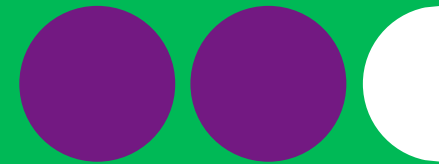


МЕГАФОН



ОЛИМПИАДА ДЛЯ ПЕРВЫХ: ДАННЫЕ

Файлы с характеристиками физических лиц

1) **gsom_interests.csv** агрегат месячных интересов за февраль, январь и декабрь.

- **Month** - месяц
- **client_id** – id абонента,
- Колонки от 'pharmacies_and_optics' до 'movies_and_series_tv' – это счетчик кол-ва возникших интересов за выбранный месяц

2) **gsom_features.csv** файл со следующими колонками:

- **month** - дата снимка витрины
- **client_id** - id абонента. Один абонент может встречаться несколько раз в витрине интересов в разных месяцах
- **city** - домашний город абонента
- **gender** - пол абонента
- **age** - возраст абонента в годах
- **mou_in** - входящий голосовой трафик абонента за месяц (month) в минутах
- **mou_out** - исходящий голосовой трафик абонента за месяц (month) в минутах
- **dou** - интернет-трафик абонента за месяц, МБ (Мегабиты)
- **admin_distr** - домашний административный район/округ города для абонента

Файлы с указанными на слайде данными
можно найти здесь ► [Data](#)



Файл с характеристиками юридических лиц

hs_dataset.csv - файл содержит в себе данные агрегат по компаниям в разрезе город-район-ОКВЭД.

- **id** – id клиента
- **cnt_large_companies** – Количество компаний с описанием «Крупные предприятия»
- **cnt_medium_companies** – Количество компаний с описанием «Средние предприятия»
- **cnt_small_companies** – Количество компаний с описанием «Малые предприятия»
- **cnt_micro_companies** – Количество компаний с описанием «Микропредприятие»
- **cnt_companies_with_workers_range_from_A_to_B** - Количество компаний с количеством сотрудников в интервале от *A* до *B*
- **lifetime_less_3_years** – Количество компаний, с регистрацией меньше 3-х лет
- **lifetime_less_5_years** – Количество компаний, с регистрацией меньше 5-ти лет по данным
- **lifetime_less_10_years** – Количество компаний, с регистрацией меньше 10-ти лет по данным
- **lifetime_great_10_years** – Количество компаний, с регистрацией больше 10-ти по данным
- **url_cnt, url_inc_volume, url_out_volume** – сумма общего, входящего и исходящего трафика
- **sk_subs_power_trans, sk_subs_power_url** - показатель от 0 до 1, где находится абонент по транзакциям , потреблению относительно других абонентов, чем больше, тем ближе к 1
- Далее идут признаки, характеризующие интересы клиентов

Файл с указанными на слайде данными
можно найти здесь ► [Data](#)

