Data Wrangling QBS 181

Instructions:

* Print random 10 rows to show these changes. Please also include the code
* Use a RMarkdown file to write your code and output answers.
* You may collaborate on the homework but you must write it up yourselves.
* Tables should be pulled into R using ODBC connection strings
* Please save the files in this format:
* Name of the file should be username(database)\_HW2\_QBS181
* Upload on to Canvas.

1. Create a new column “Enrollment group” in the table Phonecall\_Encounter
2. Insert EnrollmentGroup=Clinical Alert :code is 125060000
3. Insert EnrollmentGroup =Health Coaching :code is 125060001
4. Insert EnrollmentGroup =Technixal Question: Code is 125060002
5. Insert EnrollmentGroup =Administrative: Code is 125060003
6. Insert EnrollmentGroup =Other: Code is 125060004
7. Insert EnrollmentGroup =Lack of engagement : Code is 125060005
8. Obtain the # of records for each enrollment group
9. Merge the Phone call encounter table with Call duration table.
10. Find out the # of records for different call outcomes and call type. Use 1-Inbound and 2-Outbound, for call types; use 1-No response,2-Left voice mail and 3 successful. Please also find the call duration for each of the enrollment groups
11. Merge the tables Demographics, Conditions and TextMessages. Find the # of texts/per week, by the type of sender.
12. Obtain the count of texts based on the chronic condition over a period of time (say per week).

A picture containing graphical user interface

Description automatically generated

Graphical user interface, application, table

Description automatically generated

A picture containing text, table

Description automatically generated

library("RODBC")

library(sqldf)

library(stringr)

myconnection <- odbcConnect("dartmouth2","yzhao","yzhao@qbs181")

#Question1:

IC\_Phonecall\_Encounter<-sqlQuery(myconnection ,"select\*from PhoneCall\_Encounter")

View(IC\_Phonecall\_Encounter)

IC\_Phonecall\_Encounter$Enrollmentgroup<-as.character(IC\_Phonecall\_Encounter$EncounterCode)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060000","Clinical Alert",IC\_Phonecall\_Encounter$Enrollmentgroup)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060001","Health Coaching",IC\_Phonecall\_Encounter$Enrollmentgroup)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060002","Technixal Question",IC\_Phonecall\_Encounter$Enrollmentgroup)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060003","Administrative",IC\_Phonecall\_Encounter$Enrollmentgroup)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060004","Other",IC\_Phonecall\_Encounter$Enrollmentgroup)

IC\_Phonecall\_Encounter$Enrollmentgroup<-sub("125060005","Lack of engagement",IC\_Phonecall\_Encounter$Enrollmentgroup)

View(IC\_Phonecall\_Encounter)

#Question2:

Clinical\_Alert<-grep("Clinical Alert",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Clinical\_Alert)

#453

Health\_Coaching<-grep("Health Coaching",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Health\_Coaching)

#409

Technixal\_Question<-grep("Technixal Question",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Technixal\_Question)

#1059

Administrative<-grep("Administrative",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Administrative)

#4480

Other<-grep("Other",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Other)

#189

Lack\_of\_engagement<-grep("Lack of engagement",IC\_Phonecall\_Encounter$Enrollmentgroup, value = TRUE)

length(Lack\_of\_engagement)

#1824

#Question3:

Call\_Duration<-sqlQuery(myconnection ,"select\*from CallDuration")

Call\_Duration<-as.data.frame(Call\_Duration)

View(Call\_Duration)

merge\_dataframe<-sqldf("select\*from IC\_Phonecall\_Encounter e inner join Call\_Duration c on e.Customerid = c.tri\_CustomerIDEntityReference")

View(merge\_dataframe)

#Question4:

merge\_dataframe$CallOutcome<-as.character(merge\_dataframe$CallOutcome)

merge\_dataframe$CallOutcome<-sub("1","Inbound",merge\_dataframe$CallOutcome)

call\_outcomes\_inbound<-grep("Inbound",merge\_dataframe$CallOutcome, value = TRUE)

length(call\_outcomes\_inbound)

#5201

merge\_dataframe$CallOutcome<-sub("2","Outbound",merge\_dataframe$CallOutcome)

call\_outcomes\_outbound<-grep("Outbound",merge\_dataframe$CallOutcome, value = TRUE)

length(call\_outcomes\_outbound)

#4739

merge\_dataframe$CallType<-as.character(merge\_dataframe$CallType)

merge\_dataframe$CallType<-sub("1","No response",merge\_dataframe$CallType)

Call\_Type\_No\_response<-grep("No response",merge\_dataframe$CallType, value = TRUE)

length(Call\_Type\_No\_response)

#9872

merge\_dataframe$CallType<-sub("2","Left voice mail",merge\_dataframe$CallType)

Call\_Type\_Left\_voice\_mail <-grep("Left voice mail",merge\_dataframe$CallType, value = TRUE)

length(Call\_Type\_Left\_voice\_mail)

#832

merge\_dataframe$CallType<-sub("3","successful",merge\_dataframe$CallType)

Call\_Type\_successful <-grep("successful",merge\_dataframe$CallType, value = TRUE)

length(Call\_Type\_successful)

#0

Call\_Duration\_Clinical\_Alert<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'Clinical Alert'")

#268958

Call\_Duration\_H<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'H%' ")

#267033

Call\_Duration\_T<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'T%' ")

#442126

Call\_Duration\_A<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'A%' ")

#708638

Call\_Duration\_O<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'O%' ")

#200487

Call\_Duration\_L<-sqldf("select sum(CallDuration) from merge\_dataframe where Enrollmentgroup like'L%' ")

#380474

#Question5

#Text message = Text table?

#repetition, distinctive?

IC\_Demo<-sqlQuery(myconnection,"select distinct \*from Demographics")

IC\_Condition<-sqlQuery(myconnection,"select distinct\*from Conditions")

IC\_Text<-sqlQuery(myconnection,"select distinct \*from Text")

merge\_three\_table <- sqldf("select\* from IC\_Demo d inner join IC\_Condition c on d.contactid = c.tri\_patientid inner join IC\_Text t on d.contactid = t.tri\_contactid")

View(merge\_three\_table)

numbers\_of\_week<-(max(merge\_three\_table$TextSentDate) - min(merge\_three\_table$TextSentDate))/7

#51.71429weeks

type\_of\_sneder<-sqldf("select distinct (SenderName) from merge\_three\_table")

number\_of\_clinicians<-sqldf("select count(SenderName) from merge\_three\_table where SenderName like 'Cli%'")

number\_of\_clinicians/51.71429

#136.674

number\_of\_customer<-sqldf("select count(SenderName) from merge\_three\_table where SenderName like 'Cu%'")

number\_of\_customer/51.71429

#99.33425

number\_of\_Sysyem<-sqldf("select count(SenderName) from merge\_three\_table where SenderName like 'S%'")

number\_of\_Sysyem/51.71429

#704.7375

#Question 6:

type\_of\_chronic\_condition<-sqldf("select distinct tri\_name from merge\_three\_table")

type\_of\_chronic\_condition

number\_of\_Activity\_Monitoring<-sqldf("select count(tri\_name) from merge\_three\_table where tri\_name like 'Acti%'")

number\_of\_Activity\_Monitoring/51.71429

#679.6187

number\_of\_Hypertension<-sqldf("select count(tri\_name) from merge\_three\_table where tri\_name like 'Hyper%'")

number\_of\_Hypertension/51.71429

#159.7237

number\_of\_Diabetes<-sqldf("select count(tri\_name) from merge\_three\_table where tri\_name like 'Diab%'")

number\_of\_Diabetes/51.71429

#44.91989

number\_of\_COPD<-sqldf("select count(tri\_name) from merge\_three\_table where tri\_name like 'COPD%'")

number\_of\_COPD/51.71429

#29.70165

number\_of\_Congestive\_Heart\_Failure<-sqldf("select count(tri\_name) from merge\_three\_table where tri\_name like 'Conges%'")

number\_of\_Congestive\_Heart\_Failure/51.71429

#26.78177