SELECT COLLECTIONS, INC.

Marissa Wells, manager of collections for Select Collections, Inc., called in Marcos Kilduff, a summer intern, to describe an assignment:

I want to see if we can build a model to predict how much money we will collect from delinquent accounts. Such a model would prove very useful in deciding which accounts to purchase and how much to pay.

You learned regression during your first year, didn't you? Well, see what you can do for us. There are 3,570 accounts in a data set that I will call the training set. We purchased these accounts and then collected the amount shown in the last column, labeled *totalpay*. *Totalpay* is the variable of interest. I want you to come up with a way to forecast *totalpay*.

We've also assembled a bunch of potential predictor variables for you to use. All of these variables are "useable" in the sense that we will know them for new accounts about which we are making decisions. Your model can use any of these variables in any combination and in any form. The data dictionary [see **Exhibit 1**] explains each of these variables.

To make this assignment interesting, I want you to use the model you come up with to predict *totalpay* for the 3,570 accounts in a separate worksheet of data that I will call the test set. The accounts in this test data set are no different from the accounts in the training set except that I know the *totalpay* values and you don't. So, by tomorrow morning, I want an Excel file from you containing a single column of numbers. In cell A1, please put the adjusted R-squared for your model. In cells A2 through A3571, put your predicted values for the accounts in the test data set in ID order. Tomorrow, I'll use the actual *totalpay* values for the accounts in the test set to calculate your model's mean squared error—the average of your model's 3,570 squared errors.

And, in the interest of full disclosure, I'm giving this same assignment to all our interns. It'll be kind of fun, don't you think?

Oh, I almost forgot. We're also giving you some descriptive statistics and plots [see **Exhibit 2**] to give you a head start.

Company Background

Select Collections, Inc., was a start-up subsidiary of a major credit card company. The company purchased distressed consumer debt at discounted rates from such major credit card companies as Chase, Wells Fargo, and Bank of America and then used data-driven decision-making and dynamic value assessment to optimize the collection processes associated with the purchased accounts.

For each purchased account, the first decision was whether to resell or attempt to collect. For accounts the company decided to attempt to collect, a host of tactics was available for use in any sequence and with any frequency.

Select Collections' strategic intent was to become the best in the world at tailoring and optimizing the collection process. Like other collection companies, Select Collections used the telephone and the legal system as its two major collection tools. The company had recently opened a large, state-of-the-art call center in Topeka, Kansas, and its legal department was active in 46 states. Select Collections strictly followed all state and federal regulations regarding the collection process.

Exhibit 1

SELECT COLLECTIONS, INC.

Data Dictionary

Field	Definition
acctid	Account ID, a unique index used to identify accounts
state	State in which the account-holder lives (TEXT)
zip	ZIP code in which the accountholder lives
rollout	Card issuer from whom the account was purchased
cobal	The balance of the account at the point of "charge-off" when the account was purchased
collscr	The interpretation of this field is not fully known; however, it is produced by an in-house legacy system and exported to the data warehouse for every account. This is a text variable
cs	The left-most two digits of *collscr*. This variable is a number.
accessscr	The account's accessibility score, an a priori estimation of how likely it is that the accountholder is reached via phone.
lnacscr	The natural log of the *accessscr*
bureauscr	The output of an in-house prediction model based on the likelihood of receiving payments from the accountholder using credit-bureau attributes as predictors
eaglemod	The output of the Eagle System's proprietary model, an estimate of the risk of nonpayment of the account
numcalls	The number of telephone calls made (to date) to the accountholder
numrpcs	The number of "right-party connects" or phone calls in which the collection agent speaks with the accountholder
totalpay	The total amount of payments received from the accountholder (in dollars)

Source: Created by case writer.

Exhibit 2

SELECT COLLECTIONS, INC.
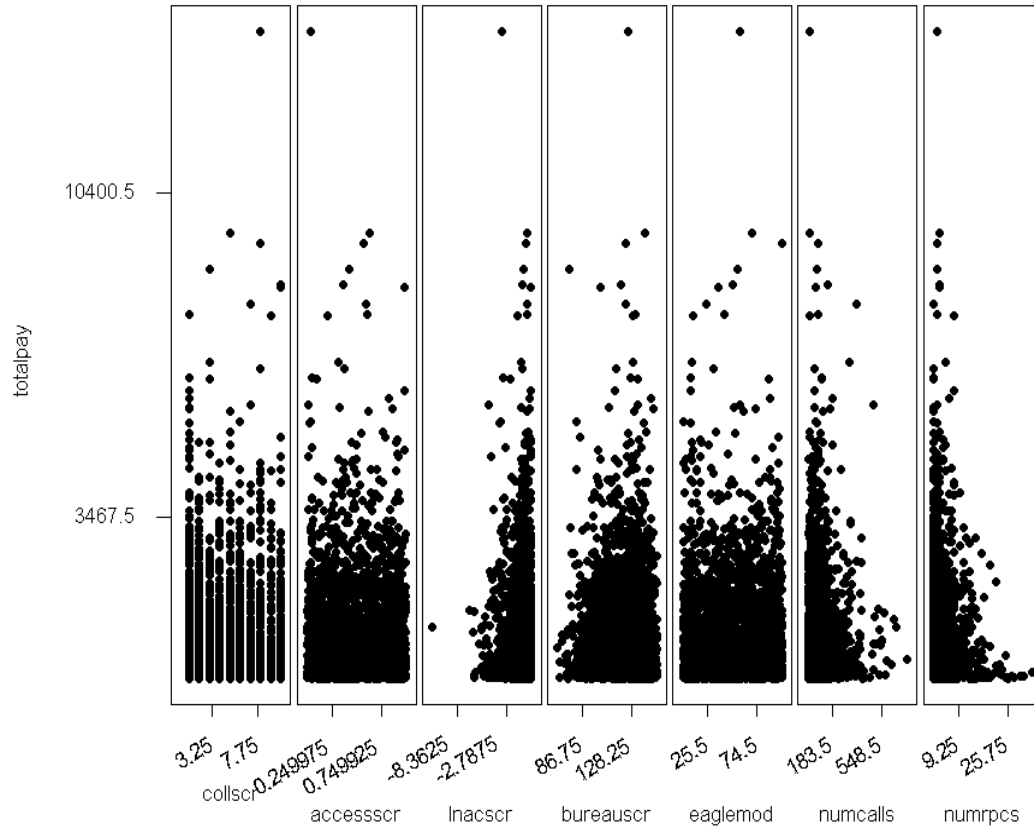
Descriptive Statistics and Plots of Training Data Set

Variable	N	Mean	Median	TrMean	StDev	Minimum	Maximum
cobal	3570	3231.7	2525	3030	2268.2	263	17534
cs	3570	4.4031	4	4.2808	2.8821	1	10
accesssc	3570	0.49524	0.4928	0.49485	0.2895	0	0.9999
lnacscr	3570	-1.0163	-0.71	-0.9052	1.0088	-11.15	0
bureausc	3570	121.27	123	121.91	15.81	66	149
eaglemod	3570	50.042	50	50.045	28.531	1	99
numcalls	3570	80.87	62	72.05	75.49	1	731
numrpcs	3570	3.3922	2	2.9312	3.3448	1	34
totalpay	3570	807.9	463.5	673.8	1016.8	1	13867

Rollout	Count
Associates	548
Bank_Of_Am	631
Chase	336
Chase_Bony	23
Chase_Rev	197
Discover	1178
Wells	614
Wells_FIB	43

TrMean refers to trimmed mean. The trimmed mean is the average calculated after removing the 5% highest and 5% lowest values.
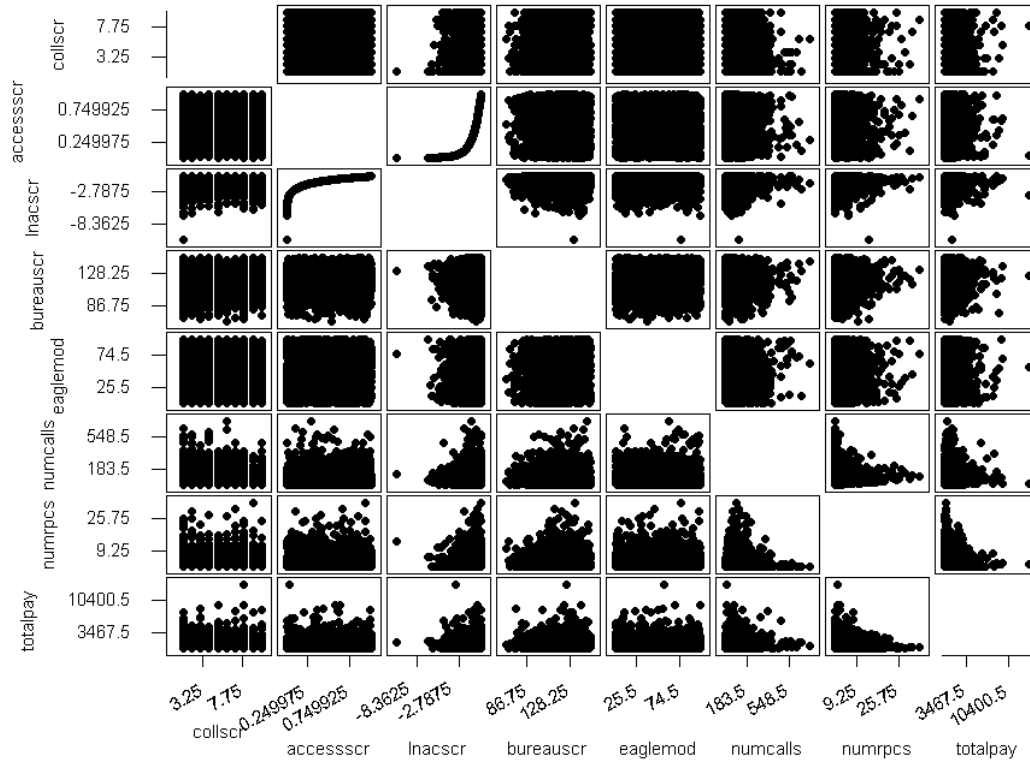
Source: Created by case writer.

Exhibit 2 (continued)

Source: Created by case writer.

Exhibit 2 (continued)

Source: Created by case writer.