Project: Analyzing a Market Test

Step 1: Plan Your Analysis

1. What is the performance metric you'll use to evaluate the results of your test?

   The performance metric used to evaluate the results would be the Gross Margin.

2. What is the test period?

   The test period is from 29th April 2016 to 21st July 2016, which would be 12 weeks.

3. At what level (day, week, month, etc.) should the data be aggregated?

   Since the experiment (test period) is measured in weeks, the data should be aggregated at a weekly level.

Step 2: Clean Up Your Data

Data preparation steps that were taken are:

1. Change the data types of the variables that need to be numerical.
2. Then, join Round Rosters transaction data to Round Rosters stores data so that the region details are captured.
3. Next, data needs to be filtered for 76 weeks, i.e. 52+12 weeks of data for the analysis and 12 weeks of data captured during the experiment. So, the data needs to be filtered for the duration 6th February 2015 to 21st July 2016.
4. Finally, the daily data is aggregated to weekly showing the total number of invoices per week per store.

Record #	StoreID	Week	Week_Start	Week_End	Count	Sum_Gross Margin	Sum_Sales
1	10018	1	2015-02-06	2015-02-12	308	2212.7105	4741.48
2	10018	2	2015-02-13	2015-02-19	288	2164.007	4571.25
3	10018	3	2015-02-20	2015-02-26	204	1560.929	3348.25
4	10018	4	2015-02-27	2015-03-05	320	2342.984	5114.96

Step 3: Match Treatment and Control Units

Apart from trend and seasonality...

1. What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.

   Average Monthly Sales & Square Feet from the RoundRoastersStore files can be considered as potential control variables, as they are numeric and it would be logical to compare stores of equal size and similar volume.

2. What is the correlation between your each potential control variable and your performance metric?

   Association Analysis is done to find the correlation between the potential control

variables (Average Monthly Sales & Square Feet) and the Performance Metric (Gross_Margin).

Pearson Correlation Analysis

Focused Analysis on Field Sum_Gross.Margin

	Association Measure	p-value	
AvgMonthSales	0.790358	0.000000	***
Sq_Ft	-0.019345	0.051796	.

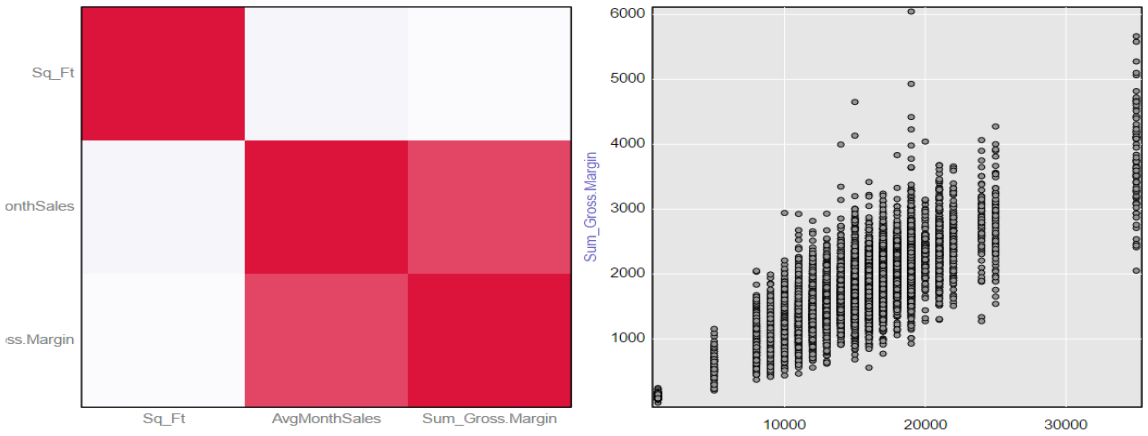
Full Correlation Matrix

	Sum_Gross.Margin	AvgMonthSales	Sq_Ft
Sum_Gross.Margin	1.000000	0.790358	-0.019345
AvgMonthSales	0.790358	1.000000	-0.046967
Sq_Ft	-0.019345	-0.046967	1.000000

Matrix of Corresponding p-values

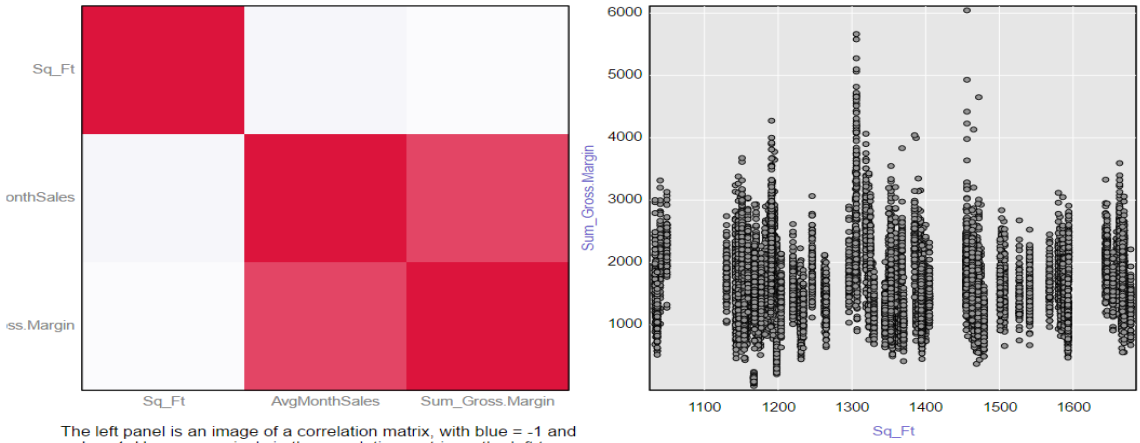
	Sum_Gross.Margin	AvgMonthSales	Sq_Ft
Sum_Gross.Margin		0.0000e+00	5.1796e-02
AvgMonthSales	0.0000e+00		2.3119e-06
Sq_Ft	5.1796e-02	2.3119e-06	

Scatterplot of Average Monthly Sales Vs. Gross Margin



The left panel is an image of a correlation matrix, with blue = -1 and red = +1. Hover over pixels in the correlation matrix on the left to see the values; click to see the corresponding scatterplot on the right. The variables have been clustered based on degree of correlation, so that highly correlated variables appear adjacent to each other.

Scatterplot of Square Feet Vs. Gross Margin



The left panel is an image of a correlation matrix, with blue = -1 and red = +1. Hover over pixels in the correlation matrix on the left to see the values; click to see the corresponding scatterplot on the right. The variables have been clustered based on degree of correlation, so that highly correlated variables appear adjacent to each other.

Comparing the correlation matrices, it can be observed that only the AvgMonthSales variable has a correlation to our performance measure whereas Sq_Ft does not correlate at all. More specifically, we see a 0.79 correlation and a 0.0000 p-value together with a *** statistical significance for the AvgMonthSales variable.

3. What control variables will you use to match treatment and control stores?

   Sq_Ft could have been used as a control variable as it is numeric. However Sq_Ft doesn't have any correlation with the target variable which is Gross_Margin in this case. Also, if two stores have similar Average month sales, they are also expected to have related gross margins. Hence, AvgMonthSales should be used as the control variable for matching the stores.

   Listed below are the final control variables that will be used to match treatment and control stores.

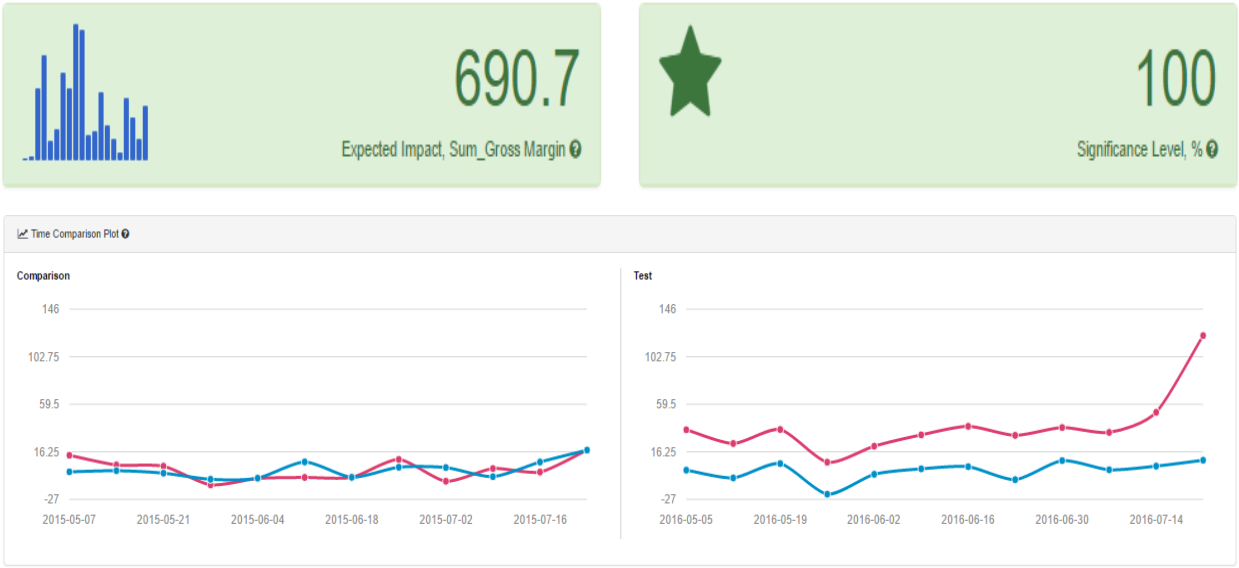
   - Trend,
   - Seasonality, and
   - AvgMonthSales.

4. Please fill out the table below with your treatment and control stores pairs:

Record #	Treatments	Controls	Distance
1	1664	7484	0.140608
2	1664	7162	0.165077
3	1675	8162	0.372912
4	1675	2114	0.3929
5	1696	7284	0.034352
6	1696	1863	0.279586
7	1700	7384	0.440351
8	1700	1662	0.756937
9	1712	7434	0.066538
10	1712	6992	0.09772
11	2288	2568	0.346735
12	2288	9081	0.392618
13	2293	9918	0.240861
14	2293	11768	0.266034
15	2301	3185	0.180416
16	2301	12536	0.310579
17	2322	9388	0.088715
18	2322	9238	0.116578
19	2341	9488	0.056892
20	2341	11368	0.096716

Step 4: Analysis and Writeup

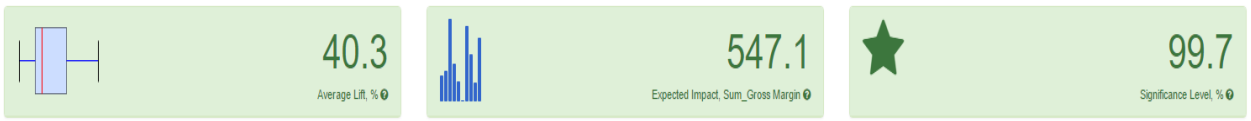
1. What is your recommendation - Should the company roll out the updated menu to all stores?



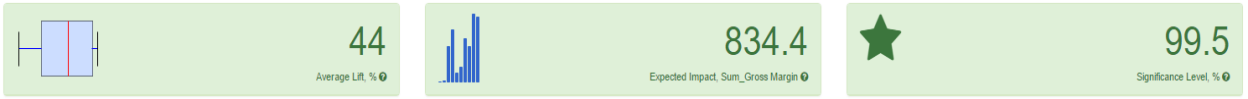
Based on the significance level in the above graph and the Time Comparison plot, my recommendation would be to roll out the updated menu to all stores.

2. What is the lift from the new menu for West and Central regions (include statistical significance)?

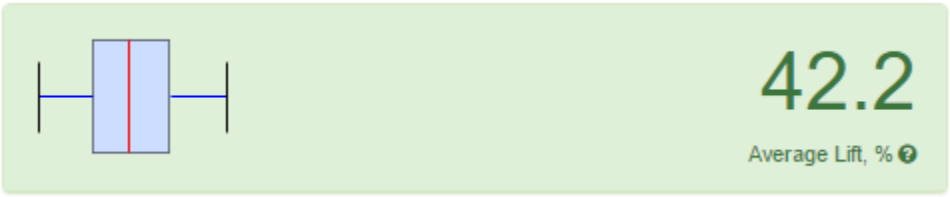
<u>Lift for Western Region:</u>



<u>Lift for Central Region:</u>

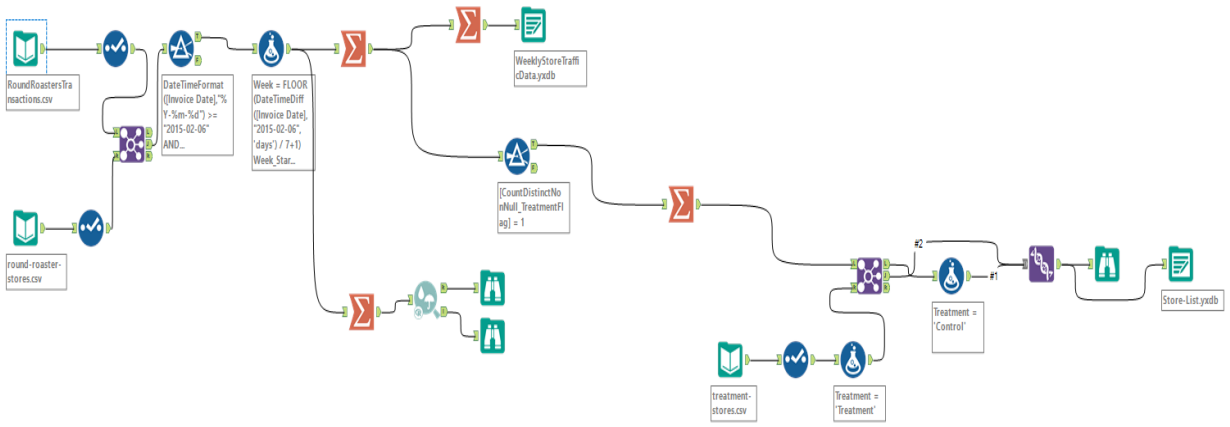


3. What is the lift from the new menu overall?

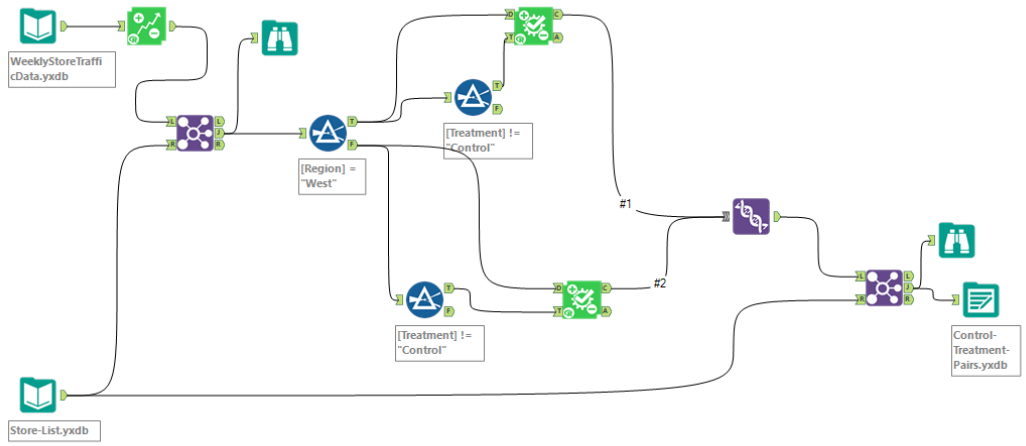


Appendix:

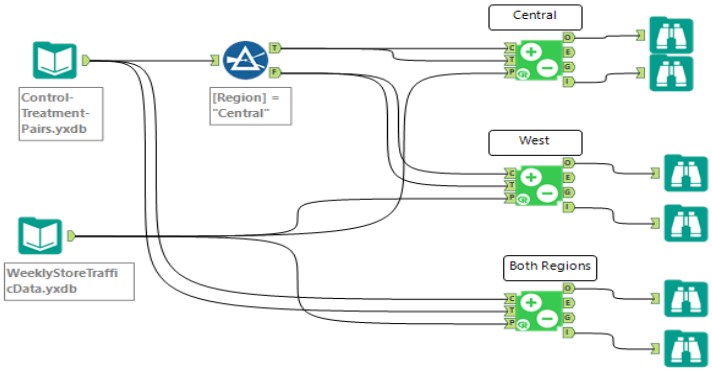
Workflow -1:



Workflow -2:



Workflow -3:


Control-Treatment-Pairs.yxdb
[Region] = "Central"
Central
West
Both Regions
WeeklyStoreTraffiData.yxdb