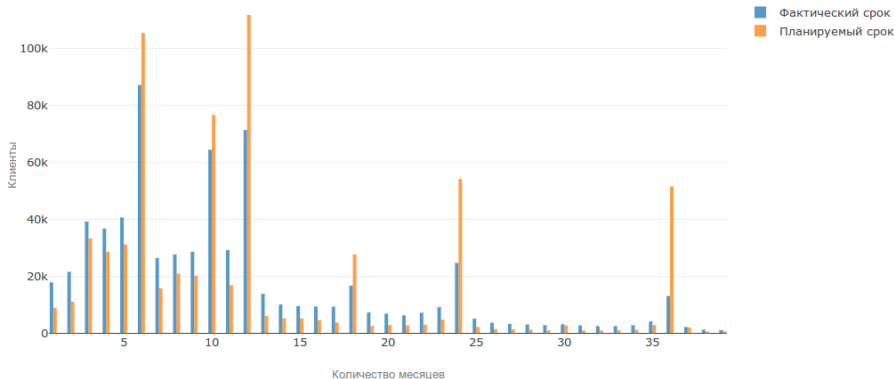


# Отчет по заданию «Закономерности в кредитном скоринге»

Таскынов Ануар, 517 группа

Курс «Прикладные задачи анализа данных»

18 октября 2017 г.






Столбец с планируемой датой окончания имеет 27% пропусков, с фактической датой окончания 4% пропусков, а столбец с датой выдачи кредита заполнен полностью. Данная гистограмма показывает, что популярным сроком кредита являются 6, 12, 18, 24, 36 месяцев. Учитывая данную закономерность можно заполнить пропуски в данных.

Одним из хороших признаков в кредитном скоринге является сумма ежемесячного платежа. Однако для **26%!** обучающей выборки данные по этому признаку неизвестны, **НО** известна сумма кредита (с точностью до 29 пропусков в трейне<sup>1</sup>). Можно по сроку кредита и сумме кредита посчитать ежемесячную сумму платежа.

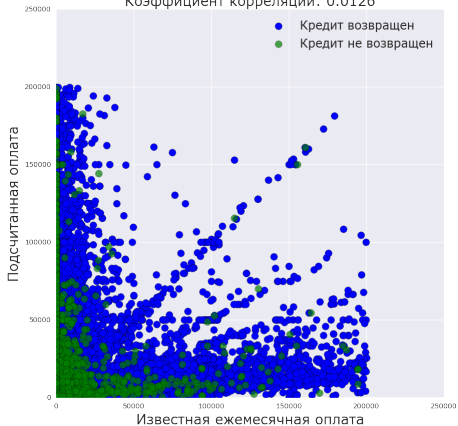
**Вопрос:** насколько скоррелированы подсчитанная ежемесячная оплата и известная? Если будет высокая корреляция, то можно будет заменить пропуски в данных.

---

<sup>1</sup>Интересно, что пропуски по сумме кредита как в трейне так и в тесте  исходят от 3 источника  

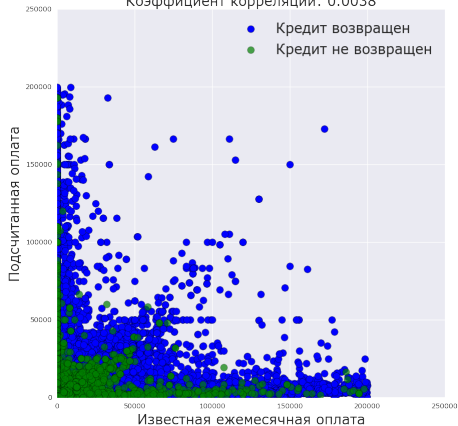
# Анализ ежемесячного платежа.

Коэффициент корреляции: 0.0126



Фактический срок

Коэффициент корреляции: 0.0038



Планируемый срок

**Неудача!**

Видна группировка не возвращенных кредитов вокруг маленьких ежемесячных сумм.

- Можно заполнить даты окончания кредитов по популярности сроков кредита.
- Нецелесообразно использовать подсчитанную ежемесячную оплату.