#The following is the savings account transactions downloaded from HDFCbank

card=read.csv("Debitstat.csv")

str(“card”)

'data.frame': 900 obs. of 8 variables:

$ Date : Factor w/ 582 levels ""," Date ",..: 537 50 135 312 312 356 507 542 25 25 ...

$ Narration : Factor w/ 495 levels "","#NAME?","00600350111791-HDFC TAXSAVER - DIVIDEND ",..: 196 180 406 38 257 180 291 196 216 220 ...

$ Value.Dat : Factor w/ 610 levels ""," HYD ",..: 453 248 561 190 209 258 408 458 458 56 ...

$ Debit.Amount : Factor w/ 341 levels "","-8000","0",..: 256 337 139 3 264 176 3 260 3 14 ...

$ Credit.Amount : Factor w/ 197 levels "","0","1","100",..: 2 2 2 86 2 2 116 2 185 2 ...

$ Chq.Ref.Number : Factor w/ 633 levels "","0","00000007970571A1 ",..: 192 569 75 2 553 471 66 200 2 139 ...

$ Closing.Balance: Factor w/ 861 levels "","-12572.35",..: 688 601 382 702 697 682 118 108 111 86 ...

$ X : num NA NA NA NA NA NA NA NA NA NA ...

#change the format and add additional columns

card$Debit.Amount=as.numeric(as.character(card$Debit.Amount))

card$Credit.Amount=as.numeric(as.character(card$Credit.Amount))

card$Date = as.Date(card$Date, format="%d/%m/%y")

card$day=format(card$Date,format="%d")

card$month=format(card$Date,format="%m")

card$year=format(card$Date,format="%y")

summary(card$Credit.Amount)

#with the below plot , you can observe three kinds of transactions(Bonus , salary, shift allowances)

plot(HCL\_sal$Credit.Amount)

#subset transactions where credit bill has beenpaid

ccard1=grep("CITIBANK CREDIT CARD",card$Narration)

ccard2=grep("AUTOPAY",card$Narration)

ccard3=grep("1540",card$Narration)

ccard4=grep("8142",card$Narration)

ccard=c(ccard1,ccard2,ccard3,ccard4)

ccard\_indx=sort(ccard)

credit=card[ccard\_indx,]

credit=unique(credit)

##aggregate more than one payment in a month

card\_ts=aggregate(Debit.Amount~month+year, data=credit, FUN=sum)

#plot the data , the trend is interesting

plot(card\_ts$Debit.Amount)

##set the missing values to zero and attach to time series

cts <- merge(expand.grid(year=unique(card\_ts$year),month=unique(card\_ts$month)),card\_ts,all=TRUE)

cts$ym = paste(cts$year, cts$month,sep="")

cts= cts[order(cts$ym),]

#Replace NAs with zero and attach to time series

cts[is.na(cts)]=0

card\_fr= cts$Debit.Amount

card\_fr <- ts(card\_fr, frequency = 12,start = c(2011, 1))

plot(card\_fr)

#fit into a time series model using <arima>

fit <- arima(card\_fr, order = c(1, 0, 0), list(order = c(2,1, 0), period = 12))

#forecast ahead of 24 months

fore <- predict(fit, n.ahead = 24)

U <- fore$pred + 2 \* fore$se #upper limit as 2 standard deviations

L <- fore$pred - 2 \* fore$se #lower limit as 2 standard deviations

ts.plot(card\_fr, fore$pred, U, L, col = c(1, 2, 4, 4), lty = c(1, 1, 2, 2))

\*###########################################\*

#working with H C L salary

\*###########################################\*

#subset transactions where HCL salary is paid

HCL\_sal\_indx=grep("HCL",card$Narration)

HCL\_sal=card[HCL\_sal\_indx,]

HCL\_sal\_indx=grep("NEFT", HCL\_sal$Narration)

HCL\_sal= HCL\_sal[HCL\_sal\_indx,]

ind=which(HCL\_sal$Credit.Amount>10000)

HCL\_sal= HCL\_sal[ind,]

#plot salary data

plot(HCL\_sal$Credit.Amount)