**Introduction**

Companies update and add new products continuously and it is important to be able to market those products effectively to the public. While there can be different marketing strategies companies use depending on the product, it is imperative to understand which ones work most effectively. Testing these different strategies helps better understand their effectiveness on the spending population, and how the sales of different products might benefit from different strategies.

**Method**

In this instance, there is a fast-food chain that wants to add a new menu item. There are 3 different marketing strategies this company could use and are undecided on which one is the most effective. 3 different marketing strategies are implemented simultaneously over the course of 4 weeks to test the effectiveness of each strategy.

The hypothesis being that there will be differences between one or all 3 of these promotional strategies. The null hypothesis then being that no differences are seen between these 3 strategies.

To evaluate these results, A/B testing is used to determine which marketing strategy works best for this product at this company.

**Results**

It is imperative to collect the important information from vast spreadsheets of information. Using BigQuery, the following code was used to condense the data spreadsheet into easier to understand segments and aggregate the data where necessary:

SELECT

  promotion,

  COUNT(location\_id) AS location\_count,

  ROUND(SUM(sales\_in\_thousands),3) AS sales\_amount,

  ROUND(AVG(sales\_in\_thousands),3) AS mean\_sales,

  ROUND(STDDEV(sales\_in\_thousands),3) AS standard\_deviation

FROM

  `turing\_data\_analytics.wa\_marketing\_campaign`

GROUP BY

  promotion

This resulted in the following table to be used for A/B testing:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| promotion | location\_count | sales\_amount | mean\_sales | standard\_deviation |
| 1 | 172 | 9993.03 | 58.099 | 16.554 |
| 2 | 188 | 8897.93 | 47.329 | 15.109 |
| 3 | 188 | 10408.52 | 55.364 | 16.766 |

For the A/B testing, Evan’s Awesome A/B Tools was used to calculate the difference, the standard error, and the p-value using 2 sample T-Testing. Since we are running pairwise comparisons across multiple marketing campaigns, it is suggested that a confidence interval of 99% is used instead of the traditional 95% confidence interval. This would ensure that there wouldn’t be an increased chance in getting a type 1 error (false positive).

A/B testing for promotions 1 and 2 were fun first:

A screenshot of a computer

Description automatically generated

The results from the A/B test showed that there is no significant difference between promotion 1 and promotion 2 for the new menu item. To signify statistical significance, we look to the p-value. If the p-value is at or below 0.05 then the result would be a statistically significant difference. Since in this case the p-value is 1, well above the 0.05 threshold, we know that the differences between the two promotional strategies are not statistically significant.

The A/B testing for promotions 1 and 3 were run next:

A screenshot of a computer

Description automatically generated

The results from the A/B testing once again showed that there is no statistical significance between the promotional strategies of 1 and 3 for the new menu item. We can see that the p-value in this instance is 0.94, still well above the threshold we want for statistical significance of 0.05.

Lastly, the A/B test was run on promotions 2 and 3:A screenshot of a computer

Description automatically generated

We see here that there is statistical significance between promotional strategy 2 and promotional strategy 3. Promotion 2 data is in the sample 1 summary section while promotion 3 data is in the sample 2 summary section. We see the p-value here is 0.001, well below the threshold of 0.05 indicating statistical significance. It is noted that the sample 2 mean is greater, meaning that promotion 3 mean is greater than the mean of promotion 2. This would suggest that promotion 3 outperformed promotion 2 as a marketing strategy.

**Discussion**

Testing for multiple metrics can get tricky as it introduces a concept known as the multiple comparisons problem. This suggests that the “larger the number of inferences made, the more likely erroneous inferences become”. This is why we increased our confidence interval to 99% to have a stricter significance threshold for each individual comparison to compensate for the number of inferences being made.

It is also noticed that promotion 1 had 16 fewer locations than promotion 2 and promotion 3, this could lead to sample ratio mismatch that leads to the results from promotion 1 being statistically more or less significant than the other 2 promotional strategies. Ideally each promotional strategy would have an equal number of locations to properly determine significance. This could possibly have pushed promotion 1 into a statistically significant difference between promotion 2.

**Decision**

From what was analyzed through A/B testing it seems that promotion 2 did not perform as well in marketing the new menu item as promotion 1 and 3. Perhaps next steps could be evaluating promotions 1 and 3 together simultaneously to see which promotion ultimately outshines the other. This would of course need to be done over another 4 weeks with an equal number of stores involved in the marketing strategy.