Fast Food Marketing Campaign A/B Test Analysis

Goal of the test

A fast-food chain wants to launch a new menu item but is unsure which of three marketing campaigns will generate the most sales. To find out, we will test each campaign in different randomly chosen markets by introducing the product and tracking weekly sales for four weeks.

➔ The experiment aims to compare the effectiveness of the three marketing strategies (through A/B testing, extended to three variants) and identify which campaign drives the highest sales, so the company can choose the best approach for a full rollout.

  About the dataset

Target metric

We aim to determine which promotion has the greatest impact on sales. Each location was assigned a different promotion, and weekly sales of the new item were recorded over a four-week period. Our analysis will focus on the metric *sales_in_thousands* column from the *wa_marketing_campaign* table in BigQuery, comparing results across promotions 1, 2, and 3. Additionally, the *location_id* column will be used to determine how many locations participated in each promotion.

Calculations

Experiment Introduction

The sampling method used for this experiment is *simple random sampling*, ensuring that each location has an equal chance of being assigned to a promotion group.

We will conduct an *A/B test (extended to three variants)* to evaluate the performance of the three marketing campaigns. An A/B test is a controlled experiment that compares different versions of a treatment - in this case, promotions - to measure their effect on a key outcome (weekly sales revenue).

The experiment will be carried out in several stages:

1. Formulating hypotheses
2. Designing the experiment
3. Running the experiment
4. Conducting validity checks
5. Interpreting the results
6. Making the final decision

We will use a confidence level of 99%, which is stricter than the traditional 95%, to address the multiple testing problem that arises from performing pairwise comparisons.

For the statistical analysis, we will use Evan Miller A/B Test Calculator to perform a two-sample t-test, since our outcome variable (sales_in_thousands) is continuous. A two-sample t-test evaluates whether the difference in mean sales between two promotions is statistically significant.

The t-test formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE}$$

where SE is the standard error of the difference between means.

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

1. Hypothesis testing

To evaluate the effectiveness of the three marketing campaigns (Promotions 1, 2, and 3) on sales of the new menu item, we define the following hypotheses for pairwise comparisons:

- **A/B test 1: *Comparison between Promotion 1 and Promotion 2***

H_0 - There is no difference in average sales revenue between Promotion 1 and Promotion 2

$$\mu1 = \mu2$$

H_a - There is a difference in average sales revenue between Promotion 1 and Promotion 2

$$\mu1 =/= \mu2$$

Significance level $\alpha = 0.01$

- **A/B test 2: *Comparison between Promotion 1 and Promotion 3***

H_0 - There is no difference in average sales revenue between Promotion 1 and Promotion 3

$$\mu1 = \mu3$$

H_a - There is a difference in average sales revenue between Promotion 1 and Promotion 3

$$\mu1 =/= \mu3$$

Significance level $\alpha = 0.01$

- **A/B test 3: *Comparison between Promotion 2 and Promotion 3***

H_0 - There is no difference in average sales revenue between Promotion 2 and Promotion 3

$$\mu2 = \mu3$$

H_a - There is a difference in average sales revenue between Promotion 2 and Promotion 3

$$\mu2 =/= \mu3$$

Significance level $\alpha = 0.01$

2. Experiment design

- Randomization unit: Location
- Target population in the experiment: Locations where promotions were run
- Duration of experiment: 4 weeks

3. <u>Run the Experiment</u>

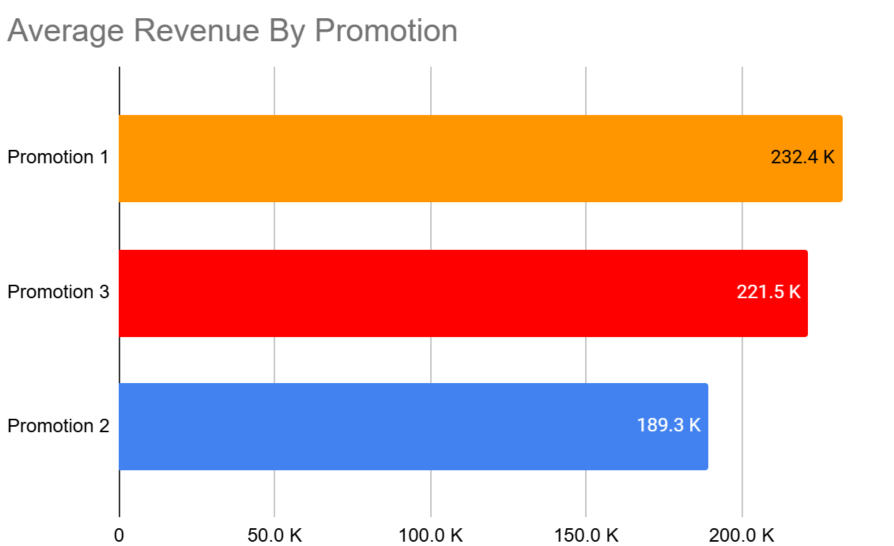
The experiment is run, and sales revenue data from promotions and their locations is collected and stored in BigQuery, and SQL queries are used to perform statistical calculations (see appendix for queries).

Using SQL query 1 in the appendix, we first aggregate by Promotion and Location ID columns in a CTE, then we calculate the count of locations, the average revenue, the standard deviation, and total revenue:

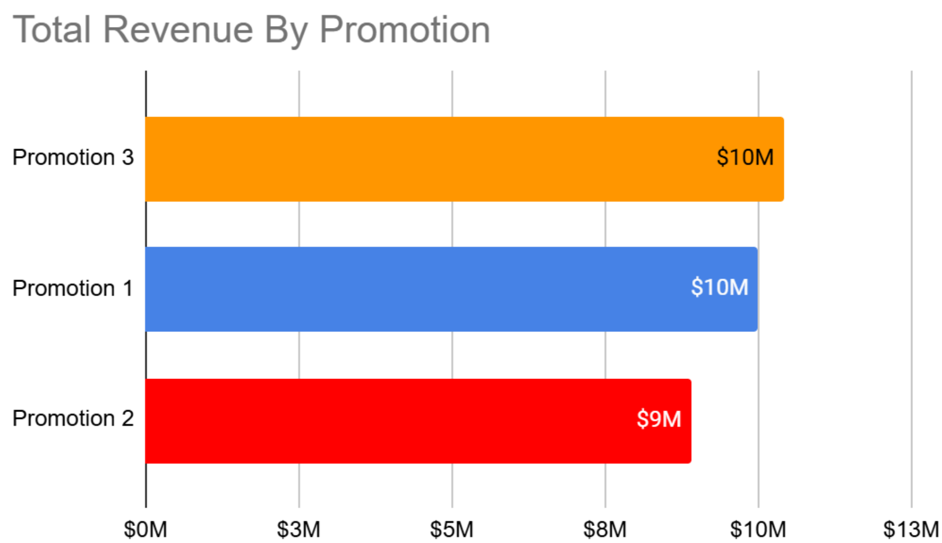
Promotion	Location Count	Average Revenue in thousands	Standard Deviation in thousands	Total Revenue Sum in thousands
Promotion 1	43	232.3960465	64.11289125	9,993.03
Promotion 2	47	189.3176596	57.98838944	8,897.93
Promotion 3	47	221.4578723	65.53546268	10,408.52

Table 1. Statistical calculations

Promotion 1 was run in 43 locations, while promotions 2 and 3 were run in 47 locations. Standard deviation is highest for Promotion 3 (65.54), then Promotion 1 (64.12), then Promotion 2 (57.99).



The average sales revenue is highest for Promotion 1 ($232.4K), followed by Promotion 3 ($221.5K), and then Promotion 2 ($189.3K).



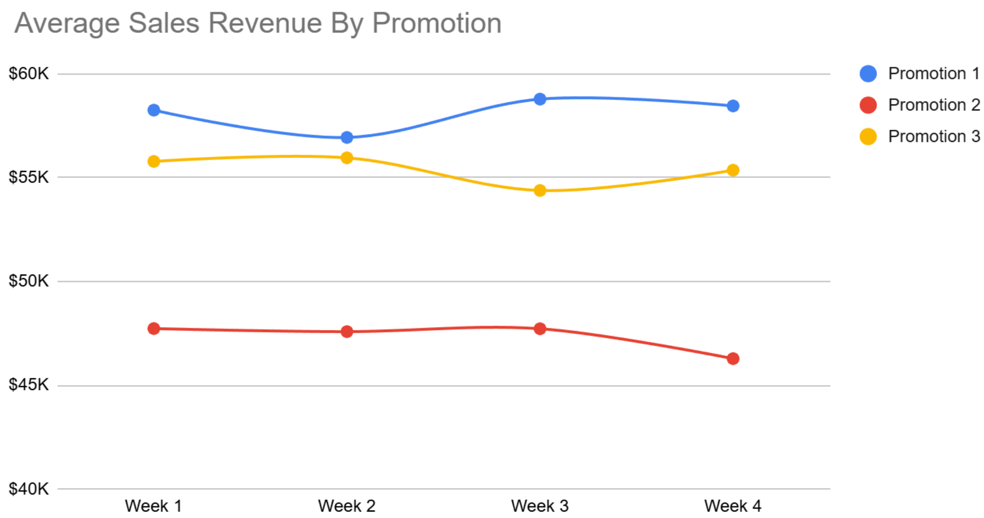
Promotion 3 generated the highest total revenue ($10.4M), followed by Promotion 1 ($9.9M) and then Promotion 2 ($8.8M).

While Promotion 3 generated the highest total revenue ($10.4M) due to more locations (47), the primary metric is average sales per location at week 4, where Promotion 1 (232.40 thousand) outperformed Promotion 2 (189.32 thousand) significantly.

Using SQL query 2 in the appendix, we retrieve weekly average revenue and weekly total revenue by promotions:

- Average Sales Revenue by Promotion in thousands

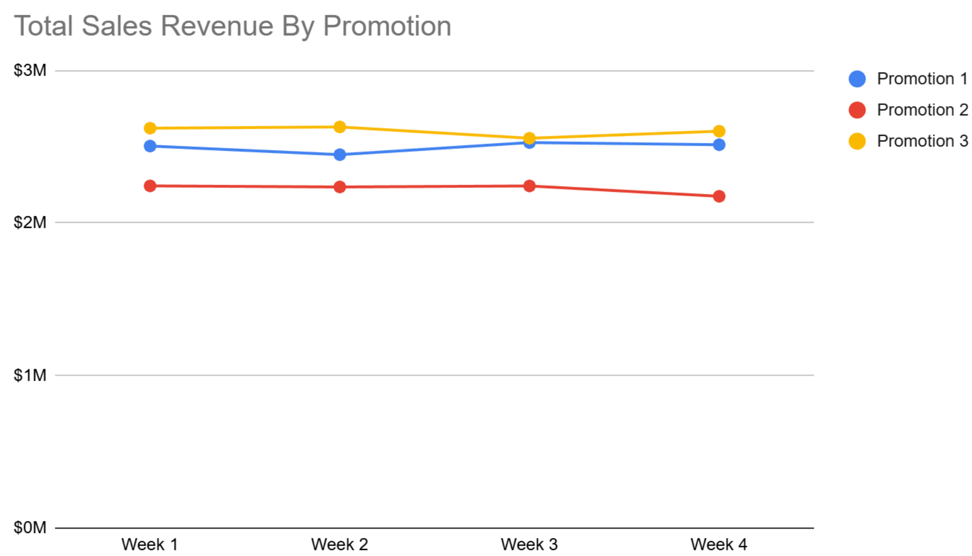
Promotion	Week 1	Week 2	Week 3	Week 4
Promotion 1	58.2444186	56.92953488	58.77488372	58.4472093
Promotion 2	47.73021277	47.58255319	47.72212766	46.28276596
Promotion 3	55.77617021	55.94914894	54.37787234	55.35468085



Promotion 1 starts at about $58K, goes down a bit to $57K in Week 2, rises to $59K in Week 3, and stays at $58K in Week 4, always the highest. Promotion 2 starts at $48K and drops to $47K, not doing well. Promotion 3 changes between $54K and $56K, ending at $55K, close to Promotion 1.

- Sum Sales Revenue by Promotion in $

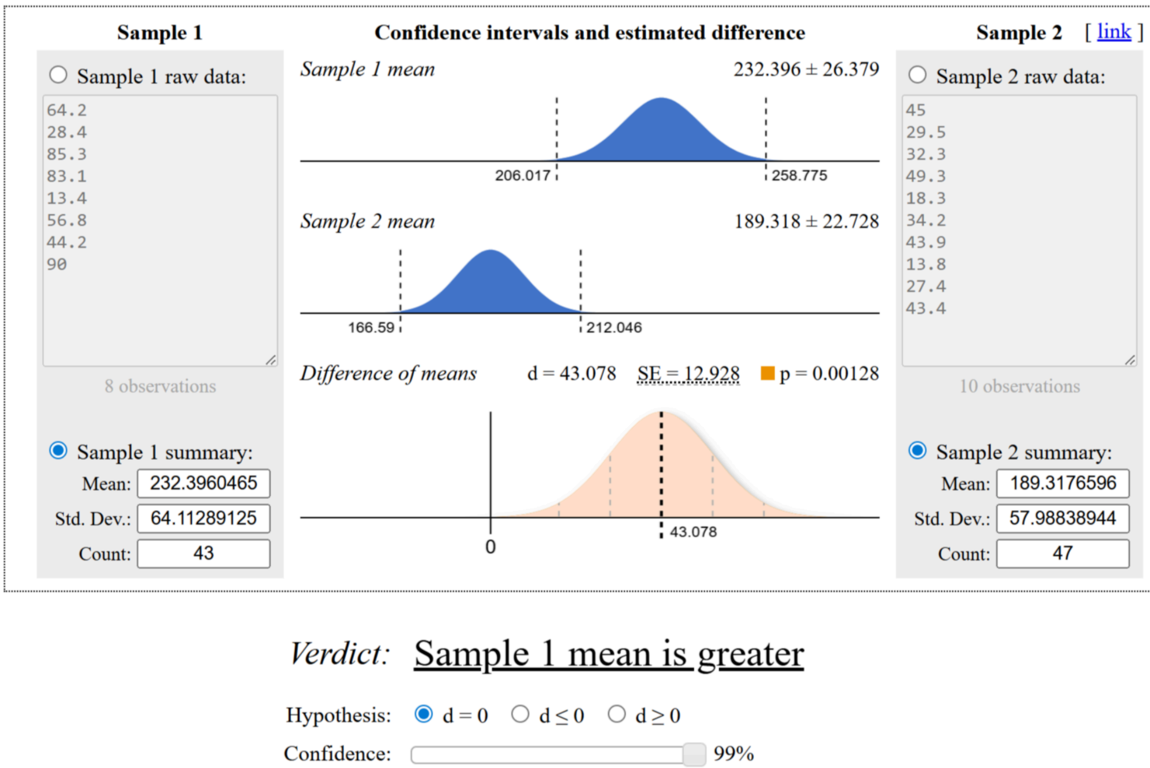
Promotion	**Week 1**	**Week 2**	**Week 3**	**Week 4**
Promotion 1	2 504.51	2 447.97	2 527.32	2 513.23
Promotion 2	2 243.32	2 236.38	2 242.94	2 175.29
Promotion 3	2 621.48	2 629.61	2 555.76	2 601.67



Promotion 3 starts at about $2.6M, dips to $2.55M in Week 3, and ends at $2.6M, thanks to 47 locations ($10.4M total). Promotion 1 stays around $2.5M, ending at $2.51M with 43 locations ($9.99M total). Promotion 2 goes from $2.24M to $2.17M with 47 locations ($8.9M total).

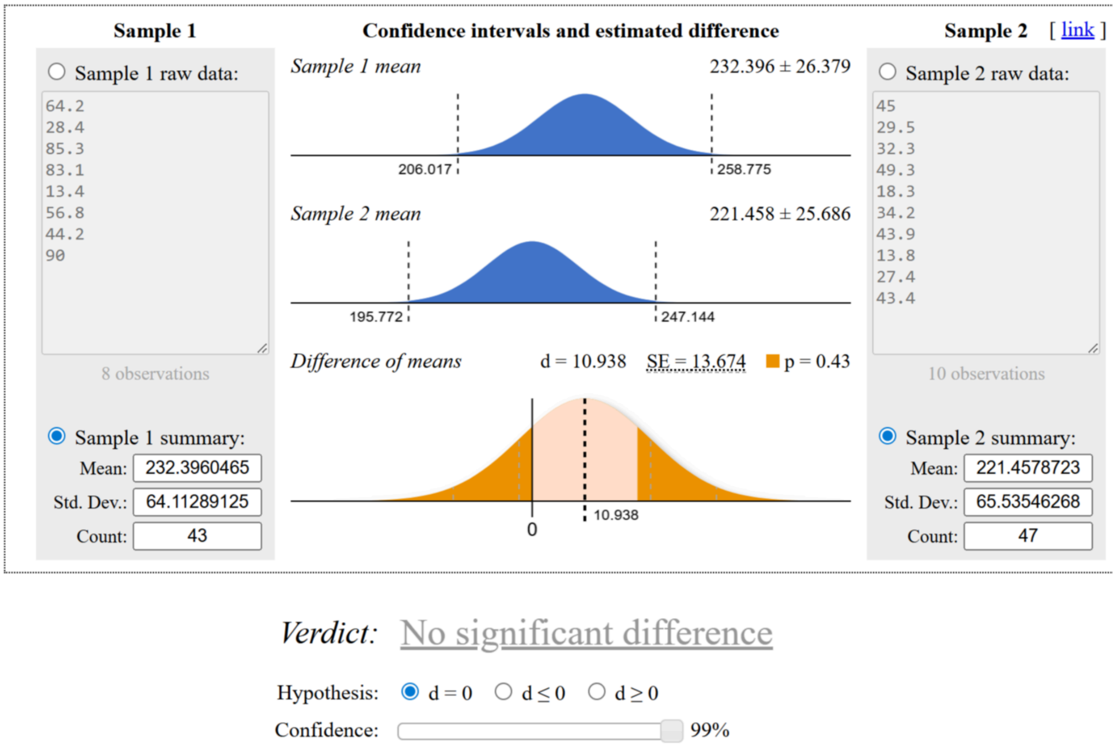
Next, we will use a two-sample t-test to compare mean sales (in thousands) across Promotions 1, 2, and 3, using *sales_in_thousands* from the *wa_marketing_campaign* table. Mean, standard deviation, and location counts per promotion will be calculated. A 99% confidence level will address multiple comparisons. The [Evan Miller A/B Test Calculator]() will perform the t-tests.

- **A/B test 1: *2 samples t-test, comparison between Promotion 1 and Promotion 2 means***



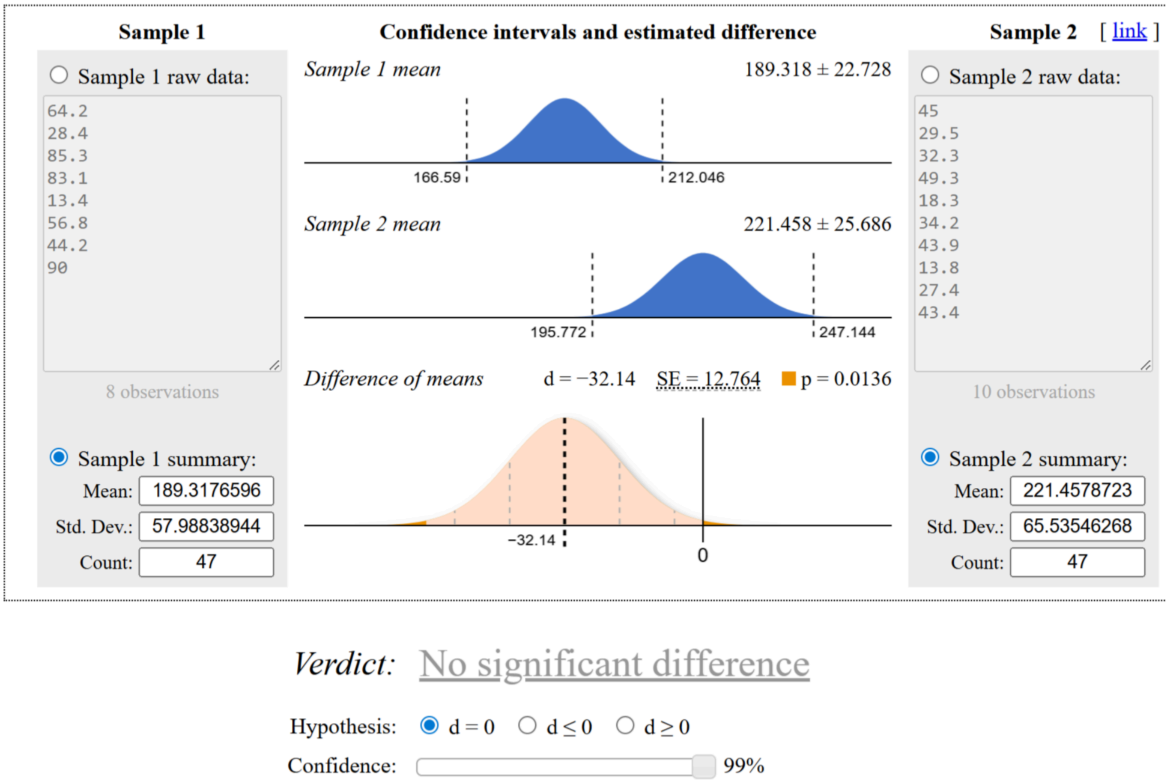
Verdict: Promotion 1 mean is greater, difference of mean is d=43.078 with a **P-value = 0.00128**

- **A/B test 2: *Comparison between Promotion 1 and Promotion 3***



Verdict: No significant difference between means of Promotion 1 and 3, and a **P-value = 0.43**

- **A/B test 3:** ***Comparison between Promotion 2 and Promotion 3***



Verdict: No significant difference between means of Promotion 2 and 3 and a **P-value = 0.0136**

4. <u>Validity checks</u>

- **External factors:** The A/B test's validity *may be* affected by external factors such as market differences (e.g., demographics or local competition), seasonal effects (e.g., holidays), operational inconsistencies (e.g., promotion execution), and customer behavior (e.g., brand loyalty), however we do not have much information at this stage about such factors.
- **Sample ratio mismatch:** Promotion 1 was tested in 43 locations, while Promotions 2 and 3 were each tested in 47 locations, creating a sample ratio mismatch. This may reduce statistical power for Promotion 1, as its smaller sample could yield less precise *sales_in_thousands* estimates. Equal location counts might have improved precision, potentially altering results.

5. <u>Result interpretation:</u>

Based on the statistical test performed in step 3. we conclude the following:

- **A/B test 1: *Comparison between Promotion 1 and Promotion 2***

$$P\text{-value} = 0.00128$$

$$P\text{-value} < \alpha$$

We reject the null hypothesis that there is no difference in average sales revenue between Promotion 1 and Promotion 2 and conclude that the average sales revenue generated by Promotion 1 and Promotion 2 are different.

- **A/B test 2: *Comparison between Promotion 1 and Promotion 3***

$$P\text{-value} = 0.43$$

$$P\text{-value} > \alpha$$

We fail to reject the null hypothesis that there is no difference in average sales revenue between Promotion 1 and Promotion 3.

- **A/B test 3: *Comparison between Promotion 2 and Promotion 3***

$$P\text{-value} = 0.0136$$

$$P\text{-value} > \alpha$$

We fail to reject the null hypothesis that there is no difference in average sales revenue between Promotion 2 and Promotion 3.

Decision

Based on the t-tests, Promotion 1 significantly outperforms Promotion 2 (p = 0.00128, mean difference = 43.08 thousand), with no significant differences between Promotions 1 and 3 (p = 0.43) or Promotions 2 and 3 (p = 0.0136).

We recommend for business ***launching Promotion 1 for the full rollout***, as it generates the highest average sales per location, potentially increasing revenue by approximately 43 thousand per location compared to Promotion 2.

Suggested solution decision: Just looking at raw numbers, campaign 1 seems to be performing the best. However, we were only able to establish statistical significance when comparing it to campaign 2. For other comparisons, the p-values were too high to rule out the possibility that the differences occurred by chance, given the confidence level.

The recommendation would be to discontinue campaign 2 due in favor of either campaign 1 or 3. However, the choice between the two campaigns should be made by running another experiment, in which these promotion campaigns would be compared against one another.

Appendix

- **Query result / Data Source**
- **SQL Query 1:**

```
WITH t1 AS (SELECT Promotion,
                   location_id,
                   SUM(sales_in_thousands) AS SUM_revenue

                   FROM tc-da-1.turing_data_analytics.wa_marketing_campaign

                   GROUP BY Promotion,
                            location_id)


SELECT Promotion,
       COUNT(DISTINCT location_id) AS Location_count,
       AVG(SUM_revenue) AS Average_revenue,
       STDDEV_SAMP(SUM_revenue) AS standard_deviation,
       SUM(SUM_revenue) AS sum_revenue,

FROM t1

GROUP BY promotion
;
```

- **SQL Query 2:**

```
WITH
  week1 AS (SELECT Promotion,
```

```
                    AVG(sales_in_thousands) AS Average_revenue_week1,
                    SUM(sales_in_thousands) AS SUM_revenue_week1

           FROM tc-da-1.turing_data_analytics.wa_marketing_campaign

          WHERE week=1

          GROUP BY Promotion),

  Week2 AS (SELECT Promotion,
                AVG(sales_in_thousands) AS Average_revenue_week2,
                SUM(sales_in_thousands) AS SUM_revenue_week2

          FROM tc-da-1.turing_data_analytics.wa_marketing_campaign

          WHERE week=2

          GROUP BY Promotion),


  Week3 AS (SELECT Promotion,
                AVG(sales_in_thousands) AS Average_revenue_week3,
                SUM(sales_in_thousands) AS SUM_revenue_week3

          FROM tc-da-1.turing_data_analytics.wa_marketing_campaign

          WHERE week=3

          GROUP BY Promotion)

SELECT mc.Promotion,
      COUNT(DISTINCT location_id) AS Location_count,
      week1.Average_revenue_week1 AS Average_revenue_week1,
      week1.SUM_revenue_week1 AS sum_revenue_week1,
      week2.Average_revenue_week2 AS Average_revenue_week2,
      week2.SUM_revenue_week2 AS sum_revenue_week2,
      week3.Average_revenue_week3 AS Average_revenue_week3,
      week3.SUM_revenue_week3 AS sum_revenue_week3,
      AVG(sales_in_thousands) AS Average_revenue_week4,
      SUM(sales_in_thousands) AS SUM_revenue_week4

FROM tc-da-1.turing_data_analytics.wa_marketing_campaign mc
```

```sql
JOIN week1
  ON week1.promotion = mc.promotion
JOIN week2
  ON week2.promotion = mc.promotion
JOIN week3
  ON week3.promotion = mc.promotion

WHERE week=4

GROUP BY Promotion,
        Average_revenue_week1,
        Average_revenue_week2,
        Average_revenue_week3,
        sum_revenue_week1,
        sum_revenue_week2,
        sum_revenue_week3
;
```