ACQUIRED VALUE SHOPPING CHALLENGE

ID5059 – Group H

SUGERE

SUGERE

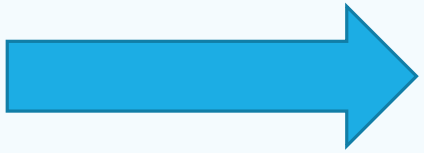
- Since 1952
- 189 shops
- 2500 employees
- 400 million £ revenue (2014), in decline of 2.3%
- Net profit = 50 million £

Mission:

To bring a smile to every customer, every day.

THE SAINT GROUP

- Alex
- Hua
- Jun
- Tom
- Victor
- Xu



Better decision-making through data mining

CONTEXT

UK = 75% customers used a coupon

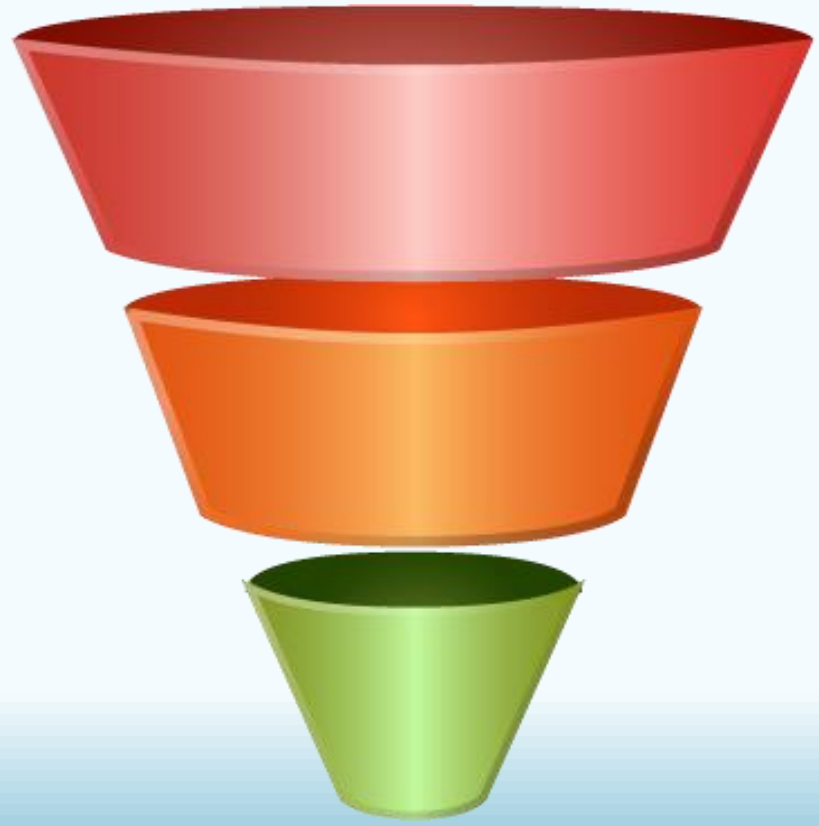
Success of coupon operations in the USA

Recurring customer increase lifetime value

How to make coupon customers recurring customers?

DATA CLEANING

- 20 GB = 380 million rows
- 1.7GB = matching company brand & category
- Features were produced per customer



```
0.id=features_train.id), boughtcategoryamount60=(select amount from temptraincategory60 where temptraincat
ain set boughtcategory90=(select count from temptraincategory90 where temptraincategory90.id=features_trai
0.id=features_train.id), boughtcategoryamount90=(select amount from temptraincategory90 where temptraincat
ain set boughtcategory180=(select count from temptraincategory180 where temptraincategory180.id=features_t
gory180.id=features_train.id), boughtcategoryamount180=(select amount from temptraincategory180 where temp

ain set boughtbrand=(select count from temptrainbrand where temptrainbrand.id=features_train.id), boughtbr
tbrandamount=(select amount from temptrainbrand where temptrainbrand.id=features_train.id);
ain set boughtbrand30=(select count from temptrainbrand30 where temptrainbrand30.id=features_train.id), bo
n.id), boughtbrandamount30=(select amount from temptrainbrand30 where temptrainbrand30.id=features_train.i
ain set boughtbrand60=(select count from temptrainbrand60 where temptrainbrand60.id=features_train.id), bo
n.id), boughtbrandamount60=(select amount from temptrainbrand60 where temptrainbrand60.id=features_train.i
ain set boughtbrand90=(select count from temptrainbrand90 where temptrainbrand90.id=features_train.id), bo
n.id), boughtbrandamount90=(select amount from temptrainbrand90 where temptrainbrand90.id=features_train.i
ain set boughtbrand180=(select count from temptrainbrand180 where temptrainbrand180.id=features_train.id),
s_train.id), boughtbrandamount180=(select amount from temptrainbrand180 where temptrainbrand180.id=feature


st set boughtcompany=(select count from temptestcompany where temptestcompany.id=features_test.id), bought
boughtcompanyamount=(select amount from temptestcompany where temptestcompany.id=features_test.id);
st set boughtcompany30=(select count from temptestcompany30 where temptestcompany30.id=features_test.id),
s_test.id), boughtcompanyamount30=(select amount from temptestcompany30 where temptestcompany30.id=feature
st set boughtcompany60=(select count from temptestcompany60 where temptestcompany60.id=features_test.id),
s_test.id), boughtcompanyamount60=(select amount from temptestcompany60 where temptestcompany60.id=feature
st set boughtcompany90=(select count from temptestcompany90 where temptestcompany90.id=features_test.id),
s_test.id), boughtcompanyamount90=(select amount from temptestcompany90 where temptestcompany90.id=feature
st set boughtcompany180=(select count from temptestcompany180 where temptestcompany180.id=features_test.id
eatures_test.id), boughtcompanyamount180=(select amount from temptestcompany180 where temptestcompany180.i

st set boughtcategory=(select count from temptestcategory where temptestcategory.id=features_test.id), bou
.id), boughtcategoryamount=(select amount from temptestcategory where temptestcategory.id=features_test.id
st set boughtcategory30=(select count from temptestcategory30 where temptestcategory30.id=features_test.id
eatures_test.id), boughtcategoryamount30=(select amount from temptestcategory30 where temptestcategory30.i
st set boughtcategory60=(select count from temptestcategory60 where temptestcategory60.id=features_test.id
eatures_test.id), boughtcategoryamount60=(select amount from temptestcategory60 where temptestcategory60.i
st set boughtcategory90=(select count from temptestcategory90 where temptestcategory90.id=features_test.id
eatures_test.id), boughtcategoryamount90=(select amount from temptestcategory90 where temptestcategory90.i
st set boughtcategory180=(select count from temptestcategory180 where temptestcategory180.id=features_test
```

MODEL

Method + Model

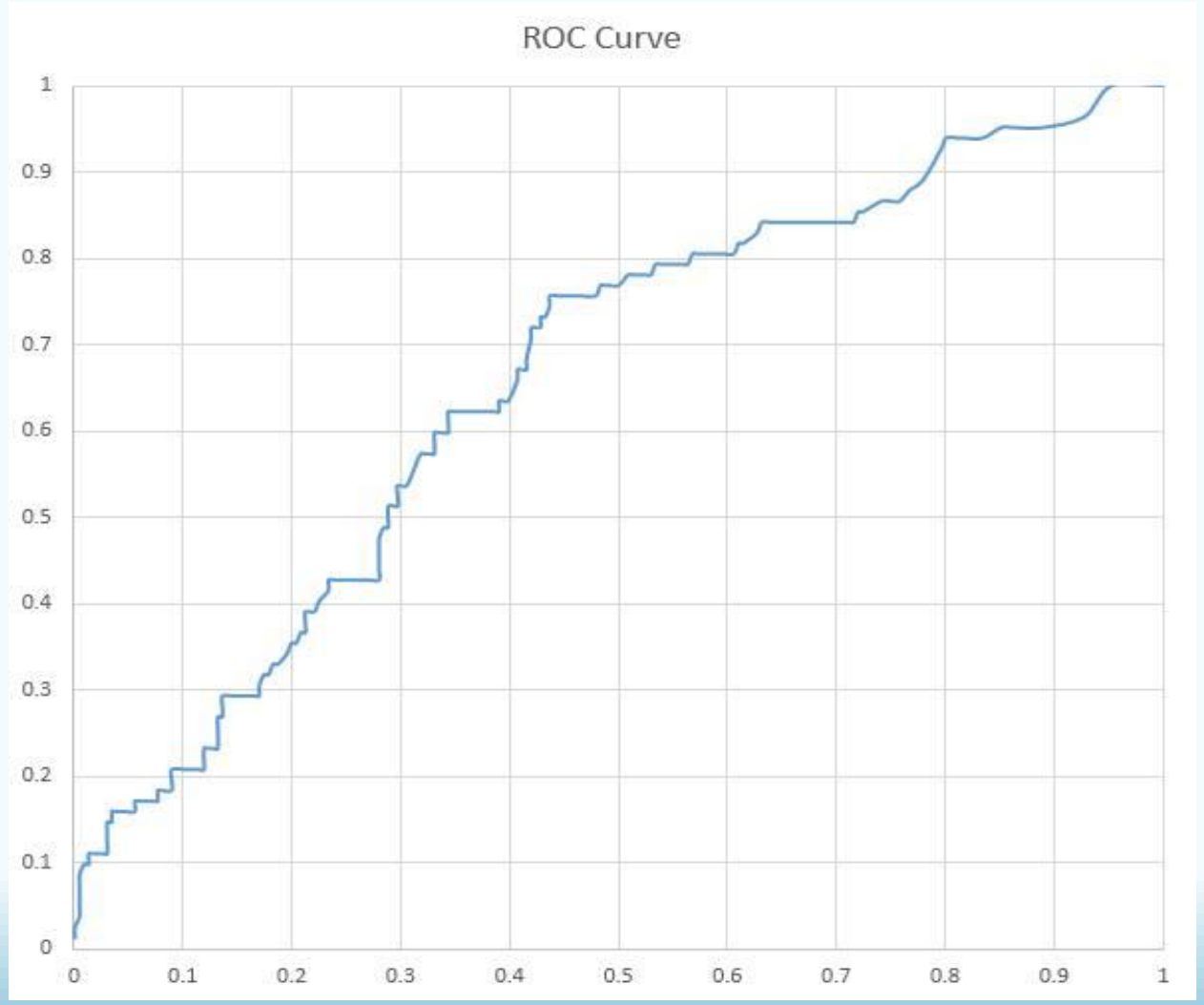
```
    self.models = list()
    progress = ProgressBar()
    for i in progress(xrange(self.num_groups)):
        inputs = self.dt.get_training_inputs(i)
        outputs = self.dt.get_training_outputs(i)
        clf = classifier(**kwargs)
        clf.fit(inputs, outputs)
        self.models.append(clf)
    return self

def predict(self, inputs):
    ret_sum = 0.0
    progress = ProgressBar()
    print "Predicting..."
    for i in progress(xrange(len(self.models))):
        ret_sum += self.models[i].predict(inputs)
    return ret_sum / len(self.models)

def compute_auc(self):
    h_inputs = self.dt.get_unseen_inputs()
    h_outputs = self.dt.get_unseen_outputs()
    pred_out = self.predict(h_inputs)
```

- Ridge Regression based on Generalized Linear Model (GLM)
- This technique adds a penalty term on the GLM coefficients which improves the result
- The features were partitioned into N chunks, and a Ridge classifier was trained on each chunk
- The final prediction was based on the average prediction by each model based on its own training data.
- The model estimation is using k-fold cross validation on their partitioned training data in order to reduce the generalization error.
- Increased the AUC score by 0.003 on Kaggle

ROC CURVE



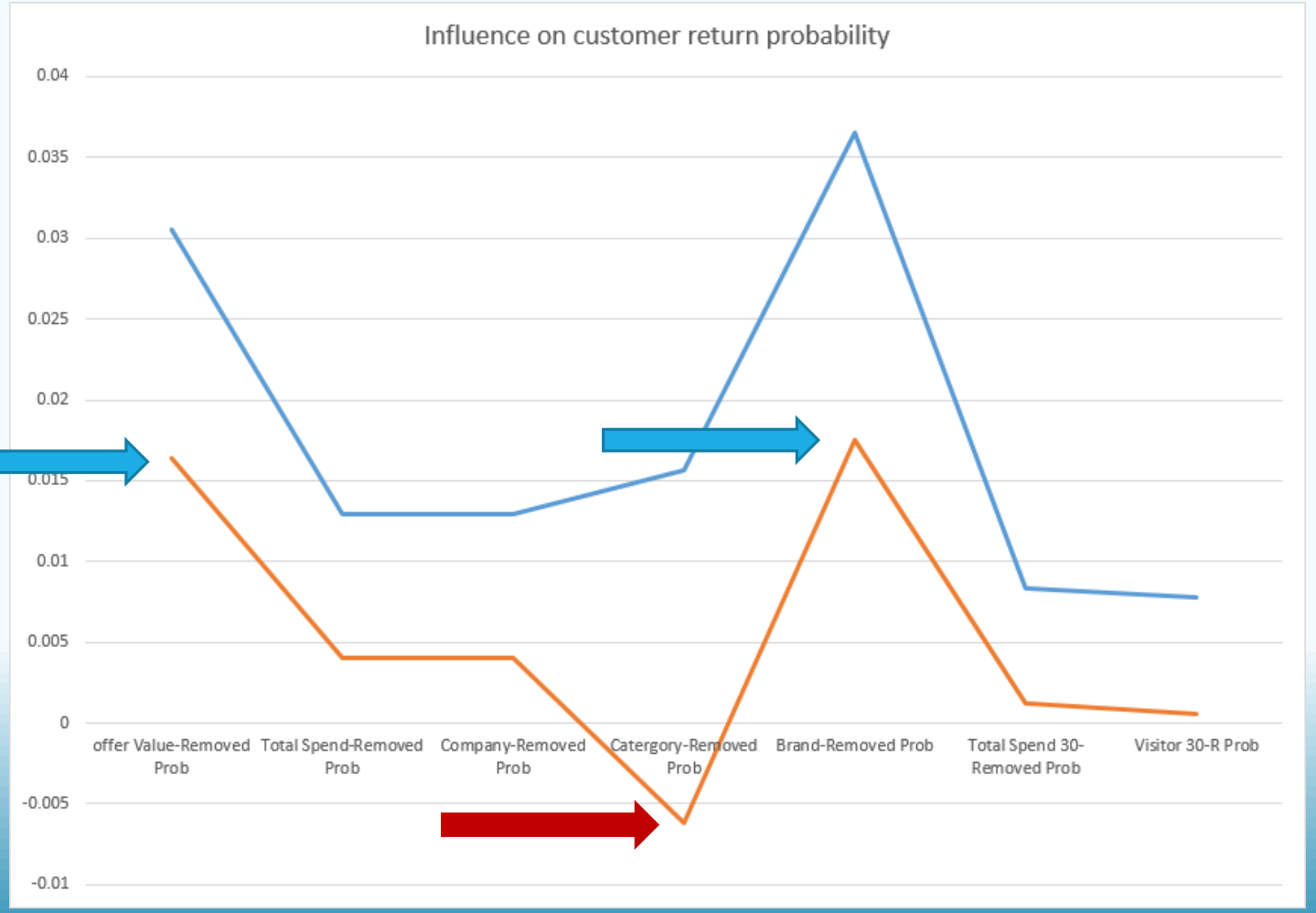
RESULTS

66	↓18	TayShin	0.60313	63	Mon, 14 Jul 2014 17:48:03 (-0.7h)	
-		**tomdalton**	**0.60303**	-	**Fri, 24 Apr 2015 08:35:02**	**Post-Deadline**

Post-Deadline Entry
If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

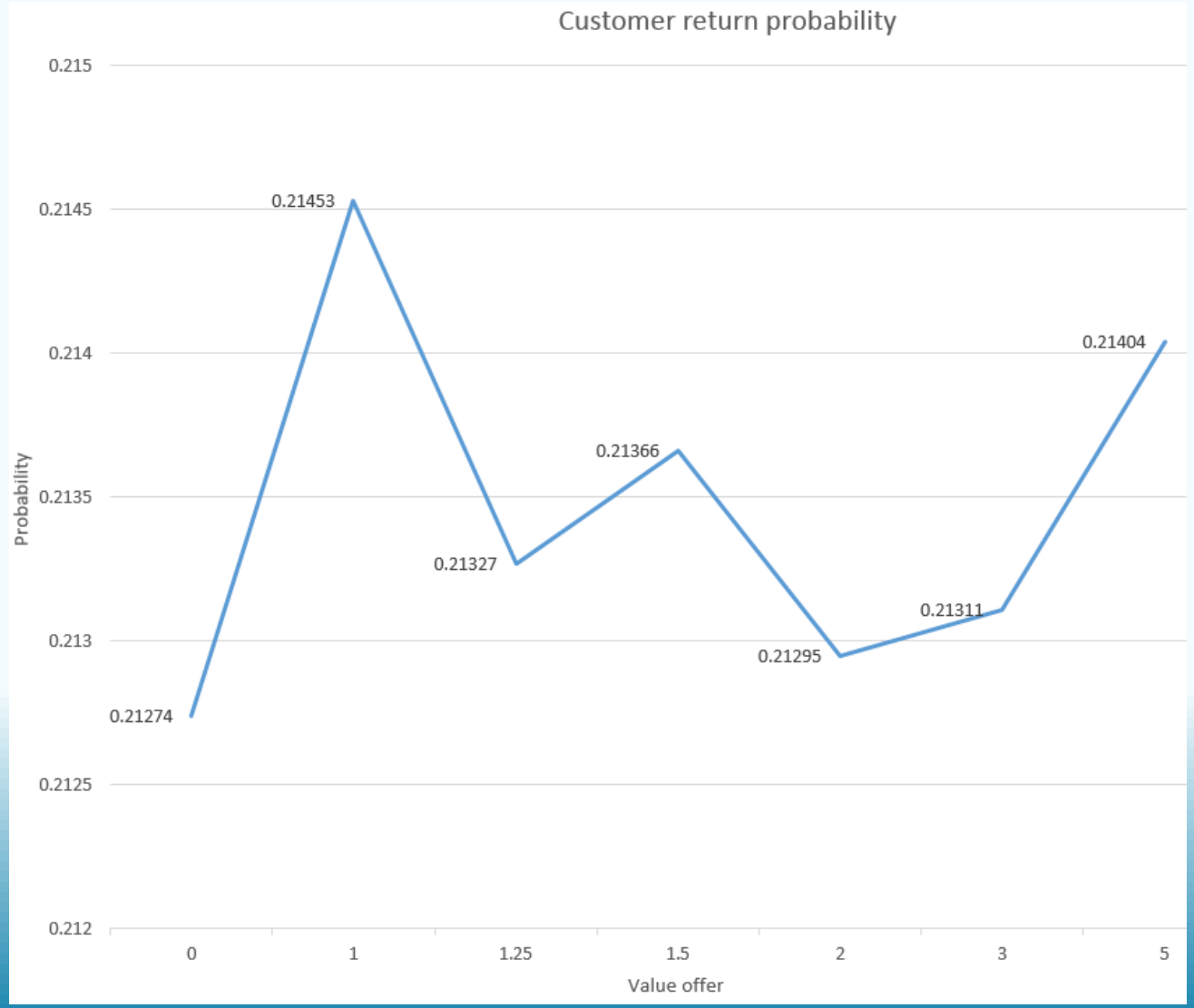
67	↓20	JustQ	0.60279	94	Mon, 14 Jul 2014 22:54:10 (-4.9h)	

FEATURES



- Value and brand influence customer return the most
- Negative influence of category.

OFFER



- Highest return probability when value=1
- A higher value does not make a returning customer

ROI

	Without coupons	With coupons
# customers	300000	300000
Return probability	0.21274	0.21453
# returning customers	63822	64359
Average yearly visits	6.5	6.9
Average spend / visit	£ 8.50	£ 9.20
Revenue	**£ 16,575,000.00**	**£ 19,044,000.00**
Offline and online marke	£ 300,000.00	£ 800,000.00
Outstanding operations	£ 200,000.00	£ 450,000.00
Marketing costs	£ 500,000.00	£ 1,250,000.00
GROSS MARGIN	**£ 16,075,000.00**	**£ 17,794,000.00**
Margin per customer	£ 53.58	£ 59.31
Difference per customer	/	**11%**

- Margin per customer: +11%

- Marketing Investment: £ 750,000
- Revenue Increase: £ 1,719,000.00
- **ROI = 229,2%**

CONCLUSION

- Ridge Regression with Cross-validation
- **Kaggle score = 0.60303**
- Value and brand are the most important features
- An offer value of 1 is best
- ROI = 223%

THANK YOU FOR YOUR ATTENTION!

Any questions?

For more information:

info@thesaintgroup.com

The Saint Group

√2 Pines St

KY16 9SX St Andrews