Mobile\_Analytics.R

DoryChen

2019-05-11

setwd("/Users/DoryChen/Desktop")  
  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

Geo\_fence <- read.csv("Geo-Fence Analytics.csv", header=T)  
  
#Create variable: "imp\_large", "cat\_entertainment", "cat\_social", "cat\_tech", "os\_ios", "distance"  
  
attach(Geo\_fence)  
  
Geo\_fence$imp\_large <- ifelse(imp\_size == "728x90", 1, 0)  
Geo\_fence$cat\_entertainment <-ifelse(app\_topcat == "IAB1" | app\_topcat == "IAB1-6", 1, 0)  
Geo\_fence$cat\_social <- ifelse(app\_topcat == "IAB14", 1, 0)  
Geo\_fence$cat\_tech <- ifelse(app\_topcat == "IAB19-6", 1, 0)  
Geo\_fence$os\_ios <- ifelse(device\_os == "iOS", 1, 0)  
  
## Create variable distance using Harvesine formula to calculate the distance   
## for a pair of latitude/longitude coordinates.  
  
library(aspace)

## Loading required package: splancs

## Loading required package: sp

##   
## Spatial Point Pattern Analysis Code in S-Plus  
##   
## Version 2 - Spatial and Space-Time analysis

##   
## Attaching package: 'splancs'

## The following object is masked from 'package:dplyr':  
##   
## tribble

## Loading required package: Hmisc

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##   
## Attaching package: 'Hmisc'

## The following object is masked from 'package:splancs':  
##   
## zoom

## The following objects are masked from 'package:dplyr':  
##   
## src, summarize

## The following objects are masked from 'package:base':  
##   
## format.pval, units

## Loading required package: shapefiles

## Loading required package: foreign

##   
## Attaching package: 'shapefiles'

## The following objects are masked from 'package:foreign':  
##   
## read.dbf, write.dbf

Geo\_fence$distance <- 6371 \* acos(cos(as\_radians(device\_lat)) \* cos(as\_radians(geofence\_lat))   
 \* cos(as\_radians(device\_lon) - as\_radians(geofence\_lon))   
 + sin(as\_radians(device\_lat)) \* sin(as\_radians(geofence\_lat)))  
  
#Create distance group and calculate click-through-rate  
attach(Geo\_fence)

## The following objects are masked from Geo\_fence (pos = 13):  
##   
## app\_id, app\_name, app\_pub, app\_review\_val, app\_review\_vol,  
## app\_topcat, device\_lat, device\_lon, device\_os, device\_zip,  
## didclick, geofence\_lat, geofence\_lon, gepfence\_radius,  
## imp\_size

Geo\_fence$distance\_group <-ifelse(between(distance, 0, 0.5), 1,   
 ifelse(between(distance, 0.5, 1), 2,   
 ifelse(between(distance, 1, 2), 3,   
 ifelse(between(distance, 2, 4), 4,   
 ifelse(between(distance, 4, 7), 5,  
 ifelse(between(distance, 7, 10), 6,   
 ifelse(distance >10, 7, 0)))))))  
  
  
library("sqldf")

## Loading required package: gsubfn

## Loading required package: proto

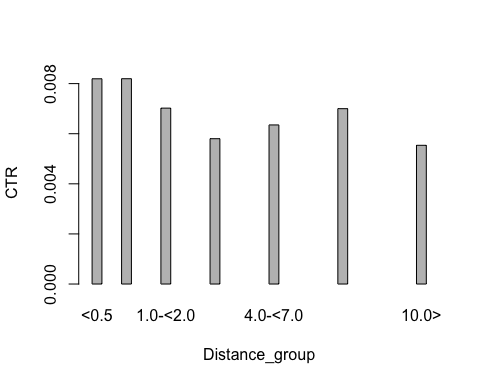
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library/Frameworks/R.framework/Resources/modules//R\_X11.so':  
## dlopen(/Library/Frameworks/R.framework/Resources/modules//R\_X11.so, 6): Library not loaded: /opt/X11/lib/libSM.6.dylib  
## Referenced from: /Library/Frameworks/R.framework/Resources/modules//R\_X11.so  
## Reason: image not found

## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):  
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/  
## Resources/library/tcltk/libs//tcltk.so'' had status 1

## Could not load tcltk. Will use slower R code instead.

## Loading required package: RSQLite

mean\_ctr <- sqldf("select distance, distance\_group, avg(didclick) as mean\_didclick   
 from Geo\_fence group by distance\_group")  
  
barplot(mean\_ctr$mean\_didclick, mean\_ctr$distance\_group,   
 xlab="Distance\_group", ylab="CTR",  
 names.arg=c("<0.5", "0.5-<1.0", "1.0-<2.0", "2.0-<4.0", "4.0-<7.0", "7.0-<10.0","10.0>"),width =500  
)



#Create variables "distance\_squared", "ln\_app\_review\_vol"  
Geo\_fence$ln\_app\_review\_vol <- log(app\_review\_vol)  
Geo\_fence$distance\_squared <- distance^2  
  
# Descriptive Statisitcs  
  
attach(Geo\_fence)

## The following objects are masked from Geo\_fence (pos = 7):  
##   
## app\_id, app\_name, app\_pub, app\_review\_val, app\_review\_vol,  
## app\_topcat, cat\_entertainment, cat\_social, cat\_tech,  
## device\_lat, device\_lon, device\_os, device\_zip, didclick,  
## distance, geofence\_lat, geofence\_lon, gepfence\_radius,  
## imp\_large, imp\_size, os\_ios

## The following objects are masked from Geo\_fence (pos = 18):  
##   
## app\_id, app\_name, app\_pub, app\_review\_val, app\_review\_vol,  
## app\_topcat, device\_lat, device\_lon, device\_os, device\_zip,  
## didclick, geofence\_lat, geofence\_lon, gepfence\_radius,  
## imp\_size

statfun<- function(x){  
 data.frame(Mean=mean(x),Median=median(x),STDEV=sd(x),Min=min(x),Max=max(x))  
}  
  
  
sum\_table<-rbind(statfun(didclick),  
 statfun(distance),  
 statfun(imp\_large),  
 statfun(cat\_entertainment),  
 statfun(cat\_social),  
 statfun(cat\_tech),  
 statfun(os\_ios),  
 statfun(ln\_app\_review\_vol),  
 statfun(app\_review\_val)  
)  
row.names(sum\_table)<-c("didclick", "distance", "imp\_large", "cat\_entertainment",   
 "cat\_social", "cat\_tech","os\_ios","ln\_app\_vol","app\_val")  
print(sum\_table)

## Mean Median STDEV Min Max  
## didclick 0.006811059 0.000000 0.08224794 0.00000000 1.00000  
## distance 2.983737139 2.020864 2.64852620 0.02075894 11.78666  
## imp\_large 0.230876800 0.000000 0.42139550 0.00000000 1.00000  
## cat\_entertainment 0.283925736 0.000000 0.45090308 0.00000000 1.00000  
## cat\_social 0.125124417 0.000000 0.33086130 0.00000000 1.00000  
## cat\_tech 0.517846126 1.000000 0.49968347 0.00000000 1.00000  
## os\_ios 0.250363997 0.000000 0.43322443 0.00000000 1.00000  
## ln\_app\_vol 10.056798904 10.087225 0.63696194 7.08086790 12.93770  
## app\_val 3.654872622 3.400000 0.36081251 1.40000000 4.70000

#Correlations between the variables  
library(corrgram)

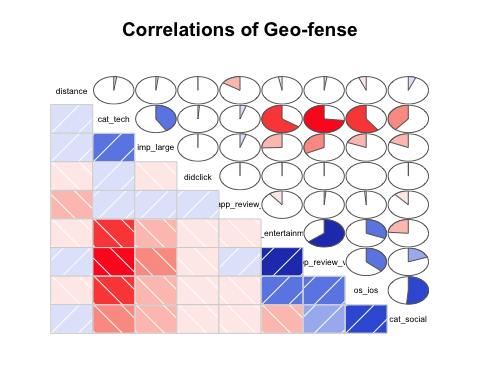
##   
## Attaching package: 'corrgram'

## The following object is masked from 'package:lattice':  
##   
## panel.fill

cor\_geofence <- subset(Geo\_fence, select = c(didclick, distance, imp\_large, cat\_entertainment, cat\_social, cat\_tech, os\_ios, ln\_app\_review\_vol, app\_review\_val))  
cor(cor\_geofence)

## didclick distance imp\_large cat\_entertainment  
## didclick 1.000000000 -0.006628356 -0.004786218 -0.007117972  
## distance -0.006628356 1.000000000 0.020024918 -0.028992663  
## imp\_large -0.004786218 0.020024918 1.000000000 -0.254731873  
## cat\_entertainment -0.007117972 -0.028992663 -0.254731873 1.000000000  
## cat\_social -0.005623417 0.060484490 -0.185311155 -0.238133905  
## cat\_tech 0.012454366 0.023499545 0.414049273 -0.652575678  
## os\_ios -0.002147325 -0.060281389 -0.190194050 0.312647684  
## ln\_app\_review\_vol 0.003982875 -0.157864184 0.049929790 -0.105545185  
## app\_review\_val -0.006523592 0.022481133 -0.321439020 0.642212363  
## cat\_social cat\_tech os\_ios ln\_app\_review\_vol  
## didclick -0.005623417 0.01245437 -0.002147325 0.003982875  
## distance 0.060484490 0.02349954 -0.060281389 -0.157864184  
## imp\_large -0.185311155 0.41404927 -0.190194050 0.049929790  
## cat\_entertainment -0.238133905 -0.65257568 0.312647684 -0.105545185  
## cat\_social 1.000000000 -0.39192721 0.513672844 -0.115376574  
## cat\_tech -0.391927215 1.00000000 -0.598919227 0.049503835  
## os\_ios 0.513672844 -0.59891923 1.000000000 -0.013523794  
## ln\_app\_review\_vol -0.115376574 0.04950383 -0.013523794 1.000000000  
## app\_review\_val 0.194394425 -0.73206714 0.366139311 0.014457854  
## app\_review\_val  
## didclick -0.006523592  
## distance 0.022481133  
## imp\_large -0.321439020  
## cat\_entertainment 0.642212363  
## cat\_social 0.194394425  
## cat\_tech -0.732067145  
## os\_ios 0.366139311  
## ln\_app\_review\_vol 0.014457854  
## app\_review\_val 1.000000000

corrgram(cor\_geofence, order = TRUE, lower.panel = panel.shade,  
 upper.panel = panel.pie, text.panel = panel.txt,  
 main = "Correlations of Geo-fense")



#Scatterplot of distance and click-through-rate  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

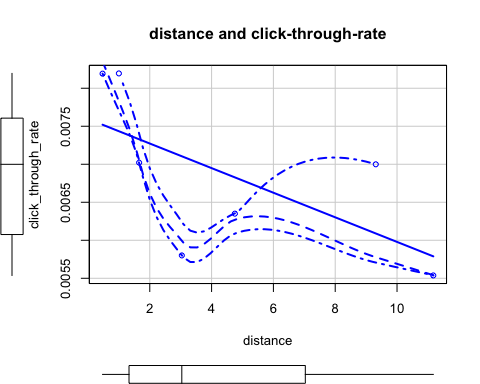
## The following object is masked from 'package:dplyr':  
##   
## recode

attach(mean\_ctr)

## The following objects are masked from Geo\_fence (pos = 6):  
##   
## distance, distance\_group

## The following object is masked from Geo\_fence (pos = 11):  
##   
## distance

scatterplot(mean\_didclick ~ distance, xlab = "distance",   
 ylab = "click\_through\_rate", main = "distance and click-through-rate")



#We found that the closer mobile device to geofence, the higher the click-through-rate.  
  
#Logistic regression of "didclick"  
fit.didclick <- glm(didclick ~ distance + imp\_large + cat\_entertainment  
 + cat\_social + cat\_tech + os\_ios + ln\_app\_review\_vol + app\_review\_val,  
 data = Geo\_fence, family = binomial())  
summary(fit.didclick)

##   
## Call:  
## glm(formula = didclick ~ distance + imp\_large + cat\_entertainment +   
## cat\_social + cat\_tech + os\_ios + ln\_app\_review\_vol + app\_review\_val,   
## family = binomial(), data = Geo\_fence)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.1439 -0.1307 -0.1144 -0.1041 3.4177   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.76438 0.89749 -7.537 4.81e-14 \*\*\*  
## distance -0.02880 0.01425 -2.020 0.043333 \*   
## imp\_large -0.34794 0.09173 -3.793 0.000149 \*\*\*  
## cat\_entertainment -0.09613 0.17893 -0.537 0.591087   
## cat\_social -0.22237 0.21137 -1.052 0.292777   
## cat\_tech 0.69306 0.17660 3.924 8.70e-05 \*\*\*  
## os\_ios 0.38421 0.12634 3.041 0.002357 \*\*   
## ln\_app\_review\_vol 0.03263 0.06301 0.518 0.604590   
## app\_review\_val 0.32515 0.18716 1.737 0.082337 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9912.5 on 121566 degrees of freedom  
## Residual deviance: 9861.4 on 121558 degrees of freedom  
## AIC: 9879.4  
##   
## Number of Fisher Scoring iterations: 8

fit.didclick\_distance\_sqrt <- glm(didclick ~ distance + distance\_squared + imp\_large + cat\_entertainment  
 + cat\_social + cat\_tech + os\_ios + ln\_app\_review\_vol + app\_review\_val,  
 data = Geo\_fence, family = binomial())  
summary(fit.didclick\_distance\_sqrt)

##   
## Call:  
## glm(formula = didclick ~ distance + distance\_squared + imp\_large +   
## cat\_entertainment + cat\_social + cat\_tech + os\_ios + ln\_app\_review\_vol +   
## app\_review\_val, family = binomial(), data = Geo\_fence)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.1510 -0.1272 -0.1148 -0.1042 3.4025   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.616818 0.896958 -7.377 1.62e-13 \*\*\*  
## distance -0.117927 0.045073 -2.616 0.008888 \*\*   
## distance\_squared 0.009166 0.004362 2.102 0.035583 \*   
## imp\_large -0.352164 0.091782 -3.837 0.000125 \*\*\*  
## cat\_entertainment -0.096143 0.178941 -0.537 0.591069   
## cat\_social -0.226695 0.211394 -1.072 0.283550   
## cat\_tech 0.687657 0.176312 3.900 9.61e-05 \*\*\*  
## os\_ios 0.385895 0.126361 3.054 0.002259 \*\*   
## ln\_app\_review\_vol 0.030512 0.063037 0.484 0.628368   
## app\_review\_val 0.323831 0.186656 1.735 0.082757 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9912.5 on 121566 degrees of freedom  
## Residual deviance: 9857.1 on 121557 degrees of freedom  
## AIC: 9877.1  
##   
## Number of Fisher Scoring iterations: 8

fit.didclick\_reduced <- glm(didclick ~ distance + distance\_squared + imp\_large   
 + cat\_tech + os\_ios,  
 data = Geo\_fence, family = binomial())  
summary(fit.didclick\_reduced)

##   
## Call:  
## glm(formula = didclick ~ distance + distance\_squared + imp\_large +   
## cat\_tech + os\_ios, family = binomial(), data = Geo\_fence)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.1471 -0.1254 -0.1152 -0.1045 3.3277   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.091004 0.110148 -46.220 < 2e-16 \*\*\*  
## distance -0.119992 0.045032 -2.665 0.007708 \*\*   
## distance\_squared 0.009158 0.004364 2.099 0.035858 \*   
## imp\_large -0.346716 0.091717 -3.780 0.000157 \*\*\*  
## cat\_tech 0.574871 0.100695 5.709 1.14e-08 \*\*\*  
## os\_ios 0.292210 0.111257 2.626 0.008628 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 9912.5 on 121566 degrees of freedom  
## Residual deviance: 9862.8 on 121561 degrees of freedom  
## AIC: 9874.8  
##   
## Number of Fisher Scoring iterations: 8