Final Assignment Solutions

Data Quality Check (Solution)
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.

1. Data Quality Check

	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	1312	0	0	0	0	0	1
5	2550	1	1	0	1	0	0

This table only shows the first 1,000 rows.

Reformat the Data
The data should look

2. Reformat the Data

	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	1312	0	test_a	2020-01-01
5	2550	1	test_a	2020-01-01

This table only shows the first 1,000 rows.

Order Binary
Computing the order binary, should result in an aggregated table with the values below.
If the number disagree slightly check the dates within the query. Small differences in the order binary based on whether the test assignment day is included are acceptable.

Below there is a subquery with item-level order binary.

3. Compute Order Binary

	test_assignment	items	ordered_items_30d
1	0	1130	341
2	1	1068	319

3.a SubQuery

	test_assignment	item_id	order_binary_30d
1	0	0	1
2	1	3313	0
3	0	2123	1
4	1	1687	0
5	0	66	0

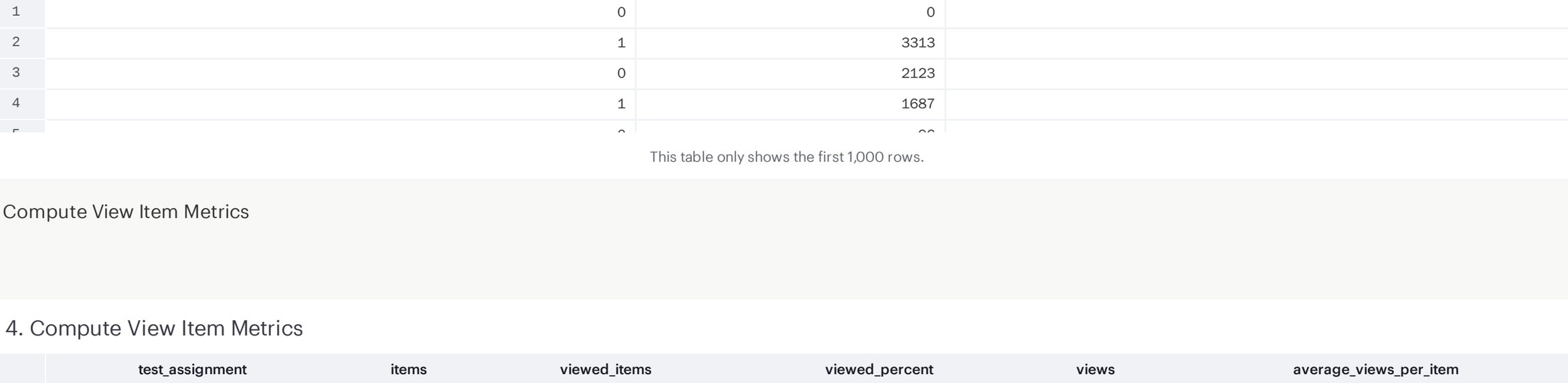
This table only shows the first 1,000 rows.

Compute View Item Metrics

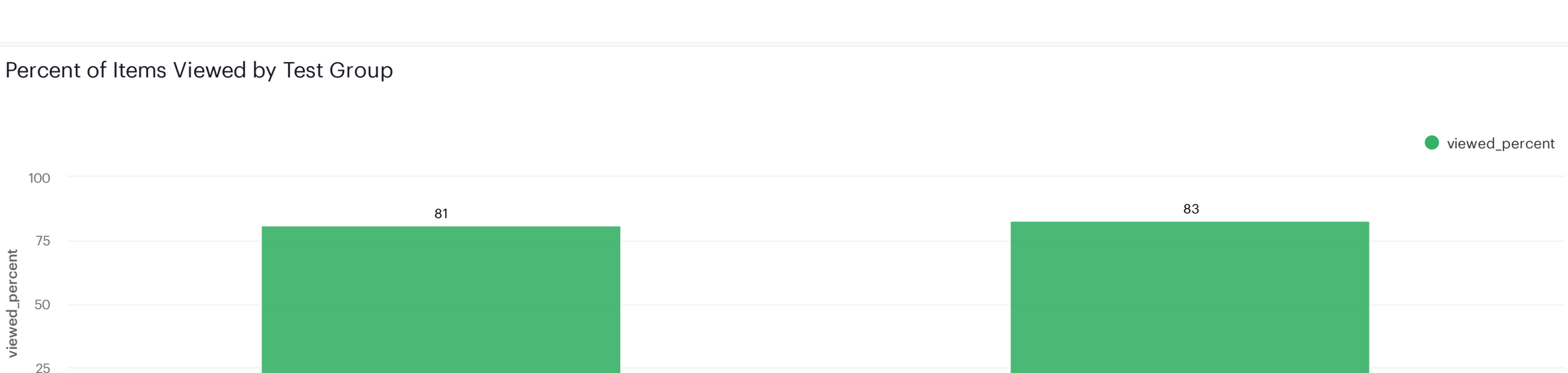
4. Compute View Item Metrics

	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

Percent of Items Viewed by Test Group



Average Views per Item by Test Group



5. Compute the lift and p-value
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.