

05.584 · Mineria de Dades · PRACTICA · 2017-2018

EEES · Estudis d'Informàtica, Multimèdia i Telecomunicació

Nom i Cognoms: Albert Costas Gutiérrez

Índex de continguts

| | |
|---|---|
| Detall composició del paquet..... | 2 |
| Definició de la practica..... | 3 |
| Anàlisis per volum de ventes..... | 4 |
| Anàlisis per import total de les compres..... | 4 |
| Anàlisis per hora i dia de la setmana..... | 4 |
| Model d'arbres de decisió..... | 4 |
| Model de regles d'associació..... | 4 |
| Accions per Gourmet..... | 4 |
| Bibliografia..... | 6 |

Detall composició del paquet

```

zip
├── R # anàlisis i models de la practica
│   ├── analisis_data_o_hora_del_dia.R
│   ├── analisis_volum_de_vendes.R
│   ├── analisis_per_import_total.R
│   ├── model_agregation_rules.R
│   └── model_decision_tree.rpart.R
├── images # directori on hi ha imatges generades amb R
├── csv # datasets per cada script .R
│   ├── analisis_data_o_hora_del_dia.csv
│   ├── analisis_volum_de_vendes.csv
│   ├── analisis_per_import_total.csv
│   ├── model_agregation_rules.csv
│   └── model_decision_tree.rpart.csv
├── SQL # sentencies sql les qual s'han creat els csv
│   ├── analisis_data_o_hora_del_dia.sql
│   ├── analisis_volum_de_vendes.sql
│   ├── analisis_per_import_total.sql
│   ├── model_agregation_rules.sql
│   └── model_decision_tree.rpart.sql
├── mysql # script per crear la base de dades, requereixen del csv de la practica
│   ├── GourmetDB.db2
│   └── GourmetDB.sql # adaptar per a mysql, importa els csv original de la practica
├── Index_documents_practica.txt # index de la practica
└── acostasg_MD_Practica.pdf # pdf explicació practica i les respostes amb gràfics per al client.
  
```

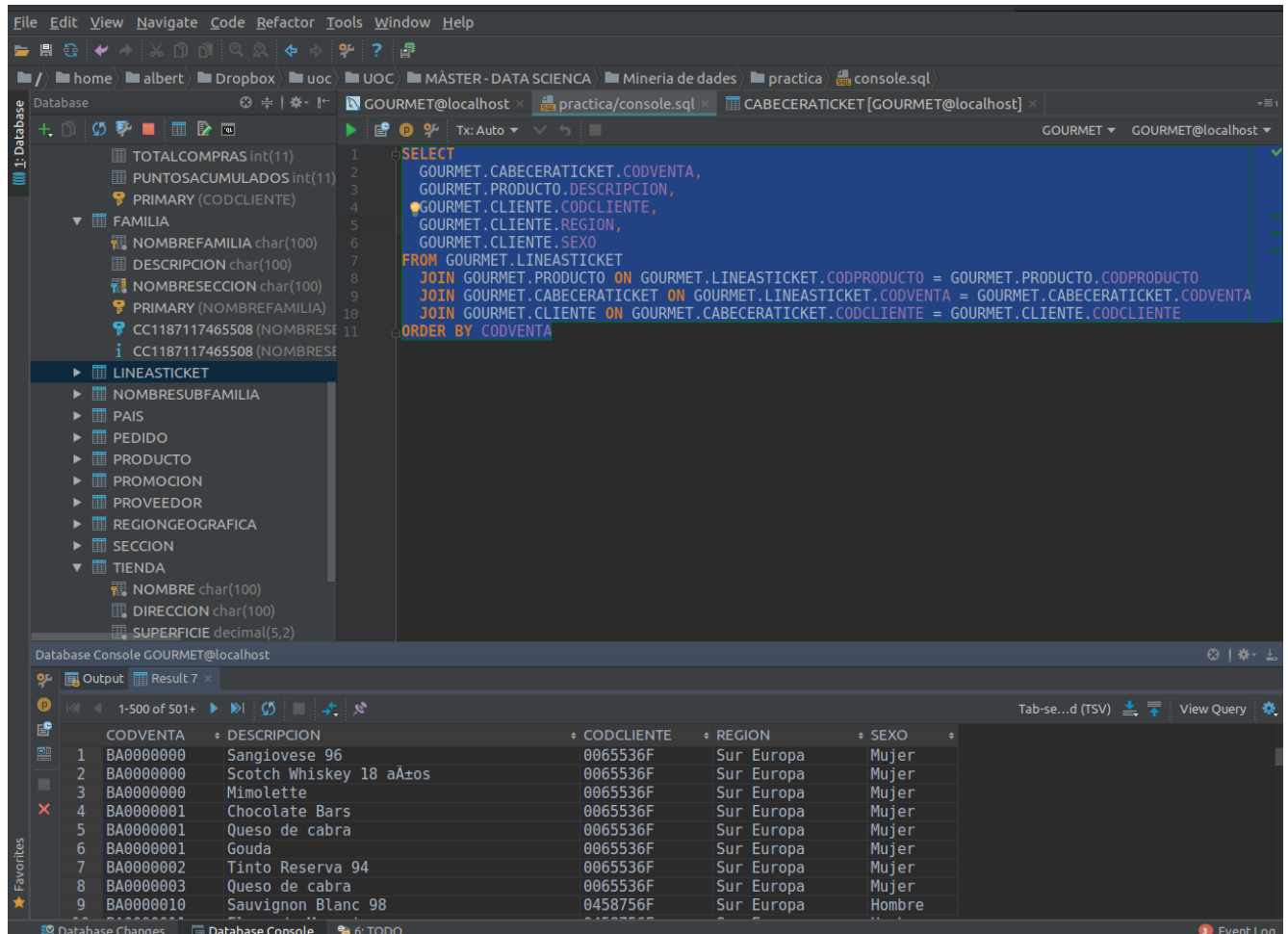
Definició de la practica

Per la realització de la pràctica i respondre a les preguntes que ens proposa Joan Schelemmer, s'ha realitzat una anàlisi (des de diferents punts de vista), neteja i transformació de dades, per arribar a realitzar els 2 models, un basat en un arbre de decisió, els quals no hems a donat bons resultats, i un molt més idoni per a grups de dades de vendes i productes, regles d'associació.

Per entendre el domini de les dades, hem instal·lat una base de dades mysql en local per desenvolupament, i executat l'script modificat per l'ocasió, està en la carpeta msqy/GourmetDB.sql



A més, a la carpeta SQL/*.sql hi han les sentències sql que ens han permès transformar les dades per utilitzar per als diferents script .R que hi ha a la carpeta R/*.R.



The screenshot shows a database console interface with a SQL query and its results. The query is a SELECT statement joining several tables: GOURMET.CABECERATICKET, GOURMET.PRODUCTO, GOURMET.CLIENTE, and GOURMET.LINEASTICKET. The results are displayed in a table with columns: CODVENTA, DESCRIPCION, CODCLIENTE, REGION, and SEXO.

```

SELECT
  GOURMET.CABECERATICKET.CODVENTA,
  GOURMET.PRODUCTO.DESCRIPCION,
  GOURMET.CLIENTE.CODCLIENTE,
  GOURMET.CLIENTE.REGION,
  GOURMET.CLIENTE.SEXO
FROM GOURMET.LINEASTICKET
JOIN GOURMET.PRODUCTO ON GOURMET.LINEASTICKET.CODPRODUCTO = GOURMET.PRODUCTO.CODPRODUCTO
JOIN GOURMET.CABECERATICKET ON GOURMET.LINEASTICKET.CODVENTA = GOURMET.CABECERATICKET.CODVENTA
JOIN GOURMET.CLIENTE ON GOURMET.CABECERATICKET.CODCLIENTE = GOURMET.CLIENTE.CODCLIENTE
ORDER BY CODVENTA
  
```

| | CODVENTA | DESCRIPCION | CODCLIENTE | REGION | SEXO |
|---|-----------|------------------------|------------|------------|--------|
| 1 | BA0000000 | Sangiovese 96 | 0065536F | Sur Europa | Mujer |
| 2 | BA0000000 | Scotch Whiskey 18 años | 0065536F | Sur Europa | Mujer |
| 3 | BA0000000 | Mimolette | 0065536F | Sur Europa | Mujer |
| 4 | BA0000001 | Chocolate Bars | 0065536F | Sur Europa | Mujer |
| 5 | BA0000001 | Queso de cabra | 0065536F | Sur Europa | Mujer |
| 6 | BA0000001 | Gouda | 0065536F | Sur Europa | Mujer |
| 7 | BA0000002 | Tinto Reserva 94 | 0065536F | Sur Europa | Mujer |
| 8 | BA0000003 | Queso de cabra | 0065536F | Sur Europa | Mujer |
| 9 | BA0000010 | Sauvignon Blanc 98 | 0458756F | Sur Europa | Hombre |

Anàlisis de les dades

Per tant, la visualització de les taules i les seves relacions: clients, proveïdors, botigues,... ens ha permès entendre el domini de les dades amb les seves dimensions, i veure els diferents atributs que utilitzarem durant l'anàlisi i els models que hem creat.

En aquest sentit, hem enfocat l'estudi de les botigues Gourmet per poder realitzar una publicitat més personalitzada. Un estudi sobre les dades pel seu volum de venda, per import d'aquestes, i en quines hores o dies es produeixen per regió, tipus de clientes per sexe, professió...

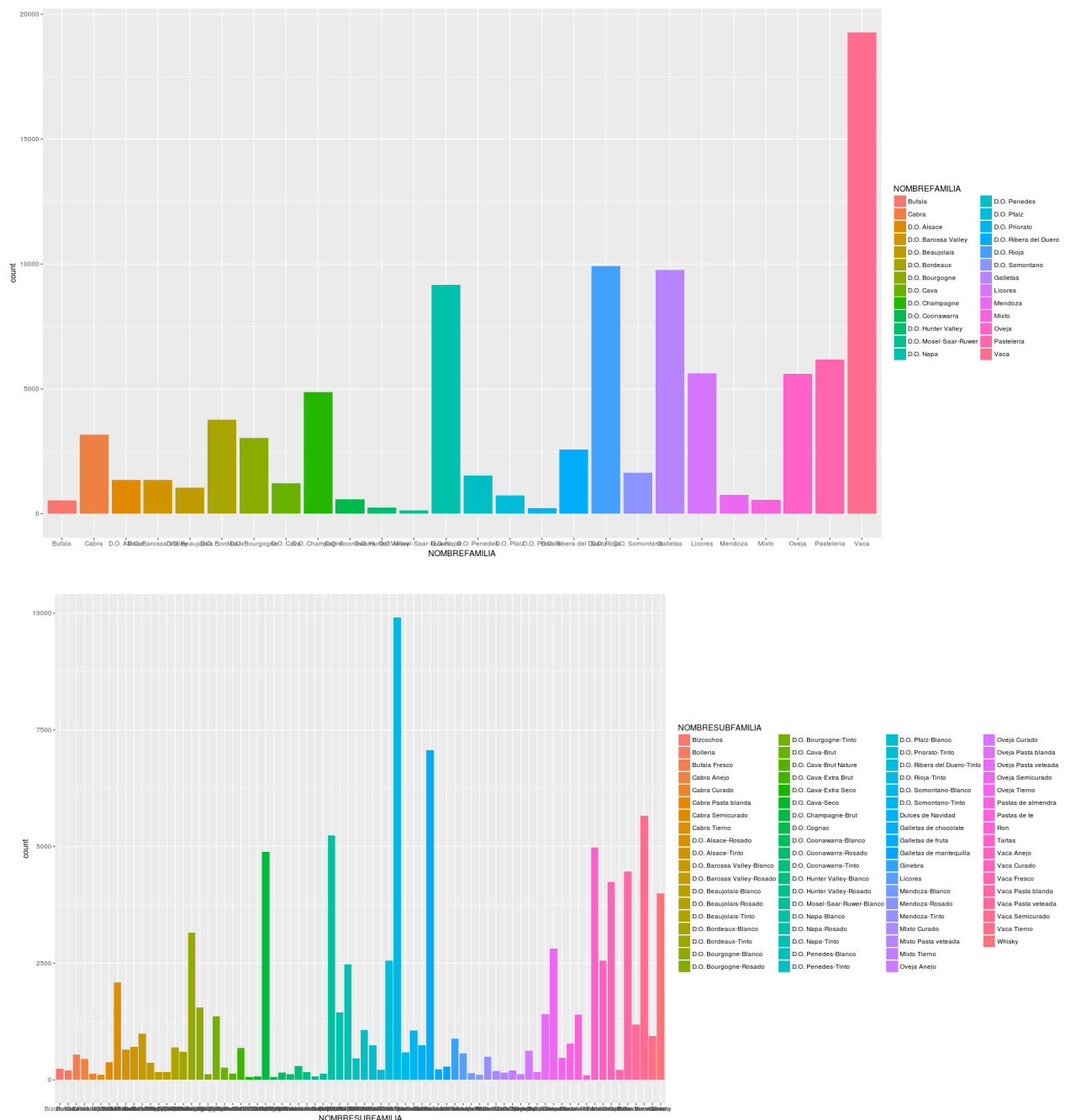
Per altra banda, en cada script on import el csv per l'anàlisi, on ja hem fet la

primera transformació amb al sentència SQL, adaptem les dades a valors qualitius i netegem de caràcters en blanc i estranys per un anàlisis més òptim.

Anàlisis per volum de ventes

Iniciam l'anàlisi important el dataset i normalitzem les dades, posteriorment reordenem les dades aleatòriament, i generem les gràfiques:

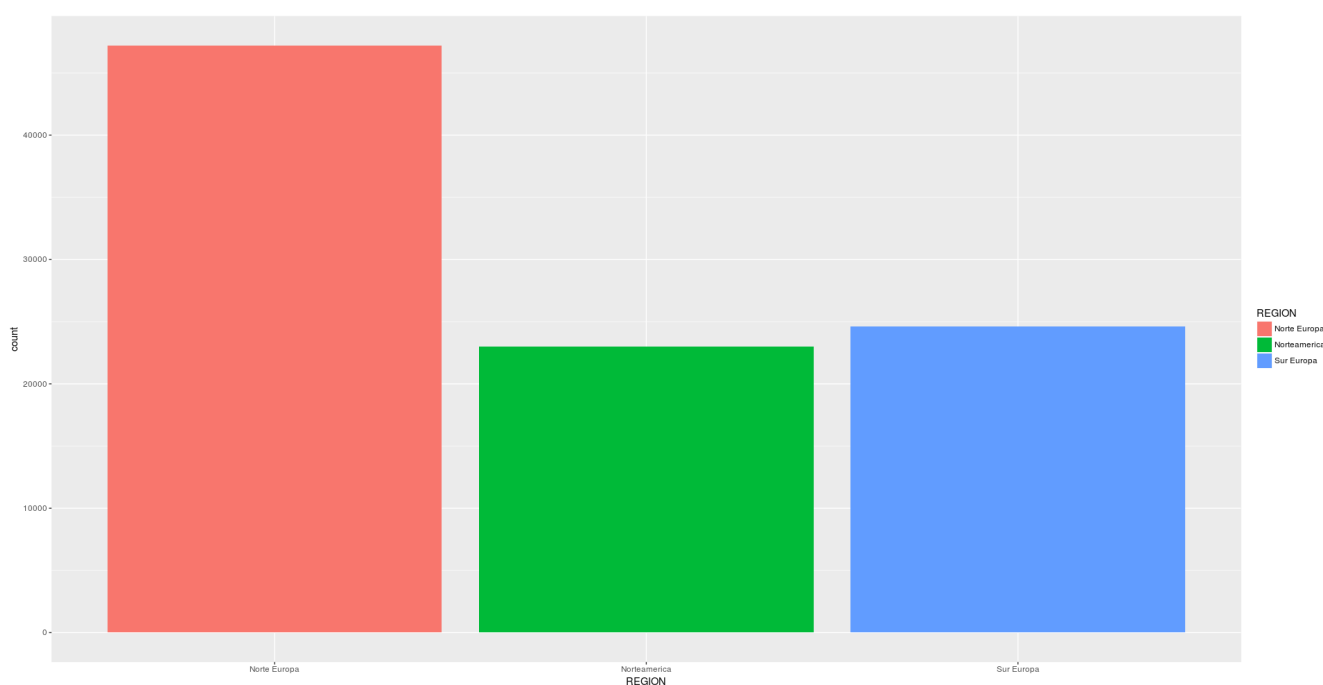
Volum de ventes per FAMÍLIA i SUBFAMÍLIES:



Per tant a nivell de volum de ventes (de freqüència), en nivell de **família** la és tracta de la **Vaca (derivats de vaca)**, per contra les **subfamílies** les que tenen un volum més alta de venta es **D.O Rioja-Tinto** i **Galletes de xocolates**, per tant les subfamílies i família més populars entre els totes les botigues del grup.

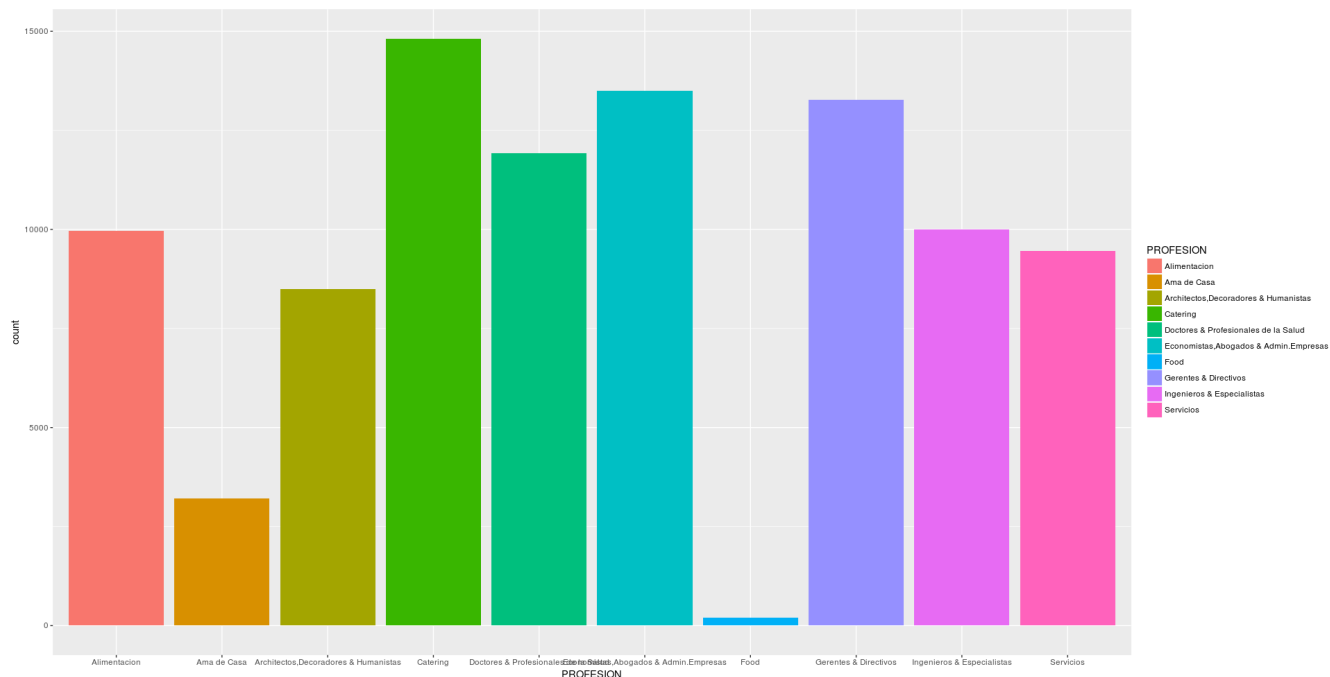
Volum de ventes per REGIÓ:

Un valor important, en cada regió la publicitat s'haurà d'orientat el seu consum:



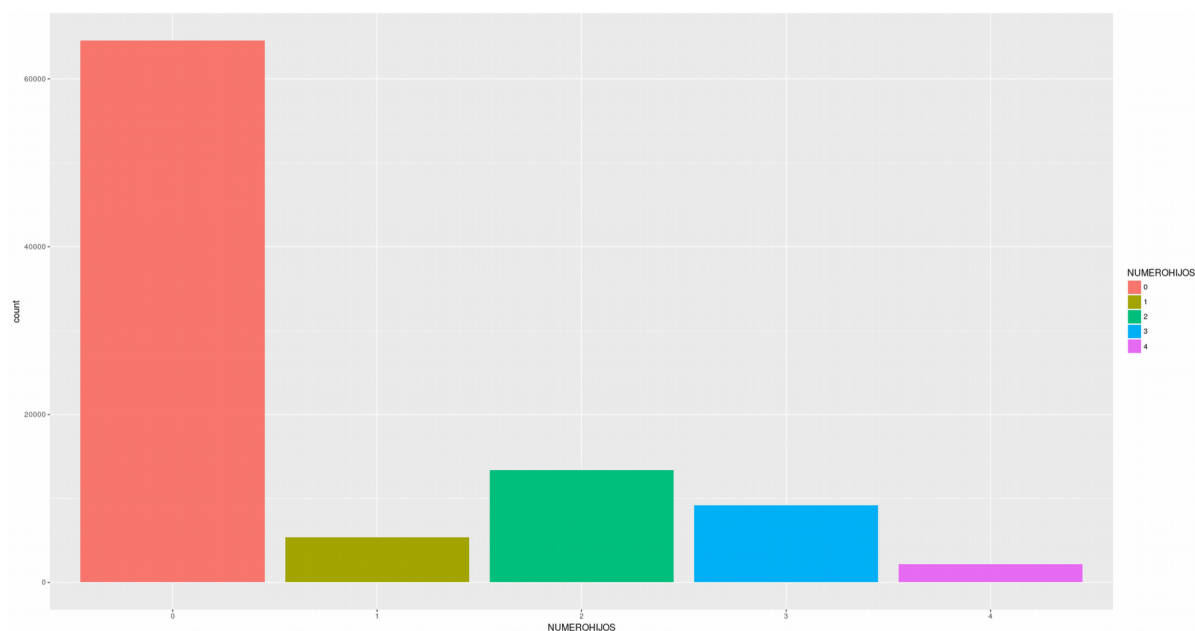
Podem veure que en **Nort Europa, es on més volum de compres es produeix**, posteriorment veure si es un més facturació hi ha, en tot cas, per augmentar o disminuir el preu del productes seria un factor a tenir present, pel que fa Nord Americà i Surt Europa el volum de compres es gairebé el mateix.

Volum de ventes per PROFESSIÓ:

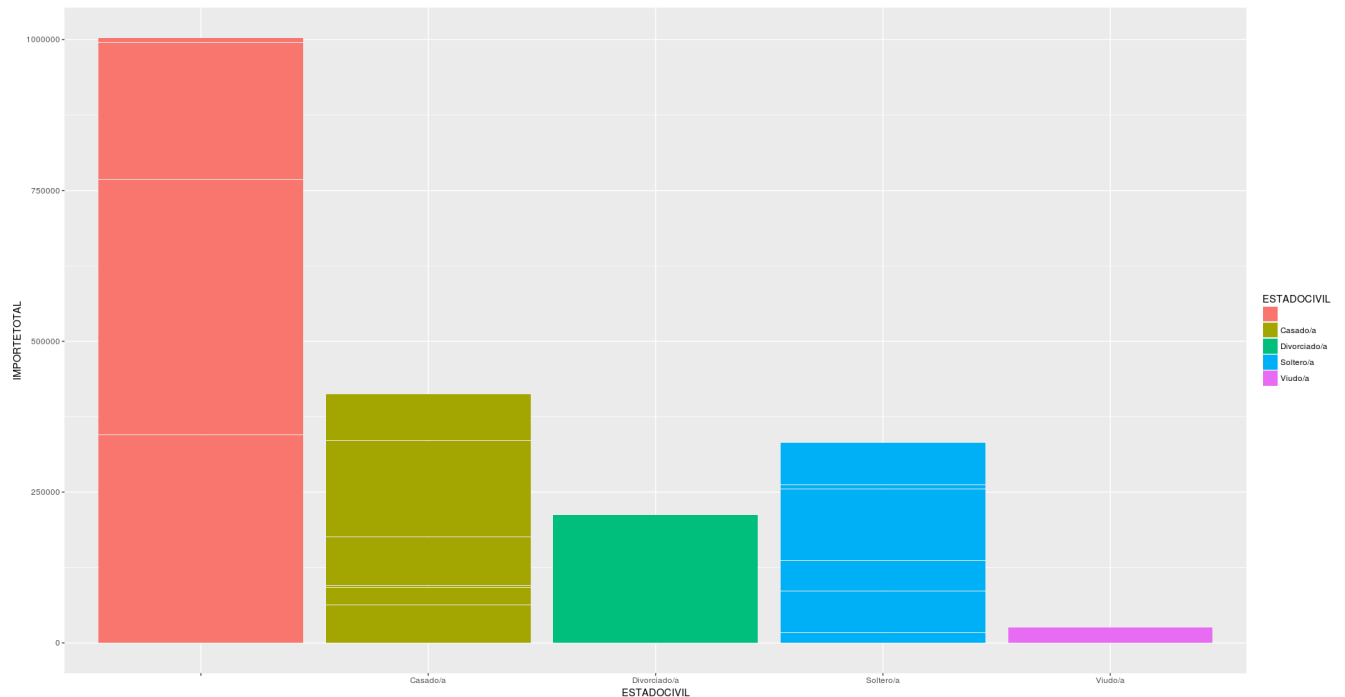


Per contra, la majoria de professions està igual, el valor **està en la diferencia**, i podem observar que hi ha dos professions que no consumeixen en les botigues, **s'hauria de realitzar alguna oferta o campanya per augmentar en sector de Food (menja) i de mestresses de casa.**

Volum de ventes per NOMBRE DE FILLS i ESTAT CIVIL:

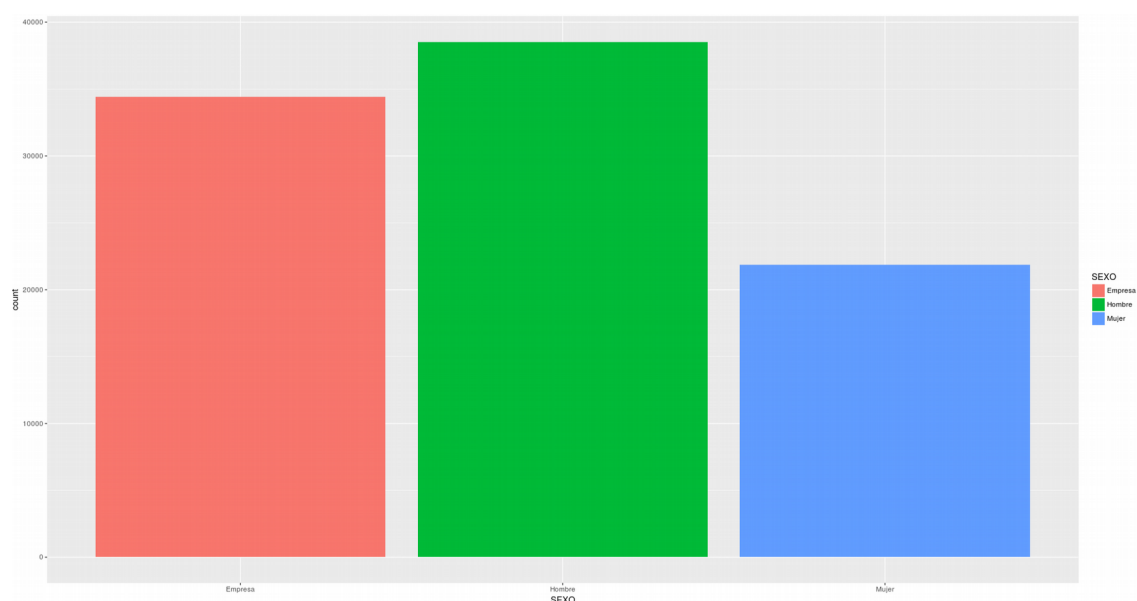


Clarament tenim un gran marge a millor, **la majoria de compres són realitzades per clients que no tenen fills**, si les dades son correctes, **hauríem de realitzar campanyes i publicitat per a les famílies.**



No podem fer un anàlisi, degut que tenim un gran volum de dades no te aquest valor, no podem valorar per estat civil, si podem veure mes endavant quin import tenen les compres que disposem d'aquesta dada.

Volum de ventes per SEXE DELS CLIENTS:



No ens aporta massa valor, podem veure que està força igualat, potser **hauríem de motivar la compra per part de les dones, potser hauríem de revisar quins productes estan orientat aquest sexe i veure s'hi han de millorar o publicitar.**

Volum de ventes per PRODUCTES:

Valor importat on podem usar per a la publicitat, o per combinar amb productes que no tenen tanta sortida al mercat:

El productes mes venuts:

| | |
|--------------------------|-------------|
| Chocolate Truffle | 4848 |
| Tiramisu | 4558 |
| Tinto Reserva 95 | 3880 |
| Camembert | 3424 |

Els productes menys venuts:

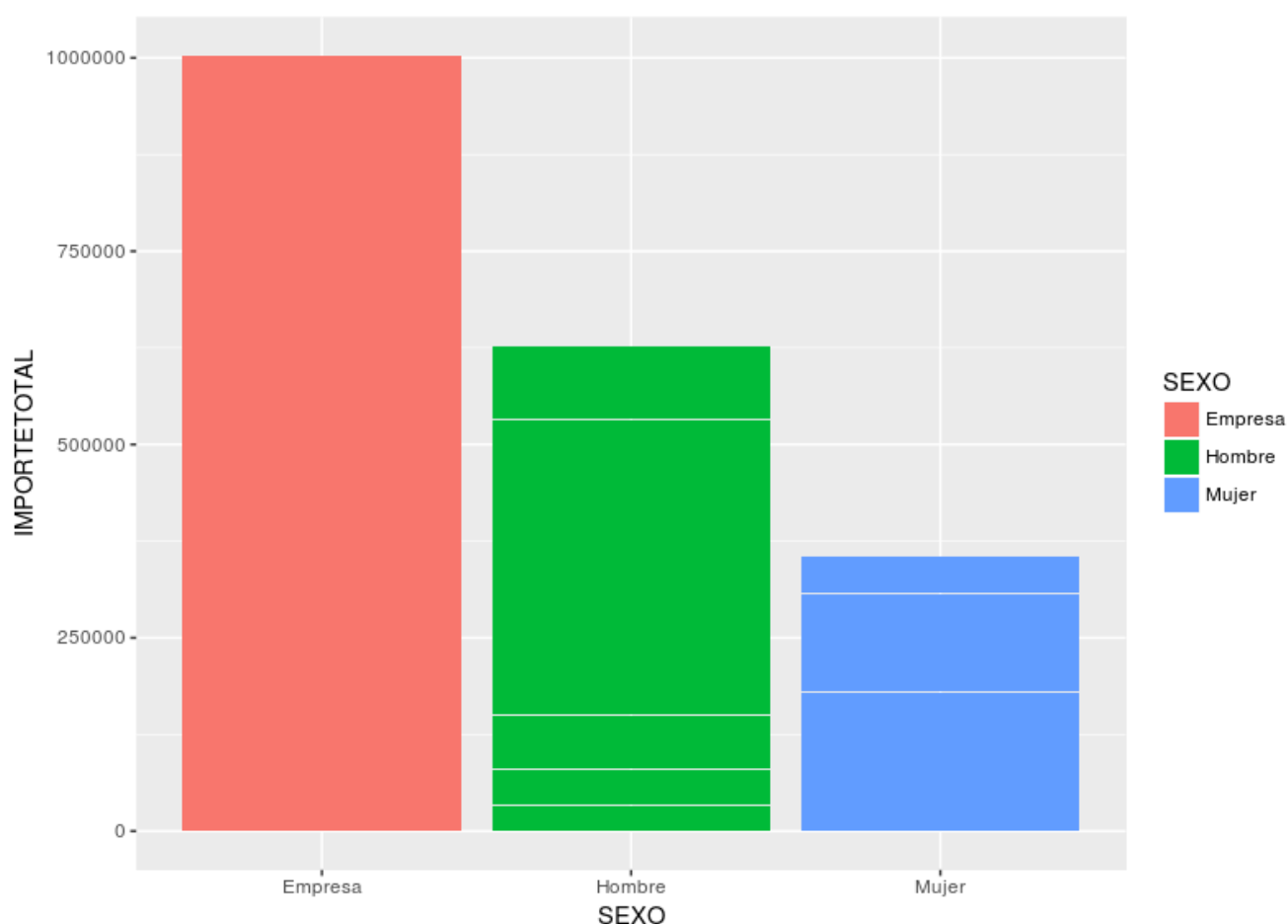
| | |
|--------------------------------------|------------|
| Cabernet Sauvignon Reserva 93 | 168 |
| Reblochon | 168 |
| White Stilton | 168 |
| Riesling 96 | 167 |
| Shiraz 97 | 167 |
| Tinto Crianza 94 | 167 |
| Brinata | 166 |
| Double Gloucester | 166 |

Per tant aquí tenim una combinació per promoció o publicitat, om podem combinar productes.

Anàlisis per import total de les compres

Amb aquest anàlisis, volem veure, si el volum de transaccions que em vist anteriorment, es materialitza amb import de les transaccions, tot i que no farem de tots els atributs dels clients, si veurem alguns, així com depenen la forma de pagament quina es la que s'utilitza per import alts o baixos.

Import de compres per SEXE DEL CLIENT:

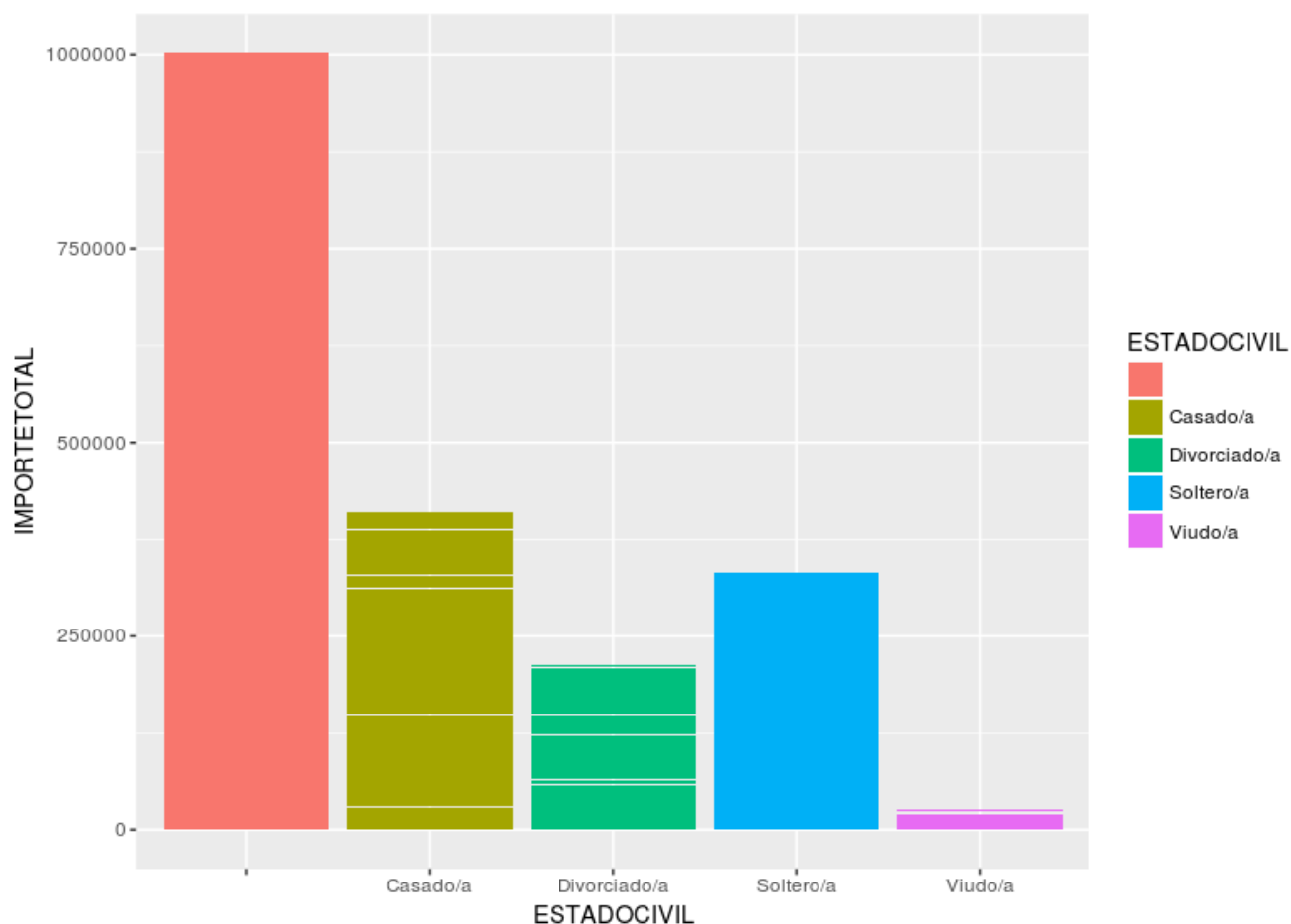


Sabem que volum de compres estava igualat, per contra, el import d'aquestes compres es molt més elevat en les empreses, després els homes i finalment les dones.

Per tant, hem de mantenir aquest import en les empreses, em de continuar amb les promocions per empreses, i en tot cas veure si podem augmentar amb les economia familiar.

Import de compres per ESTAT CIVIL:

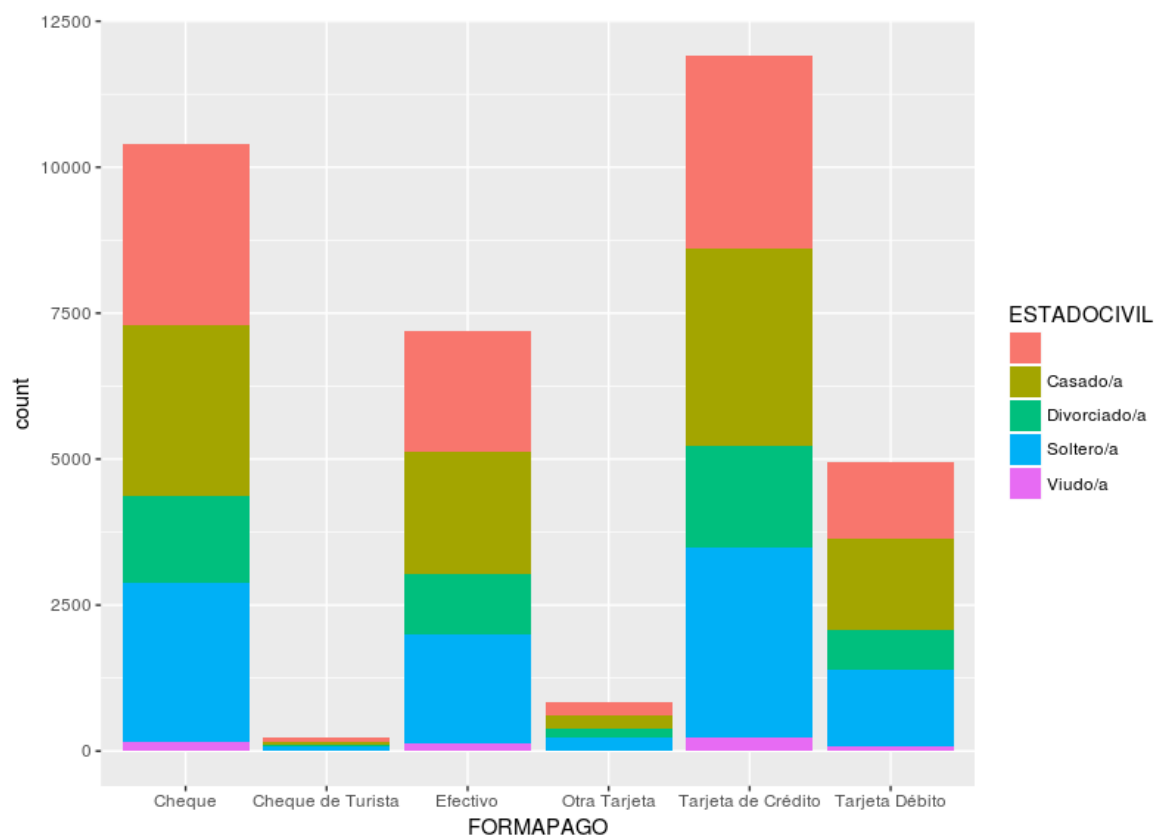
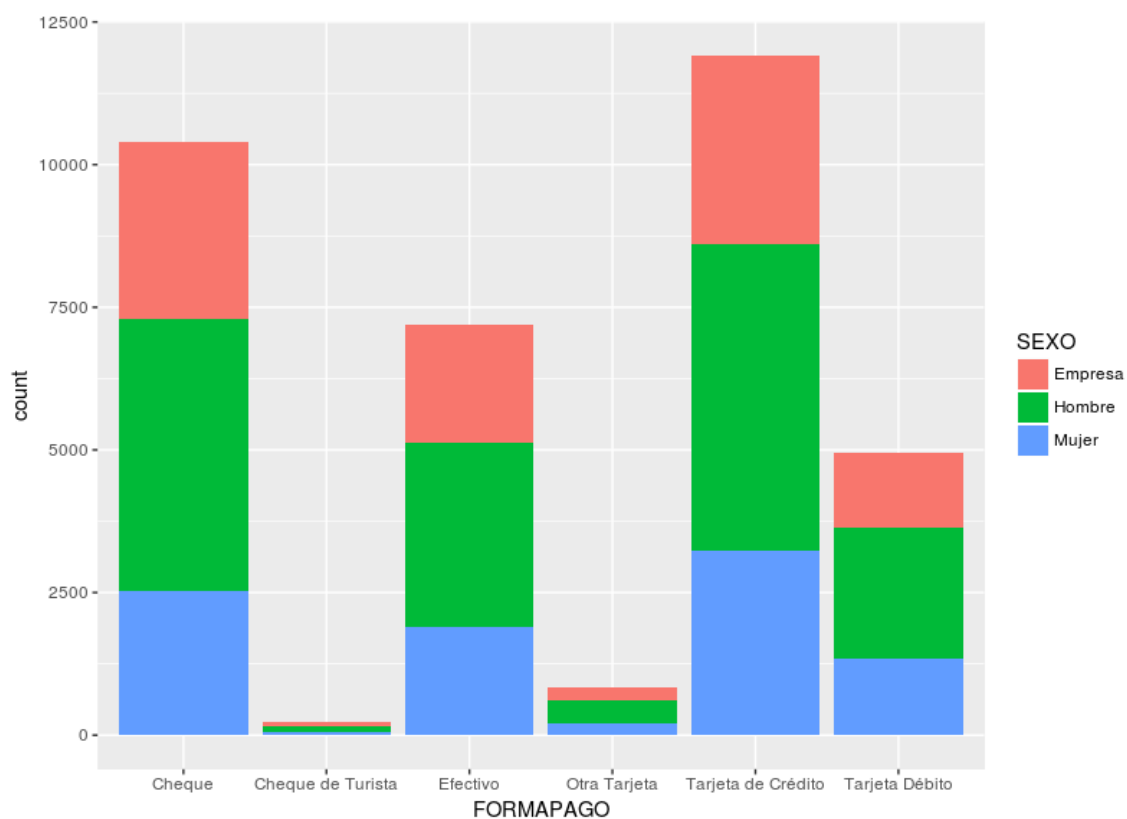
Tot i que em vist que no disposem de les dades completes, per veure el volum de les compres, si podem veure de les que tenim aquelles que gestionen un import més alt, de les dades que disposem:



De les dades que disposem, veiem que gestiona **més import el clients casats**, clarament perquè son famílies, per contra em vist abans que el volum de compres estava en amb els clientes que no tenen fills.

La conclusió es la mateixa que hem arribat abans, **hem de publicitar per famílies, per que tenen un volum de compres més alt, i perquè en freqüència està per sota de clients sense fills.**

Import de compres per PAGAMENT UTILITZAT:



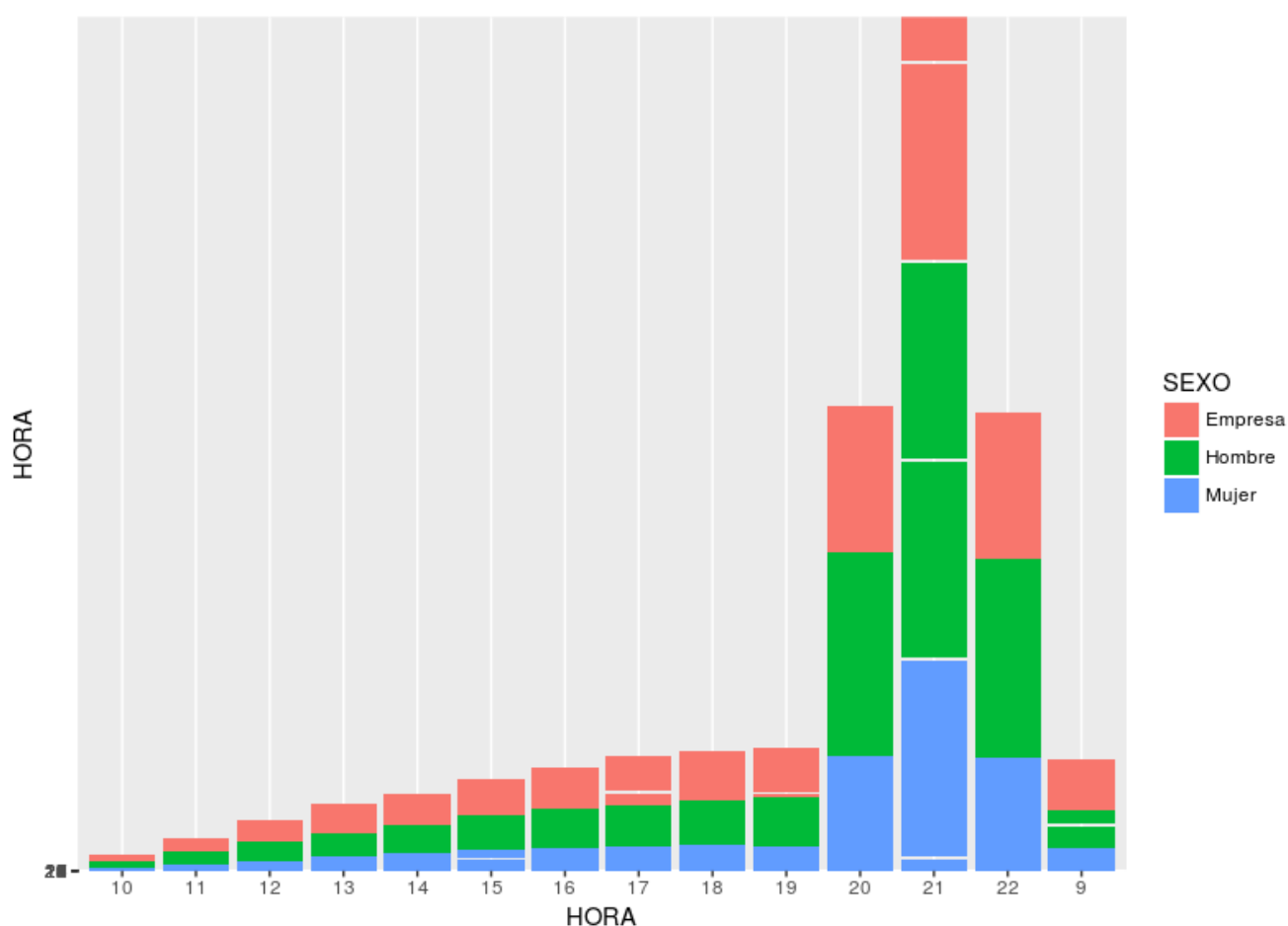
Les formes de pagament que més import gestiona es la targeta de crèdit, tant per empreses, homes i dones, l'altra és els xecs, indistintament pel sexe del client, o per l'estat civil.

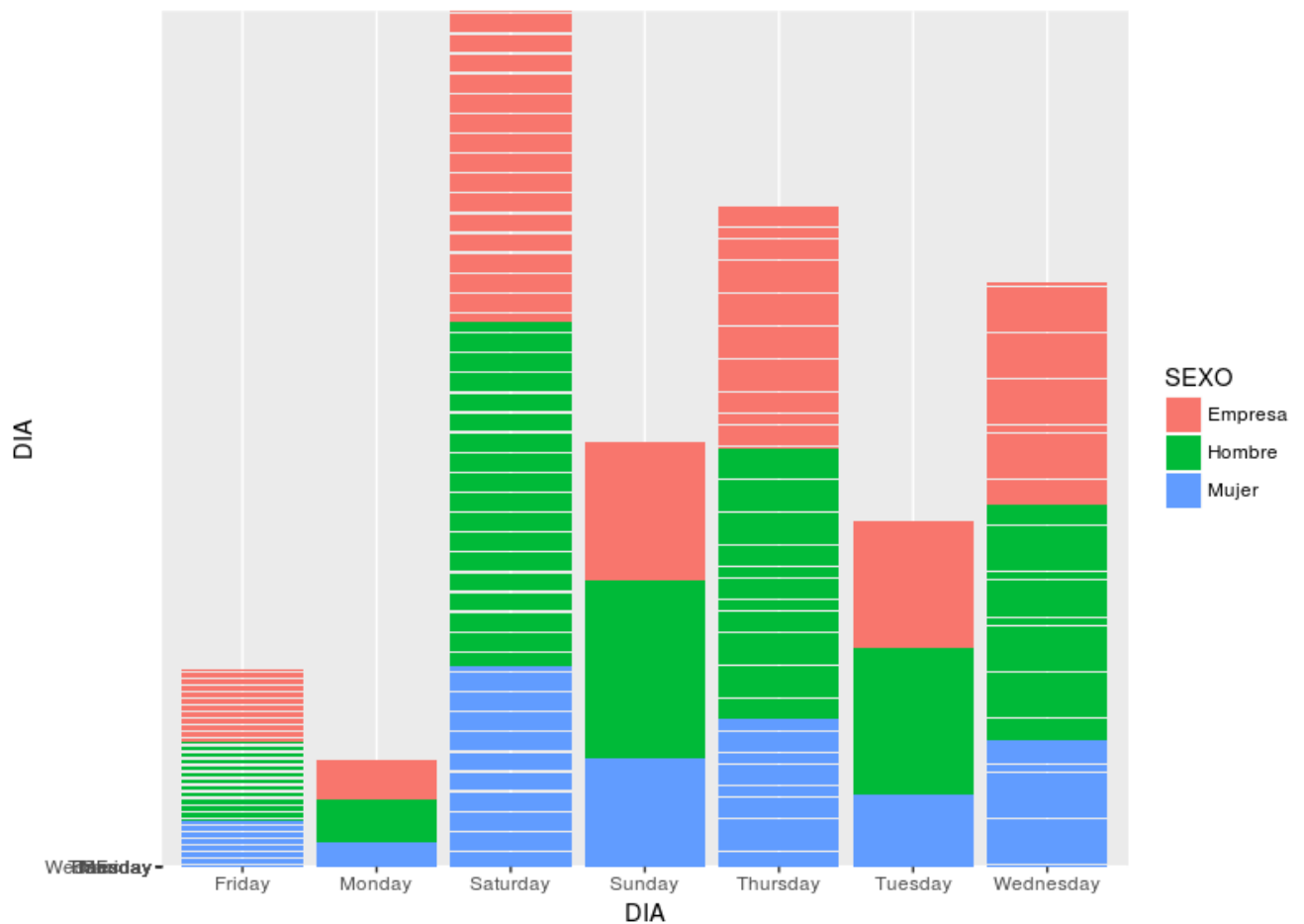
Podem per tant, intentar arribar acords o veure quines entitats finances em poden oferir millors condicions per aquest dos tipus de pagament.

Anàlisis per hora i dia de la setmana

En aquest darrer anàlisis **volem veure a quina hora del dia es produeixen la major de compres, i en quin dia de la setmana**, sempre podem per ofertes per aquelles hores de dies que hi ha menys afluència, i promocions de nous productes per als dies de major afluència.

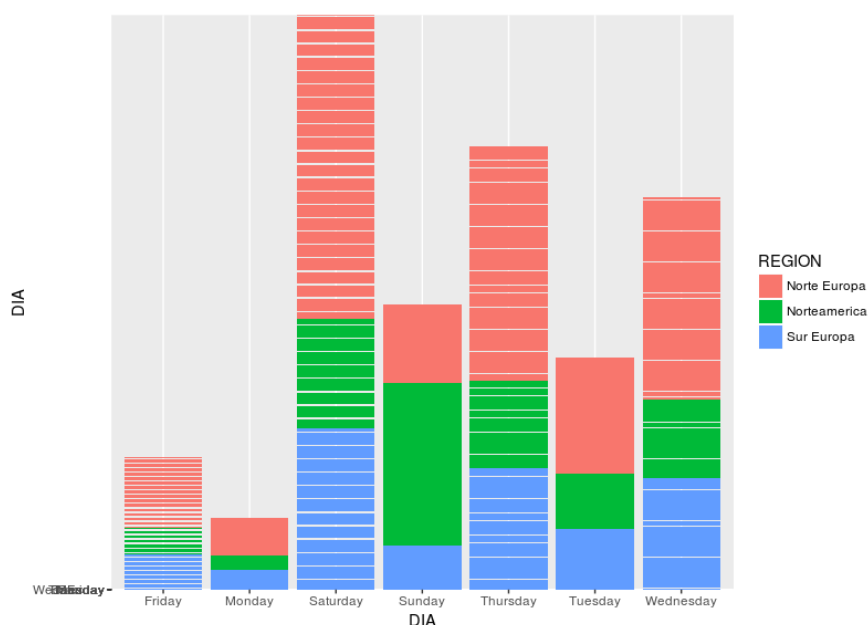
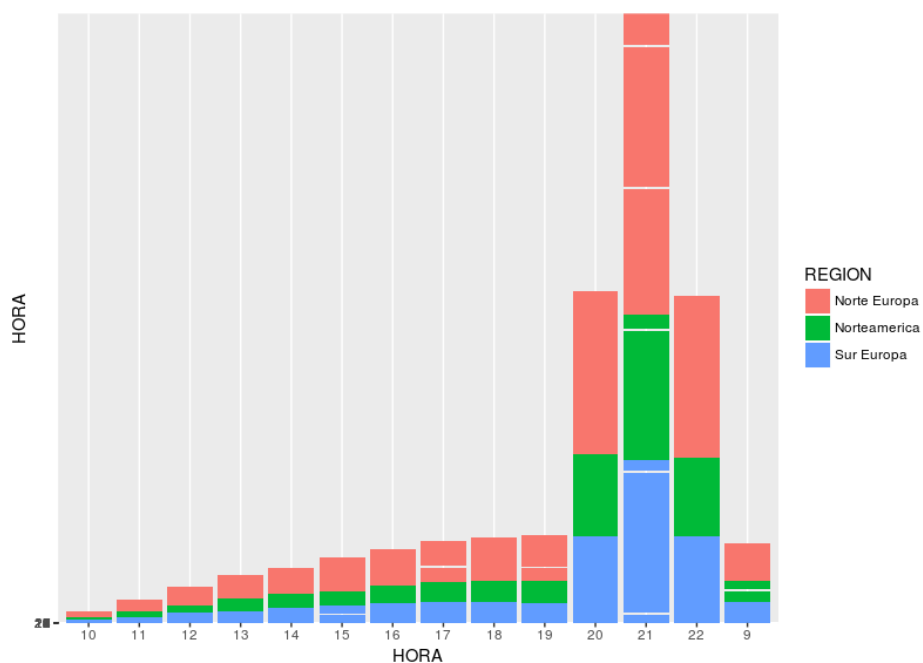
Per SEXE DEL CLIENT:





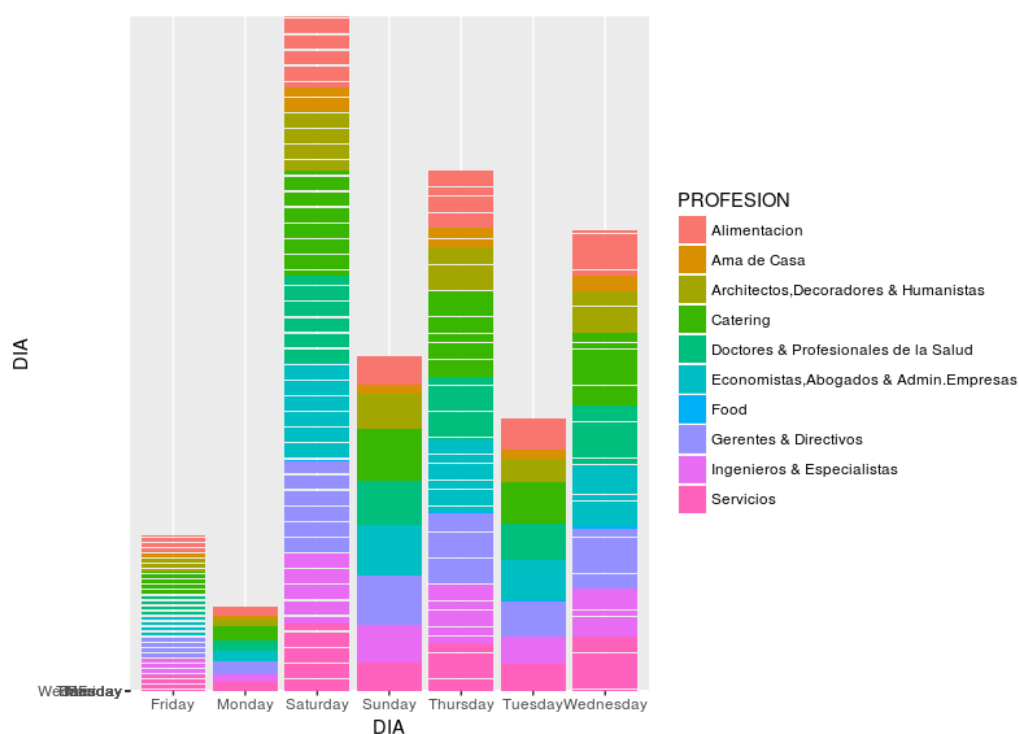
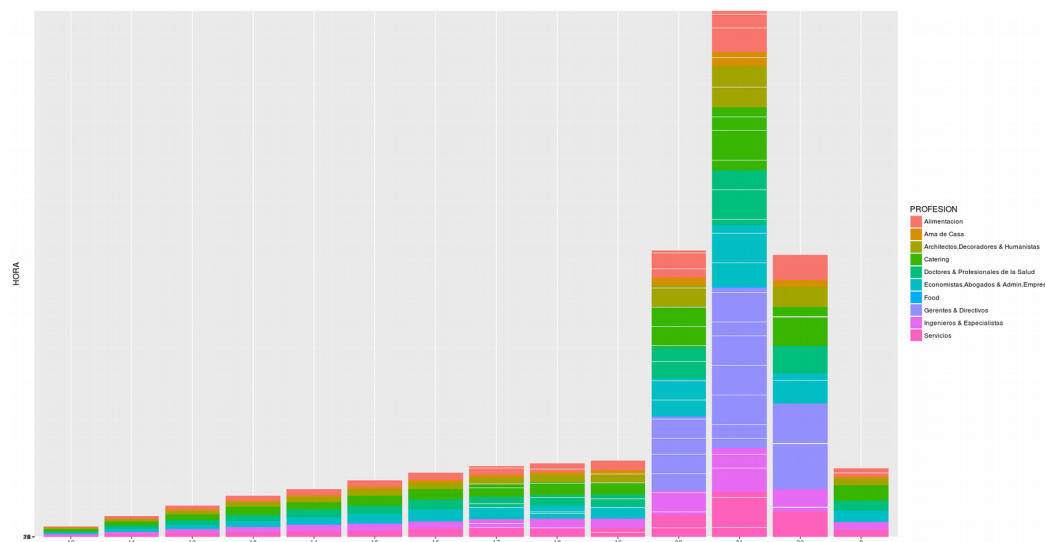
Hi ha bastanta homogeneïtat entre el sexe del client i el dimecres de màxima afluència i les 21:00. **Sols observem que a les 21:00 màxima afluència el clients de sexe home dupliquen les dones i les empreses.**

Per REGIÓ:



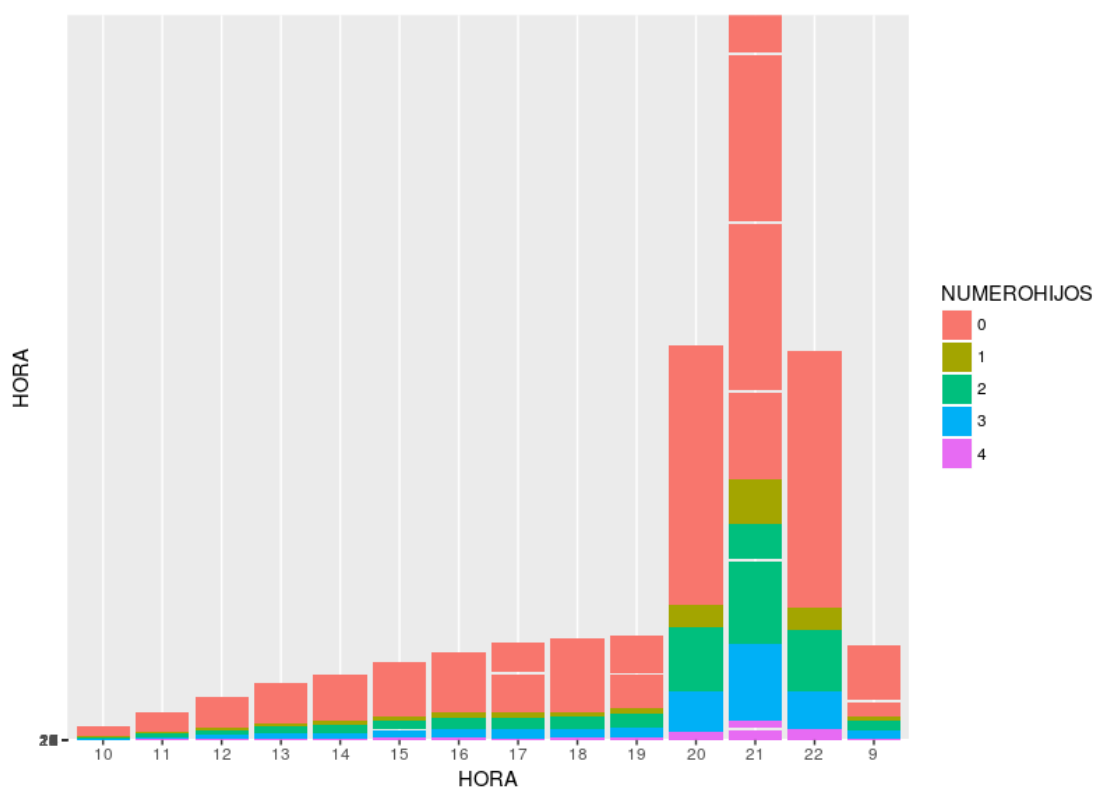
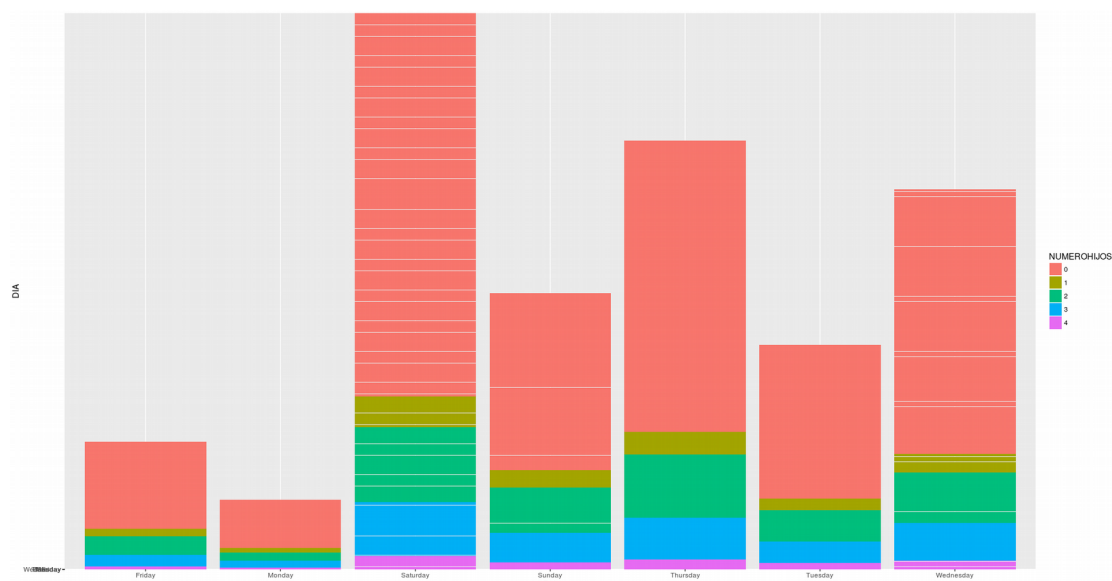
Pel que fa els rang de hora enter 20 i 22, es l'hora de maxima influencia a les 3 regions, per contra **el dissabte es el dia de màxima aflluència en Nord i Sud de Europa, i Nord Amèrica es el diumenge.**

Per PROFESSIÓ:



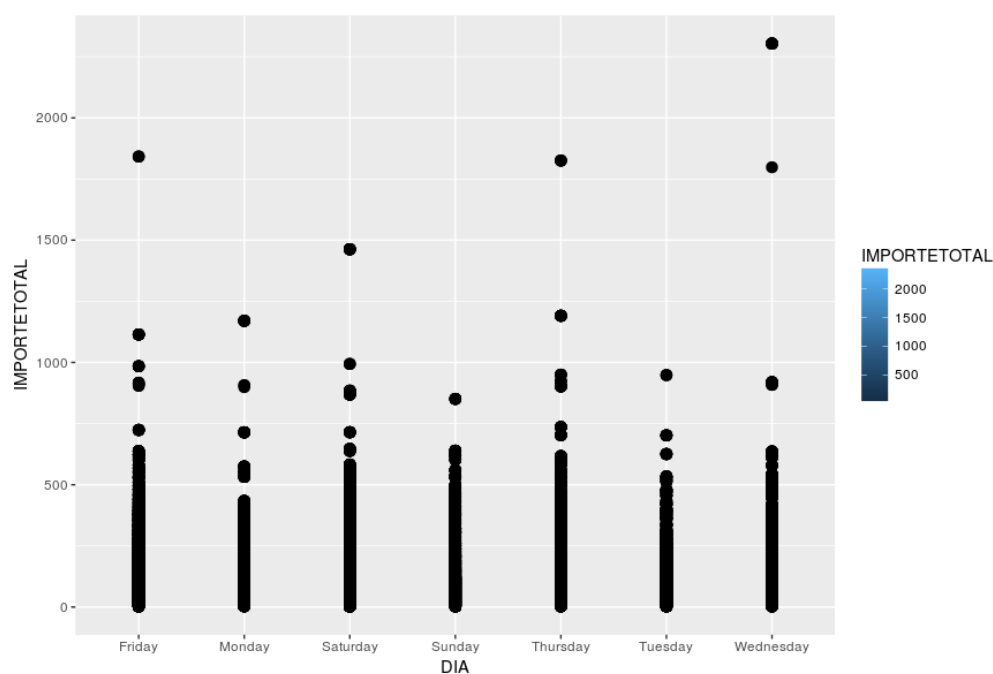
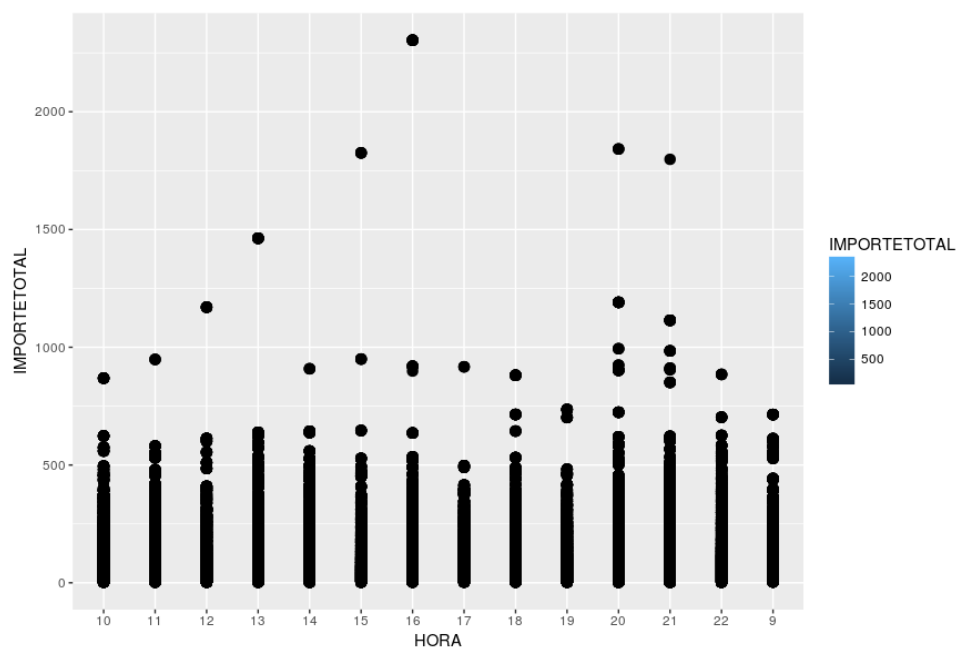
Hi ha homogeneïtat no podem traure una conclusió en la diferencia, la **professió no es un atribut que tingui influencia enter la hora i el dia de la compra.**

Per NOMBRE DE FILLS:



Ens reafirma que hem de millorar, en publicitat o promocions, les compres familiars. El major volum es produeix per clients sense fills, d'altra banda, hem vist que el import total no es el més alt. **No influeix el nombre de fills en la compra de la hora o del dia.**

Per IMPORT TOTAL:



Podem veure que els dimecres es quan es produeixen els màxims imports durant la setmana, pel que fa la hora es produeixen quan hi ha mes volum entre 20:00 i les 21:00, per contra una un valor extrem a les 16:00. **Si ignorem els valors extrems es bastant homogeni els valors de les compres, potser s'haurien d'estudiar els valors extrems que els produeix.**

En conclusió el dissabte es el dia de màxima afluència en Nord i Sud de Europa, i Nord Amèrica es el diumenge, pel que fa l'hora entre 20:00 i 22:00 hi ha màxima afluència)

Model d'arbres de decisió

El nostre primer intent utilitzarem un arbre de decisió per poder veure el perfil, o predir, de clients i els seus atributs respecte els productes, el qual podem traure algun valor o no.

Iniciam l'script important el dataset i normalitzem les dades, abans hem instal·lat i inicialitzat les llibreries, posteriorment reordenem les dades aleatòriament, i generem les gràfiques:

```
datos <- read.csv("/home/albert/Dropbox/uoc/UOC/MÀSTER - DATA CIENCIA/Mineria de
dades/practica/Entrega/csv/model_decision_tree_rpart.csv", header = TRUE, sep = ',')

datos = datos[,2:6] # clear autoincremental column

# arreglem els atributs per a que el detecti be com a qualitatiu
datos$NOMBREFAMILIA <- as.factor(sapply(datos$NOMBREFAMILIA, function (x) (chartr('áéíóúñäëïöüàèìò', 'aeiounaeiouaeiou', x))))
datos$REGION <- as.factor(sapply(datos$REGION, function (x) (chartr('áéíóúñäëïöüàèìò', 'aeiounaeiouaeiou', x))))
datos$PROFESION <- as.factor(sapply(datos$PROFESION, function (x) (chartr('áéíóúñäëïöüàèìò', 'aeiounaeiouaeiou', x))))
datos$NUMEROHIJOS <- factor(sapply(datos$NUMEROHIJOS, function (x) (chartr('áéíóúñäëïöüàèìò', 'aeiounaeiouaeiou', x))))
datos$SEXO <- as.factor(sapply(datos$SEXO, function (x) (chartr('áéíóúñäëïöüàèìò', 'aeiounaeiouaeiou', x))))

# write.csv(datos, file = "model_decision_tree_rpart.csv")

summary(datos)

nrow(datos)

#aleatoriament
datos <- datos[ sample( nrow( datos ), ) , ]

# ens assegurem que no hi ha valors a null
sapply(datos, function (x) (sum(is.na(x))))
```

Realitzem un estudi preliminar en l'script, que ja em valorat abans en l'anàlisi, per tant a continuació realitzem la binarització eliminant abans els duplicats:

```
# veiem que hi ha alguns família que la seva adquisició es mes freqüent que altres
barplot(table(datos$NOMBREFAMILIA), main = "Quantitat productes per família")

# veiem la venta de productes per professió
barplot(table(datos$PROFESION), main = "Quantitat Productes per professió")

# veiem la venta de productes per nombre de fills
barplot(table(datos$NUMEROHIJOS), main = "Quantitat Productes per nombre de fills")
```

```
# veiem la venta de productes per sexe
barplot(table(datos$SEXO), main = "Quantitat Productes per sexe")

# veiem la regió
barplot(table(datos$REGION), main = "Quantitat Productes per regió")
```

Realitzarem varies intents en diferents algorisme de arbres de decisió, per veure el grau de predicció del model i la seva qualitat, amb el grup de entrenament i el de test.

Creiem el conjunt de dades de entrenament i el de test de qualitat:

```
X <- datos[,2:5]
y <- datos[,1]

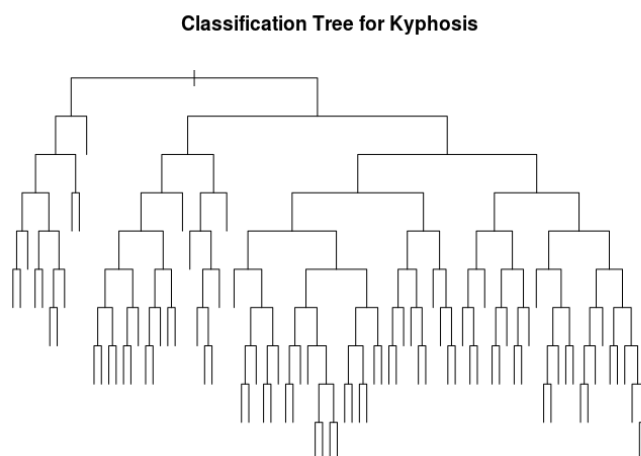
# calculate split 1/3 to test. This number is where the dataset must split the values
split <- length(datos$NOMBREFAMILIA) - round(length(datos$NOMBREFAMILIA)/3)
train <- datos[1:split,1:5]
testInputs <- X[(split + 1):length(datos$NOMBREFAMILIA),]
testOutput <- y[(split + 1):length(datos$NOMBREFAMILIA)]
```

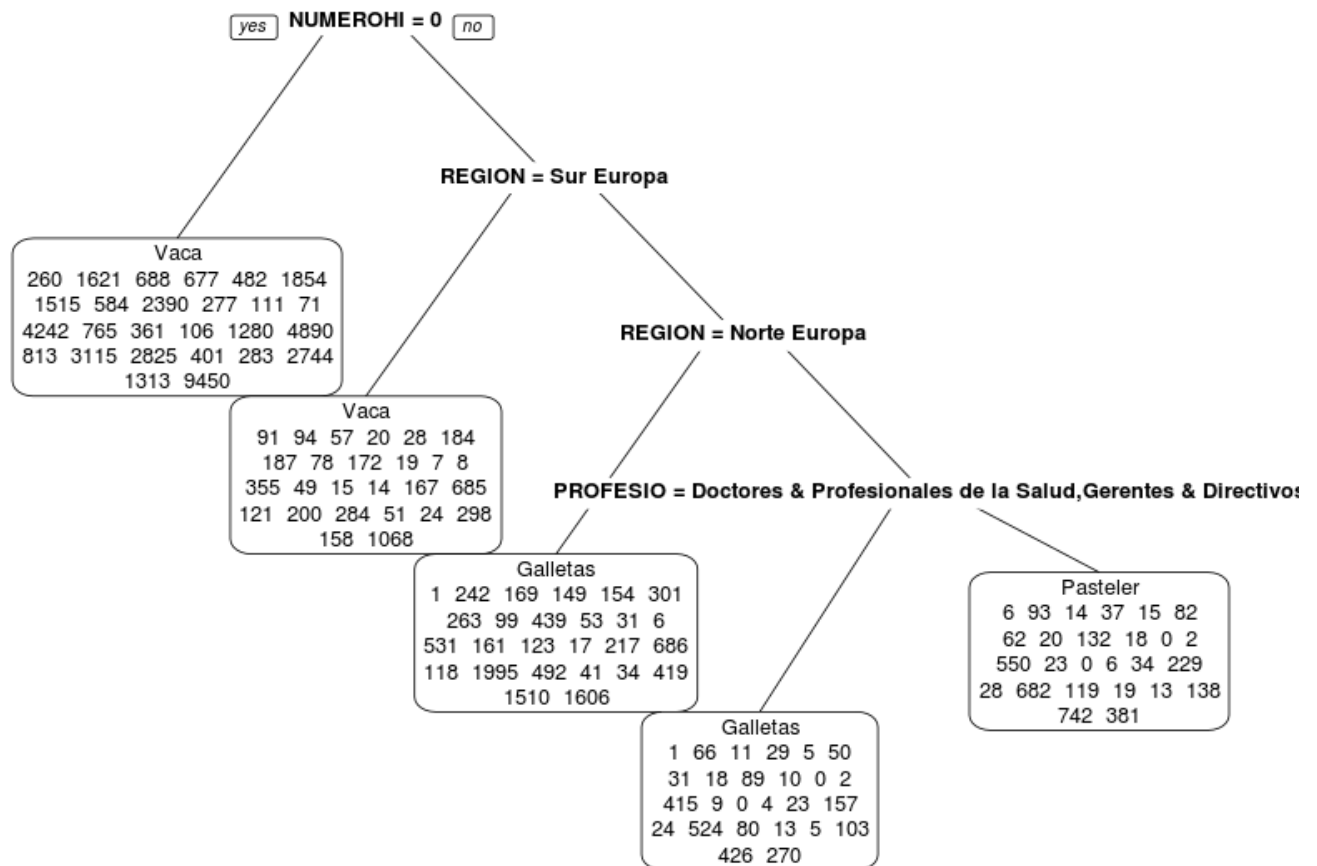
Primer model, mitjan el algorisme rpart:

```
#creem el model grow tree
model <- rpart(NOMBREFAMILIA ~., data = train, control = rpart.control(minsplit = 0.001, cp = 0))
printcp(model)

# Classification tree:
# rpart(formula = NOMBREFAMILIA ~., data = train, control = rpart.control(minsplit = 0.001,
# cp = 0))
#
# Variables actually used in tree construction:
# [1] NUMEROHIJOS PROFESION REGION SEXO
#
# Root node error: 50427/63219 = 0.79766
#
# n= 63219
```

Representem l'arbre:





Ara millorem l'arbre amb poda, i fem una validació creuada (cross-validation), per veure la qualitat:

```

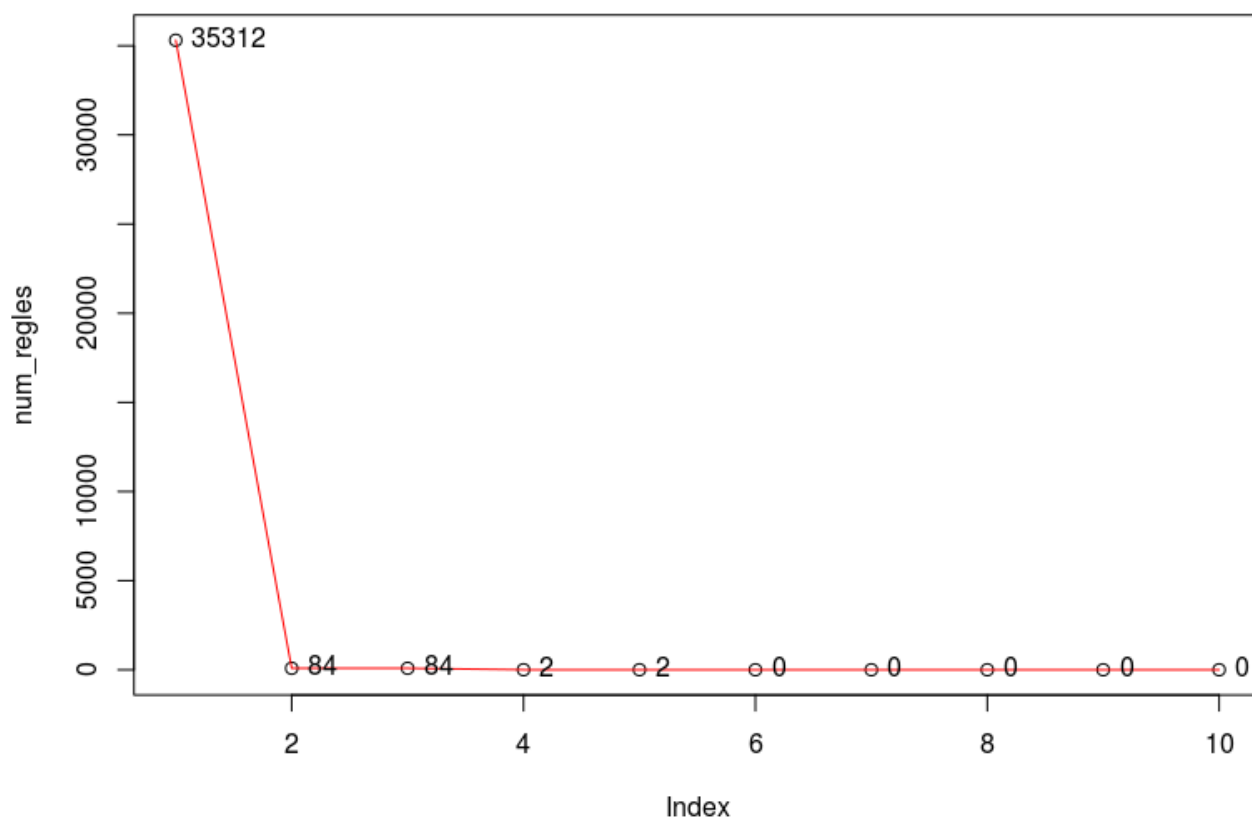
rsq.rpart(model) # visualize cross-validation results

#prune
#seleccionem el millor atribut com a separador
bestcp <- model$cpstable[which.min(model$cpstable[, "xerror"]), "CP"]
model.pruned <- prune(model, cp = bestcp)
prp(model.pruned, faclen = 0, cex = 0.8, extra = 1)

# Executem el mètode de perdició amb les dades de test
prediccio = predict(model.pruned, testInputs, type = "class")

# Podem avaluar la precisió del model utilitzant les dades de prova que hem reservat
sum( prediccio == testOutput ) / length( prediccio )
#[1] 0.218577
  
```

No és un bon model, amb el grup de dades de test, sols hem tingut un 21% de encerts, com podem veure al gràfic de cross-validation no es un bon model:



Intentarem amb random forest, un algorisme que utilitzara diferents implementació de arbres de decisions per veure quin hi ha mes encerts:

```
library(randomForest)
fit <- randomForest(NOMBREFAMILIA ~., data = datos)
print(fit) # view results
importance(fit) # importance of each predictor

# MeanDecreaseGini
# REGION      600.17721
# PROFESION   112.26541
# NUMEROHIJOS  599.87146
# SEXO        62.50985

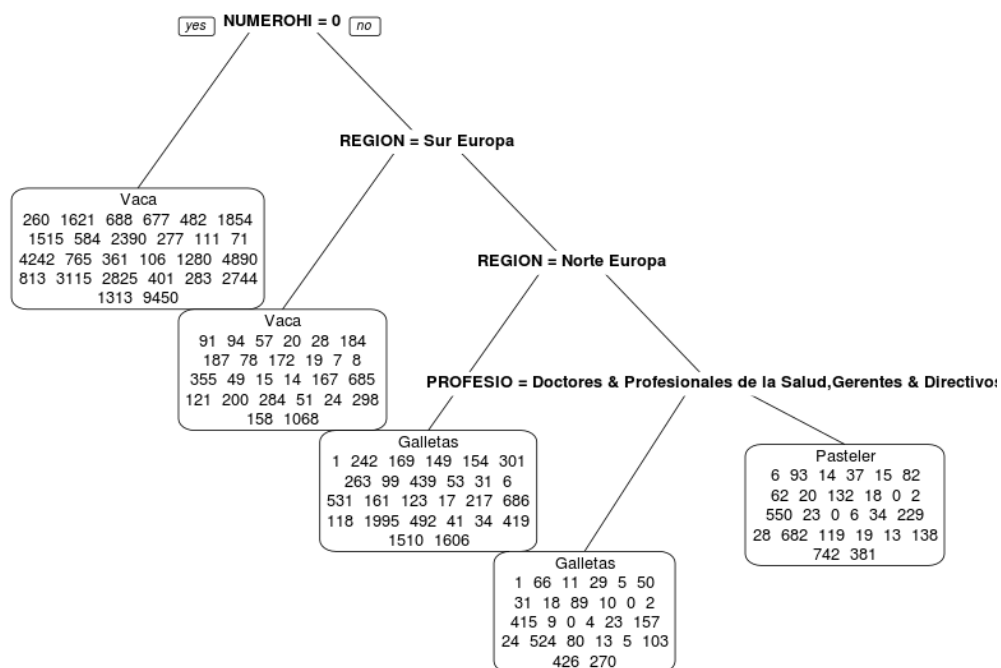
# Executem el mètode de perdició amb les dades de test
prediccio = predict(fit,testInputs ,type="class")

# Podem avaluar la precisió del model utilitzant les dades de prova que hem reservat
sum( prediccio == testOutput ) / length( prediccio )
#[1] 0.2227214
```

No obtenim millors resultats, en tot cas, com ja sabíem la regió es el factor mes

important per la classificació dels productes, així com el numero de fills.

En definitiva, model de poca qualitat, i les regles que ja sabíem, que els que no tenen fills, son els que mes compren i per tant de la família de productes de «Vaca», si tenen mes fills al surt Europa també prefereixen productes de la família «Vaca», del Nord de Europa Galletes, i si no son Doctors, professional de la Salut, segurament família «Pasteler», per desgracia sols hem pogut encertar un 21%.



Model de regles d'associació

Un model de regles d'associació **es molt útil per prendre decisions sobre màrqueting com a preus promocionals per a certs productes o on situar aquests dins del supermercat, per aquest motiu l'utilitzarem.**

Iniciam l'script important el dataset i normalitzem les dades, abans hem instal·lat i inicialitzat les llibreries, posteriorment reordenem les dades aleatòriament, i generem les gràfiques:

```

datos <- read.csv("/home/albert/Dropbox/uoc/UOC/MÀSTER - DATA CIENCIA/Mineria de
dades/practica/Entrega/csv/model_agregation_rules.csv", header = TRUE, sep = ',')

datos = datos[,2:3] # clear autoincremental column
  
```

```
summary(datos)
```

```
# arreglem els atributs CODVENTA i DESCRIPTION per a que el detecti be com a qualitatiu
datos$CODVENTA <- factor(datos$CODVENTA)
datos$DESCRIPCION <- factor(datos$DESCRIPCION)
datos$DESCRIPCION <- as.factor(sapply(datos$DESCRIPCION, function (x) (chartr("áéíóúñäëïöüàèìòù",'aeiounaeiouaeiou', x))))
```

Realitzem un estudi preliminar en l'script, que ja em valorat abans en l'anàlisi, per tant a continuació realitzem la binarització eliminant abans els duplicats:

```
# Llista per a cada usuari els productes comprats
mba <- split(x=datos[, "DESCRIPCION"], f=datos$CODVENTA)
summary(mba)
```

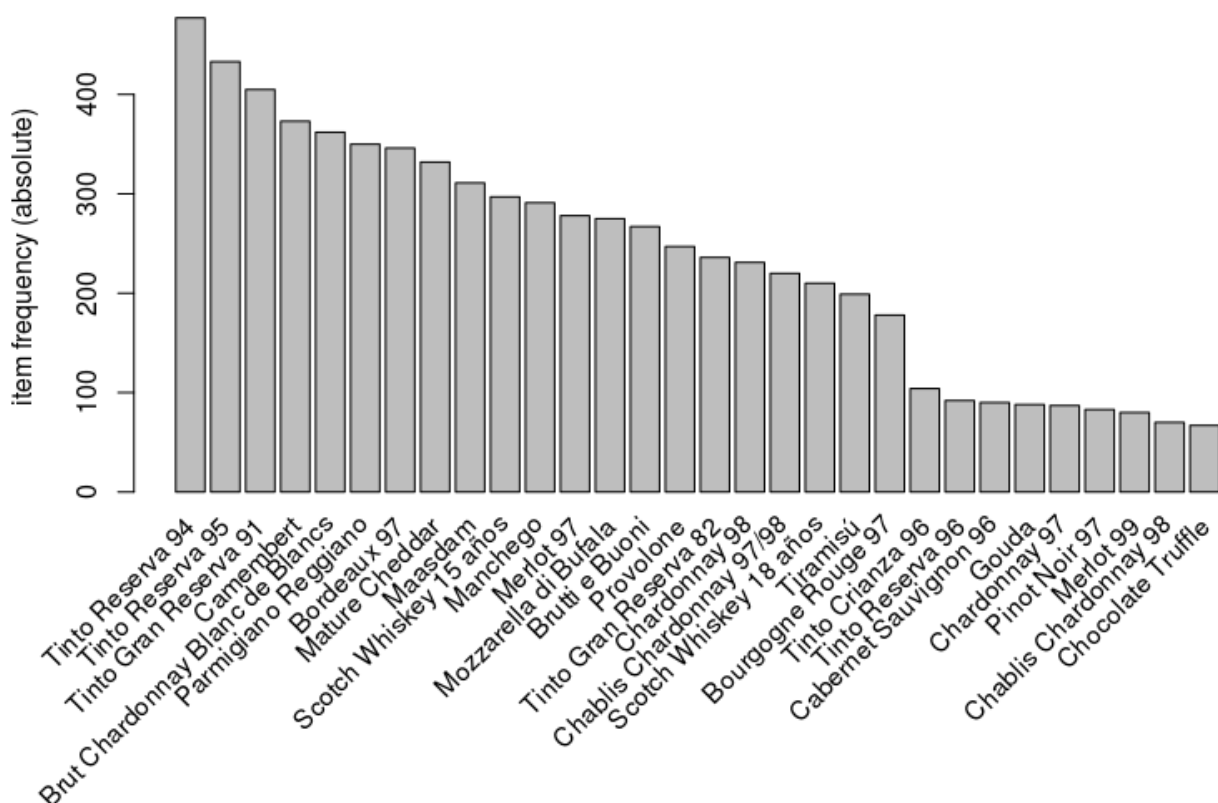
```
# total registres
sum(sapply(mba, length))
# [1] 11247
```

```
# Si hi ha duplicats els eliminen
mba <- lapply(mba, unique)
```

```
sum(sapply(mba, length))
# [1] 11042 , hem netejar 205 valors
```

```
# binaritzem
mba <- as(mba, "transactions")
class(mba)
```

```
# top 30 productes
itemFrequencyPlot(mba, topN = 30, type = "absolute")
```

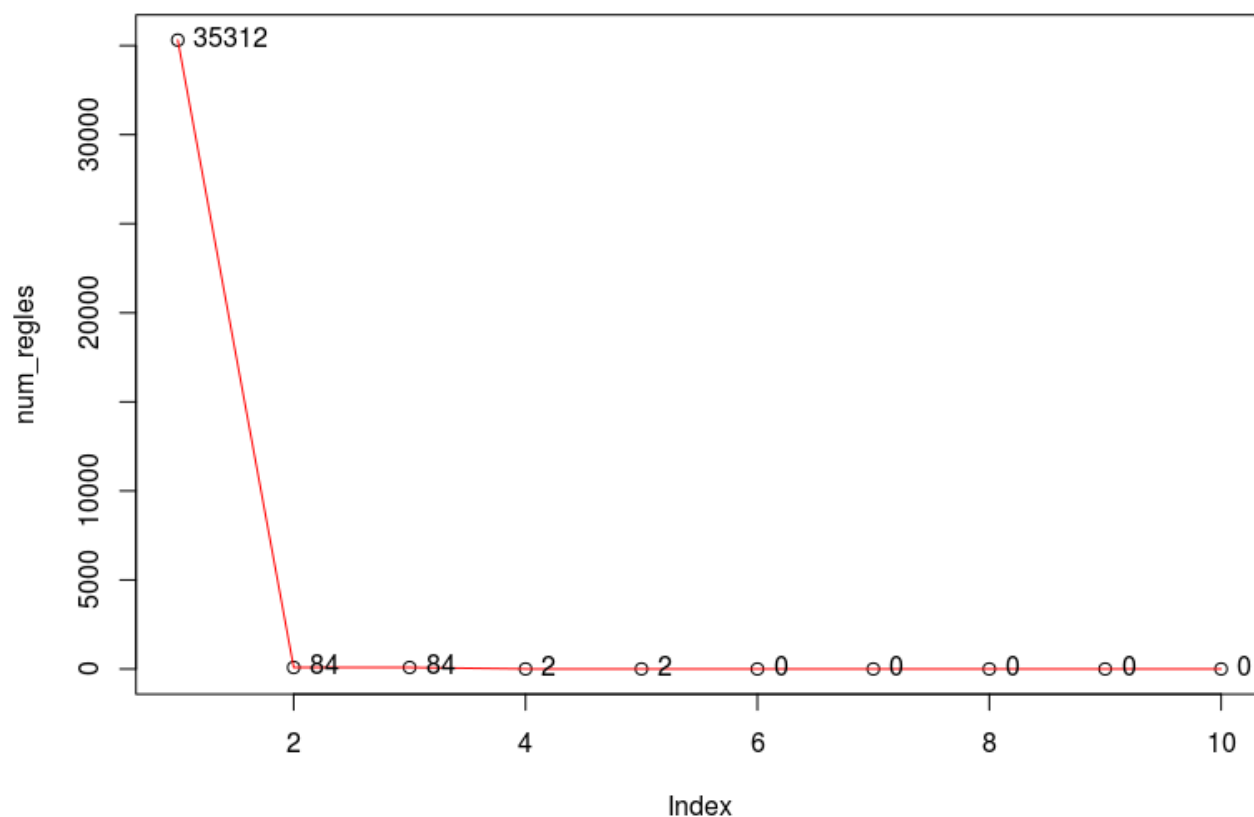


Ja sabíem del estudi que la subfamília *DO Vino Rioja*, sons els productes més venuts, ara podem veure que el *Tinto de Reserva 94, 95 i 91* son el top 3.

A continuació realitzem ajustament de suport entre 0.0001 i 0.001 per optimitzar les regles d'associació:

```
# generem varies binaritzacions per veure qui genera un numero optim de regles
regles <- list()
num_regles <- list()
serie <- seq(0.0001, 0.001, by = 0.0001)
for (k in serie)
{
  regles[[k*10000]] <- apriori(mba, parameter = list(support = k, confidence = 0.8))
  num_regles[k*10000] <- length(regles[[k*10000]])
}

num_regles <- unlist(num_regles)
plot(num_regles)
lines(num_regles, col="red")
for (i in 1:10)
{
  text(x = i, y = num_regles[i], pos = 4, num_regles[i])
}
```



Podem veure que a 0.0002 les regles generades son 84, un numero optim de regles

per agafar 0.0002 com a valor de suport.

Finalment observem les regles amb suport 0.0002:

```
regles_final <- regles[[2]]
summary(regles_final)
# set of 84 rules
#
# rule length distribution (lhs + rhs):sizes
# 3 4
# 54 30
#
# Min. 1st Qu. Median Mean 3rd Qu. Max.
# 3.000 3.000 3.000 3.357 4.000 4.000

# les regles
inspect(regles_final)
```

| # lhs | rhs | support | confidence | lift | count |
|--|--------------------------------------|----------------|-------------|------|-------|
| # [1] (Brut Extra Selection,Tinto Gran Reserva 82) | => (Tinto Reserva 95) | 0.0003871467 1 | 11.93072 2 | | |
| # [2] (Brut Chardonnay Blanc de Blancs,Danish Crumbly Blue) | => (Camembert) | 0.0003871467 1 | 13.84987 2 | | |
| # [3] (Camembert,Danish Crumbly Blue) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [4] (Brut Premier Cru,Chocolate Truffle) | => (Provolone) | 0.0003871467 1 | 20.91498 2 | | |
| # [5] (Camembert,Polvorones de Estepa) | => (Mozzarella di Bufala) | 0.0003871467 1 | 18.78545 2 | | |
| # [6] (Scotch Whiskey 15 anos,Shiraz-Cabernet 96) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [7] (Shiraz-Cabernet 96,Tinto Reserva 94) | => (Scotch Whiskey 15 anos) | 0.0003871467 1 | 17.39394 2 | | |
| # [8] (Bra,Tiramisu) | => (Tinto Gran Reserva 91) | 0.0003871467 1 | 12.75556 2 | | |
| # [9] (Chardonnay en barrica 97,Mozzarella di Bufala) | => (Scotch Whiskey 18 anos) | 0.0003871467 1 | 24.60000 2 | | |
| # [10] (Chardonnay en barrica 97,Provolone) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [11] (Chardonnay en barrica 97,Tinto Reserva 94) | => (Provolone) | 0.0003871467 1 | 20.91498 2 | | |
| # [12] (Bourgogne Rouge 97,Goat Gouda) | => (Scotch Whiskey 15 anos) | 0.0003871467 1 | 17.39394 2 | | |
| # [13] (Goat Gouda,Scotch Whiskey 15 anos) | => (Bourgogne Rouge 97) | 0.0003871467 1 | 29.02247 2 | | |
| # [14] (Sage Derby,Sylvaner 96) | => (Merlot 97) | 0.0003871467 1 | 18.58273 2 | | |
| # [15] (Merlot 97,Sage Derby) | => (Sylvaner 96) | 0.0003871467 1 | 198.69231 2 | | |
| # [16] (Iberico,Maasdam) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [17] (Gorgonzola Dolce,Tinto Reserva 95) | => (Tinto Gran Reserva 82) | 0.0003871467 1 | 21.88983 2 | | |
| # [18] (Bordeaux 97,Sauvignon Blanc 96) | => (Tinto Gran Reserva 91) | 0.0003871467 1 | 12.75556 2 | | |
| # [19] (Asiago,Parmigiano Reggiano) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [20] (Camembert,Zinfandel 96) | => (Manchego) | 0.0003871467 1 | 17.75258 2 | | |
| # [21] (Brut Chardonnay Blanc de Blancs,Reserve Chardonnay 97) | => (Tinto Reserva 95) | 0.0003871467 1 | 11.93072 2 | | |
| # [22] (Brinata,Provolone) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [23] (Postel,Tinto Reserva 95) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [24] (Brut Chardonnay Blanc de Blancs,Extra Aged Gouda) | => (Mature Cheddar) | 0.0003871467 1 | 15.56024 2 | | |
| # [25] (Monte Veronese,Semillon 99) | => (Provolone) | 0.0003871467 1 | 20.91498 2 | | |
| # [26] (Monte Veronese,Provolone) | => (Semillon 99) | 0.0003871467 1 | 198.69231 2 | | |
| # [27] (Parmigiano Reggiano,Tokay Pinot Gris 98) | => (Maasdam) | 0.0003871467 1 | 16.61093 2 | | |
| # [28] (Bordeaux 88,Tinto Gran Reserva 82) | => (Scotch Whiskey 18 anos) | 0.0003871467 1 | 24.60000 2 | | |
| # [29] (Bordeaux 88,Chardonnay 98) | => (Scotch Whiskey 15 anos) | 0.0003871467 1 | 17.39394 2 | | |
| # [30] (Bordeaux 88,Tinto Gran Reserva 82) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [31] (Chardonnay 97,Galettes) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [32] (Bourgogne Rouge 97,Yalumba Grenache 97) | => (Parmigiano Reggiano) | 0.0003871467 1 | 14.76000 2 | | |
| # [33] (Maasdam,Yalumba Grenache 97) | => (Parmigiano Reggiano) | 0.0003871467 1 | 14.76000 2 | | |
| # [34] (Brutti e Buoni,Scotch Whiskey 21 anos) | => (Merlot 97) | 0.0003871467 1 | 18.58273 2 | | |
| # [35] (Merlot 97,Scotch Whiskey 21 anos) | => (Brutti e Buoni) | 0.0003871467 1 | 19.34831 2 | | |
| # [36] (Sage Derby,Scotch Whiskey 15 anos) | => (Tinto Reserva 95) | 0.0003871467 1 | 11.93072 2 | | |
| # [37] (Swiss Gruyere,Tinto Reserva 95) | => (Camembert) | 0.0003871467 1 | 13.84987 2 | | |
| # [38] (Chablis Chardonnay 97,Tiramisu) | => (Brutti e Buoni) | 0.0003871467 1 | 19.34831 2 | | |
| # [39] (Cabralis,Chardonnay 98) | => (Brutti e Buoni) | 0.0005807201 1 | 19.34831 3 | | |
| # [40] (Brutti e Buoni,Cabralis) | => (Chardonnay 98) | 0.0005807201 1 | 22.36364 3 | | |
| # [41] (Cabernet Sauvignon 97,Passendale) | => (Chablis Chardonnay 97/98) | 0.0003871467 1 | 23.48182 2 | | |
| # [42] (Brut Chardonnay Blanc de Blancs,Pinot Noir 98) | => (Brutti e Buoni) | 0.0003871467 1 | 19.34831 2 | | |
| # [43] (Monte Veronese,Tinto Crianza 96) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [44] (Gruyere de Comte,Provolone) | => (Merlot 97) | 0.0003871467 1 | 18.58273 2 | | |
| # [45] (Brut Reserva,Tinto Reserva 96) | => (Tinto Gran Reserva 82) | 0.0003871467 1 | 21.88983 2 | | |
| # [46] (Brut Reserva,Tinto Gran Reserva 82) | => (Tinto Reserva 96) | 0.0003871467 1 | 56.15217 2 | | |
| # [47] (Manchego,Mature Blue Siltion) | => (Tinto Reserva 95) | 0.0003871467 1 | 11.93072 2 | | |
| # [48] (Mature Blue Siltion,Mature Cheddar) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [49] (Cabernet Sauvignon 97,Tinto Reserva 95) | => (Merlot 97) | 0.0003871467 1 | 18.58273 2 | | |
| # [50] (Bourgogne Rouge 97,Tinto Gran Reserva 89) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [51] (Chablis Chardonnay 97/98,Tinto Gran Reserva 89) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [52] (Scotch Whiskey 15 anos,Tinto Crianza 96/97) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [53] (Layden Gin,Provolone) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |
| # [54] (Cabernet Sauvignon 96,Gouda) | => (Tinto Gran Reserva 91) | 0.0003871467 1 | 12.75556 2 | | |
| # [55] (Bordeaux 88,Scotch Whiskey 18 anos,Tinto Gran Reserva 82) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [56] (Bordeaux 88,Bordeaux 97,Scotch Whiskey 18 anos) | => (Tinto Gran Reserva 82) | 0.0003871467 1 | 21.88983 2 | | |
| # [57] (Bordeaux 88,Bordeaux 97,Tinto Gran Reserva 82) | => (Scotch Whiskey 18 anos) | 0.0003871467 1 | 24.60000 2 | | |
| # [58] (Bordeaux 97,Gouda,Scotch Whiskey 18 anos) | => (Camembert) | 0.0003871467 1 | 13.84987 2 | | |
| # [59] (Camembert,Gouda,Scotch Whiskey 18 anos) | => (Bordeaux 97) | 0.0003871467 1 | 14.93064 2 | | |
| # [60] (Bordeaux 97,Camembert,Gouda) | => (Scotch Whiskey 18 anos) | 0.0003871467 1 | 24.60000 2 | | |
| # [61] (Bordeaux 97,Camembert,Scotch Whiskey 18 anos) | => (Gouda) | 0.0003871467 1 | 58.70455 2 | | |
| # [62] (Brutti e Buoni,Chablis Chardonnay 97/98,Tinto Crianza 96) | => (Brut Chardonnay Blanc de Blancs) | 0.0003871467 1 | 14.27072 2 | | |
| # [63] (Brut Chardonnay Blanc de Blancs,Chablis Chardonnay 97/98,Tinto Crianza 96) | => (Brutti e Buoni) | 0.0003871467 1 | 19.34831 2 | | |
| # [64] (Brutti e Buoni,Manchego,Tiramisu) | => (Merlot 97) | 0.0003871467 1 | 18.58273 2 | | |
| # [65] (Brutti e Buoni,Merlot 97,Tiramisu) | => (Manchego) | 0.0003871467 1 | 17.75258 2 | | |
| # [66] (Maasdam,Merlot 97,Tiramisu) | => (Tinto Reserva 95) | 0.0003871467 1 | 11.93072 2 | | |
| # [67] (Merlot 97,Tinto Reserva 95,Tiramisu) | => (Maasdam) | 0.0003871467 1 | 16.61093 2 | | |
| # [68] (Tinto Gran Reserva 91,Tinto Reserva 95,Tiramisu) | => (Tinto Reserva 94) | 0.0003871467 1 | 10.83019 2 | | |

| | | | |
|---|--------------------------------------|----------------|------------|
| # [69] {Tinto Gran Reserva 91,Tinto Reserva 94,Tiramisu} | => {Tinto Reserva 95} | 0.0003871467 1 | 11.93072 2 |
| # [70] {Bordeaux 97,Chablis Chardonnay 97/98,Manchego} | => {Provolone} | 0.0003871467 1 | 20.91498 2 |
| # [71] {Bordeaux 97,Manchego,Provolone} | => {Chablis Chardonnay 97/98} | 0.0003871467 1 | 23.48182 2 |
| # [72] {Brutti e Buoni,Chablis Chardonnay 97/98,Tinto Reserva 95} | => {Tinto Reserva 94} | 0.0003871467 1 | 10.83019 2 |
| # [73] {Brut Chardonnay Blanc de Blancs,Chablis Chardonnay 97/98,Parmigiano Reggiano} | => {Mature Cheddar} | 0.0003871467 1 | 15.56024 2 |
| # [74] {Brutti e Buoni,Merlot 97,Provolone} | => {Tinto Reserva 95} | 0.0003871467 1 | 11.93072 2 |
| # [75] {Brutti e Buoni,Provolone,Tinto Reserva 95} | => {Merlot 97} | 0.0003871467 1 | 18.58273 2 |
| # [76] {Brutti e Buoni,Merlot 97,Tinto Reserva 95} | => {Provolone} | 0.0003871467 1 | 20.91498 2 |
| # [77] {Bordeaux 97,Mature Cheddar,Provolone} | => {Tinto Reserva 95} | 0.0003871467 1 | 11.93072 2 |
| # [78] {Bordeaux 97,Provolone,Tinto Reserva 95} | => {Mature Cheddar} | 0.0003871467 1 | 15.56024 2 |
| # [79] {Bordeaux 97,Mature Cheddar,Tinto Reserva 95} | => {Provolone} | 0.0003871467 1 | 20.91498 2 |
| # [80] {Brut Chardonnay Blanc de Blancs,Brutti e Buoni,Mature Cheddar} | => {Camembert} | 0.0003871467 1 | 13.84987 2 |
| # [81] {Brutti e Buoni,Camembert,Mature Cheddar} | => {Brut Chardonnay Blanc de Blancs} | 0.0003871467 1 | 14.27072 2 |
| # [82] {Brut Chardonnay Blanc de Blancs,Camembert,Mature Cheddar} | => {Brutti e Buoni} | 0.0003871467 1 | 19.34831 2 |
| # [83] {Maasdam,Manchego,Tinto Reserva 95} | => {Tinto Reserva 94} | 0.0003871467 1 | 10.83019 2 |
| # [84] {Parmigiano Reggiano,Scotch Whiskey 15 anos,Tinto Reserva 95} | => {Merlot 97} | 0.0003871467 1 | 18.58273 2 |

Es una bon model, que si mes no, podria ajudar al client per realitzar distribucions de productes, ofertes o publicitat.

Accions per Gourmet

Per la realització de la publicitat, ofertes o promocions, donem les següents primícies al grup de botigues Gourmet:

- Volum de ventes (de freqüència), en nivell de família és tracta de la Vaca (derivats de vaca), per contra les subfamílies les que tenen un volum més alta de venta es D.O Rioja-Tinto i Galletes de xocolates, per tant les subfamílies i família més populars entre els totes les botigues del grup.
- Nort Europa, es on més volum de compres es produeix, s'hauria de promocionar la resta de regions.
- S'hauria de realitzar alguna oferta o campanya per augmentar en sectors de Food (menja) i de mestresses de casa, estan per sota de la resta.
- Sabem que volum de compres està igualat enter sexes, per contra, el import d'aquestes compres es molt més elevat en les empreses, després els homes i finalment les dones.
- La majoria de compres són realitzades per clients que no tenen fills. Per estat civil veiem que gestiona més import el clients casats, clarament perquè son famílies. hem de publicitar per famílies, per que tenen un volum de compres més alt, i perquè en freqüència està per sota de clients sense fills.

- Valor importat que podem usar per a la publicitat, o per combinar amb productes que no tenen tanta sortida al mercat:

El productes mes venuts:

| | |
|-------------------|------|
| Chocolate Truffle | 4848 |
| Tiramisu | 4558 |
| Tinto Reserva 95 | 3880 |
| Camembert | 3424 |

Els productes menys venuts:

| | |
|-------------------------------|-----|
| Cabernet Sauvignon Reserva 93 | 168 |
| Reblochon | 168 |
| White Stilton | 168 |
| Riesling 96 | 167 |
| Shiraz 97 | 167 |
| Tinto Crianza 94 | 167 |
| Brinata | 166 |
| Double Gloucester | 166 |

- Les formes de pagament que més import gestiona es la targeta de crèdit, tant per empreses, homes i dones, l'altra és els xecs, indistintament pel sexe del client, o per l'estat civil.

Podem per tant, s'haurien d'intentar arribar acords o veure quines entitats finances em poden oferir millors condicions per aquest dos tipus de pagament.

- El dissabte es el dia de màxima afluència en Nord i Sud de Europa, i Nord Amèrica es el diumenge, pel que fa l'hora entre 20:00 i 22:00 hi ha màxima afluència, per tant, es un bon moment per la realització de promocions o productes nous.
- En el model de regles, tenim el llistat per a que el client realitze promocions, *packs* o distribucions en les botigues per optimitzar la venda.

Bibliografia

Análisis y Decisión - <http://analisisydecision.es/trucos-r-llevar-a-sas-las-reglas-de-un-arbol-de-decision/>

Ordered categories - <https://stackoverflow.com/questions/36480309/stacked-bar-chart-with-some-negative-values-and-ordered-categories/36480947#36480947>

Validacin cruzada - <http://apuntes-r.blogspot.com.es/2014/11/validacion-cruzada.html>

Plot: Plot method to visualize association rules and itemsets - <https://rdr.io/rforge/arulesViz/man/plot.html>

Cluster Validation Essentials - <https://www.datasciencecentral.com/profiles/blogs/data-mining-association-rules-in-r-diapers-and-beer>

Introduction of Association Rules - <http://apuntes-r.blogspot.com.es/2014/11/validacion-cruzada.html>

Estructura de dades - http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/ManualR/intro_estructurasdedatos.html#intro_estructuras_arrays

R per a Principiants, Emmanuel Paradis - https://cran.r-project.org/doc/contrib/rdebuts_es.pdf

Operacions per files i columnes - <http://rparamuyprincipiantes.blogspot.com.es/2015/05/operaciones-por-filas-o-columnas-ii.html>

Gràfics estadístics amb R, Juan Carlos Correa y Nelfi González - <https://cran.r-project.org/doc/contrib/grafi3.pdf>

Based Publication Ready Plots - <http://www.sthda.com/english/rpkgs/ggpubr/>