**EDA ASSIGNMENT**

**Exploratory Data Analysis (EDA)**is an approach to analyse the data using visual techniques. It is used to discover trends, patterns, or to check assumptions with the help of statistical summary and graphical representations.

*#First the dataset is imported into R studio*

hotel\_bookings<-read.csv("C:/Users/Owner new/Downloads/hotel\_bookings - hotel\_bookings.csv")  
View(hotel\_bookings)  
dim(hotel\_bookings)

## [1] 119390 32

str(hotel\_bookings)

## 'data.frame': 119390 obs. of 32 variables:  
## $ hotel : chr "Resort Hotel" "Resort Hotel" "Resort Hotel" "Resort Hotel" ...  
## $ is\_canceled : int 0 0 0 0 0 0 0 0 1 1 ...  
## $ lead\_time : int 342 737 7 13 14 14 0 9 85 75 ...  
## $ arrival\_date\_year : int 2015 2015 2015 2015 2015 2015 2015 2015 2015 2015 ...  
## $ arrival\_date\_month : chr "July" "July" "July" "July" ...  
## $ arrival\_date\_week\_number : int 27 27 27 27 27 27 27 27 27 27 ...  
## $ arrival\_date\_day\_of\_month : int 1 1 1 1 1 1 1 1 1 1 ...  
## $ stays\_in\_weekend\_nights : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ stays\_in\_week\_nights : int 0 0 1 1 2 2 2 2 3 3 ...  
## $ adults : int 2 2 1 1 2 2 2 2 2 2 ...  
## $ children : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ babies : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ meal : chr "BB" "BB" "BB" "BB" ...  
## $ country : chr "PRT" "PRT" "GBR" "GBR" ...  
## $ market\_segment : chr "Direct" "Direct" "Direct" "Corporate" ...  
## $ distribution\_channel : chr "Direct" "Direct" "Direct" "Corporate" ...  
## $ is\_repeated\_guest : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ previous\_cancellations : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ previous\_bookings\_not\_canceled: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ reserved\_room\_type : chr "C" "C" "A" "A" ...  
## $ assigned\_room\_type : chr "C" "C" "C" "A" ...  
## $ booking\_changes : int 3 4 0 0 0 0 0 0 0 0 ...  
## $ deposit\_type : chr "No Deposit" "No Deposit" "No Deposit" "No Deposit" ...  
## $ agent : chr "NULL" "NULL" "NULL" "304" ...  
## $ company : chr "NULL" "NULL" "NULL" "NULL" ...  
## $ days\_in\_waiting\_list : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ customer\_type : chr "Transient" "Transient" "Transient" "Transient" ...  
## $ adr : num 0 0 75 75 98 ...  
## $ required\_car\_parking\_spaces : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ total\_of\_special\_requests : int 0 0 0 0 1 1 0 1 1 0 ...  
## $ reservation\_status : chr "Check-Out" "Check-Out" "Check-Out" "Check-Out" ...  
## $ reservation\_status\_date : chr "2015-07-01" "2015-07-01" "2015-07-02" "2015-07-02" ...

names(hotel\_bookings)

## [1] "hotel" "is\_canceled"   
## [3] "lead\_time" "arrival\_date\_year"   
## [5] "arrival\_date\_month" "arrival\_date\_week\_number"   
## [7] "arrival\_date\_day\_of\_month" "stays\_in\_weekend\_nights"   
## [9] "stays\_in\_week\_nights" "adults"   
## [11] "children" "babies"   
## [13] "meal" "country"   
## [15] "market\_segment" "distribution\_channel"   
## [17] "is\_repeated\_guest" "previous\_cancellations"   
## [19] "previous\_bookings\_not\_canceled" "reserved\_room\_type"   
## [21] "assigned\_room\_type" "booking\_changes"   
## [23] "deposit\_type" "agent"   
## [25] "company" "days\_in\_waiting\_list"   
## [27] "customer\_type" "adr"   
## [29] "required\_car\_parking\_spaces" "total\_of\_special\_requests"   
## [31] "reservation\_status" "reservation\_status\_date"

attributes(hotel\_bookings)

## $names  
## [1] "hotel" "is\_canceled"   
## [3] "lead\_time" "arrival\_date\_year"   
## [5] "arrival\_date\_month" "arrival\_date\_week\_number"   
## [7] "arrival\_date\_day\_of\_month" "stays\_in\_weekend\_nights"   
## [9] "stays\_in\_week\_nights" "adults"   
## [11] "children" "babies"   
## [13] "meal" "country"   
## [15] "market\_segment" "distribution\_channel"   
## [17] "is\_repeated\_guest" "previous\_cancellations"   
## [19] "previous\_bookings\_not\_canceled" "reserved\_room\_type"   
## [21] "assigned\_room\_type" "booking\_changes"   
## [23] "deposit\_type" "agent"   
## [25] "company" "days\_in\_waiting\_list"   
## [27] "customer\_type" "adr"   
## [29] "required\_car\_parking\_spaces" "total\_of\_special\_requests"   
## [31] "reservation\_status" "reservation\_status\_date"   
##   
## $class  
## [1] "data.frame"  
##   
## $row.names  
## [1] 1 2 3 4 5 6 7 8 9 10 11 12  
## [13] 13 14 15 16 17 18 19 20 21 22 23 24  
## [25] 25 26 27 28 29 30 31 32 33 34 35 36  
## [37] 37 38 39 40 41 42 43 44 45 46 47 48  
## [49] 49 50 51 52 53 54 55 56 57 58   
  
## [ reached getOption("max.print") -- omitted 19391 entries ]

sum(is.na(hotel\_bookings))

## [1] 4

head(hotel\_bookings)

## hotel is\_canceled lead\_time arrival\_date\_year arrival\_date\_month  
## 1 Resort Hotel 0 342 2015 July  
## 2 Resort Hotel 0 737 2015 July  
## 3 Resort Hotel 0 7 2015 July  
## 4 Resort Hotel 0 13 2015 July  
## 5 Resort Hotel 0 14 2015 July  
## 6 Resort Hotel 0 14 2015 July  
## arrival\_date\_week\_number arrival\_date\_day\_of\_month stays\_in\_weekend\_nights  
## 1 27 1 0  
## 2 27 1 0  
## 3 27 1 0  
## 4 27 1 0  
## 5 27 1 0  
## 6 27 1 0  
## stays\_in\_week\_nights adults children babies meal country market\_segment  
## 1 0 2 0 0 BB PRT Direct  
## 2 0 2 0 0 BB PRT Direct  
## 3 1 1 0 0 BB GBR Direct  
## 4 1 1 0 0 BB GBR Corporate  
## 5 2 2 0 0 BB GBR Online TA  
## 6 2 2 0 0 BB GBR Online TA  
## distribution\_channel is\_repeated\_guest previous\_cancellations  
## 1 Direct 0 0  
## 2 Direct 0 0  
## 3 Direct 0 0  
## 4 Corporate 0 0  
## 5 TA/TO 0 0  
## 6 TA/TO 0 0  
## previous\_bookings\_not\_canceled reserved\_room\_type assigned\_room\_type  
## 1 0 C C  
## 2 0 C C  
## 3 0 A C  
## 4 0 A A  
## 5 0 A A  
## 6 0 A A  
## booking\_changes deposit\_type agent company days\_in\_waiting\_list customer\_type  
## 1 3 No Deposit NULL NULL 0 Transient  
## 2 4 No Deposit NULL NULL 0 Transient  
## 3 0 No Deposit NULL NULL 0 Transient  
## 4 0 No Deposit 304 NULL 0 Transient  
## 5 0 No Deposit 240 NULL 0 Transient  
## 6 0 No Deposit 240 NULL 0 Transient  
## adr required\_car\_parking\_spaces total\_of\_special\_requests reservation\_status  
## 1 0 0 0 Check-Out  
## 2 0 0 0 Check-Out  
## 3 75 0 0 Check-Out  
## 4 75 0 0 Check-Out  
## 5 98 0 1 Check-Out  
## 6 98 0 1 Check-Out  
## reservation\_status\_date  
## 1 2015-07-01  
## 2 2015-07-01  
## 3 2015-07-02  
## 4 2015-07-02  
## 5 2015-07-03  
## 6 2015-07-03

tail(hotel\_bookings)

## hotel is\_canceled lead\_time arrival\_date\_year arrival\_date\_month  
## 119385 City Hotel 0 21 2017 August  
## 119386 City Hotel 0 23 2017 August  
## 119387 City Hotel 0 102 2017 August  
## 119388 City Hotel 0 34 2017 August  
## 119389 City Hotel 0 109 2017 August  
## 119390 City Hotel 0 205 2017 August  
## arrival\_date\_week\_number arrival\_date\_day\_of\_month  
## 119385 35 30  
## 119386 35 30  
## 119387 35 31  
## 119388 35 31  
## 119389 35 31  
## 119390 35 29  
## stays\_in\_weekend\_nights stays\_in\_week\_nights adults children babies meal  
## 119385 2 5 2 0 0 BB  
## 119386 2 5 2 0 0 BB  
## 119387 2 5 3 0 0 BB  
## 119388 2 5 2 0 0 BB  
## 119389 2 5 2 0 0 BB  
## 119390 2 7 2 0 0 HB  
## country market\_segment distribution\_channel is\_repeated\_guest  
## 119385 BEL Offline TA/TO TA/TO 0  
## 119386 BEL Offline TA/TO TA/TO 0  
## 119387 FRA Online TA TA/TO 0  
## 119388 DEU Online TA TA/TO 0  
## 119389 GBR Online TA TA/TO 0  
## 119390 DEU Online TA TA/TO 0  
## previous\_cancellations previous\_bookings\_not\_canceled reserved\_room\_type  
## 119385 0 0 A  
## 119386 0 0 A  
## 119387 0 0 E  
## 119388 0 0 D  
## 119389 0 0 A  
## 119390 0 0 A  
## assigned\_room\_type booking\_changes deposit\_type agent company  
## 119385 A 0 No Deposit 394 NULL  
## 119386 A 0 No Deposit 394 NULL  
## 119387 E 0 No Deposit 9 NULL  
## 119388 D 0 No Deposit 9 NULL  
## 119389 A 0 No Deposit 89 NULL  
## 119390 A 0 No Deposit 9 NULL  
## days\_in\_waiting\_list customer\_type adr required\_car\_parking\_spaces  
## 119385 0 Transient 96.14 0  
## 119386 0 Transient 96.14 0  
## 119387 0 Transient 225.43 0  
## 119388 0 Transient 157.71 0  
## 119389 0 Transient 104.40 0  
## 119390 0 Transient 151.20 0  
## total\_of\_special\_requests reservation\_status reservation\_status\_date  
## 119385 2 Check-Out 2017-09-06  
## 119386 0 Check-Out 2017-09-06  
## 119387 2 Check-Out 2017-09-07  
## 119388 4 Check-Out 2017-09-07  
## 119389 0 Check-Out 2017-09-07  
## 119390 2 Check-Out 2017-09-07

hotel\_bookings[5:10,]

## hotel is\_canceled lead\_time arrival\_date\_year arrival\_date\_month  
## 5 Resort Hotel 0 14 2015 July  
## 6 Resort Hotel 0 14 2015 July  
## 7 Resort Hotel 0 0 2015 July  
## 8 Resort Hotel 0 9 2015 July  
## 9 Resort Hotel 1 85 2015 July  
## 10 Resort Hotel 1 75 2015 July  
## arrival\_date\_week\_number arrival\_date\_day\_of\_month stays\_in\_weekend\_nights  
## 5 27 1 0  
## 6 27 1 0  
## 7 27 1 0  
## 8 27 1 0  
## 9 27 1 0  
## 10 27 1 0  
## stays\_in\_week\_nights adults children babies meal country market\_segment  
## 5 2 2 0 0 BB GBR Online TA  
## 6 2 2 0 0 BB GBR Online TA  
## 7 2 2 0 0 BB PRT Direct  
## 8 2 2 0 0 FB PRT Direct  
## 9 3 2 0 0 BB PRT Online TA  
## 10 3 2 0 0 HB PRT Offline TA/TO  
## distribution\_channel is\_repeated\_guest previous\_cancellations  
## 5 TA/TO 0 0  
## 6 TA/TO 0 0  
## 7 Direct 0 0  
## 8 Direct 0 0  
## 9 TA/TO 0 0  
## 10 TA/TO 0 0  
## previous\_bookings\_not\_canceled reserved\_room\_type assigned\_room\_type  
## 5 0 A A  
## 6 0 A A  
## 7 0 C C  
## 8 0 C C  
## 9 0 A A  
## 10 0 D D  
## booking\_changes deposit\_type agent company days\_in\_waiting\_list  
## 5 0 No Deposit 240 NULL 0  
## 6 0 No Deposit 240 NULL 0  
## 7 0 No Deposit NULL NULL 0  
## 8 0 No Deposit 303 NULL 0  
## 9 0 No Deposit 240 NULL 0  
## 10 0 No Deposit 15 NULL 0  
## customer\_type adr required\_car\_parking\_spaces total\_of\_special\_requests  
## 5 Transient 98.0 0 1  
## 6 Transient 98.0 0 1  
## 7 Transient 107.0 0 0  
## 8 Transient 103.0 0 1  
## 9 Transient 82.0 0 1  
## 10 Transient 105.5 0 0  
## reservation\_status reservation\_status\_date  
## 5 Check-Out 2015-07-03  
## 6 Check-Out 2015-07-03  
## 7 Check-Out 2015-07-03  
## 8 Check-Out 2015-07-03  
## 9 Canceled 2015-05-06  
## 10 Canceled 2015-04-22

hotel\_bookings[5:10,"adults"]

## [1] 2 2 2 2 2 2

summary(hotel\_bookings)

## hotel is\_canceled lead\_time arrival\_date\_year  
## Length:119390 Min. :0.0000 Min. : 0 Min. :2015   
## Class :character 1st Qu.:0.0000 1st Qu.: 18 1st Qu.:2016   
## Mode :character Median :0.0000 Median : 69 Median :2016   
## Mean :0.3704 Mean :104 Mean :2016   
## 3rd Qu.:1.0000 3rd Qu.:160 3rd Qu.:2017   
## Max. :1.0000 Max. :737 Max. :2017   
##   
## arrival\_date\_month arrival\_date\_week\_number arrival\_date\_day\_of\_month  
## Length:119390 Min. : 1.00 Min. : 1.0   
## Class :character 1st Qu.:16.00 1st Qu.: 8.0   
## Mode :character Median :28.00 Median :16.0   
## Mean :27.17 Mean :15.8   
## 3rd Qu.:38.00 3rd Qu.:23.0   
## Max. :53.00 Max. :31.0   
##   
## stays\_in\_weekend\_nights stays\_in\_week\_nights adults   
## Min. : 0.0000 Min. : 0.0 Min. : 0.000   
## 1st Qu.: 0.0000 1st Qu.: 1.0 1st Qu.: 2.000   
## Median : 1.0000 Median : 2.0 Median : 2.000   
## Mean : 0.9276 Mean : 2.5 Mean : 1.856   
## 3rd Qu.: 2.0000 3rd Qu.: 3.0 3rd Qu.: 2.000   
## Max. :19.0000 Max. :50.0 Max. :55.000   
##   
## children babies meal country   
## Min. : 0.0000 Min. : 0.000000 Length:119390 Length:119390   
## 1st Qu.: 0.0000 1st Qu.: 0.000000 Class :character Class :character   
## Median : 0.0000 Median : 0.000000 Mode :character Mode :character   
## Mean : 0.1039 Mean : 0.007949   
## 3rd Qu.: 0.0000 3rd Qu.: 0.000000   
## Max. :10.0000 Max. :10.000000   
## NA's :4   
## market\_segment distribution\_channel is\_repeated\_guest  
## Length:119390 Length:119390 Min. :0.00000   
## Class :character Class :character 1st Qu.:0.00000   
## Mode :character Mode :character Median :0.00000   
## Mean :0.03191   
## 3rd Qu.:0.00000   
## Max. :1.00000   
##   
## previous\_cancellations previous\_bookings\_not\_canceled reserved\_room\_type  
## Min. : 0.00000 Min. : 0.0000 Length:119390   
## 1st Qu.: 0.00000 1st Qu.: 0.0000 Class :character   
## Median : 0.00000 Median : 0.0000 Mode :character   
## Mean : 0.08712 Mean : 0.1371   
## 3rd Qu.: 0.00000 3rd Qu.: 0.0000   
## Max. :26.00000 Max. :72.0000   
##   
## assigned\_room\_type booking\_changes deposit\_type agent   
## Length:119390 Min. : 0.0000 Length:119390 Length:119390   
## Class :character 1st Qu.: 0.0000 Class :character Class :character   
## Mode :character Median : 0.0000 Mode :character Mode :character   
## Mean : 0.2211   
## 3rd Qu.: 0.0000   
## Max. :21.0000   
##   
## company days\_in\_waiting\_list customer\_type adr   
## Length:119390 Min. : 0.000 Length:119390 Min. : -6.38   
## Class :character 1st Qu.: 0.000 Class :character 1st Qu.: 69.29   
## Mode :character Median : 0.000 Mode :character Median : 94.58   
## Mean : 2.321 Mean : 101.83   
## 3rd Qu.: 0.000 3rd Qu.: 126.00   
## Max. :391.000 Max. :5400.00   
##   
## required\_car\_parking\_spaces total\_of\_special\_requests reservation\_status  
## Min. :0.00000 Min. :0.0000 Length:119390   
## 1st Qu.:0.00000 1st Qu.:0.0000 Class :character   
## Median :0.00000 Median :0.0000 Mode :character   
## Mean :0.06252 Mean :0.5714   
## 3rd Qu.:0.00000 3rd Qu.:1.0000   
## Max. :8.00000 Max. :5.0000   
##   
## reservation\_status\_date  
## Length:119390   
## Class :character   
## Mode :character   
##   
##   
##   
##

*# Evaluating the values of mean,median,IQR,range,quantile,variance and standard deviation of the dataset.*

mean(hotel\_bookings$adults)

## [1] 1.856403

median(hotel\_bookings$adults)

## [1] 2

IQR(hotel\_bookings$adults)

## [1] 0

range(hotel\_bookings$adults)

## [1] 0 55

quantile(hotel\_bookings$adults,c(0.10,0.60,0.70))

## 10% 60% 70%   
## 1 2 2

var(hotel\_bookings$adults)

## [1] 0.3355433

sd(hotel\_bookings$adults)

## [1] 0.579261

library(Hmisc)

## Warning: package 'Hmisc' was built under R version 4.2.2

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 4.2.2

##   
## Attaching package: 'Hmisc'

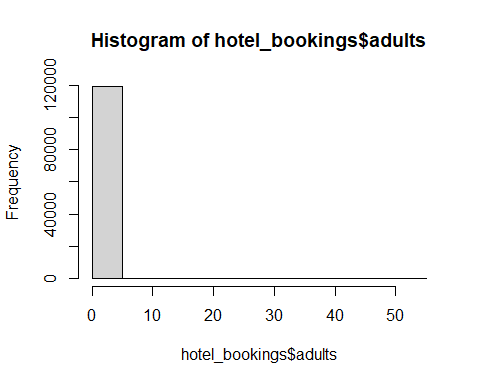
## The following objects are masked from 'package:base':  
##   
## format.pval, units

describe(hotel\_bookings)

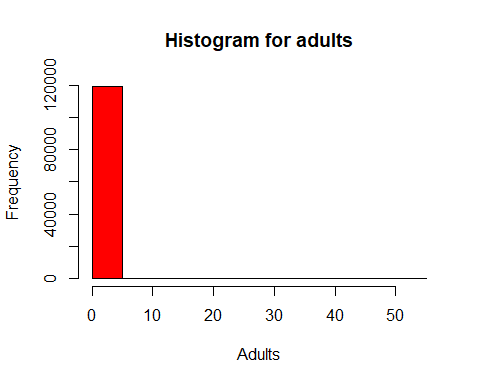
## hotel\_bookings   
##   
## 32 Variables 119390 Observations  
## --------------------------------------------------------------------------------  
## hotel   
## n missing distinct   
## 119390 0 2   
##   
## Value City Hotel Resort Hotel  
## Frequency 79330 40060  
## Proportion 0.664 0.336  
## --------------------------------------------------------------------------------  
## is\_canceled   
## n missing distinct Info Sum Mean Gmd   
## 119390 0 2 0.7 44224 0.3704 0.4664   
##   
## --------------------------------------------------------------------------------  
## lead\_time   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 479 1 104 112.5 0 3   
## .25 .50 .75 .90 .95   
## 18 69 160 265 320   
##   
## lowest : 0 1 2 3 4, highest: 622 626 629 709 737  
## --------------------------------------------------------------------------------  
## arrival\_date\_year   
## n missing distinct Info Mean Gmd   
## 119390 0 3 0.847 2016 0.7499   
##   
## Value 2015 2016 2017  
## Frequency 21996 56707 40687  
## Proportion 0.184 0.475 0.341  
## --------------------------------------------------------------------------------  
## arrival\_date\_month   
## n missing distinct   
## 119390 0 12   
##   
## lowest : April August December February January   
## highest: March May November October September  
##   
## Value April August December February January July  
## Frequency 11089 13877 6780 8068 5929 12661  
## Proportion 0.093 0.116 0.057 0.068 0.050 0.106  
##   
## Value June March May November October September  
## Frequency 10939 9794 11791 6794 11160 10508  
## Proportion 0.092 0.082 0.099 0.057 0.093 0.088  
## --------------------------------------------------------------------------------  
## arrival\_date\_week\_number   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 53 1 27.17 15.68 5 8   
## .25 .50 .75 .90 .95   
## 16 28 38 46 49   
##   
## lowest : 1 2 3 4 5, highest: 49 50 51 52 53  
## --------------------------------------------------------------------------------  
## arrival\_date\_day\_of\_month   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 31 0.999 15.8 10.13 2 4   
## .25 .50 .75 .90 .95   
## 8 16 23 28 30   
##   
## lowest : 1 2 3 4 5, highest: 27 28 29 30 31  
## --------------------------------------------------------------------------------  
## stays\_in\_weekend\_nights   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 17 0.879 0.9276 1.026 0 0   
## .25 .50 .75 .90 .95   
## 0 1 2 2 2   
##   
## lowest : 0 1 2 3 4, highest: 13 14 16 18 19  
##   
## Value 0 1 2 3 4 5 6 7 8 9 10  
## Frequency 51998 30626 33308 1259 1855 79 153 19 60 11 7  
## Proportion 0.436 0.257 0.279 0.011 0.016 0.001 0.001 0.000 0.001 0.000 0.000  
##   
## Value 12 13 14 16 18 19  
## Frequency 5 3 2 3 1 1  
## Proportion 0.000 0.000 0.000 0.000 0.000 0.000  
## --------------------------------------------------------------------------------  
## stays\_in\_week\_nights   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 35 0.953 2.5 1.865 0 1   
## .25 .50 .75 .90 .95   
## 1 2 3 5 5   
##   
## lowest : 0 1 2 3 4, highest: 35 40 41 42 50  
## --------------------------------------------------------------------------------  
## adults   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 14 0.569 1.856 0.4287 1 1   
## .25 .50 .75 .90 .95   
## 2 2 2 2 3   
##   
## lowest : 0 1 2 3 4, highest: 26 27 40 50 55  
##   
## Value 0 1 2 3 4 5 6 10 20 26 27  
## Frequency 403 23027 89680 6202 62 2 1 1 2 5 2  
## Proportion 0.003 0.193 0.751 0.052 0.001 0.000 0.000 0.000 0.000 0.000 0.000  
##   
## Value 40 50 55  
## Frequency 1 1 1  
## Proportion 0.000 0.000 0.000  
## --------------------------------------------------------------------------------  
## children   
## n missing distinct Info Mean Gmd   
## 119386 4 5 0.201 0.1039 0.1955   
##   
## lowest : 0 1 2 3 10, highest: 0 1 2 3 10  
##   
## Value 0 1 2 3 10  
## Frequency 110796 4861 3652 76 1  
## Proportion 0.928 0.041 0.031 0.001 0.000  
## --------------------------------------------------------------------------------  
## babies   
## n missing distinct Info Mean Gmd   
## 119390 0 5 0.023 0.007949 0.01578   
##   
## lowest : 0 1 2 9 10, highest: 0 1 2 9 10  
##   
## Value 0 1 2 9 10  
## Frequency 118473 900 15 1 1  
## Proportion 0.992 0.008 0.000 0.000 0.000  
## --------------------------------------------------------------------------------  
## meal   
## n missing distinct   
## 119390 0 5   
##   
## lowest : BB FB HB SC Undefined  
## highest: BB FB HB SC Undefined  
##   
## Value BB FB HB SC Undefined  
## Frequency 92310 798 14463 10650 1169  
## Proportion 0.773 0.007 0.121 0.089 0.010  
## --------------------------------------------------------------------------------  
## country   
## n missing distinct   
## 119390 0 178   
##   
## lowest : ABW AGO AIA ALB AND, highest: VGB VNM ZAF ZMB ZWE  
## --------------------------------------------------------------------------------  
## market\_segment   
## n missing distinct   
## 119390 0 8   
##   
## lowest : Aviation Complementary Corporate Direct Groups   
## highest: Direct Groups Offline TA/TO Online TA Undefined   
##   
## Value Aviation Complementary Corporate Direct  
## Frequency 237 743 5295 12606  
## Proportion 0.002 0.006 0.044 0.106  
##   
## Value Groups Offline TA/TO Online TA Undefined  
## Frequency 19811 24219 56477 2  
## Proportion 0.166 0.203 0.473 0.000  
## --------------------------------------------------------------------------------  
## distribution\_channel   
## n missing distinct   
## 119390 0 5   
##   
## lowest : Corporate Direct GDS TA/TO Undefined  
## highest: Corporate Direct GDS TA/TO Undefined  
##   
## Value Corporate Direct GDS TA/TO Undefined  
## Frequency 6677 14645 193 97870 5  
## Proportion 0.056 0.123 0.002 0.820 0.000  
## --------------------------------------------------------------------------------  
## is\_repeated\_guest   
## n missing distinct Info Sum Mean Gmd   
## 119390 0 2 0.093 3810 0.03191 0.06179   
##   
## --------------------------------------------------------------------------------  
## previous\_cancellations   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 15 0.154 0.08712 0.1682 0 0   
## .25 .50 .75 .90 .95   
## 0 0 0 0 1   
##   
## lowest : 0 1 2 3 4, highest: 19 21 24 25 26  
##   
## Value 0 1 2 3 4 5 6 11 13  
## Frequency 112906 6051 116 65 31 19 22 35 12  
## Proportion 0.946 0.051 0.001 0.001 0.000 0.000 0.000 0.000 0.000  
##   
## Value 14 19 21 24 25 26  
## Frequency 14 19 1 48 25 26  
## Proportion 0.000 0.000 0.000 0.000 0.000 0.000  
## --------------------------------------------------------------------------------  
## previous\_bookings\_not\_canceled   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 73 0.088 0.1371 0.2708 0 0   
## .25 .50 .75 .90 .95   
## 0 0 0 0 0   
##   
## lowest : 0 1 2 3 4, highest: 68 69 70 71 72  
## --------------------------------------------------------------------------------  
## reserved\_room\_type   
## n missing distinct   
## 119390 0 10   
##   
## lowest : A B C D E, highest: F G H L P  
##   
## Value A B C D E F G H L P  
## Frequency 85994 1118 932 19201 6535 2897 2094 601 6 12  
## Proportion 0.720 0.009 0.008 0.161 0.055 0.024 0.018 0.005 0.000 0.000  
## --------------------------------------------------------------------------------  
## assigned\_room\_type   
## n missing distinct   
## 119390 0 12   
##   
## lowest : A B C D E, highest: H I K L P  
##   
## Value A B C D E F G H I K L  
## Frequency 74053 2163 2375 25322 7806 3751 2553 712 363 279 1  
## Proportion 0.620 0.018 0.020 0.212 0.065 0.031 0.021 0.006 0.003 0.002 0.000  
##   
## Value P  
## Frequency 12  
## Proportion 0.000  
## --------------------------------------------------------------------------------  
## booking\_changes   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 21 0.388 0.2211 0.3919 0 0   
## .25 .50 .75 .90 .95   
## 0 0 0 1 1   
##   
## lowest : 0 1 2 3 4, highest: 16 17 18 20 21  
## --------------------------------------------------------------------------------  
## deposit\_type   
## n missing distinct   
## 119390 0 3   
##   
## Value No Deposit Non Refund Refundable  
## Frequency 104641 14587 162  
## Proportion 0.876 0.122 0.001  
## --------------------------------------------------------------------------------  
## agent   
## n missing distinct   
## 119390 0 334   
##   
## lowest : 1 10 103 104 105 , highest: 95 96 98 99 NULL  
## --------------------------------------------------------------------------------  
## company   
## n missing distinct   
## 119390 0 353   
##   
## lowest : 10 100 101 102 103 , highest: 93 94 96 99 NULL  
## --------------------------------------------------------------------------------  
## days\_in\_waiting\_list   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 128 0.09 2.321 4.559 0 0   
## .25 .50 .75 .90 .95   
## 0 0 0 0 0   
##   
## lowest : 0 1 2 3 4, highest: 236 259 330 379 391  
## --------------------------------------------------------------------------------  
## customer\_type   
## n missing distinct   
## 119390 0 4   
##   
## Value Contract Group Transient Transient-Party  
## Frequency 4076 577 89613 25124  
## Proportion 0.034 0.005 0.751 0.210  
## --------------------------------------------------------------------------------  
## adr   
## n missing distinct Info Mean Gmd .05 .10   
## 119390 0 8879 1 101.8 51.91 38.40 50.00   
## .25 .50 .75 .90 .95   
## 69.29 94.58 126.00 164.00 193.50   
##   
## lowest : -6.38 0.00 0.26 0.50 1.00  
## highest: 450.00 451.50 508.00 510.00 5400.00  
##   
## Value 0 50 100 150 200 250 300 350 400 450 500  
## Frequency 2437 35085 50975 21915 6234 2156 463 99 19 4 2  
## Proportion 0.020 0.294 0.427 0.184 0.052 0.018 0.004 0.001 0.000 0.000 0.000  
##   
## Value 5400  
## Frequency 1  
## Proportion 0.000  
##   
## For the frequency table, variable is rounded to the nearest 50  
## --------------------------------------------------------------------------------  
## required\_car\_parking\_spaces   
## n missing distinct Info Mean Gmd   
## 119390 0 5 0.175 0.06252 0.1173   
##   
## lowest : 0 1 2 3 8, highest: 0 1 2 3 8  
##   
## Value 0 1 2 3 8  
## Frequency 111974 7383 28 3 2  
## Proportion 0.938 0.062 0.000 0.000 0.000  
## --------------------------------------------------------------------------------  
## total\_of\_special\_requests   
## n missing distinct Info Mean Gmd   
## 119390 0 6 0.773 0.5714 0.7684   
##   
## lowest : 0 1 2 3 4, highest: 1 2 3 4 5  
##   
## Value 0 1 2 3 4 5  
## Frequency 70318 33226 12969 2497 340 40  
## Proportion 0.589 0.278 0.109 0.021 0.003 0.000  
## --------------------------------------------------------------------------------  
## reservation\_status   
## n missing distinct   
## 119390 0 3   
##   
## Value Canceled Check-Out No-Show  
## Frequency 43017 75166 1207  
## Proportion 0.36 0.63 0.01  
## --------------------------------------------------------------------------------  
## reservation\_status\_date   
## n missing distinct   
## 119390 0 926   
##   
## lowest : 2014-10-17 2014-11-18 2015-01-01 2015-01-02 2015-01-18  
## highest: 2017-09-08 2017-09-09 2017-09-10 2017-09-12 2017-09-14  
## --------------------------------------------------------------------------------

*# Histogram of the dataset*

hist(hotel\_bookings$adults)

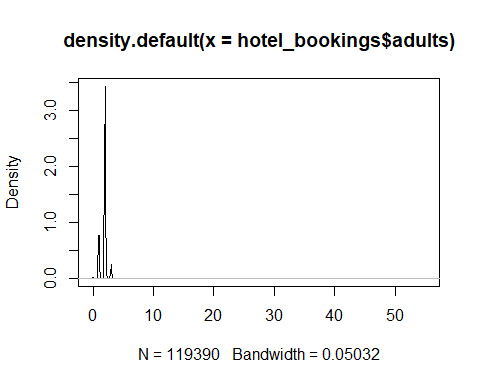


hist(hotel\_bookings$adults,col = "red",xlab = "Adults",main = "Histogram for adults")

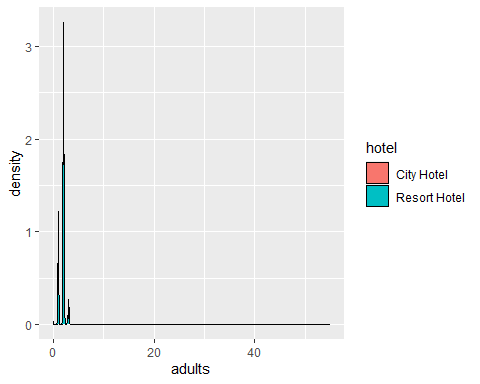


*# Plotting the dataset*

plot(density(hotel\_bookings$adults))



ggplot(data = hotel\_bookings,aes(adults,fill=hotel))+  
 geom\_density()

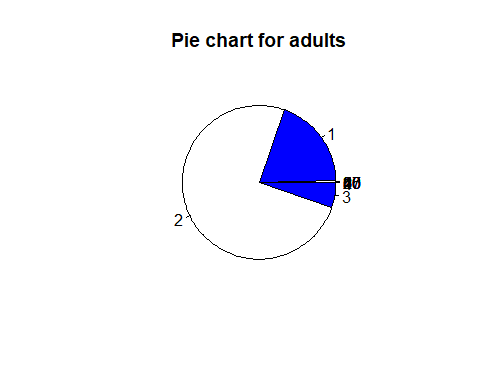


# Explore categorical variables  
table(hotel\_bookings$adults)

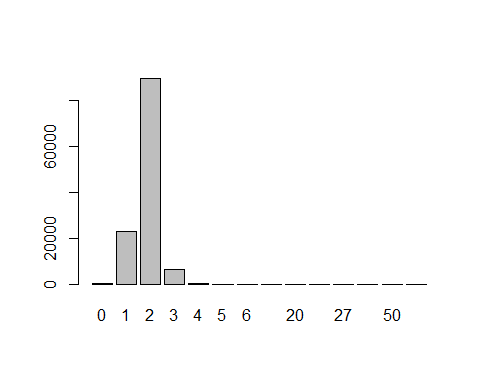
##   
## 0 1 2 3 4 5 6 10 20 26 27 40 50   
## 403 23027 89680 6202 62 2 1 1 2 5 2 1 1   
## 55   
## 1

*# Pie chart for the dataset*

pie(table(hotel\_bookings$adults),  
 main = "Pie chart for adults",col = c("white","blue"))



# Bar graph of the dataset  
barplot(table(hotel\_bookings$adults))



# Correlation  
cor(hotel\_bookings$adults,hotel\_bookings$lead\_time)

## [1] 0.1195187

plot(hotel\_bookings$adults,hotel\_bookings$lead\_time)

