Project 3 Milestone 4: A/B Testing for Fast-Food Promotions

Ajuni Sohota

Business Problem

This is an A/B testing problem dataset that focuses on a fast-food chain that needs to decide which marketing scheme will work best for promoting a new item. They have three different locations where they are testing a new item. The weekly sales are recorded for each location for the first four weeks which is what I will use to determine which promotion is the best for their sales to increase. This kind of analysis can be applied to many marketing schemes. With only a small amount of generated data, you can predict which tactic would be best to gain revenue in your business. For instance, this dataset only includes sales for four weeks after each promotion has been tried. From here we can analyze which tactic works the best, so the company will not lose any revenue trying some promotion that doesn't actually bring in revenue.

Hypothesis

I predict that one of the promotions will work best. This is, of course, not always the case. Sometimes different tactics will provide similar revenue. In this event, it wouldn't matter which tactic to use. However, more data could be collected over time to do a repeat A/B test to see if anything changed. You also must consider things like how many people go to each location (what is the foot traffic?) which might be a factor in why a promotion is working better in one location rather than in another. But normalizing the data will allow us to account for this somewhat, as well.

Methodology

For this type of test, I will use a statistic test called ANOVA. ANOVA is an abbreviation of "analysis of variance" which is a technique that is used to see if two or more groups are significantly different from each other (Singh, 2020). You can use it to check if some factor impacts the performance of the different samples, as well (Singh, 2020). For two samples, a t-test and ANOVA test have the same result. But when there are more than two samples there is a "compounded effect on the error rate of the result" (Singh, 2020). I am going to either accept or reject the null hypothesis in this analysis. I will provide a t-table where these results are shown as

well as a graph showing the revenue in sales for each promotion. The null hypothesis in this case would be:

H0 = There is no statistical difference between promotion 1, 2, or 3

And the alternate hypothesis is:

H1 = The promotions 1, 2, and 3 perform statistically differently from each other

The first step I will take is an overall ANOVA test on each promotion. This will tell me if, in general, one promotion outperforms the others in sales revenue. But what if there is another factor could be affecting the performance, such as the market size of each location promotion? To see if this changes my results, I will extract each market size from each promotion (small, medium, and large market sizes) and reperform these tests. This will tell me if there is something in particular that is driving the promotion performance.

Results

The tables below show all the t-tables for each test performed. The appendix shows the performance for each test in sales revenue for each promotion. What was discovered is that promotions 1 and 3 were not statistically different in their performance. Both promotions 1 and 3 outperformed promotion 2, however, and were also statistically different in their performance. Even with regards to market size, the promotions all performed the same as the overall ANOVA test.

This means that I reject the null hypothesis for promotions 1 and 3, but accept the null hypothesis for promotions 1 and 2/promotions 2 and 3. This could move forward either in the way the fast-food restaurant can move forward with just promotion 1 or 3, or they could move forward with 1 and 2 and test the performance of either again in four weeks to see if anything changes with promotion 2's performance. Since 1 and 3 perform the same way statistically speaking, they could also proceed with both because both drive a higher revenue, then do a retest in another four weeks for these two to see which promotion they want to keep. There are a few options and it would have to be discussed with the business team. I would recommend that they proceed with promotions 1 and 3 and do a retest after another 4 weeks of sales data is compiled.

Table 1. T-Table for All Promotions Regardless of Market Size

Multiple Comparison of Means - Tukey HSD, FWER=0.05

Group 1	Group 2	MeanDiff	P-adj	Lower	Upper	Reject
1	2	-10.7696	0.001	-14.774	-6.7652	True
1	3	-2.7345	0.2447	-6.7389	1.2698	False
2	3	8.0351	0.001	4.1207	11.9494	True

Table 2. T-Table for All Promotions with Small Market Size

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
Group 1	**Group 2**	**MeanDiff**	**P-adj**	**Lower**	**Upper**	**Reject**
1	2	-9.3519	0.001	-13.6829	-5.0208	True
1	3	-0.6483	0.9	-4.5579	3.2612	False
2	3	8.7035	0.001	4.536	12.8711	True

Table 3. T-Table for All Promotions with Medium Market Size

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
Group 1	**Group 2**	**MeanDiff**	**P-adj**	**Lower**	**Upper**	**Reject**
1	2	-8.5583	0.001	-11.3108	-5.8057	True
1	3	-2.2037	0.1356	-4.9113	0.5038	False
2	3	6.3545	0.001	3.7306	8.9785	True

Table 4. T-Table for All Promotions with Large Market Size

Multiple Comparison of Means - Tukey HSD, FWER=0.05						
Group 1	**Group 2**	**MeanDiff**	**P-adj**	**Lower**	**Upper**	**Reject**
1	2	-14.9139	0.001	-21.5284	-8.2992	True
1	3	1.9681	0.7701	-5.1424	9.0785	False
2	3	16.8819	0.001	9.9797	23.7842	True

Appendix (Supporting Documentation)

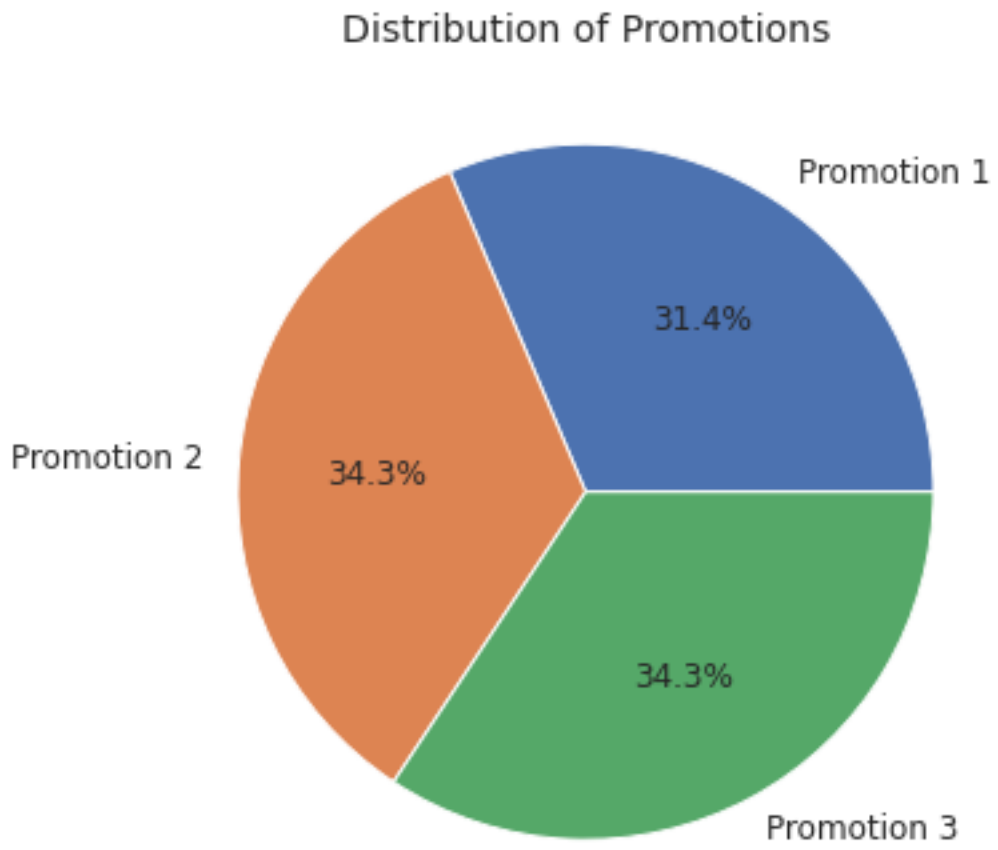

Figure 1. Distribution of Each Promotion
There is roughly 1/3 distribution for each promotions 1, 2, and 3. However, further investigation of market sizes will be performed to see if there is any other factor that could determine bias in each promotion.

Figure 2. Visualization of the sales in each promotions for each market size.

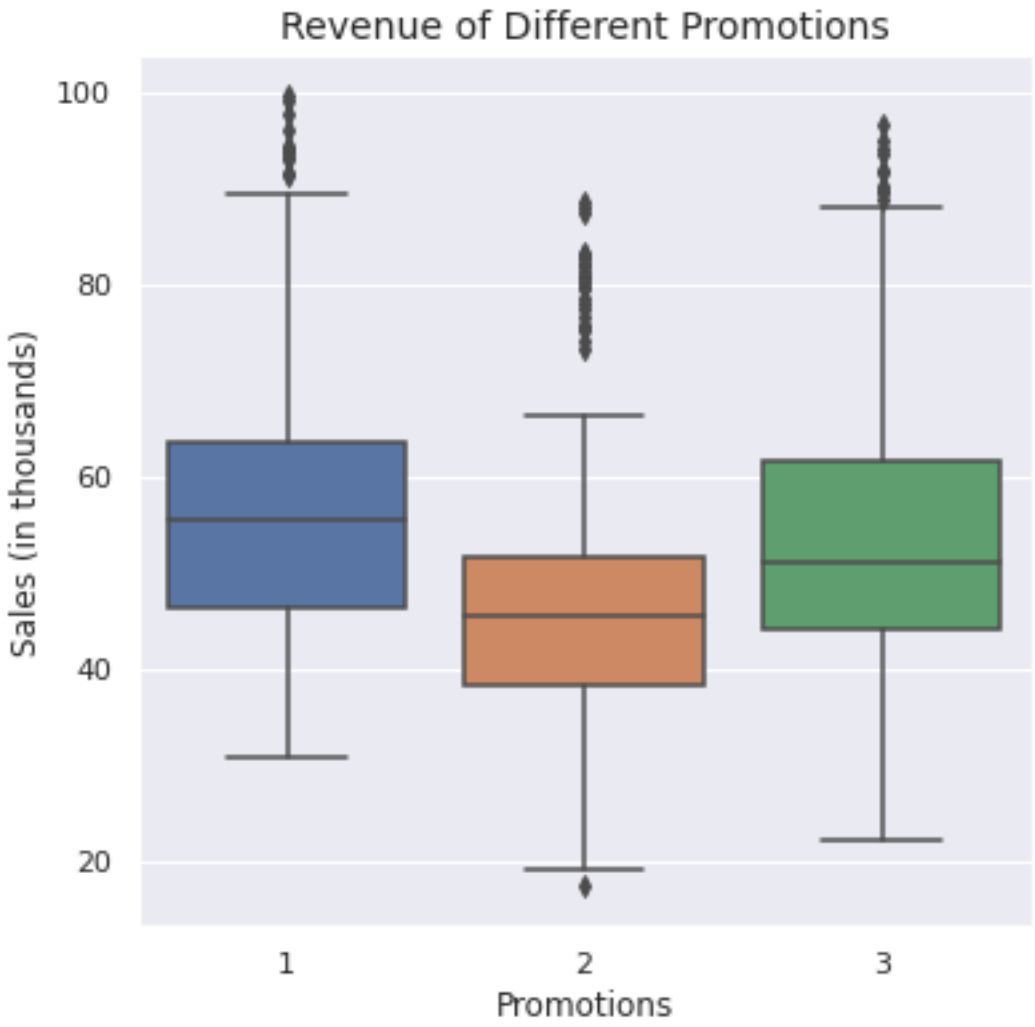

Figure 3. Sales for each promotion.

Figure 4. Performance of each promotion via ANOVA test.

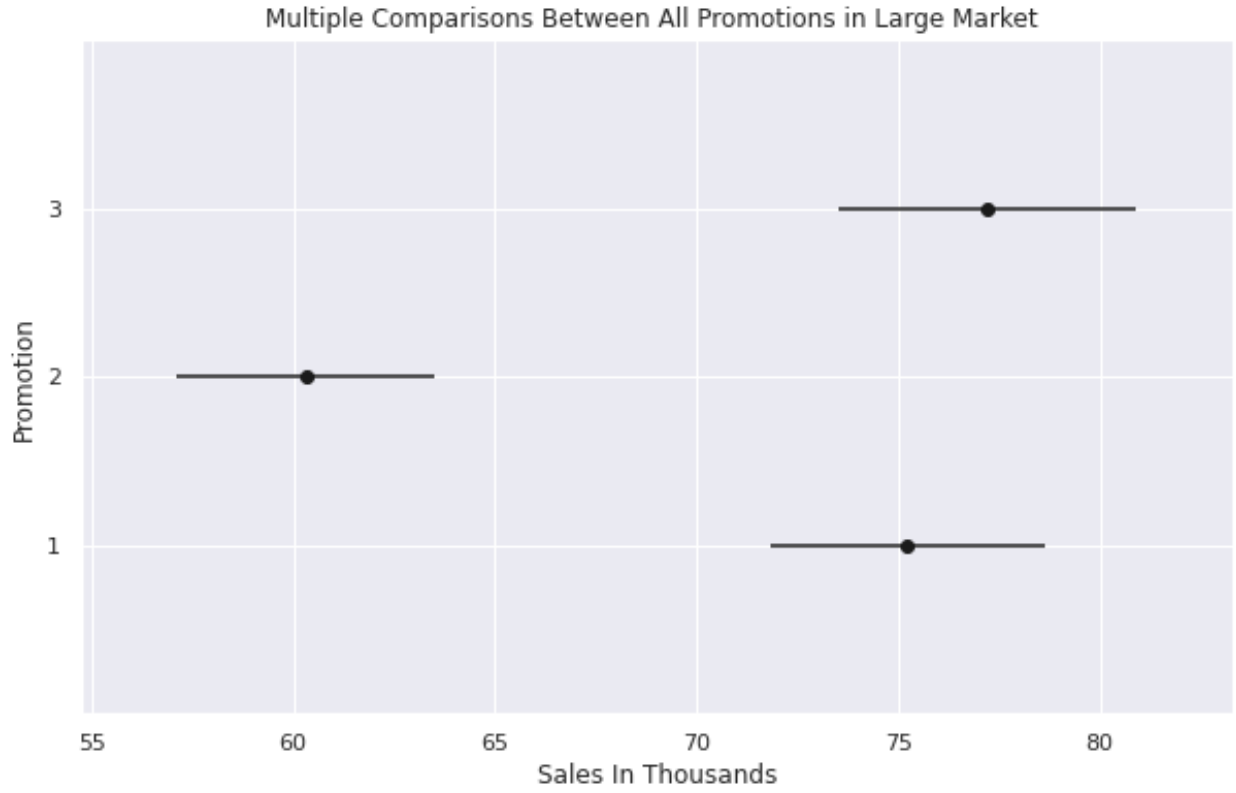

Figure 5. Performance of each promotion in the large market category.

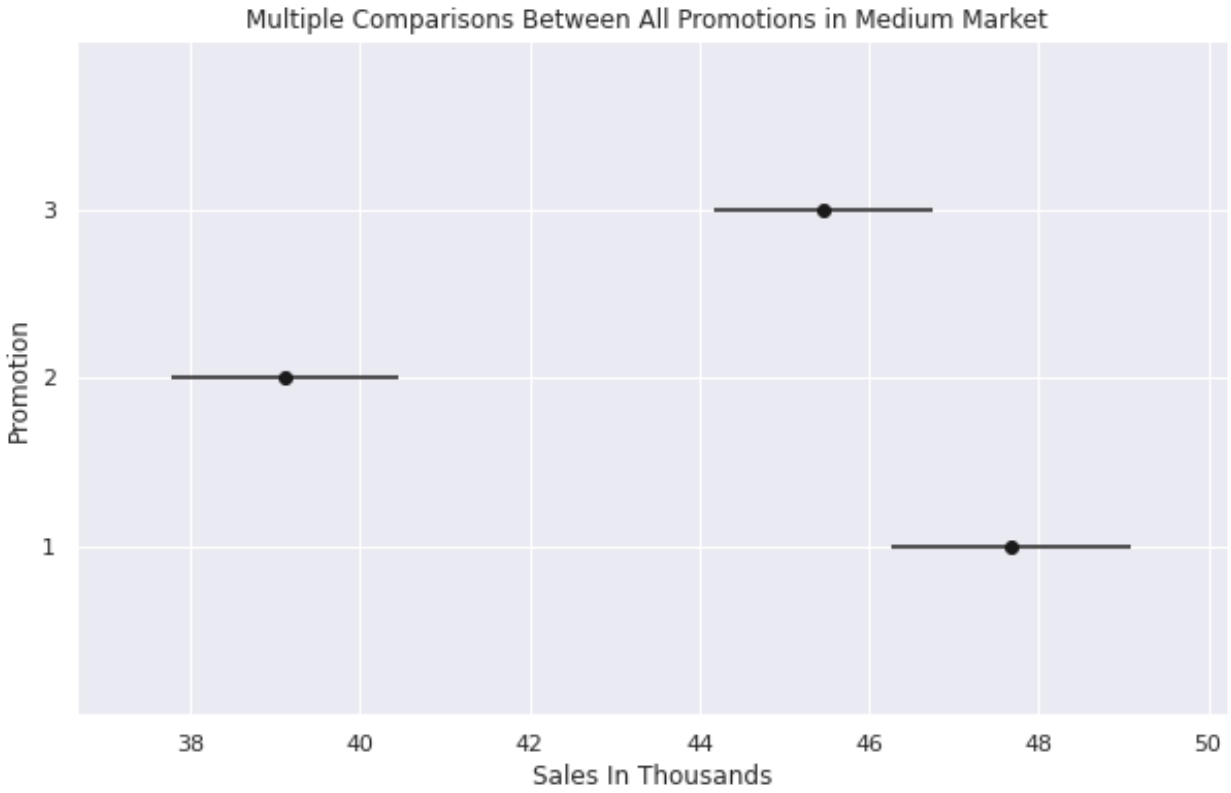

Figure 6. Performance of each promotion in the medium market category.

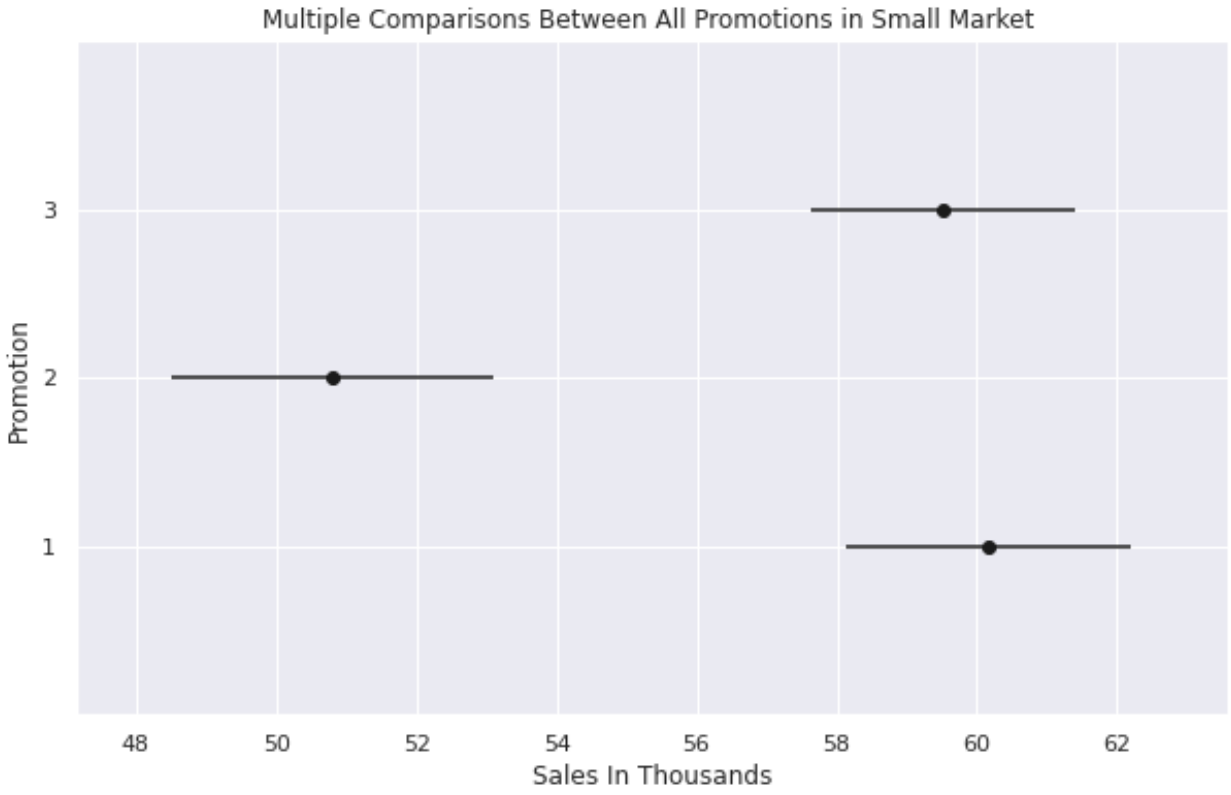

Figure 7. Performance of each promotion in the small market category.

10 Questions Audience May Ask

1. What is A/B testing? A/B testing is a statistical approach to predictive testing to see which of multiple options would result in a favorable outcome based on a subset of data provided.

2. What other markets is A/B testing used in? A/B testing is often used in marketing but it can be used for many things. Some examples are clickbait on websites, landing page design, adversitements, email notifications, calls-to-action in politics, and menu displays (Riserbato, 2021).

3. Does it matter if you use Python or R? You can perform these kinds of tests in either python or R. For me, I used Python because that is what I most familiar with, however, many people that perform statistical analysis tend to use R. If I did use R, I would have used ggplot2 to plot all my EDA and my ANOVA test results.

4. What kind of statistics do you use for A/B testing? A tukey test was performed and a t-table was provided as well as a graph of the performance of each promotion in sales revenue.

5. Would it better to use data longer than a 4 week test? The more data you have to compare the promotions, the better. For instance 1 day of results of sales wouldn't provide us as much

information as if we had the full 4 weeks, so the same can be said if we had two months of data. It's up to companies to decide when they cut off the test because they could be losing money if they keep up with a promotion doesn't work. So, it is a trade-off you have to make a decision with. Another option is to keep periodically performing these tests.

6. What kind of jobs require you to use A/B testing? So far I have seen tons of data analyst positions and data scientist positions that require you to know how to do this. Most of these positions are for web-design type of A/B testing and marketing positions.

7. How could this work at a single restaurant or a single company? This could work at a single company or restaurant by testing two or more different options on two different test groups. For instance, they could test something in the morning, and a different thing at night and then switch these after a week to account for the variance between day and night, too.

8. What kind of test is A/B testing? This is a statistical test that can be tested by ANOVA.

9. What does the A and B stand for? A and B are referring to one group (A) and another group (B) that you are testing against each other. However, you can do this test for a multitude of groups.

10. How important is visualization for A/B testing? I think A/B testing requires lots of visualizations in order to make sense of your data. First you need visualizations for the EDA part and then once you do the tests, it's good to be able to see what is occurring by graphs so that your client or company can better understand what is happening.

References:

Riserbato, R. (2021, October 12). *9 A/B testing examples from real businesses*. HubSpot Blog. Retrieved November 18, 2021, from https://blog.hubspot.com/marketing/a-b-testing-experiments-examples.

Singh, G. (2020, April 1). *Analysis of variance (ANOVA): Introduction, types & techniques*. Analytics Vidhya. Retrieved November 18, 2021, from https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/.