Project: Analyzing a Market Test BRIAN BERNS

Complete each section. When you are ready, save your file as a PDF document and submit it here.

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 used will be the sum of gross margin. From that we can analyze whether or not to introduce the sandwiches and wine to increase the growth of sales at Round Roasters.

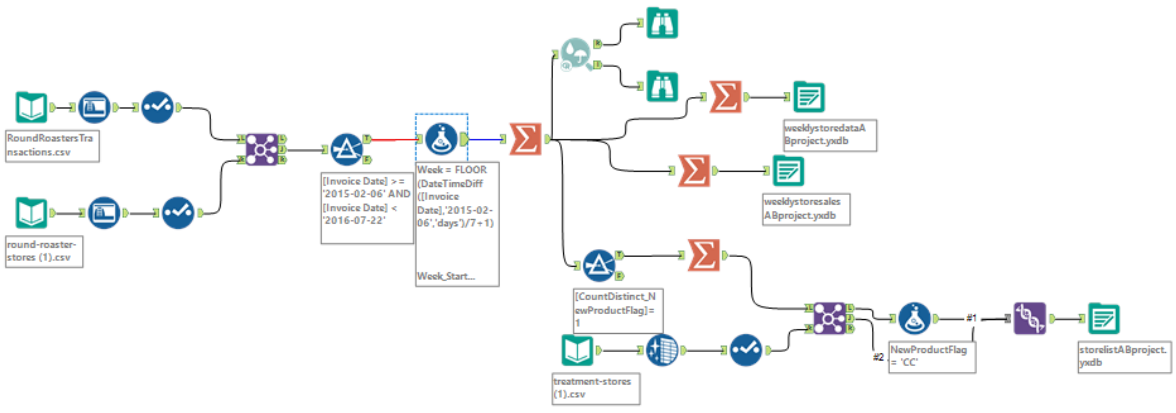
2. What is the test period?
   The test period was 12 weeks, which was from April 29th, 2016 to July 21st 2016.

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

To first prepare the data, we take the union of RoundRoasterTransaction.csv and RoundRoasterStores.csv. We would need 76 weeks of data due to A/B testing needing 52 weeks of data and then an extra 12 weeks to help calculate seasonality and trend. Since the period lasts 12 weeks for the test, we used 12 instead of the 6 weeks. I also created new columns named "week", "week_start", "week_end", and "NewProductFlag" to help calculate weekly foot traffic and sales. Once I have all this data set up and ready to go, I also added in the TreatmentStore.csv to help create treatment and control 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.
   I thought the two best control variables we can consider were AvgMonthlySales and SqFt. I used a Pearson Correlation Matrix to determine the best choice.

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

Pearson Correlation Analysis

Full Correlation Matrix

	Sq_Ft	Sum_Gross.Margin	Sum_AvgMonthSales
Sq_Ft	1.0000000	-0.0073479	-0.0294483
Sum_Gross.Margin	-0.0073479	1.0000000	0.8654716
Sum_AvgMonthSales	-0.0294483	0.8654716	1.0000000

   According to the Pearson Correlation Analysis, the sum of gross margin and AvgMonthlySales has a strong correlation of .8654, where the correlation between SqFt and AvgMonthlySales barely has a relationship (it's negative). Because of this fact, it would be safe to say we can eliminate SqFt as a control variable.

3. What control variables will you use to match treatment and control stores?
   AvgMonthly sales, along with trend and seasonality would be used to help match treatment and control stores.
4. Please fill out the table below with your treatment and control stores pairs:

Treatment Store	Control Store 1	Control Store 2
1664	1857	7484
1675	2114	8562
1696	1964	7584
1700	1508	7384
1712	7284	8212
2288	12069	9081
2293	11568	12219
2301	10018	10468
2322	2409	3102
2341	2333	11368

Step 4: Analysis and Writeup

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

recommendations. (250 words limit)

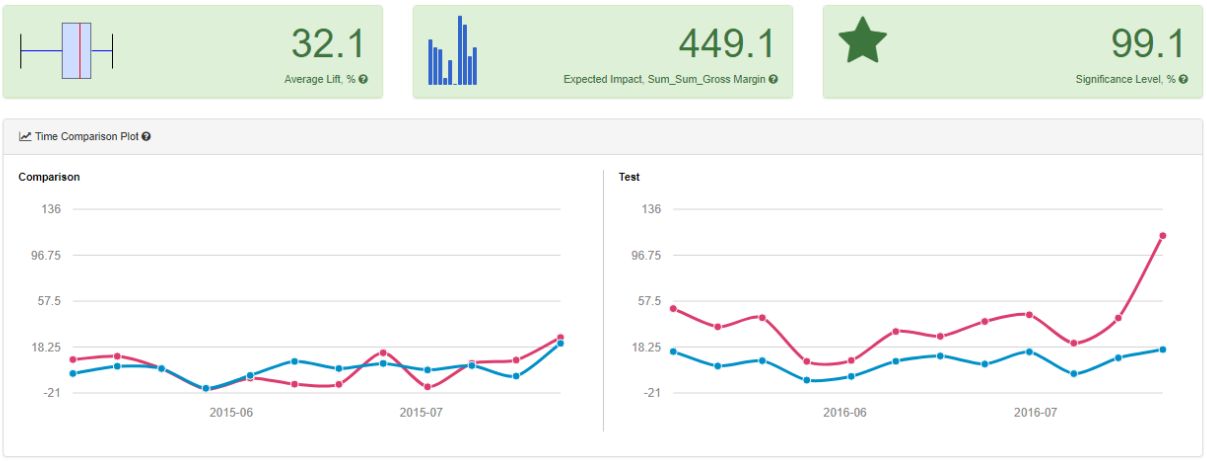
Answer these questions. Be sure to include visualizations from your analysis:

1. What is your recommendation - Should the company roll out the updated menu to all stores?
   The company SHOULD roll out the updated menu to all stores because the profit margin has increased more than 18%.

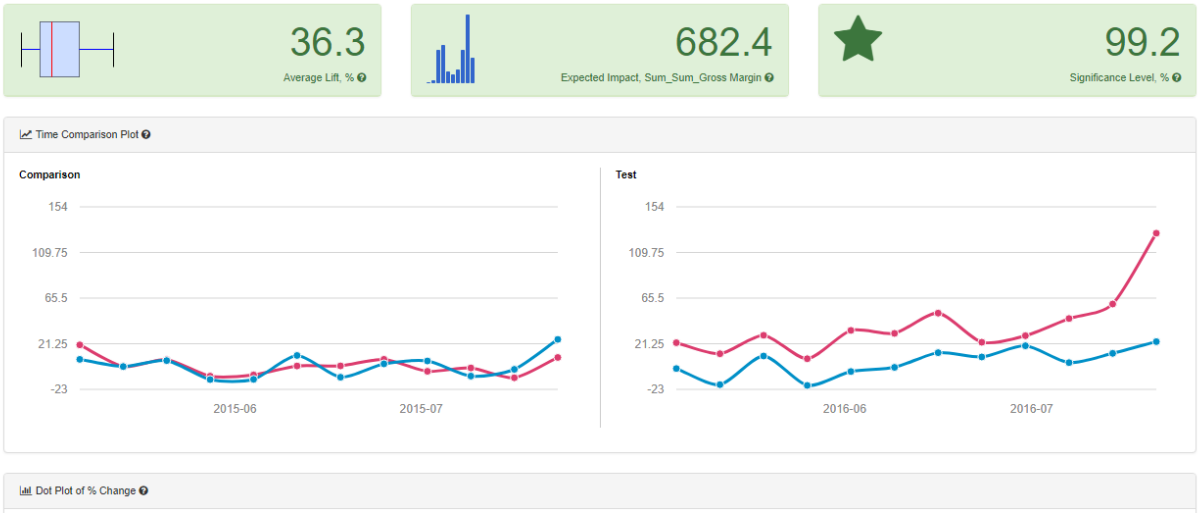
2. What is the lift from the new menu for West and Central regions (include statistical significance)?
   The lift for the Western Region is 32.1% with a 99.1% significance level, and the lift for the central region is 36.3% with a 99.2% significance level.

3. What is the lift from the new menu overall?
   The lift for the overall menu is 34.2% with a 100% significance level.

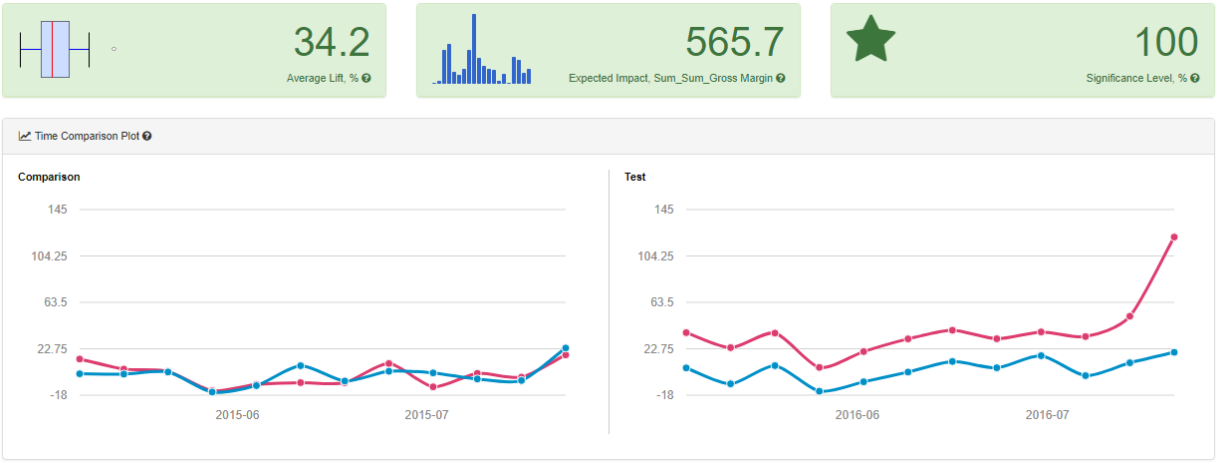
WESTERN REGION

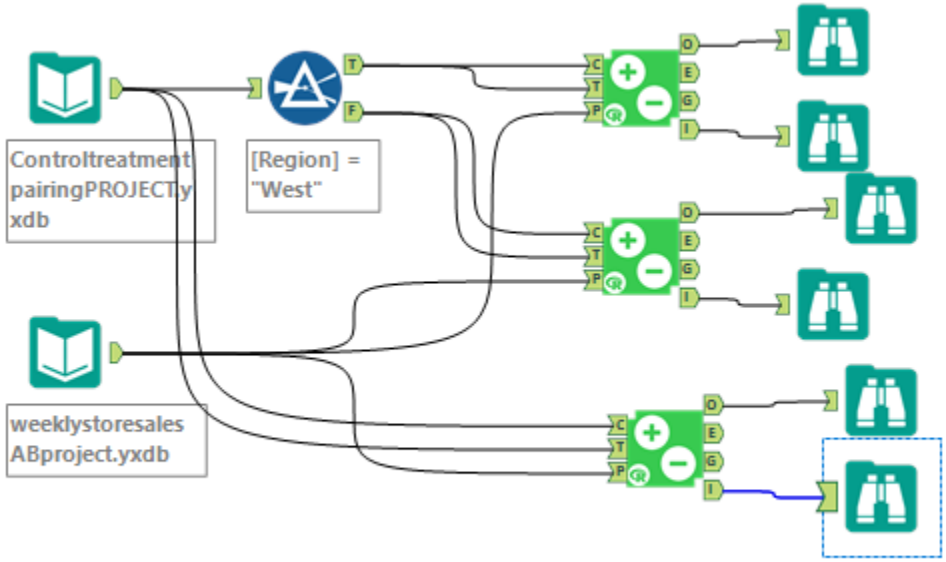


CENTRAL REGION



OVERALL





Before you Submit

Please check your answers against the requirements of the project dictated by the rubric here. Reviewers will use this rubric to grade your project.