Case study :-

A Retail company has a tracker to track sales names “Sales”. However, there are some returns also in the file names “Sales\_returns” . The company has sales managers who handle different regions in the file “Sales\_manager”.

The company wants to understand the following :-

1. What is the net sales after returns have been factored in ?
2. What % of the total sales do the returns form?
3. Is the returns linked to value of items (cost) ?

# IMPORT DATA

Sales <- read.csv("F:/springer book/Case study/CaseStudy7/Sales.csv", stringsAsFactors=FALSE)

str(Sales)

> str(Sales)

'data.frame': 8399 obs. of 21 variables:

$ Row.ID : int 1 49 50 80 85 86 97 98 103 107 ...

$ Order.ID : int 3 293 293 483 515 515 613 613 643 678 ...

$ Order.Date : chr "10/13/2010" "10/1/2012" "10/1/2012" "7/10/2011" ...

$ Order.Priority : chr "Low" "High" "High" "High" ...

$ Order.Quantity : int 6 49 27 30 19 21 12 22 21 44 ...

dim(Sales)

> dim(Sales)

[1] 8399 21

# C and O

# Merge the Returns data and the Manager name

# import the 2 datasets

Manager <- read.csv("F:/springer book/Case study/CaseStudy7/Sales\_manager.csv", stringsAsFactors=FALSE)

str(Manager)

Returns <- read.csv("F:/springer book/Case study/CaseStudy7/Sales\_returns.csv", stringsAsFactors=FALSE)

str(Returns)

# Sort data

attach(Returns)

Returns1 <- Returns[order(Order.ID),]

attach(Sales)

Sales1 <- Sales[order(Order.ID),]

total <- merge(Sales1,Returns1,by="Order.ID",all = TRUE)

dim(total)

> dim(total)

[1] 8399 22

# Visualise - Sales vs Returns (Status = Returned / NA)

total[["Status"]][is.na(total[["Status"]])] <- "NotReturned"

# frequency of returns

library(MASS)

mytable<- xtabs(~Status, data = total)

mytable

pie(mytable)

slices<- c(7527, 872)

lbls<- c("NotReturned", "Returned")

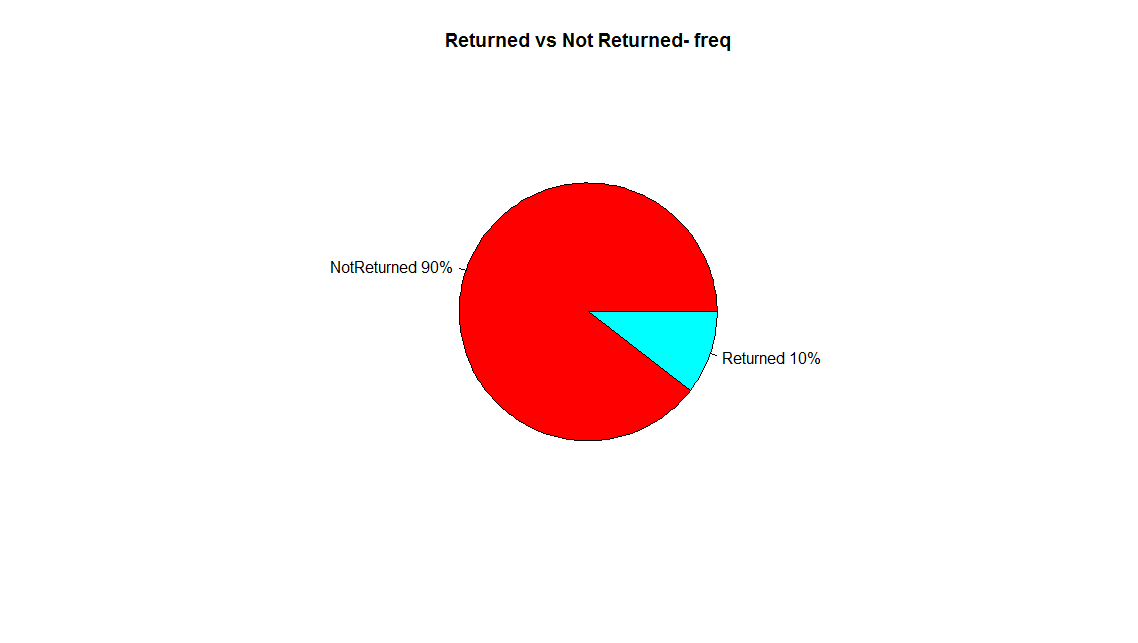
pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct)

lbls <- paste(lbls,"%",sep="")

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Returned vs Not Returned- freq")



# value of returns thru cross tab

total$Salesnum <- total$Sales

mytable2 <- aggregate(Salesnum ~ Status, total, sum)

mytable2

> mytable2

Status Salesnum

1 NotReturned 13260747

2 Returned 1654854

slices<- c(13260747, 1654854)

lbls<- c("NotReturned", "Returned")

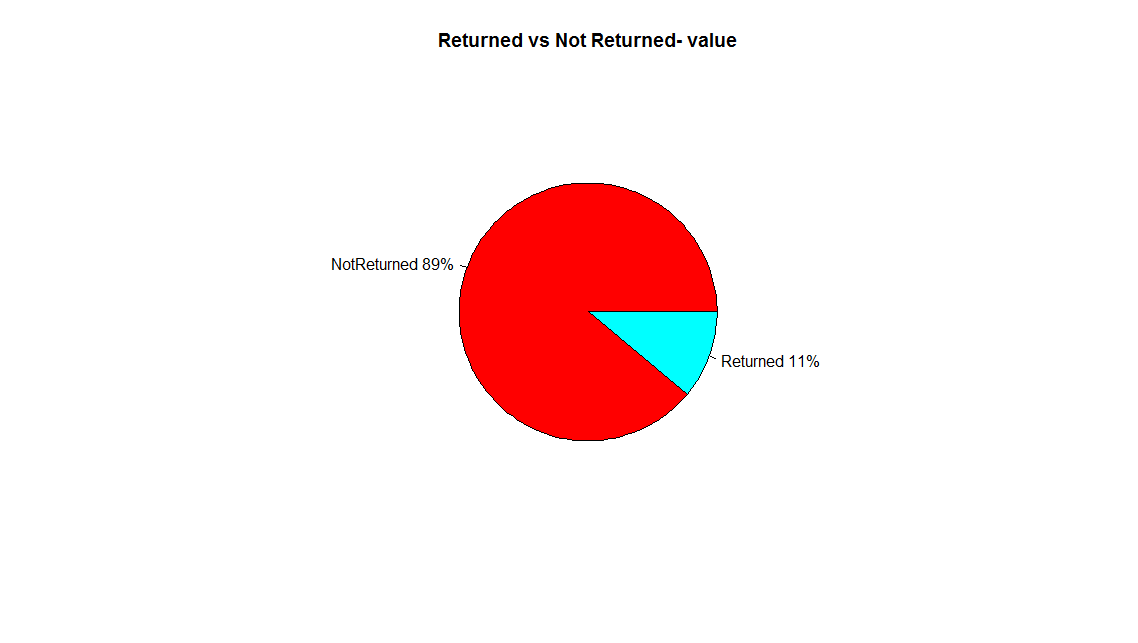
pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct)

lbls <- paste(lbls,"%",sep="")

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Returned vs Not Returned- value")



# Correlation

# Create numerical value for Status

total$statusnum[total$Status=="Returned"] <- 1

total$statusnum[total$Status=="NotReturned"] <- 0

cor(total$Salesnum,total$statusnum )

> cor(total$Salesnum,total$statusnum )

[1] 0.01157299

cor.test(total$Salesnum,total$statusnum )$p.value

> cor.test(total$Salesnum,total$statusnum )$p.value

[1] 0.2889189

# Insight

1. 10% of the total count of goods sold is returned
2. 11% of the goods is returned by value.
3. The correlation between higher sales value of goods and returns is very low (.011) . Since the p value of the correlation is high (.22) we cannot conclude that this value of correlation is not by coincidence . Therefore all evidence points to no correlation between value of goods and returns