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?

   ANS: Gross Margin, since it represents profit in our dataset.

2. What is the test period?

   ANS: 29th April 2016 to 21st July 2016. Twelve (12) weeks.

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

   ANS: 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.

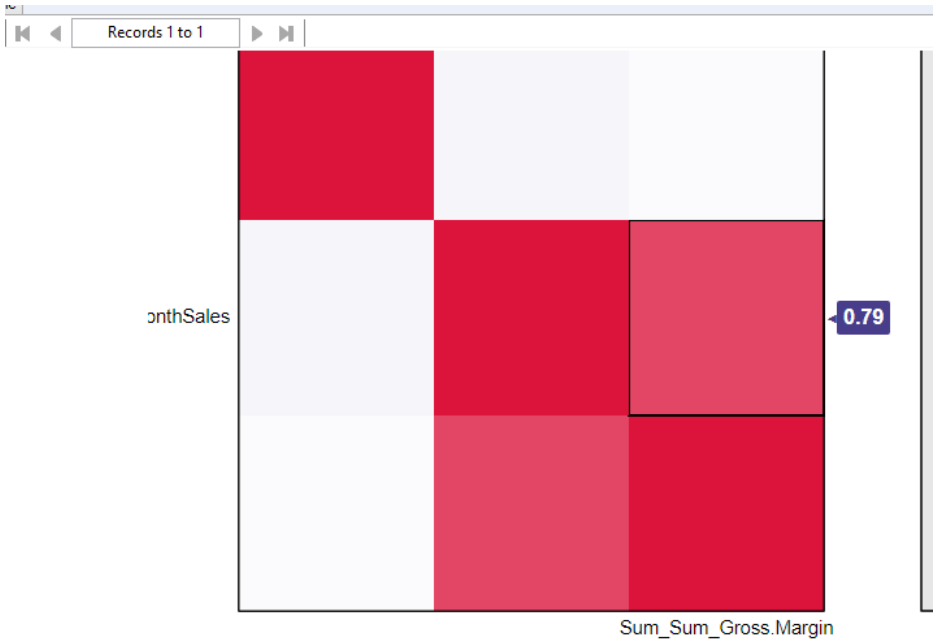
   ANS:

- Average Monthly Sales
- Size of the store(Sq_ft)
- Region

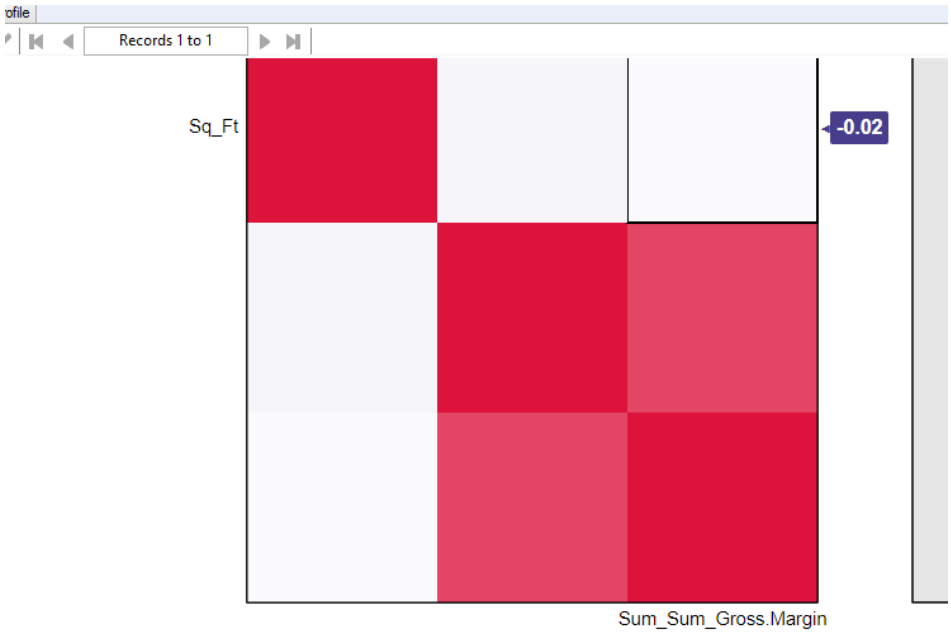
2. What is the correlation between each potential control variable and your performance metric?

   ANS: Correlation between Gross Margin and

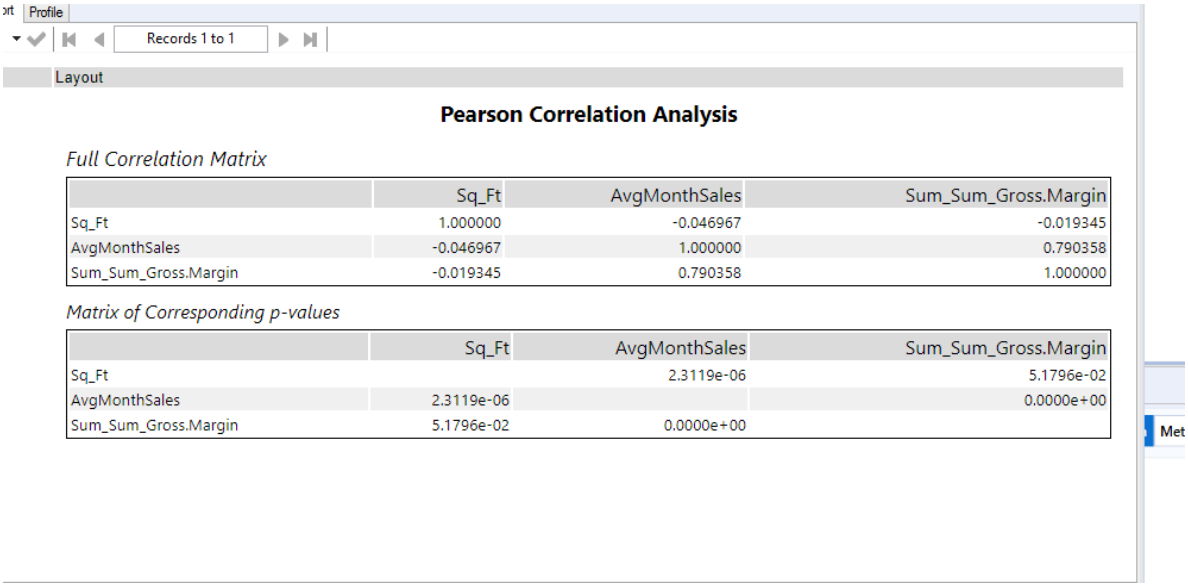
   - Average Monthly Sales ---- 0.79
   - Size of the store(Sq_ft) ----- -0.02



The left panel is an image of a correlation matrix, with blue = -1 and red = +1. Hover over pixels in the correlation matrix on the left to see the values; click to see the corresponding scatterplot on the right. The variables have been clustered based on degree of correlation, so that highly correlated variables appear adjacent to each other.



The left panel is an image of a correlation matrix, with blue = -1 and red = +1. Hover over pixels in the correlation matrix on the left to see the values; click to see the corresponding scatterplot on the right. The variables have been clustered based on degree of correlation, so that highly correlated variables appear adjacent to each other.



	Sq_Ft	AvgMonthSales	Sum_Sum_Gross.Margin
Sq_Ft	1.000000	-0.046967	-0.019345
AvgMonthSales	-0.046967	1.000000	0.790358
Sum_Sum_Gross.Margin	-0.019345	0.790358	1.000000

	Sq_Ft	AvgMonthSales	Sum_Sum_Gross.Margin
Sq_Ft		2.3119e-06	5.1796e-02
AvgMonthSales	2.3119e-06		0.0000e+00
Sum_Sum_Gross.Margin	5.1796e-02	0.0000e+00	

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

   ANS: I will use

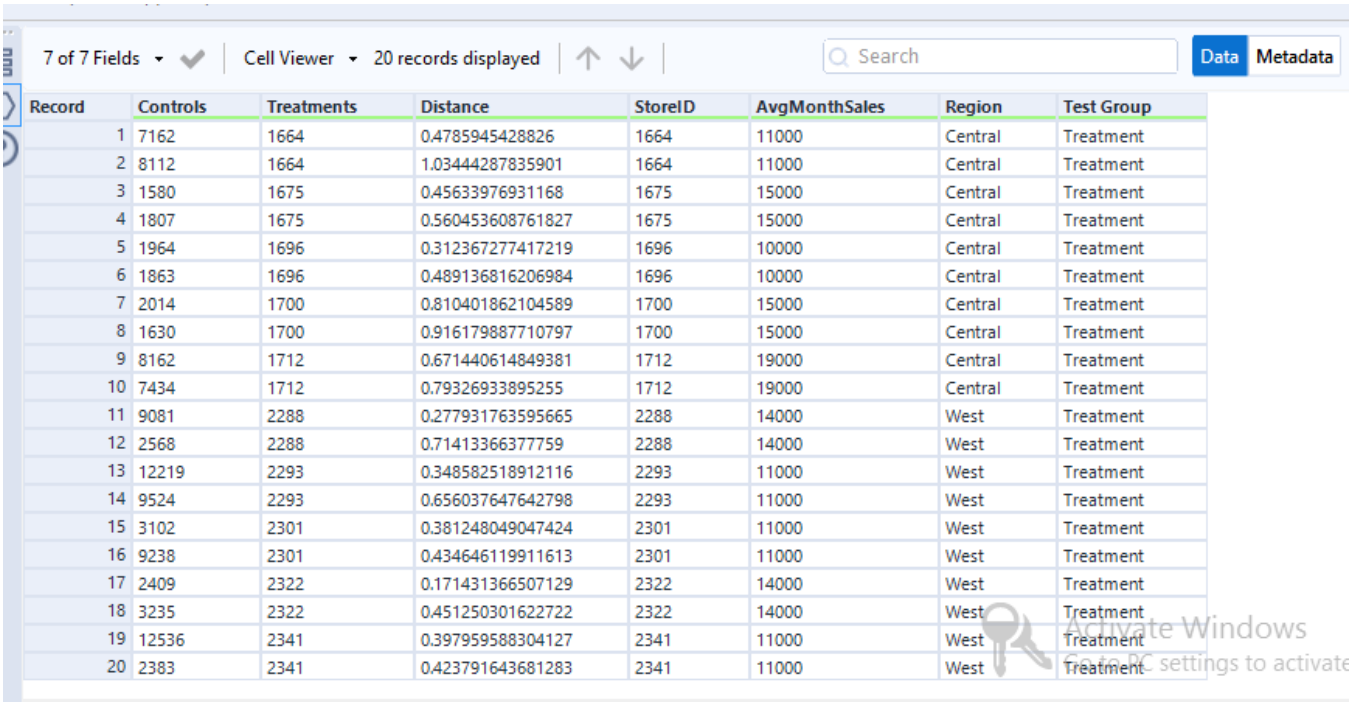
   - Average Monthly Sales because it has a high correlation to my performance metric which is Gross Margin. I wouldn't bother with Sq_ft because it has a low

correlation with my performance metric, as shown above. This is further confirmed by the low level of significance given by its P-value in the Pearson Correlation Analysis above.

- I grouped by Region as well, though I did not have a technical way of checking the correlation between region and gross margin. It makes sense logically that region might have a role to play in the changes that occur in gross margin, it would also account for stores in the same state and city, hence I decided to group by region.

4. 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



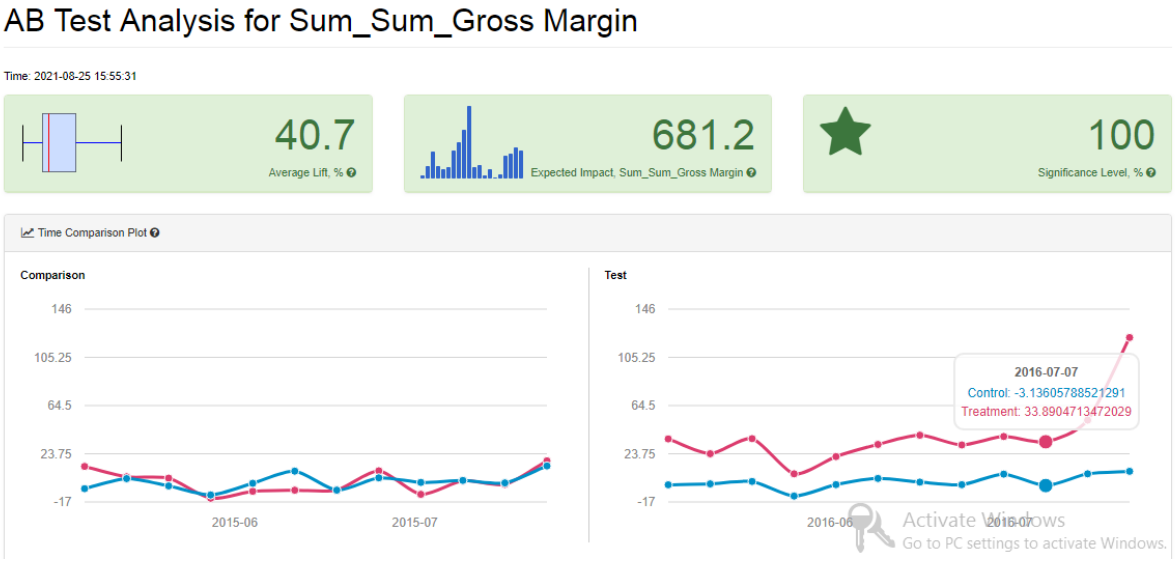
7 of 7 Fields | Cell Viewer | 20 records displayed | Search | Data | Metadata

Record	Controls	Treatments	Distance	StoreID	AvgMonthSales	Region	Test Group
1	7162	1664	0.4785945428826	1664	11000	Central	Treatment
2	8112	1664	1.03444287835901	1664	11000	Central	Treatment
3	1580	1675	0.45633976931168	1675	15000	Central	Treatment
4	1807	1675	0.560453608761827	1675	15000	Central	Treatment
5	1964	1696	0.312367277417219	1696	10000	Central	Treatment
6	1863	1696	0.489136816206984	1696	10000	Central	Treatment
7	2014	1700	0.810401862104589	1700	15000	Central	Treatment
8	1630	1700	0.916179887710797	1700	15000	Central	Treatment
9	8162	1712	0.671440614849381	1712	19000	Central	Treatment
10	7434	1712	0.79326933895255	1712	19000	Central	Treatment
11	9081	2288	0.277931763595665	2288	14000	West	Treatment
12	2568	2288	0.71413366377759	2288	14000	West	Treatment
13	12219	2293	0.348582518912116	2293	11000	West	Treatment
14	9524	2293	0.656037647642798	2293	11000	West	Treatment
15	3102	2301	0.381248049047424	2301	11000	West	Treatment
16	9238	2301	0.434646119911613	2301	11000	West	Treatment
17	2409	2322	0.171431366507129	2322	14000	West	Treatment
18	3235	2322	0.451250301622722	2322	14000	West	Treatment
19	12536	2341	0.397959588304127	2341	11000	West	Treatment
20	2383	2341	0.423791643681283	2341	11000	West	Treatment

Step 4: Analysis and Writeup

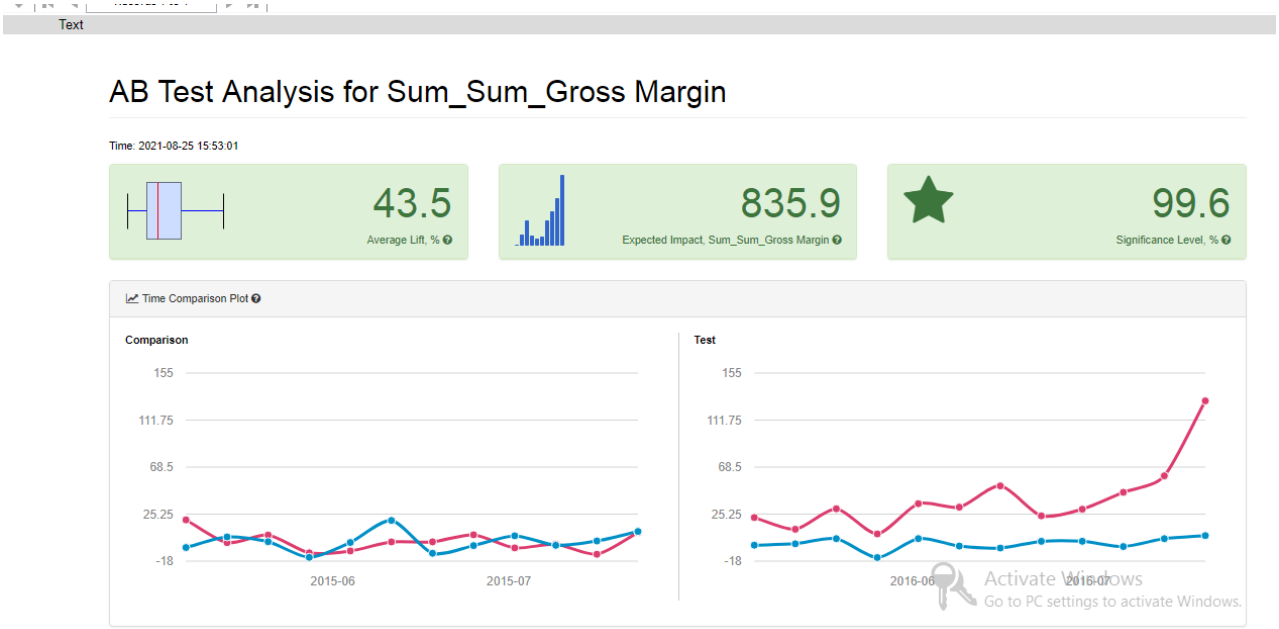
1. What is your recommendation - Should the company roll out the updated menu to all stores?

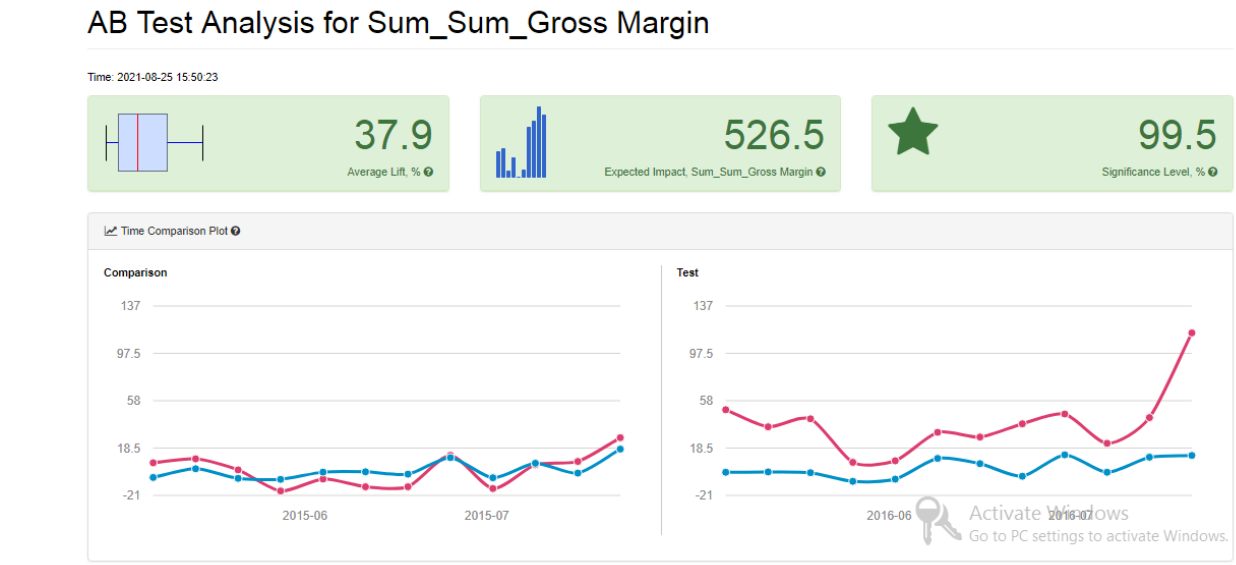
   ANS: Yes, they should. Our analysis shows a 40.7% lift with a significance of a 100%, in our performance metric which is Gross Margin. In other words, our Gross Margin will see a 40.7% increase if the new items (Gourmet Sandwiches and Limited Wine Offerings) are added to the menu. Furthermore, the problem description requires that they should be at least an 18% lift in Gross Margin and our analysis produces a lift way higher than that.



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

   ANS: The lift from the West and Central regions are 37.9% with a 99.5% significance level and 43.5% with a 99.6% significance level respectively.





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

   ANS: The lift from the new menu overall is 40.7% with a 100% significance level