

Glassdoor Test Answers

- Abhishek Pravin, Mane
San Francisco State University

Q1 - Google BigQuery

```
select h.State_Name, DATE_TRUNC(h.StartDate, MONTH) as Month ,sum(h.Total_Contract_Value) as  
Total_Contract_Value from (  
    select g.*,f.City_Name, f.State_ID, f.State_Name from  
        gdoor.gdoor_slot_performance as g inner join gdoor.gdoor_location as f on g.City_ID =  
        f.City_ID)  
as h group by h.State_Name, Month order by h.State_Name, Month;
```

Q2

```
select * from (  
    select Employer_ID, ROW_NUMBER() over (  
        partition by Employer_ID  
        order by Employer_ID, StartDate  
    ) AS ROW_NUMBER, Job_Slots, Click_Market_Value from  
    gdoor.gdoor_slot_performance  
    ) where ROW_NUMBER = 2;
```

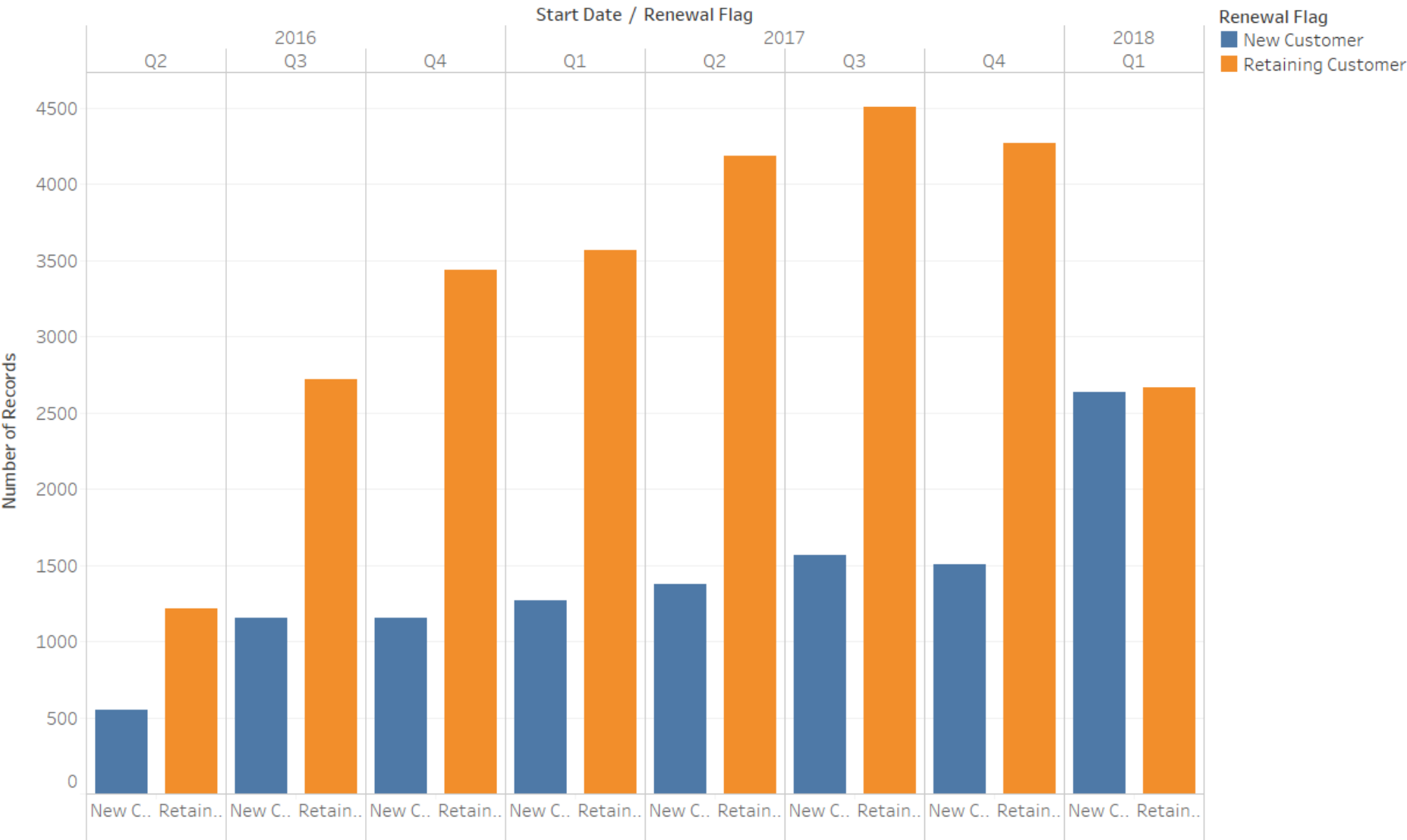
Q3 –Metric Design

- Renewal_Flag - True/false metric indicating whether the contract was renewed at the end of the contract: 1 = renewed, 0 = not renewed
- Graphs on next pages made in Tableau and other analysis in Weka –
 - Renewal_Flag over time Quarter
 - Job_Slots over time Quarter
 - Stacked Bars graph for click_market_value
 - Box Plot of Total_Contract_Value
 - Correlation matrix

- A good metric would be analysis of number of job_slots renewed over time and new customers obtained / growth rate.

- New customers should be equal or more than last quarter.
- Customers retained should be equal to new customer from last quarter + older retained customer.
- Taking into consideration job hiring process takes 1-2 months or more.

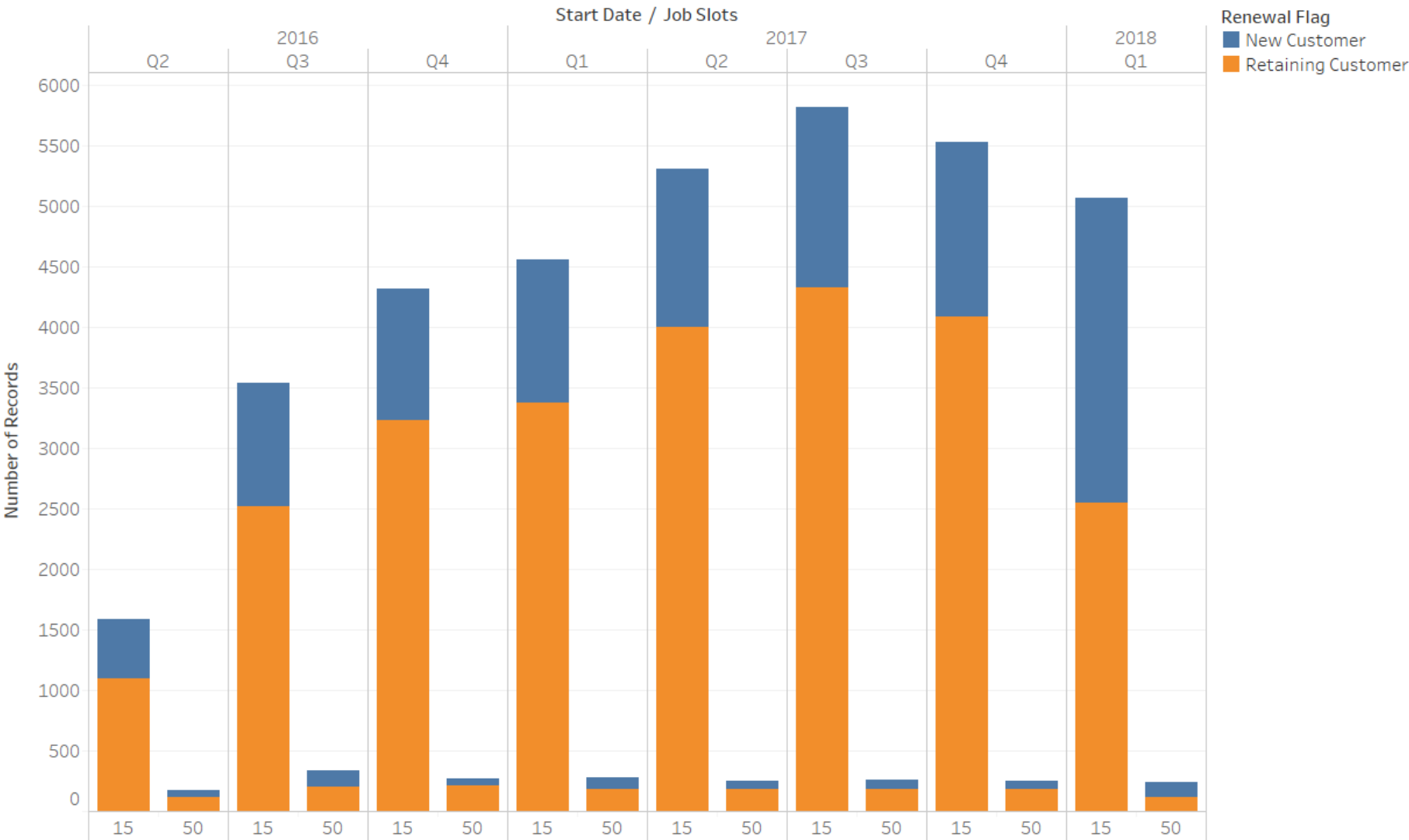
Growth Rate over time



Sum of Number of Records for each Renewal Flag broken down by Start Date Year and Start Date Quarter. Color shows details about Renewal Flag.

Performance
using Job_Slots
metrics
quarterly

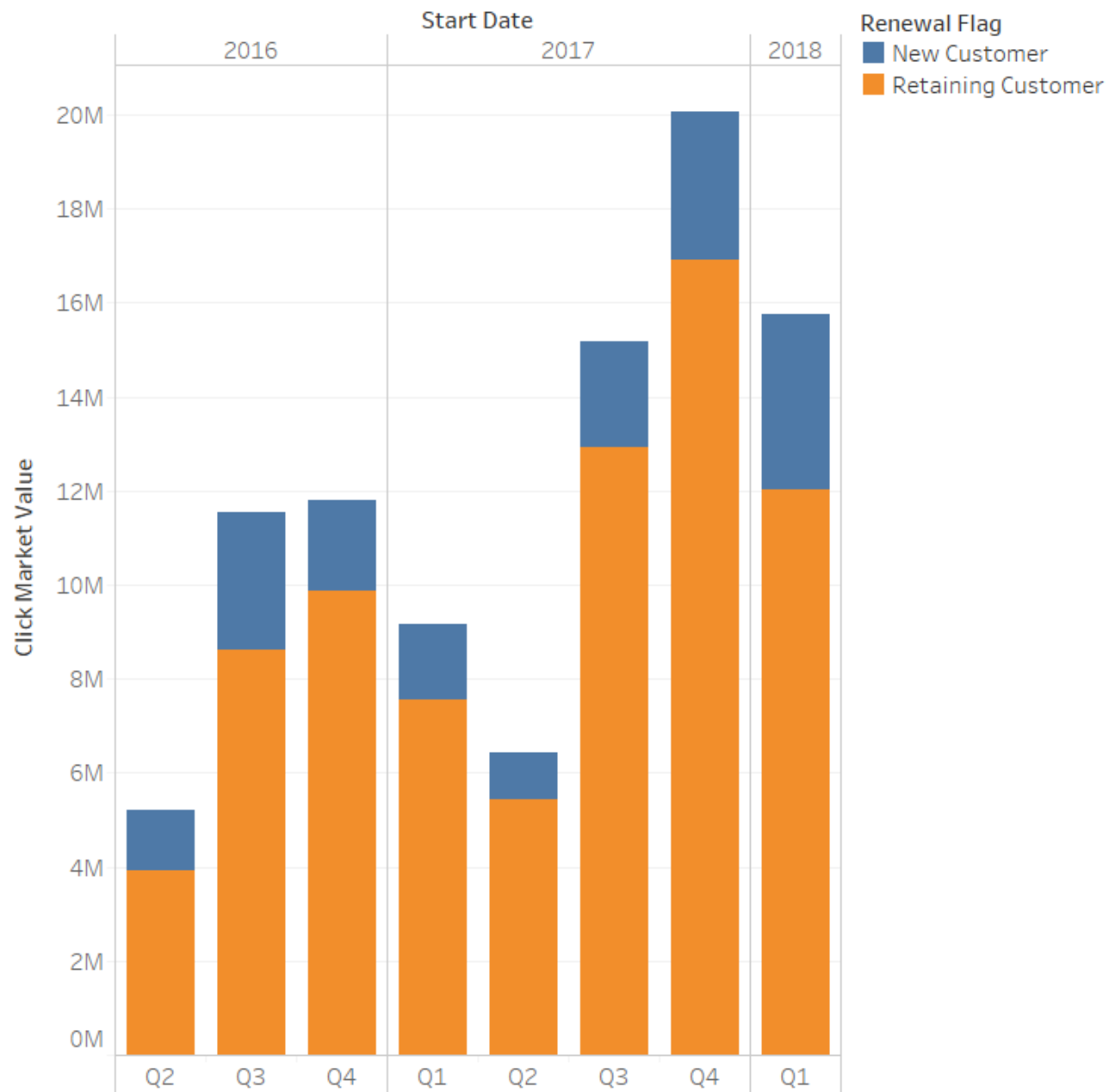
Growth Rate over time



Sum of Number of Records for each Job Slots broken down by Start Date Year and Start Date Quarter. Color shows details about Renewal Flag.

Performance
using
Click_Market
_Value
quarterly

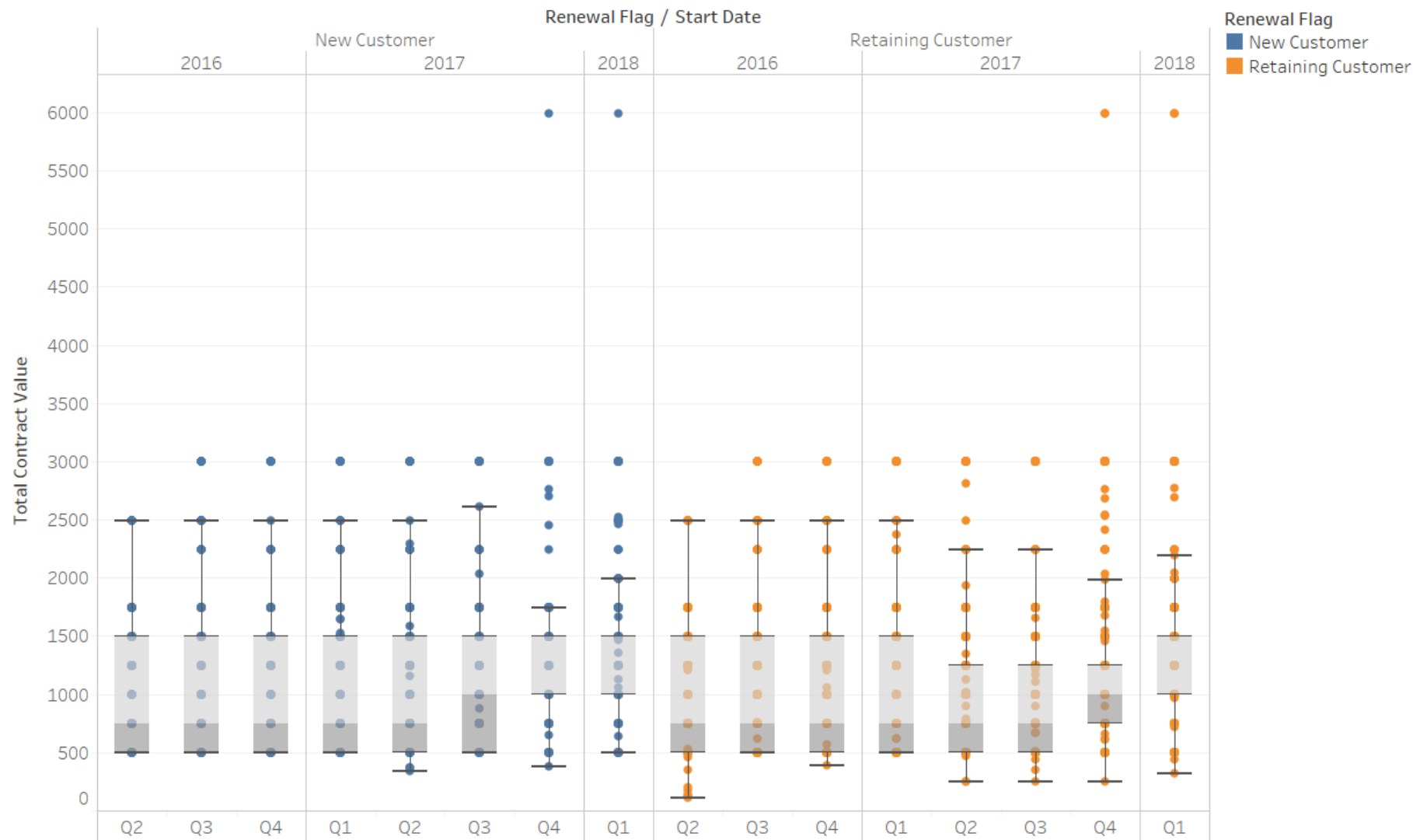
Stacked bars Plot



Click Market Value for each Quarter of Start Date broken down by Year of Start Date. Color shows details about Renewal Flag.

Performance
using
Total_Contra
ct_Value
quarterly

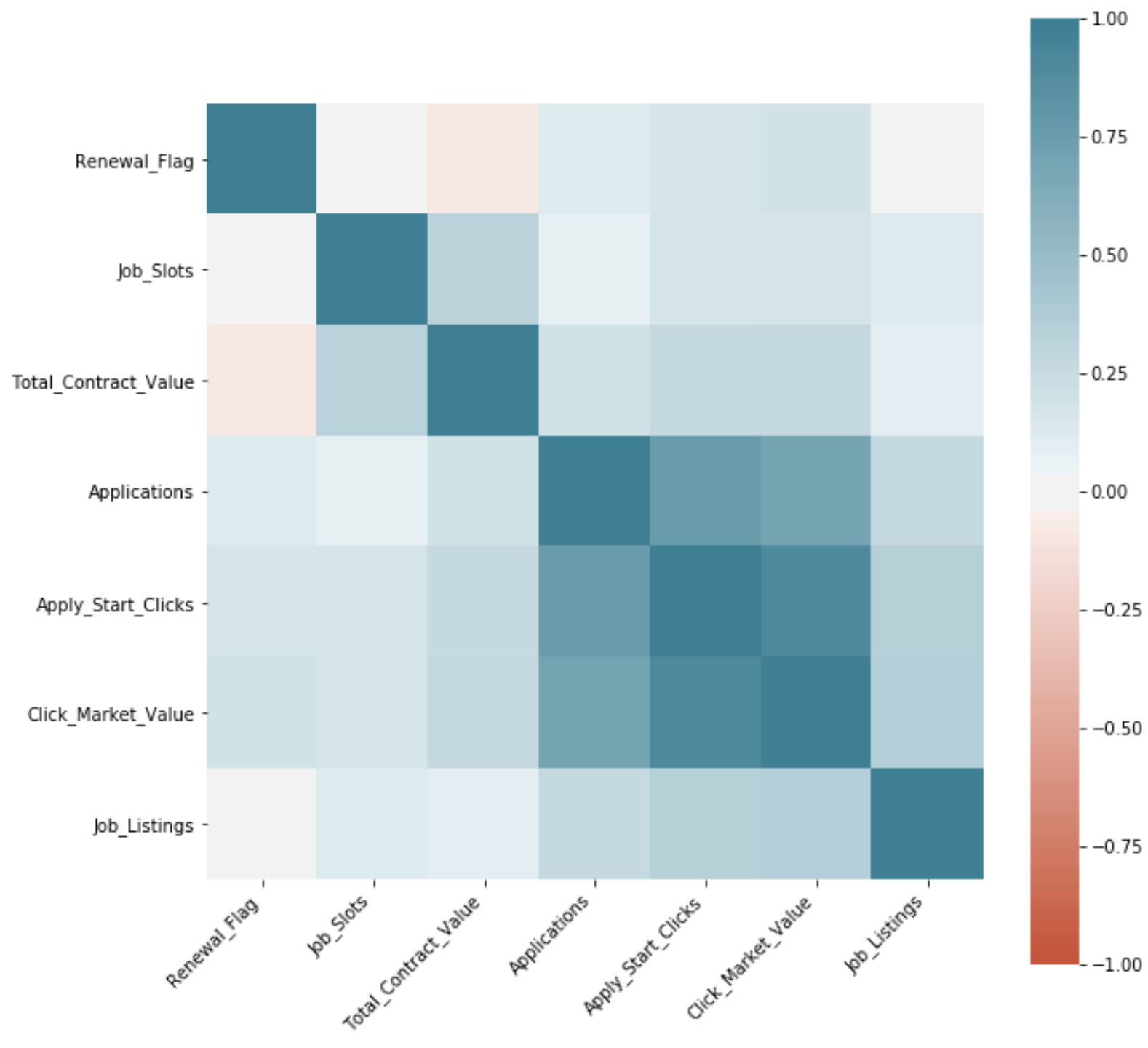
Box Plot



Total Contract Value for each Quarter of Start Date broken down by Renewal Flag and Year of Start Date. Color shows details about Renewal Flag.

Correlation Matrix

- From the matrix we can make following inferences:
- Increase in the price of `total_contract_value` may lead to a decrease in renewal of customer.
- The more the number of successful applications or `apply_start_clicks` is directly proportional to the market value they bring.



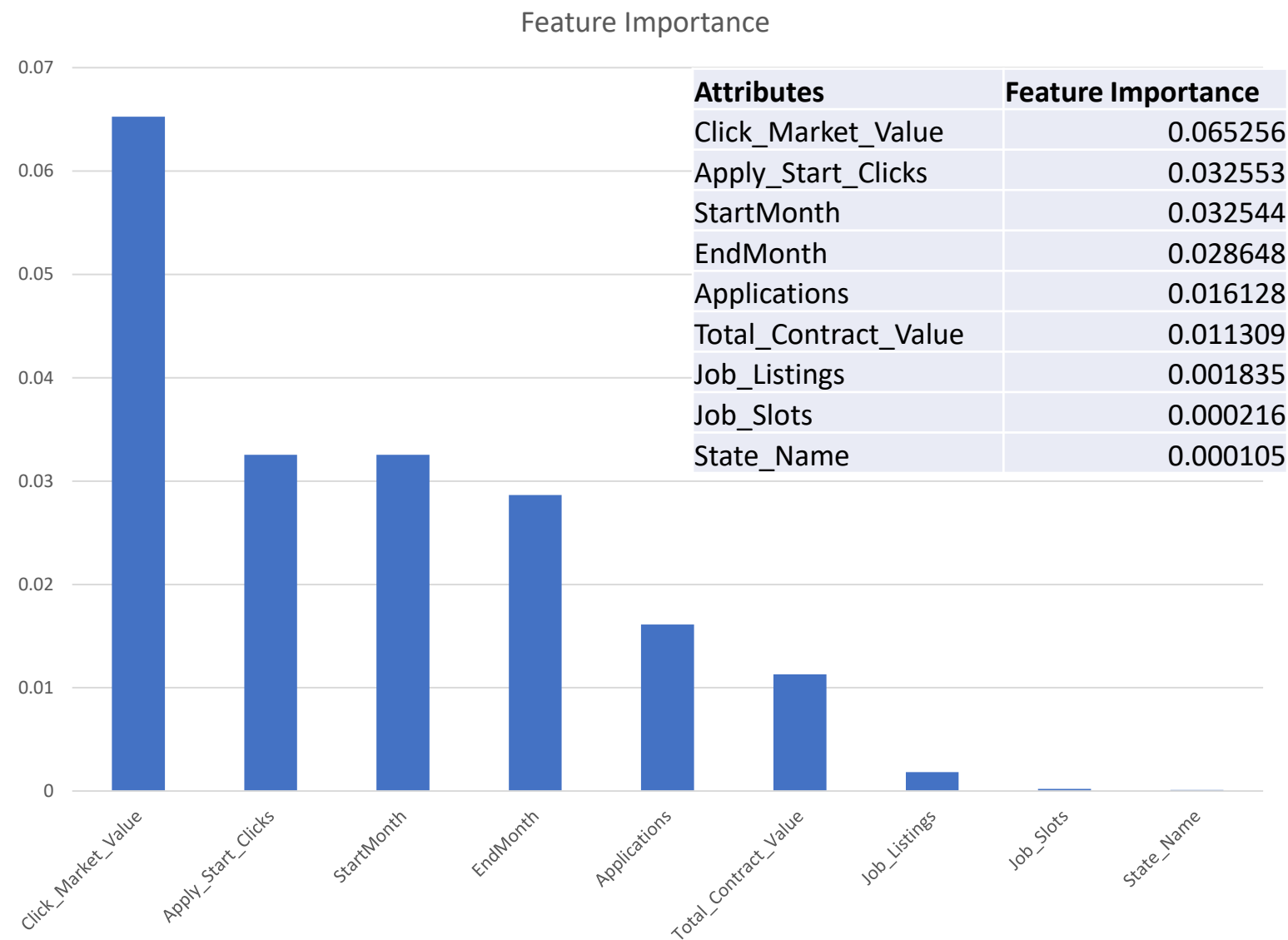
Q4 Retention Analysis

- I tried '*Information gain attribute evaluator*' and '*classifier subset evaluator*' to evaluate the information gain from each attribute. This was also backed by a decision tree J48 algorithm which chose click_market_value as the root node.
- Other features that make a significant contribution are Apply_Start_Clicks, Total_Contract_Value, StartMonth, EndMonth, Applications.
- Also tried PCA (Principal component Analysis) but wasn't good enough.

Q4 Retention Analysis (Information Gain Attribute Evaluator)

Click_Market_Value is the highest contributor to predict customer retention.

A major focus of our resources should be increasing number of apply_start_clicks as the click_market_value is directly proportional to it as seen in the correlation matrix.



Q4 Retention Analysis (Classifier Subset Evaluator)

- We also confirm click_market_value has highest value using another algorithm called classifierSubsetEvaluator as indicated in the screenshot.
- Next is screenshot of Tree obtained from training on decision tree algorithm which also shows the same.

The screenshot displays the Weka Explorer interface, specifically the 'Attribute Evaluator' tab. The 'Attribute Evaluator' section shows the 'ClassifierSubsetEval' algorithm selected. The 'Search Method' is set to 'BestFirst -D 1 -N 5'. The 'Attribute Selection Mode' is set to 'Use full training set'. The 'Result list' on the left shows the current run: '14:59:40 - BestFirst + CfsSubsetEval'. The 'Attribute selection output' pane on the right displays the following text:

```
StartMonth
EndMonth
Renewal_Flag
Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 42
  Merit of best subset found: 0.037

Attribute Subset Evaluator (supervised, Class (nominal): 10 Renewal_Flag):
  CFS Subset Evaluator
  Including locally predictive attributes

Selected attributes: 5 : 1
                    Click_Market_Value
```

The 'Selected attributes' section is circled in red, highlighting the result 'Click_Market_Value'.

Choose

Classifier output

Set...

Folds

66

More options...

(Nom) Renewal_Flag

Stop

1	1	1	EndMonth = 5: 1 (0.0)
1	1	1	EndMonth = 6: 1 (0.0)

18:36:37 - trees.J48

EndMonth	= 7 + 1	(0, 0)
----------	---------	--------

Status

OK

Log

 $\times 0$

Q5 Retention Analysis (Performance of model)

Algorithm	Precision for Class 0	Precision for Class 1	Recall for Class 0	Recall for Class 1
J48 Decision Tree	0.836	0.814	0.479	0.96
ANN Multilayer Perceptron	0.8	0.818	0.5	0.948
K-Nearest Neighbours	0.721	0.816	0.511	0.917
Random Forest	0.733	0.816	0.508	0.922

- From above we can see that J48 Decision Tree and ANN are the top 2 players.
- Next 2 slides show training results for J48 Decision Tree and Artificial Neural Network (Multilayer Perceptron) algorithms.

J48 Decision Tree - Weka

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Renewal_Flag

Start Stop

Result list (right-click for options)

12:52:28 - trees.J48

Classifier output

```
Number of Leaves :      867

Size of the tree :     1067

Time taken to build model: 1.69 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      30871           81.7623 %
Incorrectly Classified Instances    6886           18.2377 %
Kappa statistic                    0.5013
Mean absolute error                 0.2806
Root mean squared error             0.3818
Relative absolute error             67.2565 %
Root relative squared error         83.5973 %
Total Number of Instances          37757

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.479    0.040    0.836     0.479    0.609      0.534    0.765    0.674    0
                0.960    0.521    0.814     0.960    0.881      0.534    0.765    0.839    1
Weighted Avg.   0.818    0.378    0.820     0.818    0.800      0.534    0.765    0.790

=== Confusion Matrix ===

      a    b  <-- classified as
5364  5835 |    a = 0
1051 25507 |    b = 1
```

Status

OK Log

Type here to search

12:54 PM 2/22/2020

Artificial Neural Network (Multilayer Perceptron) - Weka

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Classifier

Choose **MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % **66**
More options...

(Nom) Renewal_Flag

Start Stop

Result list (right-click for options)

- 12:52:28 - trees.J48
- 12:55:35 - trees.RandomForest
- 12:59:30 - bayes.NaiveBayes
- 12:59:41 - functions.Logistic
- 13:00:05 - functions.MultilayerPerceptron**

Classifier output

```
Class 1
Input
Node 1

Time taken to build model: 202.69 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      30768      81.4895 %
Incorrectly Classified Instances    6989      18.5105 %
Kappa statistic                    0.5024
Mean absolute error                 0.2847
Root mean squared error             0.3866
Relative absolute error             68.2194 %
Root relative squared error         84.6356 %
Total Number of Instances          37757

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                -----  -----  -
                0.500    0.052    0.801     0.500    0.616     0.527    0.753    0.675    0
                0.948    0.500    0.818     0.948    0.878     0.527    0.753    0.835    1
Weighted Avg.   0.815    0.367    0.813     0.815    0.800     0.527    0.753    0.788

=== Confusion Matrix ===

      a    b  <-- classified as
5604  5595 |    a = 0
1394  25164 |    b = 1
```

Status

OK Log x0

Q6 Retention Analysis (Recommendation)

- We can observe the Statewise growth-rate and make recommendation of appropriate job_slot package to that particular state. Thus we can have different job_slot packages for different states.
- To make regional job_slot size recommendation according to state. This can be done by analyzing the correlation between job_slot and number of application.
- Based on the analysis I did, I felt there should be more Job slot variations (right now there are only 2 – 15 and 50).
- Understanding the click_market_value - how it was obtained - might also help understand how to make better analysis.

Thank You
