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 used the Weekly Gross Margin data to evaluate the test result, which indicates weekly profit of each store.
2. What is the test period?
   The test were conducted over a period of 12 weeks.
3. At what level (day, week, month, etc.) should the data be aggregated?
   The data should be aggregated weekly.

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.

Workflow capture of the cleaning and aggregation process:
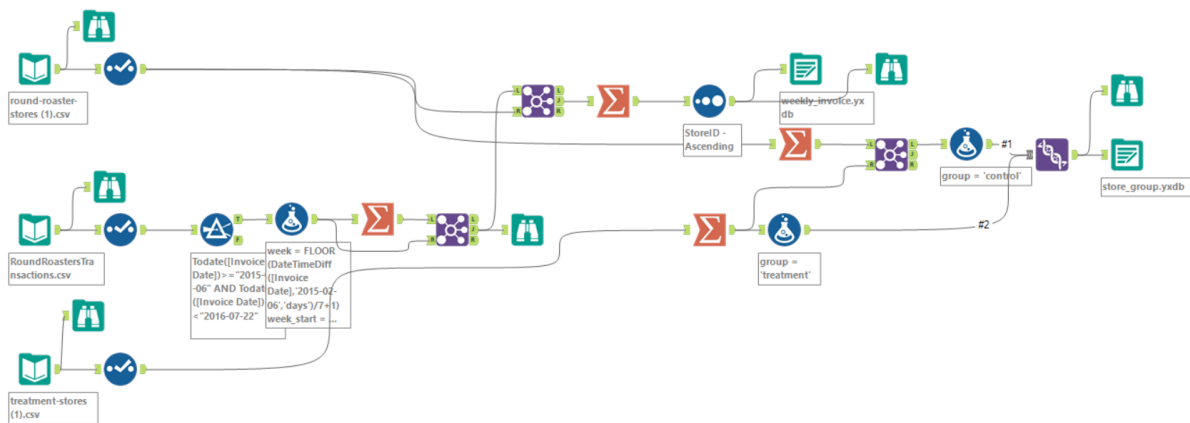
Brief steps:

*1. Selecting relevant field from initial store datasets, including: **StoreID, Region** and **AvgMonthSales.** In terms of Transactions dataset, only keep: **StoreID, InvoiceNumber/Date, Category, Product, Gross Margin** and **Sales.***
2. Then filter the transactions with the range of 76 weeks from the end of the test period 2016-July-21. Creating additional fields of week by order, start and end date of each week (only keep stores with 76 weeks of data). The output dataset contains detailed transactions date (week, start/end week) info
3. Join the resulted transaction dataset with store dataset. And group by each store and each

week with the weekly distinct invoice number counted. Then I got the dataset of weekly traffic (invoices) per store.
4. Create a dataset of weekly gross per store.
5. Create the dataset which categorize all stores into treatment and control group.

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 control variable should be considered as well: **AvgMonthlySales***

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

   *Correlation coefficient between **AvgMonthSales** and **Sum_Gross_Margin** (Weekly): is 0.79, indicates a strong positive relationship*
   Scatter plot:

Scatterplot of AvgMonthSales versus Sum_Gross_Margin
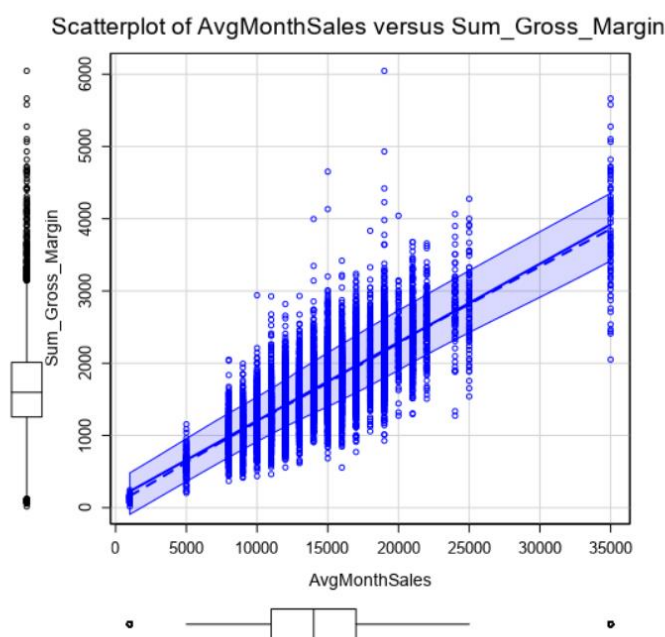
While the correlation between **Sum_Gross_Margin** and other numeric variables shows no relationship. For example: the size of a store (**Sq_ft**) relationship with **Sum_Gross_Margin** doesn't seem to have any relationship.

| Record	FieldName	AvgMonthSales	Sq_Ft	Sum_Gross Margin
1	AvgMonthSales	1	-0.046967	0.790358
2	Sq_Ft	-0.046967	1	-0.019345
3	Sum_Gross Margin	0.790358	-0.019345	1

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

   AvgMonthSales will be used along with Trend and Seasonality when paring treatment and control units.

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	11268	3102
2301	12286	12219
2322	3235	11618
2341	2572	3185
1664	1542	7584
1675	6992	7284
1696	8362	1964
1700	2014	7212
1712	8162	2114

Matching Treatment and Control Units workflow:
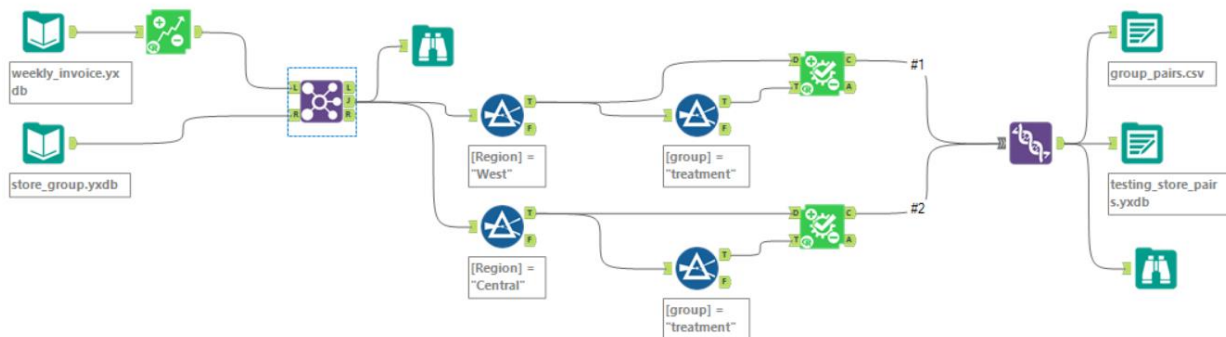
Step 4: Analysis and Writeup

Conduct your A/B analysis and create a short report outlining your results and recommendations. (250 words limit)

A/B Testing Workflow:
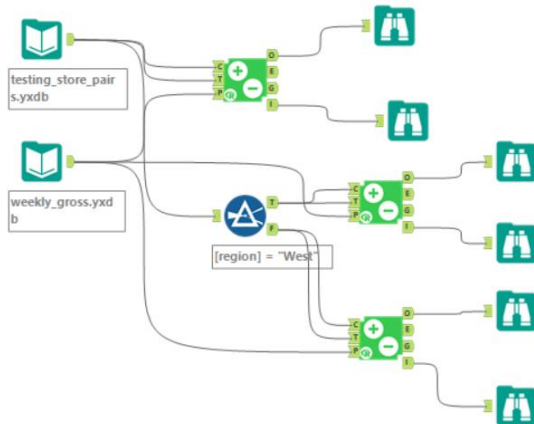
Answer these questions. Be sure to include visualizations from your analysis:

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

    Since the expected average lift is 40.8% with 100% significant level, exceeds 18% condition. Then I suggest the company should apply the new menu to all of their stores.

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

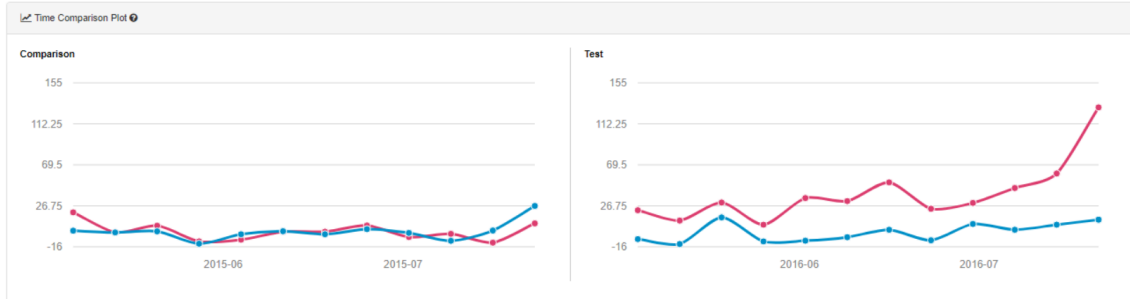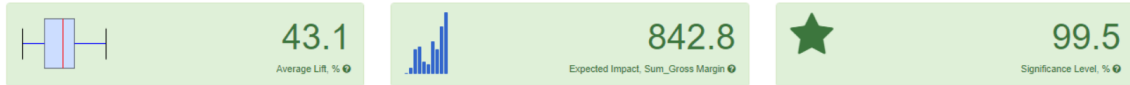   ***West Region:** 38.6% with 99.6% significant level*

AB Test Analysis for Sum_Gross Margin
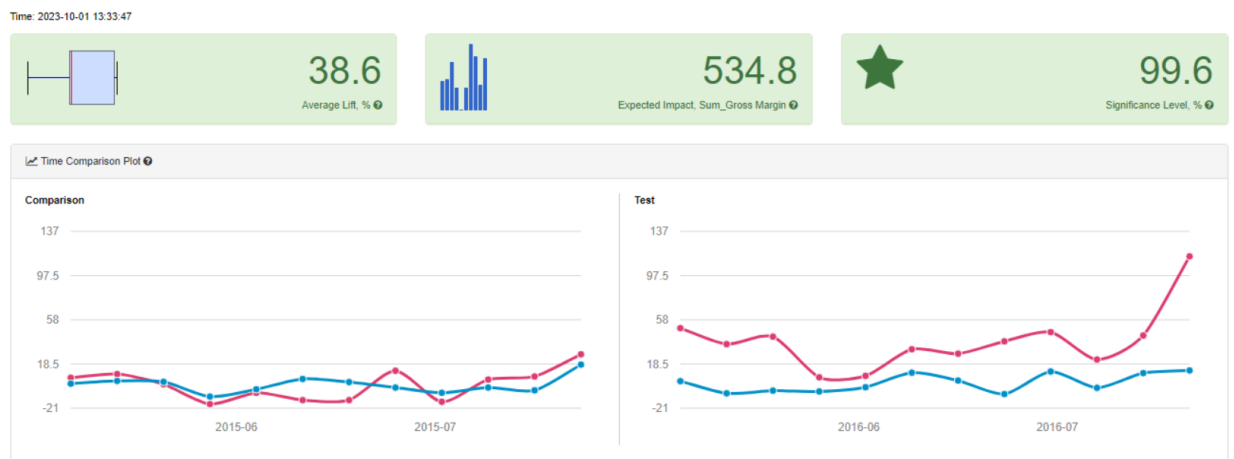
Central Region: *43.1% with 99.5% significant level*

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

   The lift from the new menu: 40.8% with 100% Significant level: