Churn data modeling Understanding categorical variables in churn data

Lecture 3 (Practical)

19/07/2021

Exploring categorical variables

- Identify the patterns in the data that will help us to reduce the proportion of churners.

##Read in the churn dataset

- churn<- read.csv(file.choose())
- names(churn)

```
 [1] "State"          "Account.Length" "Area.Code"      "Phone"
 [5] "Int.l..Plan"    "VMail.Plan"     "VMail.Message"  "Day.Mins"
 [9] "Day.Calls"      "Day.Charge"     "Eve.Mins"       "Eve.Calls"
[13] "Eve.Charge"     "Night.Mins"     "Night.Calls"
"Night.Charge"
[17] "Intl.Mins"      "Intl.Calls"     "Intl.Charge"    "CustServ.Calls"
[21] "Churn"
```

```
churn[1:10,]
# will give first 10 observations in the dataset.
###summarize the churn variable####
sum.churn<-summary(churn$Churn)
sum.churn
      False.  True.
      2850    483
# calculate proportion of churners
tab1<-table(churn$Churn)
prop.table(tab1)
        False.     True.
   0.8550855 0.1449145
```

That is 14.49% of the customers have left the company

- Make a table for counts of Churn and International Plan.

```
cnts<-
table(churn$Churn,churn$Int.l..Plan,dnn=c("Churn","International
Plan"))
cnts
```

```
                International Plan
Churn             no  yes
  False.        2664  186
  True.          346  137
```

- create a table with sums for both variables

```
sumtable<-addmargins(cnts,FUN=sum)
sumtable
```

```
                International Plan
Churn             no  yes  sum
  False.        2664  186 2850
  True.          346  137  483
  sum           3010  323 3333
```

- create a table of proportions over rows

row.margin<-round(prop.table(cnts,margin=1),4)*100

row.margin

Churn	International Plan no	yes
False.	93.47	6.53
True.	71.64	28.36

28.4% of the churners belong to the International Plan compared to 6.5% of non-churners

- create a table of proportions over columns
- col.margin<-round(prop.table(cnts,margin=2),4)*100
- col.margin

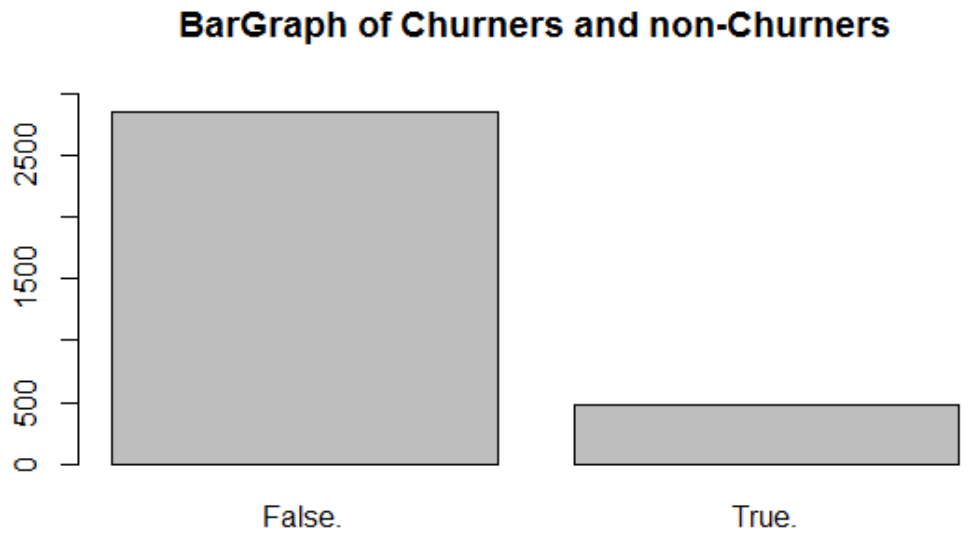
Churn	International Plan no	yes
False.	88.50	57.59
True.	11.50	42.41

42.4% of International Plan holders churned as compared to only 11.5% of those without International Plan.

Barchart of variable Churn

```
barplot(sum.churn, ylim= c(0,3000), main= "BarGraph of Churners and non-Churners")
```

```
False.  True.  Total.
  2850    483   3333
```

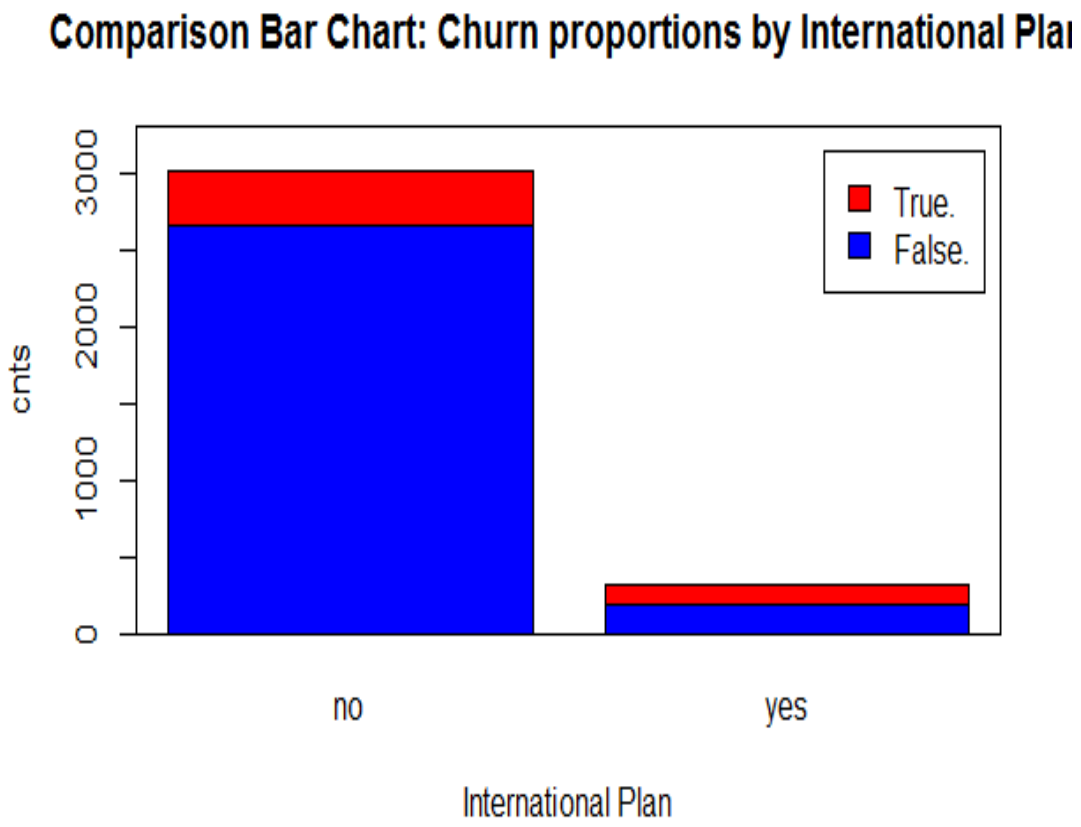


Overlayed barchart

- barplot(cnts,legend=rownames(cnts), col=c("blue","red"), ylim=c(0,3300),ylab="cnts",xlab="International Plan", main= "Comparison Bar Chart: Churn proportions by International Plan")
- box(which="plot",lty="solid",col="black")

Contingency Table of International Plan with Churn

	International Plan	
Churn	no	yes
sum False.	2664	186
2850 True.	346	137
483 sum	3010	323
3333		



Clustered bar chart with legend

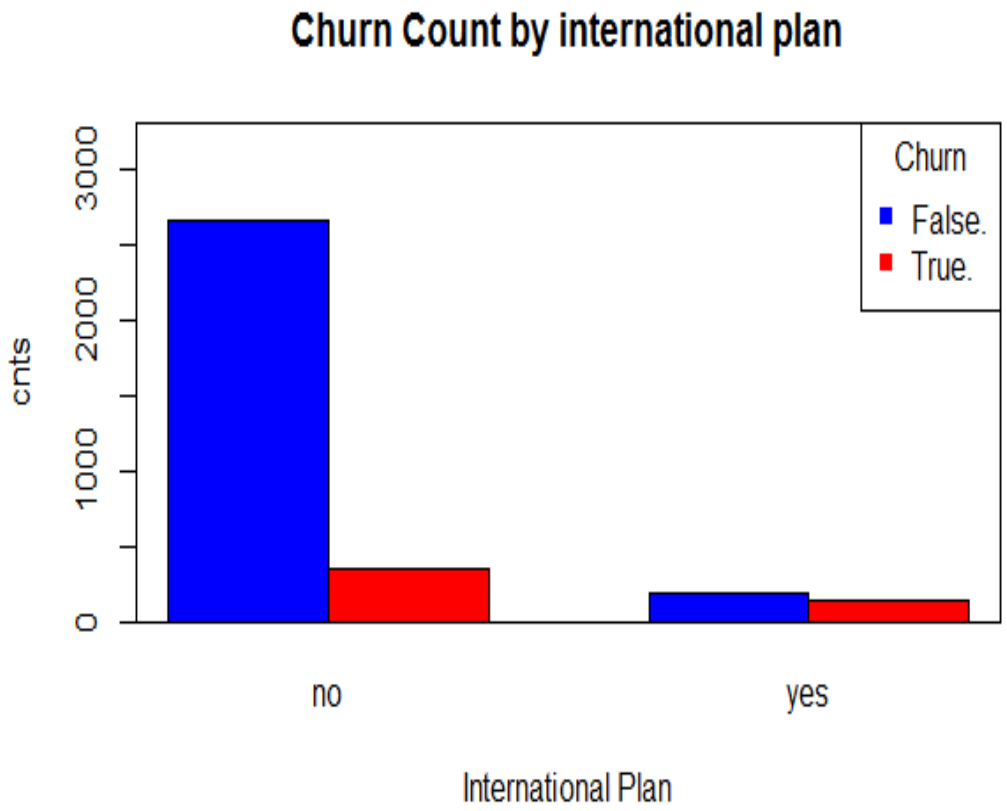
- barplot(cnts,col=c("blue","red"), ylim=c(0,3300),ylab="cnts",xlab="International Plan", main= "Churn Count by international plan",beside=TRUE)
- legend("topright",c(rownames(cnts)),col=c("blue","red"),pch = 15, title= "Churn")
- box(which= "plot",lty= "solid", col="black")

Clustered Bar Chart

The first set of two bars represents those who do not belonging to the International plan.

The second set of two bars represents those who belongs to the International plan.

Clearly proportion of churners is greater among those belonging to the plan.



Clustered barchart of Churn and International Plan with legend

- barplot(t(cnts),col=c("blue","green"), ylim=c(0,3300),ylab="cnts",xlab="Churn",

  main= "International plan Count by Churn”,

  beside=TRUE)
- legend("topright",c(rownames(cnts)),col=c("blue","green"),pch = 15, title= "International Plan")
- box(which= "plot",lty= "solid", col="black")

Clustered Bar Chart

The first set of bars represents non-churners .

Second set of bars represents churners.

Clearly proportion of International Plan holders is greater among the churners.