

Pricing Test

Goal

Pricing optimization is, non surprisingly, another area where data science can provide huge value.

The goal here is to evaluate whether a pricing test running on the site has been successful. As always, you should focus on user segmentation and provide insights about segments who behave differently as well as any other insights you might find.

Challenge Description

Company XYZ sells a software for \$39. Since revenue has been flat for some time, the VP of Product has decided to run a test increasing the price. She hopes that this would increase revenue. In the experiment, 66% of the users have seen the old price (\$39), while a random sample of 33% users a higher price (\$59).

The test has been running for some time and the VP of Product is interested in understanding how it went and whether it would make sense to increase the price for all the users.

Especially he asked you the following questions:

- **Should the company sell its software for \$39 or \$59?**
- The VP of Product is interested in having a holistic view into user behavior, especially focusing on actionable insights that might increase conversion rate. **What are your main findings looking at the data?**
- [Bonus] The VP of Product feels that the test has been running for too long and he should have been able to get statistically significant results in a shorter time. Do you agree with her intuition? **After how many days you would have stopped the test?** Please, explain why.

Data

We have two tables downloadable by clicking [here](#).

The two tables are:

"test_results" - data about the test

Columns:

- **user_id** : the Id of the user. Can be joined to user_id in user_table
- **timestamp** : the date and time when the user hit for the first time company XYZ webpage. It is in user local time
- **source** : marketing channel that led to the user coming to the site. It can be:
 - ads-["google", "facebook", "bing", "yahoo", "other"]. That is, user coming from google ads, yahoo ads, etc.
 - seo - ["google", "facebook", "bing", "yahoo", "other"]. That is, user coming from google search, yahoo, facebook, etc.
 - friend_referral : user coming from a referral link of another user
 - direct_traffic: user coming by directly typing the address of the site on the browser
- **device** : user device. Can be mobile or web
- **operative_system** : user operative system. Can be: "windows", "linux", "mac" for web, and "android", "iOS" for mobile. "Other" if it is none of the above
- **test**: whether the user was in the test (i.e. 1 -> higher price) or in control (0 -> old/lower price)
- **price** : the price the user sees. It should match test
- **converted** : whether the user converted (i.e. 1 -> bought the software) or not (0 -> left the site without buying it).

"user_table" - Information about the user

Columns:

- **user_id** : the Id of the user. Can be joined to user_id in test_results table
- **city** : the city where the user is located. Comes from the user ip address
- **country** : in which country the city is located
- **lat** : city latitude - should match user city

- **long** : city longitude - should match user city

Example

Let's check the first user

head(test_results,1)

Column Name	Value	Description
user_id	604839	The Id of the user
timestamp	2015-05-08 03:38:34	The user landed on the site on May, 8 at 3 and 38AM (and 34 seconds).
source	ads_facebook	User came via Facebook ads
device	mobile	User was using mobile
operative_system	iOS	Was using iOS
test	0	Was in the control group, i.e. seeing the old price
price	39	Indeed, the price she saw was just \$39
converted	0	Alas, left the site without purchasing the software

Let's check location info for that user

subset (user_table,user_id == 604839)

Column Name	Value	Description
user_id	604839	User id. Same user as in the previous table
city	Buffalo	She was based in Buffalo when she hit the site
country	USA	The country where Buffalo is
lat	42.89	Buffalo latitude
long	-78.86	Buffalo longitude