Hopmonk - Customer Life Time Value

CUTE03

By Sheik Basha

Agenda

- Problem Statement
- Data Set
- Plots
- Linear Regression Model & Error Metrics
- Random Forest Model & Error Metrics
- CART model & Error Metrics
- XG Boost Model & Error metrics
- SVM Model & Error Metrics
- Comparison of Best Models

Problem Statement

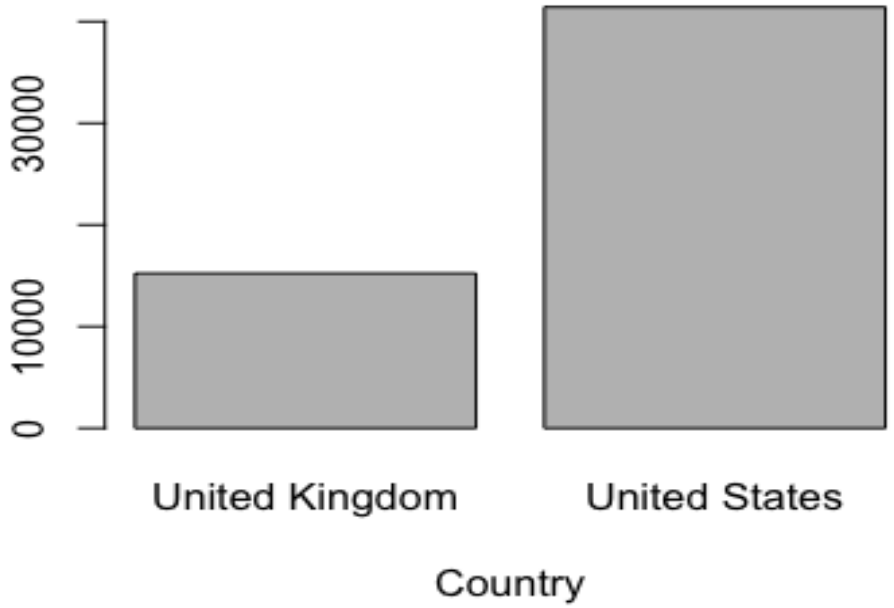
- Identify and target only specific customer groups that will turn out to be most profitable.
- “Retain and Acquire more customers and increase profits”
- Minimize operation costs
- Generate Insights from customer data to segment the customers

Data Set

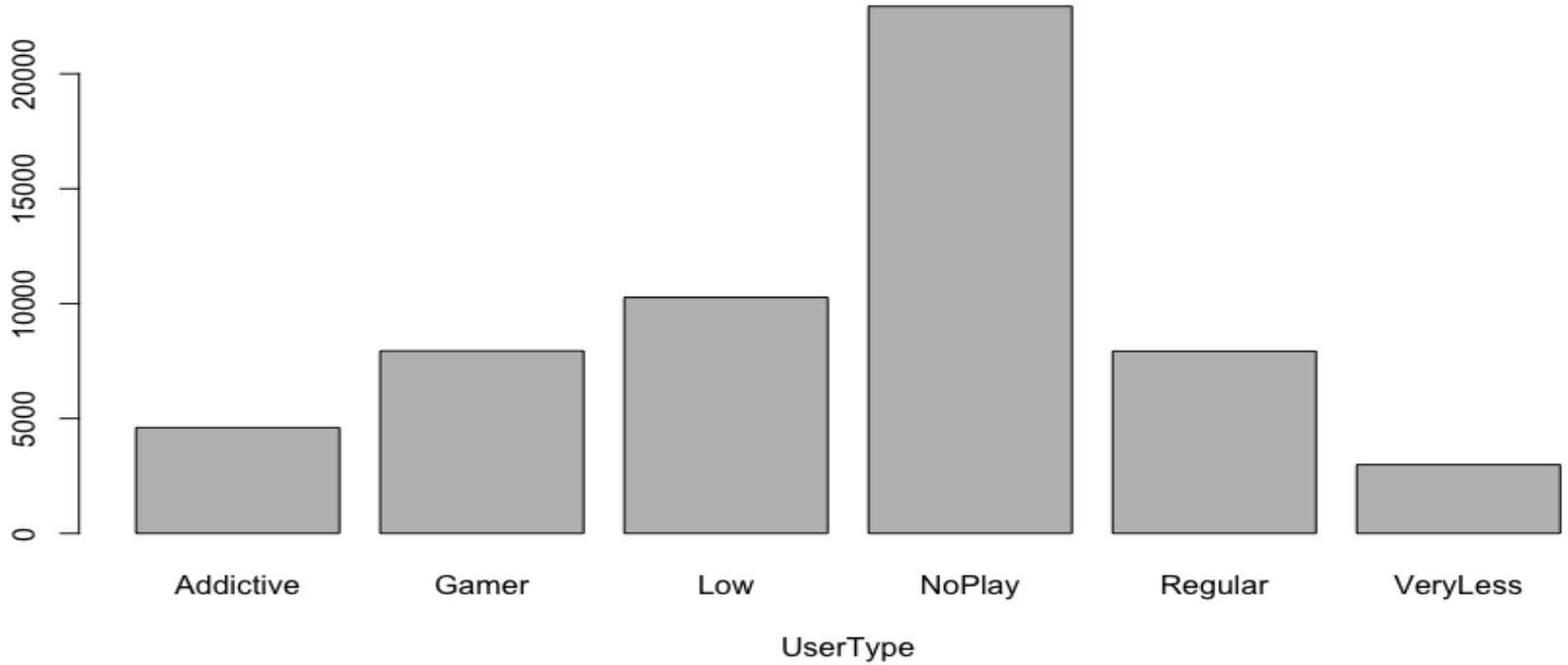
- 121 variables from the data
  - Easy: 4
  - Medium: 79 (Used functions)
  - Difficult: 38 (Derived from data given in files)
- Extracted 27 features using Step Aic
  - Reduce in dimensionality
  - Correlated variables have been removed

* File to see all variables: Group13-Cute2-ExtractedVariables.txt

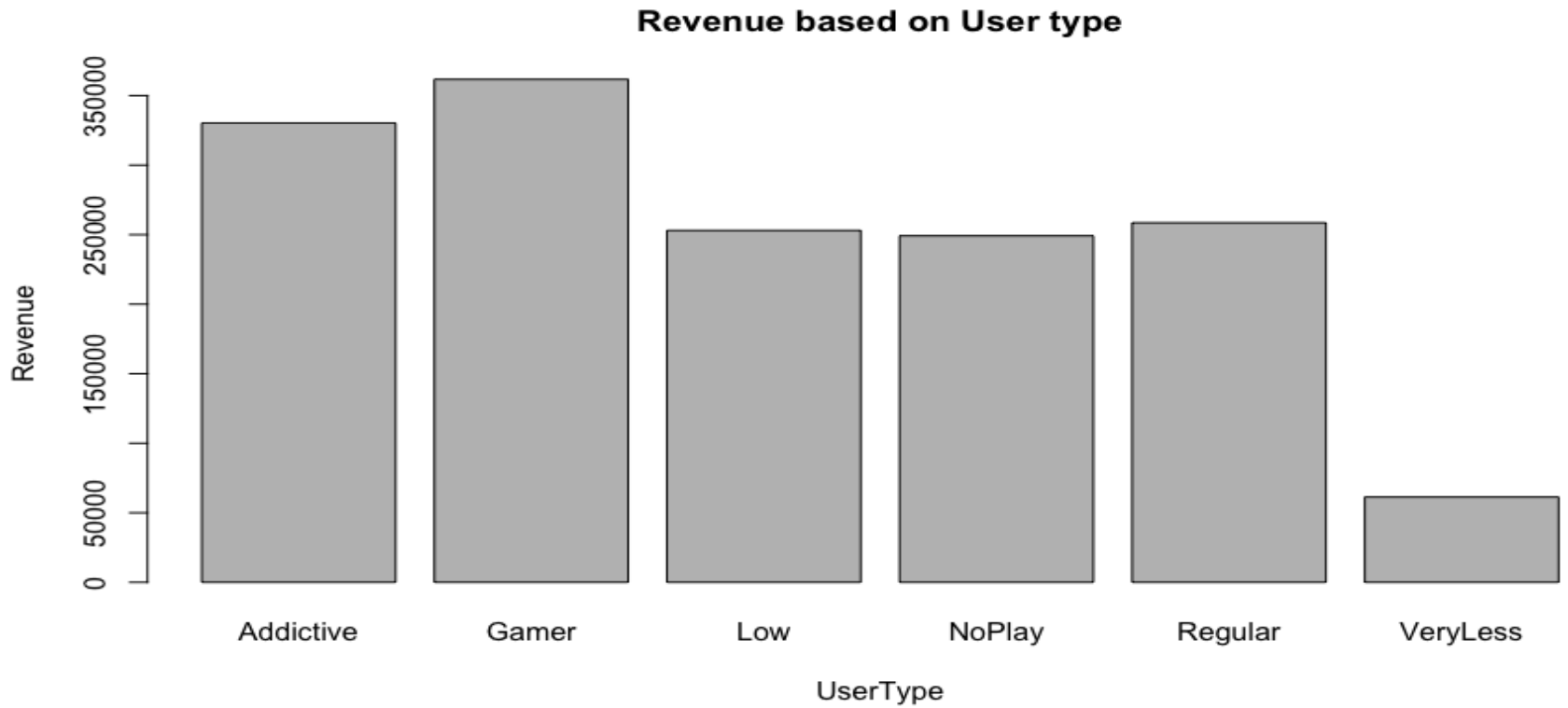
Plots



Plots


0
5000
10000
15000
20000
Addictive
Gamer
Low
NoPlay
Regular
VeryLess
UserType

Plots


Revenue based on User type
Revenue
0
50000
150000
250000
350000
Addictive
Gamer
Low
NoPlay
Regular
VeryLess
UserType

Linear Regression Model

- Predicting Total Revenue per customer based on extracted variables
- Ran on all extracted variables
- Reduced variables based on significant attributes and data knowledge
- Final model is on 27 attributes
- Median RMSE 15.17350936

	LinReg_TrainTransform	LinReg_TestTransform
mape	0.405172999	0.124943724

Regression Error Metrics

	LinReg_Train	LinReg_Test	AICLinReg_Train	AICLinReg_Test	LinReg_TrainTransform	LinReg_TestTransform
mae	9.083272156	8.938421365	9.093590099	8.940703145	1.347059523	0.393628235
mse	216.0489939	229.8263704	216.4519598	230.2353864	2.405607436	0.463730108
rmse	14.69860517	15.16002541	14.71230641	15.17350936	1.551002075	0.680977319
mape	2.630651335	2.491325444	2.659793831	2.494933761	0.405172999	0.124943724

Random Forest

- Ran Random Forest on 27 Variables extracted from Step AIC.
- Log Transformation on Target Variable
- Tuned the model using nTree and mtry.
- **Median RMSE** 0.48515

	RF_Train30_9	RF_Test30_9
mape	0.045284533	0.078922934

Random Forest Error Metrics

	RF_Train50_5	RF_Test50_5	RF_Train30_9	RF_Test30_9	RF_Train50_3	RF_Test50_3
mae	0.174194747	0.250184177	0.139657983	0.249295395	0.225101652	0.269462803
mse	0.127508339	0.227193217	0.085352658	0.232872752	0.185686578	0.246158583
rmse	0.357083098	0.476647897	0.292151773	0.482568909	0.430913655	0.496143712
mape	0.056502747	0.078153373	0.045284533	0.078922934	0.072341871	0.084047101

	RF_Train10_6	RF_Test10_6	RF_Train20_3	RF_Test20_3
mae	0.16964318	0.261219717	0.237475091	0.280226171
mse	0.116165456	0.235371322	0.20440121	0.264713494
rmse	0.340830538	0.485150824	0.452107521	0.514503153
mape	0.055051862	0.082274161	0.076280372	0.087583863

CART

- Built the Regression model using rpart.
- Log Transformed the target Variable.
- Tuned the model using complexity parameter cp and minSplit.
- MEDIAN RMSE 0.506

	TrainCP0.0001	TestCP0.0001
mape	0.079702954	0.08216484

CART Error Metrics

	TrainCP0.01	TestCP0.01	TrainCP0.001	TestCP0.001	TrainCP0.0001	TestCP0.0001
mae	0.34917524	0.347800358	0.288123444	0.290218122	0.252516151	0.261226134
mse	0.289604785	0.294674245	0.23878128	0.25132787	0.196199368	0.242374519
rmse	0.538149408	0.54283906	0.488652515	0.501326112	0.442943979	0.492315467
mape	0.105872353	0.104921509	0.08967858	0.089980155	0.079702954	0.082164844

	TrainCP0.0001_MinSp10	TestCP0.0001_MinSp10	Train0.00001_MinSp10	Test0.00001_MinSp10	Train0.00002_MinSp10	Test0.00002_MinSp10
mae	0.251188611	0.264883216	0.189893402	0.289940716	0.261253007	0.27021908
mse	0.181634815	0.257819637	0.136228109	0.292442214	0.19449948	0.25443805
rmse	0.426186362	0.507759428	0.369090922	0.540779265	0.441020952	0.504418527
mape	0.080021955	0.083961815	0.061065687	0.092627814	0.082319127	0.084790766

XG Boost

- Built the Regression model .
- Log Transformed the target Variable.
- Tuned the model using max.depth , nround and eta.
- Median RMSE 0.431

	Train3_700	Test3_700
mape	0.06800984	0.077657576

XG Boost Error Metrics

	Train3_700	Test3_700	Train2_1000	Test2_1000
mae	0.214210367	0.243071239	0.228549789	0.239007485
mse	0.169455996	0.188584722	0.199320818	0.183929308
rmse	0.411650332	0.434263425	0.446453601	0.428869803
mape	0.06800984	0.077657576	0.071885154	0.075843767

Support Vector Machines

- Built the model using SVM for 27 Variables.
- Log Transformed the Target Variable.
- **Median RMSE** 17.2206639

	SVM_Reg1_Train	SVM_Reg1_Test
mape	2.640019958	2.516680561

SVM Error Metrics

	SVM_Reg1_Train	SVM_Reg1_Test
mae	9.226218566	9.147997492
mse	280.847336	296.5512637
rmse	16.75850041	17.22066386
mape	2.640019958	2.516680561

Comparisons of Best Model

MODEL	MAPE		Parameters	Median RMSE
	TRAIN	TEST		
Linear Regression	0.405172999	0.124943724	AIC & Log Transformation	15.17
Random Forest	0.045284533	0.078922934	nTree=30, mtry = 9	0.485
CART	0.079702954	0.08216484	cp=0.0001	0.506
XG Boost	0.06800984	0.077657576	max.depth = 3,nround = 700	0.431
SVM	0.405172999	0.124943724	cost = 10, gamma = 0.1	17.22

THANK YOU