Business Case Overview:

This case requires candidate to develop draw inference and build an API for determining customer churn at "C2C," a fictitious wireless telecom company, and use insights from the data to develop an incentive plan for enticing would-be churners to remain with C2C.

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 71,000 customers in the database, and 75 potential predictors. Candidate can use whatever method they wish to develop their inference.

The data are available in one data file with 71,000 rows that combines the calibration and validation customers. "calibration" database consisting of 40,000 customers and a "validation" database consisting of 31,000 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. Inference can be estimated on the calibration data and tested on the validation data. At the time, C2C's churn rate was about 2% per month. However, a data set has been created for the calibration database so that it contained roughly 50% churners. The validation data contained 2% churners.

This case requires both data/ statistical analysis and creativity/judgment, make sure you spend sufficient time interpreting results.

Expectations from the Candidate:

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 multicollinearity.

Important variables to be included in the data for inference?

2. What can be the key factors that predict customer churn? Do these factors make sense?

3. What offers can be made to which customers to encourage them to remain with C2C?

4. Expose your code as a functional API.

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.

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