#產生百萬假會員資料  
  
Sys.setlocale(category = "LC\_ALL", locale = "cht")  
  
profile <- data.frame(  
id = paste("0000000", as.character( seq(from=1, to=1000000, by=1)),sep='' ) ,  
age = round(rnorm(1000000,25,3)),  
gender = sample(c("男生", "女生"), size = 1000000, replace = TRUE),  
group = sample(c("誠品人", "鋼鐵仁","香港人"), size = 1000000, replace = TRUE),  
zone = sample(c("台灣","香港"), size=1000000, replace=TRUE)  
)  
  
install.packages("sqldf")  
library(sqldf)  
profile$id <- sqldf("SELECT substr(id,-7,7) FROM profile")  
  
profile\_final <- data.frame(id = profile$id,  
age = profile$age,  
gender = profile$gender,  
group = profile$group,  
zone = profile$zone)  
names(profile\_final)[1] <- 'id'  
head(profile\_final,10)  
  
#檔案輸出  
write.table(profile\_final, file = "profile.csv", quote = FALSE, sep = ",", row.names = FALSE, col.names =TRUE )  
  
#產生百萬假會員一年的消費資料  
trans\_df <- data.frame(data\_month=character(),  
id=character(),  
product=character(),  
Logicals=logical(),  
Characters=character(),  
amt=integer()  
)  
  
#每月產出假消費紀錄  
for(i in 1:12)  
{  
  
id = sample(profile\_final$id, size = round(runif(1,300000,700000)), replace = TRUE)  
i\_month <- ifelse(nchar(i)==1,paste('0',i,sep=''),i)  
  
tmp\_df <- data.frame(  
data\_month = paste('2019/',i\_month,'/01',sep=''),  
id = id,  
product = sample(c("中文書", "外文書","雜誌","兒童","CD","DVD","風格文具",'生活雜貨'), size = length(id), replace = TRUE),  
amt = round(runif(length(id),50,699))  
)  
  
trans\_df <- rbind(trans\_df,tmp\_df)  
  
}  
  
#檔案輸出  
write.table(trans\_df, file = "trans.csv", quote = FALSE, sep = ",", row.names = FALSE, col.names =TRUE )