

Final Assignment (Data Wrangling, Analysis and AB Testing)

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
4	1212	0	0	0	0	0	1

This table only shows the first 1,000 rows. View complete results in [Report Details](#).

Answer:

No! This data set does contain the assignments, but it does not contain all of the information needed to analyze a test. It is important that we know the dates so that we can compare results only after the test began.

2. Reformat the Data

	item_id	test_assignment	test_number	test_start_date
	item_id	test_assignment	test_number	test_start_date
1	2512	1	test_a	2020-01-01
2	482	0	test_a	2020-01-01
3	2446	0	test_a	2020-01-01
4	1212	0	test_a	2020-01-01

This table only shows the first 1,000 rows. View complete results in [Report Details](#).

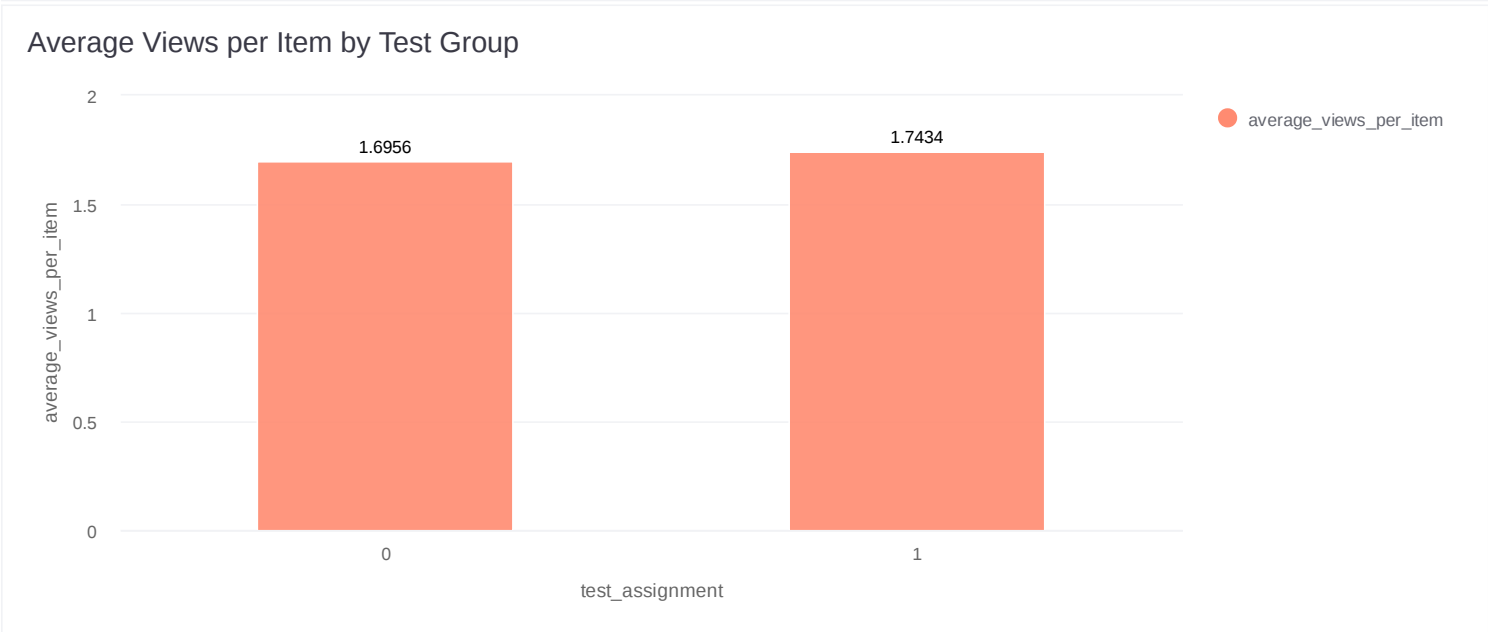
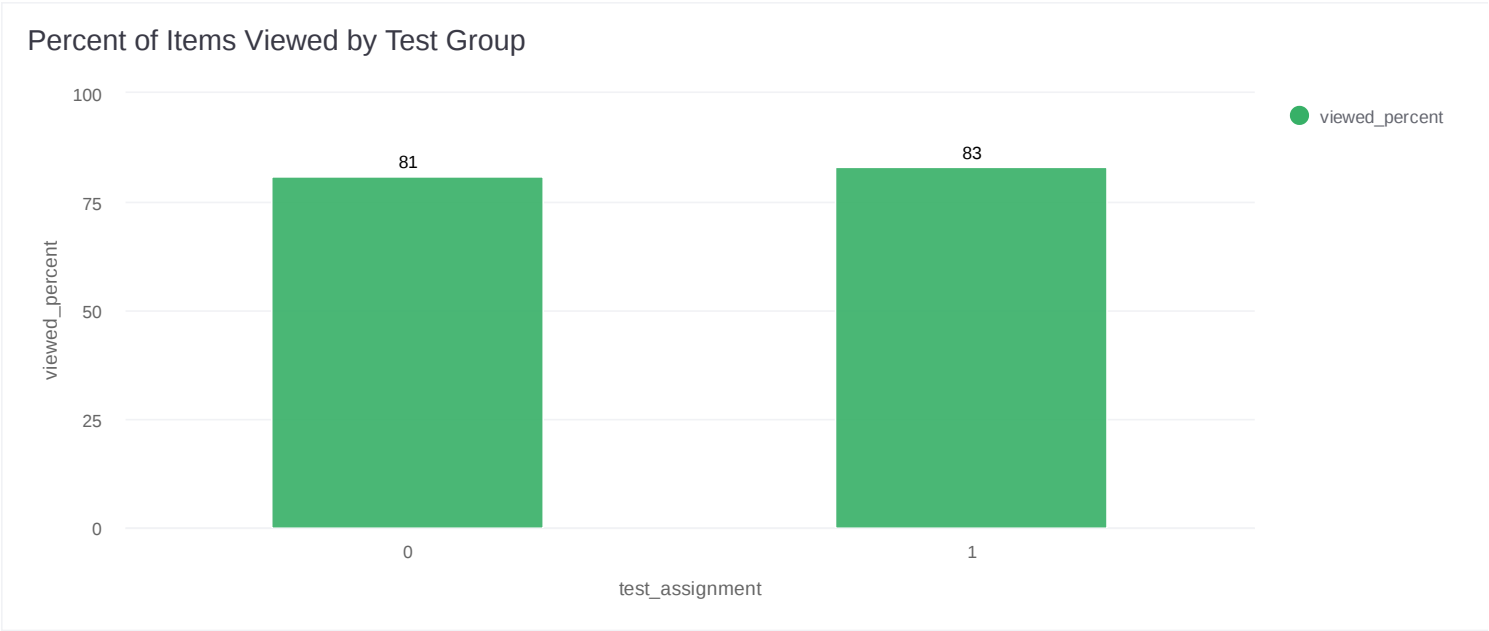
3. Compute Order Binary

	test_assignment	number_of_items	items_ordered_30d
	test_assignment	number_of_items	items_ordered_30d
1	0	1130	386
2	1	1068	363

4. Compute View Item Metrics

	test_assignment	items	viewed_items	viewed_percent	views	average_views_per_item
	test_assignment	items	viewed_items	viewed_percent	views	average_views_per_item
1	0	1130	918	81	1916	1.69557522124
2	1	1068	890	83	1862	1.74344569288

Final Assignment (Data Wrangling, Analysis and AB Testing)



5. Compute lift and p-value

	test_assignment	items	viewed_items
1	0	1130	918
2	1	1068	890

The test group 1 had 2.5% more viewed items compared to test group 0, the p-value for this lift was 0.2 which does not meet our threshold for significance. There is no statistically significant change to the metrics viewed percent as a result of the treatment.