Data Wrangling, Analysis, and AB Testing Final Project

The objective of this final project was to illustrate an AB test case study. The scenario included an AB test for product views; in which the independent variable was a new website landing page for a selected group of products, and and an AB test for orders in which a follow up email was given to a select group of users. The project was broken into scenarios; starting from re-formatting a work-in-progress test assignment table to show clear test assignment data and following up with calculating test binaries for orders and views. Finally, the AB tests were performed via a p-value calculator to see if there were was any significant difference within the tests. I was tasked to perform these calculations in SQL and to create a report with visualizations that illustrate the results of the AB tests.

1. Data Quality Check

	item_id	test_a	test_b	test_c	test_d	test_e	test_f
	item_id	test_a	test_b	test_c	test_d	test_e	test_f
1	2512	1	0	1	1	0	1
2	482	0	1	1	1	0	0
3	2446	0	1	1	0	1	0

This table only shows the first 1,000 rows. View complete results in Report Details.

2. Reformat the Data

	item_id	test_assignment	test_number	test_start_date
	item_id	test_assignment	test_number	test_start_date
1	2997	1	item_test_1	2013-01-05 00:00:00
2	3363	0	item_test_1	2013-01-05 00:00:00
3	1794	1	item_test_1	2013-01-05 00:00:00

This table only shows the first 1,000 rows. View complete results in Report Details.

3. Compute Order Binary

	item_id	test_number	test_assignment	order_binary
	item_id	test_number	test_assignment	order_binary
1	1412	item_test_2	0	0
2	207	item_test_2	0	1
3	1155	item_test_2	1	1

4. Compute View Item Metrics

	test_assignment	items	views	viewed_percent	?column?
	test_assignment	items	views	viewed_percent	?column?
1	0	1130	135332	81	119.762831858
2	1	1068	127454	83	119.338951311

Data Wrangling, Analysis, and AB Testing Final Project

ABBA
A/B testing statistics

Label	Number of successes	Number of trials	
Baseline	1130	135332	Remove
Variation 1	1068	127454	Remove

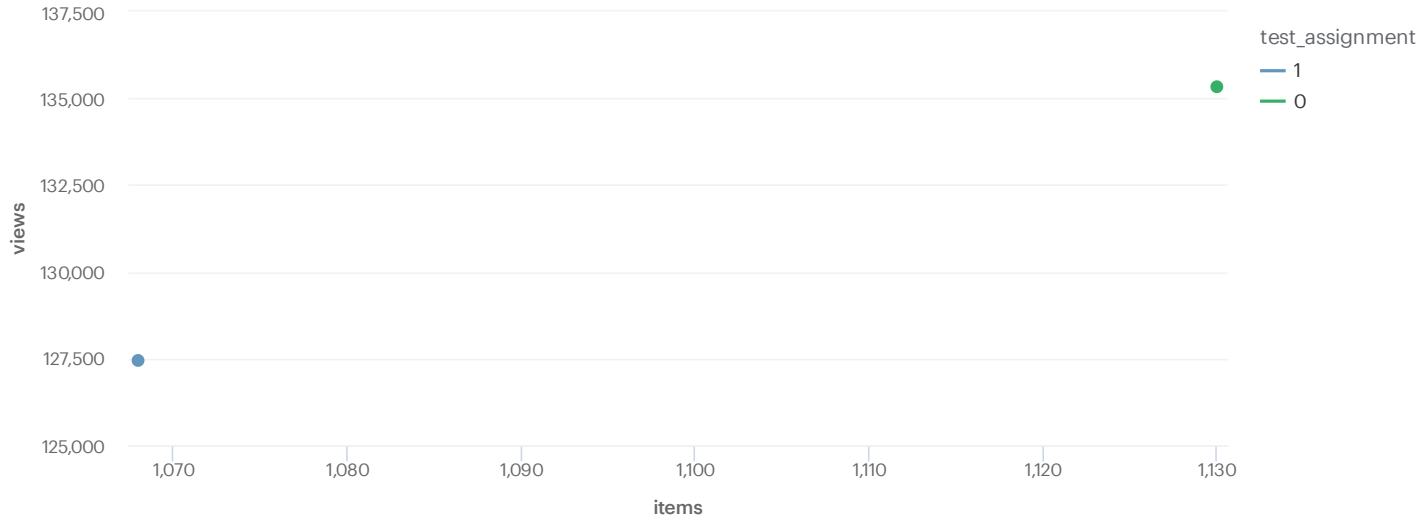
Interval confidence level: 0.95

Use multiple testing correction: ☑

Compute | Add another group

	Successes	Total	Success Rate	p-value	Improvement
Baseline	1,130	135,332	0.79% – 0.88% (0.83%)	—	—
Variation 1	1,068	127,454	0.79% – 0.89% (0.84%)	0.93	-8% – 8.7% (0.36%)

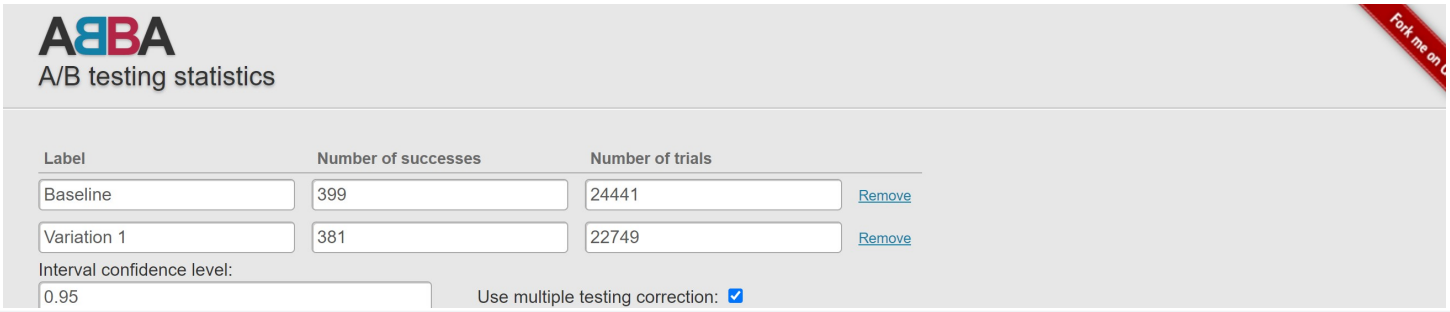
Views Binary Effect Within 30 Days

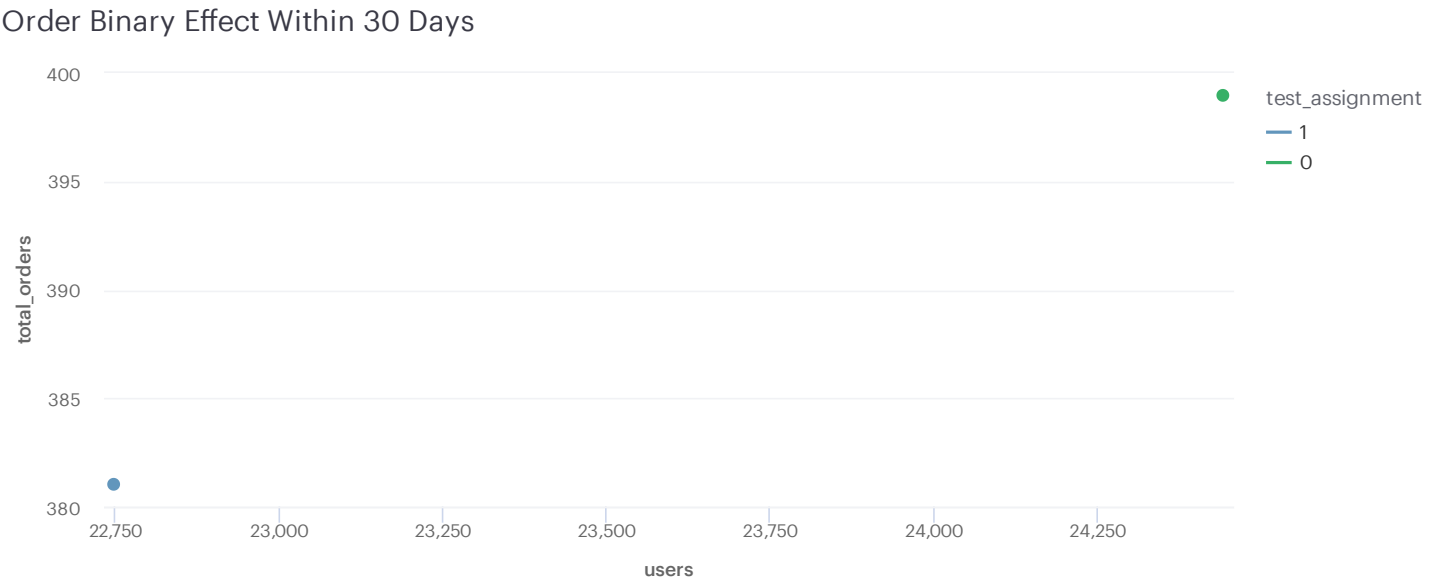


5. Compute lift and p-value

	test_assignment	users	total_orders
1	0	24441	399
2	1	22749	381

Data Wrangling, Analysis, and AB Testing Final Project





After calculating the p-value for both the orders and views binary, we are able to conclude that the effects of the independent variables for both tests were not significant. The independent variable did not produce increased views on the affected products, nor cause an increase in product orders.