In this assignment, there are three classifications chosen to test and given data to show geo dependent data

Classifications are :

Naive Bayes Classification

Decision Tree Classification

Logistic Regression Classification

Import data

```
data<-read.csv2("Data_Projects.csv", sep=";", dec=".");
```

```
summary(data)
```

```
 п.їAddressCount   CallsCount     ClicksCount       FirmsCount
 Min.   :   9    Min.   :   20   Min.   :   258   Min.   :  14.0
 1st Qu.:  81    1st Qu.:  346   1st Qu.:  2055   1st Qu.:  71.5
 Median : 371    Median :  931   Median :  6921   Median : 185.0
 Mean   :1048    Mean   : 3649   Mean   : 21826   Mean   : 305.1
 3rd Qu.:1195    3rd Qu.: 2458   3rd Qu.: 30626   3rd Qu.: 402.5
 Max.   :9552    Max.   :48497   Max.   :167155   Max.   :2379.0

               GeoPart                MobilePart   UsersCount
 0,0929166666666667: 1   0,09             : 1   Min.   :  157
 0,137857900318134 : 1   0,133974358974359: 1   1st Qu.: 1168
 0,151738923296808 : 1   0,139612188365651: 1   Median : 2934
 0,187886279357231 : 1   0,175525339925834: 1   Mean   : 9753
 0,193484698914116 : 1   0,200644166213028: 1   3rd Qu.:13265
 0,203374777975133 : 1   0,204181869211339: 1   Max.   :61127
 (Other)           :73   (Other)          :73
             Distance      IsGeo
 1004,78676794652: 1   Min.   :0.0000
 1033,11276489631: 1   1st Qu.:0.0000
 1234,54344844949: 1   Median :0.0000
 1421,72399039962: 1   Mean   :0.3544
 1423,37651183958: 1   3rd Qu.:1.0000
 1437,3055534143 : 1   Max.   :1.0000
 (Other)         :73
```

Bayes Classification

```
library(e1071)
```

```
nb<-naiveBayes(IsGeo~Distance+ClicksCount, data=data)
```

```
nb
```

```
Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
```

A-priori probabilities:
Y
        0   1
0.6455696 0.3544304

Conditional probabilities:
   Distance
Y   1004,78676794652 1033,11276489631 1234,54344844949 1421,72399039962
  0   0.00000000   0.01960784   0.00000000   0.01960784
  1   0.03571429   0.00000000   0.03571429   0.00000000
   Distance
Y   1423,37651183958 1437,3055534143 1463,57525286792 1547,69332527551
  0   0.00000000   0.00000000   0.00000000   0.01960784
  1   0.03571429   0.03571429   0.03571429   0.00000000
   Distance
Y   1576,51415402623 1600,66532620868 1661,51688251154 1673,89413344887
  0   0.00000000   0.00000000   0.00000000   0.00000000
  1   0.03571429   0.03571429   0.03571429   0.03571429
   Distance
Y   1703,52647696896 1710,21538498727 1777,74626458405 1908,10811023819
  0   0.01960784   0.00000000   0.00000000   0.01960784
  1   0.00000000   0.03571429   0.03571429   0.00000000
   Distance
Y   1971,79546984902 2071,94684987015 2080,06825959411 2110,30315695341
  0   0.01960784   0.01960784   0.01960784   0.00000000
  1   0.00000000   0.00000000   0.00000000   0.03571429
   Distance
Y   2173,7621606007 2186,3699144222 2276,31335761423 2278,16291706146
  0   0.00000000   0.01960784   0.00000000   0.01960784
  1   0.03571429   0.00000000   0.03571429   0.00000000
   Distance
Y   2289,03242434015 2337,60377736899 2519,15128538222 2586,50327354116
  0   0.01960784   0.01960784   0.01960784   0.00000000
  1   0.00000000   0.00000000   0.00000000   0.03571429
   Distance
Y   2635,88425199653 2642,20016520197 2704,49040063268 2705,56818332971
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance
Y   2721,00169183967 2766,84717399882 2775,01815203855 2780,96178910644
  0   0.01960784   0.00000000   0.01960784   0.01960784
  1   0.00000000   0.03571429   0.00000000   0.00000000
   Distance
Y   2949,85779621301 2959,44423827871 2995,88860289814 3038,6587528867
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance
Y   3061,18416680401 3061,77121200376 3201,60026197197 3356,11048392976
  0   0.01960784   0.01960784   0.01960784   0.00000000
  1   0.00000000   0.00000000   0.00000000   0.03571429
   Distance
Y   3396,56608856838 3488,72029249463 3497,97938508119 3653,40527714029
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance
Y   3746,18547386496 3772,13900324323 3861,25409028014 3867,67539940448
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance
Y   3969,15036175422 4143,34291375289 4149,00290235257 4184,7484953409
  0   0.00000000   0.01960784   0.01960784   0.01960784
  1   0.03571429   0.00000000   0.00000000   0.00000000
   Distance
Y   4279,20794032879 4333,98969073498 4511,6323033575 4522,58589659175
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance
Y   4926,34877353678 5206,83845433085 5509,38128978019 5528,0655460862
  0   0.01960784   0.01960784   0.01960784   0.01960784
  1   0.00000000   0.00000000   0.00000000   0.00000000
   Distance

```
Y   5967,70900103466 6272,75183998062 6292,20731138509 714,787236203679
  0       0.01960784       0.01960784       0.01960784       0.00000000
  1       0.00000000       0.00000000       0.00000000       0.03571429
   Distance
Y   749,966084023684 762,640024028404 813,439315154229 854,172080438245
  0       0.00000000       0.01960784       0.00000000       0.00000000
  1       0.03571429       0.00000000       0.03571429       0.03571429
   Distance
Y   878,013624111711 893,981377311841 918,713971923729 922,085394259457
  0       0.00000000       0.00000000       0.00000000       0.00000000
  1       0.03571429       0.03571429       0.03571429       0.03571429
   Distance
Y   938,766030460995 952,012290304611 989,412138569643
  0       0.00000000       0.00000000       0.01960784
  1       0.03571429       0.03571429       0.00000000

   ClicksCount
Y       [,1]     [,2]
  0 14460.80 27787.41
  1 35241.21 36437.64
```

```
predictValue <- predict(nb, data[1:40, -40], type="raw")
```

```
data[1:40, -40]
```

п.їAddressCount	CallsCount	ClicksCount	FirmsCount	GeoPart	MobilePart	UsersCount	Distance	IsGeo
156	20	1903	176	0,416104363472785	0,535762483130904	1125	749,966084023684	1
17	37	258	20	0,211678832116788	0,430656934306569	157	2289,032424340 15	0
78	56	1956	185	0,349475383373688	0,476594027441485	1195	1423,376511839 58	1
14	70	378	19	0,318718381112985	0,463743676222597	206	3396,566088568 38	0
111	90	4089	90	0,556175452096 96	0,490573297422085	2934	1576,514154026 23	1
53	96	1669	162	0,398978169995355	0,421737111007896	991	2337,603777368 99	0
18	137	373	26	0,407108239095315	0,389337641357027	180	2780,961789106 44	0
9	150	563	15	0,187886279357231	0,175525339925834	265	5206,838454330 85	0
56	153	1366	56	0,335469107551487	0,498855835240275	613	2704,490400632 68	0
371	176	5913	31	0,288145750928922	0,608773822366055	3981	4333,989690734 98	0
9552	207	101558	1473	0,383506984703 22	0,705315091276917	54098	813,4393151542 29	1
42	222	1154	64	0,223268698060942	0,139612188365651	410	4511,632303357 5	0
84	228	1031	39	0,310968494749125	0,237456242707118	578	6272,751839980 62	0
48	237	1307	14	0,203374777975133	0,669182948490231	596	3969,150361754 22	1
30	256	1139	114	0,318485523385301	0,231625835189 31	550	4149,002902352 57	0
70	270	976	132	0,246794871794872	0,133974358974359	537	4522,585896591 75	0
73	278	2805	42	0,322341513292434	0,440950920245399	1196	2775,018152038 55	0
258	310	3970	245	0,313795034533	0,458652230722	2344	893,98137731	1

п.їAddressCount	CallsCount	ClicksCount	FirmsCount	GeoPart	MobilePart	UsersCount	Distance	IsGeo
				4254	6153		1841	
61	319	1316	168	0,367379679144385	0,34331550802139	558	1033,11276489631	0
66	344	1478	35	0,092916666666667	0,09	500	5528,0655460862	0
180	348	2969	84	0,266934942991281	0,462105969148223	1358	2635,88425199653	0
45	371	2103	185	0,301743994735511	0,313919052319842	1231	2721,00169183967	0
51	378	1763	51	0,318673647469459	0,283769633507853	723	4184,7484953409	0
57	386	1568	59	0,381825760733639	0,419758232596915	892	4279,20794032879	0
103	410	2007	33	0,248215444516548	0,441920830629461	1206	5509,38128978019	0
77	461	2473	195	0,427502079290269	0,284724147490099	1140	1547,69332527551	0
133	508	3913	201	0,283261138613861	0,304146039603396	1574	3497,97938508119	0
65	563	2293	98	0,291666666666667	0,259328358208955	882	6292,20731138509	0
336	611	1383	93	0,380847953216374	0,244883040935673	1130	1437,3055534143	1
175	620	4379	48	0,306822769479209	0,654420670165523	1906	2959,44423827871	0
169	690	2760	44	0,279792746113399	0,531794630240226	1833	4926,34877353678	0
116	692	3398	157	0,301483336544018	0,374301675977654	1821	3653,40527714029	0
386	698	6814	300	0,531741541247703	0,572931990494512	3210	762,640024028404	0
27	715	1969	96	0,193484698914116	0,470220467258967	841	3861,25409028014	0
435	749	7833	341	0,438219641993227	0,490759554910498	3169	714,787236203679	1
555	750	9145	443	0,489410317251077	0,384304420423587	3900	2110,30315695341	1
397	779	7416	347	0,312048312048312	0,412731412731413	3860	1673,89413344887	1
112	798	2690	79	0,151738923296808	0,388756550738447	1289	2519,15128538222	0
120	896	2603	54	0,232898650194464	0,358727979867307	1282	3867,675399404448	0
475	931	3760	22	0,137857900318134	0,372216330858961	2492	3061,771212003376	0

Predicted values of Naive Bayes Class.¶

It show that 0 mean it is not geo dependent. But value equl show that object is geo dependent

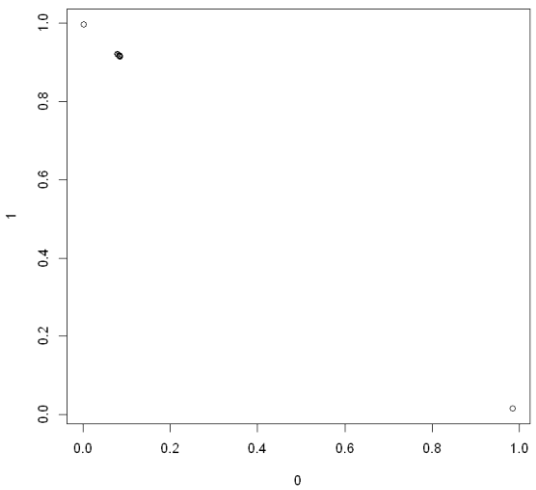
In [200]:

```
predictValue
```

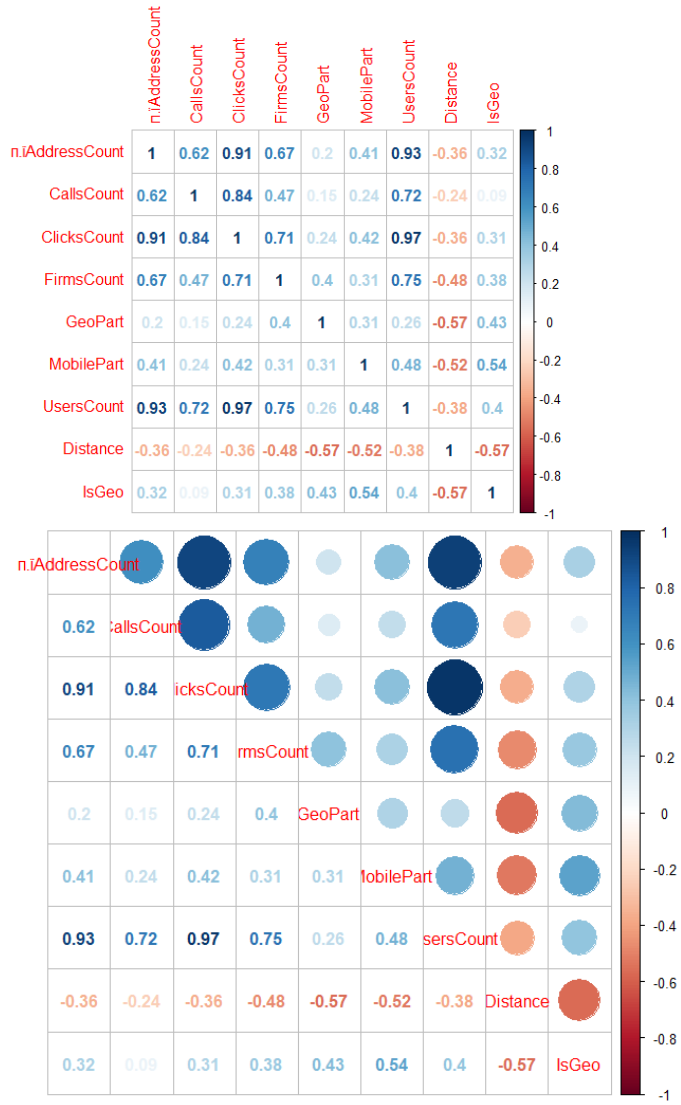
0	1
0.08405603	80.91594396
0.98488485	50.01511515
0.08401989	00.91598011
0.98487058	00.01512942
0.08248111	10.91751889
0.98470873	40.01529127

0	1
0.984871177	0.01512882
0.984848319	0.01515168
0.984748090	0.01525191
0.984065079	0.01593492
0.002570793	0.99742921
0.984775124	0.01522488
0.984790620	0.01520938
0.084455401	0.91554460
0.984777021	0.01522298
0.984797504	0.01520250
0.984553553	0.01544645
0.082571199	0.91742880
0.984754503	0.01524550
0.984733641	0.01526636
0.984530143	0.01546986
0.984650878	0.01534912
0.984696352	0.01530365
0.984721947	0.01527805
0.984663827	0.01533617
0.984600162	0.01539984
0.984390356	0.01560964
0.984624995	0.01537500
0.084405208	0.91559479
0.984318145	0.01568185
0.984559932	0.01544007
0.984467685	0.01553231
0.983905160	0.01609484
0.984668929	0.01533107
0.079405130	0.92059487
0.078222553	0.92177745
0.079770060	0.92022994
0.984569816	0.01543018
0.984582036	0.01541796
0.984413599	0.01558640

```
plot(predictValue, )
```




```
library(corrplot)
```



Decision Tree Classification

```
library(rpart)

tree <- rpart(IsGeo~Distance+ClicksCount, data=data)

tree
```

```
n= 79

node), split, n, deviance, yval
      * denotes terminal node

1) root 79 18.07595 0.3544304
  2) Distance=1033,11276489631,1421,72399039962,1547,69332527551,1703,52647696896,1908
,10811023819,1971,79546984902,2071,94684987015,2080,06825959411,2186,3699144222,2278,1
6291706146,2289,03242434015,2337,60377736899,2519,15128538222,2635,88425199653,2642,20
016520197,2704,49040063268,2705,56818332971,2721,00169183967,2775,01815203855,2780,961
78910644,2949,85779621301,2959,44423827871,2995,88860289814,3038,6587528867,3061,18416
680401,3061,77121200376,3201,60026197197,3396,56608856838,3488,72029249463,3497,979385
08119,3653,40527714029,3746,18547386496,3772,13900324323,3861,25409028014,3867,6753994
0448,4143,34291375289,4149,00290235257,4184,7484953409,4279,20794032879,4333,989690734
98,4511,6323033575,4522,58589659175,4926,34877353678,5206,83845433085,5509,38128978019
,5528,0655460862,5967,70900103466,6272,75183998062,6292,20731138509,762,640024028404,9
89,412138569643 51  0.00000 0.0000000 *
```

```
  3) Distance=1004,78676794652,1234,54344844949,1423,37651183958,1437,3055534143,1463,
57525286792,1576,51415402623,1600,66532620868,1661,51688251154,1673,89413344887,1710,2
1538498727,1777,74626458405,2110,30315695341,2173,7621606007,2276,31335761423,2586,503
27354116,2766,84717399882,3356,11048392976,3969,15036175422,714,787236203679,749,96608
4023684,813,439315154229,854,172080438245,878,013624111711,893,981377311841,918,713971
923729,922,085394259457,938,766030460995,952,012290304611 28  0.00000 1.0000000 *
```

```
predictValueOfTree<- predict(tree, data[1:20,-20])

predictValueOfTree
```

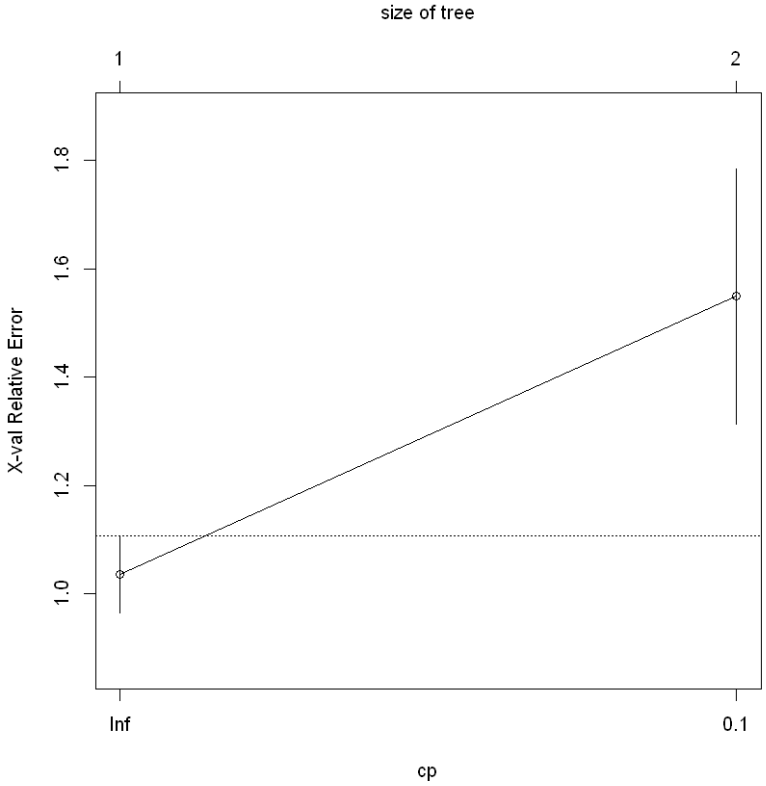
1
1
2
0
3
1
4
0
5
1
6
0
7
0
8
0
9
0
10
0
11
1
12
0
13
0
14
1
15
0
16
0
17
0
18
1
19
0
20
0

```
printcp(tree)
```

```
Regression tree:
rpart(formula = IsGeo ~ Distance + ClicksCount, data = data)

Variables actually used in tree construction:
[1] Distance

Root node error: 18.076/79 = 0.22881
```

```
n= 79

    CP nsplit rel error xerror     xstd
1 1.00      0         1 1.0228 0.070159
2 0.01      1         0 1.6597 0.238640
```

```
plotcp(tree) #plot cross-validation results
```

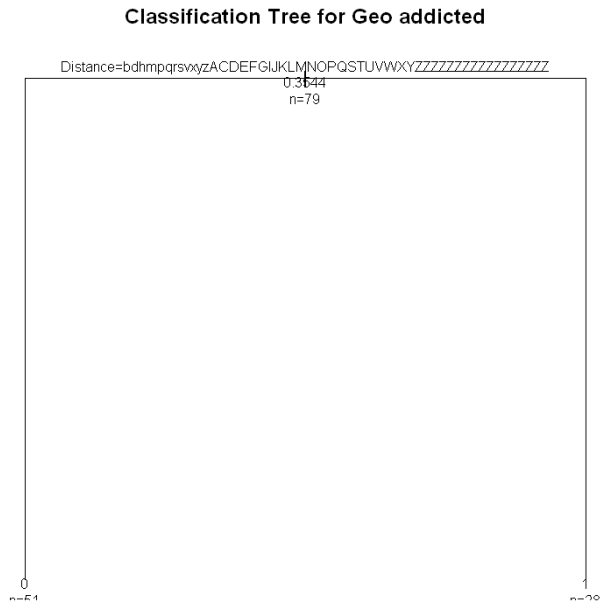



```
summary(tree) #detailed results including surrogate splitsplot decision tree
```

```
Call:
rpart(formula = IsGeo ~ Distance + ClicksCount, data = data)
  n= 79

    CP nsplit rel error   xerror       xstd
1 1.00      0         1 1.022798 0.07015851
2 0.01      1         0 1.659664 0.23864042

Variable importance
   Distance ClicksCount
         82          18

Node number 1: 79 observations,    complexity param=1
  mean=0.3544304, MSE=0.2288095
  left son=2 (51 obs) right son=3 (28 obs)
  Primary splits:
      Distance    splits as  RLRLRRRLRRRRLRRLLLLRRLRLLLLRLLLLLRLLLLLLLLLRLLLLLLLLLLRLLLL
LLLLLLLLLLLRRLRRRRRRRRL, improve=1.0000000, (0 missing)
      ClicksCount < 7399 to the left,  improve=0.2041851, (0 missing)
  Surrogate splits:
      ClicksCount < 7399 to the left,  agree=0.722, adj=0.214, (0 split)

Node number 2: 51 observations
  mean=0, MSE=0

Node number 3: 28 observations
  mean=1, MSE=0
```

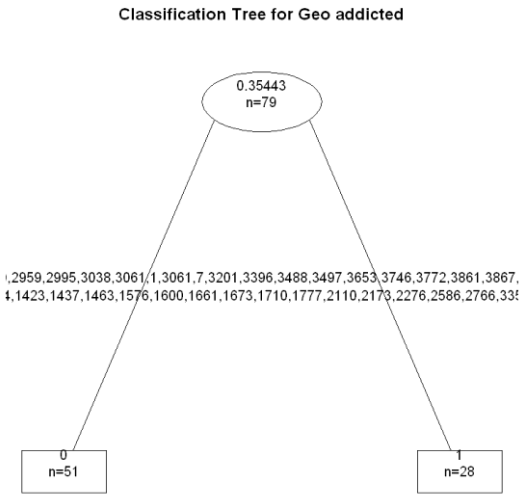
plot decision tree

```
plot(tree, uniform=TRUE,
   main="Classification Tree for Geo addicted")
text(tree, use.n=TRUE, all=TRUE, cex=.8)
```



create postscript plot of decision tree¶

```
post(tree, file = "",
   title = "Classification Tree for Geo addicted")
```



Logistic Regression Classification

```
Glm <- glm(IsGeo~Distance+ClicksCount, data=data, family=binomial())

Glm
```

```
Call:  glm(formula = IsGeo ~ Distance + ClicksCount, family = binomial(),
    data = data)

Coefficients:
             (Intercept)  Distance1033,11276489631  Distance1234,54344844949
               2.657e+01                -5.313e+01                 3.122e-09
Distance1421,72399039962  Distance1423,37651183958   Distance1437,3055534143
              -5.313e+01                 1.160e-09                 1.048e-08
Distance1463,57525286792  Distance1547,69332527551  Distance1576,51415402623
               5.752e-09                -5.313e+01                 2.861e-09
Distance1600,66532620868  Distance1661,51688251154  Distance1673,89413344887
               2.136e-07                 5.090e-09                -4.413e-06
Distance1703,52647696896  Distance1710,21538498727  Distance1777,74626458405
              -5.313e+01                 2.280e-09                 9.966e-09
Distance1908,10811023819  Distance1971,79546984902  Distance2071,94684987015
              -5.313e+01                -5.313e+01                -5.313e+01
Distance2080,06825959411  Distance2110,30315695341   Distance2173,7621606007
              -5.313e+01                 2.076e-07                -4.404e-06
 Distance2186,3699144222  Distance2276,31335761423  Distance2278,16291706146
              -5.313e+01                 6.827e-09                -5.313e+01
Distance2289,03242434015  Distance2337,60377736899  Distance2519,15128538222
              -5.313e+01                -5.313e+01                -5.313e+01
Distance2586,50327354116  Distance2635,88425199653  Distance2642,20016520197
              -3.347e-09                -5.313e+01                -5.313e+01
Distance2704,49040063268  Distance2705,56818332971  Distance2721,00169183967
              -5.313e+01                -5.313e+01                -5.313e+01
Distance2766,84717399882  Distance2775,01815203855  Distance2780,96178910644
               8.137e-09                -5.313e+01                -5.313e+01
Distance2949,85779621301  Distance2959,44423827871  Distance2995,88860289814
              -5.313e+01                -5.313e+01                -5.313e+01
 Distance3038,6587528867  Distance3061,18416680401  Distance3061,77121200376
              -5.313e+01                -5.313e+01                -5.313e+01
Distance3201,60026197197  Distance3356,11048392976  Distance3396,56608856838
              -5.313e+01                -4.411e-06                -5.313e+01
Distance3488,72029249463  Distance3497,97938508119  Distance3653,40527714029
              -5.313e+01                -5.313e+01                -5.313e+01
Distance3746,18547386496  Distance3772,13900324323  Distance3861,25409028014
              -5.313e+01                -5.313e+01                -5.313e+01
Distance3867,67539940448  Distance3969,15036175422  Distance4143,34291375289
              -5.313e+01                -4.414e-06                -5.313e+01
Distance4149,00290235257   Distance4184,7484953409  Distance4279,20794032879
              -5.313e+01                -5.313e+01                -5.313e+01
Distance4333,98969073498   Distance4511,6323033575  Distance4522,58589659175
              -5.313e+01                -5.313e+01                -5.313e+01
Distance4926,34877353678  Distance5206,83845433085  Distance5509,38128978019
              -5.313e+01                -5.313e+01                -5.313e+01
 Distance5528,0655460862  Distance5967,70900103466  Distance6272,75183998062
              -5.313e+01                -5.313e+01                -5.313e+01
Distance6292,20731138509  Distance714,787236203679  Distance749,966084023684
              -5.313e+01                 6.328e-09                -4.408e-06
Distance762,640024028404  Distance813,439315154229  Distance854,172080438245
              -5.313e+01                 2.086e-07                 1.139e-09
Distance878,013624111711  Distance893,981377311841  Distance918,713971923729
               2.082e-07                -4.408e-06                -4.409e-06
Distance922,085394259457  Distance938,766030460995  Distance952,012290304611
               2.068e-07                 2.103e-07                -4.407e-06
Distance989,412138569643               ClicksCount
              -5.313e+01                        NA

Degrees of Freedom: 78 Total (i.e. Null);  0 Residual
Null Deviance:       102.7
Residual Deviance: 4.583e-10      AIC: 158
```

```
predictValuesOfLR <- 1/(1 + exp(-predict.glm(Glm, data[1:10, -10])))
```

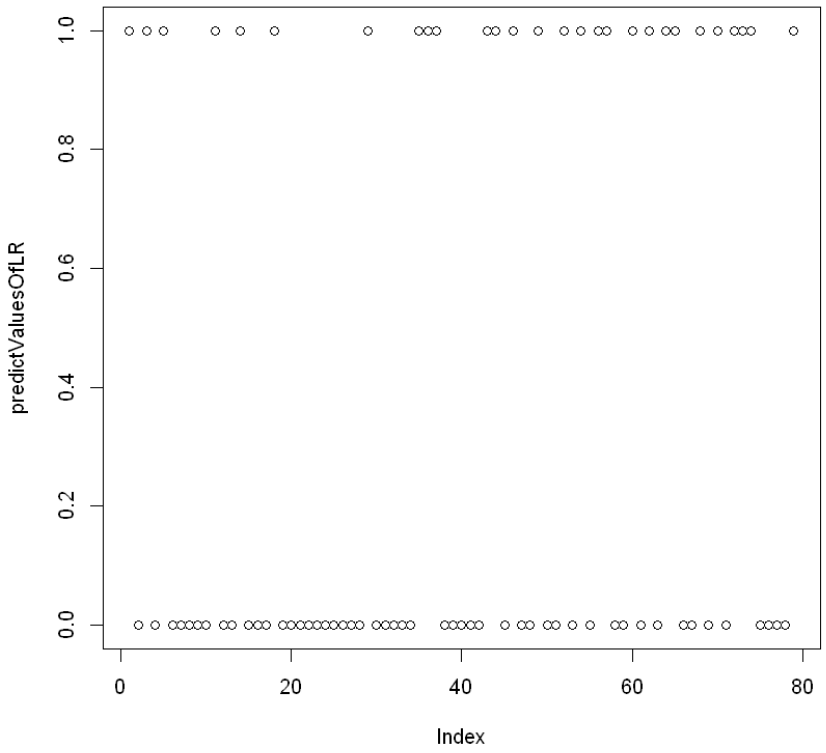
```
predictValuesOfLR
```

1
0.999999999997099
2
2.90070146546842e-12
3
0.999999999997099
4
2.90070146546863e-12
5
0.999999999997099
6
2.90070146546846e-12
7
2.9007014654683e-12
8
2.90070146546836e-12
9
2.90070146546828e-12
10
2.9007014654689e-12

```
summary(Glm) # display results
```

```
Call:
glm(formula = IsGeo ~ Distance + ClicksCount, family = binomial(),
    data = data)

Deviance Residuals:
 [1]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
[26]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
[51]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
[76]  0  0  0  0

Coefficients: (1 not defined because of singularities)
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               2.657e+01  3.561e+05       0        1
Distance1033,11276489631 -5.313e+01  5.036e+05       0        1
Distance1234,54344844949  3.122e-09  5.036e+05       0        1
Distance1421,72399039962 -5.313e+01  5.036e+05       0        1
Distance1423,37651183958  1.160e-09  5.036e+05       0        1
Distance1437,3055534143   1.048e-08  5.036e+05       0        1
Distance1463,57525286792  5.752e-09  5.036e+05       0        1
Distance1547,69332527551 -5.313e+01  5.036e+05       0        1
Distance1576,51415402623  2.861e-09  5.036e+05       0        1
Distance1600,66532620868  2.136e-07  5.036e+05       0        1
Distance1661,51688251154  5.090e-09  5.036e+05       0        1
Distance1673,89413344887 -4.413e-06  5.036e+05       0        1
Distance1703,52647696896 -5.313e+01  5.036e+05       0        1
Distance1710,21538498727  2.280e-09  5.036e+05       0        1
Distance1777,74626458405  9.966e-09  5.036e+05       0        1
Distance1908,10811023819 -5.313e+01  5.036e+05       0        1
Distance1971,79546984902 -5.313e+01  5.036e+05       0        1
Distance2071,94684987015 -5.313e+01  5.036e+05       0        1
Distance2080,06825959411 -5.313e+01  5.036e+05       0        1
Distance2110,30315695341  2.076e-07  5.036e+05       0        1
Distance2173,7621606007  -4.404e-06  5.036e+05       0        1
Distance2186,3699144222  -5.313e+01  5.036e+05       0        1
Distance2276,31335761423  6.827e-09  5.036e+05       0        1
Distance2278,16291706146 -5.313e+01  5.036e+05       0        1
Distance2289,03242434015 -5.313e+01  5.036e+05       0        1
Distance2337,60377736899 -5.313e+01  5.036e+05       0        1
Distance2519,15128538222 -5.313e+01  5.036e+05       0        1
Distance2586,50327354116 -3.347e-09  5.036e+05       0        1
```

```
plot(predictValuesOfLR)
```



You can see that object with predictValueofLR equal 1 mean that they are geo addicted other wise they are not¶