

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?

Ans: The Round Roasters coffee chain would like to know whether introducing gourmet sandwiches to the menu, along with limited wine offerings will increase their sales growth. So the **sum of gross margin** is the performance metric to be used.

2. What is the test period?

Ans: The test period for the A/B test is 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?

Ans: The data should be aggregated at **weekly** level.

Step 2: Clean Up Your Data

Ans: In this step our goal is to prepare data for the next 2 steps (**To Match Treatment & Control Units, to Perform AB Test Analysis**). The prepared data is stored in 3 output lists which are stored as Alteryx database files:


1. **Weekly_Store_Traffic.yxdb**
2. **Store_List.yxdb**
3. **Store_Sales_Analysis_Data.yxdb**

Here are the steps followed for creating the above 3 output files.

Step 1: The RoundRoastersTransactions.csv file is added via *Input* tool onto the canvas and the following fields are selected using *Select* tool:

- StoreID, Invoice Number, Invoice Date, Gross margin & Sales.

Step 2: Next *Filter* tool is used to filter the records which has exactly 76 weeks of data from 06 February 2015 to 21 July 2016.

 [Invoice Date] >= "2015-02-06" AND [Invoice Date] < "2016-07-22"

To leverage AB Analysis tool in Alteryx, at least 1 full year of data prior to the test and an

additional 12 weeks is required to calculate Trend and Seasonality. Hence, we need a total of 76 weeks of data.

Historical Data	- 52 Weeks
Trends	- 12 Weeks
Experiment	- 12 Weeks
<hr/>	
Total	- 76 Weeks
<hr/>	

Step 3: We are going to aggregate our data at week level, and to output the Weekly Store traffic. Hence 3 new fields Week, Week_Start and Week_End are generated using the following formulas using the *Formula* tool.

Week	FLOOR(DateTimeDiff([Invoice Date],'2015-02-06','days')/7+1)
Week_Start	DateTimeAdd('2015-02-06',7*([Week]-1),'days')
Week_End	DateTimeAdd([Week_Start],6,'days')

Step 4: Then, for each store, a count of weeks is created using *Summarize* tool.

 GroupBy StoreID then CountDistinct Week.

Step 5: Only the records that have exactly 76 weeks of data are extracted using *Filter* tool.

Step 6: *Join* tool is used to add back the Transaction data for stores that are having 76 weeks of data. Join is done using StoreID field. Unwanted field is removed to avoid duplicates.

Step 7: Use the *Summarize* tool to aggregate data, resulting in one invoice per row.


 Group StoreID, Invoice Number, Invoice Date, Week, Week_Start, Week_End.

 Sum Gross Margin and Sales.

Step 8: Use another *Summarize* tool again to count number of invoices per store per week.

 Group StoreID, Week, Week_Start, Week_End.

 Sum the Sum_Sales.

 Sum the Sum_Gross Margin.

Step 9: The output of this Summarize tool is saved as **Weekly_Store_Traffic.yxdb** using an *Output Data* tool for future use.

	StoreID	Week	Week_Start	Week_End	Sum_Sum_Gross Margin	Sum_Sum_Sales	CountDistinct_Invoice Nu
1	10018	1	2015-02-06	2015-02-12	2,212.7105	4,741.48	308
2	10018	2	2015-02-13	2015-02-19	2,164.007	4,571.25	288
3	10018	3	2015-02-20	2015-02-26	1,560.929	3,348.25	204
4	10018	4	2015-02-27	2015-03-05	2,342.984	5,114.96	320
5	10018	5	2015-03-06	2015-03-12	2,199.4065	4,799.48	284
6	10018	6	2015-03-13	2015-03-19	2,103.143	4,554.97	288
7	10018	7	2015-03-20	2015-03-26	1,412.927	2,999.55	194
8	10018	8	2015-03-27	2015-04-02	2,124.3715	4,519.61	286
9	10018	9	2015-04-03	2015-04-09	2,216.152	4,727.05	274
10	10018	10	2015-04-10	2015-04-16	1,686.246	3,689.33	215

Figure 1: Weekly Store Traffic

Step 10: Connect a *Summarize* tool to the output of Summarize tool used in Step 8,

- 🔧 GroupBy StoreID, Week, Week_Start, Week_End.
- 🔧 Sum the Sum_Sales and Sum_Gross Margin fields.

The output of this Summarize tool is saved as **Store_Sales_Analysis_Data.yxdb** using an *Output Data* tool for future use.

	StoreID	Week	Week_Start	Week_End	Sum_Sum_Gross Margin	Sum_Sum_Sales
1	10018	1	2015-02-06	2015-02-12	2,212.7105	4,741.48
2	10018	2	2015-02-13	2015-02-19	2,164.007	4,571.25
3	10018	3	2015-02-20	2015-02-26	1,560.929	3,348.25
4	10018	4	2015-02-27	2015-03-05	2,342.984	5,114.96
5	10018	5	2015-03-06	2015-03-12	2,199.4065	4,799.48
6	10018	6	2015-03-13	2015-03-19	2,103.143	4,554.97
7	10018	7	2015-03-20	2015-03-26	1,412.927	2,999.55
8	10018	8	2015-03-27	2015-04-02	2,124.3715	4,519.61

Figure 2: Store Sales Analysis Data

Step 11: A *Summarize* tool to the output of step 8.

- 🔧 Group by StoreID.

Step 12: The output of previous step is added to the L input of *Join* tool and the round-roaster-stores.csv is added at the R input side.

Step 13: An *Input data* tool is added onto the canvas, *treatment-stores.csv* file is connected. Then *Select* tool is used to select the fields StoreID, AvgMonthSales, Sq_Ft, City and Region.

Step 14: Next, a *Formula* tool is added to identify the Treatment candidate stores. Create a new field Test_Group, add a string value of "Treatment" for all the 10 records.

Step 15: Using a *Join* tool, add this Treatment records to the output of step 12.

Step 16: Add a *Formula* tool now. Create a new column named Test_Group which contains the string field of “CC”.

Step 17: combine these 2 tables using a *Union* tool.

The output of this *Join* tool is saved as **Store_List.yxdb** using an *Output Data* tool for future use.

	StoreID	Sq_Ft	AvgMonthSales	City	Region	Test_Group
1	10018	1,183	18,000	Long Beach	West	CC
2	10068	1,198	16,000	La Cresenta	West	CC
3	10118	1,204	13,000	Valencia	West	CC
4	10168	1,195	19,000	Burbank	West	CC
5	10218	1,193	15,000	Rancho Cucamonga	West	CC
6	10268	1,191	25,000	West Covina	West	CC
7	10318	1,147	16,000	El Cajon	West	CC
8	10368	1,159	19,000	San Marcos	West	CC
9	10418	1,146	19,000	San Diego	West	CC
10	10468	1,151	21,000	San Diego	West	CC

Figure 3: Store List

Step 18: My Alteryx Workflow for the above entire process of **Clean up Data** is given below:

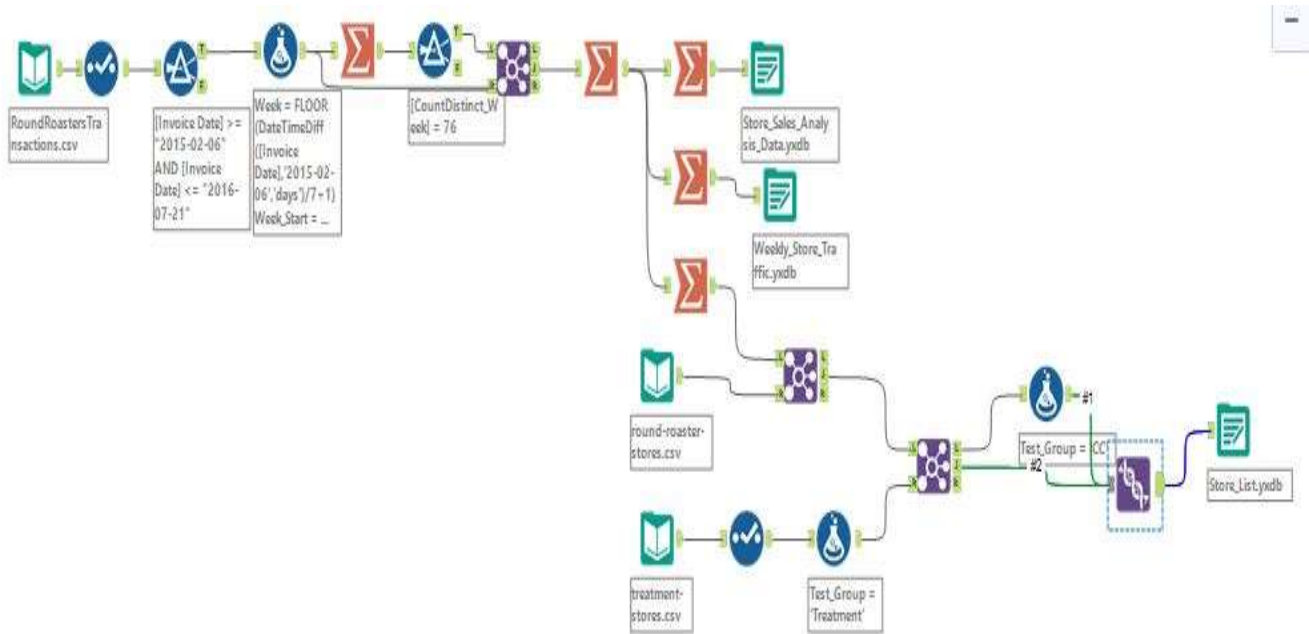


Figure 4: Alteryx Workflow Clean up Data.

Step 3: Match Treatment and Control Units

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

Ans: The **AvgMonthSales** and **Sq_Ft** are the numeric fields which are considered as control variables.

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

Ans: The Correlation Matrix is given below.

Pearson Correlation Analysis

Full Correlation Matrix

	Sq_Ft	AvgMonthSales	Sum_Gross.Margin
Sq_Ft	1.000000	-0.046967	-0.024255
AvgMonthSales	-0.046967	1.000000	0.990982
Sum_Gross.Margin	-0.024255	0.990982	1.000000

Matrix of Corresponding p-values

	Sq_Ft	AvgMonthSales	Sum_Gross.Margin
Sq_Ft		0.59138	0.78168
AvgMonthSales	0.59138		0.00000
Sum_Gross.Margin	0.78168	0.00000	

Figure 5: Pearson Correlation Analysis

The above report clearly depicts that the **AvgMonthSales** has a high correlation of 0.99 with the performance metric **Sum_Gross. Margin**, whereas Sq_Ft has a poor correlation of -0.05. Hence I chose AvgMonthSales as a control variable along with Trend with Seasonality.

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

Ans: The **AvgMonthSales** will be used a control variable along with **Trend** and **Seasonality** to match treatment and control stores.

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

Ans: There are **10 Treatment Records** and **3 control variables** (**AvgMonthSales, Trend & Seasonality**).

So there will be **30 Treatment-Control pairs**.

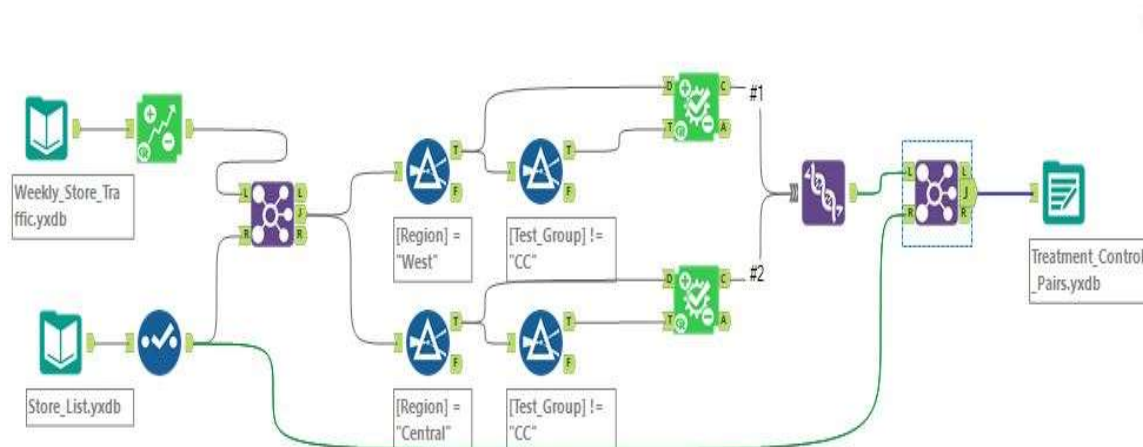
	Controls	Treatments	Distance	StoreID	Sq_Ft	AvgMonthSales	City	Region	Test_Group
20	9639	2293	0.730351	2293	1,394	11,000	Westminster	West	Treatment
21	8717	2293	0.891474	2293	1,394	11,000	Westminster	West	Treatment
22	12536	2301	0.376984	2301	1,387	11,000	Denver	West	Treatment
23	9238	2301	0.383212	2301	1,387	11,000	Denver	West	Treatment
24	2383	2301	0.420064	2301	1,387	11,000	Denver	West	Treatment
25	9388	2322	0.231362	2322	1,390	14,000	Denver	West	Treatment
26	3185	2322	0.279139	2322	1,390	14,000	Denver	West	Treatment
27	2409	2322	0.342072	2322	1,390	14,000	Denver	West	Treatment
28	2572	2341	0.26221	2341	1,400	11,000	Boulder	West	Treatment
29	12586	2341	0.459807	2341	1,400	11,000	Boulder	West	Treatment
30	3102	2341	0.485586	2341	1,400	11,000	Boulder	West	Treatment

Figure 6: Treatment-Control Pairs

Treatment Store	Control Store 1	Control Store 2
1644	1964	7162
1675	7284	2214
1696	1863	7334
1700	7037	2014
1712	8162	7434
2288	2568	9081
2293	12686	9639
2301	12536	9238
2322	9388	3185
2341	2572	12586

And the Alteryx workflow for generating the Treatment-Control pairs is given as below:

Figure 7: Alteryx Workflow (Match Treatment-Control Pairs).



Step 4: Analysis and Writeup

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

Ans: My recommendation is **YES**, the company should roll out the updated menu to all stores. According to the given project details, there should be at least 18% increase in profit growth compared over the control stores for the given test period, for justifying our recommendation. The following report clearly shows that the lift in the **profit (Sum_Gross Margin) is more than 18%, the company should roll out the updated menu.**

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

Ans: For the West Region: The Average lift is 40.4% and the Significance Level is 99.6%.

AB Test Analysis for Sum_Sum_Gross Margin

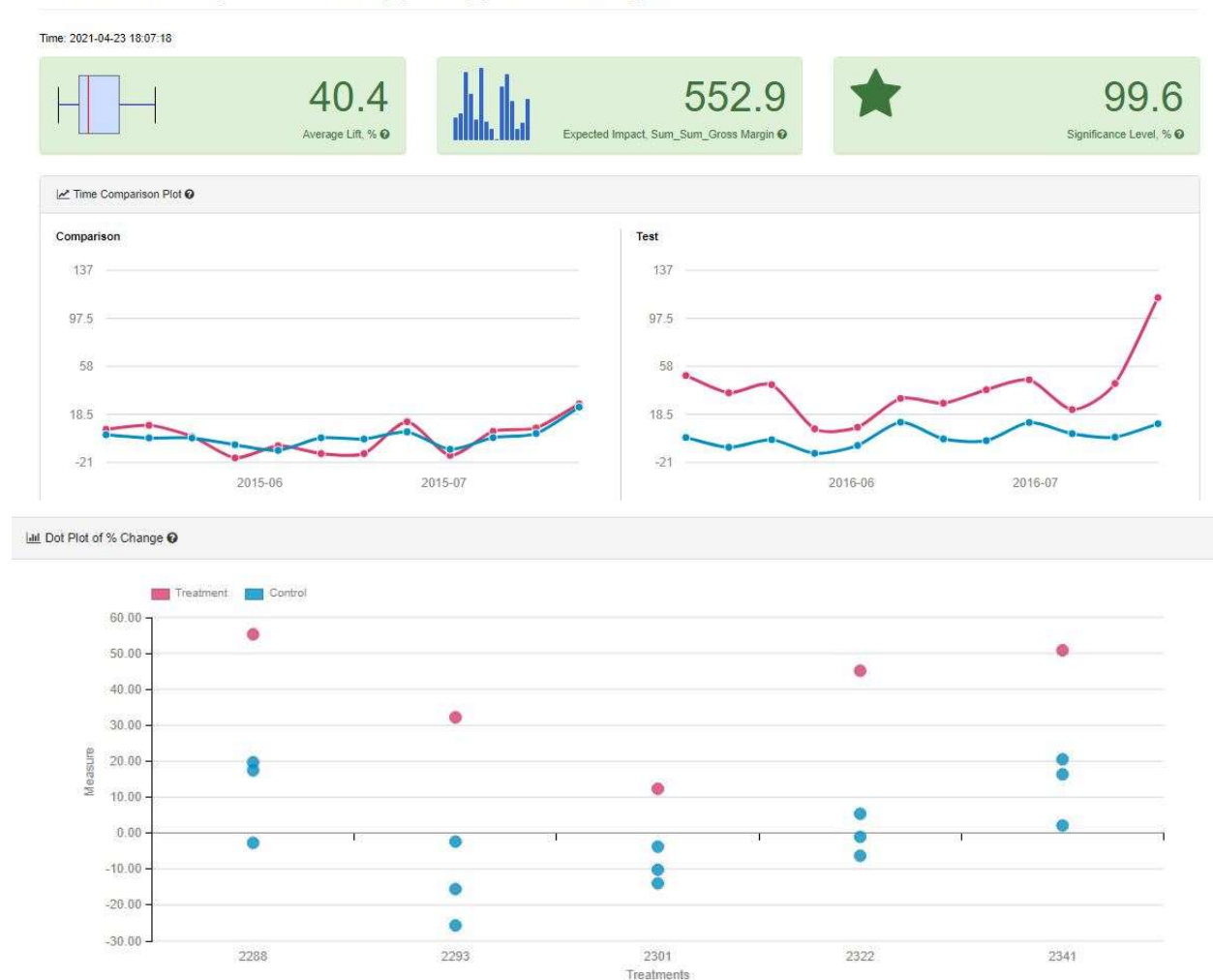


Figure 8: AB Analysis Report for WEST Region

For the Central Region: The Average lift is 42.5% and the Significance Level is 99.5%.

AB Test Analysis for Sum_Sum_Gross Margin

Time: 2021-04-23 18:07:54



Dot Plot of % Change

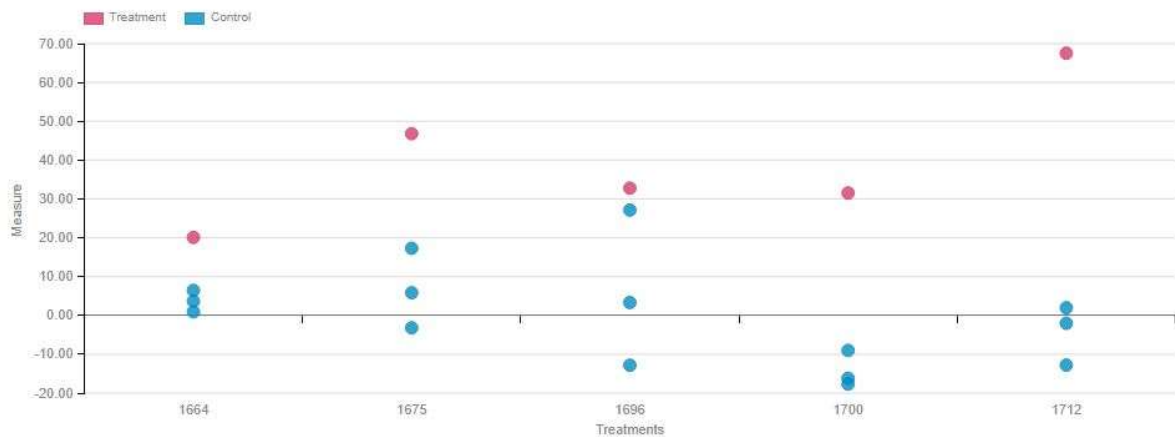


Figure 9: AB Analysis Report for CENTRAL Region

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

Ans: For the new menu the **overall lift is 41.4% and the Significance Level is 100%.**

AB Test Analysis for Sum_Sum_Gross Margin

Time: 2021-04-23 18:08:30



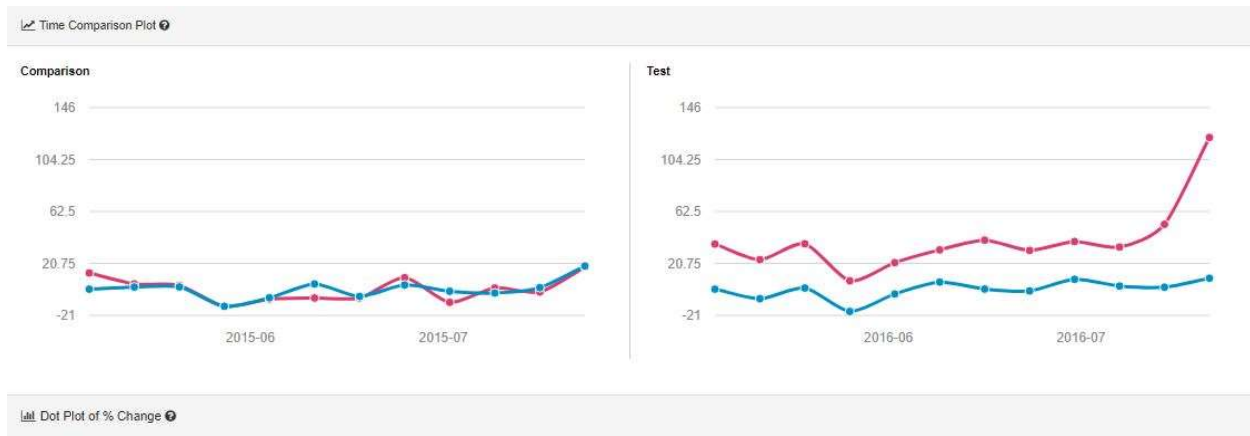


Figure 10: AB Analysis Report for Overall Lift

And the Alteryx Workflow for the AB Analysis is as follows:

