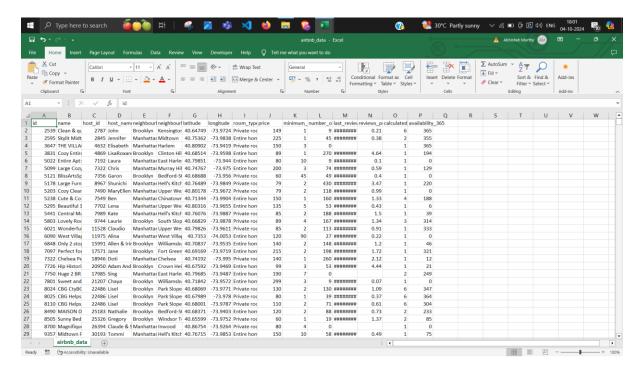
Assignment 9

Abhishek Murthy 2024-10-04

Dataset:



Code:

```
return(kmeans_result)
}
kmeans_clustering <- function() {
 file_path <- 'C:\\Users\\91984\\OneDrive\\Desktop\\VIT\\Sem7\\Exploratory Data Analysis
Lab\\airbnb data.csv'
 data <- read.csv(file path)
 k <- as.integer(readline(prompt = "Enter the number of clusters (k): "))
 clean_data_set <- clean_data(data)</pre>
 if (nrow(clean_data_set) == 0) {
  stop("No valid numeric data found after cleaning.")
 kmeans result <- perform kmeans(clean data set, k)
 data$Cluster <- as.factor(kmeans result$cluster)
 if (ncol(clean_data_set) == 2) {
  ggplot(data, aes(x = clean_data_set[, 1], y = clean_data_set[, 2], color = Cluster)) +
   geom point(size = 3) +
   labs(title = "K-Means Clustering") +
   theme_minimal()
 } else {
  print("Data has more than 2 dimensions. Cannot plot the clusters.")
 }
kmeans_clustering()
```

Output:

```
> kmeans_clustering()
Enter the number of clusters (k): 3
[1] "Cluster centers:"
         id host_id latitude longitude
                                                    price minimum_nights number_of_reviews reviews_per_month
1 22442789 111987881 40.72845 -73.94067 132.2835
2 30651742 220430165 40.72870 -73.94224 154.0320
3 14273415 18424352 40.72793 -73.95590 142.6407
                                                                   5.019526
                                                                                         28.53801
                                                                                                               1.670680
                                                                                         13.81647
                                                                   4.138884
                                                                                                                2.392406
                                                                   6.463585
                                                                                         32.74171
                                                                                                               1.078165
  calculated_host_listings_count availability_365
                             4.274754
                                                  121.5267
                            15.872095
                                                  161.7843
                             3.172917
                                                  103.2496
[1] "Cluster sizes:"
 7221 5465 26157
```