ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

DEPARTMENT OF MANAGEMENT SCIENCE AND TECHNOLOGY

MASTER OF SCIENCE IN BUSINESS ANALYTICS

Lesson: Big Data Systems

Professor: S. Safras

Assignment: Redis

Editors: Vogiatzis George, Zourou Myrsini

AM: P2821827, P2821828

1. Redis

In this task, we are going to use the **“recorded actions”** dataset in order to generate some analytics with REDIS. This dataset contains two files the emails_sent file and the modified_listing file. The first contain information about the users who received an email at least one and the last about who users had modified their listing. The months, which the files contain, was January, February and March.

1.1 Question

In this question, we should find how many users modified their listing on January. To answer that we separate the modified listing dataset into three tables, one for each month. Therefore, from the file, which contain the listing for the January, we execute the setbit in order to set one to user who modified their listing and zero in the opposite situation. The code we use was the below. The bitcount calculate that 9969 users modified their listing.

```
listing_jan<-listings[which(listings$MonthID==1),]

listing_feb<-listings[which(listings$MonthID==2),]

listing_mar<-listings[which(listings$MonthID==3),]


for (i in 1:dim(listing_jan)[1]){ if (listing_jan$ModifiedListing[i] == 1) {
r$SETBIT("ModificationsJanuary", i,"1")}}

r$BITCOUNT("ModificationsJanuary")
```

1.2 Question

In this question, we should find the users who did not modified their listing in January. Therefore, we should the inverse of the above question.

```
r$BITOP("NOT","NotModificationsJanuary","ModificationsJanuary")

r$BITCOUNT("NotModificationsJanuary")
```

The result was that 10031 users did not modified their listing on January. We know that dataset with the January listings has 19999 observations but if we add the result of the bitcounts we find 20000. The reason why we have one more observation is that bitop perform in bit and bitcount perform in bytes. We know that 1 byte was 8 bit. Therefore, 19999 observations is not create equal groups of eight values and due to fact that add one more value in order to transpose the last bits in one byte.

1.3 Question

In this question, we should define the users who received at least one e-mail per month. In order to answer this question we create new datasets which give the total number of email that a user received by month. So, in the final dataset we have unique users, months and the sum of the email that he received. Then we create three BITMAPS one for each month (“EmailsJanuary”, “EmailsFebruary”, “EmailsMarch”). Redis set one in the positions that a user received at least one email at January. The same do for the rest months. Final, to calculate the total users who received at one email in January, February and March. Therefore, we use BITOP AND to find that 2668 users received at least one email.

```
email_jan<-as.data.frame(table(emails$UserID[emails$MonthID==1]))
email_feb<-as.data.frame(table(emails$UserID[emails$MonthID==2]))
email_mar<-as.data.frame(table(emails$UserID[emails$MonthID==3]))
colnames(email_jan)<-c("user_id","received_email_jan")
colnames(email_feb)<-c("user_id","received_email_feb")
colnames(email_mar)<-c("user_id","received_email_mar")
monthly_email<-
merge(email_mar,merge(email_jan,email_feb,by="user_id",all.x=T),by="user_id",all.x = T)
monthly_email[is.na(monthly_email)]<-0

for(i in 1:dim(monthly_email)[1]) {
  if (monthly_email$received_email_jan[i]>0){
    r$SETBIT ("EmailsJanuary",i, "1")}
}
for(i in 1:dim(monthly_email)[1]) {
  if (monthly_email$received_email_feb[i]>0){
    r$SETBIT ("EmailsFebruary",i, "1")}
}
for(i in 1:dim(monthly_email)[1]) {
  if (monthly_email$received_email_mar[i]>0){
    r$SETBIT ("EmailsMarch",i, "1")}
}

r$BITOP("AND","At_least_one_email",c("EmailsJanuary","EmailsFebruary","EmailsMarch"))
r$BITCOUNT("At_least_one_email")
```

1.4 Question

In this question, we want to find the users who do not receive an email at February but he received in March and January. Therefore, we combine the information from the bitmaps we create in the previous question and we create a new one that specifically set one in the users do not receive email in February. Then, we make a bitmap that set 1 in the users that receive email in January and March but not in February. With the use of bitcount we see that 2417 users receive email at January and March but not in February

```
r$BITOP("NOT","NotEmailsFebruary","EmailsFebruary")
r$BITOP("AND","NotFebruary",c("EmailsJanuary","NotEmailsFebruary","EmailsMarch"))
```

```
r$BITCOUNT("NotFebruary")
```

1.5 Question

In this question, we should calculate how many users received an e-mail on January that they did not open but they updated their listing anyway. Before we continue in the solution of the question and the next remaining questions we create a new dataset that contain the unique users and the information if opened, received the email. We set zero if he does not receive any email and one if receive one or more emails. For the new dataset, we run the below code:

```
received_email<-as.data.frame(table(emails$UserID,emails$MonthID))
colnames(received_email)<-c("user_id","month","received")
opened_email<-
as.data.frame(table(emails$UserID[emails$EmailOpened==1],emails$MonthID[emails$Email
Opened==1]))
colnames(opened_email)<-c("user_id","month","opened")
total_emails<-merge(received_email,opened_email,by=c("user_id","month"),all.x=T)
length(which(is.na(total_emails$opened)))
total_emails<-total_emails[total_emails$received>0,]
total_emails$opened[is.na(total_emails$opened)]<-0
total_emails$opened_binary<-ifelse(total_emails$opened>0,1,0)
```

To continue in the final calculations we separate the dataset for each given months.

```
new_email_jan<-total_emails[total_emails$month==1,]
new_email_feb<-total_emails[total_emails$month==2,]
new_email_mar<-total_emails[total_emails$month==3,]
```

We create the bitmap "EmailsOpenedJanuary" for every user who received an email in January Redis set 1 if they opened at least one of them. To find if they modified their listing without opened the email we create the bitmat which set one in the users who did not opened the email in the January ("NotEmailsOpenedJanuary"). Then, we use the bitmap, which we have create for the first question "ModificationsJanuary" and create a new bitmap that set one in the users who modified their listing but did not opened the email. In total are 2006 users who satisfied the above condition.

```
for (i in 1:dim(new_email_jan)[1]){
  if (new_email_jan$opened_binary[i]==1){
    r$SETBIT("EmailsOpenedJanuary", i, "1")}}

r$BITOP("NOT","NotEmailsOpenedJanuary", "EmailsOpenedJanuary")
r$BITOP("AND","Not_open_Email_January_but_update",c("NotEmailsOpenedJanuary","Mo
dificationsJanuary"))
r$BITCOUNT("Not_open_Email_January_but_update")
```

1.6 Question

In this question, we want to find the users who received an e-mail on January and did not open it but they updated their listing or if they received an e-mail on February and did not open but they updated their listing or if they received an e-mail on March and did not open but they updated their listing.

Therefore, we create two news bitmaps that set one in every user that received an email in February (“EmailsOpenedFebruary”) and opened at least one of them the same for March (“EmailsOpenedMarch”). Then, we create the bitmaps that show the modifications that happened in February and March. To calculate the final bitmap we calculate the bitmaps that set one in the users who opened their email and update their listing for February and March. In the last step, we combine all the bitmaps that have one in the users that modified their listing without open their email. The users who satisfy the above conditions are 4768.

```
for (i in 1:dim(new_email_feb)[1]){

  if (new_email_feb$opened_binary[i]==1){  r$SETBIT("EmailsOpenedFebruary", i, "1")}}

r$BITOP("NOT","NotEmailsOpenedFebruary", "EmailsOpenedFebruary")

for (i in 1:dim(new_email_mar)[1]){  if (new_email_mar$opened_binary[i]==1){

    r$SETBIT("EmailsOpenedMarch", i, "1")}}

r$BITOP("NOT","NotEmailsOpenedMarch", "EmailsOpenedMarch")


for (i in 1:dim(listing_feb)[1]){

  if (listing_feb$ModifiedListing[i] == 1) {r$SETBIT("ModificationsFebruary", i,"1")}}

for (i in 1:dim(listing_mar)[1]){if (listing_mar$ModifiedListing[i] == 1) {

    r$SETBIT("ModificationsMarch", i,"1")}}


r$BITOP("AND","Not_open_Email_Feb_but_update",c("NotEmailsOpenedFebruary","ModificationsFebruary"))

r$BITCOUNT("Not_open_Email_Feb_but_update")

r$BITOP("AND","Not_open_Email_Mar_but_update",c("NotEmailsOpenedMarch",
"ModificationsMarch"))

r$BITCOUNT("Not_open_Email_Mar_but_update")


r$BITOP("OR","Updatewithoutopen",c("Not_open_Email_Mar_but_update",
"Not_open_Email_Feb_but_update","Not_open_Email_January_but_update"))

r$BITCOUNT("Updatewithoutopen")
```

1.7 Question

To answer the question if make sense to continue send emails to the users we calculate the bitmaps that set one in the users that opened their email and update their listing. The opposite number we find in the previous question. Ideally, if the users who modified and opened their emails are more than the users who modified without opened their email. In the previous question, we found 4768 comparison to 6247 users that modified and opened their email. Therefore, we understand that it make sense to continue send email in the users.

```
r$BITOP("AND","open_Email_Jan_but_update",
c("EmailsOpenedJanuary","ModificationsJanuary"))

r$BITCOUNT("open_Email_Jan_but_update")

r$BITOP("AND","open_Email_Feb_but_update",
c("EmailsOpenedFebruary","ModificationsFebruary"))

r$BITCOUNT("open_Email_Feb_but_update")

r$BITOP("AND","open_Email_Mar_but_update", c("EmailsOpenedMarch",
"ModificationsMarch"))

r$BITCOUNT("open_Email_Mar_but_update")

r$BITOP("OR","Updateandopen",c("open_Email_Jan_but_update",
"open_Email_Feb_but_update","open_Email_Mar_but_update"))

r$BITCOUNT("Updateandopen")
```