Business Case Overview:

This case requires trainees to develop a model for predicting customer churn at "Cell2Cell," a fictitious wireless telecom company, and use insights from the model to develop an incentive plan for enticing would-be churners to remain with Cell2Cell.

Data for the case are available in csv format. The data are a scaled down version of the full database generously donated by an anonymous wireless telephone company. There are still 71,047 customers in the database, and 75 potential predictors. Trainees can use whatever method they wish to develop their predictive model. Logistic regression is perhaps the most obvious choice and is adequate for the task.

The data are available in one data file with 71,047 rows that combines the calibration and validation customers. "calibration" database consisting of 40,000 customers and a "validation" database consisting of 31,047 customers. Each database contained (1) a "churn" variable signifying whether the customer had left the company two months after observation, and (2) a set of 75 potential predictor variables that could be used in a predictive churn model. Following usual model development procedures, the model would be estimated on the calibration data and tested on the validation data. At the time, Cell2Cell's churn rate was about 2% per month. However, data set has been created the calibration database so that it contained roughly 50% churners. The validation data contained 2% churners.

This case requires both statistical analysis and creativity/judgment. I recommend you not spend too much time on fine-tuning your predictive model; make sure you spend sufficient time interpreting results.

Expectations from the Trainees:

Your task is to execute the 3-stage process for proactive churn management. Please answer the following questions:

1. Data cleaning including missing values, outliers and multi-collinierity. Describe your predictive churn model. How did you select variables to be included in the model?
2. Demonstrate the predictive performance of the model.
3. What are the key factors that predict customer churn? Do these factors make sense?
4. What offers should be made to which customers to encourage them to remain with Cell2Cell? Assume that your objective is to generate net positive cash flow, i.e., generate additional customer revenues after subtracting out the cost of the incentive.
5. Assuming these actions were implemented, how would you determine whether they had worked?

Data Dictionary:

Position	Variable Name	Variable Descriptiion
1	revenue	Mean monthly revenue
2	mou	Mean monthly minutes of use
3	recchrge	Mean total recurring charge
4	directas	Mean number of director assisted calls
5	overage	Mean overage minutes of use
6	roam	Mean number of roaming calls
7	changem	% Change in minutes of use
8	changer	% Change in revenues
9	dropvce	Mean number of dropped voice calls
10	blckvce	Mean number of blocked voice calls
11	unansvce	Mean number of unanswered voice calls
12	custcare	Mean number of customer care calls
13	threeway	Mean number of threeway calls
14	mourec	Mean unrounded mou received voice calls
15	outcalls	Mean number of outbound voice calls
16	incalls	Mean number of inbound voice calls
17	peakvce	Mean number of in and out peak voice calls
18	opeakvce	Mean number of in and out off-peak voice calls
19	dropblk	Mean number of dropped or blocked calls
20	callfwdv	Mean number of call forwarding calls
21	callwait	Mean number of call waiting calls
22	churn	Churn between 31-60 days after obs_date
23	months	Months in Service
24	uniqsubs	Number of Uniq Subs
25	actvsubs	Number of Active Subs
26	csa	Communications Service Area
27	phones	# Handsets Issued
28	models	# Models Issued
29	eqpdays	Number of days of the current equipment
30	customer	Customer ID
31	age1	Age of first HH member
32	age2	Age of second HH member
33	children	Presence of children in HH
34	credita	Highest credit rating - a
35	creditaa	High credit rating - aa
36	creditb	Good credit rating - b

37	creditc	Medium credit rating - c
38	creditde	Low credit rating - de
39	creditgy	Very low credit rating - gy
40	creditz	Lowest credit rating - z
41	prizmrur	Prizm code is rural
42	prizmub	Prizm code is suburban
43	prizmtwn	Prizm code is town
44	refurb	Handset is refurbished
45	webcap	Hanset is web capable
46	truck	Subscriber owns a truck
47	rv	Subscriber owns a recreational vehicle
48	occprof	Occupation - professional
49	occcler	Occupation - clerical
50	occcrft	Occupation - crafts
51	occstud	Occupation - student
52	occhmkr	Occupation - homemaker
53	occret	Occupation - retired
54	occself	Occupation - self-employed
55	ownrent	Home ownership is missing
56	marryun	Marital status unknown
57	marryyes	Married
58	marryno	Not Married
59	mailord	Buys via mail order
60	mailres	Responds to mail offers
61	mailflag	Has chosen not to be solicited by mail
62	travel	Has traveled to non-US country
63	pcown	Owns a personal computer
64	creditcd	Possesses a credit card
65	retcalls	Number of calls previously made to retention team
66	retaccpt	Number of previous retention offers accepted
67	newcelly	Known to be a new cell phone user
68	newcelln	Known not to be a new cell phone user
69	refer	Number of referrals made by subscriber
70	incmiss	Income data is missing
71	income	Income (0=>missing)
72	mcycle	Owns a motorcycle
73	creditad	Number of adjustments made to customer credit rating (up or down)
74	setprcm	Missing data on handset price
75	setprc	Handset price (0=>missing)
76	retcall	Customer has made made call to retention team
77	calibrat	Calibration sample = 1; Validation sample = 0;
78	churndep	Churn (=missing for validation sample)