Output

1. library(psych)

describe(Cities.df)

Output :

vars n mean sd median trimmed mad

CityName\* 1 13232 NaN NA NA NaN NA

Population 2 13232 4416836.87 4258386.00 3046163.0 4040816.22 3846498.95

CityRank 3 13232 14.83 13.51 9.0 13.30 11.86

IsMetroCity 4 13232 0.28 0.45 0.0 0.23 0.00

IsTouristDestination 5 13232 0.70 0.46 1.0 0.75 0.00

IsWeekend 6 13232 0.62 0.48 1.0 0.65 0.00

IsNewYearEve 7 13232 0.12 0.33 0.0 0.03 0.00

Date\* 8 13232 NaN NA NA NaN NA

HotelName\* 9 13232 NaN NA NA NaN NA

RoomRent 10 13232 5473.99 7333.12 4000.0 4383.33 2653.85

StarRating 11 13232 3.46 0.76 3.0 3.40 0.74

Airport 12 13232 21.16 22.76 15.0 16.39 11.12

HotelAddress\* 13 13232 NaN NA NA NaN NA

HotelPincode 14 13232 397430.26 259837.50 395003.0 388540.47 257975.37

HotelDescription\* 15 13224 3.46 0.48 3.5 3.41 0.74

FreeWifi 16 13232 0.93 0.26 1.0 1.00 0.00

FreeBreakfast 17 13232 0.65 0.48 1.0 0.69 0.00

HotelCapacity 18 13232 62.51 76.66 34.0 46.03 28.17

HasSwimmingPool 19 13232 0.36 0.48 0.0 0.32 0.00

min max range skew kurtosis se

CityName\* Inf -Inf -Inf NA NA NA

Population 8096.0 12442373 12434277.0 0.68 -1.08 37019.65

CityRank 0.0 44 44.0 0.69 -0.76 0.12

IsMetroCity 0.0 1 1.0 0.96 -1.08 0.00

IsTouristDestination 0.0 1 1.0 -0.86 -1.26 0.00

IsWeekend 0.0 1 1.0 -0.51 -1.74 0.00

IsNewYearEve 0.0 1 1.0 2.28 3.18 0.00

Date\* Inf -Inf -Inf NA NA NA

HotelName\* Inf -Inf -Inf NA NA NA

RoomRent 299.0 322500 322201.0 16.75 582.06 63.75

StarRating 0.0 5 5.0 0.48 0.25 0.01

Airport 0.2 124 123.8 2.73 7.89 0.20

HotelAddress\* Inf -Inf -Inf NA NA NA

HotelPincode 100025.0 7000157 6900132.0 9.99 249.76 2258.86

HotelDescription\* 3.0 5 2.0 0.93 0.78 0.00

FreeWifi 0.0 1 1.0 -3.25 8.57 0.00

FreeBreakfast 0.0 1 1.0 -0.62 -1.61 0.00

HotelCapacity 0.0 600 600.0 2.95 11.39 0.67

HasSwimmingPool 0.0 1 1.0 0.60 -1.64 0.00

2) hoteltable <- table(HotelName)

hoteltable

3) citytable <- table(CityName, IsTouristDestination)

citytable

Output

CityName 0 1

Agra 0 432

Ahmedabad 424 0

Amritsar 0 136

Bangalore 656 0

Bhubaneswar 120 0

Chandigarh 336 0

Chennai 328 88

Darjeeling 0 136

Delhi 0 2048

Gangtok 0 128

Goa 0 624

Guwahati 0 48

Haridwar 0 48

Hyderabad 536 0

Indore 160 0

Jaipur 0 768

Jaisalmer 0 264

Jodhpur 0 224

Kanpur 16 0

Kochi 0 608

Kolkata 327 185

Lucknow 128 0

Madurai 0 112

Manali 0 288

Mangalore 104 0

Mumbai 0 712

Munnar 0 328

Mysore 0 160

Nainital 0 144

Ooty 0 136

Panchkula 64 0

Pune 600 0

Puri 0 56

Rajkot 128 0

Rishikesh 0 88

Shimla 0 280

Srinagar 0 40

Surat 80 0

Thiruvanthipuram 0 392

Thrissur 0 32

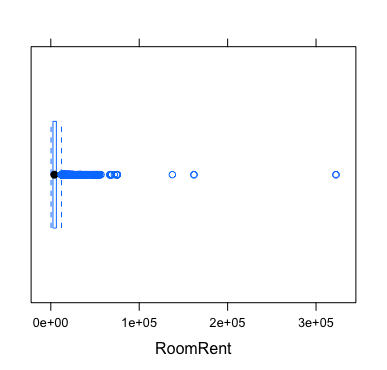
Udaipur 0 456

Varanasi 0 264

3) library(lattice)

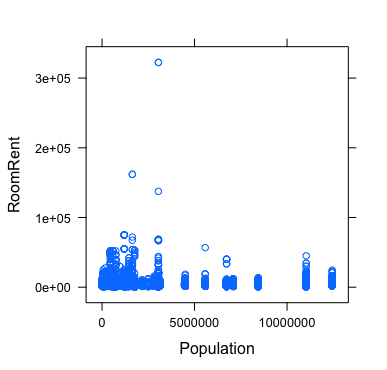
bwplot(RoomRent)

Output



4) xyplot(RoomRent~ Population)

Output



5) t.test(RoomRent, StarRating)

Output

data: RoomRent and StarRating

t = 85.813, df = 13231, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

5345.575 5595.491

sample estimates:

mean of x mean of y

5473.991838 3.458933

6) t.test(RoomRent~IsTouristDestination)

Output

data: RoomRent by IsTouristDestination

t = -19.449, df = 12888, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2152.059 -1757.983

sample estimates:

mean in group 0 mean in group 1

4111.003 6066.024

7) t.test(RoomRent~IsMetroCity)

Output

data: RoomRent by IsMetroCity

t = 10.721, df = 13224, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

888.0308 1285.4102

sample estimates:

mean in group 0 mean in group 1

5782.794 4696.073

8) t.test(RoomRent~IsWeekend)

Ouput

data: RoomRent by IsWeekend

t = -0.51853, df = 9999.4, p-value = 0.6041

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-331.2427 192.6559

sample estimates:

mean in group 0 mean in group 1

5430.835 5500.129

9) t.test(RoomRent~FreeBreakfast)

Output

data: RoomRent by FreeBreakfast

t = 0.98095, df = 6212.3, p-value = 0.3267

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-153.5017 460.9935

sample estimates:

mean in group 0 mean in group 1

5573.790 5420.044

10) t.test(RoomRent, Airport)

Output

data: RoomRent and Airport

t = 85.535, df = 13231, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

5327.875 5577.792

sample estimates:

mean of x mean of y

5473.99184 21.15874

11) t.test(RoomRent~FreeWifi)

Output

data: RoomRent by FreeWifi

t = -0.76847, df = 1804.7, p-value = 0.4423

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-360.5977 157.5701

sample estimates:

mean in group 0 mean in group 1

5380.004 5481.518

12) t.test(RoomRent, CityRank)

Output

data: RoomRent and CityRank

t = 85.635, df = 13231, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

5334.200 5584.116

sample estimates:

mean of x mean of y

5473.99184 14.83374

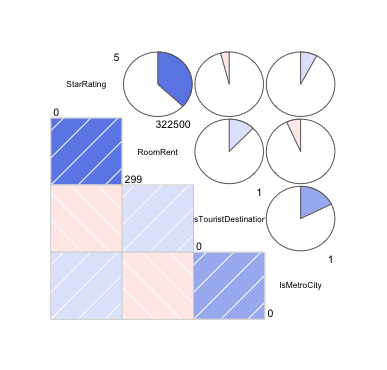
13) library(corrgram)

hotel <- c("RoomRent","StarRating","IsTouristDestination","IsMetroCity")

corrgram(Cities.df[,hotel], order=TRUE,lower.panel=panel.shade, upper.panel=panel.pie,

diag.panel=panel.minmax, text.panel=panel.txt)

Output



14) library(Hmisc)

hotel <- c("RoomRent","StarRating","IsTouristDestination","IsMetroCity")

hotelMatrix2 <- rcorr(as.matrix(Cities.df[,hotel]))

hotelMatrix2

Output

RoomRent StarRating IsTouristDestination IsMetroCity

RoomRent 1.00 0.37 0.12 -0.07

StarRating 0.37 1.00 -0.04 0.08

IsTouristDestination 0.12 -0.04 1.00 0.18

IsMetroCity -0.07 0.08 0.18 1.00

n= 13232

P

RoomRent StarRating IsTouristDestination IsMetroCity

RoomRent 0 0 0

StarRating 0 0 0

IsTouristDestination 0 0 0

IsMetroCity 0 0 0

15) test\_model<-log(RoomRent)~IsMetroCity+Date+StarRating+IsTouristDestination+IsWeekend+IsNewYearEve+HasSwimmingPool+FreeBreakfast+FreeWifi+HotelName

fit\_1<-lm(test\_model, data=Cities.df)

summary(fit\_1)

Output

Residual standard error: 0.2537 on 11538 degrees of freedom

Multiple R-squared: 0.89, Adjusted R-squared: 0.8738

F-statistic: 55.13 on 1693 and 11538 DF, p-value: < 2.2e-1