**File Name: churn\_data.xls**

**Adopted from an UCI synthetic dataset.**

“Churn” is a common phenomenon that occurs in the telecom Industry. “Churn” refers to the customers, who will leave the service in the near future. Acquiring a new customer is more expensive than retaining a customer. Therefore, companies would like to retain the customers and predict in advance the attributes that influence churn.

Your task is to identify the type of customers who will churn and prepare a plan for churn management.

The attributes are given below:

**Attribute Information:**

|  |  |  |
| --- | --- | --- |
| Attribute name | Description | Type |
| State | Customer State | discrete |
| Account Length | Tenure with the company in months | continuous. |
| Area Code | Area code | continuous. |
| Phone | phone number | discrete. |
| Int'l Plan | membership in the International plan | discrete. |
| VMail Plan | membership in the Voice plan | discrete. |
| VMail Message | number of voice mail messages | continuous. |
| Day Mins | Total Mins in Day Calls | continuous. |
| Day Calls | number of day calls made | continuous. |
| Day Charge | Total Charges for Day Calls | continuous. |
| Eve Mins | Total Mins in Eve Calls | continuous. |
| Eve Calls | number of evening calls made | continuous. |
| Eve Charge | Total Charges for Eve Calls | continuous. |
| Night Mins | Total Mins in Night Calls | continuous. |
| Night Calls | number of night calls made | continuous. |
| Night Charge | Total Charges for Night Calls | continuous. |
| Intl Mins | Total Mins in International Calls | continuous. |
| Intl Calls | number of international calls made | continuous. |
| Intl Charge | Total Charges for International Calls | continuous. |
| CustServ Calls | Number of Customer Service Calls | continuous. |
| Churn? | Churn | discrete. |
| Churn\_Training | Training Data | discrete. |

**Data Pre-processing:**

1. Summarise the data.
2. What insights does the exploratory analysis yield regarding the choice of predictors?

**Model Fit and diagnostics:**

1. Build a logistic regression model
2. Comment on the significance and the explanatory power of the model
3. Write the estimated equations in three formats:
   1. The logit as a function of the predictors
   2. The odds as a function of the predictors
   3. The probability as a function of the predictors
4. Assess model fit.
5. what would be the ideal cut-off to increase the percentage of correctly classified customers who churn?

**Recommendations**