Project: Analyzing a Market Test

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?

    - In order to determine if the new menu changes along with new television advertising has brought significant enough at least 18% increase in profit growth against the comparative period of last year, and against the same experiment period of control stores. The variable used in this case to measure the performance of profit is "gross margin" in the dataset.

2. What is the test period?

    - 04/29/2016 - 07/21/2016 (12weeks)

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

    - I chose week as the appropriate aggregated level as if by day, day is hard to compare day by day; if by month, only three-month data is too short to represent trend.

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.

    - In the RoundRoastersStore file, we can separate data type of all variables into numeric and categorical.
    - For categorical, I consider region and state only due to data availability.
    - For numerical, since our performance metric is total_sumofprofit, I then ran correlation analysis and scatterplot between sq_ft and avgmonthsales against total_sumofprofit separately as below.

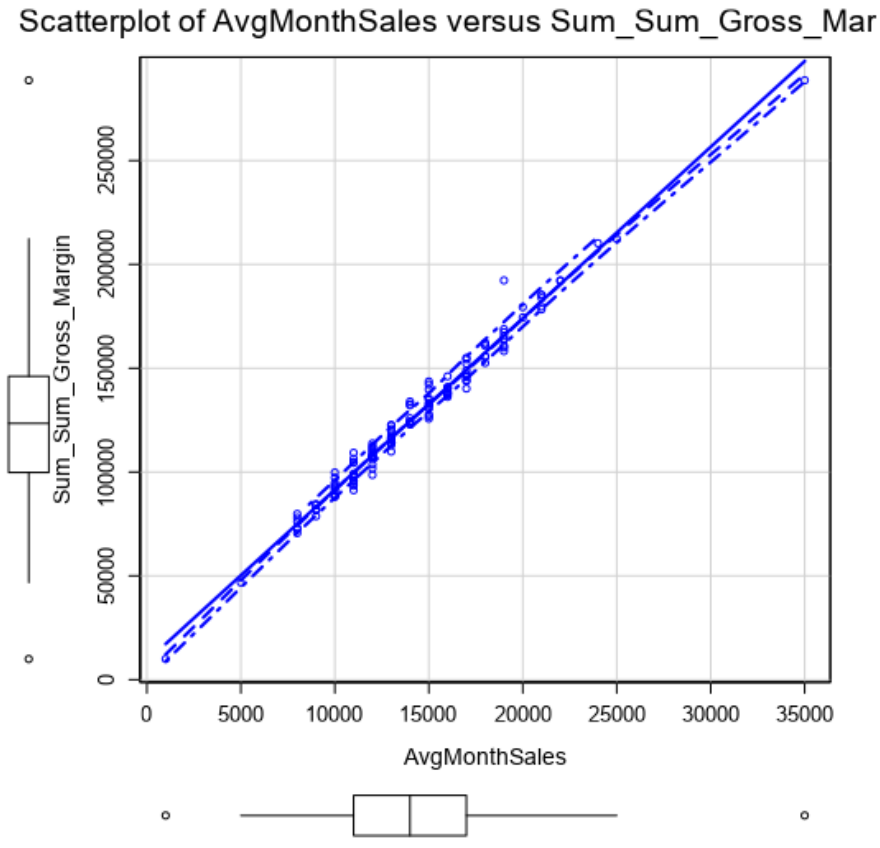
2. What is the correlation between your each potential control variable and your performance metric?

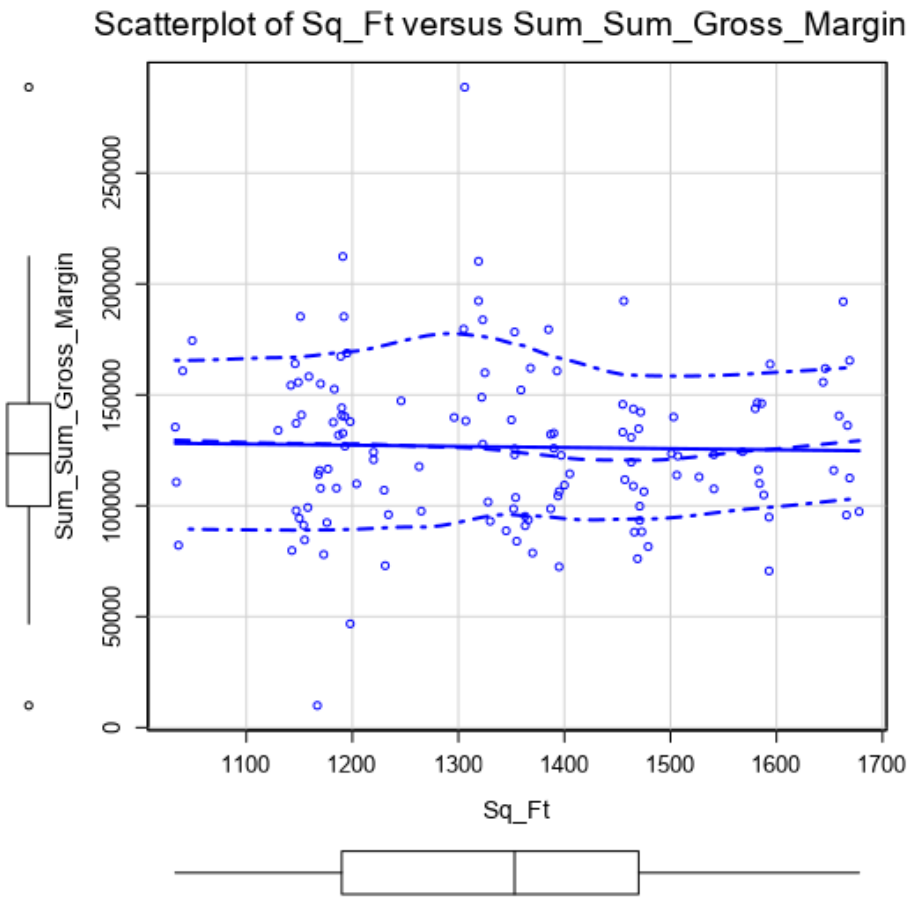
    - For categorical, since the 10 stores treatment data are in CO and IL state, and west and central region, so I then consider the same region and state as control variables.
    - For numerical, I ran correlation analysis and scatterplots between total_sumofprofit against sq_ft, and total_sumofprofit against avgmonthsales. The result shows only avgmonthsales has positive linear relationship with total_sumofprofit. The scatterplot between sq_ft and total_sumofprofit shows no relationship.

Pearson Correlation Analysis

Focused Analysis on Field Sum_Gross.Margin

	Association Measure	p-value	
AvgMonthSales	0.790358	0.0000000	***
Week	-0.026179	0.0084861	**
Sq_Ft	-0.019345	0.0517960	.





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

    - I use avgmonthsales, and calculated trend and seasonality with AB Trend tool to 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
2288	9081	2568
2293	12219	9524
2301	3102	9238
2322	2409	3235
2341	12536	2383
1664	7162	8112
1675	1580	1807
1696	1964	1863
1700	2014	1630
1712	8162	7434

Step 4: Analysis and Writeup

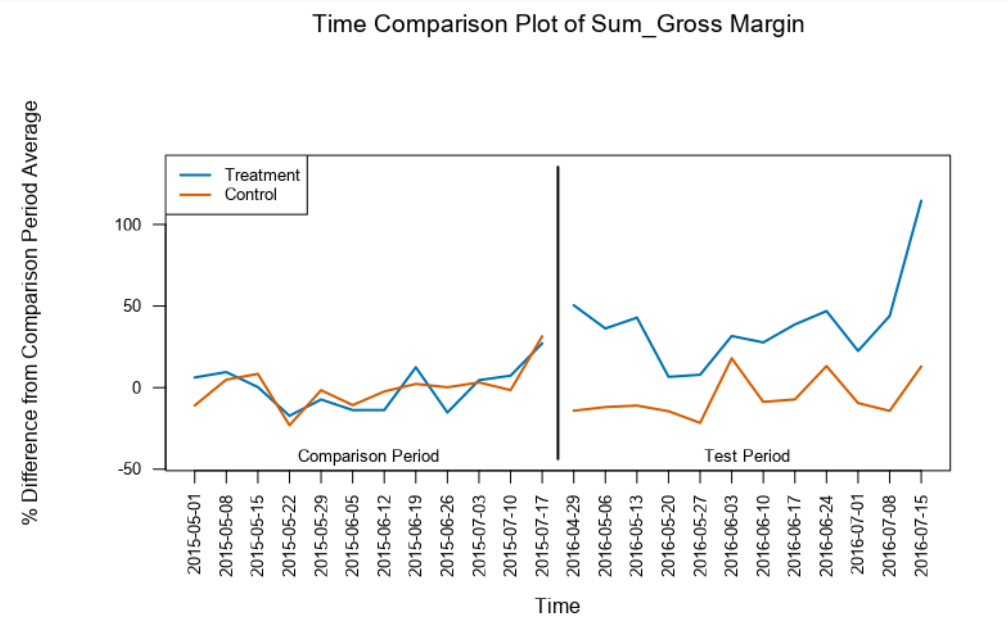
Conduct your A/B analysis and create a short report outlining your results and recommendations. (250 words limit)

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

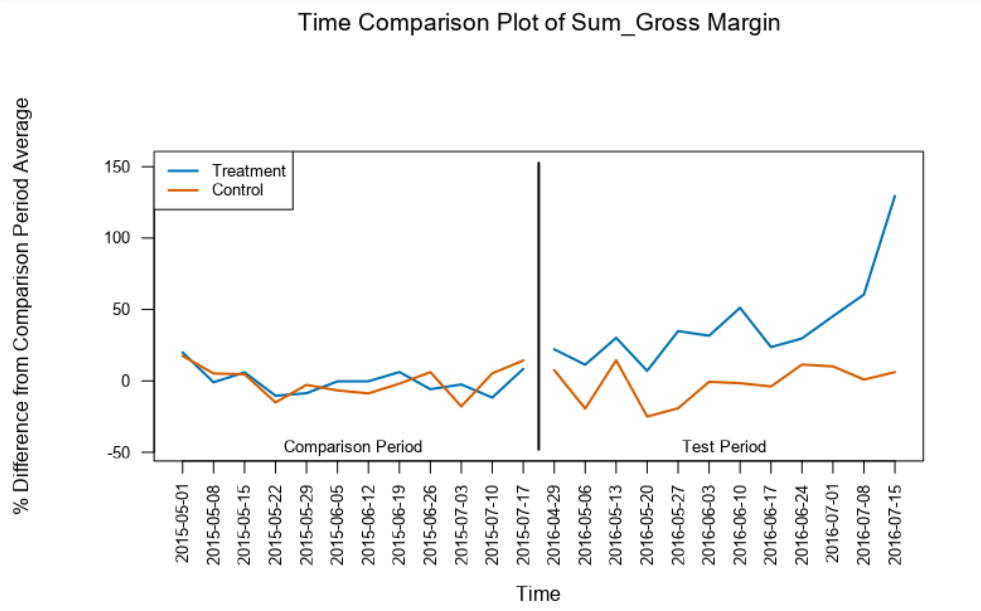
1. What is your recommendation - Should the company roll out the updated menu to all stores?

    - I recommend proceeding new menu with television advertising to all stores to boost total sales margin. The reason is as follows.
    - With 5 stores in CO (west region) and 5 stores in IL (central region) as an experimental store, the AB test analysis result indicates the percentage difference from comparison period between treatment and control has significant increase in treatment stores for both regions.
    - From the box plot, the mean of percentage change for treatment is around 45%, which is higher than our target at least 18% increase in total profit margin. Thus, it shows the new update on menu does increase total gross margin significantly and above the target performance measure.

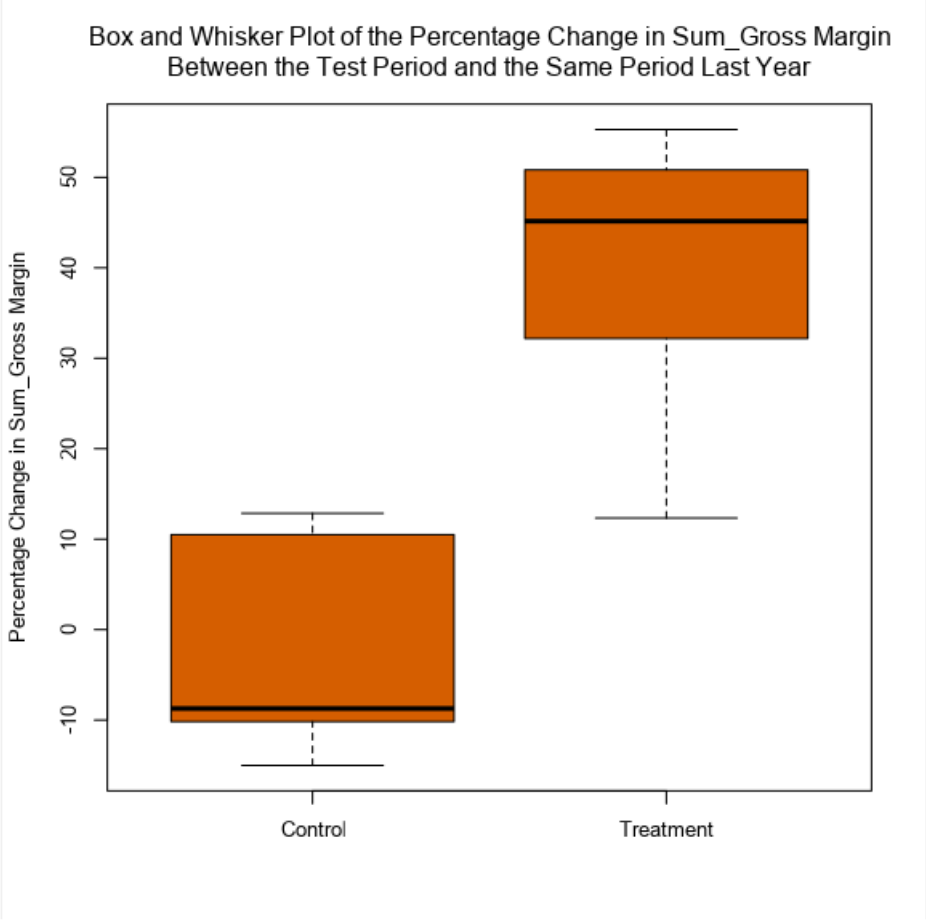
        - Line chart for west region



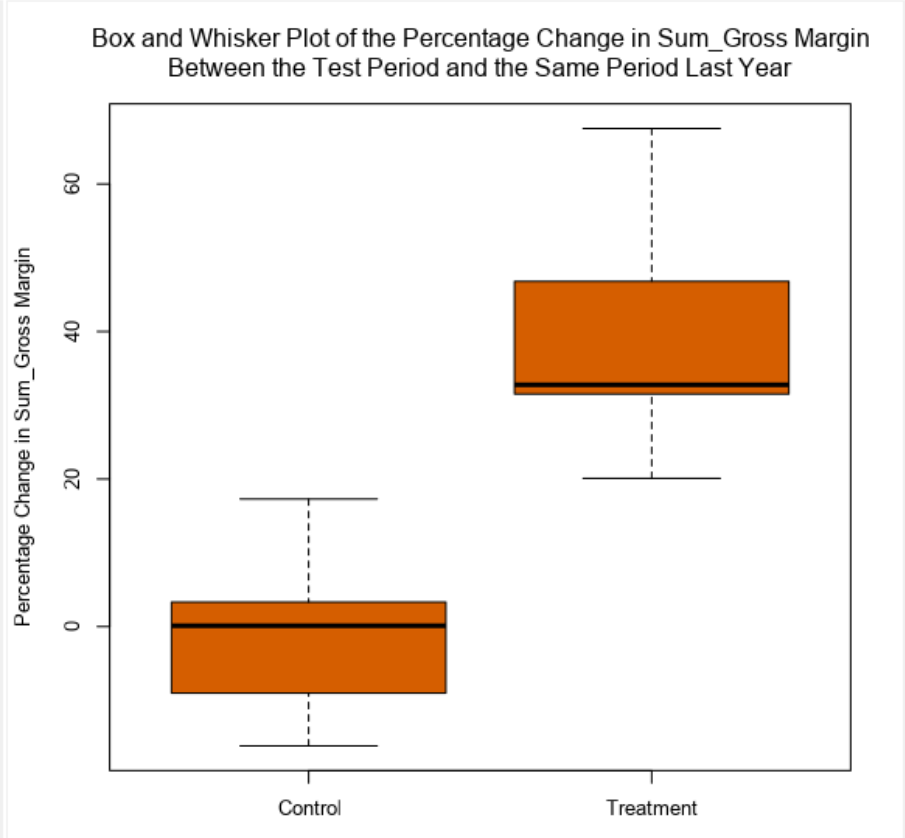
        - Line chart for central region



- Box plot for west region



- Box plot for central region



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

    - The lift for west is around 38% and the significance level is very high at 99.5%.

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
37.9%	527	99.5%

    - The lift from central is around 44% and the significance level is very high at 99.6%.

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
43.5%	836	99.6%

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

    - The overall lift from the new menu is around 41% and the significance level is very high at 100%.

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
40.7%	681	100.0%