**Описание кейса:**

Для контроля рисков расходования финансовых средств в компании необходимо прогнозировать, какой объем POS-транзакций клиентов в последующие 3 месяца после наличия информации по его транзакциям ранее.

Инженеры в компании подготовили набор данных для разработки модели. В нем содержится информация о транзакциях клиента и объеме оборота POS-транзакций клиента в последующие 3 месяца. Необходимо реализовать регрессионную модель прогноза объема оборота POS‑транзакций клиента.

**Структура входных данных (файл train.csv в** [**архиве**](https://drive.google.com/file/d/1yO8J9qAYUviDLZ-U8ny-aj7u8S49SzjU/view?usp=sharing)**):**

* PERIOD – месяц транзакций;
* cl\_id – уникальный идентификатор клиента, для которого необходимо построить прогноз (один прогноз для одного клиента);
* MCC – mcc код транзакции;
* channel\_type – зашифрованный тип канала рекламы на клиента;
* currency – код валюты транзакции;
* TRDATETIME – дата и время в месяце, когда была совершена транзакция;
* amount – сумма в валюте транзакции;
* trx\_category – тип транзакций, которые встречаются у клиента. Мы прогнозируем только POS, но может другие категории влияют на прогноз;
* target\_sum – прогнозируемая переменная (объем оборота POS-транзакций клиента в последующие 3 месяца). Для одного клиента cl\_id строим один прогноз.

**Задание:**

1. Необходимо разработать регрессионную модель прогноза объема POS-транзакций клиентов.
2. Оценить качество модели на обучающей выборке (файл train.csv в [архиве](https://drive.google.com/file/d/1yO8J9qAYUviDLZ-U8ny-aj7u8S49SzjU/view?usp=sharing)). Рассчитать меру ошибки, аргументировать ее выбор.
3. Рассчитать прогноз на тестовой выборке (файл test.csv в [архиве](https://drive.google.com/file/d/1yO8J9qAYUviDLZ-U8ny-aj7u8S49SzjU/view?usp=sharing)) и сохранить результаты прогноза в файл test\_result.csv со столбцами: cl\_id, target\_sum.
4. Рекомендуется реализовать работу в инструменте Jupyter Notebook с комментариями по реализуемой логике и визуализацией при построении выводов о данных. Пример кода приведен в файле regression\_baseline.ipynb в [архиве](https://drive.google.com/file/d/1yO8J9qAYUviDLZ-U8ny-aj7u8S49SzjU/view?usp=sharing).