



Predicting Customer Purchasing Behavior as a Function of Sport Team Performance

Howard H. Hamilton, Ph.D.

Founder and CEO

Soccermetrics Research & Consulting, LLC

Research Objectives

- How **elastic** is customer purchasing behavior?
- Can sports team performance **predict** likelihood of a customer purchase?



Presentation Outline

- Motivation
- Data
- Modeling
- Results
 - Sports Leagues
 - Teams
- Summary & Conclusions



Motivation

- Noll (1974), Hanson & Gauthier (1989)
What influences sporting event attendance?
 - Ticket price
 - Amenities
- C.-Y. Chen (2005)
Predictions of ticket purchasing behavior for a professional sports team

What are customers' purchasing behaviors in the secondary ticket market?
Are they sensitive to performance of their preferred sports team?

Customer Relationship Data

- All customer transactions for regular season events related to “Big 4” professional sports leagues
 - Event date, Purchase date/“Near-buy”
 - Number of Tickets, Total Cost
- Customer-specific information
 - “Opt-In” and “Seller” flags

League	Transactions	Genres
NFL	31,548	32
NBA	27,396	31
NHL	20,117	30
MLB	123,631	30

users · listeners · fans · donors · viewers · households · shoppers · sellers · readers · browsers · friends · followers
 travelers · patients · contributors · attendees · readers · subscribers · buyers · clients · visitors · guests · customers

Match Result Data

- Regular season match results of Big 4 sports leagues (Teams, conference/division information)
 - NFL/MLB: 2007-2010 seasons
 - NHL/NBA: 2006-07 to 2009-10 seasons
 - Sources:
 - Retrosheet.org (MLB)
 - Sports-Reference.com (NFL, NBA, NHL)

Data Pre-Processing

- Customer Relationship Data
 - Record final purchasing decision of customer (Remove duplicate records)
 - Aggregate data for each league team over all seasons in study
- Match Result Data
 - Generate updated league table as of event date
 - Calculate conference rank of customer's preferred team as of event date
 - Incorporate league tiebreakers

League Ranking & Tiebreakers

	MLB	NFL	NBA	NHL
Criteria	Win Pct	Win Pct	Win Pct	Points
Rank	Division	Conference	Conference	Conference
Tiebreakers	None	Head-to-Head Division Common Opp Conference	Head-to-Head Division Conference	Points/Game Wins

Customer Purchase Model

- Classification model

$$Buy = a_0 + a_1 Rank + a_2 Rank^2 + a_3 F_{Opt} + a_4 F_{Sell} + a_5 N_{Tix} + a_6 C_{Tix}$$

- Algorithms

- Multinomial probit
- L_2 penalized logistic regression

- Model Training

- Training set: similar distribution of team rank as overall dataset
- 10-fold cross validation (no repeats)
- Model selection criteria: Accuracy, Cohen's Kappa

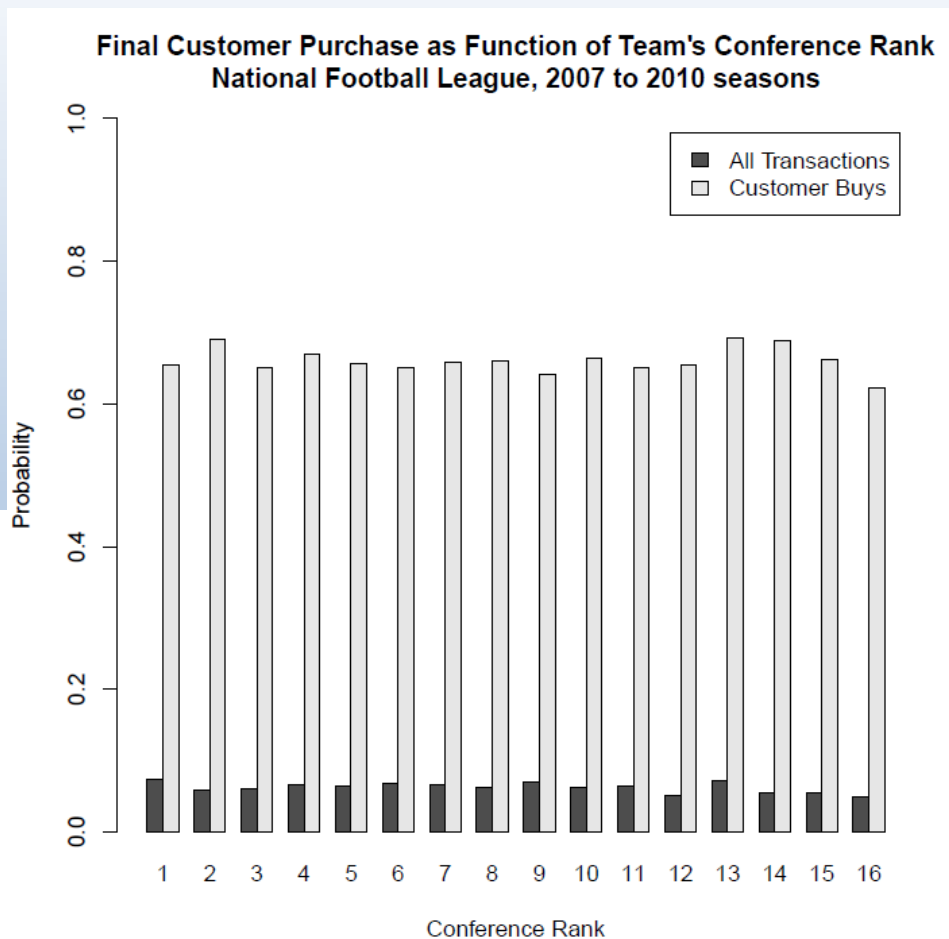
- Model Validation

- 20% of dataset reserved as validation set

Modeling Results National Football League



Customer Buys: All NFL Teams



Analysis of Deviance in NFL Model: All Customer Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			31548	40409	
Rank	1	0.04	31547	40409	0.850
Rank²	1	0.07	31546	40408	0.787
Opt	1	22.16	31545	40386	2.5×10^{-6}
Sell	1	714.22	31544	39672	~ 0
Quant	1	11.39	31543	39661	7×10^{-4}
Rank*Opt	1	0.89	31542	39660	0.345
Rank*Sell	1	0.52	31541	39659	0.470
Rank*Quant	1	0.17	31540	39658	0.280

Analysis of Deviance in NFL Model: Pittsburgh Steelers Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			802	967.29	
Rank	1	2.1896	801	965.10	0.139
Rank²	1	5.1241	800	959.67	0.020
Opt	1	1.0008	799	958.67	0.317
Sell	1	2.4977	798	956.18	0.114
Quant	1	2.6733	797	953.50	0.102
Rank*Opt	1	1.6161	796	951.89	0.204
Rank*Sell	1	0.0033	795	951.88	0.954
Rank*Quant	1	0.6302	794	951.25	0.427

Analysis of Deviance in NFL Model: Seattle Seahawks Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			714	877.73	
Rank	1	2.050	713	875.68	0.152
Rank²	1	0.343	712	875.34	0.558
Opt	1	5.599	711	869.74	0.018
Sell	1	60.221	710	809.52	~0
Quant	1	2.780	709	806.74	0.096
Rank*Opt	1	0.016	708	806.72	0.898
Rank*Sell	1	3.622	707	803.10	0.057
Rank*Quant	1	0.015	706	803.08	0.902

Predictive Quality of NFL Model: Pittsburgh Steelers Transactions

Intercept	Rank	Rank ²	Opt	Sell
2.072	-0.449	0.037	-0.198	-0.524

Performance on Test Dataset

		Reference	
		Yes	No
Prediction	Yes	0	0
	No	44	115

Accuracy	0.723
NIR	0.723
Sensitivity	0
Specificity	1.000
Detection Rate	0
Kappa	0

Predictive Quality of NFL Model: Seattle Seahawks Transactions

Intercept	Rank	Rank ²	Opt	Sell
1.027	0.147	-0.011	0.004	-1.745

Performance on Test Dataset

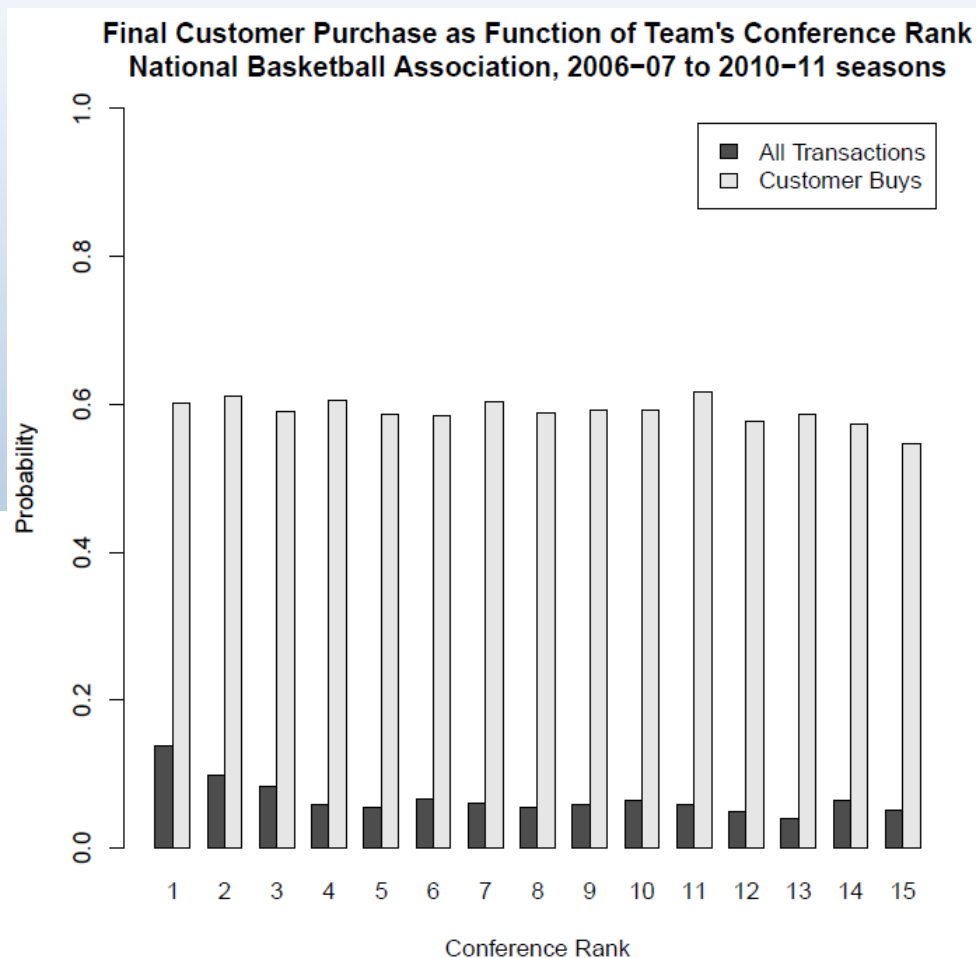
		Reference	
		Yes	No
Prediction	Yes	16	15
	No	31	80

Accuracy	0.676
NIR	0.669
Sensitivity	0.340
Specificity	0.842
Detection Rate	0.113
Kappa	0.200

Modeling Results National Basketball Association



Customer Buys: All NBA Teams



Analysis of Deviance in NBA Model: All Customer Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			27396	37034	
Rank	1	11.34	27395	37023	7×10^{-4}
Rank²	1	2.21	27394	37023	0.137
Opt	1	122.96	27393	36898	~ 0
Sell	1	552.21	27392	36345	~ 0
Quant	1	69.58	27391	36276	~ 0
Rank*Opt	1	2.20	27390	36274	0.138
Rank*Sell	1	12.46	27389	36261	4×10^{-4}
Rank*Quant	1	3.94	27388	36257	0.047

Analysis of Deviance in NBA Model: Boston Celtics Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			3085	4180.3	
Rank	1	6.309	3084	4174.0	0.012
Rank²	1	18.949	3083	4155.0	1.3×10^{-5}
Opt	1	22.358	3082	4132.7	2.3×10^{-5}
Sell	1	94.893	3081	4037.8	~ 0
Quant	1	6.608	3080	4031.2	0.010
Rank*Opt	1	0.405	3079	4030.8	0.524
Rank*Sell	1	9.066	3078	4021.7	0.002
Rank*Quant	1	2.926	3077	4018.8	0.087

Analysis of Deviance in NBA Model: Miami Heat Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			594	776.18	
Rank	1	2.226	593	773.95	0.136
Rank²	1	2.319	592	771.63	0.128
Opt	1	1.788	591	769.84	0.181
Sell	1	7.283	590	762.56	0.006
Quant	1	11.980	589	750.58	5.4×10^{-4}
Rank*Opt	1	1.089	588	749.49	0.296
Rank*Sell	1	0.410	587	749.08	0.521
Rank*Quant	1	0.255	586	748.83	0.613

Predictive Quality of NBA Model: Boston Celtics Transactions

Intercept	Rank	Rank ²	Opt	Sell
0.349	0.266	-0.020	-0.222	-0.810

Performance on Test Dataset

		Reference	
		Yes	No
Prediction	Yes	97	78
	No	143	299

Accuracy	0.642
NIR	0.611
Sensitivity	0.404
Specificity	0.793
Detection Rate	0.157
Kappa	0.208

Predictive Quality of NBA Model: Miami Heat Transactions

Intercept	Rank	Rank ²	Opt	Sell
-0.083	0.199	-0.009	-0.051	-0.965

Performance on Test Dataset

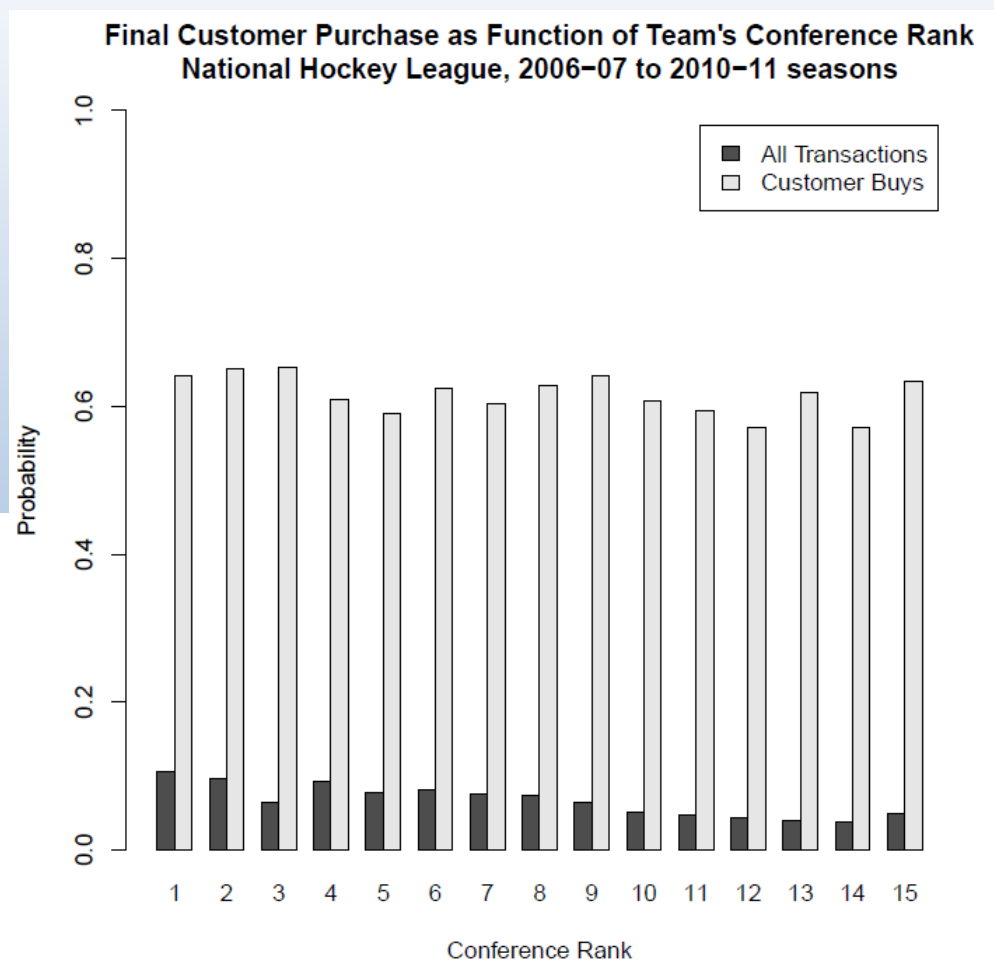
		Reference	
		Yes	No
Prediction	Yes	3	11
	No	40	64

Accuracy	0.568
NIR	0.636
Sensitivity	0.070
Specificity	0.853
Detection Rate	0.025
Kappa	-0.089

Modeling Results National Hockey League



Customer Buys: All NHL Teams



Analysis of Deviance in NHL Model: All Customer Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			20117	26716	
Rank	1	13.05	20116	26704	3×10^{-4}
Rank²	1	3.80	20115	26700	0.05
Opt	1	21.08	20114	26679	4×10^{-6}
Sell	1	432.54	20113	26246	~ 0
Quant	1	0.12	20112	26246	0.726
Rank*Opt	1	3.61	20111	26242	0.057
Rank*Sell	1	9.30	20110	26233	0.002
Rank*Quant	1	0.10	20109	26233	0.756

Analysis of Deviance in NHL Model: NY Rangers Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			2060	2676.9	
Rank	1	9.060	2059	2667.8	0.003
Rank²	1	14.727	2058	2653.1	1.2×10^{-4}
Opt	1	0.450	2057	2652.7	0.502
Sell	1	24.592	2056	2628.1	7.1×10^{-7}
Quant	1	0.861	2055	2627.2	0.354
Rank*Opt	1	2.314	2054	2624.9	0.128
Rank*Sell	1	0.148	2053	2624.8	0.701
Rank*Quant	1	1.615	2052	2623.2	0.204

Analysis of Deviance in NHL Model: Ottawa Senators Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			271	377.06	
Rank	1	1.797	270	375.26	0.180
Rank²	1	0.492	269	374.77	0.483
Opt	1	2.035	268	372.73	0.154
Sell	1	14.882	267	357.85	1.1×10^{-4}
Quant	1	0.022	266	357.83	0.883
Rank*Opt	1	3.919	265	353.91	0.048
Rank*Sell	1	0.569	264	353.34	0.451
Rank*Quant	1	2.306	263	351.04	0.128

Predictive Quality of NHL Model: NY Rangers Transactions

Intercept	Rank	Rank ²	Opt	Sell
-0.403	0.283	-0.016	0.095	-0.510

Performance on Test Dataset

		Reference	
		Yes	No
Prediction	Yes	12	7
	No	125	267

Accuracy	0.679
NIR	0.667
Sensitivity	0.088
Specificity	0.975
Detection Rate	0.029
Kappa	0.079

Predictive Quality of NHL Model: Ottawa Senators Transactions

Intercept	Rank	Rank ²	Opt	Sell
0.779	0.098	-0.011	0.063	-1.206

Performance on Test Dataset

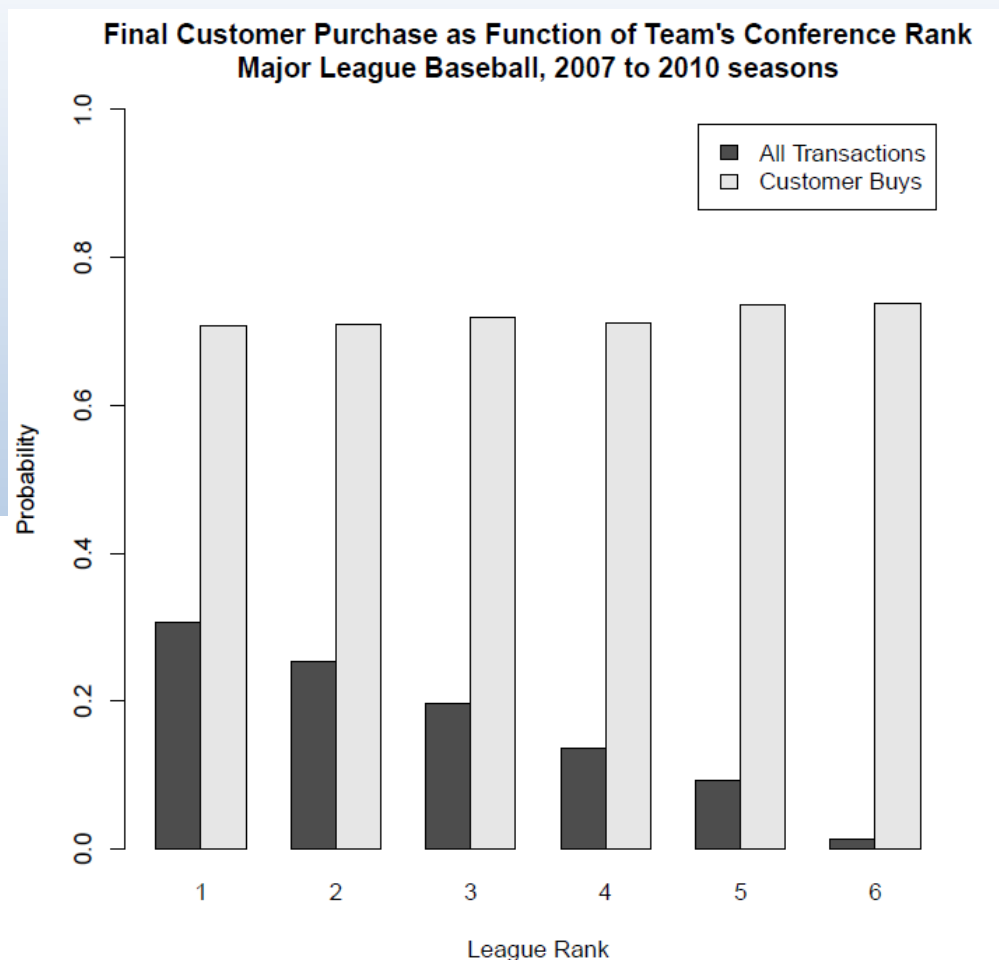
		Reference	
		Yes	No
Prediction	Yes	22	12
	No	8	11

Accuracy	0.623
NIR	0.566
Sensitivity	0.733
Specificity	0.478
Detection Rate	0.415
Kappa	0.216

Modeling Results Major League Baseball



Customer Buys: All MLB Teams



Analysis of Deviance in MLB Model: All Customer Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			123631	148074	
Rank	1	30.04	123630	148044	~0
Rank²	1	2.72	123629	148042	0.10
Opt	1	356.27	123628	147685	~0
Sell	1	2139.77	123627	145545	~0
Quant	1	185.96	123626	145360	~0
Rank*Opt	1	0.74	123625	145359	0.389
Rank*Sell	1	24.97	123624	145334	~0
Rank*Quant	1	0.06	123623	145334	0.807

Analysis of Deviance in MLB Model: Philadelphia Phillies Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			7176	8189.8	
Rank	1	19.146	7175	8170.7	1.2×10^{-5}
Rank²	1	15.073	7174	8155.6	1.0×10^{-4}
Opt	1	60.373	7173	8095.3	~ 0
Sell	1	114.502	7172	7980.8	~ 0
Quant	1	9.829	7171	7970.9	0.002
Rank*Opt	1	0.833	7170	7970.1	0.362
Rank*Sell	1	13.670	7169	7956.4	2.2×10^{-4}
Rank*Quant	1	1.063	7168	7955.4	0.303

Analysis of Deviance in MLB Model: Florida Marlins Transactions

	DF	Deviance	Residual DF	Residual Deviance	$P(\chi^2)$
NULL			684	943.43	
Rank	1	8.353	683	935.08	0.004
Rank²	1	0.948	682	934.13	0.330
Opt	1	31.225	681	902.91	2.3×10^{-8}
Sell	1	152.111	680	750.80	~ 0
Quant	1	3.127	679	747.67	0.077
Rank*Opt	1	0.295	678	747.38	0.587
Rank*Sell	1	0.273	677	747.10	0.601
Rank*Quant	1	4.467	676	742.64	0.035

Predictive Quality of MLB Model: Philadelphia Phillies Transactions

Intercept	Rank	Rank ²	Opt	Sell
0.886	0.705	-0.129	-0.467	-0.626

Performance on Test Dataset

		Reference	
		Yes	No
Prediction	Yes	0	0
	No	387	1048

Accuracy	0.731
NIR	0.731
Sensitivity	0
Specificity	1.000
Detection Rate	0
Kappa	0

Predictive Quality of MLB Model: Florida Marlins Transactions

Intercept	Rank	Rank ²	Opt	Sell
0.903	0.450	-0.031	-0.908	-2.11

Performance on Test Dataset

		Reference	
		Yes	No
Prediction	Yes	45	19
	No	16	56

Accuracy	0.743
NIR	0.552
Sensitivity	0.738
Specificity	0.747
Detection Rate	0.331
Kappa	0.482

Summary

- Developed classification model of final customer purchasing decisions for sporting events:
 - Relationship marketing transactions
 - Performance of preferred team as of event date
- Described factors that contributed significantly to data variance
- Presented model performance for customer purchases for North American Big 4 professional leagues

Conclusions

- Customer purchase behavior relatively inelastic to preferred team's league position
- Seller flag accounts for greatest variation in customer purchase decisions
- 10% of data variance captured by model significantly improves prediction of “Buy” behavior
 - Occurs most often for teams with small samples
- Interaction terms reduce false positives at cost of increased false negatives

Future Work

- Develop classification model that accounts for more of residual deviance
 - Offer type? Ticket price?
- Add factors to capture rivalry games and games with playoff implications
- Individualize prediction model (1:1 Marketing)
 - Account for previous customer purchase history
 - Default to group purchase history of customers with similar team preferences



Thank You!