

# DAT Course Project

**Steve Tang**

August 2015

# Problem & Approach

Can we predict customer retention based on product usage and business profile?

- Retention modeling currently based on trending data and ad-hoc “black box” projections
- Product usage and performance not directly used to drive retention forecasting

# Product Overview

DF (“the product”) is a patient communication tool for dental practices to:

- Create email marketing campaigns,
- Manage patient appointments, and
- Solicit customer reviews

# Customer Analysis Approach

## Business Metrics

Retention Metrics

NPS - Net  
Promoter Score

tNPS - Transaction



## Customer Benefit

Time-to-Value

ROI

Time Savings



# Data Set Overview

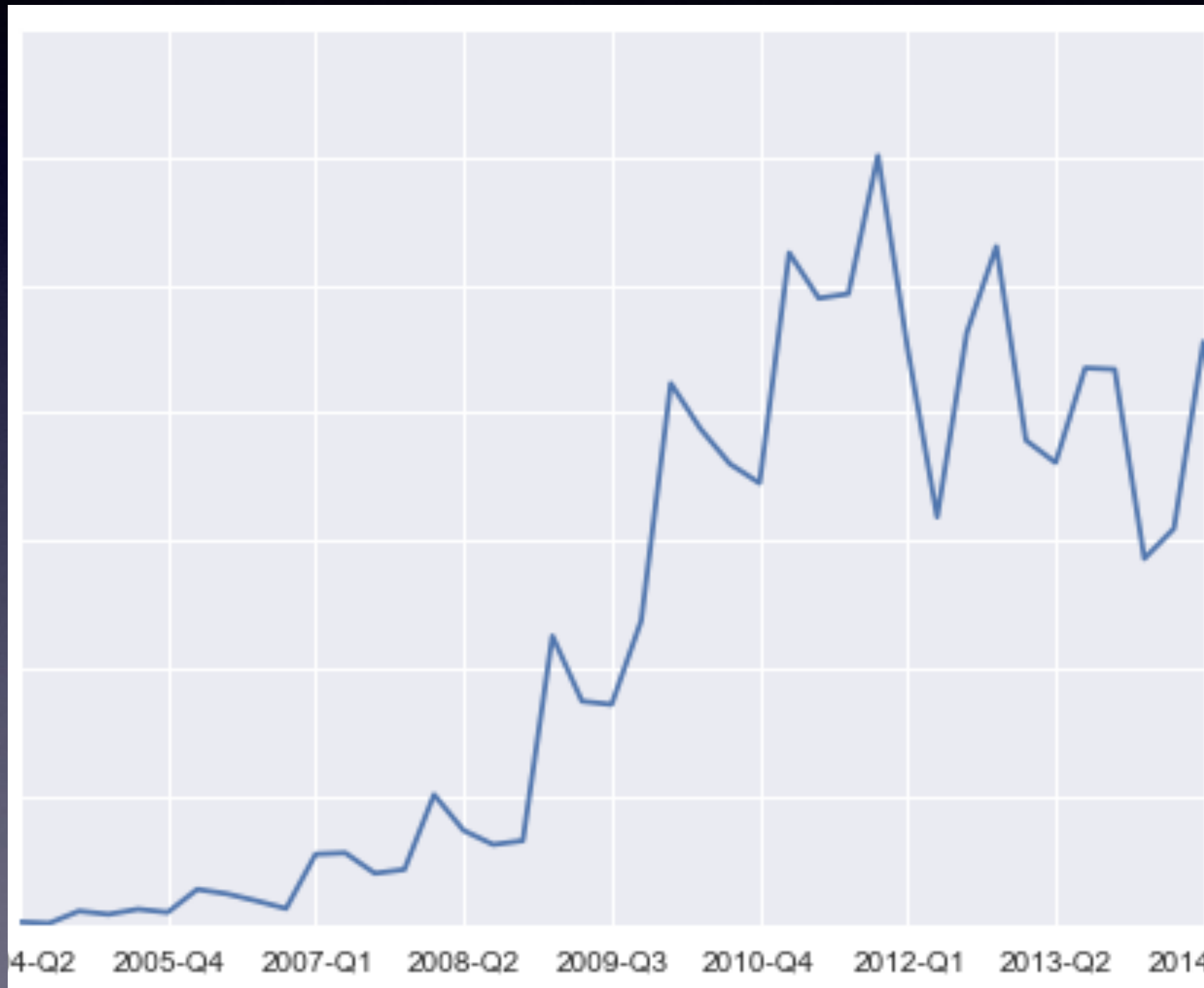
**20,049 records of both existing & cancelled accounts:**

1. **Customer engagement** - Last Login, Monthly Appointments, Monthly Campaigns, Monthly Emails Sent, Monthly Custom Campaigns
2. **Marketing effectiveness** - DF Appointments Requested, DF Appointments Confirmed, DF Appointment Contribution, Visit / Email Sent Ratio
3. **Business management** - Public Reviews, Surveys, Satisfaction Index, Appointment Confirmation Rate
4. **Integration status** - Management System
5. **Business profile** - Subscription Lifetime, Customers, Emails, Tech Savvy (Website, Google+, Yelp, FB), Lead Source Partner, Location, Specialty, Subscription Price

# Data Prep & Challenges

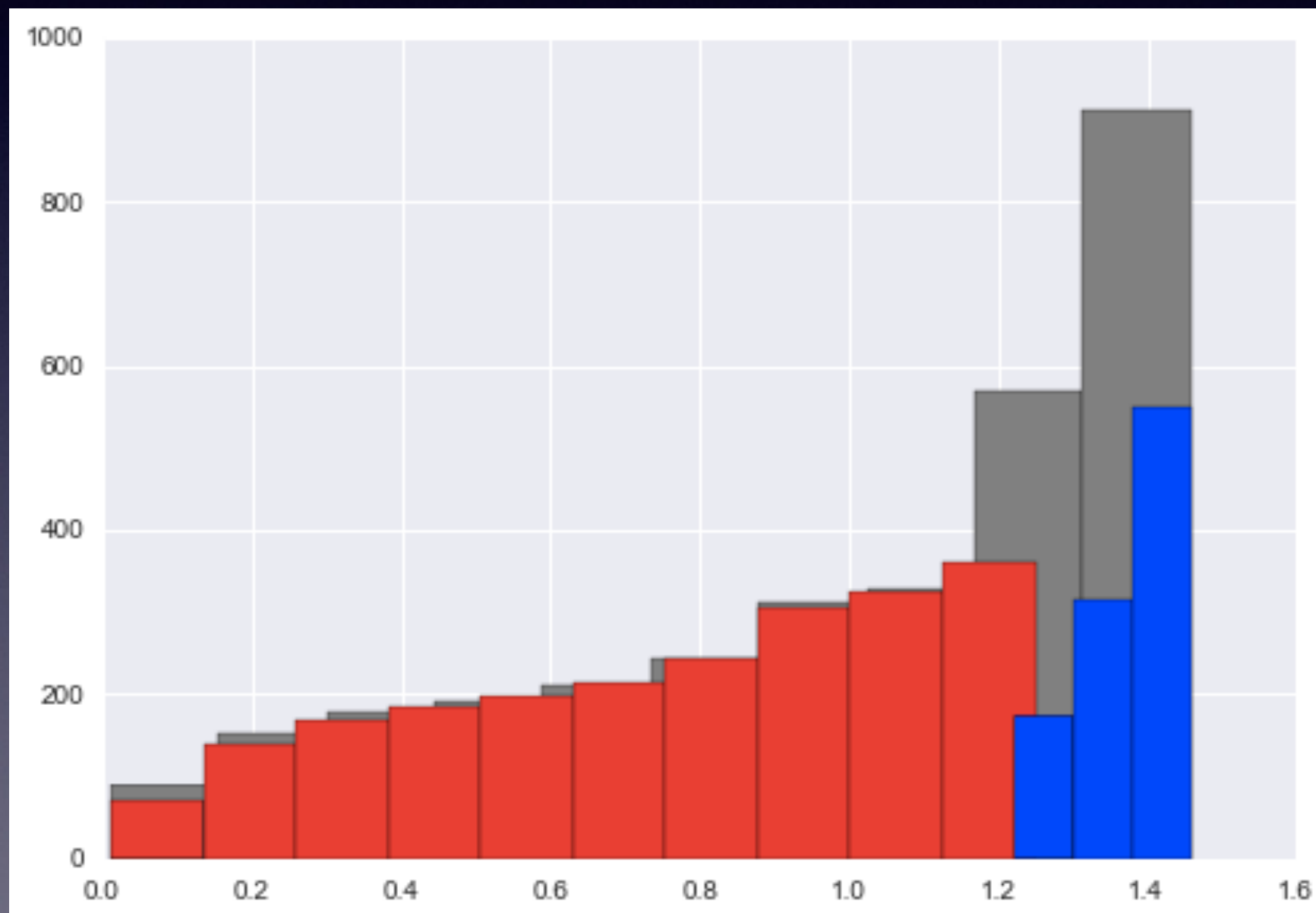
- Potential analysis limitations due to product engagement data being mostly point-in-time data (lacks time series trending data)
- Cohort analysis not clean due to changes in customer onboard processes, product reliability, and competitive landscape
- Incomplete data with unclear reason why null

# Account Adds by Cohort



# Q3'14 Cohort

## Customer Lifetime (Active & Cxl)

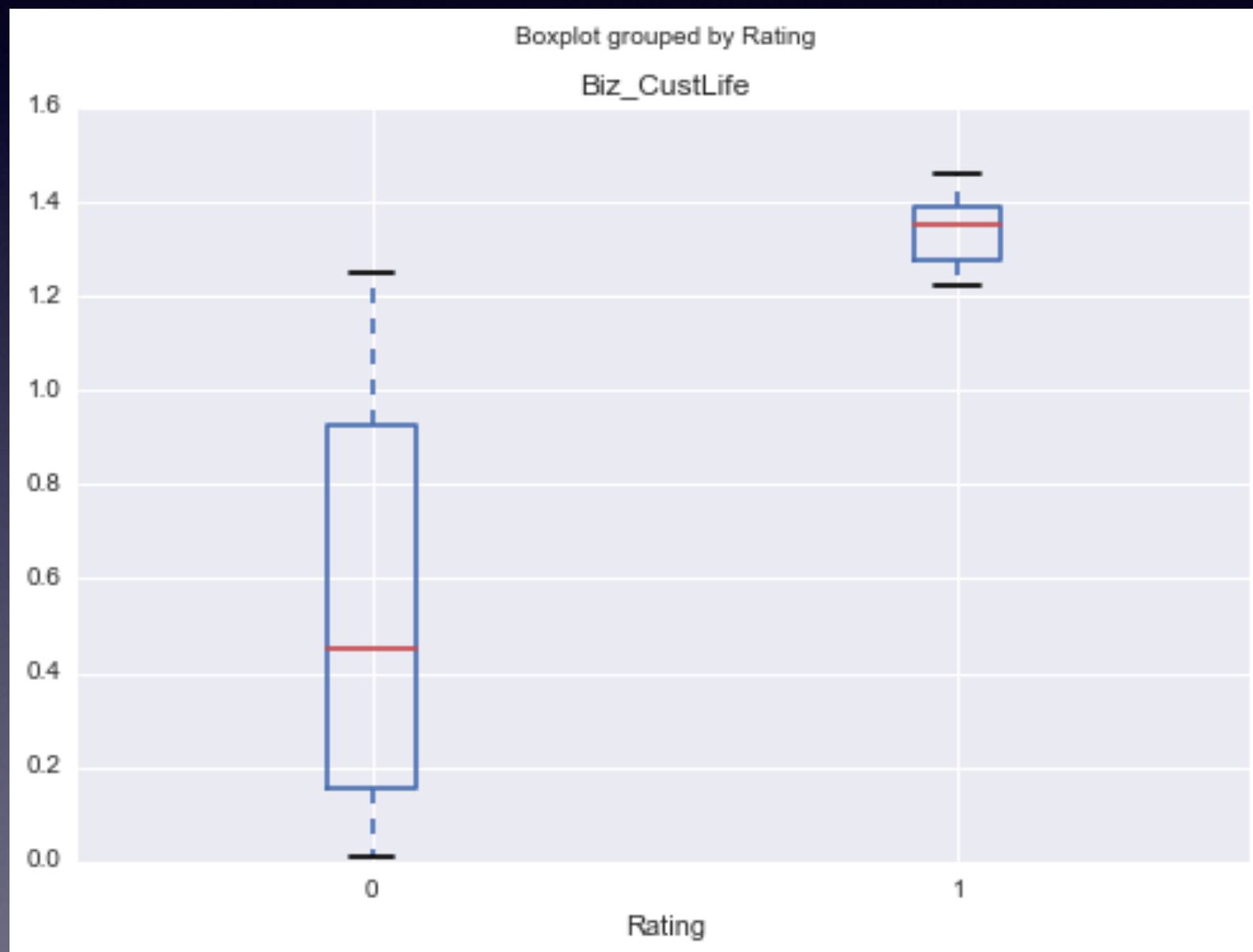


- 360 Total Cxls (35% of Q3'14 sign-ups)
- Trial Cxls <2 Mo: 27%, 98 accounts
- OOC's 2-12 Mo: 58%, 207 accounts
- 1-Yr Billable Cxl: 15%, 55 accounts



# Q3'14 Cohort

## Customer Lifetime (Active & Cxl)



### Q3'14 Retention:

- 911 sign-ups
- 360 cancels
- 65% 1-yr retention

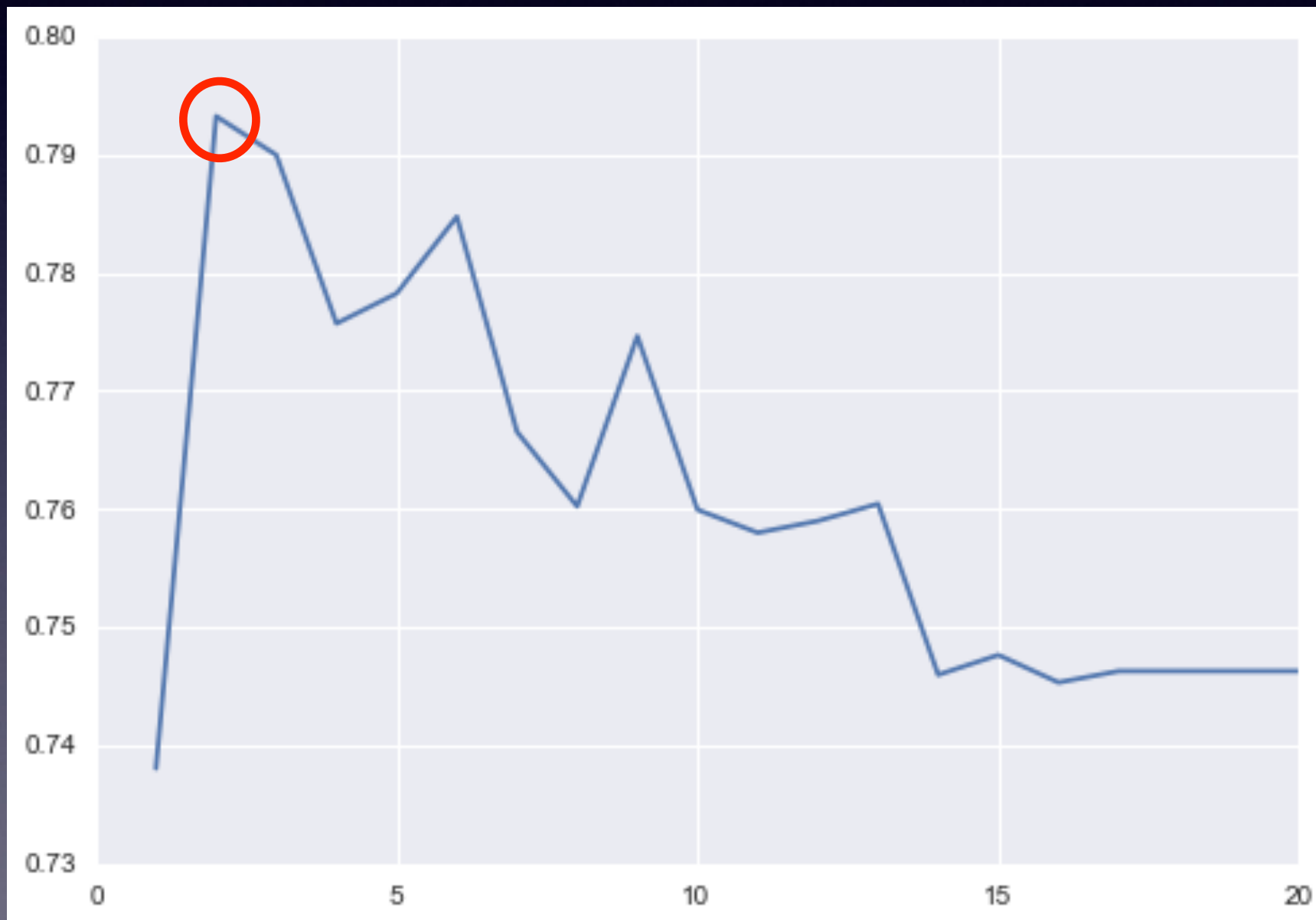
### Median Lifetime:

- Active - 1.35 yrs
- Cxl - 0.45 yrs

# Hypothesis

Customer retention can be accurately predicted by leveraging product engagement metrics, key business attributes, and level of tech savviness

# Decision Tree Classifier



## Decision Tree

(Depth: 2; 81% ROC-AUC  
Cross-Val: 79%)

**Appt Confirm Rate:** 0.873

**Campaigns Per Mo:** 0.112

**Appt Sent Per Mo:** 0.014

# Feature Prioritization

## P-Values

Features	P-Values
Cum Public Reviews	0.00E+00
Appt Sent Per Mo	0.00E+00
Campaigns Per Mo	0.00E+00
Appt Request Per Mo	0.00E+00
Email Count	0.00E+00
Customer Count	3.42E-11
Google+	2.27E-10
Appt Confirm Rate	4.96E-05
Facebook	8.44E-05
Stacked Discounting	1.44E-02
Yelp	1.76E-02
Partner Deal	2.44E-01
Website	9.23E-01
Reviews Per Mo	9.47E-01

## Decision Tree

(Depth: 13; 74% ROC-AUC)

Features	CART
Appt Confirm Rate	0.518
Campaigns Per Mo	0.152
Email Count	0.082
Appt Sent Per Mo	0.064
Appt Request Per Mo	0.058
Stacked Discounting	0.035
Customer Count	0.033
Cum Public Reviews	0.018
Reviews Per Mo	0.014
Partner Deal	0.014
Google+	0.013
Website	-
Yelp	-
Facebook	-



# Model Evaluation

**Q3'14 Cohort - Cross Val ROC-AUC**



# Logistic Regression

## Confusion Matrix

	Cxl (Pred)	Active (Pred)
Cxl (Actual)	97	7
Active (Actual)	7	117

$N_{Q3'14 \text{ Test}} = 228$

Active = 124

Cxl = 104

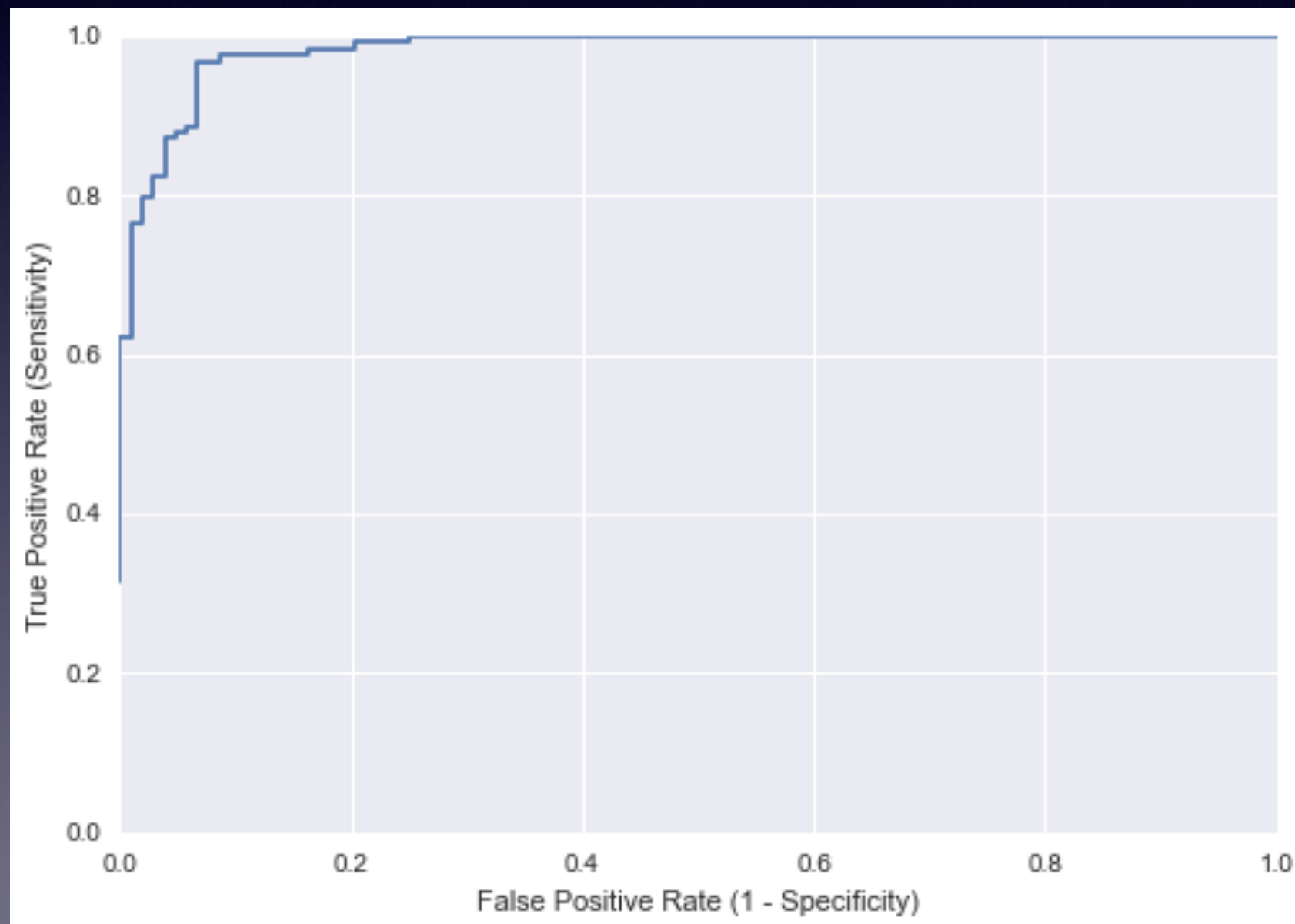
Sensitivity = 0.94

Specificity = 0.93

Accuracy = 0.94

# Logistic Regression

## ROC Curve



# Feature Reduction

Appt Confirm Rate
Campaigns Per Mo
Appt Sent Per Mo
Appt Request Per Mo
Reviews Per Mo
Cum Public Reviews
Stacked Discounting
Partner Deal
Customer Count
Email Count
Google+
Website
Yelp
Facebook



*Feature  
Reduction  
Process*



Reviews  
Per Month  
  
Cumulative  
Public  
Reviews



# Logistic Regression

## Cross Val Results by Cohort

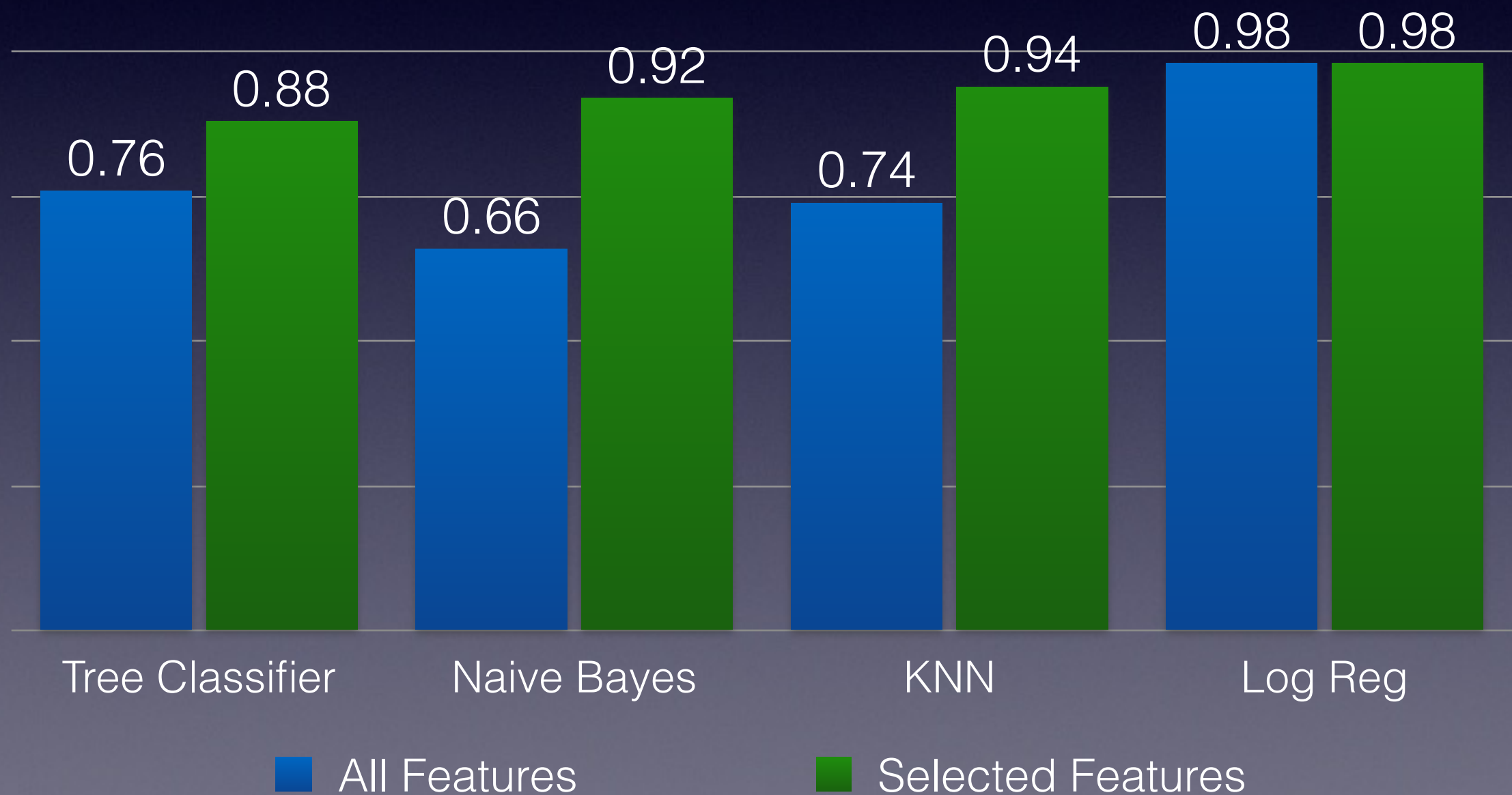
	Q3'14	Q2'14	Q4'14
N	911	619	683
% Retention	60%	67%	64%
ROC-AUC (All Features)	0.98	0.84	0.99
ROC-AUC (Select Features)	0.98	0.86	0.99

All Features (14) includes Appt Confirm Rate, Campaigns Per Mo, Appt Sent Per Mo, Appt Request Per Mo, Reviews Per Mo, Cum Public Reviews, Stacked Discounting, Email Count, Customer Count, Partner Deal, Website, Google+, Yelp, Facebook.

Select Features (2) includes Reviews Per Mo, Cum Public Reviews.

# Model Evaluation v2

## Q3'14 Cohort - Cross Val ROC-AUC



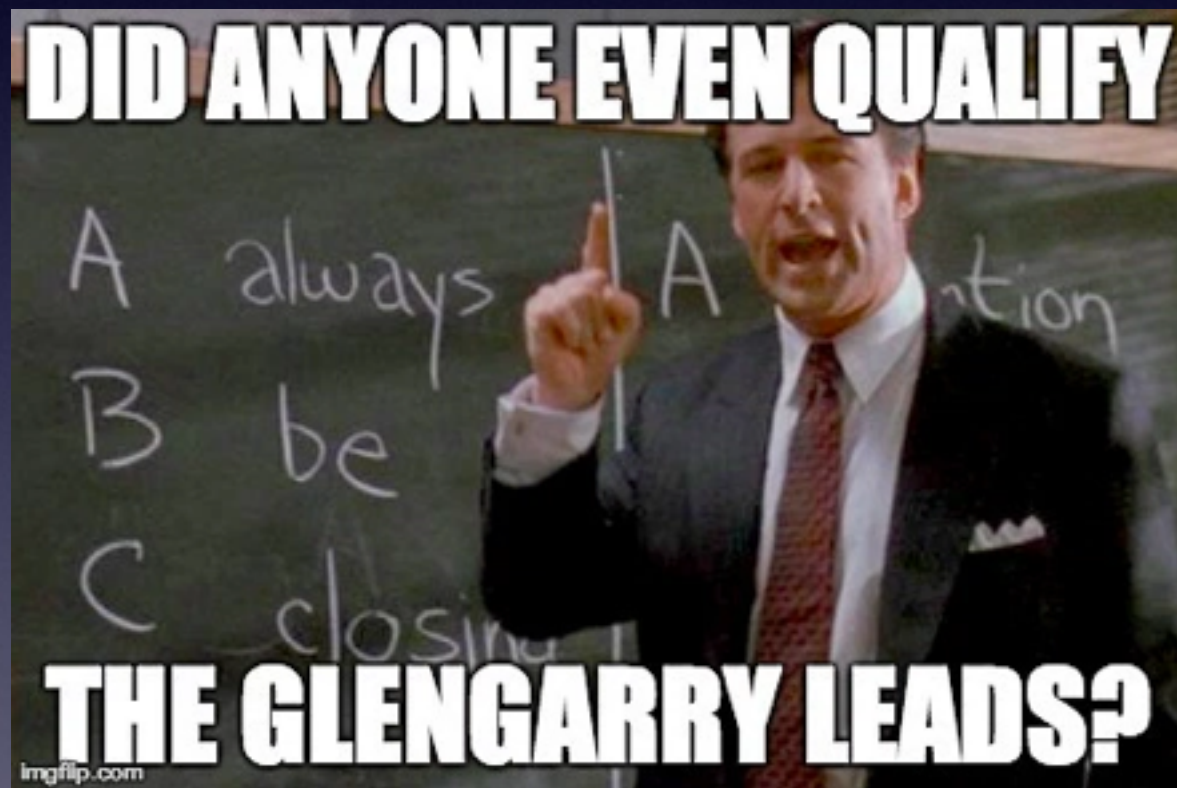
# So what?

Most important factor...

Reviews, reviews, reviews!



# Potential Issues w/ Analysis





# What next?

Create predictive model for review generation

Decouple Trial Retention and Billable Retention

NLP analysis on customer support case notes to evaluate impact on other model features

Clustering analysis for account cohorts

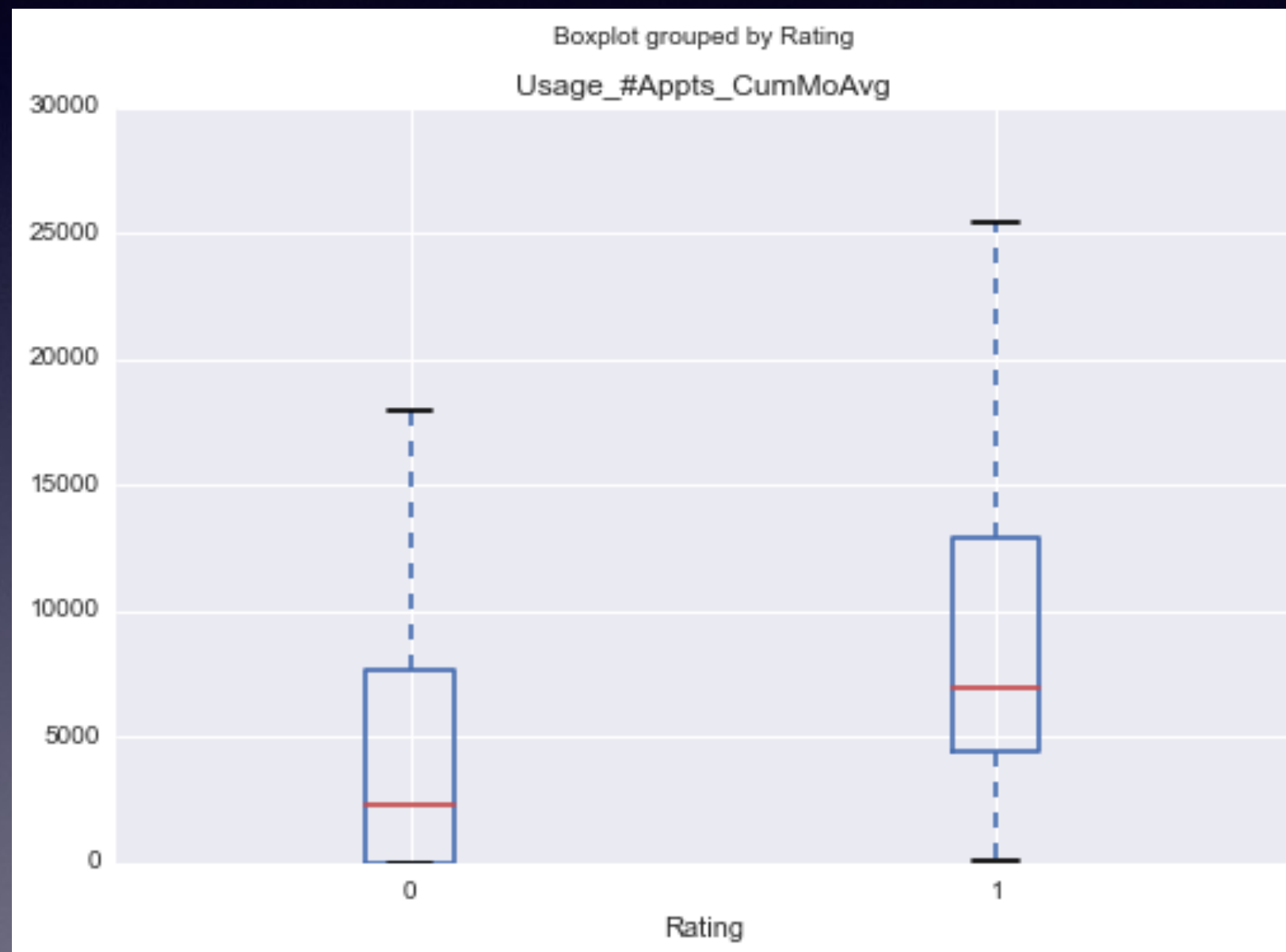
# Appendix

# Dataset Overview

- **Subscription data:** u'SubscriptionID', u'AccountID', u'Sub\_Start\_Date', u'First\_CXSent\_Date', u'MonthsUnderContract', u'StackedDiscount\_Ind', u'CancelOnFile\_Ind', u'CancelID', u'Cancel\_RecType', u'Cancellation\_Date', u'Cancellation\_Requested\_Date', u'D3 Business ID', u'Rating'
- **Product Usage data:** u'D3 Last Login Date to Portal', u'D3 Last Campaign Sent', u'D3 Email Enabled', u'D3 Number of Appointments', u'D3 Number of Campaigns Sent', u'D3 Total Emails Sent', u'D3 Custom Campaigns Emails Sent', u'D3 OOB Campaigns Enabled', u'D3 Number of Appointments Requested', u'D3 Demandforce Appointments (90 Days)', u'D3 Total Visits', u'D3 Public Reviews', u'D3 Public Reviews in Last 60 Days', u'D3 Satisfaction Index', u'D3 Survey Last 60 Days', u'D3 Confirmed Appointments (90 Days)', u'D3 Total Appointments (90 Days)', u'D3 Confirmed Appointments', u'D3 Welcomes Last 60 Days', u'D3 Number of Email Optin', u'D3 First Upload Email Count', u'Cumulative # of Email Finder'
- **User Profile data:** u'Mgmt\_Sys', u'D3 Web Alias', u'Google+', u'D3 Yelp Account', u'D3 Facebook Account', u'#Customers', u'#Emails', u'Primary Partner', u'Exclusive Distributor', u'Henry Schein Deal', u'Billing Zip/Postal Code', u'Industry', u'Specialty',
- **Engineered features:** u'Biz\_CustLife', u'Usage\_#Appts\_CumMoAvg', u'Usage\_#Camp\_CumMoAvg', u'Usage\_#DFApptReq\_CumMoAvg', u'Usage\_#PublicReviews\_CumMoAvg', u'Usage\_%ApptConfirmRate\_Cum', u'cohort'

# Q3'14 Cohort

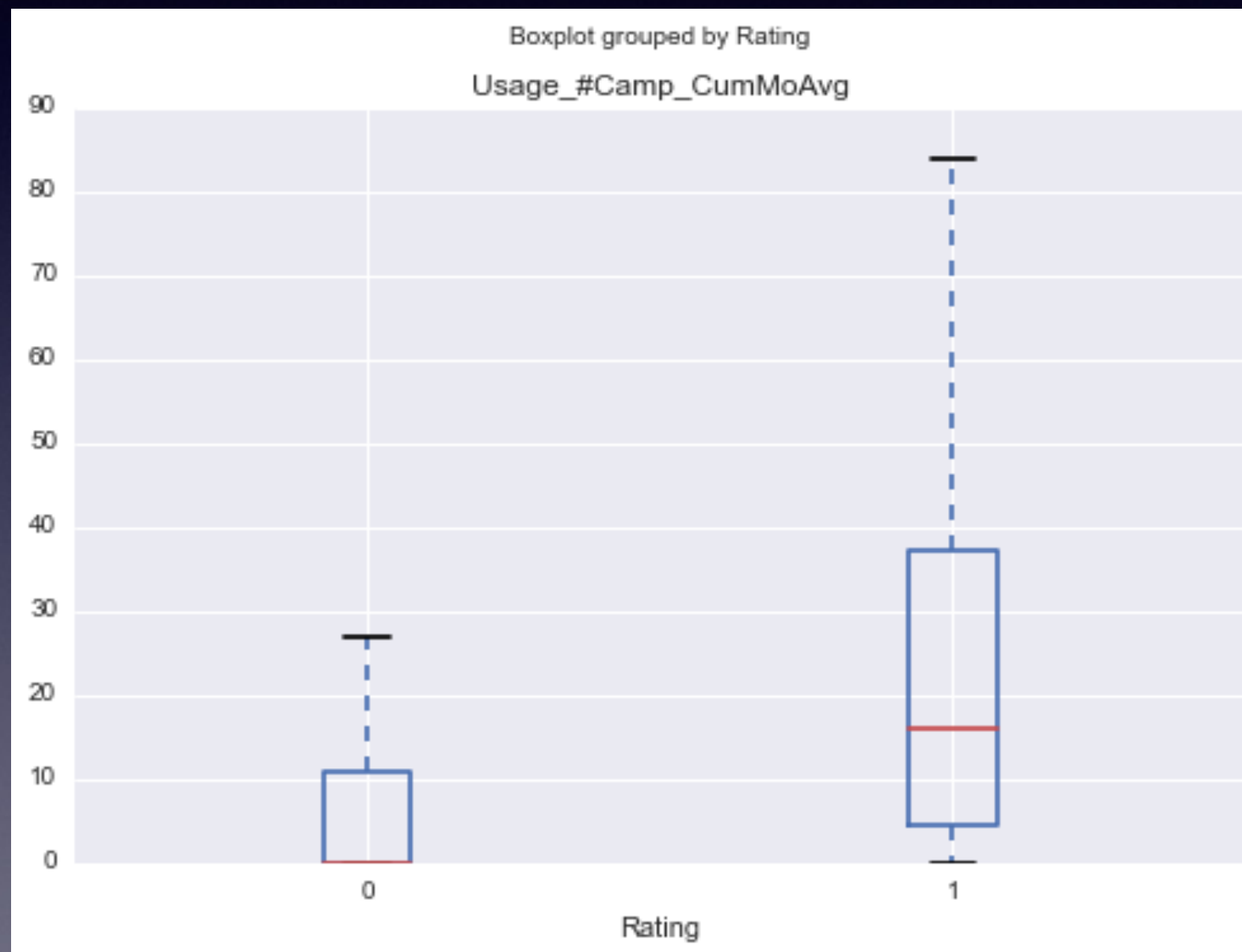
## Appointments Sent Per Mo





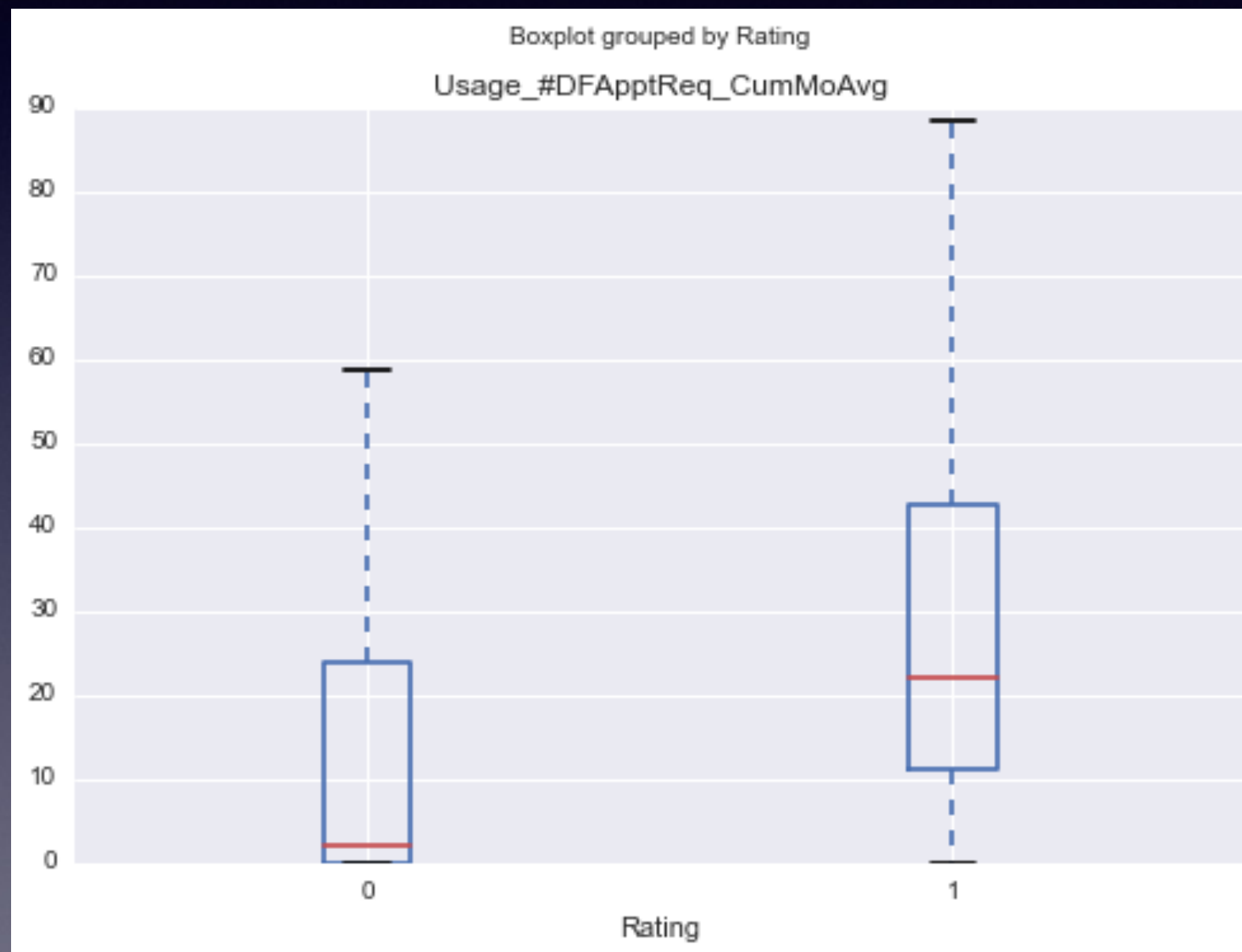
# Q3'14 Cohort

## Campaigns Sent Per Mo



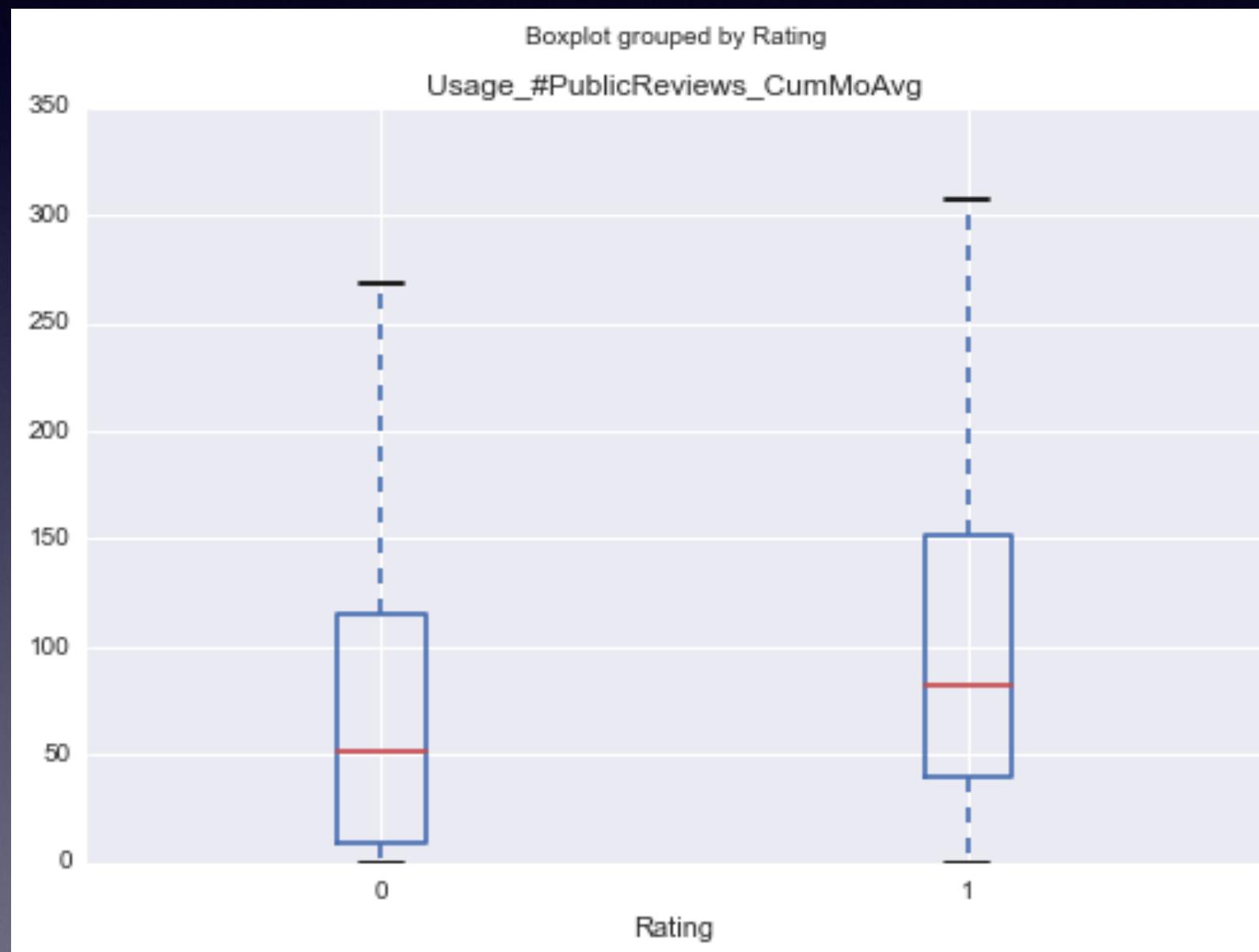
# Q3'14 Cohort

## DF Appt Requested Per Mo



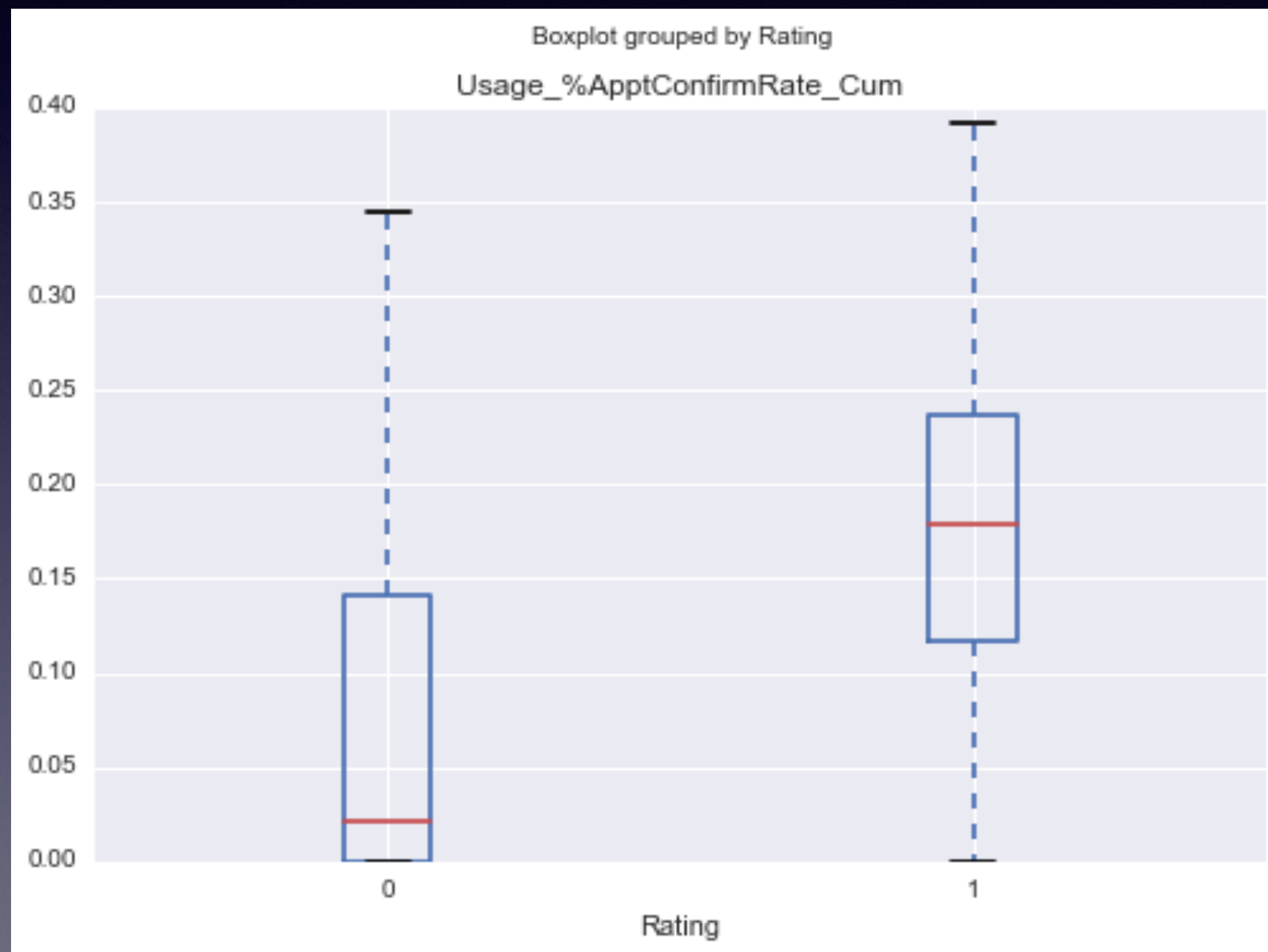
# Q3'14 Cohort

## Public Reviews Per Mo



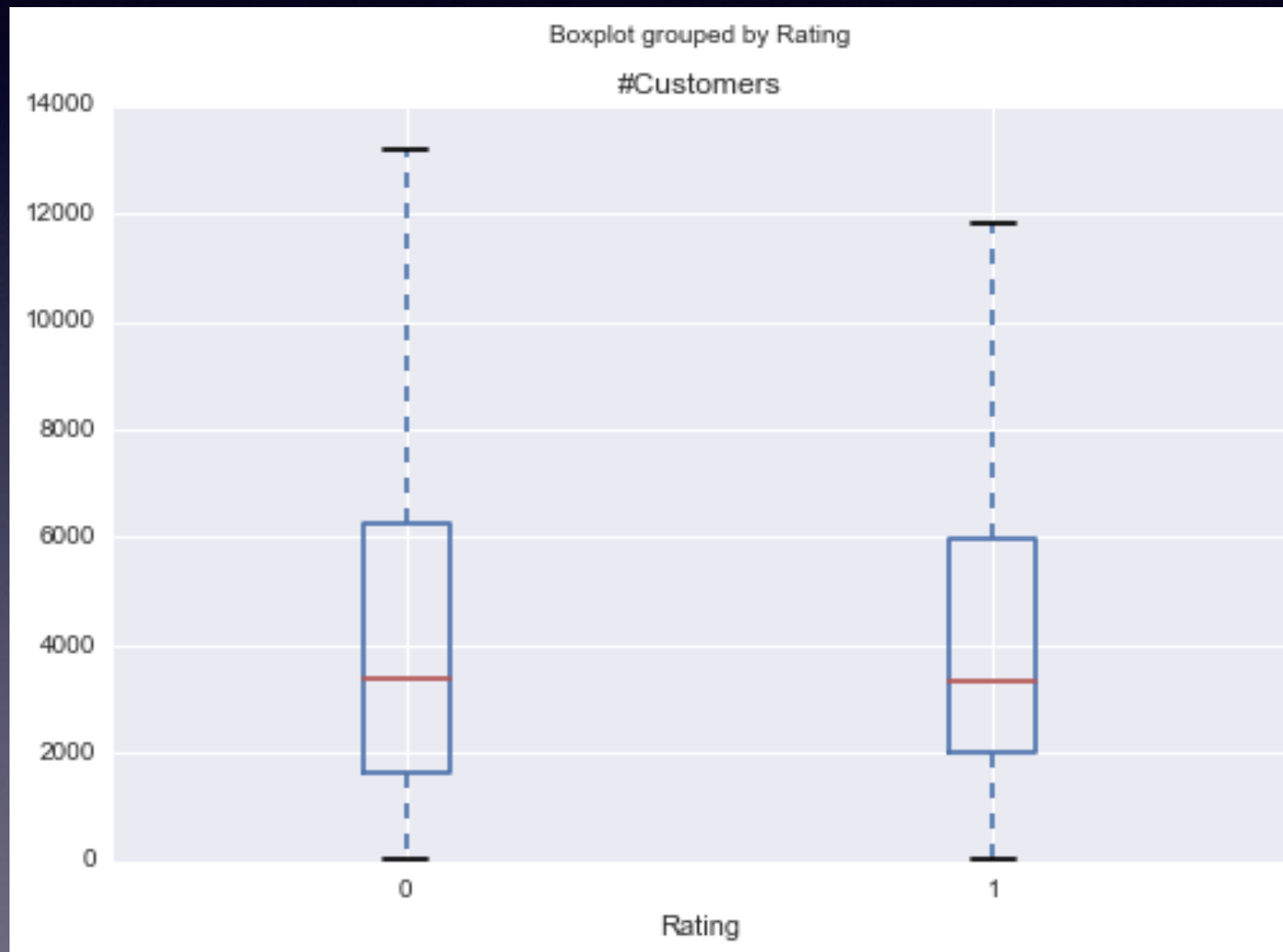
# Q3'14 Cohort

## Appt Confirmation Rate %

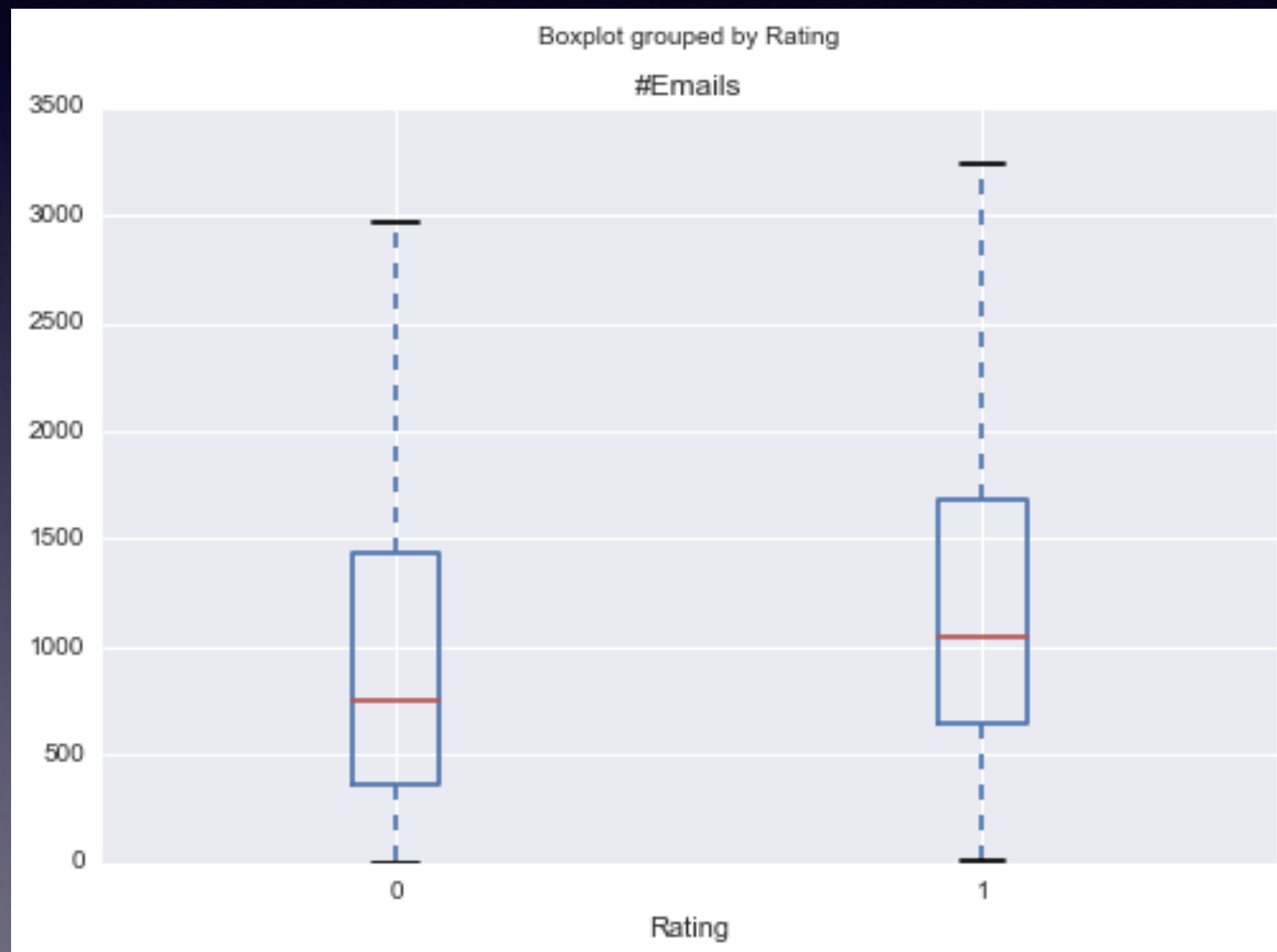




# Q3'14 Cohort Customer Count



# Q3'14 Cohort Email Count



# Q3'14 Cohort

## Additional Features

Additional Features	Active	Cancelled
Stacked Discounting	65%	79%
Website	100%	99%
Google+	66%	35%
Yelp	29%	21%
Facebook	78%	56%
Henry Schein Deal	48%	55%