

TECH ALPHARETTA

2972 Webb Bridge Rd, Alpharetta, GA 30009

**Freemium A/B Testing**

In this challenge, you'll help a B2B (business-to-business) SaaS (software-as-a-service) company analyze its recent A/B test for its explainer video.

The company uses these videos to convince free users to start a premium trial.

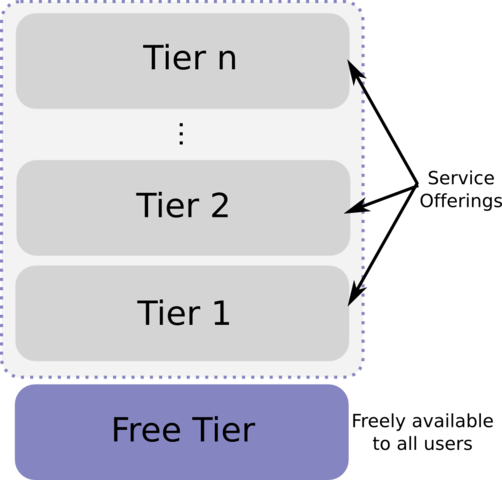
This is a very common task for data scientists in tech companies, and it's similar to the Pricing Test from Challenge #5.

**Background**

The term "freemium" refers to a service with two tiers:

* Free tier that has only basic capabilities.
* Premium tier(s) that has the full set of features.

You see this business model all the time: Trello (Gold), Amazon (Prime), and even Dropbox.



Freemium Business Model (source: Wikipedia)

The free tier is intended to lower the friction to signing up for a service and to provide a preview of what the premium tiers offer.

For many consumer-facing applications, the up-sell process is not very complicated.

However, for B2B software, it can be trickier to convince free users to upgrade to premium plans due to a variety of reasons. Two of the most common are:

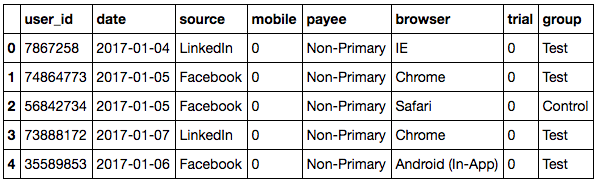
1. Users may first need approval from a manager or budgeting department.
2. Users might not understand what the premium plan offers, especially if it introduces new services and features.

To address that second obstacle, some companies will produce explainer videos that educate the user on the additional benefits for the premium tier.

**Data**

The first table is called **ab\_trial\_results.csv**.

It contains 350,000 observations from an A/B test. Those in the control group were shown the old explainer video and those in the test group were shown the new one.

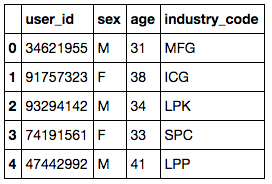


*Data Dictionary:*

* **user\_id** - Unique ID for user.
* **date** - Date the user watched explainer video.
* **source**-Marketing channel user came from.
* **mobile** - Was user on a mobile device?
* **payee** - Whether the user is the primary decision-maker for budgeting.
* **browser** - The user's browser.
* **trial** - Did the user convert, i.e. start a premium trial?
* **group** - Group (test / control).

The second table is **user\_industry.csv**.

It contains each user's industry code.



*Data Dictionary:*

* **user\_id**-Unique ID of the user.
* **sex** - User gender (M/F).
* **age** - User age (21 - 65).
* **industry\_code** - Three letter industry code.

**Objectives**

For this challenge, a "conversion" is defined as a user who began a premium trial

* Determine which sources/devices/industries had the highest conversion rates.
* Build a model that can predict conversion rate based on user data.
  + For high conversion rate users, what are the implications for the company's marketing team?
  + For low conversion rate users, what are the implications for the company's customer success teams?
* Provide actionable insights to the business. What have we learned from this test?