Programming with R – Experiential Learning Component

NAME: Chinmayi R M SRN:PES1UG22AM051

SECTION: AIML A (SEM-3)

CODE:

```
library(dplyr)
library(ggplot2)
file_path <- "C:\\Users\\chinm\\Downloads\\listings.csv" # Replace with the actual path to your CSV
file
data <- read.csv(file path)
head(data)
tail(data)
print(filter(data, is.na(data$price)))
# summary statistics
summary(data)
# checking for missing values
sum(is.na(data))
# checking for missing values in ratings and price
sum(is.na(data$review_scores_rating))
sum(is.na(data$price))
head(data$price)
# removing the "$" sign and converting to numeric
data$price <- as.numeric(gsub("\\$", "", data$price))</pre>
mean(data$price)
```

```
# removing rows with missing values in ratings and price
data <- subset(data, !(is.na(data$review scores rating) & is.na(data$price)))
# outlier detection for price which are less than Q1 – 1.5 IQR and above Q3 + 1.5 IQR
s_price <- summary(data$price)</pre>
data <- filter(data, data$price >= (s_price[2] - 1.5 * IQR(data$price, na.rm = TRUE)) &
         data$price <= (s_price[5] + 1.5 * IQR(data$price, na.rm = TRUE)))
# outlier detection for review_scores_rating which are below 0 and above 5
data <- filter(data, data$review_scores_rating >= 0 & data$review_scores_rating <= 5)
# visualization of review scores
ggplot(data, aes(x = review_scores_rating)) +
 geom_histogram(bins = 20, fill = "pink", color = "black") +
 labs(title = "Distribution of Review Scores", x = "Review Scores", y = "Frequency")
# calculating avg price by neighbourhood group
avg_price_ng <- data %>%
 group_by(neighbourhood_group_cleansed) %>%
 summarise(avg_price = mean(price, na.rm = TRUE))%>%
arrange(desc(avg price))
avg_price_ng
# we see that Bronx has the lowest average price and Manhattan has the highest average price
# finding the most common property types
most_common_property_type <- data %>%
 group by(property type) %>%
 summarise(listing_count = n()) %>%
 arrange(desc(listing_count)) %>%
```

```
head(5)
```

```
print(most_common_property_type)
```

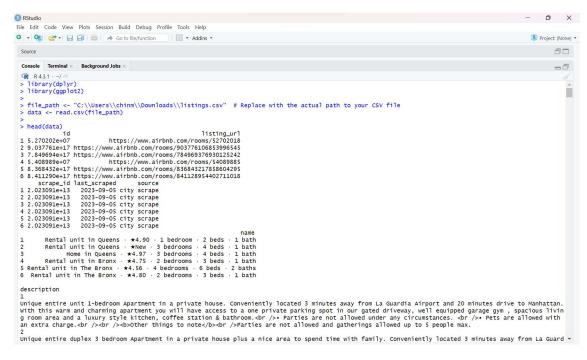
we see that the most common property type is an entire rental unit

```
# finding the neighbourhoods with most number of listings
pop_neighborhoods <- data %>%
group_by(neighbourhood_cleansed) %>%
summarise(listing_count = n()) %>%
arrange(desc(listing_count)) %>%
head(5)
```

print(pop_neighborhoods)

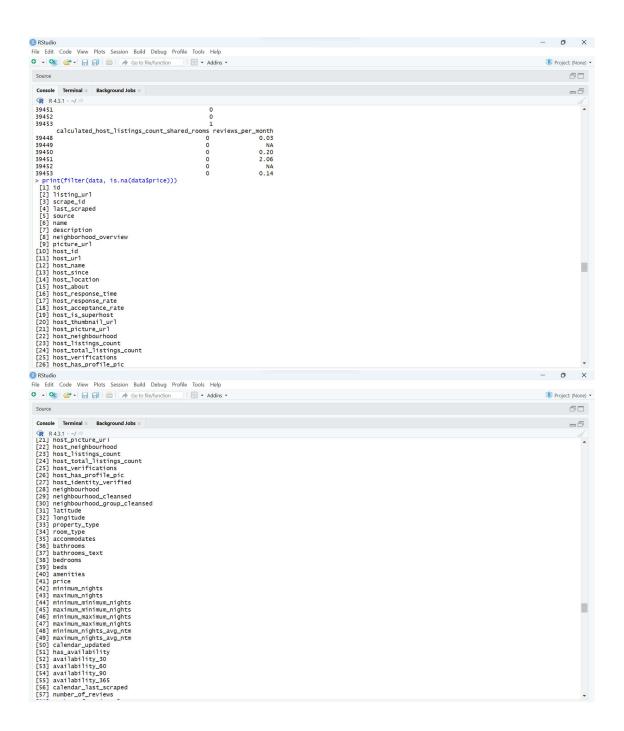
we see that Bedforf-Stuyvesant has the most number of listings

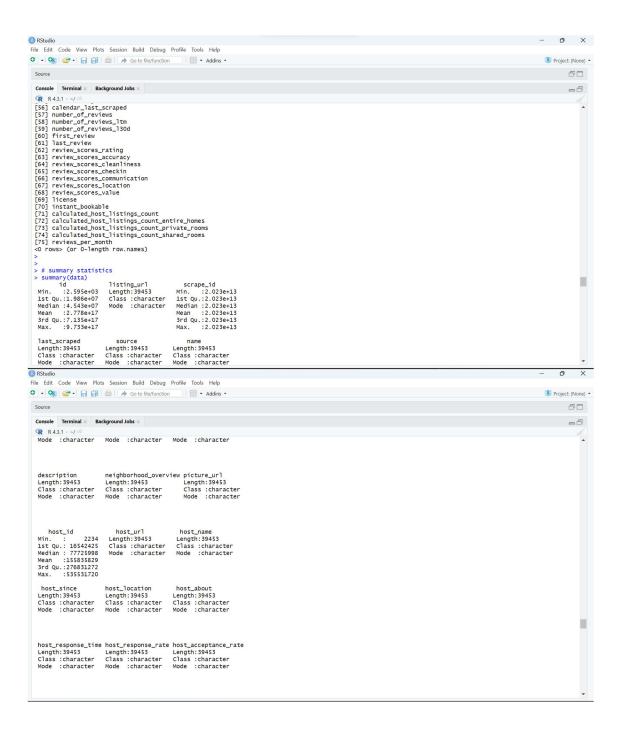
OUTPUT:

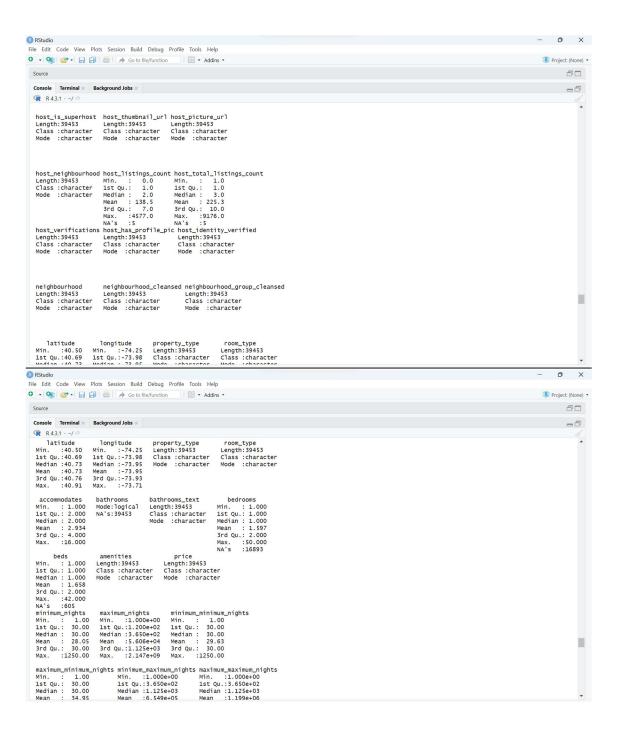


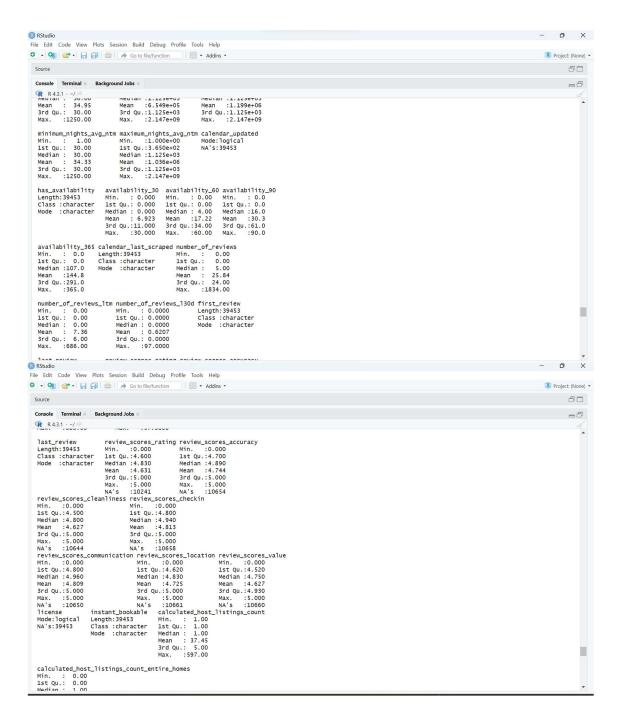
..... remaining of the first six rows of each column.....

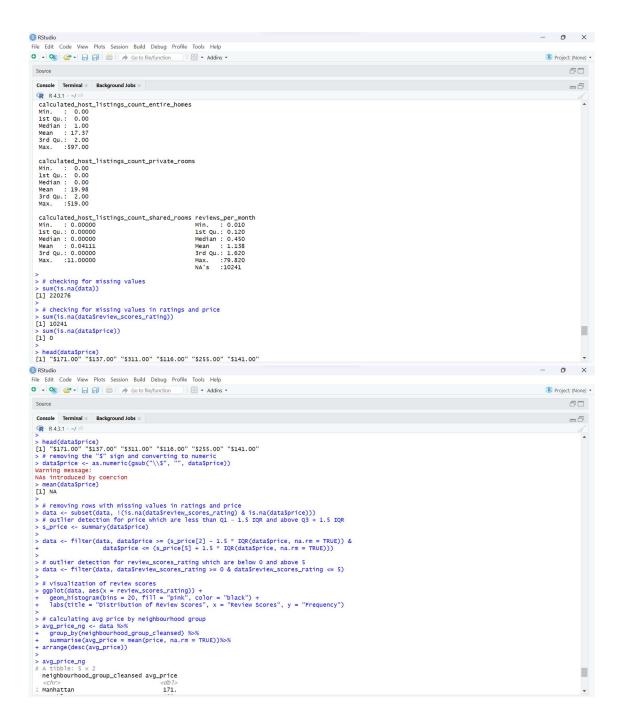
..... remaining of the last six rows of each column.....









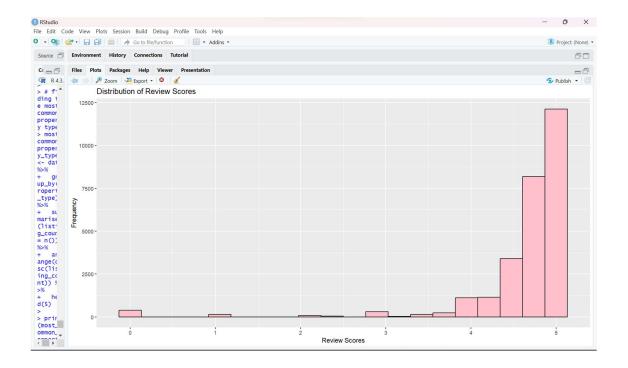


```
File Edit Code View Plots Session Build Debug Profile Tools Help

■ Project: (None) ▼

O - O Go to file/function S - Addins -
                                                                                                                                                                                                                                                             80
  Console Terminal × Background Jobs ×
  R 4.3.1 · ~/ >
neighbourhood_group_cleansed avg_price
     <chr>
Manhattan
     Brooklyn
Staten Island
                                                                  124.
119.
  > Bronx 111.
> # we see that Bronx has the lowest avg price and Manhattan has the highest avg price
  > # finding the most common property types
> most_common_property_type <- data %>%
+ group_by(property_type) %>%
+ summarise(listing_count = n()) %>%
+ arrange(desc(listing_count)) %>%
+ head(s)
    <<hr/>

  > # finding the neighbourhoods with most number of listings > pop_neighborhoods <- data %>% group_by(neighbourhood_cleansed) %>% summarise(listing_count = n()) %>% arrange(desc(listing_count)) %>% head($)
  > print(pop_neighborhoods)
# A tibble: 5 × 2
neighbourhood_cleansed listing_count
                                                                                                                                                                                                                                                                                   П
                                                                                                                                                                                                                                                                        O
File Edit Code View Plots Session Build Debug Profile Tools Help
○ • 👒 💣 • 🔒 🧁 🍙 🍙 Addins •
                                                                                                                                                                                                                                                        80
  Console Terminal × Background Jobs ×
  R 4.3.1 · ~/ ≈
  ># finding the most common property types
> most_common_property_type <- data %>%
    group_b/(property_type) %>%
+ summarise(listing_count = n()) %>%
+ summarise(listing_count)) %>%
+ head(S)
  > print(most_common_property_type)
# A tibhle: 5 > 2
    A tibble: 5 x 2
property_type
                                   listing_count
    cohire rental unit 11809
Private room in rental unit 2420
Private room in home 1923
Entire home 1170
Entire condo 901
# we see that the most common property type is an entire rental unit
    # finding the neighbourhoods with most number of listings
  > pop_neighborhoods <- data %%
+ group_by(neighbourhood_cleansed) %>%
+ summarise(listing_count = n()) %>%
+ arrange(desc(listing_count)) %>%
+ head(5)
  > print(pop_neighborhoods)
    A tibble: 5 × 2
neighbourhood_cleansed listing_count
     Bedford-Stuyvesant 2076
                                                        1730
1384
1196
     Williamsburg
     Harlem
Bushwick
     Upper East Side
        pper East Side 948
we see that Bedforf-Stuyvesant has the most number of listings
```



Three interesting pieces of information from this analysis:

- 1) Bronx has the lowest average price and Manhattan has the highest average price
- 2) The most common property type is an entire rental unit
- 3) Bedford-Stuyvesant has the most number of listings