Problem statement: - A company has an automated Service Request system. All its customers can call into the help line number and the service request is logged into the system by the customer care executive. An automated Ticket id or Service Request Number (SR number) is generated. This Ticket id is shared with the customer. Once the request is successfully resolved, the customer service executive gives the customer a status call and closes the same on the system.

Other data points captured include

1. Entitlement- Type of Request
2. Impact- What is the impact on the customer’s business because of the problem
3. Billable – Customers can be on a Billable service plan OR on a Free service plan
4. Date Opened – Time stamp when the Service Request was put in the system
5. Closed Date - Time stamp when the Service Request was closed in the system

The Company wants to understand the average time it takes to resolve queries.

1. Does this average time vary by Impact ?
2. What is the data that should be used to make commitment on resolution time to customers for different Impact ?

Solution :-

# SET WORKINg directory

setwd('H://springer book//Case study//CaseStudy4')

#IMPORT DATA FILE

Resolution<- read.csv("H:/springer book/Case study/CaseStudy4/Resolution time for Service request.csv", stringsAsFactors=FALSE)

#CHECK FORMAT

str(Resolution)

> str(Resolution)

'data.frame': 117 obs. of 7 variables:

$ SR.number : chr "1-7657336422" "1-7658643852" "1-7735438423" "1-7880118403" ...

$ SR.Type : chr "Incident" "Incident" "Incident" "Incident" ...

$ Entitlement: chr "Requested Services SESP Basic - Problem Management - 52" "Requested Services SESP Basic - Problem Management - 52" "Requested Services SESP Basic - Problem Management - 52" "Requested Services SESP Basic - Problem Management - 52" ...

$ Impact : chr "Moderate" "Significant" "Moderate" "Moderate" ...

$ Billable : chr "N" "N" "N" "N" ...

$ Date.Opened: chr "9/11/2014 11:14" "9/11/2014 20:02" "10/8/2014 8:04" "11/13/2014 13:19" ...

$ Closed.Date: chr "10/16/2014 17:33" "10/22/2014 15:04" "4/15/2015 17:28" "12/19/2014 13:45" ...

# D = CREATE Y VARIABLE = CREATE RESOLUTON TIME

# CONVERT Date.Opened AND Closed.Date INTO DATE TIME FORMAT

Resolution$Open <- as.Date(Resolution$Date.Opened, "%m/%d/%Y %H:%M")

Resolution$closed <- as.Date(Resolution$Closed.Date, "%m/%d/%Y %H:%M")

str(Resolution$Open)

str(Resolution$closed)

> str(Resolution$Open)

Date[1:117], format: "2014-09-11" "2014-09-11" "2014-10-08" "2014-11-13" "2014-11-21" ...

> str(Resolution$closed)

Date[1:117], format: "2014-10-16" "2014-10-22" "2015-04-15" "2014-12-19" "2014-12-07" ...

# DERIVED VARIABLE Y = RESOLUTION TIME

Resolution$resolution.time=(Resolution$closed -Resolution$Open)

View(Resolution)

str(Resolution$resolution.time)

> str(Resolution$resolution.time)

Class 'difftime' atomic [1:117] 35 41 189 36 16 16 57 24 10 9 ...

..- attr(\*, "units")= chr "days"

# CONVERT RESOLUTION TIME INTO NUMERIC

Resolution$resolution.time= as.numeric(Resolution$resolution.time)

str(Resolution$resolution.time)

# Box plot

boxplot(Resolution$resolution.time, horizontal=TRUE, main="RESOLUTION\_TIME")

Note :- Some Service Requests have very long time to closure

# DENSITY PLOT

d<- density(Resolution$resolution.time)

plot(d)

Note :- Most of the values have a resolution time of <=100 days. Some values MAY have Resolution time of less than 0 days. Those are outliers and should be removed from the study

# CHECK IF THERE ARE ANY VALUE LESS THAN 0 FOR RESOLUTION TIME

attach(Resolution)

Resolution$resolution.time.cat[resolution.time<0]<-0

Resolution$resolution.time.cat[resolution.time>=0]<-1

# CREATE TABLE TO CHECK FREQUENCY

mytable<- table(Resolution$resolution.time.cat)

mytable

> mytable

1

117

# C & O - COLLECT AND ORGANISE THE DATA

# CHECK FOR MISSING VALUES

#EXPLORE THE DATA TO UNDERSTAND NA AND OTHER SEGMENTS FOR DISCRETE VARIABLES

mytable<- table(Resolution$SR.Type)

mytable

> mytable

Incident Problem Request for Fulfillment

80 29 8

> mytable<- table(Resolution$Entitlement)

> mytable

45

Operations & Alarm Management - BGP - 14

1

Preventive Maintenance - 1

1

Preventive Maintenance - 15

2

Preventive Maintenance - 3

1

Process History & Analytics - BGP - 7

1

Process Optimization - BGP

1

Process Optimization - BGP - 1

1

Process Optimization - BGP - 15

2

Process Optimization - BGP - 18

1

Process Optimization - BGP - 2

1

Process Optimization - BGP - 7

2

Requested Services - Fulfilment

14

Requested Services - Fulfilment - 1

1

Requested Services - Fulfilment - 2

7

Requested Services - Fulfilment - 3

4

Requested Services - Incident Support - 1

6

Requested Services Hiway Care Full - Fulfilment - 25

3

Requested Services SESP Basic - Fulfilment - 5

1

Requested Services SESP Basic - Fulfilment - 8

1

Requested Services SESP Basic - Incident Support - 5

1

Requested Services SESP Basic - Problem Management - 52

11

Requested Services SESP Remote - Fulfilment - 13

8

Requested Services SESP Remote - Incident Support - 5

1

> mytable<- table(Resolution$Impact)

> mytable

Minor Moderate Significant

9 63 45

> mytable<- table(Resolution$Billable)

> mytable

N Y

107 10

Note :- No need to check for Primary key – Service Request Number (SR number)

# MISSING VALUES FOR CONTINUOUS DATA IN 10TH COLUMN- RESOLUTION TIME

Resolution[!complete.cases(resolution.time),10]

> Resolution[!complete.cases(resolution.time),10]

numeric(0)

# OUTLIERS FOR CONTINUOUS VARIABLE -RESOLUTION TIME

# I HAVE DEFINED OUTLIERS AS BEING IN THE TOP .003% OF THE NORMAL DISTRIBUTION POPULATION

install.packages("outliers")

library(outliers)

outs <- scores(Resolution$resolution.time, type="chisq", prob=0.997)

Resolution$resolution.time[outs]

> Resolution$resolution.time[outs]

numeric(0)

Note :– No outlier detected

# A – ANALYZE :- FREQUENCY TABLE OF AVG. RESOLUTION TIME BY IMPACT

library("plyr")

table1<- ddply(Resolution, c("Impact"), summarise,

N = length(resolution.time),

mean = mean(resolution.time),

median =median(resolution.time),

sd = sd(resolution.time)

)

table1

> table1

Impact N mean median sd

1 Minor 9 13.11111 8 12.81059

2 Moderate 63 58.28571 24 65.47849

3 Significant 45 53.04444 30 55.27738

Note :- most of the cases fall under Moderate Impact . The Resolution time for Moderate Impact is very high

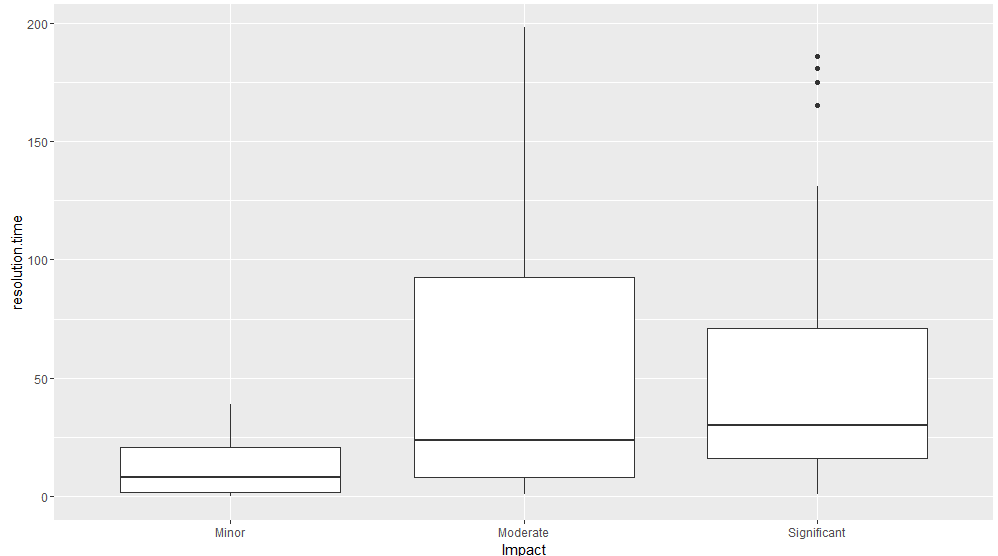
# CREATE BOXPLOT TO CHECK DETAILS OF SEGMENTS UNDER IMPACT

library(ggplot2)

bp1 <- ggplot(Resolution, aes(x=Impact, y=resolution.time)) +

geom\_boxplot()

bp1

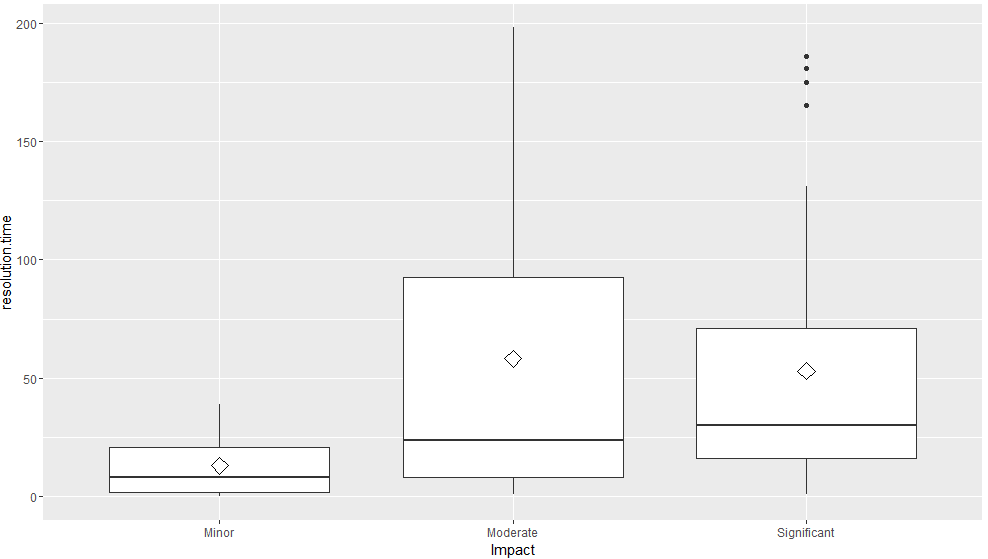


# PUT A DIAMOND WHERE THE MEAN IS

bp2<- ggplot(Resolution, aes(x=Impact, y=resolution.time)) + geom\_boxplot() +

stat\_summary(fun.y=mean, geom="point", shape=5, size=4)

bp2



# FIND PERCENTILES

quantile(Resolution$resolution.time, c(.25, .50, .75, .90, .99))

> quantile(Resolution$resolution.time, c(.25, .50, .75, .90, .99))

25% 50% 75% 90% 99%

9.00 24.00 71.00 165.80 191.52

summary(Resolution$resolution.time)

> summary(Resolution$resolution.time)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.00 9.00 24.00 52.79 71.00 198.00

# MAKE SUBSETS FOR VARIBLE IMPACT AND RUN QUARTILES

str(Resolution)

data1 <- subset(Resolution,Resolution$Impact=="Minor")

quantile(data1$resolution.time, c(.25, .50, .75, .90, .99))

> quantile(data1$resolution.time, c(.25, .50, .75, .90, .99))

25% 50% 75% 90% 99%

2.00 8.00 21.00 24.60 37.56

Data2 <- subset(Resolution,Resolution$Impact=="Significant")

quantile(data2$resolution.time, c(.25, .50, .75, .90, .99))

> quantile(data1$resolution.time, c(.25, .50, .75, .90, .99))

25% 50% 75% 90% 99%

16.0 30.0 71.0 151.4 186.0

# REMOVE OUTLIER FOR SIGNIFICANT >99%

data2.1 <- subset(data2,data2$resolution.time <186.0)

quantile(data2.1$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data2.1$resolution.time)

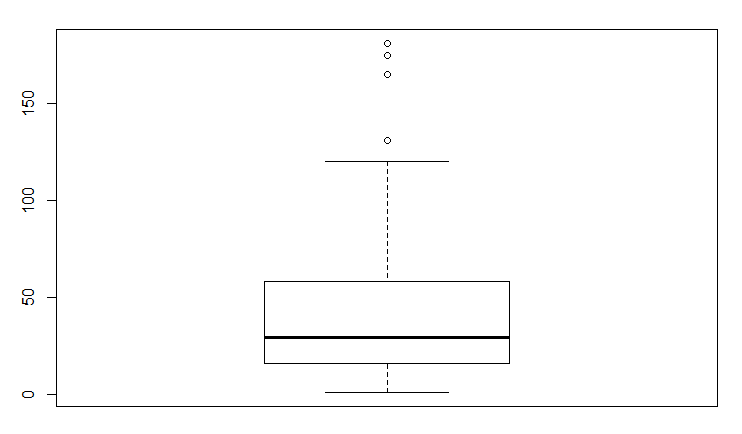
> data2.1 <- subset(data2,data2$resolution.time <186.0)

> quantile(data2.1$resolution.time, c(.25, .50, .75, .90, .99))

25% 50% 75% 90% 99%

16.00 29.00 58.00 119.60 178.48

> boxplot(data2.1$resolution.time)



data1 <- subset(Resolution,Resolution$Impact=="Moderate")

quantile(data1$resolution.time, c(.25, .50, .75, .90, .99))

> quantile(data1$resolution.time, c(.25, .50, .75, .90, .99))

25% 50% 75% 90% 99%

8.00 24.00 92.50 178.20 194.28

# REMOVE OUTLIER FOR MODERATE >99% AND REWORK BECAUSE MODERATE HAS A LONG TAIL IN THE BOXPLOT

dim(data1)

data1.1 <- subset(data1,data1$resolution.time <=194.28)

quantile(data1.1$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data1.1$resolution.time)

# REDO OUTLIER TREATMENT

data1.2 <- subset(data1.1,data1.1$resolution.time <=190.17)

quantile(data1.2$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data1.2$resolution.time)

# REDO OUTLIER TREATMENT

data1.3 <- subset(data1.2,data1.2$resolution.time <=187.8)

quantile(data1.3$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data1.3$resolution.time)

dim(data1.3)

# REDO OUTLIER TREATMENT

data1.4 <- subset(data1.3,data1.3$resolution.time <=185.82)

quantile(data1.4$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data1.4$resolution.time)

dim(data1.4)

# REDO OUTLIER TREATMENT

data1.5 <- subset(data1.4,data1.4$resolution.time <=185.82)

quantile(data1.5$resolution.time, c(.25, .50, .75, .90, .99))

boxplot(data1.5$resolution.time)

dim(data1.5)

> data1.5 <- subset(data1.4,data1.4$resolution.time <=185.82)

> quantile(data1.5$resolution.time, c(.25, .50, .75, .90, .99))

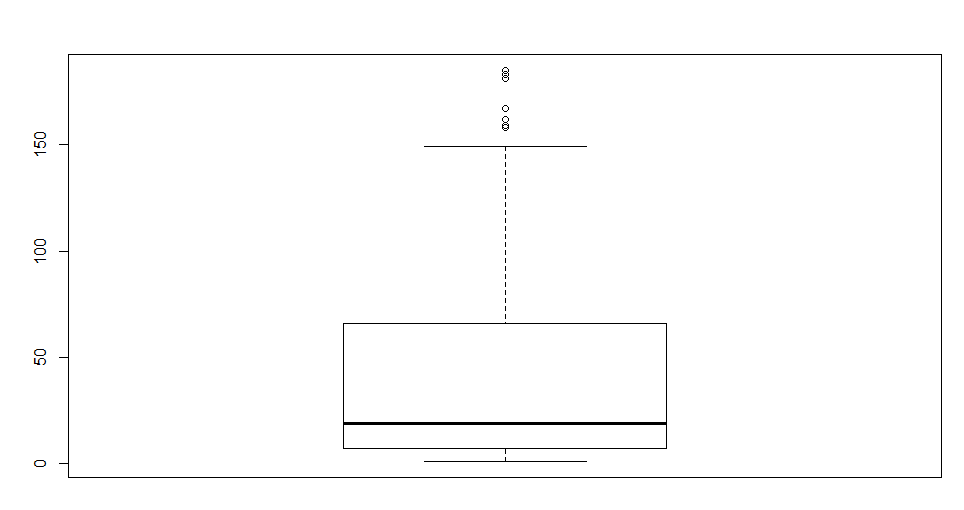
25% 50% 75% 90% 99%

7.00 19.00 66.00 158.20 183.84

> boxplot(data1.5$resolution.time)

> dim(data1.5)

[1] 59 11



summary(data1.5$resolution.time)

> summary(data1.5$resolution.time)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.00 7.00 19.00 49.25 66.00 185.00

sd(data1.5$resolution.time)

> sd(data1.5$resolution.time)

[1] 57.2284

# I – INSIGHT :-

INSIGHT :-

1. The average Resolution Time varies significantly by IMPACT
2. The SLA can be derived using Chebychev’s theorem and Empirical Rule :-
   1. For Moderate Impact cases
      1. By Checbyshev’s theorem
         1. Atleast 75% of the cases get resolved between 0 – 163 days
         2. Atleast 88.89% of the cases get resolved between 0-220 days
      2. By Emperical Rule
         1. 95% of the cases get resolved between 0 – 163 days
         2. 99.97% of the cases get resolved between 0-220 days

Work to do :- Define the significant ranges where IMPACT is Minor and Significant