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. (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 I will use is the profit or gross margin.
2. What is the test period?
   The test period is 2016-April-29 to 2016-July-21 for the treatment and 2015-February-06 to 2016-July-21 for control.
3. At what level (day, week, month, etc.) should the data be aggregated?
   The data should be aggregated on a weekly basis.

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...

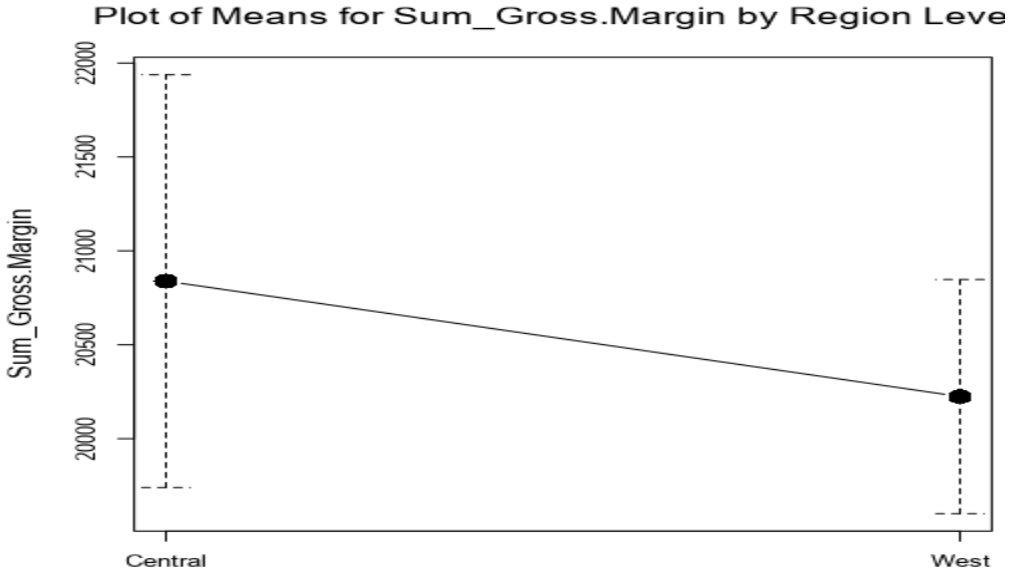
1. What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.
   The Sq_Ft, AvgMonthSales, and possibly location information from State, Region, Longitude, Latitude, or Timezone. Store Id will be used as a unique identifier.
2. What is the correlation between your each potential control variable and your performance metric?
   Upon using an Association Analysis to determine if Gross Margin is correlated with Sq Ft, AvgMonthSales, Latitude, and/or Longitude of the store, I observed correlation values of 0.08, 0.79, 0.08, and 0.08, respectively. The AvgMonthSales is highly correlated with Gross Margin.

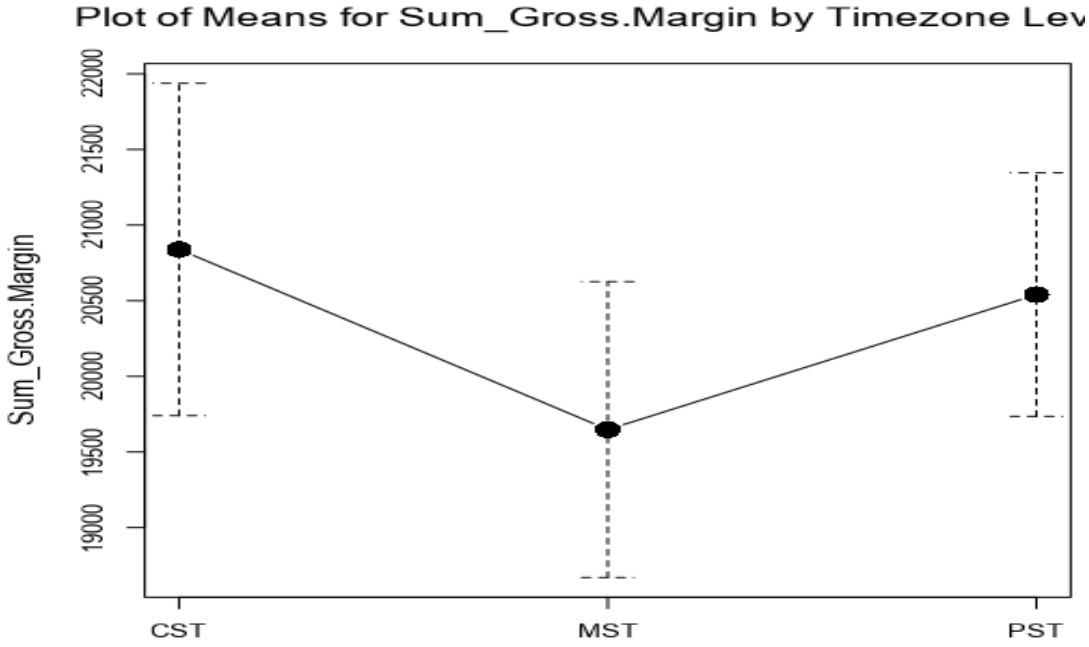
Pearson Correlation Analysis

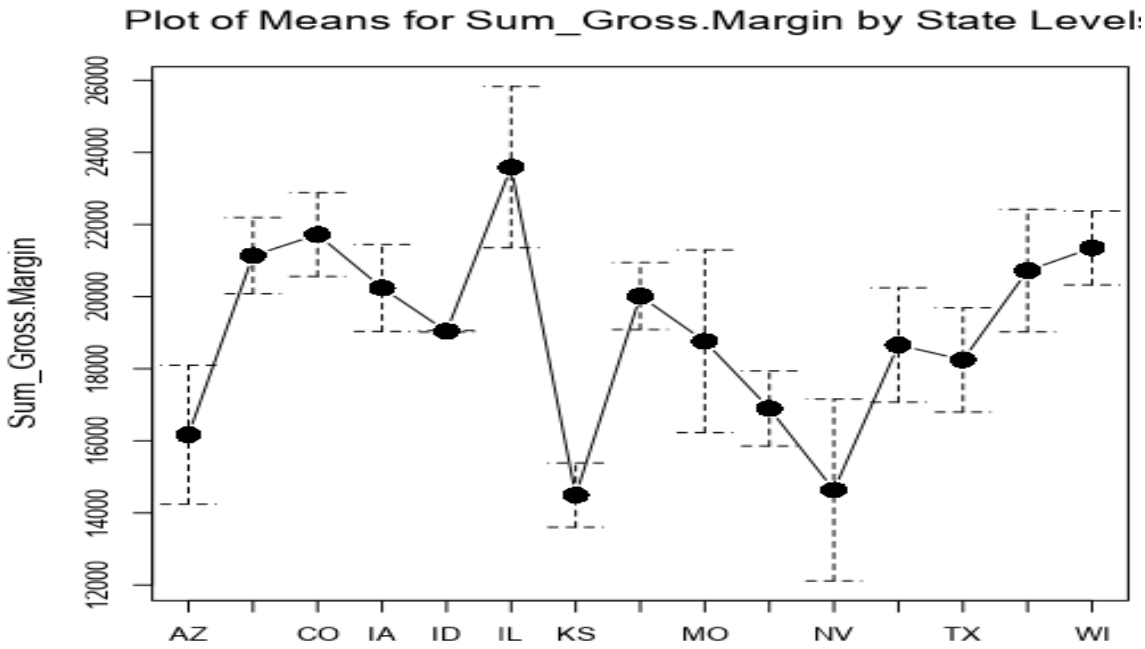
Full Correlation Matrix

	Sq_Ft	AvgMonthSales	Latitude	Longitude	Sum_Gross.Margin
Sq_Ft	1.000000	-0.052161	0.999998	0.083126	0.082776
AvgMonthSales	-0.052161	1.000000	-0.052132	-0.256769	0.788853
Latitude	0.999998	-0.052132	1.000000	0.082817	0.082564
Longitude	0.083126	-0.256769	0.082817	1.000000	0.080610
Sum_Gross.Margin	0.082776	0.788853	0.082564	0.080610	1.000000

Using Plot of Means Tools to determine if Gross Margin is correlated with State, Region, or Timezone, I did not observe any differences with Region and Timezone. All confidence intervals encompassed one another in the plots. The State variable did produce some States that were significantly different than the other states. I will say the state is correlated with Gross Margin.







However, upon looking at the rest of the questions in the document, it is apparent I should use Region instead of State. Region will be used

3. What control variables will you use to match treatment and control stores?
   I will use AvgMonthSales and Region.
4. Please fill out the table below with your treatment and control stores pairs:

Treatment Store	Control Store 1	Control Store 2
1664	8112	1542
1675	1807	1580
1696	1964	7334
1700	2014	1508
1712	8162	7434
2288	2568	3185
2293	9589	9639
2301	9238	11268
2322	2409	8817
2341	12286	9524

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 Lift Analysis and T-test below highlights why the company should roll out the updated menu. The p-value is 0.000004, providing a significance of 99.999996 (rounded

up to 100.0% in the significance level of the lift analysis). With the new updated menu, each shop can expect an increase of 44.5% in their profits per week. For the restaurants in the analysis, this is an additional 718 dollars per week in profit (gross margin).

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
44.5%	718	100.0%

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.45	-2.25
Minimum	12.34	-25.67
Maximum	67.52	19.70
Standard Deviation	16.30	12.73

Welch's Two Sample t-test(s) of Sum_Gross Margin by Test Group

Test	t-Statistic	Degrees of Freedom	p-Value
Control vs Treatment	-6.93436	14.37974	0.000006

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

Below, for the West region, the expected lift is 37.0% with a p-value of 0.006011. This gives a significance level of 99.4%, which is statistically significant.

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
37.0%	509	99.4%

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.17	1.98
Minimum	12.34	-11.22
Maximum	55.30	19.70
Standard Deviation	16.34	9.10

Welch's Two Sample t-test(s) of Sum_Gross Margin by Test Group

Test	t-Statistic	Degrees of Freedom	p-Value
Control vs Treatment	-4.49836	5.13488	0.006011

Below, for the Central region, the expected lift is 44.6% with a p-value of 0.003754. This gives a significance level of 99.6%, which is statistically significant.

Lift Analysis for Sum_Gross Margin

Lift	Expected Impact	Significance Level
44.6%	846	99.6%

Summary Statistics for Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.74	-1.92
Minimum	20.09	-16.18
Maximum	67.52	19.65
Standard Deviation	17.15	12.38

Welch's Two Sample t-test(s) of Sum_Gross Margin by Test Group

Test	t-Statistic	Degrees of Freedom	p-Value
Control vs Treatment	-4.61399	5.92702	0.003754

3. What is the lift from the new menu overall?
   In the 133 restaurants, there are 91 in the West and 42 in the Central regions. Therefore, by averaging the lift based on regional lift, $(91 * 37.0\% + 42 * 44.6\%)/133 = 39.4\%$.

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.