### Project Overview

You're a business analyst for Round Roasters, a coffee restaurant in the United States of America. 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. Your job is to analyze the A/B test and write up a recommendation to whether the Round Roasters chain should launch this new menu.

#### How Do I Complete this Project?

This project uses skills learned throughout the "A/B Testing ” course. To complete this project:

* Go through the course
* Apply the skills learned in the course to solve the business problem given in the project details section.
* Use our guidelines and rubric to help build your project.
* When you're ready, submit it to us for review using the submission template found in the supporting materials section.

#### Skills Required

In order to complete this project, you must be able to:

* Cleanup, format, and blend a wide range of data sources
* Plan and analyze A/B tests

NEXT

### 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](http://www.investopedia.com/terms/r/returnoninvestment.asp)). 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.

### Steps to Success

#### Step 1: Plan Your Analysis

To perform the correct analysis, you will need to prepare a data set. Prior to rolling up your sleeves and preparing the data, it’s a good idea to have a plan of what you need to do in order to prepare the correct data set. A good plan will help you with your analysis. Here are a few questions to get you started:

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

Gross margin or total sales per week

* What is the test period?

12 weeks

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

weekly

#### Step 2: Clean Up Your Data

In this step, you should prepare the data for steps 3 and 4. You should aggregate the transaction data to the appropriate level and filter on the appropriate data ranges. You can assume that there is no missing, incomplete, duplicate, or dirty data. You’re ready to move on to the next step when you have weekly transaction data for all stores.

#### Step 3: Match Treatment and Control Units

In this step, you should create the trend and seasonality variables, and use them along with your other control variable(s) to match two control units to each treatment unit. Treatment stores should be matched to control stores in the same region. Note: Calculate the number of transactions per store per week and use 12 periods to calculate trend and seasonality.

Apart from trend and seasonality...

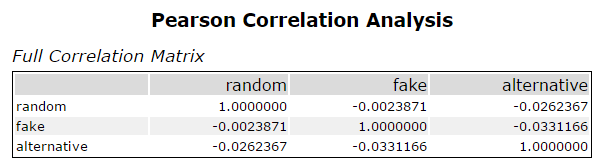
* What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.
* What is the correlation between your each potential control variable and your performance metric? (Example of correlation matrix below)
* What control variables will you use to match treatment and control stores?

Store Traffic

Weekly Total Sales

#### Step 4: Analysis and Writeup

Conduct your A/B analysis and create a short report outlining your results and recommendations.

[[](https://classroom.udacity.com/nanodegrees/nd008t/parts/ddc3c4f0-c44a-4c70-8372-0522a4c2b81c/modules/1d02aa02-80ab-4aee-8090-ae3dd40abb46/lessons/48312ab7-ffe9-415c-9cf1-7598f96bb361/concepts/b793edbe-f56b-492b-b42a-88b06a7f1b92)](https://classroom.udacity.com/nanodegrees/nd008t/parts/ddc3c4f0-c44a-4c70-8372-0522a4c2b81c/modules/1d02aa02-80ab-4aee-8090-ae3dd40abb46/lessons/48312ab7-ffe9-415c-9cf1-7598f96bb361/concepts/b793edbe-f56b-492b-b42a-88b06a7f1b92)

In an AB Analysis we use the correlation matrix to find the most correlated variable to the performance metric to include in the AB controls tool to help find the best matches.

# Tips

#### Aggregate and Export

We recommend you save your aggregated transaction database as a separate file to further reduce your development time. You shouldn't need to re-aggregate the transaction database every time you want to test out a new workflow.

#### Creating a week variable

When calculating the week column, assign week 1 to the first week of the test period, so the week\_begin variable for week 1 would be your test start date. This will cause weeks prior to the test period to be negative, which is perfectly fine. See [here](https://classroom.udacity.com/nanodegrees/nd008-ent/parts/840e51ff-4c1b-41bd-bc02-0a659921b756/modules/bbde5353-d851-4624-bca9-75e8a07e6f10/lessons/ee1d5344-e0f5-41bd-90af-30695d0458a8/concepts/e93ee789-fb71-486c-8b0f-cd0de9b080e0) . If the first link doesn't work, please click [here](https://help.alteryx.com/current/Formula.htm) for an example of how to calculate the week variable using a formula tool in Alteryx.

#### Number of Weeks

The trend tool is used to create trend and seasonality variables to use as control variables. To do this, you need at least 52 weeks of data, plus the number of weeks you select in the tool to calculate trend, before the beginning of the test start date. In lesson 4, you used 6 weeks to calculate the trend, so you needed 58 weeks prior to the test start date. For the project, you are asked to use 12 weeks to calculate trend, so you'll need 64 weeks of data prior to the test start date. Since the test lasts for 12 weeks, this means you'll need a total 76 weeks of data.

Your filter at the beginning of your workflow should go back 76 weeks from the end of the test period 2016-July-21. In Alteryx the expression shuold look like [Invoice Date]>="2015-02-06" AND [Invoice Date]<"2016-07-22"

All stores should have 76 weeks of Data.

### AB Trend Tool

The Test start date is 2016-April-29

The performance metric for this tool is the invoice count per week which represent weekly foot traffic. You had to create this variable with a summarize tool.

### AB Controls Tool

You should use 3 numeric measures to match treatment and control stores.

1. Trend
2. Seasonality
3. AvgMonthSales (This should be determined by looking at the correlation between the appropriate numeric variables in the round roasters stores file AvgMonthSales and Sq\_ft with the performance metric gross margin.) - This variable is in the round-roaster-stores.csv file

### AB Analysis Tool

Make sure to use weekly gross margin per store in all lift calculations and not total sales. Your data has gross margin in it but you will have to use a summarize tool to get weekly gross margin per store.

# Project: Analyzing a Market Test

Complete each section. When you are ready, save your file as a PDF document and submit it [here](https://classroom.udacity.com/nanodegrees/nd008/parts/11a7bf4c-2b69-47f3-9aec-108ce847f855/project).

## Step 1: Plan Your Analysis

*To perform the correct analysis, you will need to prepare a data set. (500 word limit)*

*Answer the following questions to help you plan out your analysis:*

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

The performance metric to evaluate the new menu is Gross Margin by Week

1. What is the test period?

The test period is for 12 weeks when the new menu will be implemented in the treatment stores.

And all the required data is:

76 weeks of data:

* 12 weeks for test (2016-04-29 to 2016-07-21)
* 12 weeks for trend (2016-02-05 to 2016-04-28)
* 52 weeks of historical data (2015-02-06 to 2016-02-06)

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

Gross margin and store traffic must be aggregated at weekly level

## Step 2: Clean Up Your Data

*In this step, you should prepare the data for steps 3 and 4. You should aggregate the transaction data to the appropriate level and filter on the appropriate data ranges. You can assume that there is no missing, incomplete, duplicate, or dirty data. You’re ready to move on to the next step when you have weekly transaction data for all stores.*

## Step 3: Match Treatment and Control Units

*In this step, you should create the trend and seasonality variables, and use them along with you other control variable(s) to match two control units to each treatment unit. Note: Calculate the number of transactions per store per week to calculate trend and seasonality.*

*Apart from trend and seasonality...*

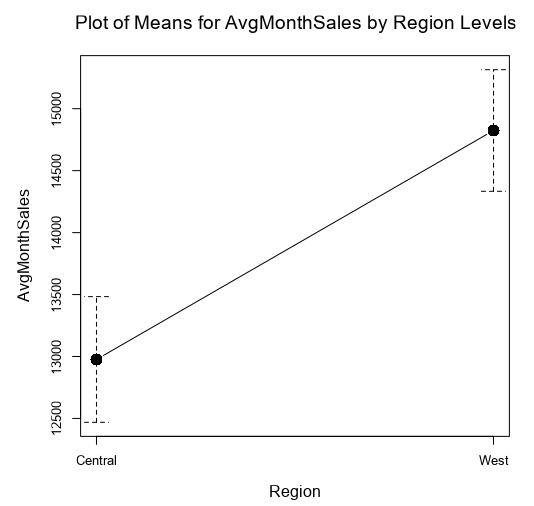
1. What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.

The variables that has been used from the file are:

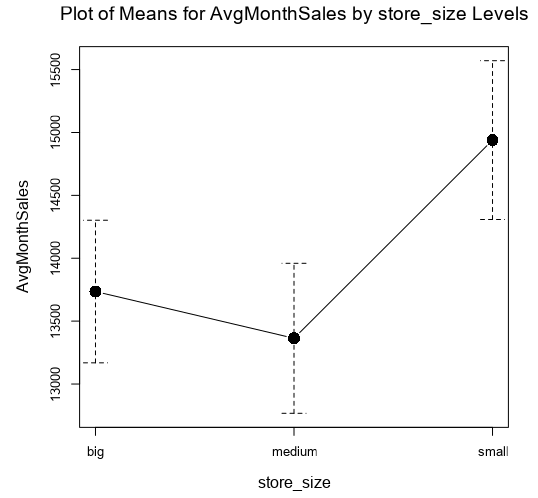
* Region
* AvgMonthSales
* Sq\_Ft

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

The next plot of means shows there is positive correlation for Region and Avg\_MonthSales



Also, Sq\_ft was evaluated with a plot of means using the ranges from interquartile values, and were created three categories small, medium and big. But the plot of means does not show correlation. Then Sq\_ft has been discarded as control variable.



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

* Region
* AvgMonthSales
* Seasonality
* Trend

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

|  |  |  |
| --- | --- | --- |
| Treatment Store | Control Store 1 | Control Store 2 |
| 1664 | 7162 | 1964 |
| 1675 | 1807 | 7584 |
| 1696 | 1863 | 7334 |
| 1700 | 7037 | 2014 |
| 1712 | 8162 | 7434 |
| 2288 | 2568 | 9081 |
| 2293 | 12686 | 9639 |
| 2301 | 12019 | 9238 |
| 2322 | 3235 | 9388 |
| 2341 | 2572 | 3102 |

## Step 4: Analysis and Writeup

*Conduct your A/B analysis and create a short report outlining your results and recommendations. (250 words limit)*

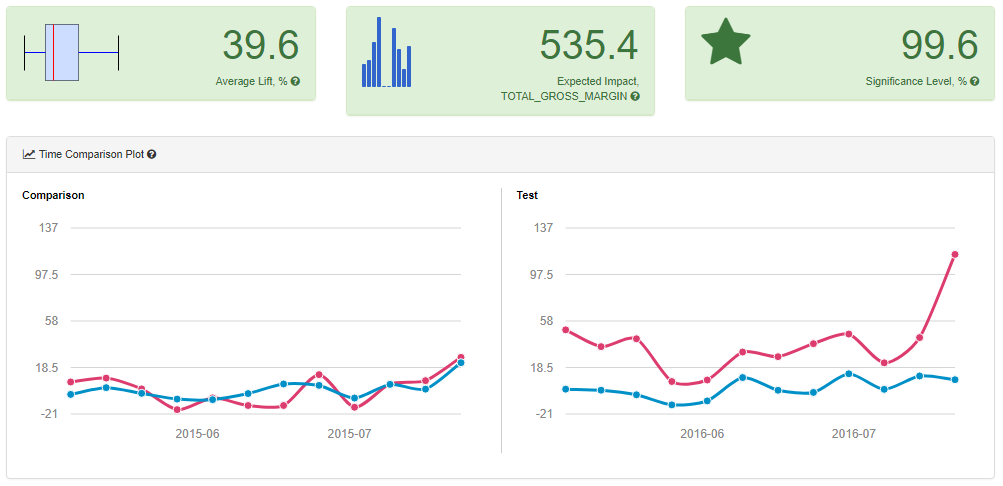
*Answer these questions. Be sure to include visualizations from your analysis:*

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

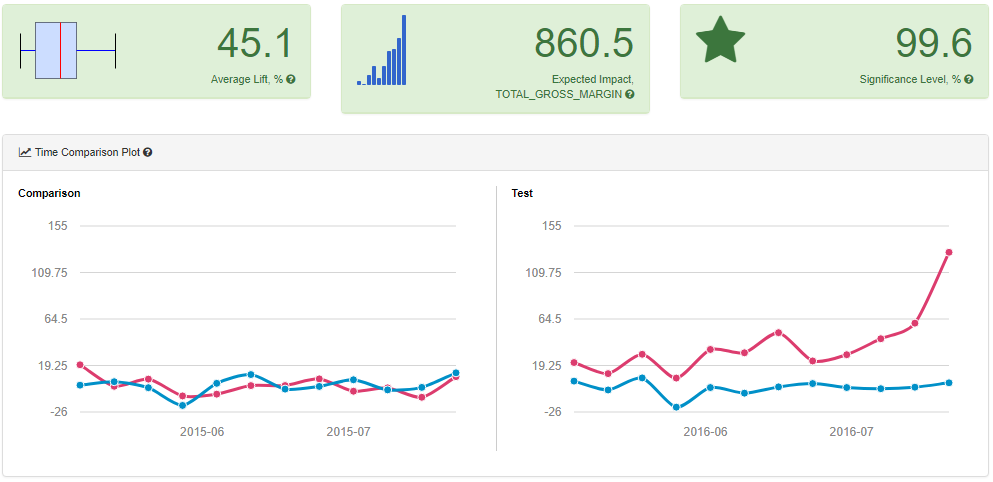
Yes, there is an increase of gross margin in the test period of 42.3 % with statistically significance of 100 %. Therefore, the new menu should be implemented in all stores.

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

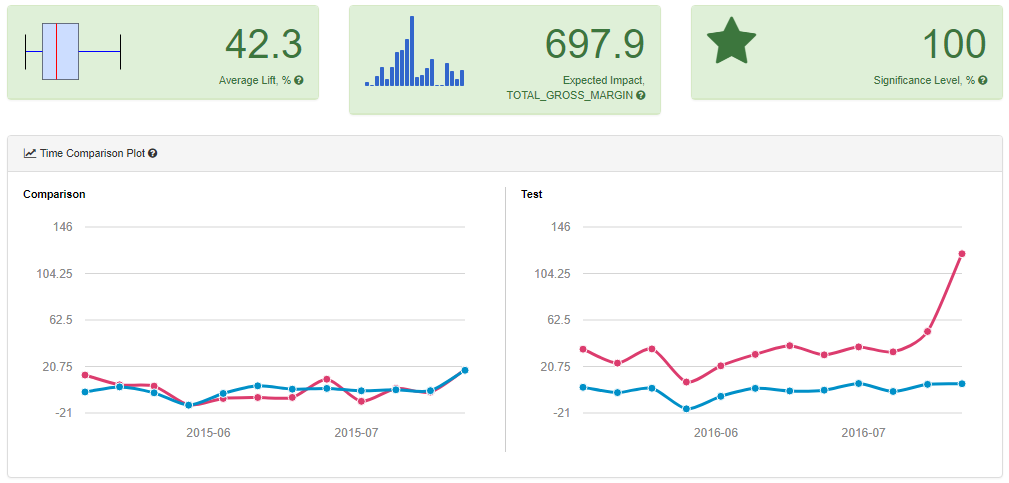
For the Stores in West region has the next lift and statistical significance:



For the Stores in Central region has the next lift and statistical significance:



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



**Alteryx Workflows**

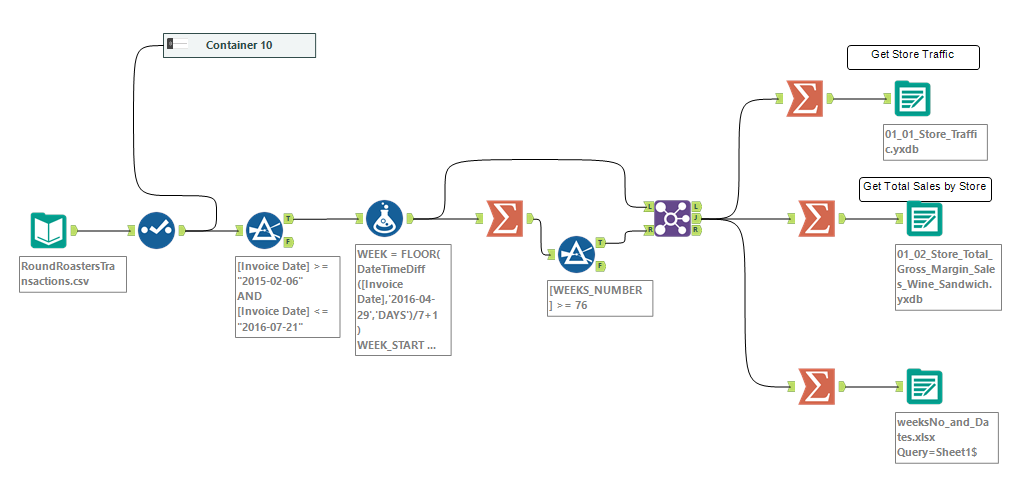


Figure . Data Preparation

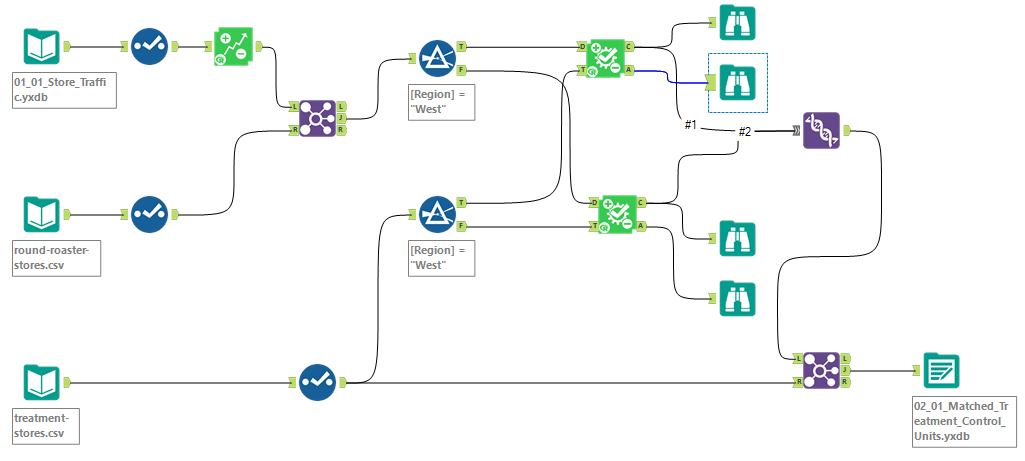


Figure Match Treatment and Control Stores

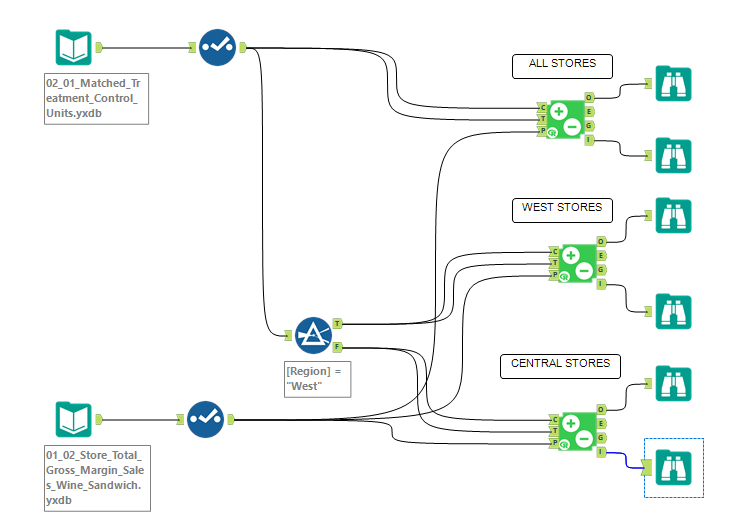


Figure AB Test Analysis