12/5/2018

Analyzing a Market Test

Predictive Analytics for Business Project

Udacity Nanodegree

Tool: Alteryx

Katerina Bosko, PhD
WWW.CROSS-VALIDATED.COM

Business Problem:

Round Roasters is an upscale coffee chain with locations in the western United States of America. The past few years have resulted in stagnant growth at the coffee chain, and a new management team was put in place to reignite growth at their stores.

The first major growth initiative is to introduce gourmet sandwiches to the menu, along with limited wine offerings. The new management team believes that a television advertising campaign is crucial to drive people into the stores with these new offerings.

However, the television campaign will require a significant boost in the company's marketing budget, with an unknown return on investment (ROI). Additionally, there is concern that current customers will not buy into the new menu offerings.

To minimize risk, the management team decides to test the changes in two cities with new television advertising. Denver and Chicago cities were chosen to participate in this test because the stores in these two cities (or markets) perform similarly to all stores across the entire chain of stores; performance in these two markets would be a good proxy to predict how well the updated menu performs.

The test ran for a period of 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.

The comparative period is the test period, but for last year (2015-April-29 to 2015-July-21).

You've been asked to analyze the results of the experiment to 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; otherwise known as *incremental lift*. In the data, profit is represented in the *gross_margin* variable.

You have been able to gather three data files to use for your analysis:

- Transaction data for all stores from 2015-January-21 to 2016-August-18
- A listing of all Round Roasters stores
- A listing of the 10 stores (5 in each market) that were used as test markets.

Step 1: Plan Your Analysis

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

weekly gross margin per store

2. What is the test period?

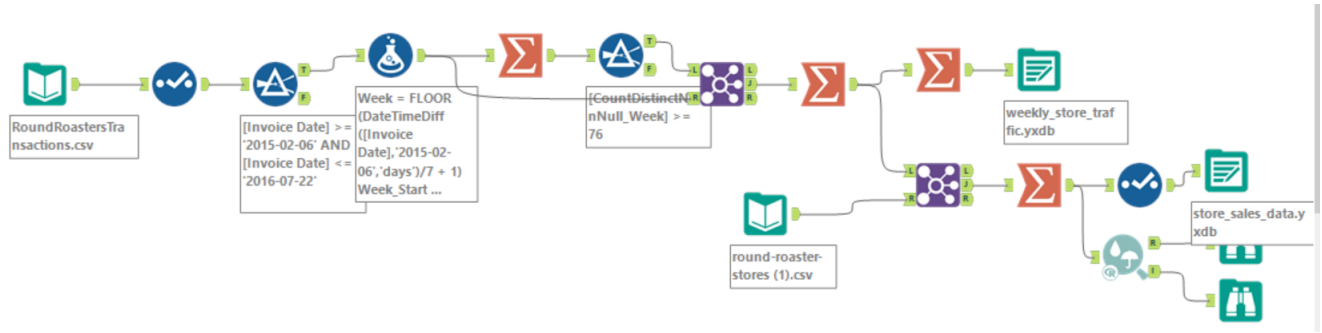
The test ran for a period of 12 weeks (2016-April-29 to 2016-July-21)

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

Week

Step 2: Clean Up Your Data

The workflow:



1. Imported RoundRoastersTransactions.csv file onto workflow.
2. Via Select tool, changed the type of certain variables as following:

	Field	Type	Size	Rename
▶	StoreID	V_String	254	
	Invoice Number	Int64	8	
	Invoice Date	Date	10	
	SKU	V_String	254	
	Category	V_String	254	
	Product	V_String	254	
	QTY	Int32	4	
	Size	V_String	254	
	Gross Margin	Double	8	
	Sales	Double	8	
	*Unknown	Unknown	0	

3. Filtered data, so that invoice date is between '2015-02-06' and '2016-07-22'. This is because we need 76 weeks altogether for the analysis.

4. Using Formula tool, calculated the week number, week start dates and week end dates.

5. Using Summarize tool, grouped the data by StoreID and counted number of weeks.

6. Using Filter tool, made sure that only those stores are included that have enough data (>= 76 weeks).

7. Joined the tables to include information from Step 2.

8. Using Summarize tool, grouped the data by StoreID, Invoice Number, Week, Week Start, Week End and summed the data for Gross Margin and Sales.

9. Aggregated the data once again by using the third Summarize tool, now counting the number of transactions per store per week.

10. Saved data into separate file ('weekly_store_traffic.yxdb). Here's the preview:

Record #	StoreID	Week	Week_Start	Week_End	Count
1	10018	1	2015-02-06	2015-02-12	308
2	10018	2	2015-02-13	2015-02-19	288
3	10018	3	2015-02-20	2015-02-26	204
4	10018	4	2015-02-27	2015-03-05	320
5	10018	5	2015-03-06	2015-03-12	284
6	10018	6	2015-03-13	2015-03-19	288
7	10018	7	2015-03-20	2015-03-26	194
8	10018	8	2015-03-27	2015-04-02	286
9	10018	9	2015-04-03	2015-04-09	274
10	10018	10	2015-04-10	2015-04-16	215

11. Via Join tool, added information on Sq_Ft, AvgMonthSales, Region from the round-roaster-stores.csv file.

12. Used Summarize tool once again to sum up Gross Margin and Sales at the level of Weeks.

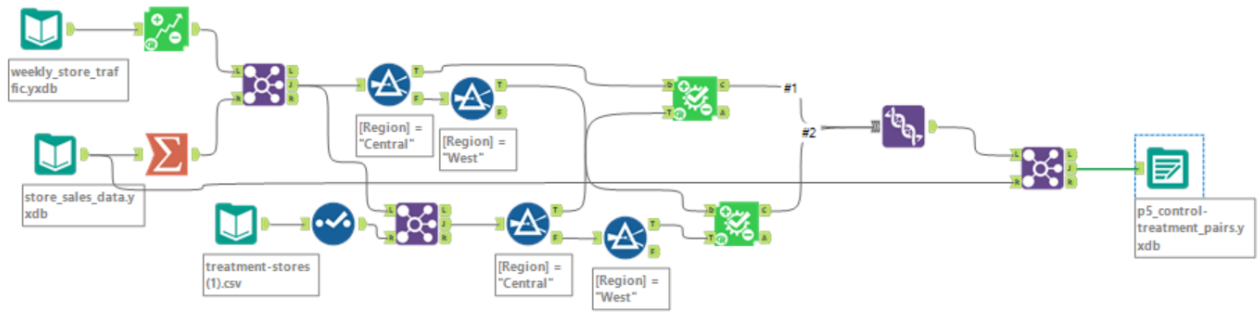
13. To choose the control variables, checked the correlation between variables Gross Margin, Sq_Ft and AvgMonthSales via Association Analysis tool. Here's the correlation matrix:

14. Saved the output into a separate file ('store_sales_data.yxdb'). Here's a preview:

Record #	StoreID	Week	Week_Start	Week_End	AvgMonthSales	Region	Sum_Gross Margin
1	10018	1	2015-02-06	2015-02-12	18000	West	2212.7105
2	10018	2	2015-02-13	2015-02-19	18000	West	2164.007
3	10018	3	2015-02-20	2015-02-26	18000	West	1560.929
4	10018	4	2015-02-27	2015-03-05	18000	West	2342.984
5	10018	5	2015-03-06	2015-03-12	18000	West	2199.4065
6	10018	6	2015-03-13	2015-03-19	18000	West	2103.143
7	10018	7	2015-03-20	2015-03-26	18000	West	1412.927
8	10018	8	2015-03-27	2015-04-02	18000	West	2124.3715
9	10018	9	2015-04-03	2015-04-09	18000	West	2216.152
10	10018	10	2015-04-10	2015-04-16	18000	West	1686.246

Step 3: Match Treatment and Control Units

The workflow:



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

AvgMonthSales

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

As determined in Step 2, there is strong correlation between Gross Margin and AvgMonthSales = 0.79
Since there is no correlation between Gross Margin and Sq_Ft, this variable was excluded from the further analysis.

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

Trend, Seasonality, AvgMonthSales and Region

4. 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

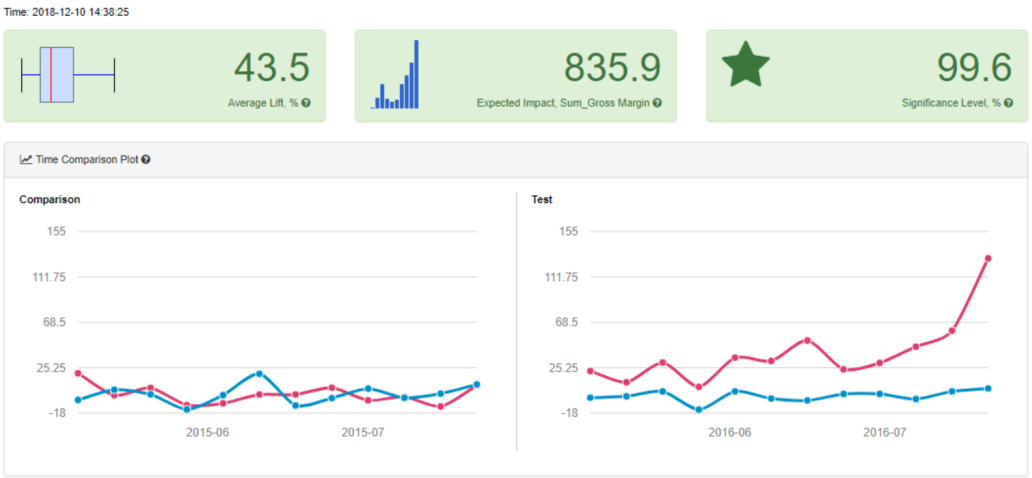
Step 4: Analysis and Writeup

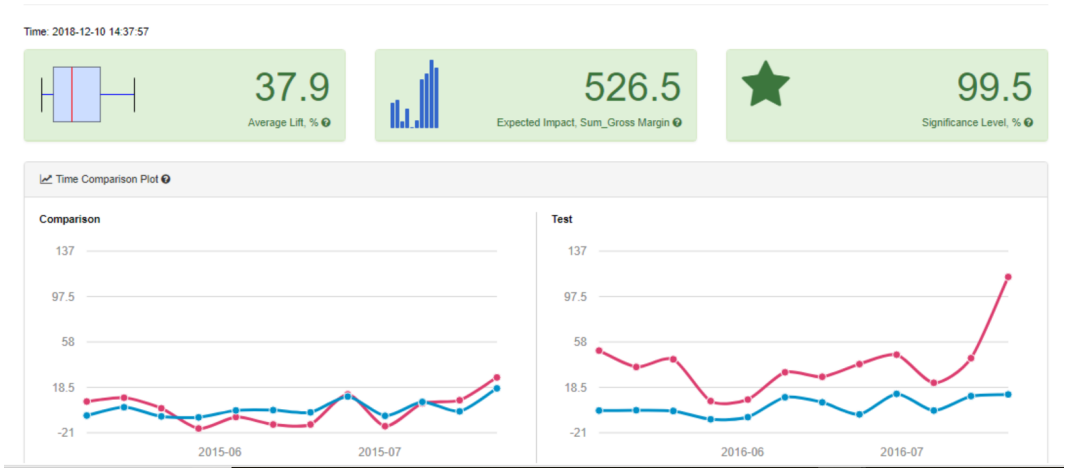
1. Recommendation

Based on the A/B analysis results, Round Roasters should roll out the updated menu to all stores.

2. Lift from the new menu for West and Central regions

The lift for the Central region is 43.5%, statistically significant at the level p=0.04.
The lift for the West region is 37.9%, statistically significant at the level p=0.05 (see pictures below).





3. Lift from the new menu overall

The overall lift per store per week from introducing the new menu is 40.7%, which is highly statistically significant. Expected impact on gross margin is $ 681.2 per week per store.