

Assignment 9

Abhishek Murthy

2024-10-04

Dataset:

Type here to search

airbnb_data - Excel

Abhishek Murthy

30°C Partly sunny

1801

04-10-2024

FileHomeInsertPage LayoutFormulasDataReviewViewDeveloperHelp

Tell me what you want to do

CutCopyFormat Painter

Clipboard

Calibri11

B

I

U

A

Font

Wrap Text

Merge & Center

Alignment

General

Number

Conditional Formatting

Format as Table

Cell Styles

Insert

Delete Format

Cells

AutoSum

Fill

Clear

Editing

Sort & Filter

Find & Select

Add-ins

Add-ins

A1

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

id

name

host_id

host_name

neighbourhood

longitude

room_type

price

minimum_number_of_reviews

last_review

reviews_per_calculated_availability_365

1	id	name	host_id	host_name	neighbourhood	longitude	room_type	price	minimum_number_of_reviews	last_review	reviews_per_calculated_availability_365												
2	2539	Clean & qu	2787	John	Brooklyn	Kensington	40.64749	-73.9724	Private roc	149	1	9	#####	0.21	6	365							
3	2595	Skylit Midt	2845	Jennifer	Manhattan	Midtown	40.75362	-73.9838	Entire hom	225	1	45	#####	0.38	2	355							
4	3647	THE VILLA	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.9419	Private roc	150	3	0			1	365							
5	3831	Cozy Entir	4869	LisaRoxan	Brooklyn	Clinton Hill	40.68514	-73.9598	Entire hom	89	1	270	#####	4.64	1	194							
6	5022	Entire Apt:	7192	Laura	Manhattan	East Harle	40.79851	-73.944	Entire hom	80	10	9	#####	0.1	1	0							
7	5099	Large Cozy	7322	Chris	Manhattan	Murray Hill	40.74767	-73.975	Entire hom	200	3	74	#####	0.59	1	129							
8	5121	BlissArtsSp	7356	Garon	Brooklyn	Bedford-Si	40.68688	-73.956	Private roc	60	45	49	#####	0.4	1	0							
9	5178	Large Furn	8967	Shunichi	Manhattan	Hell's Kitc	40.76489	-73.9849	Private roc	79	2	430	#####	3.47	1	220							
10	5203	Cozy Clear	7490	MaryEllen	Manhattan	Upper We	40.80178	-73.9672	Private roc	79	2	118	#####	0.99	1	0							
11	5238	Cute & Co.	7549	Ben	Manhattan	Chinatown	40.71344	-73.9904	Entire hom	150	1	160	#####	1.33	4	188							
12	5295	Beautiful 1	7702	Lena	Manhattan	Upper We	40.80316	-73.9655	Entire hom	135	5	53	#####	0.43	1	6							
13	5441	Central Mi	7989	Kate	Manhattan	Hell's Kitc	40.76076	-73.9887	Private roc	85	2	188	#####	1.5	1	39							
14	5803	Lovely Roc	9744	Laurie	Brooklyn	South Slop	40.66829	-73.9878	Private roc	89	4	167	#####	1.34	3	314							
15	6021	Wonderful	11528	Claudio	Manhattan	Upper We	40.79826	-73.9611	Private roc	85	2	113	#####	0.91	1	333							
16	6090	West Villaj	11975	Alina	Manhattan	West Villaj	40.7353	-74.0053	Entire hom	120	90	27	#####	0.22	1	0							
17	6848	Only 2 sto	15991	Allen & Iri	Brooklyn	Williamsb	40.70837	-73.9535	Entire hom	140	2	148	#####	1.2	1	46							
18	7097	Perfect for	17571	Jane	Brooklyn	Fort Green	40.69169	-73.9719	Entire hom	215	2	198	#####	1.72	1	321							
19	7322	Chelsea Pe	18946	Doti	Manhattan	Chelsea	40.74192	-73.995	Private roc	140	1	260	#####	2.12	1	12							
20	7726	Hip Histori	20950	Adam And	Brooklyn	Crown Hei	40.67592	-73.9469	Entire hom	99	3	53	#####	4.44	1	21							
21	7750	Huge 2 BR	17985	Sing	Manhattan	East Harle	40.79685	-73.9487	Entire hom	190	7	0			2	249							
22	7801	Sweet and	21207	Chaya	Brooklyn	Williamsb	40.71842	-73.9572	Entire hom	299	3	9	#####	0.07	1	0							
23	8024	CBG CtyBC	22486	Lisel	Brooklyn	Park Slope	40.68069	-73.9771	Private roc	130	2	130	#####	1.09	6	347							
24	8025	CBG Helps	22486	Lisel	Brooklyn	Park Slope	40.67989	-73.978	Private roc	80	1	39	#####	0.37	6	364							
25	8110	CBG Helps	22486	Lisel	Brooklyn	Park Slope	40.68001	-73.9787	Private roc	110	2	71	#####	0.61	6	304							
26	8490	MAISON D	25183	Nathalie	Brooklyn	Bedford-Si	40.68371	-73.9403	Entire hom	120	2	88	#####	0.73	2	233							
27	8505	Sunny Bed	25326	Gregory	Brooklyn	Windsor T	40.65599	-73.9752	Private roc	60	1	19	#####	1.37	2	85							
28	8700	Magnifiqu	26394	Claude & S	Manhattan	Inwood	40.86754	-73.9264	Private roc	80	4	0			1	0							
29	9357	Midtown F	30193	Tommi	Manhattan	Hell's Kitc	40.76715	-73.9853	Entire hom	150	10	58	#####	0.49	1	75							

airbnb_data

Accessibility: Unavailable

Code:

```
library(ggplot2)

clean_data <- function(data) {
  numeric_data <- data[sapply(data, is.numeric)]
  clean_data <- numeric_data[complete.cases(numeric_data) &
    !apply(numeric_data, 1, function(x) any(is.infinite(x))), ]

  return(clean_data)
}

perform_kmeans <- function(data, k) {
  kmeans_result <- kmeans(data, centers = k)
  print("Cluster centers:")
  print(kmeans_result$centers)
  print("Cluster sizes:")
  print(table(kmeans_result$cluster))
}
```

```

    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)

  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      5.019526      28.53801      1.670680
2 30651742 220430165 40.72870 -73.94224 154.0320      4.138884      13.81647      2.392406
3 14273415 18424352 40.72793 -73.95590 142.6407      6.463585      32.74171      1.078165
      calculated_host_listings_count availability_365
1              4.274754              121.5267
2              15.872095              161.7843
3              3.172917              103.2496
[1] "Cluster sizes:"
      1      2      3
7221 5465 26157

```
