### **Смысл uplift**

В решении о том, кому стоит отправлять СМС, возможны 3 различные ситуации с различными эффектами:

**+1**: клиент совершил покупку после коммуникации, но если бы ее не было — не совершил бы;

**0**: клиент совершит/не совершит покупку независимо от коммуникации;

**-1**: клиент не совершил покупку после коммуникации, но если бы коммуникации не было — совершил бы.

**Задача**

**X5** имеет возможность отправлять СМС клиентам, для того чтобы стимулировать их совершать покупки. Понятно, что имеет смысл делать коммуникацию только по тем клиентам, кто без нее не совершил бы покупку, а после нее - совершит. Необходимо разработать алгоритм, который сможет успешно предсказывать, каким клиентам стоит отправлять СМС, а каким нет.

**Данные**

Участникам предоставляется несколько наборов данных для решения этой задачи. Общие данные:

* clients.csv: информация о клиентах;
* products.csv: информация о товарах;
* purchases.csv: история покупок клиентов до момента рассылки.

Данные, относящиеся непосредственно к задаче:

* uplift\_train.csv: набор клиентов для обучения, с указанием treatment\_flg — была ли совершена коммуникация, target — была ли совершена покупка после совершения коммуникации;
* uplift\_test.csv: список клиентов, для которых необходимо оценить **uplift**;
* uplift\_sample\_submission.csv: пример файла с предсказаниями.
* **Формат решений**

Для построения модели, участникам предоставляется обучающий набор клиентов uplift\_train.csv с информацией о наличии коммуникации (treatment\_flg) и совершение покупки (target).

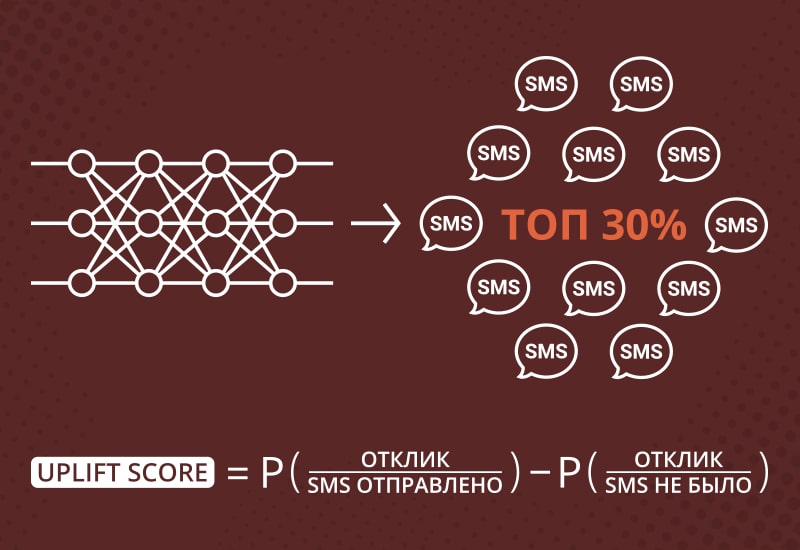
Необходимо для каждого клиента из тестовой выборки uplift\_test.csv оценить эффективность коммуникации (**uplift**). Имеет значение порядок оценок, а не их абсолютные значения. Результат должен быть представлен в виде CSV-файла с колонками client\_id и uplift.

  client\_id,uplift  
008fb49e3a,0.1149912020897228  
0095340acc,0.8353208872466903  
015c0b4d79,0.3085840952650095  
...  
ff70c360ad,0.0809048695228205  
ff86a1311b,0.4815832858531034  
ffcccc2cc4,0.10523347182011245

Участникам также предоставляется базовый пример решения от организаторов.

**Критерий качества**

В задачах моделирования **uplift-а** клиенты из тестовой выборки ранжируются по убыванию эффективности коммуникации. Из ранжированного списка выбирается **топ 30%** (наиболее перспективные). По выбранным 30% оценивается средняя добавленная конверсия. Простым языком, вычисляется средний прирост отклика при воздействии на клиента.



### **Сайт соревнования**

<https://retailhero.ai/c/uplift_modeling/overview>