Telecommunication Customer Churn Analysis

## Authors

Ali El-Sharif, Neha Panchal, Sagnik Adusumilli, Sarmad Shubber

## Background

Nowadays, the telecom industry faces fierce competition in satisfying its customers. When a business loses customers, its profit and growth rate may be negatively affected. The churn rate of a company is percentage of subscribers to a service who discontinue their subscriptions within a given time period.

## Objective

We will create a system that groups customers by their traits to better serve their individual needs.

## Data Analysis

The data has been sourced from Kaggle <https://www.kaggle.com/jpacse/datasets-for-churn-telecom> The dataset is preprocessed and a balanced consisting of 71,047 instances and 58 attributes.

## Data Dictionary

|  |  |
| --- | --- |
| Column Name | Column Description |
| CustomerID | Unique customer ID |
| Churn | Customer cut ties with service during a given time period (Yes/No) |
| MonthlyRevenue | Monthly revenue |
| MonthlyMinutes | Mean monthly minute of use |
| TotalRecurringCharge | Mean total recurring charge |
| DirectorAssistedCalls | Mean number of director assisted calls |
| OverageMinutes | Mean overage minute of use |
| RoamingCalls | Number of call were made and received outside of home area network operator |
| PercChangeMinutes | % change in minute of use |
| PercChangeRevenues | % change in revenues |
| DroppedCalls | Number of telephone calls which, due to technical reasons were cut off |
| BlockedCalls | Number of incoming subscriber’s blocked calls from specific telephone numbers |
| UnansweredCalls | Number of calls which are not picked up by an agent |
| CustomerCareCalls | Number of calls made by customer care to the subscriber. |
| ThreewayCalls | Number of conference call with two other two callers with local numbers |
| ReceivedCalls | Number of incoming calls that were answered |
| OutboundCalls | Number of call initiated by a call center agent to a customer |
| InboundCalls | Number of call initiated by a customer to call center agent |
| PeakCallsInOut | Mean of peak in and out voice calls |
| OffPeakCallsInOut | Mean number of in and out off peak voice calls |
| DroppedBlockedCalls | Mean number of dropped blocked calls |
| CallForwardingCalls | Number of redirected calls to another available telephone numbers |
| CallWaitingCalls | Number of suspended calls already in progress to accept a second call |
| MonthsInService | Months in Service |
| UniqueSubs | Number of unique subs |
| ActiveSubs | Number of active subs |
| ServiceArea | Area covered by the signal of a broadcasting station |
| Handsets | Outer part of mobile phone that does not include the battery or the sim card |
| HandsetModels | Models of mobile phones |
| CurrentEquipmentDays | Number of days of the current equipment |
| AgeHH1 | Age of first HH member |
| AgeHH2 | Age of second HH member |
| ChildrenInHH | Presence of children in HH (Yes/No) |
| HandsetRefurbished | Pre-owned handset sent back due to a fault and has been repaired for re-sale (Yes/No) |
| HandsetWebCapable | Handset is web capable (Yes/No) |
| TruckOwner | Subscriber owns a Truck (Yes/No) |
| RVOwner | Subscriber owns a recreational vehicle (Yes/No) |
| Homeownership | Subscriber owns a home (Yes/No) |
| BuysViaMailOrder | Buys via mail order (Yes/No) |
| RespondsToMailOffers | Number of time Subscriber responds to a mail offers (Yes/No) |
| OptOutMailings | Chosen by not to be solicited by Mails (Yes/No) |
| NonUSTravel | Rravelled to non-US country (Yes/No) |
| OwnsComputer | Owns a personal computer (Yes/No) |
| HasCreditCard | Has a credit card (Yes/No) |
| RetentionCalls | Customer made call to retention team |
| RetentionOffersAccepted | Number of previous retention offers accepted |
| NewCellphoneUser | Known to be a new cell phone user (Yes/No) |
| NotNewCellphoneUser | Known not to be a new cell phone user (Yes/No) |
| ReferralsMadeBySubscriber | Number of referals made by subscriber |
| IncomeGroup | Subscriber income group |
| OwnsMotorcycle | Subscriber owns a motor cycle (Yes/No) |
| AdjustmentsToCreditRating | Number of adjustments were made to customer’s credit rating (Up or Down) |
| HandsetPrice | Price of outer part of mobile phone not including battery or sim card |
| MadeCallToRetentionTeam | Number of calls were made to retention team (Yes/No) |
| CreditRating | Credit Rating |
| PrizmCode | PRIZM (Potential Rating Index for Zip Market) - census data market segmentation |
| Occupation | Occupation |
| MaritalStatus | Marital status |

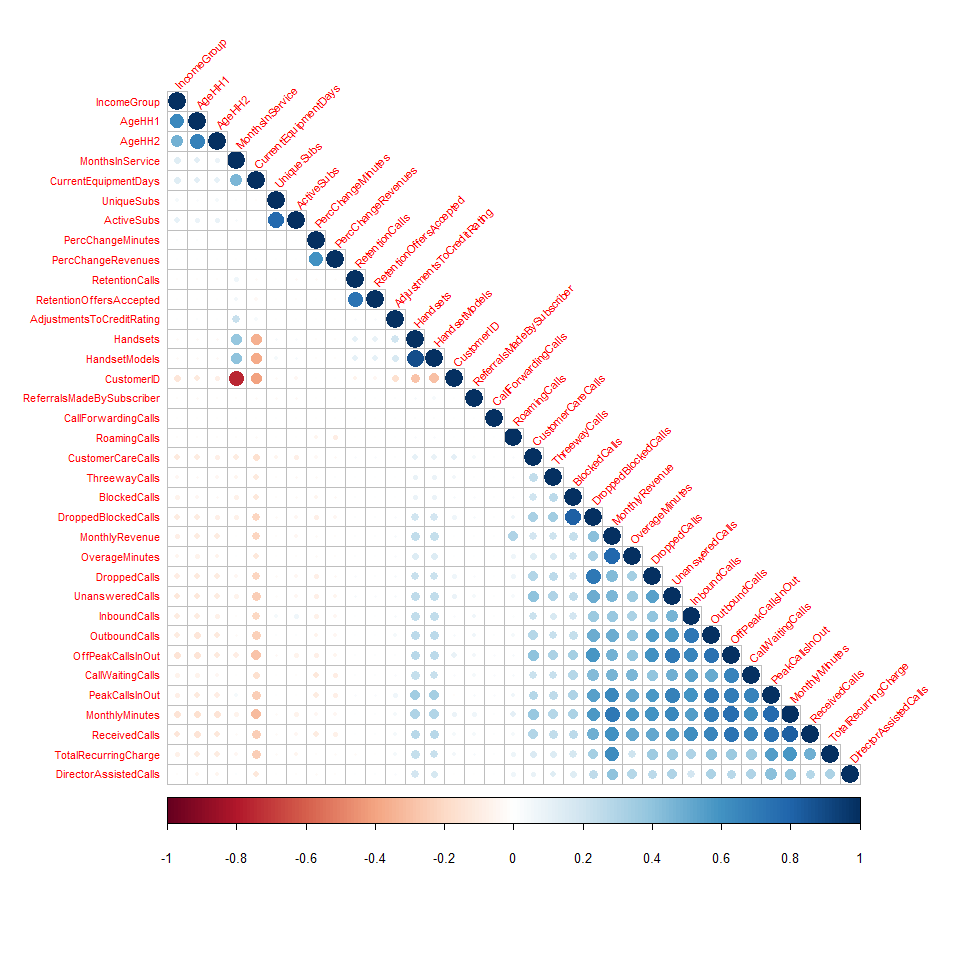
## Data Exploartion

Summary:

## CustomerID Churn MonthlyRevenue MonthlyMinutes   
## Min. :3000002 No :36336 Min. : -6.17 Min. : 0.0   
## 1st Qu.:3100632 Yes:14711 1st Qu.: 33.61 1st Qu.: 158.0   
## Median :3201534 Median : 48.46 Median : 366.0   
## Mean :3201957 Mean : 58.83 Mean : 525.7   
## 3rd Qu.:3305376 3rd Qu.: 71.06 3rd Qu.: 723.0   
## Max. :3399994 Max. :1223.38 Max. :7359.0   
## NA's :156 NA's :156   
## TotalRecurringCharge DirectorAssistedCalls OverageMinutes   
## Min. :-11.00 Min. : 0.0000 Min. : 0.00   
## 1st Qu.: 30.00 1st Qu.: 0.0000 1st Qu.: 0.00   
## Median : 45.00 Median : 0.2500 Median : 3.00   
## Mean : 46.83 Mean : 0.8952 Mean : 40.03   
## 3rd Qu.: 60.00 3rd Qu.: 0.9900 3rd Qu.: 41.00   
## Max. :400.00 Max. :159.3900 Max. :4321.00   
## NA's :156 NA's :156 NA's :156   
## RoamingCalls PercChangeMinutes PercChangeRevenues   
## Min. : 0.000 Min. :-3875.00 Min. :-1107.700   
## 1st Qu.: 0.000 1st Qu.: -83.00 1st Qu.: -7.100   
## Median : 0.000 Median : -5.00 Median : -0.300   
## Mean : 1.236 Mean : -11.55 Mean : -1.192   
## 3rd Qu.: 0.300 3rd Qu.: 66.00 3rd Qu.: 1.600   
## Max. :1112.400 Max. : 5192.00 Max. : 2483.500   
## NA's :156 NA's :367 NA's :367   
## DroppedCalls BlockedCalls UnansweredCalls CustomerCareCalls  
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.700 1st Qu.: 0.000 1st Qu.: 5.30 1st Qu.: 0.000   
## Median : 3.000 Median : 1.000 Median : 16.30 Median : 0.000   
## Mean : 6.011 Mean : 4.086 Mean : 28.29 Mean : 1.869   
## 3rd Qu.: 7.700 3rd Qu.: 3.700 3rd Qu.: 36.30 3rd Qu.: 1.700   
## Max. :221.700 Max. :384.300 Max. :848.70 Max. :327.300   
##   
## ThreewayCalls ReceivedCalls OutboundCalls InboundCalls   
## Min. : 0.0000 Min. : 0.0 Min. : 0.00 Min. : 0.000   
## 1st Qu.: 0.0000 1st Qu.: 8.3 1st Qu.: 3.30 1st Qu.: 0.000   
## Median : 0.0000 Median : 52.8 Median : 13.70 Median : 2.000   
## Mean : 0.2988 Mean : 114.8 Mean : 25.38 Mean : 8.178   
## 3rd Qu.: 0.3000 3rd Qu.: 153.5 3rd Qu.: 34.00 3rd Qu.: 9.300   
## Max. :66.0000 Max. :2692.4 Max. :644.30 Max. :519.300   
##   
## PeakCallsInOut OffPeakCallsInOut DroppedBlockedCalls  
## Min. : 0.00 Min. : 0.00 Min. : 0.00   
## 1st Qu.: 23.00 1st Qu.: 11.00 1st Qu.: 1.70   
## Median : 62.00 Median : 35.70 Median : 5.30   
## Mean : 90.55 Mean : 67.65 Mean : 10.16   
## 3rd Qu.: 121.30 3rd Qu.: 88.70 3rd Qu.: 12.30   
## Max. :2090.70 Max. :1474.70 Max. :411.70   
##   
## CallForwardingCalls CallWaitingCalls MonthsInService UniqueSubs   
## Min. : 0.00000 Min. : 0.000 Min. : 6.00 Min. : 1.000   
## 1st Qu.: 0.00000 1st Qu.: 0.000 1st Qu.:11.00 1st Qu.: 1.000   
## Median : 0.00000 Median : 0.300 Median :16.00 Median : 1.000   
## Mean : 0.01228 Mean : 1.841 Mean :18.76 Mean : 1.532   
## 3rd Qu.: 0.00000 3rd Qu.: 1.300 3rd Qu.:24.00 3rd Qu.: 2.000   
## Max. :81.30000 Max. :212.700 Max. :61.00 Max. :196.000   
##   
## ActiveSubs ServiceArea Handsets HandsetModels   
## Min. : 0.000 NYCBRO917: 1684 Min. : 1.000 Min. : 1.000   
## 1st Qu.: 1.000 HOUHOU281: 1510 1st Qu.: 1.000 1st Qu.: 1.000   
## Median : 1.000 DALDAL214: 1498 Median : 1.000 Median : 1.000   
## Mean : 1.354 NYCMAN917: 1182 Mean : 1.806 Mean : 1.559   
## 3rd Qu.: 2.000 APCFCH703: 783 3rd Qu.: 2.000 3rd Qu.: 2.000   
## Max. :53.000 (Other) :44366 Max. :24.000 Max. :15.000   
## NA's : 24 NA's :1 NA's :1   
## CurrentEquipmentDays AgeHH1 AgeHH2 ChildrenInHH  
## Min. : -5.0 Min. : 0.00 Min. : 0.00 No :38679   
## 1st Qu.: 205.0 1st Qu.: 0.00 1st Qu.: 0.00 Yes:12368   
## Median : 329.0 Median :36.00 Median : 0.00   
## Mean : 380.5 Mean :31.34 Mean :21.14   
## 3rd Qu.: 515.0 3rd Qu.:48.00 3rd Qu.:42.00   
## Max. :1812.0 Max. :99.00 Max. :99.00   
## NA's :1 NA's :909 NA's :909   
## HandsetRefurbished HandsetWebCapable TruckOwner RVOwner   
## No :43956 No : 5001 No :41524 No :46894   
## Yes: 7091 Yes:46046 Yes: 9523 Yes: 4153   
##   
##   
##   
##   
##   
## Homeownership BuysViaMailOrder RespondsToMailOffers OptOutMailings  
## Known :33987 No :32596 No :31821 No :50295   
## Unknown:17060 Yes:18451 Yes:19226 Yes: 752   
##   
##   
##   
##   
##   
## NonUSTravel OwnsComputer HasCreditCard RetentionCalls   
## No :48168 No :41583 No :16544 Min. :0.0000   
## Yes: 2879 Yes: 9464 Yes:34503 1st Qu.:0.0000   
## Median :0.0000   
## Mean :0.0372   
## 3rd Qu.:0.0000   
## Max. :4.0000   
##   
## RetentionOffersAccepted NewCellphoneUser NotNewCellphoneUser  
## Min. :0.00000 No :41223 No :44012   
## 1st Qu.:0.00000 Yes: 9824 Yes: 7035   
## Median :0.00000   
## Mean :0.01828   
## 3rd Qu.:0.00000   
## Max. :3.00000   
##   
## ReferralsMadeBySubscriber IncomeGroup OwnsMotorcycle  
## Min. : 0.00000 Min. :0.000 No :50368   
## 1st Qu.: 0.00000 1st Qu.:0.000 Yes: 679   
## Median : 0.00000 Median :5.000   
## Mean : 0.05207 Mean :4.325   
## 3rd Qu.: 0.00000 3rd Qu.:7.000   
## Max. :35.00000 Max. :9.000   
##   
## AdjustmentsToCreditRating HandsetPrice MadeCallToRetentionTeam  
## Min. : 0.00000 Unknown:28982 No :49302   
## 1st Qu.: 0.00000 30 : 7328 Yes: 1745   
## Median : 0.00000 150 : 4115   
## Mean : 0.05391 130 : 2105   
## 3rd Qu.: 0.00000 80 : 1960   
## Max. :25.00000 10 : 1928   
## (Other): 4629   
## CreditRating PrizmCode Occupation MaritalStatus   
## 1-Highest: 8522 Other :24655 Other :37637 No :12696   
## 2-High :18993 Rural : 2425 Professional: 8755 Unknown:19700   
## 3-Good : 8410 Suburban:16378 Crafts : 1519 Yes :18651   
## 4-Medium : 5357 Town : 7589 Clerical : 986   
## 5-Low : 6499 Self : 879   
## 6-VeryLow: 1152 Retired : 733   
## 7-Lowest : 2114 (Other) : 538

Structure:

## 'data.frame': 51047 obs. of 58 variables:  
## $ CustomerID : int 3000002 3000010 3000014 3000022 3000026 3000030 3000038 3000042 3000046 3000050 ...  
## $ Churn : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 1 1 1 1 1 ...  
## $ MonthlyRevenue : num 24 17 38 82.3 17.1 ...  
## $ MonthlyMinutes : int 219 10 8 1312 0 682 26 98 24 1056 ...  
## $ TotalRecurringCharge : int 22 17 38 75 17 52 30 66 35 75 ...  
## $ DirectorAssistedCalls : num 0.25 0 0 1.24 0 0.25 0.25 2.48 0 0 ...  
## $ OverageMinutes : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ RoamingCalls : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ PercChangeMinutes : int -157 -4 -2 157 0 148 60 24 20 43 ...  
## $ PercChangeRevenues : num -19 0 0 8.1 -0.2 -3.1 4 6.8 -0.3 2.4 ...  
## $ DroppedCalls : num 0.7 0.3 0 52 0 9 0 0 0 0 ...  
## $ BlockedCalls : num 0.7 0 0 7.7 0 1.7 1 0.3 0 0 ...  
## $ UnansweredCalls : num 6.3 2.7 0 76 0 13 2.3 4 1 0 ...  
## $ CustomerCareCalls : num 0 0 0 4.3 0 0.7 0 4 0 0 ...  
## $ ThreewayCalls : num 0 0 0 1.3 0 0 0 0 0 0 ...  
## $ ReceivedCalls : num 97.2 0 0.4 200.3 0 ...  
## $ OutboundCalls : num 0 0 0.3 370.3 0 ...  
## $ InboundCalls : num 0 0 0 147 0 0 0 0 1.7 0 ...  
## $ PeakCallsInOut : num 58 5 1.3 555.7 0 ...  
## $ OffPeakCallsInOut : num 24 1 3.7 303.7 0 ...  
## $ DroppedBlockedCalls : num 1.3 0.3 0 59.7 0 10.7 1 0.3 0 0 ...  
## $ CallForwardingCalls : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ CallWaitingCalls : num 0.3 0 0 22.7 0 0.7 0 0 0 0 ...  
## $ MonthsInService : int 61 58 60 59 53 53 57 59 53 55 ...  
## $ UniqueSubs : int 2 1 1 2 2 1 2 2 3 1 ...  
## $ ActiveSubs : int 1 1 1 2 2 1 2 2 3 1 ...  
## $ ServiceArea : Factor w/ 747 levels "AIRAIK803","AIRAND864",..: 659 611 353 611 564 564 564 561 631 611 ...  
## $ Handsets : int 2 2 1 9 4 3 2 3 4 9 ...  
## $ HandsetModels : int 2 1 1 4 3 2 2 3 3 5 ...  
## $ CurrentEquipmentDays : int 361 1504 1812 458 852 231 601 464 544 388 ...  
## $ AgeHH1 : int 62 40 26 30 46 28 52 46 36 46 ...  
## $ AgeHH2 : int 0 42 26 0 54 0 58 46 34 68 ...  
## $ ChildrenInHH : Factor w/ 2 levels "No","Yes": 1 2 2 1 1 1 1 2 2 1 ...  
## $ HandsetRefurbished : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ HandsetWebCapable : Factor w/ 2 levels "No","Yes": 2 1 1 2 1 2 2 2 2 2 ...  
## $ TruckOwner : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ RVOwner : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Homeownership : Factor w/ 2 levels "Known","Unknown": 1 1 2 1 1 1 1 1 1 1 ...  
## $ BuysViaMailOrder : Factor w/ 2 levels "No","Yes": 2 2 1 2 2 1 2 2 1 1 ...  
## $ RespondsToMailOffers : Factor w/ 2 levels "No","Yes": 2 2 1 2 2 1 2 2 1 1 ...  
## $ OptOutMailings : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ NonUSTravel : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 2 1 1 1 ...  
## $ OwnsComputer : Factor w/ 2 levels "No","Yes": 2 2 1 1 2 1 1 1 1 2 ...  
## $ HasCreditCard : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...  
## $ RetentionCalls : int 1 0 0 0 0 0 0 0 0 0 ...  
## $ RetentionOffersAccepted : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ NewCellphoneUser : Factor w/ 2 levels "No","Yes": 1 2 2 2 1 2 1 2 2 2 ...  
## $ NotNewCellphoneUser : Factor w/ 2 levels "No","Yes": 1 1 1 1 2 1 2 1 1 1 ...  
## $ ReferralsMadeBySubscriber: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ IncomeGroup : int 4 5 6 6 9 1 9 6 9 5 ...  
## $ OwnsMotorcycle : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ AdjustmentsToCreditRating: int 0 0 0 0 1 1 1 0 0 1 ...  
## $ HandsetPrice : Factor w/ 16 levels "10","100","130",..: 9 9 16 1 1 9 9 9 15 9 ...  
## $ MadeCallToRetentionTeam : Factor w/ 2 levels "No","Yes": 2 1 1 1 1 1 1 1 1 1 ...  
## $ CreditRating : Factor w/ 7 levels "1-Highest","2-High",..: 1 4 3 4 1 3 1 1 1 3 ...  
## $ PrizmCode : Factor w/ 4 levels "Other","Rural",..: 3 3 4 1 1 1 1 1 1 1 ...  
## $ Occupation : Factor w/ 8 levels "Clerical","Crafts",..: 5 5 2 4 5 4 7 5 4 5 ...  
## $ MaritalStatus : Factor w/ 3 levels "No","Unknown",..: 1 3 3 1 3 3 3 1 3 1 ...

Creating a correlation matrix  The given correlation matrix shows the correlation coefficients between sets of variables. This allows us to see which pairs have the highest correlation. A correlation matrix shows the correlation coefficients for combinations of 36 variables. Positive correlations are displayed in blue, and negative correlations in red color. Color intensity and the size of the circle are proportional to the correlation coefficients. In the bottom side of the correlogram, the legend color shows the correlation coefficients and the corresponding colors. In the above plot, 1) correlations with p-value > 0.01 are considered as insignificant. In this case, the correlation coefficient values are left blank or crosses are added. 2) )A positive correlation exists, when one variable increases as the other variable increase or one variable decreases while the other decrease. For example: If you look at MonthlyMinutes and Monthly revenue variables, they are positively correlated. If there is an increase in monthly minutes, then it directly results in increased Monthly revenue. 3) A negative correlation, the variables move in inverse, or opposite, directions. In other words, as one variable increases, the other variable decreases. For example: Customer ID and MonthsInService are negatively correlated. When new customer joins in, then he has not months in service. So there in no Correlation between these two variables and thatâs why it is showing in dark red color in matrix. 4) A value of zero indicates that there is no relationship between the two variables. In our case, dropped calls and CallForwarding calls variables are observed and found to have a correlation of 0; this means that there is no correlation, or relationship, between these two variables. 5) A weak correlation means that as one variable increase or decreases, there is a lower likelihood of being affected the second variable. For example: DirectorAssisted calls and CurrentEquipment days are loosely correted.If there is one increase in directorassisted call,it doesnât much affect CurrentEquipment days.

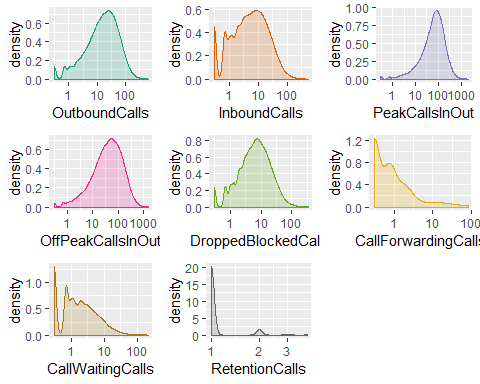
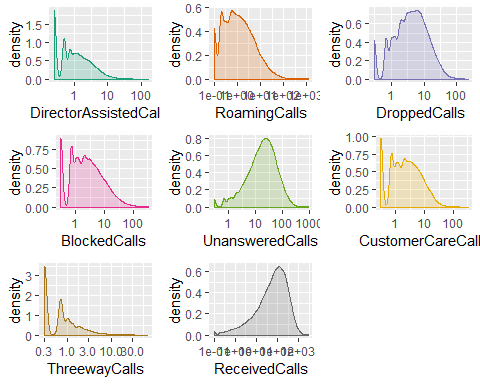
## Missing Data

## AgeHH1 AgeHH2   
## 909 909   
## PercChangeMinutes PercChangeRevenues   
## 367 367   
## MonthlyRevenue MonthlyMinutes   
## 156 156   
## TotalRecurringCharge DirectorAssistedCalls   
## 156 156   
## OverageMinutes RoamingCalls   
## 156 156   
## ServiceArea Handsets   
## 24 1   
## HandsetModels CurrentEquipmentDays   
## 1 1   
## CustomerID Churn   
## 0 0   
## DroppedCalls BlockedCalls   
## 0 0   
## UnansweredCalls CustomerCareCalls   
## 0 0   
## ThreewayCalls ReceivedCalls   
## 0 0   
## OutboundCalls InboundCalls   
## 0 0   
## PeakCallsInOut OffPeakCallsInOut   
## 0 0   
## DroppedBlockedCalls CallForwardingCalls   
## 0 0   
## CallWaitingCalls MonthsInService   
## 0 0   
## UniqueSubs ActiveSubs   
## 0 0   
## ChildrenInHH HandsetRefurbished   
## 0 0   
## HandsetWebCapable TruckOwner   
## 0 0   
## RVOwner Homeownership   
## 0 0   
## BuysViaMailOrder RespondsToMailOffers   
## 0 0   
## OptOutMailings NonUSTravel   
## 0 0   
## OwnsComputer HasCreditCard   
## 0 0   
## RetentionCalls RetentionOffersAccepted   
## 0 0   
## NewCellphoneUser NotNewCellphoneUser   
## 0 0   
## ReferralsMadeBySubscriber IncomeGroup   
## 0 0   
## OwnsMotorcycle AdjustmentsToCreditRating   
## 0 0   
## HandsetPrice MadeCallToRetentionTeam   
## 0 0   
## CreditRating PrizmCode   
## 0 0   
## Occupation MaritalStatus   
## 0 0

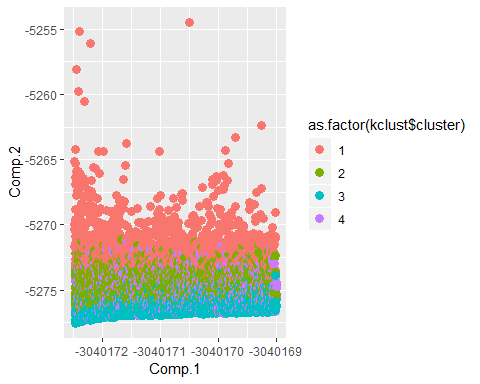
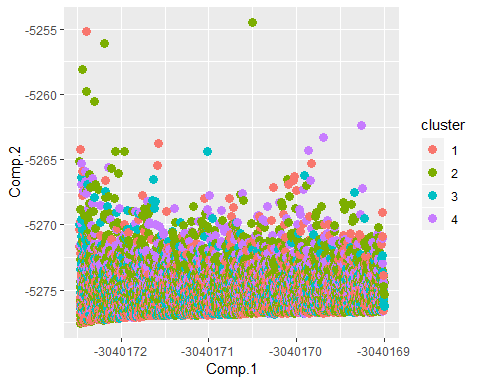
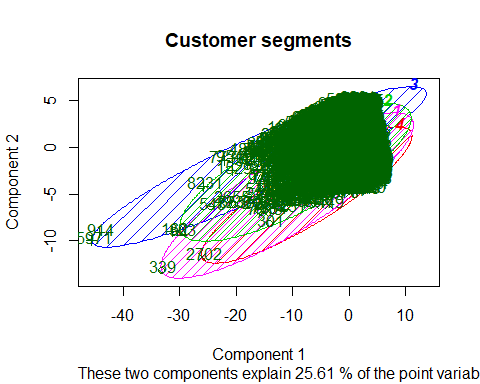
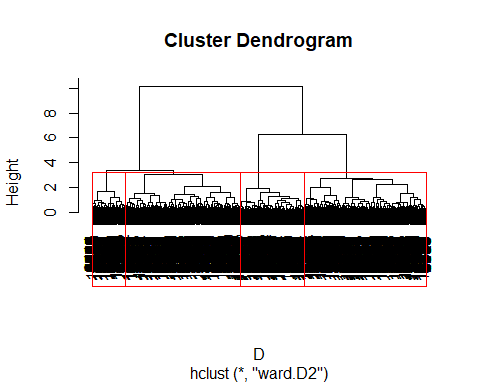
Columns with NA: MonthlyRevenue, MonthlyMinutes, TotalRecurringCharge, DirectorAssistedCalls, OverageMinutes, RoamingCalls, PercChangeMinutes, PercChangeRevenues, Handsets, HandsetModels, CurrentEquipmentDays, AgeHH1, AgeHH2.

The maximum number of records with NA columns is 909 for columns AgeHH1 and AgeHH2. We are going to eliminate these rows with NA values.

After running the na.omit command, number of rows was 51,047 and now it is 49,752 rows.



## Modeling And Data Prep



## [1] 2

## [1] 2