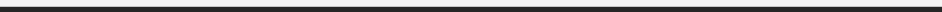


Анализ POS кредитов.

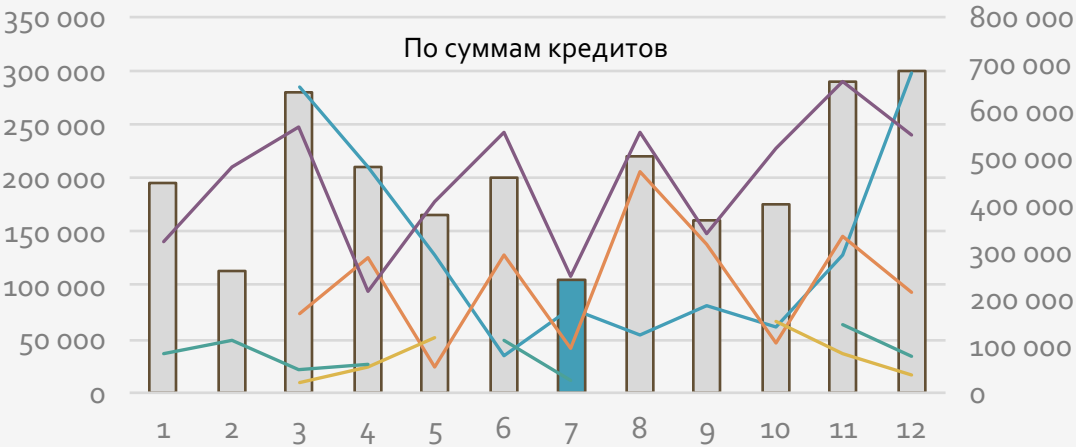


Саммари анализа.

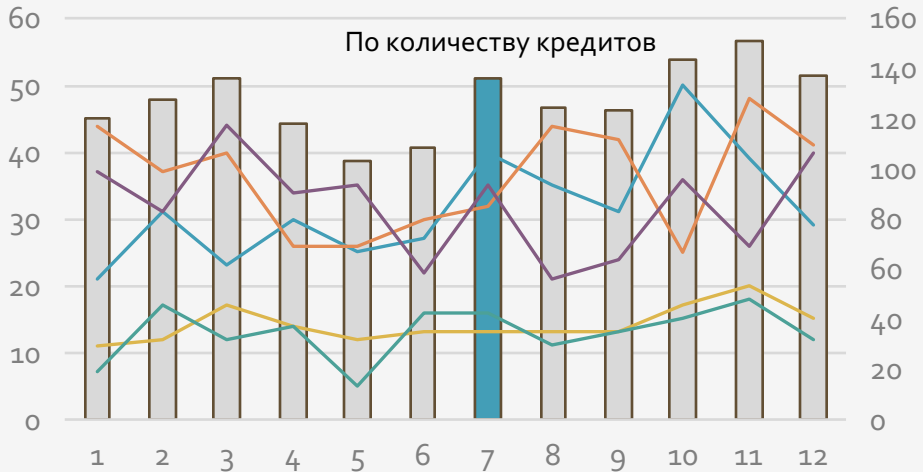
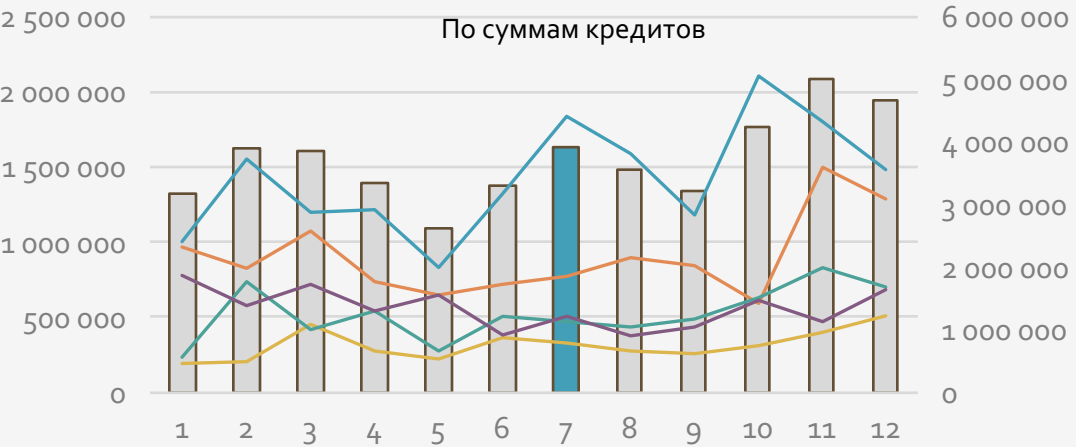
- Неожиданный вывод - люди, которые были клиентами банка - на 7,5% чаще уходили в дефолт. Но таких людей в 3 раза больше, чем не клиентов банка.
 - Наибольшее количество дефолтов в категории мобильных телефонов и составляет 113 случаев, что составляет 57% от общего количества дефолтов. При сроках более 12 месяцев и зарплате 20-40к чаще уходят в дефолт.
 - В 7 месяце покупали более дешевые товары, тк количество в штуках выше остальных месяцев, а по сумме выдач - меньше. И при этом уровень дефолта в 7 месяце значительно ниже.
 - С 9 месяца стабильно растёт дефолтность у женщин. (Как далее увидим, со средним специальным образованием). Так же заметим, что в 7 месяце женщины брали меньше кредитов, чем мужчины с меньшим средним чеком.
 - Основой POS портфеля являются люди со средним и средним специальным образованием и составляют 60% от общего количества. Как ранее наблюдали, с 9 по 12 месяц рост дефолтов в сегменте со средним специальным образованием.
 - Основным регионами выдач являются «других регионов» которые составляют 82% от общего количества выдач.
 - В 5 месяце было проседание в выдачах кредитов у мужчин, что сказалась на категории высшего образования. Так же мужчины склонны брать более крупные кредиты с чеками 40-70к (в 2 раза чаще чем женщины, что по году 129 раз).
 - Интересно, что людям старше 79 лет выдавались кредиты, уровень дефолтности = 0 , а в 20-25 лет составляет 15%.
-

В 7 месяце покупали более дешевые товары , тк количество в штуках выше остальных месяцев, а по сумме выдач - меньше. И при этом уровень дефолта в 7 месяце значительно ниже.

Дефолт



Нет дефолта

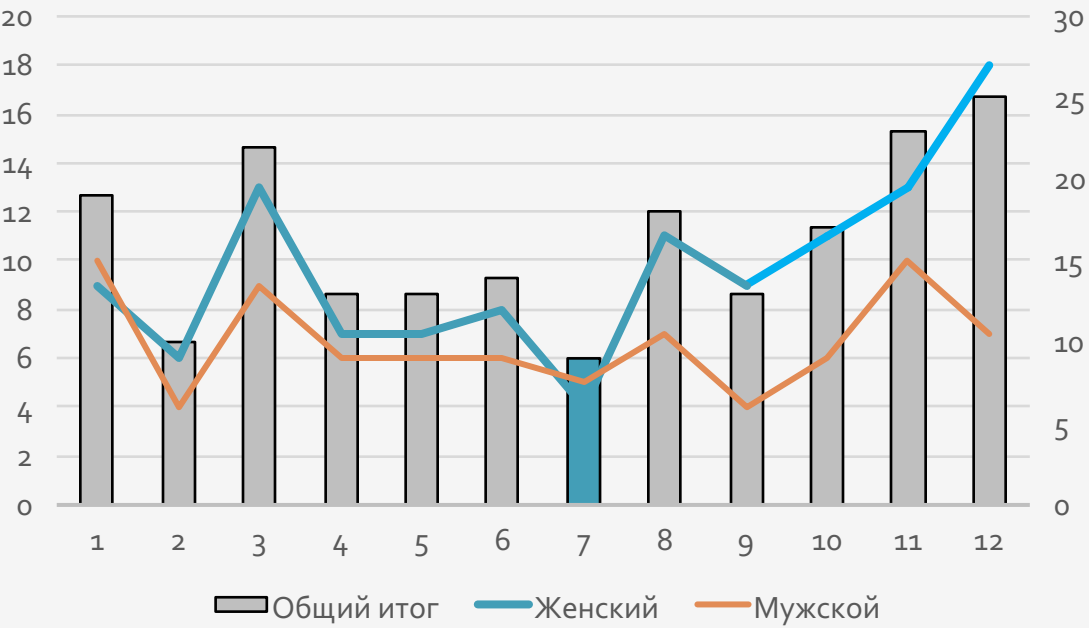


- Общий итог
- Бытовая техника
- Мебель

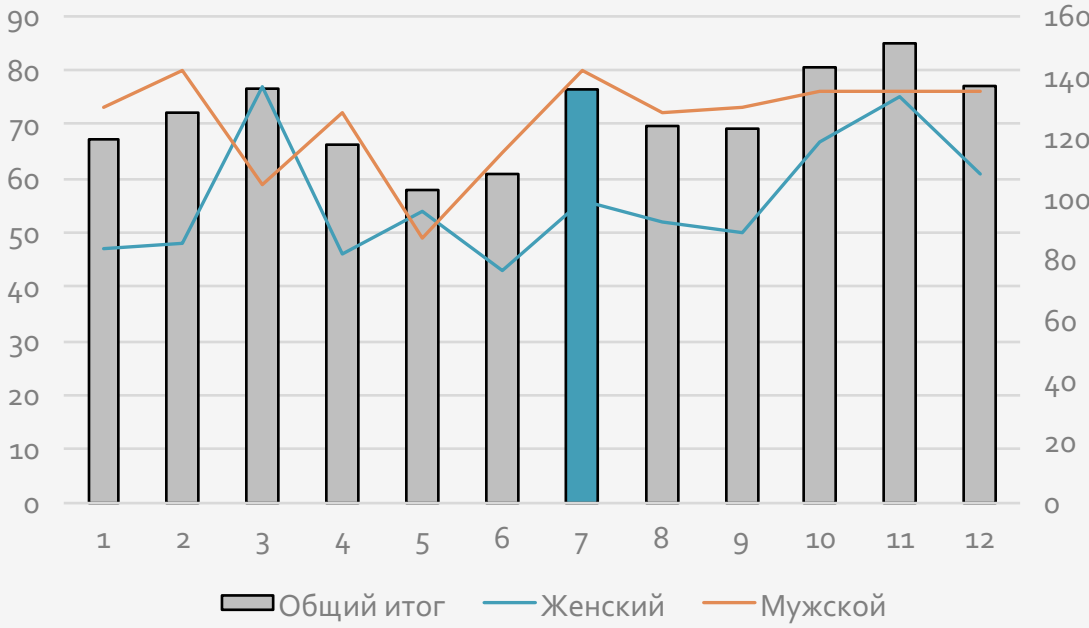
- Остальное
- Компьютеры, комплектующие
- Мобильные телефоны

С 9 месяца стабильно растёт дефолтность у женщин. (Как далее увидим, со средним специальным образованием).
Так же заметим, что в 7 месяце женщины брали меньше кредитов, чем мужчины с меньшим средним чеком

Дефолт

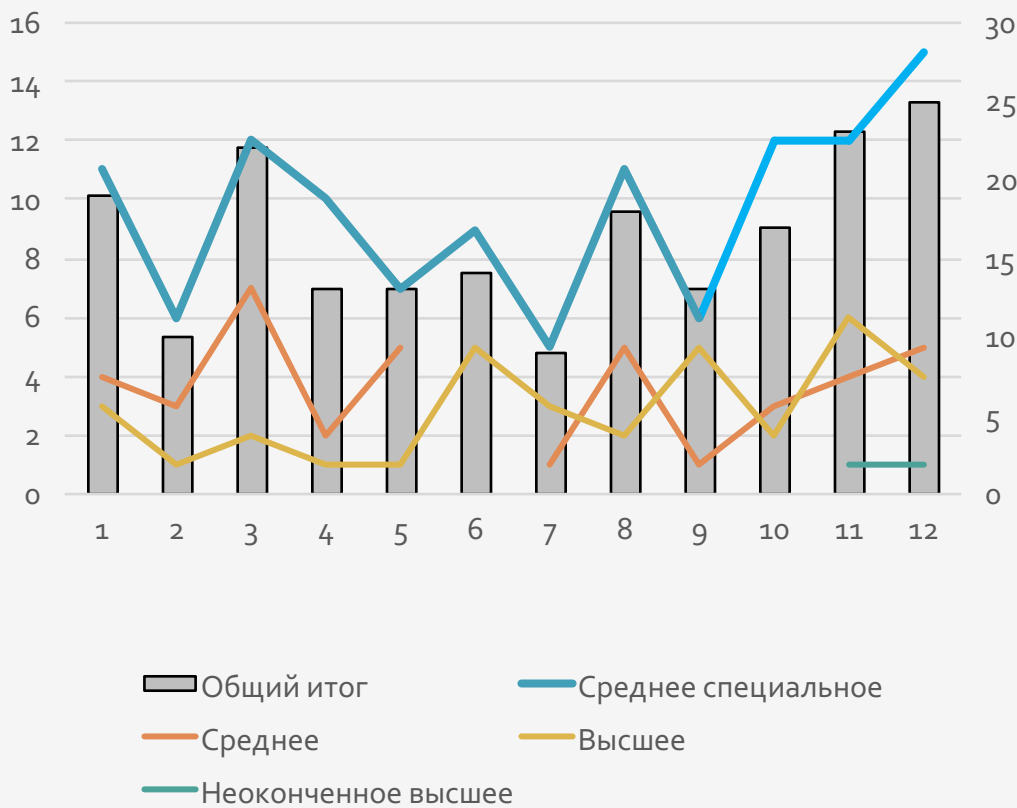


Нет дефолта



Основой POS портфеля являются люди со средним и средним специальным образованием и составляют 60% от общего количества. Как ранее наблюдали, с 9 по 12 месяц рост дефолтов в сегменте со средним специальным образованием.

Дефолт

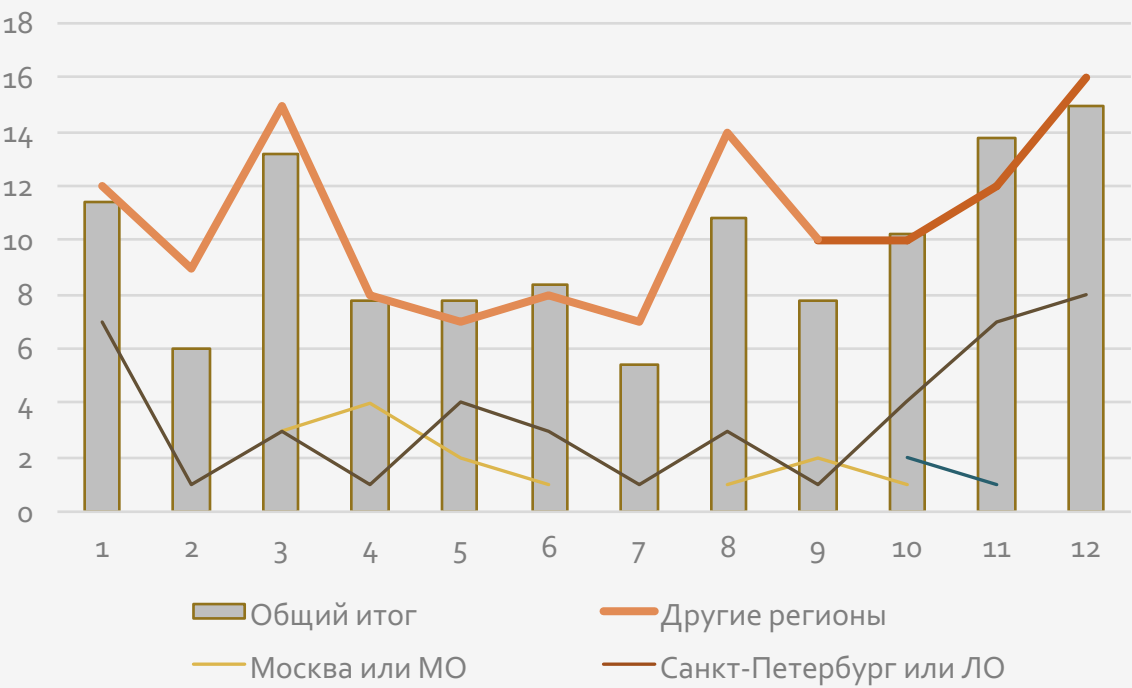


Нет дефолта

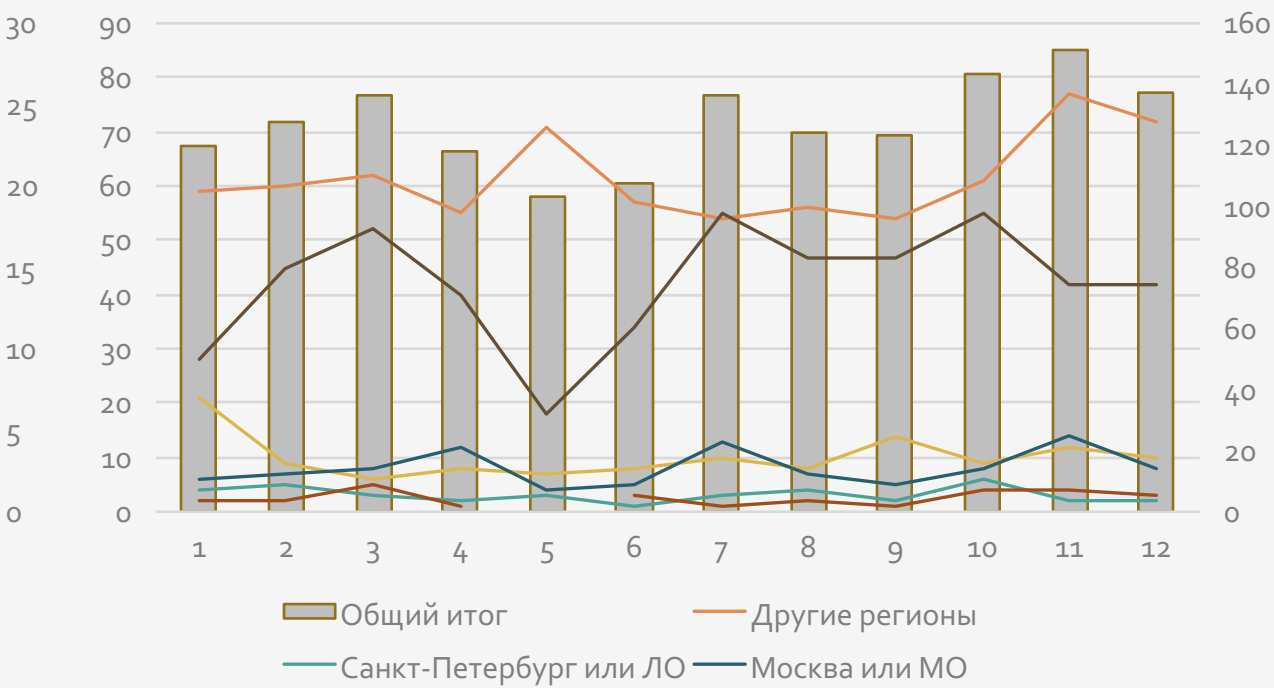


Основным регионами выдач являются «других регионов» которые составляют 82% от общего количества выдач.

Дефолт



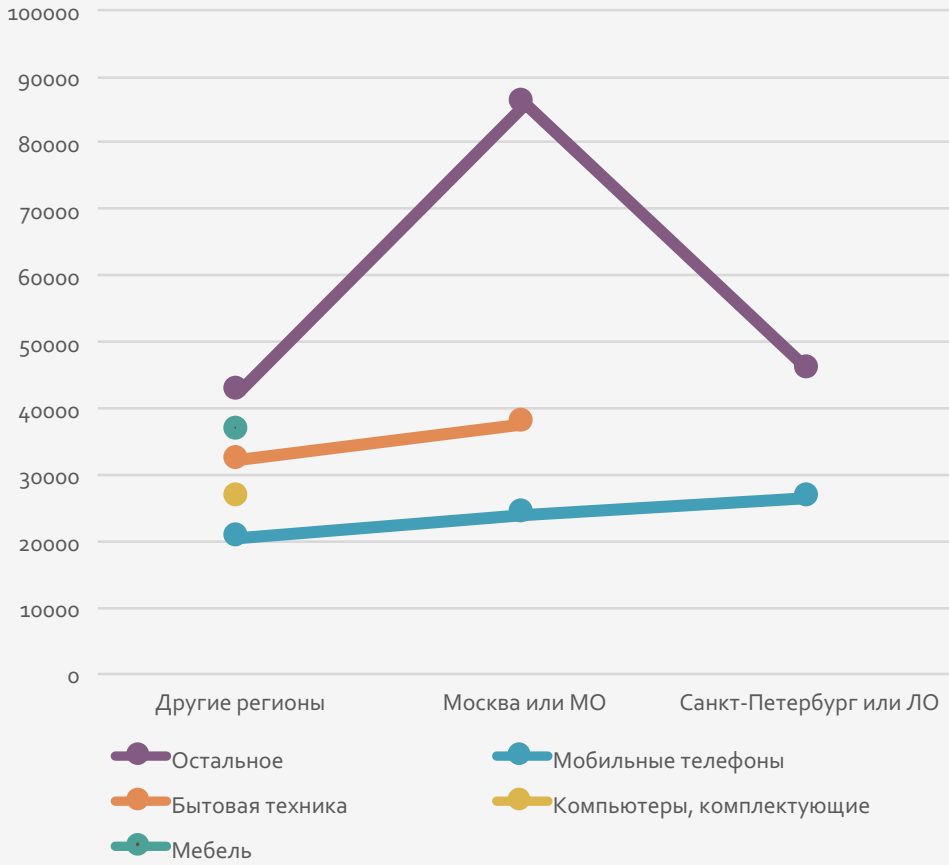
Нет дефолта



Москвичи ушли в дефолт в категории другое, со средним чеком 85.000₽, что на 112% больше среднего по категории в регионах. В Москве люди склонны брать в кредит более дорогую мебель, чем в других регионах(на 97%).

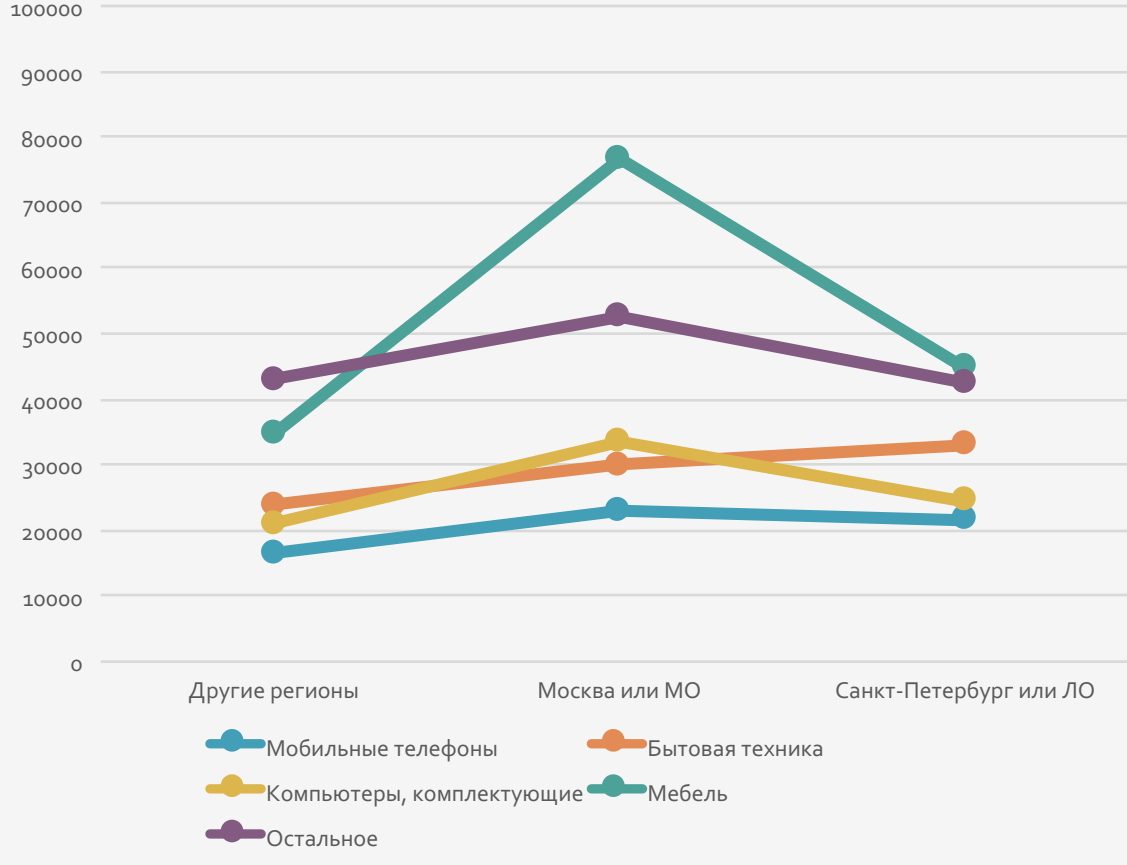
Дефолт

Средний чек

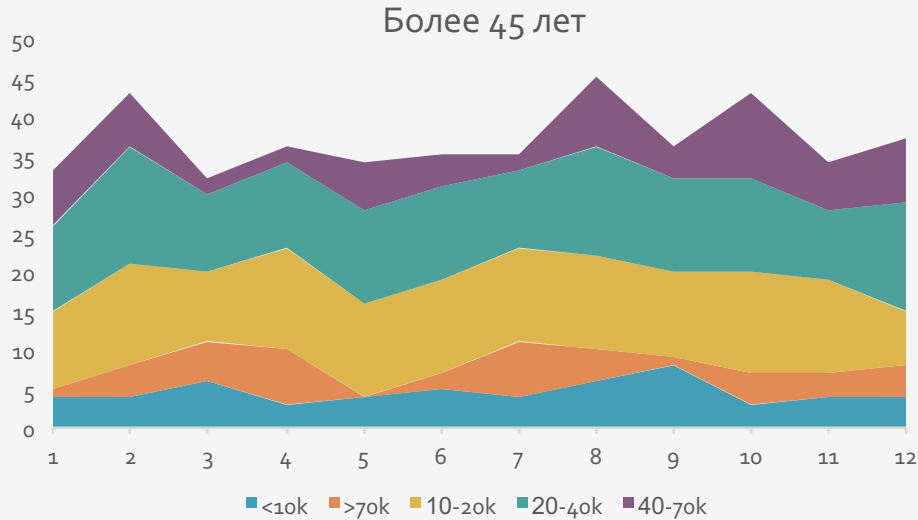
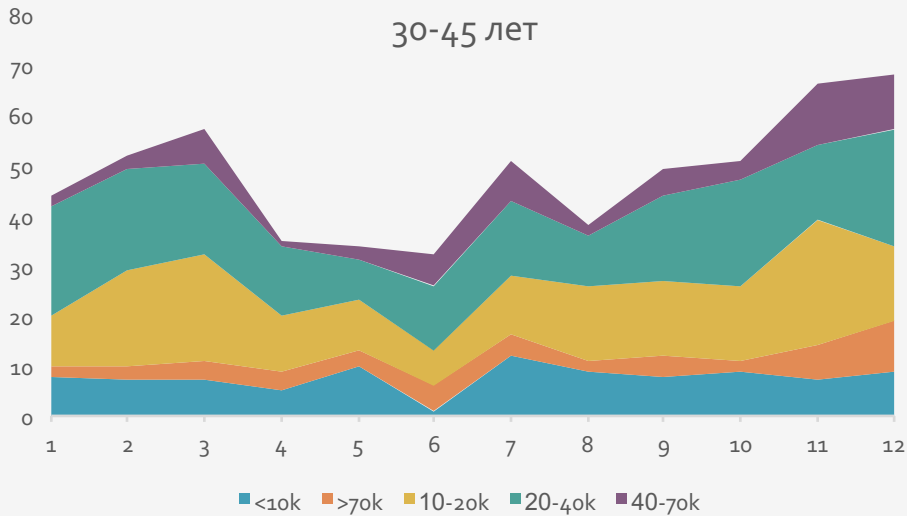
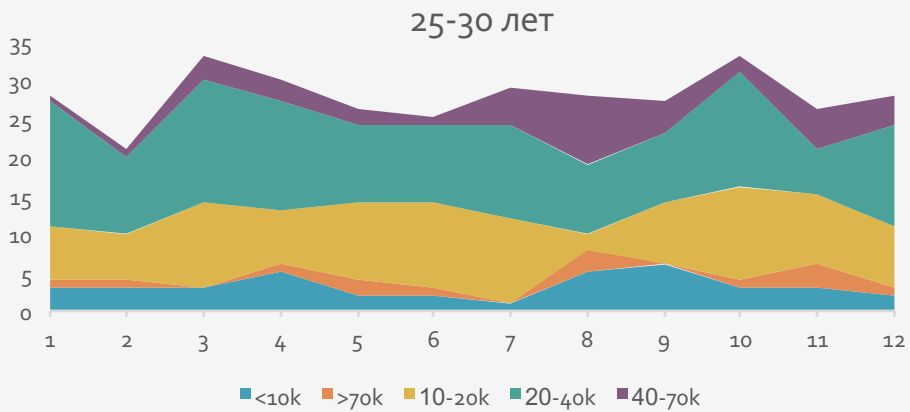
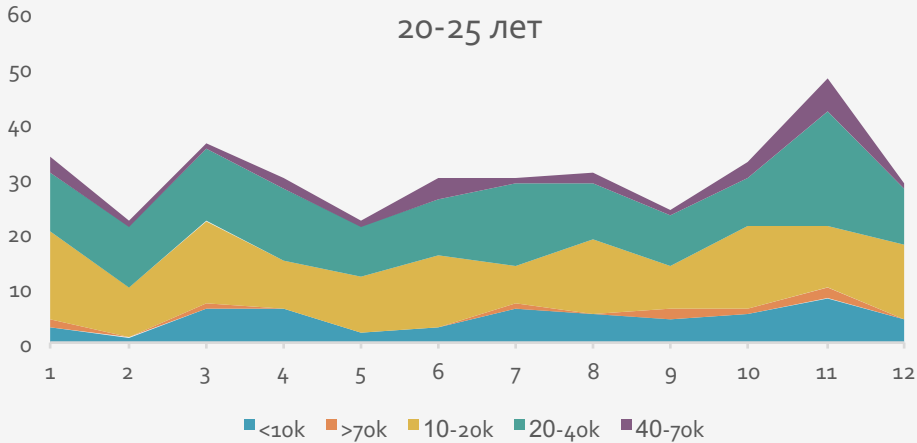


Нет дефолта

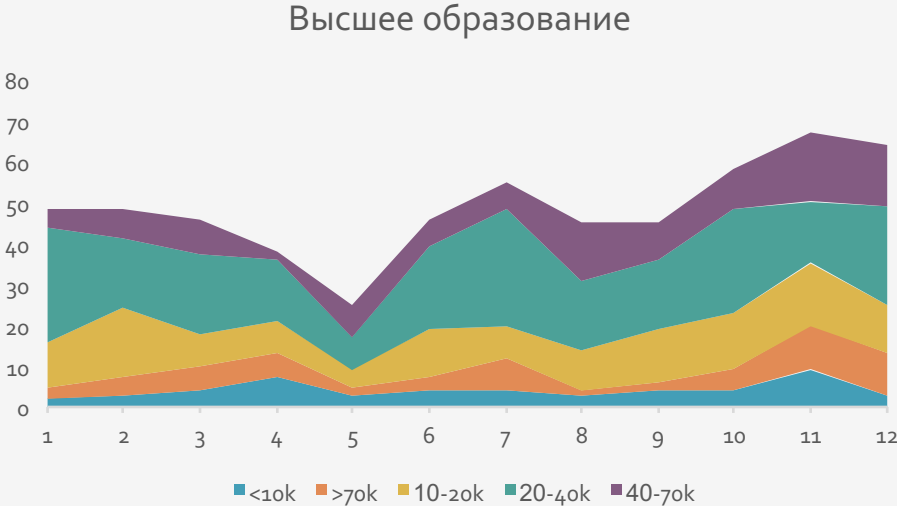
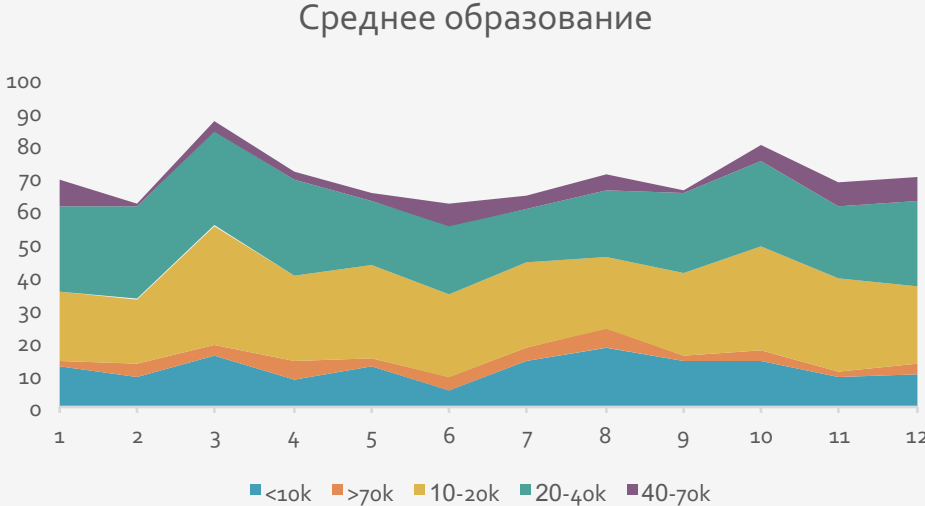
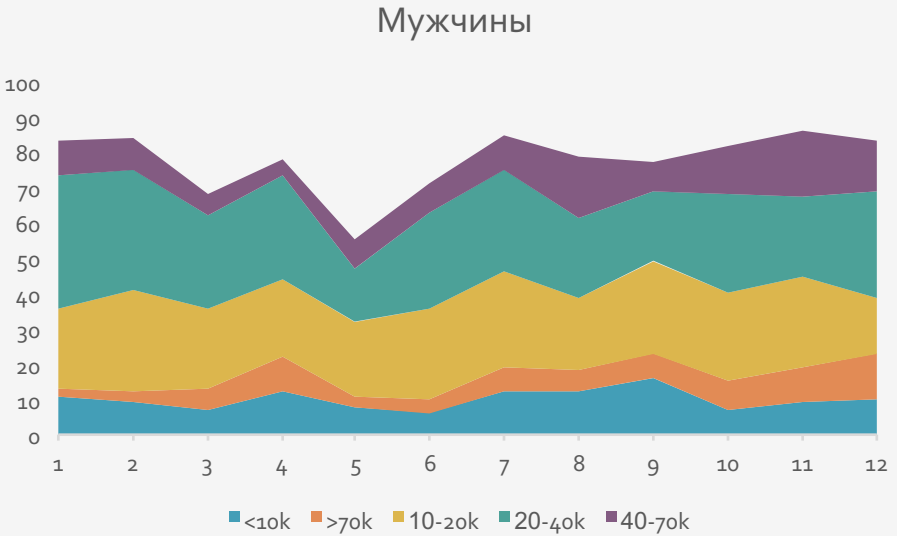
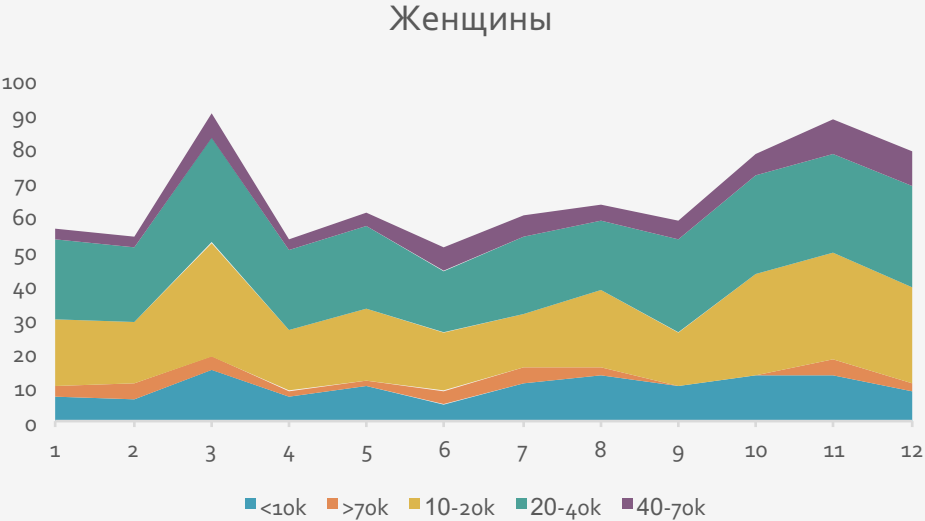
Средний чек



Распределение суммы кредита от возраста по 12 месяцам достаточно стабильное во всех сегментах. В сегменте 30-45 лет с 8 месяца начало увеличиваться количество выдач и в 12 месяце в 2 раза выше, чем в середине года (68шт).



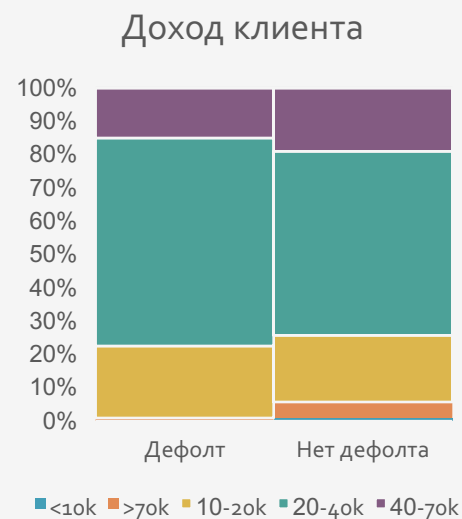
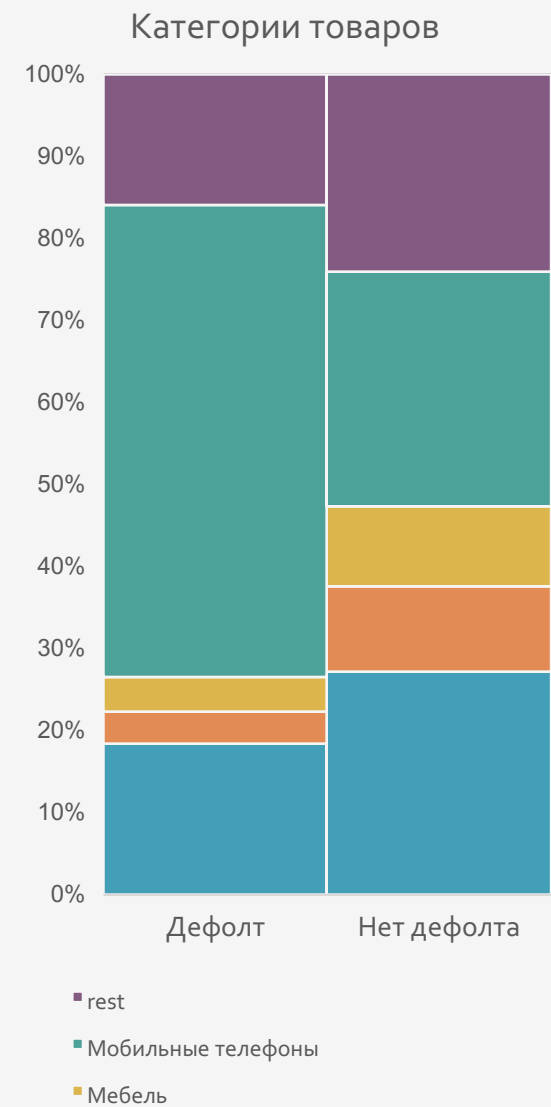
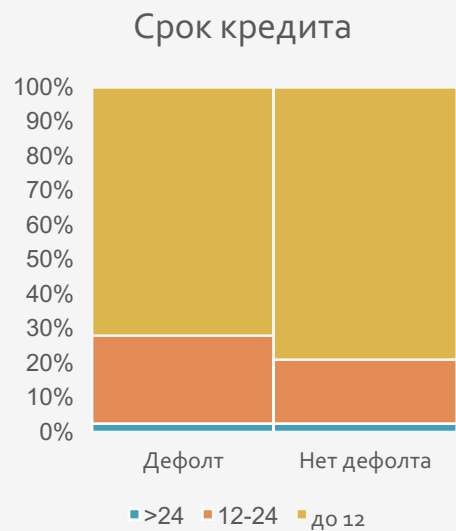
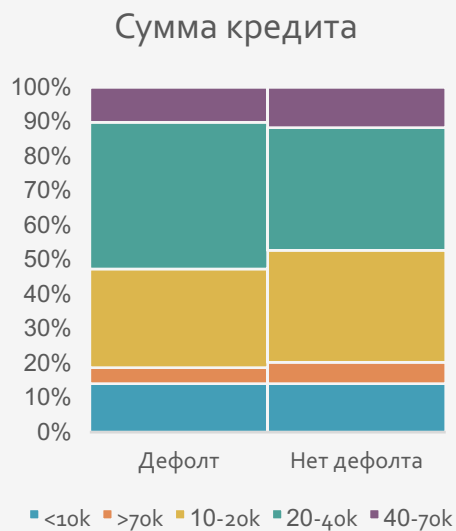
В 5 месяце было проседание в выдачах кредитов у мужчин, что сказалась на категории высшего образования. Так же мужчины склонны брать более крупные кредиты с чеками 40-70k (в 2 раза чаще чем женщины, что по году 129 раз).



Более ответственные люди это мужчины, люди у которых есть дети, в статусе «другое», и в возрасте более 40 лет.
Интересно, что людям старше 79 лет выдавались кредиты, уровень дефолтности = 0 , а в 20-25 лет составляет 15%.



Наибольшее количество дефолтов в категории мобильных телефонов и составляет 113 случаев, что составляет 57% от общего количества дефолтов. При сроках более 12 месяцев и зарплате 20-40к чаще уходят в дефолт.



Основные показатели дефолта.

Сумма кредита	Дефолт по кредиту	Всего	% дефолта
<10k	27	238	11,3%
>70k	10	114	8,8%
10-20k	56	554	10,1%
20-40k	83	619	13,4%
40-70k	20	198	

Возраст	Дефолт по кредиту	Всего	% дефолта
>40	38	443	8,6%
20-25	55	369	14,9%
25-30	41	334	12,3%
30-45	62	577	10,7%

Платеж к доходу	Дефолт по кредиту	Всего	% дефолта
<10%	150	1178	12,7%
>50%	0	9	0,0%
10-50%	46	536	8,6%

Срок кредита	Дефолт по кредиту	Всего	% дефолта
>24	5	46	10,9%
12-24	50	314	15,9%
до 12	141	1363	10,3%

Категория	Дефолт по кредиту	Всего	% дефолта
Бытовая техника	36	471	7,6%
Компьютеры	8	178	4,5%
Мебель	8	164	4,9%
Мобильные телефоны	113	498	22,7%
Остальное	31	412	7,5%

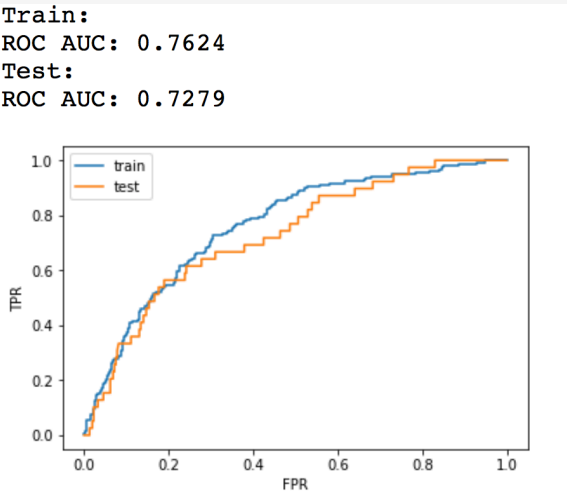
Является клиентом	Дефолт по кредиту	Всего	% дефолта
Клиент банка	145	1042	13,9%
Не клиент банка	51	681	7,5%
Общий итог	196	1723	

Создание модели для снижения дефолта.



1. Сделаем бинизацию числовых признаков (сгруппируем их) для того, чтобы легче было работать.
2. Создадим функцию для вероятностей Weight of evidence (woe). И далее на ней будем строить модель.
3. Разобьем данные равномерно по группам данных и проверим флаг дефолта в трейне и тесте и в итоге получим по 11% в каждой группе.
4. Проверим корреляции между признаками.
5. Так как после применения функции woe у нас остались пропущенные значения, заменим их медианами.
6. Далее построим несколько моделей для определения лучшего результата.

С логистической регрессией получили ROC AUC 0.7279 и accuracy score 0.8869



Запустим подбор по сетке GridSearch и запросим оптимальные параметры: {'C': 10, 'max_iter': 100, 'penalty': 'l2', 'tol': 0.1.}

При данных параметрах получаем ROC AUC на тесте равным 0.7279. и уравнение регрессии:

$$P = -2.348 + x_1 * 0.0443 + x_2 * (-0.229) + x_3 * 1.0256 + x_4 * 0.441 + x_5 * 0.450 + x_6 * 0.806 + x_7 * 0.790 + x_8 * 0.502 + x_9 * 0.592 + x_{10} * 0.104 + x_{11} * 0.281 + x_{12} * 0.416 + x_{13} * 0.832 + x_{14} * 0.535$$

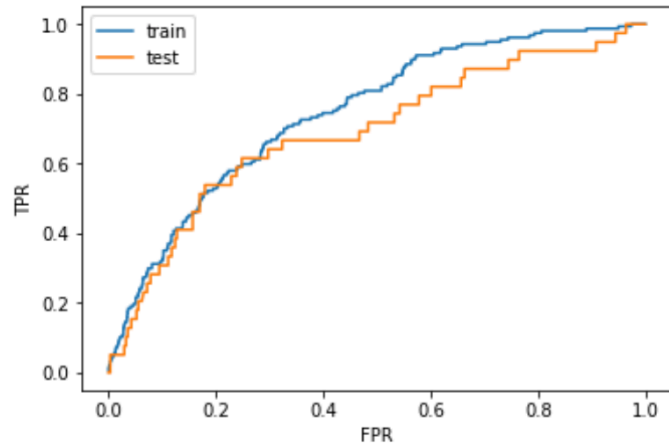
, где x от 1 до 14 – наши признаки.

x1	Месяц выдачи кредита
x2	Сумма кредита
x3	Срок кредита
x4	Возраст клиента
x5	Пол клиента
x6	Образование клиента
x7	Тип товара
x8	Наличие детей у клиента
x9	Регион выдачи кредита
x10	Доход клиента
x11	Семейное положение
x12	Оператор связи
x13	Является ли клиентом банка
x14	Платеж к доходу

Из 3 моделей лучшей оказалась модель AdaBoosting с ROC AUC 0.7473 и accuracy score 0.878

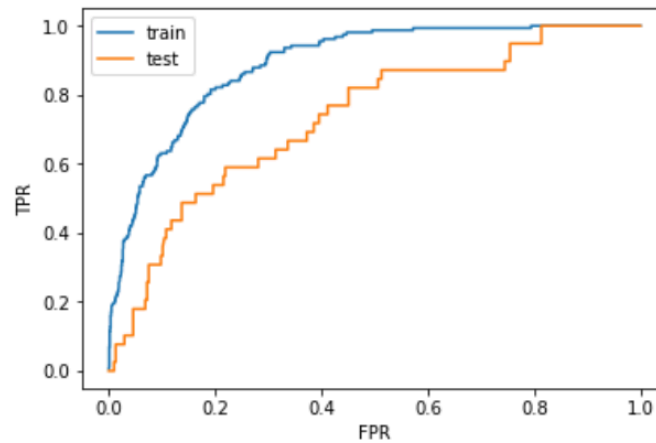
С логистической регрессией и полиномиальными фичами во второй степени получили ROC AUC 0.690 и accuracy score 0.8898

Train:
ROC AUC: 0.7460
Test:
ROC AUC: 0.6906



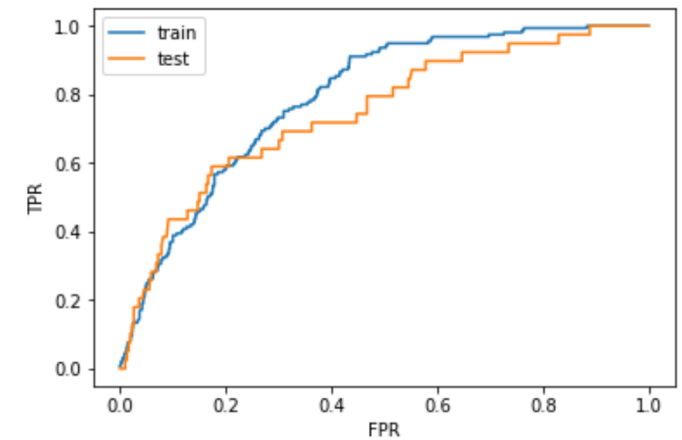
Catboost немного переобучается, но ROC AUC выше и составляет 0.731 и accuracy score 0.878

Train:
ROC AUC: 0.8901
Test:
ROC AUC: 0.7317



Метод композиции слабых моделей AdaBoosting показал лучший ROC AUC 0.7473 и accuracy score 0.878

Train:
ROC AUC: 0.7908
Test:
ROC AUC: 0.7473



Качество построенных моделей достаточно низкое из-за малого количества данных, большой смещенности в регионах, категориях.

В качестве рекомендаций и дальнейших исследований:

1. Необходимо пересмотреть стратегию выдач кредитов в сегменте мобильных телефонов, т.к. Там сосредоточено более 50% всех дефолтов, в частности женщин со средним специальным и средним образованием проживающих в «других регионах».
 2. Создать скоркарту с большими весами на образование и категорию, установить пороги на сегмент мобильных телефонов. Тогда отрезем часть выдач, но общий прv будет значительно выше.
-