This is an attribution case study for online advertisement.

## **TASK**

An advertiser launched advertisement campaigns for its product across different online channels such as display (e.g. Yahoo, US News), paid search (e.g. Google, Bing), social networks (e.g. Facebook, Pinterest), affiliate (e.g. coupon websites), etc. The advertiser wants to understand which online channel (e.g. paid search) has contributed to conversion the most.

You are given the data containing information of users who were exposed to any of advertiser’s advertisements such as which advertisement(s) she/he was exposed to, when she/he was exposed to the advertisement(s), attributes of the advertisement(s), whether she/he converted or not, etc.

Your task is to analyze the data and provide attribution insights at the channel level (i.e. which channel drove how much conversions).

## **DATA**

In the tab-separated-value (TSV) file attached, there are 50,000 samples with 6 fields as follows:

* user\_id – Unique user ID.
* conversion\_id – Unique conversion ID. Empty if the user hasn’t converted
* repeat – You can ignore this field.
* event\_timestamp – Time of conversion if user has converted (i.e. conversion\_id is not empty). Otherwise it is time of when this entry is recorded.
* run\_timestamp – You can ignore this field.
* advertisements – List of advertisements that the user has been exposed. It is formatted as follows:

|  |
| --- |
| [            [              {'channel':'web', 'dvceType':'Tablet', ‘action’: ‘imp’, …},              [                ['ad\_id1', 'ad\_timestamp1'],                ['ad\_id2', 'ad\_timestamp2']              ]            ],            [              {'channel':'search', 'publisherId':'google', ‘action’: ‘click’, …},              [                ['ad\_id3', 'ad\_timestamp3']              ]            ]          ] |
|  |

* + In this example, the user was exposed to two different advertisements: she/he saw (‘action’: ‘imp’ - i.e. impression) the first advertisement twice and clicked (‘action’: ‘click’) the second advertisement once.

## **HINT**

You can build a conversion model, which is similar to click-through-rate (CTR) prediction model, first that predicts whether a user would convert or not given advertisements. Then you can use the conversion model to get the contribution of each channel to conversion.

You don’t need to use all advertisement attributes for this task – channel level attribution.