**DATA MANIPULATION WITH PANDAS**

**A/B Testing for ShoeFly.com**

Our favorite online shoe store, ShoeFly.com is performing an A/B Test. They have two different versions of an ad, which they have placed in emails, as well as in banner ads on Facebook, Twitter, and Google. They want to know how the two ads are performing on each of the different platforms on each day of the week. Help them analyze the data using aggregate measures.

If you get stuck during this project or would like to see an experienced developer work through it, click “**Get Unstuck**“ to see a **project walkthrough video**.

### Tasks

**11/11 Complete**

Mark the tasks as complete by checking them off

## Analyzing Ad Sources

**1.**

Examine the first few rows of ad\_clicks.

Stuck? Get a hint

**2.**

Your manager wants to know which ad platform is getting you the most views.

How many views (i.e., rows of the table) came from each utm\_source?

Stuck? Get a hint

**3.**

If the column ad\_click\_timestamp is not null, then someone actually clicked on the ad that was displayed.

Create a new column called is\_click, which is True if ad\_click\_timestamp is not null and False otherwise.

Stuck? Get a hint

**4.**

We want to know the percent of people who clicked on ads from each utm\_source.

Start by grouping by utm\_source and is\_click and counting the number of user\_id‘s in each of those groups. Save your answer to the variable clicks\_by\_source.

Stuck? Get a hint

**5.**

Now let’s pivot the data so that the columns are is\_click (either True or False), the index is utm\_source, and the values are user\_id.

Save your results to the variable clicks\_pivot.

Stuck? Get a hint

**6.**

Create a new column in clicks\_pivot called percent\_clicked which is equal to the percent of users who clicked on the ad from each utm\_source.

Was there a difference in click rates for each source?

Stuck? Get a hint

## Analyzing an A/B Test

**7.**

The column experimental\_group tells us whether the user was shown Ad A or Ad B.

Were approximately the same number of people shown both adds?

Stuck? Get a hint

**8.**

Using the column is\_click that we defined earlier, check to see if a greater percentage of users clicked on Ad A or Ad B.

Stuck? Get a hint

**9.**

The Product Manager for the A/B test thinks that the clicks might have changed by day of the week.

Start by creating two DataFrames: a\_clicks and b\_clicks, which contain only the results for A group and B group, respectively.

Stuck? Get a hint

**10.**

For each group (a\_clicks and b\_clicks), calculate the percent of users who clicked on the ad by day.

Stuck? Get a hint

**11.**

Compare the results for A and B. What happened over the course of the week?

Do you recommend that your company use Ad A or Ad B?