Business Case:
Scaling CatskillProvisions.com

CATSKILL
provisions

Honey Maple Syrup Sauces Etc. Gifts Honey Whiskey Wholesale About

BACKGROUND

- Small ecommerce website with growth intentions
- Originated in 2010 out of a passion for bee keeping
- First product sold online was honey
  - Honey continues to be the primary product sold along with other gift items
- This analysis is the first time anyone has looked at the backend of the website: customers/traffic/sales conversion

CATSKILL provisions

TALE OF TWO BUSINESSES

- Two parallel businesses for Catskill Provisions
  - Primary: selling products (honey, whiskey) wholesale to restaurants, liquor stores, specialty shops
  - Secondary: ecommerce website
- OPPORTUNITY: pursue ecommerce website given interest to scale web traffic, converting traffic to sales and delivering revenue improvement

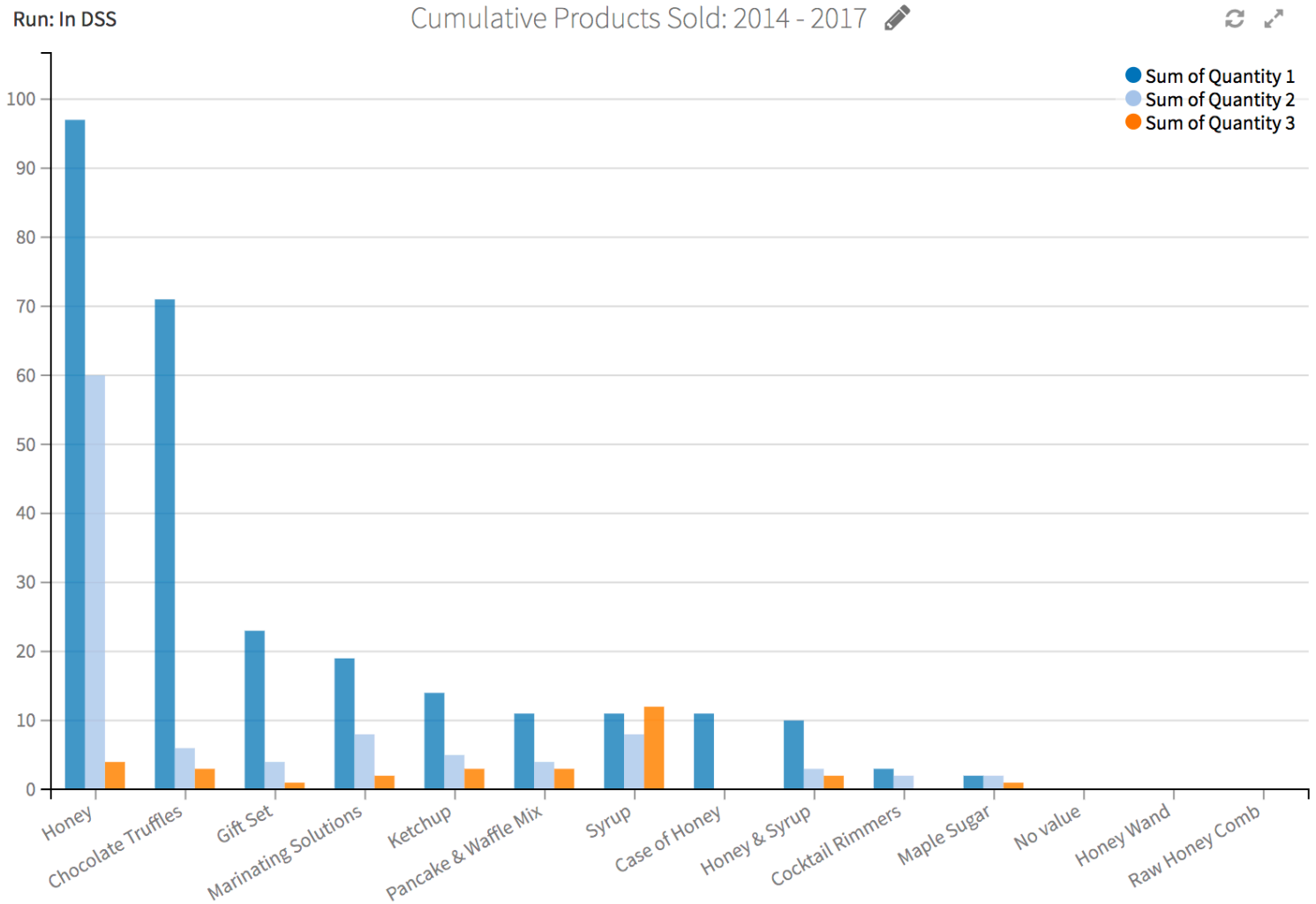
CATSKILL provisions

EDA: WEB ECOMMERCE CUSTOMER PROFILE

- Female (76%)
- Region: East (NY, NJ), New England (67%)
- 1x purchaser (85%)
- Tuesday/Thursday web shopper (shopping from work)
- Most purchased products: Honey, Truffles, Gift Sets
- Product associations: Honey wands with Honey; Maple Syrup with Honey (Apriori algorithm)
- Email Domain: Gmail; Yahoo; AOL; work domain

CATSKILL provisions

EDA: WEB STORE PRODUCTS



CATSKILL provisions

EDA: BUSINESS SNAPSHOT

Year	Customers	Avg Sales	Repeat Customer % Total Sales	Visits Conversion	Unique Visits Conversion
1Q2017	91	$22.52	30.7%	1.56%	2.17%
FY2016	312	$40.10	26.3%	1.03%	0.79%
FY2015	320	$37.50	16.7%	0.90%	0.77%
FY2014	125	$54.86	6.6%	0.70%	0.39%

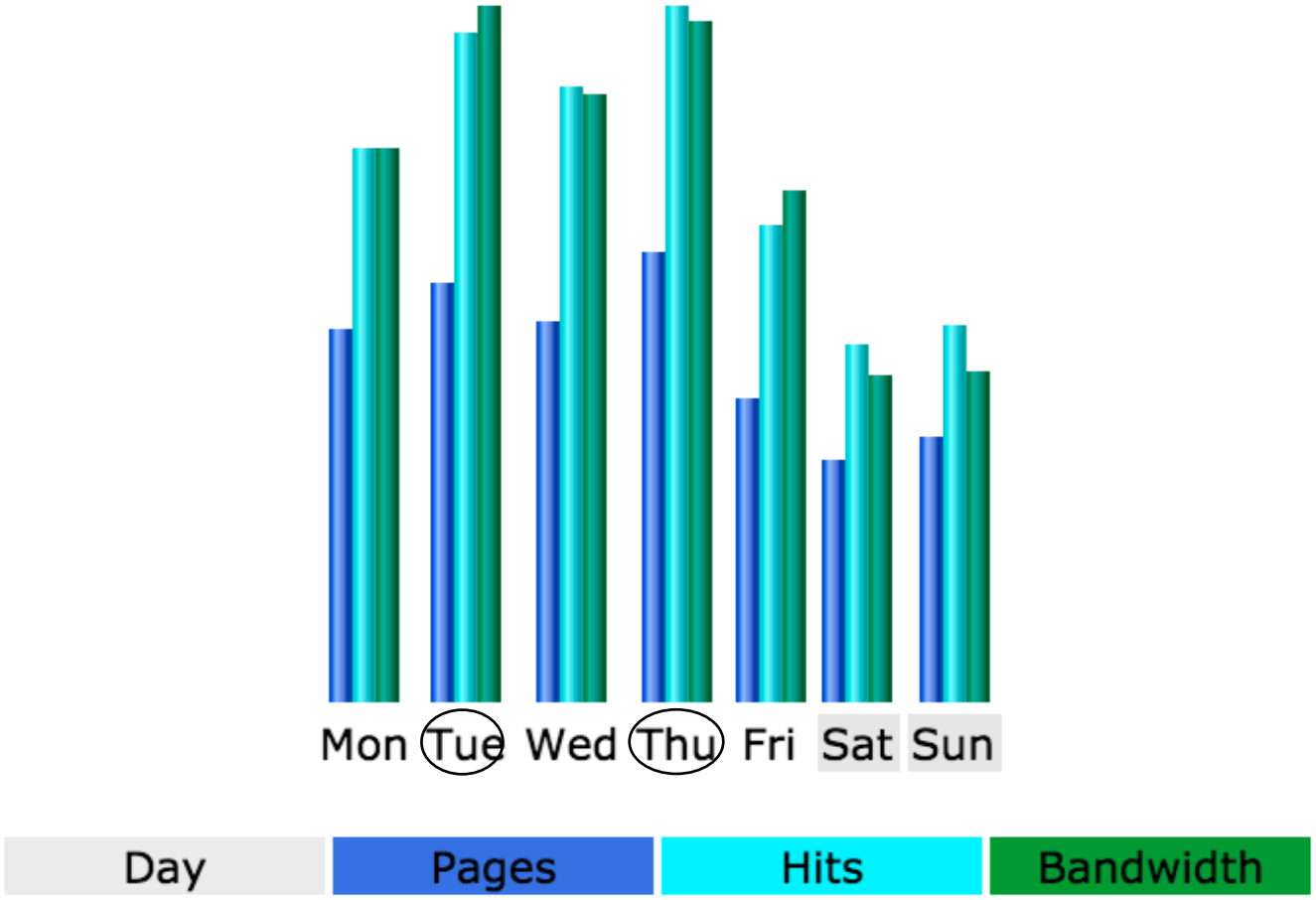
CATSKILL provisions

EDA: WEB TRAFFIC CHANGES 2015 – YTD

Year	Unique Visitors % Chg	No. of Visitors % Chg	Page Views % Chg	Site Hits % Chg
1Q2017	-53.7%	-59.2%	-59.0%	-49.5%
FY2016	-5.3%	-7.7%	-8.8%	-11.7%
FY2015	29.2%	32.9%	79.1%	30.9%

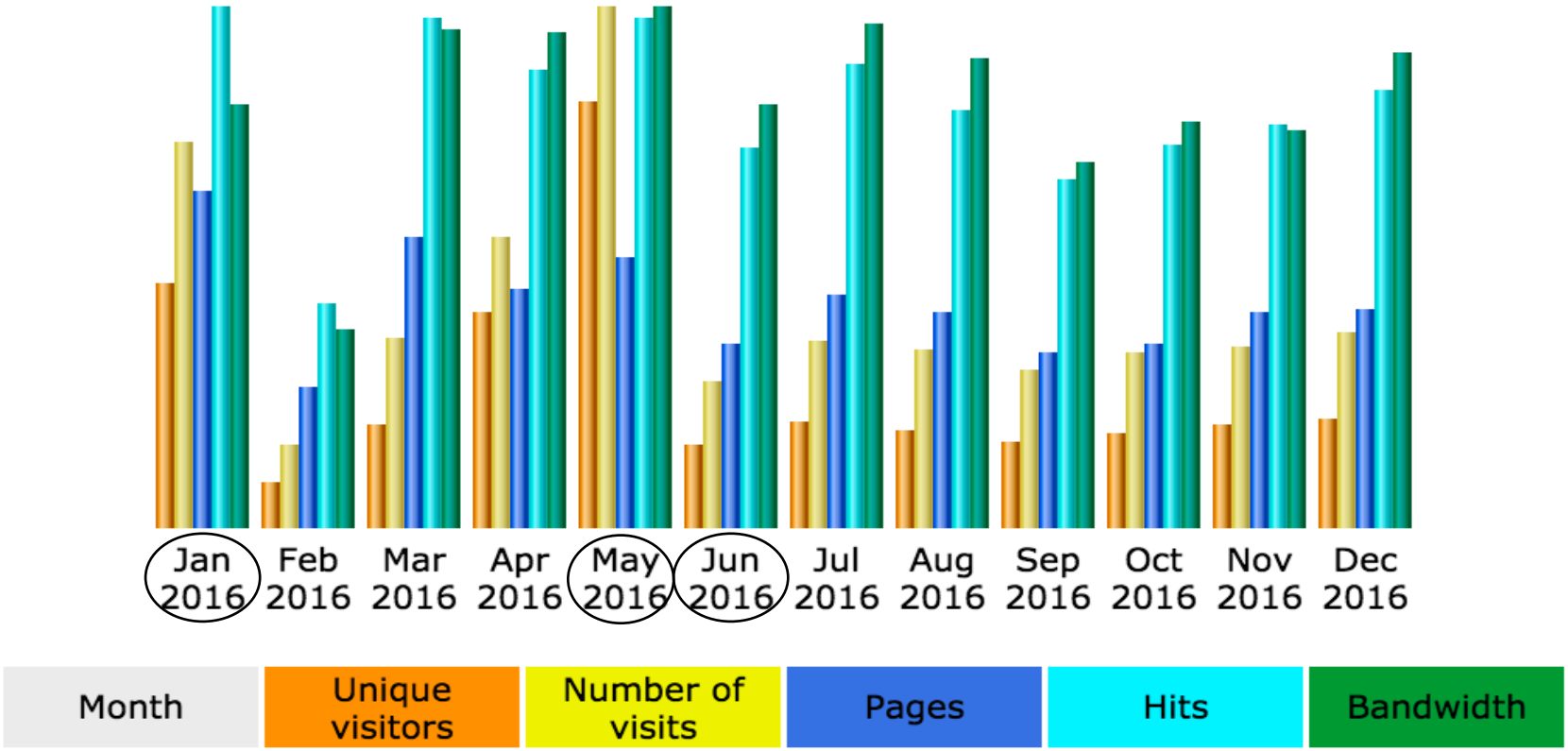
CATSKILL provisions

EDA:
PURCHASE BEHAVIOR – DAY OF WEEK



CATSKILL provisions

EDA:
PURCHASE BEHAVIOR – TIME OF YEAR



CATSKILL provisions

EDA: SEO WORDCLOUD – GOOGLE PRIMARY REFERER

CATSKILL provisions

EDA: TOP SEARCH PHRASES

do bees eat nectar
what do bees eat
were do the bees store the necter after collecting it from the flowers answer
how bees eat poklen
are bees can eat nectar
about catskill provisions
what do bees eat for food
what is the food of honey bees
what do bees do with nectar
searching of food in honey bee
how do bee s eat
how does a bee feed her babies
what do bumble bees eat and drink
how do bees eat

CATSKILL provisions

DATA SET VARIABLES

- Small website, small data set – trained
  - Combination of transactional and web traffic data
  - Each row represents transactional data
- Key variables in data set
  - Transaction date/day of the week
  - Shipping & Billing State & Region
  - Repeat Purchases
  - Email domain
  - Daily Web Visits/Web Hits
  - Purchase Total
  - Total Order Quantity
  - Product for Sale
  - Sales
  - Sales to Web Visit Conversions

CATSKILL provisions

DATA SET VARIABLES

- Certain features highly correlated
- Clustering and PCA not useful
- Modeled Features:
  - Shipping Region
  - Repeat Purchasers
  - Gender
- Other Features Modeled:
  - Total Purchase Quantity
  - Email Domain
  - Web Visits
  - Sales

CATSKILL *provisions*

SHIPPING REGION: MODEL OUTPUT

ACTIONS	Name	Trained	Accuracy	Precision	Recall	F1 Score	Log Loss	ROC AUC
	Logistic Regression	2017-03-27 17:24:08	0.69	0.72	0.41	0.45	1.35	0.76
	SVM	2017-03-27 17:24:09	0.58	0.51	0.42	0.38	1.22	0.75
	Random forest	2017-03-27 17:24:08	0.46	0.46	0.41	0.34	1.60	0.71
	XGBoost	2017-03-27 17:24:29	0.60	0.50	0.34	0.33	1.39	0.70
	Gradient Tree boosting	2017-03-27 17:24:10	0.44	0.34	0.35	0.31	1.57	0.68
	K Nearest Neighbors (k=5)	2017-03-27 17:24:29	0.53	0.42	0.28	0.27	7.41	0.65
	Decision Tree	2017-03-27 17:24:28	0.61	0.27	0.25	0.25	2.54	0.58

CATSKILL
provisions

SHIPPING REGION BY PRODUCT QUANTITY

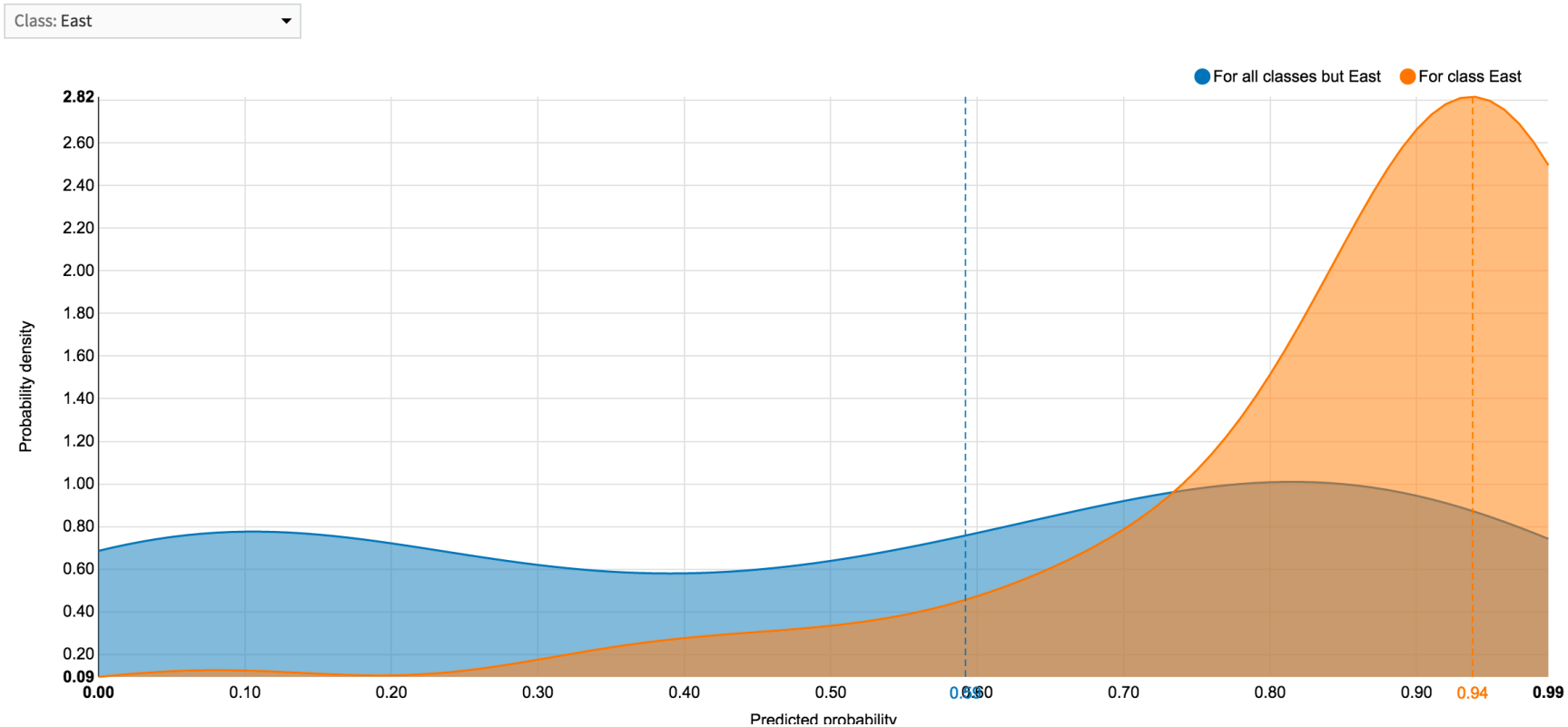
CATSKILL provisions

SHIPPING REGION: CONFUSION MATRIX

	Predicted						
Actual	East	South	West	MidWest	New England	MIdAtlantic	
East	95 %	0 %	3 %	2 %	0 %	0 %	100 %
South	67 %	33 %	0 %	0 %	0 %	0 %	100 %
West	56 %	0 %	44 %	0 %	0 %	0 %	100 %
MidWest	25 %	8 %	25 %	42 %	0 %	0 %	100 %
New England	78 %	0 %	0 %	0 %	22 %	0 %	100 %
MIdAtlantic	80 %	10 %	0 %	0 %	0 %	10 %	100 %

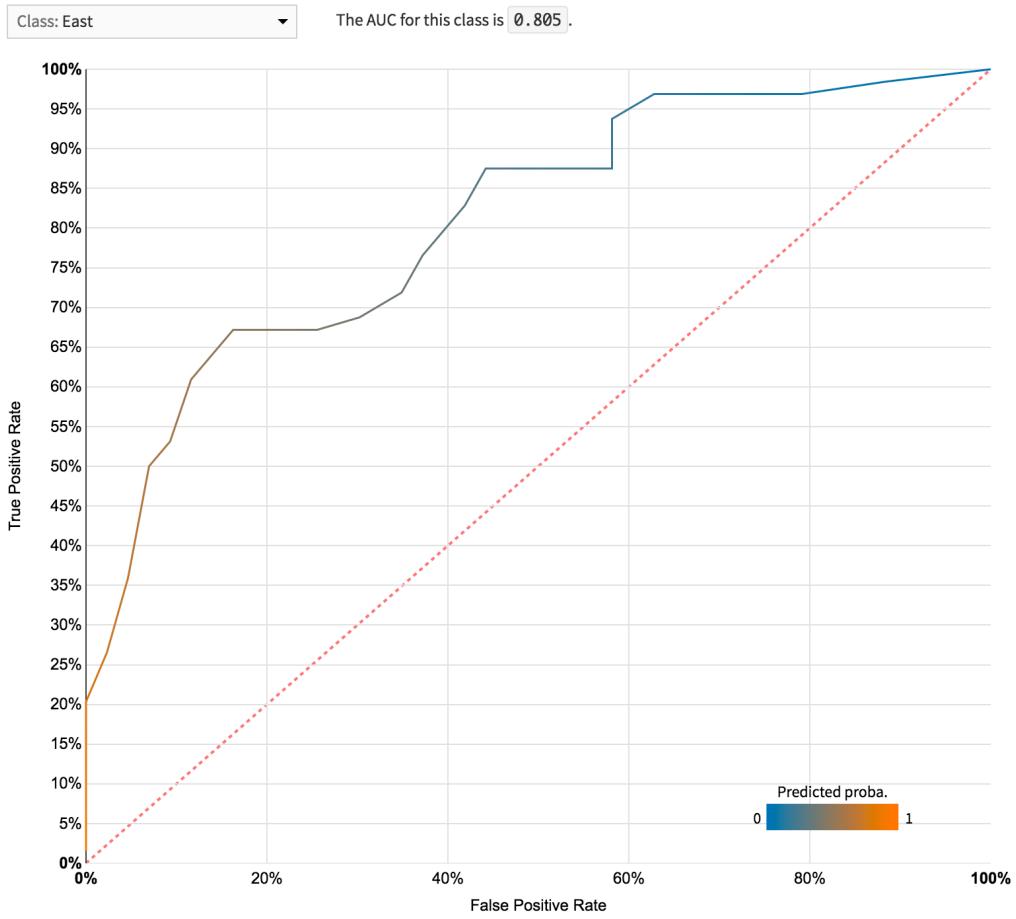
CATSKILL provisions

SHIPPING REGION: LOGISTIC DENSITY CURVE



CATSKILL provisions

SHIPPING REGION: LOGISTIC ROC AUC



Reading tips

The Receiver Operating Characteristic (or ROC) curve shows the true positive rate vs. the false positive resulting from different cutoffs in the predictive model. The "faster" the curve climbs, the better it is.

On the contrary, a curve close to the diagonal line is worse.

The MAUC (Multi-class Area Under the Curve) for this model is 0.759, which is **good**.

EXPORT DATA

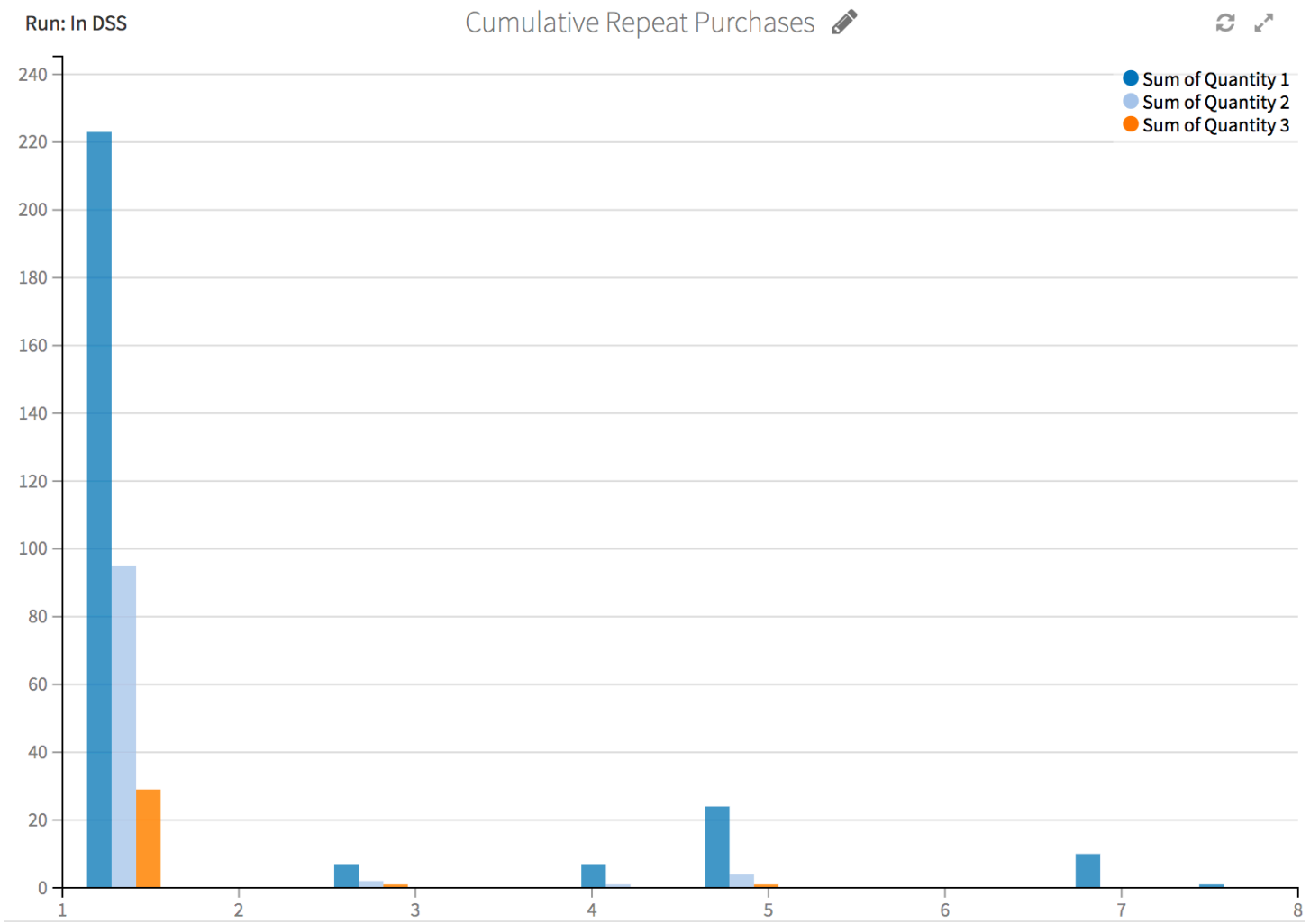
CATSKILL provisions

REPEAT PURCHASES: MODEL OUTPUT

ACTIONS	Name	Trained	EVS	MAPE	MAE	MSE	RMSE	RMSLE	R2 Score	Correlation
	Lasso (L1) regression	2017-03-26 17:28:29	0.89	11.8%	0.23	0.34	0.58	0.16	0.89	0.95
	XGBoost	2017-03-26 17:28:32	0.89	6.6%	0.17	0.36	0.60	0.17	0.88	0.94
	Ridge (L2) regression	2017-03-27 11:19:53	0.86	36.0%	0.46	0.46	0.68	0.23	0.85	0.94
	K Nearest Neighbors (k=3)	2017-03-27 11:19:54	0.80	8.1%	0.19	0.63	0.79	0.21	0.79	0.89
	Random forest	2017-03-26 17:28:28	0.75	34.5%	0.54	0.76	0.87	0.26	0.75	0.87
	Gradient Tree boosting	2017-03-26 17:28:29	0.71	39.3%	0.64	0.90	0.95	0.25	0.71	0.92
	Decision Tree	2017-03-26 17:28:31	0.57	40.8%	0.65	1.30	1.14	0.31	0.57	0.76

CATSKILL provisions

CUMULATIVE REPEAT PURCHASES

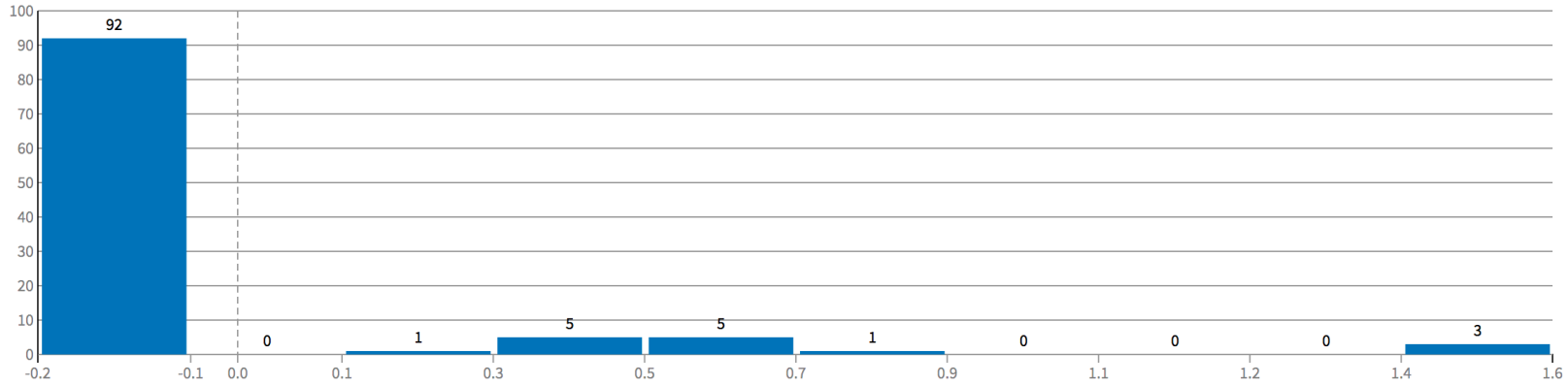


CATSKILL provisions

REPEAT PURCHASES: LASSO ERROR DISTRIBUTION

Minimum	25th perc.	Median	75th perc.	90th perc.	Maximum
-0.24403	-0.092757	-0.092757	-0.092757	0.40574	1.6177
	Average	0.019206		**Standard deviation**	0.34044

The errors (difference between predicted and actual values) should be centered around zero, and the distribution should be "narrow", i.e the spread of the error should be limited. More generally, the errors should be "normally" distributed around zero (the curve should look like a bell).

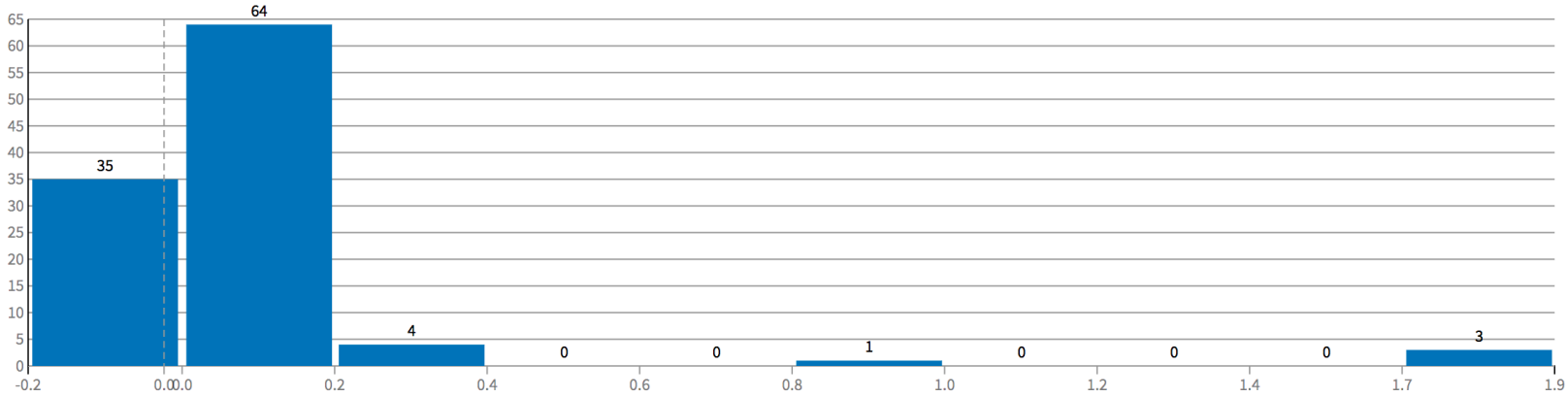


CATSKILL provisions

REPEAT PURCHASES: XGBOOST ERROR DISTRIBUTION

Minimum	25th perc.	Median	75th perc.	90th perc.	Maximum
-0.17948	-0.030143	0.024462	0.024462	0.16853	1.8572
	Average	0.072093		**Standard deviation**	0.33173

The errors (difference between predicted and actual values) should be centered around zero, and the distribution should be "narrow", i.e the spread of the error should be limited. More generally, the errors should be "normally" distributed around zero (the curve should look like a bell).



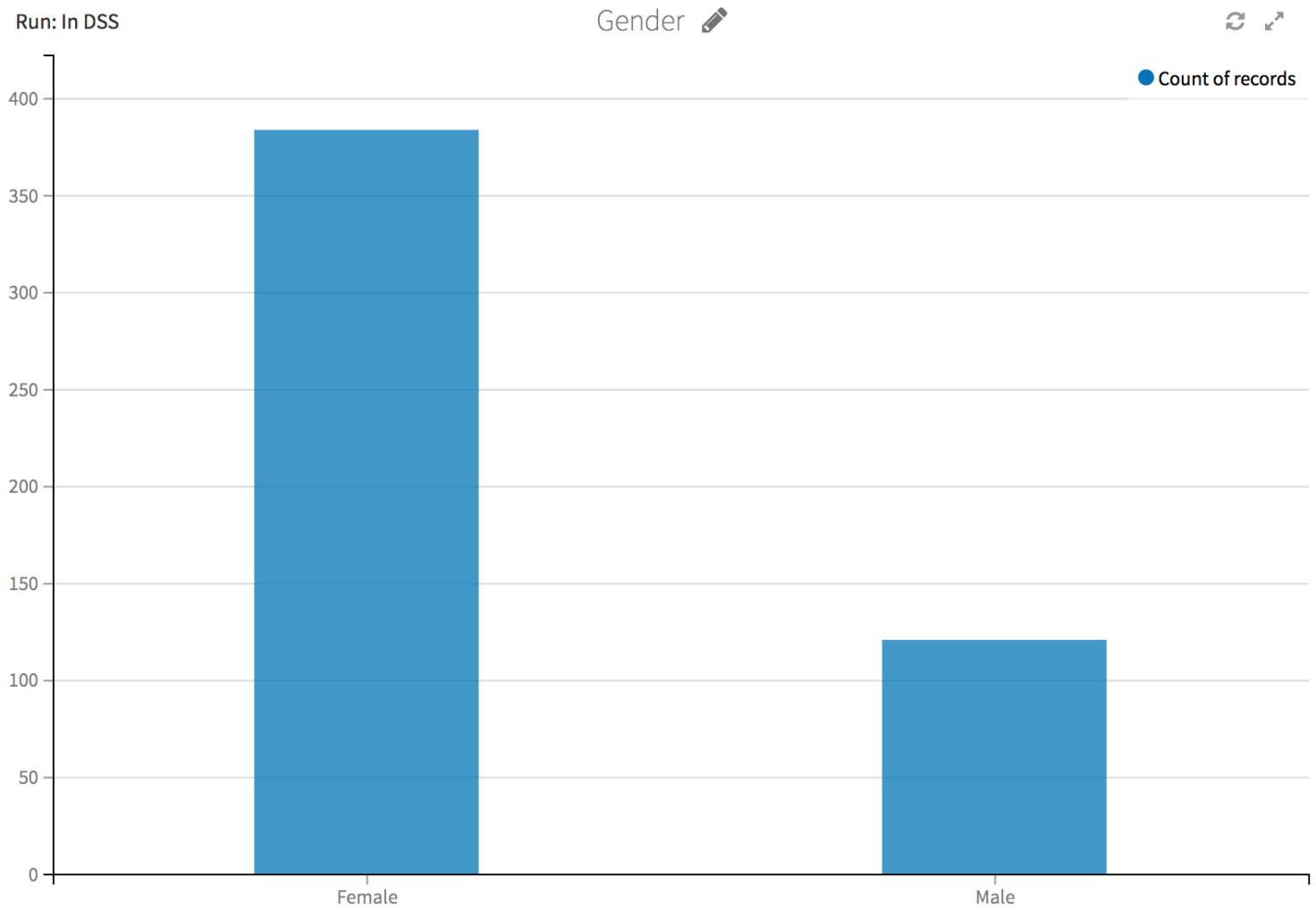
CATSKILL provisions

GENDER: MODEL OUTPUT

ACTIONS	Name	Trained	Accuracy	Precision	Recall	F1 Score	Cost Matrix Gain	Log Loss	ROC AUC	Lift
	SVM	2017-03-27 17:07:58	0.84	0.64	0.61	0.62	0.11	0.66	0.87	2.10
	Logistic Regression	2017-03-27 17:07:57	0.79	0.50	0.61	0.55	0.09	0.44	0.81	1.88
	K Nearest Neighbors (k=5)	2017-03-27 17:08:13	0.71	0.40	0.74	0.52	0.09	0.73	0.79	1.88
	XGBoost	2017-03-27 17:08:13	0.69	0.39	0.78	0.52	0.09	0.44	0.78	1.88
	Gradient Tree boosting	2017-03-27 17:07:58	0.80	0.55	0.48	0.51	0.08	0.48	0.73	1.77
	Random forest	2017-03-27 17:07:57	0.78	0.47	0.35	0.40	0.05	0.66	0.62	1.44
	Decision Tree	2017-03-27 17:08:12	0.21	0.21	1.00	0.35	-0.02	0.51	0.52	1.00
	Random forest	2017-03-27 17:05:42	0.48	0.25	0.70	0.36	0.01	0.66	0.60	1.33
	Decision Tree	2017-03-27 17:05:59	0.21	0.21	1.00	0.35	-0.02	0.51	0.52	1.00

CATSKILL provisions

GENDER

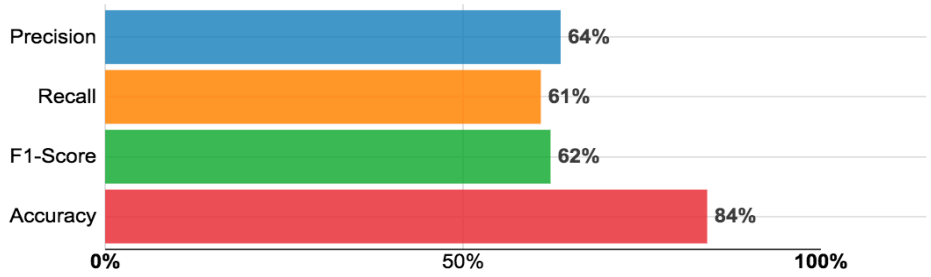


CATSKILL provisions

GENDER:
SVM CONFUSION MATRIX & ACCURACY

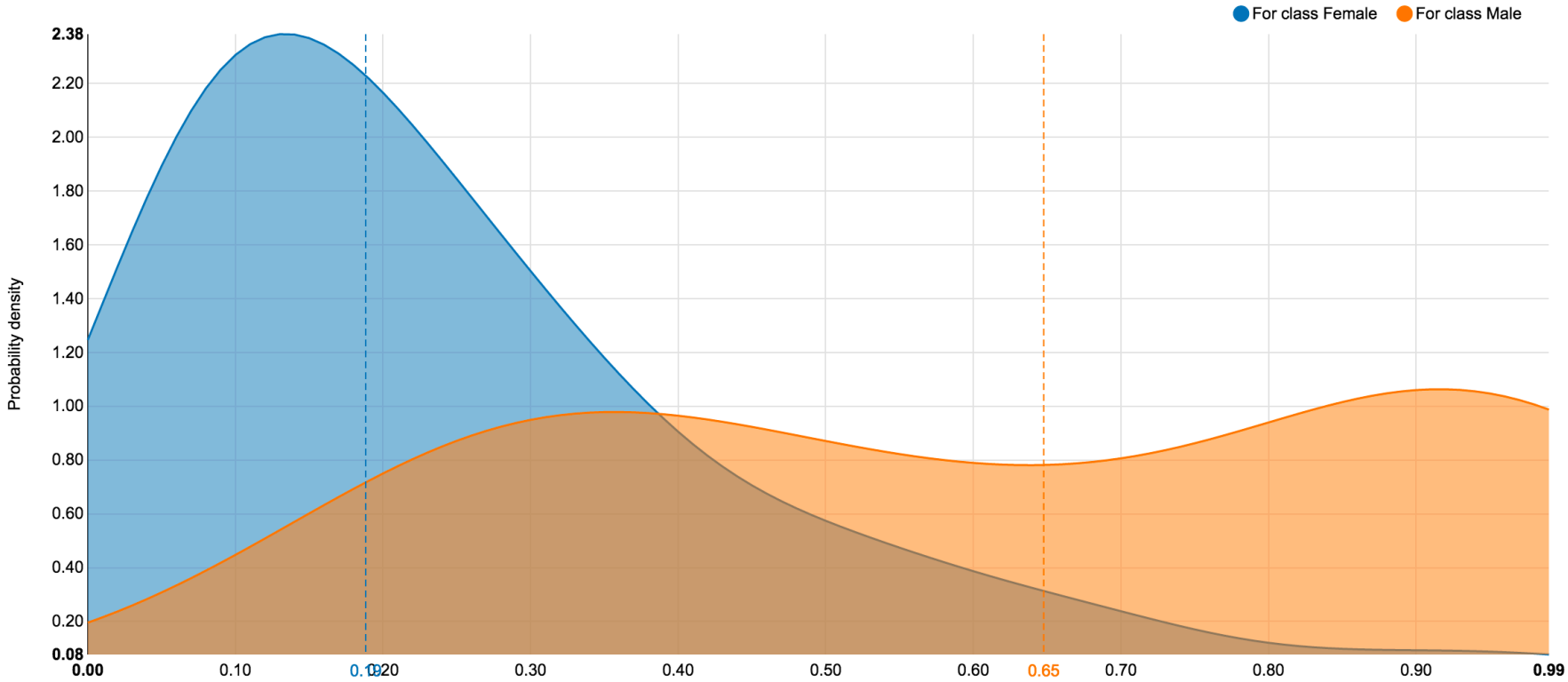
Threshold (cut-off) 0 1 0.525

	Predicted Male	Predicted Female
Total	22	85
Actually Male	14	9
Actually Female	8	76



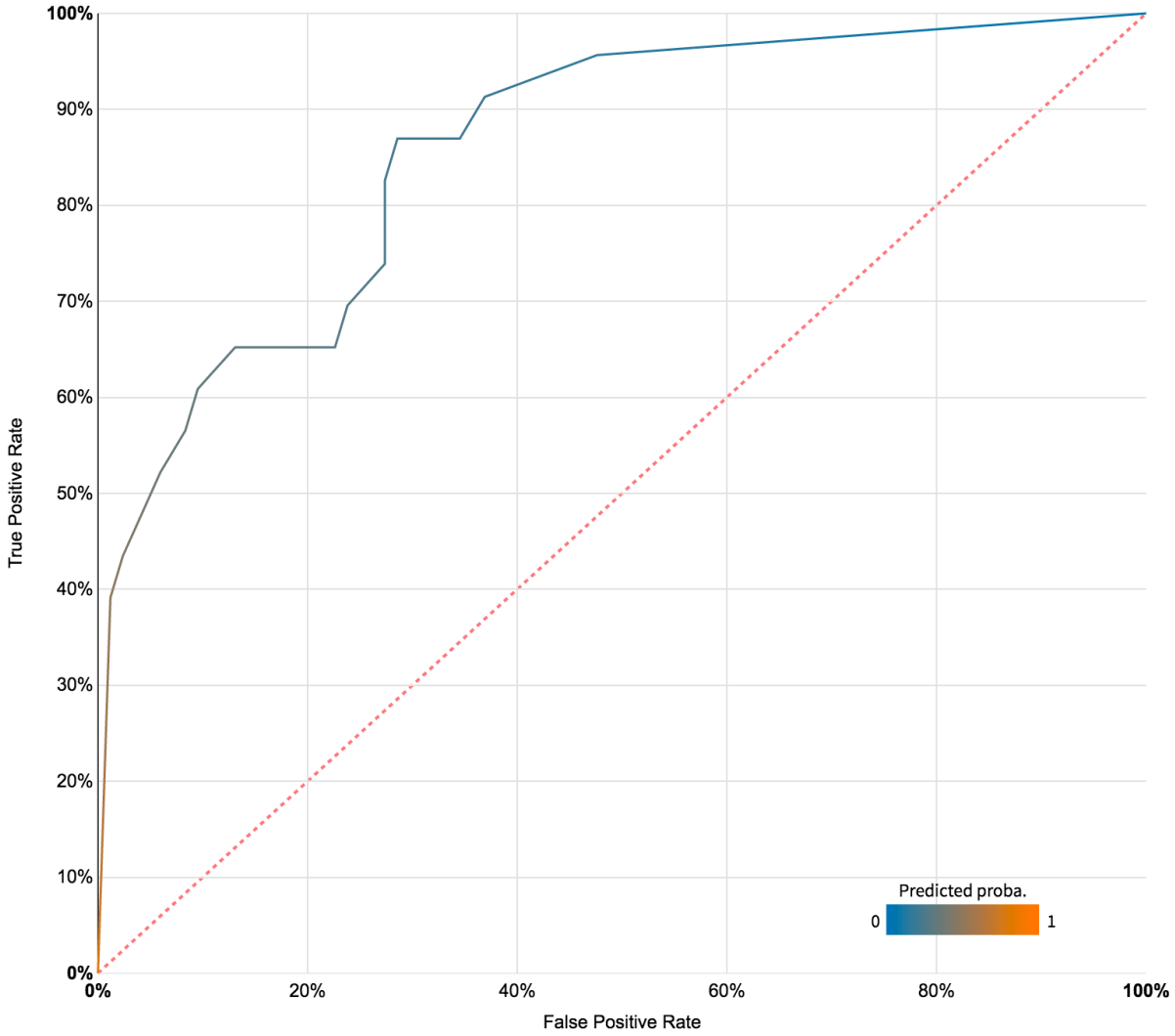
CATSKILL provisions

GENDER: SVM DENSITY CURVE



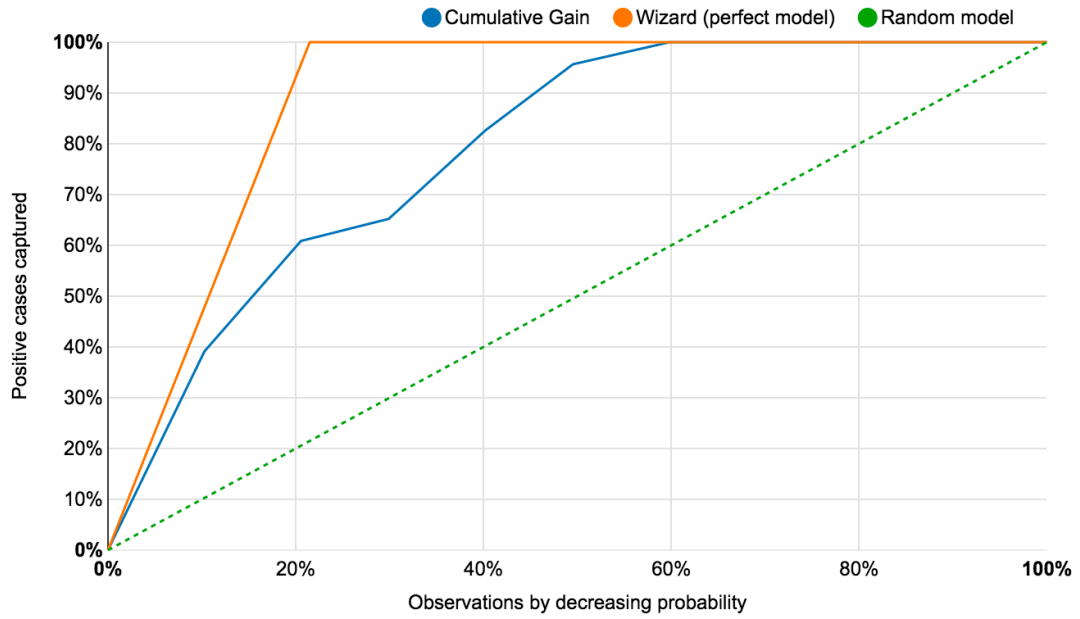
CATSKILL provisions

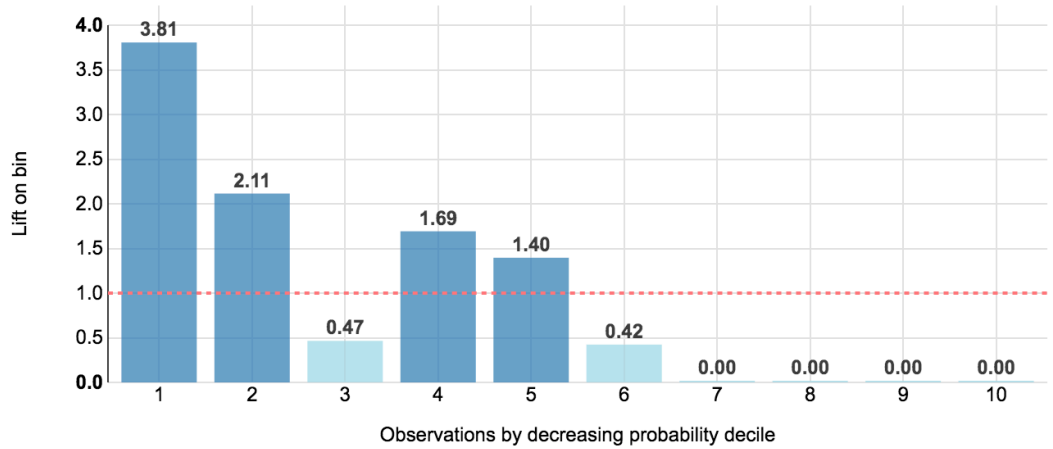
GENDER: SVM ROC AUC CURVE



CATSKILL provisions

GENDER: SVM LIFT CHARTS





CATSKILL provisions

BUSINESS RECOMENDATIONS

- Web traffic and sales conversions coming from New England and Eastern US
  - Focus traffic driving tactics to grow 'sweet spot'
  - Target marketing for gender
  - Grow other 'bee' regions
- Make 1x purchasers repeat-purchasers
  - Constant contact CRM solution using email promotion
  - Investment in search engine marketing

CATSKILL provisions