Рубежный Контроль 1

**Описание проекта**

Нужно выяснить, влияет ли семейное положение и количество детей клиента на факт погашения кредита в срок.

Есть ли связь между наличием детей и возвратом кредита в срок?

Есть ли связь между семейным положением клиента и возвратом его/ее кредита в срок?

Влияет ли уровнем дохода на возврат кредита в срок?

Какие типы кредитов больше всего возвращаются в срок, а какие нет?

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

У вас есть данные в файле data.csv

**Краткая инструкция**

1. Соответствие данных к типам данных по каждой колонке.
2. Определить пропущенные значения, как думаете из-за чего есть пропуски в данных?
3. В данных с числами нужно проверить соответствует ли данные к типу int или float, исправить если есть несоответствия.
4. Определить дубликаты в данных, и исправить, как думаете из-за чего возможны дубликаты в данных, именно в данных с которым работаете?
5. Замена данных при их отсутствии, при их не точности (к примеру в данных могут быть отрицательные значения), чем можем заменить? Нужно описать возможные причины появления таких данных и обработать их.
6. Что можно делать с не исправляемыми данными, удаляем, заменяем или ваши варианты? Если удалить можно ли делать выводы (ответить на вопрос проекта) по оставшимся данным?

**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

Буду проверять:

Как описываете каждый проделанный шаг, оставляете ли комментарий к шагам?

Как описываете найденные в данных проблемы, ошибки, и как их решаете

Какие выводы делаете?