

University School of Management Studies



Academic Year: 2021~22

Assignment on: Wiretel Customer Churning Analysis

Post Graduate Diploma

In

Data Analytics

Under supervision of

Mr. Satyajeet Azad

Submitted By:-

Nishant

Date of Submission:-

22nd Jan 2022

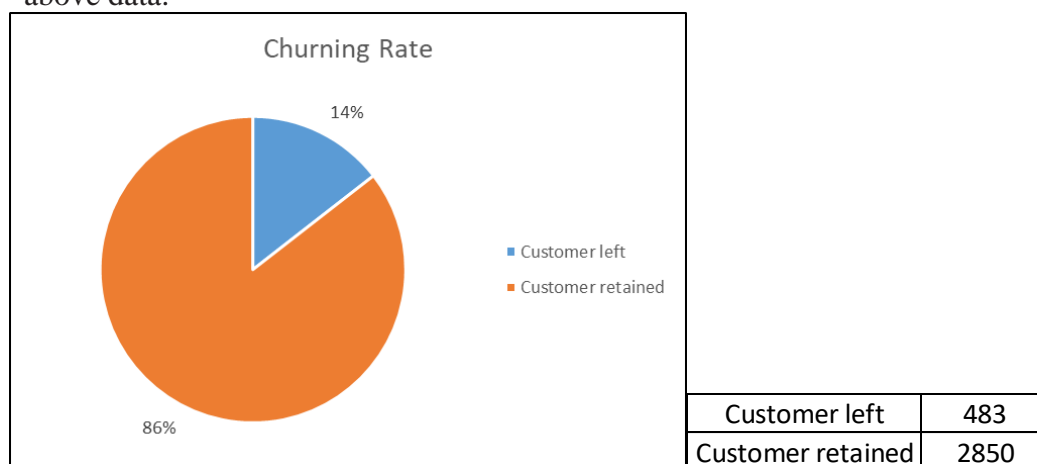
Wiretel is a newly established telecom provider in XYZ country. And as other telecom providers, wiretel also has a problem of customer churning. Churning is basically people switching their network provider to some other company due to any possible reason. So, as a data analyst, we will study the below mentioned attributes to find out factors responsible for customer churning

Parameters Available:

- *State*: Categorical, for the 50 states and the District of Columbia.
- *Account length*: Integer-valued, how long account has been active.
- *Area code*: Categorical
- *Phone number*: Essentially a surrogate for customer ID.
- *International plan*: Dichotomous categorical, yes or no.
- *Voice mail plan*: Dichotomous categorical, yes or no.
- *Number of voice mail messages*: Integer-valued.
- *Total day minutes*: Continuous, minutes customer used service during the day.
- *Total day calls*: Integer-valued.
- *Total day charge*: Continuous, perhaps based on above two variables.
- *Total eve minutes*: Continuous, minutes customer used service during the evening.
- *Total eve calls*: Integer-valued.
- *Total eve charge*: Continuous, perhaps based on above two variables.
- *Total night minutes*: Continuous, minutes customer used service during the night.
- *Total night calls*: Integer-valued.
- *Total night charge*: Continuous, perhaps based on above two variables.
- *Total international minutes*: Continuous, minutes customer used service to make international calls.
- *Total international calls*: Integer-valued.
- *Total international charge*: Continuous, perhaps based on above two variables.
- *Number of calls to customer service*: Integer-valued.
- *Churn*: Target. Indicator of whether the customer has left the company (true or false).

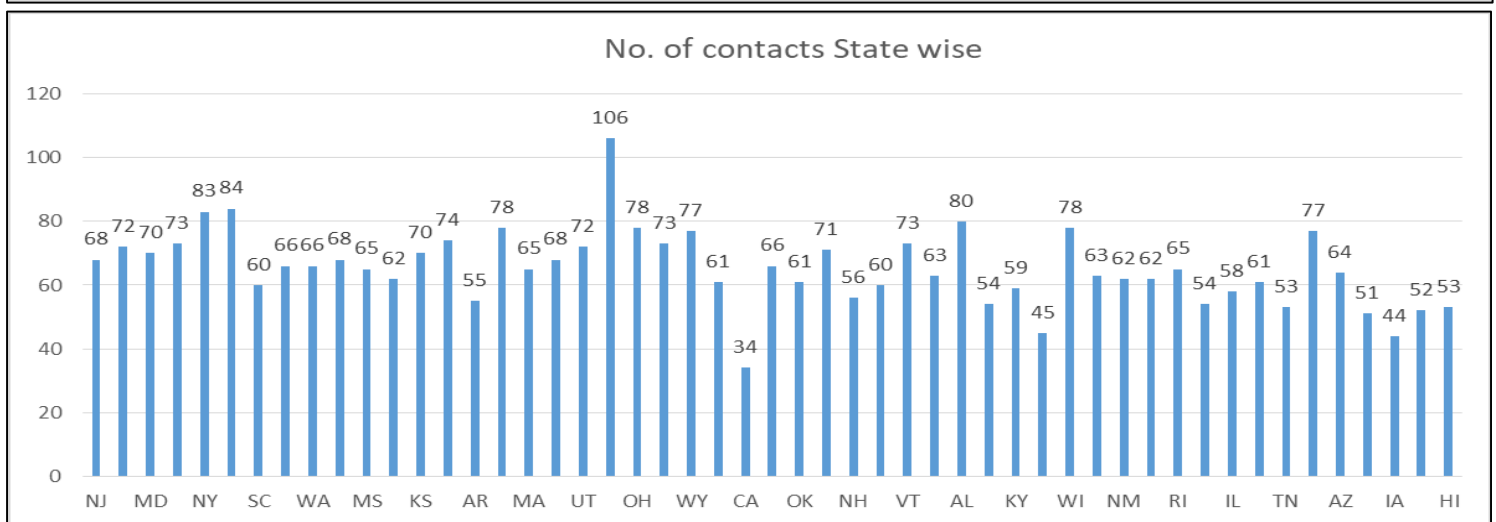
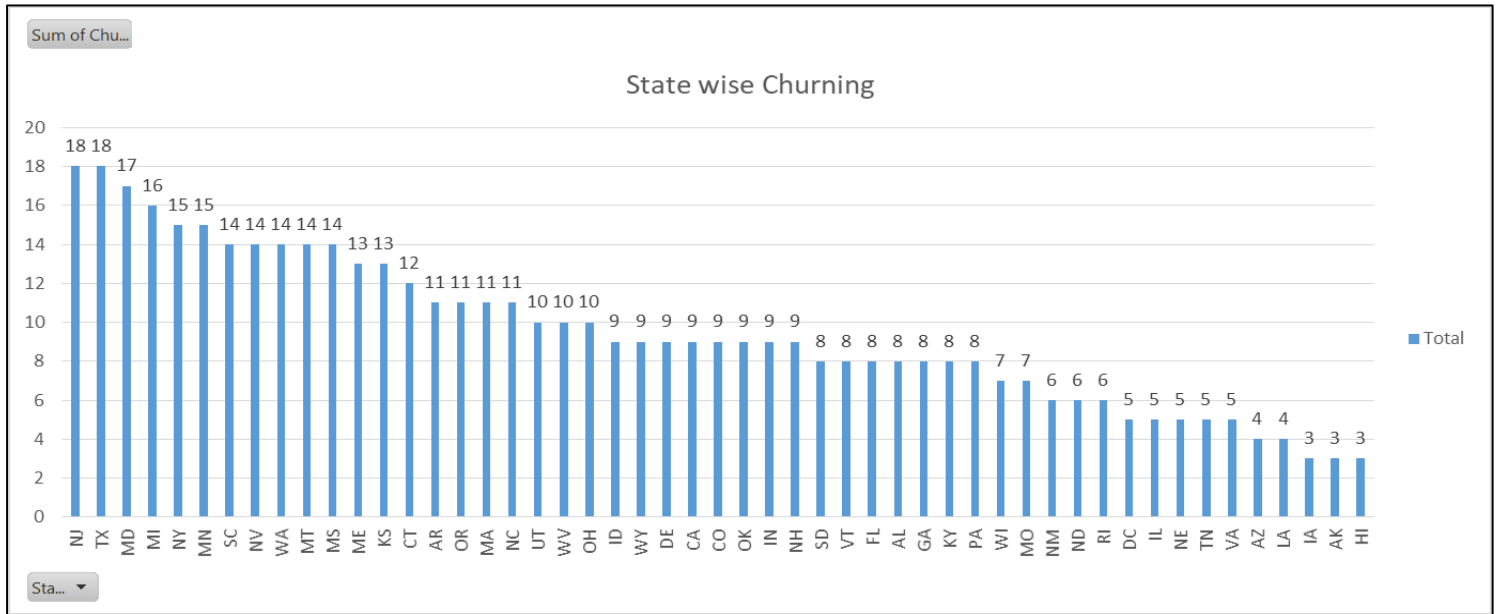
Depicting the data.

1. First we have calculated the total percentage of customer left the company as per above data.



It is clear with the chart that churning percentage is 14% of total customer.

2. In second step we try to measure customer churning state-wise.



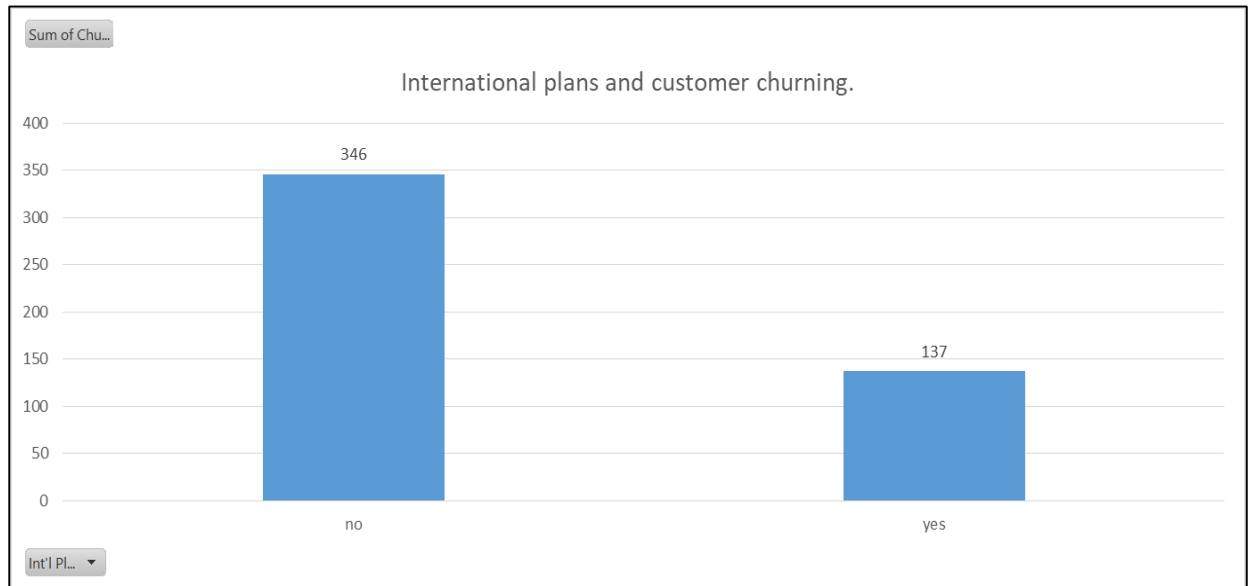
Applying descriptive statistics to analyse population distribution state-wise

We get

Column1		State Churning %	
Mean	65.35294118	NJ	26.47
Standard Error	1.65252594	TX	25.00
Median	65	MD	24.29
Mode	68	MI	21.92
Standard Deviation	11.80139573	NY	18.07
Sample Variance	139.2729412		
Kurtosis	2.223966771		
Skewness	0.382741357		
Range	72		
Minimum	34		
Maximum	106		
Sum	3333		
Count	51		
Confidence Level(95.0%)	3.319196036		

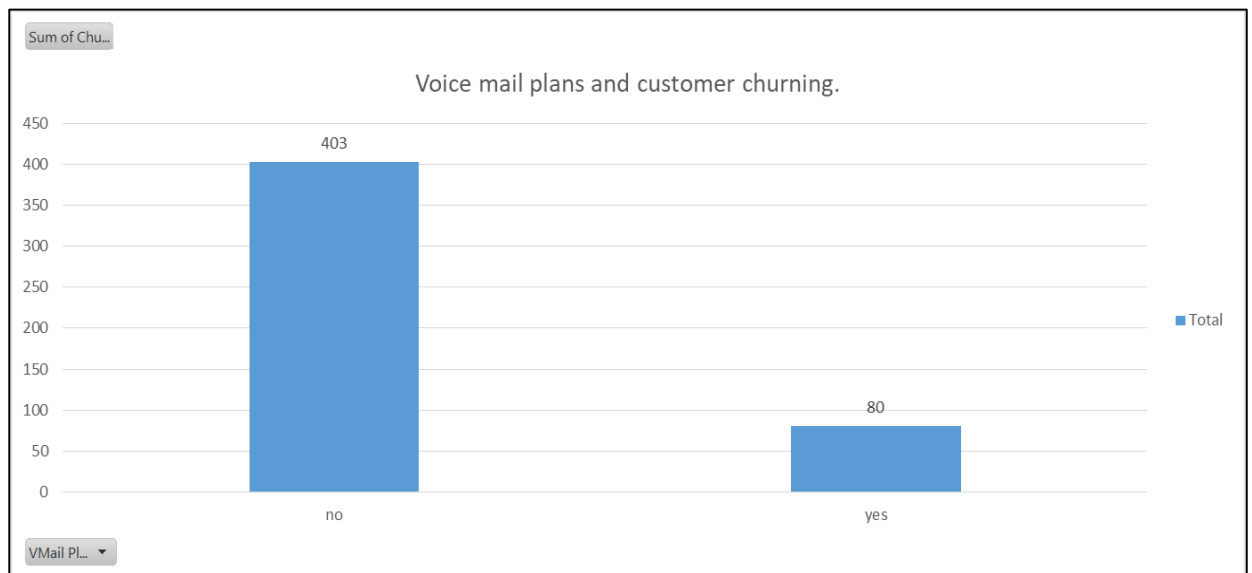
We can see states with highest churning value(NJ, TX ...) has count of contacts near to the mean value. This indicates that there is a possibility of more customer leaving in particular states.

3. Comparing customer churn with International plans.



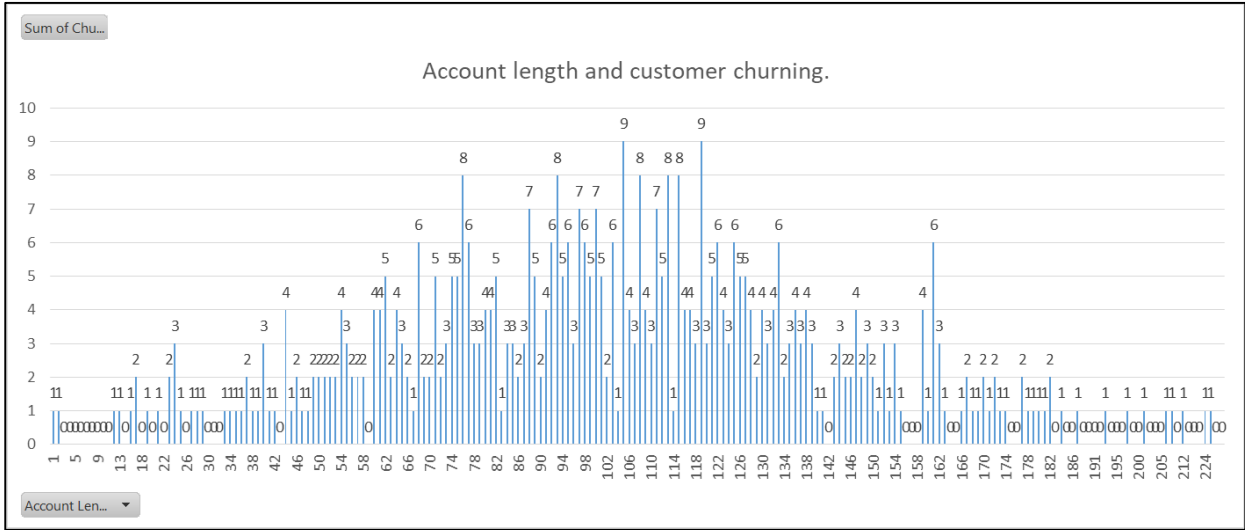
We can see that people without international plans are more likely to change its telecom company as compared to people with international plans.

4. Comparing customer churn with Voice mail plans.



It can be seen that people without voicemail plans are leaving the company similar to people with international plans.

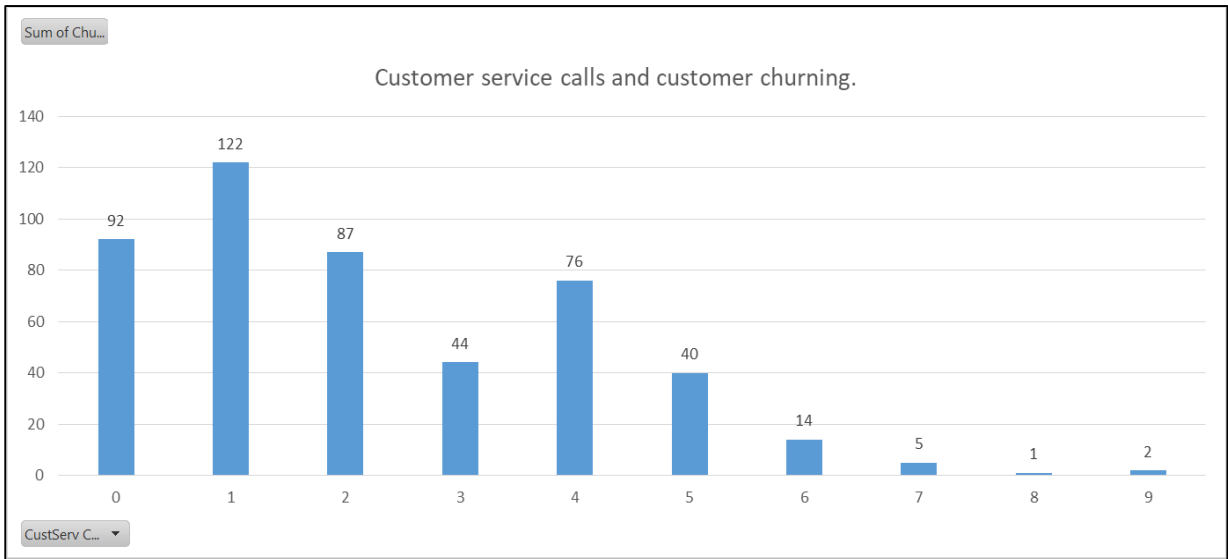
5. Comparing Account length and customer churning.



There we can see a normal distribution while comparing customer churn with Account length.

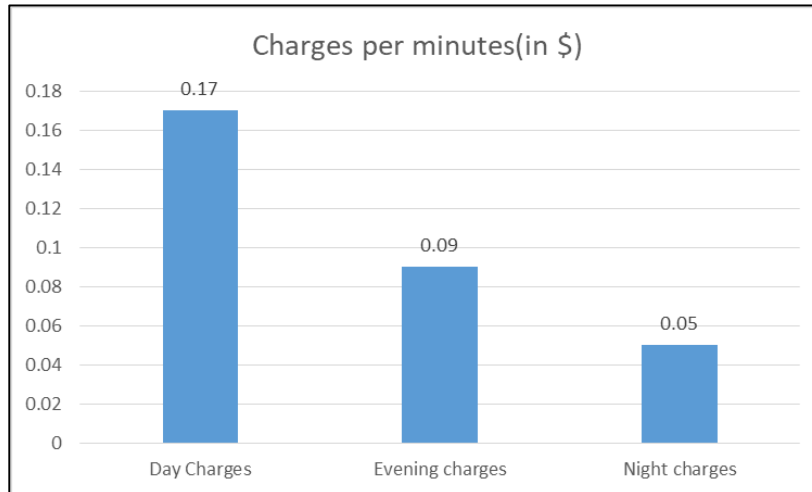
From the graph, we can see that mostly customer leaves between the time span of 2 to 5 months.

6. Comparing number of calls to customer service and customer churning.



It can be seen that customers are leaving with initial communication with customer services only. And the plot is rightward skewed.

7. Comparing day, evening, and night charges per minutes.



It is seen that calling in day time is much more expensive than calling in evening or night.

Understanding the factors that explain customer churn.

1. Preprocessing of data/ data preparation.

- First, we will transform the categorical variables in the form of 0,1 so that further modelling can be performed on these variables.
- Then, all the continuous variables are on different scale. So for numerical stability, we will normalize and standardize the data.

We have used Z-score standardization for this

$$Z = (X - \mu) / \sigma$$

- Then, we have made a correlation matrix to see related attributes.

	Int'l Plan	VMail Plan	Account Length	VMail Message	Day Mins	Day Calls	Day Charge	Eve Mins	Eve Calls	Eve Charge	Night Mins	Night Calls	Night Charge	Intl Mins	Intl Calls	Intl Charge	CustServ Calls	Churn?
Int'l Plan	1																	
VMail Plan	0.006006371	1																
Account Length	0.024734655	0.002918409	1															
VMail Message	0.008745486	0.956926642	-0.004627824	1														
Day Mins	0.049395824	-0.001684069	0.006216021	0.000778274	1													
Day Calls	0.003754626	-0.011085902	0.038469882	-0.009548068	0.006750414	1												
Day Charge	0.049398061	-0.001685629	0.006214135	0.000775523	0.999999952	0.006752962	1											
Eve Mins	0.019100013	0.021545046	-0.006757142	0.017562034	0.007042511	-0.021451408	0.007049607	1										
Eve Calls	0.006113577	-0.006444444	0.019259967	-0.005864351	0.015768993	0.006462114	0.015769282	-0.011430108	1									
Eve Charge	0.019106152	0.021558501	-0.006745302	0.01757778	0.007029035	-0.021449263	0.007036131	0.999999776	-0.011422894	1								
Night Mins	-0.028904754	0.006078657	-0.008955192	0.007681136	0.004323367	0.022937845	0.004323879	-0.012583678	-0.002092768	-0.01259202	1							
Night Calls	0.012451186	0.015552528	-0.013176275	0.007123063	0.022972456	-0.019556965	0.02297242	0.007585643	0.007709706	0.007595843	0.011203856	1						
Night Charge	-0.028912765	0.006063709	-0.008959535	0.00766329	0.004300357	0.022926638	0.004300861	-0.012592806	-0.002055984	-0.012601142	0.999999215	0.01118782	1					
Intl Mins	0.045870743	-0.001317871	0.009513902	0.002856196	-0.010154586	0.021564794	-0.010156862	-0.011034714	0.008702881	-0.011042582	-0.015207297	-0.013604996	-0.015213526	1				
Intl Calls	0.017366343	0.007617595	0.020661428	0.013957339	0.008033357	0.004574268	0.008031572	0.002541292	0.017433692	0.002541458	-0.012353432	0.00030458	-0.012329215	0.032303884	1			
Intl Charge	0.04578013	-0.001276267	0.009545675	0.002883658	-0.010091974	0.021666095	-0.010094257	-0.011066621	0.008673858	-0.011074499	-0.015179849	-0.01363017	-0.015186139	0.999992742	0.032372145	1		
CustServ Calls	-0.024521956	-0.017823944	-0.003795939	-0.013262583	-0.013423186	-0.01894193	-0.013426969	-0.012984553	0.002422575	-0.012987407	-0.009287613	-0.012801927	-0.009276954	-0.00963968	-0.017560599	-0.009674732	1	
Churn?	0.259851847	-0.102148141	0.016540742	-0.08972797	0.205150829	0.018459312	0.205150743	0.09279579	0.009233132	0.092786039	0.035492853	0.006141203	0.035495556	0.068238776	-0.052844336	0.068258632	0.208749999	1

It is clear that few attributes are highly correlational (i.e. R value near to 1). So, we can drop on attribute from each pair.

2. Running Regression Model.

Logistic regression is implemented in R- Studio for classification variable to find the probability of event success or failure.

The output of logistic regression comes in the range 0 to 1. With values >0.5 considered as 1, and values <0.5 as 0.

```

<Input>
> Reg1 =
glm(Data$Churn~Data$Int.l.Plan+Data$Account.Length+Data$VMail.Message+Data$Day.
Calls+Data$Day.Charge+Data$Eve.Calls+Data$Eve.Charge+Data$Night.Calls+Data$Night.
Charge+Data$Intl.Calls+Data$Intl.Charge+Data$CustServ.Calls)
> summary(Reg1)
<Output>
Call:
glm(formula = Data$Churn ~ Data$Int.l.Plan + Data$Account.Length +
    Data$VMail.Message + Data$Day.Calls + Data$Day.Charge + Data$Eve.Calls +
    Data$Eve.Charge + Data$Night.Calls + Data$Night.Charge +
    Data$Intl.Calls + Data$Intl.Charge + Data$CustServ.Calls)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.68355  -0.16873  -0.08514   0.01849   1.08106

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.1156782   0.0058314   19.837 < 2e-16 ***
Data$Int.l.Plan  0.3016863   0.0187944   16.052 < 2e-16 ***
Data$Account.Length 0.0036092   0.0055486    0.650 0.515434
Data$VMail.Message -0.0318056   0.0055425   -5.739 1.04e-08 ***
Data$Day.Calls    0.0066467   0.0055503    1.198 0.231183
Data$Day.Charge   0.0688380   0.0055499   12.404 < 2e-16 ***
Data$Eve.Calls    0.0016826   0.0055432    0.304 0.761492
Data$Eve.Charge   0.0327118   0.0055447    5.900 4.01e-09 ***
Data$Night.Calls  0.0007273   0.0055448    0.131 0.895652
Data$Night.Charge  0.0161225   0.0055460    2.907 0.003673 **
Data$Intl.Calls   -0.0196611   0.0055471   -3.544 0.000399 ***
Data$Intl.Charge  0.0225319   0.0055519    4.058 5.05e-05 ***
Data$CustServ.Calls 0.0767661   0.0055454   13.843 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.1022844)

    Null deviance: 413.01  on 3332  degrees of freedom
Residual deviance: 339.58  on 3320  degrees of freedom
AIC: 1874.4

Number of Fisher Scoring iterations: 2

```

Now, from above regression model, value of Account length, Day calls, Night calls and Evening call are having probability more than 0.05. So, these values do-not have significant effect on churning.
So, we have further reduced our model to.

```

<Input>
> Reg2 <- glm(Data$Churn~ Data$Int.l.Plan+Data$VMail.Message +
Data$Day.Charge+Data$Eve.Charge+Data$Night.Charge+Data$Intl.Calls+Data$Intl.Charge
+Data$CustServ.Calls)
> summary(Reg2)

```

<Output>

Call:

```
glm(formula = Data$Churn. ~ Data$Int.l.Plan + Data$VMail.Message +  
    Data$Day.Charge + Data$Eve.Charge + Data$Night.Charge + Data$Intl.Calls +  
    Data$Intl.Charge + Data$CustServ.Calls)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.68023	-0.16905	-0.08360	0.01838	1.07463

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.115638	0.005830	19.837	< 2e-16 ***
Data\$Int.l.Plan	0.302102	0.018782	16.085	< 2e-16 ***
Data\$VMail.Message	-0.031894	0.005540	-5.757	9.36e-09 ***
Data\$Day.Charge	0.068942	0.005546	12.431	< 2e-16 ***
Data\$Eve.Charge	0.032531	0.005541	5.871	4.77e-09 ***
Data\$Night.Charge	0.016252	0.005542	2.932	0.003388 **
Data\$Intl.Calls	-0.019534	0.005544	-3.524	0.000431 ***
Data\$Intl.Charge	0.022705	0.005548	4.092	4.37e-05 ***
Data\$CustServ.Calls	0.076627	0.005542	13.826	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.1022239)

Null deviance: 413.01 on 3332 degrees of freedom
Residual deviance: 339.79 on 3324 degrees of freedom
AIC: 1868.4

Number of Fisher Scoring iterations: 2

And the regression equation will be:

Churn = 0.115 + 0.302*International_plan – 0.032*Voicemail_messages +
0.068*Day_charges + 0.032*Evening_charges + 0.016*Night_charges –
0.019*International_calls + 0.022*International_charges + 0.077*Customer_service_calls.

Why Customers are churning from wiretel?

Based on above analysis, there are following significant factors effecting the customer churning.

- International plan
- Voicemail msgs
- Day charges
- Evening charges
- Night charges
- International calls
- International charges
- Calls to customer services

Also there are some less significant factors as mentioned below.

- Voice mail plan.
- Account length
- State

Recommendation to the management/ Retention strategy.

The model describes the significant attributes for churning of customers. The company can predict churning of customers and offer a more attractive deals at the right time to retain their customers.

- A more focused approach on states with higher churning rate to find out reason behind customers churning from these states.
- Company can offer a standardized charges plan for Day, evening and night time to attain customer satisfaction.
- Company shall offer discount offers or other scheme to customer with account length between 2-5 months to ensure retention of customers. Also, some legacy offers can be offered to customer who are using wiretel's phone number from more than 9 months or 1 year with a period of 1 year or every six months.
- Company should review their voice mail plans and international plans to attract more customer's opting for these plans.