**#####################**

**# REGLES D'ASSOCIACIÓ #**

**#####################**

**### 1. Importació de la base de dades ###**

**#install.packages("arules")**

**library(arules)**

**dd <- read.table("Booking\_data\_preprocessada.csv",header=T, sep=",");**

**dcat<-dd[,sapply(dd, is.factor)] # Selecció de les variables categòriques**

**dtrans<-as(dcat, "transactions")# Transformación a transacciones**

**### 2. Itemsets més freqüents ###**

**foo<-function(x){length(levels(x))}**

**sum(sapply(dcat, foo)) # suma de les categories que hi ha en tots els factors de la base de dades.**

**## 2.1. Image de dtrans ##**

**dtrans**

**inspect(head(dtrans,10))**

**summary(dtrans)**

**dim(dtrans)**

**image(dtrans)**

**dcat$Review\_Date\_Year <- as.factor(substring(dcat$Review\_Date,1,4)) # Deixem només l'any**

**# Eliminació de camps no necessaris**

**dcat$Review\_Date <- NULL**

**dcat$Hotel\_Name <- NULL**

**dtrans<-as(dcat, "transactions") # Transformació a transaccions**

**dtrans**

**summary(dtrans)**

**dim(dtrans)**

**image(dtrans[sample(1:length(dtrans),100)])**

**levels(dcat$Hotel\_Country)[4]**

**levels(dcat$Hotel\_City)[4]**

**summary(dtrans)**

**itemFrequencyPlot(dtrans, support=0.3, cex.names = 1, col= 'lightsteelblue')**

**## 2.2. Algorisme Eclat ##**

**eclatdtrans<-eclat(dtrans, parameter = list(support=0.4, minlen=1, maxlen=5))**

**inspect(head(eclatdtrans,n=14, by="support"))**

**library(arulesViz)**

**plot(eclatdtrans, method="graph") #tots els itemsets**

**plot(head(eclatdtrans,n=14, by="support"), method="graph")**

**### 3. Regles d'Associació (algorisme a priori). ###**

**## Confidence ##**

**rulesdtrans <- apriori(dtrans, parameter = list(sup = 0.02, conf = 0.05, maxlen=5))**

**inspect(head(rulesdtrans,n=15, by="confidence"))**

**## Lift ##**

**rulesdtrans <- apriori(dtrans, parameter = list(sup = 0.01, conf = 0.05, maxlen=5))**

**inspect(head(rulesdtrans,n=15, by="lift"))**