**R CODE**

#Setting home directory

setwd("~/Data Analytics/workingdirectory/hotel")

#Uploading the CSV

hotel.df <- read.csv(paste("Cities42.csv",sep = ""))

#Viewing the dataframe

View(hotel.df)

#Finding mean, variance and standard deviation of all features in dataframe

summary(hotel.df)

library(psych)

library(car)

#Run regression test RoomRent vs Others

fithotel <- lm(hotel.df$RoomRent~hotel.df$X+hotel.df$CityName+hotel.df$Population+hotel.df$CityRank+hotel.df$IsMetroCity+hotel.df$IsTouristDestination+hotel.df$IsWeekend+hotel.df$IsNewYearEve+hotel.df$Date)

fithotel2 <- lm(hotel.df$RoomRent~hotel.df$StarRating+hotel.df$Airport+hotel.df$HotelPincode+hotel.df$FreeWifi+hotel.df$FreeBreakfast+hotel.df$HotelCapacity+hotel.df$HasSwimmingPool)

fithotel3 <- lm(hotel.df$RoomRent~hotel.df$HotelName)

coef(fithotel)

coef(fithotel2)

coef(fithotel3)

#Let Y = RoomRent

#Let X1 = Date

#Let X2 = Population

#Let X3 = CityName

#Plotting graph for continous independent variables

boxplot(hotel.df$RoomRent~hotel.df$Population,data=hotel.df, main="RoomRent vs Population",

xlab="Population", ylab="Room Rent")

#Drawing table for discontinous independent variables

tableDate <- xtabs(hotel.df$RoomRent~hotel.df$Date)

View(tableDate)

tableCity <- xtabs(hotel.df$RoomRent~hotel.df$CityName)

View(tableCity)

#Scatter Plots for RoomRent Vs Independent Variables

plot(hotel.df$RoomRent, hotel.df$Population, main="Room Rent Vs Population",

xlab="Room Rent ", ylab="Population")

plot(hotel.df$RoomRent, hotel.df$CityName, main="Room Rent vs City Name ",

xlab="Room Rent ", ylab="City Name")

plot(hotel.df$RoomRent, hotel.df$Date, main="Room Rent vs Date",

xlab="Room Rent ", ylab="Date")

#Creating new dataframe with only RoomRent and independent variable

hotelnew.df <- data.frame(hotel.df$RoomRent ,hotel.df$CityName, hotel.df$Population, hotel.df$Date)

View(hotelnew.df)

#Converting All new features into numeric

RoomRent = as.numeric(hotelnew.df$hotel.df.RoomRent)

Population = as.numeric(hotelnew.df$hotel.df.Population)

Date = as.numeric(hotelnew.df$hotel.df.Date)

CityName = as.numeric(hotelnew.df$hotel.df.CityName)

#Binding them to a table

M <- cbind(RoomRent,Population,Date,CityName)

#Drawing corrgram for all the new features of Dataset

library(corrgram)

corrgram(M, order=TRUE, lower.panel=panel.shade,

upper.panel=panel.pie, text.panel=panel.txt,

main="Hotel Pricing")

#Drawing the Variance - Covariance Matrix for Y, X1, X2, X3

cor(M)

#Hypothesis Testing Between RoomRent and Population

t.test(RoomRent,Population)

#Hypothesis Testing Between RoomRent and Date

t.test(RoomRent,Date)

#Hypothesis Testing Between RoomRent and CityName

t.test(RoomRent,CityName)

#Fitting Of Linear Regeression Model

fitFinal <- lm(RoomRent~CityName+Date+Population)

coef(fitFinal)