GPShopper® APPLIED REGRESSION

AUGUST 2013

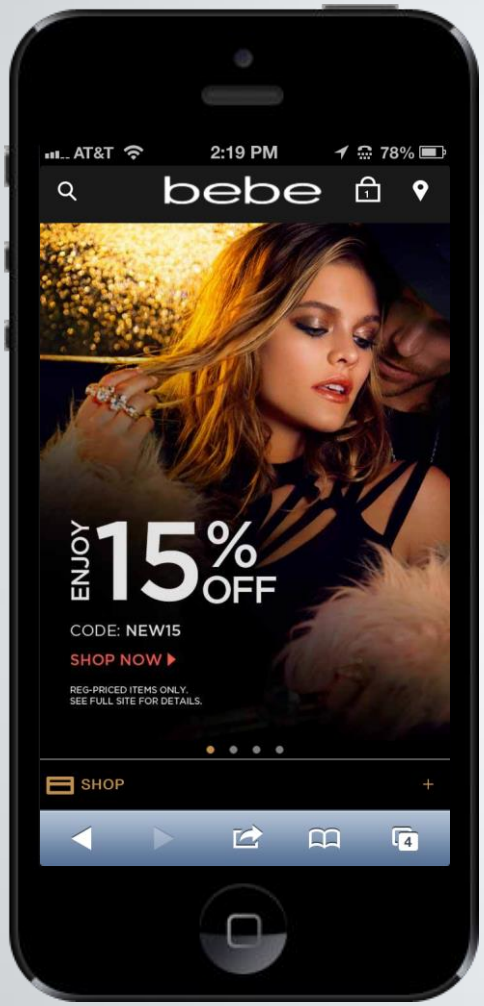
RYAN MOORE
KIM RUST
DEREK SZOPA

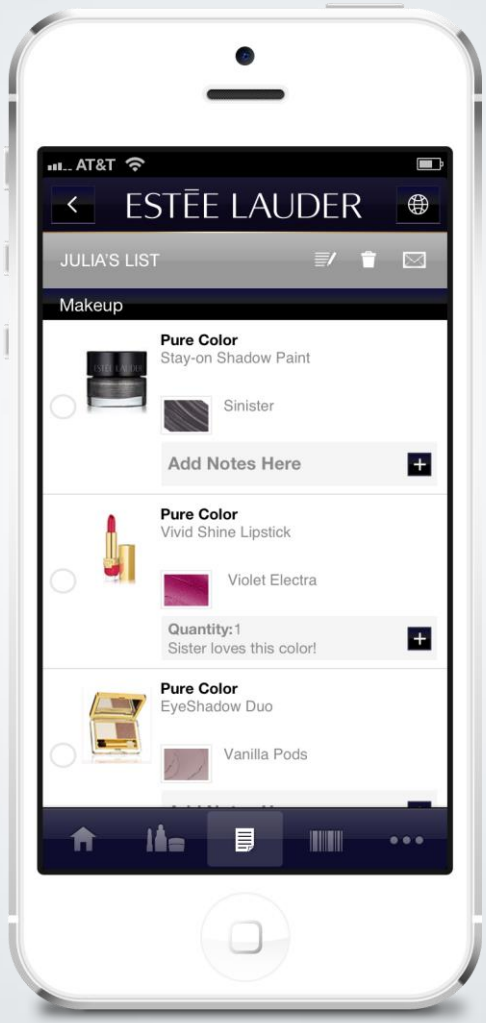
THE OVERVIEW

- THE BUSINESS & THE BUSINESS PROBLEM
- THE DATA
- FULL MODEL & DISCOVERIES
- TAKE 2!
- TAKEAWAYS

THE BUSINESS
& THE BUSINESS PROBLEM

A SAMPLE OF WHAT WE BUILD



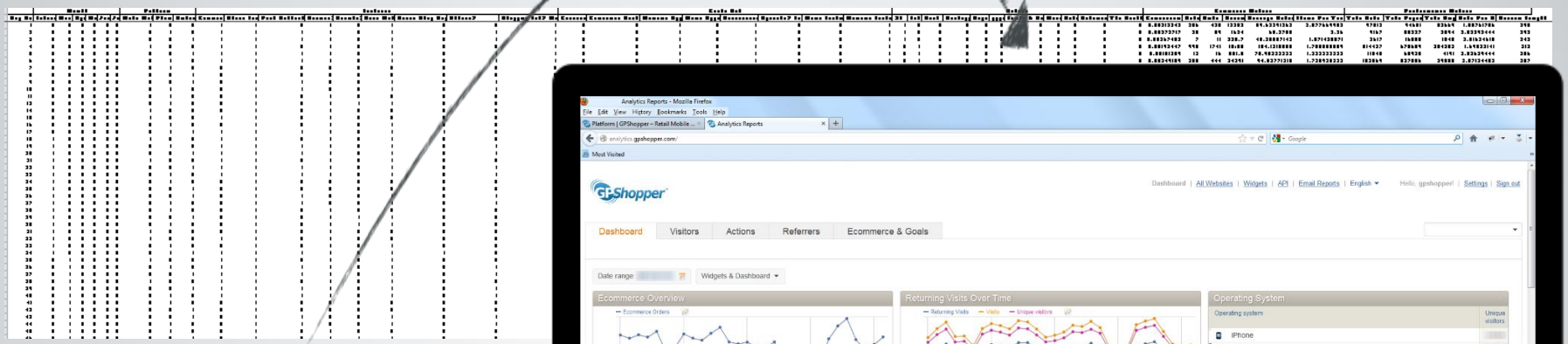


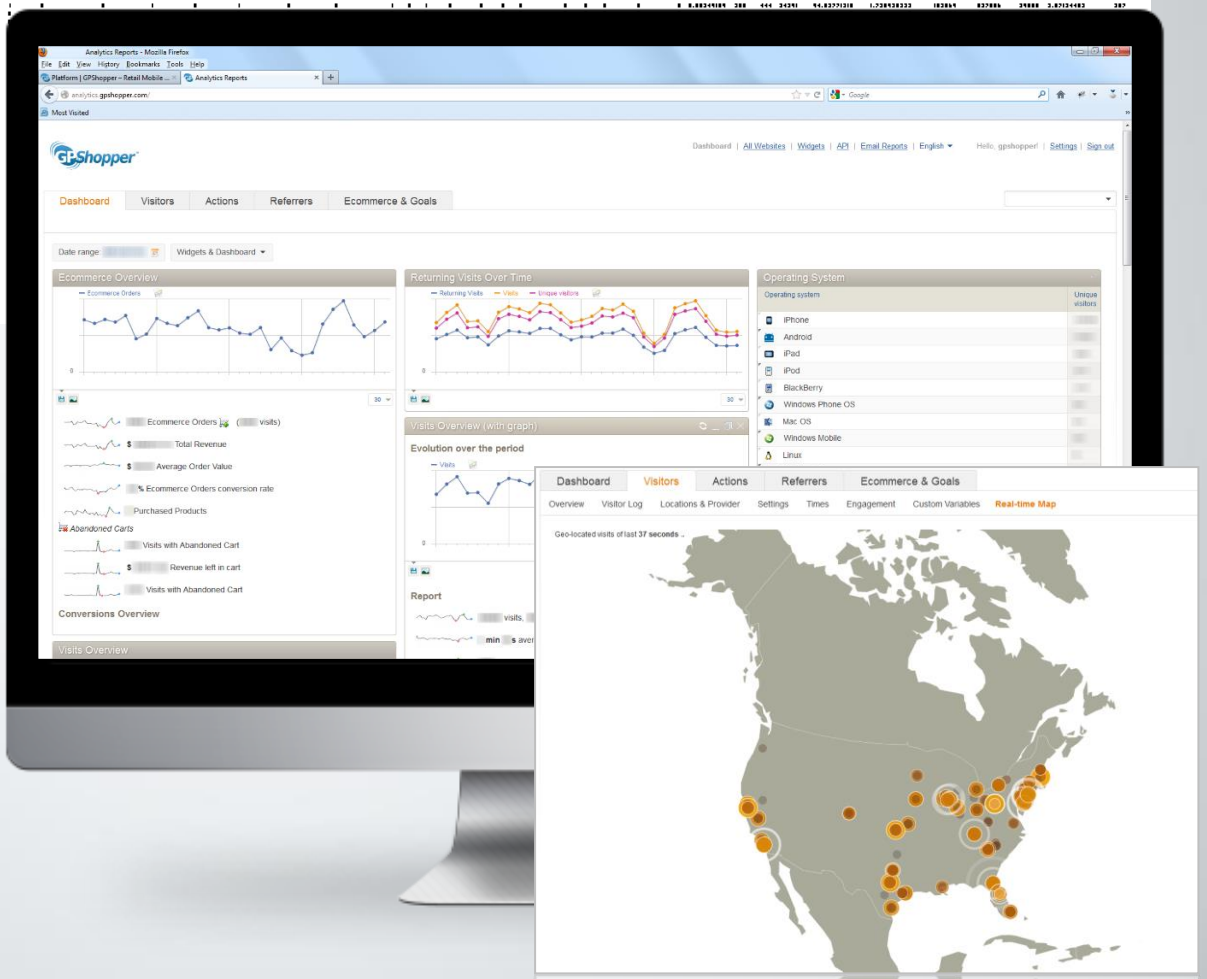
WHAT DRIVES COMMERCE?
WHAT DRIVES ENGAGEMENT?

...WE SPECULATE, BUT...

THE DATA

A SAMPLE OF THE DATA WE WORK WITH





...IT COMES FROM HERE

WHAT THE MODEL CONTAINED

EACH ROW = 1 RETAILER, 1 PLATFORM, 1 MONTH OF DATA (JAN-JULY 2013)

MONTH

PLATFORM

FEATURES

GOODS SOLD

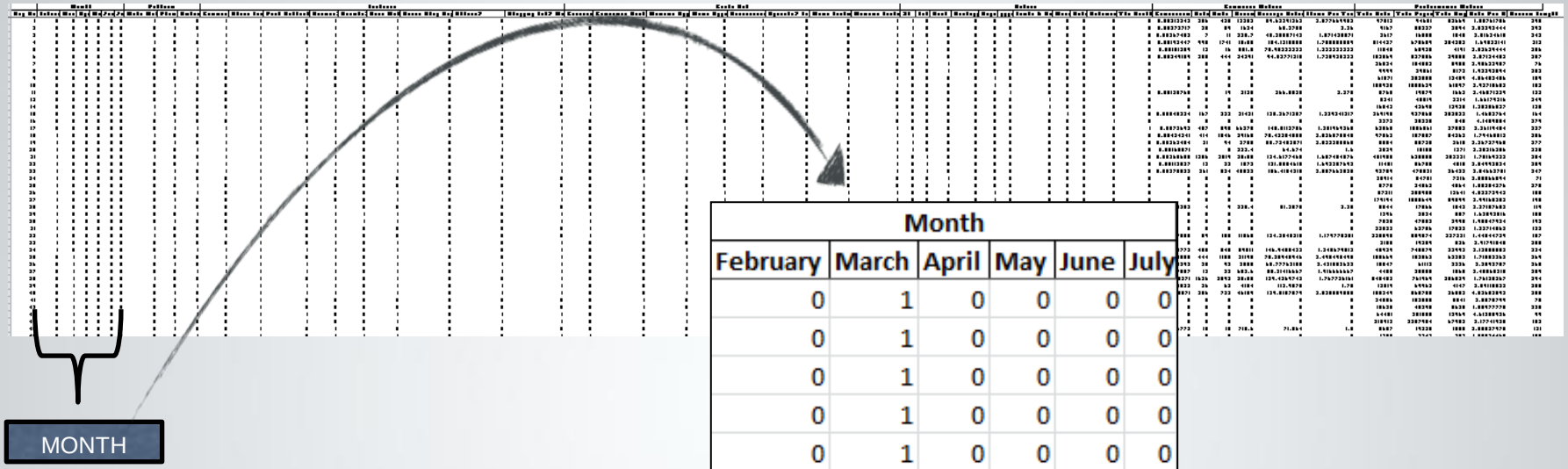
RETAILER

COMMERCE METRICS

PERFORMANCE METRICS

GPShopper

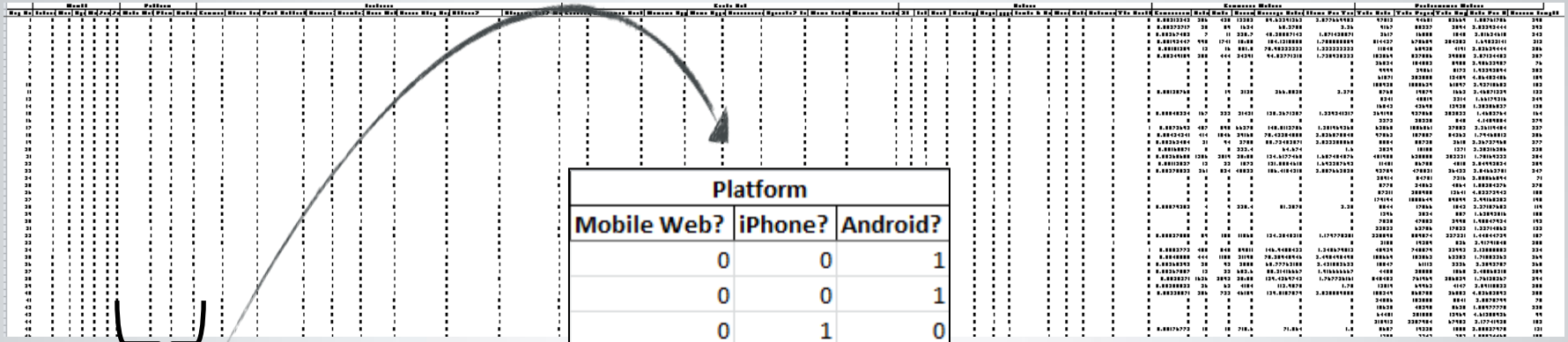
WHAT THE MODEL CONTAINED: CONTROL FOR SEASONALITY



MONTH

Month					
February	March	April	May	June	July
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0

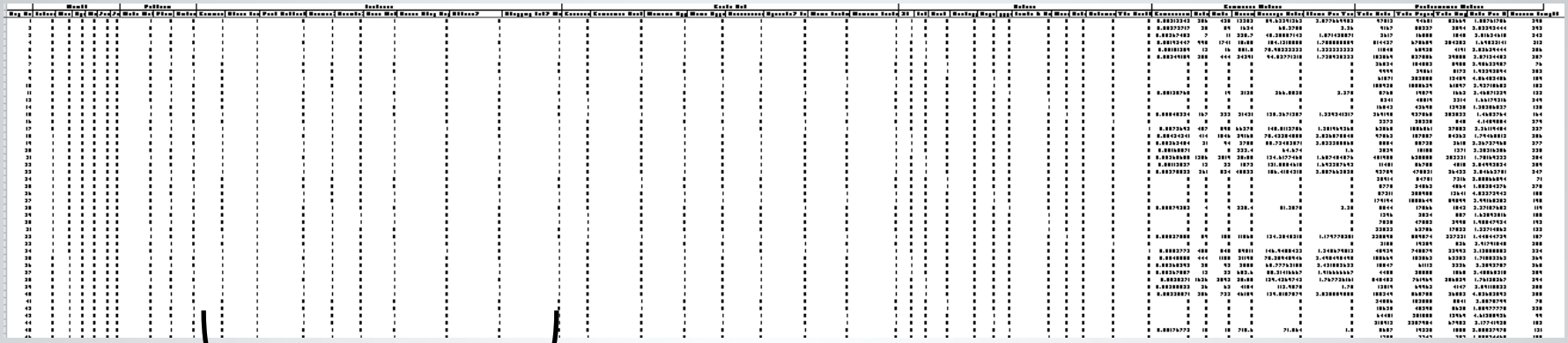
WHAT THE MODEL CONTAINED: CONTROL FOR PLATFORM



PLATFORM

Platform		
Mobile Web?	iPhone?	Android?
0	0	1
0	0	1
0	1	0
0	1	0
0	0	1
0	1	0
1	0	0
0	1	0
0	1	0
0	0	1
0	0	1
0	1	0
0	1	0
0	0	1
0	1	0

WHAT THE MODEL CONTAINED: CONTROL FOR FEATURE SET



FEATURES

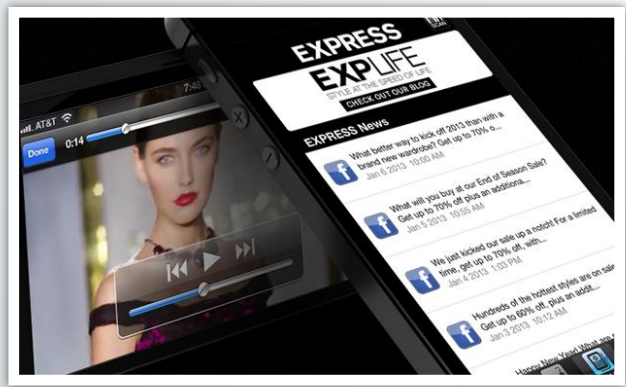
Features								
Commerce?	Store Locator?	Push Notifications?	Browse?	Events?	Social Media?	Social product Share?	Offers?	Shopping List? Wish List?
0	1	0	0	0	0	0	0	0
0	1	1	0	0	1	0	1	1
0	1	1	0	0	1	0	1	1
0	1	0	1	0	0	0	0	0
0	1	0	1	1	1	0	0	1
0	1	0	1	1	1	0	0	1
0	1	0	1	0	0	0	0	0
0	1	0	1	0	1	0	0	1
0	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0

WHAT IS A FEATURE?

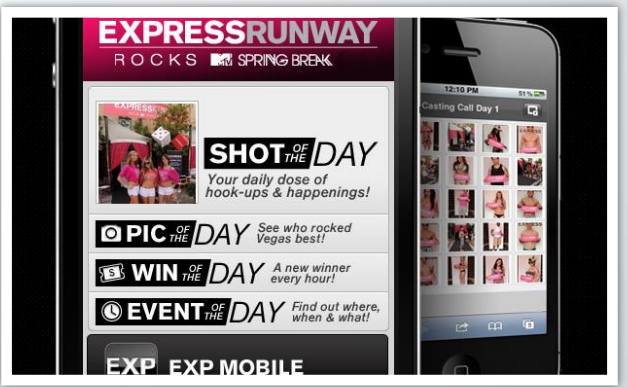
COMMERCE

PUSH NOTIFICATIONS

OFFERS



SOCIAL

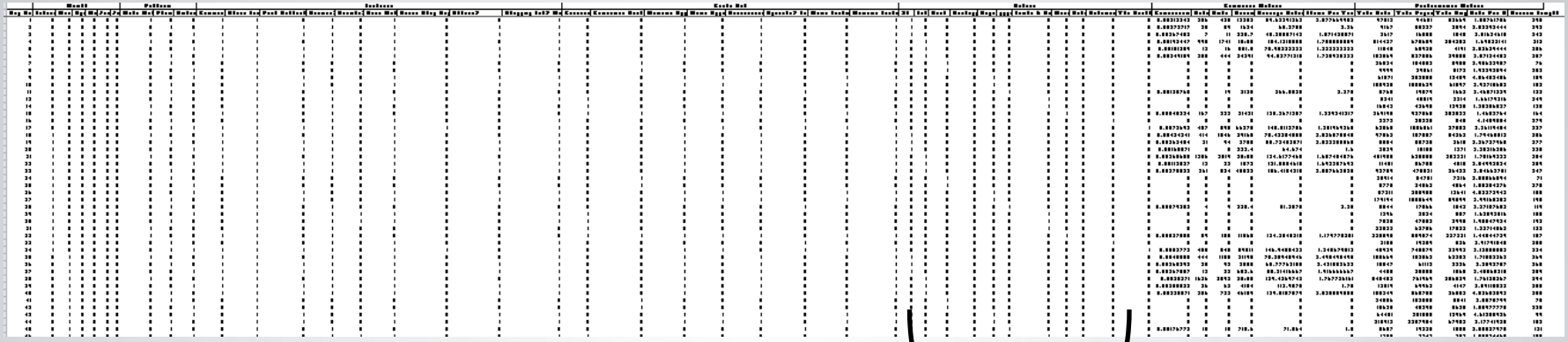


EVENTS

WHAT THE MODEL CONTAINED: CONTROL FOR GOODS SOLD

Goods Sold							
Grocery?	Consumer Electronics?	Women's Apparel?	Men's Apparel?	Accessories?	Specialty? Luxury?	Men's Footwear?	Women's Footwear?
0	1	0	0	0	0	0	0
0	0	1	1	1	0	1	1
0	0	1	1	1	0	1	1
0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	0

WHAT THE MODEL CONTAINED: CONTROL FOR RETAILER



RETAILER

Retailer										
A	B	C	D	E	F	G	H	I	J	K
0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0

WHAT THE MODEL CONTAINED: COMMERCE METRICS

Commerce Metrics					
Conversion Rate	Orders	Units	Revenue	Average Order Value	Items Per Transaction
0.001013086	12	16	851.8	70.98333333	1.333333333
0.002491093	258	444	24390.73	94.53771318	1.720930233
0.000452342	167	222	21420.61	128.2671257	1.329341317
0.007269199	457	595	66270.2	145.0113786	1.301969365
0.004243412	414	1046	29159.61	70.43384058	2.526570048
0.003624036	31	94	2750.47	88.72483871	3.032258065
0.00165071	5	8	323.37	64.674	1.6
0.002606052	1256	2019	169079.89	134.6177468	1.607484076
0.001135272	13	22	1573.11	121.0084615	1.692307692
0.002785218	261	524	40823.12	156.4104215	2.007662835
0.00027085	89	105	11067.58	124.3548315	1.179775281
0.0082773	405	545	59510.92	146.9405432	1.345679012
0.004085802	444	1108	31195.2	70.25945946	2.495495495
0.00360292	38	92	2499.55	65.77763158	2.421052632
0.002678571	12	23	602.57	50.21416667	1.916666667
0.003027096	1636	2892	211742.53	129.4269743	1.767726161
0.002808331	36	63	4103.55	113.9875	1.75

COMMERCE METRICS

WHAT THE MODEL CONTAINED: PERFORMANCE METRICS

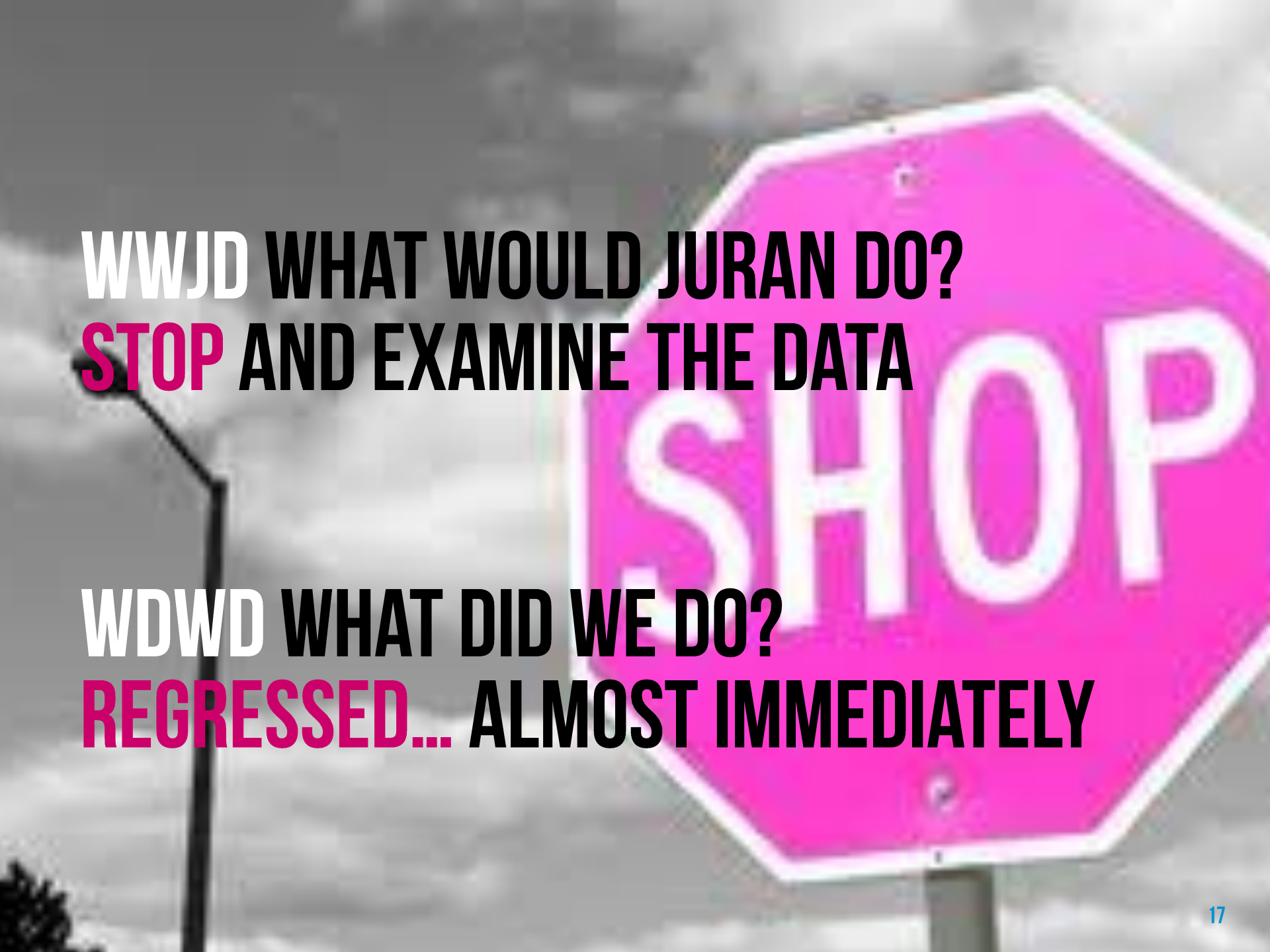
Performance Metrics				
Total Visits	Total Pageviews	Total Uniques	Visits Per Uniqu	Session Length
11845	60930	4191	2.826294	306
103569	527806	29000	3.5713448	257
369190	927860	252822	1.46027	164
62868	1006061	27803	2.2611948	337
97563	107887	54363	1.7946581	286
8554	50725	2618	3.2673798	277
3029	18100	1271	2.3831628	325
481955	630850	283221	1.7016923	304
11451	56700	4018	2.8499253	289
93709	470831	26422	3.546627	247
328595	809874	227331	1.4454473	157
48929	745079	22992	2.1280880	324
108669	152062	63252	1.7180326	269
10547	61113	3236	3.2592707	268
4480	25085	1860	2.4086021	289
540452	761969	306839	1.7613536	294
12819	69962	4147	3.0911502	305
108249	560708	26882	4.0268209	250
385372	944339	264069	1.4593	147

PERFORMANCE METRICS

20
MINUTES PER MONTH

5.4
MINUTES

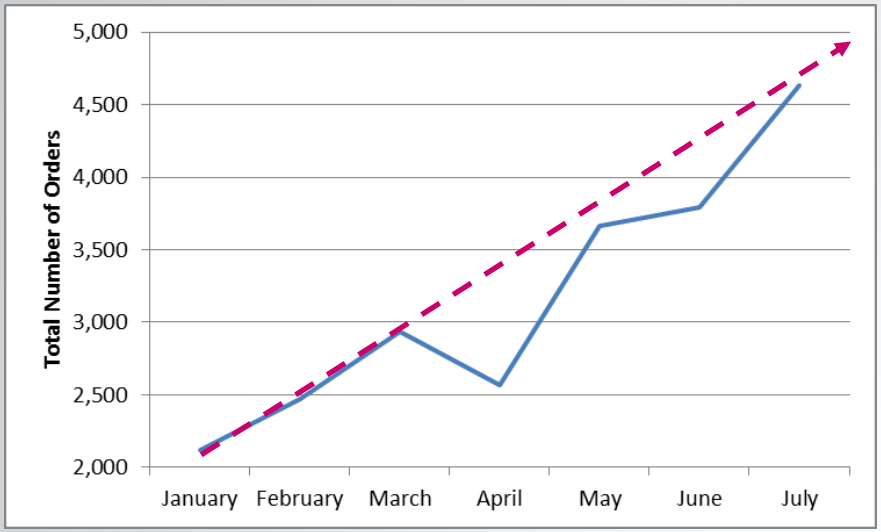
3.6
SESSIONS

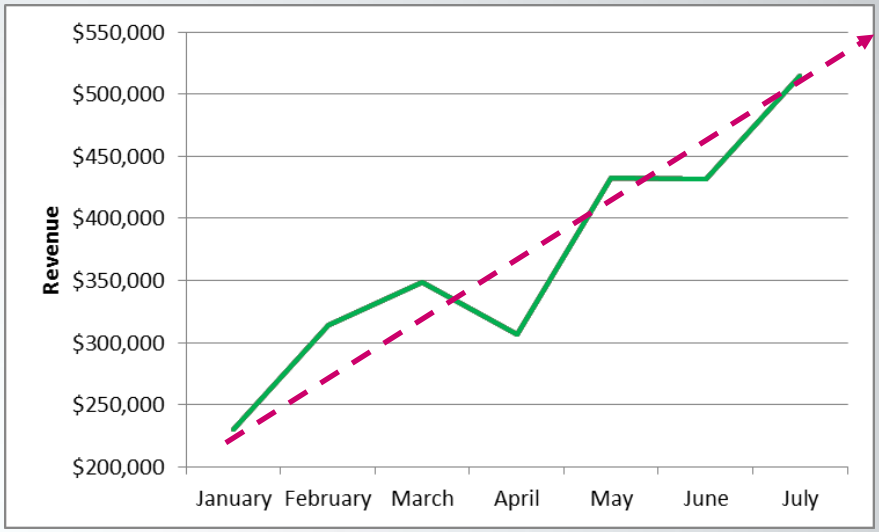


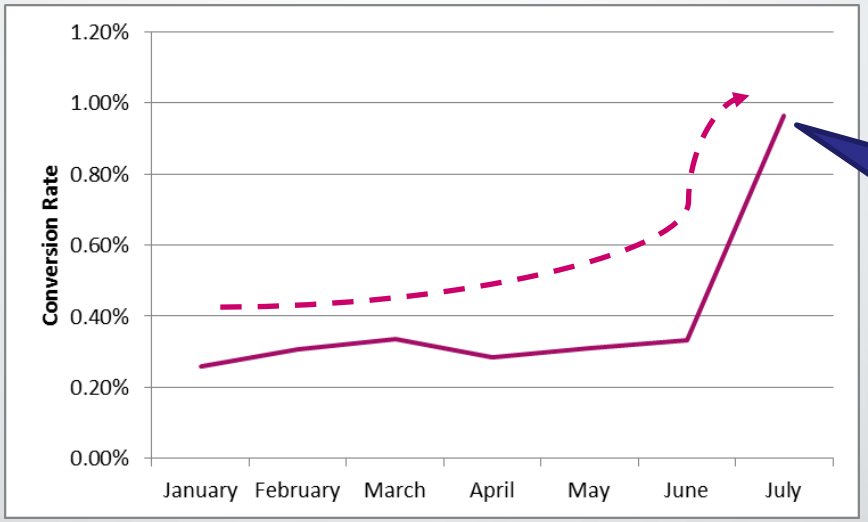
WWJD WHAT WOULD JURAN DO?
STOP AND EXAMINE THE DATA

WDWD WHAT DID WE DO?
REGRESSED… ALMOST IMMEDIATELY

DID WE NEED TO CONTROL FOR SEASONALITY?





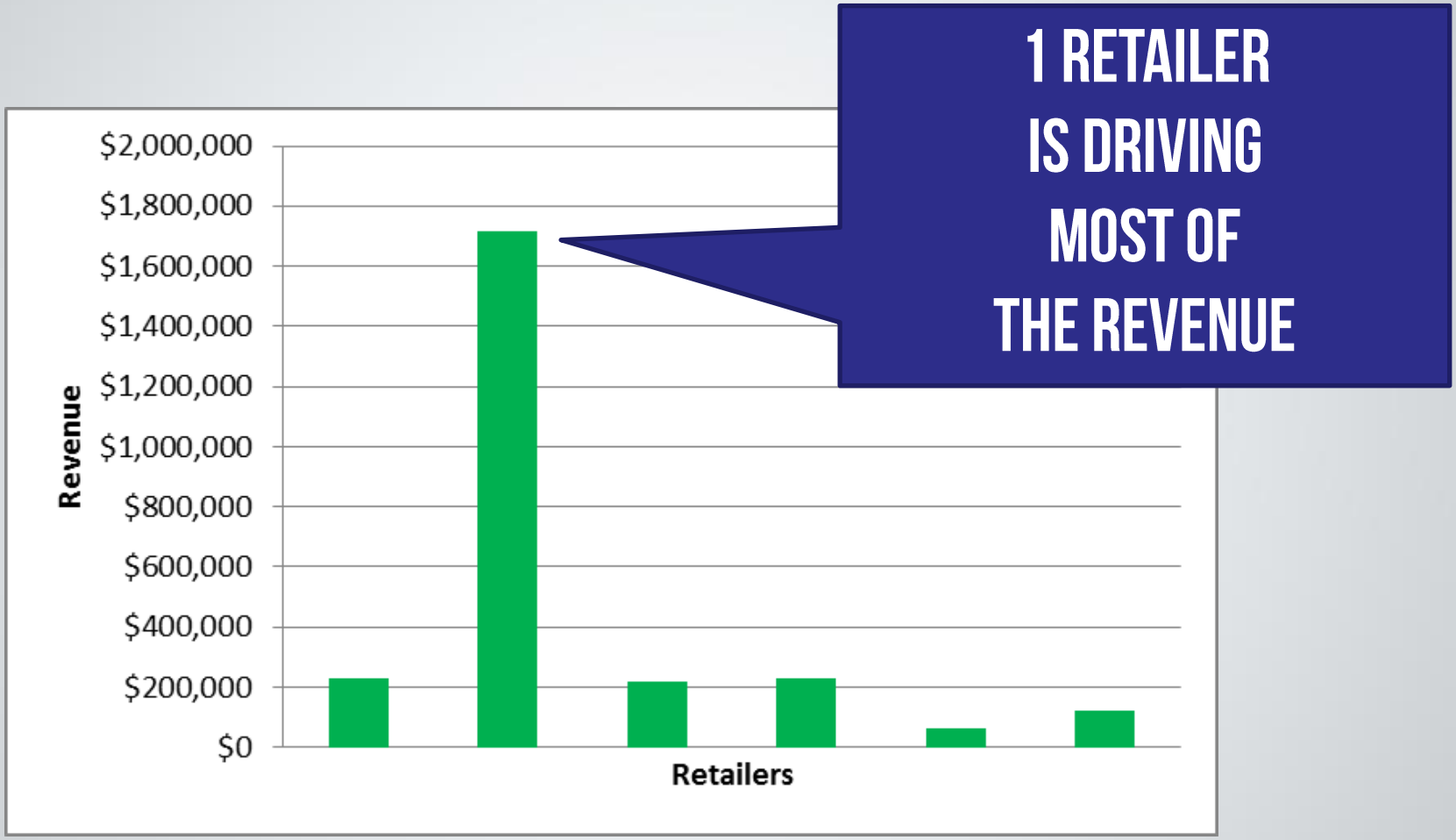


...WE'LL COME BACK TO THIS

DID WE NEED TO CONTROL FOR SEASONALITY?

- TOO LITTLE DATA - ONLY 7 MONTHS OF DATA
- COULD BE INTUITIVE – JAN/FEB SLOW IN RETAIL, POST-HOLIDAY
- GROWTH OVER TIME IS NORMAL…
  - SMARTPHONE USER ADOPTION
  - APP USER BASE GROWS OVER TIME (DOWNLOADS), SO DOES REVENUE
  - LAUNCHING NEW RETAILERS ADDS REVENUE OVER TIME

DID WE NEED TO CONTROL FOR RETAILER?



WE RAN THE FULL MODEL ANYWAY...

THE FULL MODEL & DISCOVERIES

INITIAL ROADBLOCKS

DEPPENDENT: REVENUE
INDEPENDENT: ALL DATA

- LOADS OF DATA TOSSED OUT OF MODEL
  - NEARLY EVERY RETAILER
  - NEARLY EVERY FEATURE

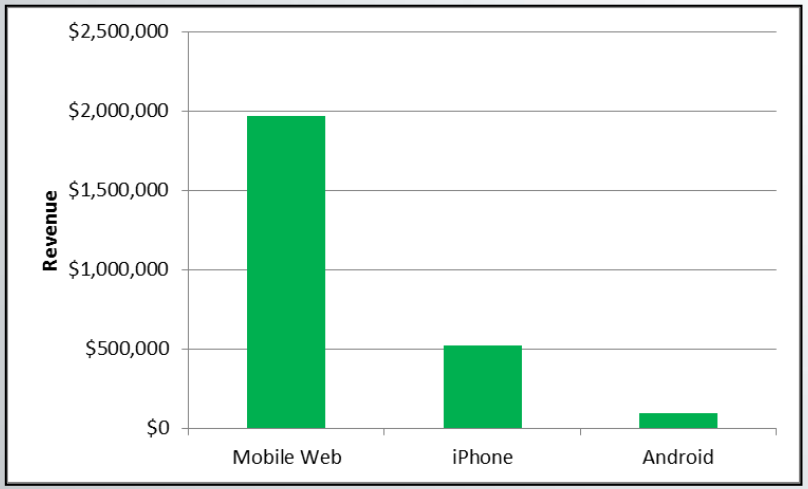
```
* Store Locator? is (essentially) constant
* Store Locator? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Shopping List? Wish List? is highly correlated with other X variables
* Shopping List? Wish List? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Grocery? is highly correlated with other X variables
* Grocery? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Consumer Electronics? Is highly correlated with other X variables
* Consumer Electronics? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Women_s Apparel? is highly correlated with other X variables
* Women_s Apparel? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Men_s Apparel? is highly correlated with other X variables
* Men_s Apparel? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* Women_s Footwear? is highly correlated with other X variables
* Women_s Footwear? has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* B is highly correlated with other X variables
* B has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* C is highly correlated with other X variables
* C has been removed from the equation.

* NOTE * All values in column are identical.
* NOTE * All values in column are identical.


* D has all values = 0
* D has been removed from the equation.
```

GPShopper

INITIAL ROADBLOCKS

DEPPENDENT: REVENUE
INDEPENDENT: ALL DATA

- LOADS OF DATA TOSSED OUT OF MODEL
  - NEARLY EVERY RETAILER
  - NEARLY EVER FEATURE
- NEGATIVE COEFFICIENTS WHERE POSITIVE RELATIONSHIPS ARE EXPECTED
  - REVENUE VS. PUSH NOTIFICATIONS/MW



Predictor	Coef	SE Coef	T	P
Constant	131418	61399	2.14	0.039
February	1895	2431	0.78	0.441
March	1935	2735	0.71	0.484
April	7390	7503	0.98	0.331
May	5146	8004	0.64	0.524
June	5231	8034	0.65	0.519
July	4554	8242	0.55	0.584
Mobile Web?	-145656	65688	-2.22	0.033
iPhone?	13605	7595	1.79	0.082
Android?	13541	8802	1.54	0.133
Push Notifications?	-146065	68268	-2.14	0.039
Social Media?	-5677	11295	-0.50	0.618
Shopping List? Wish List?	-36049	8109	-4.45	0.000
Men_s Apparel?	-6783	6762	-1.00	0.323
Ã	11156	4446	2.51	0.017
Conversion Rate	-1760520	951413	-1.85	0.072
Orders	189.82	13.19	14.39	0.000
Units	-20.497	6.033	-3.40	0.002
Average Order Value	176.23	54.00	3.26	0.002
Items Per Transaction	-3861	1796	-2.15	0.038
Total Visits	-0.03339	0.05378	-0.62	0.539
Total Pageviews	-0.008489	0.005404	-1.57	0.125
Total Uniques	0.00441	0.08050	0.05	0.957
Visits Per Unique	5170	3583	1.44	0.158
Session Length	56.93	57.14	1.00	0.326

S = 3683.34 R-Sq = 99.8% R-Sq(adj) = 99.6%

INITIAL ROADBLOCKS

DEPPENDENT: REVENUE
INDEPENDENT: ALL DATA

- LOADS OF DATA TOSSED OUT OF MODEL
  - NEARLY EVERY RETAILER
  - NEARLY EVER FEATURE
- NEGATIVE CORRELATIONS WHERE POSITIVE RELATIONSHIPS ARE EXPECTED
  - REVENUE VS. PUSH NOTIFICATIONS/MW
- HIGH ADJUSTED R-SQUARE

Predictor	Coef	SE Coef	T	P
Constant	131418	61399	2.14	0.039
February	1895	2431	0.78	0.441
March	1935	2735	0.71	0.484
April	7390	7503	0.98	0.331
May	5146	8004	0.64	0.524
June	5231	8034	0.65	0.519
July	4554	8242	0.55	0.584
Mobile Web?	-145656	65688	-2.22	0.033
iPhone?	13605	7595	1.79	0.082
Android?	13541	8802	1.54	0.133
Push Notifications?	-146065	68268	-2.14	0.039
Social Media?	-5677	11295	-0.50	0.618
Shopping List? Wish List?	-36049	8109	-4.45	0.000
Men_s Apparel?	-6783	6762	-1.00	0.323
A	11156	4446	2.51	0.017
Conversion Rate	-1760520	951413	-1.85	0.072
Orders	189.82	13.19	14.39	0.000
Units	-20.497	6.033	-3.40	0.002
Average Order Value	176.23	54.00	3.26	0.002
Items Per Transaction	-3861	1796	-2.15	0.038
Total Visits	-0.03339	0.05378	-0.62	0.539
Total Pageviews	-0.008489	0.005404	-1.57	0.125
Total Uniques	0.00441	0.08050	0.05	0.957
Visits Per Unique	5170	3583	1.44	0.158
Session Length	56.93	57.14	1.00	0.326

S = 3683.34 R-Sq = 99.8% R-Sq(adj) = 99.6%

INITIAL ROADBLOCKS

DEPPENDENT: REVENUE
INDEPENDENT: ALL DATA

- LOADS OF DATA TOSSED OUT OF MODEL
  - NEARLY EVERY RETAILER
  - NEARLY EVER FEATURE
- NEGATIVE CORRELATIONS WHERE POSITIVE RELATIONSHIPS ARE EXPECTED
  - REVENUE VS. PUSH NOTIFICATIONS/MW
- HIGH ADJUSTED R-SQUARE
- HIGH P-VALUES

Predictor	Coef	SE Coef	T	P
Constant	131418	61399	2.14	0.039
February	1895	2431	0.78	0.441
March	1935	2735	0.71	0.484
April	7390	7503	0.98	0.331
May	5146	8004	0.64	0.524
June	5231	8034	0.65	0.519
July	4554	8242	0.55	0.584
Mobile Web?	-145656	65688	-2.22	0.033
iPhone?	13605	7595	1.79	0.082
Android?	13541	8802	1.54	0.133
Push Notifications?	-146065	68268	-2.14	0.039
Social Media?	-5677	11295	-0.50	0.618
Shopping List? Wish List?	-36049	8109	-4.45	0.000
Men_s Apparel?	-6783	6762	-1.00	0.323
A	11156	4446	2.51	0.017
Conversion Rate	-1760520	951413	-1.85	0.072
Orders	189.82	13.19	14.39	0.000
Units	-20.497	6.033	-3.40	0.002
Average Order Value	176.23	54.00	3.26	0.002
Items Per Transaction	-3861	1796	-2.15	0.038
Total Visits	-0.03339	0.05378	-0.62	0.539
Total Pageviews	-0.008489	0.005404	-1.57	0.125
Total Uniques	0.00441	0.08050	0.05	0.957
Visits Per Unique	5170	3583	1.44	0.158
Session Length	56.93	57.14	1.00	0.326

S = 3683.34 R-Sq = 99.8% R-Sq(adj) = 99.6%

TAKE 2!

ADJUST THE DATA BASED ON LEARNINGS

TAKING OUT THE TRASH...

- REMOVE MONTHS
  - SEASONALITY IS NOT A FACTOR
- REMOVE VARIABLES WHERE ALL VALUES =
  - STORE LOCATOR
  - BROWSE
- VARIABLES WITH BIG P-VALUES

CHANGE APPROACH...

1. DRIVING COMMERCE
   - REVENUE
   - CONVERSION RATE = ORDERS/VISITS
2. DRIVING ENGAGEMENT
   - SESSION LENGTH

ISOLATE MOBILE WEB DATA ONLY

CORRELATION MATRIX (SAMPLE)

	June	July	Commerce?	Store Locator?	Browse?	Events?	Shopping List? Wish List?	Grocery?	Consumer Electronics?	Womens Apparel?
Accessories?	0.019518001	0.060522753	1	#DIV/0!	#DIV/0!	#DIV/0!	0.466666667	-0.6	0.28	1
Specialty? Luxury?	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Mens Footwear?	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	1	0.28
Womens Footwear?	-0.050964719	-0.158034885	0.731126155	#DIV/0!	#DIV/0!	#DIV/0!	0.638284739	-0.290129427	0.382970843	0.731126155
A	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	0.6	-0.466666667	-0.28	0.28
B	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	0.6	-0.466666667	-0.28	0.28
C	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
D	0.156863684	0.120172781	0.170192589	#DIV/0!	#DIV/0!	#DIV/0!	-0.283654314	-0.283654314	-0.170192589	0.170192589
E	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
F	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
G	-0.019518001	-0.060522753	-1	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	-0.28	-1
H	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	1	0.28
I	-0.077289816	0.373878251	0.095038193	#DIV/0!	#DIV/0!	#DIV/0!	-0.158396988	-0.158396988	-0.095038193	0.095038193
J	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
K	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Conversion Rate	0.096439107	0.199172938	0.61005944	#DIV/0!	#DIV/0!	#DIV/0!	0.488664143	-0.892152917	-0.460524061	0.61005944
Orders	0.062665428	0.093408232	0.504276656	#DIV/0!	#DIV/0!	#DIV/0!	0.668405536	-0.652195615	-0.278358081	0.504276656
Units	0.080824866	0.202471925	0.56092539	#DIV/0!	#DIV/0!	#DIV/0!	0.648864502	-0.79649	-0.39486261	0.56092539
Revenue	0.051335925	0.058942214	0.429751156	#DIV/0!	#DIV/0!	#DIV/0!	0.587747256	-0.535346214	-0.2126643	0.429751156
Average Order Value	0.01049364	-0.015226637	0.867482553	#DIV/0!	#DIV/0!	#DIV/0!	0.26977814	-0.312761698	0.492168515	0.867482553
Items Per Transaction	0.058588313	0.242066855	0.795260995	#DIV/0!	#DIV/0!	#DIV/0!	0.408192	-0.79368712	-0.157163549	0.795260995
Total Visits	0.01515169	0.010485802	0.571415494	#DIV/0!	#DIV/0!	#DIV/0!	0.308093479	-0.12655852	0.41954527	0.571415494
Total Pageviews	0.029776832	0.225199721	0.661811803	#DIV/0!	#DIV/0!	#DIV/0!	-0.131388055	-0.053654943	0.597425872	0.661811803
Total Uniques	0.001551968	-0.019019531	0.564254686	#DIV/0!	#DIV/0!	#DIV/0!	0.23812802	-0.024630069	0.534698603	0.564254686
Visits Per Unique	0.108126401	0.307218051	0.585135136	#DIV/0!	#DIV/0!	#DIV/0!	0.231880106	-0.741602113	-0.304787399	0.585135136
Session Length	-0.172265369	-0.206478113	0.542166959	#DIV/0!	#DIV/0!	#DIV/0!	0.762169711	-0.730648376	-0.334611093	0.542166959

ISOLATE MOBILE WEB DATA ONLY

- RUN A CORRELATION OF FULL SET, MOBILE WEB ONLY
- #DIVO QUICKLY IDENTIFIES VARIABLES WHERE ALL VALUES = SAME, CAN BE REMOVED FROM MODEL
- RESULTS
  - STRONG NEGATIVE CORRELATION BETWEEN GROCERY AND CONVERSION RATE (-.89) BECAUSE…
    - GROCERY CLIENT 1 IS NOT COMMERCE ENABLED
    - GROCERY CLIENT 2 IS COMMERCE ENABLED AND ONLY GROCERY PRODUCTS CANNOT BUY ONLINE
    - PEOPLE MOSTLY STILL BUY GROCERIES IN-STORE
  - STRONG POSITIVE CORRELATION BETWEEN SHOPPING LIST/WISH LIST AND SESSION LENGTH (.76)
    - INTUITIVELY MAKES SENSE – A USER WHO USES THE FEATURE HAS A LONG SESSION LENGTH
  - WOMEN'S APPAREL IS STRONGLY POSITIVELY CORRELATED WITH COMMERCE METRICS (.79-.86)
    - INTUITIVELY MAKES SENSE – ALL COMMERCE ENABLED SELL WOMEN'S APPAREL

ISOLATE MOBILE WEB DATA ONLY

	June	July	Commerce?	Store Locator?	Browse?	Events?	Shopping List? Wish List?	Grocery?	Consumer Electronics?	Womens Apparel?
Accessories?	0.019518001	0.060522753	1	#DIV/0!	#DIV/0!	#DIV/0!	0.466666667	-0.6	0.28	1
Specialty? Luxury?	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Mens Footwear?	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	1	0.28
Womens Footwear?	-0.050964719	-0.158034885	0.731126155	#DIV/0!	#DIV/0!	#DIV/0!	0.638284739	-0.290129427	0.382970843	0.731126155
A	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	0.6	-0.466666667	-0.28	0.28
B	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	0.6	-0.466666667	-0.28	0.28
C	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
D	0.156863684	0.120172781	0.170192589	#DIV/0!	#DIV/0!	#DIV/0!	-0.283654314	-0.283654314	-0.170192589	0.170192589
E	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
F	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
G	-0.019518001	-0.060522753	-1	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	-0.28	-1
H	-0.019518001	-0.060522753	0.28	#DIV/0!	#DIV/0!	#DIV/0!	-0.466666667	0.6	1	0.28
I	-0.077289816	0.373878251	0.095038193	#DIV/0!	#DIV/0!	#DIV/0!	-0.158396988	-0.158396988	-0.095038193	0.095038193
J	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
K	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
Conversion Rate	0.096439107	0.199172938	0.61005944	#DIV/0!	#DIV/0!	#DIV/0!	0.488664143	-0.892152917	-0.460524061	0.61005944
Orders	0.062665428	0.093408232	0.504276656	#DIV/0!	#DIV/0!	#DIV/0!	0.668405536	-0.652195615	-0.278358081	0.504276656
Units	0.080824866	0.202471925	0.56092539	#DIV/0!	#DIV/0!	#DIV/0!	0.648864502	-0.79649	-0.39486261	0.56092539
Revenue	0.051335925	0.058942214	0.429751156	#DIV/0!	#DIV/0!	#DIV/0!	0.587747256	-0.535346214	-0.2126643	0.429751156
Average Order Value	0.01049364	-0.015226637	0.867482553	#DIV/0!	#DIV/0!	#DIV/0!	0.26977814	-0.312761698	0.492168515	0.867482553
Items Per Transaction	0.058588313	0.242066855	0.795260995	#DIV/0!	#DIV/0!	#DIV/0!	0.408192	-0.79368712	-0.157163549	0.795260995
Total Visits	0.01515169	0.010485802	0.571415494	#DIV/0!	#DIV/0!	#DIV/0!	0.308093479	-0.12655852	0.41954527	0.571415494
Total Pageviews	0.029776832	0.225199721	0.661811803	#DIV/0!	#DIV/0!	#DIV/0!	-0.131388055	-0.053654943	0.597425872	0.661811803
Total Uniques	0.001551968	-0.019019531	0.564254686	#DIV/0!	#DIV/0!	#DIV/0!	[illegible]	[illegible]	0.534698603	0.564254686
Visits Per Unique	0.108126401	0.307218051	0.585135136	#DIV/0!	#DIV/0!	#DIV/0!	0.231880106	-0.741602113	-0.304787399	0.585135136
Session Length	-0.172265369	-0.206478113	0.542166959	#DIV/0!	#DIV/0!	#DIV/0!	0.762169711	-0.730648376	-0.334611093	0.542166959

ISOLATE COMMERCE-ENABLED ONLY

AFTER FULL MODEL…INDEPENDENT VS. DEPENDENT VARIABLES

- REGRESSION 6: REVENUE VS ALL PERFORMANCE METRICS
- REGRESSION 7: REVENUE VS ALL PERFORMANCE METRICS, CONTROL FOR RETAILER
- REGRESSION 8: REVENUE VS PERFORMANCE METRICS, VISITS PER UNIQUES
- REGRESSION 9: REVENUE VS TOTAL VISITS, TOTAL UNIQUES, SESSION LENGTH
- REGRESSION 10: REVENUE VS TOTAL VISITS, TOTAL PAGEVIEWS, SESSION LENGTH
- REGRESSION 11: REVENUE SESSION LENGTH AND VISITS
- REGRESSION 12: REVENUE VS SESSION LENGTH
- REGRESSION 13: REVENUE VS TOTAL VISITS
- REGRESSION 14: REVENUE VS TOTAL UNIQUES
- REGRESSION 15: REVENUE VS VISITS PER UNIQUE
- REGRESSION 16: REVENUE VS TOTAL PAGEVIEWS
- REGRESSION 17: CONVERSION RATE VS FEATURES
- REGRESSION 18: CONVERSION RATE VS FEATURES (REMOVAL OF ROW 58)

ISOLATE COMMERCE-ENABLED ONLY

REGRESSION 13: REVENUE VS TOTAL VISITS

- INTUITIVELY WOULD HAVE EXPECTED VISITS PER UNIQUE TO EXPLAIN REVENUE (REGRESSION 15)

Regression 13 Analysis: Revenue versus Total Visits

The regression equation is
Revenue = 4942 + 0.233 Total Visits

Predictor	Coef	SE Coef	T	P
Constant	4942	6154	0.80	0.425
Total Visits	0.23347	0.02373	9.84	0.000

S = 37847.6 R-Sq = 62.1% R-Sq(adj) = 61.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	1.38656E+11	1.38656E+11	96.80	0.000
Residual Error	59	84514097109	1432442324		
Total	60	2.23171E+11			

Regression 15 Analysis: Revenue versus Visits Per Unique

The regression equation is
Revenue = 90344 - 18672 Visits Per Unique

Predictor	Coef	SE Coef	T	P
Constant	90344	22624	3.99	0.000
Visits Per Unique	-18672	8282	-2.25	0.028

S = 59013.1 R-Sq = 7.9% R-Sq(adj) = 6.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	17700145115	17700145115	5.08	0.028
Residual Error	59	2.05470E+11	3482548923		
Total	60	2.23171E+11			

Unusual Observations

Obs	Visits Per Unique	Revenue	Fit	SE Fit	Residual	St Resid
20	1.76	211743	57457	10123	154285	2.65R
28	1.90	190956	54907	9408	136049	2.34R
36	1.92	214682	54408	9278	160274	2.75R
45	1.93	231448	54309	9253	177138	3.04R
54	1.88	246423	55183	9482	191240	3.28R

R denotes an observation with a large standardized residual.

ISOLATE COMMERCE-ENABLED ONLY

REGRESSION 17: CONVERSION RATE VS PLATFORM AND FEATURES

- P VALUE LOOKS GOOD
- VERY HIGH ADJUSTED R-SQUARE

Regression 17 Analysis: Conversion Rate versus Mobile Web?, iPhone?, ...

```
* Never Stop Exploring? is highly correlated with other X variables
* Never Stop Exploring? has been removed from the equation.


* Offers? is highly correlated with other X variables
* Offers? has been removed from the equation.


The regression equation is
Conversion Rate = 0.0670 - 0.0665 Mobile Web? + 0.00416 iPhone?
                  + 0.00230 Android? - 0.0701 Push Notifications?
                  + 0.00443 Social Media? - 0.00188 Shopping List? Wish List?
```

Predictor	Coef	SE Coef	T	P
Constant	0.067038	0.001148	58.40	0.000
Mobile Web?	-0.066532	0.001089	-61.09	0.000
iPhone?	0.0041555	0.0004446	9.35	0.000
Android?	0.0023042	0.0004446	5.18	0.000
Push Notifications?	-0.0700694	0.0009942	-70.48	0.000
Social Media?	0.0044295	0.0006020	7.36	0.000
Shopping List? Wish List?	-0.0018768	0.0005445	-3.45	0.001

S = 0.000960480 R-Sq = 99.0% R-Sq(adj) = 98.9%

Analysis of Variance

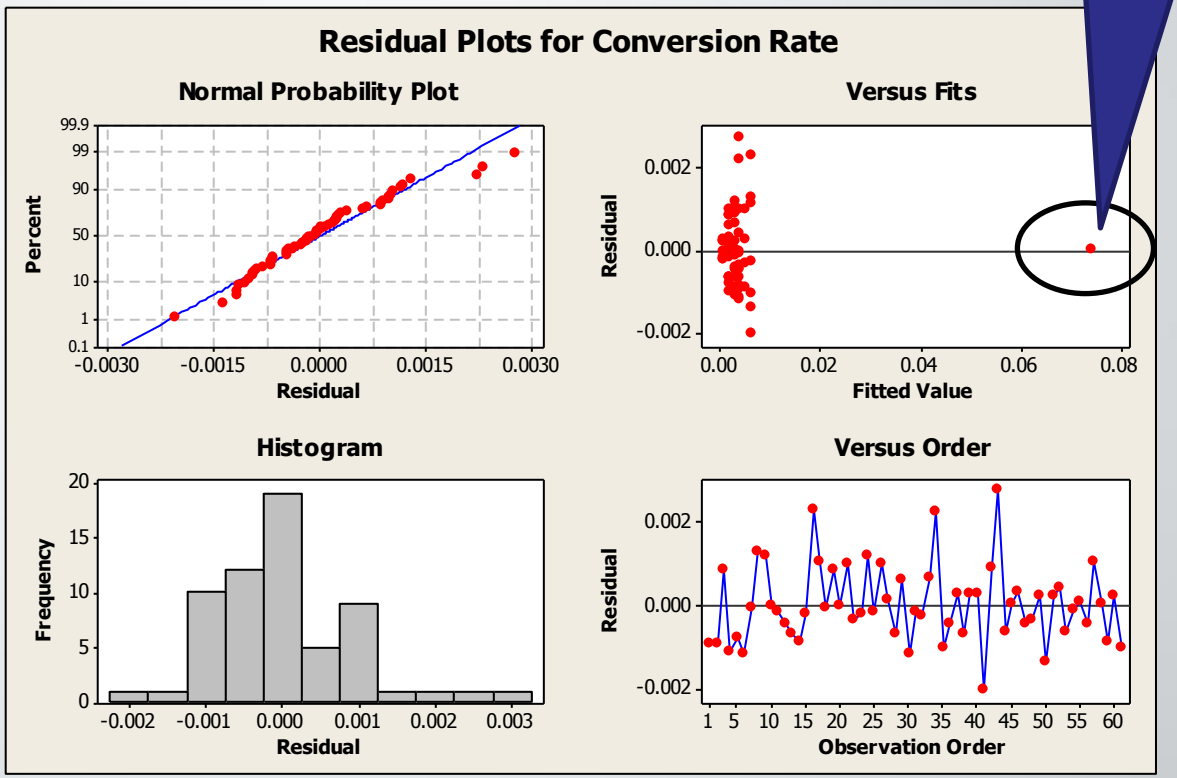
Source	DF	SS	MS	F	P
Regression	6	0.00505743	0.00084290	913.70	0.000
Residual Error	54	0.00004982	0.00000092		
Total	60	0.00510724			

ISOLATE COMMERCE-ENABLED ONLY

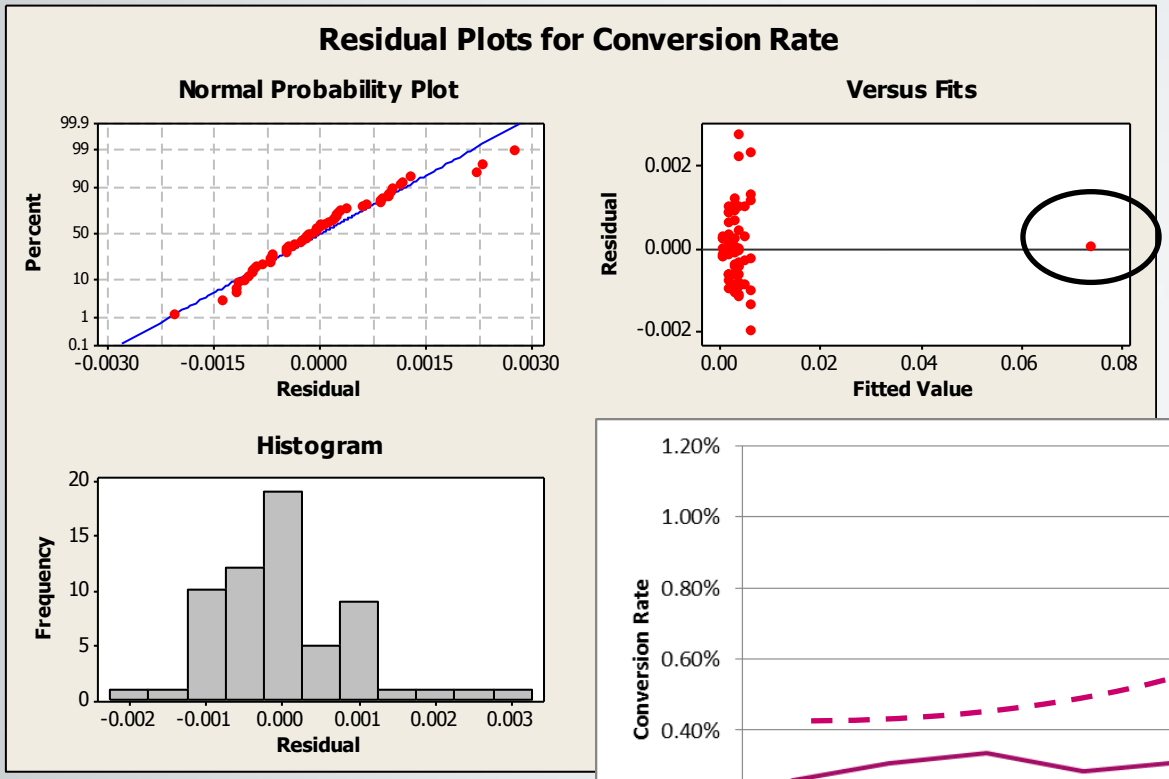
REGRESSION 17: CONVERSION RATE VS PLATFORM AND FEATURES

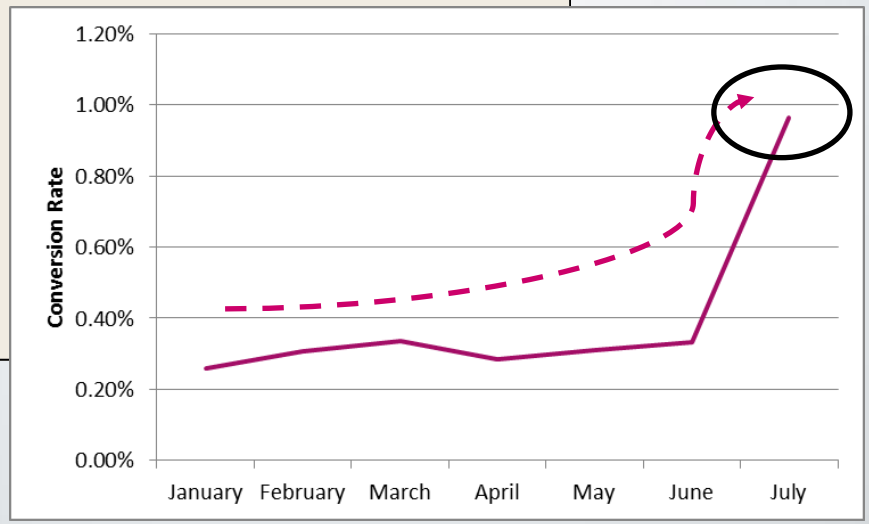
- P VALUE LOOKS GOOD
- VERY HIGH ADJUSTED R-SQUARE
- ...BUT LOOK AT THE 4 IN 1!

REMEMBER ME?



ISOLATE COMMERCE-ENABLED ONLY





A SINGLE OBSERVATION WITH A HIGH CONVERSION RATE IS SKEWING THE DATA

ISOLATE COMMERCE-ENABLED ONLY

REGRESSION 18: CONVERSION RATE VS FEATURES (REMOVAL OF OUTLIER)

- P VALUE LOOKS GOOD
- REASONABLE ADJUSTED R-SQUARE
- IMPORTANCE OF SOCIAL MEDIA TO CONVERSION
  - NEW PRODUCTS ANNOUNCED
- ON-GOING CHALLENGES:
  - SOME SEEMINGLY VALUABLE VARIABLES GOT KICKED OUT
    - SOCIAL PRODUCT SHARE
    - OFFERS
    - PUSH NOTIFICATIONS

Regression 18 Analysis: Conversion Rate versus Mobile Web?, iPhone?, ...

```
* Push Notifications? is highly correlated with other X variables
* Push Notifications? has been removed from the equation.

* Social Product Share? is highly correlated with other X variables
* Social Product Share? has been removed from the equation.

* Offers? is highly correlated with other X variables
* Offers? has been removed from the equation.

The regression equation is
Conversion Rate = - 0.00303 + 0.00354 Mobile Web? + 0.00416 iPhone?
                  + 0.00230 Android? + 0.00443 Social Media?
                  - 0.00188 Shopping List? Wish List?
```

Predictor	Coef	SE Coef	T	P
Constant	-0.0030313	0.0006792	-4.46	0.000
Mobile Web?	0.0035378	0.0005740	6.16	0.000
iPhone?	0.0041555	0.0004446	9.35	0.000
Android?	0.0023042	0.0004446	5.18	0.000
Social Media?	0.0044295	0.0006020	7.36	0.000
Shopping List? Wish List?	-0.0018768	0.0005445	-3.45	0.001

S = 0.000960480 R-Sq = 74.6% R-Sq(adj) = 72.2%

TAKEAWAYS

KEY TAKEAWAYS FOR DATA MODEL

1. KISS = KEEP IT SIMPLE, STUPID!
   - START WITH A SIMPLE MODEL AND BUILD UP CONTROLS AS NEEDED
   - OVER DOCTORING THE DATA TOO SOON
2. CHALLENGES OF TOO MANY BINARY VARIABLES
3. COMPARING DATA BEFORE AND AFTER
   - FOR EXAMPLE, COMPARE PERFORMANCE BEFORE A FEATURE IS INTRODUCED AND AFTER
4. NEED LONGER TIMEFRAME OF DATA TO ASSESS SEASONALITY
5. TOO DIVERSE (AND SMALL) OF A SAMPLE POOL
6. LOOK FOR OUTLIERS EARLY & OFTEN
7. SEEING EFFECT OF BRAND STRENGTH, CONSIDER WAYS TO NORMALIZE AGAINST IT
   - DIVIDING DATA BY NUMBER OF STORES

AND DON'T FORGET - YOU NEED AS MANY ROWS AS COLUMNS WHEN USING MINITAB!

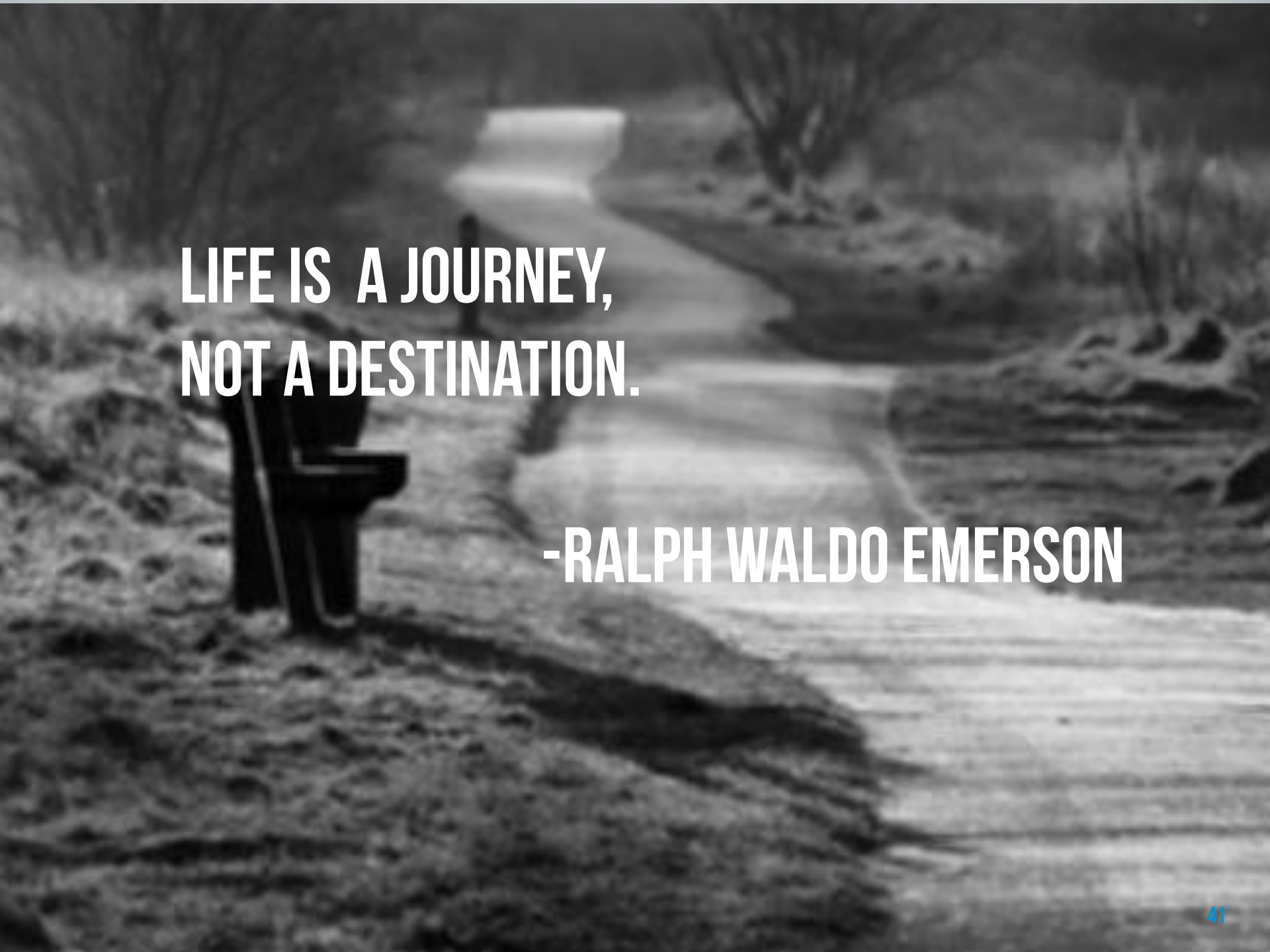
KEY TAKEAWAYS FOR THE BUSINESS

1. CONVERSION RATE LESS IMPORTANT FOR A GROCER
   - GROCERY IS BROWSE AND BUY IN-STORE FOR OUR CLIENT

2. POSITIVE CORRELATION BETWEEN WISHLIST AND SESSION LENGTH

3. TOTAL VISITS VS VISITS PER UNIQUE IN EXPLAINING REVENUE
   - THE STICKY CUSTOMER MATTERS LESS THAN ONE WOULD ASSUME

4. SOCIAL MEDIA'S IMPACT ON CONVERSION RATE
   - IMPORTANT FEATURE FOR DRIVING CONVERSION RATE
   - NEW PRODUCT ANNOUNCEMENTS, FREE MASS MARKETING


LIFE IS A JOURNEY,
NOT A DESTINATION.
-RALPH WALDO EMERSON

THANKYOU