Project: Analyzing a Market Test

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. (250 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?

   In this case, it would logically be better to use gross margin as it is the performance metric that represents profit. Total sales on the other hand just shows the total amount sold. We would like to know how much we earned as profit over the amount we made in total.

2. What is the test period?

   The test ran for 12 weeks from April 29 2016 to July 21 2016.

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

   A week would be the best level to aggregate the data by. Since it's a coffee chain, a week would be enough to show one cycle of customers. Also, the test ran for 12 weeks and aggregating it by months would mean that the data we have is not sufficient. Aggregating the data by day would not be enough to show a cycle of customers.

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.

   There are only 2 possible control variables from the file. Sq_ft being one and

Avgmonthsales the other. These are the only variables that could reasonably impact gross margin.

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

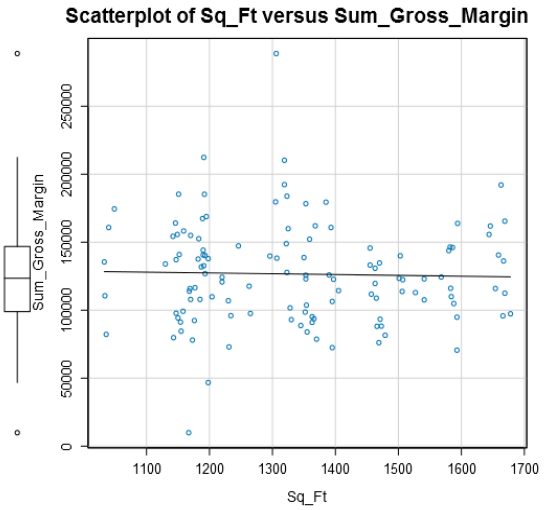
Pearson Correlation Analysis

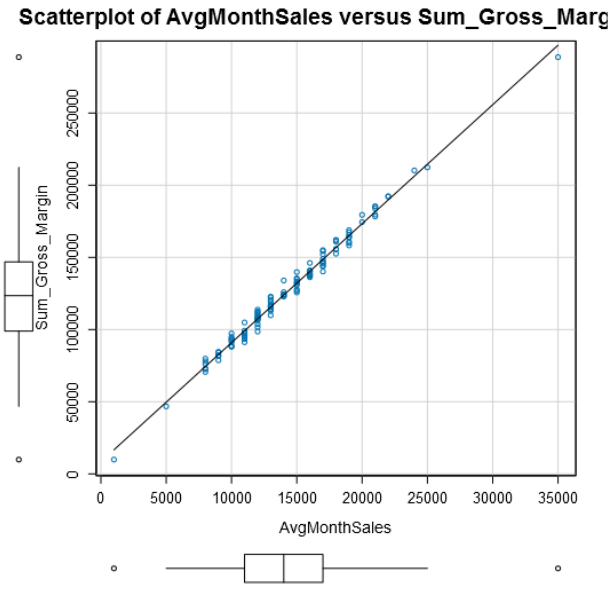
Full Correlation Matrix

	AvgMonthSales	Sq_Ft	Sum_Gross.Margin
AvgMonthSales	1.000000	-0.040063	0.994571
Sq_Ft	-0.040063	1.000000	-0.028444
Sum_Gross.Margin	0.994571	-0.028444	1.000000

Matrix of Corresponding p-values

	AvgMonthSales	Sq_Ft	Sum_Gross.Margin
AvgMonthSales		0.65997	0.00000
Sq_Ft	0.65997		0.75481
Sum_Gross.Margin	0.00000	0.75481	





There appears to be a negative correlation between square feet and gross margin as seen from the scatterplot. The p-value is 0.75 and as such is not a good control variable. On the other hand, average month sales has a really high correlation with gross margin

at 99%. The p-value is 0.00 and as such could be used as a control variable.

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

   The only control variable I will be using is avgmonthsales.

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

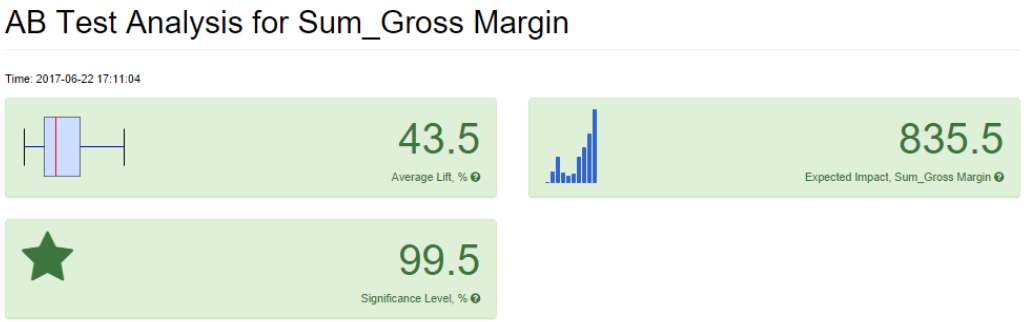
Conduct your A/B analysis and create a short report outlining your results and recommendations. (250 words limit)

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?

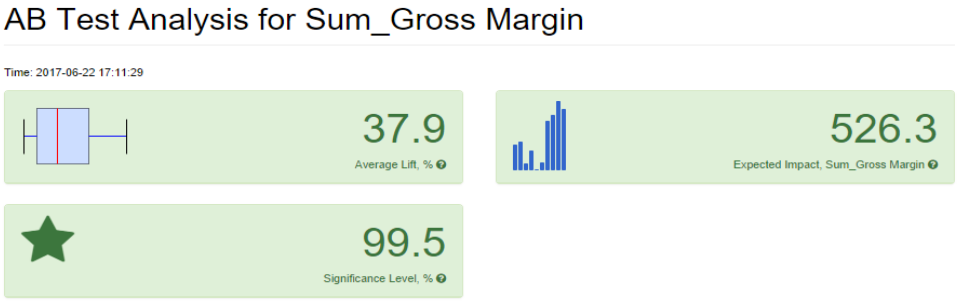
   Based on the A/B analysis results, it would be best to roll out the updated menu to all stores as soon as possible.

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



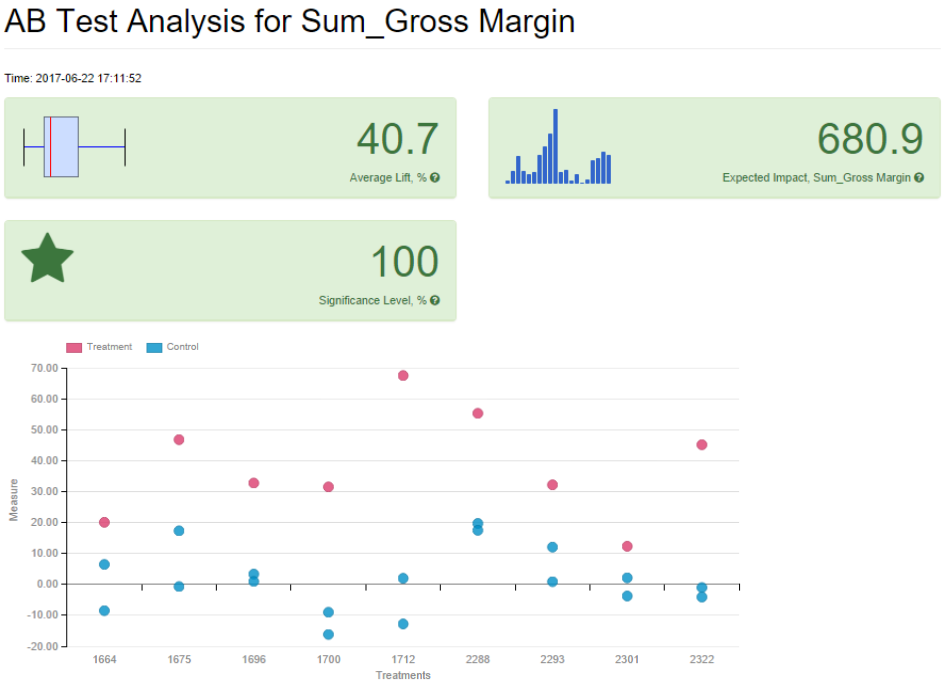
   Central has the lift at 43.5% and its significance level is 99.5%. This means that the new

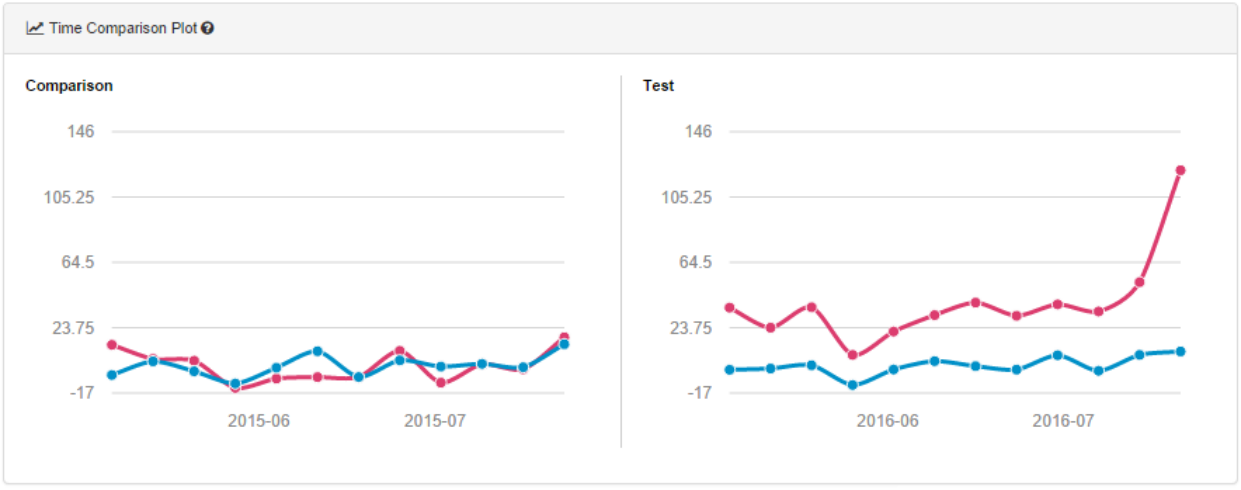
menu was very effective on this region.



West has the lift at 37.9% and its significance level is 99.5%. The menu was also effective in this region but it is still lower than Central.

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





The overall lift is 40.7% with a 100% significance level. One can see from the dot plot that the all the treatment stores outperform the control stores. The comparison plot shows the improvement on the treatment stores in comparison to the previous year on the same period. The same goes with the control stores but in this case the control stores did just as poorly as the previous year. With this, there is no doubt that rolling out the new menu would be profitable to the coffee chain.

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.