**Тестовое задание на позицию «скоринг-аналитика»**

Вам нужно:

1. Провести исследовательский анализ данных по анкетным признакам клиентов.
2. Найти по анкетным признакам сегмент клиентов, у которых дефолтность за весь период менее 25% (количество клиентов в сегменте должно быть больше 1000).
3. Взять, найденный в предыдущем задании, сегмент клиентов и построить по нему таблицу, в которой будет отображены дефолтность и доходность по каждой неделе.

Результат сохранить как excel-файл.

1. Построить ML-модель бинарной классификации, целевым событием для которой является наступление дефолта с просрочкой в 30 дней. Результат оцените метрикой качества Gini.

Результат работы следует предоставить как тетрадку Jupyter Notebook с последовательно выполненными ячейками.

**Вам поможет:**

**Дефолтность** – показатель, который характеризует какой процент клиентов ничего не заплатил до определённого дня просрочки. Для выполнения задания исследуйте дефолтность на 30-ый день просрочки.

**Доходность** – отношение того сколько собрали денег с клиентов к тому, сколько выдали.

Известно, что преимущественно обращаются за кредитом мужчины

**Описание данных**

Вам предоставлены данные по первичным клиентам за первую половину 21-го года.

Данные представлены в трёх таблицах:

**Anket:**

**DraftId** – уникальный индентификатор заявки

**formula\_id** – номер модели, которая скорила заявку

**flagFC** – флаг наличия ответа от партнера1

**flagEF**– флаг наличия ответа от партнера2

**flagCKB**– флаг наличия ответа от партнера3

**Amount** – сумма займа, которую выдали клиенту

**ActivityType** – вид деятельности

**Education** – образование

**Gender** – пол

**LoanPurpose** – цель займа

**MaritalStatus** – семейное положение

**MonthlyIncome** – ежемесячный доход клиента

**Post** – занимаемая должность

**Payments:**

**DraftId** – уникальный индентификатор заявки

**ApprovalCreateTime** – время, когда заявка была подана

**ApprovalAmount** – одобренная сумма займа

**RequestAmount** – запрашиваемая сумма займа

**FormulaId** - номер модели, которая скорила заявку

**Answer** – ответ по скорингу

**LoanAmount** – сумма выданного займа

**IssueDate** – дата выдачи займа

**Duration** – количество дней на сколько выдан займ

**RetireDate** – дата погашения по договору

**RealRetireDate** – реальная дата погашения

**LoanStatus** – статус займа

**Balance** – сумма задолженность по займу

**Payment** – сумма всех выплат по займу

**DayOfDelay** – количество дней просрочки по займу

**Amount00** – сумма, выплаченная клиентом до нулевого дня просрочки

**Amount15** – сумма, выплаченная клиентом между 0 и 15 днем просрочки

**Amount30** – сумма, выплаченная клиентом между 15 и 30 днем просрочки

Данные в таблице **ML** являются признаками из кредитной истории клиента, которые помогут при построении модели.