Applied Regression Analysis – Final Project

17th June, 2016

If you have any questions regarding this project please contact cpereira17@gsb.columbia.edu

Caio Pereira
Lorenzo Prada

Overall description of the project (past consulting project)

Pricing

- Estimation of the most profitable price changes for a beverage company in Spain/Portugal (bottler) with ~10 brands and ~100 SKUs
- Understanding of the switching between brands and elasticities (reactions to price changes) based on household / cluster observations

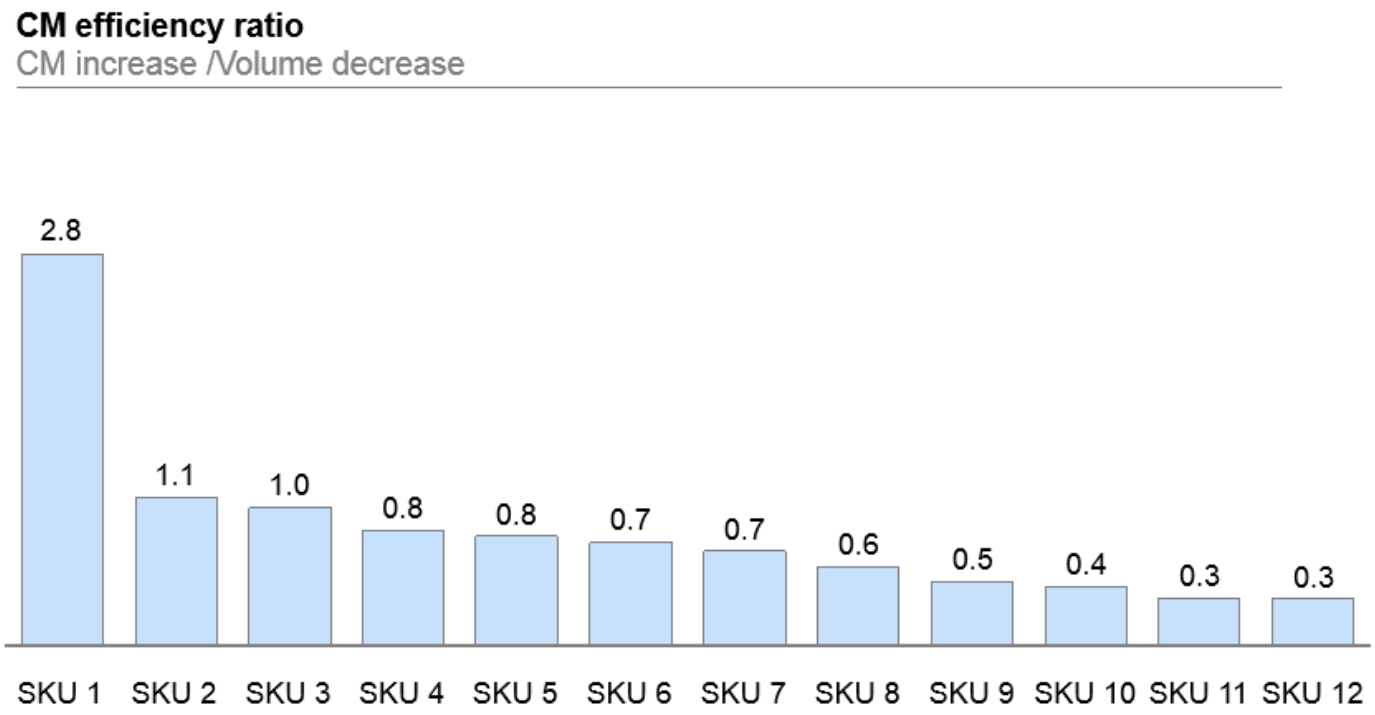
Assortment

- Assess the productivity of the SKUs and evaluate which the potential to reduce the portfolio (generating cost reductions and increased operation efficiency)

Promotions

- Understand the effectiveness and profitability of promotional actions / events carried out in Spain/Portugal (markets where promotions are very significant due to recent crisis and sudden reduction of purchasing power / perception of increased prices)

Example of some pricing analysis



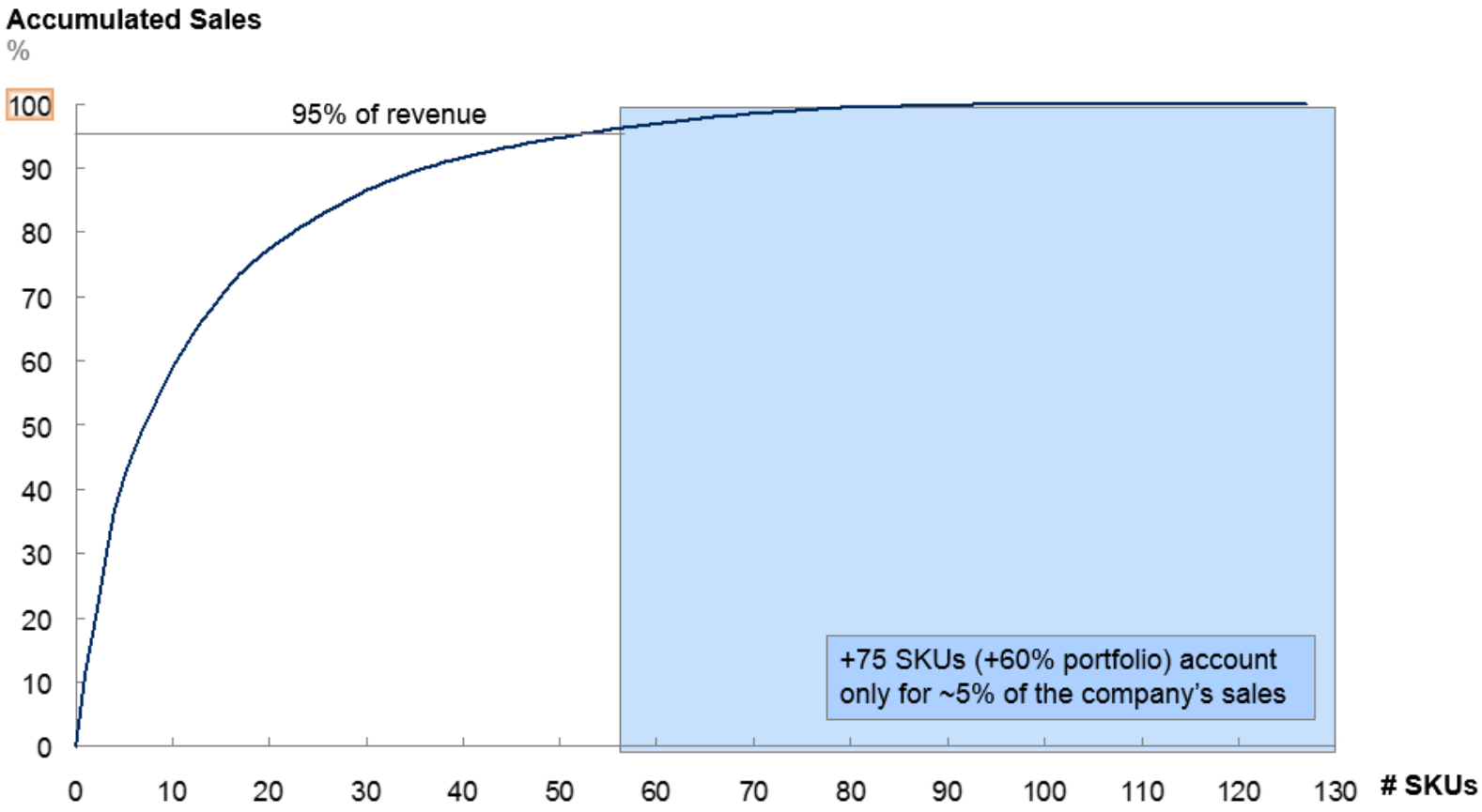
For a given price change:

Contribution margin (net revenue) ratio $= \dfrac{\Delta_{CM} \text{ (or } \Delta_{\text{net revenue}})}{\Delta_{\text{Volume}}} =$

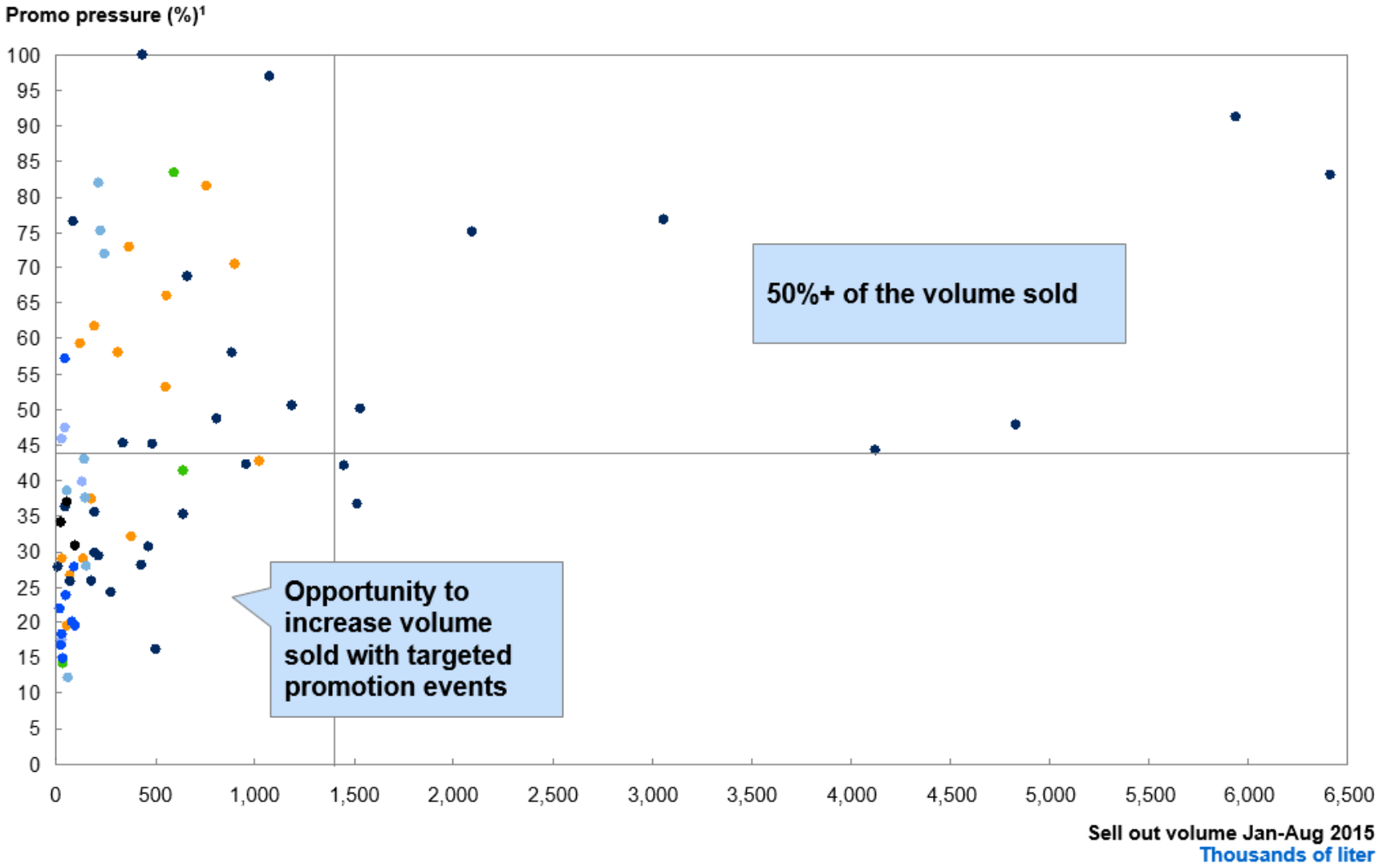
Gross profit dollars gained per liter of volume lost in the company's portfolio

- Higher ratio means that the company overall gained more Euros (of CM or Net revenue) for each liter lost

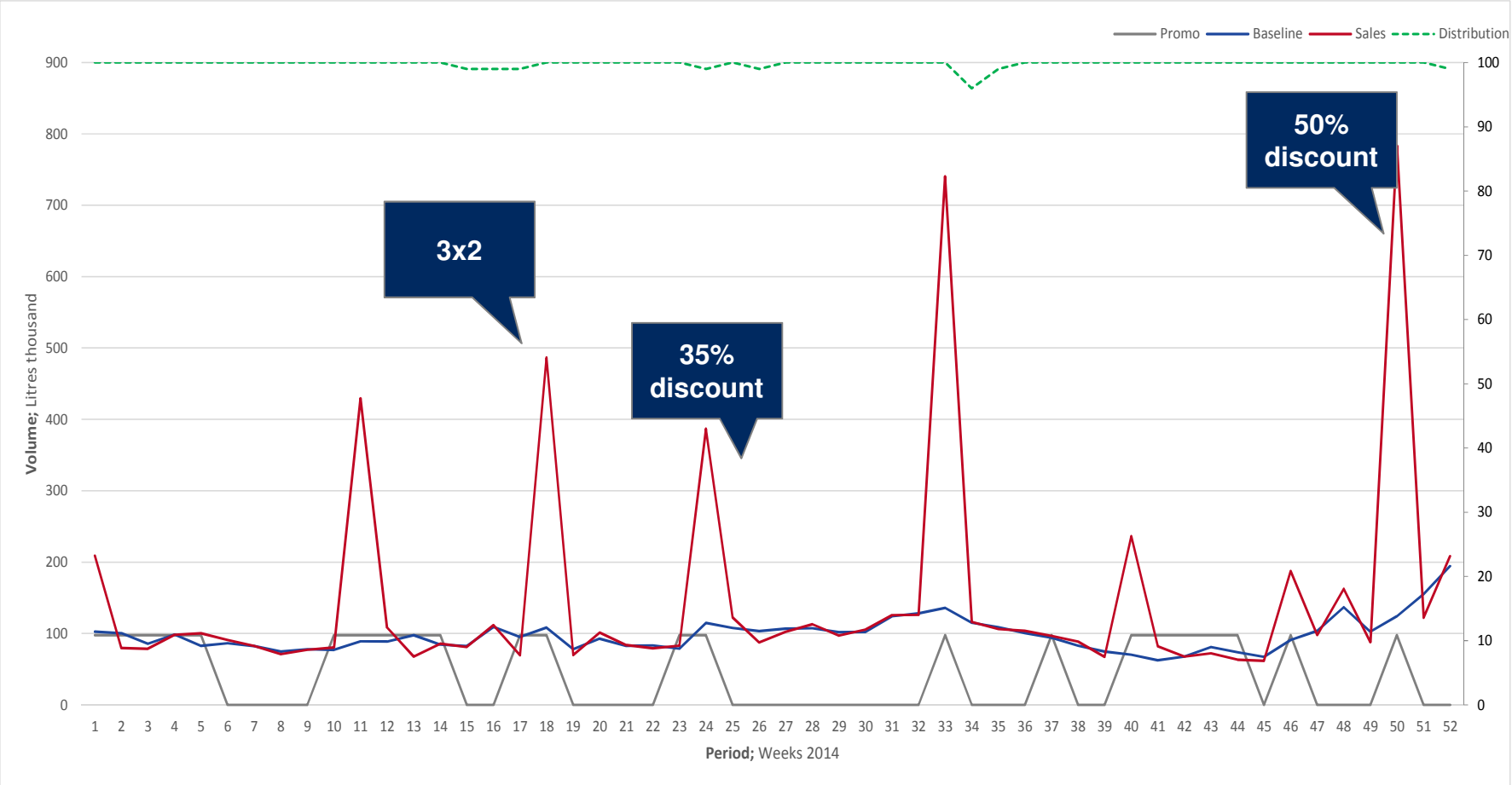
Example of some assortment analysis



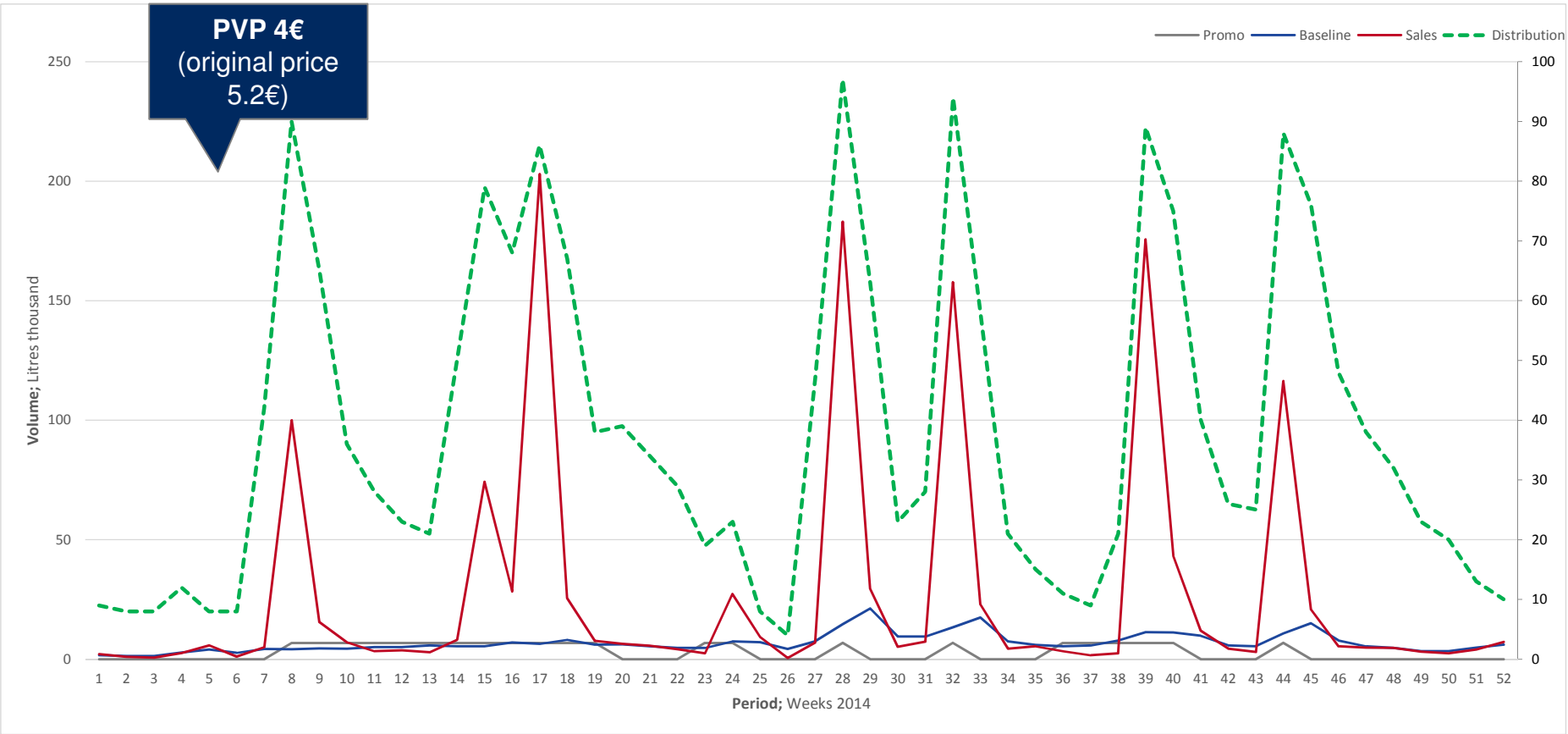
Later observed that the sales were highly influenced by promotions (specially in Portugal)



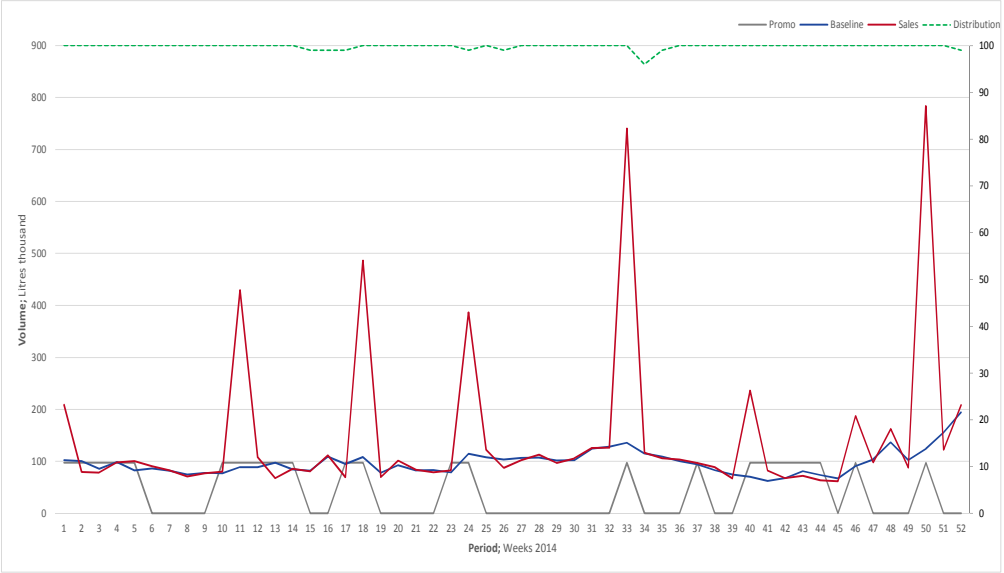
Example of weekly sales in 2014 for a given continuous SKU



Example of weekly sales in 2014 for an in-and-out SKU



Analysis performed in this class' project



- Focus of this project in Applied Regression Analysis
  1. **Regression** (how much of the variance in weekly sales is explained by the dependent variables)
  2. **Exponential smoothing** (forecast the baseline sales in promotion periods to assess the effectiveness and economics of the promotions)
  3. **Logistic regression** (simple tool to help predict the likelihood of success of a given promotion type in a given period)

1 Regression

How the data look like

- **2014 weekly sales information for 64 SKUs of 8 different brands for one account (eg., Morton Williams)**
- **3328 observations** (after cleaning outliers and SKUs with unrepresentative / missing information)

- **Volume** (liters), Nr of units, Sales (Euro)
- **Period** (week in the year of 2014)
- **Weighted distribution - a measure of the presence of the SKU in the stores** (eg., SKU present in all stores in all weeks has distribution of 100%, SKU present in 50% stores 100% of weeks has 50% distribution,, SKU present in 50% stores 50% weeks has 25% distribution)
- **Promotion pressure – measure of the volume sold under any kind of promotion** (Volume sold under promotion in a given week / total volume sold in the same week)

BRAND	SKU	WeekOf	Volume (lite	Nr of units	Sales (Euro	Weighted distribution	Promo pressure	Year	Month	Quarter
Brand F	SKU 1	06/01/2014	57	113	112	20	0	2014	1	Q1
Brand F	SKU 1	13/01/2014	37	73	70	17	14	2014	1	Q1
Brand F	SKU 1	20/01/2014	59	118	117	25	0	2014	1	Q1
Brand F	SKU 1	27/01/2014	49	98	93	21	22	2014	1	Q1
Brand F	SKU 1	03/02/2014	58	115	111	22	21	2014	2	Q1
Brand F	SKU 1	10/02/2014	49	98	95	21	12	2014	2	Q1
Brand F	SKU 1	17/02/2014	55	109	105	21	20	2014	2	Q1
Brand F	SKU 1	24/02/2014	55	110	109	23	0	2014	2	Q1
Brand F	SKU 1	03/03/2014	70	139	135	21	11	2014	3	Q1
Brand F	SKU 1	10/03/2014	70	139	134	23	4	2014	3	Q1
Brand F	SKU 1	17/03/2014	70	140	135	21	10	2014	3	Q1
Brand F	SKU 1	24/03/2014	63	125	123	22	5	2014	3	Q1
Brand F	SKU 1	31/03/2014	64	127	124	27	8	2014	3	Q1
Brand F	SKU 1	07/04/2014	93	185	183	23	0	2014	4	Q2
Brand F	SKU 1	14/04/2014	80	159	155	25	0	2014	4	Q2
Brand F	SKU 1	21/04/2014	75	149	147	24	0	2014	4	Q2
Brand F	SKU 1	28/04/2014	107	213	211	23	0	2014	4	Q2
Brand F	SKU 1	05/05/2014	77	154	152	22	3	2014	5	Q2

1 Regression

Preparation of the data for excel regression tool

REGRESSION					
Dependent	Independent				
Volume	Weighted distr	Promo pressure	Q2	Q3	Q4
57	20	0	0	0	0
37	17	14	0	0	0
59	25	0	0	0	0
49	21	22	0	0	0
58	22	21	0	0	0
49	21	12	0	0	0
55	21	20	0	0	0
55	23	0	0	0	0
70	21	11	0	0	0
70	23	4	0	0	0
70	21	10	0	0	0
63	22	5	0	0	0
64	27	8	0	0	0
93	23	0	1	0	0
80	25	0	1	0	0
75	24	0	1	0	0
107	23	0	1	0	0
77	22	3	1	0	0
89	27	21	1	0	0
56	20	4	1	0	0

- **Reminder of the goal of this regression**: How much of the variance in the sales can be explained by the weighted distribution, promo pressure and period in the year?
- If we achieve a low R-square, what could be the other factors that explain the variance in the sales?

1 Regression

Preparation of the data for excel regression tool

Excel

	Volume	Weighted distr	Promo pressure
Volume	1.00		
Weighted distr	0.27	1.00	
Promo pressure	0.31	0.34	1.00

Minitab

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-5578	795	-7,02	0,000	
Weighted distr	142,8	13,3	10,70	0,000	1,13
Promo pressure	243,2	17,1	14,21	0,000	1,13

- **There is no strong evidence of multi-collinearity**
  - **Variance inflation factors (VIF) relatively low**
  - **Correlations relatively low**

1 Regression

Results – Model with 2 variables (promotion pressure and weighted distribution)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.35336111
R Square	**12.5%**
Adjusted R Square	**12.4%**
Standard Error	27 556
Observations	3328

ANOVA

	df	*SS*	*MS*	*F*	*p-value*
Regression	2	360228366458	180114183229	237	0.0000%
Residual	3325	2524735693362	759319006		
Total	3327	2884964059820			

	Coefficients	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*
Intercept	-5577.8	794.6	-7.0	**0.0000%**	-7135.7	-4019.9
Weighted distr	142.8	13.3	10.7	**0.0000%**	116.6	169.0
Promo pressure	243.2	17.1	14.2	**0.0000%**	209.6	276.7

1 Regression

Results – Model with 5 variables (promotion pressure, weighted distribution and 3 dummy variables for the quarters)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.354414742
R Square	**12.6%**
Adjusted R Square	**12.4%**
Standard Error	27 556
Observations	3328

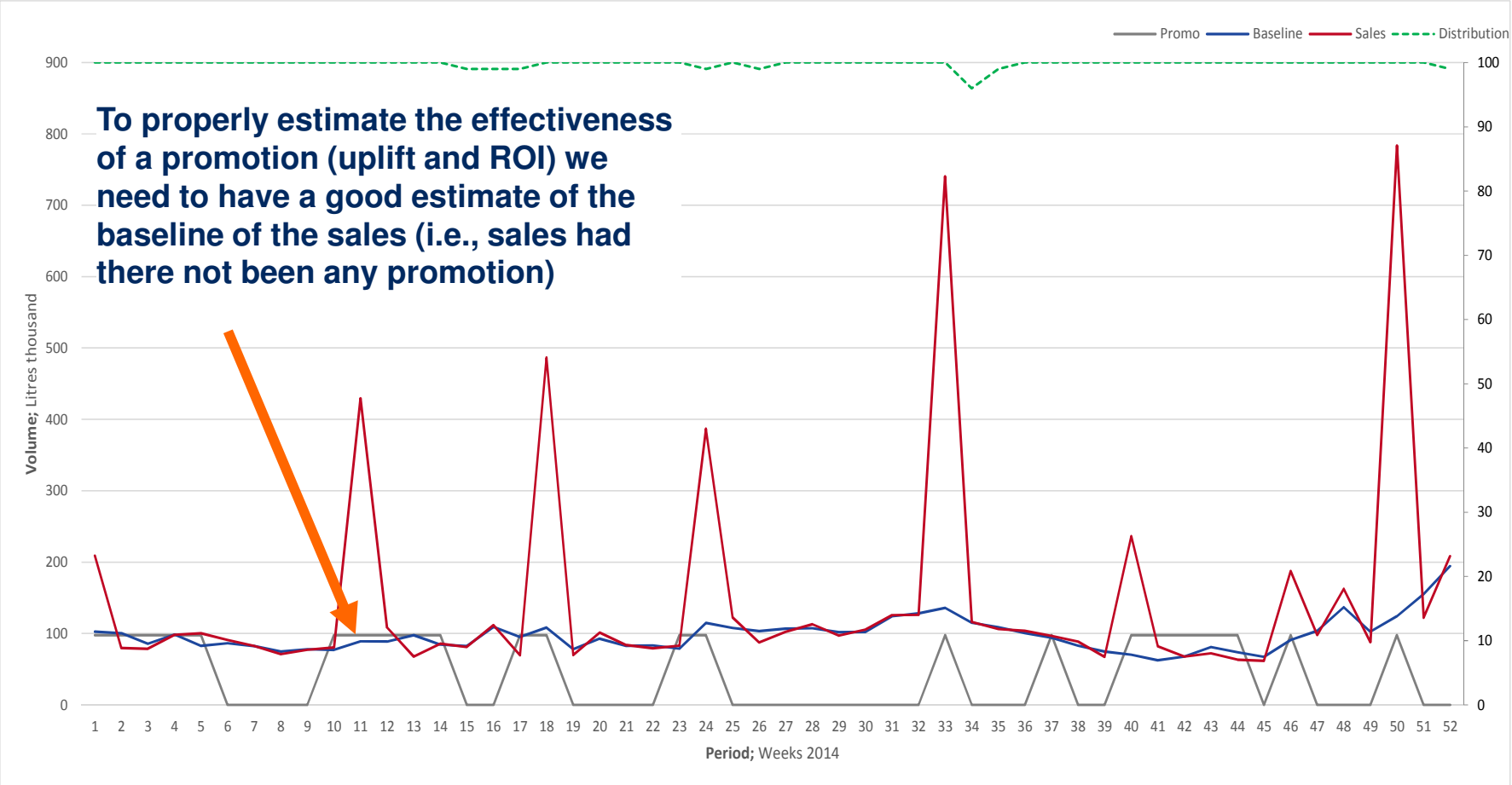
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	362379784584	72475956917	95	0.0000%
Residual	3322	2522584275235	759357097		
Total	3327	2884964059820			

	Coefficients	*Standard Error*	*t Stat*	***P-value***	*Lower 95%*	*Upper 95%*
Intercept	-6515.4	1130.0	-5.8	**0.0000%**	-8731.0	-4299.9
Weighted distr	140.5	13.4	10.5	**0.0000%**	114.2	166.9
Promo pressure	247.3	17.4	14.2	**0.0000%**	213.2	281.5
Q2	300.4	1353.4	0.2	**82.4335%**	-2353.1	2954.0
Q3	1819.7	1353.9	1.3	**17.9032%**	-834.9	4474.4
Q4	1703.1	1365.0	1.2	**21.2233%**	-973.3	4379.5

2 Exponential Smoothing

Introduction



2 Exponential Smoothing

Calculations

CALCULATION STEPS								
Sales	Week	Seasonality index	Actual sales adj by seasonality	Events	Sell-through, excl promo weeks	(used in calculations): Acct-prod key	ExpAverage through remaining points	Re-seasonalized baseline
62.0	51	1.39	369.0		369.0	SKU 64	381.2	528.9
62.0	52	1.30	756.5		756.5	SKU 64	756.5	983.9
62.0	1	0.96	1 144.5		1 144.5	SKU 64	1 136.5	1 088.4
62.0	2	0.88	71 951.0	413		SKU 64	1 127.0	994.5
62.0	3	0.71	8 455.7	470		SKU 64	661.0	472.1
62.0	4	0.81	649.7		649.7	SKU 64	651.5	529.5
62.0	5	0.68	740.8		740.8	SKU 64	739.4	503.0
62.0	6	0.71	8 744.7	482		SKU 64	1 145.3	814.4
62.0	7	0.75	1 562.8		1 562.8	SKU 64	1 551.1	1 159.2
62.0	8	0.75	932.6		932.6	SKU 64	1 003.0	748.6
62.0	9	0.79	3 962.7		3 962.7	SKU 64	3 849.5	3 054.2
62.0	10	0.79	1 200.5		1 200.5	SKU 64	1 259.0	990.1
62.0	11	0.91	1 490.7		1 490.7	SKU 64	1 485.5	1 347.3
62.0	12	0.91	2 693.3	560		SKU 64	969.5	879.7
62.0	13	1.00	447.2		447.2	SKU 64	453.4	454.2
62.0	14	0.87	737.6		737.6	SKU 64	728.3	631.9
62.0	15	0.85	548.8		548.8	SKU 64	550.7	465.6
62.0	16	1.09	464.0				505.5	505.7
62.0	17	0.99	453.3				454.8	449.4
62.0	18	1.26	513.8		513.8	SKU 64	513.4	647.5
62.0	19	0.95	548.5		548.5	SKU 64	551.8	523.1
62.0	20	0.96	757.1		757.1	SKU 64	746.8	718.1
62.0	21	0.89	440.8		440.8	SKU 64	445.2	395.9
62.0	22	0.96	356.6		356.6	SKU 64	354.1	341.5
62.0	23	0.88	135.8		135.8	SKU 64	139.3	123.0

=ExpAverageLabel(cell , range , sales values, ALPHA)

Which is the best ALPHA (eg., minimum RMSE)

2 Exponential Smoothing

Calculations

		Estimated sales, seasonality adjusted								
		Alpha								
SKU	**Actual sales**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
SKU 1	74.5	75.3	75.9	76.2	76.2	75.8	74.8	73.1	70.9	69.9
SKU 1	84.6	82.3	80.5	79.1	77.9	76.5	74.9	72.9	70.6	69.8
SKU 1	67.2	69.6	71.3	72.3	72.8	72.7	72.0	70.8	69.5	69.4
SKU 1	78.7	76.6	75.3	74.2	73.2	72.2	71.0	69.7	68.6	69.1
SKU 1	65.0	67.1	68.3	68.8	68.7	68.4	67.9	67.3	67.2	68.6
SKU 1	78.2	74.7	71.8	69.5	67.7	66.5	65.7	65.5	66.1	68.2
SKU 1	51.5	53.2	54.5	55.9	57.2	58.6	60.2	62.1	64.4	67.7
SKU 1	43.2	44.9	47.0	49.3	51.8	54.4	57.2	60.2	63.5	67.4
SKU 1	49.2	50.3	51.5	52.7	54.1	55.8	57.9	60.5	63.8	67.5
SKU 1	66.8	64.9	63.1	61.7	60.8	60.5	61.0	62.4	64.8	67.9
SKU 1	64.6	63.7	62.9	62.4	62.0	62.0	62.5	63.6	65.7	68.4
SKU 1	51.8	54.9	57.4	59.4	61.0	62.3	63.4	64.7	66.6	68.8
SKU 1	72.6	70.8	69.7	68.9	68.3	67.8	67.5	67.5	68.2	69.5
SKU 1	72.0	72.6	72.6	72.2	71.6	70.9	70.1	69.6	69.7	70.0
SKU 1	81.6	79.3	77.5	75.9	74.6	73.4	72.3	71.4	70.9	70.6
SKU 1	65.6	67.6	69.2	70.4	71.3	71.8	72.0	71.9	71.7	71.0
SKU 1	73.7	72.9	72.6	72.8	73.1	73.3	73.4	73.2	72.7	71.5
SKU 1	69.3	71.3	72.7	73.7	74.3	74.6	74.7	74.4	73.7	71.9
SKU 1	89.0	86.2	83.8	81.8	80.0	78.6	77.4	76.3	74.8	72.4
SKU 1	77.2	78.0	78.1	78.0	77.7	77.4	77.1	76.5	75.3	72.6

2 Exponential Smoothing

Calculations

Squared differences
Alpha

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.72	1.96	2.82	2.78	1.62	0.11	1.82	12.74	21.45
5.46	16.65	30.09	45.56	65.18	93.52	136.95	194.74	218.84
5.81	16.35	25.62	30.63	29.82	23.15	13.07	5.00	4.95
4.16	11.71	20.06	29.70	42.17	59.22	81.23	101.48	91.77
4.44	10.70	14.07	14.00	11.57	8.24	5.48	5.02	13.14
12.85	41.89	76.93	110.90	138.73	156.80	161.62	146.31	100.05
2.68	9.13	18.72	32.06	50.43	75.77	111.52	166.70	261.24
2.83	14.08	36.77	73.00	124.72	194.46	286.58	411.81	583.90
1.27	5.07	12.04	23.88	43.63	76.13	128.57	212.52	336.10
3.79	13.74	26.24	36.51	39.98	34.00	19.74	4.15	1.23
0.88	2.80	4.95	6.51	6.56	4.41	0.93	1.15	14.17
9.37	31.11	57.88	84.83	110.02	135.32	167.47	218.61	290.04
3.07	8.54	13.92	18.80	23.18	26.30	25.82	18.94	9.64
0.36	0.35	0.04	0.17	1.34	3.58	5.78	5.62	3.94
5.12	16.78	31.85	48.86	67.14	86.06	103.16	113.15	119.92
4.34	13.35	23.44	32.46	38.71	41.13	40.05	37.14	29.98
0.69	1.09	0.81	0.39	0.15	0.10	0.26	1.02	4.90

2 Exponential Smoothing

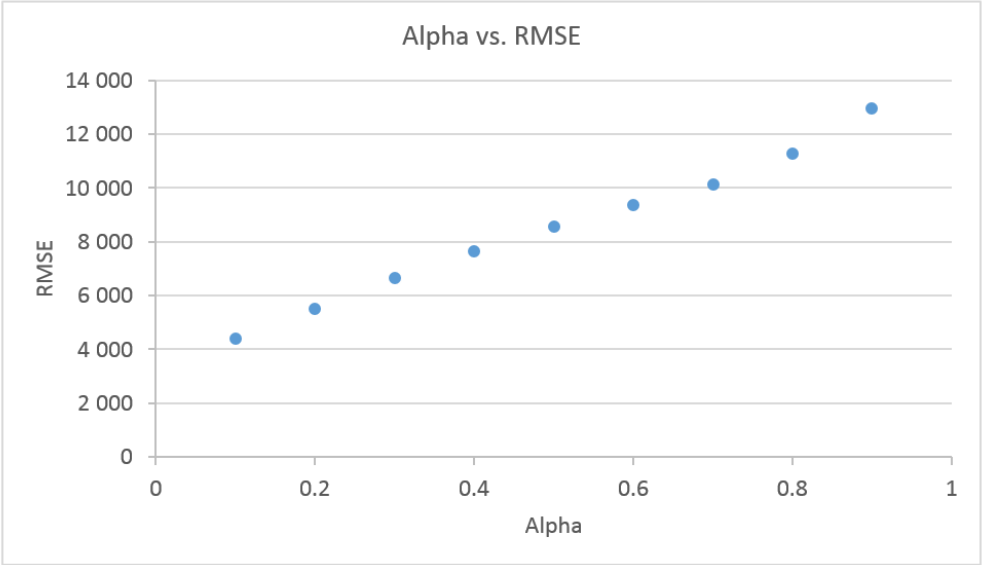
Results

$$
\begin{aligned}
L_t &= L_{t-1} + \alpha(\varepsilon_t) \\
&= \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \alpha(1-\alpha)^3 Y_{t-3} + \ldots
\end{aligned}
$$

- In this case the higher alpha smoothing seems to work better
- Higher alpha means higher weight on most recent / close observations

Alpha	RMSE
0.1	4 395
0.2	5 508
0.3	6 641
0.4	7 659
0.5	8 557
0.6	9 359
0.7	10 152
0.8	11 276
0.9	12 960

=SQRT (sum of squared diffrences / count of square differences)



3 Logistic regression

Data

Can we build a simple model that can quantitatively indicate if a promotion event (with specific characteristics) will be successful?

Brand	Event ID	SKU	Week start date	Week end date	Sales volume (L)	Baseline volume (L)	Sales units (	Baseline units (#)	Sales value (€)	Adjusted uplift (%)	Promotion type	Discount depth
Brand C	31	SKU 1	30/12/2013	02/03/2014	477 198	405 802	477 198	405 802	446 793	17%	Other	Other
Brand C	38	SKU 2	20/01/2014	02/02/2014	11 546	8 952	34 988	27 127	17 028	30%	Volume	> 25% < 40%
Brand C	39	SKU 3	20/01/2014	02/02/2014	58 478	13 170	14 767	3 326	69 320	322%	Volume	> 25% < 40%
Brand C	40	SKU 4	30/12/2013	02/02/2014	17 423	13 463	34 844	26 925	29 612	23%	Volume	> 25% < 40%
Brand C	41	SKU 1	30/12/2013	02/02/2014	299 848	227 947	299 848	227 947	272 094	31%	Volume	> 25% < 40%
Brand A	42	SKU 6	30/12/2013	02/02/2014	15 580	14 248	10 386	9 498	10 916	2%	Volume	> 25% < 40%
Brand C	45	SKU 7	20/01/2014	02/02/2014	112 096	36 824	56 048	18 412	63 732	203%	Volume	> 25% < 40%
Brand C	46	SKU 8	30/12/2013	02/02/2014	565 908	469 884	141 477	117 471	380 222	20%	Volume	> 25% < 40%
Brand D	47	SKU 9	30/12/2013	02/02/2014	6 006	4 326	18 198	13 106	10 469	51%	Volume	> 25% < 40%
Brand D	48	SKU 10	30/12/2013	02/02/2014	23 255	15 004	11 745	7 578	38 790	46%	Volume	> 25% < 40%
Brand D	49	SKU 11	30/12/2013	02/02/2014	8 600	7 014	17 197	14 026	14 702	16%	Volume	> 25% < 40%
Brand D	50	SKU 12	20/01/2014	02/02/2014	34 295	26 908	34 295	26 908	29 381	27%	Volume	> 25% < 40%
Brand D	51	SKU 13	30/12/2013	02/02/2014	4 356	4 326	1 089	1 081	4 083	-12%	Volume	> 25% < 40%
Brand D	53	SKU 14	20/01/2014	02/02/2014	35 788	22 474	17 894	11 237	22 285	54%	Volume	> 25% < 40%
Brand B	55	SKU 15	30/12/2013	02/02/2014	4 841	3 770	14 669	11 424	7 412	4%	Volume	> 25% < 40%
Brand B	56	SKU 16	30/12/2013	02/02/2014	4 601	3 690	2 323	1 863	7 239	40%	Volume	> 25% < 40%
Brand B	57	SKU 17	20/01/2014	02/02/2014	8 849	6 967	5 899	4 645	5 569	24%	Volume	> 25% < 40%
Brand B	58	SKU 18	30/12/2013	02/02/2014	5 382	5 374	897	896	3 825	-7%	Volume	> 25% < 40%
Brand A	59	SKU 19	20/01/2014	02/02/2014	18 655	14 494	12 436	9 662	11 725	22%	Volume	> 25% < 40%
Brand A	61	SKU 20	30/12/2013	02/02/2014	2 528	1 848	7 664	5 602	3 862	35%	Volume	> 25% < 40%
Brand H	70	SKU 21	30/12/2013	02/02/2014	1 049	189	3 180	572	976	386%	Volume	> 25% < 40%
Brand H	71	SKU 22	30/12/2013	02/02/2014	4 082	3 065	12 371	9 289	5 699	24%	Volume	> 25% < 40%
Brand H	72	SKU 23	30/12/2013	02/02/2014	7 369	3 656	3 722	1 847	9 176	95%	Volume	> 25% < 40%
Brand H	74	SKU 24	30/12/2013	02/02/2014	14 350	11 911	14 350	11 911	12 789	17%	Volume	> 25% < 40%
Brand H	75	SKU 25	30/12/2013	02/02/2014	512	489	128	122	455	61%	Volume	> 25% < 40%
Brand H	79	SKU 26	30/12/2013	02/02/2014	1 626	1 114	813	557	784	24%	Volume	> 25% < 40%
Brand H	82	SKU 27	20/01/2014	02/02/2014	14 494	4 153	7 247	2 077	6 643	204%	Volume	> 25% < 40%
Brand C	86	SKU 28	17/02/2014	02/03/2014	115 551	8 660	38 517	2 887	76 587	468%	Direct price	> 40% <60%
Brand D	87	SKU 29	24/02/2014	02/03/2014	55 788	9 562	37 192	6 375	36 779	30%	Direct price	< 25%
Brand E	91	SKU 30	17/02/2014	23/02/2014	8 560	916	1 070	114	4 536	293%	Direct price	> 40% <60%
Brand C	92	SKU 8	03/02/2014	02/03/2014	321 632	321 255	80 408	80 314	238 264	0%	Other	Other

3 Logistic regression

Model

Can we build a simple model that can quantitatively indicate if a promotion event (with specific characteristics) will be successful?

Brand	Event ID	SKU	Week start date	Week end date	Sales volume (L)	Baseline volume (L)	Sales units (	Baseline units (#)	Sales value (€)	Adjusted uplift (%)	Promotion type	Discount depth	Success?
Brand C	31	SKU 1	30/12/2013	02/03/2014	477 198	405 802	477 198	405 802	446 79	17%	Other	Other	0
Brand C	38	SKU 2	20/01/2014	02/02/2014	11 546	8 952	34 988	27 127	17 02	30%	Volume	> 25% < 40%	0
Brand C	39	SKU 3	20/01/2014	02/02/2014	58 478	13 170	14 767	3 326	69 32	322%	Volume	> 25% < 40%	1
Brand C	40	SKU 4	30/12/2013	02/02/2014	17 423	13 463	34 844	26 925	29 61	23%	Volume	> 25% < 40%	0
Brand C	41	SKU 1	30/12/2013	02/02/2014	299 848	227 947	299 848	227 947	272 09	31%	Volume	> 25% < 40%	0
Brand A	42	SKU 6	30/12/2013	02/02/2014	15 580	14 248	10 386	9 498	10 91	2%	Volume	> 25% < 40%	0
Brand C	45	SKU 7	20/01/2014	02/02/2014	112 096	36 824	56 048	18 412	63 73	203%	Volume	> 25% < 40%	1
Brand C	46	SKU 8	30/12/2013	02/02/2014	565 908	469 884	141 477	117 471	380 22	20%	Volume	> 25% < 40%	0
Brand D	47	SKU 9	30/12/2013	02/02/2014	6 006	4 326	18 198	13 106	10 46	51%	Volume	> 25% < 40%	0
Brand D	48	SKU 10	30/12/2013	02/02/2014	23 255	15 004	11 745	7 578	38 79	46%	Volume	> 25% < 40%	0
Brand D	49	SKU 11	30/12/2013	02/02/2014	8 600	7 014	17 197	14 026	14 70	16%	Volume	> 25% < 40%	0
Brand D	50	SKU 12	20/01/2014	02/02/2014	34 295	26 908	34 295	26 908	29 38	27%	Volume	> 25% < 40%	0
Brand D	51	SKU 13	30/12/2013	02/02/2014	4 356	4 326	1 089	1 081	4 08	-12%	Volume	> 25% < 40%	0
Brand D	53	SKU 14	20/01/2014	02/02/2014	35 788	22 474	17 894	11 237	22 28	54%	Volume	> 25% < 40%	0
Brand B							669	11 424	7 41	4%	Volume	> 25% < 40%	0
Brand B							323	1 863	7 23	40%	Volume	> 25% < 40%	0
Brand B							899	4 645	5 56	24%	Volume	> 25% < 40%	0
Brand B							897	896	3 82	-7%	Volume	> 25% < 40%	0
Brand A							436	9 662	11 72	22%	Volume	> 25% < 40%	0
Brand A							664	5 602	3 86	35%	Volume	> 25% < 40%	0
Brand H							180	572	97	386%	Volume	> 25% < 40%	1
Brand H							371	9 289	5 69	24%	Volume	> 25% < 40%	0
Brand H							722	1 847	9 17	95%	Volume	> 25% < 40%	0
Brand H							350	11 911	12 78	17%	Volume	> 25% < 40%	0
Brand H							128	122	45	61%	Volume	> 25% < 40%	0
Brand H							813	557	78	24%	Volume	> 25% < 40%	0
Brand H	82	SKU 27	20/01/2014	02/02/2014	14 494	4 153	7 247	2 077	6 64	204%	Volume	> 25% < 40%	1
Brand C	86	SKU 28	17/02/2014	02/03/2014	115 551	8 660	38 517	2 887	76 58	468%	Direct price	> 40% <60%	1
Brand D	87	SKU 29	24/02/2014	02/03/2014	55 788	9 562	37 192	6 375	36 77	30%	Direct price	< 25%	0
Brand E	91	SKU 30	17/02/2014	23/02/2014	8 560	916	1 070	114	4 53	293%	Direct price	> 40% <60%	1
Brand C	92	SKU 8	03/02/2014	02/03/2014	321 632	321 255	80 408	80 314	238 26	0%	Other	Other	0

- 1) define success factor: if uplift > 100%, successful
- Adjust data to have binary outcome

3 Logistic regression

Model

Can we build a simple model that can quantitatively indicate if a promotion event (with specific characteristics) will be successful?

Brand	Event ID	SKU	Week start date	Week end date	Sales volume (L)	Baseline volume (L)	Sales units (	Baseline units (#)	Sales value (€)	Adjusted uplift (%)	Promotion type	Discount depth	Success?
Brand C	31	SKU 1	20/12/2013	02/03/2014	477 198	405 802	477 198	405 802	446 79	17%	Other	Other	0
Brand C	38	SKU 2	20/01/2014	02/02/2014	11 546	8 952	34 988	27 127	17 02	30%	Volume	> 25% < 40%	0
Brand C	39	SKU 3	20/01/2014	02/02/2014	58 478	13 170	14 767	3 326	69 32	322%	Volume	> 25% < 40%	1
Brand C	40	SKU 4	30/12/2013	02/02/2014	17 423	13 463	34 844	26 925	29 61	23%	Volume	> 25% < 40%	0
Brand C	41	SKU 1	30/12/2013	02/02/2014	299 848	227 947	299 848	227 947	272 09	31%	Volume	> 25% < 40%	0
Brand A	42	SKU 6	30/12/2013	02/02/2014	15 580	14 248	10 386	9 498	10 91	2%	Volume	> 25% < 40%	0
Brand C	45	SKU 7	20/01/2014	02/02/2014	112 096	36 824	56 048	18 412	63 73	203%	Volume	> 25% < 40%	1
Brand C	46	SKU 8	30/12/2013	02/02/2014	565 908	469 884	141 477	117 471	380 22	20%	Volume	> 25% < 40%	0
Brand D	47	SKU 9	30/12/2013	02/02/2014	6 006	4 326	18 198	13 106	10 46	51%	Volume	> 25% < 40%	0
Brand D	48	SKU 10	30/12/2013	02/02/2014	23 255	15 004	11 745	7 578	38 79	46%	Volume	> 25% < 40%	0
Brand D	49	SKU 11	30/12/2013	02/02/2014	8 600	7 014	17 197	14 026	14 70	16%	Volume	> 25% < 40%	0
Brand D	50	SKU 12	20/01/2014	02/02/2014	34 295	26 908	34 295	26 908	29 38	27%	Volume	> 25% < 40%	0
Brand D	51	SKU 13	30/12/2013	02/02/2014	4 356	4 326	1 089	1 081	4 08	-12%	Volume	> 25% < 40%	0
Brand D	53	SKU 14	20/01/2014	02/02/2014	35 788	22 474	17 894	11 237	22 28	54%	Volume	> 25% < 40%	0
Brand B							669	11 424	7 41	4%	Volume	> 25% < 40%	0
Brand B							323	1 863	7 23	40%	Volume	> 25% < 40%	0
Brand B							899	4 645	5 56	24%	Volume	> 25% < 40%	0
Brand B							897	896	3 82	-7%	Volume	> 25% < 40%	0
Brand A							436	9 662	11 72	22%	Volume	> 25% < 40%	0
Brand A							664	5 602	3 86	35%	Volume	> 25% < 40%	0
Brand H							180	572	97	386%	Volume	> 25% < 40%	1
Brand H							371	9 289	5 69	24%	Volume	> 25% < 40%	0
Brand H							722	1 847	9 17	95%	Volume	> 25% < 40%	0
Brand H							350	11 911	12 78	17%	Volume	> 25% < 40%	0
Brand H							128	122	45	61%	Volume	> 25% < 40%	0
Brand H							813	557	78	24%	Volume	> 25% < 40%	0
Brand H	82	SKU 27	20/01/2014	02/02/2014	14 494	4 153	7 247	2 077	6 64	204%	Volume	> 25% < 40%	1
Brand C	86	SKU 28	17/02/2014	02/03/2014	115 551	8 660	38 517	2 887	76 58	468%	Direct price	> 40% <60%	1
Brand D	87	SKU 29	24/02/2014	02/03/2014	55 788	9 562	37 192	6 375	36 77	30%	Direct price	< 25%	0
Brand E	91	SKU 30	17/02/2014	23/02/2014	8 560	916	1 070	114	4 53	293%	Direct price	> 40% <60%	1
Brand C	92	SKU 8	03/02/2014	02/03/2014	321 632	321 255	80 408	80 314	238 26	0%	Other	Other	0

- 1) define success factor: if uplift > 100%, successful
- Adjust data to have binary outcome

3 Logistic regression

Model and results

Logistic regression model

=+LogisticRegTrain(binary outcome , predictors

Success? (7,8)	Constant	Volume	Direct price	Cross	Q1	Q2	Q3
Coefficients	-2.11	0.03	1.67	-10.70	0.41	0.93	0.65
Std error	0.78	0.52	0.47	382.25	0.88	0.76	0.84
p-value	1%	96%	0%	98%	65%	22%	43%
Log-likelihood	-90.0						
Number valid obs	167						
Total obs	167						

Possible future events

	Volume	Direct price	Cross	Q1	Q2	Q3
Promo 1	1	0	0	1	0	0
Promo 2	1	0	0	0	1	0
Promo 3	1	0	0	0	0	1
Promo 4	0	1	0	1	0	0

Likelihood of success (according to our pre-established threshold for the uplift)

=+LogisticRegPredict (model coefficients , event information)

	Prediction
Promo 1	0.16
Promo 2	0.24
Promo 3	0.19
Promo 4	0.49

Thank you for your attention