**Исследование надёжности заёмщиков**

## **Описание проекта**

Заказчик — кредитный отдел банка. Нужно разобраться, влияет ли семейное положение и количество детей клиента на факт погашения кредита в срок. Входные данные от банка — статистика о платёжеспособности клиентов.

Результаты исследования будут учтены при построении модели **кредитного скоринга** — специальной системы, которая оценивает способность потенциального заёмщика вернуть кредит банку.

## **Инструкция по выполнению**

**Шаг 1**. **Откройте таблицу и изучите общую информацию о данных**

Путь к файлу: /datasets/data.csv. [Скачать датасет](https://code.s3.yandex.net/datasets/data.csv)

**Шаг 2**. **Предобработка данных**

1. определите и заполните пропущенные значения:
   * опишите, какие пропущенные значения вы обнаружили;
   * приведите возможные причины появления пропусков в данных;
   * объясните, по какому принципу заполнены пропуски;
2. замените вещественный тип данных на целочисленный:
   * поясните, как выбирали метод для изменения типа данных;
3. удалите дубликаты:
   * поясните, как выбирали метод для поиска и удаления дубликатов в данных;
   * приведите возможные причины появления дубликатов;
4. выделите леммы в значениях столбца с целями получения кредита:
   * опишите, как вы проводили лемматизацию целей кредита;
5. категоризируйте данные:
   * перечислите, какие «словари» вы выделили для этого набора данных, и объясните, почему.

В данных могут встречаться артефакты — значения, которые не отражают действительность. Например, отрицательное количество дней трудового стажа. Для реальных данных — это нормально. Нужно описать возможные причины появления таких данных и обработать их.

**Шаг 3. Ответьте на вопросы**

* Есть ли зависимость между наличием детей и возвратом кредита в срок?
* Есть ли зависимость между семейным положением и возвратом кредита в срок?
* Есть ли зависимость между уровнем дохода и возвратом кредита в срок?
* Как разные цели кредита влияют на его возврат в срок?

Ответы сопроводите интерпретацией — поясните, о чём именно говорит полученный вами результат.

**Шаг 4.** **Напишите общий вывод**

**Оформление:** Задание выполните в Jupyter Notebook. Программный код заполните в ячейках типа code, текстовые пояснения — в ячейках типа markdown. Примените форматирование и заголовки.

## **Описание данных**

* children — количество детей в семье
* days\_employed — общий трудовой стаж в днях
* dob\_years — возраст клиента в годах
* education — уровень образования клиента
* education\_id — идентификатор уровня образования
* family\_status — семейное положение
* family\_status\_id — идентификатор семейного положения
* gender — пол клиента
* income\_type — тип занятости
* debt — имел ли задолженность по возврату кредитов
* total\_income — ежемесячный доход
* purpose — цель получения кредита