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?
   I'll be using the 'Gross Margin' value to evaluate the results of the test. Since, it is the direct representation of profit. Profit is the only factor the management is concerned about in this scenario
2. What is the test period?
   The test was run for 12 weeks (2016-April-29 to 2016-July-21). The test period is more than enough to capture the complete business performance, since it involves people who visit the shop in daily basis, weekly or on weekend basis and those who visit on a monthly basis like on family outing kind of occasions.
3. At what level (day, week, month, etc.) should the data be aggregated?
   The data should be aggregated in a weekly level, since it fairly captures all the customer types in a week.

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.

I have used multiple files to perform the analysis. Because of personal preferences and ease of operation and editing. I will include all the screenshots of my remaining flow diagrams in the end of this file

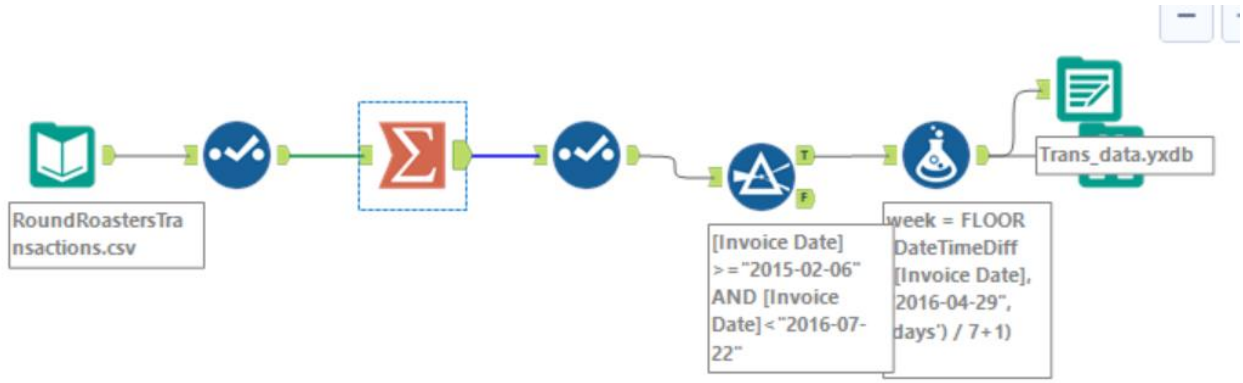


Figure 1: Cleaning the transaction data

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.

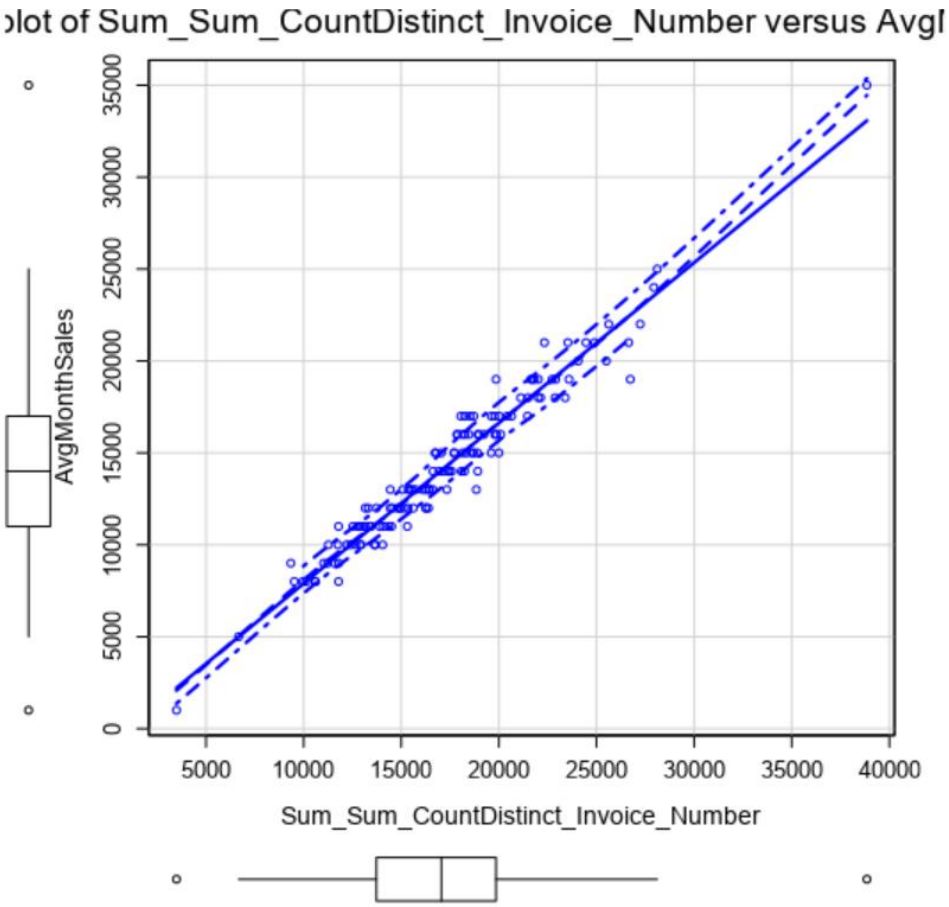
   Region, Sq_ft, Average Monthly Sales may be the possible control variables from the RoundRoastersStore file.
2. What is the correlation between your each potential control variable and your performance metric?

   Full Correlation Matrix

	Sum_Sum_Sum_Gross.Margin	Sq_Ft	AvgMonthSales	Sum_Sum_CountDistinct_Invoice.Number
Sum_Sum_Sum_Gross.Margin	1.0000000	-0.0242551	0.9909824	0.9850953
Sq_Ft	-0.0242551	1.0000000	-0.0469674	0.0044595
AvgMonthSales	0.9909824	-0.0469674	1.0000000	0.9746709
Sum_Sum_CountDistinct_Invoice.Number	0.9850953	0.0044595	0.9746709	1.0000000

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

   I will use average monthly sales along with the trend and seasonality that was determined. Invoice numbers not considered because there is a linear relation with average monthly sales. Hence number of invoices and average monthly sales are duplicates. Hence one of them is not considered.



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

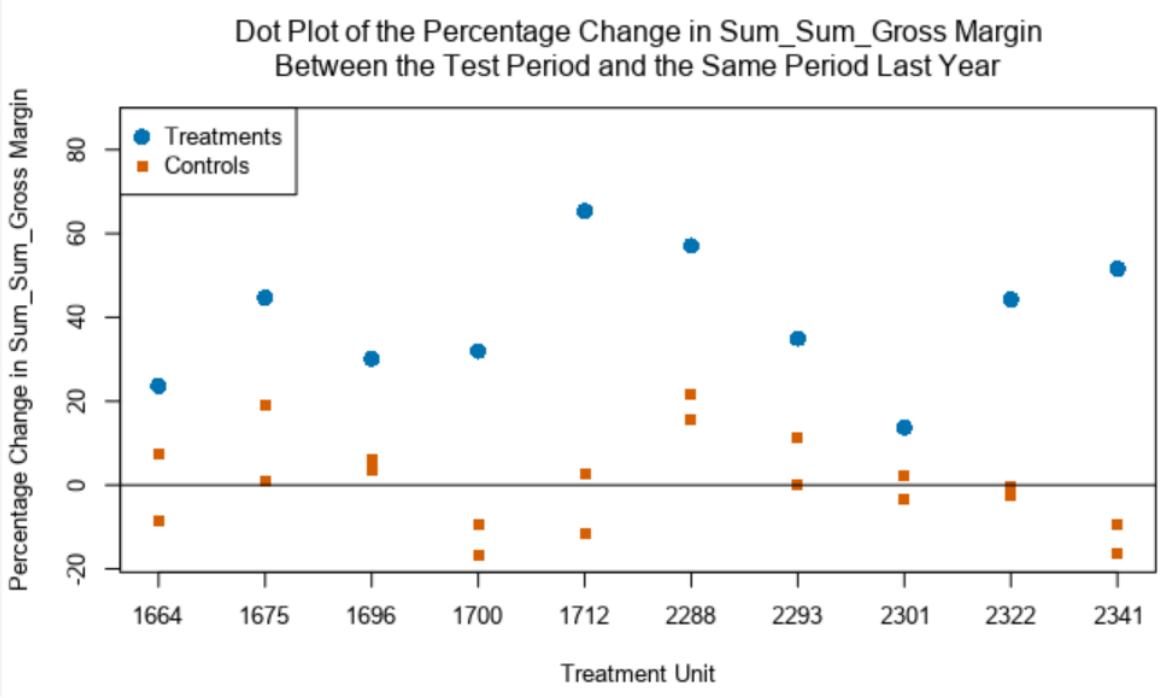
Treatment Store	Control Store 1	Control Store 2
2288	9081	2568
2293	12219	9524
2301	3102	9238
2322	2409	3235
2341	12536	2383
1664	7162	8112
1675	1580	1807
1696	1964	1863
1700	2014	1630
1712	8162	7434

Step 4: Analysis and Writeup

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?

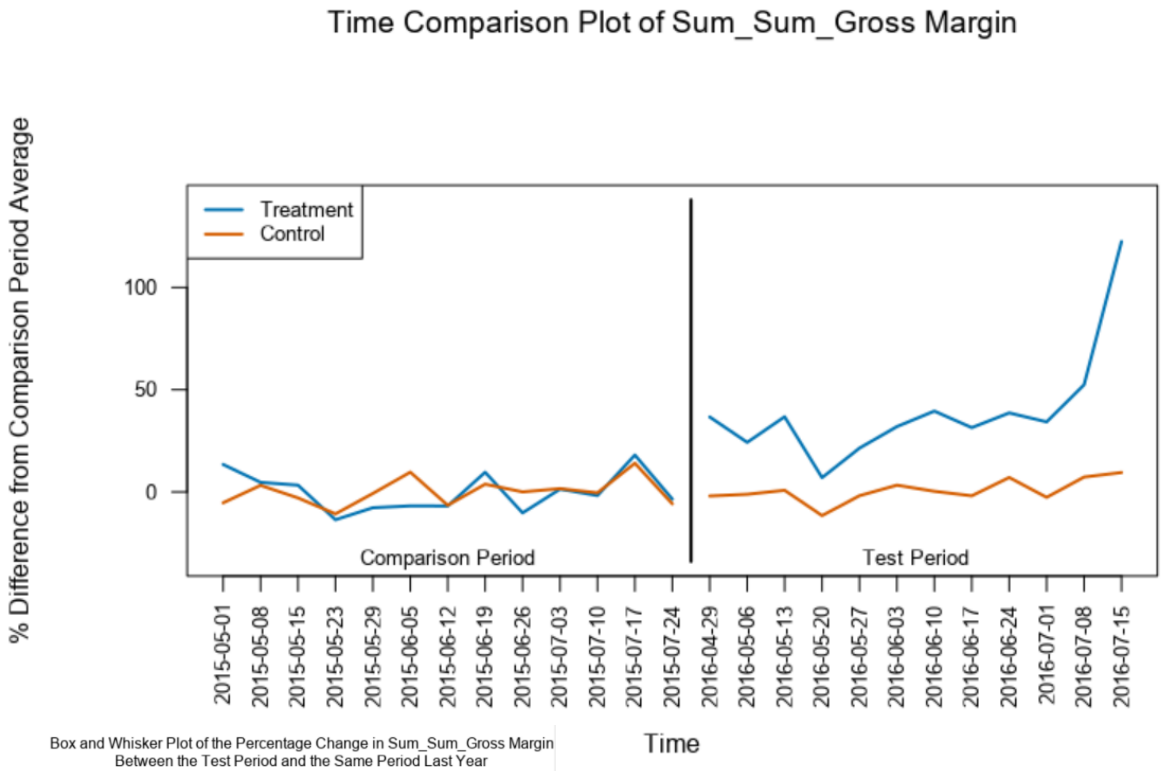
   Yes, From the analysis, it is recommended to the company to roll out the updated menu to all stores. There is a considerable increase of 40% in the profits i.e. the gross margin due to the adoption of new menu. The Dot plot and the line graphs represent that in every store the positive change is observed due to new menu. Also, the marketing cost of 18% is well justified by the increase in profits. Thus, it is profitable and feasible to implement the change across all the stores.

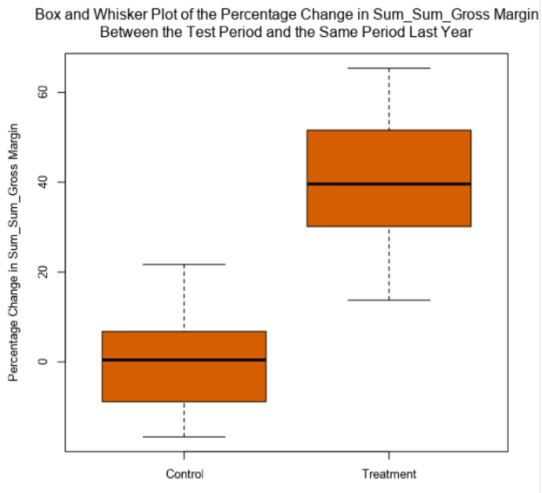


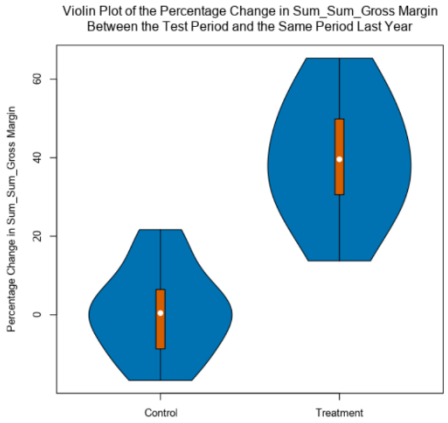
Test Summary

The average percentage change in Sum_Sum_Gross Margin was 39.7% for the treatment units in the test period relative to the comparison period. This same measure was 0.6% for the control units, with the difference between the treatment and control units being 39.1%, which is highly statistically significant. More detailed summary statistics for the treatment and control groups are contained in the first table (which immediately follows), while the details of the hypothesis test of a significant difference in the mean average percentage change in Sum_Sum_Gross Margin is contained in a table at the end of this report.

A comparison of the treatment-control pairs indicates an average lift in Sum_Sum_Gross Margin for the treatment units over the control units of 40.4%, which results in an expected impact of 672 on Sum_Sum_Gross Margin, with 100.0% of the treatment-control pairs exhibiting a positive lift for the treatment units.







From the Time Comparison Plot, we observe that the trend in the gross margin value is increasing as the time passes, which indicate that the adoption of new menu will increase the gross margin in the future more than the experimental values. This shows the possibility of long-term profit generation and the increase in profit is not just due to the initial hype of marketing.

From the Box and whisker plot, we see how the percentage raise is distributed among the members of the treatment and the control groups. We see that the lowest raise in the treatment group is about 18% and the highest raise in profit is close to 65%. Hence The new menu can be implemented and no stores are expected to give losses, even the least possible raise is considerably closer or higher than the marketing costs.

The Violin plot shows the increase in profit is more around the 40% value in most of the stores and there seems to be no outlier results to skew the overall result.

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

The Lift from new menu for West region is:

Lift Analysis for Sum_Sum_Gross Margin

Lift	Expected Impact	Significance Level
39.1%	538	99.6%

Summary Statistics for Sum_Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	40.32	1.88
Minimum	13.74	-16.32
Maximum	57.06	21.69
Standard Deviation	16.04	11.49

For Central Region:

Lift Analysis for Sum_Sum_Gross Margin

Lift	Expected Impact	Significance Level
41.6%	806	99.7%

Summary Statistics for Sum_Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.16	-0.67
Minimum	23.65	-16.71
Maximum	65.36	19.01
Standard Deviation	15.57	10.76

We see that the Lift percentage in both the regions are almost similar. Hence, both the regions can be considered for the adoption of the new menu.

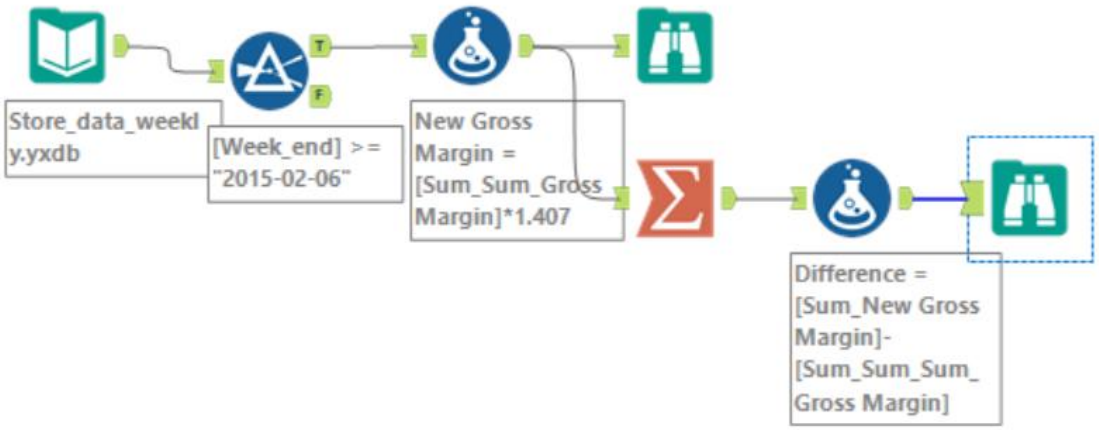
3. What is the lift from the new menu overall?

Lift Analysis for Sum_Sum_Gross Margin

Lift	Expected Impact	Significance Level
40.4%	672	100.0%

Summary Statistics for Sum_Sum_Gross Margin by Test Group

Statistic	Treatment	Control
Average	39.74	0.61
Minimum	13.74	-16.71
Maximum	65.36	21.69
Standard Deviation	15.40	10.91



Record	Sum_Sum_Sum_Gross Margin	Sum_New Gross Margin	Difference
1	16825466.973	23673432.031011	6847965.058011

We see an increase of $ 6,847,965 in gross margin with the implementation of the new menu

Flow Diagrams:

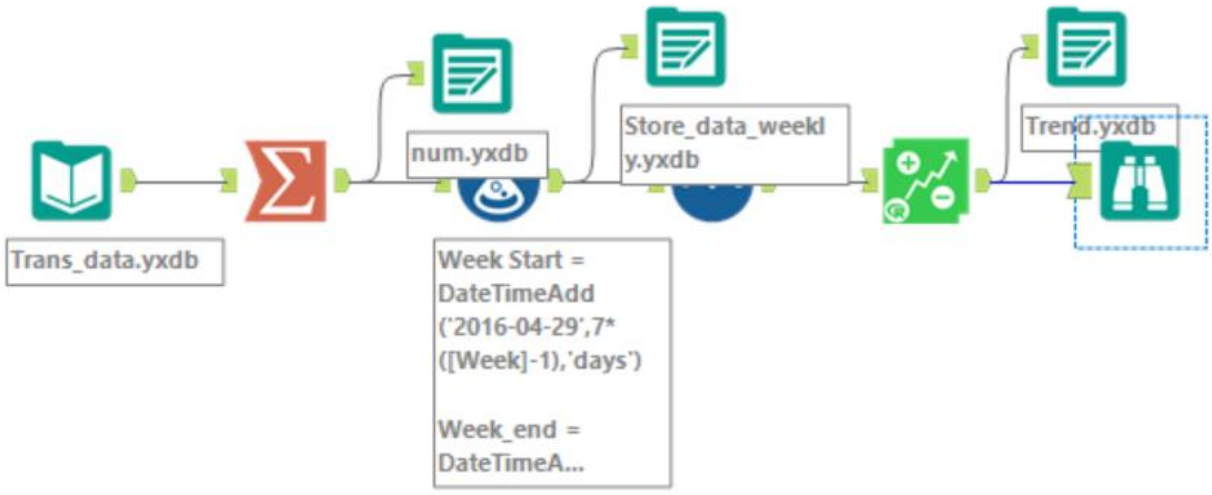


Figure 2: Obtaining Trend and Seasonality

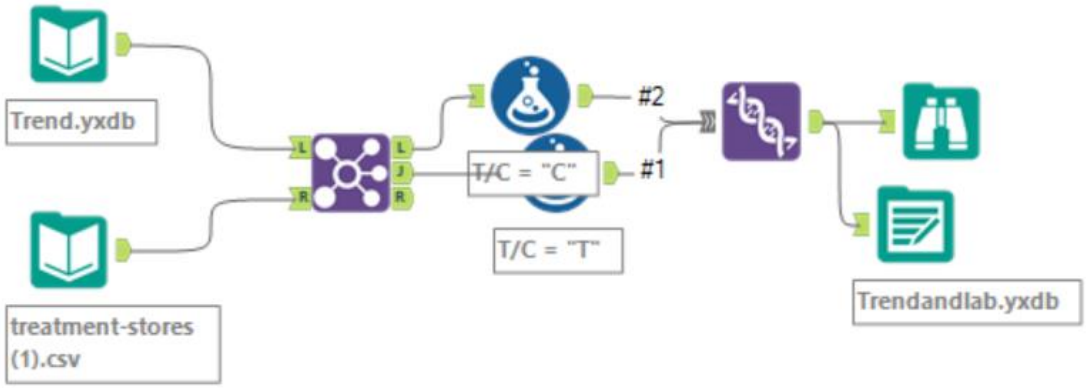


Figure 3: Labelling the stores as Treatment

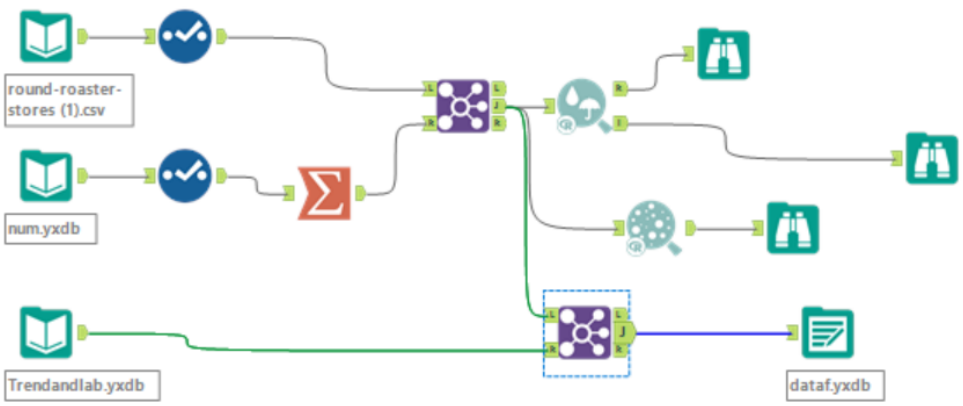


Figure 4: For selecting variables and blending the data

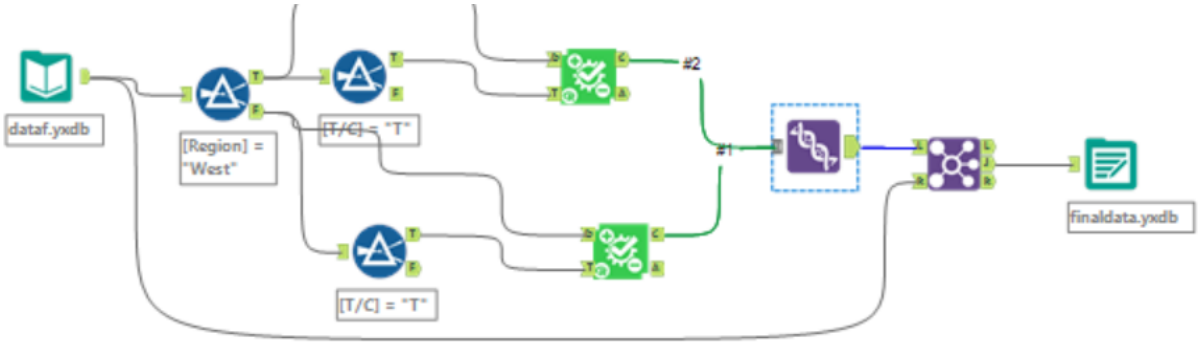


Figure 5: Region specific linking of treatment and control units

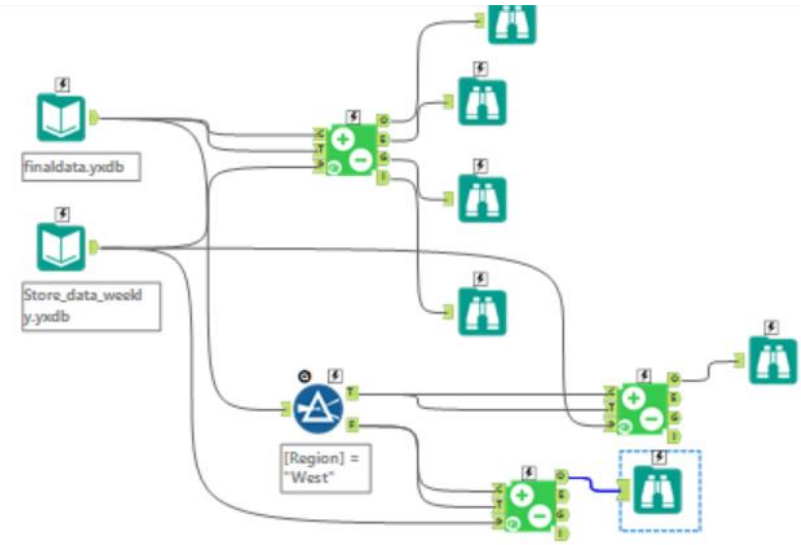


Figure 6: Performing AB Analysis