Fast Food Marketing Campaign A/B Test Analysis

Goal of the Test

The goal of this A/B test is to evaluate which one of the three marketing campaigns (Promotion 1, 2, 3) showed the best results for promoting a newly introduced menu item. The evaluation will be made in terms of total **sales** per group.

There are three groups that need to be A/B tested against each other. Therefore a confidence level of 99% will be used. Setting a higher confidence level allows to compensate for the number of inferences that will have to be made, hence lowering the impact of multiplicity.

The descriptive measures for the test group are provided in the following table:

Group	n	Mean	St. dev
Promotion 1	43	232.39	64.11
Promotion 2	47	189.32	57.98
Promotion 3	47	221.45	65.53

Test hypotheses:

$\mathbf{H_0}$ - there are no significant differences among the promotion campaigns in terms of effectiveness.
$\mathbf{H_1}$ - there are significant differences among the promotion campaigns in terms of effectiveness.

Target Metrics

Since the goal of this A/B test is to evaluate the effectiveness of the three different marketing campaigns based on generated income, a single key metric will be used:

- **sales** - the total income generated per unique location throughout the whole 4-week testing period. It is a continuous measure, therefore it will be tested using a *Two-Tailed Independent T-Test.*

However, before we proceed looking for significant differences among the groups based on **sales**, we need to make sure that all 3 test groups are truly homogeneous. To achieve that, we will evaluate the following metrics:

- **locations_count** - represents total count of unique locations within each test group. The metric will be evaluated using *Chi-square Goodness of Fit Test* to check for a potential ***Sample Ratio Mismatch (SRM)***.
- **market_size** - a parameter that assigns each location a category based on its size of market area by sales. The metric will be evaluated with *Chi-square Goodness of Fit Test* to check for potential **Sample Distribution Mismatch (SDM)**.
- **age_of_store** - a parameter that shows the age of each location expressed in years. The performance of the older locations can differ from the newer ones in unforeseen ways (e.g. Novelty effect if the store just opened). Therefore, all 3 test groups will be evaluated against each other using *Two-Tailed Independent T-Test* to check for potential **SDM**.

SQL code used for data extraction is provided in the **Appendix**.

Calculations

1. Testing for potential Sample Ratio Mismatch (α = 0.05)

Promotion 1 group has **n = 43**, while **Promotion 2** & **Promotion 3** groups have **n = 47**. To check for potential ***SRM*** a *Chi-square Goodness of Fit Test* was performed. At 95% confidence, the analysis showed no significant variation in sample sizes across the three groups (**p = 0.089**). Therefore, in terms of **n** all 3 groups can be considered homogeneous.

NB: **p** value was derived from $\mathbf{x^2}$ using this online P Value from Chi-Square Calculator. The steps for calculating $\mathbf{x^2}$ are provided in the table below:

Promotion	Observed (O)	Expected (E)	O – E	(O – E)²	(O – E)² / E
Promotion 1	43	45.67	−2.67	7.11	0.156
Promotion 2	47	45.67	1.33	1.77	0.039
Promotion 3	47	45.67	1.33	1.77	0.039
$\chi^2 = 0.234$					
p = 0.890					

2. Testing for potential Sample Distribution Mismatch in terms of market_size (α = 0.05)

Testing the three test groups (Promotion 1, 2, 3) using the *Chi-square Goodness of Fit Test* showed no significant difference in terms of distribution of locations across **market_size** categories (Large, Medium, Small) (**p = 0.88**).

NB: evaluations were made using this online Chi-Square Calculator.

Chi-Square Calculator

Success! The contingency table below provides the following information: the observed cell totals, (the expected cell totals) and [the chi-square statistic for each cell].

The chi-square statistic, *p*-value and statement of significance appear beneath the table. Blue means you're dealing with dependent variables; red, independent.

Results						
	Large	Medium	Small			***Row Totals***
Promo 1	14 (13.18) [0.05]	24 (25.11) [0.05]	5 (4.71) [0.02]			43
Promo 2	16 (14.41) [0.18]	27 (27.45) [0.01]	4 (5.15) [0.26]			47
Promo 3	12 (14.41) [0.40]	29 (27.45) [0.09]	6 (5.15) [0.14]			47
Column Totals	42	80	15			**137 (Grand Total)**

The chi-square statistic is 1.1885. The *p*-value is .879993. The result is *not* significant at $p < .05$.

3. Testing for potential Sample Distribution Mismatch in terms of age_of_store (α = 0.05)

No significant sample distribution mismatches in terms of **age_of_store** have been identified among the three test groups:

- **Promotion 1** vs **Promotion 2** (**p = 0.83**; link to test)
- **Promotion 1** vs **Promotion 3** (**p = 0.5**; link to test)
- **Promotion 2** vs **Promotion 3** (**p = 0.36**; link to test)

NB: *Two-Tailed Independent T-Tests* was performed using Evan Miller A/B Test Calculator

4. The evaluation of total sales per each test group (α = 0.01)

The three test groups were compared in terms of **sales** using a *Two-Tailed Independent T-Test.* A total of three comparisons were performed using Evan Miller A/B Test Calculator, and the results are as follows:

- **Promotion 1** group showed significantly better **sales** results compared to **Promotion 2** group (**p = 0.00128**; link to test).
- There was no significant difference in terms of **sales** between **Promotion 1** group and **Promotion 3** group (**p = 0.43**; link to test).
- At **α = 0.01**, no significant difference in terms of **sales** between **Promotion 2** group and **Promotion 3** group has been identified. However, the difference would be significant at **α = 0.05** (**p = 0.0136**; link to test).

The results are summarised in the below table:

Comparison	p-value	Significant at α = 0.01?	Interpretation
1 vs 2	0.00128	✅	Promotion 1 had significantly better sales than Promotion 2.
1 vs 3	0.43	❌	No significant difference in sales between Promotion 1 and 3.
2 vs 3	0.0136	❌	Not significant at α = 0.01, but significant at α = 0.05.

Decision

Promotion 1 group is the most effective group in terms of **sales**, because at **α = 0.01** it significantly outperforms **Promotion 2** group and is statistically equivalent to **Promotion 3** group.

Additionally, the reliability of the above findings can be reinforced by the results of the secondary tests. These tests showed no significant differences among the three test groups neither in terms of **Sample Ratio Mismatch** ($n_1 \approx n_2 \approx n_3$), nor in terms of **Sample Distribution Mismatch** across **market_size** and **age_of_store** variables**.**

Appendix

- SQL code used for extraction of **continuous** data that is required for *Independent T-Test*:

```
SELECT
 promotion,
 location_id,
 age_of_store,
 ROUND(SUM(sales_in_thousands),2) AS sales
FROM tc-da-1.turing_data_analytics.wa_marketing_campaign
GROUP BY 1,2,3
ORDER BY 1
```

- SQL code used for extraction of **categorical** data that is required for *Chi-square Goodness of Fit Test:*

```
SELECT
 promotion,
 market_size,
 COUNT(DISTINCT location_id) AS locations_count,
 FROM tc-da-1.turing_data_analytics.wa_marketing_campaign
GROUP BY 1,2
ORDER BY 1
```