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

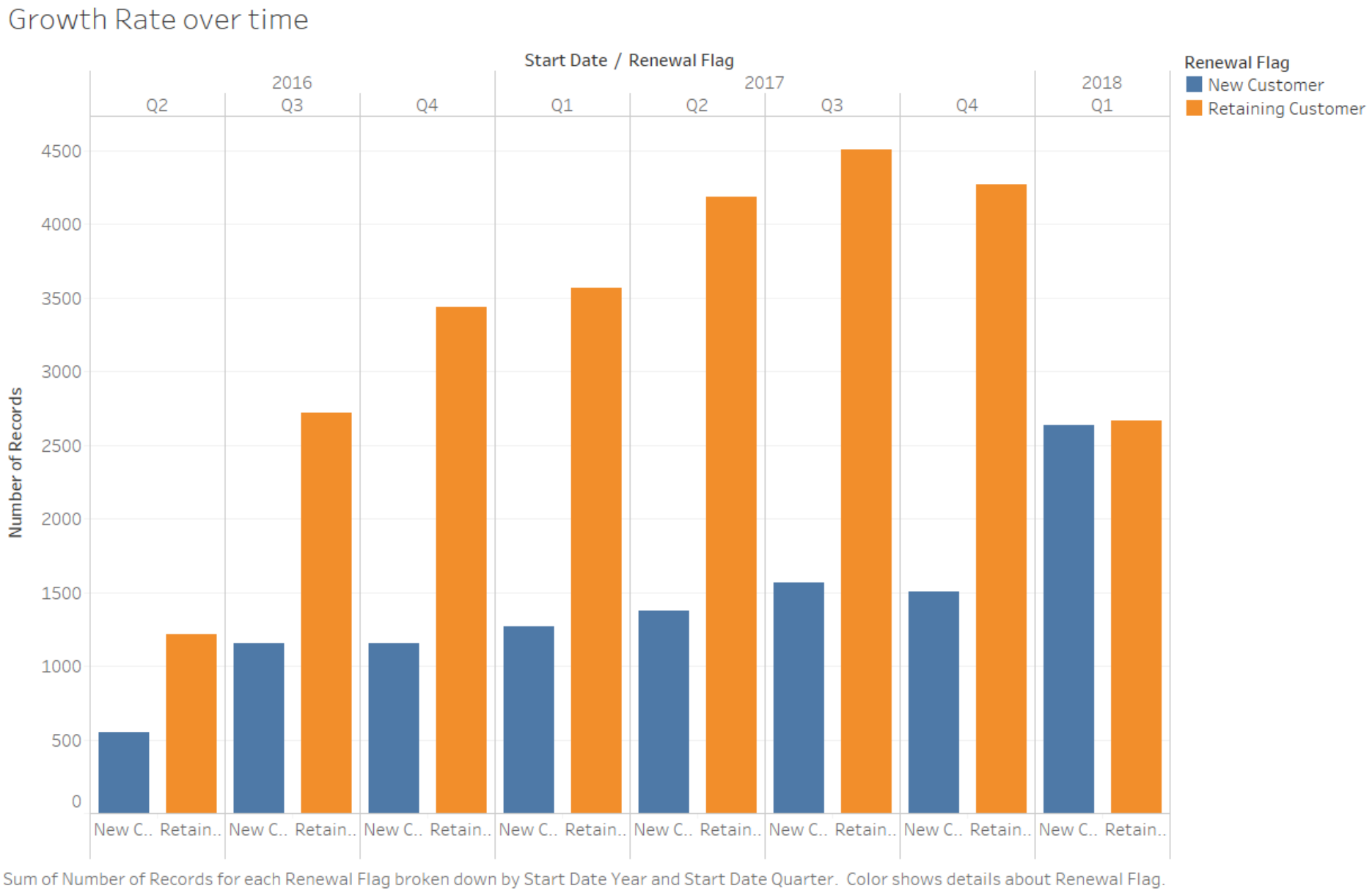
- Graphs on next pages made in Tableau and other analysis in Weka –
  - Renewal_Flag over time Quarter
  - Job_Slots over time Quarter
  - Stacked Bars graph for click_market_value
  - Box Plot of Total_Contract_Value
  - Correlation matrix

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

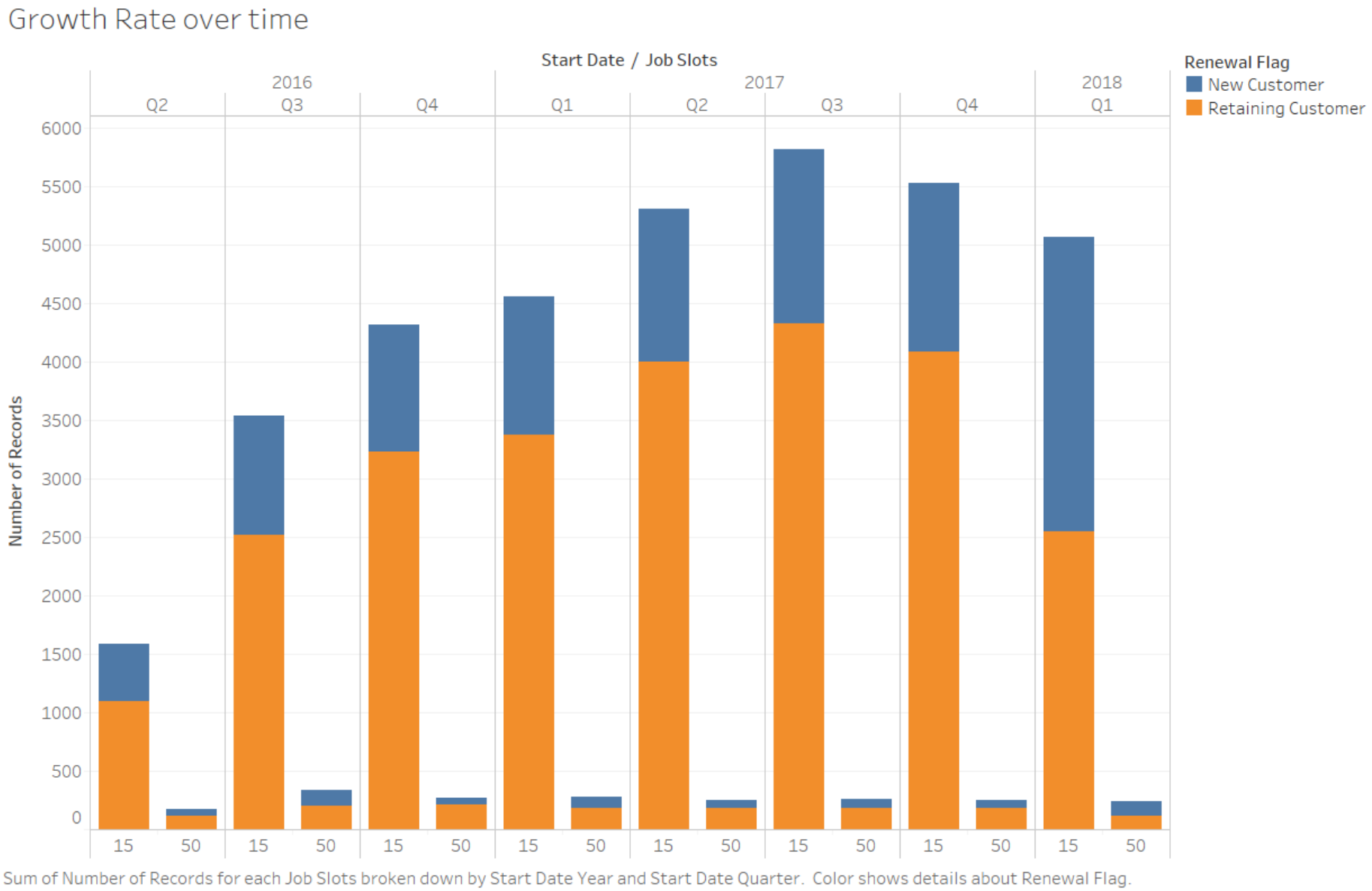
- New customers should be equal or more than last quarter.

- 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.

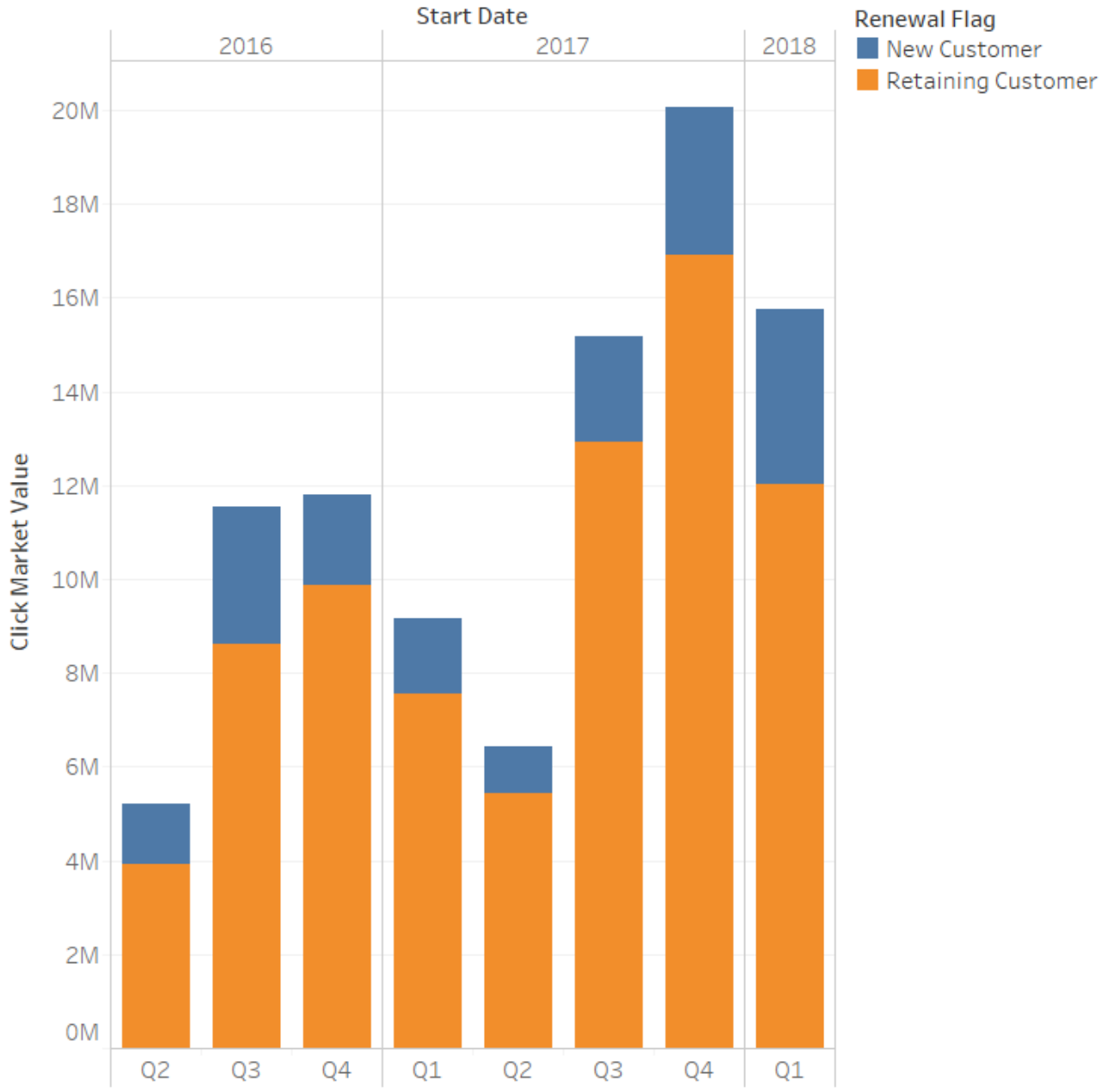


Performance using Job_Slots metrics quarterly



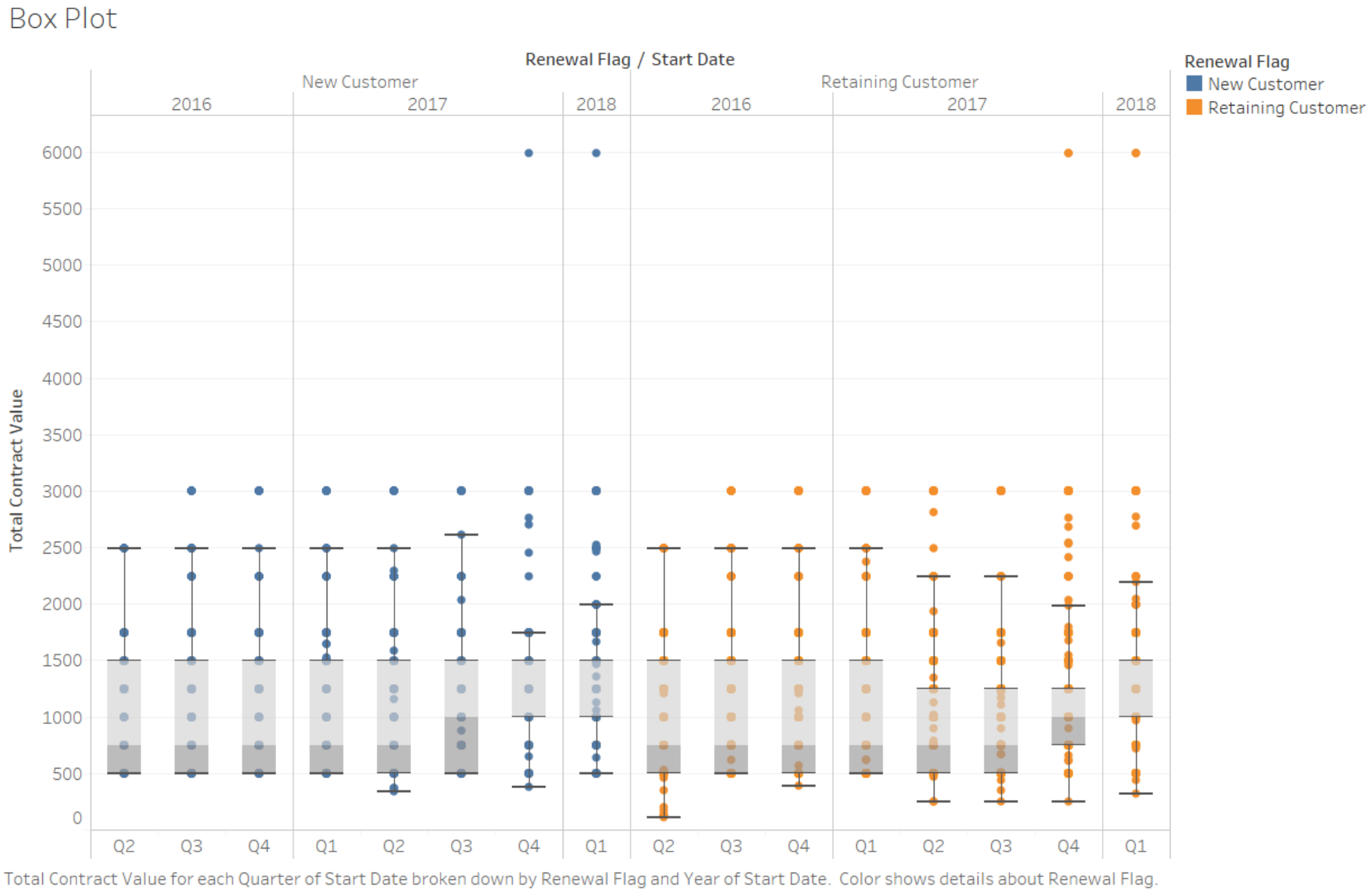
Performance using Click_Market_Value quarterly

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_Contract_Value quarterly

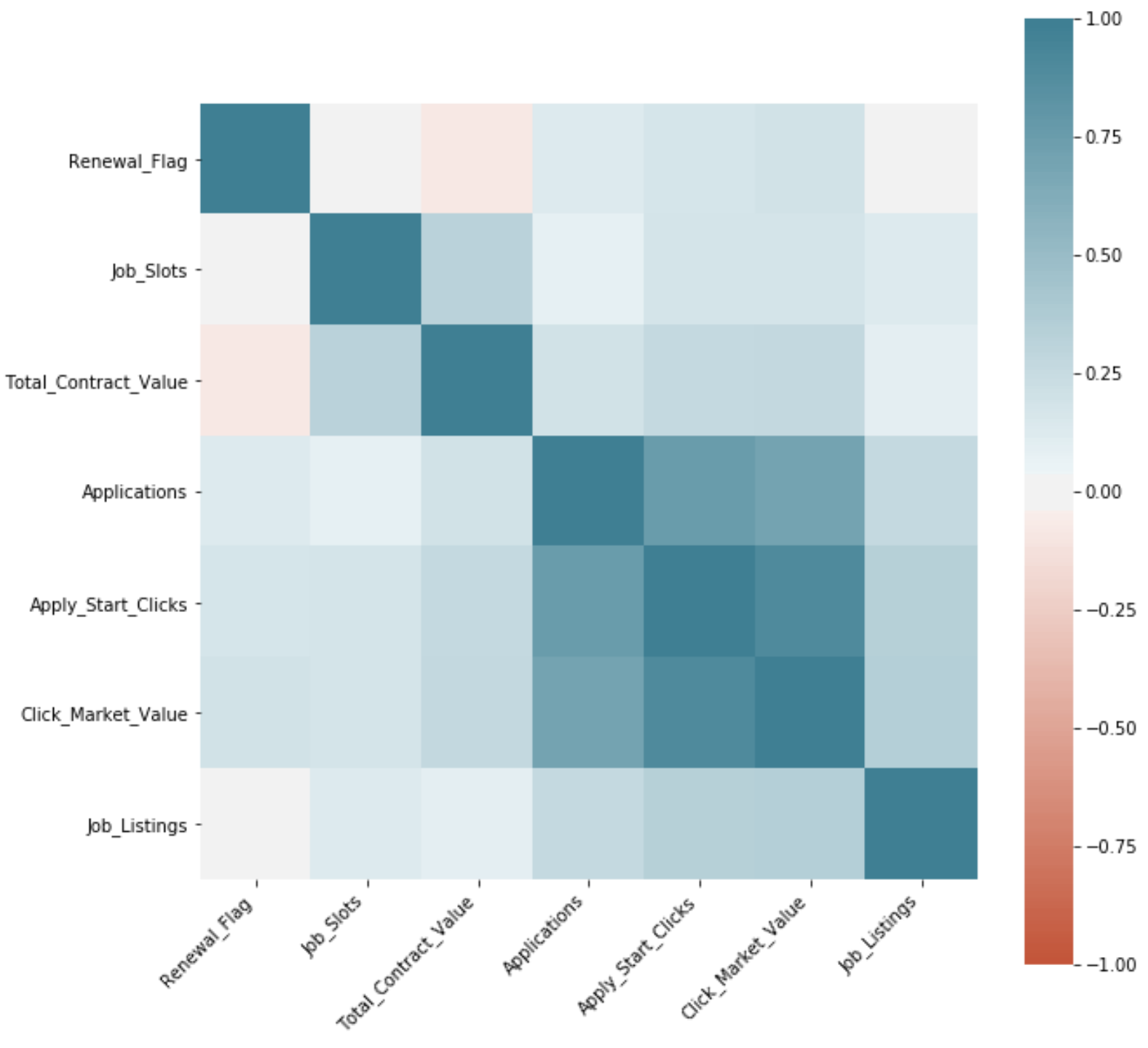


Correlation Matrix

- 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.



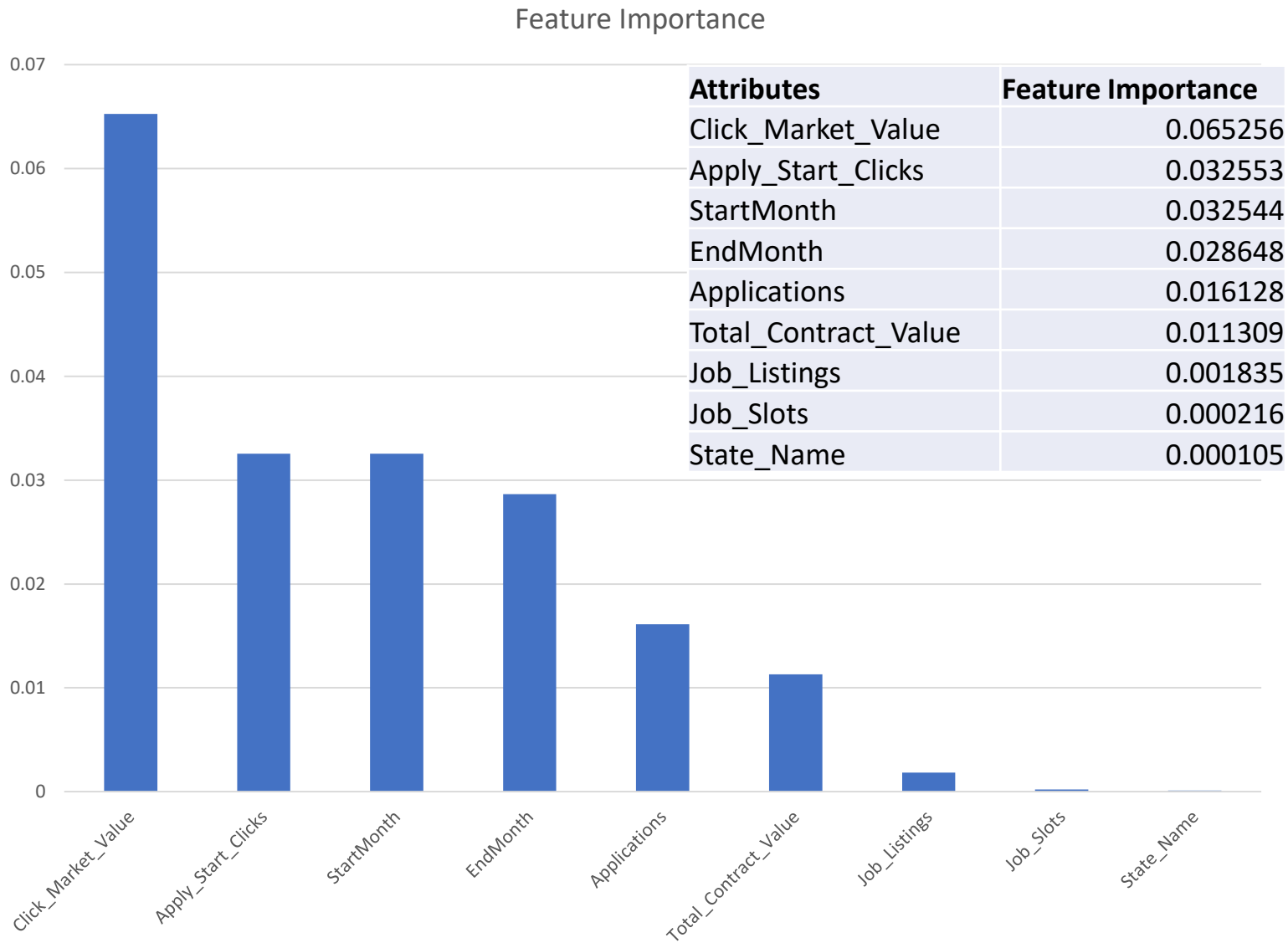
Q4 Retention Analysis

- 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)



Attributes	Feature Importance
Click_Market_Value	0.065256
Apply_Start_Clicks	0.032553
StartMonth	0.032544
EndMonth	0.028648
Applications	0.016128
Total_Contract_Value	0.011309
Job_Listings	0.001835
Job_Slots	0.000216
State_Name	0.000105

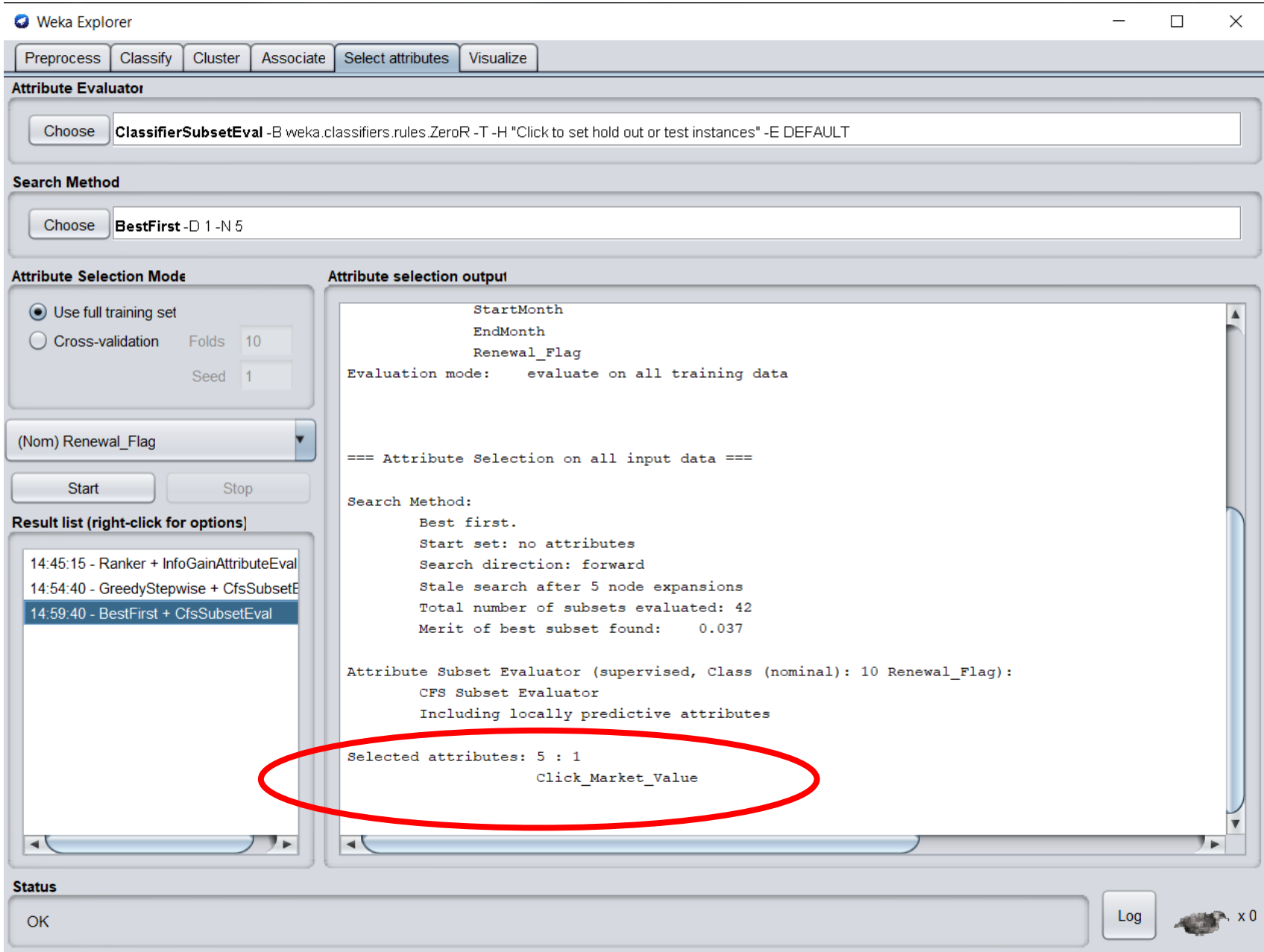
Click_Market_Value is the highest contributer to predict customer retention.

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.



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)

18:36:37 - trees.J48

Classifier output

```
=== Classifier model (full training set) ===

J48 pruned tree
------------------

Click_Market_Value <= 1048.8
|   Total_Contract_Value <= 785
|   |   StartMonth = 1
|   |   |   EndMonth = 1: 0 (105.0)
|   |   |   EndMonth = 2: 1 (527.0/99.0)
|   |   |   EndMonth = 3: 1 (0.0)
|   |   |   EndMonth = 4: 1 (0.0)
|   |   |   EndMonth = 5: 1 (0.0)
|   |   |   EndMonth = 6: 1 (0.0)
|   |   |   EndMonth = 7: 1 (0.0)
|   |   |   EndMonth = 8: 1 (0.0)
|   |   |   EndMonth = 9: 1 (0.0)
|   |   |   EndMonth = 10: 1 (0.0)
|   |   |   EndMonth = 11: 1 (0.0)
|   |   |   EndMonth = 12: 1 (0.0)
|   |   StartMonth = 2
|   |   |   EndMonth = 1: 1 (0.0)
|   |   |   EndMonth = 2: 0 (84.0/3.0)
|   |   |   EndMonth = 3
|   |   |   |   Apply_Start_Clicks <= 35
|   |   |   |   |   Total_Contract_Value <= 620: 1 (329.0/49.0)
|   |   |   |   |   Total_Contract_Value > 620
|   |   |   |   |   |   Applications <= 5: 1 (148.0/25.0)
|   |   |   |   |   |   Applications > 5: 0 (12.0/3.0)
|   |   |   |   Apply_Start_Clicks > 35: 1 (48.0/2.0)
|   |   |   EndMonth = 4: 1 (0.0)
|   |   |   EndMonth = 5: 1 (0.0)
|   |   |   EndMonth = 6: 1 (0.0)
|   |   |   EndMonth = 7: 1 (0.0)
```

Status

OK

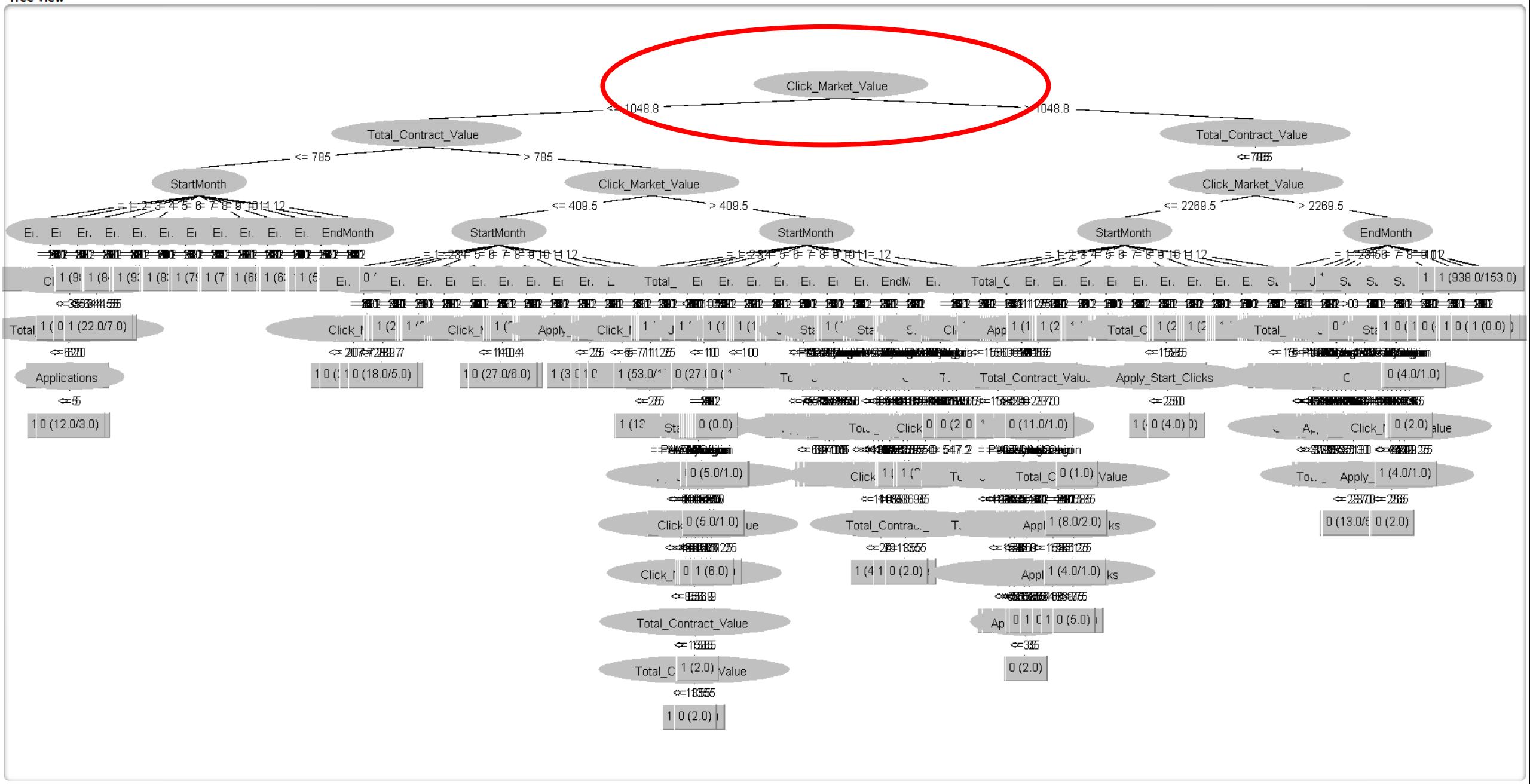
Log | x 0

Type here to search

6:37 PM
2/22/2020

Weka Classifier Tree Visualizer: 18:36:37 - trees.J48 (combined_slot_loc3-weka.filters.unsupervised.attribute.Remove-R1-5,12-13-weka.filters.unsupervised.attribute.NumericToNominal-R8,9,10)

Tree View



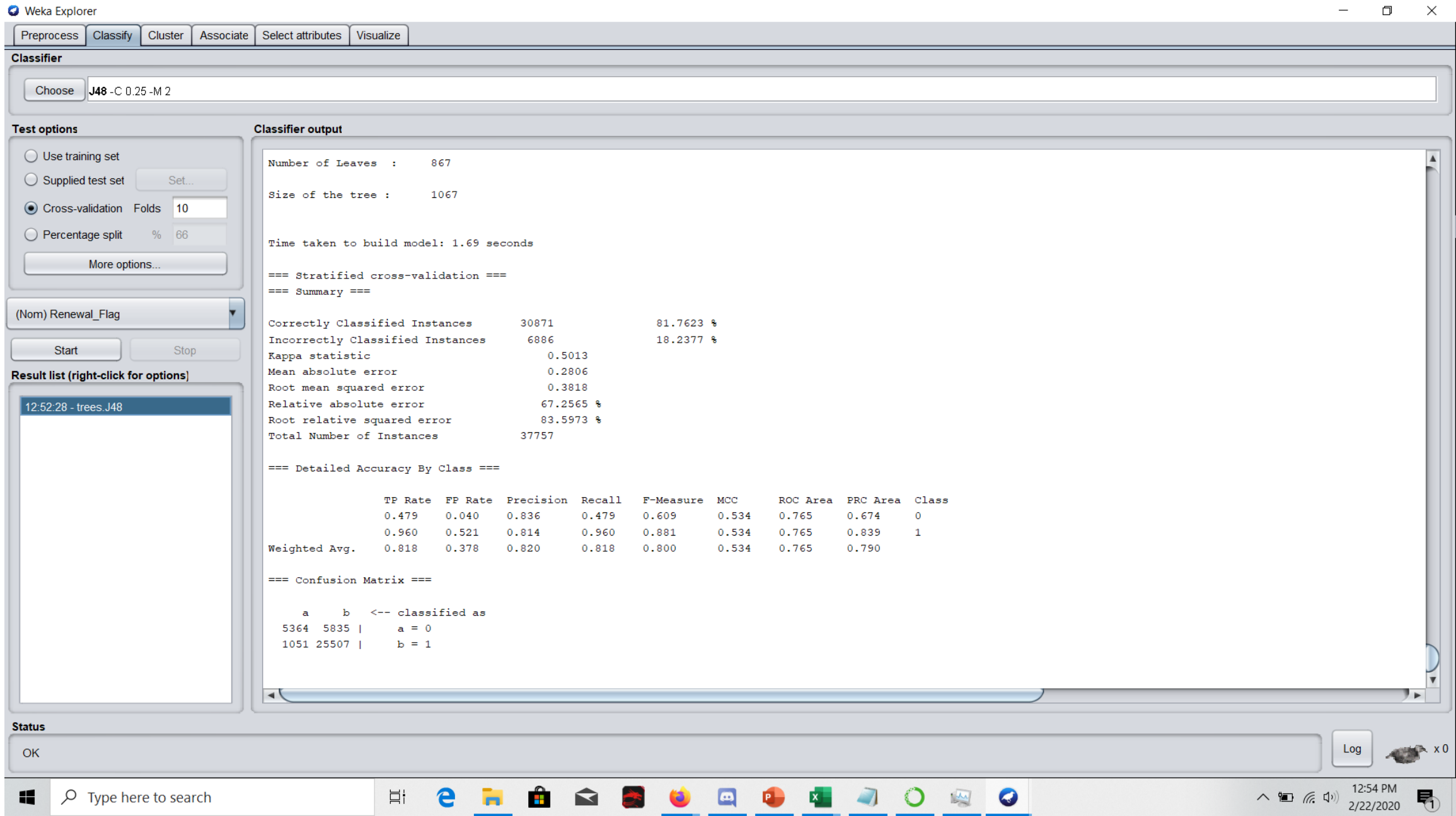
Q5 Retention Analysis (Performance of model)

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

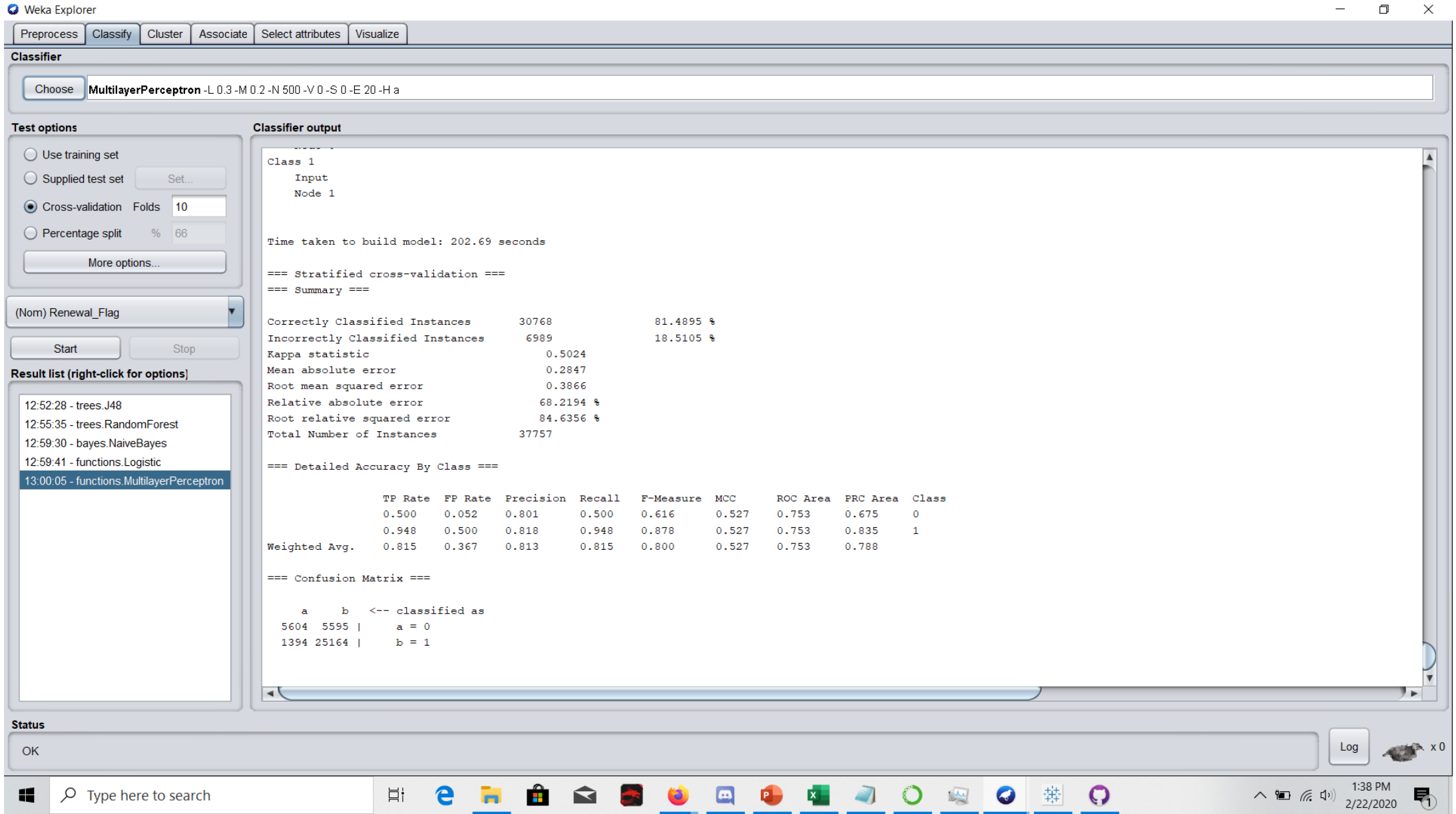
- From above we can see that J48 Decision Tree and ANN are the top 2 players.

- Next 2 slides show training results for J48 Decision Tree and Artificial Neural Network (Multilayer Perceptron) algorithms.

J48 Decision Tree - Weka



Artificial Neural Network (Multilayer Perceptron) - Weka



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