NAME: Dineshkumar Anbalagan

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?*
2. *What is the test period?*
3. *At what level (day, week, month, etc.) should the data be aggregated?*

" " "

The main purpose of this project is to AB test a new menu launch in order to determine whether or not to implement the new menu at all the round roaster stores across United States. So, before starting our analysis, we have to predefine a set of variables and the performance metric which are going to use in the entire scope of this project. For this project, we are going to use **sum of gross margin** as our performance metric to determine whether to implement the new menu or not. Furthermore, the test period of this project is for **12 weeks** starting from **2016-April-29 to 2016-July-21** and since we are gonna use week level analysis, the dataset should be aggregated at **week level.**

STEP 2 : CLEAN UP YOUR DATA

" " "

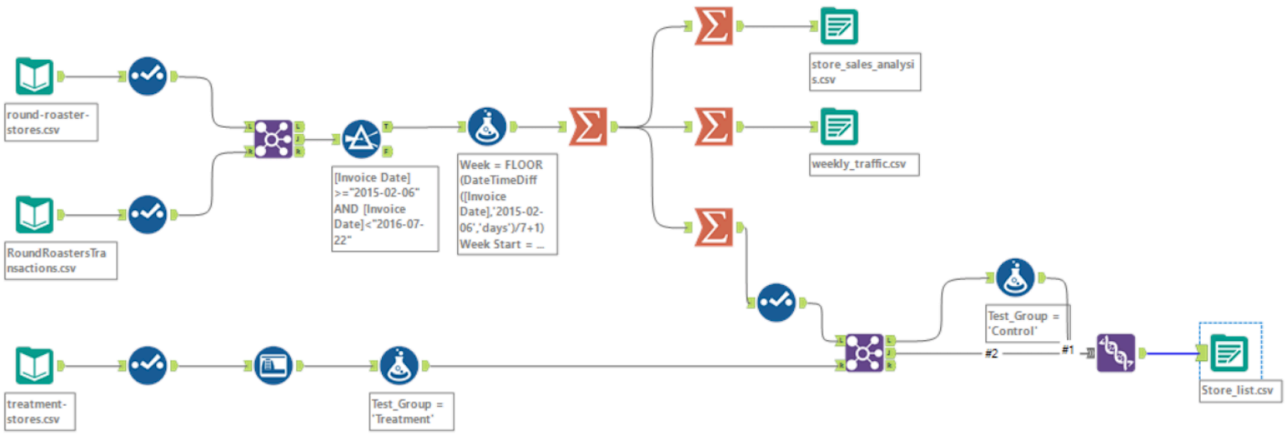
In this step, you should prepare the data for steps 3 and 4. You should aggregate the transaction data to the appropriate level and filter on the appropriate data ranges. You can assume that there is no missing, incomplete, duplicate, or dirty data. You're ready to move on to the next step when you have weekly transaction data for all stores.

" " "

As mentioned in the above context, this is the step where we cleanse, format and blend our datasets to create a new dataset from the existing ones which is gonna be used to perform AB test in the upcoming steps. The workflow which I have designed to cleanse and format the dataset is shown in the image attached below.

1. **Round_roaster_stores** and **round_roaster_transactions** datasets are inner joined by keeping **store id** as a primary and foreign key. Then the data is filtered to keep it within the range from**"2015-02-06"** to**"2016-07-22"** while introducing a new column named **weeks** to keep track of the date's week number. Then the columns are aggregated with respect to **region, storeID, average month sales, invoice number, invoice date, week, week start, week end, gross margin, sales.** All those aggregated columns are further aggregated to get the datasets such as **store_sales_analysis** and **weekly_traffic**.

2. **Treatment_stores** dataset is then merged with one of the aggregated dataset from the above workflow by introducing a new column called **Test_group,** which contains a string **"Treatment"** for all the records from treatment_stores dataset. Meanwhile all the records except those records in treatment stores are marked with the string **"Control".** This results in a dataset called **store_list.**



STEP 3 : MATCH TREATMENT AND CONTROL UNITS

" " "

Apart from trend and seasonality…

1. *What control variables should be considered? Note: Only consider variables in the RoundRoastersStore file.*
2. *What is the correlation between your each potential control variable and your performance metric?*
3. *What control variables will you use to match treatment and control stores?*
4. *Please fill out the table below with your treatment and control stores pairs*

" " "

Variables like **square feet** and **average month sales** appears to be a better control variables here. But, considering the correlation factor (in the below image), the square feet variable cannot be used as a control variable, since it has **lesser correlation (-0.019)** with performance metric variable. Furthermore, variables such as **average month sales, trend and seasonality** will be used to match treatment and control stores.

Pearson Correlation Analysis

Full Correlation Matrix

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

5 of 5 Fields | Cell Viewer | 20 records displayed

Record	Controls	Treatments	Distance	AvgMonthSales	Region
1	1964	1664	0.294607	11,000	Central
2	7162	1664	0.34613	11,000	Central
3	7284	1675	0.663814	15,000	Central
4	2214	1675	0.703483	15,000	Central
5	1863	1696	0.413188	10,000	Central
6	7334	1696	0.661492	10,000	Central
7	7037	1700	0.918504	15,000	Central
8	2014	1700	1.007526	15,000	Central
9	8162	1712	0.487861	19,000	Central
10	7434	1712	0.612087	19,000	Central
11	2568	2288	0.407287	14,000	West
12	9081	2288	0.446997	14,000	West
13	12686	2293	0.6895	11,000	West
14	9639	2293	0.730351	11,000	West
15	12536	2301	0.376984	11,000	West
16	9238	2301	0.383212	11,000	West
17	9388	2322	0.231362	14,000	West
18	3185	2322	0.279139	14,000	West
19	2572	2341	0.26221	11,000	West
20	12586	2341	0.459807	11,000	West

STEP 4 : ANALYSIS AND WRITE UP

" " "

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

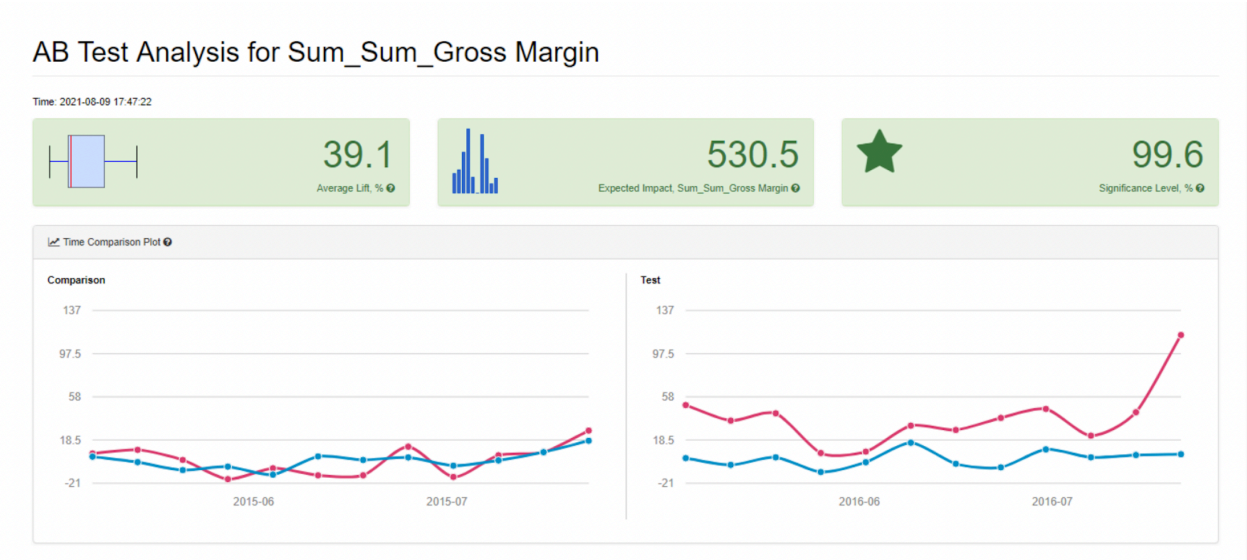
1. *What is your recommendation - Should the company roll out the updated menu to all stores?*
2. *What is the lift from the new menu for West and Central regions (include statistical significance)?*
3. *What is the lift from the new menu overall?*

" " "

Based on the AB analysis which I have performed using the dataset provided with me, I would recommend the company to to roll out their updated menu to all stores since the `profit growth is **more than 18 percent** compared to the comparative period while compared to the control stores.

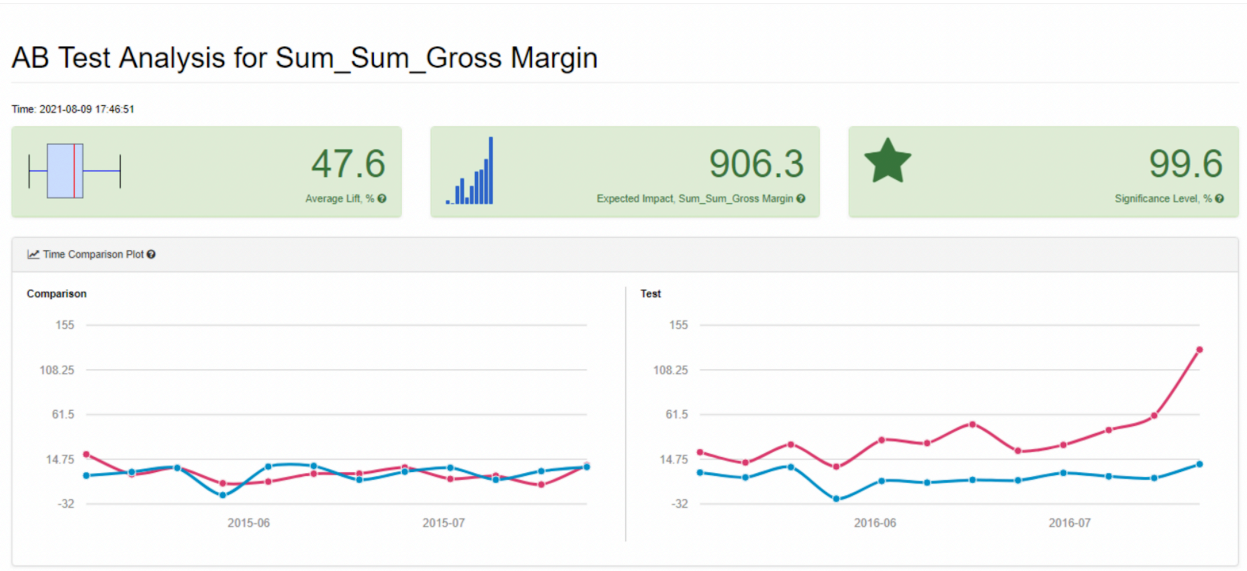
You've been asked to analyze the results of the experiment to determine whether the menu changes should be applied to all stores. The predicted impact to profitability should be enough to justify the increased marketing budget: at least 18% increase in profit growth compared to the comparative period while compared to the control stores; otherwise known as *incremental lift*. In the data, profit is represented in the *gross_margin* variable.

WESTERN REGION:



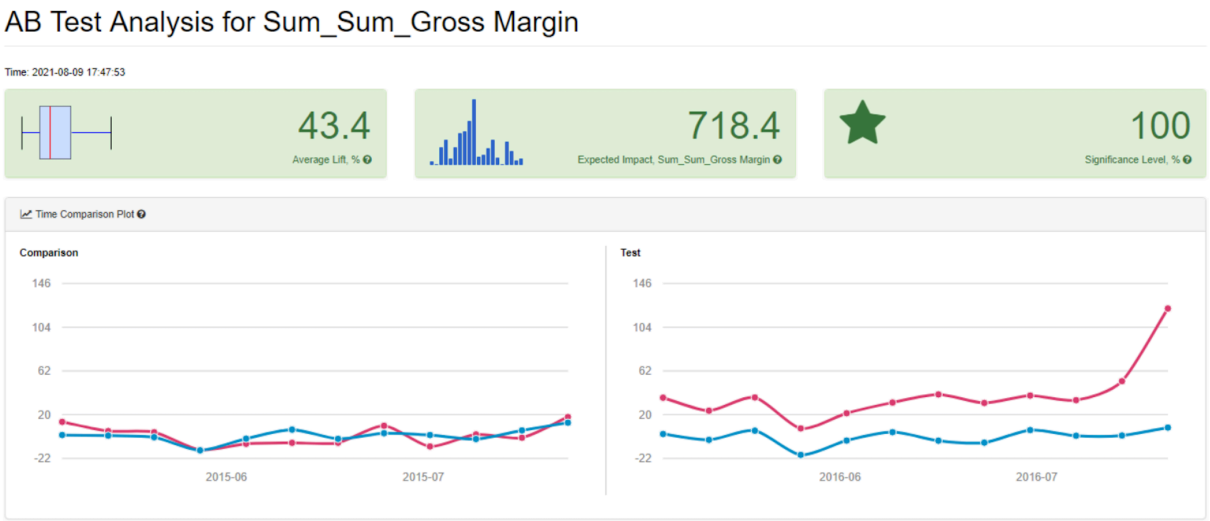
Stores in the western region shows **39.1 percent increase** in average lift after the implementation of their new menu with **99.6 percent increase** in the significance level.

CENTRAL REGION:



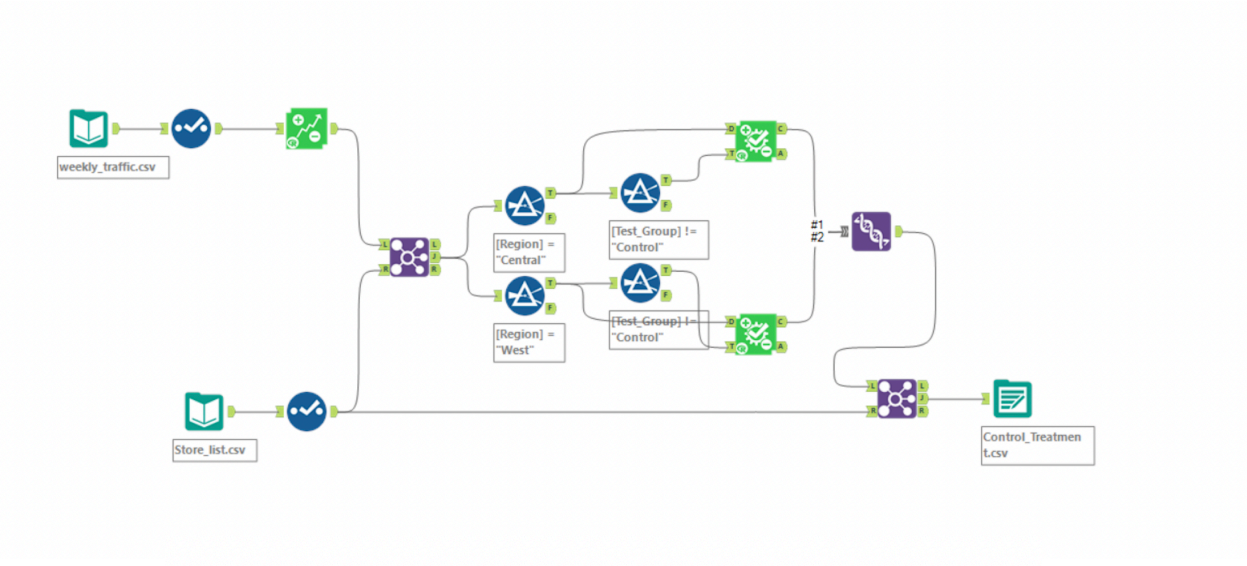
Stores in the central region shows **39.1 percent increase** in average lift after the implementation of their new menu with **99.6 percent increase** in the significance level.

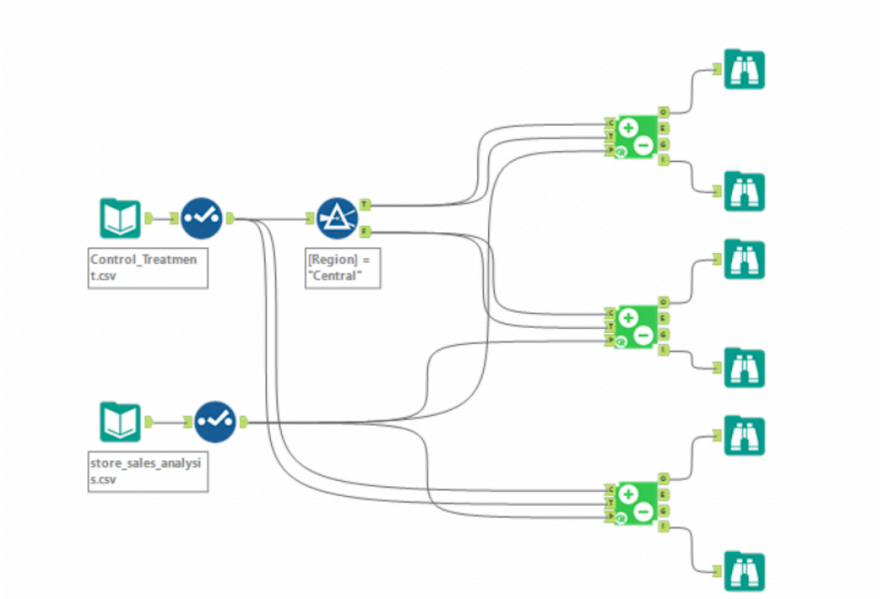
OVERALL REGION:



While looking at the overall region, implementation of new menu shows **43.4 percent increase** in average lift with **99.6 percent increase** in the significance level.

PROCESS WORKFLOW:





Reference:

- Udacity - Predictive analytics for business - AB testing module - Matched pair contents.
- Udacity AB testing (free course).