AB Testing Project

The codes used will be written out or screenshots provided. The report builder print out will be at the end. Thank you for your review.

Exercise 1: Data Quality Check

--We are running an experiment at an item-level, which means all users who visit will see the same page, but the layout of different item pages may differ.

--Compare this table to the assignment events we captured for user_level_testing.

--Does this table have everything you need to compute metrics like 30-day view-binary?

```
SELECT
  *
FROM
  dsv1069.final_assignments_qa
```

No, the final_assignment_qa table does not have everything needed to compute metrics like 30-day view-binary on it's own. I would need to join it with another table to capture the data requested.

Exercise 2: Reformat Data

--Reformat the final_assignments_qa to look like the final_assignments table, filling in any missing values with a placeholder of the appropriate data type.

```
SELECT
  item_id,
  test_a                              AS test_assignment,
  'item_test_1'                           AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa

UNION ALL

SELECT
  item_id,
  test_b                              AS test_assignment,
  'item_test_2'                           AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa

UNION ALL

SELECT
  item_id,
  test_c                              AS test_assignment,
  'item_test_3'                           AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa
```

```
UNION ALL

SELECT
  item_id,
  test_d                                AS test_assignment,
  'item_test_4'                         AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa

UNION ALL

SELECT
  item_id,
  test_e                                AS test_assignment,
  'item_test_5'                         AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa

UNION ALL

SELECT
  item_id,
  test_f                                AS test_assignment,
  'item_test_6'                         AS test_number,
  CAST('2013-01-05 00:00:00' AS timestamp)        AS  test_start_date
FROM
  dsv1069.final_assignments_qa;
```

Exercise 3: Compute Order Binary

-- Use this table to
-- compute order_binary for the 30 day window after the test_start_date
-- for the test named item_test_2

```
SELECT
  order_binary.test_assignment,
  COUNT(order_binary.item_id) AS num_items,
  COUNT(order_binary.orders_binary_30) AS num_orders_30
FROM (
  SELECT
    test_events.test_number,
    test_events.test_assignment,
    test_events.item_id,
    MAX(CASE
        WHEN orders.created_at > test_events.test_start_date
        THEN 1
        ELSE NULL
      END) AS orders_binary,
    MAX(CASE
        WHEN orders.created_at > test_events.test_start_date
          AND orders.created_at <= test_events.test_start_date + INTERVAL '30 days'
        THEN 1
        ELSE NULL
      END) AS orders_binary_30
  FROM dsv1069.final_assignments test_events
  LEFT JOIN dsv1069.orders orders
    ON test_events.item_id = orders.item_id
  GROUP BY
    test_events.test_number,
```

```
   test_events.test_assignment,
   test_events.item_id
) order_binary
WHERE order_binary.test_number = 'item_test_2'
GROUP BY order_binary.test_assignment;
```

Exercise 4: Compute View Item Binary

-- Use this table to
-- compute view_binary for the 30 day window after the test_start_date
-- for the test named item_test_2

```
SELECT
  views_binary.test_assignment,
  COUNT(views_binary.item_id) AS num_views,
  COUNT(views_binary_30) AS num_views_30
FROM
  (
    SELECT
      views.test_number,
      views.test_assignment,
      views.item_id,
      MAX(
        CASE
          WHEN test_events.event_time > views.test_start_date THEN 1
          ELSE NULL
        END
      ) AS views_binary,
      MAX(
        CASE
          WHEN test_events.event_time > views.test_start_date
              AND test_events.event_time < views.test_start_date + INTERVAL '30 days'
          THEN 1
          ELSE NULL
        END
      ) AS views_binary_30
    FROM
      (
        SELECT
          event_id,
          event_time,
          event_name,
          MAX(
            CASE
              WHEN parameter_name = 'item_id' THEN CAST(parameter_value AS INT)
              ELSE NULL
            END
          ) AS item_id
        FROM
          dsv1069.events
        GROUP BY
          event_id,
          event_time,
          event_name
      ) test_events
      LEFT OUTER JOIN (
        SELECT
          *
        FROM
          dsv1069.final_assignments
      ) views
```

```
    ON test_events.item_id = views.item_id
  GROUP BY
    views.test_number,
    views.test_assignment,
    views.item_id
 ) views_binary
WHERE
 views_binary.test_number = 'item_test_2'
GROUP BY
 views_binary.test_assignment;
```

Exercise 5: P-Values for Binary Metrics

Use the https://thumbtack.github.io/abba/demo/abba.html to compute the lifts in metrics and the p-values for the binary metrics (30 day order binary and 30 day view binary) using a interval 95% confidence.

Orders

Baseline: 332/1130, 27-32% (29%)
Variable: 297/1068, 25-31% (28%), p-value 0.42, -18-7.5% (-5.3%)

Views

Baseline: 909/1130, 78-83% (80%)
Variable: 878/1068, 80-84% (82%), p-value 0.29, -1.9-6.2% (2.2%)

AB Testing Project

Exercise 1: Data Quality Check

	item_id COUNT DISTINCT: 2198	test_a SUM: 1086 AVG: 0.49 MIN: 0 MAX: 1	test_b SUM: 1068 AVG: 0.49 MIN: 0 MAX: 1	test_c SUM: 1123 AVG: 0.51 MIN: 0 MAX: 1	test_d SUM: 1105 AVG: 0.50 MIN: 0 MAX: 1	test_e SUM: 1078 AVG: 0.49 MIN: 0 MAX: 1	test_f SUM: 1081 AVG: 0.49 MIN: 0 MAX: 1
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	3556	1	1	0	1	0	0
6	131	0	0	0	0	1	1
7	1178	1	0	1	0	1	1
8	110	0	1	1	1	1	0

Page 1 of 22 Showing rows 1-100 of 2,198

Exercise 2: Reformat Data

	item_id COUNT DISTINCT: 2198	test_number COUNT DISTINCT: 6	test_assignment SUM: 6541 AVG: 0.50 MIN: 0 MAX: 1	test_start_date MIN: 2013-01-05 MAX: 2013-01-05
1	2512	item_test_1	1	2013-01-05 00:00:00
2	482	item_test_1	0	2013-01-05 00:00:00
3	2446	item_test_1	0	2013-01-05 00:00:00
4	1312	item_test_1	0	2013-01-05 00:00:00
5	3556	item_test_1	1	2013-01-05 00:00:00
6	131	item_test_1	0	2013-01-05 00:00:00
7	1178	item_test_1	1	2013-01-05 00:00:00
8	110	item_test_1	0	2013-01-05 00:00:00

Page 1 of 132 Showing rows 1-100 of 13,188

Exercise 3: Compute Order Binary

	test_assignment SUM: 1 AVG: 0.50 MIN: 0 MAX: 1	num_items SUM: 2198 AVG: 1099 MIN: 1068 MAX: 1130	num_orders_30 SUM: 629 AVG: 314.50 MIN: 297 MAX: 332
1	0	1130	332
2	1	1068	297

Page 1 of 1 Showing rows 1-2 of 2

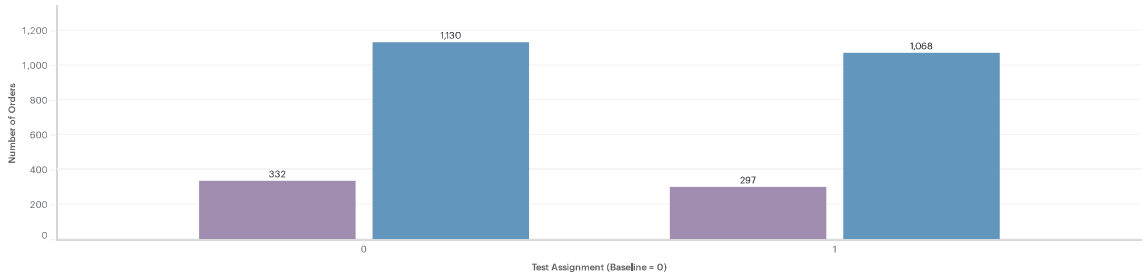
Exercise 4: Compute View Item Binary

	test_assignment SUM: 1 AVG: 0.50 MIN: 0 MAX: 1	num_views SUM: 2198 AVG: 1099 MIN: 1068 MAX: 1130	num_views_30 SUM: 1787 AVG: 893.50 MIN: 878 MAX: 909
1	0	1130	909
2	1	1068	878

Page 1 of 1 Showing rows 1-2 of 2

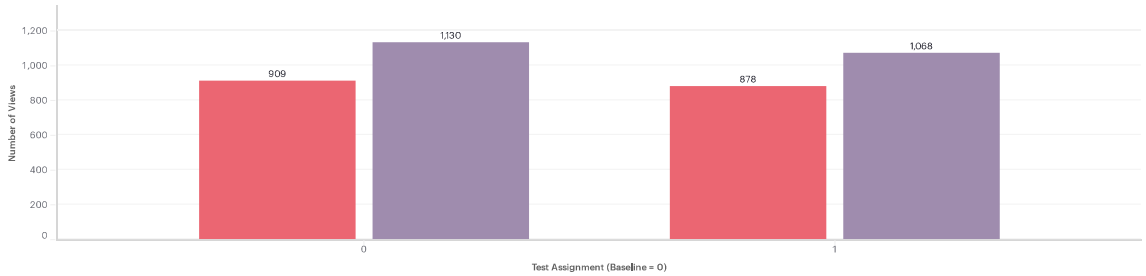
Exercise 5: Order Binary

P-value = 0.42, indicating 58% confidence level. Lift is assessed at -18-7.5%, with observed lift being -5.3%.



Exercise 5: View Items Binary

P-value = 0.29, indicating 71% confidence level. Lift is assessed at -1.9-6.2%, with observed lift being 2.2%.



Report run on May 18, 2025 at 7:35 PM UTC