Telecom Churn Analysis

1. **Data Source:** churn\_data.txt
2. **Summary of the project:**

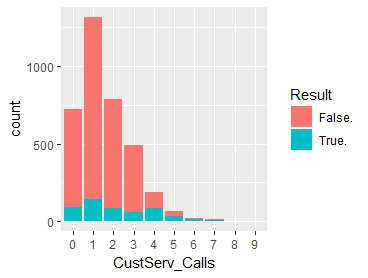
Churn (loss of customers to competition) is a problem for telecom companies because it is expensive to acquire a new customer and companies want to retain their existing customers. The goal of this project is to reduce the churn using analytics.

1. **Why to predict:**

Most telecom companies suffer from voluntary churn. Churn rate has strong impact on the life time value of the customer because it affects the length of service and the future revenue of the company. For example if a company has 25% churn rate then, the average customer lifetime is 4 years; similarly a company with a churn rate of 50%, has an average customer lifetime of 2 years.

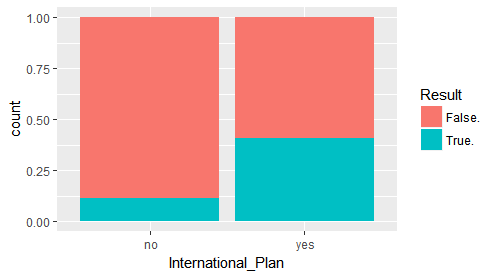
Telecom companies spend hundreds of dollars to acquire a new customer and when that customer leaves, the company not only loses the future revenue from that customer but also the resources spend to acquire that customer. Churn erodes profitability.

1. **Objective:**
2. To predict customer churn.
3. To develop a predictive model and checking accuracy of that model.
4. **Modelling techniques and tools used:**
5. R statistical language and R Studio IDE
6. Support Vector Machine algorithm using RBF kernal
7. **Exploratory Data Analysis:**
8. customer\_data contains 4617 observations and 21 variables.
9. No missing value in any column.
10. Alloting column names to each column
11. Column “CustServ\_Calls”and “Area\_Code” has been converted into factor.
12. Creating new features like bill,avg\_day\_call,avg\_eve\_call , avg\_night\_call,avg\_intl\_call,phn\_no\_cat
13. Consider 3600 rows as train dataset and rest of the data as test data sets.Here we are using random sampling of data.



From fig 1.1 we can make a conclusion that:

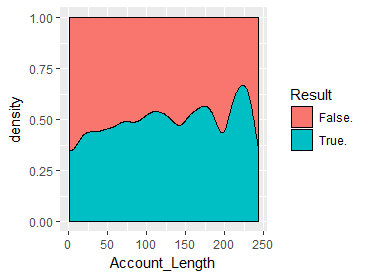
1. Most of the user called helpdesk 0-3 times.



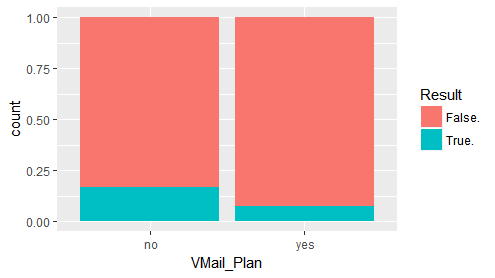
From fig 1.2 we can conclude that:

1. Customers those who have internatinal plan are more likely to churn.This is a major concern for any telecom operator because

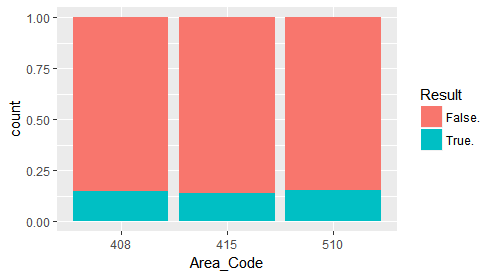
Customers those who have international plan (It’s a general assumption) generate more “per customer revenue” rather than normal customer.So losing these segment customers is not good for company.



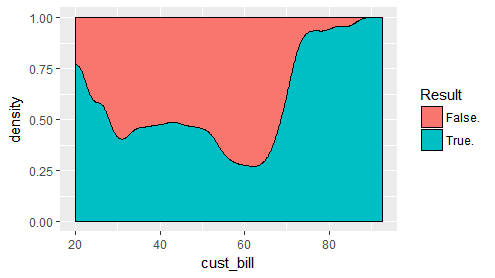
1. From fig 1.3 Old customers are more likely to discontinue service compared to new customer.



1. From fig 1.4 we can see that customers those who have Vmail plan are less likely to churn.



1. From fig 1.5 we can say that customer loss pattern is same across all the areas.



1. From fig 1.6 we can see that those who have bill range 30-70 are less likely to churn.
2. From fig. ‘avg\_min\_per\_call’,’No\_of\_calls’ and ‘Total\_minute’ we can we can make a assumption that in Daytime there maybe some problem with service quality as customers who are more active in daytime are more likely to discontinue service.

**B. Train the model:**

We are going to apply SVM to this dataset.(RBF kernal).After train the model with different cost and gamma value we get the best model with cost value 10 and gamma value 0.02

For this model training error is 7.5%(approx)

* **Model performance on test data:**

Test data contains 1017 observations.From confusion matrix table we can get around 93% accuracy.