.00 5.00 4.71 4.25
[780] 4.88 4.86 4.67 4.57 4.67 4.60 4.78 5.00 4.81 4.75 4.80 5.00 4.75 0.00 4.33 0.00 4.77 4.68 4.74
[799] 0.00 5.00 4.76 4.44 4.97 4.70 4.62 4.64 4.59 4.50 4.83 4.51 4.86 4.78 4.59 4.25 4.68 4.78 4.75
[818] 0.00 4.70 4.83 4.51 4.33 4.78 4.68 4.86 4.65 4.70 4.73 4.61 4.77 4.71 4.43 4.57 4.58 4.90 4.44
[837] 4.69 4.78 4.91 4.88 4.67 5.00 4.63 4.64 4.83 0.00 5.00 4.95 5.00 4.86 4.97 4.85 5.00 4.98 4.75
[856] 4.90 4.83 4.95 4.96 4.86 4.95 4.81 4.91 4.90 4.91 4.91 4.89 4.88 4.99 4.98 5.00 5.00 4.93 5.00
[875] 4.92 4.96 0.00 0.00 0.00 4.91 0.00 0.00 4.56 5.00 4.91 5.00 5.00 4.98 4.80 4.92 4.88 5.00 4.54
[894] 5.00 0.00 4.97 5.00 4.63 5.00 0.00 5.00 0.00 4.60 0.00 4.85 4.62 4.85 4.65 4.55 4.59 4.64 4.68
[913] 4.80 4.89 4.73 4.89 4.80 5.00 4.76 4.79 4.84 4.98 5.00 4.78 4.97 5.00 5.00 4.94 4.93 4.92 4.70
[932] 4.82 4.86 5.00 4.82 4.76 4.68 4.69 4.92 4.95 4.99 4.93 0.00 3.87 4.94 4.93 4.96 4.97 4.58 4.91
[951] 4.93 0.00 4.77 4.87 4.74 4.96 4.94 4.97 4.93 5.00 4.95 5.00 5.00 4.32 4.91 5.00 5.00 4.97 4.90
23°C
Partly cloudy
Search
ENG
IN
23:00
22-11-2023

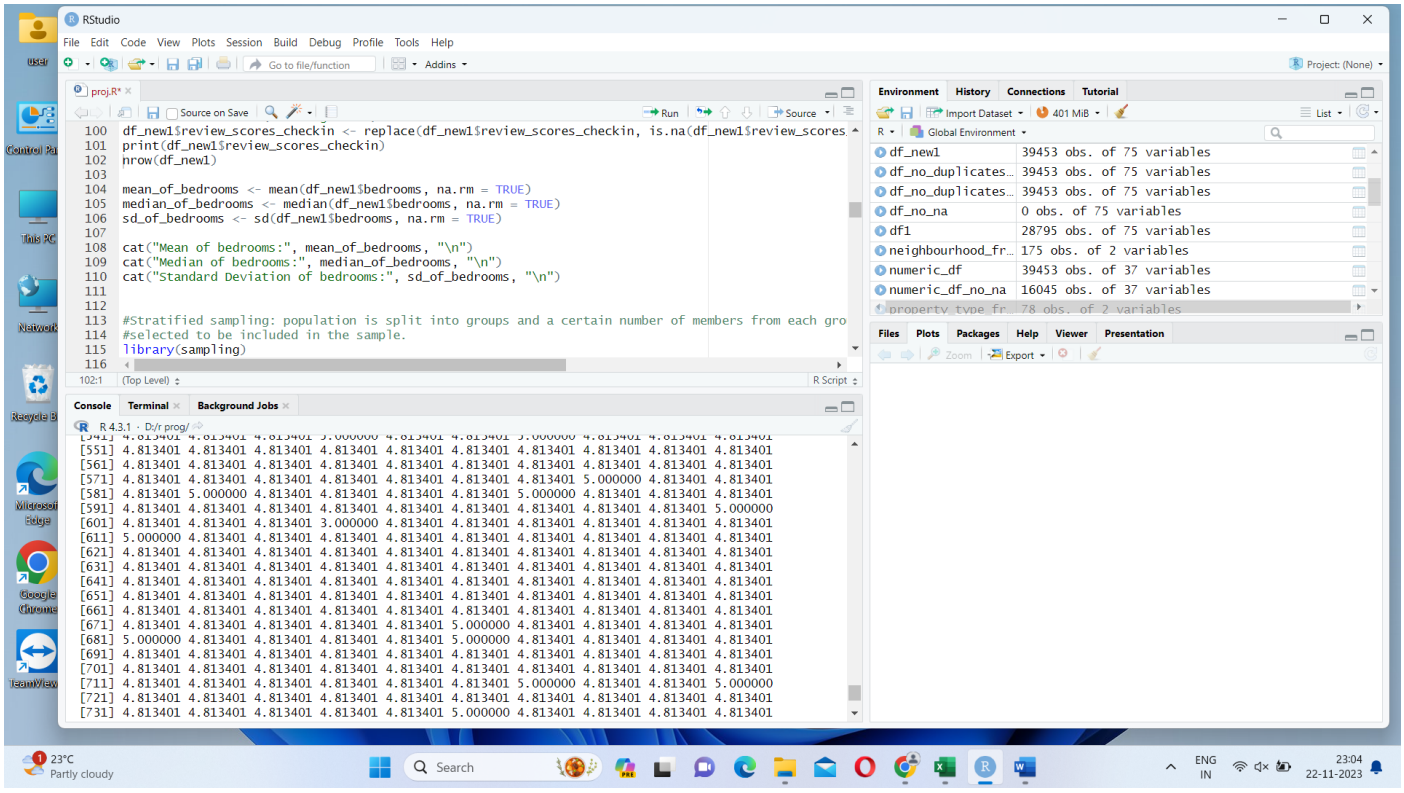

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
#Data cleaning - missing values - Filling in a constant value (replacing NA with 0)
df_new <- read.csv("listings.csv")
df_new$review_scores_checkin <- replace(df$review_scores_checkin, is.na(df$review_scores_checkin),
print(df_new$review_scores_checkin)
nrow(df_new)
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in mean value (replacing NA with mean)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[723] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 0.00 0.00 0.00 0.00
[742] 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 5.00 0.00 5.00 0.00 4.98 5.00 0.00
[761] 4.73 4.62 4.66 4.75 4.75 0.00 0.00 4.93 0.00 5.00 4.69 4.59 4.80 4.29 4.25 5.00 5.00 4.71 4.25
[780] 4.88 4.86 4.67 4.57 4.67 4.60 4.78 5.00 4.81 4.75 4.80 5.00 4.75 0.00 4.33 0.00 4.77 4.68 4.74
[799] 0.00 5.00 4.76 4.44 4.97 4.70 4.62 4.64 4.59 4.50 4.83 4.51 4.86 4.78 4.59 4.25 4.68 4.78 4.75
[818] 0.00 4.70 4.83 4.51 4.33 4.78 4.68 4.86 4.65 4.70 4.73 4.61 4.77 4.71 4.43 4.57 4.58 4.90 4.44
[837] 4.69 4.78 4.91 4.88 4.67 5.00 4.63 4.64 4.83 0.00 5.00 4.95 5.00 4.86 4.97 4.85 5.00 4.98 4.75
[856] 4.90 4.83 4.95 4.96 4.86 4.95 4.81 4.91 4.90 4.91 4.91 4.89 4.88 4.99 4.98 5.00 5.00 4.93 5.00
[875] 4.92 4.96 0.00 0.00 0.00 4.91 0.00 0.00 4.56 5.00 4.91 5.00 5.00 4.98 4.80 4.92 4.88 5.00 4.54
[894] 5.00 0.00 4.97 5.00 4.63 5.00 0.00 5.00 0.00 4.60 0.00 4.85 4.62 4.85 4.65 4.55 4.59 4.64 4.68
[913] 4.80 4.89 4.73 4.89 4.80 5.00 4.76 4.79 4.84 4.98 5.00 4.78 4.97 5.00 5.00 4.94 4.93 4.92 4.70
[932] 4.82 4.86 5.00 4.82 4.76 4.68 4.69 4.92 4.95 4.99 4.93 0.00 3.87 4.94 4.93 4.96 4.97 4.58 4.91
[951] 4.93 0.00 4.77 4.87 4.74 4.96 4.94 4.97 4.93 5.00 4.95 5.00 5.00 4.32 4.91 5.00 5.00 4.97 4.90
[970] 4.96 4.92 4.96 5.00 4.94 4.95 5.00 5.00 5.00 4.96 4.93 5.00 0.00 4.98 4.99 4.96 5.00 4.68 4.59
[989] 4.81 4.92 5.00 4.97 4.64 4.97 4.92 4.91 0.00 4.76 5.00 4.90
[reached getOption("max.print") -- omitted 38453 entries]
> nrow(df_new)
[1] 39453
Environment History Connections Tutorial
Import Dataset
407 MiB
Global Environment
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
23:01 22-11-2023

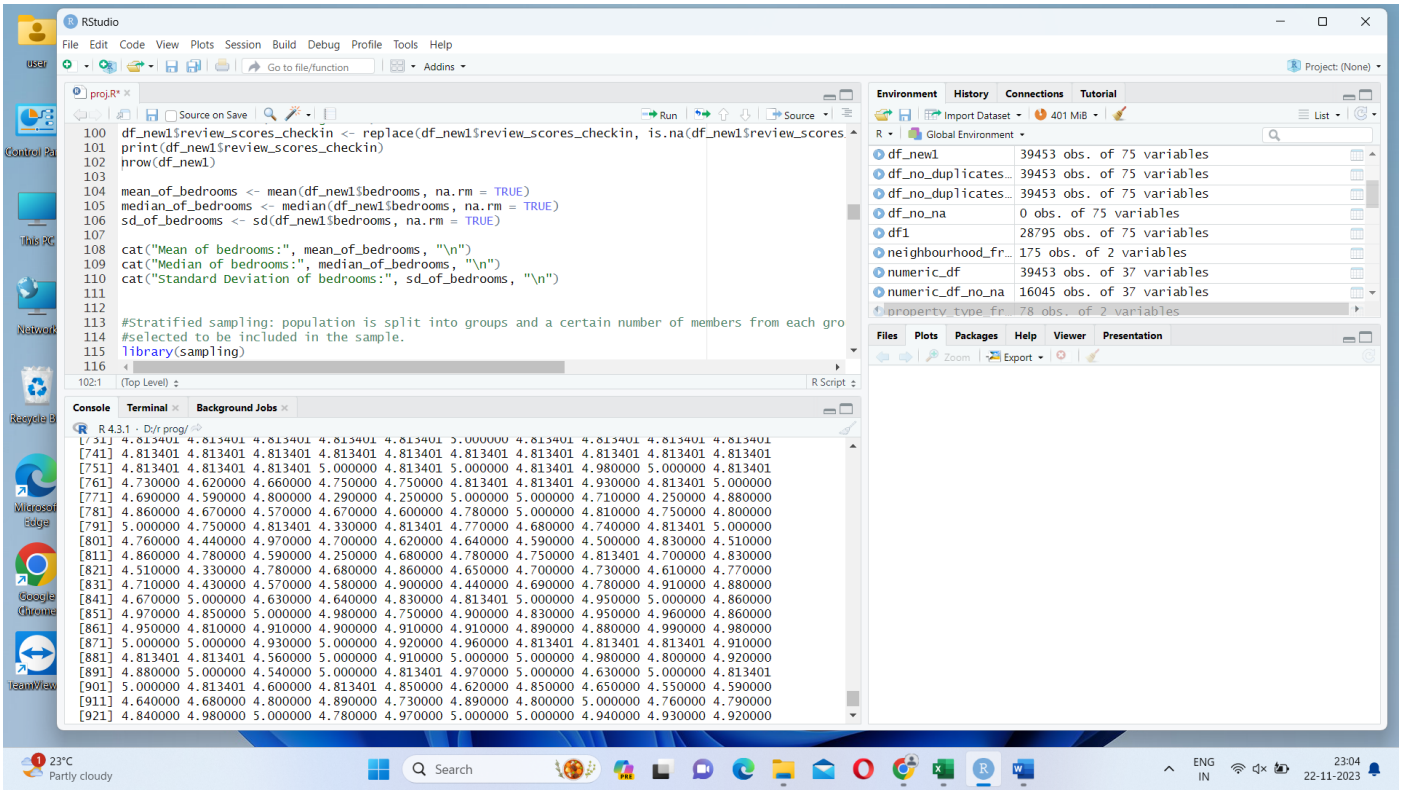

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Data cleaning - missing values - Filling in mean value (replacing NA with mean)
df_new1 <- read.csv("listings.csv")
df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
print(df_new1$review_scores_checkin)
nrow(df_new1)
mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[970] 4.96 4.92 4.96 5.00 4.94 4.95 5.00 5.00 5.00 4.96 4.93 5.00 0.00 4.98 4.99 4.96 5.00 4.68 4.59
[989] 4.81 4.92 5.00 4.97 4.64 4.97 4.92 4.91 0.00 4.76 5.00 4.90
[reached getOption("max.print") -- omitted 38453 entries]
> nrow(df_new)
[1] 39453
> mean_of_bedrooms <- mean(df_new$bedrooms, na.rm = TRUE)
> median_of_bedrooms <- median(df_new$bedrooms, na.rm = TRUE)
> sd_of_bedrooms <- sd(df_new$bedrooms, na.rm = TRUE)
> cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
Mean of bedrooms: 1.596986
> cat("Median of bedrooms:", median_of_bedrooms, "\n")
Median of bedrooms: 1
> cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
Standard Deviation of bedrooms: 0.9769223
> #Data cleaning - missing values - Filling in mean value (replacing NA with mean)
> df_new1 <- read.csv("listings.csv")
> df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores_chec
kin), mean(df_new1$review_scores_checkin, na.rm = TRUE))
Environment History Connections Tutorial
Import Dataset
409 MiB
Global Environment
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
23:01 22-11-2023

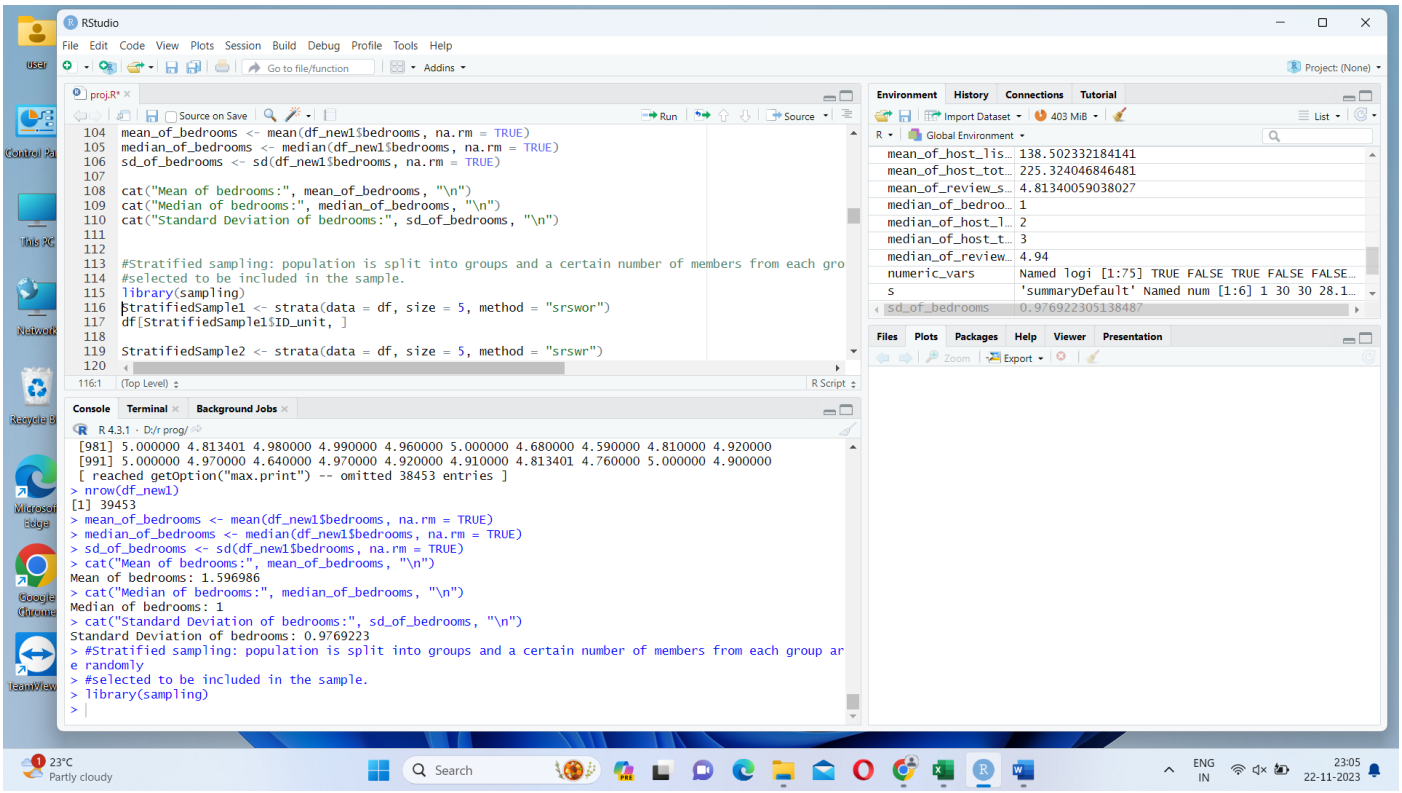

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
100 df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
101 print(df_new1$review_scores_checkin)
102 nrow(df_new1)
103
104 mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
105 median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
106 sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
107
108 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
109 cat("Median of bedrooms:", median_of_bedrooms, "\n")
110 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
111
112
113 #Stratified sampling: population is split into groups and a certain number of members from each gro
114 #selected to be included in the sample.
115 library(sampling)
116
102:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> print(df_new1$review_scores_checkin)
[1] 4.950000 4.813401 4.950000 4.680000 4.440000 5.000000 4.850000 4.813401 4.710000 5.000000
[11] 4.813401 4.830000 5.000000 4.813401 4.750000 5.000000 5.000000 4.670000 4.000000 4.380000
[21] 5.000000 5.000000 4.830000 4.750000 4.330000 4.750000 5.000000 4.920000 4.970000 4.813401
[31] 4.900000 4.813401 4.813401 4.940000 5.000000 5.000000 4.910000 4.960000 4.820000 4.790000
[41] 4.940000 4.900000 4.910000 4.880000 4.730000 4.450000 4.813401 4.510000 4.720000 4.790000
[51] 4.910000 4.880000 5.000000 5.000000 4.813401 5.000000 4.820000 4.500000 5.000000 4.750000
[61] 5.000000 4.900000 4.880000 4.980000 4.800000 4.930000 4.813401 5.000000 4.860000 4.813401
[71] 5.000000 5.000000 5.000000 5.000000 5.000000 4.813401 4.500000 5.000000 5.000000
[81] 4.110000 5.000000 4.813401 4.813401 4.813401 4.800000 5.000000 4.780000 4.920000 4.780000
[91] 4.813401 4.813401 4.813401 5.000000 5.000000 4.860000 4.840000 4.790000 4.940000 4.820000
[101] 4.750000 4.960000 4.950000 4.840000 4.790000 4.940000 4.940000 4.970000 5.000000 4.000000
[111] 4.900000 4.290000 4.500000 5.000000 5.000000 5.000000 4.890000 5.000000 4.813401 4.980000
[121] 4.970000 4.880000 5.000000 4.860000 4.850000 4.730000 4.813401 4.813401 4.710000 4.950000
[131] 5.000000 4.400000 4.760000 4.660000 4.790000 4.900000 4.813401 4.813401 5.000000 4.500000
[141] 5.000000 5.000000 5.000000 4.870000 4.890000 5.000000 4.830000 5.000000 5.000000 4.770000
[151] 4.920000 4.820000 4.740000 4.730000 5.000000 5.000000 5.000000 5.000000 5.000000 5.000000
[161] 5.000000 4.950000 4.830000 4.750000 5.000000 4.930000 4.890000 4.930000 4.870000 4.813401
[171] 4.870000 4.870000 4.800000 4.820000 4.740000 4.940000 4.780000 4.860000 4.730000 4.810000
[181] 4.670000 5.000000 4.680000 4.500000 5.000000 4.813401 4.813401 4.690000 4.920000 5.000000
Environment History Connections Tutorial
Import Dataset
401 MiB
List
R
Global Environment
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Zoom
Export
23°C
Partly cloudy
Search
ENG
IN
23:02
22-11-2023

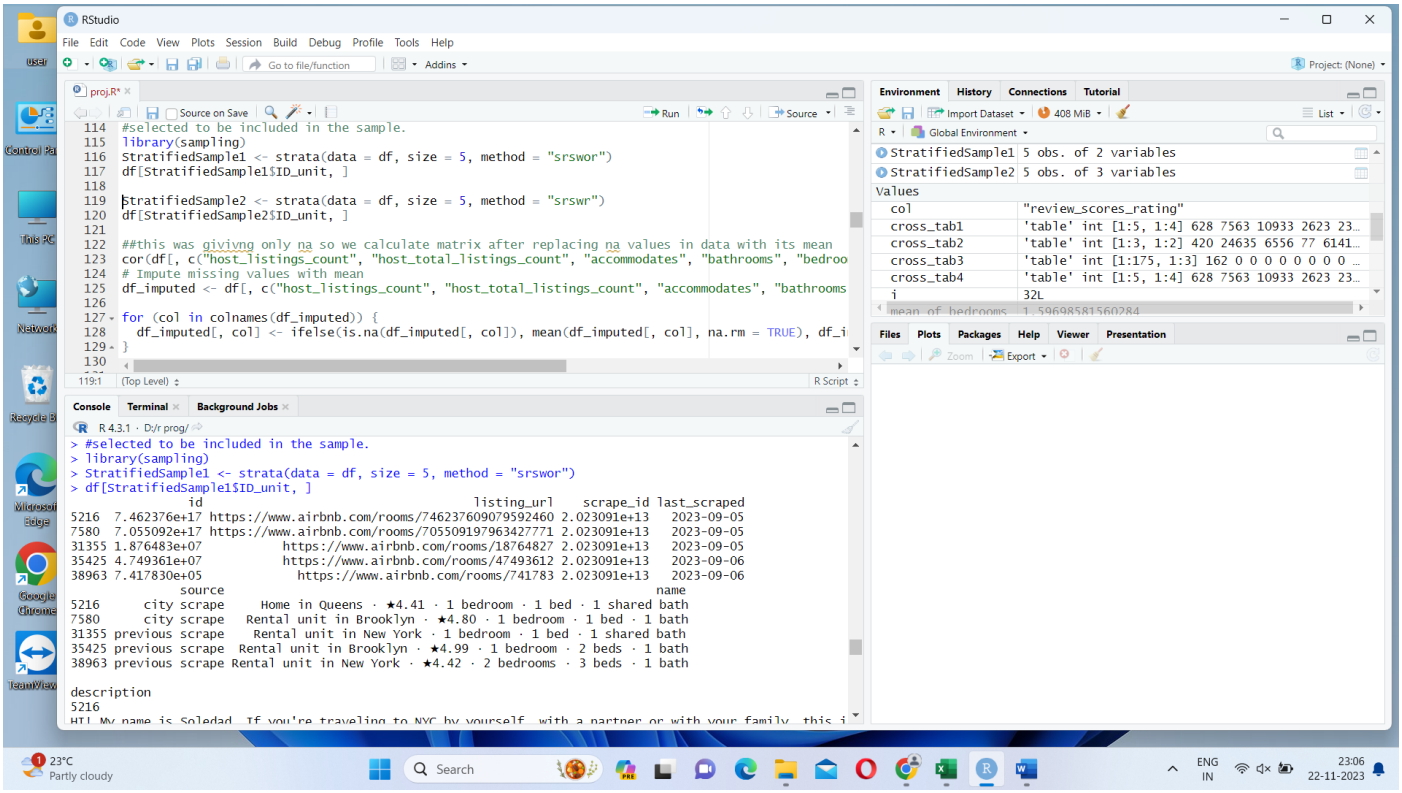

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
100 df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
101 print(df_new1$review_scores_checkin)
102 nrow(df_new1)
103
104 mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
105 median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
106 sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
107
108 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
109 cat("Median of bedrooms:", median_of_bedrooms, "\n")
110 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
111
112
113 #Stratified sampling: population is split into groups and a certain number of members from each gro
114 #selected to be included in the sample.
115 library(sampling)
116
102:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[181] 4.670000 5.000000 4.680000 4.500000 5.000000 4.813401 4.813401 4.690000 4.920000 5.000000
[191] 5.000000 4.813401 4.950000 4.813401 5.000000 4.880000 5.000000 5.000000 4.920000 4.560000
[201] 4.940000 4.850000 4.850000 4.720000 5.000000 4.880000 4.660000 4.680000 4.000000 5.000000
[211] 5.000000 4.813401 5.000000 4.750000 4.813401 4.000000 4.813401 4.813401 4.860000 4.860000
[221] 4.870000 4.830000 4.600000 4.880000 4.960000 4.740000 5.000000 4.700000 4.560000 4.670000
[231] 4.400000 5.000000 4.813401 4.600000 4.580000 4.600000 4.640000 4.790000 4.813401 4.813401
[241] 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[251] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[261] 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[271] 4.813401 4.813401 3.000000 4.813401 5.000000 4.813401 4.813401 2.000000 4.813401 5.000000
[281] 4.813401 5.000000 4.813401 5.000000 4.000000 4.813401 4.813401 4.813401 5.000000 4.813401
[291] 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 5.000000
[301] 5.000000 4.813401 4.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[311] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[321] 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401
[331] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[341] 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[351] 5.000000 4.813401 4.813401 5.000000 4.813401 5.000000 4.813401 5.000000 5.000000 5.000000
[361] 5.000000 4.000000 4.813401 5.000000 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401
Environment History Connections Tutorial
Import Dataset
401 MiB
List
R
Global Environment
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Zoom
Export
23°C
Partly cloudy
Search
ENG
IN
23:03
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
100 df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
101 print(df_new1$review_scores_checkin)
102 nrow(df_new1)
103
104 mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
105 median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
106 sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
107
108 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
109 cat("Median of bedrooms:", median_of_bedrooms, "\n")
110 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
111
112
113 #Stratified sampling: population is split into groups and a certain number of members from each gro
114 #selected to be included in the sample.
115 library(sampling)
116
Environment History Connections Tutorial
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[371] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000
[381] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.000000
[391] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[401] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[411] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[421] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401
[431] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[441] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401
[451] 4.813401 4.813401 4.813401 4.000000 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401
[461] 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401
[471] 5.000000 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[481] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.000000
[491] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000
[501] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[511] 4.813401 4.813401 3.500000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[521] 5.000000 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[531] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[541] 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 5.000000 4.813401 4.813401
[551] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
23°C Partly cloudy
Search
ENG IN
23:03 22-11-2023

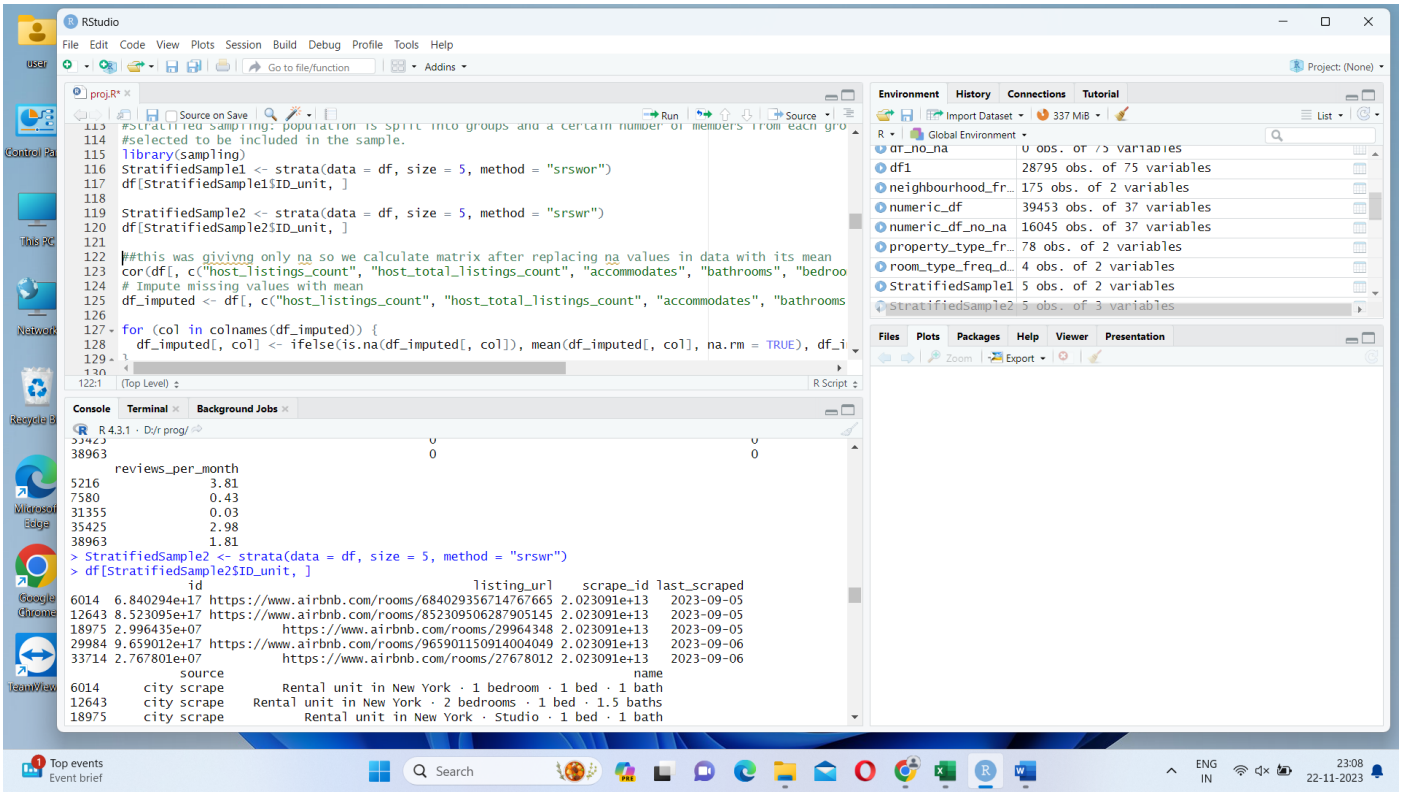

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
100 df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
101 print(df_new1$review_scores_checkin)
102 nrow(df_new1)
103
104 mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
105 median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
106 sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
107
108 cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
109 cat("Median of bedrooms:", median_of_bedrooms, "\n")
110 cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
111
112
113 #Stratified sampling: population is split into groups and a certain number of members from each gro
114 #selected to be included in the sample.
115 library(sampling)
116
Environment History Connections Tutorial
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
[551] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[561] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[571] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401
[581] 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401
[591] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000
[601] 4.813401 4.813401 4.813401 3.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[611] 5.000000 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[621] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[631] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[641] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[651] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[661] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[671] 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401
[681] 5.000000 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401
[691] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[701] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[711] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 5.000000
[721] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[731] 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401
23°C Partly cloudy
Search
ENG IN
23:04 22-11-2023

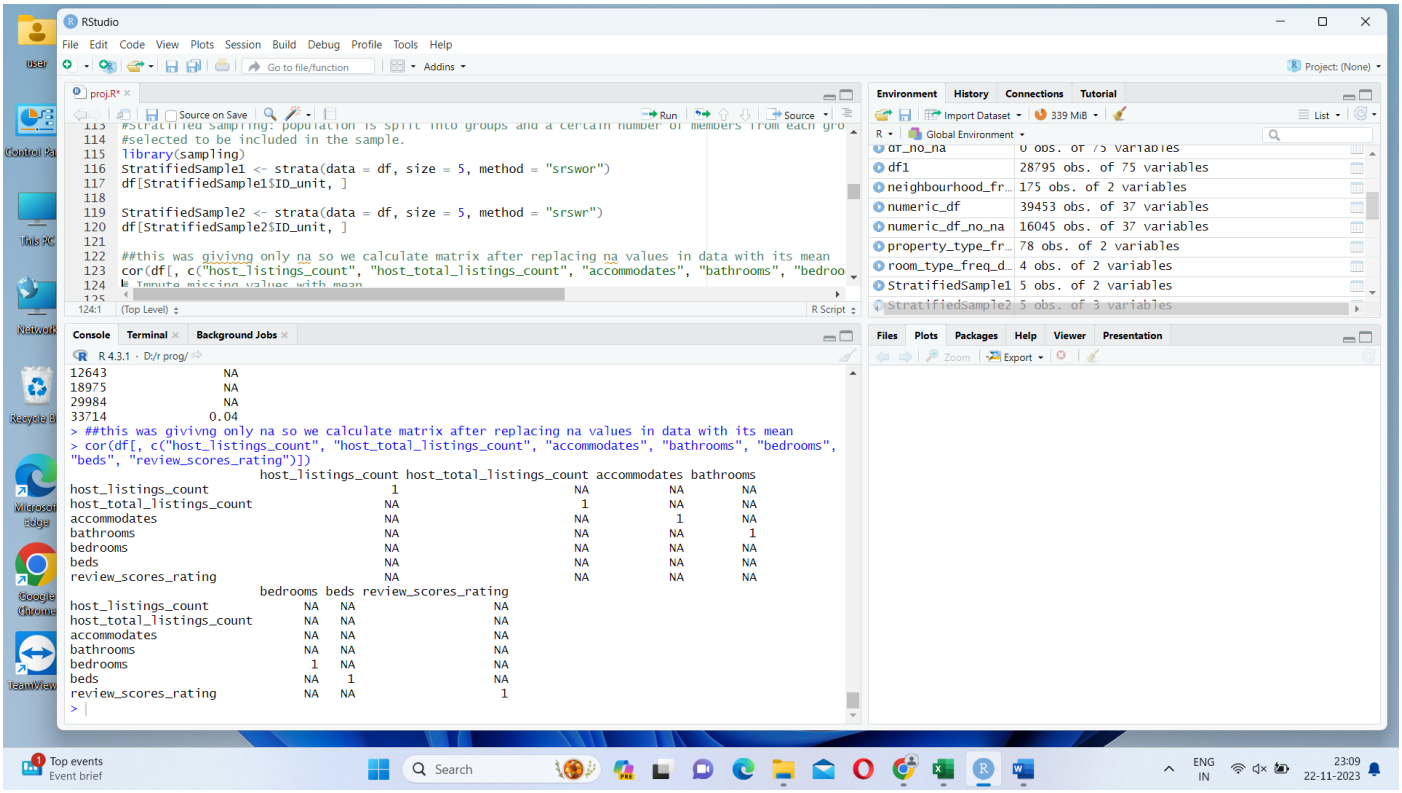

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
df_new1$review_scores_checkin <- replace(df_new1$review_scores_checkin, is.na(df_new1$review_scores
print(df_new1$review_scores_checkin)
nrow(df_new1)
mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Stratified sampling: population is split into groups and a certain number of members from each gro
#selected to be included in the sample.
library(sampling)
102:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.1 · D:/r prog/
[731] 4.813401 4.813401 4.813401 4.813401 4.813401 5.000000 4.813401 4.813401 4.813401 4.813401
[741] 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401 4.813401
[751] 4.813401 4.813401 4.813401 5.000000 4.813401 5.000000 4.813401 4.980000 5.000000 4.813401
[761] 4.730000 4.620000 4.660000 4.750000 4.750000 4.813401 4.813401 4.930000 4.813401 5.000000
[771] 4.690000 4.590000 4.800000 4.290000 4.250000 5.000000 5.000000 4.710000 4.250000 4.880000
[781] 4.860000 4.670000 4.570000 4.670000 4.600000 4.780000 5.000000 4.810000 4.750000 4.800000
[791] 5.000000 4.750000 4.813401 4.330000 4.813401 4.770000 4.680000 4.740000 4.813401 5.000000
[801] 4.760000 4.440000 4.970000 4.700000 4.620000 4.640000 4.590000 4.500000 4.830000 4.510000
[811] 4.860000 4.780000 4.590000 4.250000 4.680000 4.780000 4.750000 4.813401 4.700000 4.830000
[821] 4.510000 4.330000 4.780000 4.680000 4.860000 4.650000 4.700000 4.730000 4.610000 4.770000
[831] 4.710000 4.430000 4.570000 4.580000 4.900000 4.440000 4.690000 4.780000 4.910000 4.880000
[841] 4.670000 5.000000 4.630000 4.640000 4.830000 4.813401 5.000000 4.950000 5.000000 4.860000
[851] 4.970000 4.850000 5.000000 4.980000 4.750000 4.900000 4.830000 4.950000 4.960000 4.860000
[861] 4.950000 4.810000 4.910000 4.900000 4.910000 4.910000 4.890000 4.880000 4.990000 4.980000
[871] 5.000000 5.000000 4.930000 5.000000 4.920000 4.960000 4.813401 4.813401 4.813401 4.910000
[881] 4.813401 4.813401 4.560000 5.000000 4.910000 5.000000 5.000000 4.980000 4.800000 4.920000
[891] 4.880000 5.000000 4.540000 5.000000 4.813401 4.970000 5.000000 4.630000 5.000000 4.813401
[901] 5.000000 4.813401 4.600000 4.813401 4.850000 4.620000 4.850000 4.650000 4.550000 4.590000
[911] 4.640000 4.680000 4.800000 4.890000 4.730000 4.890000 4.800000 5.000000 4.760000 4.790000
[921] 4.840000 4.980000 5.000000 4.780000 4.970000 5.000000 5.000000 4.940000 4.930000 4.920000
Environment
History
Connections
Tutorial
Import Dataset
401 MiB
List
R
Global Environment
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
23°C
Partly cloudy
Search
ENG
IN
23:04
22-11-2023

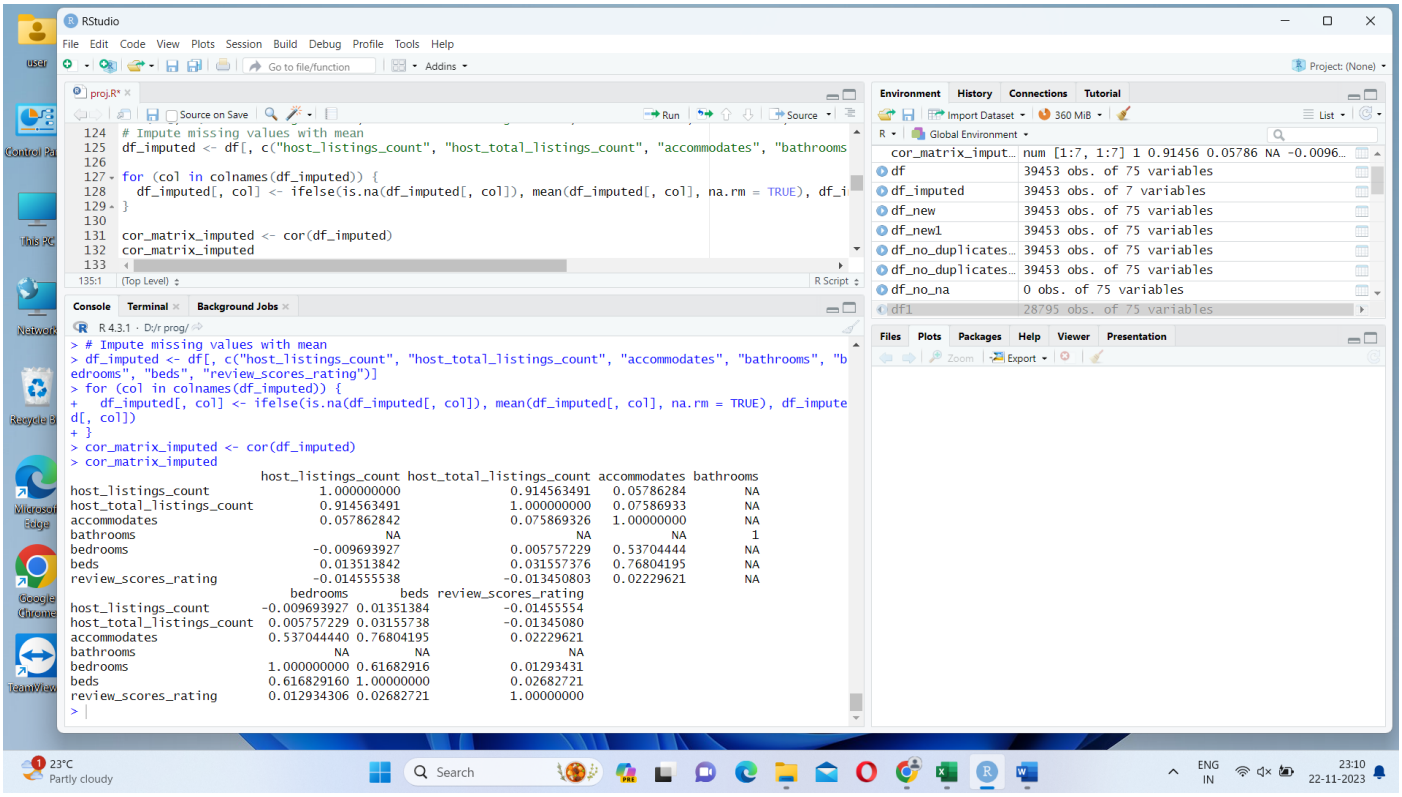

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
cat("Median of bedrooms:", median_of_bedrooms, "\n")
cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
#Stratified sampling: population is split into groups and a certain number of members from each gro
#selected to be included in the sample.
library(sampling)
stratifiedSample1 <- strata(data = df, size = 5, method = "srswor")
df[stratifiedSample1$ID_unit,]
StratifiedSample2 <- strata(data = df, size = 5, method = "srswr")
116:1
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.3.1 · D:/r prog/
[981] 5.000000 4.813401 4.980000 4.990000 4.960000 5.000000 4.680000 4.590000 4.810000 4.920000
[991] 5.000000 4.970000 4.640000 4.970000 4.920000 4.910000 4.813401 4.760000 5.000000 4.900000
[reached getOption("max.print") -- omitted 38453 entries]
> nrow(df_new1)
[1] 39453
> mean_of_bedrooms <- mean(df_new1$bedrooms, na.rm = TRUE)
> median_of_bedrooms <- median(df_new1$bedrooms, na.rm = TRUE)
> sd_of_bedrooms <- sd(df_new1$bedrooms, na.rm = TRUE)
> cat("Mean of bedrooms:", mean_of_bedrooms, "\n")
Mean of bedrooms: 1.596986
> cat("Median of bedrooms:", median_of_bedrooms, "\n")
Median of bedrooms: 1
> cat("Standard Deviation of bedrooms:", sd_of_bedrooms, "\n")
Standard Deviation of bedrooms: 0.9769223
> #Stratified sampling: population is split into groups and a certain number of members from each group ar
e randomly
> #selected to be included in the sample.
> library(sampling)
>
Environment
History
Connections
Tutorial
Import Dataset
403 MiB
List
R
Global Environment
mean_of_host_lis... 138.502332184141
mean_of_host_tot... 225.324046846481
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.976922305138487
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
23°C
Partly cloudy
Search
ENG
IN
23:05
22-11-2023

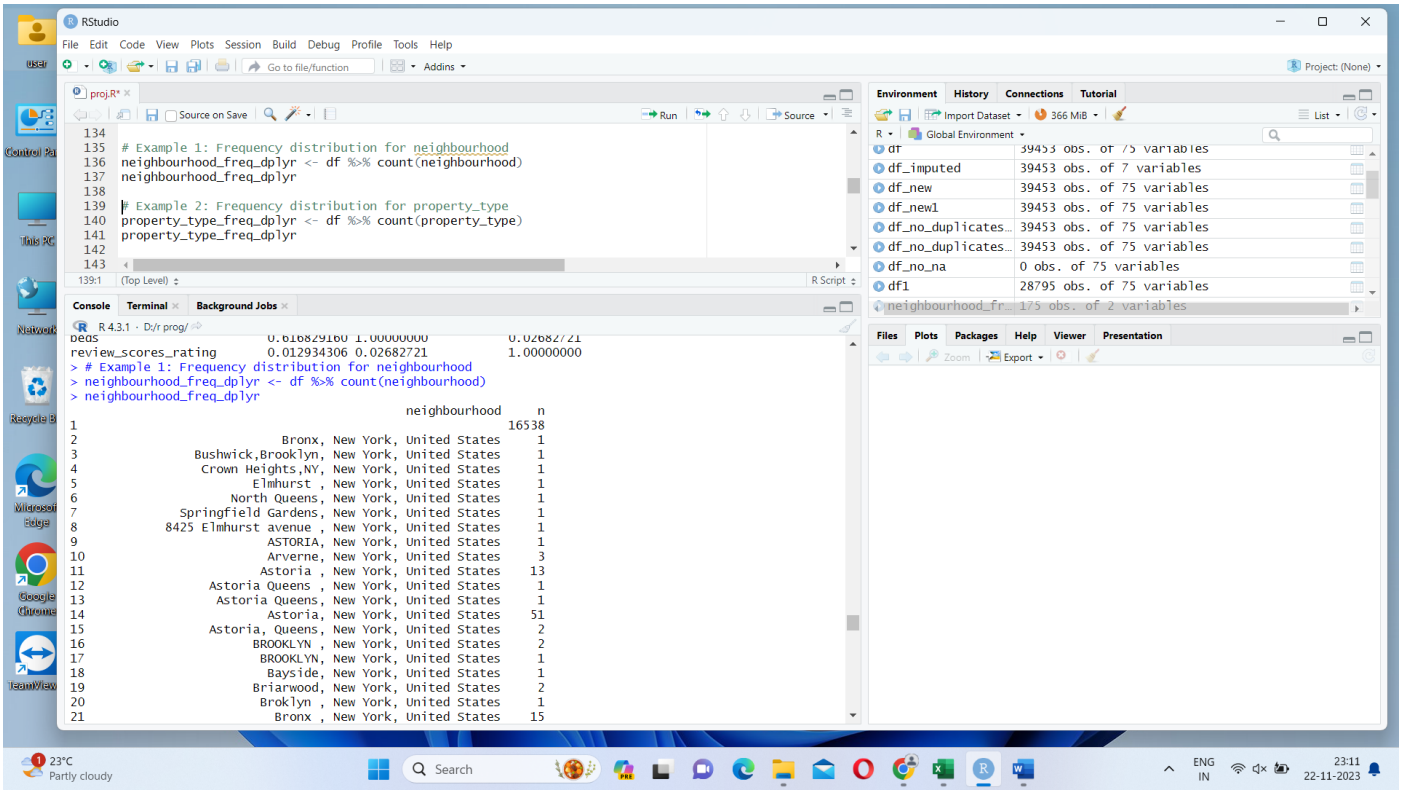


5 random rows are printed all screenshots of row data printed is not included in both sample 1 and 2

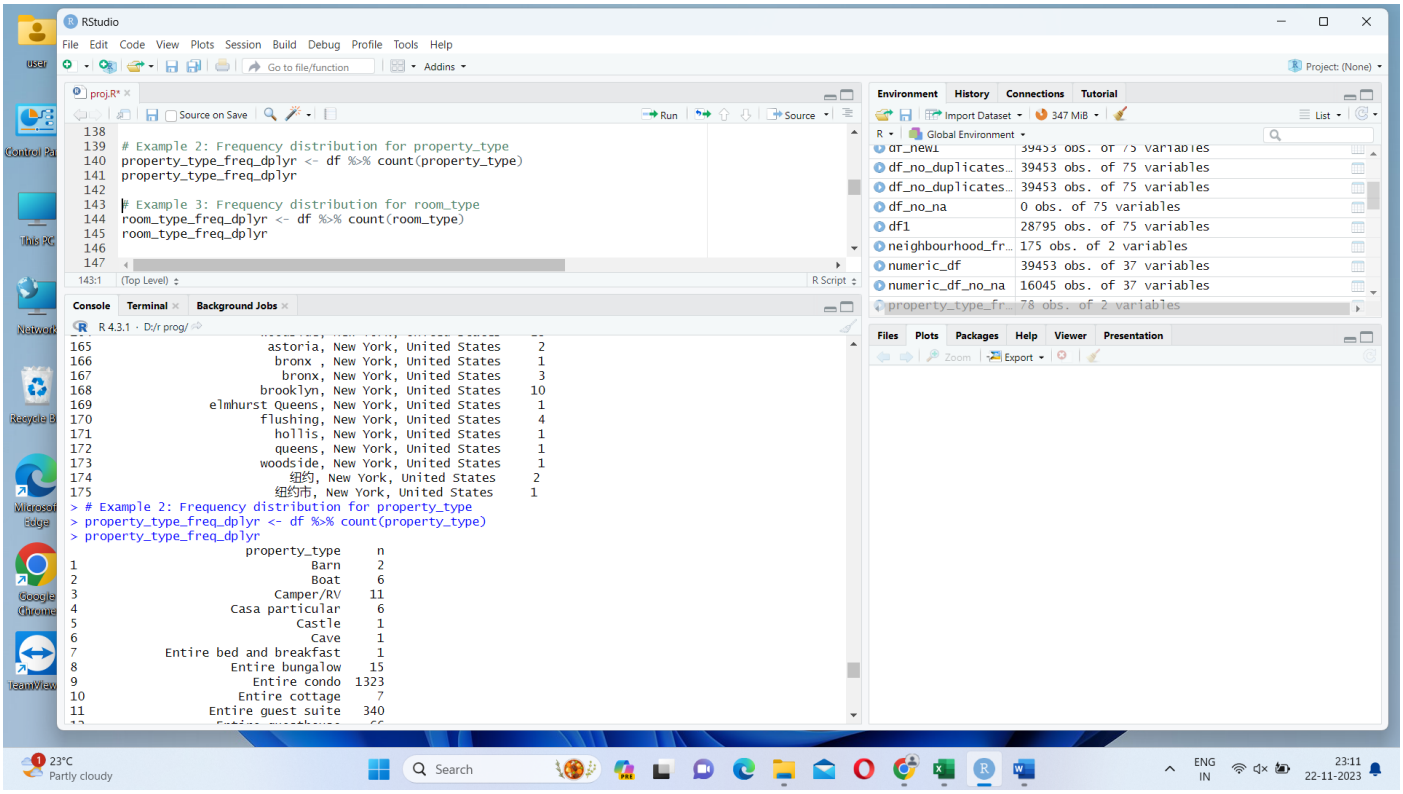



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
#selected to be included in the sample.
library(sampling)
StratifiedSample1 <- strata(data = df, size = 5, method = "srswor")
df[StratifiedSample1$ID_unit,]
StratifiedSample2 <- strata(data = df, size = 5, method = "srswr")
df[StratifiedSample2$ID_unit,]
##this was givivng only na so we calculate matrix after replacing na values in data with its mean
cor(df[, c("host_listings_count", "host_total_listings_count", "accommodates", "bathrooms", "bedroo
Impute missing values with mean
Environment History Connections Tutorial
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
numeric_df 39453 obs. of 37 variables
numeric_df_no_na 16045 obs. of 37 variables
property_type_fr... 78 obs. of 2 variables
room_type_freq_d... 4 obs. of 2 variables
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
12643 NA
18975 NA
29984 NA
33714 0.04
> ##this was givivng only na so we calculate matrix after replacing na values in data with its mean
> cor(df[, c("host_listings_count", "host_total_listings_count", "accommodates", "bathrooms", "bedrooms",
"beds", "review_scores_rating")])
host_listings_count host_total_listings_count accommodates bathrooms
host_listings_count 1 NA NA NA
host_total_listings_count NA 1 NA NA
accommodates NA NA 1 NA
bathrooms NA NA NA 1
bedrooms NA NA NA NA
beds NA NA NA NA
review_scores_rating NA NA NA NA
bedrooms beds review_scores_rating
host_listings_count NA NA NA
host_total_listings_count NA NA NA
accommodates NA NA NA
bathrooms NA NA NA
bedrooms 1 NA NA
beds NA 1 NA
review_scores_rating NA NA 1
Files Plots Packages Help Viewer Presentation
Top events
Event brief
Search
ENG
IN
23:09
22-11-2023

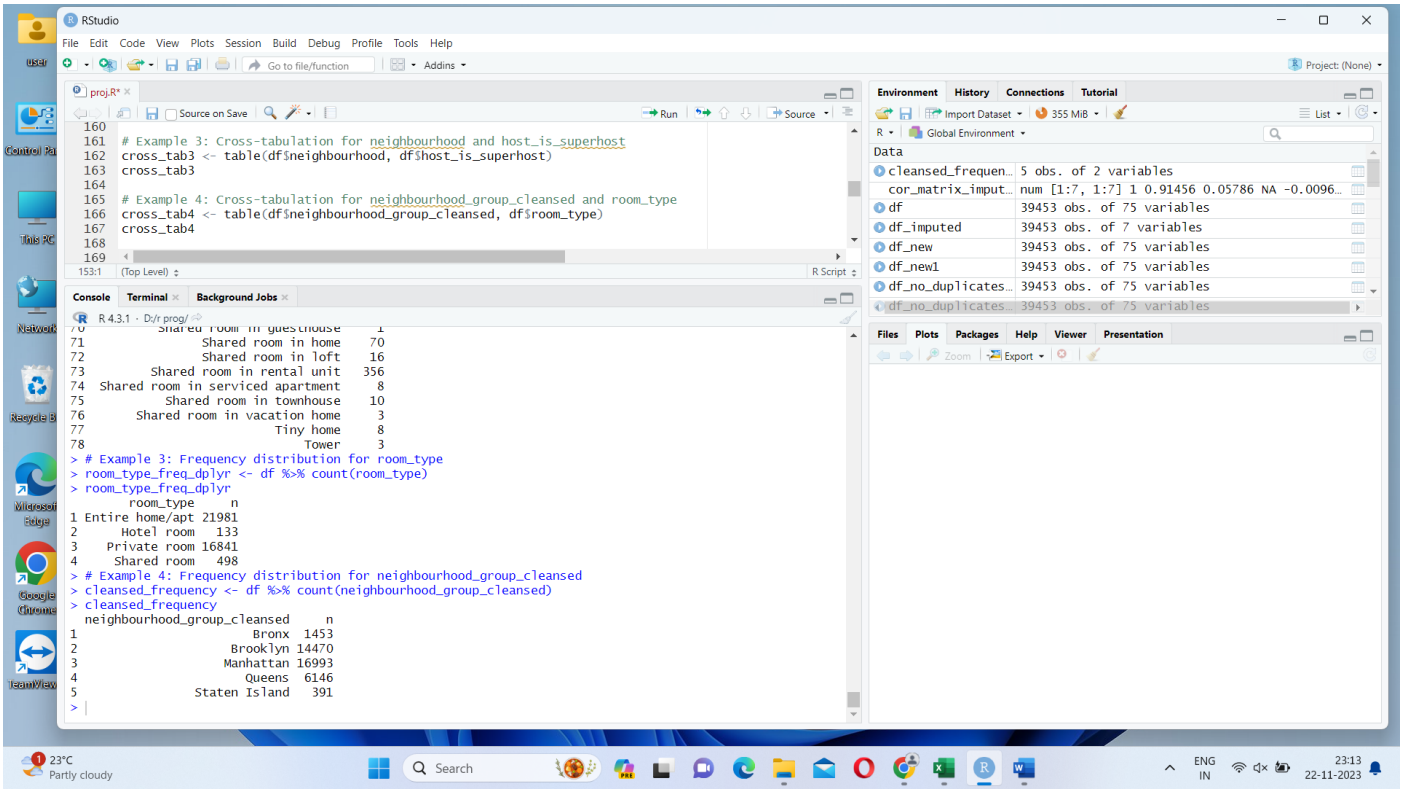

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
# Impute missing values with mean
df_imputed <- df[, c("host_listings_count", "host_total_listings_count", "accommodates", "bathrooms
for (col in colnames(df_imputed)) {
df_imputed[, col] <- ifelse(is.na(df_imputed[, col]), mean(df_imputed[, col], na.rm = TRUE), df_i
}
cor_matrix_imputed <- cor(df_imputed)
cor_matrix_imputed
Environment History Connections Tutorial
cor_matrix_imput... num [1:7, 1:7] 1 0.91456 0.05786 NA -0.0096...
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> # Impute missing values with mean
> df_imputed <- df[, c("host_listings_count", "host_total_listings_count", "accommodates", "bathrooms", "b
edrooms", "beds", "review_scores_rating")]
> for (col in colnames(df_imputed)) {
+ df_imputed[, col] <- ifelse(is.na(df_imputed[, col]), mean(df_imputed[, col], na.rm = TRUE), df_impute
d[, col])
+ }
> cor_matrix_imputed <- cor(df_imputed)
> cor_matrix_imputed
host_listings_count host_total_listings_count accommodates bathrooms
host_listings_count 1.000000000 0.914563491 0.05786284 NA
host_total_listings_count 0.914563491 1.000000000 0.07586933 NA
accommodates 0.057862842 0.075869326 1.00000000 NA
bathrooms NA NA NA 1
bedrooms -0.009693927 0.005757229 0.53704444 NA
beds 0.013513842 0.031557376 0.76804195 NA
review_scores_rating -0.014555538 -0.013450803 0.02229621 NA
bedrooms beds review_scores_rating
host_listings_count -0.009693927 0.01351384 -0.01455554
host_total_listings_count 0.005757229 0.03155738 -0.01345080
accommodates 0.537044440 0.76804195 0.02229621
bathrooms NA NA NA
bedrooms 1.000000000 0.61682916 0.01293431
beds 0.616829160 1.00000000 0.02682721
review_scores_rating 0.012934306 0.02682721 1.00000000
Files Plots Packages Help Viewer Presentation
23°C
Partly cloudy
Search
ENG
IN
23:10
22-11-2023



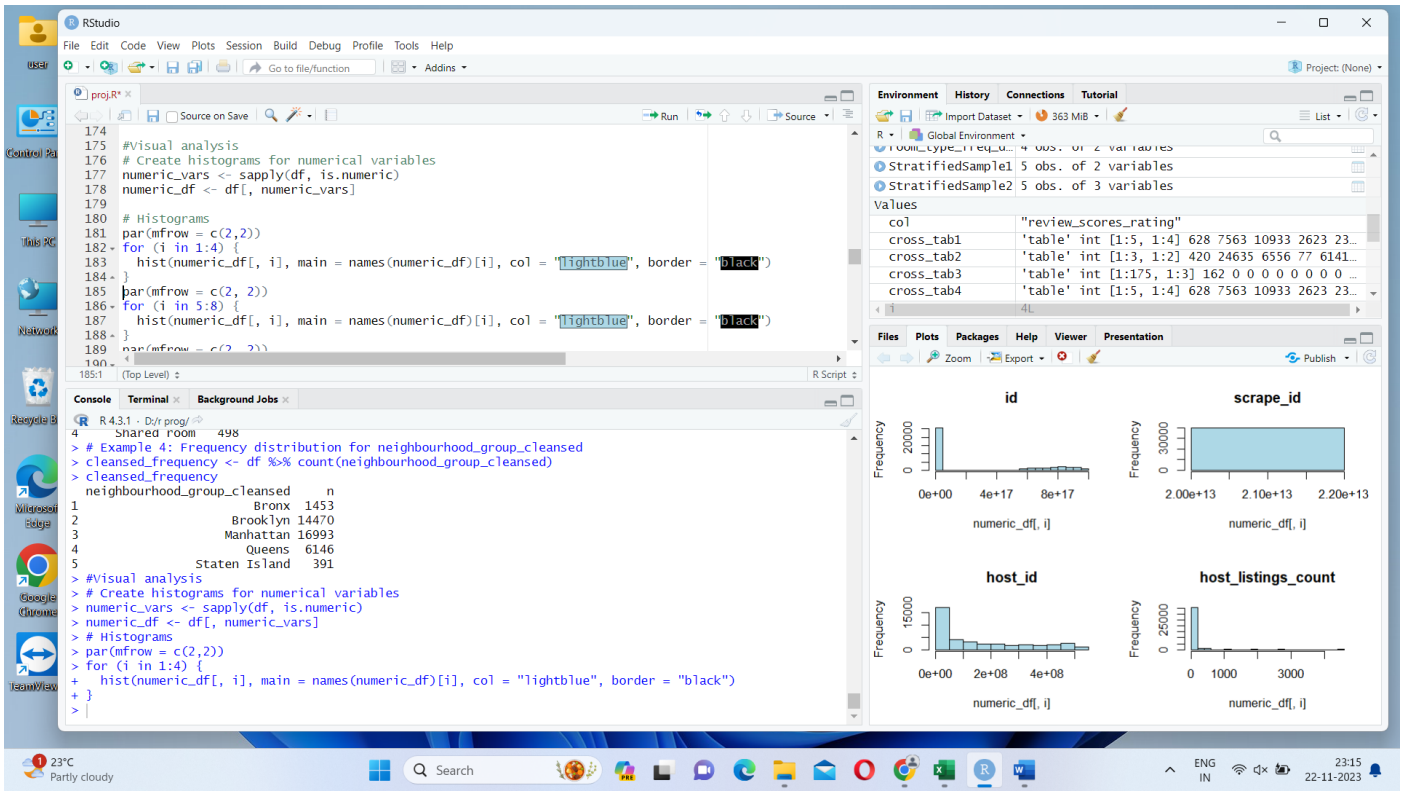
all 175 rows not included in screenshot



all 78 rows not included in screenshots



Visual Analysis

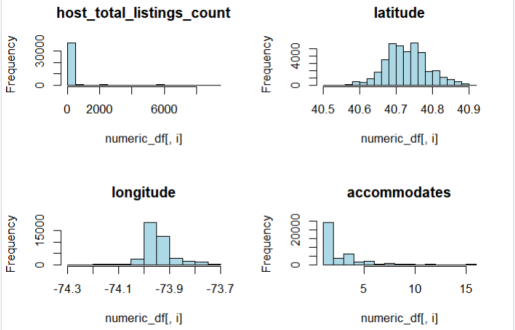


```
numeric_df <- df[, numeric_vars]

# Histograms
par(mfrow = c(2,2))
for (i in 1:4) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 5:8) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 9:12) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
```

```
  neighbourhood_group_cleansed     n
1                        Bronx  1453
2                     Brooklyn 14470
3                    Manhattan 16993
4                       Queens  6146
5                Staten Island   391
> #Visual analysis
> # Create histograms for numerical variables
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Histograms
> par(mfrow = c(2,2))
> for (i in 1:4) {
+   hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 5:8) {
+   hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
>
```

col	"review_scores_rating"
cross_tab1	'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2	'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3	'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4	'table' int [1:5, 1:4] 628 7563 10933 2623 23…

StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables

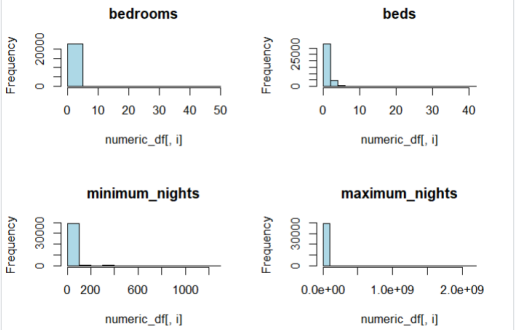


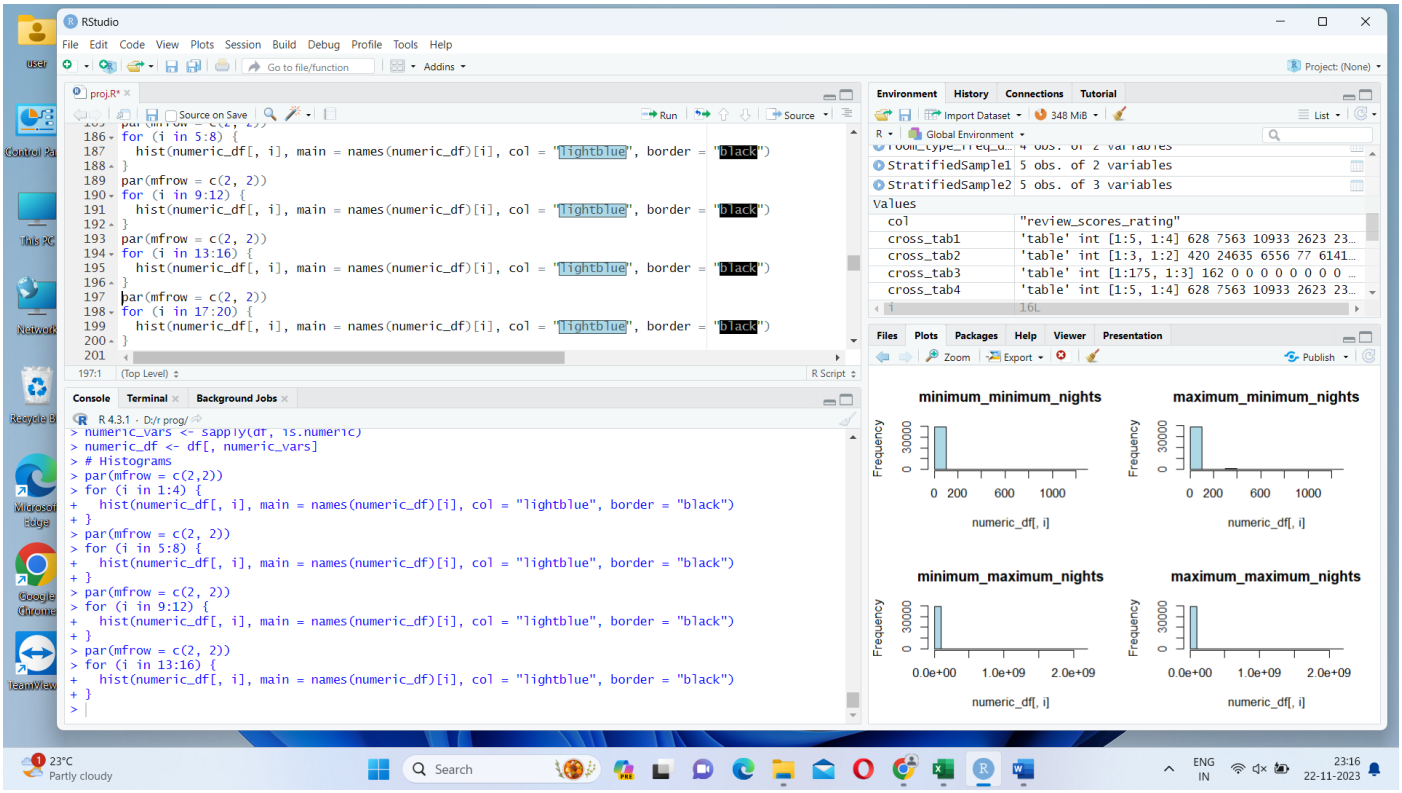

```
numeric_df <- df[, numeric_vars]

# Histograms
par(mfrow = c(2,2))
for (i in 1:4) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 5:8) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 9:12) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 13:16) {
  hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
```

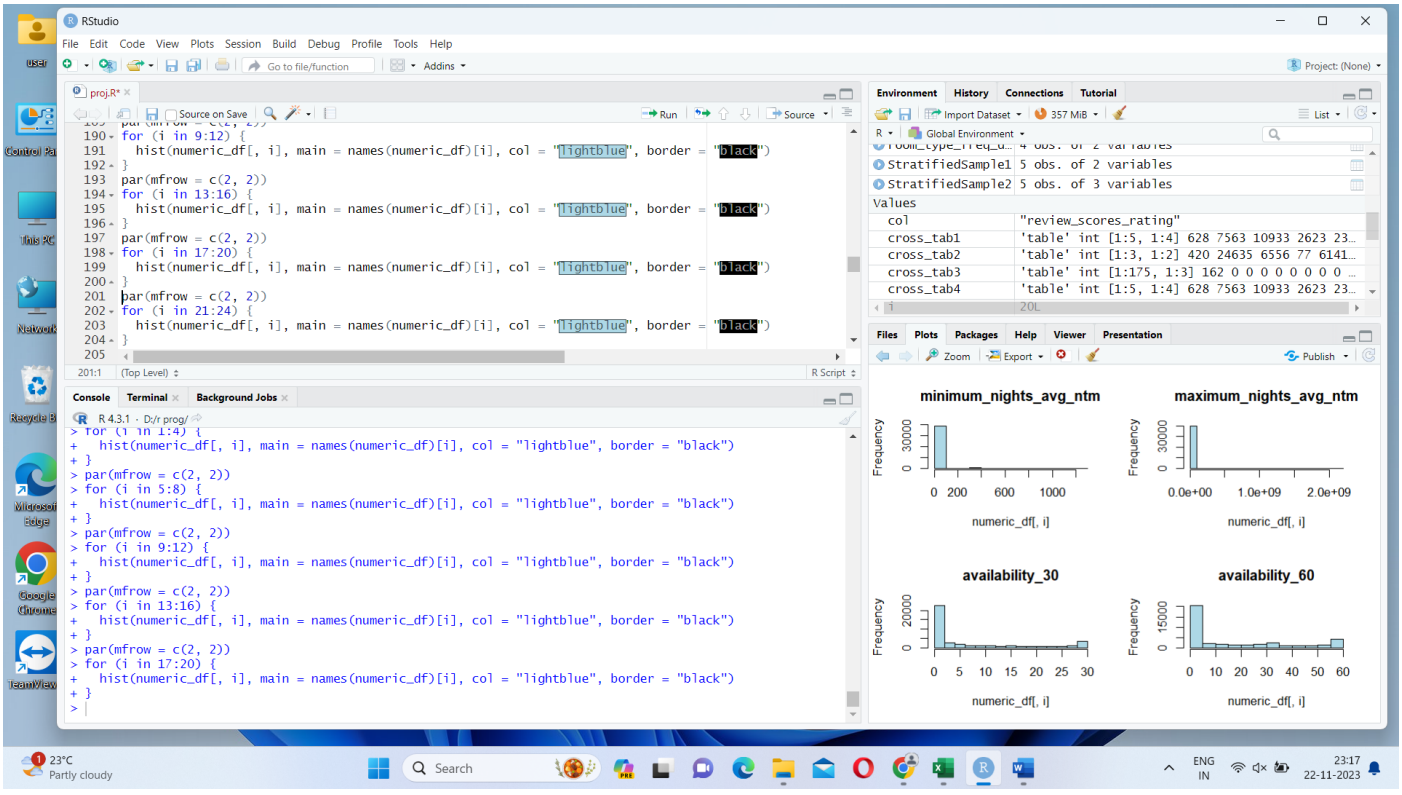
```
4                       Queens  6146
5                Staten Island   391
> #Visual analysis
> # Create histograms for numerical variables
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Histograms
> par(mfrow = c(2,2))
> for (i in 1:4) {
+   hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 5:8) {
+   hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 9:12) {
+   hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
>
```

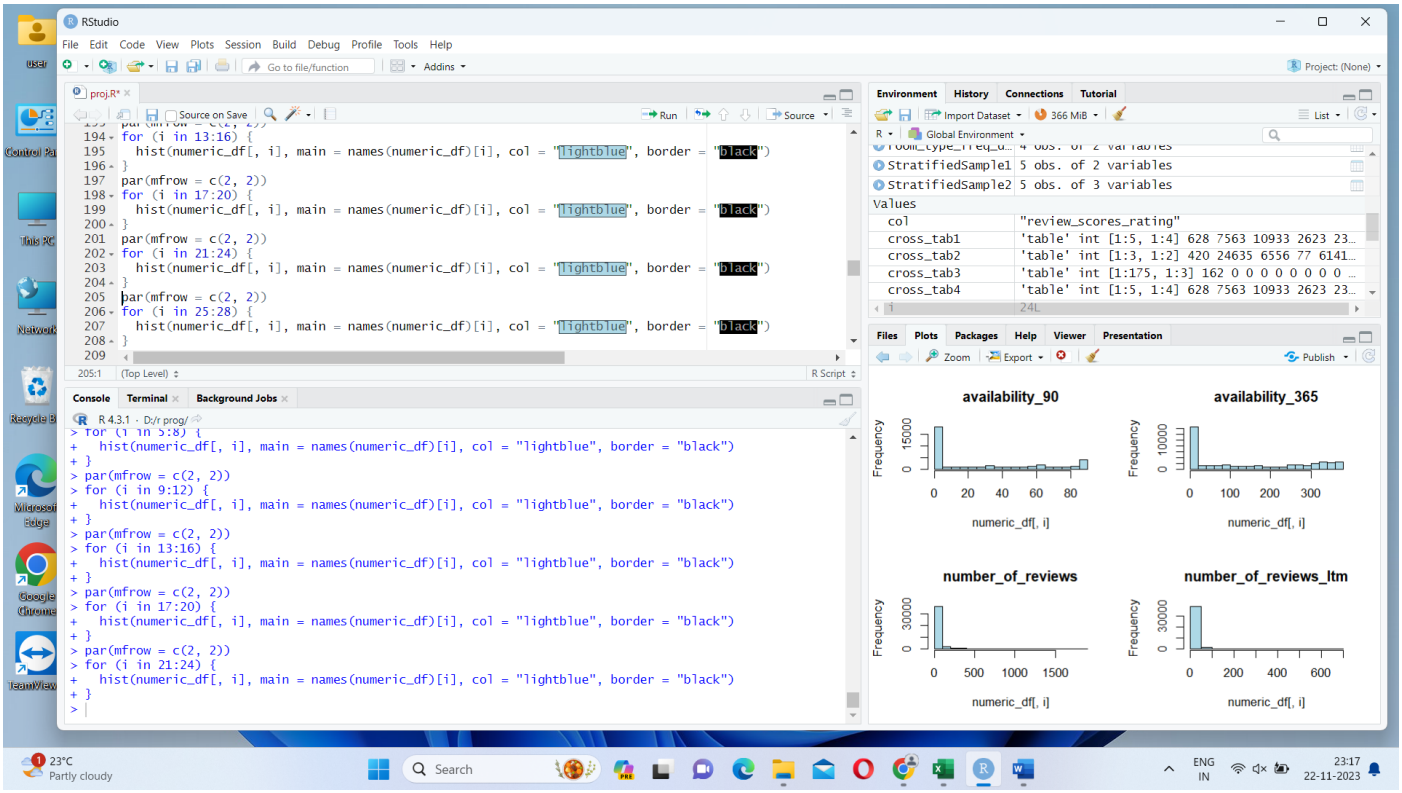
col	"review_scores_rating"
cross_tab1	'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2	'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3	'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4	'table' int [1:5, 1:4] 628 7563 10933 2623 23…

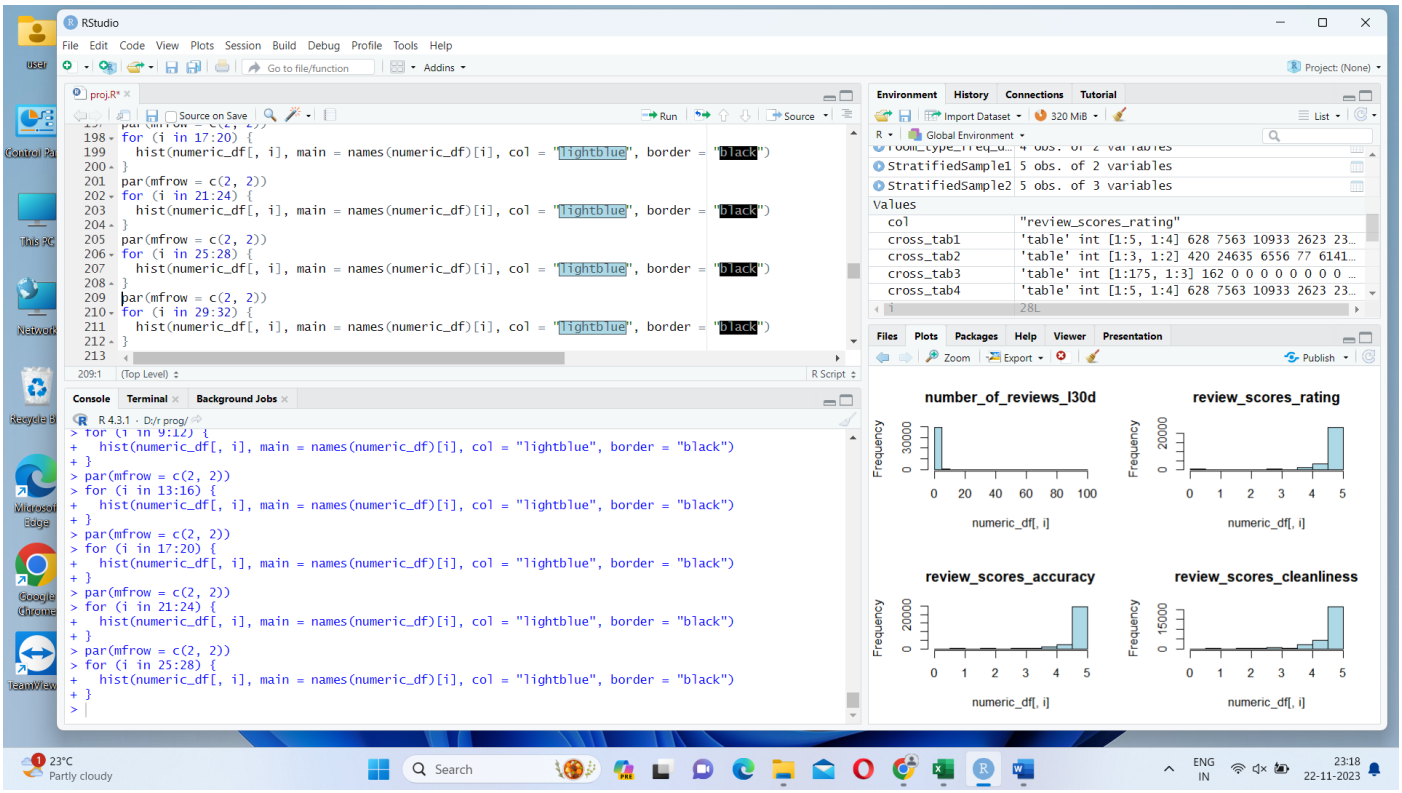
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables

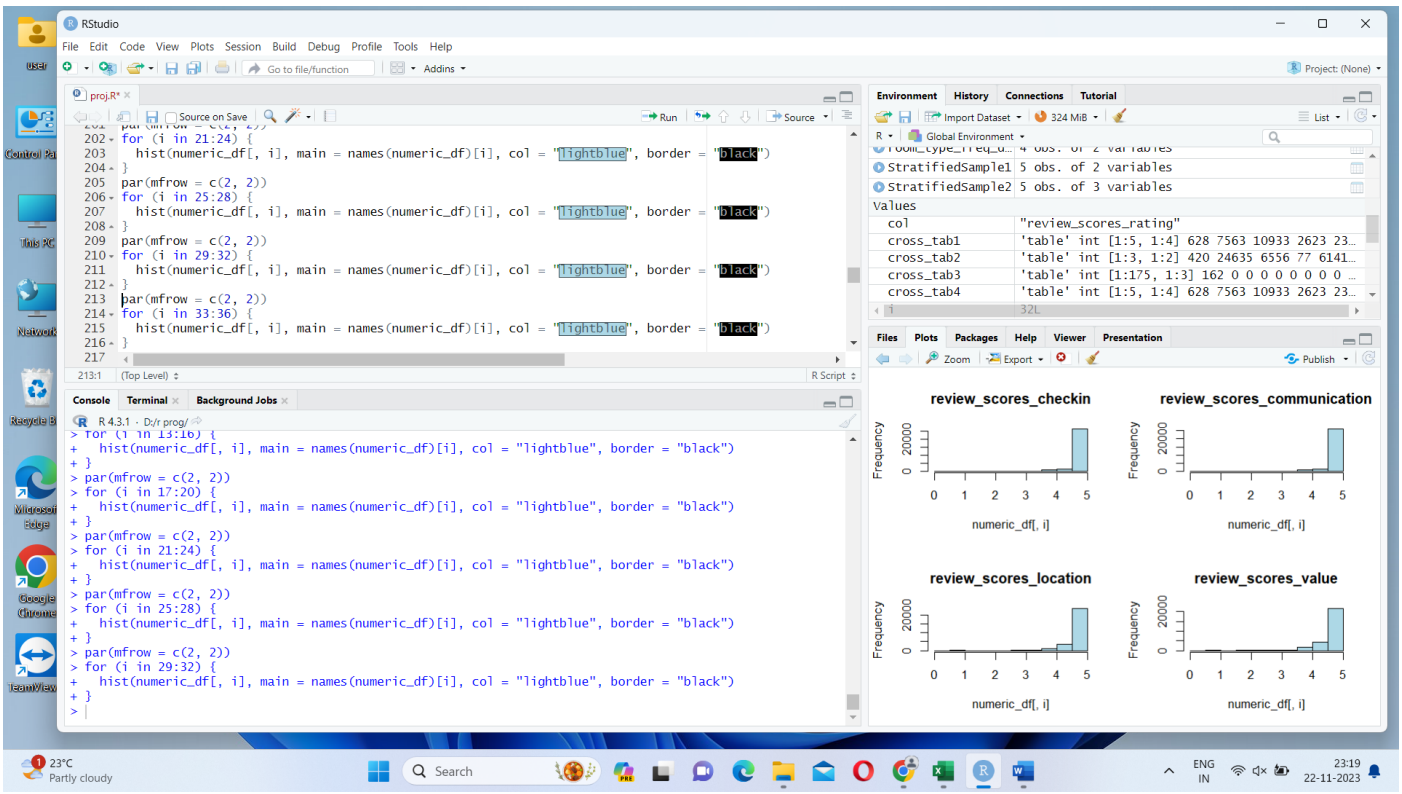


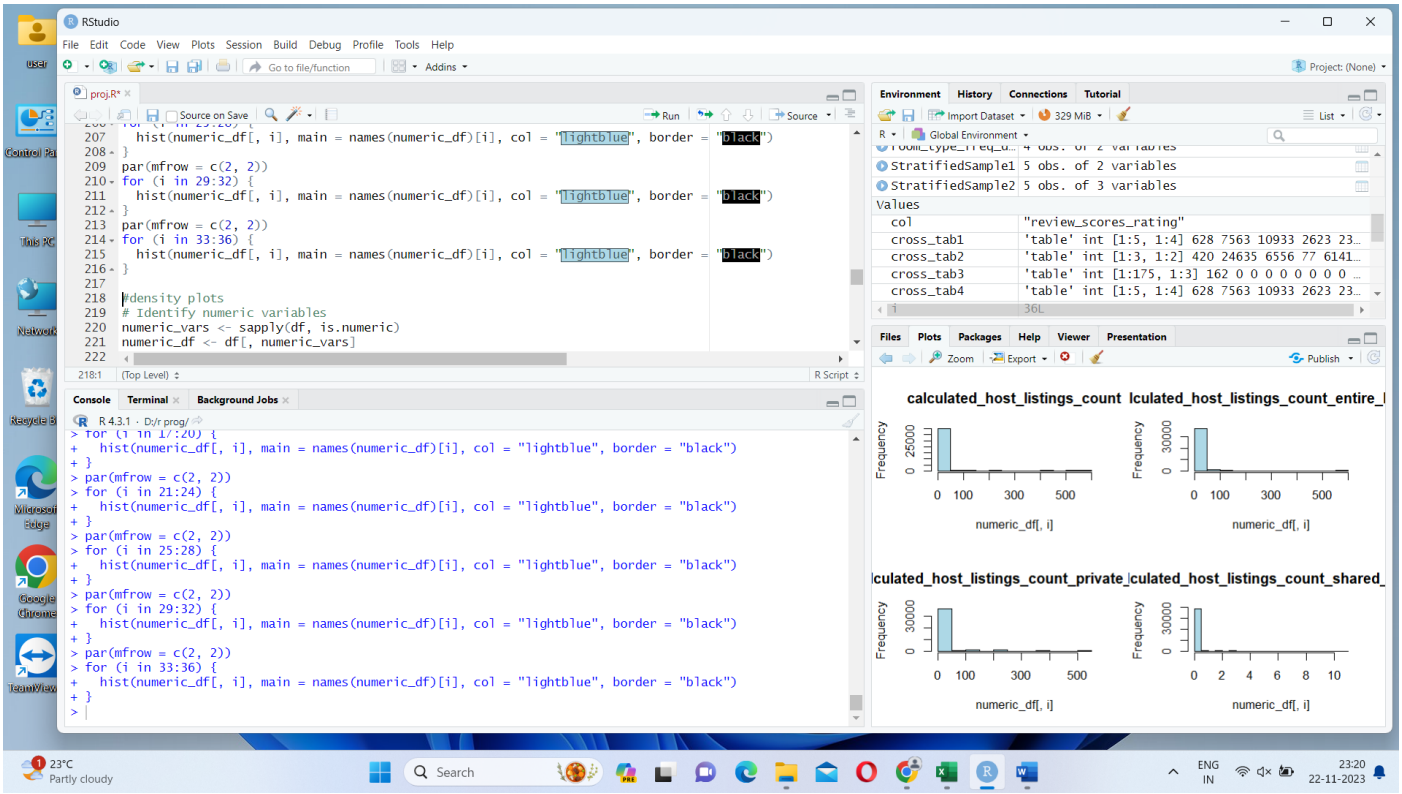

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
for (i in 5:8) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 9:12) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 13:16) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 17:20) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
Environment History Connections Tutorial
Import Dataset 348 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
minimum_minimum_nights
maximum_minimum_nights
minimum_maximum_nights
maximum_maximum_nights
Frequency
numeric_df[, i]
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Histograms
> par(mfrow = c(2,2))
> for (i in 1:4) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 5:8) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 9:12) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 13:16) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
23°C Partly cloudy
Search
ENG IN
23:16 22-11-2023

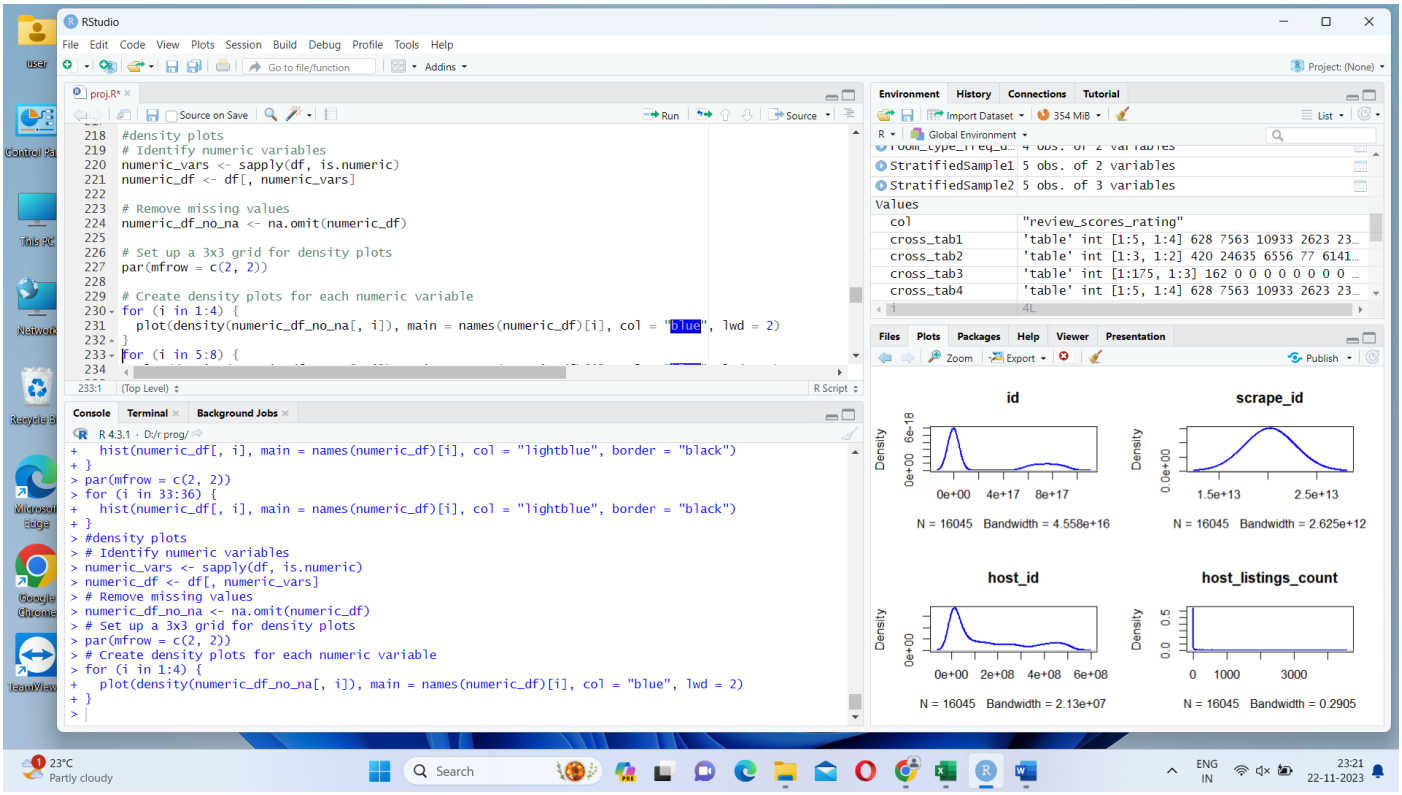

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
for (i in 9:12) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 13:16) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 17:20) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 21:24) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
Environment History Connections Tutorial
Import Dataset 357 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
minimum_nights_avg_ntm
maximum_nights_avg_ntm
availability_30
availability_60
Frequency
numeric_df[, i]
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 1:4) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 5:8) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 9:12) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 13:16) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 17:20) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
23°C Partly cloudy
Search
ENG IN
23:17 22-11-2023

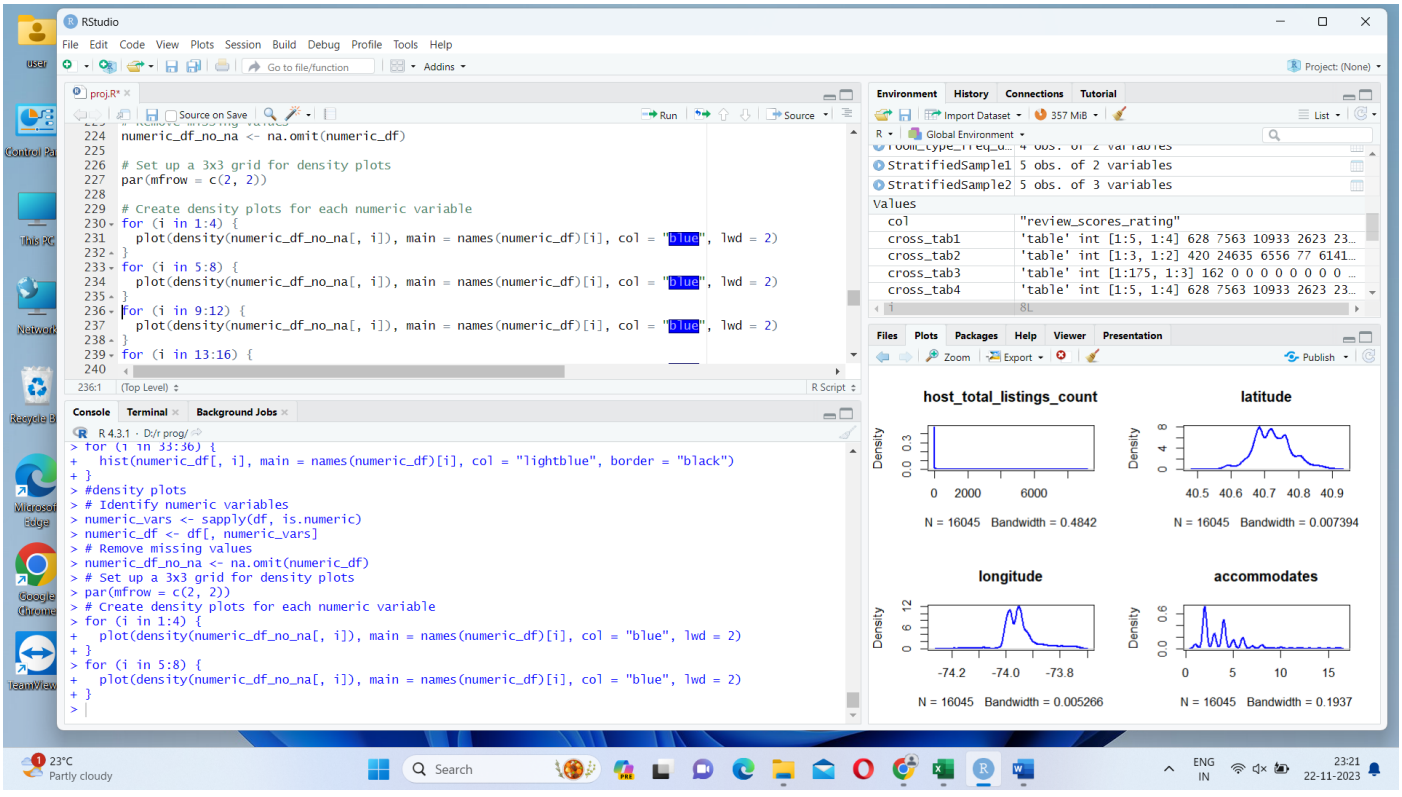

RStudio
availability_90
availability_365
number_of_reviews
number_of_reviews_ltm
numeric_df[, i]
Frequency


RStudio
number_of_reviews_l30d
review_scores_rating
review_scores_accuracy
review_scores_cleanliness
numeric_df[, i]
Frequency


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
for (i in 21:24) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 25:28) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 29:32) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 33:36) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
213:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 13:16) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 17:20) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 21:24) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 25:28) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 29:32) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
Environment History Connections Tutorial
Import Dataset 324 MiB
R Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
review_scores_checkin
review_scores_communication
review_scores_location
review_scores_value
Frequency
numeric_df[, i]
23°C Partly cloudy
Search
ENG IN
23:19 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 29:32) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 33:36) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
#density plots
# Identify numeric variables
numeric_vars <- sapply(df, is.numeric)
numeric_df <- df[, numeric_vars]
218:1 (Top Level) R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 17:20) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 21:24) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 25:28) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 29:32) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 33:36) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
Environment History Connections Tutorial
Import Dataset 329 MiB
R Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
calculated_host_listings_count
lculated_host_listings_count_entire_l
culated_host_listings_count_private_
culated_host_listings_count_shared_
Frequency
numeric_df[, i]
23°C Partly cloudy
Search
ENG IN
23:20 22-11-2023

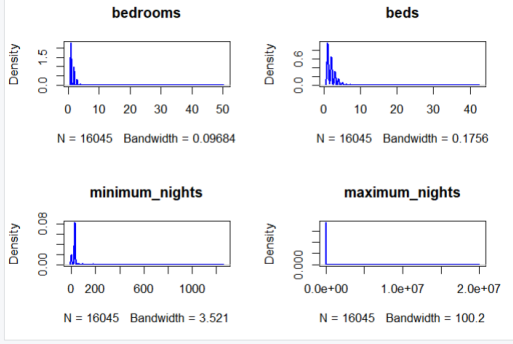

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
#density plots
# Identify numeric variables
numeric_vars <- sapply(df, is.numeric)
numeric_df <- df[, numeric_vars]
# Remove missing values
numeric_df_no_na <- na.omit(numeric_df)
# Set up a 3x3 grid for density plots
par(mfrow = c(2, 2))
# Create density plots for each numeric variable
for (i in 1:4) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 5:8) {
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> par(mfrow = c(2, 2))
> for (i in 33:36) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> #density plots
> # Identify numeric variables
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Remove missing values
> numeric_df_no_na <- na.omit(numeric_df)
> # Set up a 3x3 grid for density plots
> par(mfrow = c(2, 2))
> # Create density plots for each numeric variable
> for (i in 1:4) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
Environment History Connections Tutorial
Import Dataset 354 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
id
Density
0e+00 4e+17 8e+17
N = 16045 Bandwidth = 4.558e+16
scrape_id
1.5e+13 2.5e+13
N = 16045 Bandwidth = 2.625e+12
host_id
0e+00 2e+08 4e+08 6e+08
N = 16045 Bandwidth = 2.13e+07
host_listings_count
0 1000 3000
N = 16045 Bandwidth = 0.2905
23°C
Partly cloudy
Search
ENG
IN
23:21
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
numeric_df_no_na <- na.omit(numeric_df)
# Set up a 3x3 grid for density plots
par(mfrow = c(2, 2))
# Create density plots for each numeric variable
for (i in 1:4) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 5:8) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 9:12) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 13:16) {
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 33:36) {
+ hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
+ }
> #density plots
> # Identify numeric variables
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Remove missing values
> numeric_df_no_na <- na.omit(numeric_df)
> # Set up a 3x3 grid for density plots
> par(mfrow = c(2, 2))
> # Create density plots for each numeric variable
> for (i in 1:4) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 5:8) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
Environment History Connections Tutorial
Import Dataset 357 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
host_total_listings_count
Density
0 2000 6000
N = 16045 Bandwidth = 0.4842
latitude
40.5 40.6 40.7 40.8 40.9
N = 16045 Bandwidth = 0.007394
longitude
-74.2 -74.0 -73.8
N = 16045 Bandwidth = 0.005266
accommodates
0 5 10 15
N = 16045 Bandwidth = 0.1937
23°C
Partly cloudy
Search
ENG
IN
23:21
22-11-2023

```
par(mfrow = c(2, 2))

# Create density plots for each numeric variable
for (i in 1:4) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 5:8) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 9:12) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 13:16) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 17:20) {
```

```
> #density plots
> # Identify numeric variables
> numeric_vars <- sapply(df, is.numeric)
> numeric_df <- df[, numeric_vars]
> # Remove missing values
> numeric_df_no_na <- na.omit(numeric_df)
> # Set up a 3x3 grid for density plots
> par(mfrow = c(2, 2))
> # Create density plots for each numeric variable
> for (i in 1:4) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 5:8) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 9:12) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
>
```

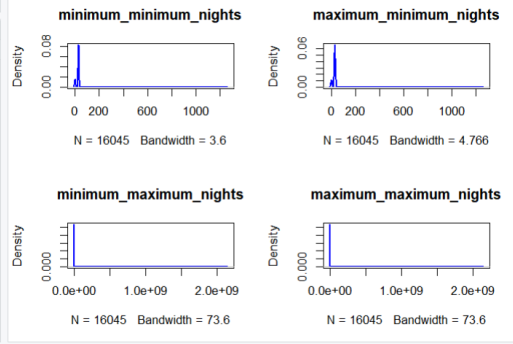
```
Environment | History | Connections | Tutorial
Import Dataset | 358 MiB
Global Environment
StratifiedSample1  5 obs. of 2 variables
StratifiedSample2  5 obs. of 3 variables
Values
  col          "review_scores_rating"
  cross_tab1   'table' int [1:5, 1:4] 628 7563 10933 2623 23…
  cross_tab2   'table' int [1:3, 1:2] 420 24635 6556 77 6141…
  cross_tab3   'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
  cross_tab4   'table' int [1:5, 1:4] 628 7563 10933 2623 23…
```

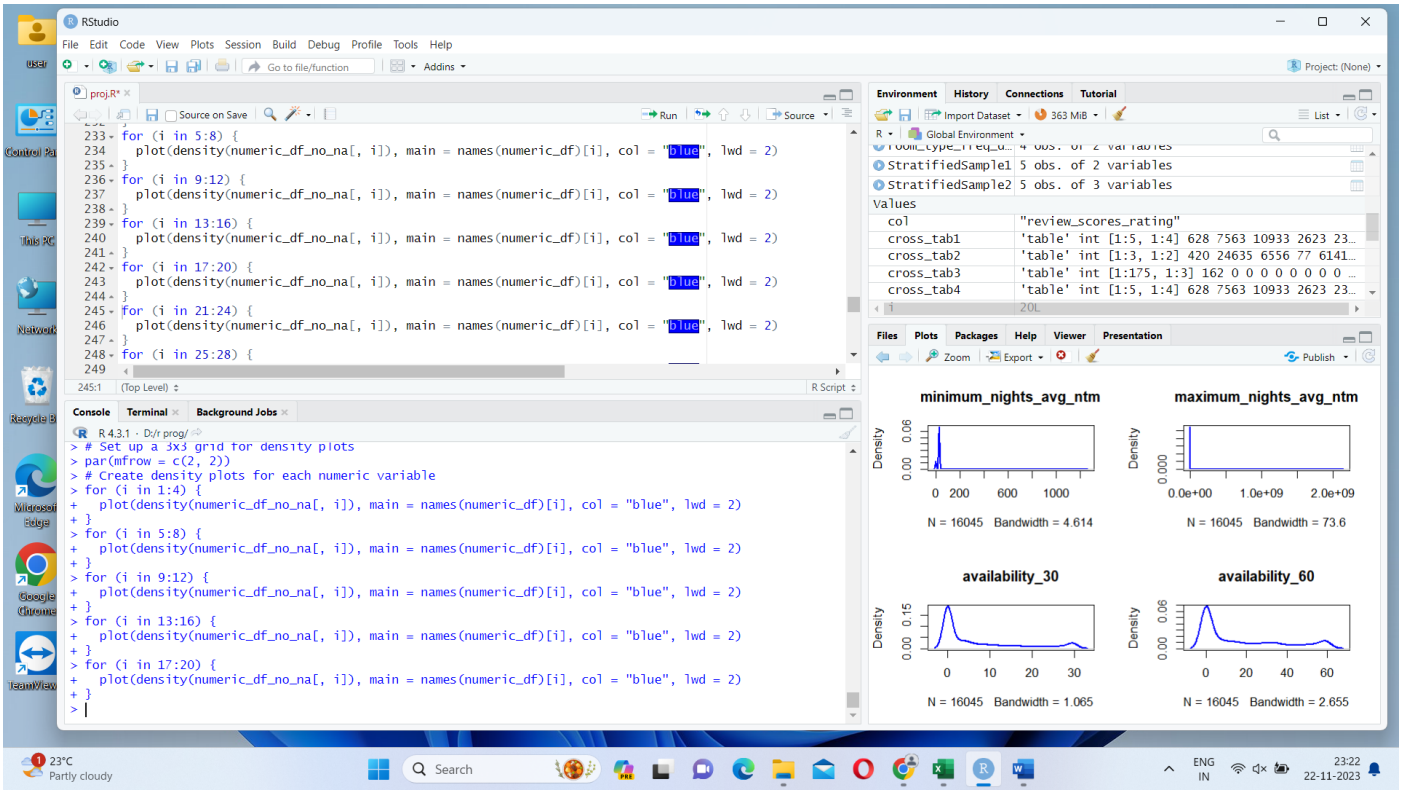


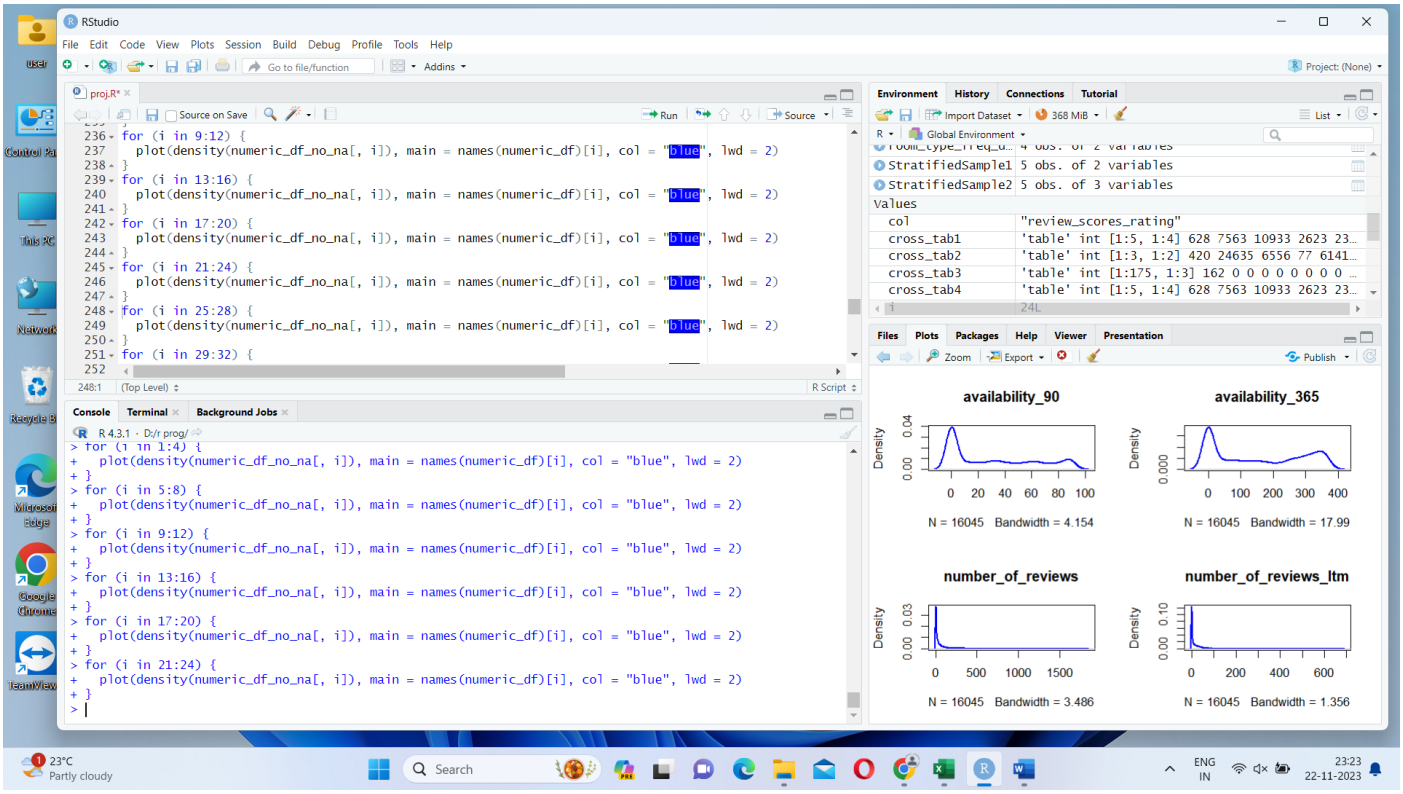

```
for (i in 1:4) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 5:8) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 9:12) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 13:16) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 17:20) {
  plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 21:24) {
```

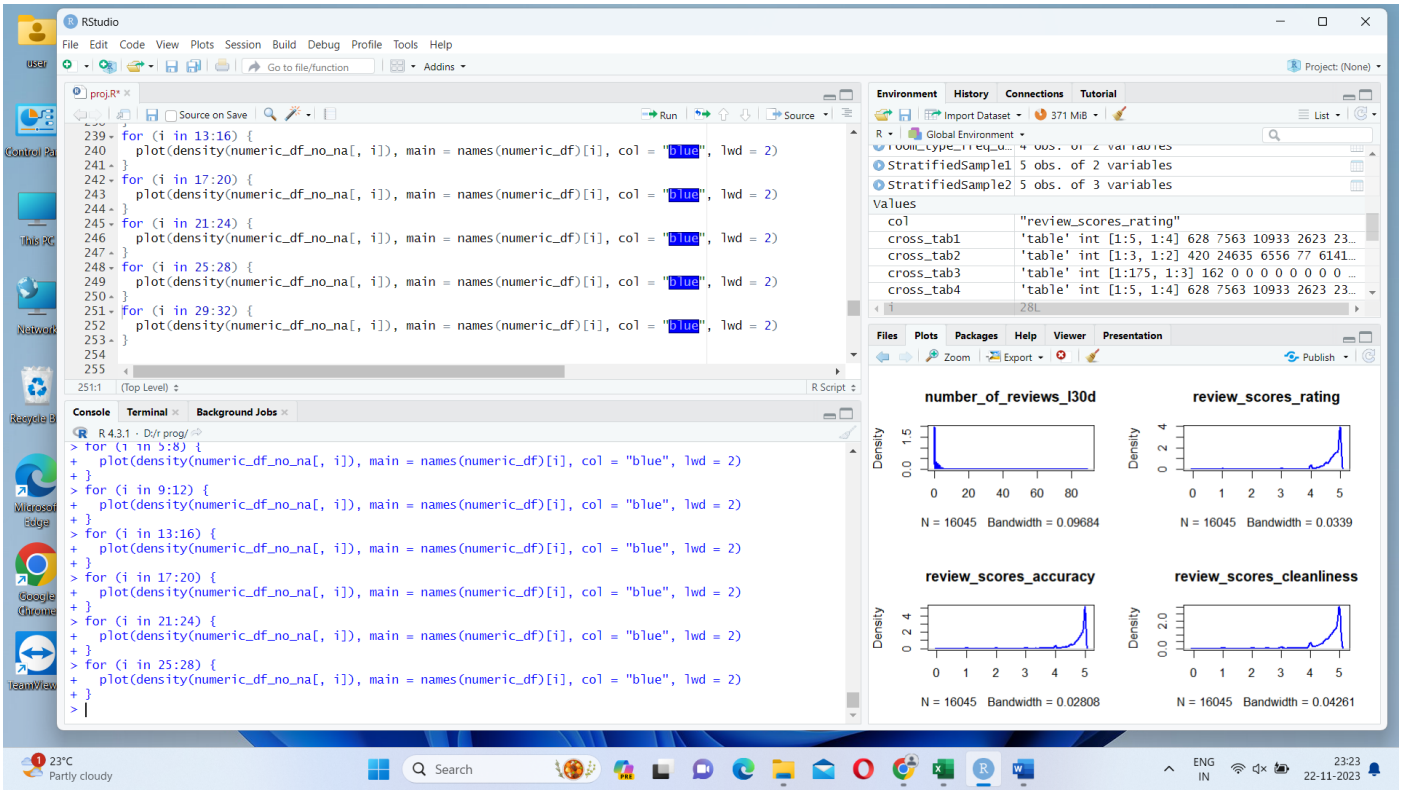
```
> numeric_df <- df[, numeric_vars]
> # Remove missing values
> numeric_df_no_na <- na.omit(numeric_df)
> # Set up a 3x3 grid for density plots
> par(mfrow = c(2, 2))
> # Create density plots for each numeric variable
> for (i in 1:4) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 5:8) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 9:12) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 13:16) {
+   plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
>
```

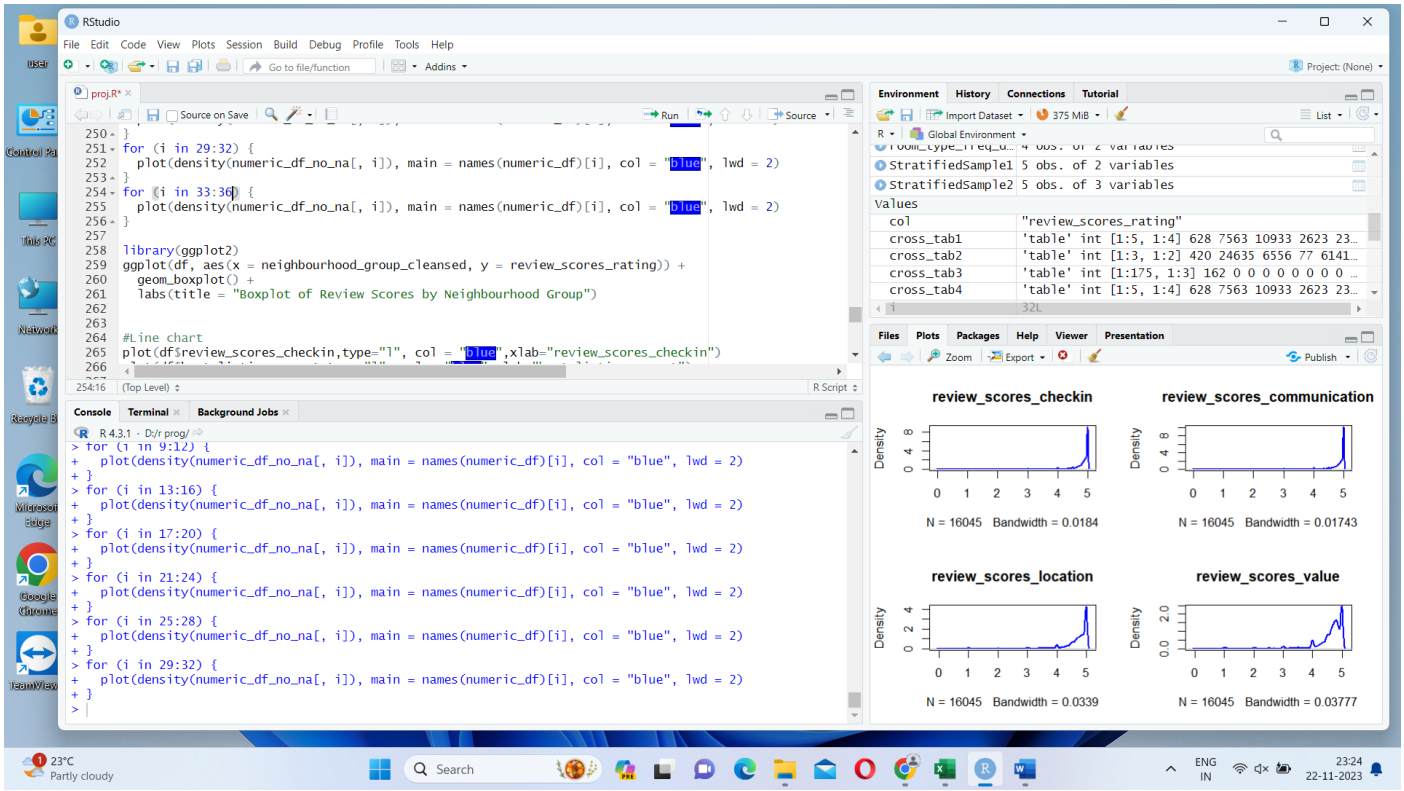
```
Environment | History | Connections | Tutorial
Import Dataset | 362 MiB
Global Environment
StratifiedSample1  5 obs. of 2 variables
StratifiedSample2  5 obs. of 3 variables
Values
  col          "review_scores_rating"
  cross_tab1   'table' int [1:5, 1:4] 628 7563 10933 2623 23…
  cross_tab2   'table' int [1:3, 1:2] 420 24635 6556 77 6141…
  cross_tab3   'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
  cross_tab4   'table' int [1:5, 1:4] 628 7563 10933 2623 23…
```

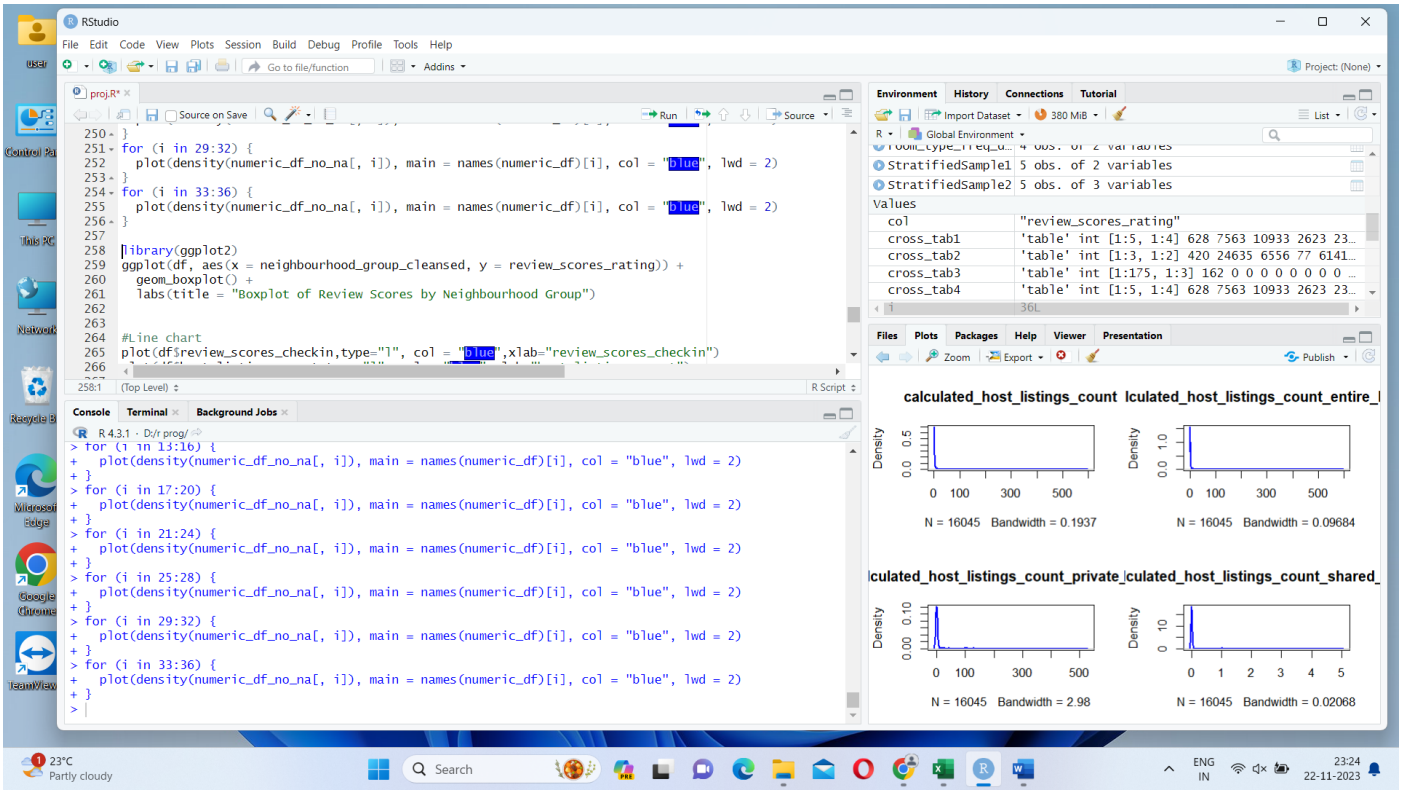


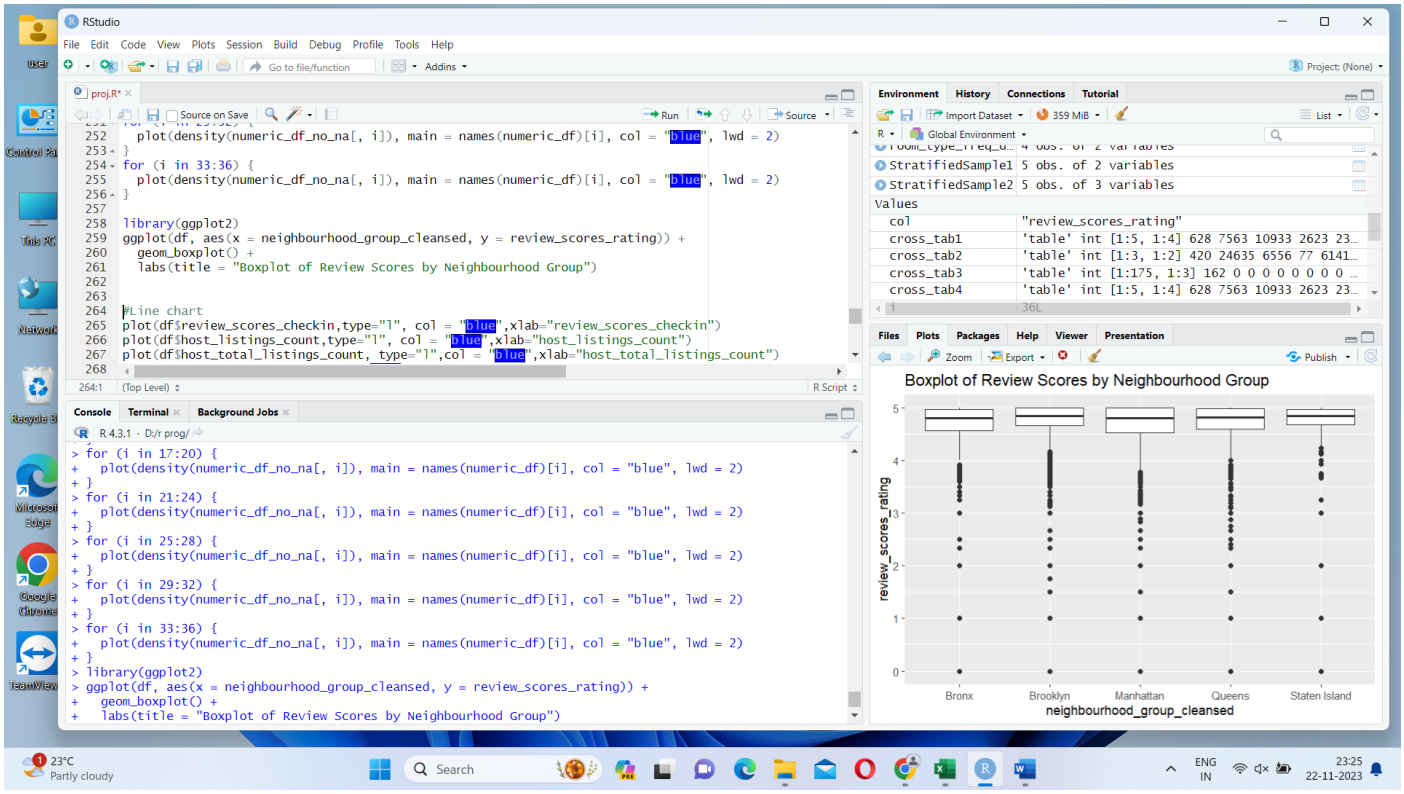

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
for (i in 5:8) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 9:12) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 13:16) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 17:20) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 21:24) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 25:28) {
245:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> # Set up a 3x3 grid for density plots
> par(mfrow = c(2, 2))
> # Create density plots for each numeric variable
> for (i in 1:4) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 5:8) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 9:12) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 13:16) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 17:20) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
Environment History Connections Tutorial
Import Dataset
363 MiB
R
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
minimum_nights_avg_ntm
Density
N = 16045 Bandwidth = 4.614
maximum_nights_avg_ntm
N = 16045 Bandwidth = 73.6
availability_30
N = 16045 Bandwidth = 1.065
availability_60
N = 16045 Bandwidth = 2.655
23°C
Partly cloudy
Search
ENG
IN
23:22
22-11-2023

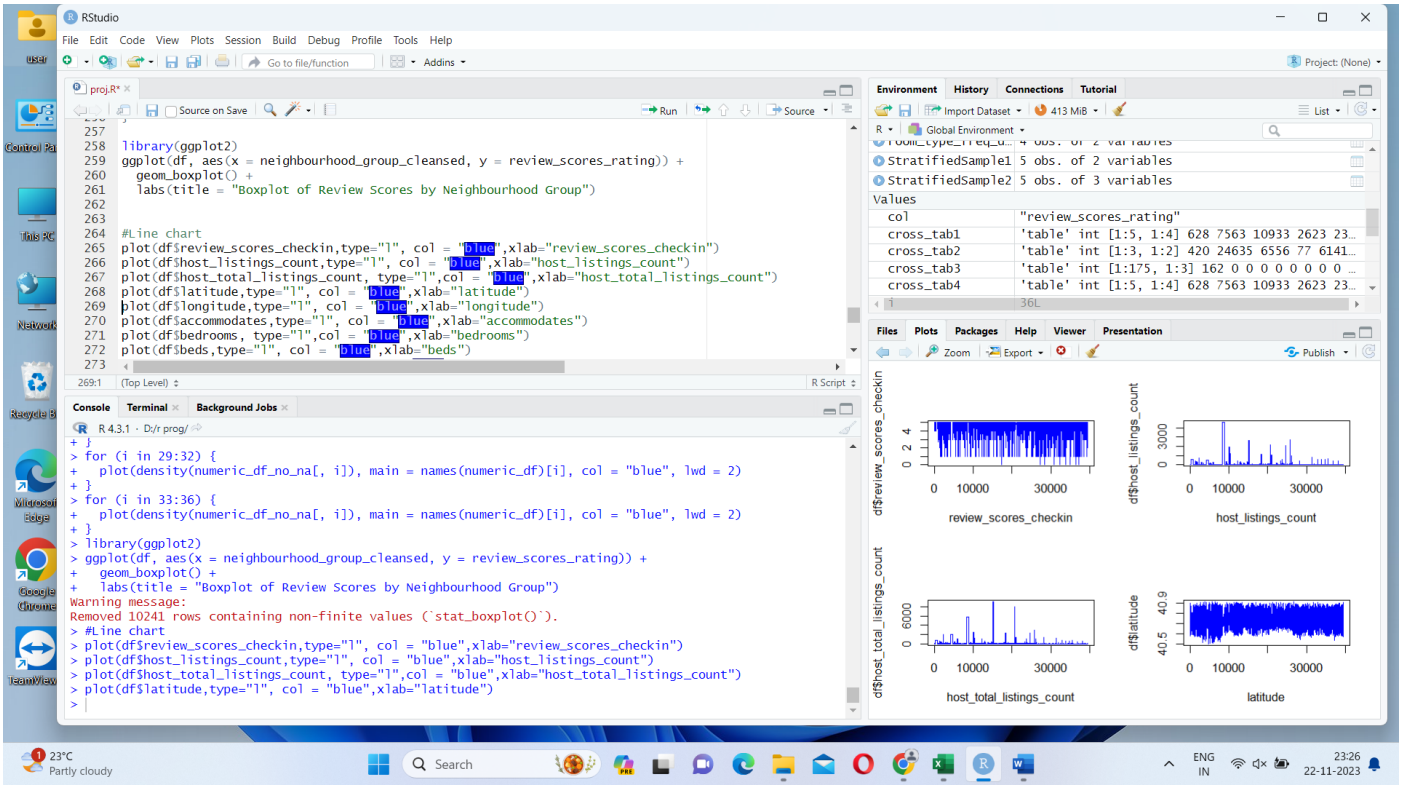

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
for (i in 9:12) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 13:16) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 17:20) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 21:24) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 25:28) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 29:32) {
248:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 1:4) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 5:8) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 9:12) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 13:16) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 17:20) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 21:24) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
Environment History Connections Tutorial
Import Dataset
368 MiB
R
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
availability_90
Density
N = 16045 Bandwidth = 4.154
availability_365
N = 16045 Bandwidth = 17.99
number_of_reviews
N = 16045 Bandwidth = 3.486
number_of_reviews_ltm
N = 16045 Bandwidth = 1.356
23°C
Partly cloudy
Search
ENG
IN
23:23
22-11-2023

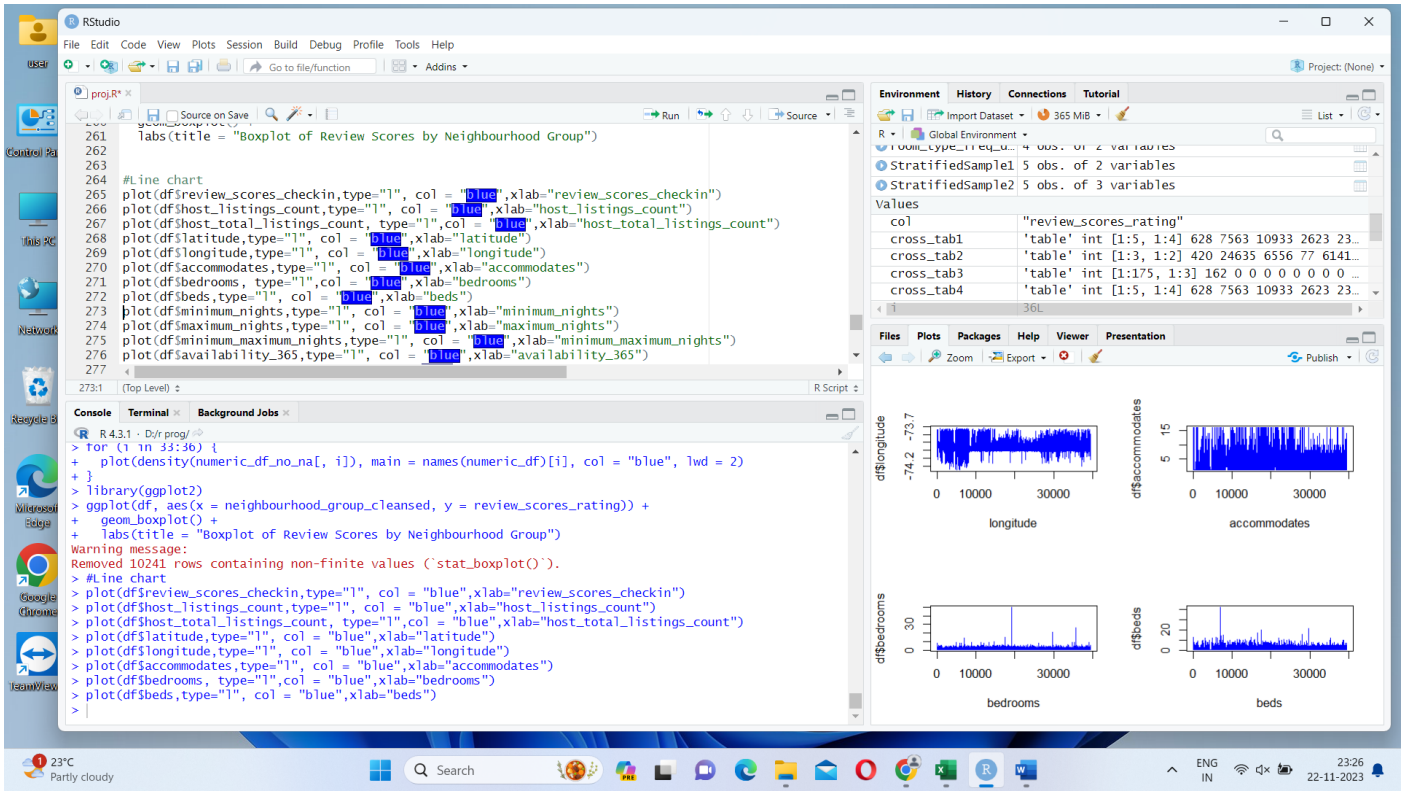

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
for (i in 13:16) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 17:20) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 21:24) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 25:28) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 29:32) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Environment History Connections Tutorial
Import Dataset 371 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
number_of_reviews_l30d
review_scores_rating
review_scores_accuracy
review_scores_cleanliness
Density
N = 16045 Bandwidth = 0.09684
N = 16045 Bandwidth = 0.0339
N = 16045 Bandwidth = 0.02808
N = 16045 Bandwidth = 0.04261
23°C
Partly cloudy
Search
ENG
IN
23:23
22-11-2023

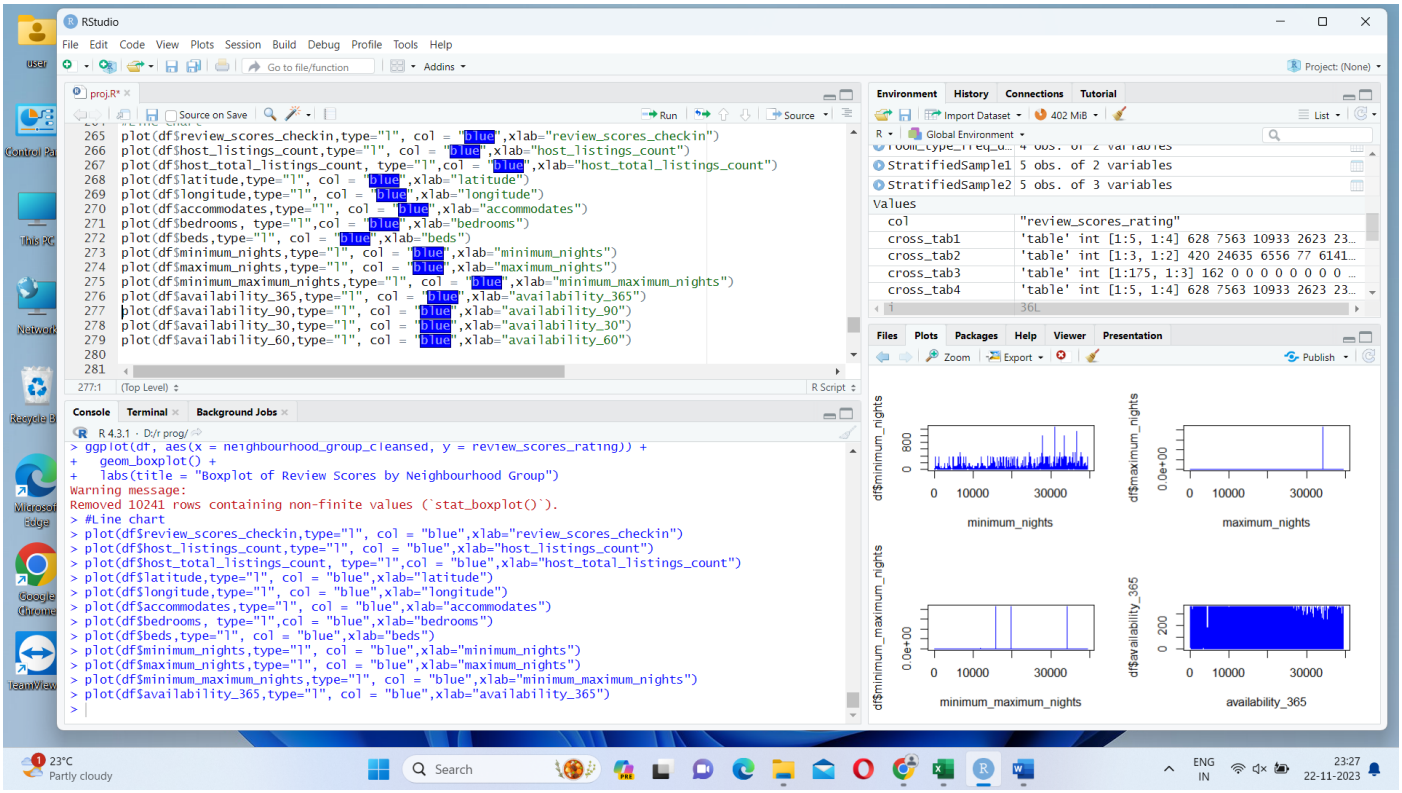

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
for (i in 29:32) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 33:36) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Environment History Connections Tutorial
Import Dataset 375 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
review_scores_checkin
review_scores_communication
review_scores_location
review_scores_value
Density
N = 16045 Bandwidth = 0.0184
N = 16045 Bandwidth = 0.01743
N = 16045 Bandwidth = 0.0339
N = 16045 Bandwidth = 0.03777
23°C
Partly cloudy
Search
ENG
IN
23:24
22-11-2023

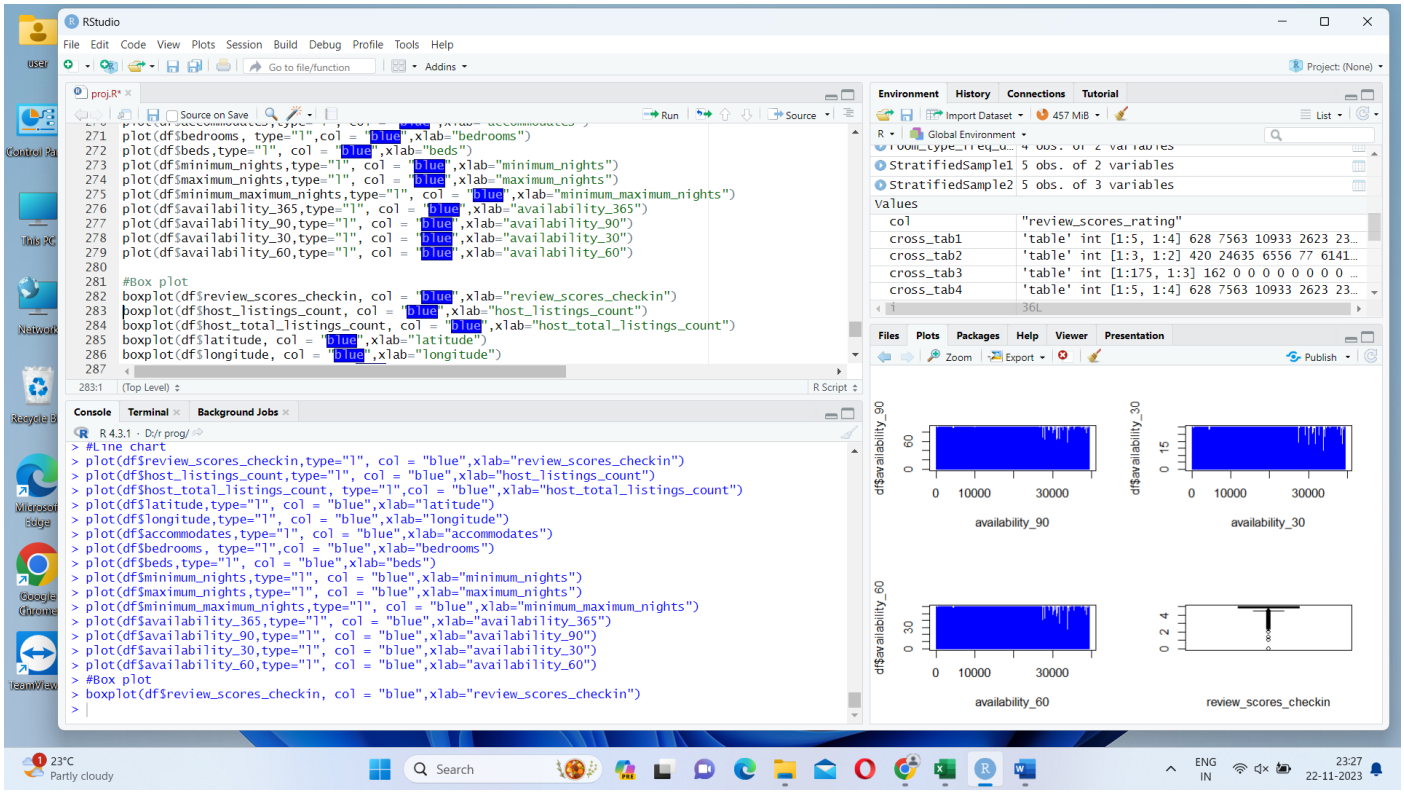

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
for (i in 29:32) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 33:36) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 13:16) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 17:20) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 21:24) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 25:28) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 29:32) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 33:36) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
Environment History Connections Tutorial
Import Dataset
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
calculated_host_listings_count
lculated_host_listings_count_entire_
Density
N = 16045 Bandwidth = 0.1937
N = 16045 Bandwidth = 0.09684
culated_host_listings_count_private
culated_host_listings_count_shared_
N = 16045 Bandwidth = 2.98
N = 16045 Bandwidth = 0.02068
23°C
Partly cloudy
Search
ENG
IN
23:24
22-11-2023

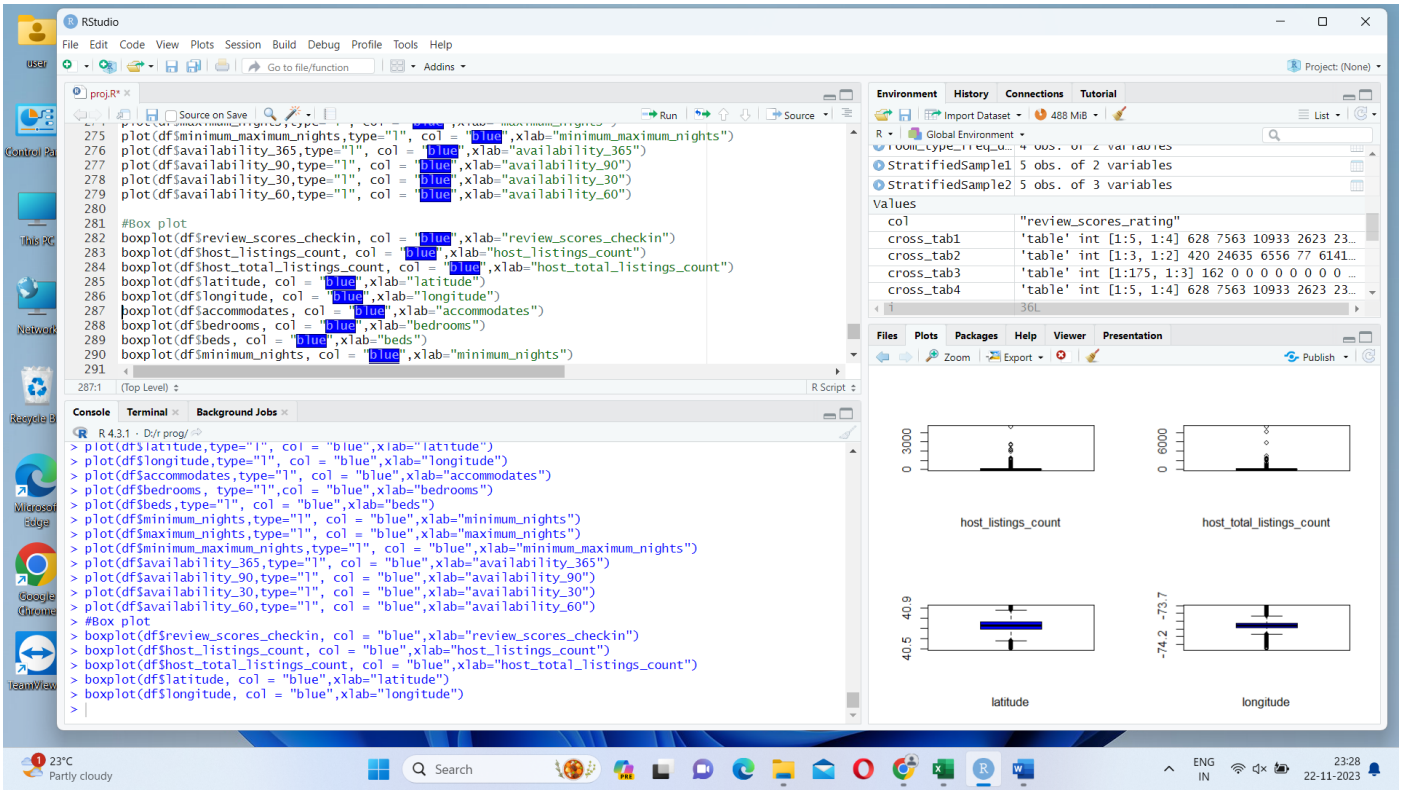

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 33:36) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> for (i in 17:20) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 21:24) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 25:28) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 29:32) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> for (i in 33:36) {
+ plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
+ }
> library(ggplot2)
> ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
+ geom_boxplot() +
+ labs(title = "Boxplot of Review Scores by Neighbourhood Group")
Environment History Connections Tutorial
Import Dataset
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Boxplot of Review Scores by Neighbourhood Group
review_scores_rating
Bronx
Brooklyn
Manhattan
Queens
Staten Island
neighbourhood_group_cleansed
23°C
Partly cloudy
Search
ENG
IN
23:25
22-11-2023

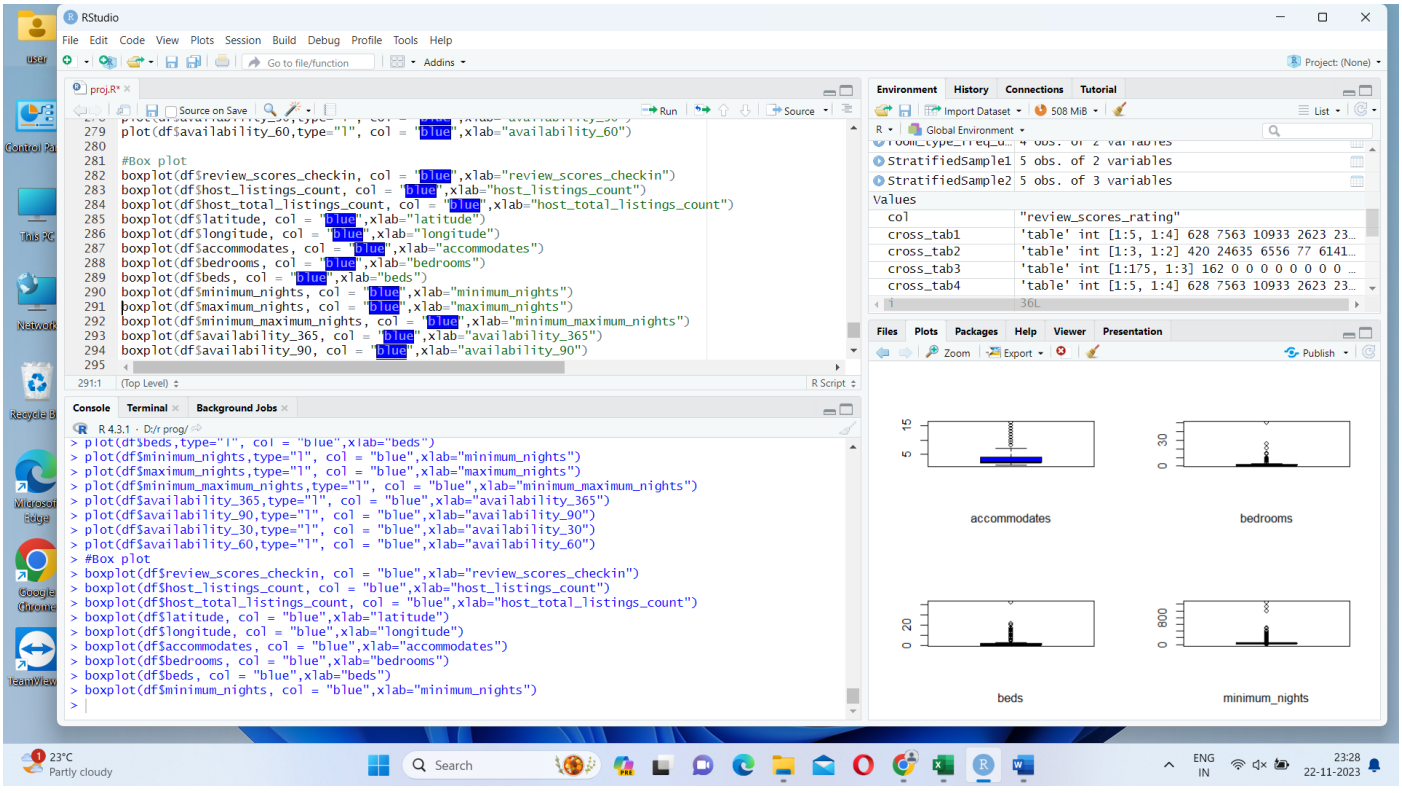

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
plot(df$latitude,type="l", col = "blue",xlab="latitude")
plot(df$longitude,type="l", col = "blue",xlab="longitude")
plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
plot(df$beds,type="l", col = "blue",xlab="beds")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
for (i in 29:32) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
for (i in 33:36) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
Warning message:
Removed 10241 rows containing non-finite values (`stat_boxplot()`).
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
plot(df$latitude,type="l", col = "blue",xlab="latitude")
Environment History Connections Tutorial
Import Dataset
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
df$review_scores_checkin
review_scores_checkin
df$host_listings_count
host_listings_count
df$host_total_listings_count
host_total_listings_count
df$latitude
latitude
23°C Partly cloudy
Search
ENG IN
23:26 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
plot(df$latitude,type="l", col = "blue",xlab="latitude")
plot(df$longitude,type="l", col = "blue",xlab="longitude")
plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
plot(df$beds,type="l", col = "blue",xlab="beds")
plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
for (i in 33:36) {
plot(density(numeric_df_no_na[, i]), main = names(numeric_df)[i], col = "blue", lwd = 2)
}
library(ggplot2)
ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
geom_boxplot() +
labs(title = "Boxplot of Review Scores by Neighbourhood Group")
Warning message:
Removed 10241 rows containing non-finite values (`stat_boxplot()`).
#Line chart
plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
plot(df$latitude,type="l", col = "blue",xlab="latitude")
plot(df$longitude,type="l", col = "blue",xlab="longitude")
plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
plot(df$beds,type="l", col = "blue",xlab="beds")
Environment History Connections Tutorial
Import Dataset
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
df$longitude
longitude
df$accommodates
accommodates
df$bedrooms
bedrooms
df$beds
beds
23°C Partly cloudy
Search
ENG IN
23:26 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
265 plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
266 plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
267 plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
268 plot(df$latitude,type="l", col = "blue",xlab="latitude")
269 plot(df$longitude,type="l", col = "blue",xlab="longitude")
270 plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
271 plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
272 plot(df$beds,type="l", col = "blue",xlab="beds")
273 plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
274 plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
275 plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
276 plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
277 plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
278 plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
279 plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> ggplot(df, aes(x = neighbourhood_group_cleansed, y = review_scores_rating)) +
+ geom_boxplot() +
+ labs(title = "Boxplot of Review Scores by Neighbourhood Group")
Warning message:
Removed 10241 rows containing non-finite values (`stat_boxplot()`).
> #Line chart
> plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
> plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
> plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
> plot(df$latitude,type="l", col = "blue",xlab="latitude")
> plot(df$longitude,type="l", col = "blue",xlab="longitude")
> plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
> plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
Environment History Connections Tutorial
Import Dataset 402 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
df$minimum_nights
minimum_nights
df$maximum_nights
maximum_nights
df$minimum_maximum_nights
minimum_maximum_nights
df$availability_365
availability_365
23°C Partly cloudy
Search
ENG IN
23:27 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
271 plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
272 plot(df$beds,type="l", col = "blue",xlab="beds")
273 plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
274 plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
275 plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
276 plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
277 plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
278 plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
279 plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
281 #Box plot
282 boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
283 boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
284 boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
285 boxplot(df$latitude, col = "blue",xlab="latitude")
286 boxplot(df$longitude, col = "blue",xlab="longitude")
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> #Line chart
> plot(df$review_scores_checkin,type="l", col = "blue",xlab="review_scores_checkin")
> plot(df$host_listings_count,type="l", col = "blue",xlab="host_listings_count")
> plot(df$host_total_listings_count, type="l",col = "blue",xlab="host_total_listings_count")
> plot(df$latitude,type="l", col = "blue",xlab="latitude")
> plot(df$longitude,type="l", col = "blue",xlab="longitude")
> plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
> plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
> plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
> plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
> plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
> #Box plot
> boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
Environment History Connections Tutorial
Import Dataset 457 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
df$availability_90
availability_90
df$availability_30
availability_30
df$availability_60
availability_60
review_scores_checkin
23°C Partly cloudy
Search
ENG IN
23:27 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
#Box plot
boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
boxplot(df$latitude, col = "blue",xlab="latitude")
boxplot(df$longitude, col = "blue",xlab="longitude")
boxplot(df$accommodates, col = "blue",xlab="accommodates")
boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
boxplot(df$beds, col = "blue",xlab="beds")
boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
287:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> plot(df$latitude,type="l", col = "blue",xlab="latitude")
> plot(df$longitude,type="l", col = "blue",xlab="longitude")
> plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
> plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
> plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
> plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
> plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
> #Box plot
> boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
> boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
> boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
> boxplot(df$latitude, col = "blue",xlab="latitude")
> boxplot(df$longitude, col = "blue",xlab="longitude")
Environment History Connections Tutorial
Import Dataset
488 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
host_listings_count
host_total_listings_count
latitude
longitude
23°C
Partly cloudy
Search
ENG
IN
23:28
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
Source on Save
Run
Source
plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
#Box plot
boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
boxplot(df$latitude, col = "blue",xlab="latitude")
boxplot(df$longitude, col = "blue",xlab="longitude")
boxplot(df$accommodates, col = "blue",xlab="accommodates")
boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
boxplot(df$beds, col = "blue",xlab="beds")
boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
boxplot(df$maximum_nights, col = "blue",xlab="maximum_nights")
boxplot(df$minimum_maximum_nights, col = "blue",xlab="minimum_maximum_nights")
boxplot(df$availability_365, col = "blue",xlab="availability_365")
boxplot(df$availability_90, col = "blue",xlab="availability_90")
291:1 (Top Level)
R Script
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
> plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
> plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
> plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
> #Box plot
> boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
> boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
> boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
> boxplot(df$latitude, col = "blue",xlab="latitude")
> boxplot(df$longitude, col = "blue",xlab="longitude")
> boxplot(df$accommodates, col = "blue",xlab="accommodates")
> boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
> boxplot(df$beds, col = "blue",xlab="beds")
> boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
Environment History Connections Tutorial
Import Dataset
508 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23…
Files Plots Packages Help Viewer Presentation
Zoom
Export
Publish
accommodates
bedrooms
beds
minimum_nights
23°C
Partly cloudy
Search
ENG
IN
23:28
22-11-2023

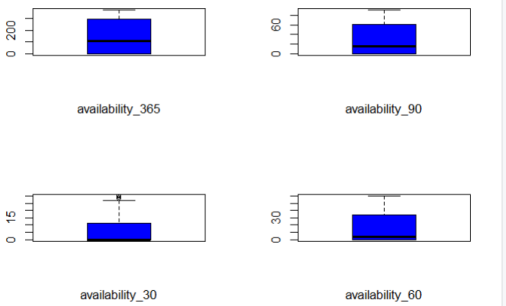
RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

```
boxplot(df$latitude, col = "blue",xlab="latitude")
boxplot(df$longitude, col = "blue",xlab="longitude")
boxplot(df$accommodates, col = "blue",xlab="accommodates")
boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
boxplot(df$beds, col = "blue",xlab="beds")
boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
boxplot(df$availability_365, col = "blue",xlab="availability_365")
boxplot(df$availability_90, col = "blue",xlab="availability_90")
boxplot(df$availability_30, col = "blue",xlab="availability_30")
boxplot(df$availability_60, col = "blue",xlab="availability_60")
```

```
> boxplot(df$availability_365, col = "blue",xlab="availability_365")
> boxplot(df$availability_90, col = "blue",xlab="availability_90")
> boxplot(df$availability_30, col = "blue",xlab="availability_30")
> boxplot(df$availability_60, col = "blue",xlab="availability_60")
>
```

StratifiedSample1	5 obs. of 2 variables
StratifiedSample2	5 obs. of 3 variables
Values	
col	"review_scores_rating"
cross_tab1	'table' int [1:5, 1:4] 628 7563 10933 2623 23…
cross_tab2	'table' int [1:3, 1:2] 420 24635 6556 77 6141…
cross_tab3	'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 …
cross_tab4	'table' int [1:5, 1:4] 628 7563 10933 2623 23…



23°C Partly cloudy

Search

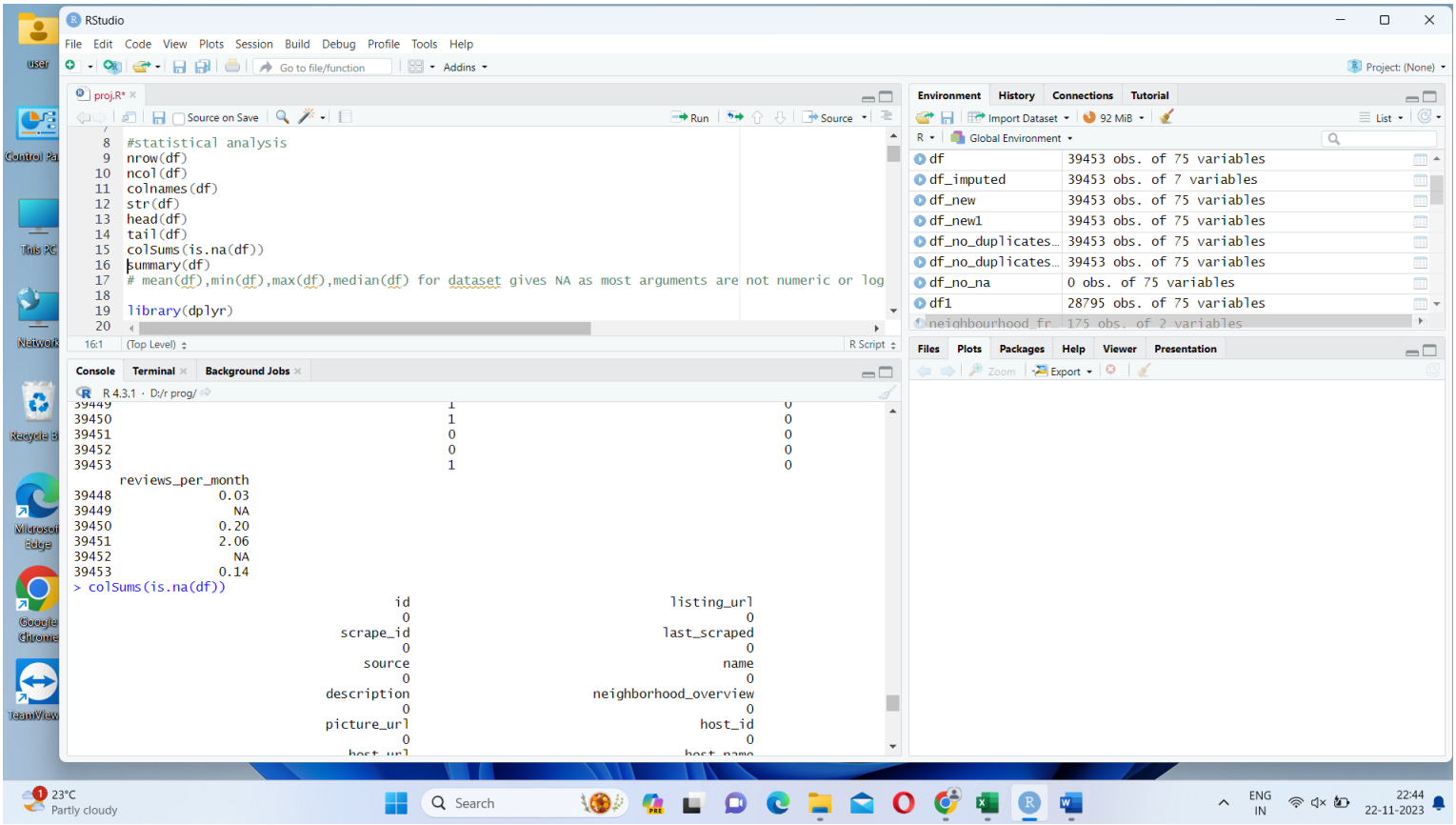
ENG IN

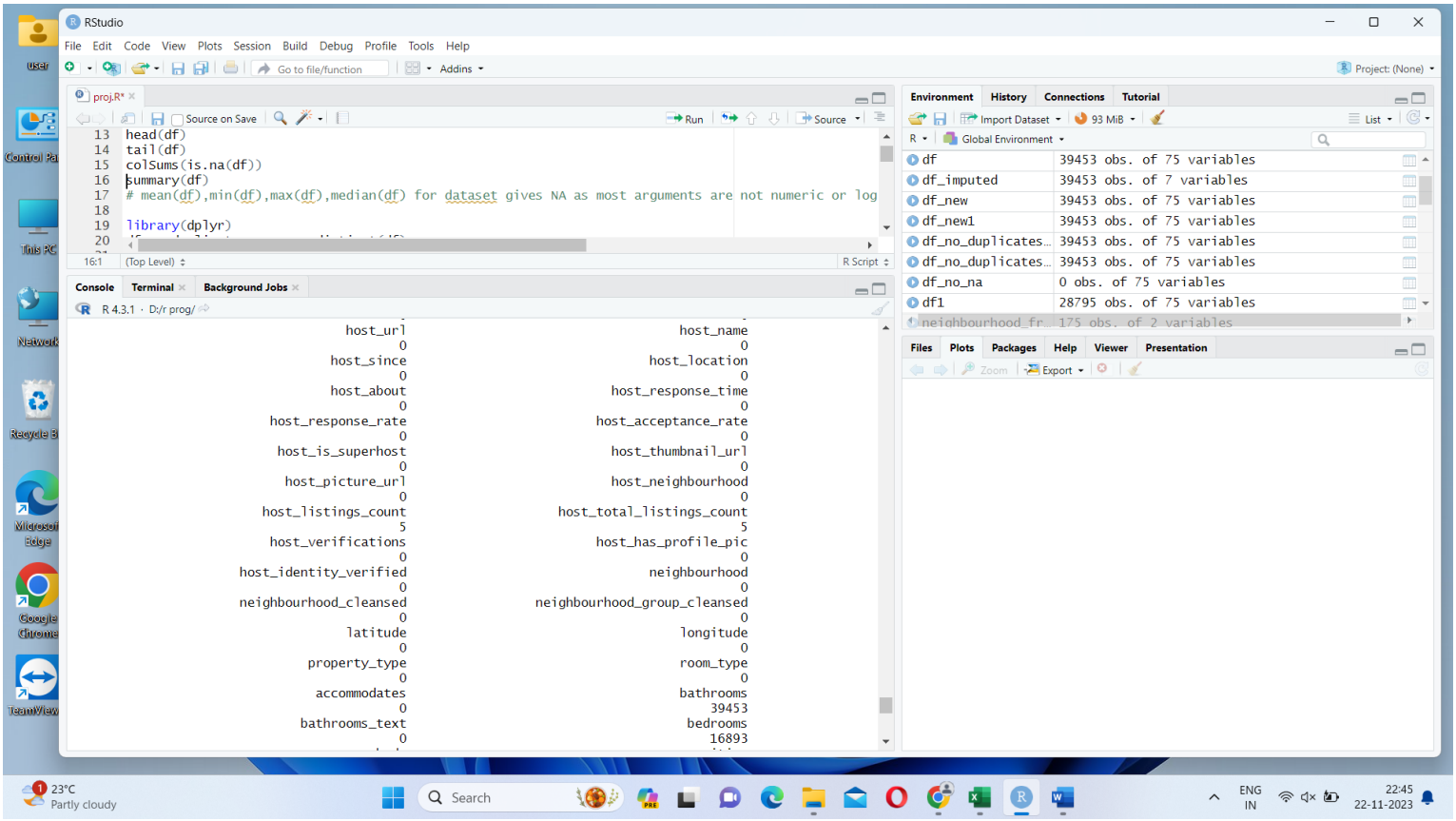
23:30 22-11-2023

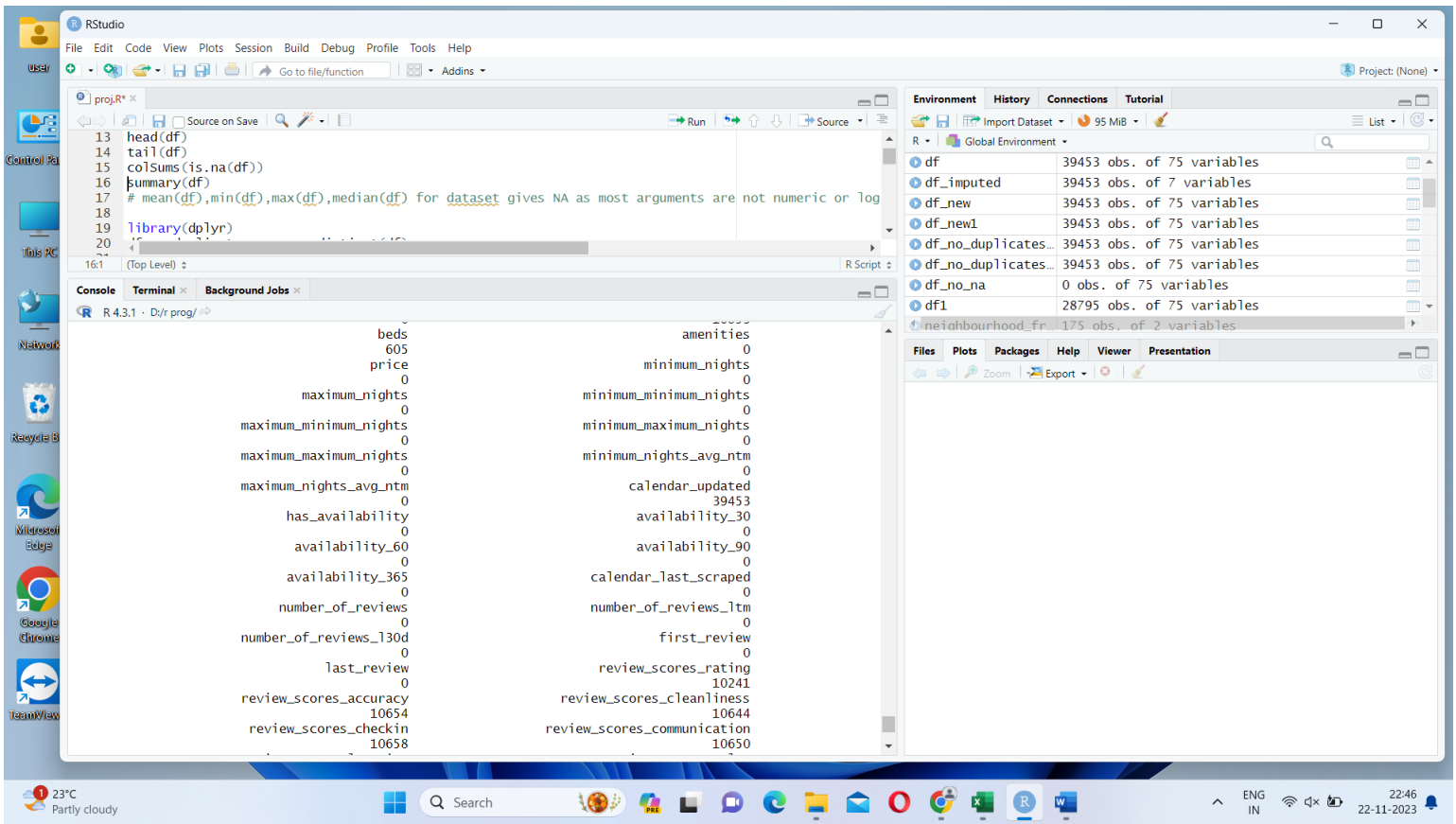
Interesting insights:

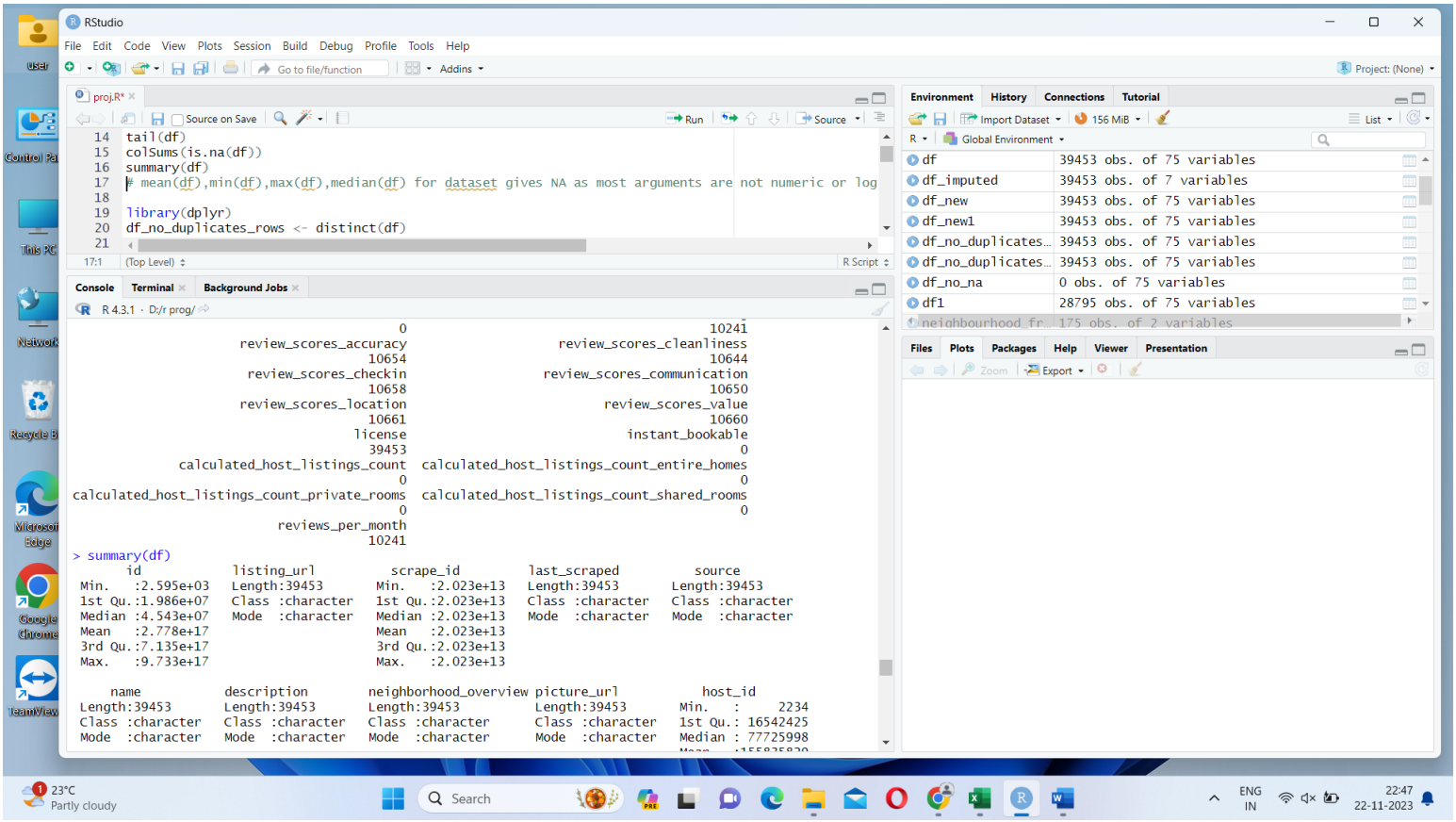
1)The data given in the dataset has a lot of junk values and has not been framed effectively

Eg: the columns bathrooms,calendar_updated and license have only na values. Many other columns like bedrooms and all the review columns have a lot of na values



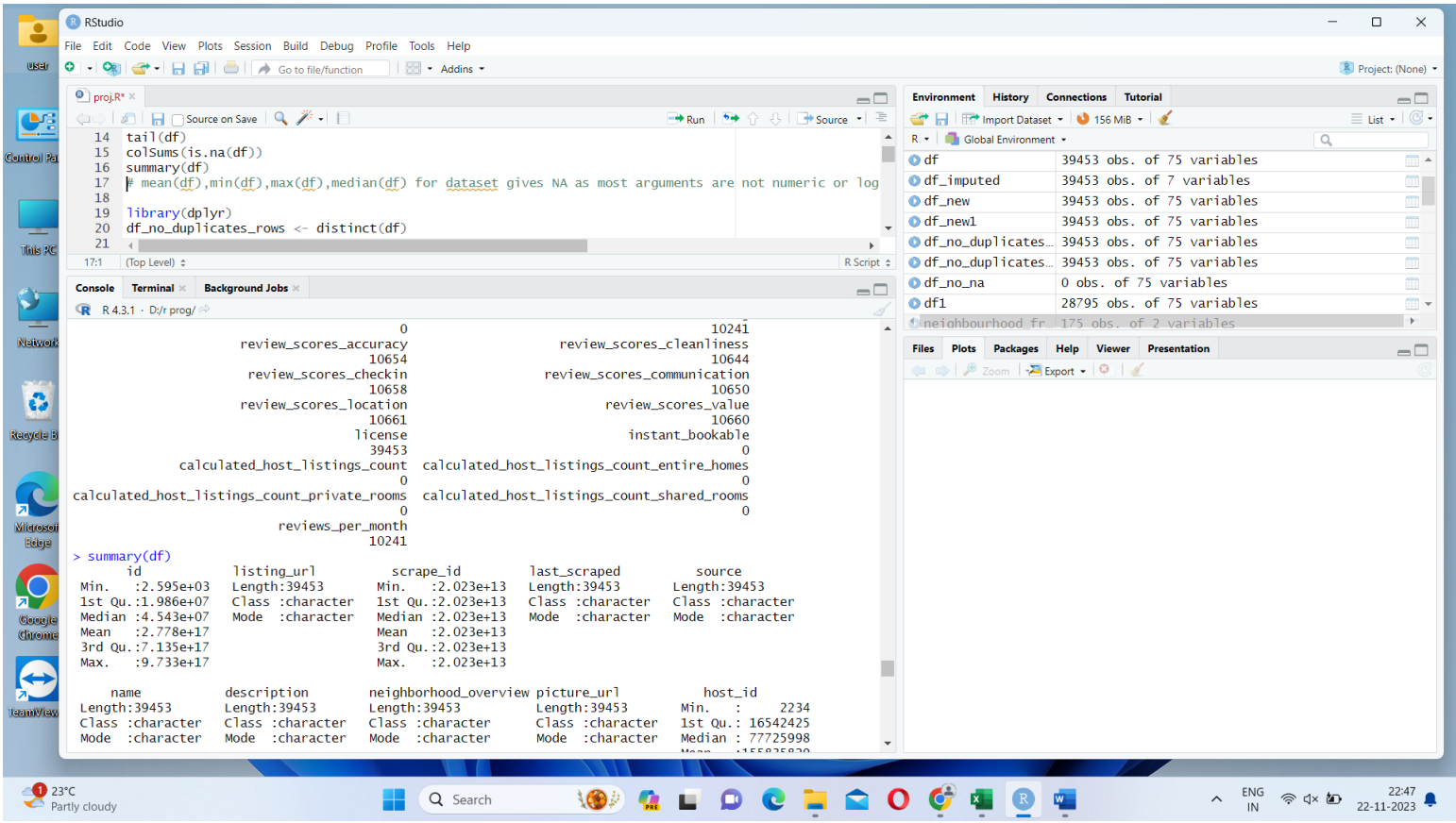

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
13 head(df)
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
host_url 0
host_name 0
host_since 0
host_location 0
host_about 0
host_response_time 0
host_response_rate 0
host_acceptance_rate 0
host_is_superhost 0
host_thumbnail_url 0
host_picture_url 0
host_neighbourhood 0
host_listings_count 5
host_total_listings_count 5
host_verifications 0
host_has_profile_pic 0
host_identity_verified 0
neighbourhood 0
neighbourhood_cleansed 0
neighbourhood_group_cleansed 0
latitude 0
longitude 0
property_type 0
room_type 0
accommodates 0
bathrooms 39453
bathrooms_text 0
bedrooms 16893
Environment History Connections Tutorial
Import Dataset
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:45 22-11-2023

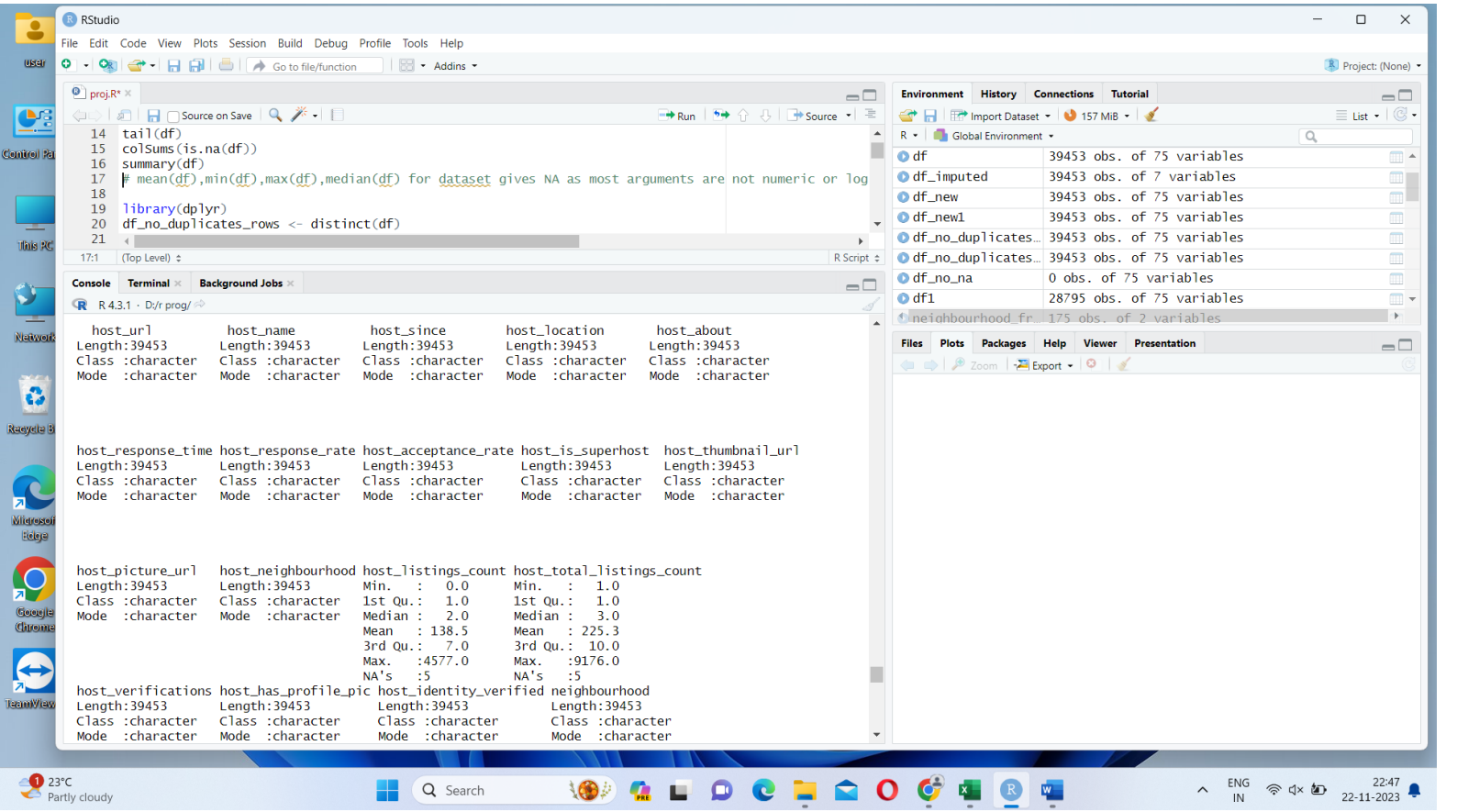

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
13 head(df)
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
beds 605
amenities 0
price 0
minimum_nights 0
maximum_nights 0
minimum_minimum_nights 0
maximum_minimum_nights 0
minimum_maximum_nights 0
maximum_maximum_nights 0
minimum_nights_avg_ntm 0
maximum_nights_avg_ntm 0
calendar_updated 39453
has_availability 0
availability_30 0
availability_60 0
availability_90 0
availability_365 0
calendar_last_scraped 0
number_of_reviews 0
number_of_reviews_ltm 0
number_of_reviews_l30d 0
first_review 0
last_review 0
review_scores_rating 10241
review_scores_accuracy 10654
review_scores_cleanliness 10644
review_scores_checkin 10658
review_scores_communication 10650
Environment History Connections Tutorial
Import Dataset
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:46 22-11-2023

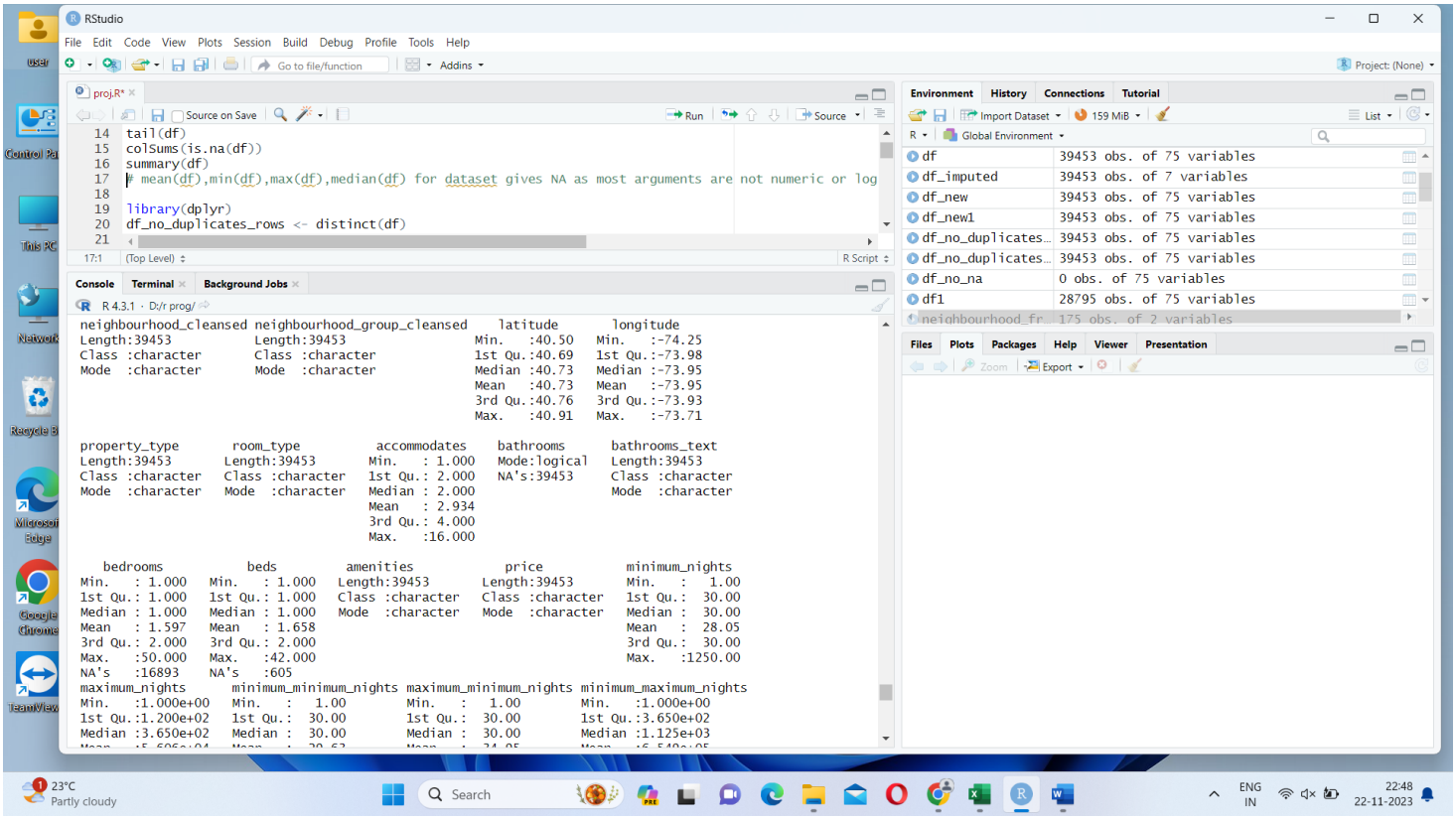


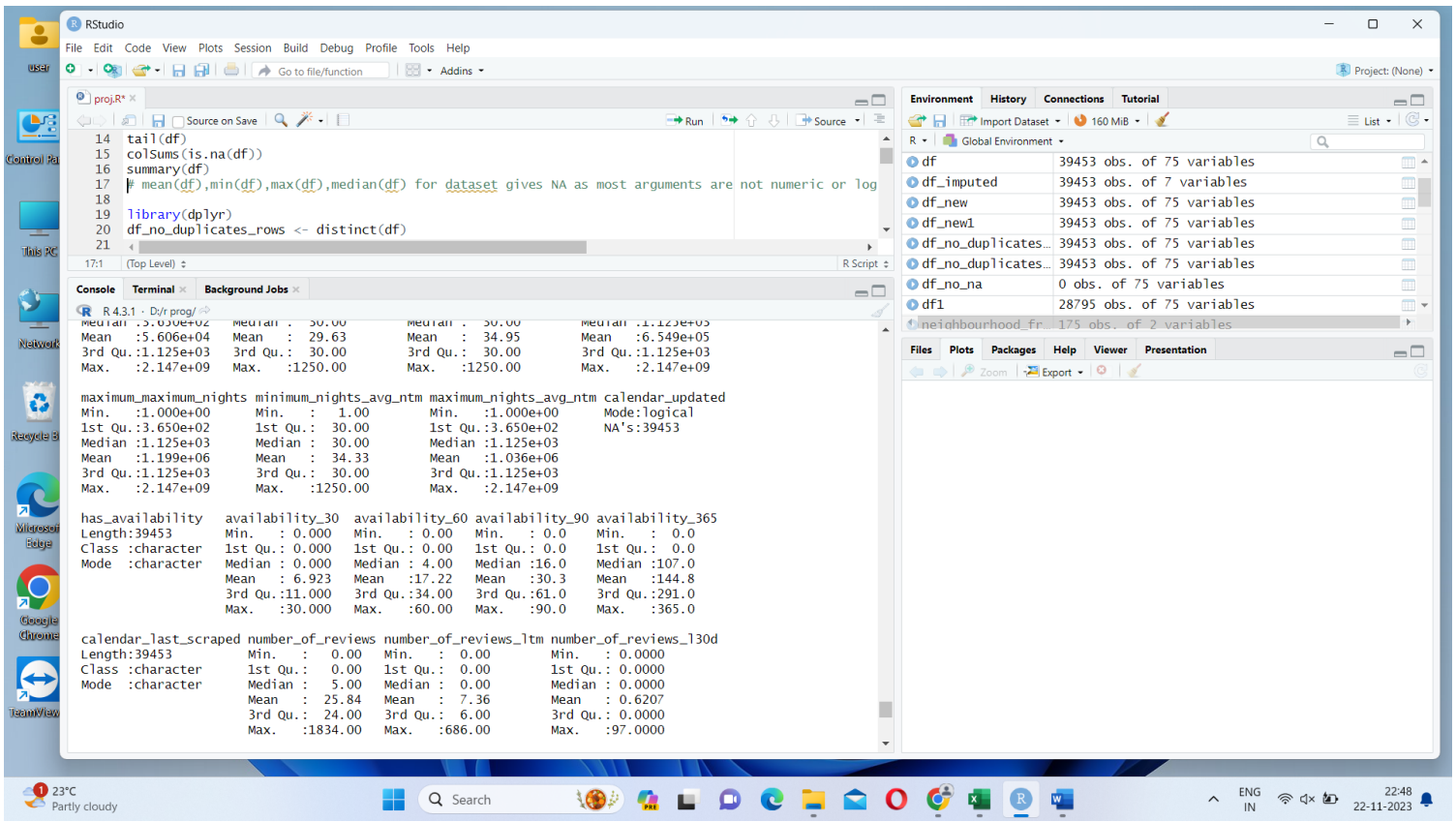
2)The dataset columns like host_listings_count ,host_total_listings_count,accommodates.bedrooms ,beds,minimum_minimum_nights ,maximum_minimum_nights, minimum_nights, minimum_nights_avg_ntm,number_of_reviews ,number_of_reviews_ltm ,number_of_reviews_l30d,calculated_host_listings_count, calculated_host_listings_count_entire_homes, calculated_host_listings_count_private_rooms, calculated_host_listings_count_shared_rooms reviews_per_month have outliers since the maximum or mimimum values lie much beyond 1.5+/-IQR from the 3 and 1st quartile respectively

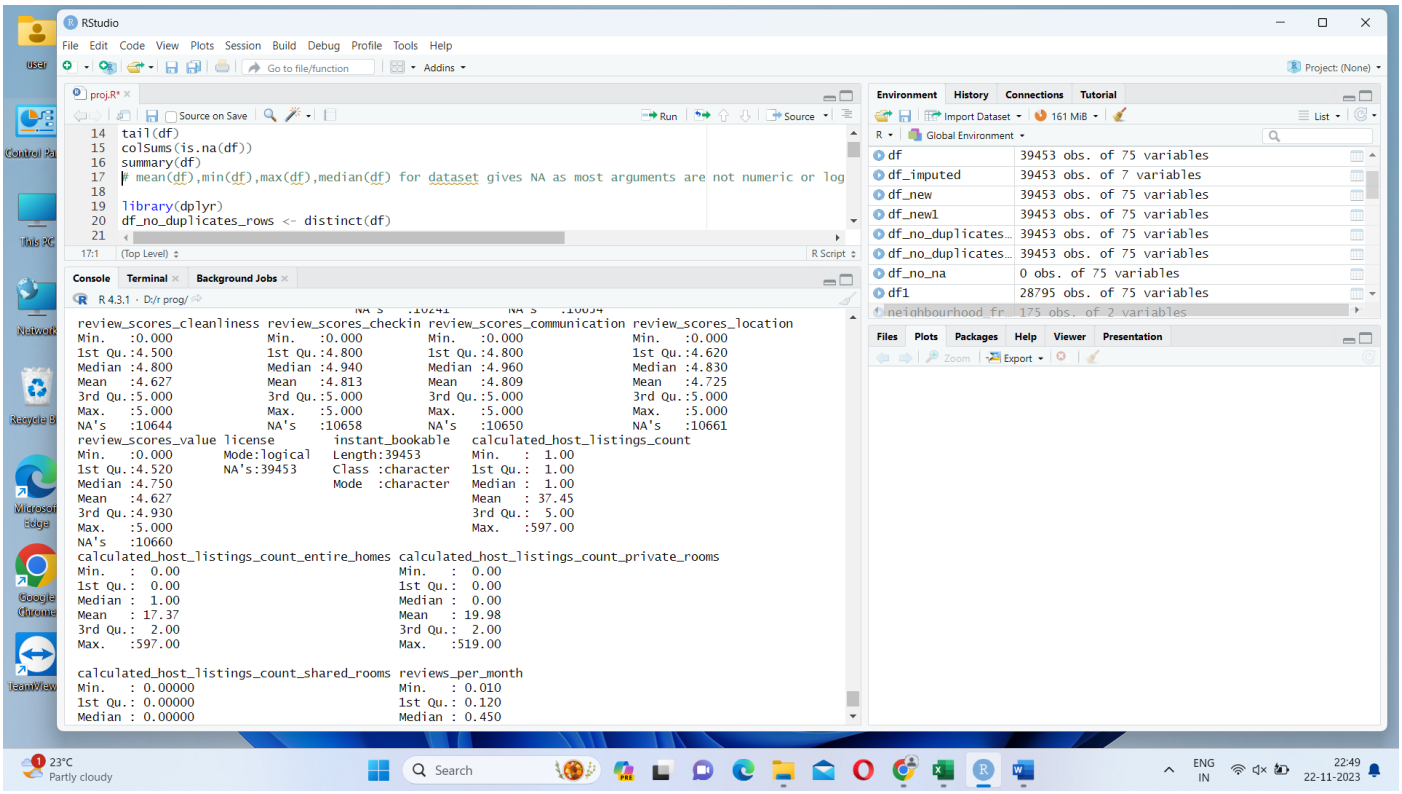
The boxplots too infer this as the dots seen there indicate outliers

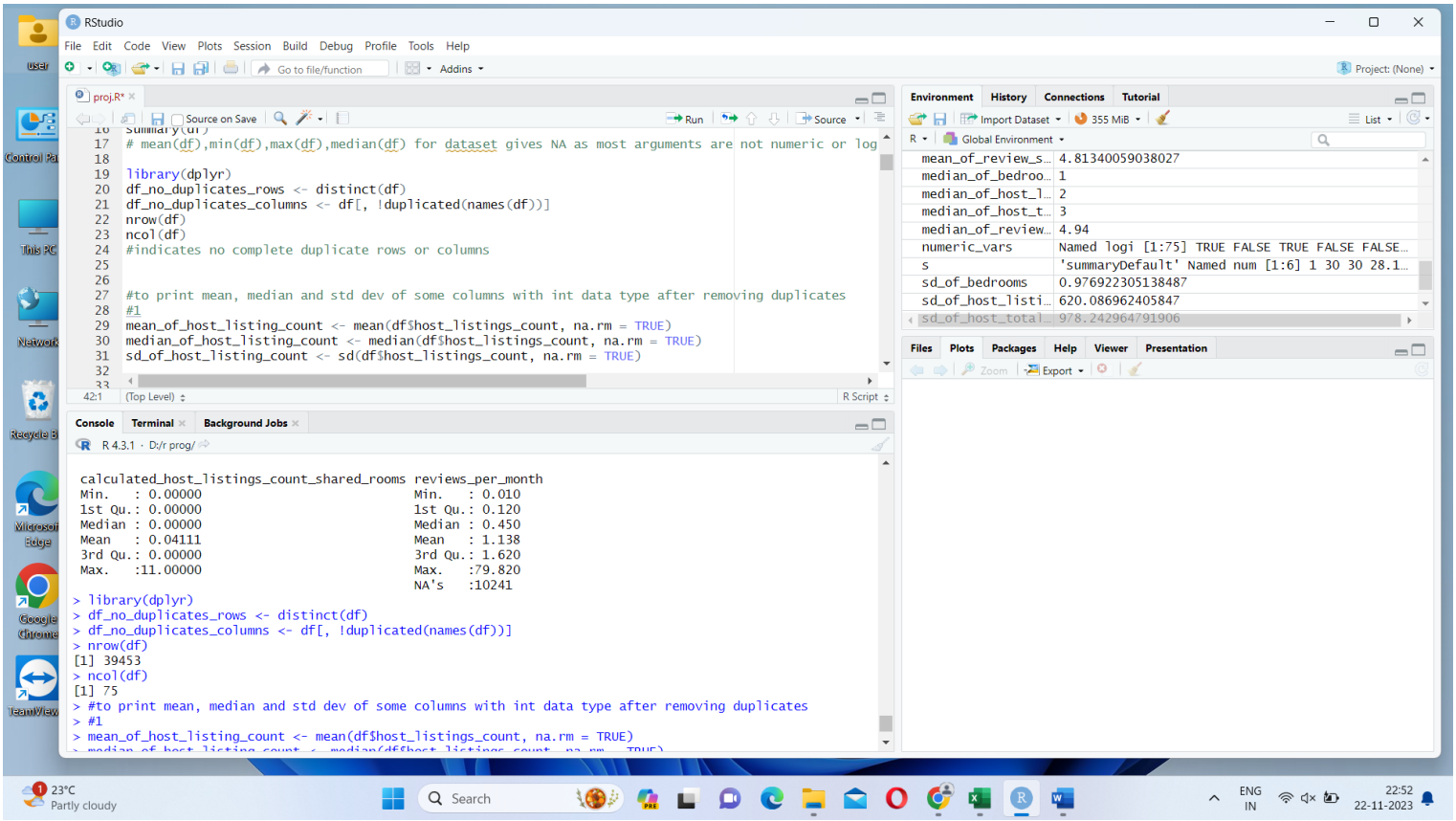

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
0 10241
review_scores_accuracy review_scores_cleanliness
10654 10644
review_scores_checkin review_scores_communication
10658 10650
review_scores_location review_scores_value
10661 10660
license instant_bookable
39453 0
calculated_host_listings_count calculated_host_listings_count_entire_homes
0 0
calculated_host_listings_count_private_rooms calculated_host_listings_count_shared_rooms
0 0
reviews_per_month
10241
> summary(df)
id listing_url scrape_id last_scraped source
Min. :2.595e+03 Length:39453 Min. :2.023e+13 Length:39453 Length:39453
1st Qu.:1.986e+07 Class :character 1st Qu.:2.023e+13 Class :character Class :character
Median :4.543e+07 Mode :character Median :2.023e+13 Mode :character Mode :character
Mean :2.778e+17 Mean :2.023e+13
3rd Qu.:7.135e+17 3rd Qu.:2.023e+13
Max. :9.733e+17 Max. :2.023e+13
name description neighborhood_overview picture_url host_id
Length:39453 Length:39453 Length:39453 Length:39453 Min. : 2234
Class :character Class :character Class :character Class :character 1st Qu.: 16542425
Mode :character Mode :character Mode :character Mode :character Median : 77725998
Environment History Connections Tutorial
Import Dataset 156 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:47 22-11-2023

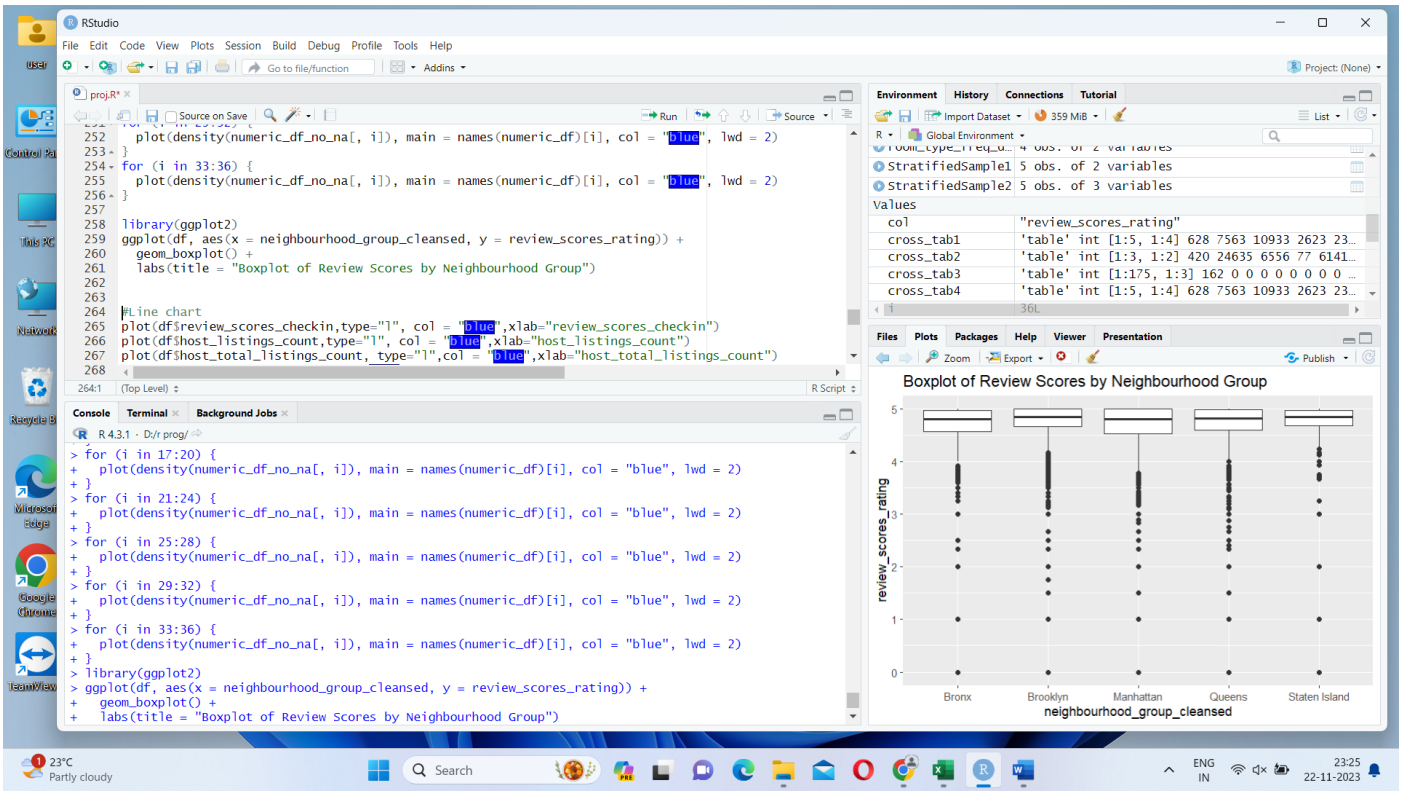

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
host_url host_name host_since host_location host_about
Length:39453 Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character Mode :character
host_response_time host_response_rate host_acceptance_rate host_is_superhost host_thumbnail_url
Length:39453 Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character Mode :character
host_picture_url host_neighbourhood host_listings_count host_total_listings_count
Length:39453 Length:39453 Min. : 0.0 Min. : 1.0
Class :character Class :character 1st Qu.: 1.0 1st Qu.: 1.0
Mode :character Mode :character Median : 2.0 Median : 3.0
Mean : 138.5 Mean : 225.3
3rd Qu.: 7.0 3rd Qu.: 10.0
Max. :4577.0 Max. :9176.0
NA's :5 NA's :5
host_verifications host_has_profile_pic host_identity_verified neighbourhood
Length:39453 Length:39453 Length:39453 Length:39453
Class :character Class :character Class :character Class :character
Mode :character Mode :character Mode :character Mode :character
Environment History Connections Tutorial
Import Dataset 157 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Export
23°C Partly cloudy
Search
ENG IN
22:47 22-11-2023

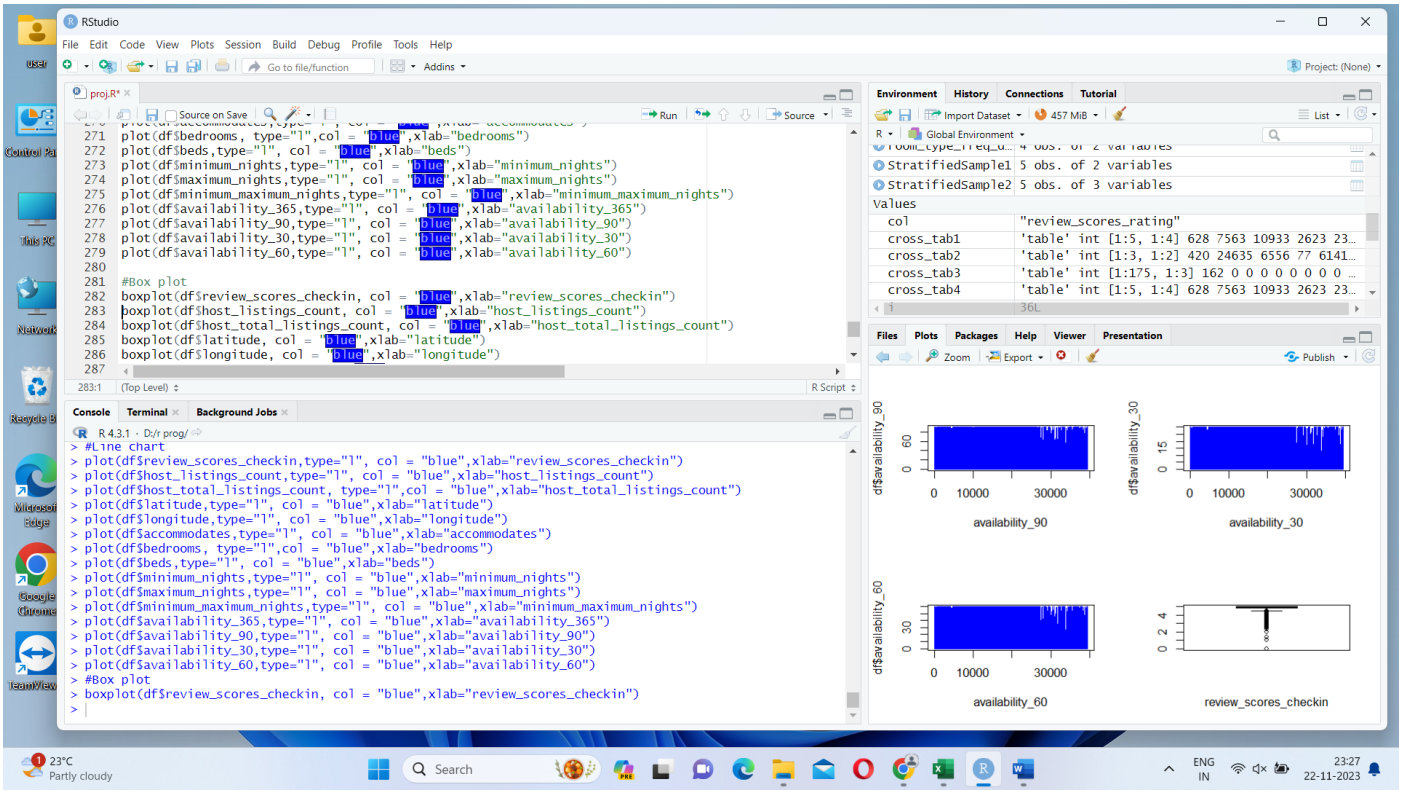

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
neighbourhood_cleansed neighbourhood_group_cleansed latitude longitude
Length:39453 Length:39453 Min. :40.50 Min. :-74.25
Class :character Class :character 1st Qu.:40.69 1st Qu.:-73.98
Mode :character Mode :character Median :40.73 Median :-73.95
Mean :40.73 Mean :-73.95
3rd Qu.:40.76 3rd Qu.:-73.93
Max. :40.91 Max. :-73.71
property_type room_type accommodates bathrooms bathrooms_text
Length:39453 Length:39453 Min. : 1.000 Mode:logical Length:39453
Class :character Class :character 1st Qu.: 2.000 NA's:39453 Class :character
Mode :character Mode :character Median : 2.000 Mode :character
Mean : 2.934
3rd Qu.: 4.000
Max. :16.000
bedrooms beds amenities price minimum_nights
Min. : 1.000 Min. : 1.000 Length:39453 Length:39453 Min. : 1.00
1st Qu.: 1.000 1st Qu.: 1.000 Class :character Class :character 1st Qu.: 30.00
Median : 1.000 Median : 1.000 Mode :character Mode :character Median : 30.00
Mean : 1.597 Mean : 1.658 Mean : 28.05
3rd Qu.: 2.000 3rd Qu.: 2.000 3rd Qu.: 30.00
Max. :50.000 Max. :42.000 Max. :1250.00
NA's :16893 NA's :605
maximum_nights minimum_minimum_nights maximum_minimum_nights minimum_maximum_nights
Min. :1.000e+00 Min. : 1.00 Min. : 1.00 Min. :1.000e+00
1st Qu.:1.200e+02 1st Qu.: 30.00 1st Qu.: 30.00 1st Qu.:3.650e+02
Median :3.650e+02 Median : 30.00 Median : 30.00 Median :1.125e+03
Environment History Connections Tutorial
Import Dataset 159 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C
Partly cloudy
Search
ENG
IN
22:48
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Mean :5.606e+04 Mean : 29.63 Mean : 34.95 Mean :6.549e+05
3rd Qu.:1.125e+03 3rd Qu.: 30.00 3rd Qu.: 30.00 3rd Qu.:1.125e+03
Max. :2.147e+09 Max. :1250.00 Max. :1250.00 Max. :2.147e+09
maximum_maximum_nights minimum_nights_avg_ntm maximum_nights_avg_ntm calendar_updated
Min. :1.000e+00 Min. : 1.00 Min. :1.000e+00 Mode:logical
1st Qu.:3.650e+02 1st Qu.: 30.00 1st Qu.:3.650e+02 NA's:39453
Median :1.125e+03 Median : 30.00 Median :1.125e+03
Mean :1.199e+06 Mean : 34.33 Mean :1.036e+06
3rd Qu.:1.125e+03 3rd Qu.: 30.00 3rd Qu.:1.125e+03
Max. :2.147e+09 Max. :1250.00 Max. :2.147e+09
has_availability availability_30 availability_60 availability_90 availability_365
Length:39453 Min. : 0.000 Min. : 0.00 Min. : 0.0 Min. : 0.0
Class :character 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 0.0
Mode :character Median : 0.000 Median : 4.00 Median :16.0 Median :107.0
Mean : 6.923 Mean :17.22 Mean :30.3 Mean :144.8
3rd Qu.:11.000 3rd Qu.:34.00 3rd Qu.:61.0 3rd Qu.:291.0
Max. :30.000 Max. :60.00 Max. :90.0 Max. :365.0
calendar_last_scraped number_of_reviews number_of_reviews_ltm number_of_reviews_l30d
Length:39453 Min. : 0.00 Min. : 0.00 Min. : 0.0000
Class :character 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.0000
Mode :character Median : 5.00 Median : 0.00 Median : 0.0000
Mean : 25.84 Mean : 7.36 Mean : 0.6207
3rd Qu.: 24.00 3rd Qu.: 6.00 3rd Qu.: 0.0000
Max. :1834.00 Max. :686.00 Max. :97.0000
Environment History Connections Tutorial
Import Dataset 160 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
neighbourhood_fr... 175 obs. of 2 variables
Files Plots Packages Help Viewer Presentation
Export
23°C
Partly cloudy
Search
ENG
IN
22:48
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
14 tail(df)
15 colSums(is.na(df))
16 summary(df)
17 # mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
18
19 library(dplyr)
20 df_no_duplicates_rows <- distinct(df)
21
Environment History Connections Tutorial
Import Dataset 161 MiB
Global Environment
df 39453 obs. of 75 variables
df_imputed 39453 obs. of 7 variables
df_new 39453 obs. of 75 variables
df_new1 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_duplicates... 39453 obs. of 75 variables
df_no_na 0 obs. of 75 variables
df1 28795 obs. of 75 variables
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
review_scores_cleanliness review_scores_checkin review_scores_communication review_scores_location
Min. :0.000 Min. :0.000 Min. :0.000 Min. :0.000
1st Qu.:4.500 1st Qu.:4.800 1st Qu.:4.800 1st Qu.:4.620
Median :4.800 Median :4.940 Median :4.960 Median :4.830
Mean :4.627 Mean :4.813 Mean :4.809 Mean :4.725
3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.:5.000 3rd Qu.:5.000
Max. :5.000 Max. :5.000 Max. :5.000 Max. :5.000
NA's :10644 NA's :10658 NA's :10650 NA's :10661
review_scores_value license instant_bookable calculated_host_listings_count
Min. :0.000 Mode:logical Length:39453 Min. : 1.00
1st Qu.:4.520 NA's:39453 Class :character 1st Qu.: 1.00
Median :4.750 Mode :character Median : 1.00
Mean :4.627 Mean : 37.45
3rd Qu.:4.930 3rd Qu.: 5.00
Max. :5.000 Max. :597.00
NA's :10660
calculated_host_listings_count_entire_homes calculated_host_listings_count_private_rooms
Min. : 0.00 Min. : 0.00
1st Qu.: 0.00 1st Qu.: 0.00
Median : 1.00 Median : 0.00
Mean : 17.37 Mean : 19.98
3rd Qu.: 2.00 3rd Qu.: 2.00
Max. :597.00 Max. :519.00
calculated_host_listings_count_shared_rooms reviews_per_month
Min. : 0.00000 Min. : 0.010
1st Qu.: 0.00000 1st Qu.: 0.120
Median : 0.00000 Median : 0.450
23°C
Partly cloudy
Search
ENG
IN
22:49
22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
# mean(df),min(df),max(df),median(df) for dataset gives NA as most arguments are not numeric or log
library(dplyr)
df_no_duplicates_rows <- distinct(df)
df_no_duplicates_columns <- df[, !duplicated(names(df))]
nrow(df)
ncol(df)
#indicates no complete duplicate rows or columns
#to print mean, median and std dev of some columns with int data type after removing duplicates
#1
mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
median_of_host_listing_count <- median(df$host_listings_count, na.rm = TRUE)
sd_of_host_listing_count <- sd(df$host_listings_count, na.rm = TRUE)
Environment History Connections Tutorial
Import Dataset 355 MiB
Global Environment
mean_of_review_s... 4.81340059038027
median_of_bedroo... 1
median_of_host_l... 2
median_of_host_t... 3
median_of_review... 4.94
numeric_vars Named logi [1:75] TRUE FALSE TRUE FALSE FALSE...
s 'summaryDefault' Named num [1:6] 1 30 30 28.1...
sd_of_bedrooms 0.976922305138487
sd_of_host_listi... 620.086962405847
sd_of_host_total... 978.242964791906
Files Plots Packages Help Viewer Presentation
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
calculated_host_listings_count_shared_rooms reviews_per_month
Min. : 0.00000 Min. : 0.010
1st Qu.: 0.00000 1st Qu.: 0.120
Median : 0.00000 Median : 0.450
Mean : 0.04111 Mean : 1.138
3rd Qu.: 0.00000 3rd Qu.: 1.620
Max. :11.00000 Max. :79.820
NA's :10241
> library(dplyr)
> df_no_duplicates_rows <- distinct(df)
> df_no_duplicates_columns <- df[, !duplicated(names(df))]
> nrow(df)
[1] 39453
> ncol(df)
[1] 75
> #to print mean, median and std dev of some columns with int data type after removing duplicates
> #1
> mean_of_host_listing_count <- mean(df$host_listings_count, na.rm = TRUE)
23°C
Partly cloudy
Search
ENG
IN
22:52
22-11-2023

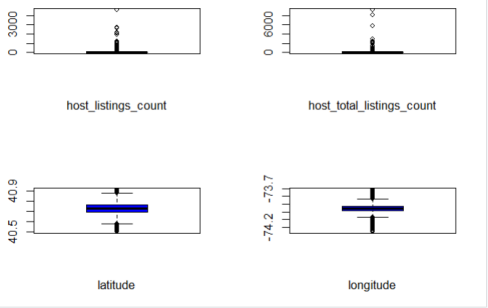




Last one in this image

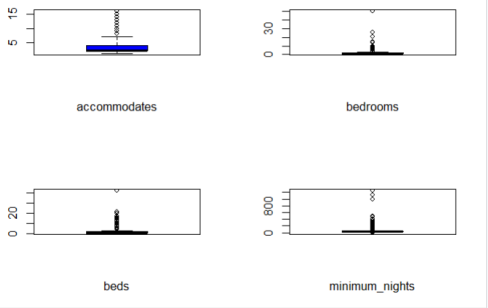
```
plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
plot(df$availability_60,type="l", col = "blue",xlab="availability_60")

#Box plot
boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
boxplot(df$latitude, col = "blue",xlab="latitude")
boxplot(df$longitude, col = "blue",xlab="longitude")
boxplot(df$accommodates, col = "blue",xlab="accommodates")
boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
boxplot(df$beds, col = "blue",xlab="beds")
boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
```

```
> plot(df$latitude,type="l", col = "blue",xlab="latitude")
> plot(df$longitude,type="l", col = "blue",xlab="longitude")
> plot(df$accommodates,type="l", col = "blue",xlab="accommodates")
> plot(df$bedrooms, type="l",col = "blue",xlab="bedrooms")
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
> plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
> plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
> plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
> #Box plot
> boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
> boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
> boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
> boxplot(df$latitude, col = "blue",xlab="latitude")
> boxplot(df$longitude, col = "blue",xlab="longitude")
>
```




```
plot(df$availability_60,type="l", col = "blue",xlab="availability_60")

#Box plot
boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
boxplot(df$latitude, col = "blue",xlab="latitude")
boxplot(df$longitude, col = "blue",xlab="longitude")
boxplot(df$accommodates, col = "blue",xlab="accommodates")
boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
boxplot(df$beds, col = "blue",xlab="beds")
boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
boxplot(df$maximum_nights, col = "blue",xlab="maximum_nights")
boxplot(df$minimum_maximum_nights, col = "blue",xlab="minimum_maximum_nights")
boxplot(df$availability_365, col = "blue",xlab="availability_365")
boxplot(df$availability_90, col = "blue",xlab="availability_90")
```

```
> plot(df$beds,type="l", col = "blue",xlab="beds")
> plot(df$minimum_nights,type="l", col = "blue",xlab="minimum_nights")
> plot(df$maximum_nights,type="l", col = "blue",xlab="maximum_nights")
> plot(df$minimum_maximum_nights,type="l", col = "blue",xlab="minimum_maximum_nights")
> plot(df$availability_365,type="l", col = "blue",xlab="availability_365")
> plot(df$availability_90,type="l", col = "blue",xlab="availability_90")
> plot(df$availability_30,type="l", col = "blue",xlab="availability_30")
> plot(df$availability_60,type="l", col = "blue",xlab="availability_60")
> #Box plot
> boxplot(df$review_scores_checkin, col = "blue",xlab="review_scores_checkin")
> boxplot(df$host_listings_count, col = "blue",xlab="host_listings_count")
> boxplot(df$host_total_listings_count, col = "blue",xlab="host_total_listings_count")
> boxplot(df$latitude, col = "blue",xlab="latitude")
> boxplot(df$longitude, col = "blue",xlab="longitude")
> boxplot(df$accommodates, col = "blue",xlab="accommodates")
> boxplot(df$bedrooms, col = "blue",xlab="bedrooms")
> boxplot(df$beds, col = "blue",xlab="beds")
> boxplot(df$minimum_nights, col = "blue",xlab="minimum_nights")
>
```



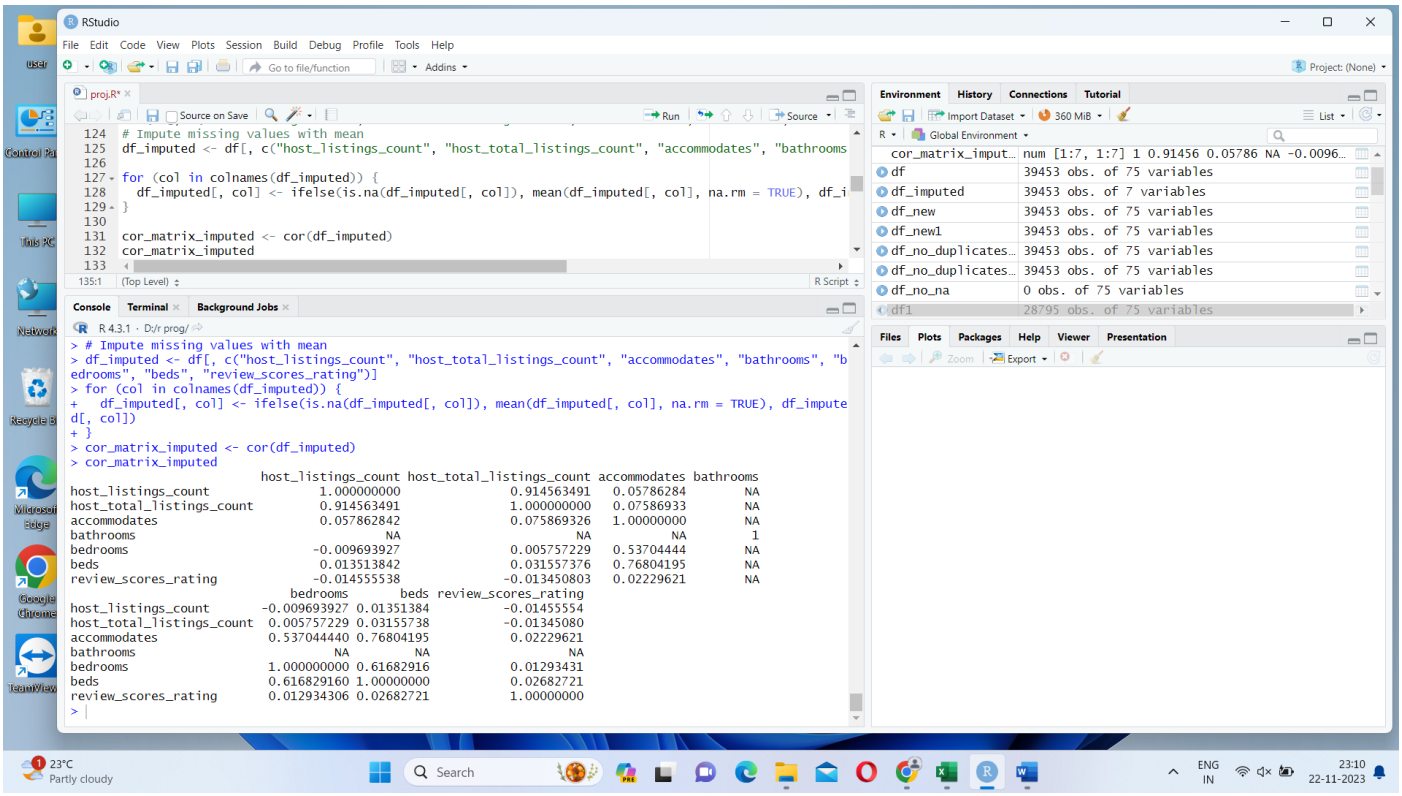
3) The coorelation matrix for the columns host_listings_count , host_total_listings_count ,accommodates, bathrooms,bedrooms ,beds, review_scores_rating was tabulated

Bedrooms and review score rating showed negative correlation with host_listings_count

Review score rating also showed negative correlation with host_total_listings_count

All other pairs showed a positive correlation

Negative correlation indicates that as one variable decreases the other one decreases too and vice versa for positive
Thus it is seen that as host listings count decreased so did the bed room and review score rating decrease and when host total listing count decreased the review score rating also decreased.



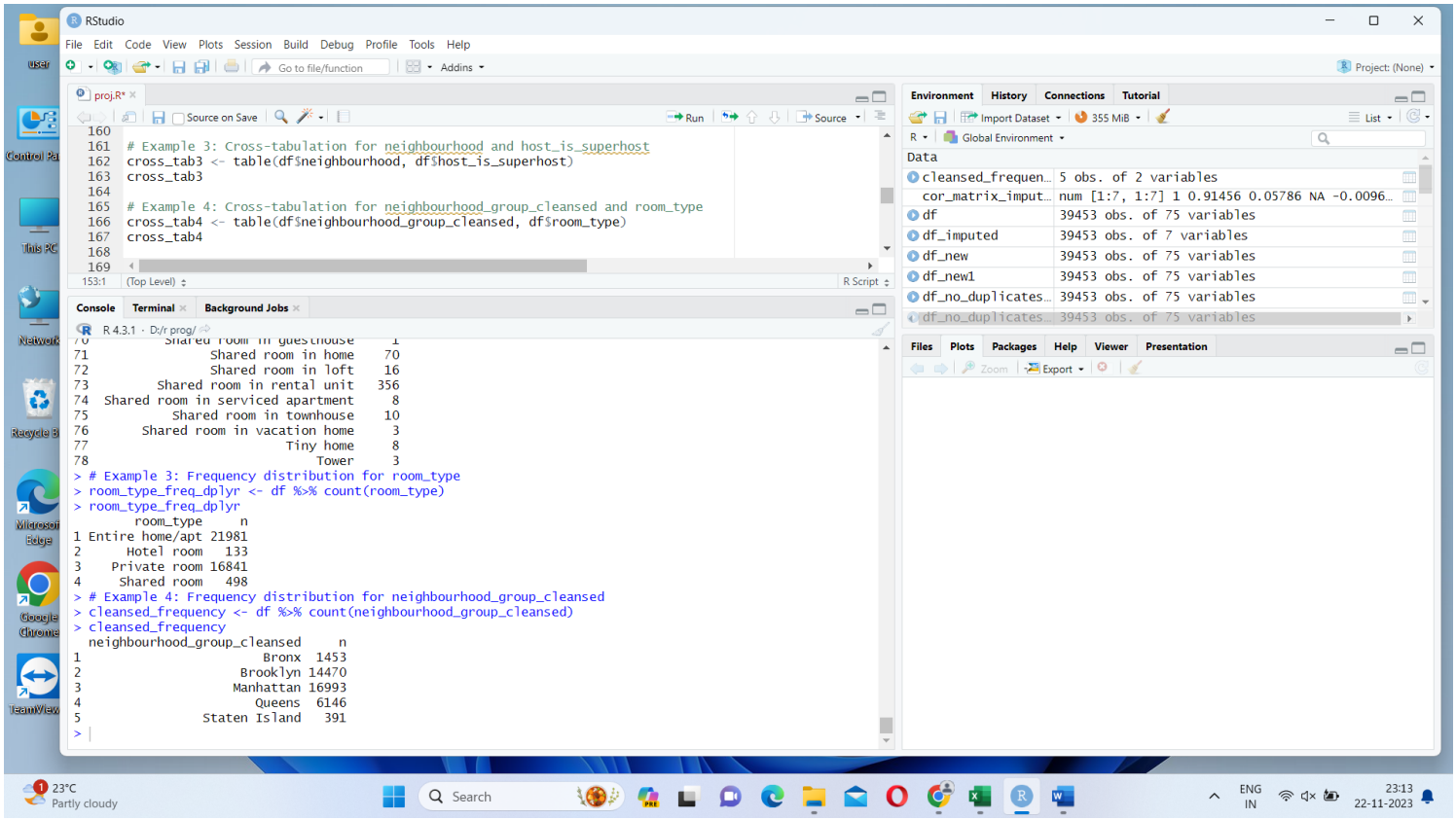
4)When frequency distributions for some variables were tabula ted we saw that New York ,United States had highest count fo r neighbourhood.

```
101                 New York city, New York, United States     1
102                     New York, New York , United States     1
103                           New York, Ny, United States     1
104                  New York, US, New York, United States     1
105                              New York, United States  9780
106                     New york, New York, United States     2
107                     New-York, New York, United States     1
108                      Newyork, New York, United States     1
109                    Nueva York, New York, United States     1
```

Entire rental unit had the highest property type listing

```
11                  Entire guest suite   340
12                   Entire guesthouse    66
13                         Entire home  1577
14                     Entire home/apt    11
15                         Entire loft   596
16                        Entire place    83
17                  Entire rental unit 16607
18          Entire serviced apartment   673
19                    Entire townhouse   537
20                Entire vacation home    40
21                        Entire villa     7
22                           Farm stay     1
```

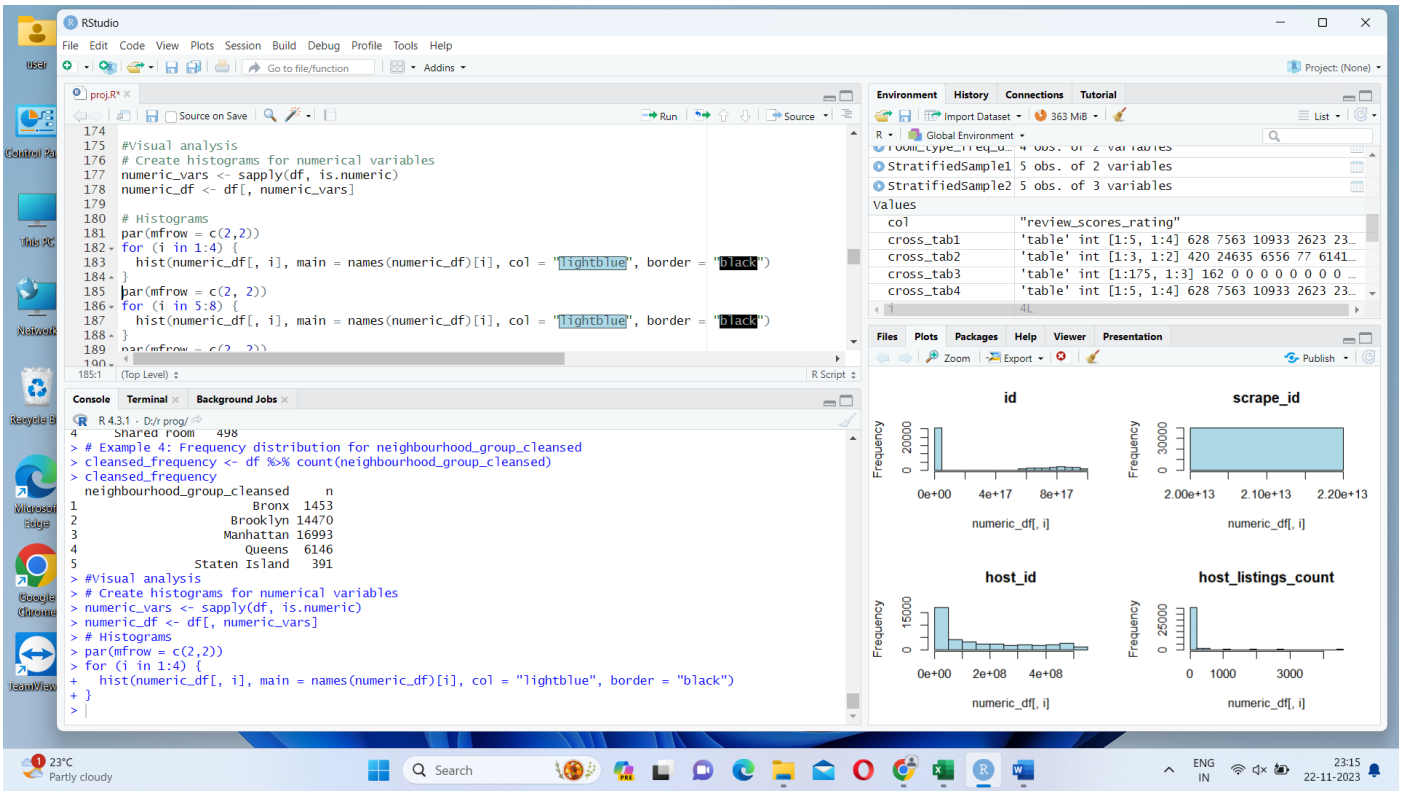
Entire home/apt had the highest room type and hotel room had le ast in dataset



Manhattan had the highest neighbourhood_group_cleansed listi ng and Staten island had the least in the dataset

5) From the histograms and density plots :

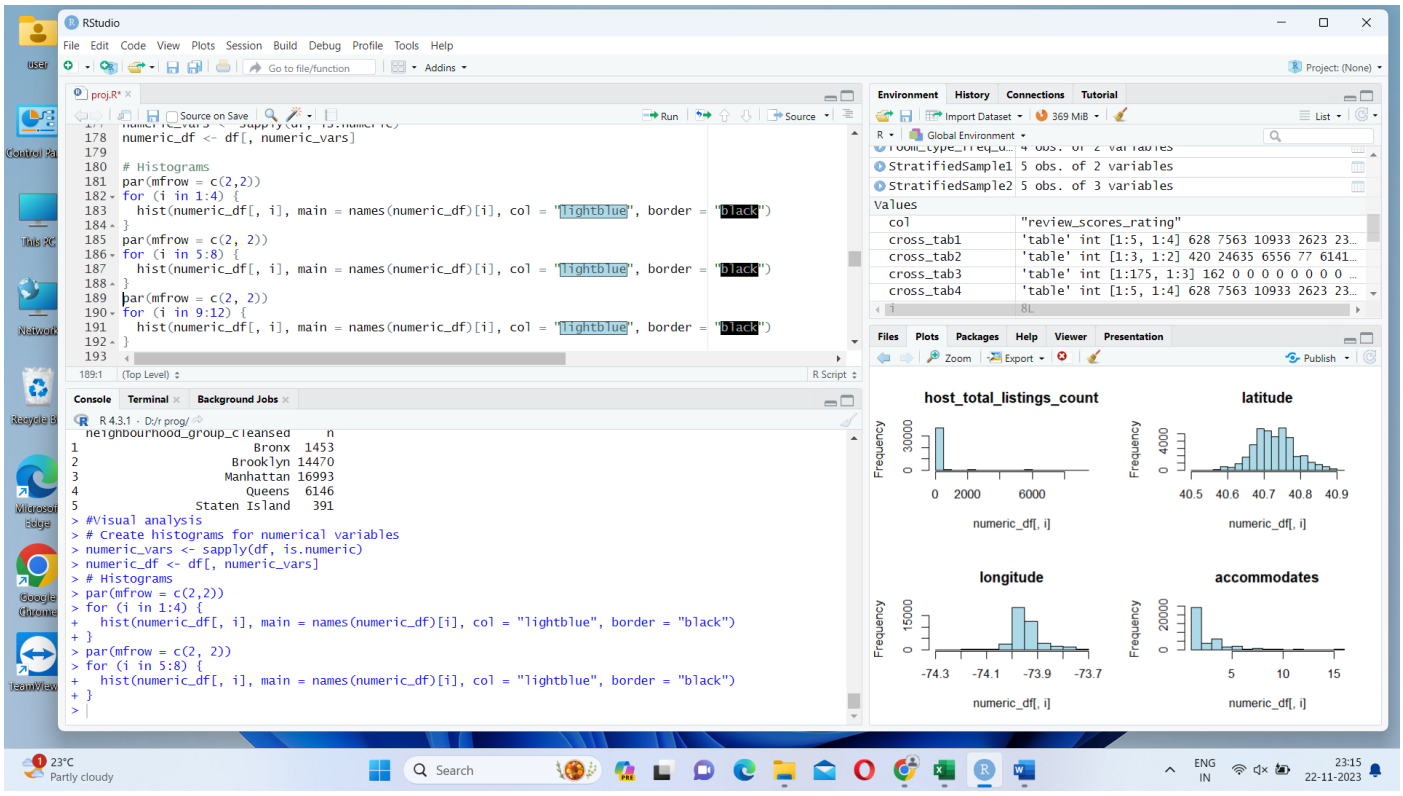
i)Most host listings are less than 1000



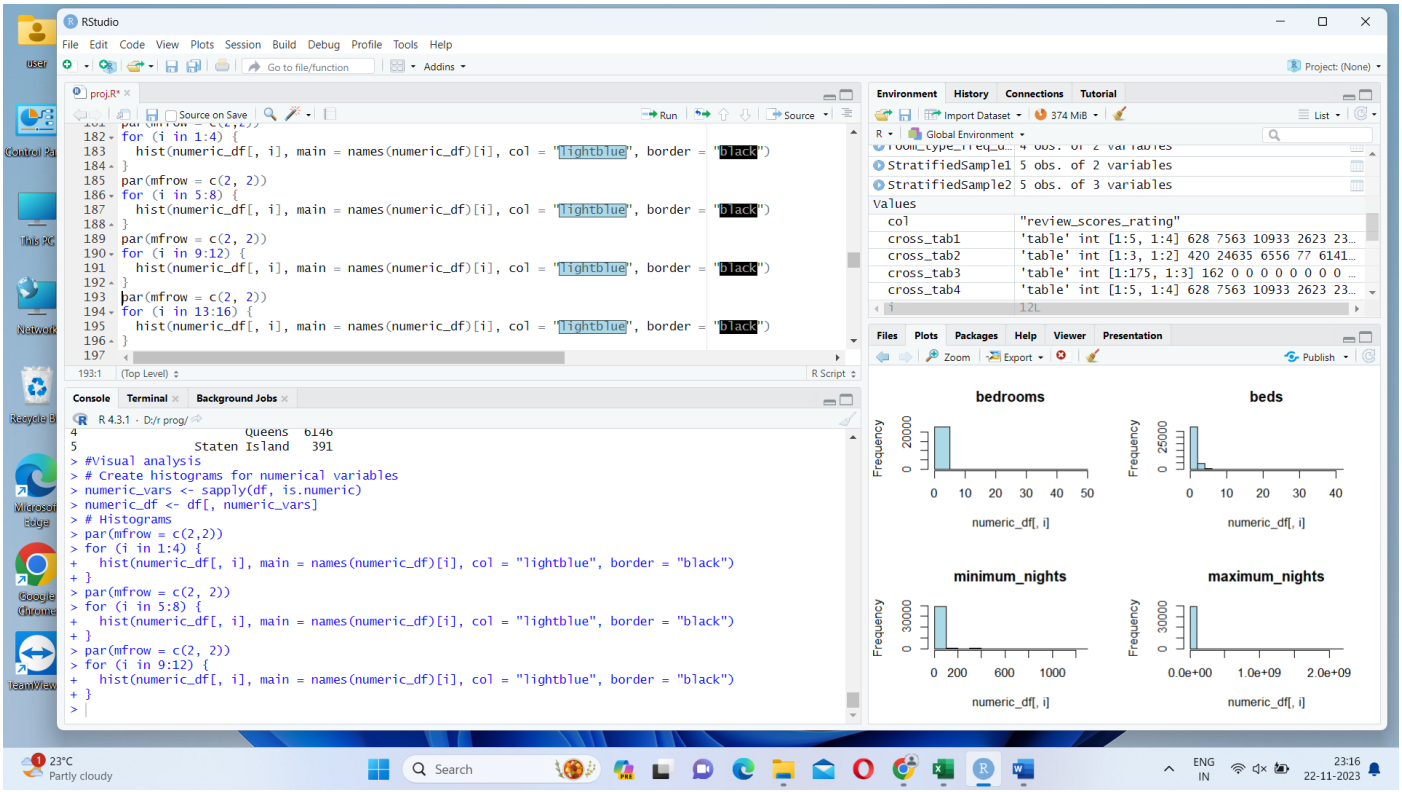
ii)Most host total listings are less than 1000

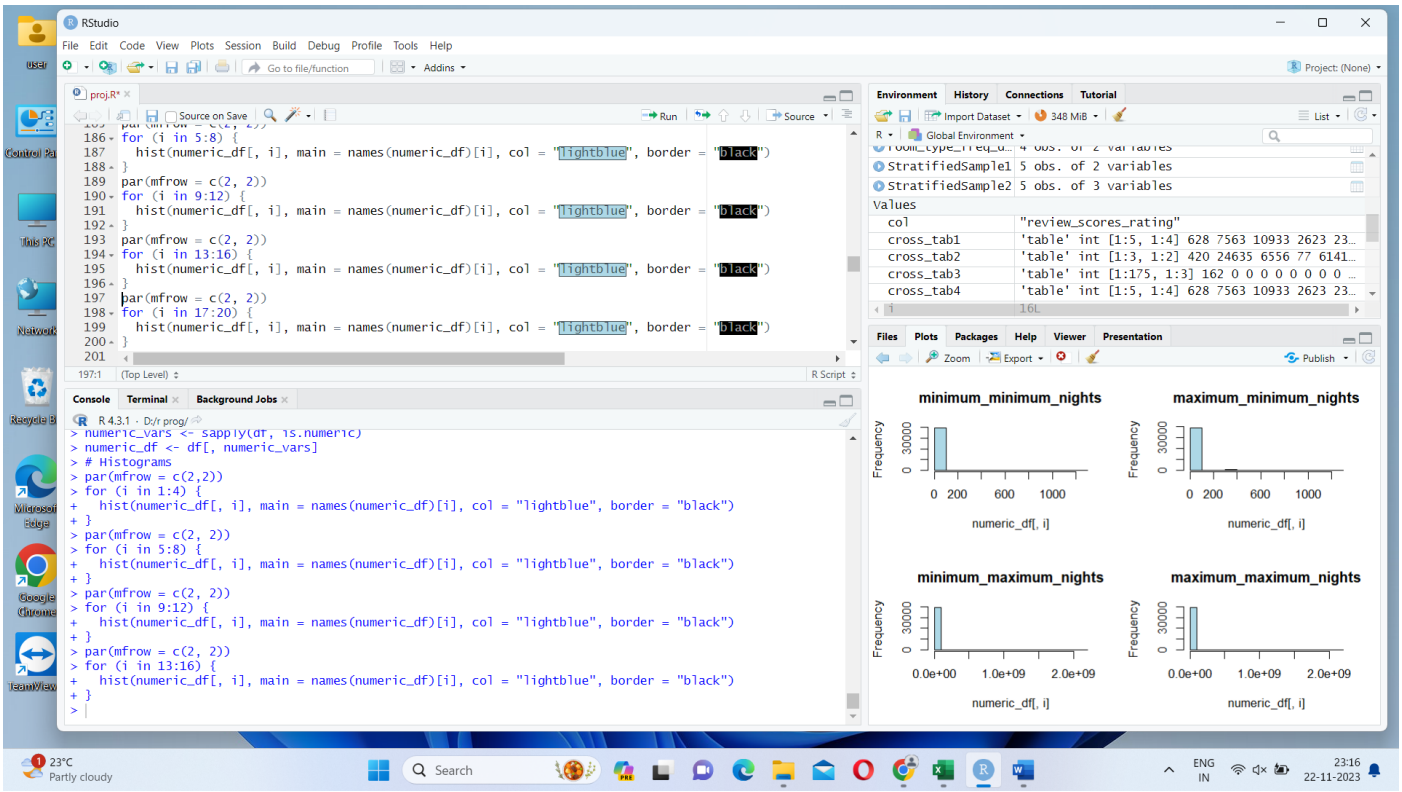
iii)Latitude values show a normal distribution

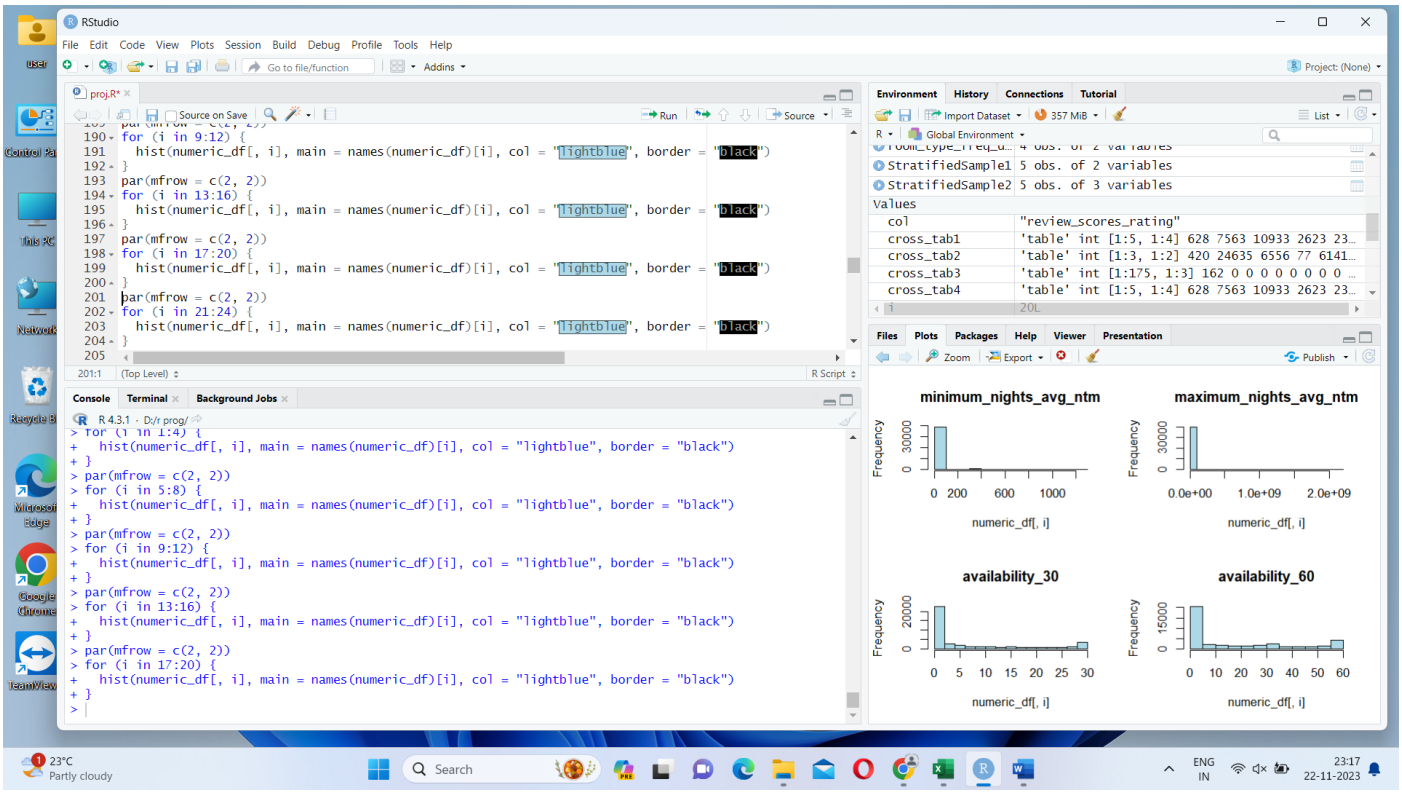
iv)Most listings have accommodates less than 5

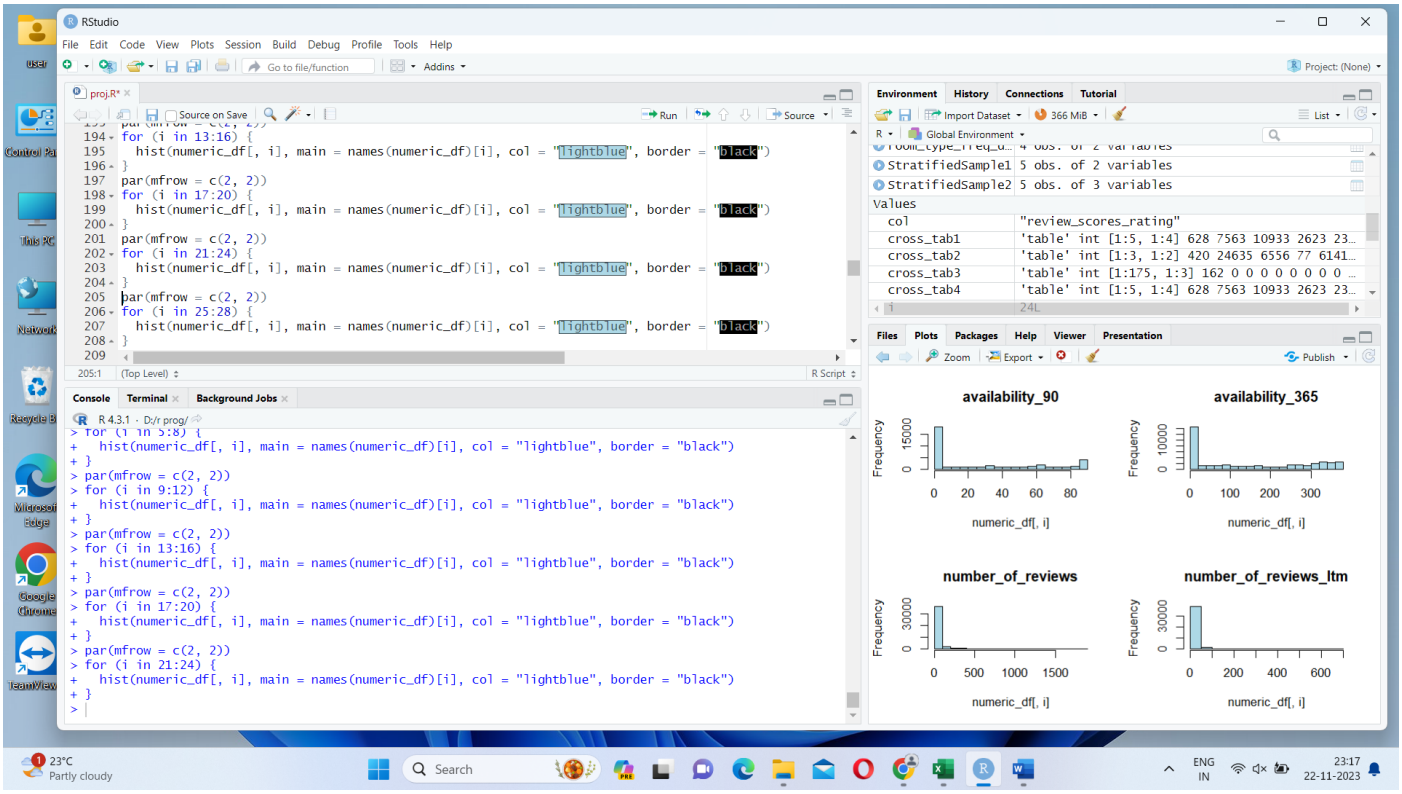


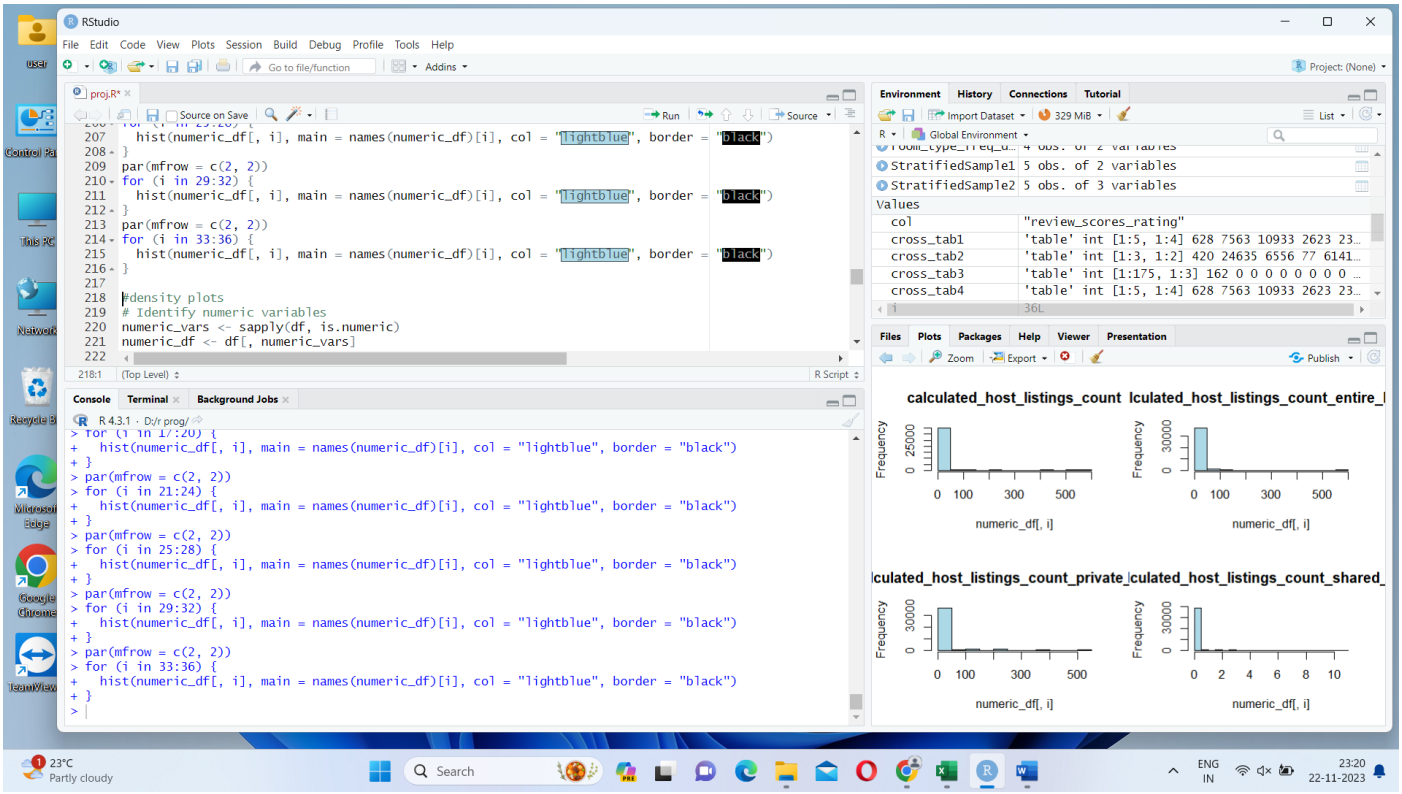
v) Most categorical variable values are all concentrated below a particular value(right skewed (right tail)plot)



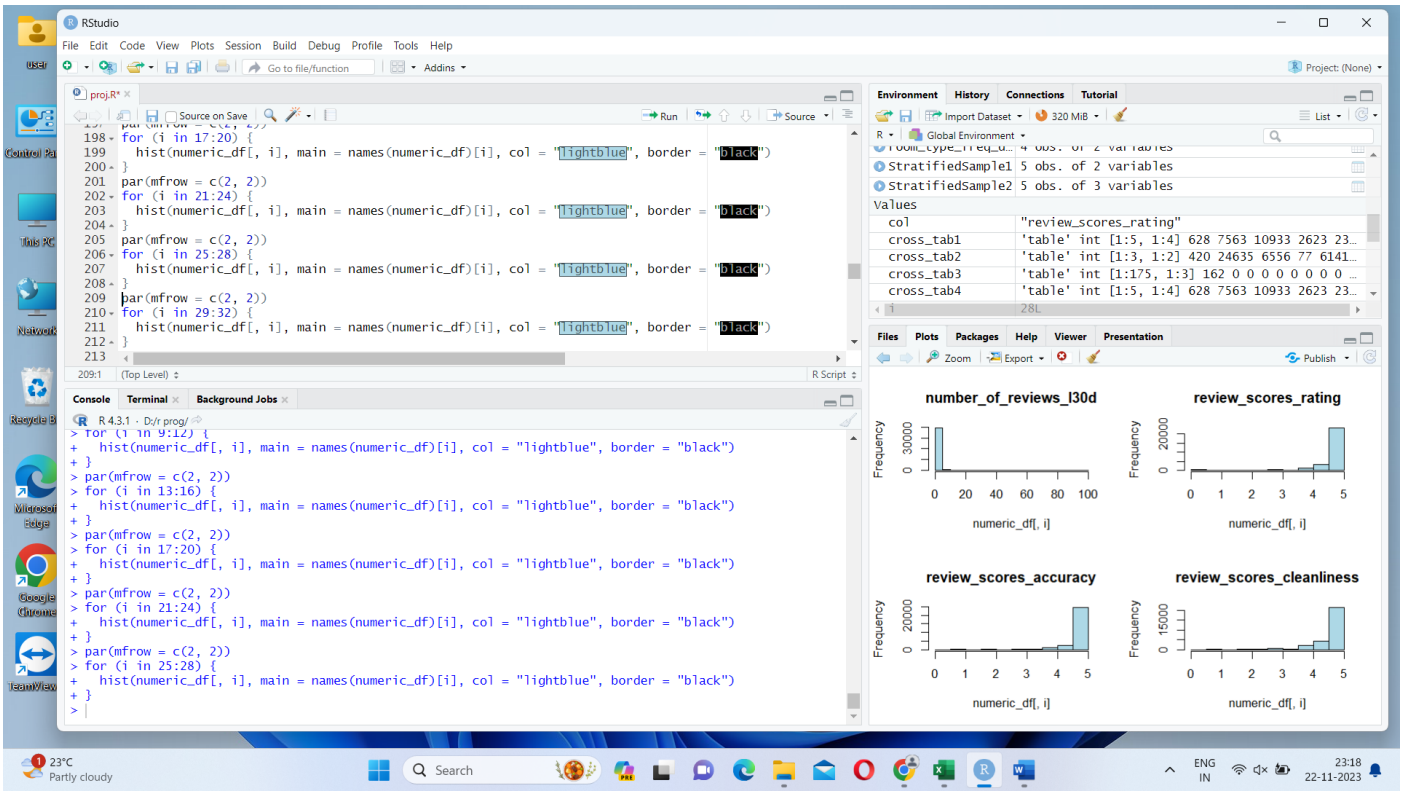


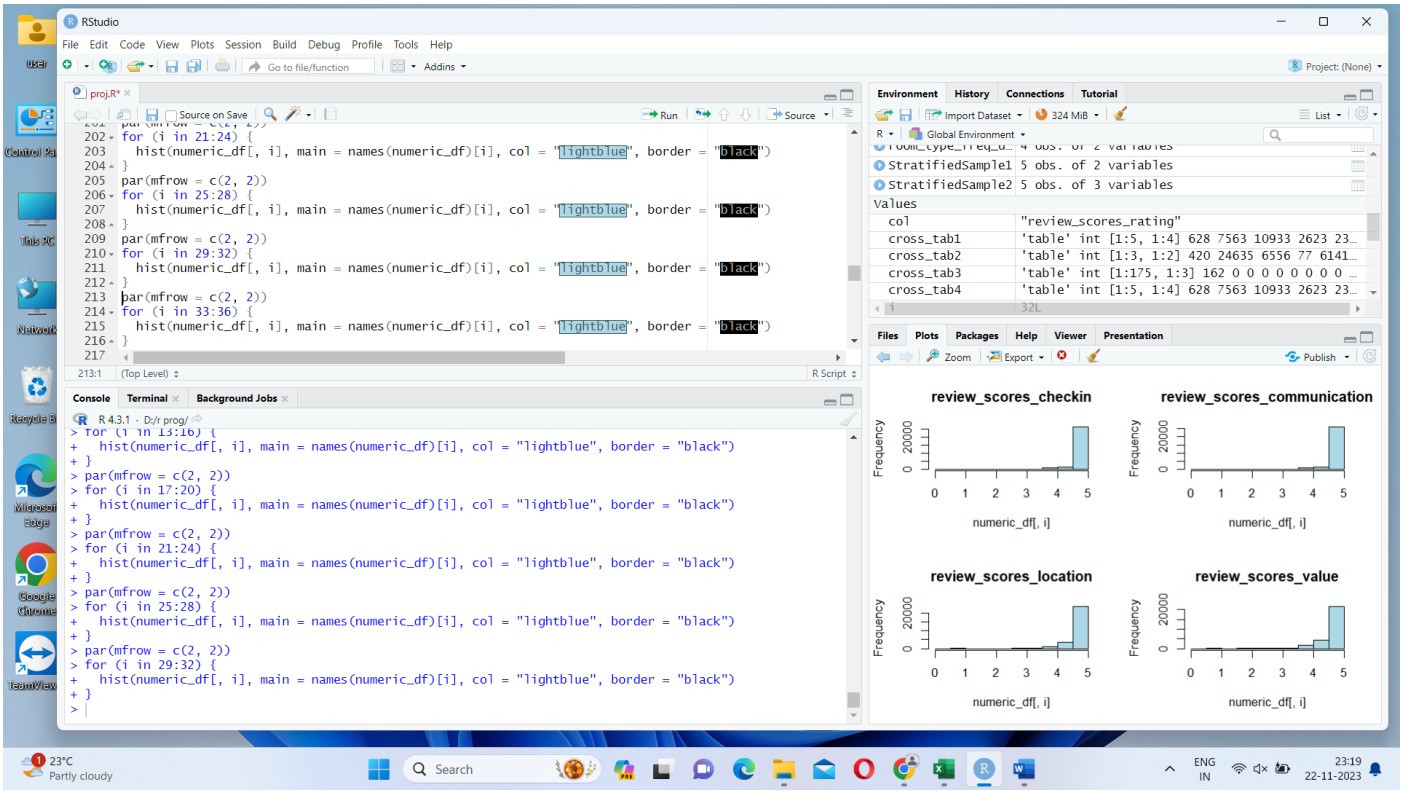

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
for (i in 9:12) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 13:16) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 17:20) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 21:24) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Environment History Connections Tutorial
Import Dataset 357 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
minimum_nights_avg_ntm
maximum_nights_avg_ntm
availability_30
availability_60
Frequency
numeric_df[, i]
23°C Partly cloudy
Search
ENG IN
23:17 22-11-2023


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
proj.R*
for (i in 13:16) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 17:20) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 21:24) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
par(mfrow = c(2, 2))
for (i in 25:28) {
hist(numeric_df[, i], main = names(numeric_df)[i], col = "lightblue", border = "black")
}
Console Terminal Background Jobs
R 4.3.1 · D:/r prog/
Environment History Connections Tutorial
Import Dataset 366 MiB
Global Environment
StratifiedSample1 5 obs. of 2 variables
StratifiedSample2 5 obs. of 3 variables
Values
col "review_scores_rating"
cross_tab1 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
cross_tab2 'table' int [1:3, 1:2] 420 24635 6556 77 6141...
cross_tab3 'table' int [1:175, 1:3] 162 0 0 0 0 0 0 0 0 ...
cross_tab4 'table' int [1:5, 1:4] 628 7563 10933 2623 23...
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
availability_90
availability_365
number_of_reviews
number_of_reviews_ltm
Frequency
numeric_df[, i]
23°C Partly cloudy
Search
ENG IN
23:17 22-11-2023

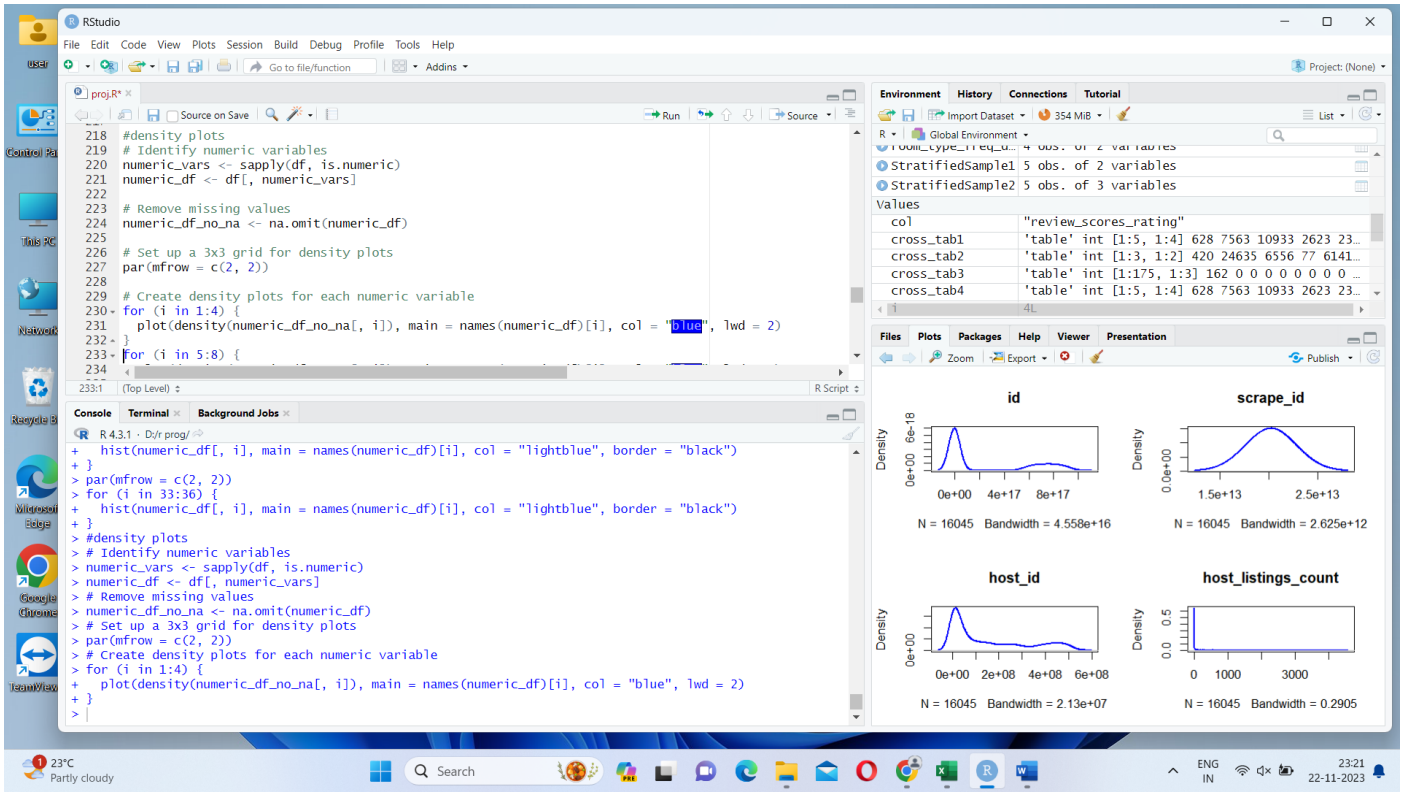


vi) all review related columns have most values above a particu lar value(left skewed plot)

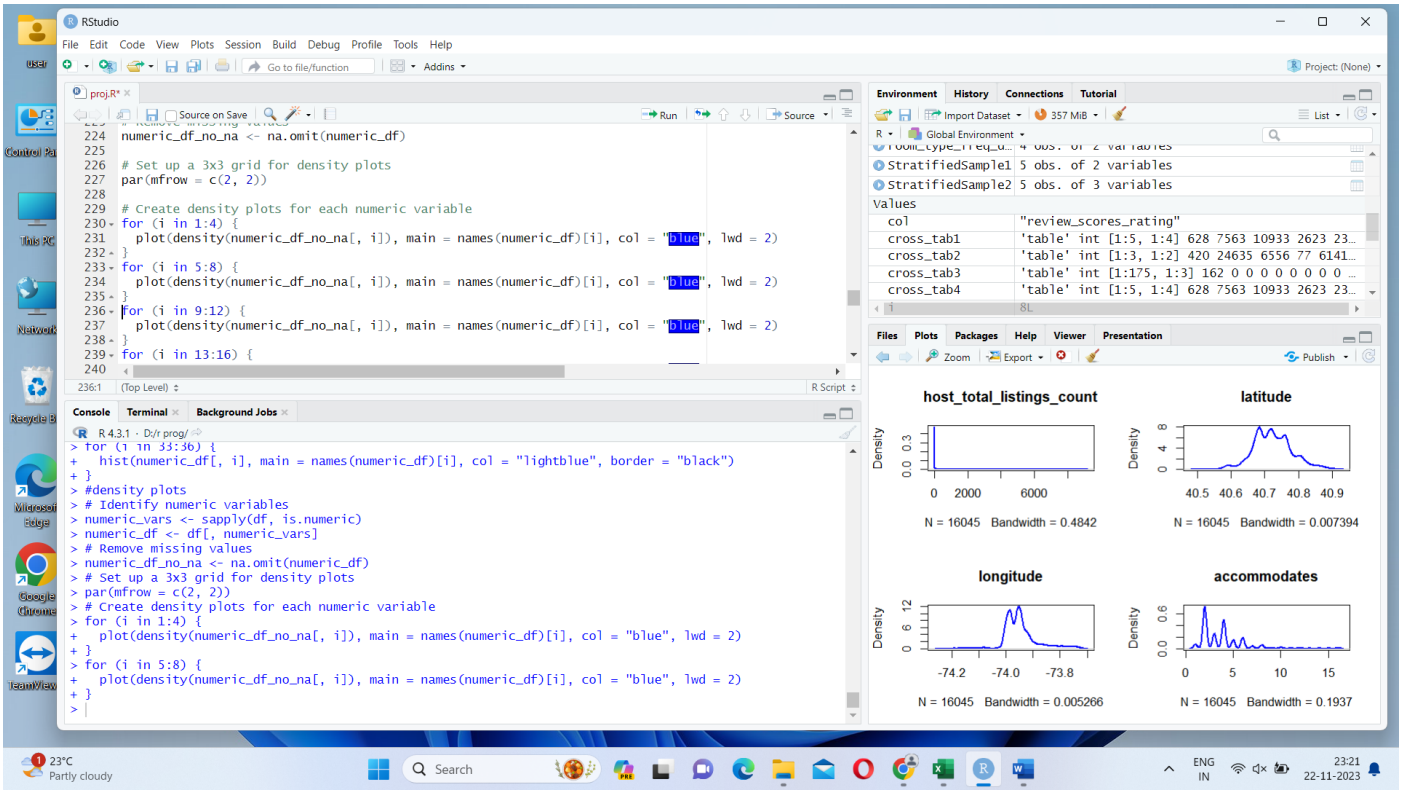




vii) Host id and id show right skewed while scrape id is normally distributed



viii) Longitude is bimodal(2 peaks) and latitude is almost normal with 3 peaks near the highest point of the curve



the density plots too show the same inference screenshots are given above from page number 26 to 30

6)There are no duplicates (rows/columns)seen in this dataset