

CAPSTONE PROJECT R CODE

```
# CAPSTONE PROJECT
# Analysis of Hotel Room Pricing In The Indian Market
# Name: Shrey Shailesh Shah
# E-Mail: shreyss99@gmail.com
# College: Sardar Patel Institute of Technology
```

1a. Setting up current working directory

```
# =====

setwd("D:/Master/data analytics")
```

1b. Reading the data in cities.df dataframe

```
# =====

cities.df <- read.csv(paste("Cities42.csv"))
```

1c. Viewing data

```
# =====

library("utils", lib.loc="C:/Program Files/R/R-3.3.3/library")

View(cities.df)
```

2a. Summarising the data

```
# =====

View(summary(cities.df))
```

2b. Describing data

```
# =====

View(describe(cities.df))
```

3. Problem as $Y = F(x_1, x_2, x_3..)$

4. Dependent variable of $Y = F(x_1, x_2, x_3..)$

=====

Y is Hotel Room Rent

5. Independent variables of $Y = F(x_1, x_2, x_3..)$

=====

x1 is StarRating, x2 is HotelCapacity, x3 is HasSwimmingPool

6. Visualization of Independent variables

=====

For x1 is StarRating

boxplot(cities.df\$StarRating, horizontal = 1, xlab="Star Rating", ylab="Frequency", main="Star Rating of different hotels in different countries")

hist(cities.df\$StarRating, xlab="Star Rating", ylab="Count", main="Star Rating of different hotels in different countries")

For x2 is Hotel Capacity

boxplot(cities.df\$HotelCapacity, horizontal=1, xlab="Capacity", ylab="Frequency", main="Boxplot of Hotel Capacity")

hist(cities.df\$HotelCapacity, xlab="Capacity", ylab="Frequency", main="Hotel Capacity")

```
# For x2 is Swimming Pool
```

```
table <- xtabs(~HasSwimmingPool, data=cities.df)
```

```
table
```

```
# For y is RoomRent
```

```
boxplot(cities.df$RoomRent, horizontal = 1, xlab="Roomrent", ylab="Frequency",  
main="Boxplot of Room Rent of Hotels")
```

7. Scatterplots

```
# =====
```

```
# library("car", lib.loc="~/R/win-library/3.3")
```

```
# Scatterplot Matrix
```

```
scatterplotMatrix(~RoomRent+StarRating+HotelCapacity+HasSwimmingPool, data=cities.df)
```

```
# Scatterplot of Room rent against Hotel capacity
```

```
plot(~RoomRent+HotelCapacity, xlab="Rent of hotel rooms",main="Room rent against Hotel  
Capacity", data=cities.df, log=("xy"))
```

```
# Scatterplot of Room rent against Star Rating of hotels
```

```
plot(~RoomRent+StarRating, xlab="Rent of hotel rooms",main="Room rent against Star  
Rating", data=cities.df, log=("xy"))
```

```
# Scatterplot of Room Rent against is a tourist destination
```

```
plot(~RoomRent+HasSwimmingPool, xlab="Rent of hotel rooms",main="Room rent against  
hotel has a swimming pool", data=cities.df, log=("xy"))
```

8. Corrgram of x1, x2, x3

=====

```
corrgram(cities.df[,new], order=TRUE, lower.panel=panel.pys, upper.panel=panel.pie,  
diag.panel = panel.minmax, text.panel=panel.txt)
```

```
corrgram(cities.df[,new], order=TRUE, lower.panel=panel.pts, upper.panel=panel.pie, diag.panel  
= panel.minmax, text.panel=panel.txt)
```

9. Variance - Covariance Matrix

=====

```
hotel.df <- cities.df[,c(6,11,12,13)] # selecting the chosen variables column
```

```
cov(hotel.df, y=NULL, use="everything", method=c("pearson", "kendall", "spearman")) # using  
the cov function to generate covariance matrix
```

```
cor(hotel.df, y=NULL, use="everything", method=c("pearson", "kendall", "spearman")) # using  
the cor function to generate correlation matrix
```