Project: Analyzing a Market Test

We are approaching this project as a Business Analyst for Round Roasters, a US based coffee restaurant. The executive team conducted a market test with a new menu and needs to figure whether the new menu can drive enough sales to offset the cost of marketing the new menu. Our task is to analyze the A/B test and recommend whether Round Roasters should launch this new menu or not.

Step 1: Analysis

We will break up our analysis into three parts:

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

   We are trying to analyze the results and determine whether the menu changes should be applied to all stores. The predicted impact to profitability should be enough to justify the increased marketing budget: at least 18% increase in profit growth compared to the comparative period while compared to the control stores. So in this case we will use the profit or sum of **gross_margin** as the performance metric.

2. **What is the test period?**

   The test period is **12 weeks (2016-April-29 to 2016-July-21)** where five stores in each of the test markets offered the updated menu along with television advertising.

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

   Weeks

Step 2: Data Cleaning

1. We first combined the RoundRoasterTransaction and the round-roaster-store files.
2. The data between 02-06-2015 and 07-21-2016 (76 weeks) is used as A/B test requires 52 weeks of data and 12 weeks extra to calculate the period of seasonality.
3. We then add fields for week, week_start, week_end and new_product flag to calculate the weekly traffic and sales for each store.
4. We then add Treatment_store dataset to create a list of control and treatment stores.

Step 3: Matching the Treatment and Control Units

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

   AvgMonthSales should be considered as constant variables. ***Square Feet*** should be ignored.

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

   We performed the Pearson Correlation Analysis and ***AvgMonthSales*** has a very high correlation of 0.99 with the performance metric : the Sum of Gross Margin. Square Feet which we have decided to ignore has a correlation coefficient of -0.02.

Pearson Correlation Analysis *Full Correlation Matrix*

	Sum_Gross.Margin	AvgMonthSales	Trend	Seasonality	Sq_Ft
Sum_Gross.Margin	1.000000	0.988207	-0.066564	0.088777	-0.020453
AvgMonthSales	0.988207	1.000000	-0.132040	0.153059	-0.046967
Trend	-0.066564	-0.132040	1.000000	-0.772280	0.197444
Seasonality	0.088777	0.153059	-0.772280	1.000000	-0.229980
Sq_Ft	-0.020453	-0.046967	0.197444	-0.229980	1.000000

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

The best control variables to match treatment and Control stores will be AvgMonthSales, Trend and Seasonality.

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

Treatment Store	Control Store 1	Control Store 2
1664	7162	8112
1675	1580	1807
1696	1964	1863
1700	2014	1630
1712	8162	7434
2288	9081	2568
2293	12219	9524
2301	3102	9238
2322	2409	3235
2341	12536	2383

Controls	Treatments	Distance	Test Group	Region	AvgMonthSales
7162	1664	0.478595	Barrington	Central	11000
8112	1664	1.034443	Barrington	Central	11000
1580	1675	0.45634	Northwest Hwy & Elmhurst Rd	Central	15000
1807	1675	0.560454	Northwest Hwy & Elmhurst Rd	Central	15000
1964	1696	0.312367	Higgins & Meacham	Central	10000
1863	1696	0.489137	Higgins & Meacham	Central	10000
2014	1700	0.810402	Roosevelt & Summit	Central	15000
1630	1700	0.91618	Roosevelt & Summit	Central	15000
8162	1712	0.671441	159th & LaGrange	Central	19000
7434	1712	0.793269	159th & LaGrange	Central	19000
9081	2288	0.277932	S. Parker Rd & E. Quincy Ave	West	14000
2568	2288	0.714134	S. Parker Rd & E. Quincy Ave	West	14000
12219	2293	0.348583	King Sooper Arvada # 55	West	11000
9524	2293	0.656038	King Sooper Arvada # 55	West	11000
3102	2301	0.381248	Hampden & Santa Fe, Sheridan	West	11000
9238	2301	0.434646	Hampden & Santa Fe, Sheridan	West	11000
2409	2322	0.171431	King Soopers - Denver #1	West	14000
3235	2322	0.45125	King Soopers - Denver #1	West	14000
12536	2341	0.39796	1352 College Ave - Boulder	West	11000
2383	2341	0.423792	1352 College Ave - Boulder	West	11000

Step 4: Analysis and Writeup

1. What is your recommendation - Should the company roll out the updated menu to all stores? There was an overall lift of about ~41% at a significance of 100%. The company should roll out the new menu to all stores.

Lift Analysis for Sum_Gross Margin

Significance Level	Lift	Expected Impact
100.0%	40.7%	681

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.45	0.09
Minimum	12.34	-16.18
Maximum	67.52	19.70
Standard Deviation	16.30	10.54

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

   We see a lift of 43.5% for Central Region with a significance of 99.5%.

Significance Level	Lift	Expected Impact
99.6%	43.5%	836

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.74	-1.73
Minimum	20.09	-16.18
Maximum	67.52	17.29
Standard Deviation	17.15	10.03

   We see a lift of ~38% for West Region with a significance of 99.5%.

Significance Level	Lift	Expected Impact
99.5%	37.9%	527

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.17	1.92
Minimum	12.34	-13.96
Maximum	55.30	19.70
Standard Deviation	16.34	11.24

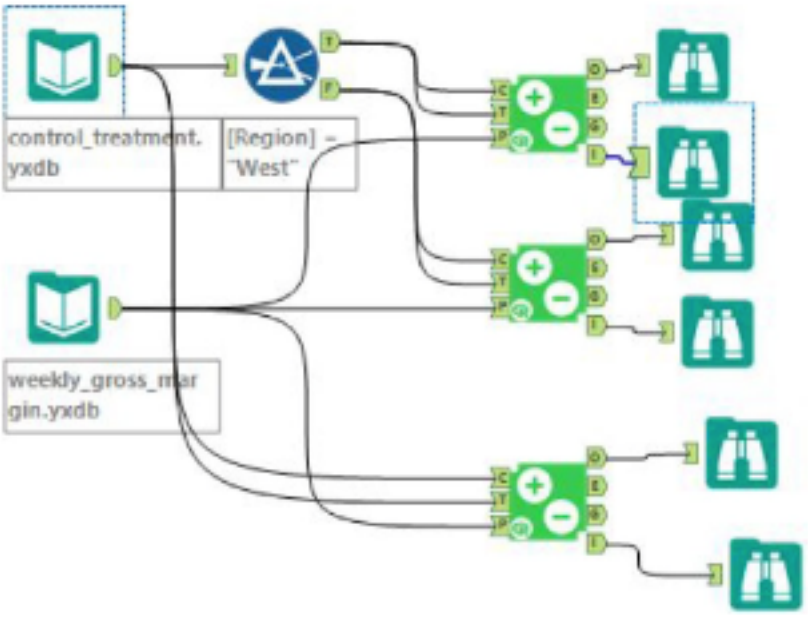
3. What is the lift from the new menu overall?

   The lift for the new menu is 40.7% with a statistical significance of 100%.

Significance Level	Lift	Expected Impact
100.0%	40.7%	681

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.45	0.09
Minimum	12.34	-16.18
Maximum	67.52	19.70
Standard Deviation	16.30	10.54


control_treatment.yxdb
[Region] = "West"
weekly_gross_margin.yxdb