Project 2.1

Project - Churn Prediction

ABSTRACT

Customer value analysis is critical for a good marketing and a customer relationship management strategy. An important component of this strategy is the customer retention rate. Customer retention rate has a strong impact on the customer lifetime value, and understanding the true value of a possible customer churn will help the company in its customer relationship management. Conventional statistical methods are very successful in predicting a customer churn. The goal of this study is to apply logistic regression techniques to predict a customer churn and analyze the churning and no-churning customers by using data provided for the project

Introduction

The subject of customer retention, loyalty, and churn is receiving attention in many industries. This is important in the customer lifetime value context. A company will have a sense of how much is really being lost because of the customer churn and the scale of the efforts that would be appropriate for retention campaign. The mass marketing approach cannot succeed in the diversity of consumer business today. Customer value analysis along with customer churn predictions will help marketing programs target more specific groups of customers. Personal retail banking sector is characterized by customers who stays with a company very long time. Customers usually give their financial business to one company and they won't switch the provider of their financial help very often. In the company's perspective this produces a stable environment for the customer relationship management. Although the continuous relationships with the customers the potential loss of revenue because of customer churn in this case can be huge.

This paper will present a customer churn analysis for the data provided in the project 2. The goal of this paper is twofold. First the churning customers are analyzed in R – Logistic Regression model after applying appropriate preprocessing techniques . The second stage we need to connect with Tableau and R for visualization, probability and prediction in Tableau using the trained and testing set.

Logistic regression Binomial (binary) logistic regression is a form of regression which is used in a situation when dependent is not a continuous variable but a state which may or may not happen, or a category in a specific classification . Logistic regression can be used to predict a discrete outcome on the basis of continuous and/or categorical variables. Multinomial logistic regression exists to handle the case of dependents with more classes than two. In the logistic regression there can be only one dependent variable. Logistic regression applies maximum likelihood estimation after transforming the dependent into a logistic variable [8]. Unlike the normal regression model the dependent variable in logistic regression is usually dichotomous: the dependent variable can take value 1 with probability q and value 0 with probability 1-q.

Preprocessing

The Churn data provided for the project 2 is checked for the preprocessing requirements if any. Normally, the missing values requires creates many problems during analysis and this requires preprocessing and imputation etc., In this project the Amelia library is used to identify the missing values and as per the graph given below there is no missing value in the data.

```
library(Amelia)
```

```
## Loading required package: Rcpp
```

```
## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.5, built: 2018-05-07)
## ## Copyright (C) 2005-2018 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```
any(is.na(Churn))
```

```
## [1] FALSE
```

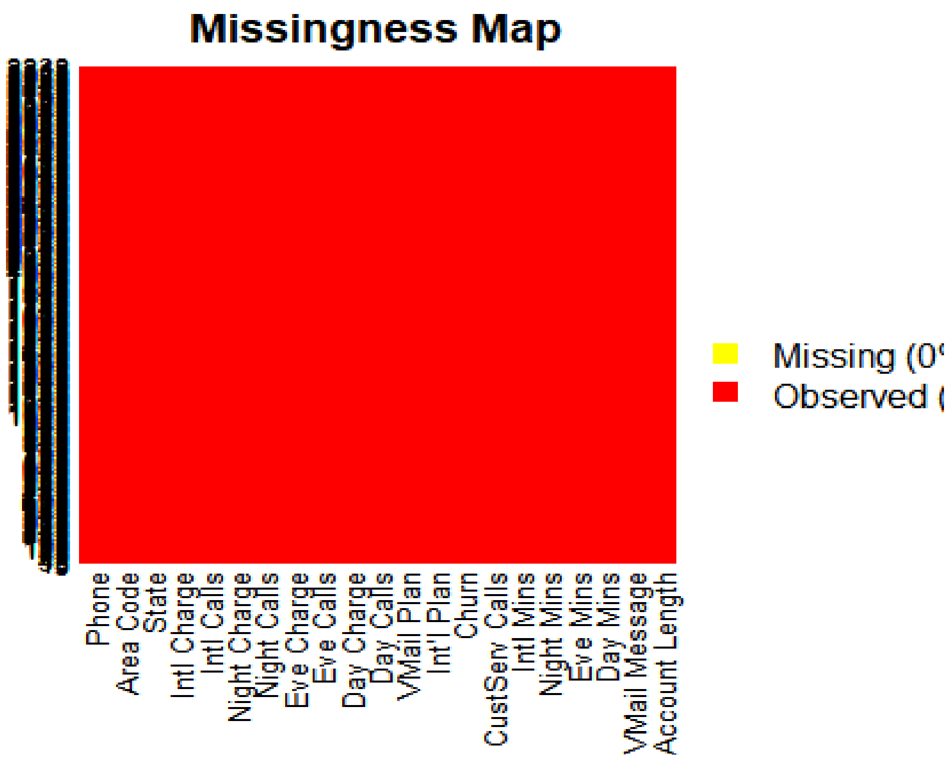
```
# visualize the missing values using the missing map from the Amelia package
missmap(Churn,col=c("yellow","red"))
```

```
## Warning in if (class(obj) == "amelia") {: the condition has length > 1 and
## only the first element will be used
```

```
## Warning: Unknown or uninitialised column: 'arguments'.
```

```
## Warning: Unknown or uninitialised column: 'arguments'.
```

```
## Warning: Unknown or uninitialised column: 'imputations'.
```



Excluding phone and state from the data set as they are not very important

```
mydata2<-Churn[,-21]
mydata<-mydata2[,-19]
sapply(mydata, function(x) sum(is.na(x)))

## Account Length  VMail Message       Day Mins       Eve Mins     Night Mins
##              0              0              0              0              0
##      Intl Mins CustServ Calls          Churn     Int'l Plan     VMail Plan
##              0              0              0              0              0
##      Day Calls     Day Charge      Eve Calls     Eve Charge    Night Calls
##              0              0              0              0              0
```

```
##   Night Charge     Intl Calls    Intl Charge      Area Code
##              0              0              0              0

mydata <- mydata[complete.cases(mydata), ]
intrain<- createDataPartition(mydata$Churn,p=0.8,list=FALSE)
set.seed(2017)
training<- mydata[intrain,]
testing<- mydata[-intrain,]
dim(training); dim(testing)

## [1] 2667  19

## [1] 666  19

library (data.table)

library (plyr)
library (stringr)

##
## Attaching package: 'stringr'

## The following object is masked from 'package:strucchange':
##
##     boundary

LogModel <- glm(Churn ~ .,family=binomial(link="logit"),data=training)
print(summary(LogModel))

##
## Call:
## glm(formula = Churn ~ ., family = binomial(link = "logit"), data = training)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1332  -0.5222  -0.3425  -0.1992   3.2941
##
## Coefficients:
```

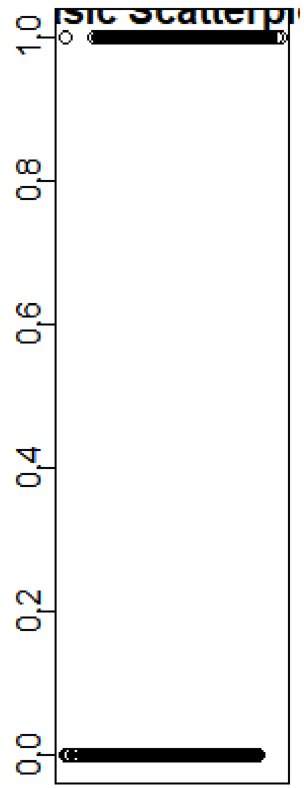
```
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -7.420e+00  1.024e+00  -7.249  4.2e-13 ***
## `Account Length`    1.141e-03  1.533e-03   0.744 0.456828
## `VMail Message`     4.207e-02  2.027e-02   2.075 0.037983 *
## `Day Mins`         -1.341e+00  3.624e+00  -0.370 0.711274
## `Eve Mins`         -6.519e-01  1.813e+00  -0.360 0.719209
## `Night Mins`       -4.968e-01  9.739e-01  -0.510 0.610007
## `Intl Mins`        -7.267e-01  5.875e+00  -0.124 0.901560
## `CustServ Calls`  5.299e-01  4.387e-02   12.079   < 2e-16 ***
## `Int'l Plan`       2.096e+00  1.610e-01   13.022   < 2e-16 ***
## `VMail Plan`     -2.319e+00  6.533e-01  -3.550 0.000385 ***
## `Day Calls` 2.586e-03 3.042e-03 0.850 0.395283
## `Day Charge`      7.962e+00  2.132e+01   0.373 0.708808
## `Eve Calls`       8.044e-04  3.055e-03   0.263 0.792319
## `Eve Charge`      7.744e+00  2.133e+01   0.363 0.716583
## `Night Calls`    -2.353e-03  3.160e-03  -0.745 0.456435
## `Night Charge`      1.109e+01  2.164e+01   0.513 0.608260
## `Intl Calls`       -1.044e-01  2.812e-02  -3.712 0.000206 ***
## `Intl Charge`       2.960e+00  2.176e+01   0.136 0.891786
## `Area Code`        -6.509e-05  1.462e-03  -0.045 0.964500
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2230.1  on 2666  degrees of freedom
## Residual deviance: 1755.5 on 2648 degrees of freedom
## AIC: 1793.5
##
## Number of Fisher Scoring iterations: 6
```

```
anova(LogModel, test="Chisq")
```

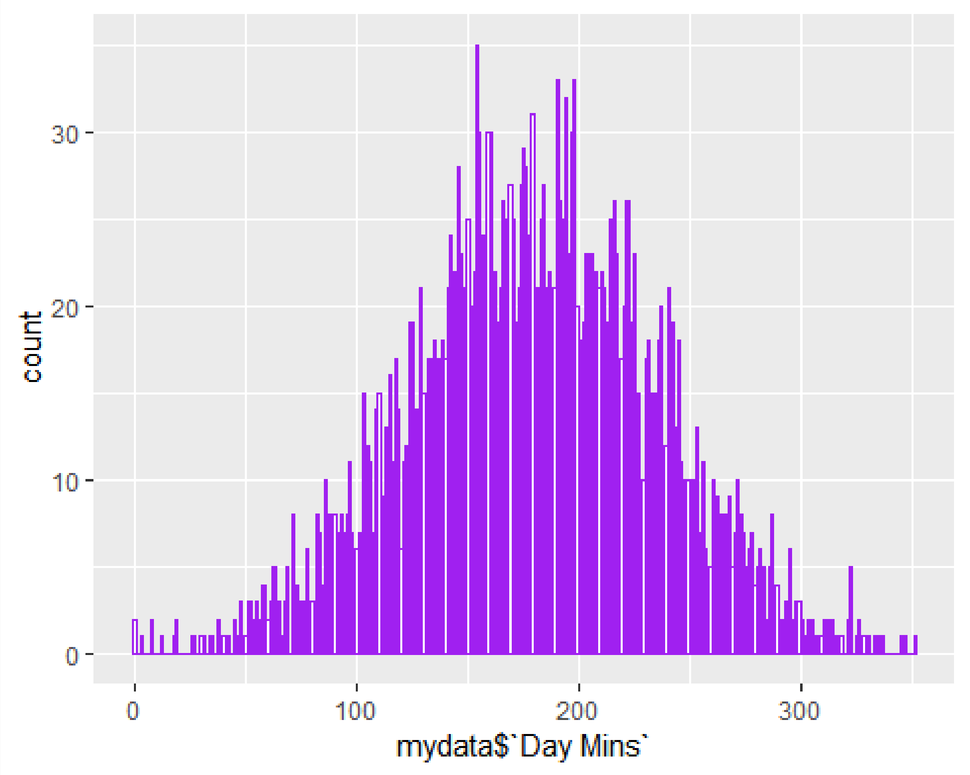
```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
```

```
##
## Response: Churn
##
## Terms added sequentially (first to last)
##
##
##                  Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                             2666     2230.1
## `Account Length`  1    1.550      2665     2228.6 0.2132053
## `VMail Message`   1   26.557      2664     2202.0 2.559e-07 ***
## `Day Mins`        1   92.031      2663     2110.0 < 2.2e-16 ***
## `Eve Mins`        1   17.926      2662     2092.1 2.297e-05 ***
## `Night Mins`      1    1.924      2661     2090.1 0.1654485
## `Intl Mins`       1    9.890      2660     2080.2 0.0016614 **
## `CustServ Calls`  1  130.903      2659     1949.3 < 2.2e-16 ***
## `Int'l Plan`      1  163.266      2658     1786.1 < 2.2e-16 ***
## `VMail Plan`      1   14.094      2657     1772.0 0.0001739 ***
## `Day Calls`       1    0.719      2656     1771.3 0.3966047
## `Day Charge`      1    0.051      2655     1771.2 0.8214162
## `Eve Calls`       1    0.017      2654     1771.2 0.8948703
## `Eve Charge`      1    0.132      2653     1771.1 0.7158769
## `Night Calls`     1    0.536      2652     1770.5 0.4641722
## `Night Charge`    1    0.190      2651     1770.3 0.6629360
## `Intl Calls`      1   14.804      2650     1755.5 0.0001193 ***
## `Intl Charge`     1    0.019      2649     1755.5 0.8911511
## `Area Code`       1    0.002      2648     1755.5 0.9644909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

testing$Churn <- as.character(testing$Churn)
testing$Churn[testing$Churn=="No"] <- "0"
testing$Churn[testing$Churn=="Yes"] <- "1"
fitted.results <- predict(LogModel,newdata=testing,type='response')
fitted.results

sapply(mydata, sd)
```

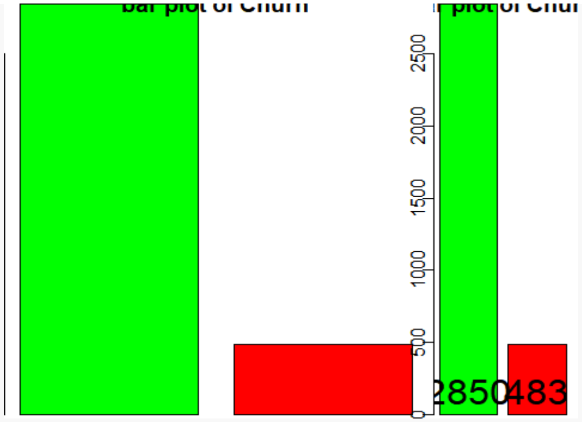
```
## Account Length  VMail Message       Day Mins       Eve Mins     Night Mins
##      39.8221059     13.6883654     54.4673892     50.7138444     50.5738470
##       Intl Mins CustServ Calls          Churn     Int'l Plan     VMail Plan
##       2.7918395      1.3154910      0.3520674      0.2958791      0.4473979
##       Day Calls     Day Charge      Eve Calls     Eve Charge    Night Calls
##      20.0690842      9.2594346     19.9226253      4.3106676     19.5686093
##    Night Charge     Intl Calls    Intl Charge      Area Code
##       2.2758728      2.4612143      0.7537726     42.3712905

plot.new()
plot(mydata$Churn ~mydata$`Day Mins`)
title('Basic Scatterplot')
```



```
ggplot(mydata, aes(x=mydata$`Day Mins`)) + geom_histogram(binwidth = 1, fill = "white", color = "purple")
```




```
#Randomly split data into train and test set
#80% will be ssigned to train set, 20% will be assigned to tst set
barplot(table(mydata$Churn), col= c("green", "red"), main='bar plot of Churn')
text(barplot(table(mydata$Churn), col =c('green' , 'red'), main='bar plot of Churn'), 0,table(mydata$Churn)
, cex =2 , pos =3)
```




```
#proportion
round(prop.table(table(mydata$Churn))*100,digits = 2)

##
##     0     1 (As per this churn is around 15%)
## 85.51 14.49

mydata_train<-mydata_n[1:2666,]
mydata_test<-mydata_n[2667:3333,]
mydata_train_labels<-mydata_n[1:2666,7]
mydata_test_labels<-mydata_n[2667:3333,7]

sapply(mydata_n, sd)

## Account.Length  VMail.Message       Day.Mins       Eve.Mins     Night.Mins
##      0.1645542      0.2683993      0.1552662      0.1394387      0.1360243
##      Intl.Mins CustServ.Calls          Churn     Int.l.Plan     VMail.Plan
##      0.1395920      0.1461657      0.3520674      0.2958791      0.4473979
##      Day.Calls     Day.Charge      Eve.Calls     Eve.Charge    Night.Calls
##      0.1216308      0.1552554      0.1171919      0.1394587      0.1378071
##   Night.Charge     Intl.Calls    Intl.Charge
##      0.1360354      0.1230607      0.1395875
```

```r
#Forward elimination
#Lower AIC indicates a better model
forward <- step(glm(Churn ~ 1, data = mydata_train), direct
essage+Day.Mins + Eve.Mins +
                  Night.Mins + Intl.Mins + CustServ.Calls +
                  Day.Calls + Day.Charge + Eve.Calls + Eve.
                  Night.Charge + Intl.Calls + Intl.Charge)

logit<- glm(Churn ~Account.Length+Day.Mins+ Day.Charge +Cus
il.Message+Day.Calls +Eve.Calls+ Intl.Mins + Night.Calls+In
l")
summary(logit)

##
## Call:
## glm(formula = Churn ~ Account.Length + Day.Mins + Day.Ch
##     CustServ.Calls + VMail.Plan + Eve.Mins + Eve.Charge
##     Day.Calls + Eve.Calls + Intl.Mins + Night.Calls + In
##     family = "binomial", data = mydata_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7136  -0.5583  -0.3989  -0.2492   3.0223
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -7.4168     0.6846 -10.834  < 2e-16 ***
## Account.Length    0.1336     0.3696   0.362  0.71771
## Day.Mins        191.3463  1231.6229   0.155  0.87654
## Day.Charge     -187.0655  1231.7103  -0.152  0.87929
## CustServ.Calls    3.9702     0.3798  10.455  < 2e-16 ***
## VMail.Plan       -1.7024     0.6098  -2.792  0.00524 **
## Eve.Mins        462.5882   642.8816   0.720  0.47180
## Eve.Charge     -460.1275   642.7876  -0.716  0.47410
## VMail.Message     1.5122     0.9823   1.539  0.12371
## Day.Calls         0.3629     0.4940   0.735  0.46264
```

```
## Eve.Calls          0.4090     0.5063   0.808  0.41921
## Intl.Mins          2.0353     0.4447   4.576 4.73e-06 ***
## Night.Calls        0.1657     0.4403   0.376  0.70665
## Intl.Calls        -1.6524     0.5357  -3.085  0.00204 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##   Null deviance: 2162.0  on 2665  degrees of freedom
## Residual deviance: 1867.4  on 2652  degrees of freedom
## AIC: 1895.4
##
## Number of Fisher Scoring iterations: 5
```

```r
                                                    # Confidence interval using log-likelihood
confint(logit)
```

```
## Waiting for profiling to be done...
```

```
##                        2.5 %       97.5 %
## (Intercept)       -8.7761881   -6.0914407
## Account.Length    -0.5919165    0.8578438
## Day.Mins       -2224.7782913 2605.8924991
## Day.Charge     -2601.7748293 2229.2388877
## CustServ.Calls     3.2299992    4.7198281
## VMail.Plan        -2.9307458   -0.5373346
## Eve.Mins        -796.3103582 1725.2392867
## Eve.Charge     -1722.5889512  798.5919581
```

```
## VMail.Message     -0.4038228    3.4527140
## Day.Calls         -0.6043867    1.3330979
## Eve.Calls         -0.5823325    1.4036271
## Intl.Mins          1.1687246    2.9129356
## Night.Calls       -0.6973052    1.0293210
## Intl.Calls        -2.7201822   -0.6195282
```

```
exp(logit$coefficients)
```

```
##    (Intercept) Account.Length       Day.Mins     Day.Charge CustServ.Calls
##   6.010934e-04   1.142977e+00   1.260826e+83   5.734114e-82   5.299764e+01
##     VMail.Plan       Eve.Mins     Eve.Charge  VMail.Message      Day.Calls
##   1.822392e-01  7.933806e+200  1.476321e-200   4.536688e+00   1.437439e+00
##      Eve.Calls      Intl.Mins    Night.Calls     Intl.Calls
##   1.505333e+00   7.654355e+00   1.180211e+00   1.915949e-01
```

```
# logistic regression model:
fit <- glm(Churn~.,data =mydata_train ,family = binomial(link='logit'))
summary(fit)
```

```
##
## Call:
## glm(formula = Churn ~ ., family = binomial(link = "logit"), data = mydata_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9680  -0.5111  -0.3376  -0.1979   3.1864
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -8.42021    0.76774 -10.967  < 2e-16 ***
## Account.Length     0.01841    0.38853   0.047 0.962208
## VMail.Message      1.53996    1.01473   1.518 0.129114
## Day.Mins         213.42834 1293.28315   0.165 0.868922
## Eve.Mins         592.02181  674.48943   0.878 0.380088
## Night.Mins      -110.40724  368.95516  -0.299 0.764755
## Intl.Mins       -287.85749  120.73173  -2.384 0.017113 *
```

```
## CustServ.Calls    4.51834    0.40343  11.200  < 2e-16 ***
## Int.l.Plan        2.06924    0.16257  12.729  < 2e-16 ***
## VMail.Plan       -1.87056    0.63215  -2.959 0.003086 **
## Day.Calls         0.44168    0.51623   0.856 0.392222
## Day.Charge     -209.05873 1293.37170  -0.162 0.871590
## Eve.Calls         0.43463    0.53648   0.810 0.417852
## Eve.Charge     -589.41466  674.38935  -0.874 0.382120
## Night.Calls       0.01869    0.46116   0.041 0.967668
## Night.Charge    111.62067  368.91878   0.303 0.762224
## Intl.Calls       -1.92925    0.56673  -3.404 0.000664 ***
## Intl.Charge     289.76049  120.71956   2.400 0.016383 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##   Null deviance: 2162.0  on 2665  degrees of freedom
## Residual deviance: 1702.5 on 2648 degrees of freedom
## AIC: 1738.5
##
## Number of Fisher Scoring iterations: 6
```

```r
library(MASS)
step_fit <- stepAIC(fit,method='backward')
```

```r
summary(step_fit)
```

```
##
## Call:
## glm(formula = Churn ~ VMail.Message + Day.Mins + Eve.Mins + Intl.Mins +
##     CustServ.Calls + Int.l.Plan + VMail.Plan + Night.Charge +
##     Intl.Calls + Intl.Charge, family = binomial(link = "logit"),
##     data = mydata_train)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9778  -0.5124  -0.3392  -0.2020   3.1476
```

```
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -7.8769     0.5686 -13.853  < 2e-16 ***
## VMail.Message      1.5052     1.0120   1.487 0.136929
## Day.Mins           4.4084     0.4297  10.259  < 2e-16 ***
## Eve.Mins           2.5099     0.4622   5.430 5.64e-08 ***
## Intl.Mins       -291.5210   120.5655  -2.418 0.015608 *
## CustServ.Calls     4.5206     0.4022  11.240  < 2e-16 ***
## Int.l.Plan         2.0630     0.1622  12.721  < 2e-16 ***
## VMail.Plan        -1.8555     0.6300  -2.945 0.003230 **
## Night.Charge       1.2494     0.4652   2.686 0.007235 **
## Intl.Calls        -1.9404     0.5652  -3.433 0.000596 ***
## Intl.Charge      293.4337   120.5539   2.434 0.014931 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2162.0  on 2665  degrees of freedom
## Residual deviance: 1704.9  on 2655  degrees of freedom
## AIC: 1726.9
##
## Number of Fisher Scoring iterations: 6
```

```
confint(step_fit)
```

```
## Waiting for profiling to be done...
```

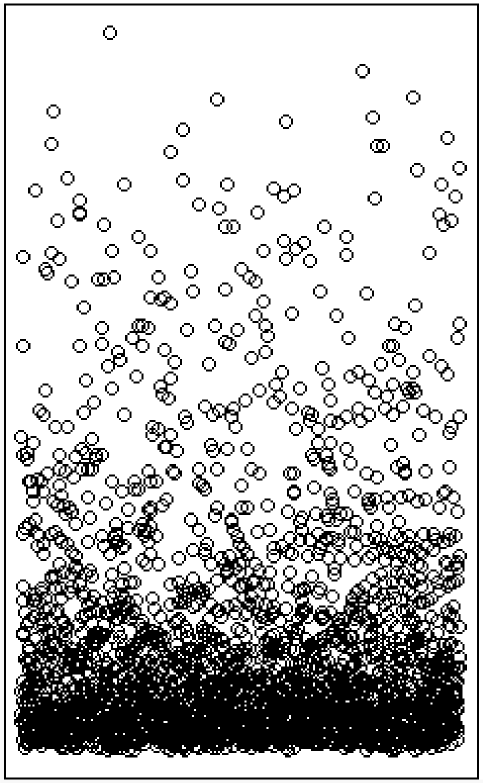
```
##                        2.5 %      97.5 %
## (Intercept)       -9.0107622  -6.7808248
## VMail.Message     -0.4650469   3.5078770
## Day.Mins           3.5752571   5.2606938
## Eve.Mins           1.6088317   3.4218238
## Intl.Mins       -528.7346134 -55.8150001
## CustServ.Calls     3.7372142   5.3151528
## Int.l.Plan         1.7458847   2.3822001
```

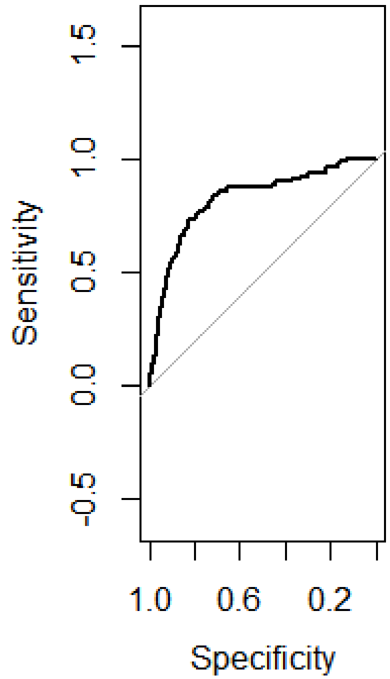
```
## VMail.Plan       -3.1260647  -0.6530832
## Night.Charge      0.3391697   2.1636548
## Intl.Calls       -3.0675068  -0.8511819
## Intl.Charge      57.7541777 530.6286565
```

```r
#ANOVA on base model
anova(fit,test = 'Chisq')
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Churn
##
## Terms added sequentially (first to last)
##
##
##                Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                            2665     2162.0
## Account.Length  1    0.277      2664     2161.8 0.5985480
## VMail.Message   1   19.675      2663     2142.1 9.180e-06 ***
## Day.Mins        1  103.623      2662     2038.5 < 2.2e-16 ***
## Eve.Mins        1   23.581      2661     2014.9 1.197e-06 ***
## Night.Mins      1    3.199      2660     2011.7 0.0736818 .
## Intl.Mins       1   17.379      2659     1994.3 3.062e-05 ***
## CustServ.Calls  1  111.293      2658     1883.0 < 2.2e-16 ***
## Int.l.Plan      1  152.056      2657     1730.9 < 2.2e-16 ***
## VMail.Plan      1    8.313      2656     1722.6 0.0039361 **
## Day.Calls       1    1.004      2655     1721.6 0.3164303
## Day.Charge      1    0.101      2654     1721.5 0.7509001
## Eve.Calls       1    0.705      2653     1720.8 0.4010942
## Eve.Charge      1    0.752      2652     1720.1 0.3857120
## Night.Calls     1    0.000      2651     1720.1 0.9948372
## Night.Charge    1    0.088      2650     1720.0 0.7668220
## Intl.Calls      1   11.638      2649     1708.3 0.0006464 ***
## Intl.Charge     1    5.795      2648     1702.5 0.0160706 *
```

```
## ---
pred <- predict(fit,newdata = mydata_test,type ='response')
#check the AUC curve
library(pROC)

g <- roc( Churn~ pred, data = mydata_test)
g

##
## Call:
## roc.formula(formula = Churn ~ pred, data = mydata_test)
##
## Data: pred in 558 controls (Churn 0) < 109 cases (Churn 1).
## Area under the curve: 0.8266

plot(g)
```






```
library(caret)
#with default prob cut 0.50
mydata_test$pred_Churn <- ifelse(pred<0.8,'yes','no')

table(mydata_test$pred_Churn,mydata_test$Churn)

##
##          0   1
##    no    1   3
##    yes 557 106

#training split of churn classes
round(table(mydata_train$Churn)/nrow(mydata_train),2)*100
```

```
##
##  0  1
## 86 14
```

```r
# test split of churn classes
round(table(mydata_test$Churn)/nrow(mydata_test),2)*100
```

```
##
##  0  1
## 84 16
```

```r
#predicted split of churn classes
round(table(mydata_test$pred_Churn)/nrow(mydata_test),2)*100
```

```
##
##  no yes
##   1  99
```

```r
#create confusion matrix
#confusionMatrix(mydata_test$Churn,mydata_test$pred_Churn)
#how do we create a cross validation scheme
control <- trainControl(method = 'repeatedcv',
                        number = 10,
                        repeats = 3)
seed <-7
metric <- 'Accuracy'
set.seed(seed)
fit_default <- train(Churn~.,
                     data = mydata_train,
                     method = 'glm',
                     metric = NaN,
                     trControl = control)
```

```
## Warning in train.default(x, y, weights = w, ...): You are trying to do
## regression and your outcome only has two possible values Are you trying to
## do classification? If so, use a 2 level factor as your outcome column.
```

```
## Warning in train.default(x, y, weights = w, ...): The metric "NaN" was not
## in the result set. RMSE will be used instead.
```

```r
print(fit_default)
```

```
## Generalized Linear Model
##
## 2666 samples
##   17 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold, repeated 3 times)
## Summary of sample sizes: 2399, 2400, 2399, 2399, 2400, 2399, ...
## Resampling results:
##
##   RMSE       Rsquared   MAE
##   0.3171573  0.1668586  0.2165651
```

```r
library(caret)
varImp(step_fit)
```

```
##                  Overall
## VMail.Message   1.487325
## Day.Mins       10.258636
## Eve.Mins        5.429922
## Intl.Mins       2.417948
## CustServ.Calls 11.239954
## Int.l.Plan     12.720766
## VMail.Plan      2.945000
## Night.Charge    2.685836
## Intl.Calls      3.433245
## Intl.Charge     2.434045
```

```r
varImp(fit_default)
```

```
## glm variable importance
##
```

```
##                    Overall
## Int.l.Plan     100.00000
## CustServ.Calls  79.36943
## Intl.Calls      21.54437
## VMail.Plan      17.99874
## Intl.Charge     15.22120
## Intl.Mins       15.11822
## VMail.Message    6.87083
## Day.Calls        5.43601
## Eve.Mins         5.43158
## Eve.Charge       5.40848
## Eve.Calls        4.87659
## Night.Charge     3.45603
## Night.Mins       3.43761
## Day.Mins         1.44730
## Day.Charge       1.42506
## Account.Length   0.02108
## Night.Calls      0.00000
```

```
library(devtools)
library(woe)

library(riv)

iv_df <- iv.mult(mydata_train, y="Churn", summary=TRUE, verbose=TRUE)

iv_df
```

```
##          Variable InformationValue Bins ZeroBins    Strength
## 1      Day.Charge      0.643151413    6        0 Very strong
## 2        Day.Mins      0.643151413    6        0 Very strong
## 3  CustServ.Calls      0.158681659    2        0     Average
## 4        Eve.Mins      0.149576165    5        0     Average
## 5      Eve.Charge      0.149310982    5        0     Average
## 6     Intl.Charge      0.097357797    4        0        Weak
## 7       Intl.Mins      0.097357797    4        0        Weak
## 8   VMail.Message      0.081622650    2        0        Weak
```

```
## 9      VMail.Plan      0.081622650    2        0        Weak
## 10     Intl.Calls      0.068633851    2        0        Weak
## 11   Night.Charge      0.060709508    6        0        Weak
## 12 Account.Length      0.033794008    4        0        Weak
## 13       Day.Calls      0.028673937    3        0        Weak
## 14      Night.Mins      0.022784889    2        0        Weak
## 15       Eve.Calls      0.010611328    2        0   Wery weak
## 16     Night.Calls      0.009978104    2        0   Wery weak
## 17      Int.l.Plan      0.000000000    1        0   Wery weak

iv <- iv.mult(mydata_train, y="Churn", summary=FALSE, verbose=TRUE)

# Plot information value summary
```

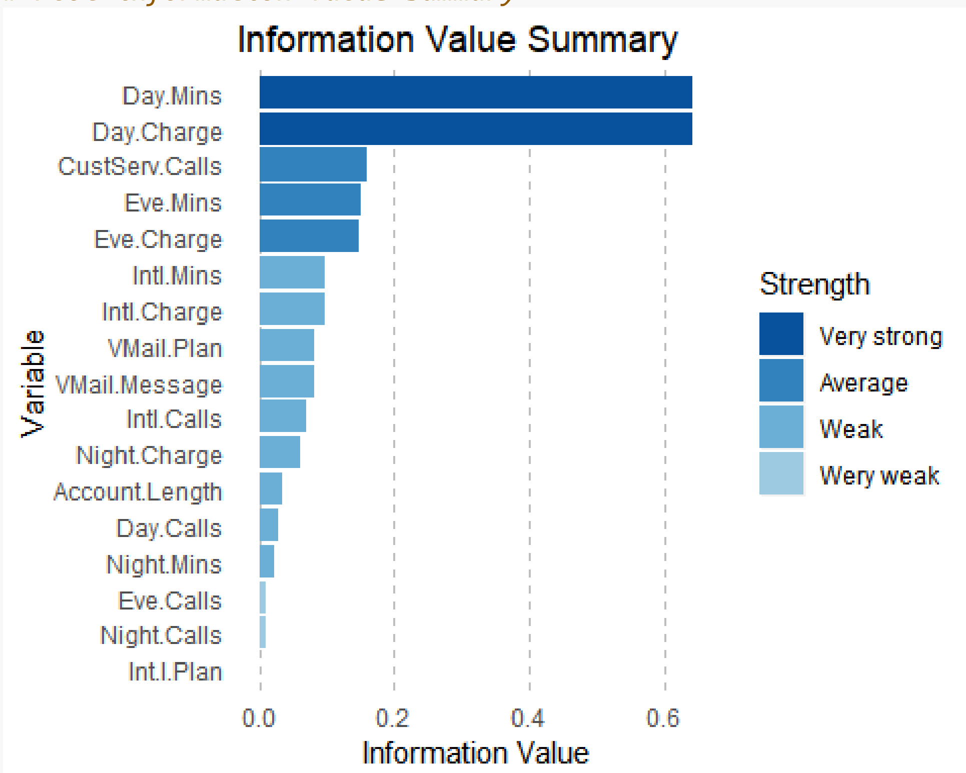


TABLEAU VISUALIZATION

Library(Rserve)

Rserve()

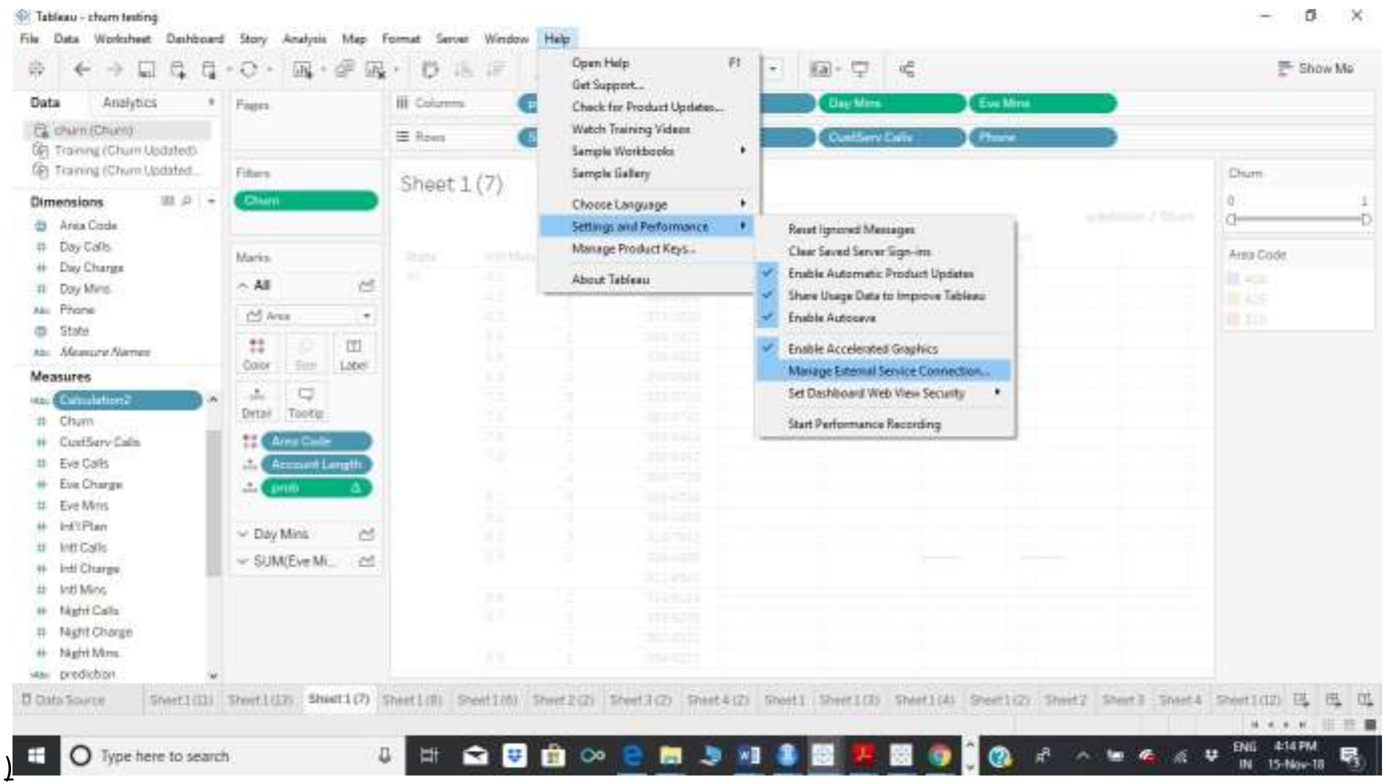
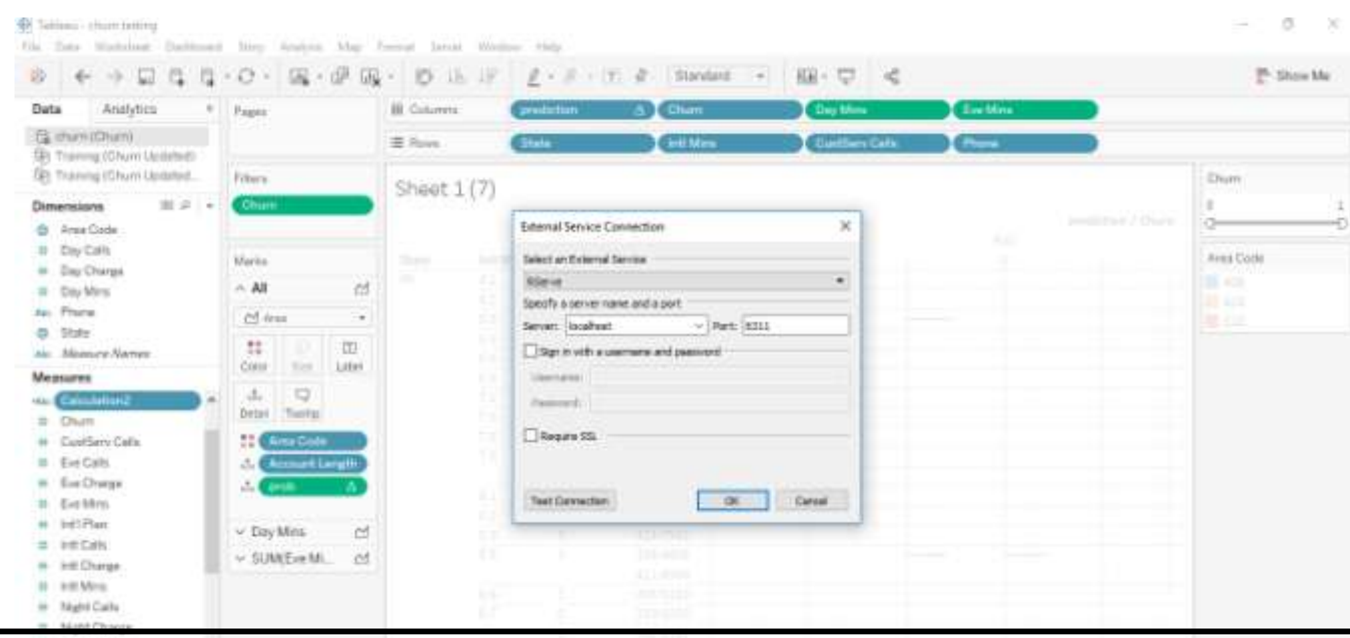


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Open Help
Get Support...
Check for Product Updates...
Watch Training Videos
Sample Workbooks
Sample Gallery
Choose Language
Settings and Performance
Manage Product Keys...
About Tableau
Reset Ignored Messages
Clear Saved Server Sign-ins
Enable Automatic Product Updates
Share Usage Data to Improve Tableau
Enable Autosave
Enable Accelerated Graphics
Manage External Service Connection...
Set Dashboard Web View Security
Start Performance Recording
Sheet 1 (7)
Show Me
Day Mins
Eve Mins
CustServ Calls
Phone
Data
Analytics
Dimensions
Measures
Calculation2
Churn
Area Code
Account Length
prob
Type here to search
)


External Service Connection
Select an External Service
RServe
Specify a server name and a port
Server: localhost
Port: 6311
Sign in with a username and password
Require SSL
Test Connection
OK
Cancel
Sheet 1 (7)

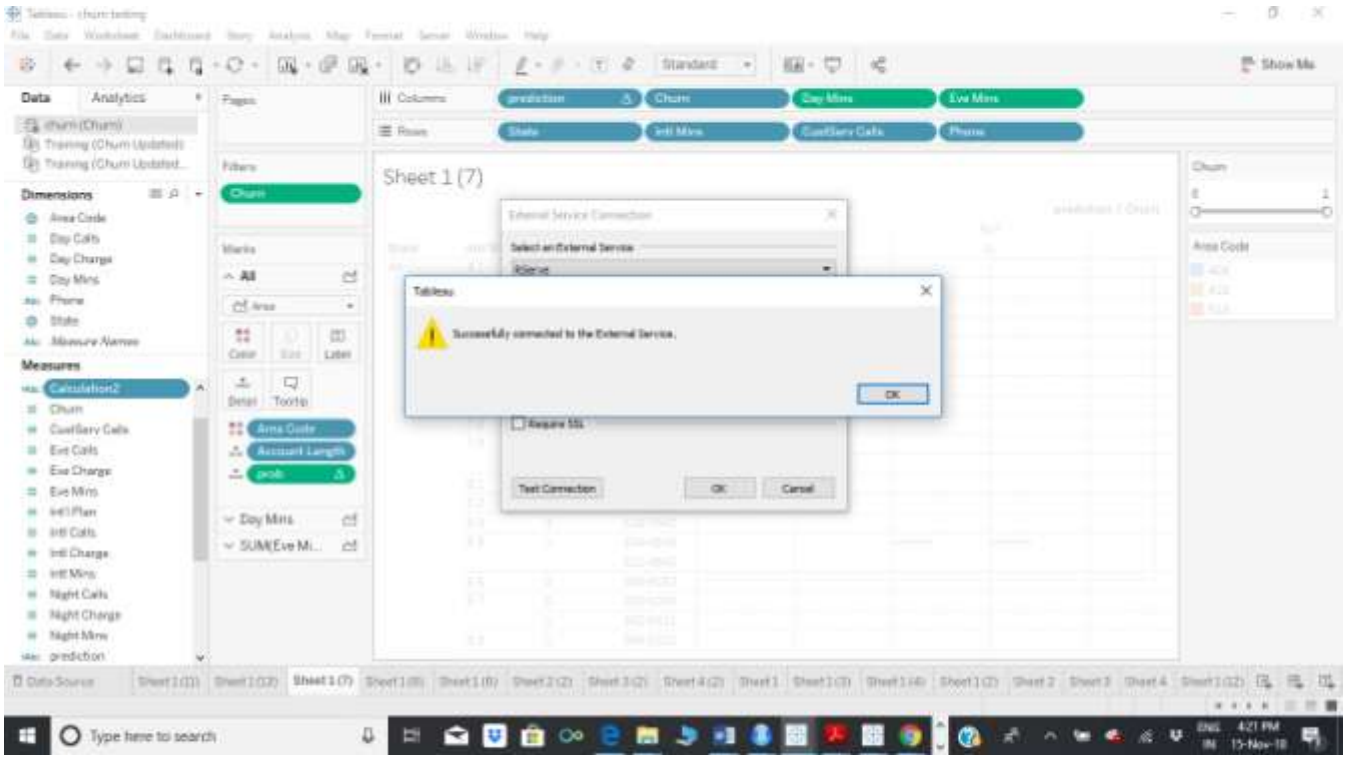
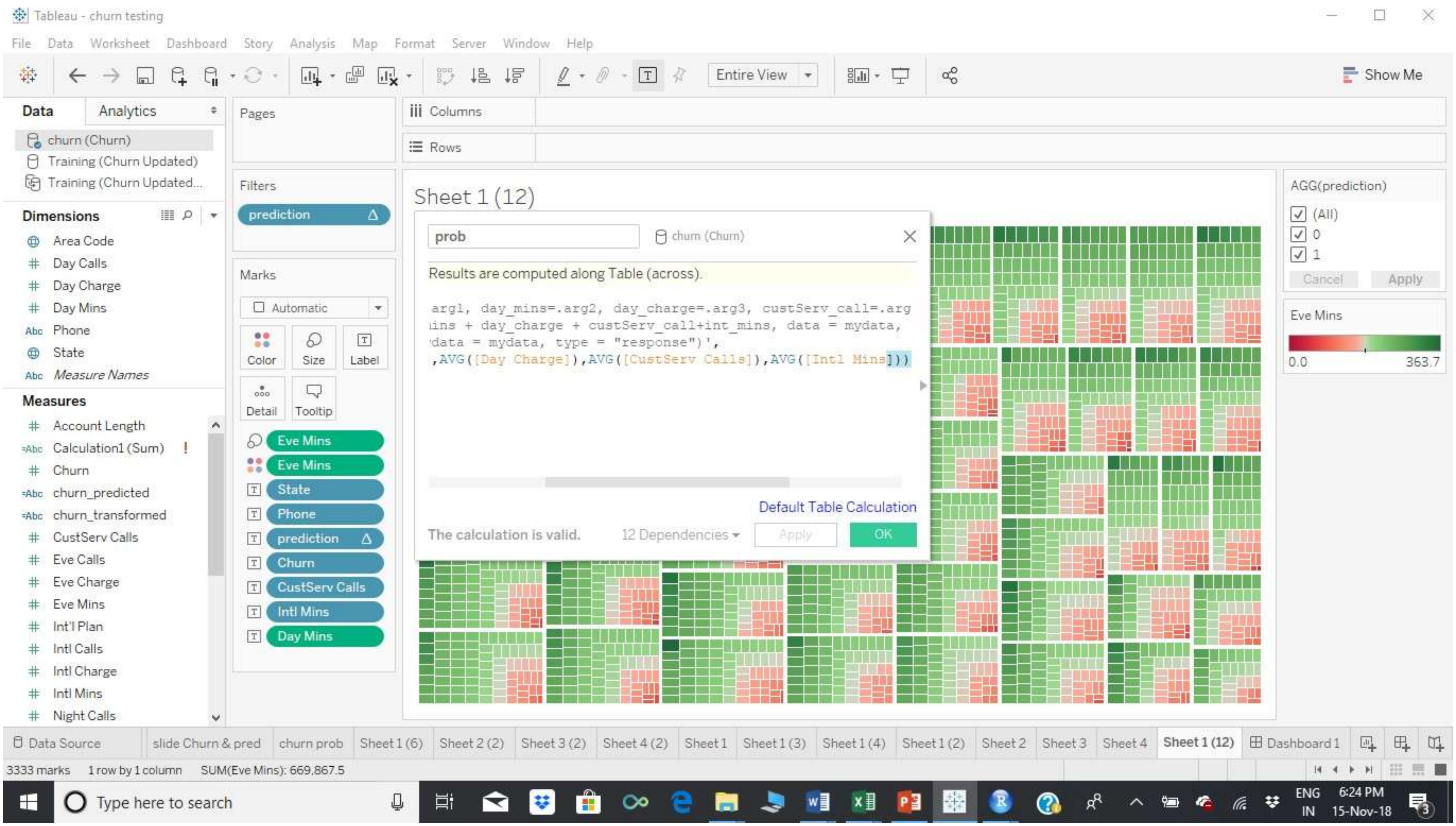


Tableau
Successfully connected to the External Service.
OK



Calculated field (prob)

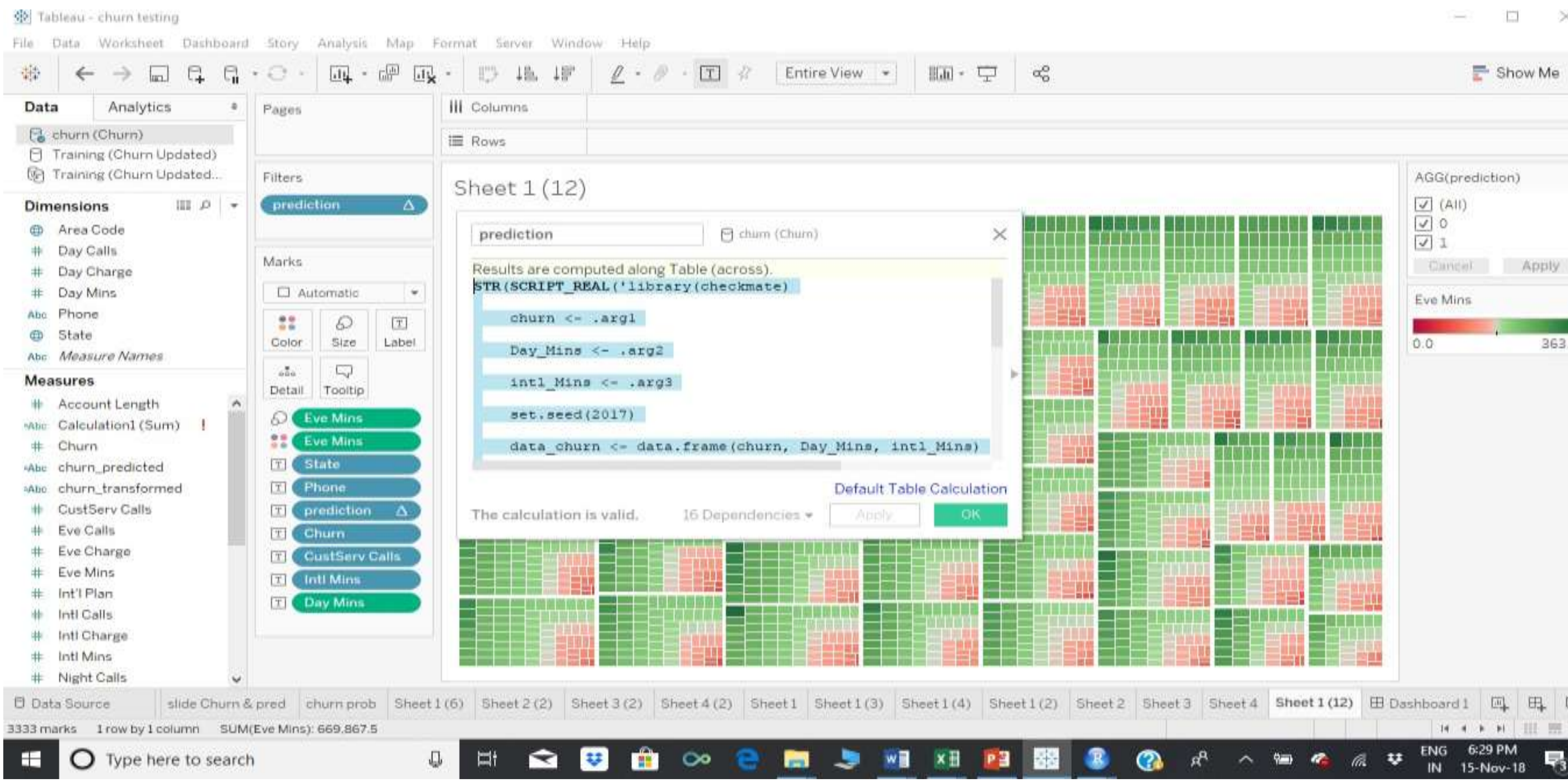
SCRIPT_REAL('library(dplyr);

mydata <- data.frame(churn=.arg1, day_mins=.arg2, day_charge=.arg3, custServ_call=.arg4, int_mins=.arg5);

lrmodel <- glm(churn ~ day_mins + day_charge + custServ_call+int_mins, data = mydata, family = "binomial");

prob <- predict(lrmodel, newdata = mydata, type = "response")',

AVG([Churn]),AVG([Day Mins]),AVG([Day Charge]),AVG([CustServ Calls]),AVG([Intl Mins]))



STR(SCRIPT_REAL('library(checkmate)

```
churn <- .arg1

Day_Mins <- .arg2

intl_Mins <- .arg3

set.seed(2017)

data_churn <- data.frame(churn, Day_Mins, intl_Mins)

intrain<-sample(1:nrow(data_churn),.7*nrow(data_churn))

training<- data_churn[intrain,]

testing<- data_churn[-intrain,]

new_data<-rbind(training,testing)

LogModel <- glm(churn ~ .,family=binomial(link="logit"),data=training)

fitted.results <- predict(LogModel,newdata=new_data,type="response")

pred_val <- ifelse(fitted.results > 0.5,1,0)

pred_val

',ATTR([Churn]),SUM([Day Mins]),sum([Intl Mins])))
```

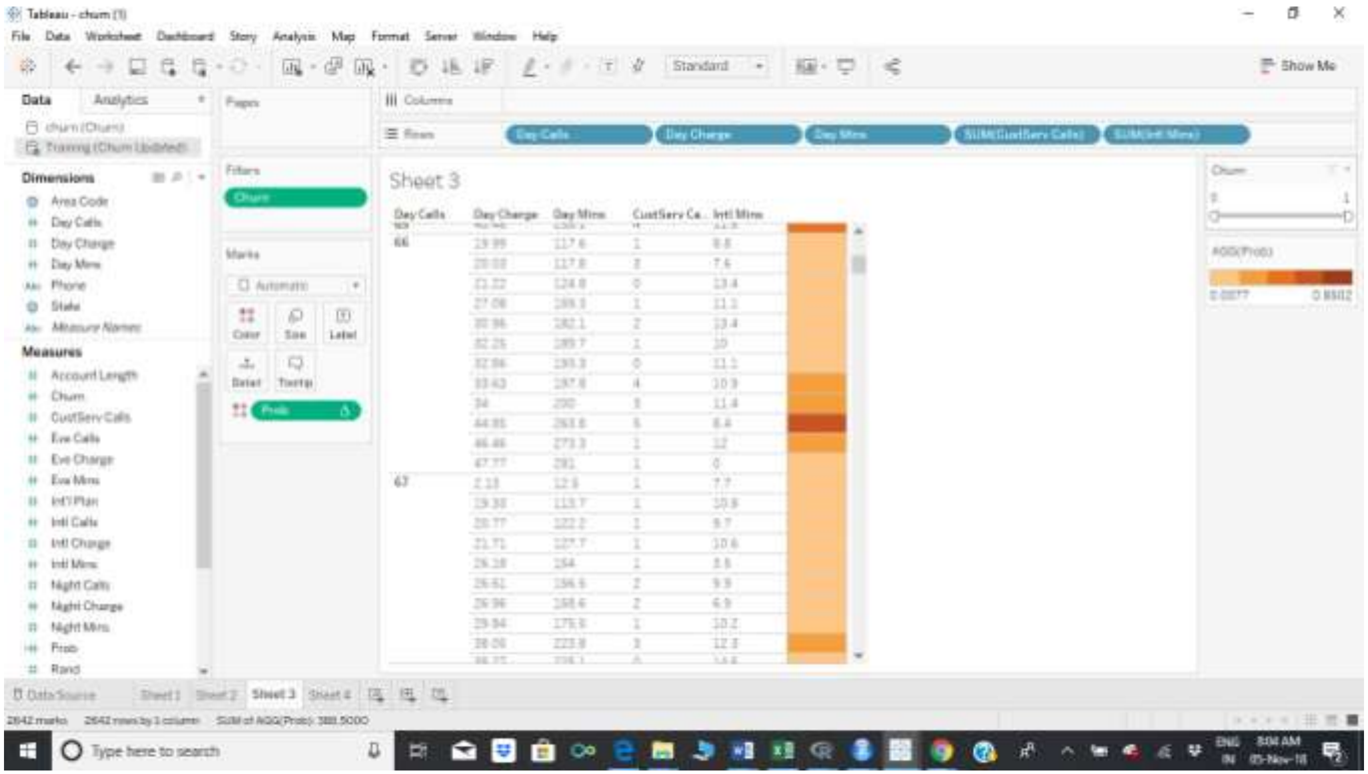


Tableau - churn (1)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Pages
Columns
Rows
Day Calls
Day Charge
Day Mins
Filters
Churn
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Prob
Dimensions
Area Code
Phone
State
Measure Names
Measures
Account Length
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Rand
Sheet 3
Data Source
Sheet 1
Sheet 2
Sheet 4
Type here to search

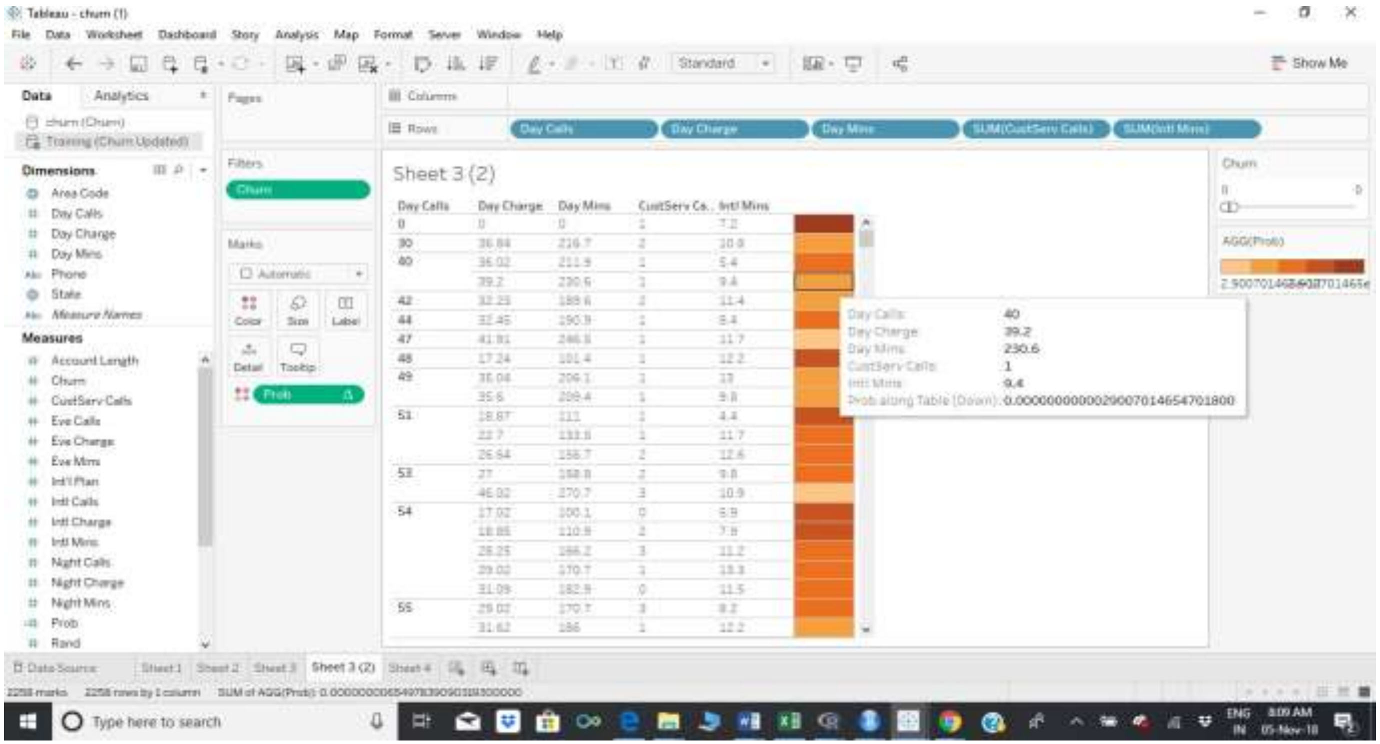


Tableau - churn (1)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Dimensions
Measures
Pages
Filters
Marks
Columns
Rows
Sheet 3 (2)
Day Calls
Day Charge
Day Mins
Intl Mins
Churn
AGG(Prob)
Sheet 3 (2)
Type here to search

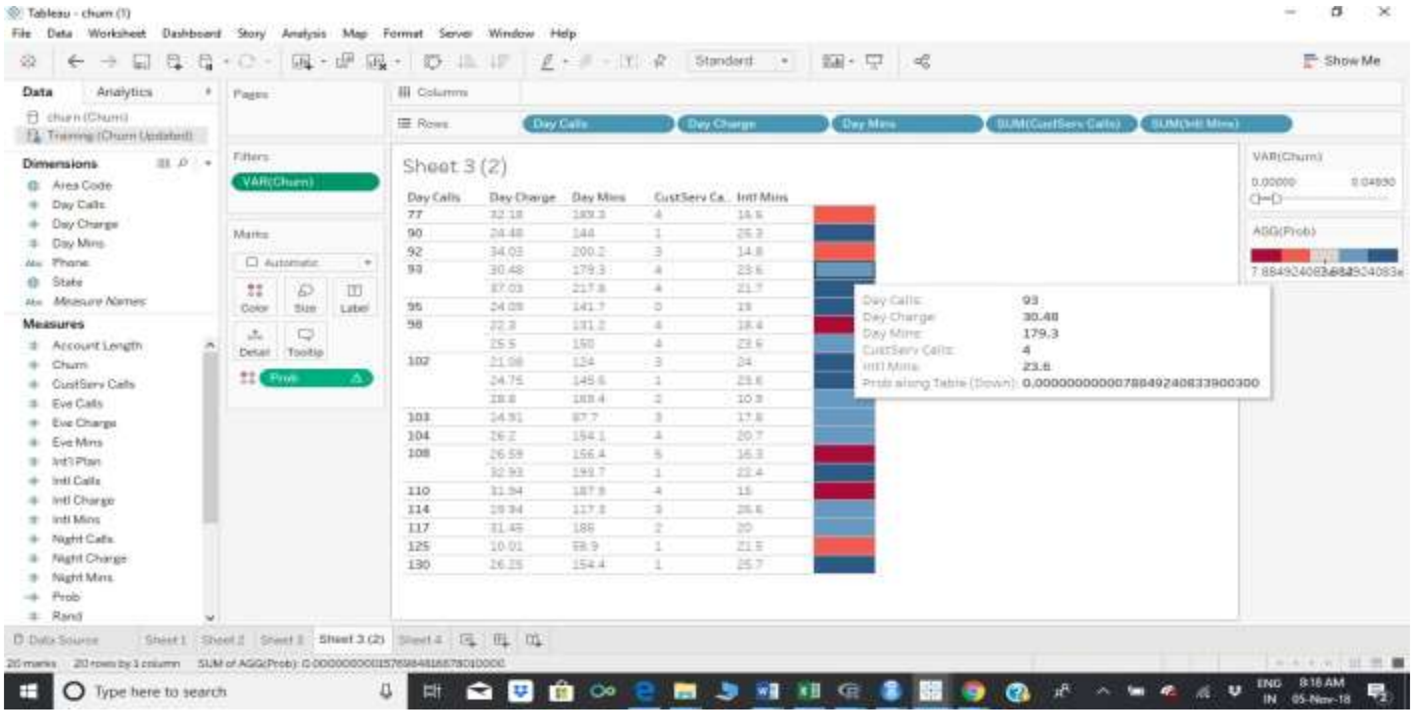


Tableau - churn (1)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
Pages
Filters
VAR(Churn)
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Prob
Columns
Rows
Day Calls
Day Charge
Day Mins
SUM(CustServ Calls)
SUM(Intl Mins)
Sheet 3 (2)
Day Calls
Day Charge
Day Mins
CustServ Ca..
Intl Mins
VAR(Churn)
AGG(Prob)
Day Calls: 93
Day Charge: 30.48
Day Mins: 179.3
CustServ Calls: 4
Intl Mins: 23.6
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 3 (2)
Sheet 4
Type here to search

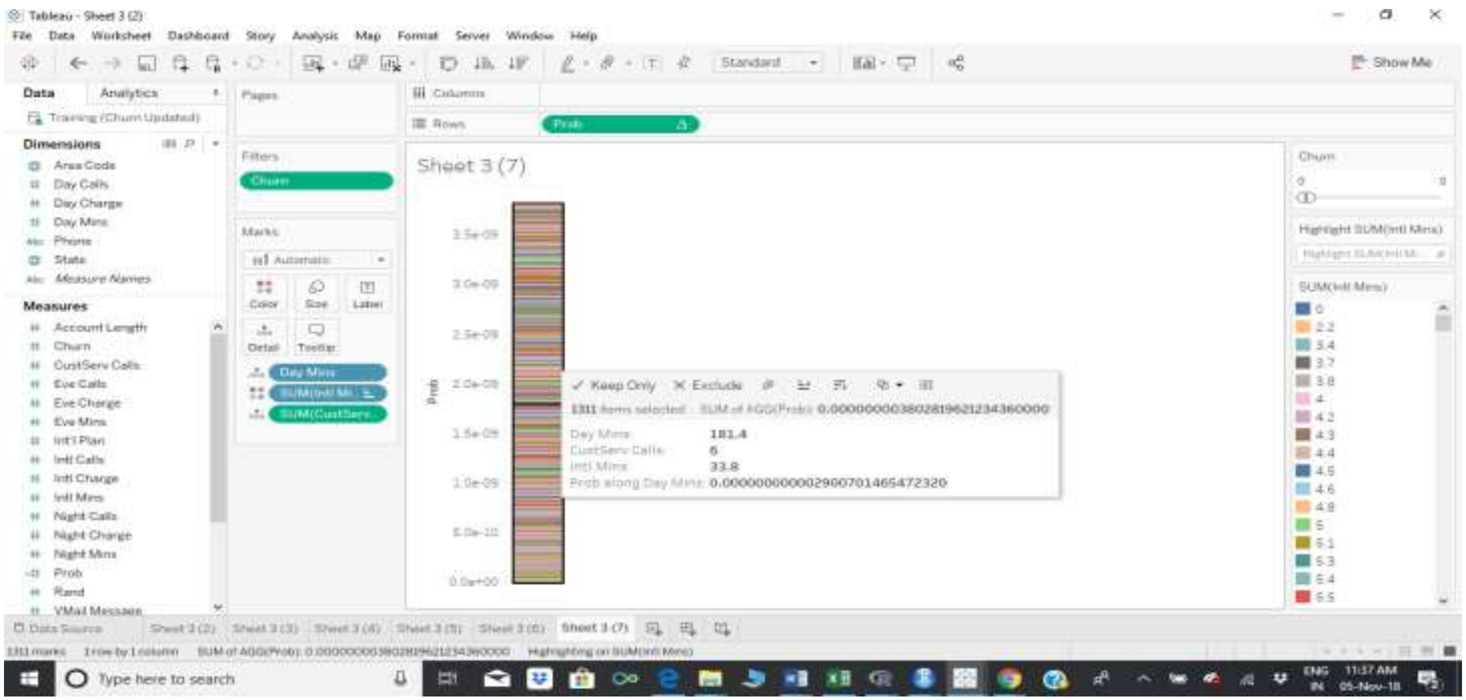


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Pages
Columns
Rows
Prob
Dimensions
Filters
Churn
Sheet 3 (7)
Marks
Measures
Keep Only
Exclude
Day Mins
181.4
6
33.8
Type here to search

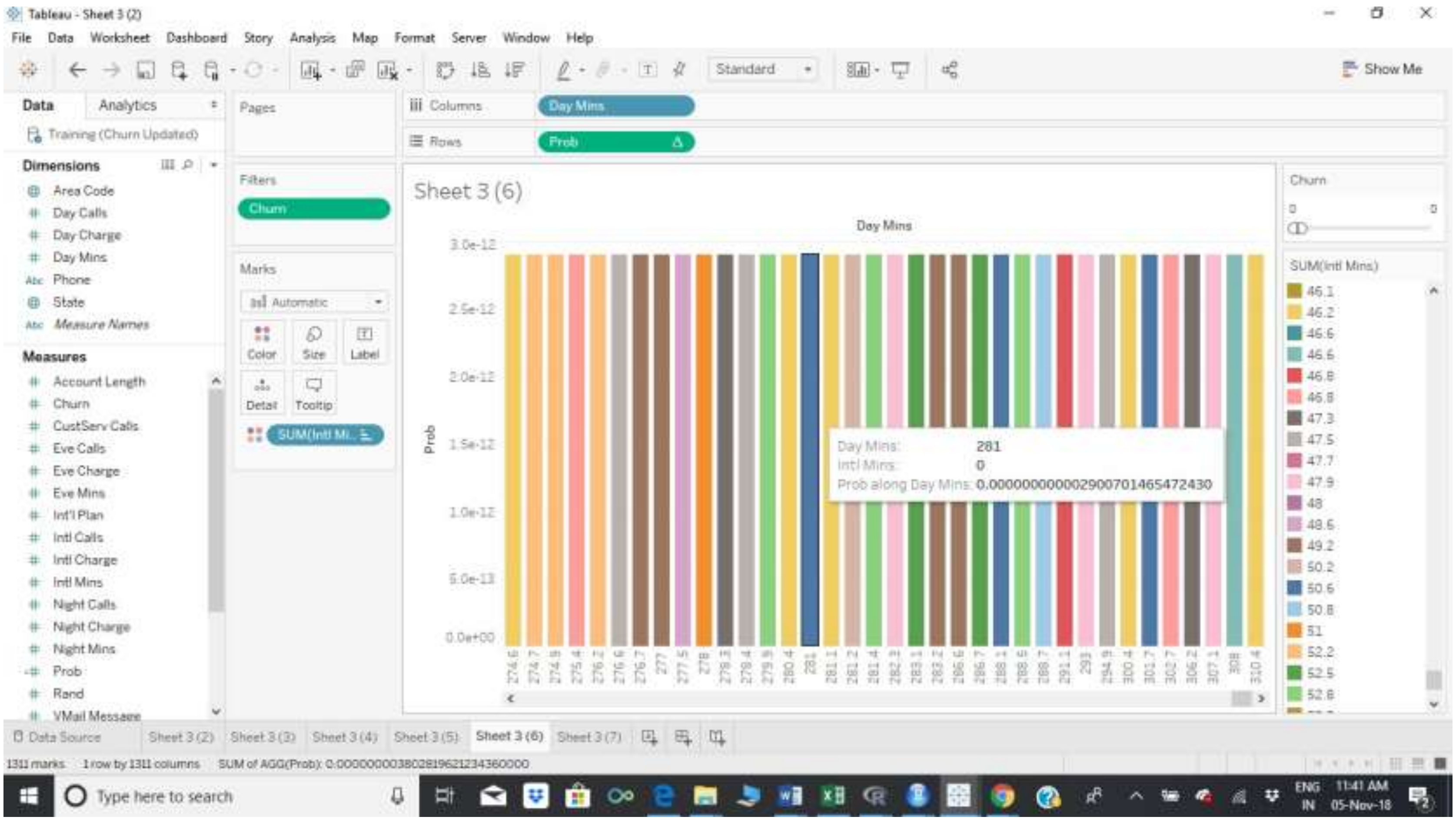


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
Pages
Filters
Churn
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Columns
Day Mins
Rows
Prob
Sheet 3 (6)
Day Mins
Prob
3.0e-12
2.5e-12
2.0e-12
1.5e-12
1.0e-12
5.0e-13
0.0e+00
Day Mins: 281
Intl Mins: 0
Prob along Day Mins: 0.000000000002900701465472430
Churn
SUM(Intl Mins)
46.1
46.2
46.6
46.6
46.8
46.8
47.3
47.5
47.7
47.9
48
48.6
49.2
50.2
50.6
50.8
51
52.2
52.5
52.8
Data Source
Sheet 3 (2)
Sheet 3 (3)
Sheet 3 (4)
Sheet 3 (5)
Sheet 3 (6)
Sheet 3 (7)
1311 marks
1 row by 1311 columns
SUM of AGG(Prob): 0.0000000003802819621234360000
Type here to search
ENG
IN
11:41 AM
05-Nov-18

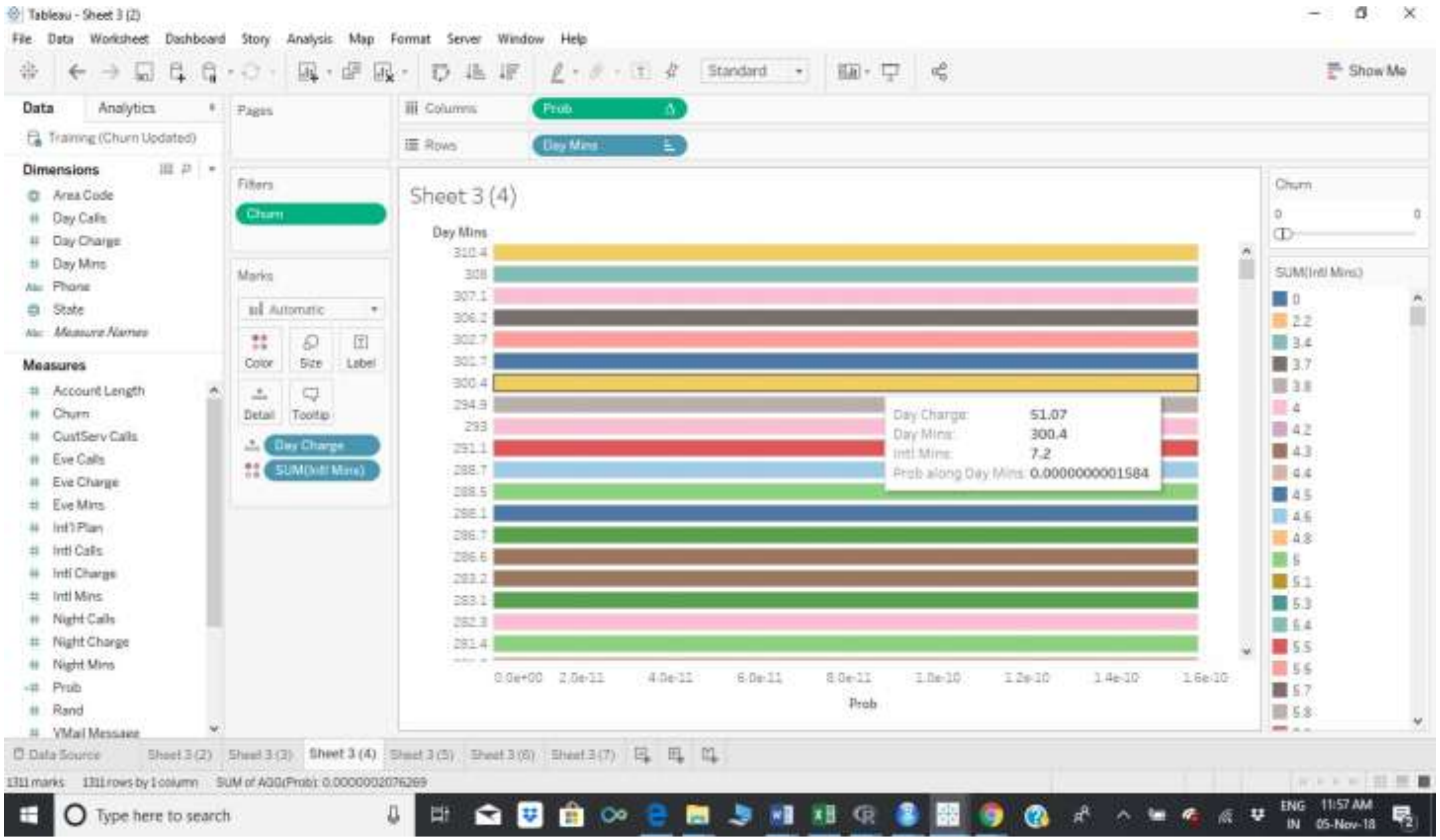


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
VMail Message
Pages
Filters
Churn
Marks
Automatic
Color Size Label
Detail Tooltip
Day Charge
SUM(Intl Mins)
Columns Prob
Rows Day Mins
Sheet 3 (4)
Day Mins
Day Charge: 51.07
Day Mins: 300.4
Intl Mins: 7.2
Prob along Day Mins: 0.0000000001584
Prob
Churn
SUM(Intl Mins)
Data Source Sheet 3 (2) Sheet 3 (3) Sheet 3 (4) Sheet 3 (5) Sheet 3 (6) Sheet 3 (7)
1311 marks 1311 rows by 1 column SUM of AGG(Prob): 0.0000002076269
Type here to search
11:57 AM 05-Nov-18

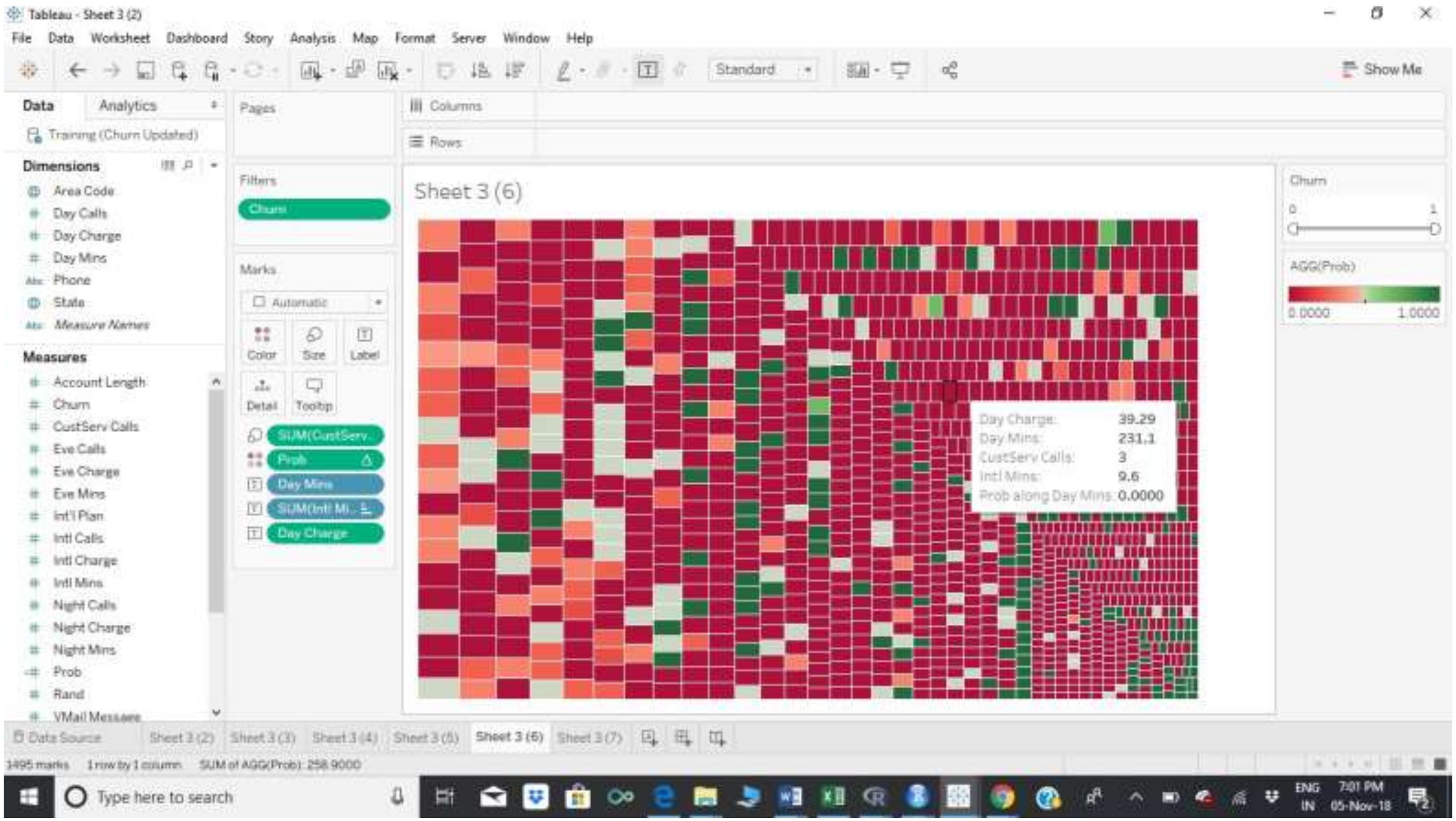


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
VMail Messages
Pages
Filters
Churn
Marks
Automatic
Color Size Label
Detail Tooltip
SUM(CustServ..
Prob
Day Mins
SUM(Intl Mi..
Day Charge
Columns
Rows
Sheet 3 (6)
Day Charge: 39.29
Day Mins: 231.1
CustServ Calls: 3
Intl Mins: 9.6
Prob along Day Mins: 0.0000
Churn
0 1
AGG(Prob)
0.0000 1.0000
Data Source Sheet 3 (2) Sheet 3 (3) Sheet 3 (4) Sheet 3 (5) Sheet 3 (6) Sheet 3 (7)
1495 marks 1 row by 1 column SUM of AGG(Prob): 258.9000
Type here to search
ENG IN 7:01 PM 05-Nov-18

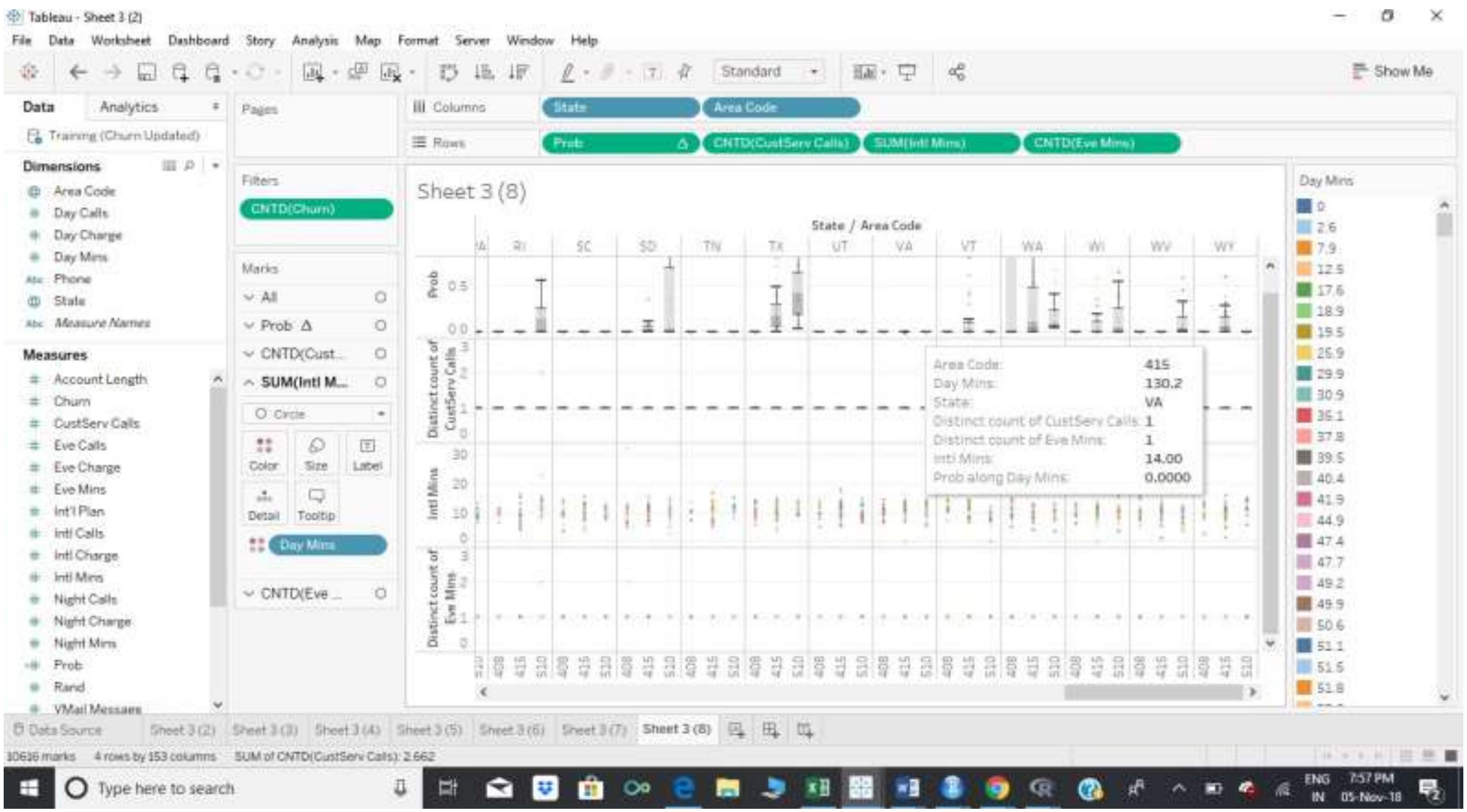


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
VMail Messages
Pages
Filters
CNTD(Churn)
Marks
All
Prob Δ
CNTD(Cust...
SUM(Intl M...
Circle
Color Size Label
Detail Tooltip
Day Mins
CNTD(Eve ...
Columns
State
Area Code
Rows
Prob
CNTD(CustServ Calls)
SUM(Intl Mins)
CNTD(Eve Mins)
Sheet 3 (8)
State / Area Code
Prob
Distinct count of CustServ Calls
Intl Mins
Distinct count of Eve Mins
Area Code: 415
Day Mins: 130.2
State: VA
Distinct count of CustServ Calls: 1
Distinct count of Eve Mins: 1
Intl Mins: 14.00
Prob along Day Mins: 0.0000
Day Mins
Data Source
Sheet 3 (2)
Sheet 3 (3)
Sheet 3 (4)
Sheet 3 (5)
Sheet 3 (6)
Sheet 3 (7)
Sheet 3 (8)
10616 marks
4 rows by 153 columns
SUM of CNTD(CustServ Calls): 2.662
Type here to search
ENG IN
7:57 PM
05-Nov-18

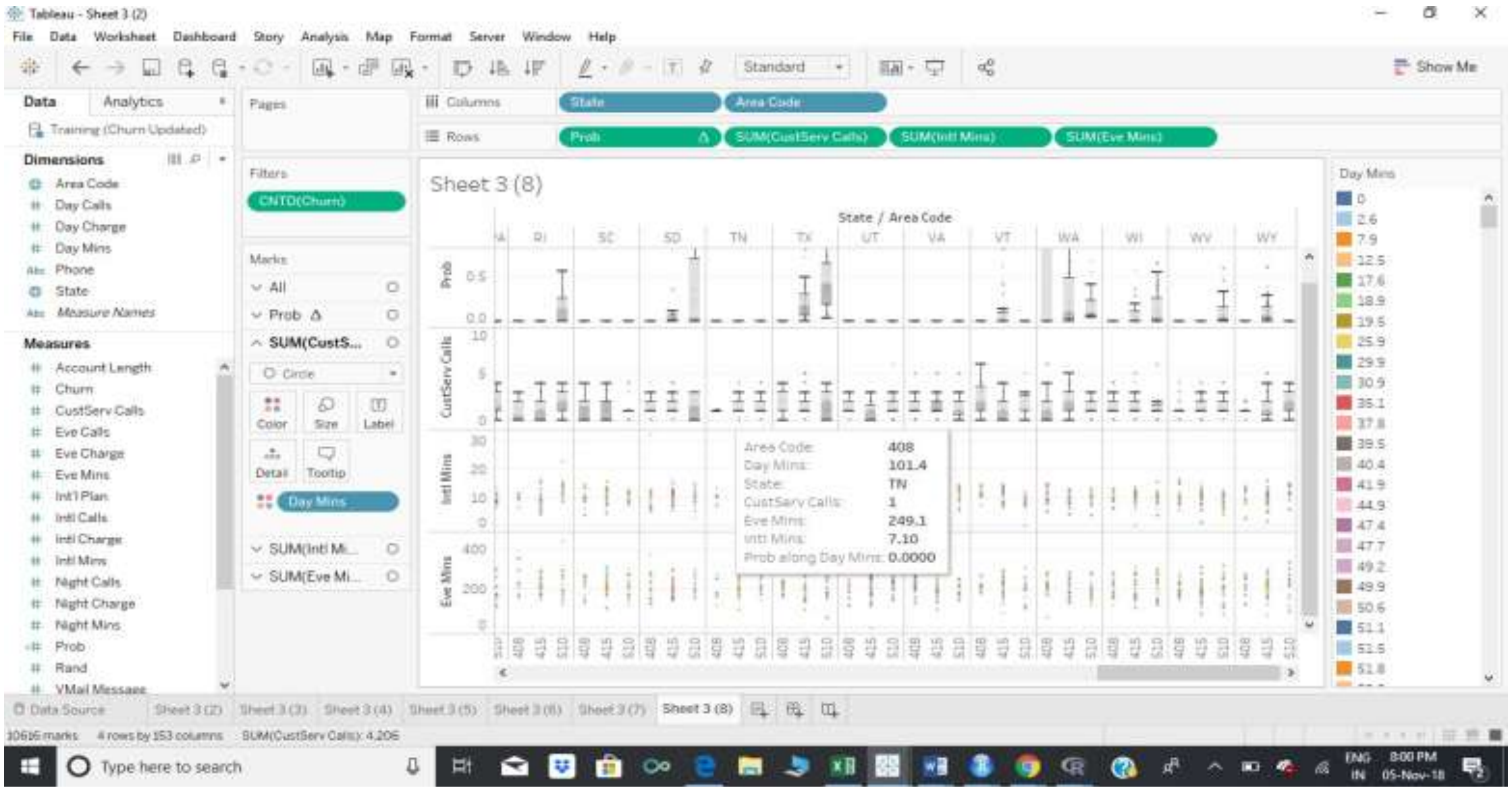


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
Pages
Filters
CNTD(Churn)
Marks
All
Prob
Color Size Label
Detail Tooltip
Day Mins
Columns
State
Area Code
Rows
Prob
SUM(CustServ Calls)
SUM(Intl Mins)
SUM(Eve Mins)
Sheet 3 (8)
State / Area Code
Area Code: 408
Day Mins: 101.4
State: TN
CustServ Calls: 1
Eve Mins: 249.1
Intl Mins: 7.10
Prob along Day Mins: 0.0000
Day Mins
Data Source
Sheet 3 (8)
Type here to search

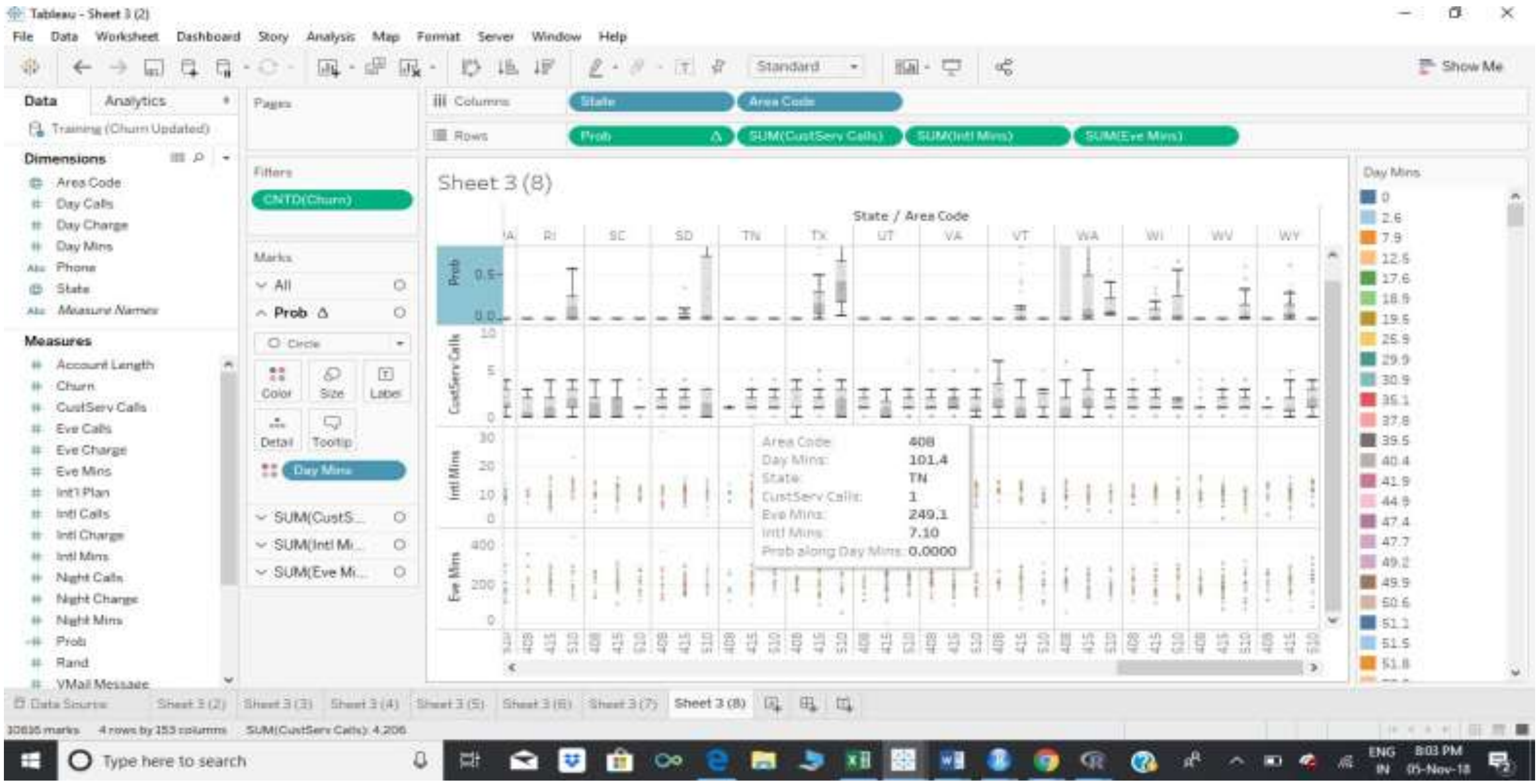


Tableau - Sheet 3 (2)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
VMail Message
Pages
Filters
CNTD(Churn)
Marks
All
Prob
Circle
Color Size Label
Detail Tooltip
Day Mins
SUM(CustS...
SUM(Intl Mi...
SUM(Eve Mi...
Columns State Area Code
Rows Prob SUM(CustServ Calls) SUM(Intl Mins) SUM(Eve Mins)
Sheet 3 (8)
State / Area Code
Area Code: 408
Day Mins: 101.4
State: TN
CustServ Calls: 1
Eve Mins: 249.1
Intl Mins: 7.10
Prob along Day Mins: 0.0000
Day Mins
Sheet 3 (2) Sheet 3 (3) Sheet 3 (4) Sheet 3 (5) Sheet 3 (6) Sheet 3 (7) Sheet 3 (8)
4 rows by 153 columns SUM(CustServ Calls): 4,206
Type here to search
8:03 PM 05-Nov-18

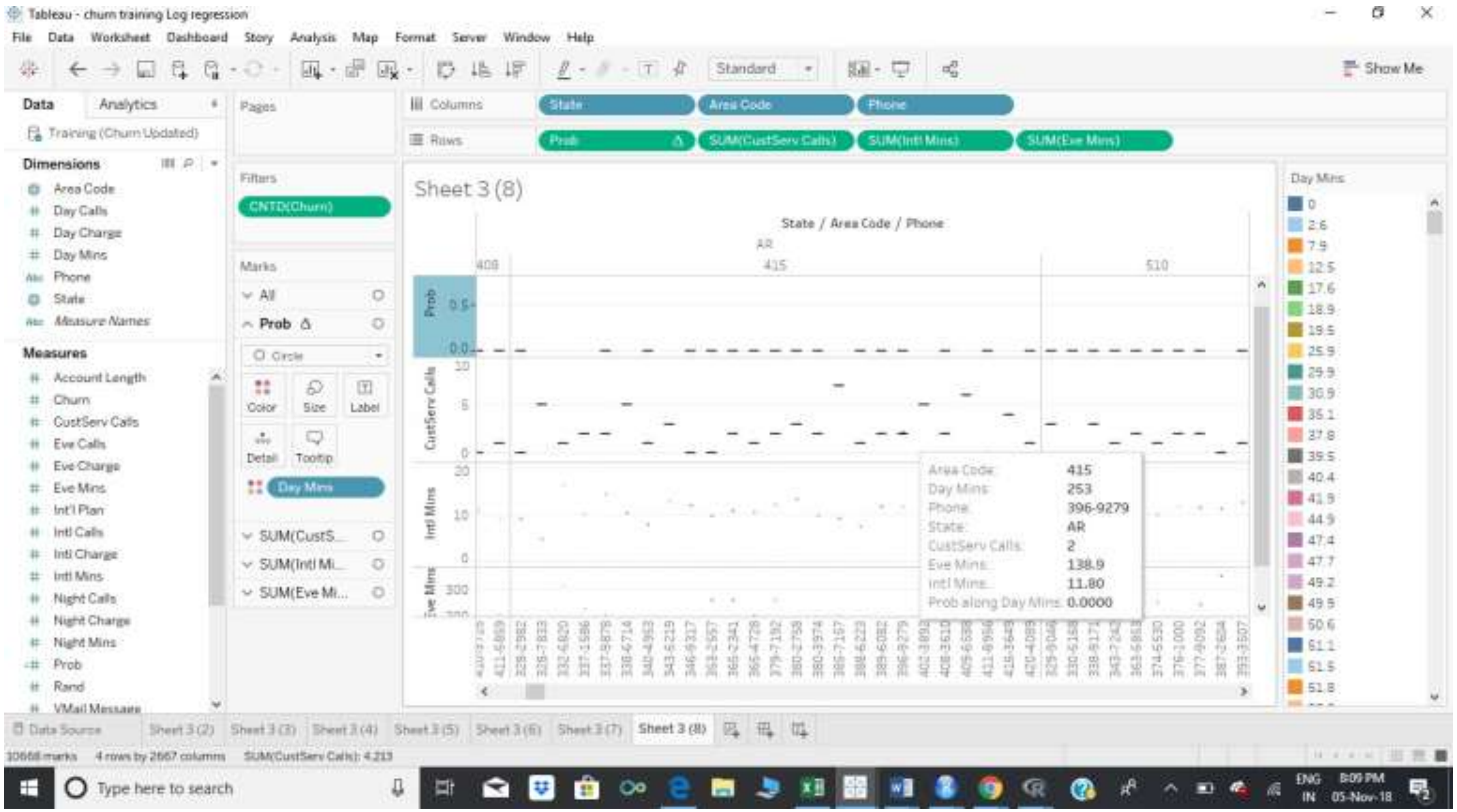


Tableau - churn training Log regression
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
Pages
Filters
CNTD(Churn)
Marks
All
Prob
Circle
Color Size Label
Detail Tooltip
Day Mins
SUM(CustS...
SUM(Intl Mi...
SUM(Eve Mi...
Columns
State
Area Code
Phone
Rows
Prob
SUM(CustServ Calls)
SUM(Intl Mins)
SUM(Eve Mins)
Sheet 3 (8)
State / Area Code / Phone
AR
408
415
510
Prob
CustServ Calls
Intl Mins
Eve Mins
Area Code: 415
Day Mins: 253
Phone: 396-9279
State: AR
CustServ Calls: 2
Eve Mins: 138.9
Intl Mins: 11.80
Prob along Day Mins: 0.0000
Day Mins
0
2.6
7.9
12.5
17.6
18.9
19.5
25.9
29.9
30.9
35.1
37.8
39.5
40.4
41.9
44.9
47.4
47.7
49.2
49.5
50.6
51.1
51.5
51.8
Data Source
Sheet 3 (2)
Sheet 3 (3)
Sheet 3 (4)
Sheet 3 (5)
Sheet 3 (6)
Sheet 3 (7)
Sheet 3 (8)
4 rows by 2667 columns
SUM(CustServ Calls): 4,213
Type here to search
ENG
IN
05-Nov-18

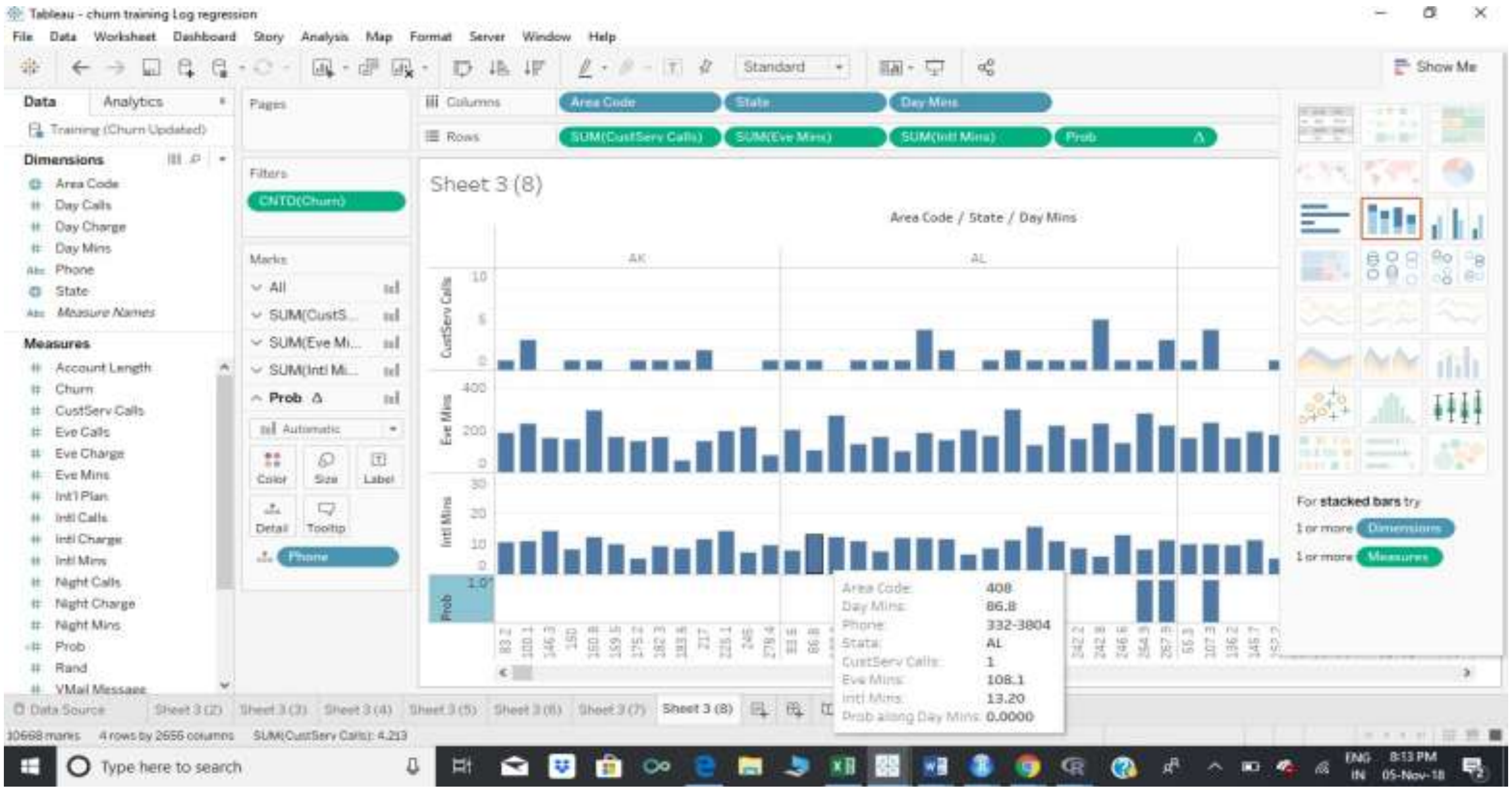


Tableau - churn training Log regression
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Prob
Rand
VMail Message
Pages
Filters
CNTD(Churn)
Marks
All
SUM(CustS...
SUM(Eve Mi...
SUM(Intl Mi...
Prob
Automatic
Color Size Label
Detail Tooltip
Phone
Columns
Area Code
State
Day Mins
Rows
SUM(CustServ Calls)
SUM(Eve Mins)
SUM(Intl Mins)
Prob
Sheet 3 (8)
Area Code / State / Day Mins
AK
AL
CustServ Calls
Eve Mins
Intl Mins
Prob
Area Code: 408
Day Mins: 86.8
Phone: 332-3804
State: AL
CustServ Calls: 1
Eve Mins: 108.1
Intl Mins: 13.20
Prob along Day Mins: 0.0000
For stacked bars try
1 or more Dimensions
1 or more Measures
Data Source
Sheet 3 (8)
10668 marks 4 rows by 2656 columns SUM(CustServ Calls): 4,213
Type here to search
8:13 PM
05-Nov-18

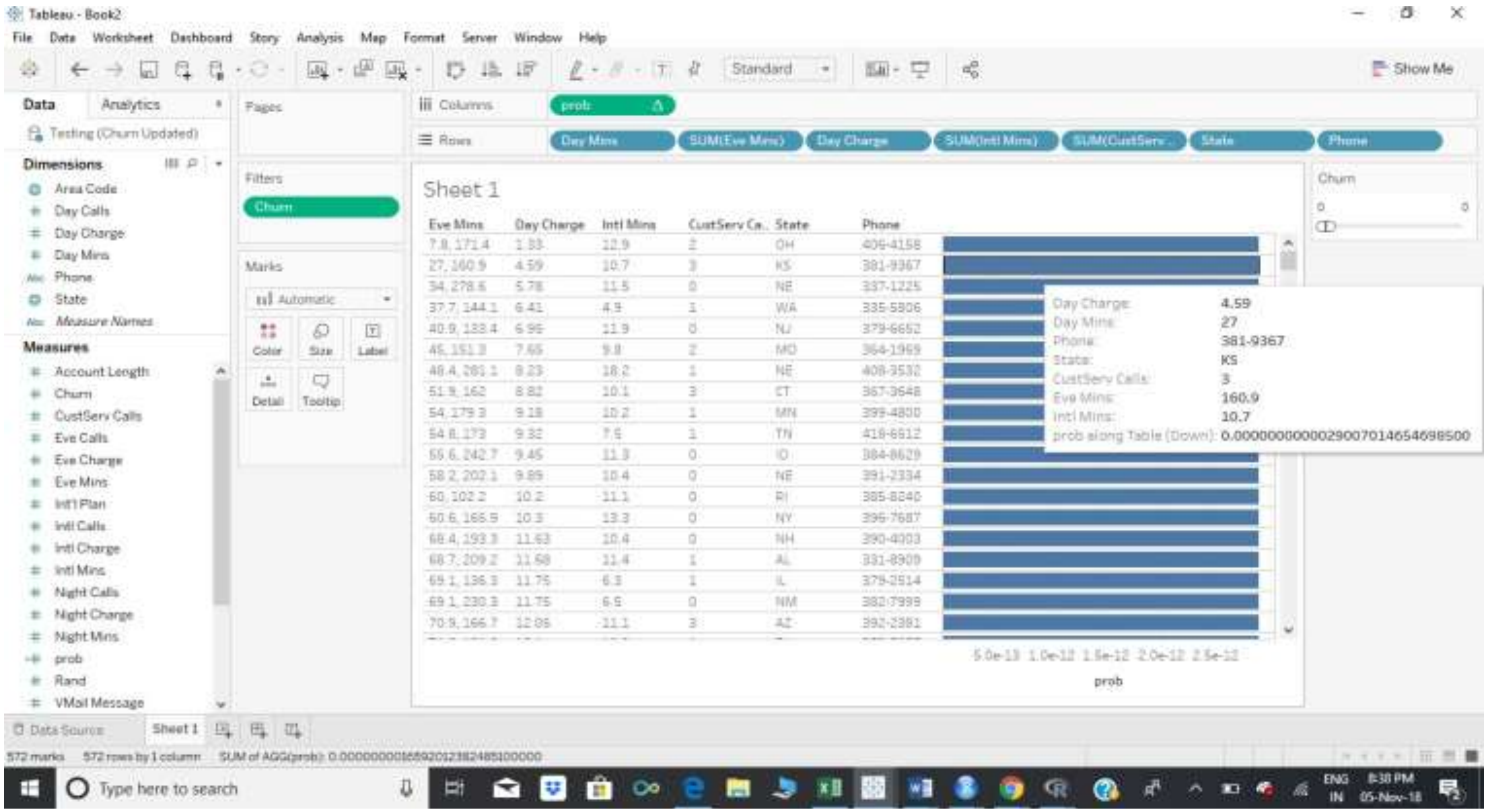


Tableau - Book2
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Testing (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prob
Rand
VMail Message
Pages
Filters
Churn
Marks
Automatic
Color Size Label
Detail Tooltip
Columns
prob
Rows
Day Mins
SUM(Eve Mins)
Day Charge
SUM(Intl Mins)
SUM(CustServ...
State
Phone
Sheet 1
Eve Mins Day Charge Intl Mins CustServ Ca.. State Phone
Day Charge: 4.59
Day Mins: 27
Phone: 381-9367
State: KS
CustServ Calls: 3
Eve Mins: 160.9
Intl Mins: 10.7
prob along Table (Down): 0.0000000000002900701465469850
prob
Churn
Data Source
Sheet 1
Type here to search

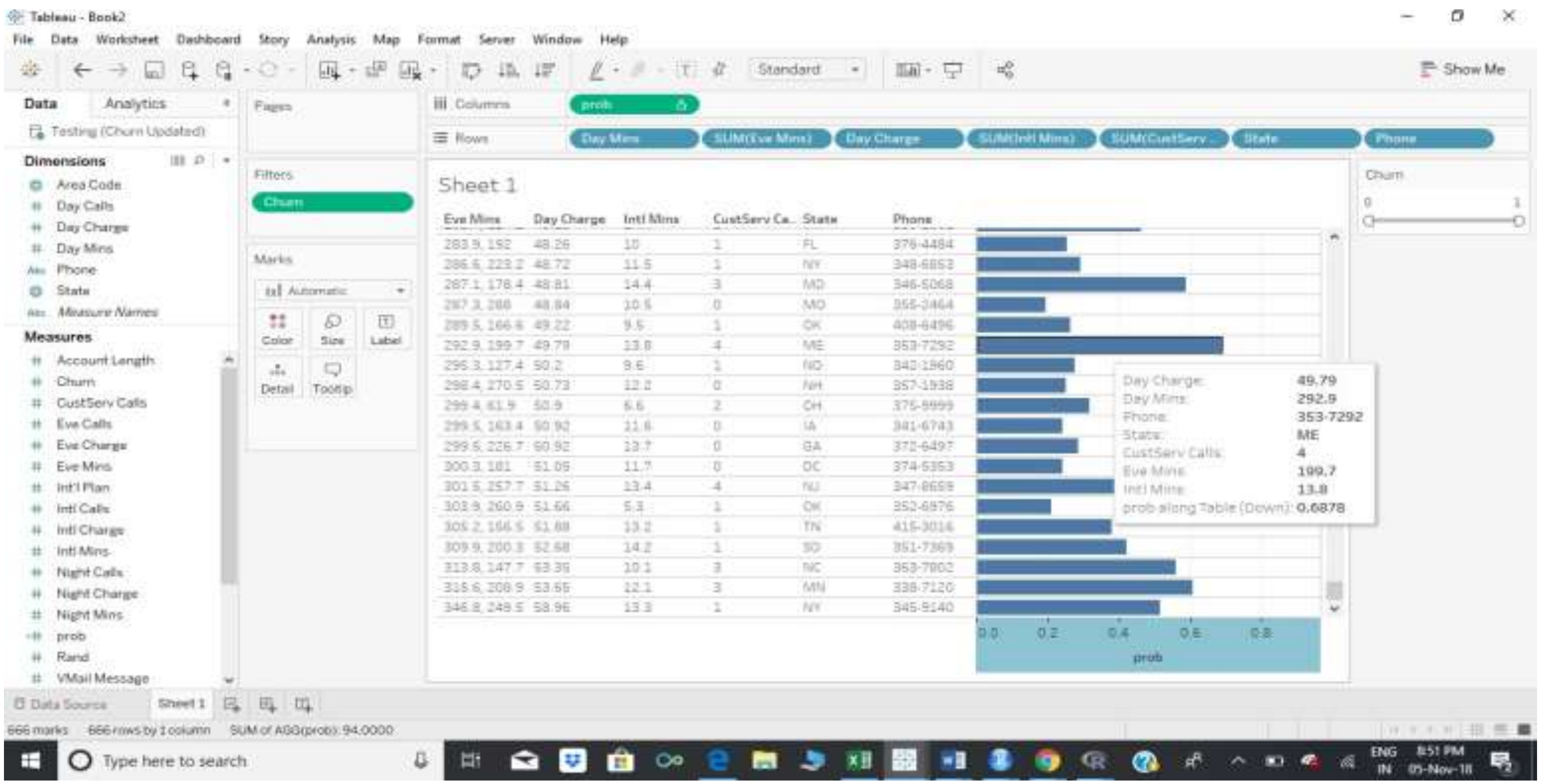
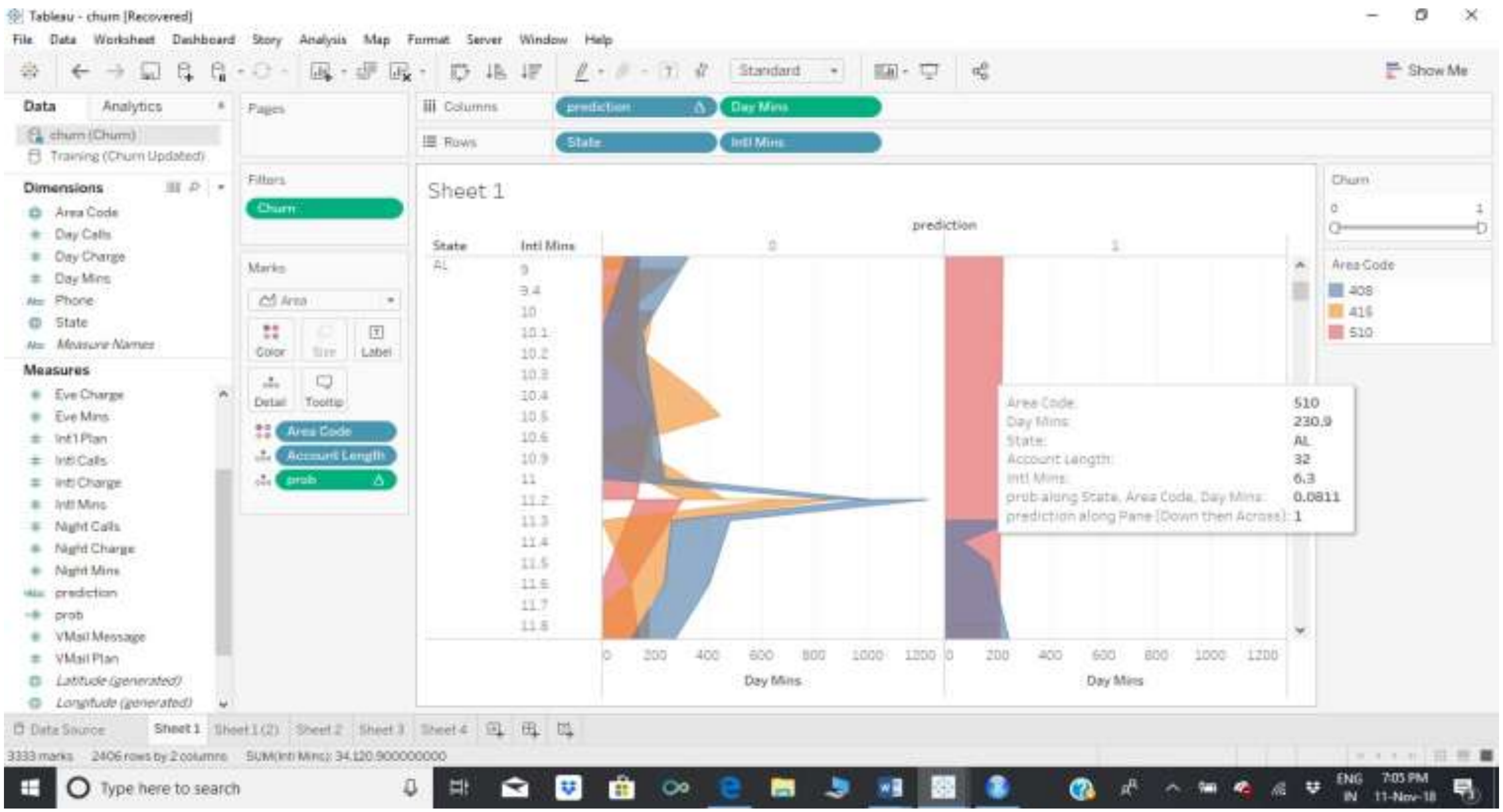


Tableau - Book2
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
Testing (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prob
Rand
VMail Message
Pages
Filters
Churn
Marks
Automatic
Color Size Label
Detail Tooltip
Columns
prob
Rows
Day Mins
SUM(Eve Mins)
Day Charge
SUM(Intl Mins)
SUM(CustServ ...
State
Phone
Sheet 1
Eve Mins Day Charge Intl Mins CustServ Ca.. State Phone
283.9, 192 48.26 10 1 FL 376-4484
286.6, 223.2 48.72 11.5 1 NY 348-6853
287.1, 178.4 48.81 14.4 3 MD 346-5068
292.9, 199.7 49.79 13.8 4 ME 353-7292
296.3, 127.4 50.2 9.6 1 ND 342-1960
298.4, 270.5 50.73 12.2 0 NH 357-1938
299.4, 61.9 50.9 6.6 2 OH 375-9999
299.6, 226.7 50.92 13.7 0 GA 372-6497
300.3, 181 51.05 11.7 0 DC 374-5353
301.5, 257.7 51.26 13.4 4 NJ 347-8659
303.9, 260.9 51.66 5.3 1 OK 352-6976
313.8, 147.7 53.35 10.1 3 NC 353-7802
315.6, 208.9 53.65 12.1 3 MN 338-7120
346.8, 249.5 58.96 13.3 1 NY 345-9140
0.0 0.2 0.4 0.6 0.8
prob
Day Charge: 49.79
Day Mins: 292.9
Phone: 353-7292
State: ME
CustServ Calls: 4
Eve Mins: 199.7
Intl Mins: 13.8
prob along Table (Down): 0.6878
Churn
0 1
Data Source
Sheet 1
666 marks 666 rows by 1 column SUM of AGG(prob): 94.0000
Type here to search
ENG IN 8:51 PM 05-Nov-18



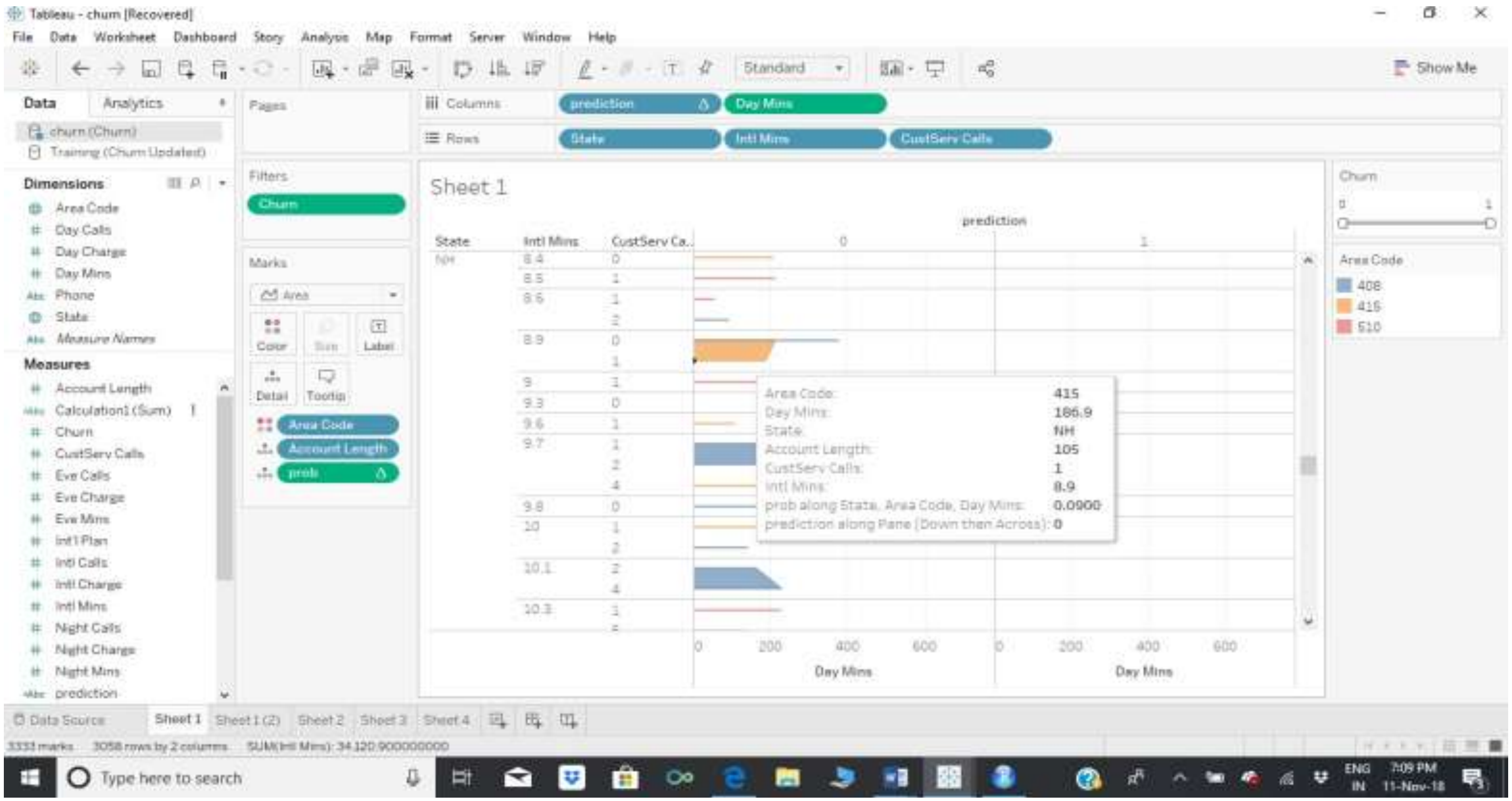


Tableau - churn [Recovered]
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Pages
Columns prediction Day Mins
Rows State Intl Mins CustServ Calls
Filters
Churn
Marks
Area
Color Size Label
Detail Tooltip
Area Code
Account Length
prob
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Sheet 1
prediction
State Intl Mins CustServ Ca..
Area Code: 415
Day Mins: 186.9
State: NH
Account Length: 105
CustServ Calls: 1
Intl Mins: 8.9
prob along State, Area Code, Day Mins: 0.0900
prediction along Pane (Down then Across): 0
Day Mins
Day Mins
Churn
Area Code
408
415
510
Data Source Sheet 1 Sheet 1 (2) Sheet 2 Sheet 3 Sheet 4
Type here to search

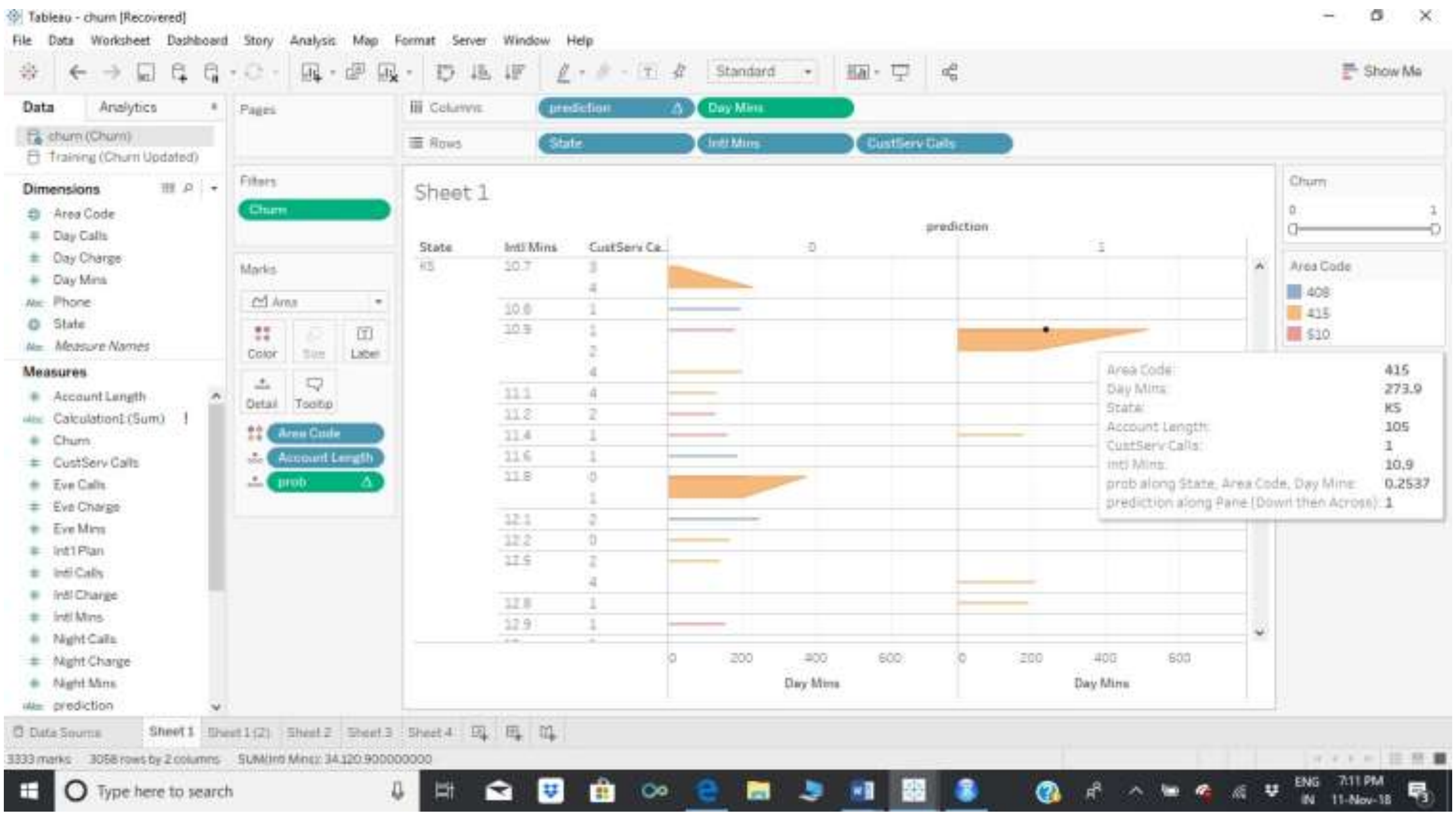


Tableau - churn [Recovered]
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Pages
Columns
prediction
Day Mins
Rows
State
Intl Mins
CustServ Calls
Filters
Churn
Marks
Sheet 1
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Color
Label
Detail
Tooltip
Area Code
Account Length
prob
Area Code
408
415
510
Area Code: 415
Day Mins: 273.9
State: KS
Account Length: 105
CustServ Calls: 1
Intl Mins: 10.9
prob along State, Area Code, Day Mins: 0.2537
prediction along Pane (Down then Across): 1
Day Mins
Day Mins
Data Source
Sheet 1
Sheet 2
Sheet 3
Sheet 4
3333 marks
Type here to search

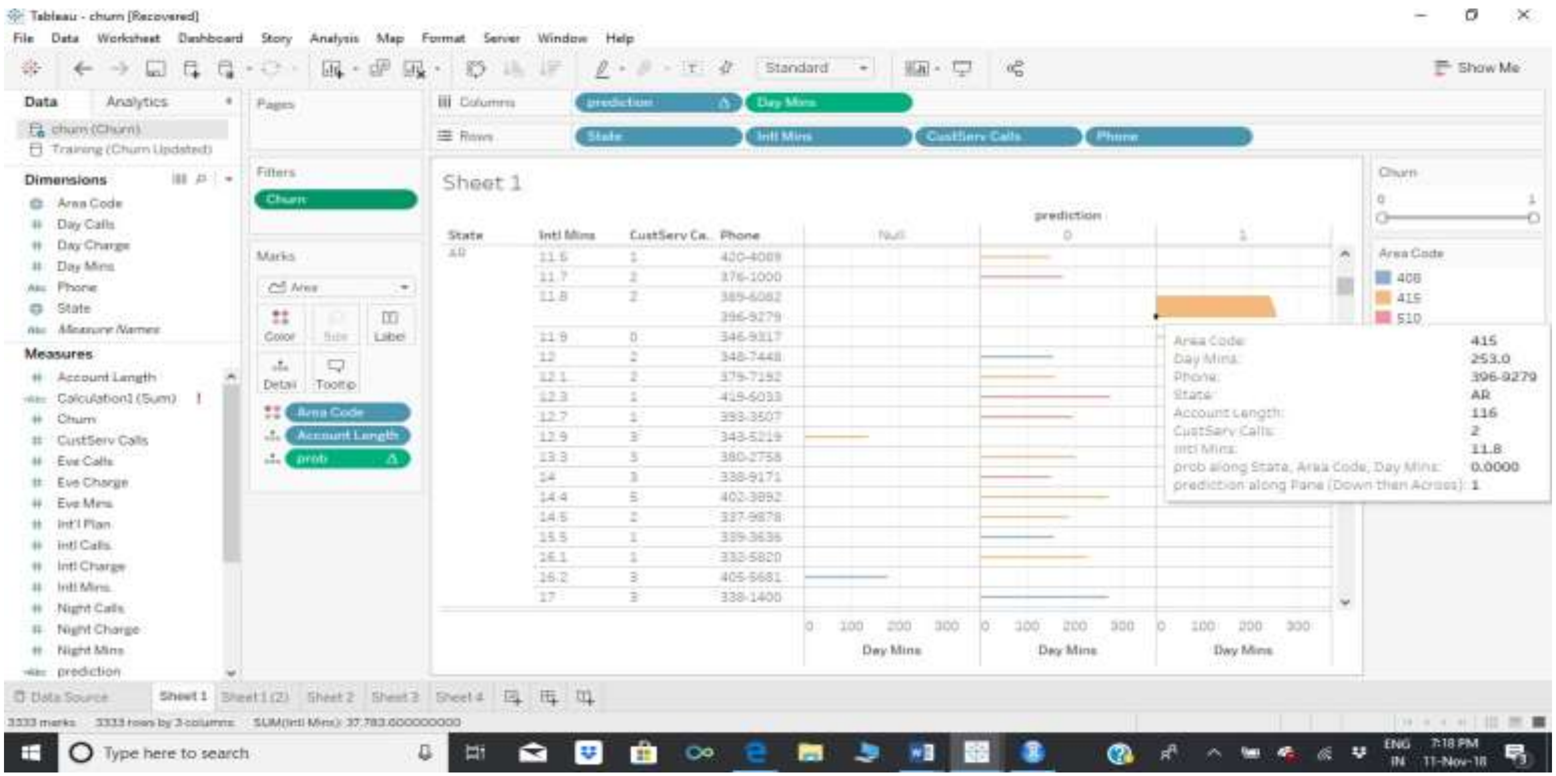


Tableau - churn (Recovered)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Pages
Columns
prediction
Day Mins
Rows
State
Intl Mins
CustServ Calls
Phone
Filters
Churn
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Marks
Area
Color
Size
Label
Detail
Tooltip
Area Code
Account Length
prob
Sheet 1
prediction
State
Intl Mins
CustServ Ca..
Phone
Day Mins
Churn
Area Code
408
415
510
Area Code: 415
Day Mins: 253.0
Phone: 396-9279
State: AR
Account Length: 116
CustServ Calls: 2
Intl Mins: 11.8
prob along State, Area Code, Day Mins: 0.0000
prediction along Pane (Down then Across): 1
Data Source
Sheet 1
Sheet 1 (2)
Sheet 2
Sheet 3
Sheet 4
3333 marks
3333 rows by 3 columns
Type here to search
7:18 PM
11-Nov-18

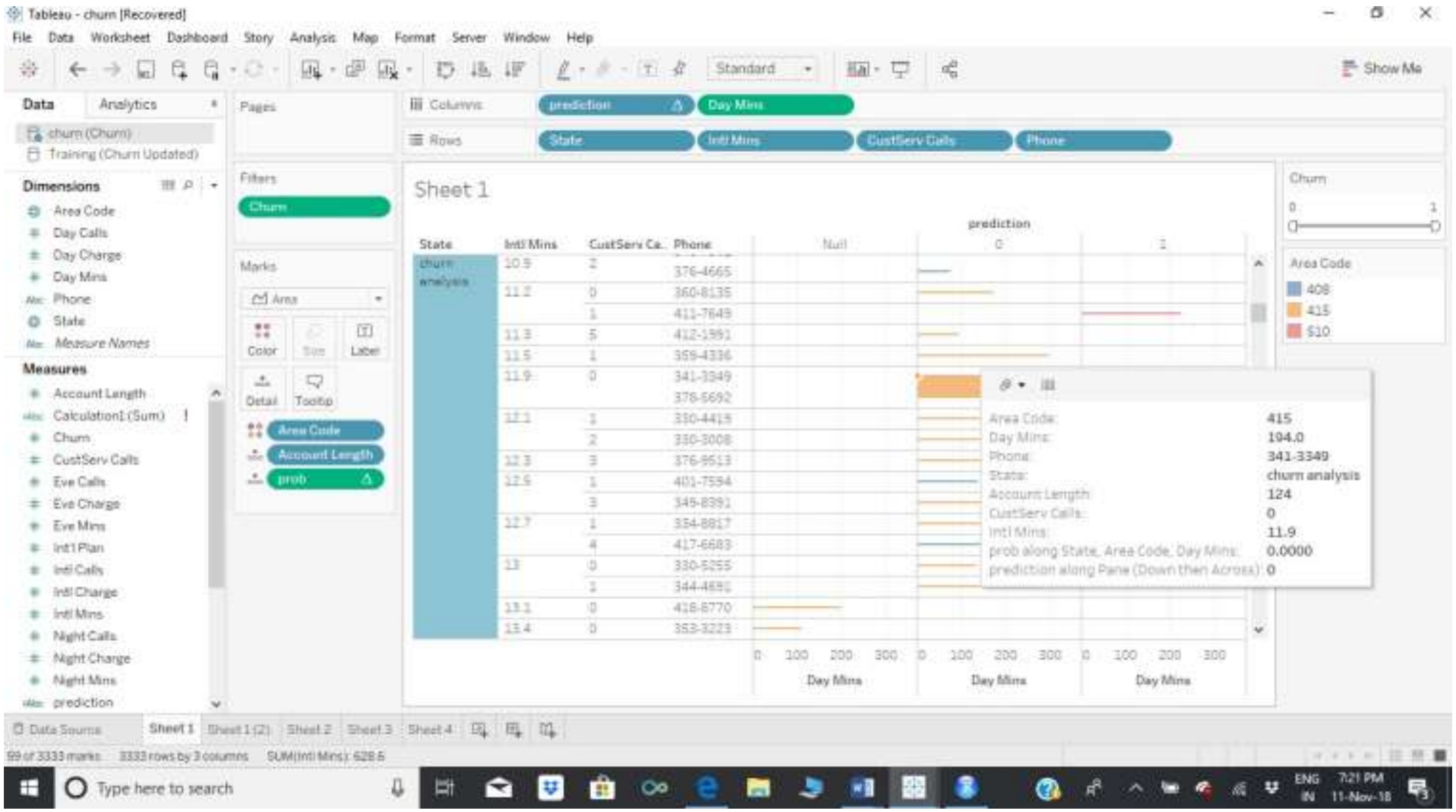


Tableau - churn [Recovered]
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
churn (Churn)
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Pages
Filters
Churn
Marks
Area
Color Size Label
Detail Tooltip
Area Code
Account Length
prob
Columns prediction Day Mins
Rows State Intl Mins CustServ Calls Phone
Sheet 1
prediction
State Intl Mins CustServ Ca.. Phone
Null 0 1
churn analysis
Area Code: 415
Day Mins: 194.0
Phone: 341-3349
State: churn analysis
Account Length: 124
CustServ Calls: 0
Intl Mins: 11.9
prob along State, Area Code, Day Mins: 0.0000
prediction along Pane (Down then Across): 0
Day Mins Day Mins Day Mins
Churn
0 1
Area Code
408
415
510
Data Source Sheet 1 Sheet 1 (2) Sheet 2 Sheet 3 Sheet 4
99 of 3333 marks 3333 rows by 3 columns SUM(Intl Mins): 628.6
Type here to search
ENG IN 7:21 PM 11-Nov-18

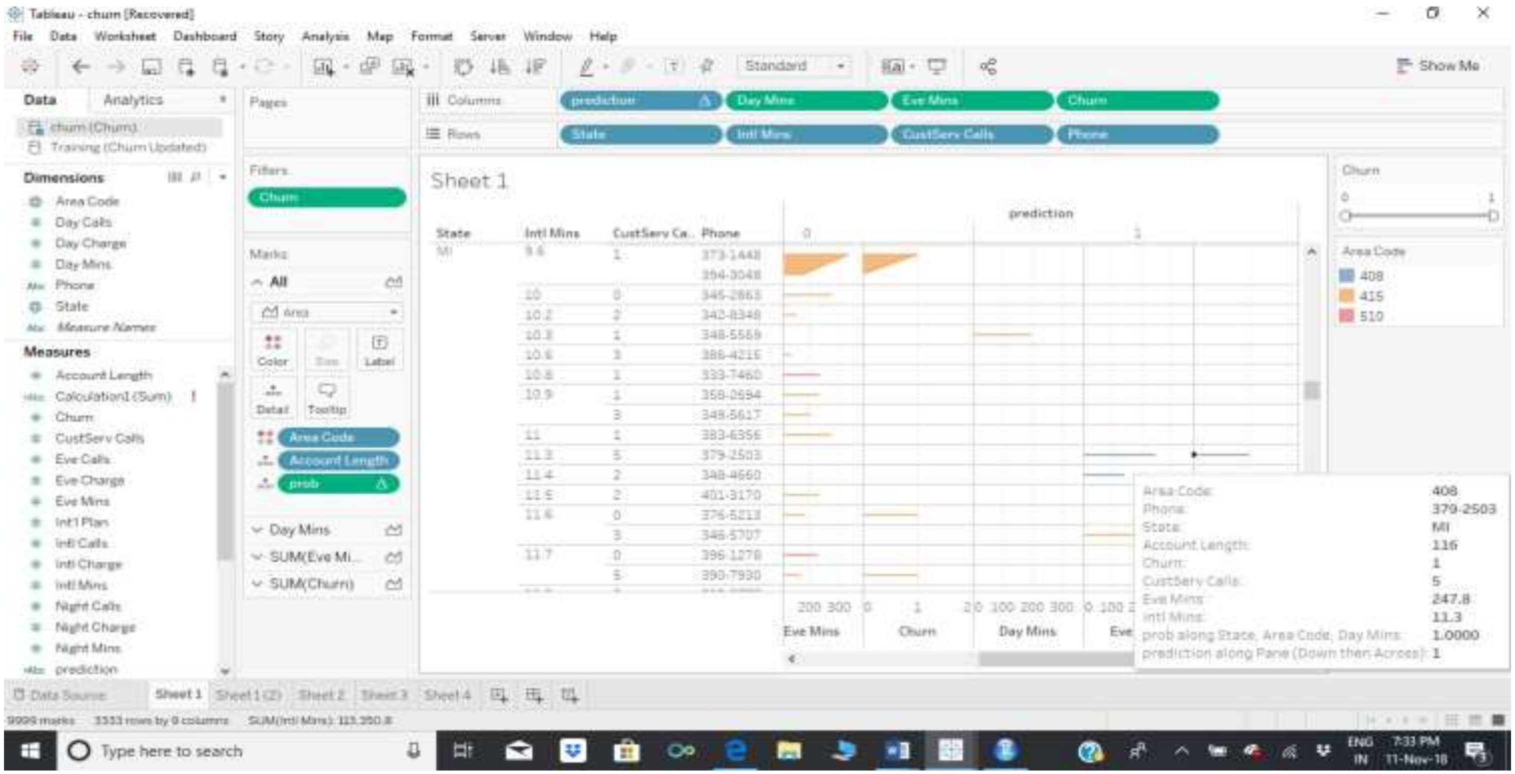


Tableau - churn [Recovered]
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Columns
prediction
Day Mins
Eve Mins
Churn
Rows
State
Intl Mins
CustServ Calls
Phone
Sheet 1
prediction
Area Code
408
415
510
Area Code: 408
Phone: 379-2503
State: MI
Account Length: 116
Churn: 1
CustServ Calls: 5
Eve Mins: 247.8
Intl Mins: 11.3
prob along State, Area Code, Day Mins: 1.0000
prediction along Pane (Down then Across): 1
Eve Mins
Churn
Day Mins
Type here to search

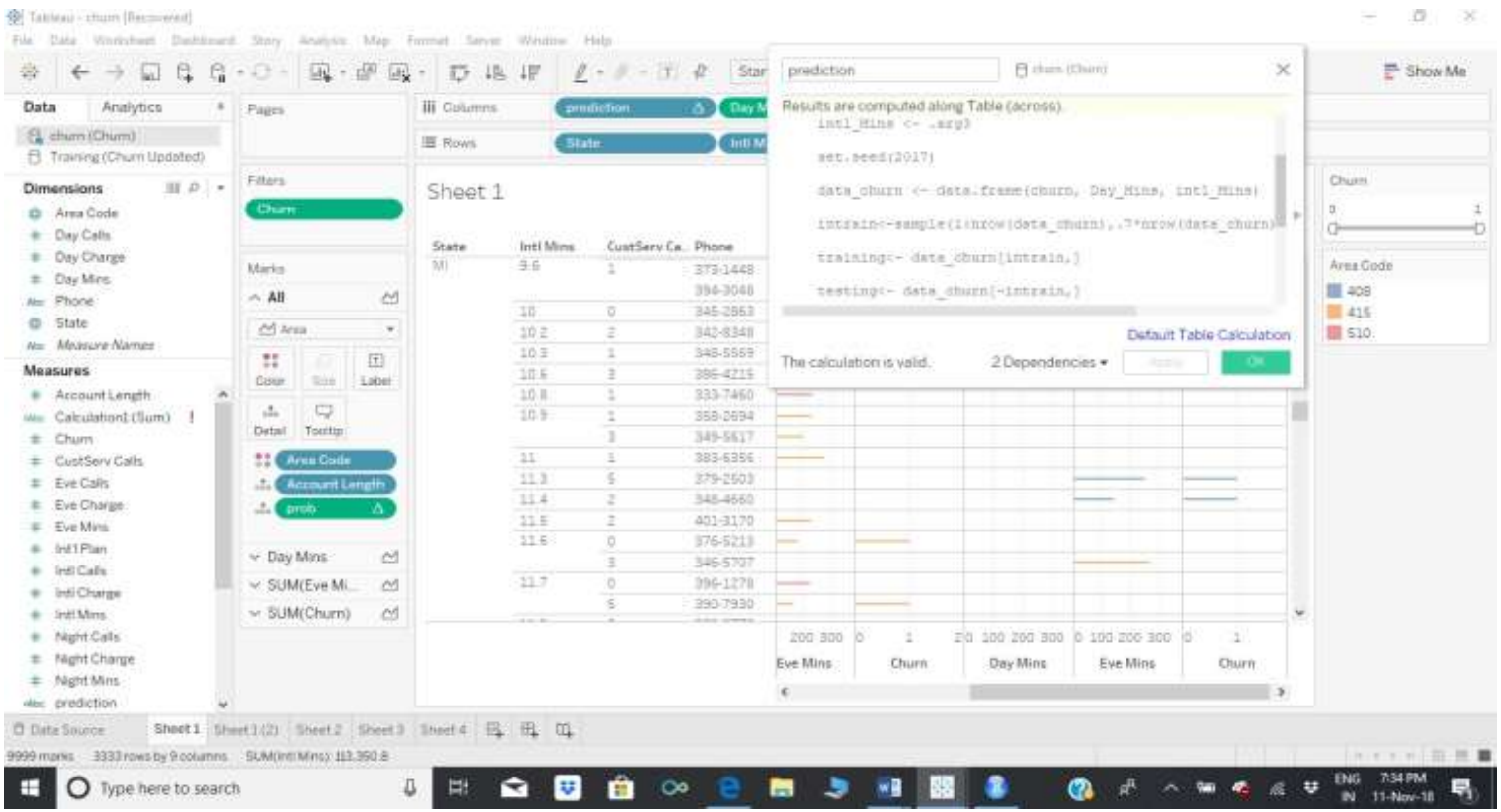
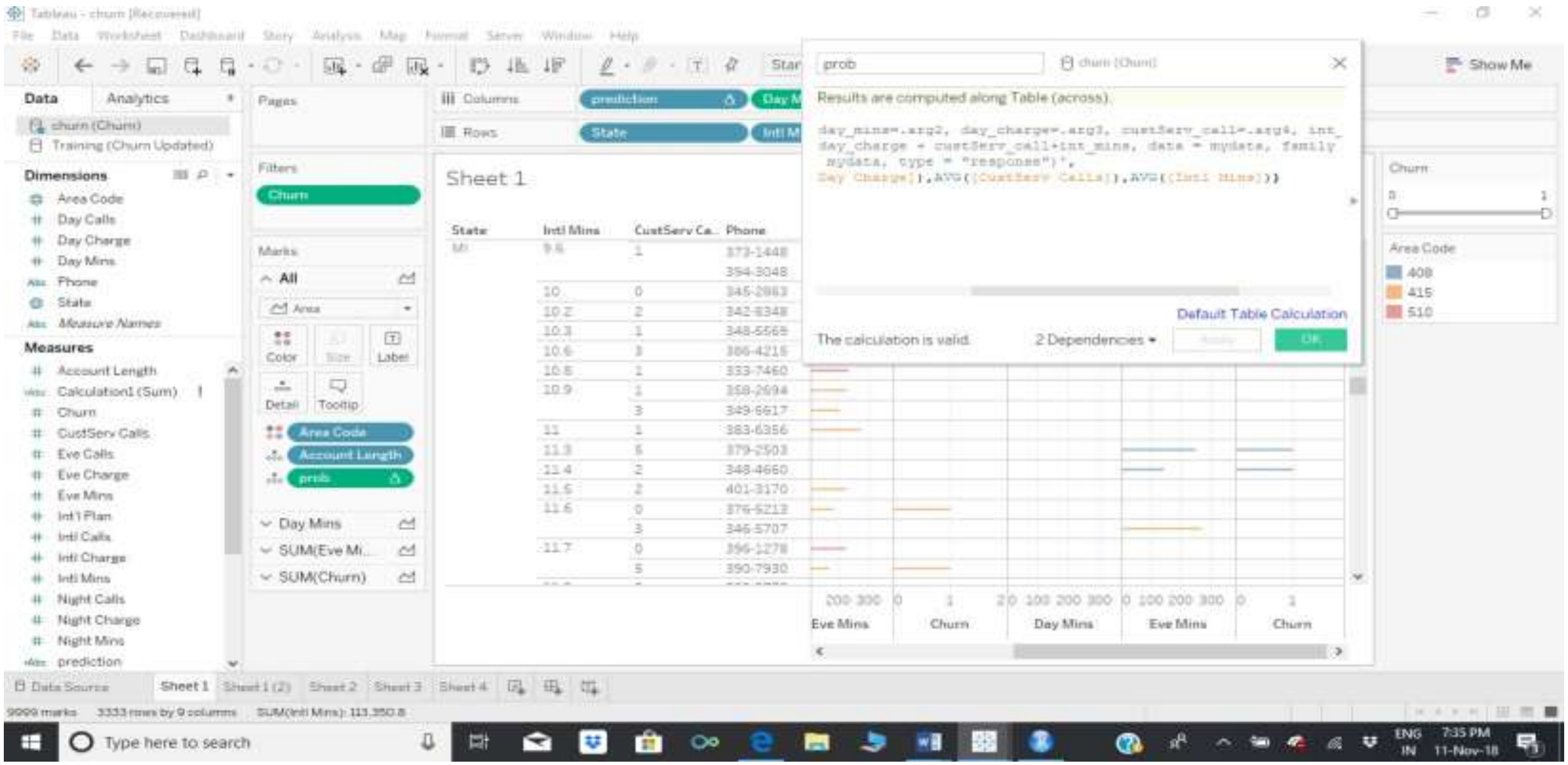


Tableau - churn (Recovered)
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
churn (Churn)
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Pages
Filters
Churn
Marks
All
Area
Color Size Label
Detail Tooltip
Area Code
Account Length
prob
Day Mins
SUM(Eve Mi...
SUM(Churn)
Columns
prediction
Rows
State
Sheet 1
State Intl Mins CustServ Ca.. Phone
prediction
Results are computed along Table (across).
intl_Mins <- .arg3
set.seed(2017)
data_churn <- data.frame(churn, Day_Mins, intl_Mins)
intrain<-sample(1:nrow(data_churn),.7*nrow(data_churn)
training<- data_churn[intrain,]
testing<- data_churn[-intrain,]
Default Table Calculation
The calculation is valid.
2 Dependencies
OK
Churn
Area Code
408
415
510
Eve Mins Churn Day Mins Eve Mins Churn
Data Source Sheet 1 Sheet 1 (2) Sheet 2 Sheet 3 Sheet 4
9999 marks 3333 rows by 9 columns SUM(Intl Mins): 113,350.8
Type here to search
7:34 PM
11-Nov-18


Show Me
Data
Analytics
Pages
Columns
prediction
Rows
State
prob
Results are computed along Table (across).
Default Table Calculation
The calculation is valid.
2 Dependencies
OK
Filters
Churn
Marks
All
Color
Size
Label
Detail
Tooltip
Area Code
Account Length
Day Mins
SUM(Churn)
Sheet 1
State
Intl Mins
Phone
Dimensions
Day Calls
Day Charge
Measures
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Night Calls
Night Charge
Night Mins
408
415
510
Data Source
Sheet 2
Sheet 3
Sheet 4
Type here to search

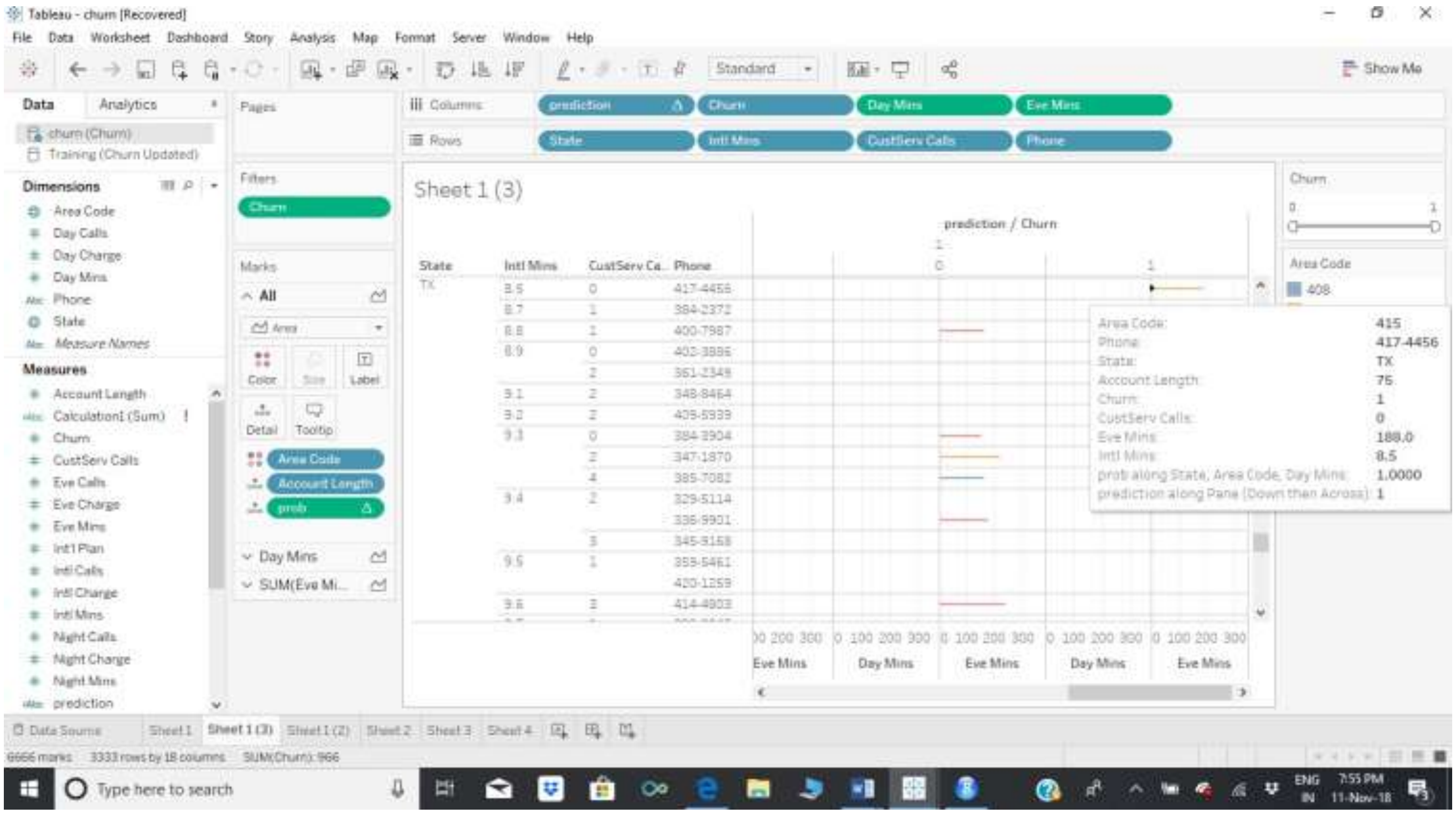
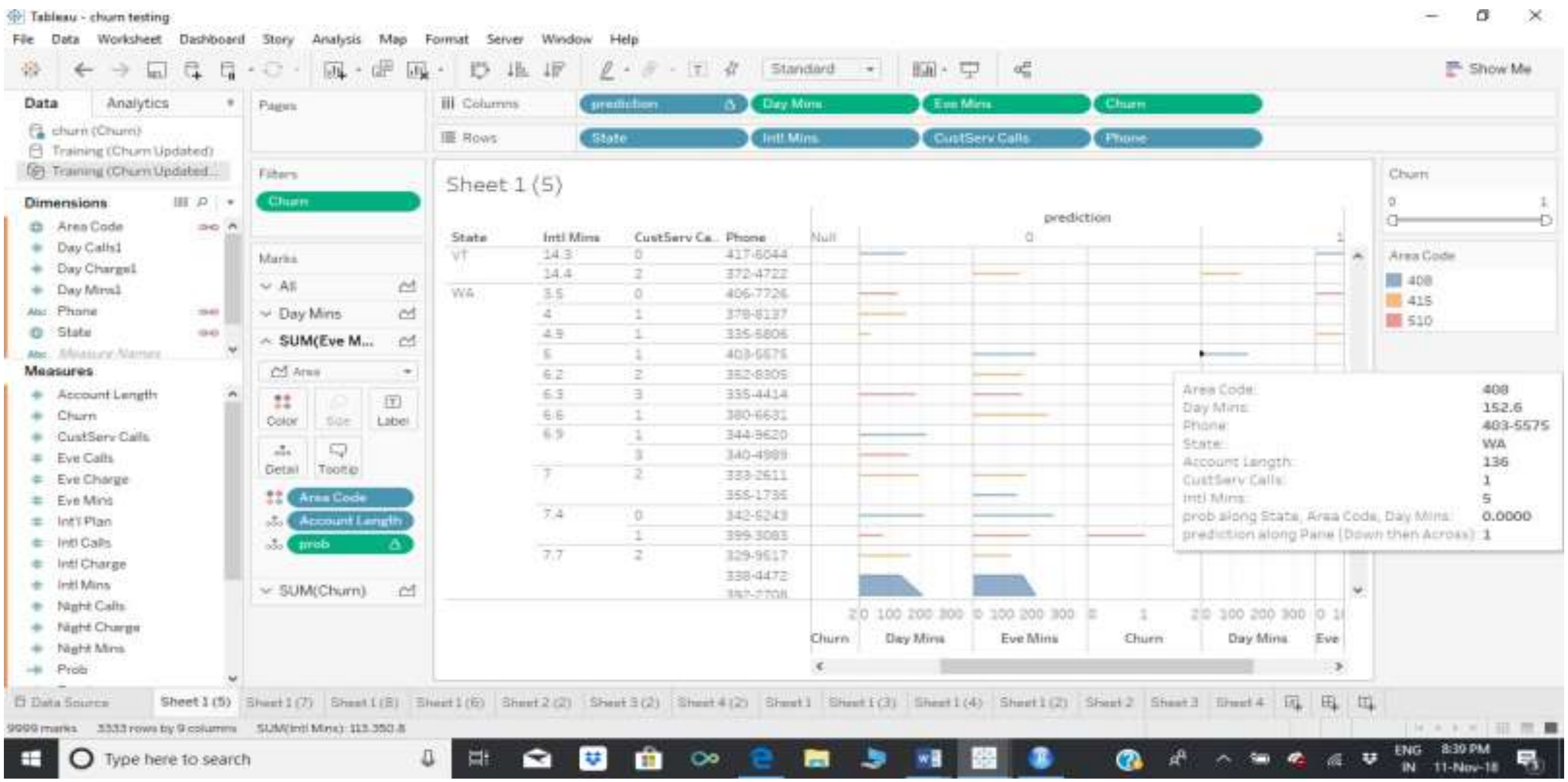
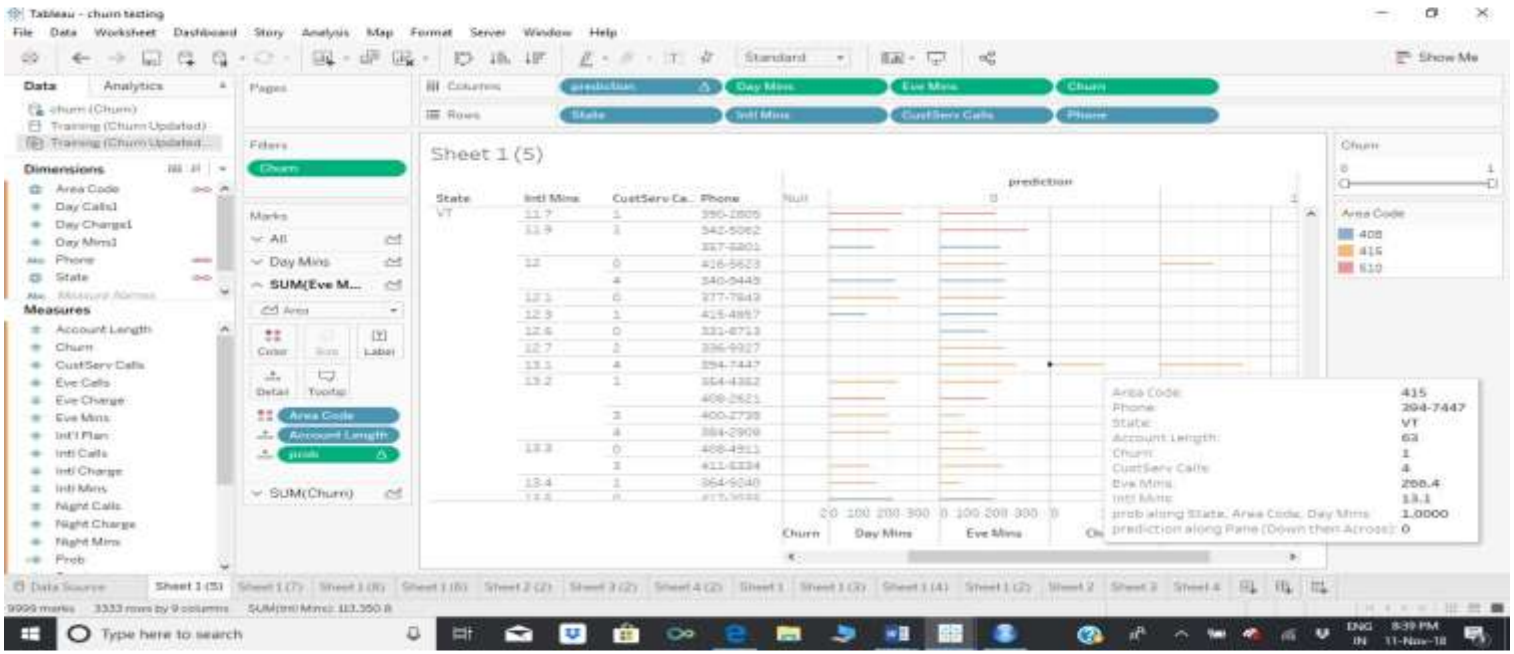
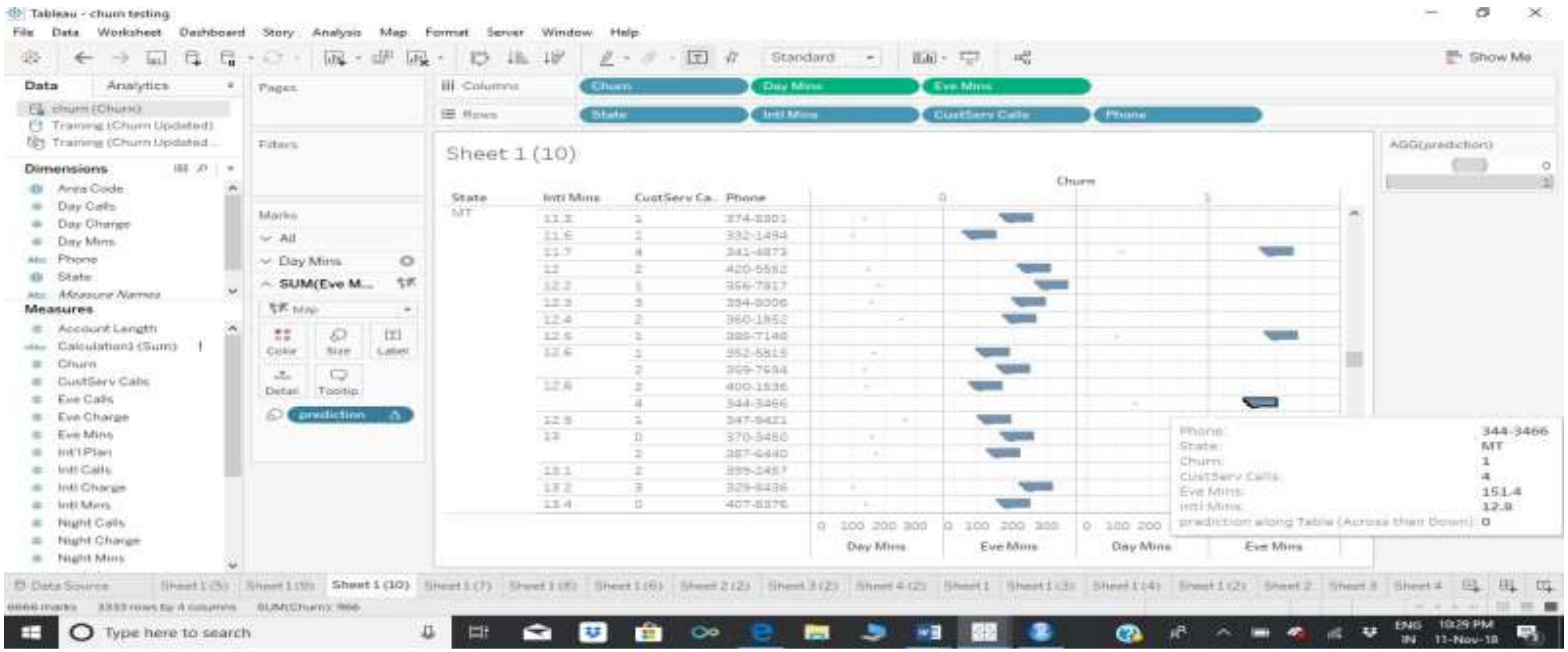


Tableau - churn [Recovered]
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
churn (Churn)
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Pages
Filters
Churn
Marks
All
Area
Color
Size
Label
Detail
Tooltip
Area Code
Account Length
prob
Day Mins
SUM(Eve Mi...
Columns
prediction
Churn
Day Mins
Eve Mins
Rows
State
Intl Mins
CustServ Calls
Phone
Sheet 1 (3)
prediction / Churn
State
Intl Mins
CustServ Ca..
Phone
TX
Eve Mins
Day Mins
Churn
Area Code
408
Area Code: 415
Phone: 417-4456
State: TX
Account Length: 75
Churn: 1
CustServ Calls: 0
Eve Mins: 188.0
Intl Mins: 8.5
prob along State, Area Code, Day Mins: 1.0000
prediction along Pane (Down then Across): 1
Data Source
Sheet 1
Sheet 1 (3)
Sheet 1 (2)
Sheet 2
Sheet 3
Sheet 4
6666 marks
3333 rows by 18 columns
SUM(Churn): 966
Type here to search
ENG
IN
7:55 PM
11-Nov-18







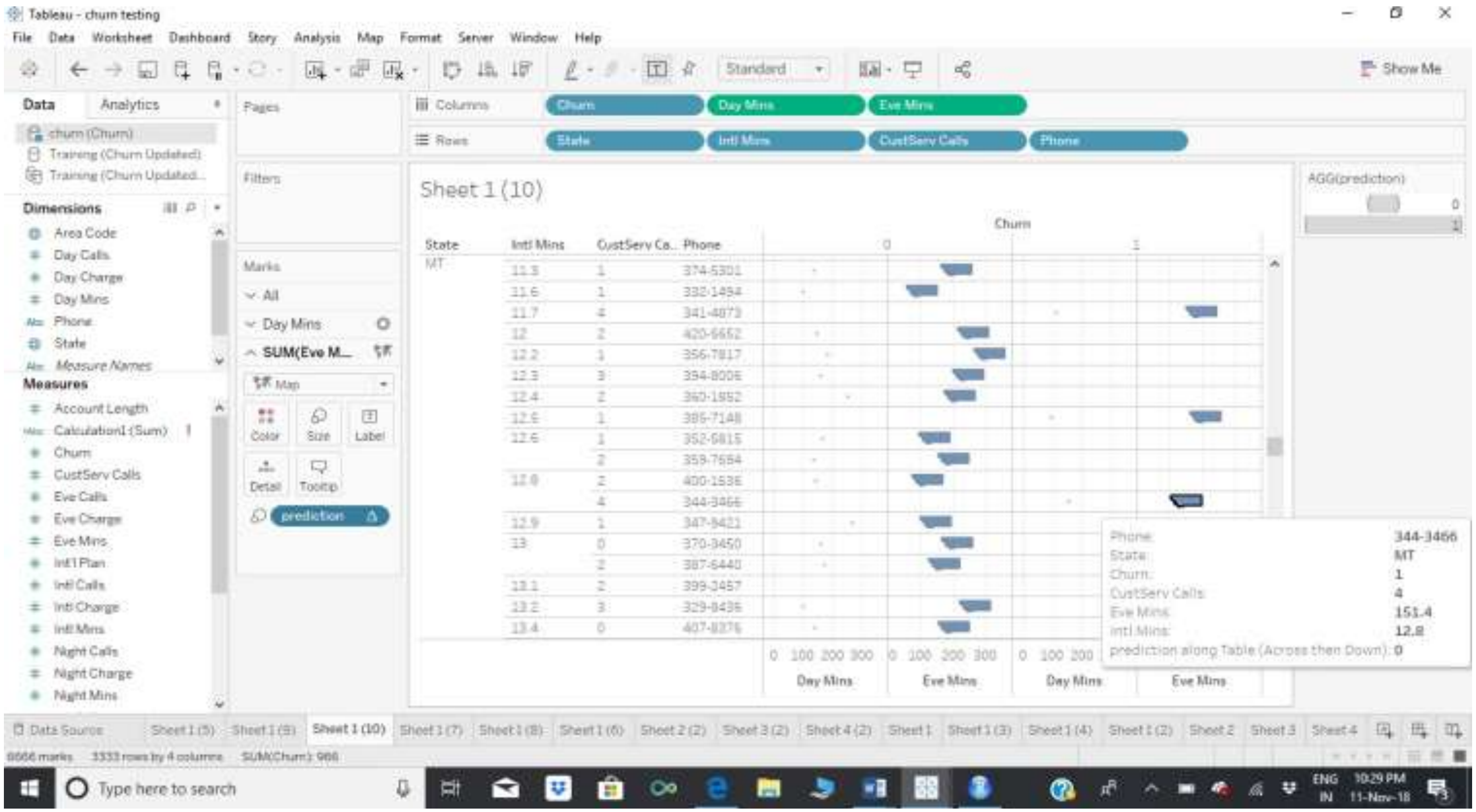


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Columns
Churn
Day Mins
Eve Mins
Rows
State
Intl Mins
CustServ Calls
Phone
Sheet 1 (10)
Churn
Phone: 344-3466
State: MT
Churn: 1
CustServ Calls: 4
Eve Mins: 151.4
Intl Mins: 12.8
prediction along Table (Across then Down): 0
Day Mins
Eve Mins

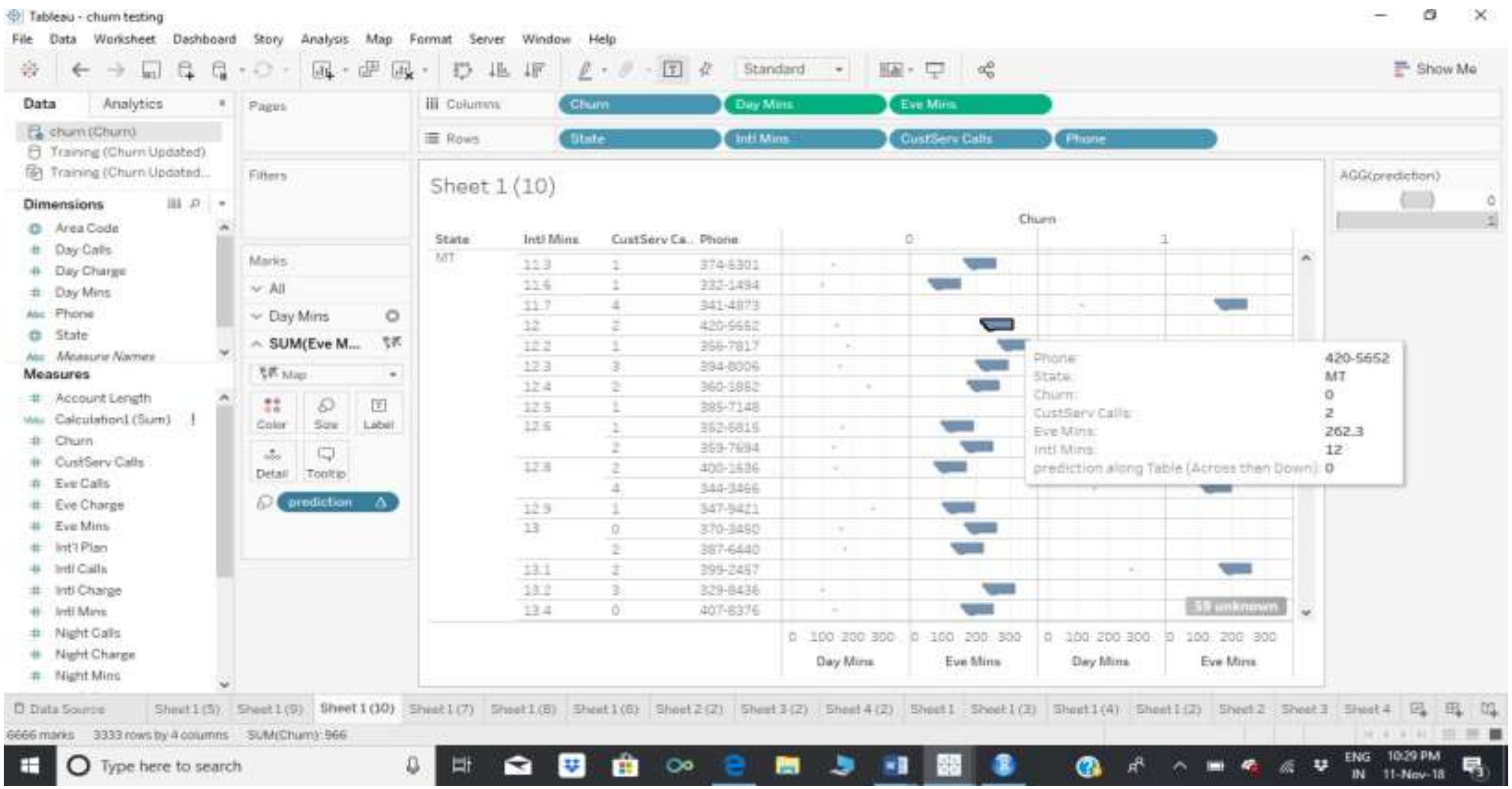


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
churn (Churn)
Training (Churn Updated)
Pages
Columns Churn Day Mins Eve Mins
Rows State Intl Mins CustServ Calls Phone
Filters
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Account Length
Calculation1 (Sum)
Churn
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
Marks
All
Day Mins
SUM(Eve M...
Map
Color Size Label
Detail Tooltip
prediction
Sheet 1 (10)
Churn
State Intl Mins CustServ Ca.. Phone
MT
11.3 1 374-5301
11.6 1 332-1494
11.7 4 341-4873
12 2 420-5652
12.2 1 356-7817
12.3 3 394-8006
12.4 2 360-1852
12.5 1 385-7148
12.6 1 352-6815
2 359-7694
12.8 2 400-1636
4 344-3466
12.9 1 347-9421
13 0 370-3450
2 387-6440
13.1 2 399-2457
13.2 3 329-8436
13.4 0 407-8376
Phone: 420-5652
State: MT
Churn: 0
CustServ Calls: 2
Eve Mins: 262.3
Intl Mins: 12
prediction along Table (Across then Down): 0
59 unknown
0 100 200 300
Day Mins
Eve Mins
AGG(prediction)
Data Source
Sheet 1 (5) Sheet 1 (9) Sheet 1 (10) Sheet 1 (7) Sheet 1 (8) Sheet 1 (6) Sheet 2 (2) Sheet 3 (2) Sheet 4 (2) Sheet 1 Sheet 1 (3) Sheet 1 (4) Sheet 1 (2) Sheet 2 Sheet 3 Sheet 4
6666 marks 3333 rows by 4 columns SUM(Churn): 966
Type here to search
ENG IN 10:29 PM 11-Nov-18

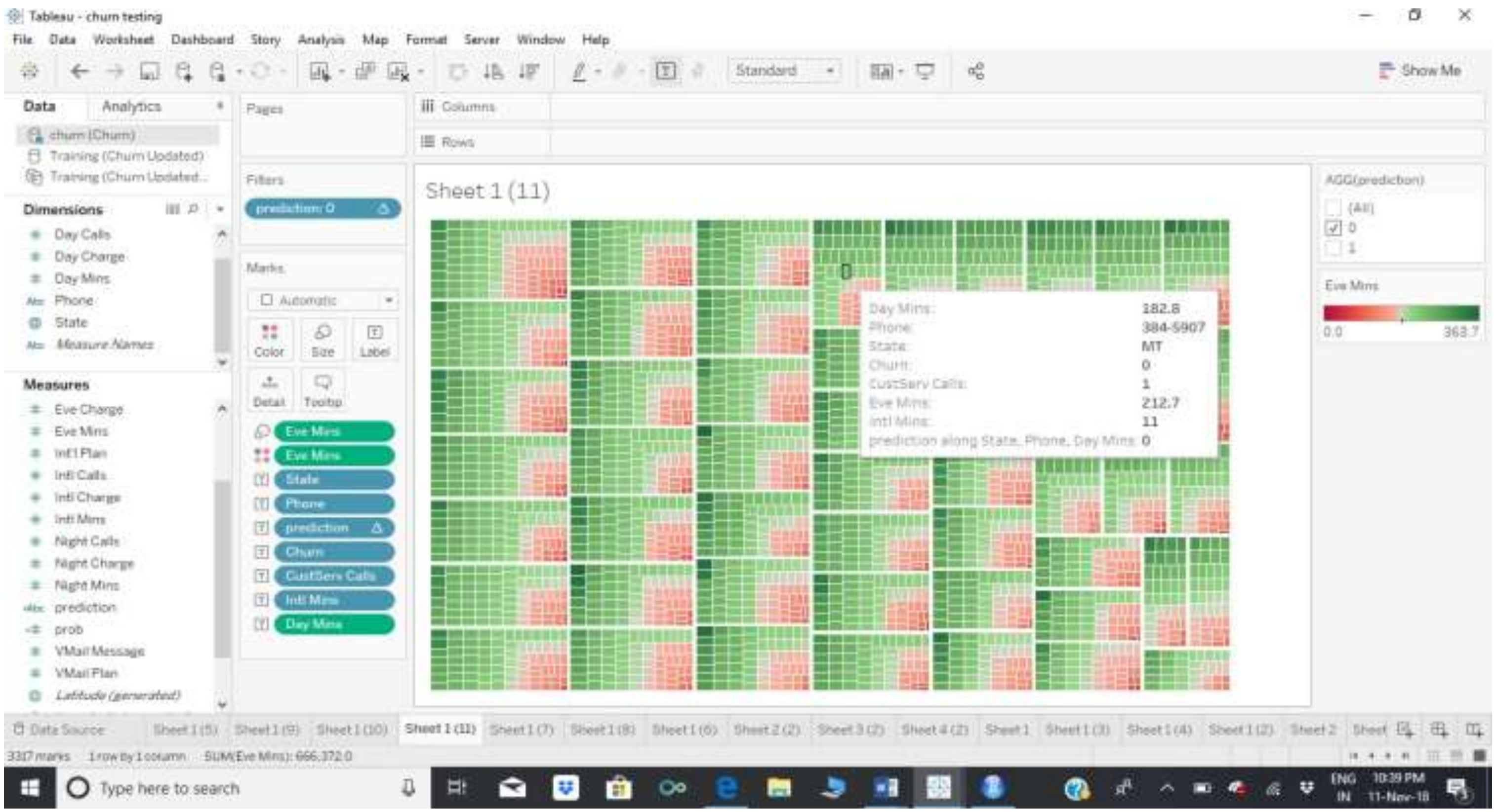
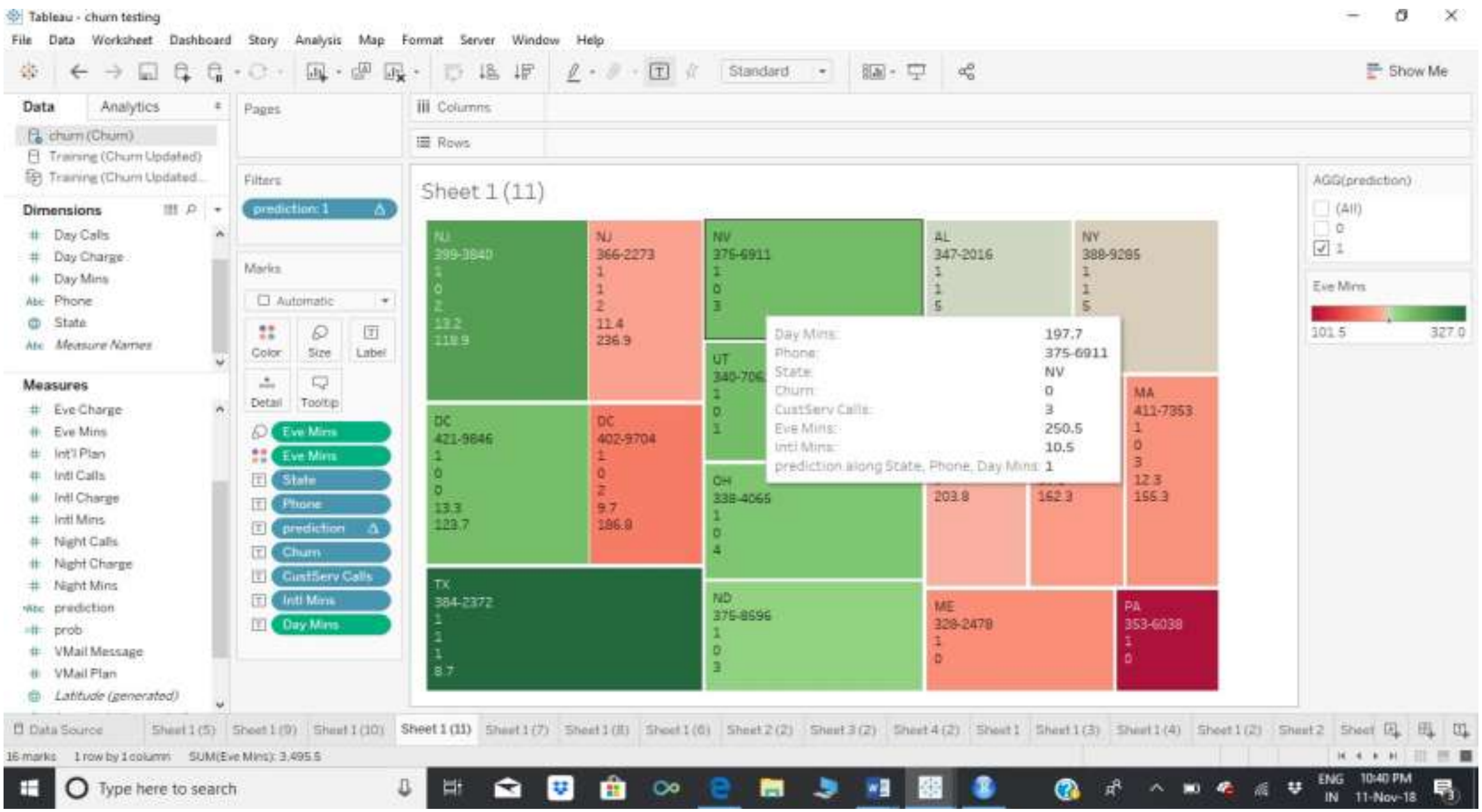


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data Analytics
Pages
Columns
Rows
Filters
prediction: 0
Marks
Automatic
Color Size Label
Detail Tooltip
Sheet 1 (11)
AGG(prediction)
(All)
0
1
Eve Mins
0.0
363.7
Dimensions
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Eve Charge
Eve Mins
Intl Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
prob
VMail Message
VMail Plan
Latitude (generated)
Eve Mins
Eve Mins
State
Phone
prediction
Churn
CustServ Calls
Intl Mins
Day Mins
Day Mins: 182.8
Phone: 384-5907
State: MT
Churn: 0
CustServ Calls: 1
Eve Mins: 212.7
Intl Mins: 11
prediction along State, Phone, Day Mins: 0
Data Source
Sheet 1 (11)
Type here to search



Prediction in slider

we can select 0,1 or All

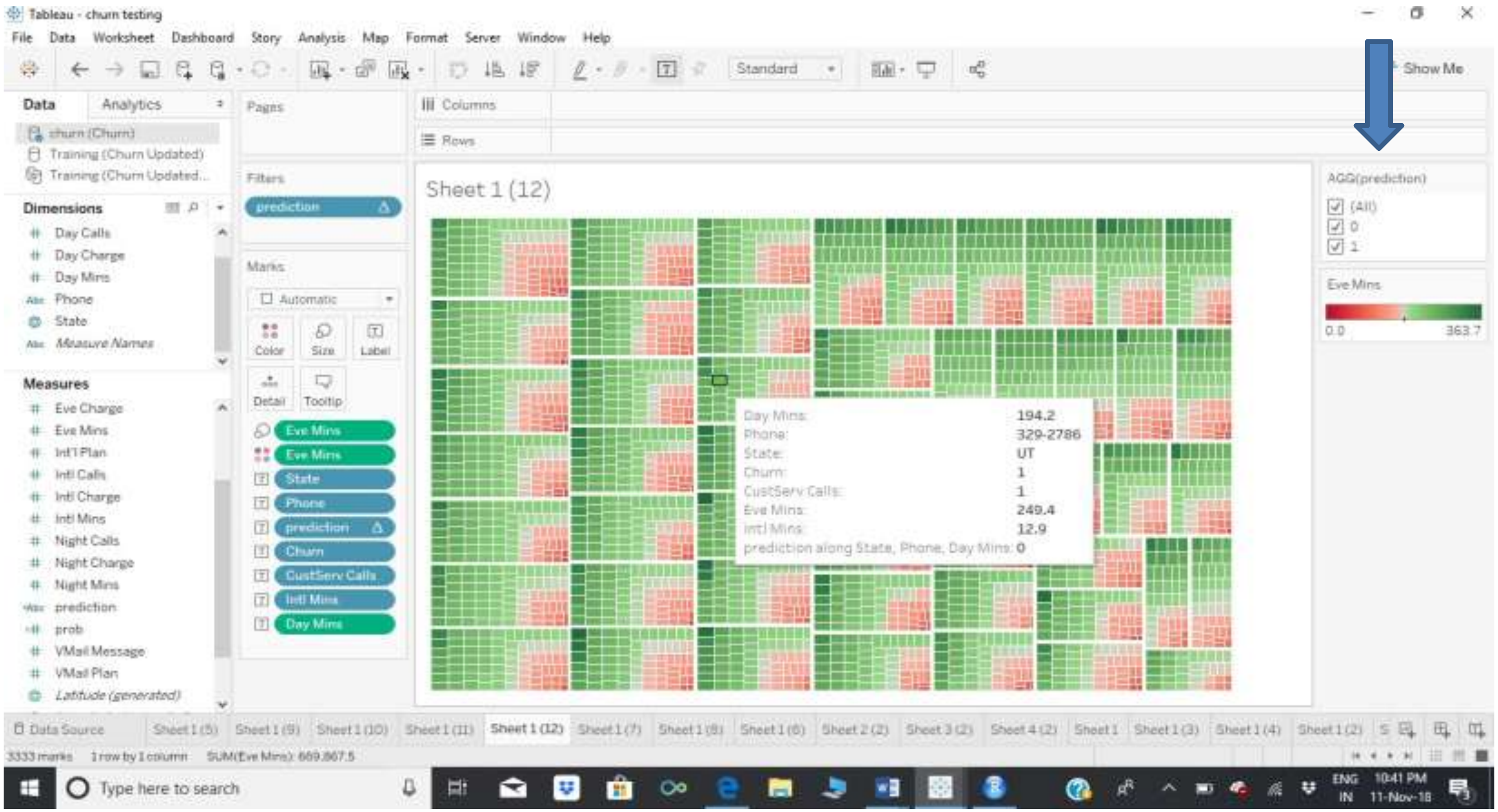


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data Analytics
churn (Churn)
Training (Churn Updated)
Pages
Columns
Rows
Filters
prediction
Marks
Automatic
Color Size Label
Detail Tooltip
Eve Mins
Eve Mins
State
Phone
prediction
Churn
CustServ Calls
Intl Mins
Day Mins
Dimensions
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
prob
VMail Message
VMail Plan
Latitude (generated)
Sheet 1 (12)
Day Mins: 194.2
Phone: 329-2786
State: UT
Churn: 1
CustServ Calls: 1
Eve Mins: 249.4
Intl Mins: 12.9
prediction along State, Phone, Day Mins: 0
AGG(prediction)
(All)
0
1
Eve Mins
0.0
363.7
Data Source
3333 marks 1 row by 1 column SUM(Eve Mins): 669,867.5
Type here to search
ENG IN
10:41 PM 11-Nov-18

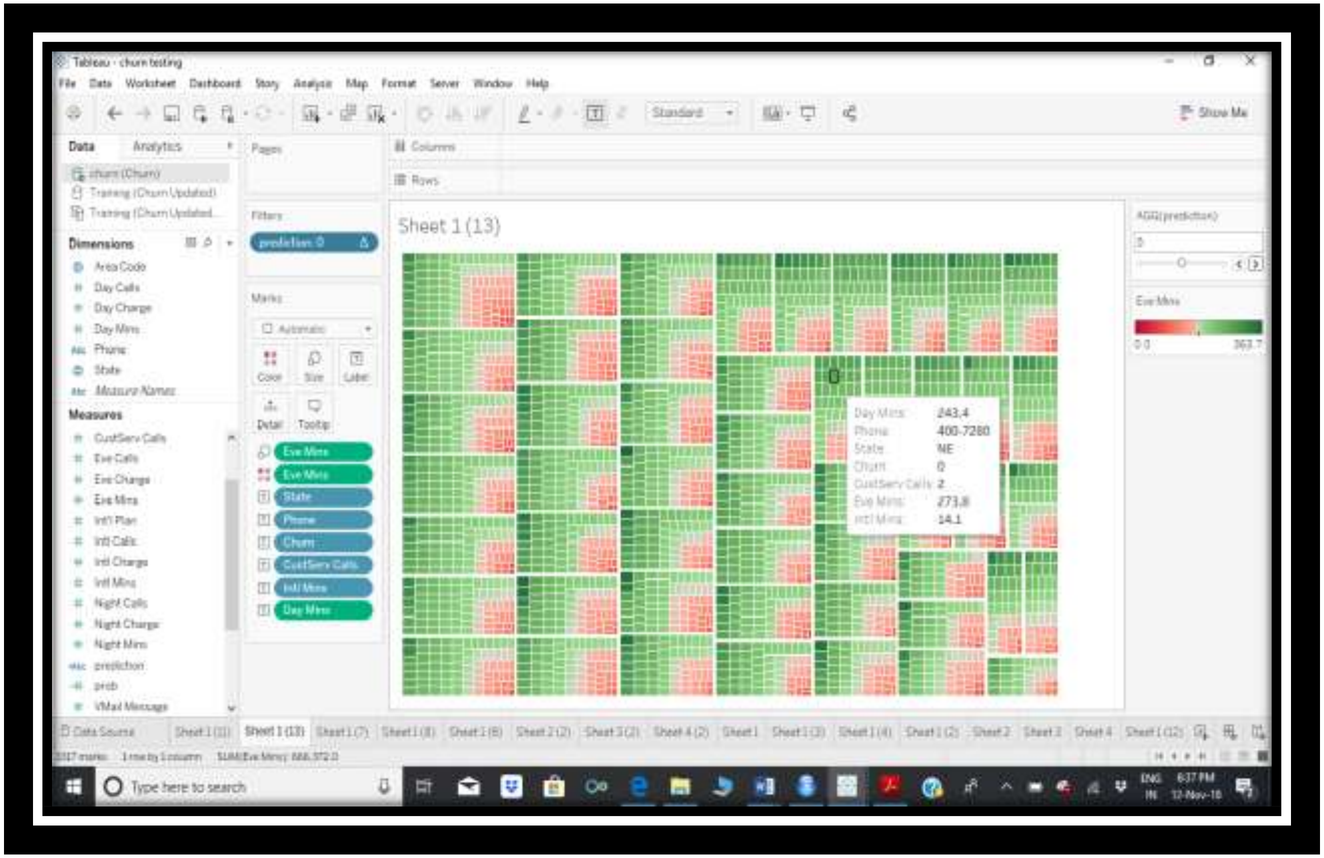


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
Pages
Columns
Rows
Filters
Marks
Dimensions
Measures
Sheet 1 (13)
Day Mins: 243.4
Phone: 400-7280
State: NE
Churn: 0
CustServ Calls: 2
Eve Mins: 273.8
Intl Mins: 14.1
Type here to search

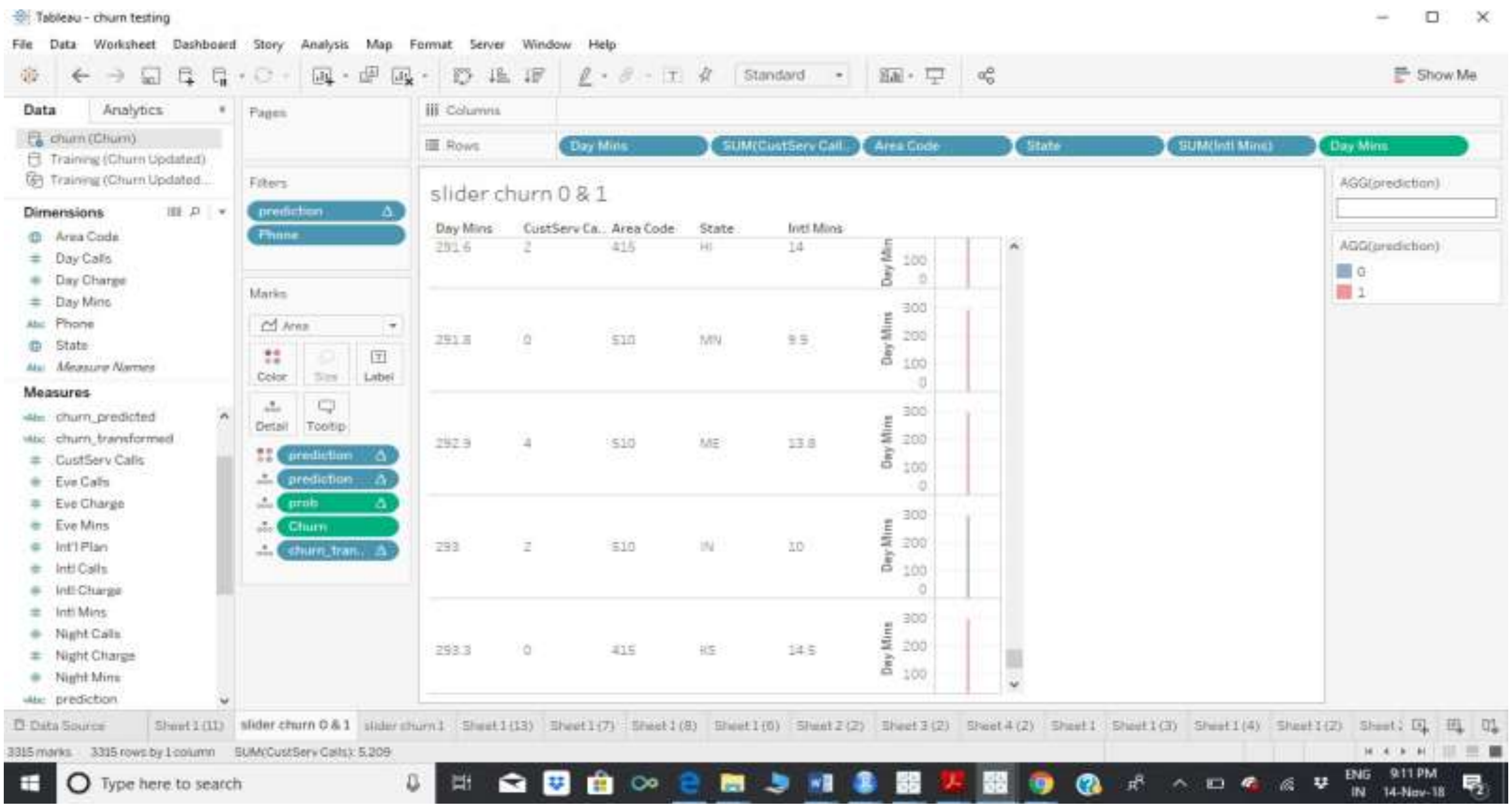
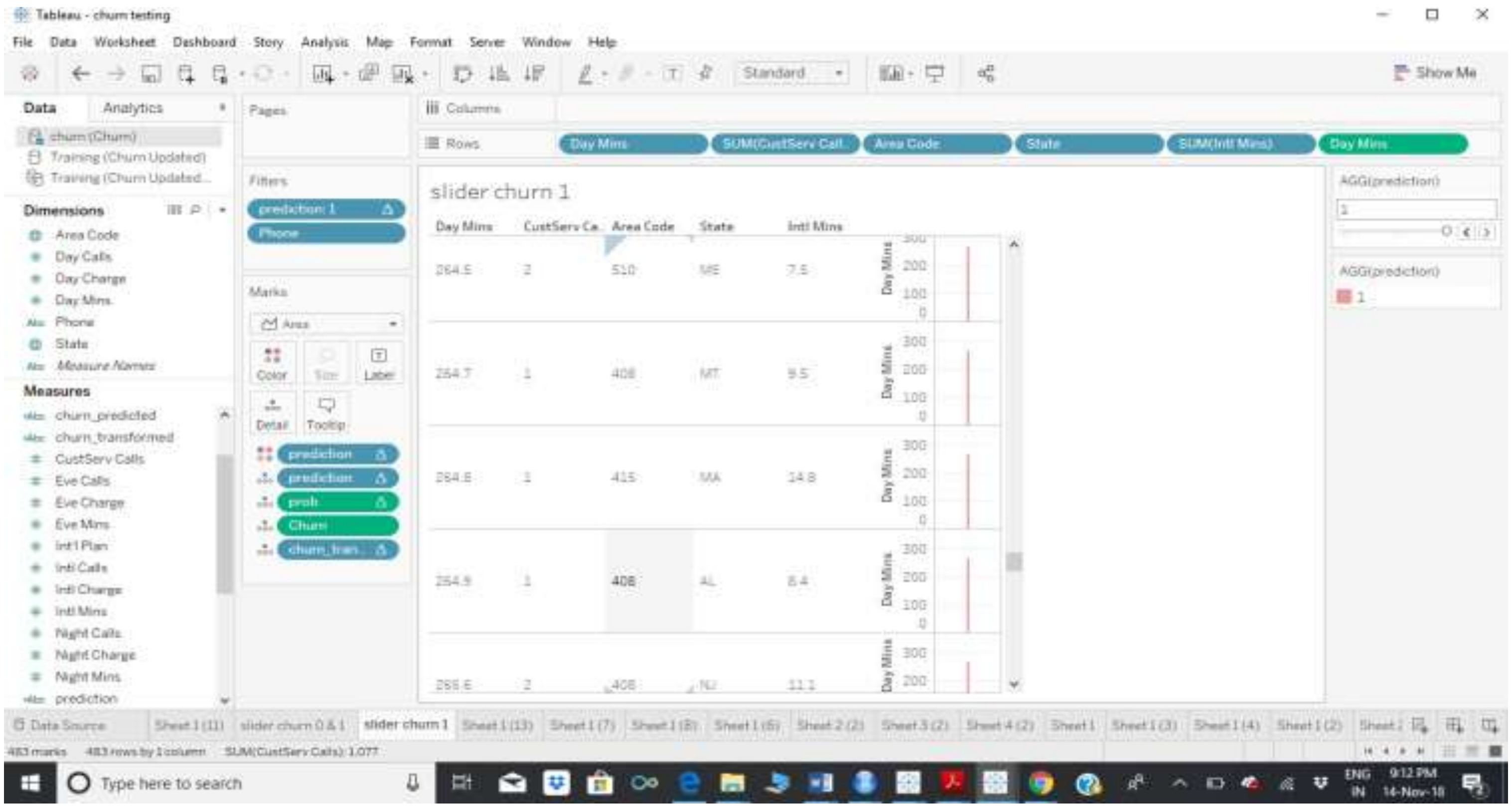


Tableau - churn testing
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Data
Analytics
churn (Churn)
Training (Churn Updated)
Dimensions
Area Code
Day Calls
Day Charge
Day Mins
Phone
State
Measure Names
Measures
churn_predicted
churn_transformed
CustServ Calls
Eve Calls
Eve Charge
Eve Mins
Int'l Plan
Intl Calls
Intl Charge
Intl Mins
Night Calls
Night Charge
Night Mins
prediction
Pages
Filters
prediction
Phone
Marks
Color
Label
Detail
Tooltip
prob
Churn
Columns
Rows
Day Mins
Area Code
State
SUM(Intl Mins)
slider churn 0 & 1
Day Mins
Area Code
State
Intl Mins
AGG(prediction)
slider churn 0 & 1
3315 marks
3315 rows by 1 column
Type here to search
9:11 PM
14-Nov-18



churn prediction.csv

churn all data.csv

churn 0 data.csv

All churn1 data.csv

In this project a customer churn analysis was presented for churn data provided in the project 2. The analysis focused on churn prediction based on logistic regression. The different models predicted the actual churners relatively well.

The differences between the models input data (the significance level in case of each of the variables) indicates the dynamic nature of the churning customer profile. This makes it hard to formulate one standard model that could be used as the predictive model in the future. The findings of this study indicate that, in case of logistic regression model, the user should update the model to be able to produce predictions with high accuracy. It is interesting for a company's perspective whether the churning customers are worth retaining or not. And also in marketing perspective what can be done to retain them. For Visualization Tableau desktop is used. Calculated fields is used to calculate the probability and the fitted results prediction. Using the Library Reserve in R Logistic regression model is connected with Tableau.

Through this model we can identify easily the churn 1 or 0 and the probabilities through a slider to visualize the churn 1, churn 0 and all. The visualization in Tableau provides separately state wise, Area code, specific telephone no and their probability to predict and visualize them nicely. The data can be imported and stored as excel data for further analysis and churn predictions. These files are exported and stored and attached in this project file.

Acknowledgement

This is a quite interesting project and I have gained a lot of knowledge. I thank the institute Acadgild and the Mentors Dr.Vinod and Miss Pooja who taught us the R coding and other subjects to understand the Algorithms. I thank the support coordinator for guiding me to understand the project related queries and complete the project on time.

Thank you Acadgild.