Challenge 1: Lead Generation

In this first challenge, you'll combine 3 key skills - Exploratory Analysis, Conversion Rate Analysis, and Machine Learning - to draw fresh insights for an age-old problem.

Background

Lead generation is a challenge that businesses have faced since the dawn of capitalism. To understand lead generation, let's first introduce the idea of a **"sales funnel."**

Example Sales Funnel

A sales funnel is simply a series of expected steps that a prospect undergoes before buying. Most businesses require sales funnels because visitors are not ready to become customers right away.

Here's an example for a task-management SaaS business:

- **Awareness:** Karen, a project manager at an IT company, stumbles upon a blog post detailing 10 strategies for boosting the efficiency of meetings.
- **Lead:** She then enters her email into the opt-in box at the bottom of the blog post, which promises a PDF with 5 more strategies.
- **Prospect:** 2 weeks later, Karen receives an invitation to a webinar on how a better task management software can make her job easier. She attends.
- **Sale:** The webinar is informative and effective, so she agrees to a 3-month pilot trial for the software.

As you might guess, this process looks different for every business. Some require longer funnels that can span months (or even years) while others ask for the sale immediately.

Lead generation, or *capturing the contact info of people who could be interested in your company's product*, is often regarded as the gateway to the rest of the sales funnel.

That's why **lead generation optimization** will be likely be one of your most requested and valuable jobs as a data scientist.

Data

We have one table called **lead_gen.csv**.

It contains 280,000 observations from a multi-channel marketing campaign. Each observation represents one session from one visitor. The table excludes data from visitors who are already leads.

	source	returning	mobile	country	pages_viewed	lead
0	Taboola	1	1	US	2	0
1	Taboola	0	1	US	6	0
2	Facebook	0	0	US	2	0
3	Taboola	0	0	US	5	0
4	Taboola	0	0	Canada	5	0

Data Dictionary:

- **source -** Marketing channel that visitor came from.
- **returning** - Has the visitor been to the website before?
- **mobile** - Device (mobile / desktop). Tablets count as mobile.
- **country** - Visitor country based on IP address.
- **pages_viewed** - Number of pages viewed in the session.
- **lead** - Did the visitor opt into email list during the session?

Objectives

For this challenge, a "conversion" is defined as a visitor who became a lead.

- First, determine which sources/countries/devices had the highest conversion rates.
- Next, build a model that can predict conversion rate based on visitor information.
    - What insights can you draw from your model? Which features were the most impactful?
- Finally, provide actionable insights to the business. What have we learned from this campaign?