Project: Analyzing a Market Test

Step 1: Plan Your Analysis

Round Roasters is an upscale coffee chain with locations in the western United States of America and the new management want to reignite growth the store, The first major growth initiative is to introduce gourmet sandwiches to the menu, along with limited wine offerings, in order to support the growth of store the management believes that a television advertising campaign is crucial to drive people into the stores with these new offering.

However, the television campaign will require a significant boost in the company's marketing budget, with an unknown return on investment (ROI). Additionally, there is concern that current customers will not buy into the new menu offerings.

In order to minimize the risk, the management decides to test in two cities with new television advertising, due to the television campaign will growth the traffic on the store, the company has to be set the performance metric for profitability in each store and the impact of profitability should be enough to justify marketing budget at least 18% increase in profit growth. In the data, profit is represented in the gross_margin variable.

The test period will run for a period of 12 weeks (2016-April-29 to 2016-July-21) where five stores in each of the test markets (two cities) offered the updated menu along with television advertising.

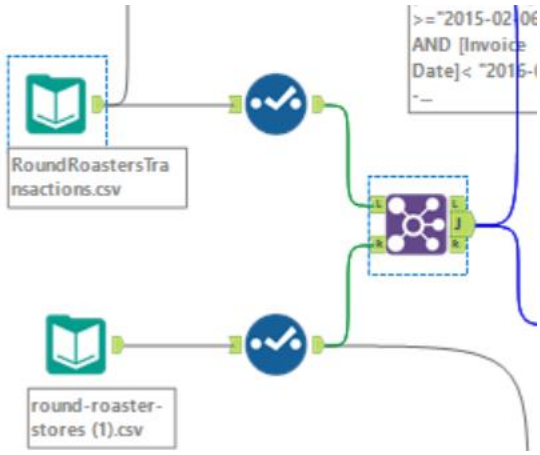
The preparing dataset for run this experiment will require transaction data for all stores which is available from 2015-January-21 to 2016-August-18, the list of all Round Roasters stores and some listing of 10 stores that will use for test markets. Since the transactional data has different granularity at time level whereas the transactional on daily and the test will have level on week, in this case the should be aggregated to be week level.

Step 2: Clean Up Your Data

Record	StoreID	Invoice Number	Invoice Date	Gross Margin	Sales
1	10018	16296643	2015-01-21	6.7365	14.97
2	10018	16296643	2015-01-21	1.1	2.75
3	10018	16297717	2015-01-21	4.185	8.37
4	10018	16297717	2015-01-21	5.98	11.96

At figure above shown Transactional data with day level, since we need week level for testing experiment so the next step is should be aggregating to week level, and due to there is no region in this transactional data firstly we need to join the data with the list of all Round Roasters to get region which is will be need for test markets.

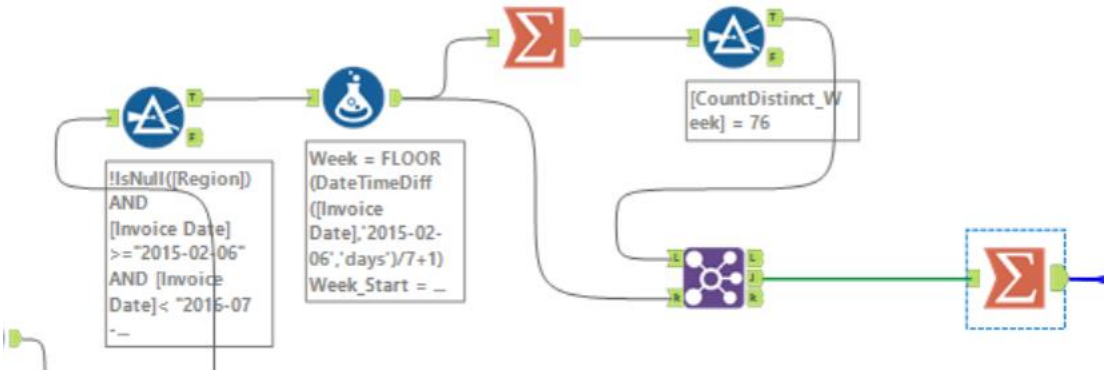
Join the Transactional data and the list of all Round Roasters to get region by tools join in alteryx :



In Join tools where J output will result like this :

Record	StoreID	Region	Invoice Number	Invoice Date	Gross Margin	Sales
1	10018	West	16296643	2015-01-21	6.7365	14.97
2	10018	West	16296643	2015-01-21	1.1	2.75
3	10018	West	16297717	2015-01-21	4.185	8.37
4	10018	West	16297717	2015-01-21	5.98	11.96

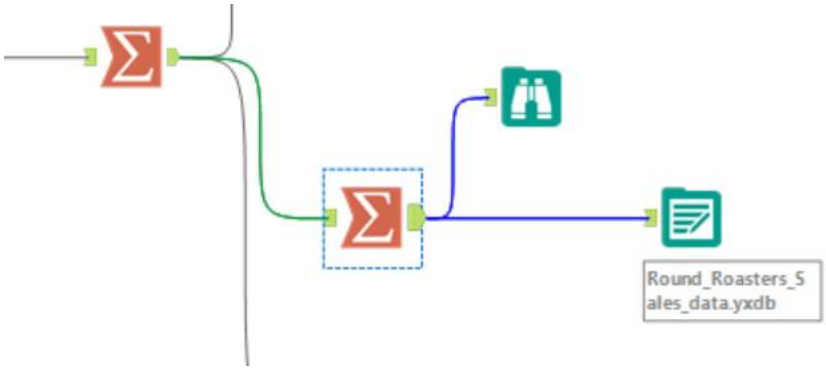
After the region columns is exist on the transactional data, now move to next step for aggregating data to the week level.



This step aggregating is quite tricky, since the test will run 12 weeks and for creating trend and seasonality will require 52 weeks data plus 12 weeks to calculate trend this means will need 76 weeks past data. Firstly filter data from 2015-February-06 to 2016-July-22 and ensure there is no missing value in region on each store, next make aggregating to week level by calculating week, week_start and week_end, after that get some calculation to count week for all store should be 76, and next summarize the total sales on each invoice number by grouping it.

Record	StoreID	Region	Invoice Number	Invoice Date	Week	Week_Start	Week_End	Sum_Gross Margin
1	10018	West	16551067	2015-02-06	1	2015-01-21	2015-01-27	27.191
2	10018	West	16555940	2015-02-06	1	2015-01-21	2015-01-27	23.6625
3	10018	West	16558368	2015-02-06	1	2015-01-21	2015-01-27	1.1
4	10018	West	16558471	2015-02-06	1	2015-01-21	2015-01-27	6.7365

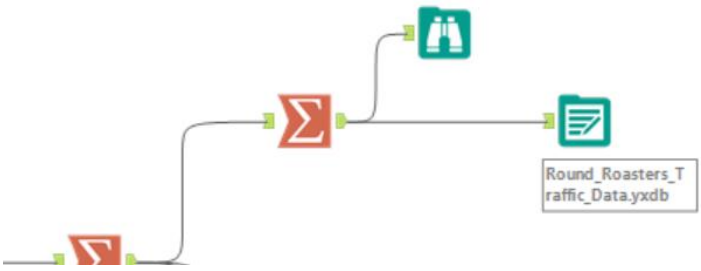
After data has the total Gross Margin one each invoice, the next step is to calculate the total Gross Margin on each week on each store.



Record	StoreID	Region	Week	Week_Start	Week_End	Sum_Sum_Gross Margin
1	10018	West	1	2015-01-21	2015-01-27	2,212.710512
2	10018	West	2	2015-01-28	2015-02-03	2,164.007015
3	10018	West	3	2015-02-04	2015-02-10	1,560.929009
4	10018	West	4	2015-02-11	2015-02-17	2,342.984016

Finally now the transactional data has week level for testing the experiment, with total Gross Margin on each week on each store. Now let just export the data in order to do AB testing analysis soon,

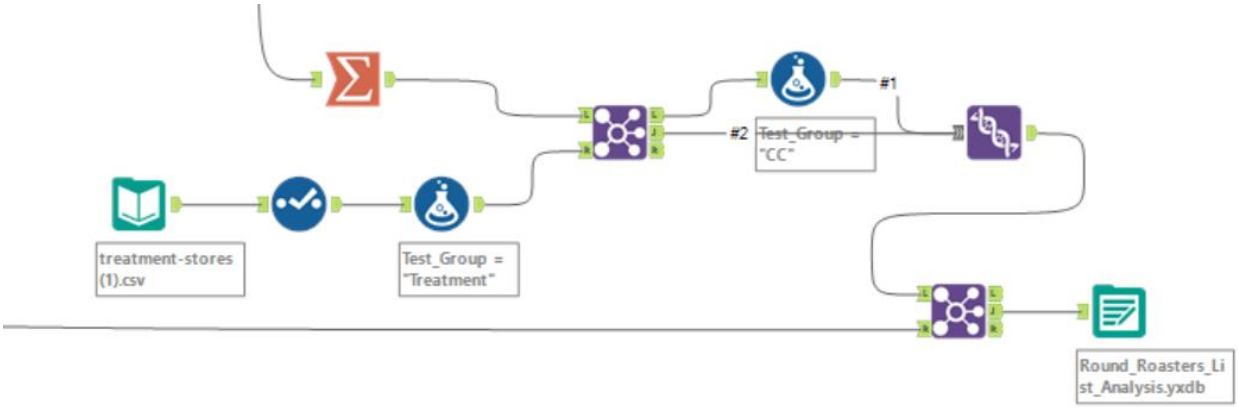
Since we want to calculate the trend and seasonality in each store this means weed to calculate the traffic in each store on each week by summarize the total invoice in each week.



The result will comes like this bellow, and save the dataset in order to decide the trend and seasonality in each store.

Record	StoreID	Region	Week	Week_Start	Week_End	Total Invoice
1	10018	West	1	2015-01-21	2015-01-27	308
2	10018	West	2	2015-01-28	2015-02-03	288
3	10018	West	3	2015-02-04	2015-02-10	204
4	10018	West	4	2015-02-11	2015-02-17	320

Lastly we need to labeled all of the store to be Treatment or CC (Control), grouping the transactional data by store and region, then open the treatment-store dataset with adding new column ‘Test Group’ with value Treatment.

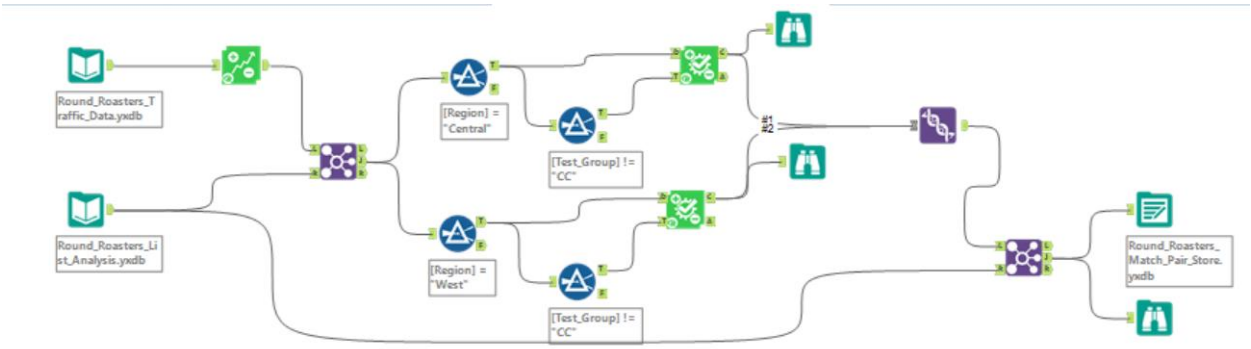


Next join the two dataset by store ID, the first one output J will be appears only from treatment-store dataset, the second one the output L will be appears the rest of store that non treatment group, since the output L doesn’t have column ‘Test_Group’ so we can take formula for that, now lets join the two output which is J and L so we could know the all of store have labeled now, and just export this dataset to be store list for analysis soon.

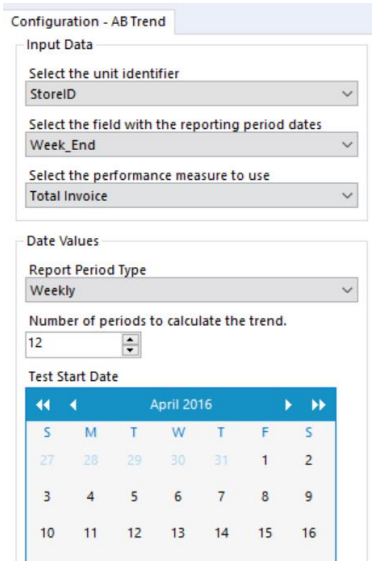
Record	StoreID	Region	AvgMonthSales	Test_Group
44	1630	Central	17,000	CC
45	1662	Central	11,000	CC
46	1664	Central	11,000	Treatment
47	1675	Central	15,000	Treatment
48	1696	Central	10,000	Treatment

There is an additional column like AvgMonthSales in this dataset, and the feature in so potential to become another control variable.

Step 3: Match Treatment and Control Units



In order to calculate the trend and seasonality, firstly bring down the traffic dataset of Round Roasters and connect AB Trend.



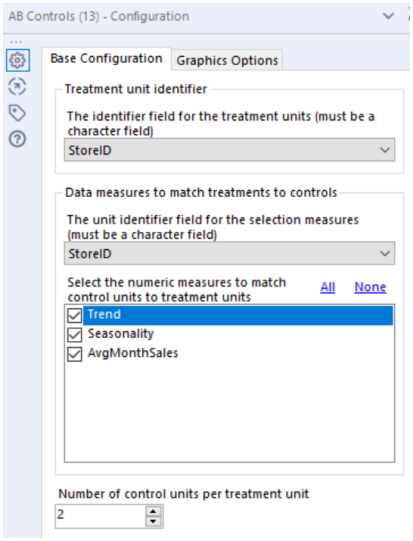
Select the unit identifier to storeID because the store id is unit that we want to measure each trend and seasonality, next select week_end as reporting period dates then performance metric for calculating trend and seasonality by the Total Invoice in each store in each week, since the test experiment will run for 12 weeks, so the level of report period should on the same level this means we choose weekly for period type and the number of periods is 12 too. Test start date should be 29 April 2016.

After calculate the trend and seasonality on each store based on traffic on weekly, next step is join the data from the output of AB Trend to store list with label dataset.

The result will looks like this :

Record	StoreID	Region	Trend	Seasonality	Test_Group	AvgMonthSales
43	1580	Central	0.115568	-0.002121	CC	14,000
44	1630	Central	-0.326265	0.005148	CC	17,000
45	1662	Central	-0.248433	0.010416	CC	11,000
46	1664	Central	0.433152	-0.012974	Treatment	11,000

Now the data are ready to do match and pair analysis, and since the test require two market strategy in different city so we need to separate by region Central and West before doing match and pair analysis, and also we need to filter the test_group store by ‘Treament’ and ‘CC’ in order to know which the treatment store.



Setting up the AB Control by selection measure to Store ID and select the numeric measure to match the control unit and treatment units by 'Trend' , 'Seasonality' and AvgMonthSales which has strong correlation to sum of gross margin, and last assign 2 control unit per treatment unit.

Finally the result will look like table bellow, it will generate the paired store with each treatment unit store has two controls unit store in same region.

Record	Controls	Treatments	Distance	Region	Test_Group
9	8162	1712	0.798071	Central	Treatment
10	7434	1712	0.89303	Central	Treatment
11	9081	2288	0.278834	West	Treatment
12	3185	2288	0.471271	West	Treatment

Apart from trend and seasonality...

1. What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.
    - Sq_ft
    - AvgMonthSales
2. What is the correlation between your each potential control variable and your performance metric?

Pearson Correlation Analysis

Full Correlation Matrix

	Sq_Ft	AvgMonthSales	Sum_Gross.Margin
Sq_Ft	1.000000	-0.046967	-0.020353
AvgMonthSales	-0.046967	1.000000	0.988219
Sum_Gross.Margin	-0.020353	0.988219	1.000000

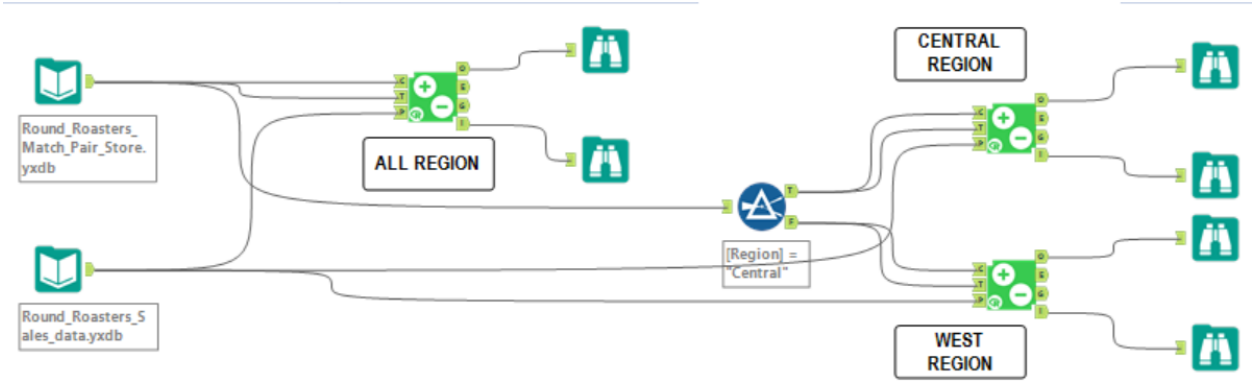
The Table Person Correlation shown that AvgMonthSales has signigicantly strong correlated to sum of Gross Margin with r about 0.98 and the other one which is Sq_ft doesn't have strong correlated to any numerical feature especially with sum of Gross margin, it indicated that AvgMonthSales could be become control variables for analysis AB Match Pair store soon.

3. What control variables will you use to match treatment and control stores?
    - AvgMonthSales due to have a strong correlation to Groos Margin s

4. Please fill out the table below with your treatment and control stores pairs:

Treatment Store	Control Store 1	Control Store 2
1664	7162	1542
1675	1580	7284
1696	1863	1964
1700	2014	7212
1712	8162	7434
2288	9081	3185
2293	8817	1221
2301	3102	9238
2322	9388	2568
2341	2572	1228

Step 4: Analysis and Writeup



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

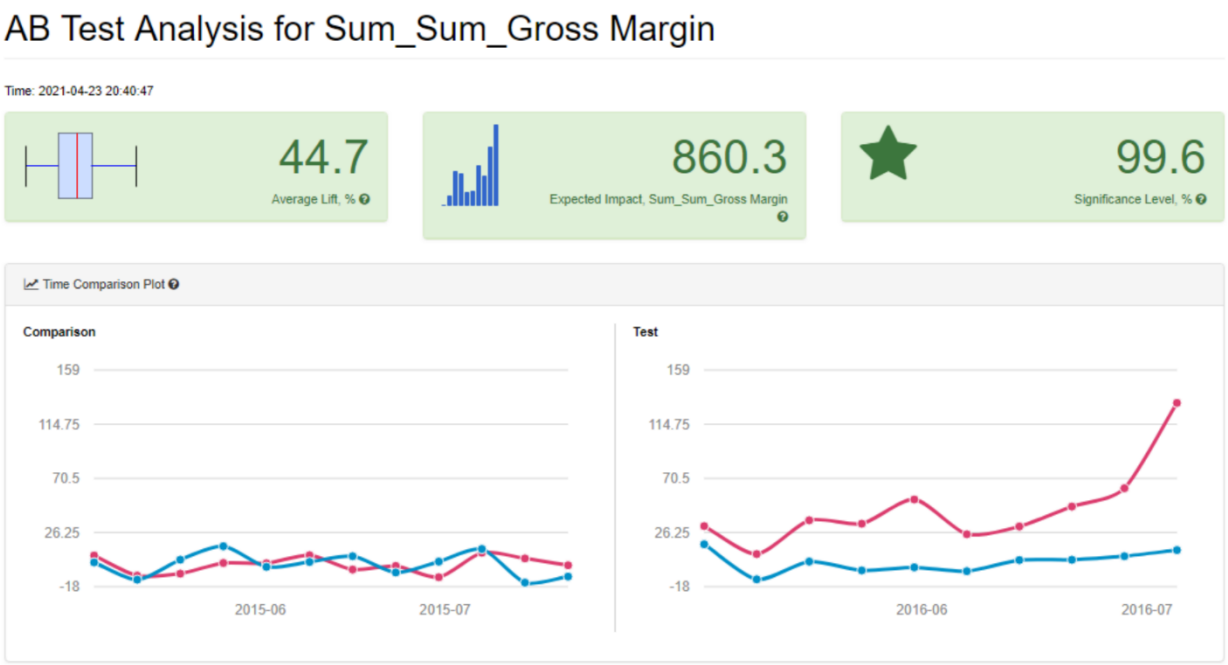
1. What is your recommendation - Should the company roll out the updated menu to all stores?

   The requirement of increase in profit growth compared to the comparative period is at least minimum on 18%, After doing the AB testing on five stores in each of the test markets with a period of 12 weeks (2016-April-29 to 2016-July-21) (two cities) The performance on all treatment stores have profitability on the treatment units over control units is about 38% increased, and also by observation period (in weeks) the treatment units store shows that average impacted gross margin increased to about 648.1 per period with significant level 100%.

   This means by the result shown, company should roll out the new menu which is gourmet sandwiches in all store to reignite the growth in each store.

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

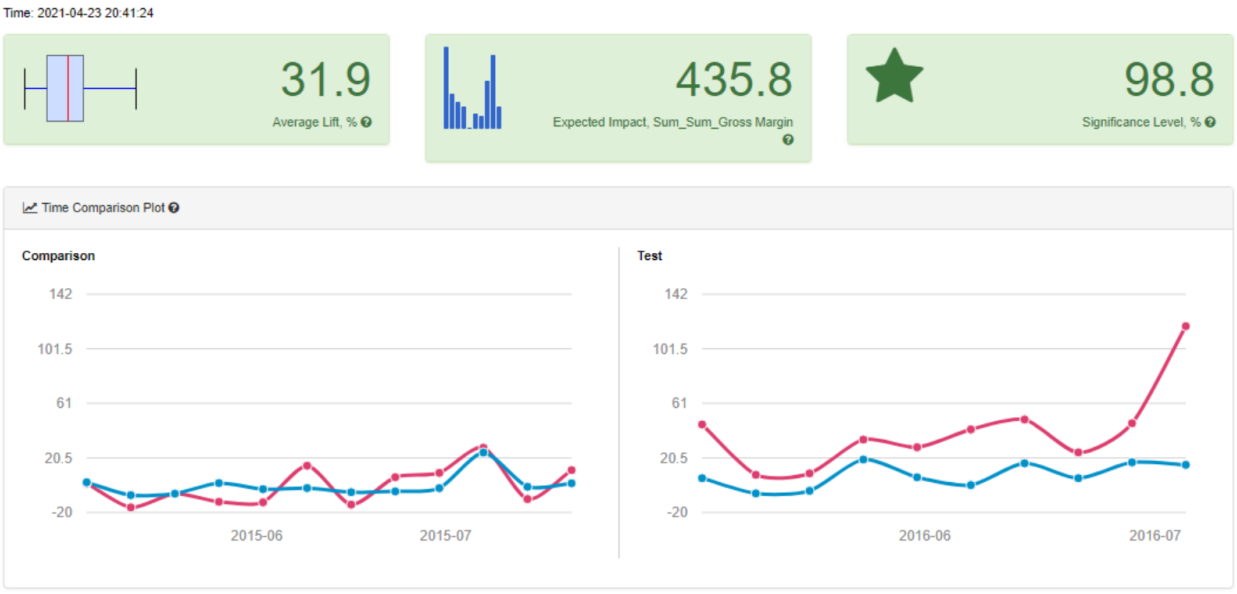
- Reporting Treatment Store at Central Region



As shown the visualization above about the performance of treatment store at Central Region with new menu offered, we could interpreted that the impacted of gross margin could lifting up about 44.7% on average and with significance level by 99.6%.

Looking at Time Comparison Plot the Test period time compared to Last year period(Comparison), the performance on treatment store has growth significantly after the new menu launch in that store.

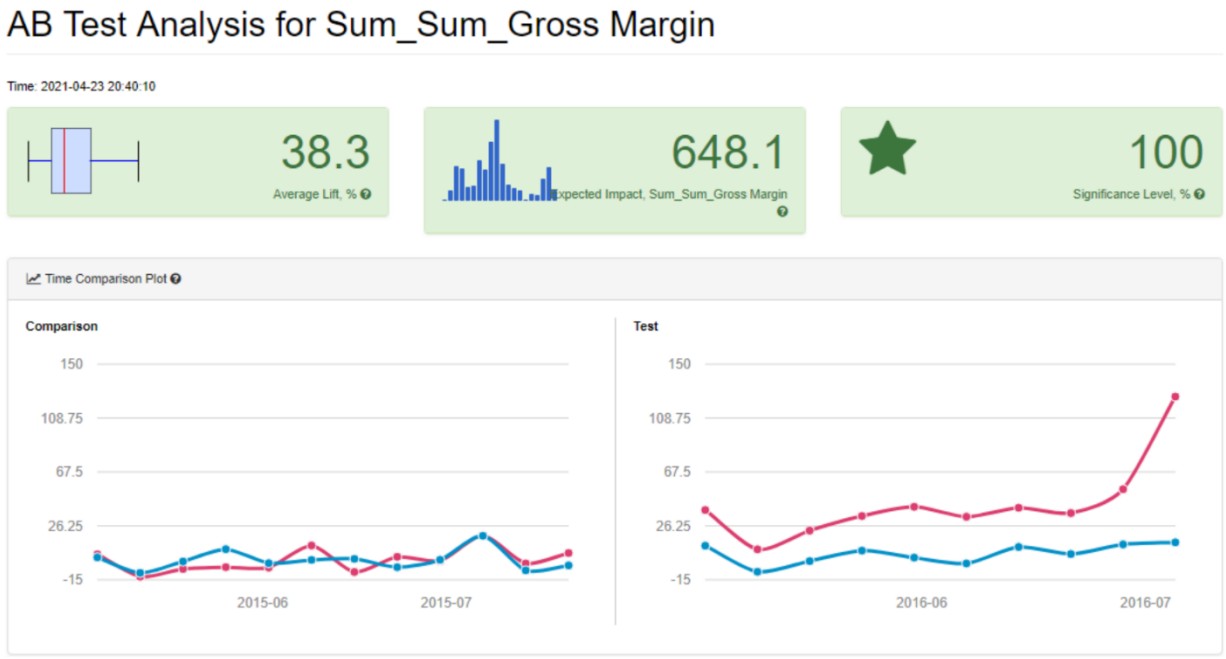
- Reporting Treatment Store at West Region



As shown the visualization above about the performance of treatment store at West Region with new menu offered, it still the same result that in West Region the impacted of new offered in this region could lifting up the gross margin about 31.9% on average and with significance level by 98.8%, less lifting up gross margin than Central Region but the lifting up gross margin is enough to match the requirement increase in profit growth compared to the comparative period is at least minimum on 18%.

Looking at Time Comparison Plot the Test period time compared to Last year period(Comparison), the performance on treatment store has growth significantly after the new menu launch in that store.

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



The new menu on overall performance within a period of 12 weeks (2016-April-29 to 2016-July-21) on all of treatment store shows that it could lifting up the average gross margin by 38%, and could increase average gross margin 648.1 dollar on each one period time. The Dot plot on above tells almost all of treatment store is outperformed by control store.