Challenge 5: Pricing Test

Pricing optimization is another valuable use for data scientists.

You'll need to know how to identify or calculate the most relevant KPI and analyze which price level performed better for it.

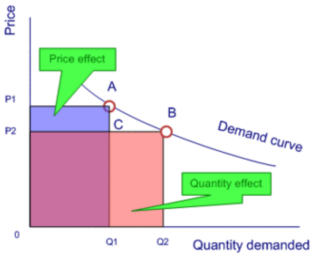
Background

Pricing is one of toughest challenges for most companies.

To begin, there are different pricing methodologies depending on the industry, product, brand power, and so on. And to make things trickier, a product can command drastically different prices depending on the context.

Let's take a bottle of water for example. For $2, you could buy a 6-pack of bottled water at a supermarket. However, those same $2 might only afford you one bottle at a movie theatre.

As a result, companies will try to find the optimal price, that which maximizes earnings (a.k.a. "gross revenue"), for their market.



Total Revenue Test (Source: Wikipedia)

When a company changes prices for a product, the quantity sold changes as well. In economics, this relationship is depicted as a demand curve (shown above).

In practice, you'll usually have 2-3 price points to analyze. You'll be fine as long as you remember that Earnings = Price * Quantity.

Data

We have one two tables.

The first is called **promo_results.csv**.

It contains 270,000 observations from a pricing test ran by a small rafting company. The company sells rafting trips down the Delaware river.

	user_id	date	source	device	OS	test	price	booked
0	74942195	2017-06-28	Unknown	desktop	windows	0	199	0
1	16114183	2017-04-23	Facebook	mobile	iOS	0	199	0
2	78795944	2017-06-27	Facebook	desktop	windows	0	199	0
3	37142290	2017-06-07	Unknown	mobile	iOS	1	299	0
4	18437714	2017-06-20	Google	desktop	windows	0	199	0

Data Dictionary:

- **user_id -** Unique ID of visitor.
- **date** - Date the visitor received the promotion.
- **source** - Marketing channel that the visitor came from.
- **device** - Mobile or desktop.
- **OS** - Visitor device's operating system.
- **test** - Did the visitor belong in the test group?
- **price** - The price seen for the rafting trip.
- **booked** - Did the visitor book a trip?

The second table is called **user_data.csv**.

It contains any information we have about the site visitor.

	user_id	state	country
0	89785143	NJ	USA
1	13450647	PA	USA
2	79577941	NY	USA
3	49333063	PA	USA
4	39218008	PA	USA

Data Dictionary:

- **user_id -** Unique ID of visitor.
- **state** - State of origin.
- **country** - Country of origin.

Objectives

For many take-home challenges, you won't be given all the details about the dataset or the background. Instead, you'll be left to discover them yourself.

For example, you haven't been told how long the test lasted or what percentage of visitors were in the test group.

- Start by understanding the dataset and the test.
- Ensure the dataset is clean before analyzing the test results.
- Which price point should the company use in the future?
- Provide actionable insights to the business. What have we learned from this test?
- (Bonus) Are the results statistically significant? *Note: Business significance is not always statistical significance, and vice-versa.*
- (Bonus) Could the company have ended the test earlier and still gotten statistically significant results?