Minseok (Richard) Park
Predictive Analytics for Business Nanodegree

The 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.

Analysis Planning

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

   The performance metric that will be used to evaluate the sales growth from introduction of gourmet sandwiches and limited wine offerings is the sum of gross margin.

2. What is the test period?

   The test period will be 12 weeks, from April 29, 2016 to July 21, 2016.

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

The data will be aggregated at a week level, as the test period is in weeks.

Data Clean Up

First, RoundRoasters Transaction data and round-roaster-stores data are combined. 76 weeks of data (52 weeks of data + minimum of 12 weeks as the test period is 12 weeks) will be used to calculate seasonality and trend for AB Testing.

From the combined data with appropriate level and filtering, three different datasets are created for AB testing/analysis: '*Weekly_Store_Traffic'*, data containing average month sales and number of invoices per week per storeID, '*Store_Sales_by_Week'*, data containing weekly sum of gross margin and sum of sales along with regional information, and '*Store_List'* containing information of treatment and control stores with average month sales data for each store.

In the procedure, new variables '*Week*', '*Week_start*', '*Week_End*', and '*NewProduct_Flag*' will be added to calculate the weekly traffic and sales for each store.
Lastly, '*treatment_stores'* dataset is introduced to generate list of control and treatment stores.

Matching Treatment and Control Units

1. What control variables should be considered?

   Variable *'AvgMonthSales'* should be considered as the control variable, and *'Sq_Ft'* variable will not be used as the control variable

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

   Looking at the correlations between each potential control variables and the performance metric, variable *'AvgMonthSales'* is highly correlated with Sum of Gross margin variable with 0.99, but *Square Feet* has poor correlation with the sum of gross margin variable with -0.047. Therefore, we use *'AvgMonthSales'* as our control variables for the AB Controls.

Pearson Correlation Analysis

Full Correlation Matrix

	Sum_Sum_Gross.Margin	AvgMonthSales	Sq_Ft
Sum_Sum_Gross.Margin	1.000000	0.990978	-0.024224
AvgMonthSales	0.990978	1.000000	-0.046967
Sq_Ft	-0.024224	-0.046967	1.000000

Figure1. Pearson Correlation Analysis of possible control variables

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

   ***'AvgMonthSales'* will be used to match treatment and control stores, along with trend and seasonality.**

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

Table1. Treatment store with control stores

Analysis and Writeup

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

   In conclusion, the company should roll out the updated menu to all stores, as the average change in gross margin was 39.5% increase for treatment stores during the test period, which exceeds the company's minimum qualification of 18% increase in gross margin.

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

   For West regions, the average lift in gross margin is 37.9%, with statistical significance of 99.5%.
   For Central regions, the average lift in gross margin is 43.5%, with statistical significance of 99.6%.

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

   Overall, the average lift for the new menu launch is 40.7%, with statistical significance of 100%.

West Region

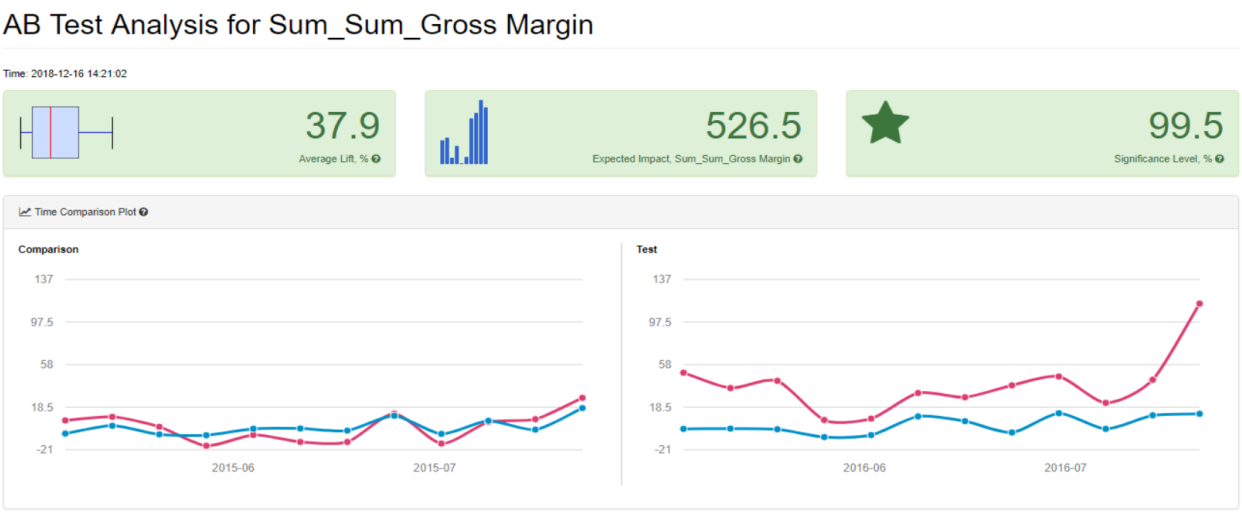

Figure 2. A/B Test Analysis for West Regions

Central Region

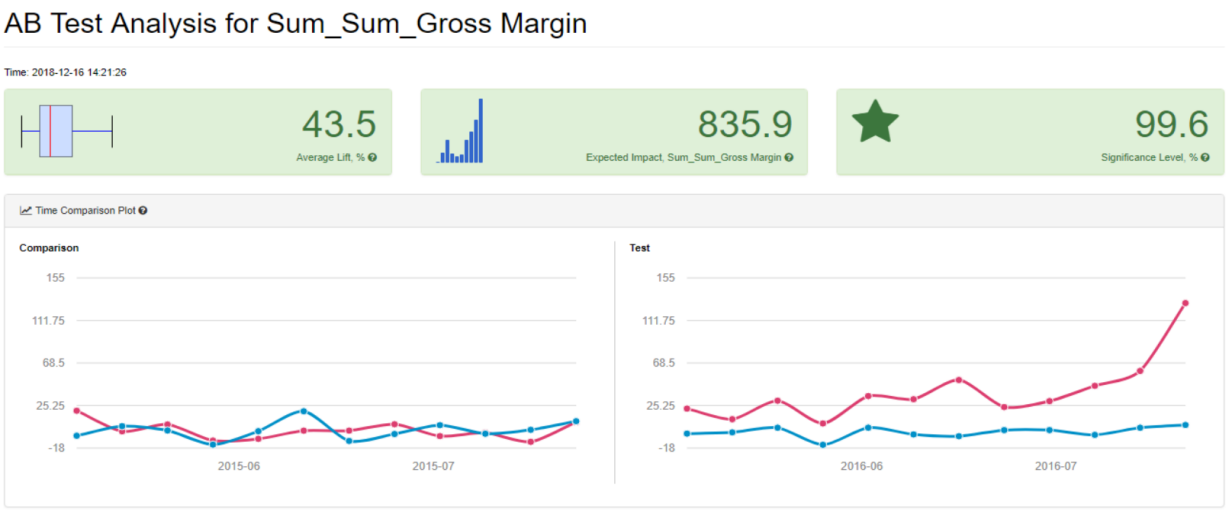

Figure 3. A/B Test Analysis for Central Region

Minseok (Richard) Park
Predictive Analytics for Business Nanodegree

Overall

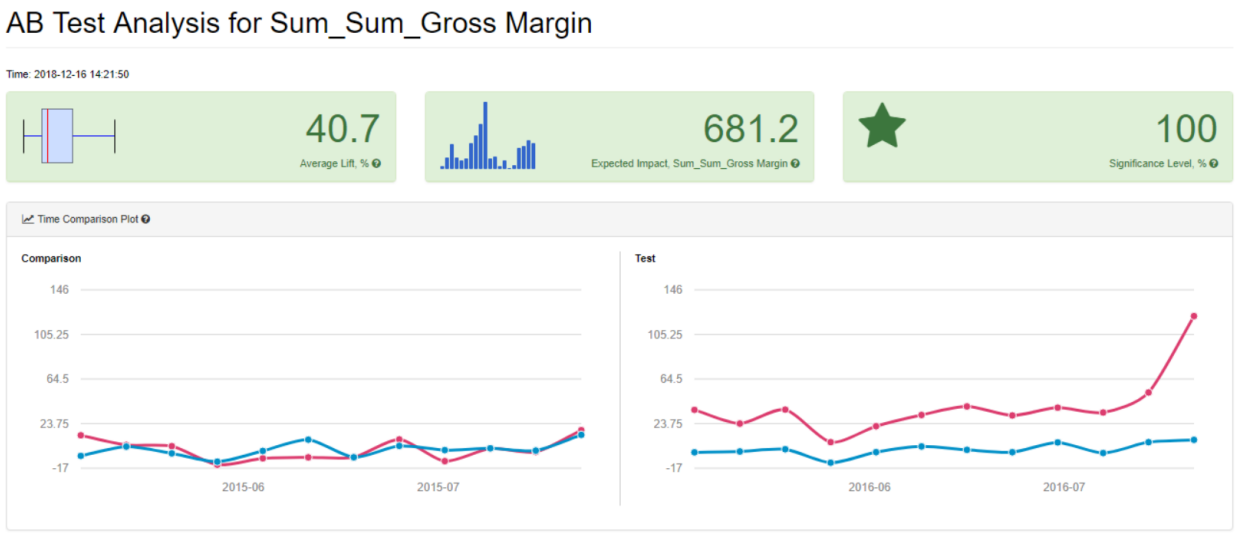

Figure 4. Overall A/B Test Analysis

Alteryx Workflow

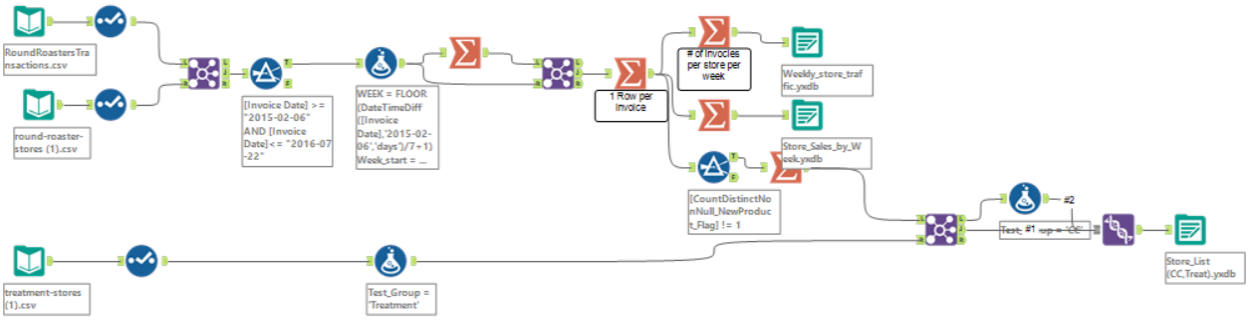

Figure 5. Data Preparation Workflow

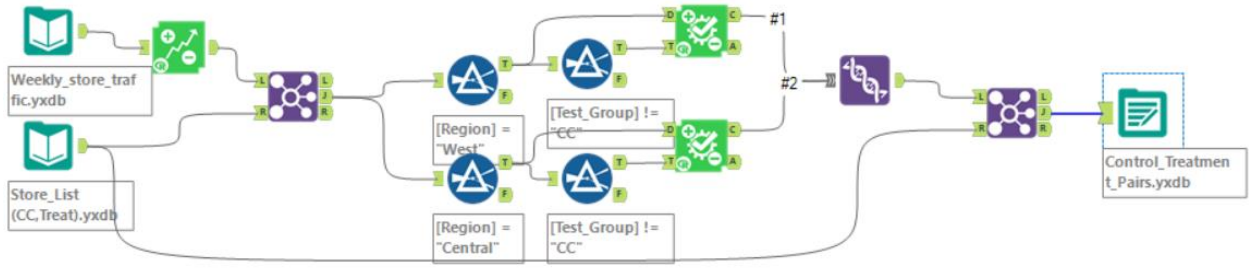

Figure 6. Control and Treatment Matching Pairs Workflow

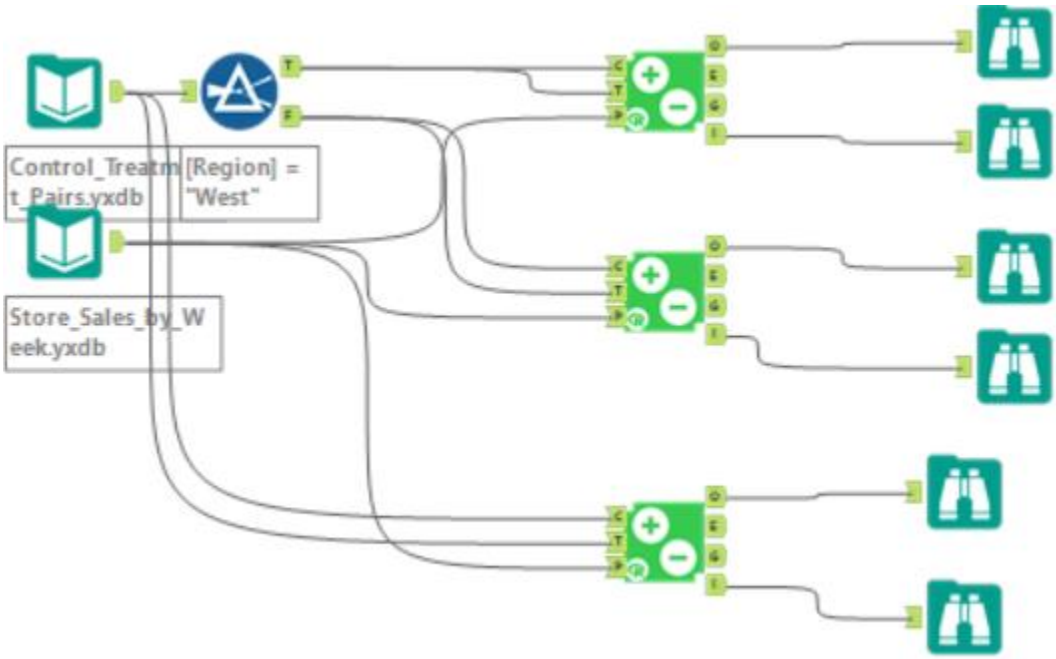

Figure 7. A/B Analysis Workflow