SAS Capstone Project

Harshith Reddy V
Jigsaw ID: JIG15137

Question 1. What are the top five factors driving likelihood of churn at Mobicom?
Solution: The top 5 factors that are driving the likelihood of churn at Mobicom are,

a) Average monthly minutes of use over the life of the customer (**Avgmou**)
b) Number of days (age) of current equipment (**Eqpdays**)
c) Mean number of roaming calls (**roam_Mean**)
d) Mean overage minutes of use (**ovrmou_Mean**)
e) Range of number of minutes of use (**mou_Range**)

Question 2. Validation of survey findings. a) Whether "cost and billing" and "network and service quality" are important factors influencing churn behaviour. b) Are data usage connectivity issues turning out to be costly? In other words, is it leading to churn?
Solution: Validation of survey findings

a) **Cost and billing** - Cost and billing does influence churn at Mobicom. Mean overage minutes of use is one of the important factors driving likelihood of churn (ovrmou_Mean) as per the logistic regression model which leads to excess charges in billing. It can also be noticed that customers on non optimal rate plans (nonop_rateplan) have a higher probability of churn and it is one of the significant variables influencing churn. **Network and service quality** - Network and service quality is an important factor influencing churn behaviour according to the logistic regression model. If the mean number of dropped or blocked data and voice calls (drop_blk_Mean) variable is higher, the probability of churn increases.

b) **Data usage connectivity** issues are indirectly turning out to be costly since they are adding up to the mean number of dropped or blocked **data** and voice calls (drop_blk_Mean) which is causing churn at Mobicom.

Question 3. Would you recommend rate plan migration as a proactive retention strategy?
Solution: Rate plan migration is a good proactive retention strategy because the customers with non optimal rate plans (nonop_rateplan) have a higher probability of influencing churn rates at mobicom indicating that rate plan migration is a good strategy to move subscribers from non optimal plans to optimal plans which reduces the possibility of churn. It is also noticed that customers with account spending limit (asl_flag_y) have a lower probability of churn.

Question 4. What would be your recommendation on how to use this churn model for prioritisation of customers for a proactive retention campaigns in the future?
Solution: Proactive retention campaigns should prioritise the following customers,
a) Subscribers from NW_rockymountain and South Florida regions.
b) Customers on non optimal rate plans
c) Customers with higher number of unique subscribers in the household
d) Subscribers from Asian, Hispanic and Arab ethnicity.
e) Subscribers with a higher number of days (age) of current equipment (Eqpdays)

Question 5. What would be the target segments for proactive retention campaigns? Falling ARPU forecast is also a concern and therefore, Mobicom would like to save their high revenue customers besides managing churn. Given a budget constraint of a contact list of 20% of the subscriber pool, which subscribers should prioritized if "revenue saves" is also a priority besides controlling churn. In other words, controlling churn is the primary objective and revenue saves is the secondary objective.
Solution: From a contact list of the first 20% of the given subscriber pool the following segments are to be targeted for proactive retention campaigns,
a) Subscribers with a higher number of days (age) of current equipment (Eqpdays)
b) Subscribers with higher mean monthly revenue (Rev_mean)
c) Subscribers with a high Average monthly minutes of use over the life of the customer (Avgmou)
d) Customers whose handset price is less then 60 (Hnd_price_60)
e) Customers with high Range of number of minutes of use (mou_range)
f) Customers with higher Mean number of dropped or blocked calls (drop_blk_mean)

Logistic regression model for Training dataset

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-1.1426	0.1865	37.5180	<.0001	
mou_Mean	1	-0.00086	0.000112	59.7039	<.0001	-0.2555
totmrc_Mean	1	-0.00434	0.00124	12.3381	0.0004	-0.0577
rev_Range	1	0.000583	0.000491	1.4075	0.2355	0.0236
mou_Range	1	0.000227	0.000080	8.1490	0.0043	0.0538
drop_blk_Mean	1	0.00600	0.00165	13.2388	0.0003	0.0509
months	1	-0.0179	0.00378	22.3095	<.0001	-0.0957
eqpdays	1	0.00119	0.000144	68.5950	<.0001	0.1671
custcare_Mean	1	-0.0160	0.00615	6.7623	0.0093	-0.0494
ovrmou_Mean	1	0.00152	0.000354	18.5206	<.0001	0.0805
avgmou	1	0.000620	0.000110	31.5804	<.0001	0.1536
models	1	0.1216	0.0382	10.1461	0.0014	0.0606
uniqsubs	1	0.1077	0.0233	21.3038	<.0001	0.0520
roam_Mean	1	0.0123	0.00395	9.6439	0.0019	0.0336
avgcrdtscore	1	-0.2501	0.0541	21.3660	<.0001	-0.0595
datovr_Range	1	-0.00733	0.00346	4.4866	0.0342	-0.0349
Asian	1	0.2910	0.1019	8.1547	0.0043	0.0311
AfroAmerican	1	-0.4032	0.1140	12.5215	0.0004	-0.0471
foreign	1	-0.2007	0.0959	4.3792	0.0364	-0.0261
hndwebcap_wcmb	1	-0.3686	0.0765	23.2100	<.0001	-0.0846
hndwebcap_wc	1	-0.1796	0.0816	4.8443	0.0277	-0.0331
mailresp_new	1	-0.1682	0.0460	13.3897	0.0003	-0.0451
single	1	-0.1057	0.0496	4.5395	0.0331	-0.0252
mtrcycle_new	1	0.2578	0.1294	3.9718	0.0463	0.0243
new_hndset	1	-0.1440	0.0643	5.0215	0.0250	-0.0274
avgrev_30	1	-0.2081	0.0761	7.4719	0.0063	-0.0349
agegr60_bkt	1	-0.1948	0.0900	4.6869	0.0304	-0.0267
hnd_price_200	1	-0.2265	0.0777	8.4903	0.0036	-0.0384

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mou_Mean	0.999	0.999	0.999
totmrc_Mean	0.996	0.993	0.998
rev_Range	1.001	1.000	1.002
mou_Range	1.000	1.000	1.000
drop_blk_Mean	1.006	1.003	1.009
months	0.982	0.975	0.990
eqpdays	1.001	1.001	1.001
custcare_Mean	0.984	0.972	0.996
ovrmou_Mean	1.002	1.001	1.002
avgmou	1.001	1.000	1.001
models	1.129	1.048	1.217
uniqsubs	1.114	1.064	1.166
roam_Mean	1.012	1.005	1.020
avgcrdtscore	0.779	0.700	0.866
datovr_Range	0.993	0.986	0.999
Asian	1.338	1.096	1.633
AfroAmerican	0.668	0.534	0.835
foreign	0.818	0.678	0.987
hndwebcap_wcmb	0.692	0.595	0.804
hndwebcap_wc	0.836	0.712	0.981
mailresp_new	0.845	0.772	0.925
single	0.900	0.816	0.992
mtrcycle_new	1.294	1.004	1.668
new_hndset	0.866	0.763	0.982
avgrev_30	0.812	0.700	0.943
agegr60_bkt	0.823	0.690	0.982
hnd_price_200	0.797	0.685	0.929

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	63.5	Somers' D	0.271
Percent Discordant	36.5	Gamma	0.271
Percent Tied	0.0	Tau-a	0.098
Pairs	31615792	c	0.635

Partition for the Hosmer and Lemeshow Test					
		churn = 1		churn = 0	
Group	Total	Observed	Expected	Observed	Expected
1	1322	151	146.64	1171	1175.36
2	1322	202	203.29	1120	1118.71
3	1322	217	237.18	1105	1084.82
4	1322	238	262.66	1084	1059.34
5	1322	276	287.53	1046	1034.47
6	1322	322	313.53	1000	1008.47
7	1322	347	341.38	975	980.62
8	1322	391	375.88	931	946.12
9	1322	459	426.32	863	895.68
10	1324	531	539.60	793	784.40

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
10.9326	8	0.2055

Logistic regression model for Validation dataset

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-1.8431	0.1124	213.8781	<.0001	
mou_Mean	1	-0.00074	0.000100	54.9441	<.0001	-0.2218
totmrc_Mean	1	-0.00746	0.00163	20.9772	<.0001	-0.0946
rev_Range	1	-0.00040	0.000543	0.5470	0.4595	-0.0165
mou_Range	1	0.000167	0.000072	5.3966	0.0202	0.0423
drop_blk_Mean	1	0.00452	0.00146	9.6536	0.0019	0.0426
months	1	-0.0290	0.00619	22.0240	<.0001	-0.0987
totcalls	1	0.000019	0.000026	0.5467	0.4597	0.0211
eqpdays	1	0.00149	0.000222	45.1081	<.0001	0.1285
iwylis_vce_Mean	1	-0.00456	0.00179	6.4800	0.0109	-0.0408
rev_Mean	1	0.00334	0.00114	8.5408	0.0035	0.0796
models	1	0.1621	0.0447	13.1557	0.0003	0.0599
uniqsubs	1	0.1233	0.0262	22.1260	<.0001	0.0592
adjmou	1	0.000064	0.000012	29.3069	<.0001	0.1870
los_angeles	1	-0.1628	0.0836	3.7953	0.0514	-0.0249
NW_rockymountain	1	0.2287	0.1038	4.8604	0.0275	0.0243
south_florida	1	0.2805	0.1259	4.9669	0.0258	0.0245
Asl_flag_y	1	-0.3063	0.0553	30.7220	<.0001	-0.0742
Hispanic	1	0.1559	0.0613	6.4610	0.0110	0.0299
Asian	1	0.3326	0.1004	10.9644	0.0009	0.0363
Arab	1	0.4951	0.2014	6.0427	0.0140	0.0258
AfroAmerican	1	-0.5173	0.1128	21.0216	<.0001	-0.0629
avgrev_30	1	-0.2000	0.0817	5.9959	0.0143	-0.0342
avgrev_70	1	0.1130	0.0496	5.1812	0.0228	0.0271
age25_bkt	1	0.2090	0.0446	21.9914	<.0001	0.0553
hnd_price_60	1	0.1841	0.0534	11.8746	0.0006	0.0413
hnd_price_200	1	-0.2522	0.0668	14.2408	0.0002	-0.0475
actvsubs_3	1	-0.2600	0.1209	4.6296	0.0314	-0.0272
Nonop_rateplan	1	0.0259	0.00990	6.8308	0.0090	0.0434

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mou_Mean	0.999	0.999	0.999
totmrc_Mean	0.993	0.989	0.996
rev_Range	1.000	0.999	1.001
mou_Range	1.000	1.000	1.000
drop_blk_Mean	1.005	1.002	1.007
months	0.971	0.960	0.983
totcalls	1.000	1.000	1.000
eqpdays	1.001	1.001	1.002
iwylis_vce_Mean	0.995	0.992	0.999
rev_Mean	1.003	1.001	1.006
models	1.176	1.077	1.284
uniqsubs	1.131	1.075	1.191
adjmou	1.000	1.000	1.000
los_angeles	0.850	0.721	1.001
NW_rockymountain	1.257	1.026	1.540
south_florida	1.324	1.034	1.694
Asl_flag_y	0.736	0.661	0.820
Hispanic	1.169	1.036	1.318
Asian	1.395	1.145	1.698
Arab	1.641	1.106	2.435
AfroAmerican	0.596	0.478	0.744
avgrev_30	0.819	0.698	0.961
avgrev_70	1.120	1.016	1.234
age25_bkt	1.232	1.129	1.345
hnd_price_60	1.202	1.083	1.335
hnd_price_200	0.777	0.682	0.886
actvsubs_3	0.771	0.608	0.977
Nonop_rateplan	1.026	1.006	1.046

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	62.9	Somers' D	0.257
Percent Discordant	37.1	Gamma	0.257
Percent Tied	0.0	Tau-a	0.092
Pairs	31124112	c	0.629

Partition for the Hosmer and Lemeshow Test					
		churn = 1		churn = 0	
Group	Total	Observed	Expected	Observed	Expected
1	1322	156	156.61	1166	1165.39
2	1322	206	210.32	1116	1111.68
3	1322	199	239.67	1123	1082.33
4	1322	235	263.27	1087	1058.73
5	1323	292	284.69	1031	1038.31
6	1322	312	305.58	1010	1016.42
7	1322	335	329.95	987	992.05
8	1322	402	358.50	920	963.50
9	1322	409	400.99	913	921.01
10	1323	518	514.43	805	808.57

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
20.3588	8	0.0091

Logistic regression model for Complete dataset

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-1.4932	0.0892	280.1443	<.0001	
mou_Mean	1	-0.00091	0.000083	121.3997	<.0001	-0.2697
totmrc_Mean	1	-0.00303	0.000930	10.6287	0.0011	-0.0400
rev_Range	1	0.000169	0.000363	0.2183	0.6403	0.0105
mou_Range	1	0.000216	0.000056	14.8684	0.0001	0.0604
change_mou	1	-0.00017	0.000061	8.0007	0.0047	-0.0305
drop_blk_Mean	1	0.00380	0.00117	10.4711	0.0012	0.0322
months	1	-0.0180	0.00269	44.7808	<.0001	-0.0965
eqpdays	1	0.00122	0.000103	140.3153	<.0001	0.1696
iwylis_vce_Mean	1	-0.00280	0.00125	5.0043	0.0253	-0.0249
ovrmou_Mean	1	0.00115	0.000270	18.2354	<.0001	0.0609
avgmou	1	0.000773	0.000084	84.0294	<.0001	0.1913
models	1	0.1131	0.0269	17.7139	<.0001	0.0571
uniqsubs	1	0.1220	0.0192	40.4874	<.0001	0.0581
roam_Mean	1	0.00591	0.00246	7.8791	0.0050	0.0906
avgcrdtscore	1	-0.1783	0.0424	17.6640	<.0001	-0.0426
NW_rockymountain	1	0.2203	0.0718	9.4191	0.0021	0.0238
south_florida	1	0.1907	0.0823	5.3736	0.0204	0.0184
Asl_flag_y	1	-0.2555	0.0556	21.1316	<.0001	-0.0503
Asian_Non	1	0.3165	0.1199	6.9705	0.0083	0.0202
South_european	1	0.3247	0.1570	4.2789	0.0386	0.0161
Hispanic	1	0.1497	0.0442	11.4966	0.0007	0.0281
Asian	1	0.3733	0.0720	26.8758	<.0001	0.0397
Arab	1	0.3104	0.1407	4.8666	0.0274	0.0170
AfroAmerican	1	-0.3536	0.0796	19.7377	<.0001	-0.0419
hndwebcap_unkw	1	0.1336	0.0533	6.2827	0.0122	0.0215
mailresp_new	1	-0.0825	0.0355	5.3989	0.0201	-0.0220
town	1	0.0891	0.0414	4.6439	0.0312	0.0174
new_hndset	1	-0.1881	0.0459	16.8021	<.0001	-0.0356
avgrev_30	1	-0.1583	0.0554	8.1616	0.0043	-0.0263
avgrev_70	1	0.0742	0.0362	4.2039	0.0403	0.0177
avgrev_101	1	-0.0990	0.0716	1.9097	0.1670	-0.0162
age25_bkt	1	0.1683	0.0366	21.1941	<.0001	0.0427
agegr60_bkt	1	-0.1767	0.0641	7.5979	0.0058	-0.0241
hnd_price_60	1	0.1359	0.0392	11.9977	0.0005	0.0353
hnd_price_200	1	-0.2047	0.0541	14.2958	0.0002	-0.0352
actvsubs_3	1	-0.2378	0.0840	8.0076	0.0047	-0.0255
actvsubs_5	1	-0.7827	0.3211	5.9431	0.0148	-0.0222
Nonop_rateplan	1	0.0281	0.00831	11.4720	0.0007	0.0391

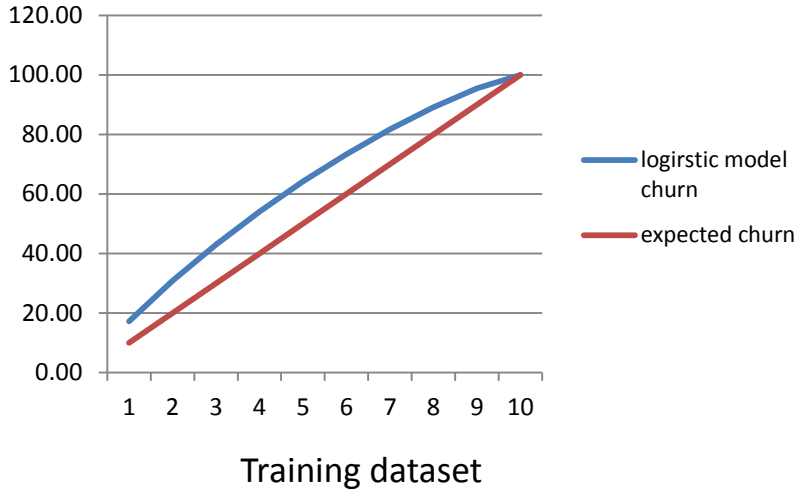
Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mou_Mean	0.999	0.999	0.999
totmrc_Mean	0.997	0.995	0.999
rev_Range	1.000	0.999	1.001
mou_Range	1.000	1.000	1.000
change_mou	1.000	1.000	1.000
drop_blk_Mean	1.004	1.001	1.006
months	0.982	0.977	0.987
eqpdays	1.001	1.001	1.001
iwylis_vce_Mean	0.997	0.995	1.000
ovrmou_Mean	1.001	1.001	1.002
avgmou	1.001	1.001	1.001
models	1.120	1.062	1.180
uniqsubs	1.130	1.088	1.173
roam_Mean	1.007	1.002	1.012
avgcrdtscore	0.837	0.770	0.909
NW_rockymountain	1.246	1.083	1.435
south_florida	1.210	1.030	1.422
Asl_flag_y	0.774	0.695	0.864
Asian_Non	1.372	1.085	1.736
South_european	1.384	1.017	1.882
Hispanic	1.161	1.065	1.266
Asian	1.453	1.261	1.673
Arab	1.364	1.035	1.797
AfroAmerican	0.702	0.601	0.821
hndwebcap_unkw	1.143	1.030	1.269
mailresp_new	0.921	0.859	0.987
town	1.093	1.008	1.186
new_hndset	0.829	0.757	0.907
avgrev_30	0.854	0.766	0.952
avgrev_70	1.077	1.003	1.156
avgrev_101	0.906	0.787	1.042
age25_bkt	1.183	1.101	1.271
agegr60_bkt	0.838	0.739	0.950
hnd_price_60	1.146	1.061	1.237
hnd_price_200	0.815	0.733	0.906
actvsubs_3	0.788	0.669	0.930
actvsubs_5	0.457	0.244	0.858
Nonop_rateplan	1.029	1.012	1.045

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	63.3	Somers' D	0.266
Percent Discordant	36.7	Gamma	0.266
Percent Tied	0.0	Tau-a	0.097
Pairs	127073184	c	0.633

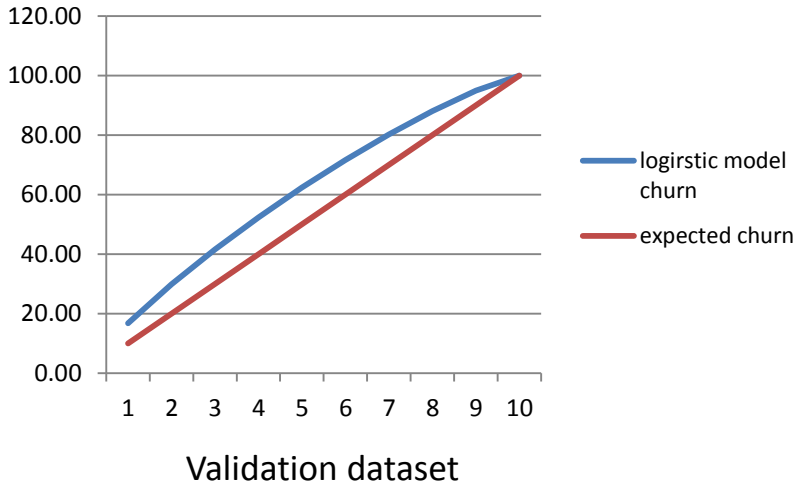
Partition for the Hosmer and Lemeshow Test					
		churn = 1		churn = 0	
Group	Total	Observed	Expected	Observed	Expected
1	2644	295	293.38	2349	2350.62
2	2644	384	407.30	2260	2236.70
3	2644	474	475.97	2170	2168.03
4	2644	497	532.89	2147	2111.11
5	2644	580	584.21	2064	2059.79
6	2644	633	637.31	2011	2006.69
7	2644	719	695.56	1925	1948.44
8	2644	797	764.08	1847	1879.92
9	2644	905	854.72	1739	1789.28
10	2648	1028	1066.58	1620	1581.42

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
14.4751	8	0.0702

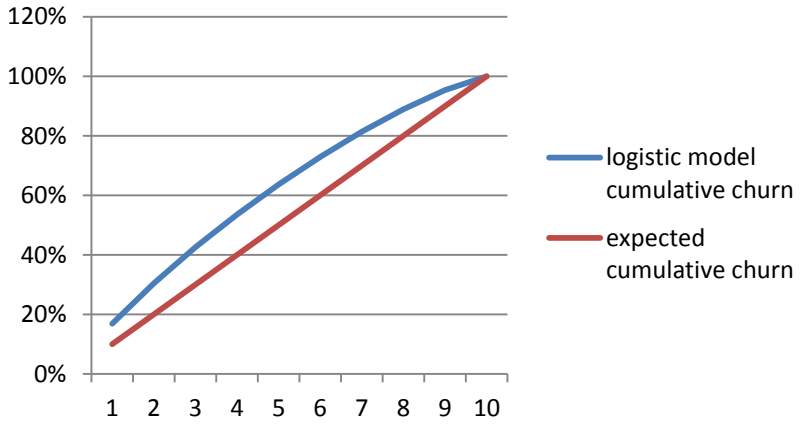
Gain charts



Training dataset



Validation dataset



Complete dataset

Logistic regression model for 20% of subscriber list

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-1.3081	0.1663	61.8889	<.0001	
mou_Mean	1	-0.00104	0.000174	35.5582	<.0001	-0.3168
totmrc_Mean	1	-0.00923	0.00198	21.7352	<.0001	-0.1225
mou_Range	1	0.000240	0.000102	5.5142	0.0189	0.0975
drop_blk_Mean	1	0.00828	0.00255	10.5644	0.0012	0.0706
months	1	-0.0183	0.00582	9.8789	0.0017	-0.0974
eqpdays	1	0.00133	0.000224	35.5940	<.0001	0.1839
rev_Mean	1	0.00574	0.00126	20.7367	<.0001	0.2177
avgmou	1	0.000815	0.000192	18.1007	<.0001	0.1984
models	1	0.1319	0.0574	5.2723	0.0217	0.0666
avgcrdtscore	1	-0.3522	0.0820	18.4580	<.0001	-0.0842
recv_sms_Mean	1	0.0614	0.0331	3.4386	0.0637	0.0306
datovr_Range	1	-0.0132	0.00921	2.0477	0.1524	-0.0791
South_european	1	0.8464	0.3278	6.6674	0.0098	0.0409
married_unkw	1	0.2588	0.0684	14.3406	0.0002	0.0690
new_hndset	1	-0.3192	0.1016	9.8663	0.0017	-0.0598
avgrev_70	1	0.1682	0.0801	4.4077	0.0358	0.0398
avgrev_101	1	-0.3390	0.1671	4.1177	0.0424	-0.0549
hnd_price_60	1	0.1943	0.0827	5.5191	0.0188	0.0501
actvsubs_4	1	0.6882	0.3000	5.2633	0.0218	0.0367

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
mou_Mean	0.999	0.999	0.999
totmrc_Mean	0.991	0.987	0.995
mou_Range	1.000	1.000	1.000
drop_blk_Mean	1.008	1.003	1.013
months	0.982	0.971	0.993
eqpdays	1.001	1.001	1.002
rev_Mean	1.006	1.003	1.008
avgmou	1.001	1.000	1.001
models	1.141	1.019	1.277
avgcrdtscore	0.703	0.599	0.826
recv_sms_Mean	1.063	0.997	1.135
datovr_Range	0.987	0.969	1.005
South_european	2.331	1.226	4.432
married_unkw	1.295	1.133	1.481
new_hndset	0.727	0.595	0.887
avgrev_70	1.183	1.011	1.384
avgrev_101	0.712	0.514	0.988
hnd_price_60	1.214	1.033	1.428
actvsubs_4	1.990	1.105	3.583

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	64.2	Somers' D	0.283
Percent Discordant	35.8	Gamma	0.283
Percent Tied	0.0	Tau-a	0.104
Pairs	5128790	c	0.642

Partition for the Hosmer and Lemeshow Test					
		churn = 1		churn = 0	
Group	Total	Observed	Expected	Observed	Expected
1	529	61	58.87	468	470.13
2	529	72	82.82	457	446.18
3	529	85	96.37	444	432.63
4	529	103	106.68	426	422.32
5	529	111	117.05	418	411.95
6	529	140	127.24	389	401.76
7	529	155	139.91	374	389.09
8	529	151	154.69	378	374.31
9	529	194	174.91	335	354.09
10	528	207	220.45	321	307.55

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
12.5088	8	0.1299