MARKETING ANALYTICS PROJECT

DATA SET

MarketID	MarketSize	LocationID	AgeOfStore	Promotion	week	SalesInThousands
1	Medium	1	4	3	1	33.73
1	Medium	1	4	3	2	35.67
1	Medium	1	4	3	3	29.03
1	Medium	1	4	3	4	39.25
1	Medium	2	5	2	1	27.81
1	Medium	2	5	2	2	34.67
1	Medium	2	5	2	3	27.98
1	Medium	2	5	2	4	27.72
1	Medium	3	12	1	1	44.54
1	Medium	3	12	1	2	37.94
1	Medium	3	12	1	3	45.49
1	Medium	3	12	1	4	34.75
1	Medium	4	1	2	1	39.28
1	Medium	4	1	2	2	39.8
1	Medium	4	1	2	3	24.77
1	Medium	4	1	2	4	30.98
1	Medium	5	10	2	1	30.37
1	Medium	5	10	2	2	24.82
1	Medium	5	10	2	3	37.47
1	Medium	5	10	2	4	23.35
1	Medium	6	10	3	1	32.9
1	Medium	6	10	3	2	22.18
1	Medium	6	10	3	3	42.98
1	Medium	6	10	3	4	26.68
1	Medium	7	15	1	1	42.92
1	Medium	7	15	1	2	42.16
1	Medium	7	15	1	3	51.72

Columns	Description
MarketID	unique identifier for market
MarketSize	size of market area by sales (small, medium, or large)
LocationID	unique identifier for store location
AgeOfStore	age of store in years
Promotion	one of three promotions that were tested (1,2, or 3)
week	one of four weeks when the promotions were run (1,2,3, or 4)
SalesInThousands	sales amount for a specific *LocationID*, *Promotion*, and *week*

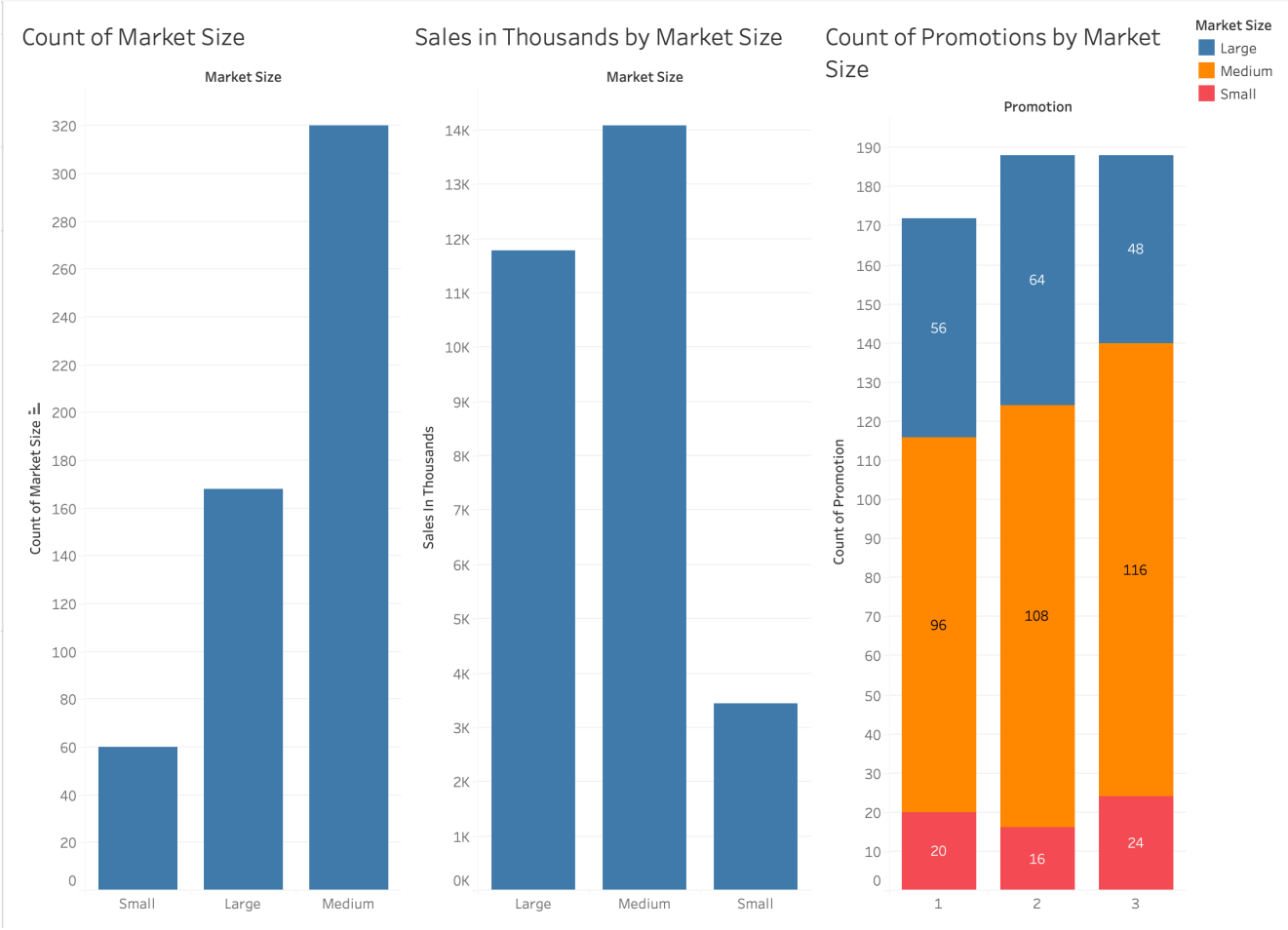
550 observations

Log-Linear and Linear regression

- The main goal was to predict which promotion had the greatest effect on sales
- Is there a relationship between sales, promotions, and different number of weeks?

Data Exploration

- The data has more Medium markets, followed by Large, and Small Markets
- Medium markets have the most Sales followed by Large and Small markets
- Promotions were approximately spread evenly across market sizes



Methodology

- Created dummy variables for each promotion and week
- Started by creating linear regression models for all markets and after discovering a low adjusted R^2 value I decided to run linear regressions for each market individually reaching the same result
- Decided to run log-linear models for the same scenarios as above looking for an improvement
- Further improved the model by building a stacked log-linear model including interactions
- **Important**: Sales was my dependent variable. Promotion 1 and Week 1 was my baseline

Linear Regression models

ALL MARKET $R^2 = 0.066$

Model parameters (SalesInThousands):

Source	Value	Standard error	t	Pr > |t|	Lower bound	Upper bound
Intercept	58.423	1.720	33.962	**<0.0001**	55.044	61.803
Promotion 2	-10.770	1.708	-6.304	**<0.0001**	-14.125	-7.414
Promotion 3	-2.735	1.708	-1.601	0.110	-6.090	0.621
Week 2	-0.404	1.956	-0.207	0.836	-4.247	3.439
Week 3	-0.316	1.956	-0.162	0.872	-4.159	3.527
Week 4	-0.578	1.956	-0.295	0.768	-4.420	3.265

SMALL MARKET $R^2 = 0.379$

Model parameters (SalesInThousands):

Source	Value	Standard error	t	Pr > |t|	Lower bound	Upper bound
Intercept	58.999	1.653	35.690	**<0.0001**	55.684	62.313
Promotion 2	-9.352	1.753	-5.334	**<0.0001**	-12.867	-5.837
Promotion 3	-0.648	1.583	-0.410	0.684	-3.821	2.525
Week 2	2.430	1.909	1.273	0.208	-1.397	6.257
Week 3	-0.897	1.909	-0.470	0.640	-4.724	2.930
Week 4	3.123	1.909	1.636	0.108	-0.704	6.950

MEDIUM MARKET $R^2 = 0.128$

Model parameters (SalesInThousands):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	47.251	1.179	40.086	**<0.0001**	44.932	49.571
Promotion	-8.218	1.192	-6.895	**<0.0001**	-10.564	-5.873
Promotion	-2.706	1.153	-2.347	**0.02**	-4.975	-0.437
Week 2	1.074	1.315	0.817	0.415	-1.513	3.661
Week 3	0.967	1.315	0.736	0.463	-1.62	3.555
Week 4	-0.243	1.315	-0.185	0.853	-2.83	2.344

LARGE MARKET $R^2 = 0.184$

Model parameters (SalesInThousands):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	75.913	2.91	26.085	**<0.0001**	70.166	81.659
Promotion 2	-14.914	2.818	-5.293	**<0.0001**	-20.478	-9.349
Promotion 3	1.968	3.029	0.65	0.517	-4.014	7.95
Week 2	-2.137	3.36	-0.636	0.526	-8.773	4.499
Week 3	-0.146	3.36	-0.043	0.965	-6.782	6.49
Week 4	-0.424	3.36	-0.126	0.9	-7.06	6.212

ALL MARKET LOG LINEAR

- Better adjusted R^2 compared to linear regression model
- In all markets promotion 1 and Week 1 seems to be the best combination to increase sales
- There seems to be an indifference between Promotion 1 and 3
- Significant coefficients: Intercept, Promotion 2. Promotion 3 borderline depending on significance level
- Insignificant coefficients : Week 2,3, and 4

Goodness of fit statistics (Log (SalesInThousands)):

Observations	548
Sum of weights	548
DF	542
R^2	0.089
Adjusted R^2	0.081
MSE	0.087
RMSE	0.295
MAPE	5.744
DW	0.364
Cp	6
AIC	-1332.2
SBC	-1306.3
PC	0.931

Model parameters (Log (SalesInThousands)):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	4.033	0.031	128.688	<0.0001	3.971	4.094
Promotion 2	-0.216	0.031	-6.929	<0.0001	-0.277	-0.154
Promotion 3	-0.054	0.031	-1.739	0.083	-0.115	0.007
Week 2	-0.008	0.036	-0.219	0.827	-0.078	0.062
Week 3	-0.009	0.036	-0.264	0.792	-0.079	0.061
Week 4	-0.017	0.036	-0.464	0.643	-0.087	0.053

SMALL MARKET LOG LINEAR

- Better adjusted R^2 compared to linear regression model
- In small markets promotion 1 and Week 1 seems to be the best combination to increase sales
- There seems to be an indifference between Promotion 1 and 3, but 3 is insignificant
- Significant coefficients: Intercept, Promotion 2
- Insignificant coefficients : Promotion 3 and Week 2,3, and 4

Goodness of fit statistics (Log (SalesInThousands)):

Observations	60
Sum of weights	60
DF	54
R^2	0.442
Adjusted R^2	0.39
MSE	0.009
RMSE	0.095
MAPE	1.713
DW	1.825
Cp	6
AIC	-276.67
SBC	-264.11
PC	0.682

Model parameters (Log (SalesInThousands)):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	4.075	0.03	135.502	<0.0001	4.014	4.135
Promotion 2	-0.172	0.032	-5.399	<0.0001	-0.236	-0.108
Promotion 3	-0.011	0.029	-0.388	0.7	-0.069	0.047
Week 2	0.042	0.035	1.202	0.234	-0.028	0.111
Week 3	-0.023	0.035	-0.662	0.511	-0.093	0.047
Week 4	0.057	0.035	1.641	0.107	-0.013	0.127

Medium Market Log Linear

- Similar adjusted R^2 to linear regression model
- In medium markets promotion 1 and Week 1 seems to be the best combination to increase sales
- There seems to be an indifference between Promotion 1 and 3
- Significant coefficients: Intercept, Promotion 2, and Promotion 3
- Insignificant coefficients : Week 2,3, and 4

Goodness of fit statistics (Log (SalesInThousands)):

Observations	320
Sum of weights	320
DF	314
R^2	0.141
Adjusted R^2	0.127
MSE	0.043
RMSE	0.207
MAPE	4.497
DW	0.976
Cp	6
AIC	-1001.713
SBC	-979.103
PC	0.892

Model parameters (Log (SalesInThousands)):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	3.844	0.029	132.241	<0.0001	3.787	3.901
Promotion 2	-0.202	0.029	-6.857	<0.0001	-0.259	-0.144
Promotion 3	-0.062	0.028	-2.194	0.029	-0.118	-0.006
Week 2	0.019	0.032	0.572	0.568	-0.045	0.082
Week 3	0.018	0.032	0.552	0.581	-0.046	0.082
Week 4	-0.012	0.032	-0.377	0.707	-0.076	0.052

LARGE MARKET LOG LINEAR

- Better adjusted R^2 compared to linear regression model
- In large markets promotion 1 and Week 1 seems to be the best combination to increase sales
- There seems to be an indifference between Promotion 1 and 3, but 3 is insignificant
- Significant coefficients: Intercept, Promotion 2
- Insignificant coefficients : Promotion 3, and Week 2,3, and 4

Goodness of fit statistics (Log (SalesInThousands)):

Observations	168
Sum of weights	168
DF	162
R^2	0.221
Adjusted R^2	0.197
MSE	0.052
RMSE	0.229
MAPE	4.815
DW	0.291
Cp	6
AIC	-489.53
SBC	-470.79
PC	0.837

Model parameters (Log (SalesInThousands)):

Source	Value	Standard error	t	Pr > |t|	Lower bound (95%)	Upper bound (95%)
Intercept	4.305	0.043	99.513	**<0.0001**	4.219	4.39
Promotion 2	-0.232	0.042	-5.53	**<0.0001**	-0.314	-0.149
Promotion 3	0.028	0.045	0.617	0.538	-0.061	0.117
Week 2	-0.026	0.05	-0.521	0.603	-0.125	0.073
Week 3	0.003	0.05	0.061	0.952	-0.096	0.102
Week 4	0	0.05	0.004	0.997	-0.098	0.099

Results from log linear models

- I came to the realization that the week in which a promotion ran had no relationship to sales, so I took it out for the next model
- The only variables affecting sales were promotions and the size of the market, so I created interactions
- I had to create dummy variables for the market sizes
- **Important**: Sales was my dependent variable. Large Market and Promotion 1 was my baseline

Large Market Log Linear

- Best adjusted R^2
- Large market and promotion 1 seem to increase sales the most
- There seems to be an indifference between Promotion 1 and 3, but 3 is insignificant
- There doesn't seem to be a relationship between markets and promotions
- Significant coefficients: Intercept, Promotion 2
- Insignificant coefficients : Promotion 3, and Week 2,3, and 4

Observations	548
Sum of weig	548
DF	539
R^2	0.565
Adjusted R^2	0.558
MSE	0.042
RMSE	0.204
MAPE	4.281
DW	0.768
Cp	9.000
AIC	-1730.738
SBC	-1691.982
PC	0.450

Model parameters (Log (SalesInThousands)):

Source	Value	Standard error	t	Pr > |t|	Lower bound	Upper bound
Intercept	4.299	0.027	157.333	**<0.0001**	4.245	4.353
Small	-0.206	0.053	-3.858	**0.000**	-0.310	-0.101
Medium	-0.449	0.034	-13.066	**<0.0001**	-0.517	-0.382
Promotion 2	-0.232	0.037	-6.190	**<0.0001**	-0.305	-0.158
Promotion 3	0.028	0.040	0.691	0.490	-0.051	0.107
Small P2	0.059	0.078	0.760	0.447	-0.094	0.213
Small P3	-0.039	0.074	-0.527	0.598	-0.184	0.106
Medium P2	0.021	0.047	0.437	0.662	-0.072	0.113
Medium P3	-0.078	0.049	-1.588	0.113	-0.175	0.018

Summary: Business insights

- We found that promotion 1 and large market were the most effective at increasing sales
  - However, this could be biased considering market size is broken down by sales and not income or population size
- The week a promotion ran did not have an effect of the total sales made by that specific promotion
- Based on the interactions between market size and promotions we concluded that promotion 1 seems to be the most effective at increasing sales
  - However, market has no significant effect
- **Recommendation:**
  - Run and market promotion 1 the same way in all markets because they all respond equally
  - No need to spend money making different marketing campaigns for each market (less cost)

Limitations

- Next time the organization conducts an A/B Test, they should randomize all variables equally (same number of markets, same number of promotions, etc.) to get better results
- Results were skewed due to the low sample size of certain markets & promotions
- Market size should be based on a better metric such as income level or population because basing it on sales skews the results since our dependent variable is sales