Choosing Intro

School of Information Studies
Syracuse University

Patterns in Retail Behavior

- Obtain
- Scrub
- **Explore**
- **Model**
- **iNterpret**

School of Information Studies
Syracuse University

Our Challenge This Week?

School of Information Studies
Syracuse University

What Do We Recommend?

School of Information Studies
Syracuse University

How Do We Retain Customers?

School of Information Studies
Syracuse University

But How to Choose?

School of Information Studies
Syracuse University
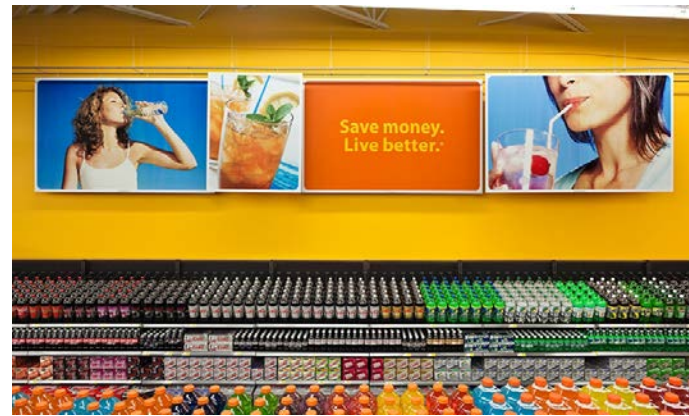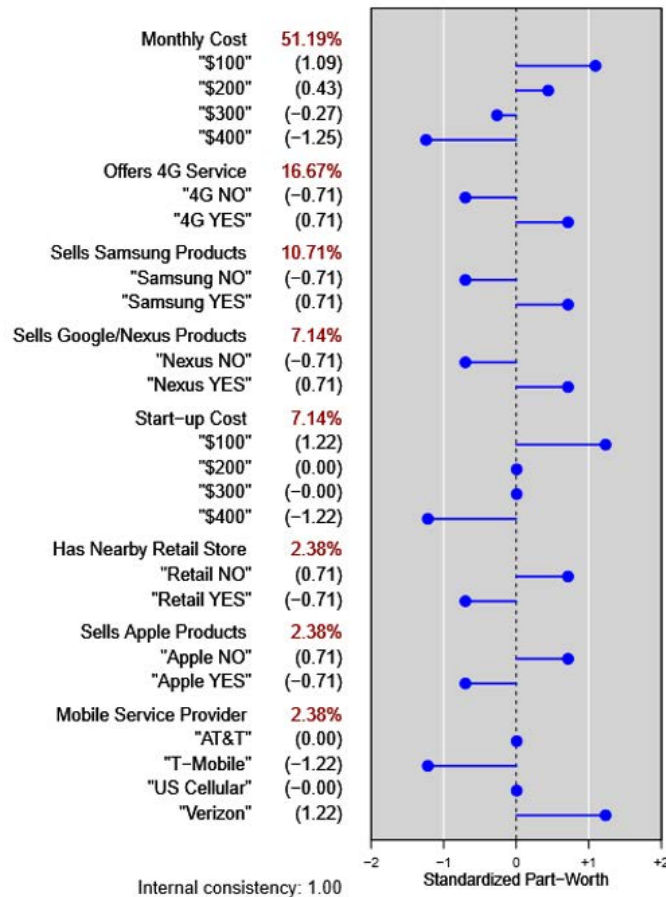
Data Review

School of Information Studies
Syracuse University

Choosing Data

citrus fruit	semi-finished bread	margarine	ready soups	
tropical fruit	yogurt	coffee		
whole milk				
pip fruit	yogurt	cream cheese	meat spreads	
other vegetables	whole milk	condensed milk	long-life bakery product	
whole milk	butter	yogurt	rice	abrasive cleaner
rolls/buns				
other vegetables	UHT-milk	rolls/buns	bottled beer	liquor (appetizer)
pot plants				
whole milk	cereals			
tropical fruit	other vegetables	white bread	bottled water	chocolate
citrus fruit	tropical fruit	whole milk	butter	curd
beef				
frankfurter	rolls/buns	soda		
chicken	tropical fruit			
butter	sugar	fruit/vegetable juice	newspapers	

age	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	Response
30	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
33	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
35	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
30	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
59	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
35	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
36	no	307	yes	no	cellular	14	may	341	1	330	2	other	no
39	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	no
41	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	no

School of Information Studies
Syracuse University

Preference Data

Consumer Description

School of Information Studies
Syracuse University
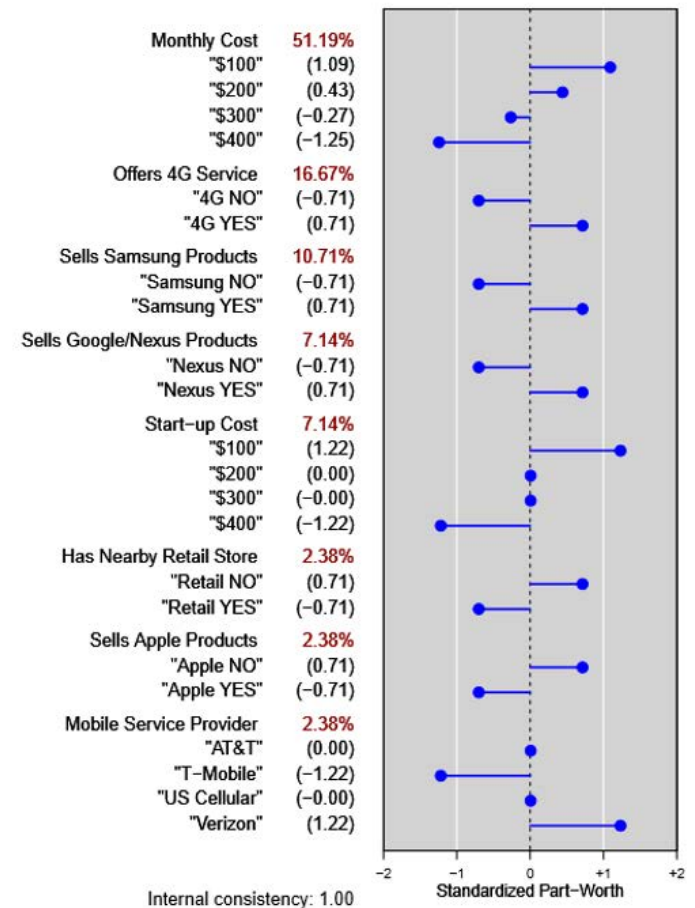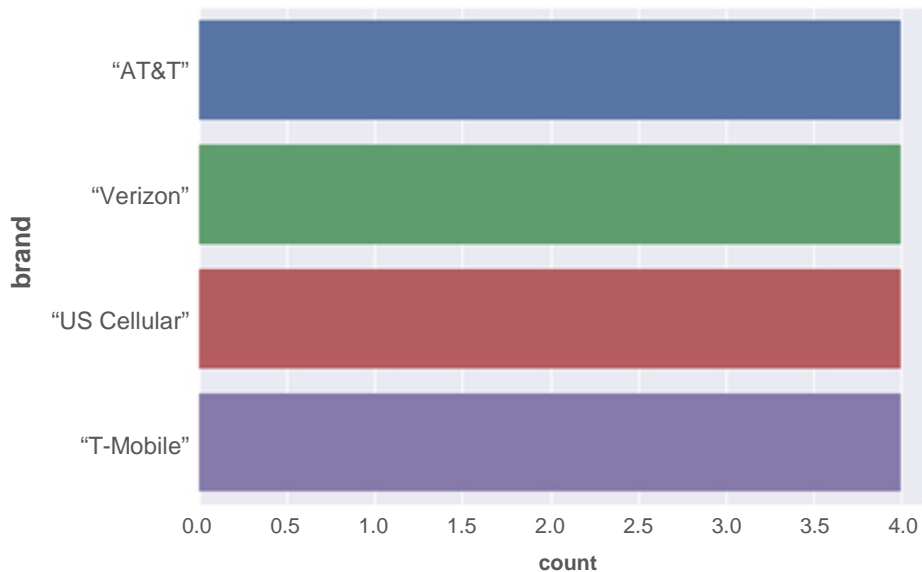
Consumer Description (cont.)

School of Information Studies
Syracuse University
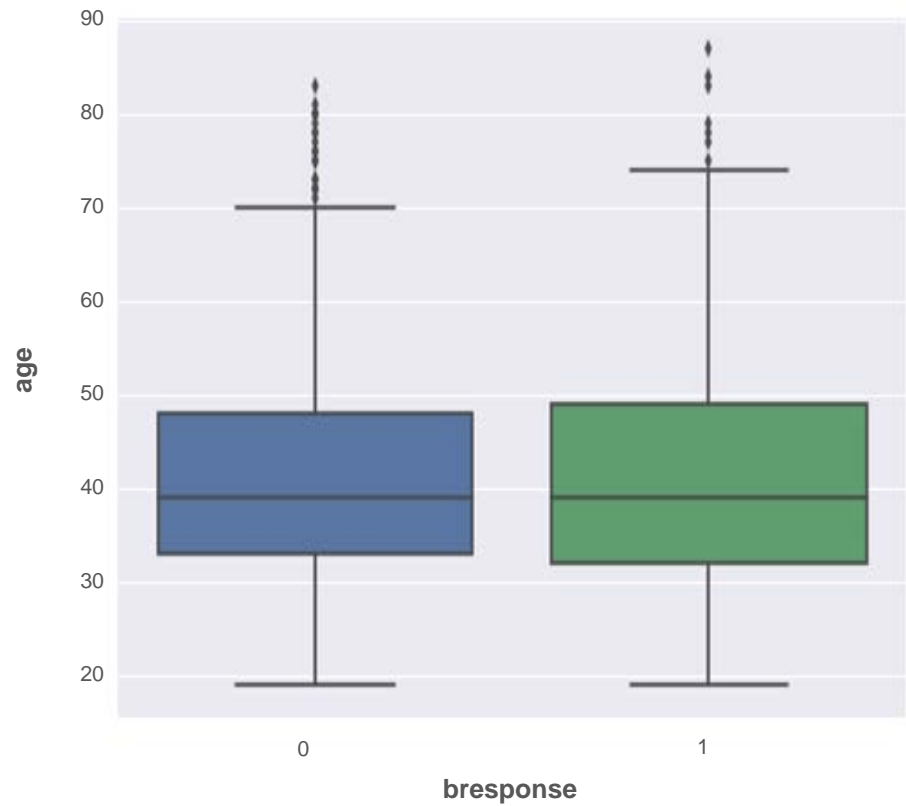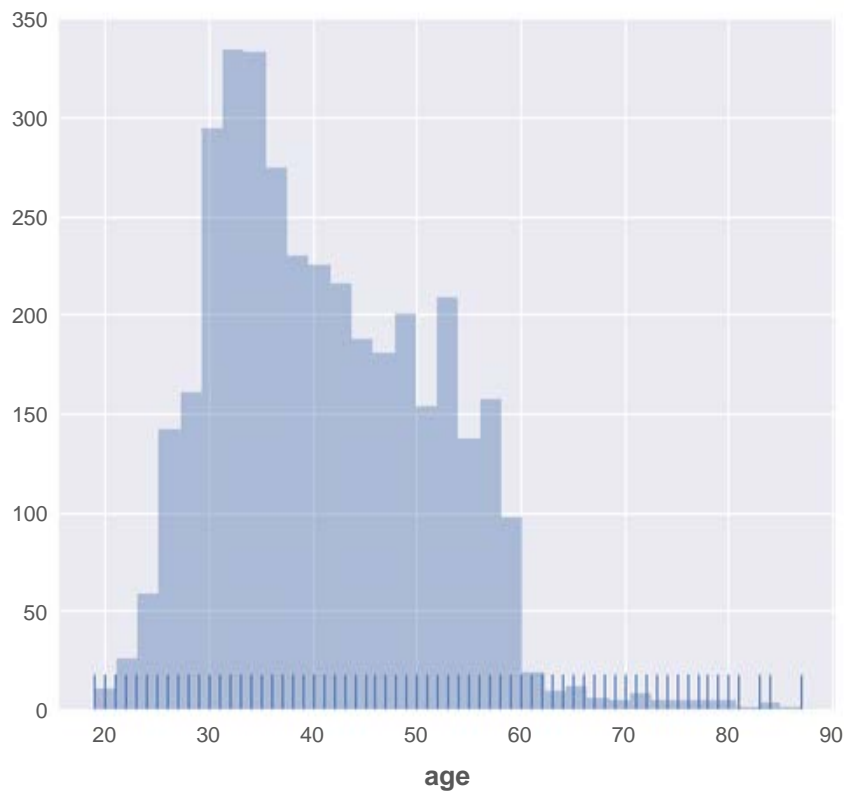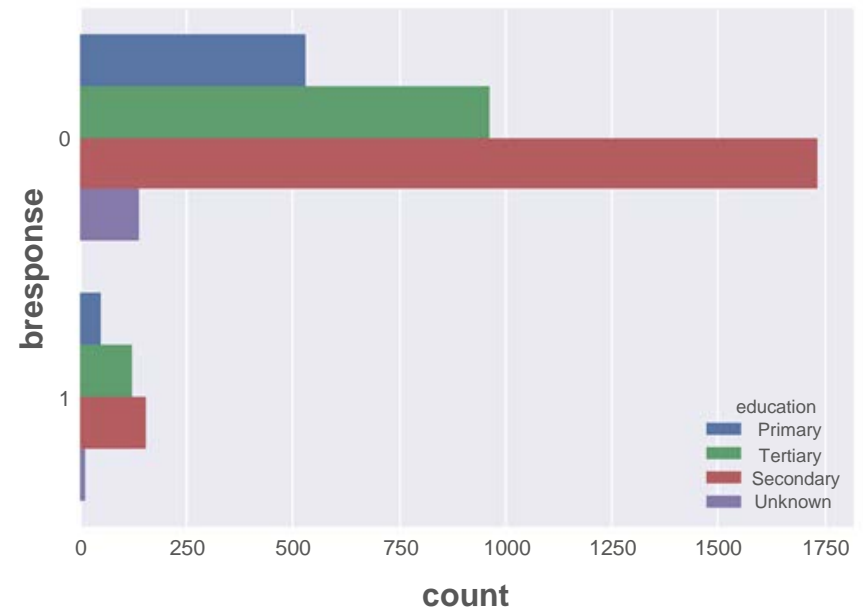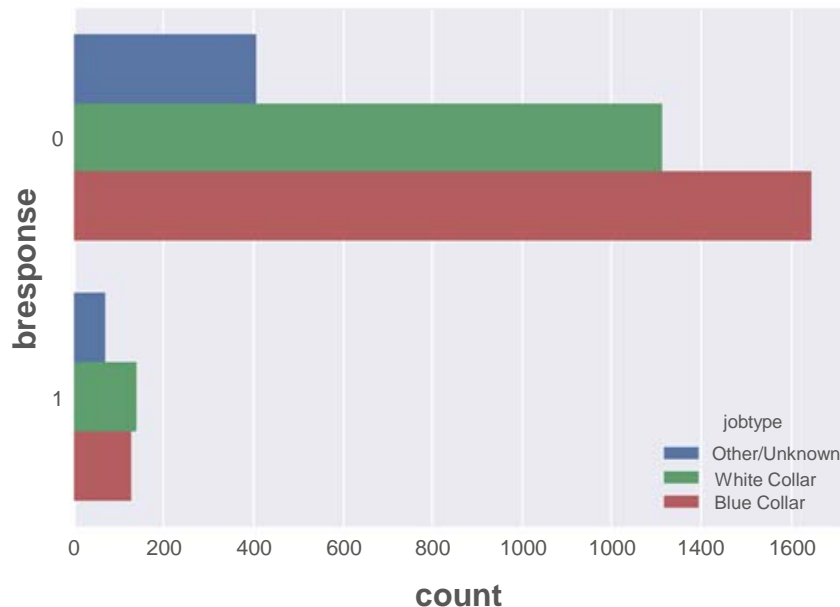
Product Description

citrus fruit	semi-finished bread	margarine	ready soups	
tropical fruit	yogurt	coffee		
whole milk				
pip fruit	yogurt	cream cheese	meat spreads	
other vegetables	whole milk	condensed milk	long life bakery product	
whole milk	butter	yogurt	rice	abrasive cleaner
rolls/buns				
other vegetables	UHT-milk	rolls/buns	bottled beer	liquor (appetizer)
pot plants				
whole milk	cereals			
tropical fruit	other vegetables	white bread	bottled water	chocolate
citrus fruit	tropical fruit	whole milk	butter	curd
beef				
frankfurter	rolls/buns	soda		
chicken	tropical fruit			
butter	sugar	fruit/vegetable juice	newspapers	

Proportion of Market Baskets Containing Item
(Item Relative Frequency or Support)

School of Information Studies
Syracuse University

Product Association

Grouped Matrix for 344 Rules
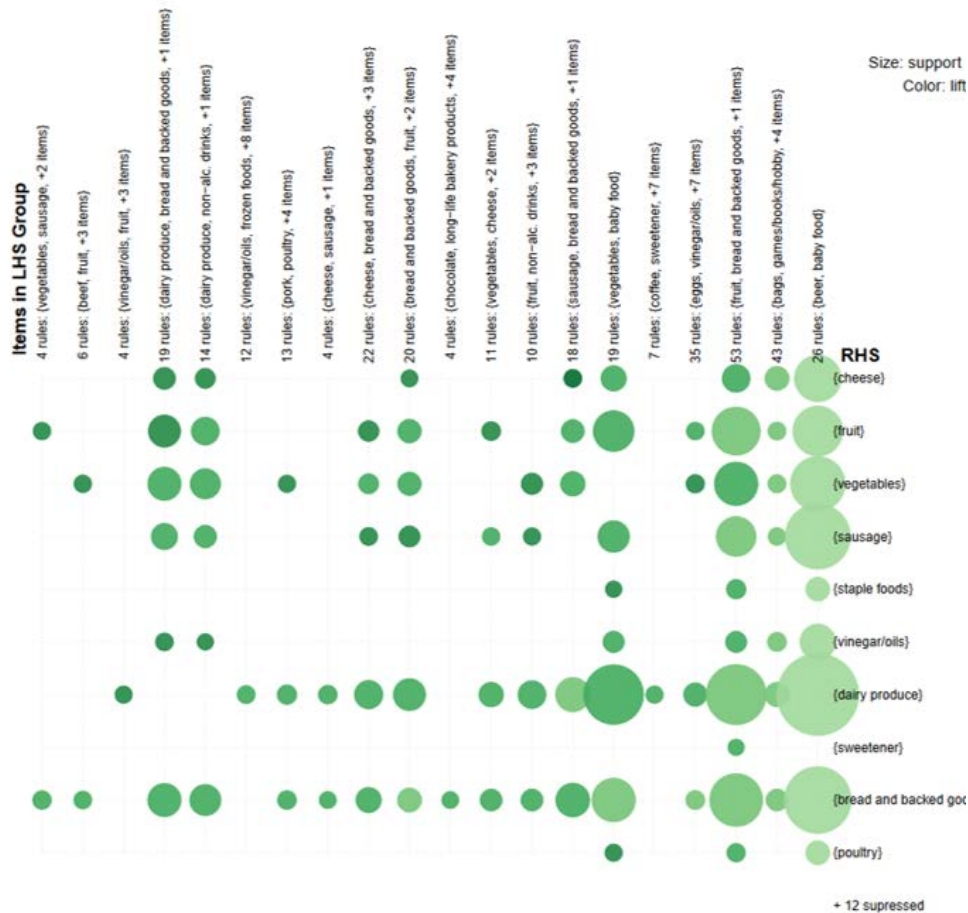
Graph for 10 rules
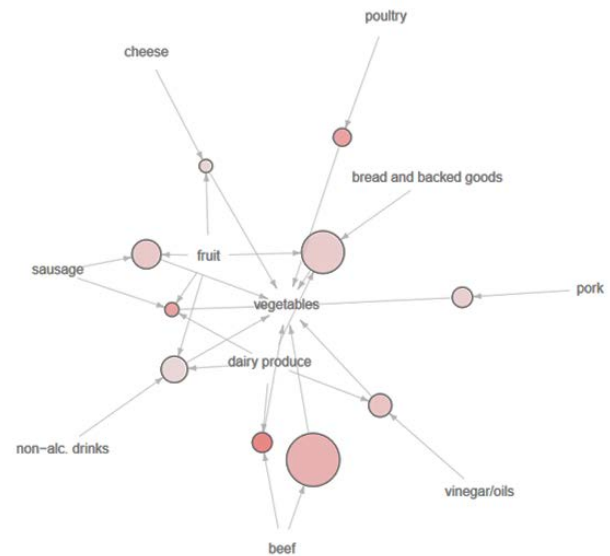
School of Information Studies
Syracuse University

Recommendation

School of Information Studies
Syracuse University

Our Challenge Was …

School of Information Studies
Syracuse University

How Do We Retain Customers?

School of Information Studies
Syracuse University

Recommendation

- Low cost is better

- Smaller start-up cost is better

- Regional carrier preference is strong

- Google / Samsung preferred

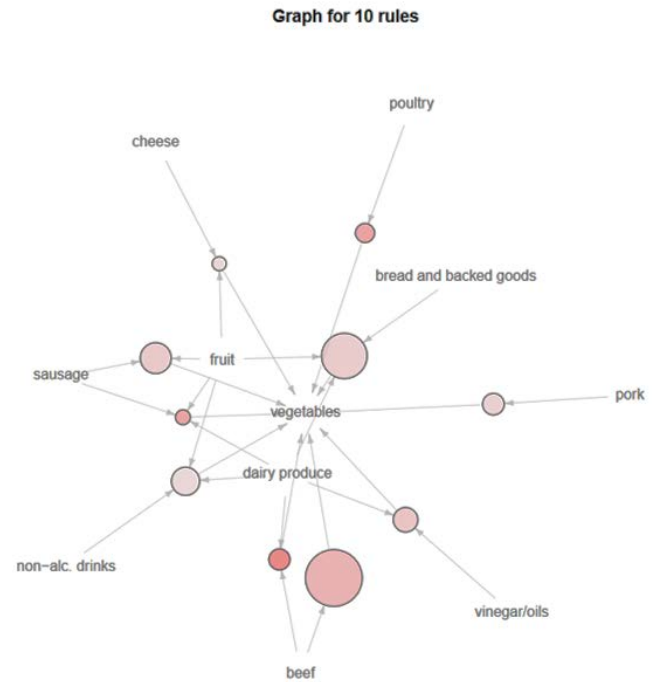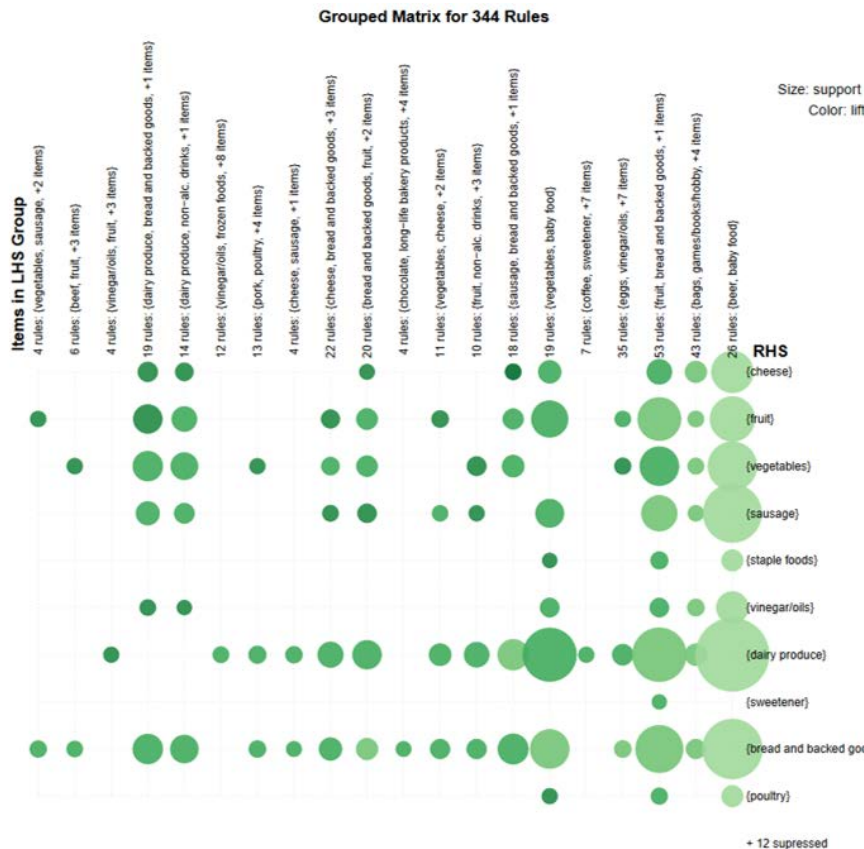School of Information Studies
Syracuse University

What Do We Recommend?

School of Information Studies
Syracuse University
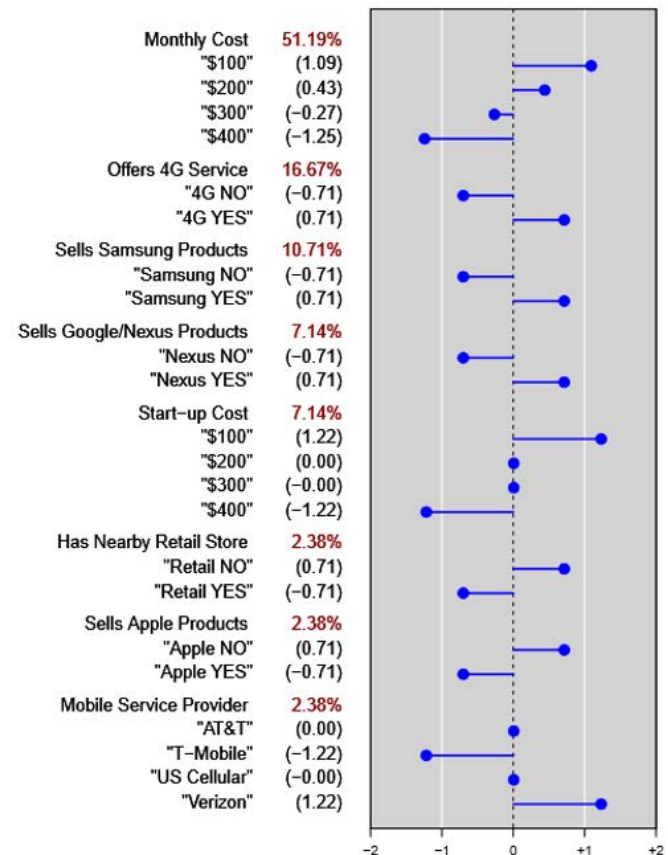
Product Recommendations

Grouped Matrix for 344 Rules

Graph for 10 rules

School of Information Studies
Syracuse University

Consumer Choice
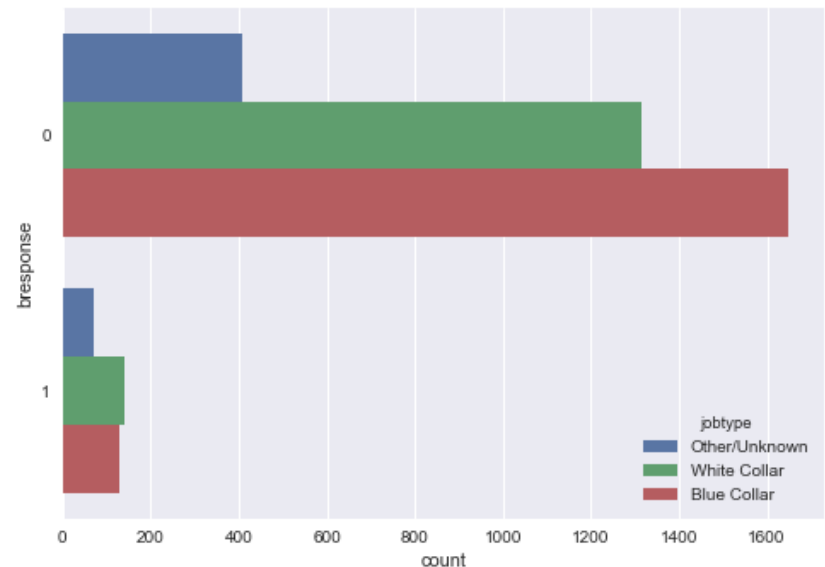
School of Information Studies
Syracuse University

Consumer Choice

Summary Statistics
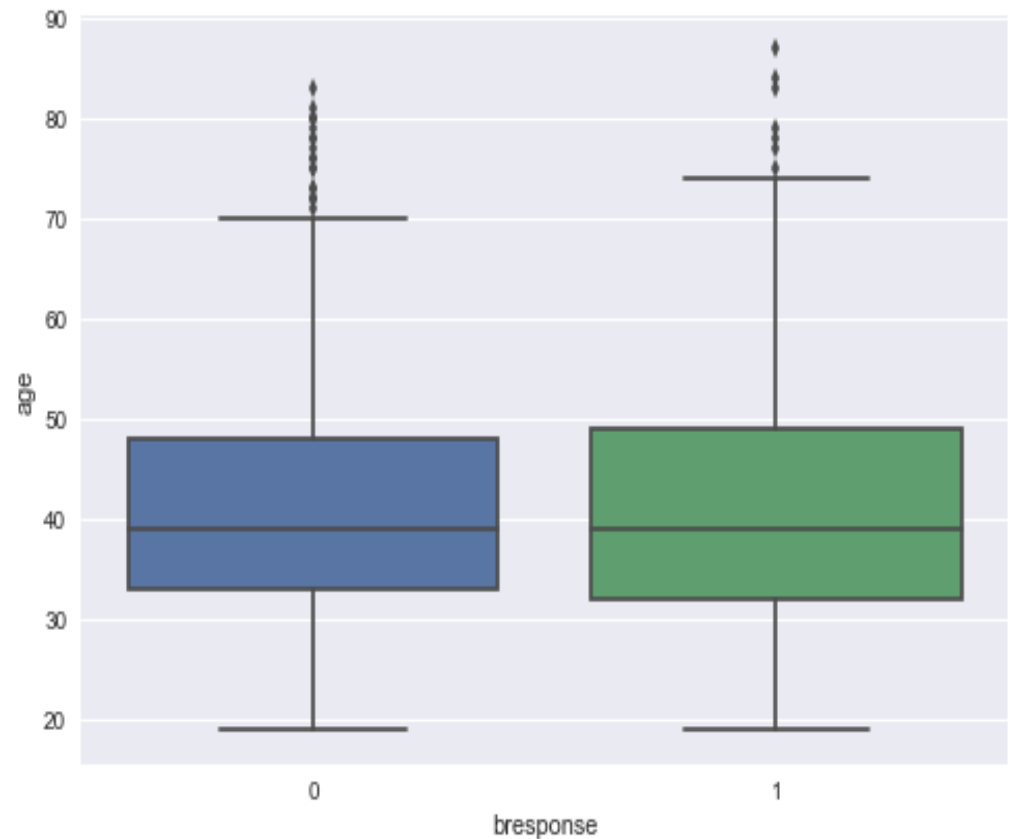
Account Balance

- Low – (-) 3,313

- High – 71,188

- Mean – 1,422

- Number of records: 4521

  - Blue-collar – 946

  - Management – 969

  - Technician – 768

  - Retired – 230

Summary Statistics

- Counts – per factor
- Range – per factor
- Deviation – by response
- Conjoint analysis

Choosing | School of Information Studies
Syracuse University

Consumer Knowledge

age	job	marital	education	default	balance	rentown	repairplan	contact	plannum	months	devices	bresponse
30	unemployed	married	primary	no	1787	rent	no	cellular	19	79	1	0
33	services	married	secondary	no	4789	own	yes	cellular	11	220	5	0
35	management	single	tertiary	no	1350	own	no	cellular	16	185	2	0
30	management	married	tertiary	no	1476	own	yes	unknown	3	199	1	0
59	blue-collar	married	secondary	no	0	own	no	unknown	5	226	1	0
35	management	single	tertiary	no	747	rent	no	cellular	23	141	4	0
36	self-employed	married	tertiary	no	307	own	no	cellular	14	341	3	0
39	technician	married	secondary	no	147	own	no	cellular	6	151	1	0
41	entrepreneur	married	tertiary	no	221	own	no	unknown	14	57	1	0
43	services	married	primary	no	-88	own	yes	cellular	17	313	3	0
39	services	married	secondary	no	9374	own	no	unknown	20	273	1	0
43	admin.	married	secondary	no	264	own	no	cellular	17	113	1	0
36	technician	married	tertiary	no	1109	rent	no	cellular	13	328	1	0
20	student	single	secondary	no	502	rent	no	cellular	30	261	1	1

School of Information Studies
Syracuse University

Consumer Choice

- Response variable
  - What do we do to predict?
  - Continuous
  - Categorical
- Explanatory variables
  - What do we know?
  - Continuous
  - Categorical
- Factors that prompt a choice
  - Geographic
  - Demographic
  - Behavioral
  - Psychographic

balance	rentown	repairplan
1787	rent	no
4789	own	yes
1350	own	no
1476	own	yes
0	own	no
747	rent	no
307	own	no
147	own	no
221	own	no
-88	own	yes
9374	own	no
264	own	no
1109	rent	no
502	rent	no

School of Information Studies
Syracuse University

Logistic Regression

- What if my response variable is not continuous?
  - Dead or alive
  - Approve or disapprove
  - Switch plans or stay
  - Renew or cancel
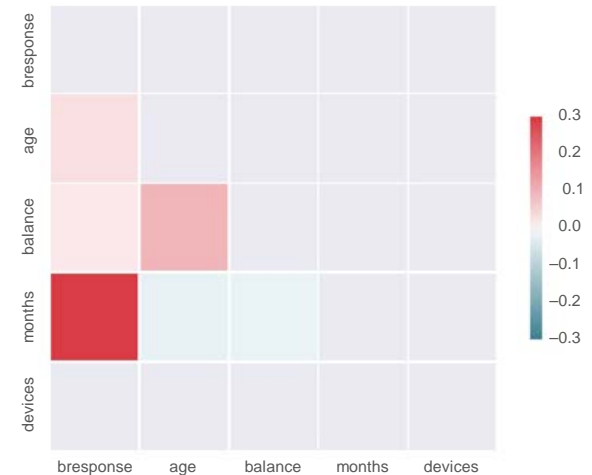- Assumption is Bernoulli distribution

$$P(y) = p^y(1-p)^{(1-y)}$$

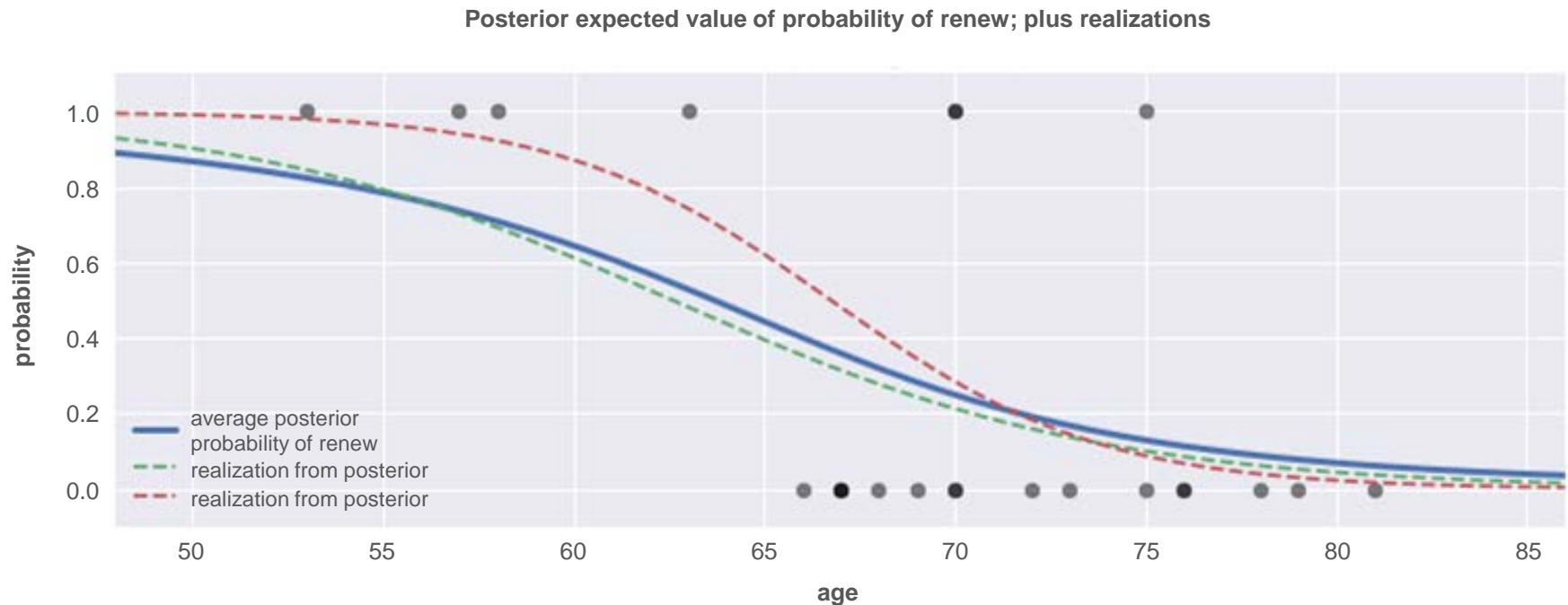School of Information Studies
Syracuse University

Logistic Regression (cont.)

- More classification than regression
- Link function
- Probability

Consumer Choice Splits

Posterior expected value of probability of renew; plus realizations

School of Information Studies
Syracuse University

Consumer Choice
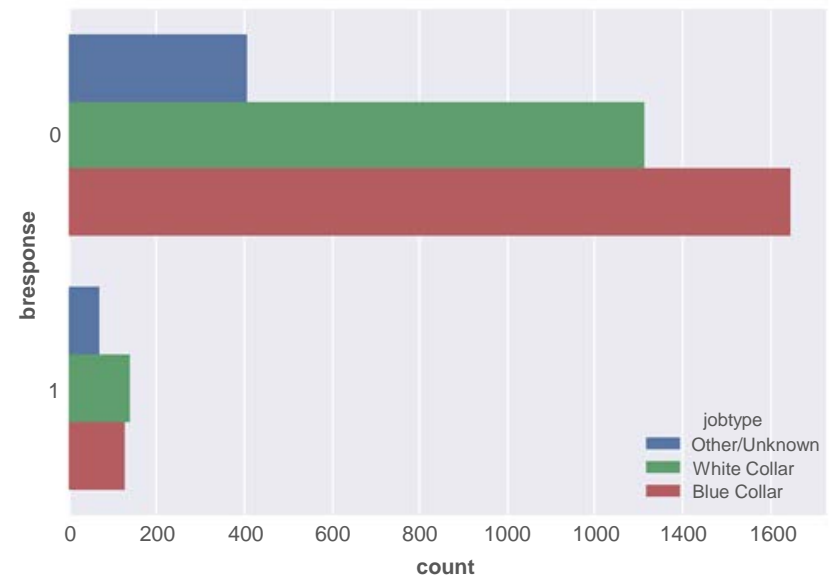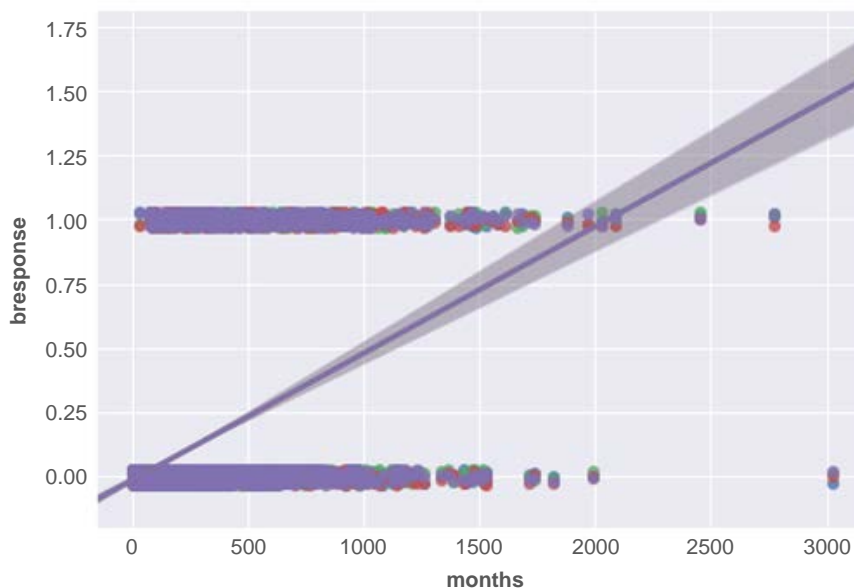
School of Information Studies
Syracuse University

Market Basket

School of Information Studies
Syracuse University

Market Basket


```
GV BRWN RICE  007874212222 F      0.70 N
GV LONG RICE  007874235205 F      1.28 N
GV 2   RF MLK 007874235187 F      3.08 O
RAMEN-BEEF12  004178900232 F      1.82 O
                    SUBTOTAL     10.30
                       TOTAL     10.30
           SHOPPING CARD TEND    10.00
                  CASH  TEND      0.50
                  CHANGE DUE      0.20

SHOP.CARD REDEMPTION            10.00
ACCOUNT 605214515600
APPR. CODE = 037453
REF #0571931
```

School of Information Studies
Syracuse University

Market Basket (cont.)

- What goes with what?
- What products are ordered together?
- What activities go together?
- What products should be featured?
- What products should be recommended?

School of Information Studies
Syracuse University
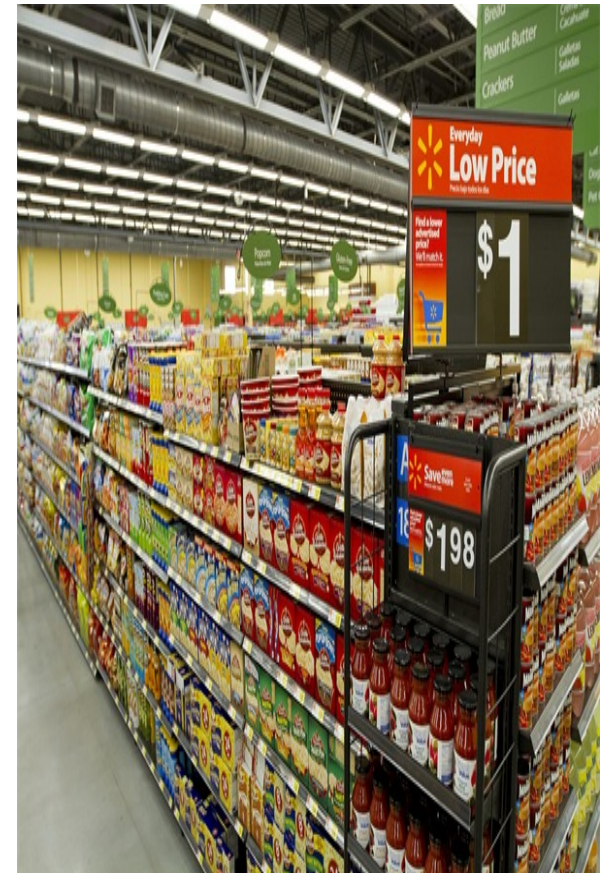
Data Conversion

School of Information Studies
Syracuse University

Terminology

- Item set
- Association rule
  - Antecedent
  - Consequent
- Selection criteria
  - Support
  - Confidence
  - Lift

Proportion of Market Baskets Containing Item
(Item Relative Frequency or Support)

School of Information Studies
Syracuse University
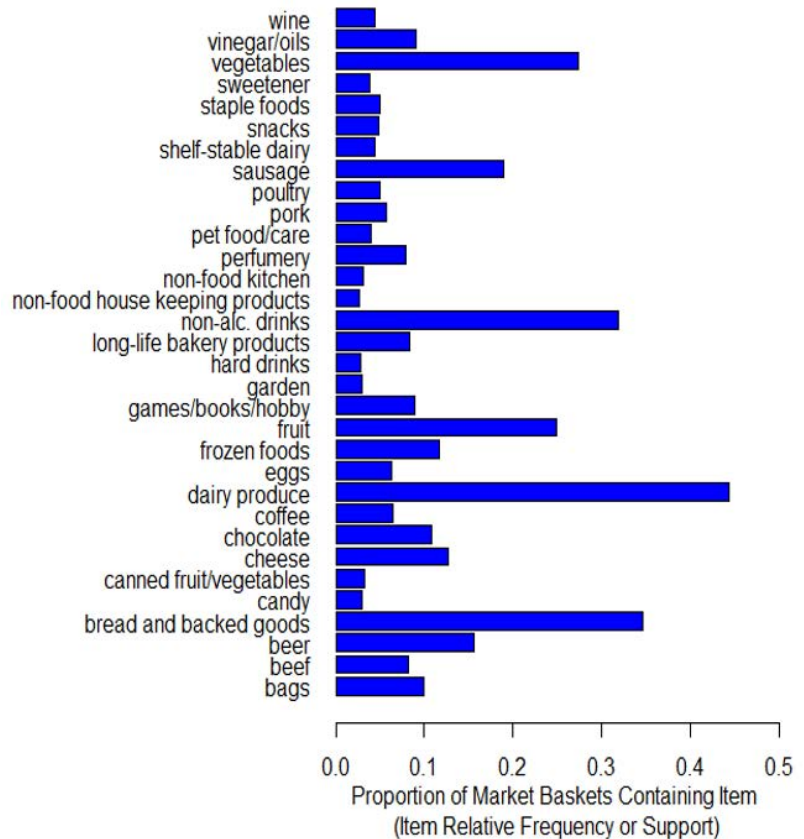
Use Cases

- Store layout
- Feature promotions
- Product bundling
- Recommended items

School of Information Studies
Syracuse University
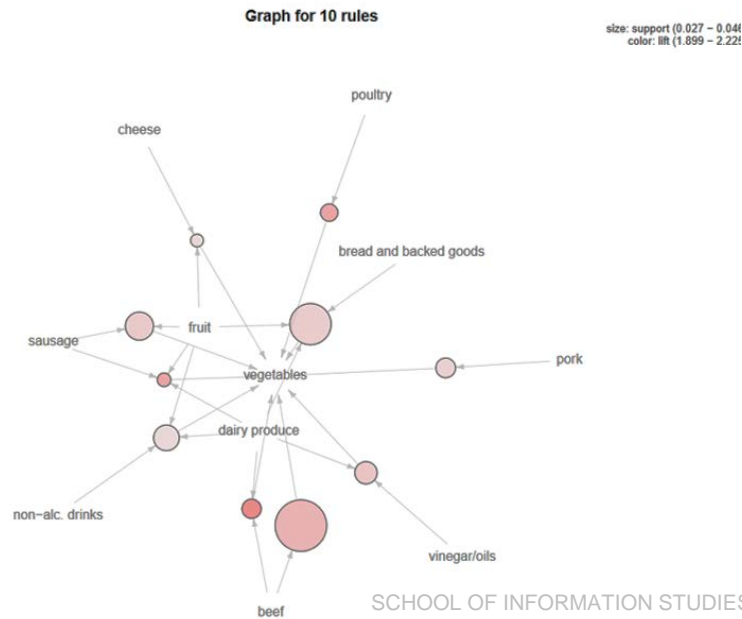
Evaluation Methods

School of Information Studies
Syracuse University

Evaluation Methods

Time

$$y = 20.047x + 199.65$$
$$R^2 = 0.4344$$

Number of Friends
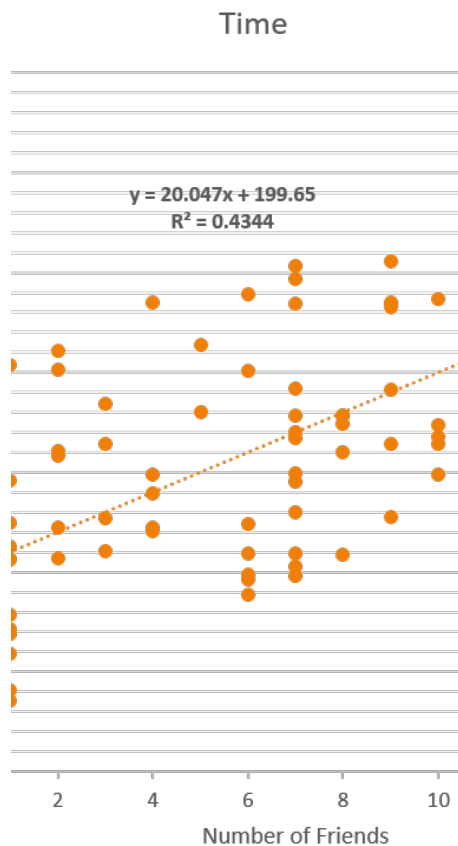
School of Information Studies
Syracuse University

Agenda

- Intro
- R-squared
- P-values
- Model validation

School of Information Studies
Syracuse University
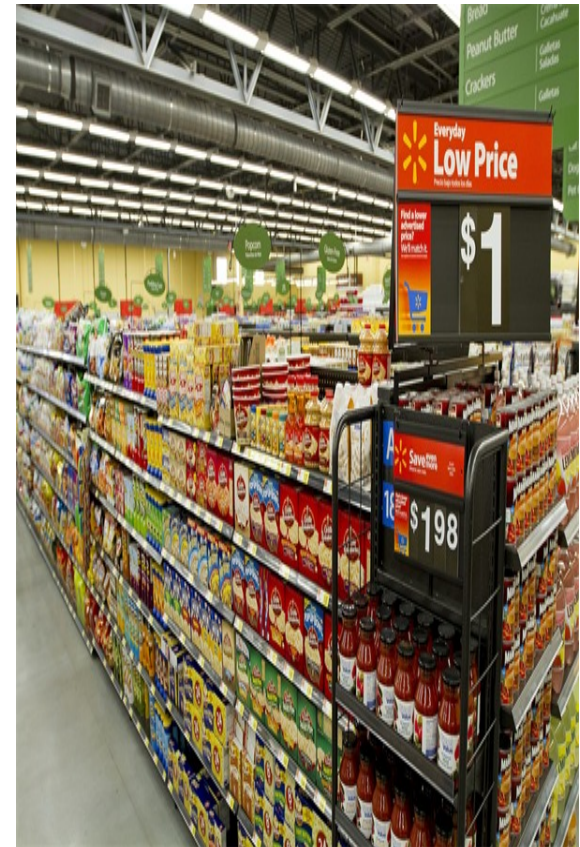
R-Squared

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y_i})^2}{\sum_i (y_i - \overline{y_i})^2}$$

lm(formula = hardness ~ dens, data = hardness)

Residuals:
   Min   1Q Median   3Q   Max
-338.40 -96.98 -15.71   92.71 625.06

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
(Intercept) -1160.500   108.580 -10.69 2.07e-12 ***
dens   57.507   2.279   25.24 < 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 183.1 on 34 degrees of freedom
Multiple R-squared: 0.9493,
Adjusted R-squared: 0.9478
F-statistic: 637 on 1 and 34 DF, p-value: < 2.2e-16

School of Information Studies
Syracuse University

R-Squared (cont.)

Revenue

y = 25x
R² = 1

Time

y = 20.047x + 199.65
R² = 0.4344

School of Information Studies
Syracuse University

P-Values

- Low p-value?
  - Highly unlikely to occur randomly, therefore significant

- High p-value?
  - Coefficient might actually be zero, therefore consider removing from model

lm(formula = hardness ~ dens, data = hardness)

Residuals:
    Min   1Q Median   3Q   Max
-338.40 -96.98 -15.71   92.71 625.06

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1160.500   108.580 -10.69 *2.07e-12* ***
dens   57.507   2.279   25.24 *< 2e-16* ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 183.1 on 34 degrees of freedom
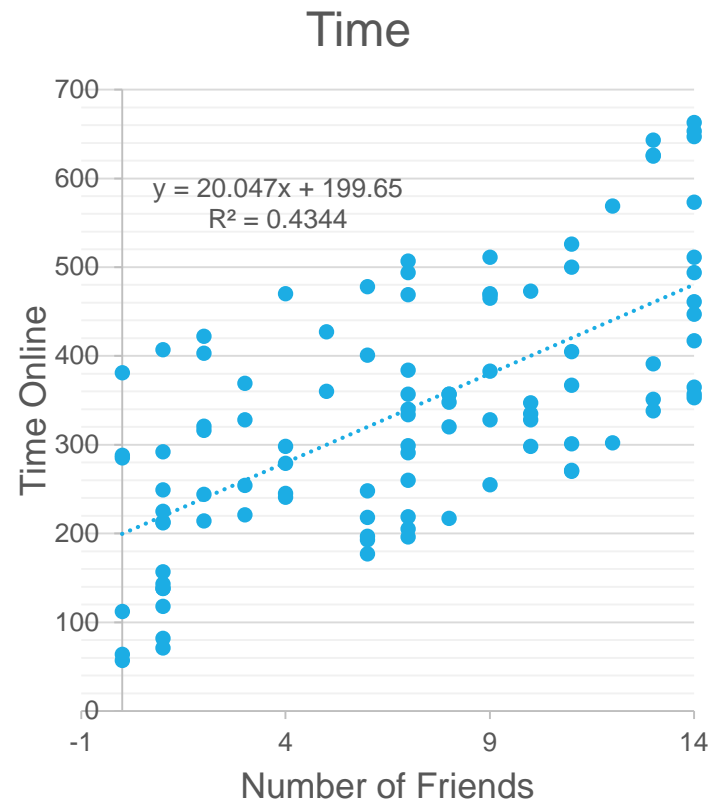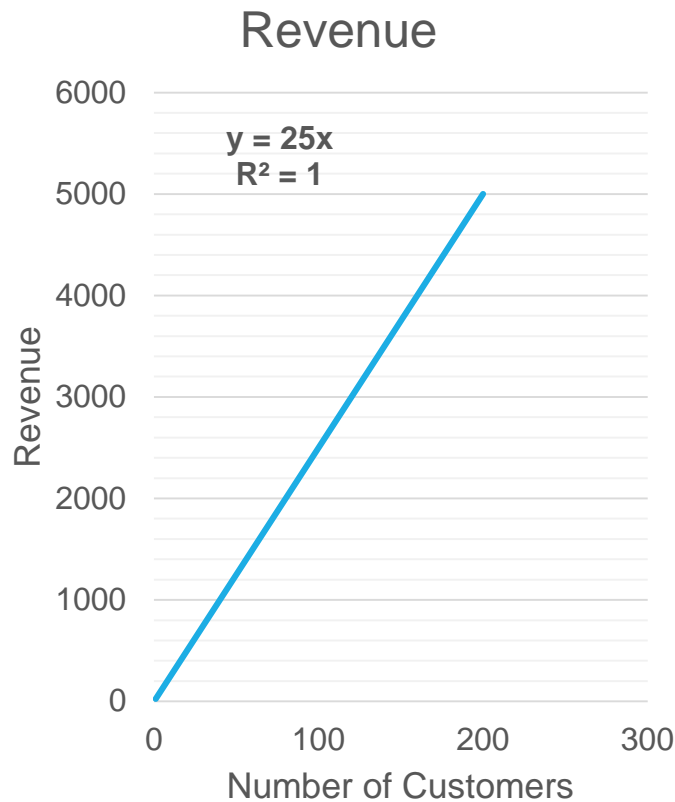Multiple R-squared: 0.9493,
Adjusted R-squared: 0.9478
F-statistic: 637 on 1 and 34 DF, p-value: < 2.2e-16

School of Information Studies
Syracuse University

Model Validation

- Collect new data

- Compare the results with:

  - Theoretical expectation (how much should a 0-bedroom house cost?)

  - Earlier empirical studies

  - Simulation (see GPA example from text)

- Split the original data with one portion for training and one for testing

School of Information Studies
Syracuse University

Confusion Matrix

- True positive

- False positive

- Accuracy

- Precision

- Specificity

School of Information Studies
Syracuse University