Spring 2013

Customer Age as a Predictor of Contact Volume

Tollan Renner
University of New Hampshire - Main Campus

Recommended Citation

Customer Age as a Predictor of Contact Volume

Tollan Renner
University of New Hampshire
tdb26@unh.edu
May 2013

Abstract

A two stage modeling approach for modeling customer age as a predictor of contact volume was conducted using a real-world data set of approximately 2,000,000 contacts from a company call center. Two models were constructed in the first stage, one a straightforward regression and the other a series of regressions. One was selected as better performing and scaled up to predict calls received from calls answered. The second stage of the modeling included a day of the week covariate and performed the best of the models created. This model uses age bins as model effects, of which the youngest age bin had the largest effect. The model predicts the volume of calls received by the company.

Table of Contents

Acknowledgments

I would like to thank my co-adviser Matthew Shump, who provided continuous help and advice on the project's completion. I would also like to thank my thesis adviser, Dr. Ernst Linder, of the Department of Mathematics and Statistics at the University of New Hampshire for providing guidance on this project, helpful feedback on this paper, and additional modeling ideas. Furthermore, I am appreciative of Dr. Phil Ramsey, also of the University, whose courses with JMP were very valuable for performing the theoretical and empirical analyses portions of this paper.

Customer Age as a Predictor of Contact Volume

1. Introduction

The primary objective of this report is to outline my undergraduate research on customer age as a predictor of contact volume for a business analytics project in accordance with the honor's thesis requirements of the Department of Mathematics and Statistics. To clarify the technical terms in the title, customer age is the time a customer has been active with the company after signing up, and contact volume is the amount of contacts a company receives. Modern companies usually record these contacts, which may be in form of calls, emails, etc., in computer databases, and this paper uses calls and contacts interchangeably. My project employed data mining and regression techniques to model customer contacts volume for an online services company. Matthew Shump and Dr. Ernst Linder provided theoretical analysis support for the project.

The primary focus of my work is the role that age of a customer plays in customer contact volume. A descriptive mathematical model that describes and predicts this relationship for contact volume was developed. The JMP statistical software package, a product of the SAS Institute Inc., was used extensively to analyze the data and the JMP Scripting Language (JSL) was also employed to create the data table for the regressions. The outcome of this project described in final sections of this report is designed to be useful for the company in assessing the required level of detail for their models and data collection. From a theoretical standpoint, this project attempts to determine if matching each individual call to a single customer of the company will be a more accurate method of predicting the contacts received response than the traditional approach of using a regression model with multiple input variables.

Customers may be categorized by their age with the company, that is the time they have been with the company since they signed up. One goal of the modeling efforts is to determine if the ways the customer contact age groups are interacting with the company are changing from age group to age group.

1

One model addresses the potential differences in the behavior of customers at different life stages with the company, the alternative being treating them as a single approximately homogenous group. Methods for forecasting the number of customer sign ups and the survival of customers after they sign up are already fairly accurate, and are not considered in this report.

The modeling techniques employed in this paper are two-stage. The first stage of the modeling is a two model approach to uncover age determined characteristics of the data without considering other covariates. The goal of these models is to predict total contacts received. The first model, Total Contacts Received by Active Customer Age Bins, is designed to capture the age factor that drives contact volume by dividing the contacts into the age-at-contact bins. This first method is a standard regression that assumes the contacts associated with customers are not easily separable by those customer's ages. The model finds the inputs that are significant contributors to the response variable's variation. The second method uses the additional information of linking contacts to a specific customer. This method helps identify the age of the customer at the time of the contact. The basis of this model is multiple regressions between each individual input of customer age group and the contact volume directly attributable to this cohort. Each regression model explains the unique variation in the total contacts received response variable and those regression models are then added together. This model is then be compared to the model found through method one to find the model with the better prediction potential. For both methods total contacts answered is used as the response and the final model is scaled up to predict total contacts received. The second stage was to consider time covariates in addition to the age bins, and the stage two model had the best prediction performance.

Three data files were employed in the project. The first file, the contacts logger, contains close to two million contact answered observations, with the customer identification number, contact date, and contact identification number recorded for each observation. The second file, denoted customers, contains

2

information on over half a million customers. For each observation of this file, the customer identification number, sign up date, and delete date where applicable are recorded. Finally, the third file, contacts received, records the number of contacts received each day from December, 2011, to November, 2012. An overview of these files can be found in Appendix A.

The final product of the project is a statistical model with inputs of customer age in terms of bins determined from the analysis and the output of total contact volume. The research questions explored are what are the features of the models and how can they be used to predict call volumes based on predicted user sign ups and age cohorts? Furthermore, how accurate are the models? An important aspect of the models are their applicability to predicting call volumes based on predicted user sign ups and age cohorts.

2. Literature

Several customer contact-oriented research papers comprise the literature related to this paper's research objective. First, The Institute for Operations Research and the Management Sciences (INFORMS) publishes several operations research papers, some of relevance to this project. "Improving Customer Service Operations at Amazon.com," by Keblis and Chen, and, "Design and Control of a Large Call Center," by Bassamboo et al., are of particular interest. Other relevant operations research papers include Duder and Rosenwein (2001), Whitt (2006), and the American Statistical Association's "Statistical Analysis of a Telephone Call Center," by Brown et al. Keblis and Chen (2006) develop a mathematical model to plan capacity and reduce the average cost of handling a customer contact for Amazon.com, similar to the statistical model developed in this report. Similarly, Bassamboo et al. (2006) analyzes a call center model, and the authors break the customers into pools similar to how customers are organized into into age bins in this report. Brown et al. (2005) considers contact volume as a function of time of day during the weekday, and my paper attempts to predict contact volume as a function of the existing

customer base. Duder and Rosenwein (2001) are especially concerned with achieving "zero abandonments" in call center performance by ensuring customer service centers are adequately staffed to never cause customers to wait long enough to become discouraged and leave the company. From a sales perspective, retaining the leads who contact the company is vital to building a business. Finally, Whitt (2006) optimizes call center staffing using two methods, a deterministic fluid approximation, and a numerical algorithm based on a stochastic Markovian birth-and-death model. The author, an operations researcher at Columbia University, considers such factors as uncertain arrival rates, customer abandonment and waiting costs, and employee absenteeism. These papers provide some background for this paper's research objective.

3. Theoretical Analysis
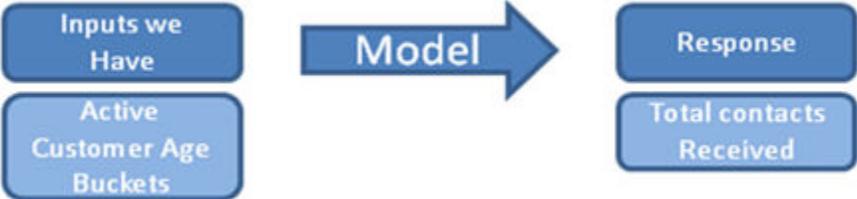
Figure 1. General Modeling Approach

The methodology used to complete the data analysis portion of this report included data manipulations and modeling. After the data has been cleaned up, the first step is determining the customer age bins to be used as the initial model inputs. The size of the customer age bins are determined to be roughly evenly partitioned based on a distribution of the contacts by contact age. After the age bins have been determined, the next step is the creation of a data table of contact volume answered by day and by customer age. This table functions as the response variable for the regression model. Next, a data table of the total active customers by age bins by day is generated using the age bins determined in step one. This

table is the input file to the regression models. These first three steps prepare the data through manipulations for the final modeling steps.
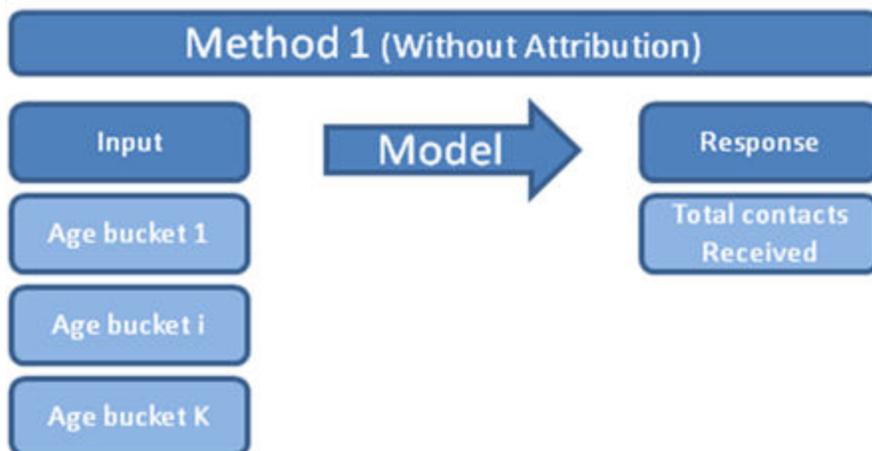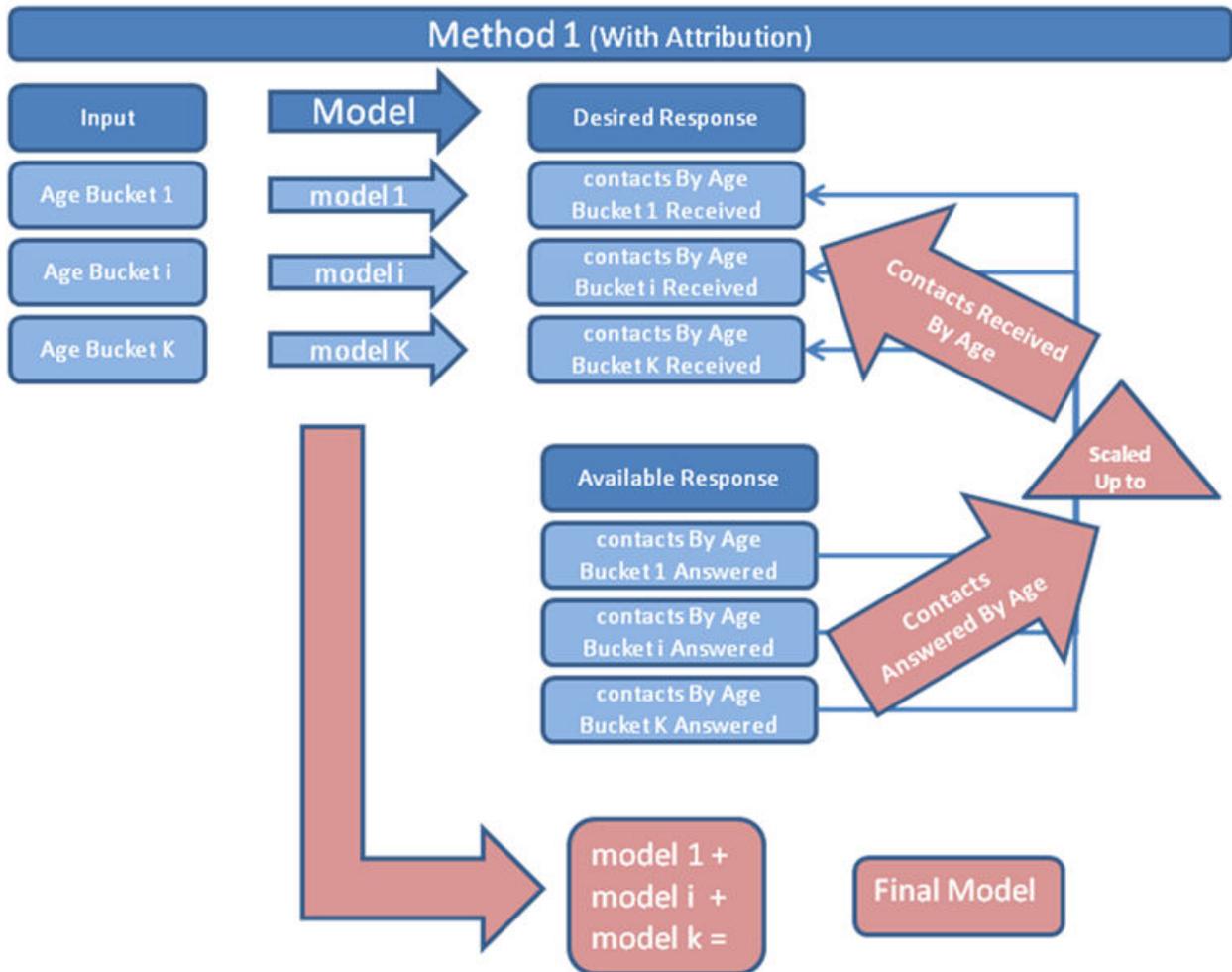

Figure 2. Method I Diagram

Steps four through seven comprise the first stage of the modeling effort, one that ignores the day of the week covariate. The models attempt to describe the relationship between the number of customers who have signed up and have not canceled their membership, i.e. the active customers, and the total contact volume answered. The first modeling attempt, step four, is a regression model testing the relationship between the inputs of active customers by age bin and output of total contact volume answered. The second modeling approach attempts to describe the relationship between the number of active customers per day and the corresponding contact volume answered per age group. This model is formed by the addition of a series of regression models made from each individual active customer age bin and its relationship to the total contacts answered from customers of that age. Each of the age bins has a corresponding regression formula which attempts to explain the relationship between customers at that age and the number of calls answered from those customers. Each of these regression formulas are

mutually exclusive and attempt to explain separate segments of variation within the overall calls answered response variable. The summation of these individual regression formulas is the final regression formula that describes the variation for all of the calls answered. Assuming the age bins are mutually exclusive but inclusive of all customer ages, the model is as follows:

β_1 * Age bin 1 customers + intercept1 = calls answered from Age bin 1

+ β_2 * Age bin 2 customers + intercept2 = calls answered from Age bin 2

+ …

+ β_8 * Age bin 8 customers + intercept8 = calls answered from Age bin 8

= β_1 * Age bin 1 customers + β_2 * Age bin 2 customers + ... + β_8 * Age bin 8 customers + interceptF = All calls answered

(where interceptF = intercept1 + intercept2 + ... + intercept8)

Figure 3. Method II Model Diagram

Once the two models are estimated in steps four and five, the next step is to compare the two models in performance using training and validation sets. The better model can then be selected for use. The next step, step seven, is to scale up the better model to account for the total contacts that were received, not just answered. The scaled up model predicts how many contacts the customers attempt, whether or not the company is able to answer them. To scale up the better model, the average daily percent difference between the total calls answered and the total calls received can be found and this

factor applied to the better regression model. To accomplish this a bi-variate regression of calls answered to calls received is be estimated, and the calculated intercept and coefficient applied to the output of the total calls answered model to scale up the model to total calls received. Stage two of the modeling effort, and step eight, is a final modeling approach that considers a day of the week factor as a covariate and estimates calls received directly. This multivariate regression has the greatest potential for explaining the variability in the response of the three models created.

4. Empirical Testing

As a precursor to the analysis, the data was first cleaned up and prepared by dealing with data anomalies. A review of the contacts variables revealed several inconsitencies within the data set. For example, 159 contacts were dated 2099 and one 2098, all of which I converted to 1999 and 1998, respectively. Furthermore, five contacts were dated 2020, which I assumed to be a switching of the last two digits as the year 2020 has not occurred yet, and modified as 2002. In addition, once age at contact was calculated for the customers by subtracting the customer sign up date from their contact date, 13,010 contact observations were found to have a negative age, meaning customer contacts were recorded before the customer signed up. These customers and their approximately 13,010 contact observations were deleted for data consistency purposes. Furthermore, 180 customers that did not have a sign up date were deleted and customers that ended up deleting their accounts were also deleted in order to limit the analysis to active customers only. For the customer's file, file number two, 553,807 customers were initially recorded. Of those, 125,412 were deleted, leaving the 428,395 active customers. In total 335,103 contact anomalies were deleted. Finally, the calls received scatterplot showed one significant outlier, the November 12th observation, that was excluded from the analysis due to its huge calls received value. In

summary, the first sign up date from the data sets was March 25th, 1999, and the last contact date was January 1st, 2013.

After preparing the data, the analysis may begin. Once the age at contact has been calculated, a distribution is appropriate to understand the data. Appendix A contains the JMP output for the distribution, which shows that there are yearly peaks, meaning roughly every 365 days from their start up date, customers are more likely to contact the company. In general, most contacts are early in a customer's life cycle, and diminish rapidly with age. The main portion of the project was the determination and set up of the customer age bins. The first step to determining the age bins was to calculate the customer age at contact for each of the contacts. After converting the data to a JMP data table, a column was added to convert the contact dates given in day and time to just date, a format that was more easily comparable to the other variables. The customer sign up date was then subtracted from this column to find the age at contact for each of the contacts. Then, the JMP distribution platform was used to find the contact age percentiles. From this distribution the following eight age bins in days were determined: (1) < 1 days, (2) 1-4 days, (3) 5-9 days, (4) 10-49 days, (5) 50-99 days, (6) 100-499 days, (7) 500-999 days, (8) 1000-4999 days; as noted in Appendix B. The second step was to create a data table with contact volume answered per day. The JMP table summary was an effective method for obtaining these figures, which were then ordered chronologically from December 1st, 2011 to November 30th, 2012.

The most challenging part of the project was also the most time intensive. A table of total active customers per day by age bin is required for the modeling techniques. Several scripting attempts to create this table were unsuccessful due to the large data set and computing limitations. Through extensive programming and using a method that avoided loops and minimized calculations via list vectors, eventually a script was prepared to generate the required data table, and this table was combined with the other necessary data tables to create the base table for the models. The JSL script for this program can be

9

found in Appendix B.

The first model was formed by using the JMP data table formed in step three. Each of the total active age bins columns were compared to the total contacts received, and a regression model was calculated. Appendix C contains the JMP output for this model. The second model was calculated as a series of regression models with each of the total active age bins per day categories compared to their respective total contacts received age bins columns. This approach generated eight regression formulas, which were then added to get one model with eight effects, one intercept, and the response of total contacts answered. Alternatively, the step four and stage two models were classical multivariate regressions.

For step six, training, validation, and testing sets were created from the master data table used in modeling by using a column of random numbers and breaking the categories into 60 percent for the training set, and 20 percent each for the remaining two sets. Steps four and five, the modeling methods, were repeated for the training and validation sets, and the residuals and normal quantile plots were examined to determine the model that performed better. The normal quantile plots showed a clear divergence between the two models, with the plots of the model that was the sum of regressions displaying non-linear tendencies. This trait and the residuals plots are evidence that the first model with each of the age bins as inputs and one output of contacts received was the better model.

With the better model selected, the next step was to scale up this model to account for all the contacts received rather than just the contacts answered. In order to accomplish this a bi-variate regression of calls answered to calls received was performed, and since this model captured a lot of variability in the response it was useful for scaling up the original model. This bi-variate regression can be found in Appendix C. From the results of this regression, the coefficient was multiplied by the original models terms and intercept was added to scale up the model to the final model of calls answered. For

comparison purposes, the three models are listed below as relationships between age bins and the predicted contacts response, but it should be noted that without including the day of the week covariate, these models underperformed.

Step 4: Basic Age Bins Model

-2621.635 + 6.860764 * Age1 +

-0.945631 * Age2 + 0.3033863 * Age3 +

-0.022427 * Age4 + 0.0051306 * Age5 +

-0.014962 * Age6 + 0.0032280 * Age7 +

0.1100191 * Age8 = Predicted Calls Answered

Step 5: Summation of a Series of Age Bin Models

1856.0 + 0.6703035 * Age1 +

0.0718025 * Age2 + 0.0642127 * Age3 +

0.0177125 * Age4 + 0.0060867 * Age5 +

0.0031718 * Age6 + 0.0038747 * Age7 +

0.0042196 * Age8 = Predicted Calls Answered

Step 7: Scaled Step 4 Model

-5457.5 + 14.7013 * Age1 +

-2.0263 * Age2 + 0.6501 * Age3 +

-0.0481 * Age4 + 0.0110 * Age5 +

-0.0321 * Age6 + 0.0069 * Age7 +

0.2357 * Age8 = Predicted Calls Received

The second stage of the modeling effort, and step eight, is a final modeling attempt that includes a day of the week covariate. Since the graph of total calls received by day of the week clearly shows a discrepancy between the days of the week versus the weekend, it is appropriate to consider day of the week as a regressor. This model directly estimates calls received from the active customer age bins and a categorical variable of day of the week. The model report is given at the end of Appendix C, and results of the models are discussed in the conclusion.

5. Conclusions

For the first modeling efforts ignoring the day of the week covariate the better of two models using different modeling approaches were compared. From the Analysis of Variance report of the model from step four, the model tests as statistically significant, with the Age1, Age2, and Age3 estimates being statistically significant with p-values less than 0.05. However, the remaining effects are not statistically significant and have p-values greater than 0.05, and the R-squared value shows that only about 65 percent of the variation in the response is explained by the model. The full report of this model can be found in Appendix C. This model was scaled up from only contacts answered to the full count of contacts received using a bi-variate regression of the two as the scaling factor. This bi-variate regression, which has a high R-squared value of 0.90 and explains a lot of the variability in calls received based on calls answered, can also be found in Appendix C.

The scaled model from step four predicts that for every additional new customer less than a day old, the predicted number of contacts received will increase by 14.7 contacts. Likewise, the somewhat unintuitive negative coefficient of the Age2 bin says that for every customer between one and four days old, the predicted contacts received will drop by 2.0 contacts. Similarly, the Age3 bin, customers between five and nine days old, contribute 0.7 contacts received per customer. Since the model ignores the day of

the week covariate and does not explain a lot of the variability in the response, the model's biggest contribution is the indication that the largest predicted effect on contacts received is the number of new customers less than a day old. New customers are the most likely to contact the company.

A far better model at explaining the data than steps four through seven is the modeling attempt from step eight that includes a day of the week covariate. The results of this model can be found at the end of Appendix C, and from this JMP report the R-squared value shows this model explains 90 percent of the variation in the response, the total calls received. In this model five of the eight age bin effects are statistically significant and every day of the week is a statistically significant predictor of calls received. The day effects are the largest, with the weekend days of Saturday and Sunday having a large negative effect on the response and the days of the business week having the largest positive effects, with Friday's being the smallest of these. Each predicts an increase of between 456-1753 calls per day. Of the significant age bins, five, six and seven have a negative effect and one and eight have a positive effect. The largest age bin effects were age bin one and six, with the coefficient from one indicating that for every new customer with the company, predicted calls received is expected to increase by approximately 3.8. The plots show the response is grouped into week and weekend days. This model is the best predictor of all the models tried.

The modeling attempts used in this paper were by no means exhaustive of all the possibilities. For instance, age could be used as covariate instead of age bins. Other opportunities for further research include a time series analysis approach to modeling where time dependencies and autocorrelations are considered. Month and either a season variable or a polynomial effect could potentially be added to the final model to explain some of the additional variation in the response, as it is evident from the graph of calls received by day in Appendix A that there is a periodicity effect in the response. Finally, the reason for contact is another potential factor to be considered for future projects.

References

Bassamboo, Archal, J. M. Harrison, and Assaf Zeevi. "Design and Control of a Large Call Center: Asymptotic Analysis of an LP-Based Method." *Operations Research* 54.3 (2006): 419-35. *INFORMS*. Web.

Brown, Lawrence, Noah Gans, Avishai Mandelbaum, Anat Sakov, Haipeng Shen, Sergey Zeltyn, and Linda Zhao. "Statistical Analysis of a Telephone Call Center: A Queuing-Science Perspective." *Journal of the American Statistical Association* 100.469 (2005): 36-50. Web.

Duder, John C., and Moshe B. Rosenwein. "Towards 'Zero Abandonments' in Call Center Performance." *European Journal of Operational Research* 135 (2001): 50-56. Web.

Keblis, Matthew F., and Maomao Chen. "Improving Customer Service Operations at Amazon.com." *Interfaces* 36.5 (2006): 433-45. *INFORMS*. Web.

Whitt, Ward. "Staffing a Call Center with Uncertain Arrival Rate and Absenteeism."*Production and Operations Management* 15.1 (2006): 88-102. Web.

Appendix A: Data

Table 1. Data Files

File 1: Contact Logger (Contacts Answered and Classified)

Columns: Customer ID, User Sign Up Date, Contact Date, Unique Contact ID

Application: Determine call volumes by age of customer

File 2: Customer Sign ups and Deletes

Columns: Customer ID, Sign Up Date, Cancel Date (if exists)

Application: Determine how many customers of particular ages are active during each day

File 3: Daily Contacts Received

Columns: Date, Day of the Week, Total Contacts Received

Application: Scaling up the final model of contacts that the support center was able to answer versus what was actually received

Age at Contact (Days)
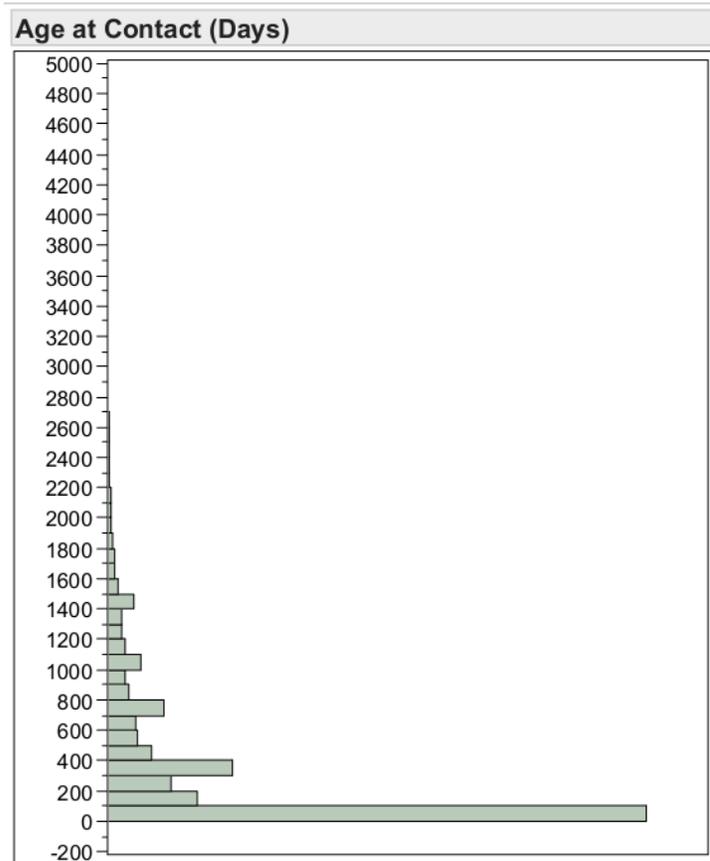
Quantiles		
100.0%	maximum	4822
99.5%		2854
97.5%		1815
90.0%		1148
75.0%	quartile	566
50.0%	median	148
25.0%	quartile	8
10.0%		1
2.5%		0
0.5%		0
0.0%	minimum	0

Summary Statistics	
Mean	390.95036
Std Dev	554.92457
Std Err Mean	0.4493185
Upper 95% Mean	391.831
Lower 95% Mean	390.06971
N	1525314

Figure 4. Contacts by Customer Age at Contact

16
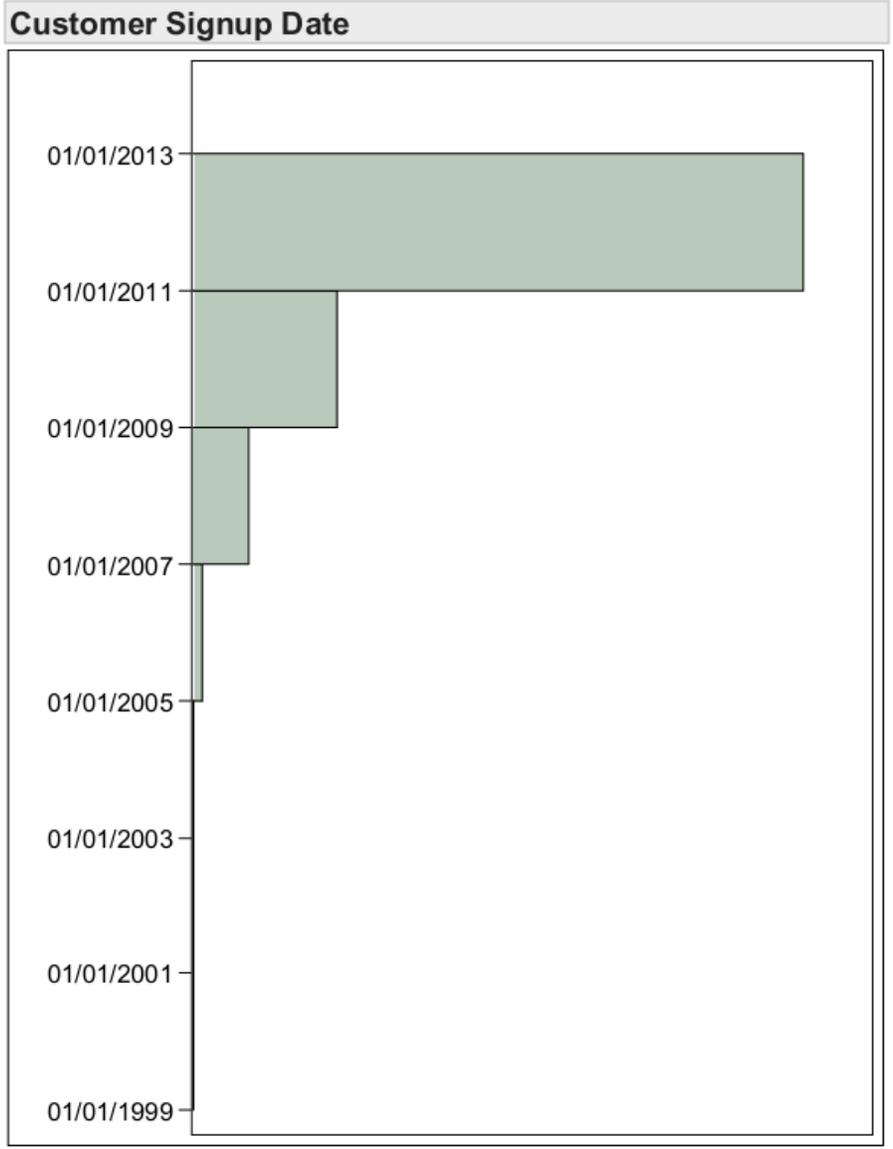
Figure 5. Customers: By Sign Up Date
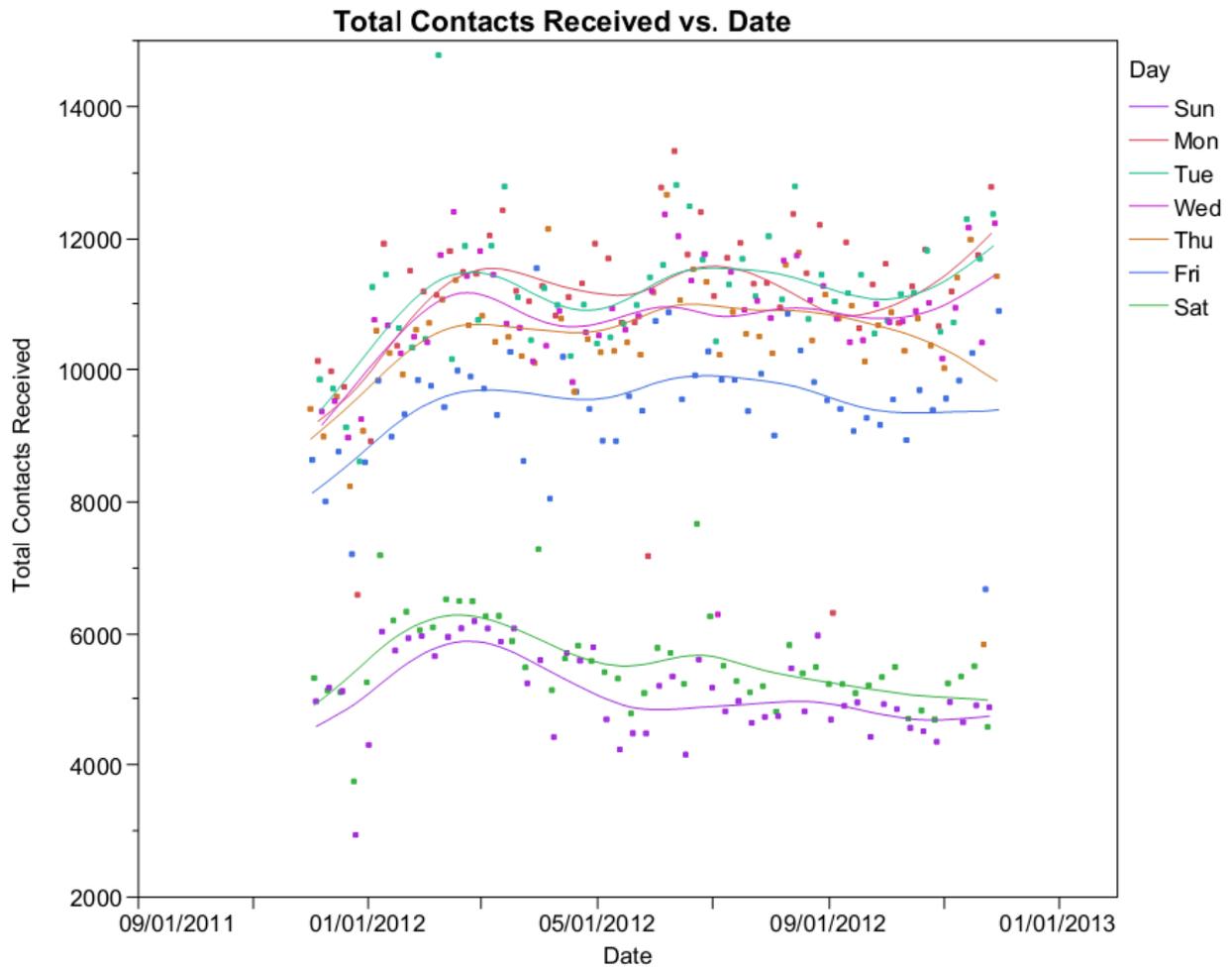
Figure 6. Contacts Received Per Day in 2012

(Weekend days are the bottom lines.)

Table 2. Selected Age Bins

1 : <1 day
2 : 1-4 days
3 : 5-9 days
4 : 10-49 days
5 : 50-99 days
6 : 100-499 days
7 : 500-999 days
8 : 1000-4999 days

Table 3. Final JSL Script for Age Bins

```
dt = open("5 Customers.jmp");

// Create age bins
new column("Age1", formula(0));
new column("Age2", formula(0));
new column("Age3", formula(0));
new column("Age4", formula(0));
new column("Age5", formula(0));
new column("Age6", formula(0));
new column("Age7", formula(0));
new column("Age8", formula(0));

// resets columns
Age1<<Eval Formula;
Age2<<Eval Formula;
Age3<<Eval Formula;
Age4<<Eval Formula;
Age5<<Eval Formula;
Age6<<Eval Formula;
Age7<<Eval Formula;
Age8<<Eval Formula;

// intialize variables
last_date = 5000;
age_days = {};
num_cust = 0;
new_cust = 0;
age_8 = 0;

For(i = 1, i < last_date, i++,
```

```
/* increment age of customers one day */
If ( i != 1 & :Date[i] != :Date[i-1], age_days++; new_cust = 0);

/* add a new customer */
If (!IsMissing(:customerid[i]), Insert Into(age_days, 0); num_cust++; new_cust++);

/* update age bins */
If ( i != 1 & :Date[i] == :Date[i-1],
        :Age1[i] = new_cust;
        :Age2[i] = :Age2[i-1];
        :Age3[i] = :Age3[i-1];
        :Age4[i] = :Age4[i-1];
        :Age5[i] = :Age5[i-1];
        :Age6[i] = :Age6[i-1];
        :Age7[i] = :Age7[i-1];
        :Age8[i] = :Age8[i-1],

        For (k=(age_8+1), k<=num_cust, k++,
                If(age_days[k] < 1, :Age1[i]++, age_days[k] < 5, :Age2[i]++,
                age_days[k] < 10, :Age3[i]++, age_days[k] < 50, :Age4[i]++,
                age_days[k] < 100, :Age5[i]++, age_days[k] < 500, :Age6[i]++,
                age_days[k] < 1000, :Age7[i]++, age_8++)
                );
        Age8[i] = age_8;
        );
);
```
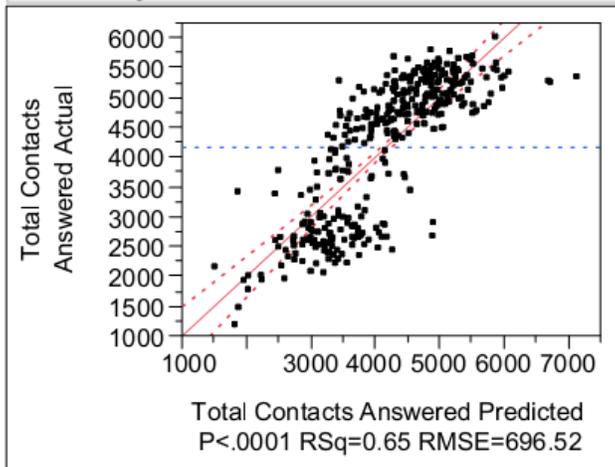
Appendix C: Model Reports

Response Total Contacts Answered

Whole Model

Actual by Predicted Plot

Total Contacts Answered Predicted
P<.0001 RSq=0.65 RMSE=696.52

Summary of Fit

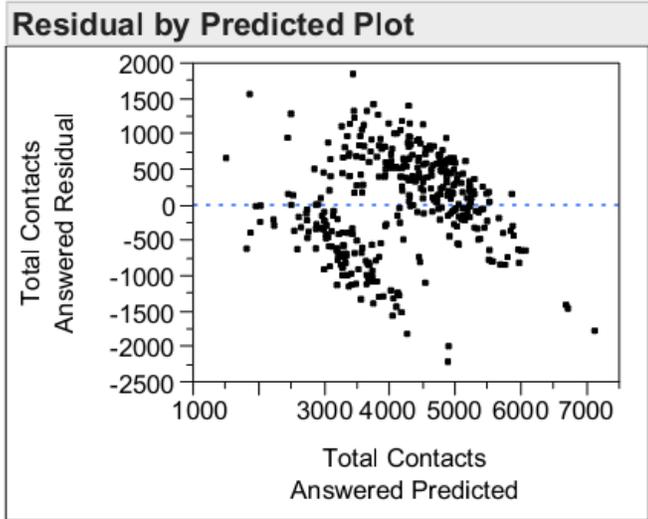|	
RSquare	0.648958
RSquare Adj	0.64107
Root Mean Square Error	696.5177
Mean of Response	4165.786
Observations (or Sum Wgts)	365

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	319280389	39910049	82.2655
Error	356	172708758	485136.96	Prob > F
C. Total	364	491989147		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2621.635	1441.167	-1.82	0.0697
Age1	6.860764	0.316053	21.71	<.0001*
Age2	-0.945631	0.112283	-8.42	<.0001*
Age3	0.3033863	0.103361	2.94	0.0035*
Age4	-0.022427	0.020558	-1.09	0.2761
Age5	0.0051306	0.016267	0.32	0.7527
Age6	-0.014962	0.011378	-1.32	0.1893
Age7	0.003228	0.03881	0.08	0.9338
Age8	0.1100191	0.092322	1.19	0.2342

21

Response Total Contacts Answered

Whole Model

Residual by Predicted Plot

Response Total Contacts Received

Whole Model

Actual by Predicted Plot

Total Contacts Received Predicted
P<.0001 RSq=0.90 RMSE=820.48
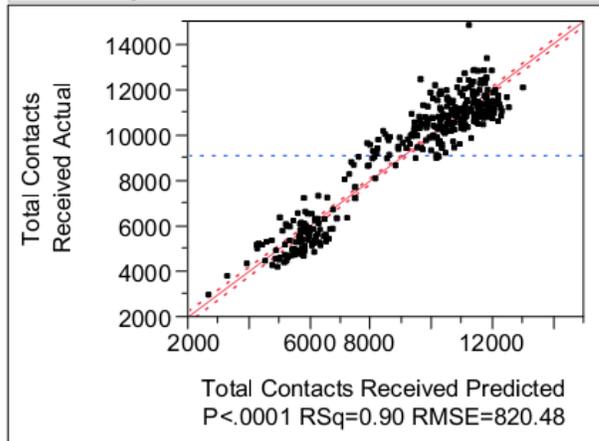
Summary of Fit

RSquare	0.902386
RSquare Adj	0.902117
Root Mean Square Error	820.4799
Mean of Response	9086.592
Observations (or Sum Wgts)	365

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2259022428	2.259e+9	3355.712
Error	363	244366996	673187.32	Prob > F
C. Total	364	2503389424		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	338	235931951	698024	2.0688
Pure Error	25	8435046	337402	Prob > F
Total Error	363	244366996		0.0158*
				Max RSq
				0.9966

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	160.12697	159.9671	1.00	0.3175
Total Contacts Answered	2.1428043	0.03699	57.93	<.0001*

23

Response Total Contacts Received

Whole Model

Actual by Predicted Plot

Total Contacts Received Predicted
P<.0001 RSq=0.90 RMSE=833.5
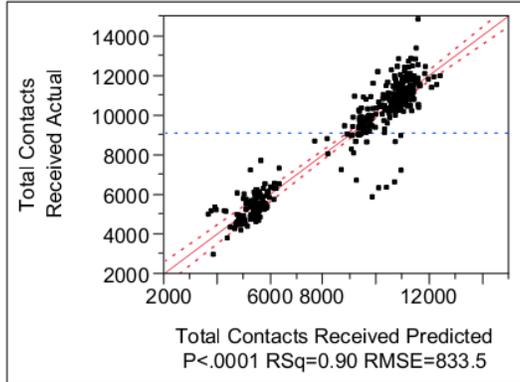
Summary of Fit

RSquare	0.902871
RSquare Adj	0.898985
Root Mean Square Error	833.5001
Mean of Response	9086.592
Observations (or Sum Wgts)	365

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	14	2260236584	161445470	232.3885
Error	350	243152840	694722.4	Prob > F
C. Total	364	2503389424		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1292.856	1761.043	-0.73	0.4634
Age1	3.8397637	0.612952	6.26	<.0001*
Age2	-0.084914	0.198797	-0.43	0.6695
Age3	-0.056537	0.139334	-0.41	0.6852
Age4	0.0398571	0.025602	1.56	0.1204
Age5	-0.076871	0.019476	-3.95	<.0001*
Age6	-0.044758	0.013764	-3.25	0.0013*
Age7	-0.126731	0.046587	-2.72	0.0068*
Age8	0.421028	0.111547	3.77	0.0002*
Day[Fri]	456.14102	124.597	3.66	0.0003*
Day[Mon]	1639.8497	135.7397	12.08	<.0001*
Day[Sat]	-3034.323	149.4664	-20.30	<.0001*
Day[Sun]	-3521.328	132.845	-26.51	<.0001*
Day[Thu]	1267.7075	108.6391	11.67	<.0001*
Day[Tue]	1753.8887	132.1066	13.28	<.0001*

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Age1	1	1	27262566	39.2424	<.0001*
Age2	1	1	126751	0.1824	0.6695
Age3	1	1	114384	0.1646	0.6852

24

Response Total Contacts Received

Whole Model

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Age6	1	1	7346390	10.5746	0.0013*
Age7	1	1	5141112	7.4002	0.0068*
Age8	1	1	9897335	14.2465	0.0002*
Day	6	6	671951139	161.2038	<.0001*

Residual by Predicted Plot
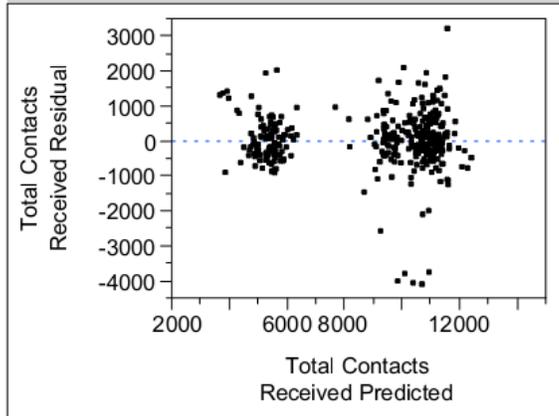
Day

Leverage Plot
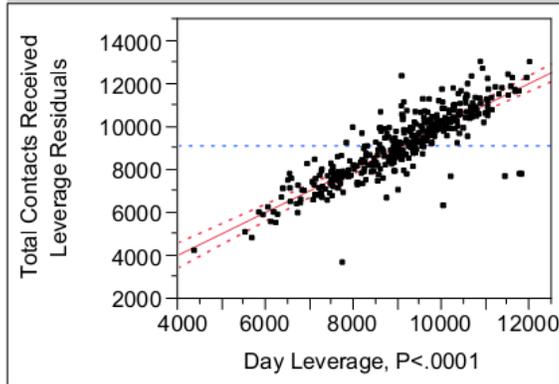
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Fri	9542.503	131.60165	9478.3
Mon	10726.211	143.07790	11081.7
Sat	6052.039	155.51547	5547.5
Sun	5565.033	139.71944	5085.1
Thu	10354.069	116.81179	10497.8
Tue	10840.250	139.35524	11144.1
Wed	10524.426	135.26238	10775.3