Ads Analysis

Goal

Maybe the first industry to heavily rely on data science was the online ads industry. Data Science is used to choose which ads to show, how much to pay, optimize the ad text and the position as well as in countless of other related applications.

Optimizing ads is one of the most intellectually challenging jobs a data scientist can do. It is a really complex problem given the huge (really really huge) size of the datasets as well as number of features that can be used. Moreover, companies often spend huge amounts of money in ads and a small ad optimization improvement can be worth millions of dollars for the company.

The goal of this project is to look at a few ad campaigns and analyze their current performance as well as predict their future performance.

Challenge Description

Company XYZ is a food delivery company. Like pretty much any other site, in order to get customers, they have been relying significantly on online ads, such as those you see on Google or Facebook.

At the moment, they are running 40 different ad campaigns and want you to help them understand their performance.

Specifically, you are asked to:

- If you had to identify the 5 best ad groups, which ones would they be? Which metric did you choose to identify the best ad groups? Why? Explain the pros of your metric as well as the possible cons.
- For each group, predict how many ads will be shown on Dec, 15 (assume each ad group keeps following its trend).
- Cluster ads into 3 groups: the ones whose avg_cost_per_click is going up, the ones whose avg_cost_per_click is flat and the ones whose avg_cost_per_click is going down.

Data

We have 1 table downloadable by clicking **here**.

The table is:

```
"ad_table" - aggregate information about ads
```

Columns:

- **date** : all data are aggregated by date
- **shown** : the number of ads shown on a given day all over the web. Impressions are free. That is, companies pay only if a user clicks on the ad, not to show it
- **clicked** : the number of clicks on the ads. This is what companies pay for. By clicking on the ad, the user is brought to the site
- **converted** : the number of conversions on the site coming from ads. To be counted, a conversion has to happen on the same day as the ad click.
- **avg_cost_per_click** : on an average, how much it cost each of those clicks
- **total_revenue** : how much revenue came from the conversions
- **ad** : we have several different ad groups. This shows which ad group we are considering

Example

```
    Let's check one ad group for one day:
```

head(ad_table, 1)

Column Name	Value	Description
date	2015-10-01	aggregate stats are about Oct, 1.
shown	65877	all over the web, this ad group was shown 65877 times
clicked	2339	it received 2339 clicks. Clicks/Shown, often called click-through-rate, was therefore about 3.5%
converted	43	of those 2339 users coming to the site on Oct,1 via the ad, 43 converted on the same day. That is, they bought something on Oct, 1.
avg_cost_per_click	0.9	on an average, each click cost 0.9 USD
total_revenue	641.62	those 43 conversions, in total, generated 641.62 USD
ad	ad_group_1	this is the ad group 1. It is one of the many different ad variations we have.