A/B Test: Landing Page Experiment + Logistic Regression (Case Study)

by Konrad Kuleta

Scenario: E-commerce landing page test for new visitors from marketing channels.

Objective: Evaluate whether the new landing page (**Variant B**) increases purchase probability (**conversion**) versus **Variant A**.

Scope: End-to-end A/B test workflow in Python:

- **Sanity checks:** Sample Ratio Mismatch (SRM) to validate the random split
- **Descriptive results:** overall conversion rates (A vs B) + uplift
- **Segmentation:** uplift by **acquisition_channel** (Paid Search, Paid Social, Organic, Email, Affiliate)
- **Model-based estimate:** logistic regression to estimate the Variant B effect while controlling for acquisition channel
- **Decision framing:** ship / don't ship / run longer + key limitations (synthetic data)

Sanity checks (SRM + basic stats)

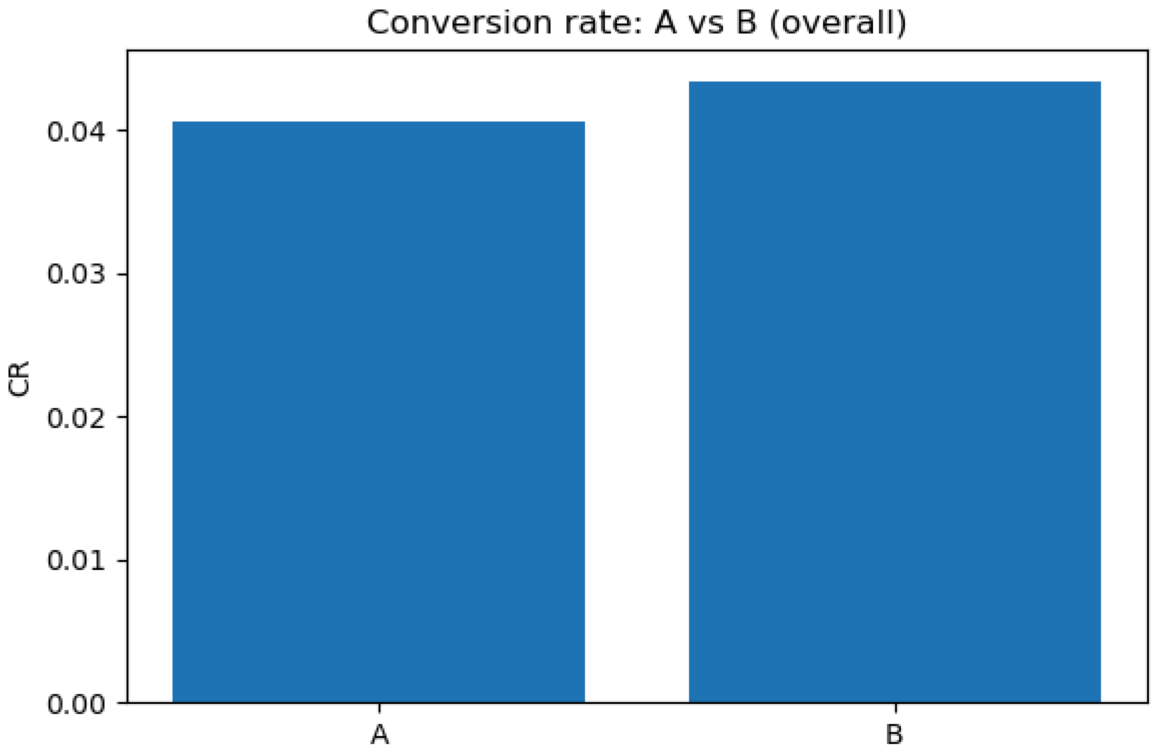
SRM (Sample Ratio Mismatch) checks whether the A/B split is ~50/50 (requirement for the test validity)

```
A/B sample sizes (number of users/sessions per variant):
variant
B    30079
A    29921
Name: count, dtype: int64
Total N = 60,000

SRM chi2: 0.4161 p-value: 0.5189062549479662
```

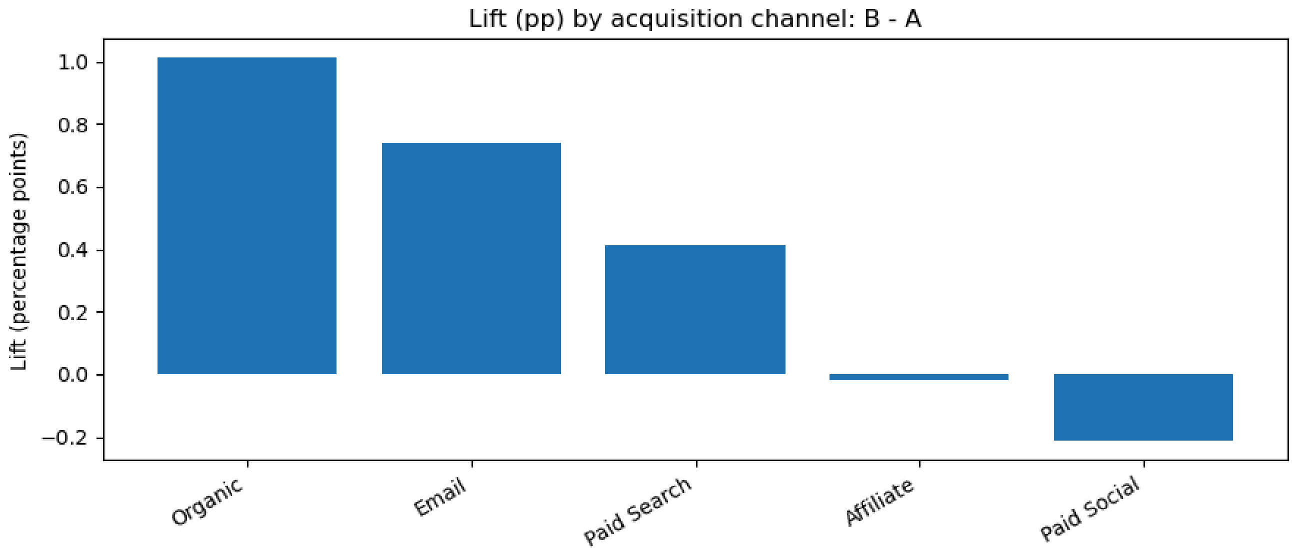
Conversion Rate per variant and acquisition channel

variant	count	sum	cr
A	29921	1213	0.040540
B	30079	1306	0.043419




```
Absolute lift (pp): 0.29 pp
Relative lift: 7.1 %
```

		count	sum	cr
acquisition_channel	**variant**			
Affiliate	**A**	3014	136	0.045123
	B	3070	138	0.044951
Email	**A**	3022	212	0.070152
	B	2992	232	0.077540
Organic	**A**	4465	213	0.047704
	B	4446	257	0.057805
Paid Search	**A**	10587	451	0.042599
	B	10552	493	0.046721
Paid Social	**A**	8833	201	0.022756
	B	9019	186	0.020623



Statistical modelling (logistic regression)

A baseline model (**Model 1**) was estimated to quantify the raw association between Variant B and conversion. Next, an adjusted model (**Model 2**) was estimated with acquisition channel included as a categorical control variable. This adjustment isolates the variant effect from differences in traffic mix, since conversion rates vary across channels. Comparing Model 1 vs Model 2 helps verify whether the observed uplift is attributable to the landing page change rather than channel composition.

Model 1 (unadjusted A/B):

`converted ~ variant_B` Measures the raw difference between variants.

```
                           Logit Regression Results
==============================================================================
Dep. Variable:              converted   No. Observations:                60000
Model:                          Logit   Df Residuals:                    59998
Method:                           MLE   Df Model:                            1
Date:                Sat, 17 Jan 2026   Pseudo R-squ.:               0.0001479
Time:                        17:27:10   Log-Likelihood:                -10450.
converged:                       True   LL-Null:                       -10452.
Covariance Type:            nonrobust   LLR p-value:                   0.07869
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -3.1641      0.029   -107.942      0.000      -3.222      -3.107
variant_B      0.0716      0.041      1.758      0.079      -0.008       0.151
==============================================================================
```

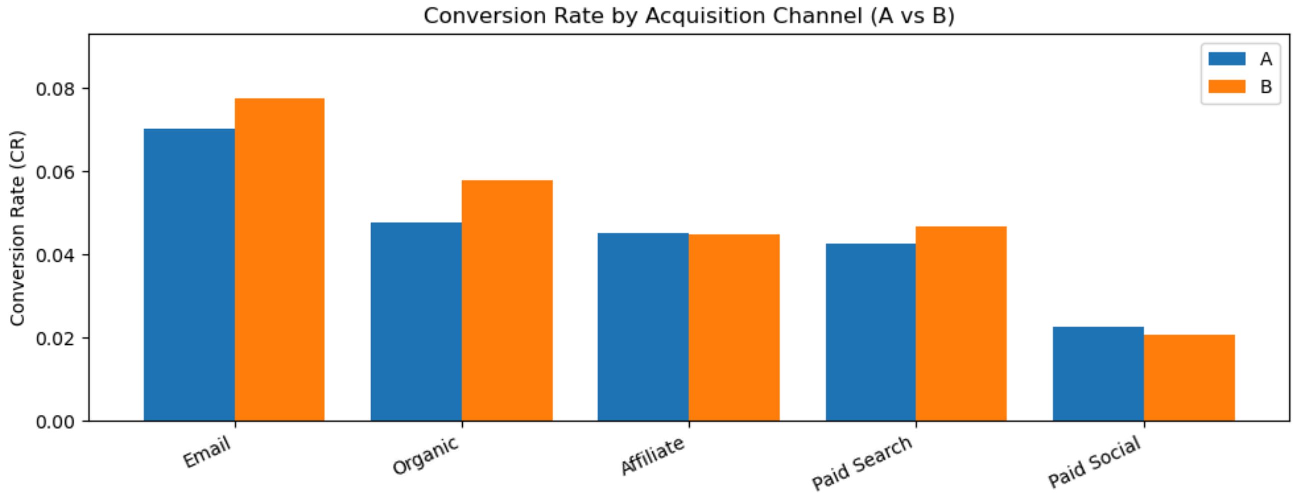
Model 2 (adjusted for acquisition channel):

`converted ~ variant_B + C(acquisition_channel)`

```
                           Logit Regression Results
==============================================================================
Dep. Variable:              converted   No. Observations:                60000
Model:                          Logit   Df Residuals:                    59994
Method:                           MLE   Df Model:                            5
Date:                Sat, 17 Jan 2026   Pseudo R-squ.:                 0.01808
Time:                        17:27:12   Log-Likelihood:                -10263.
converged:                       True   LL-Null:                       -10452.
Covariance Type:            nonrobust   LLR p-value:                 1.637e-79
==================================================================================
=======================
                                            coef    std err          z      P>|z|
[0.025      0.975]
----------------------------------------------------------------------------------
-----------------------
Intercept                                -3.0933      0.065    -47.299      0.000
-3.221      -2.965
C(acquisition_channel)[T.Email]           0.5255      0.079      6.644      0.000
0.370       0.680
C(acquisition_channel)[T.Organic]         0.1665      0.078      2.138      0.033
0.014       0.319
C(acquisition_channel)[T.Paid Search]    -0.0084      0.070     -0.120      0.904
-0.146       0.129
C(acquisition_channel)[T.Paid Social]    -0.7554      0.080     -9.396      0.000
-0.913      -0.598
variant_B                                 0.0761      0.041      1.862      0.063
-0.004       0.156
==================================================================================
=======================

Model 1 (A/B only): OR(B vs A) = 1.074 (~7.4% higher odds), 95% CI [0.992, 1.164],
p=0.079
Model 2 (+ acquisition_channel): OR(B vs A) = 1.079 (~7.9% higher odds), 95% CI
[0.996, 1.169], p=0.063
```

Channel-related CR heterogeneity



Summary

Key findings:

- **Randomization looks clean (SRM):** traffic split between A and B is ~50/50, so the test setup doesn't show obvious allocation issues.
- **Overall effect is small:** Variant **B** looks *slightly* better than **A**, but the uplift is not clearly "proven" (results are borderline, could be noise).
- **Channel matters more than the variant:** conversion rates differ a lot by **acquisition_channel** — channel effects explain more of the outcome than the landing page variant alone.
- **No universal winner by channel:** some channels show higher CR for B, others don't — treat channel splits as directional, not final truth.

Recommendations:

- *If it were real:* **no full-rollout** based on this alone. Either run longer / collect more data, or a cautious rollout with monitoring.
- **Prioritize learning where it matters most:** focus next iteration on the highest-impact channels (volume or business value).

Limitations:

- **Synthetic dataset:** patterns are simulated and may not reflect real-world tracking noise, attribution, seasonality, or user behavior.
- **Multiple comparisons risk:** slicing by channel increases the chance of seeing "differences" that are just randomness.
- **Single outcome metric:** only conversion (no revenue/AOV/LTV), so we can't judge business impact beyond CR.
- **Unobserved factors:** no explicit controls for time effects, device, returning users, etc. (possible hidden confounders).