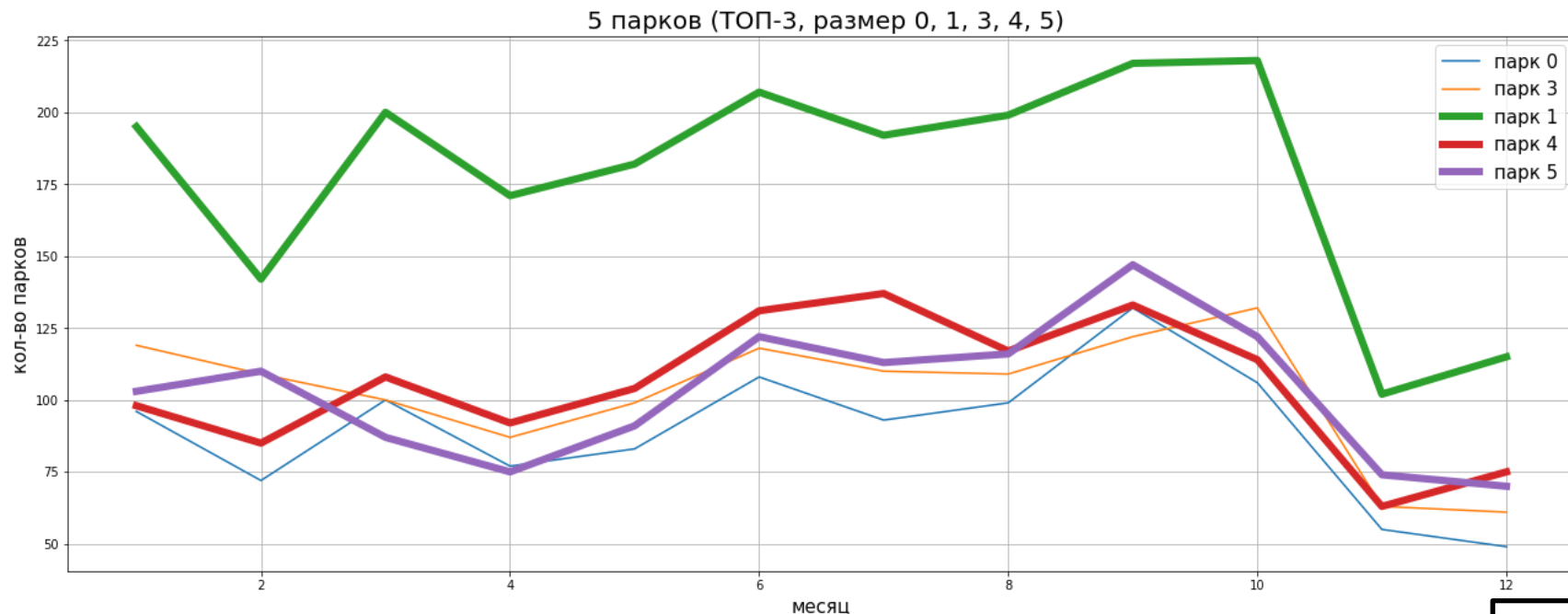


Презентация к тестовому заданию

ресурс:
pm_test.ipynb

Частота компаний с определенным размером парка машин

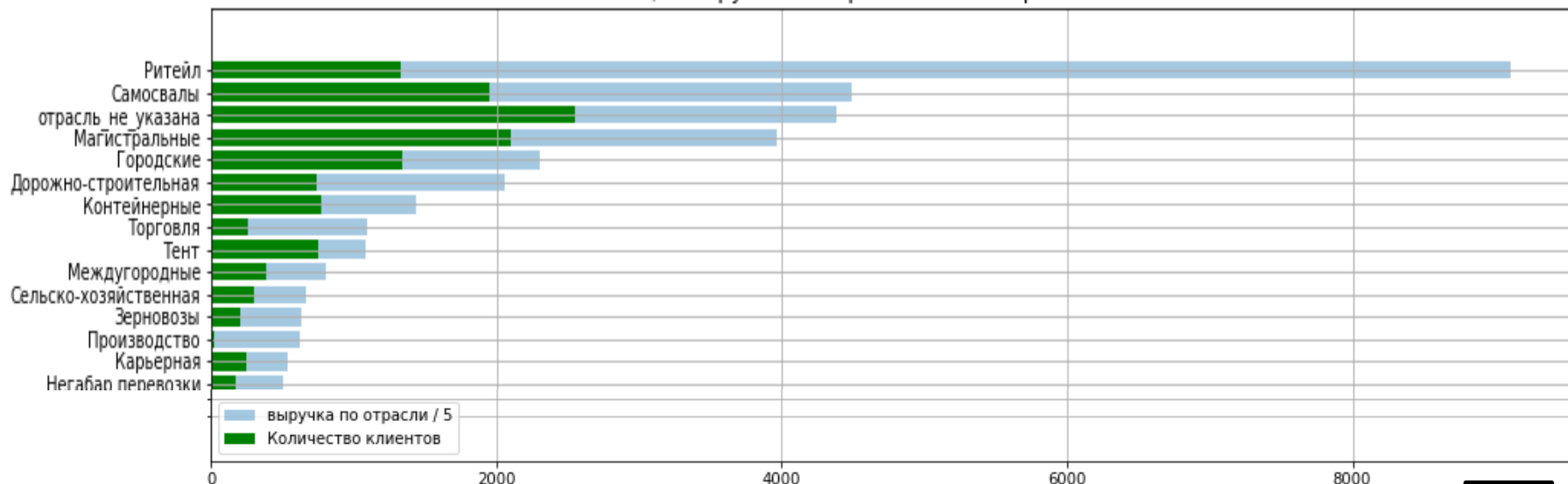
- По всем месяцам преобладают парки размером '1';
- На втором месте чаще всего парки размером '4' (3, 4, 5, 6, 7, 8 и 12 месяцы);
- На третьем месте половину года преобладают парки размером '5' (1, 6, 7, 8, 10, 12 месяцы), остальные парки реже.



Частота компаний из определенной отрасли

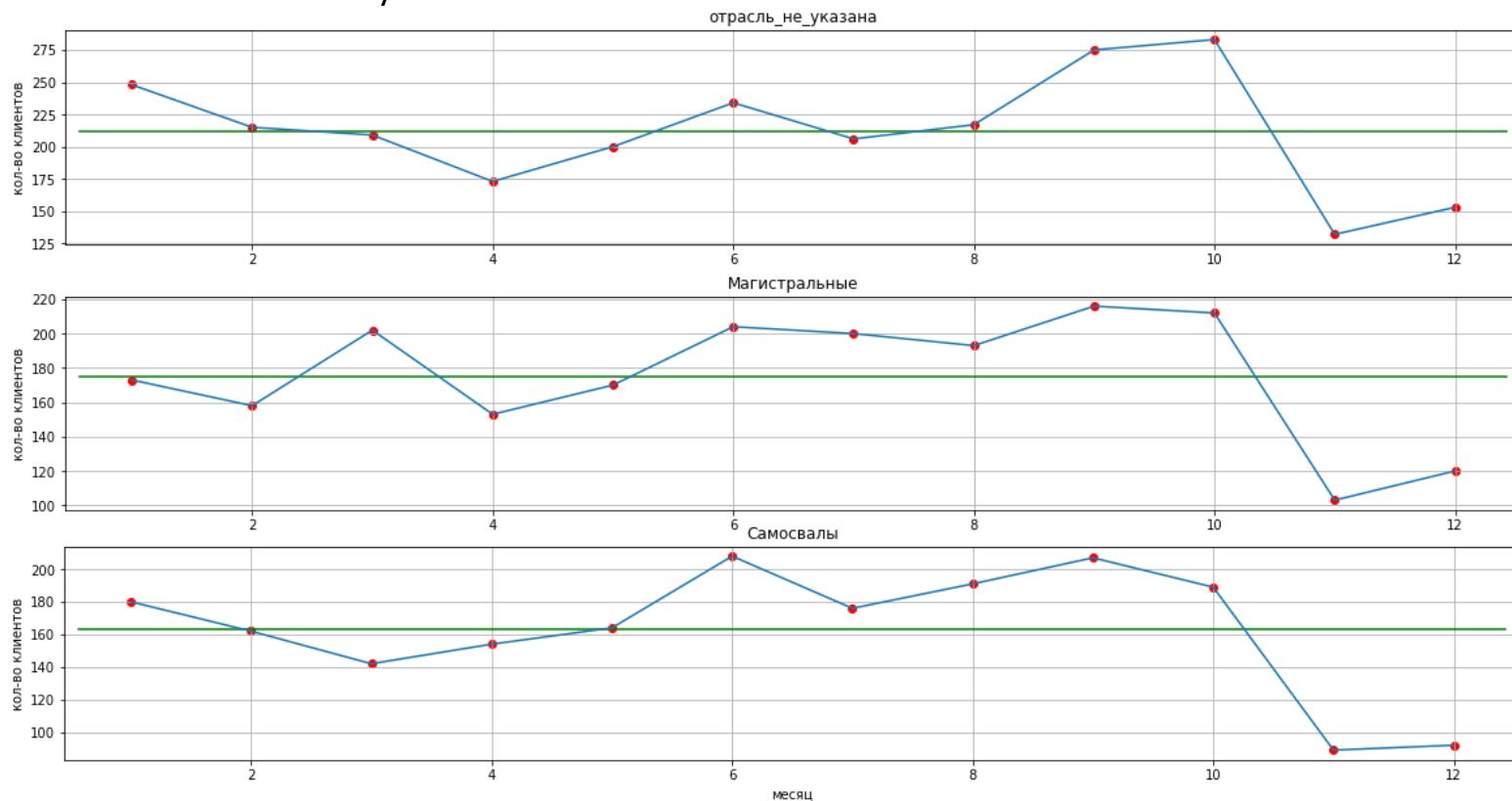
- По количеству клиентов по всем месяцам превалирует отрасль без названия ("отрасль_не_указана"), второе и третье место с переменным успехом делят отрасли "Магистральные" и "Самосвалы".
- Наибольшую выручку в каждом из месяцев приносит отрасль "Ритейл", которая входит в ТОП-6 отраслей по количеству клиентов в месячном разрезе.

Общая выручка по отраслям за всё время



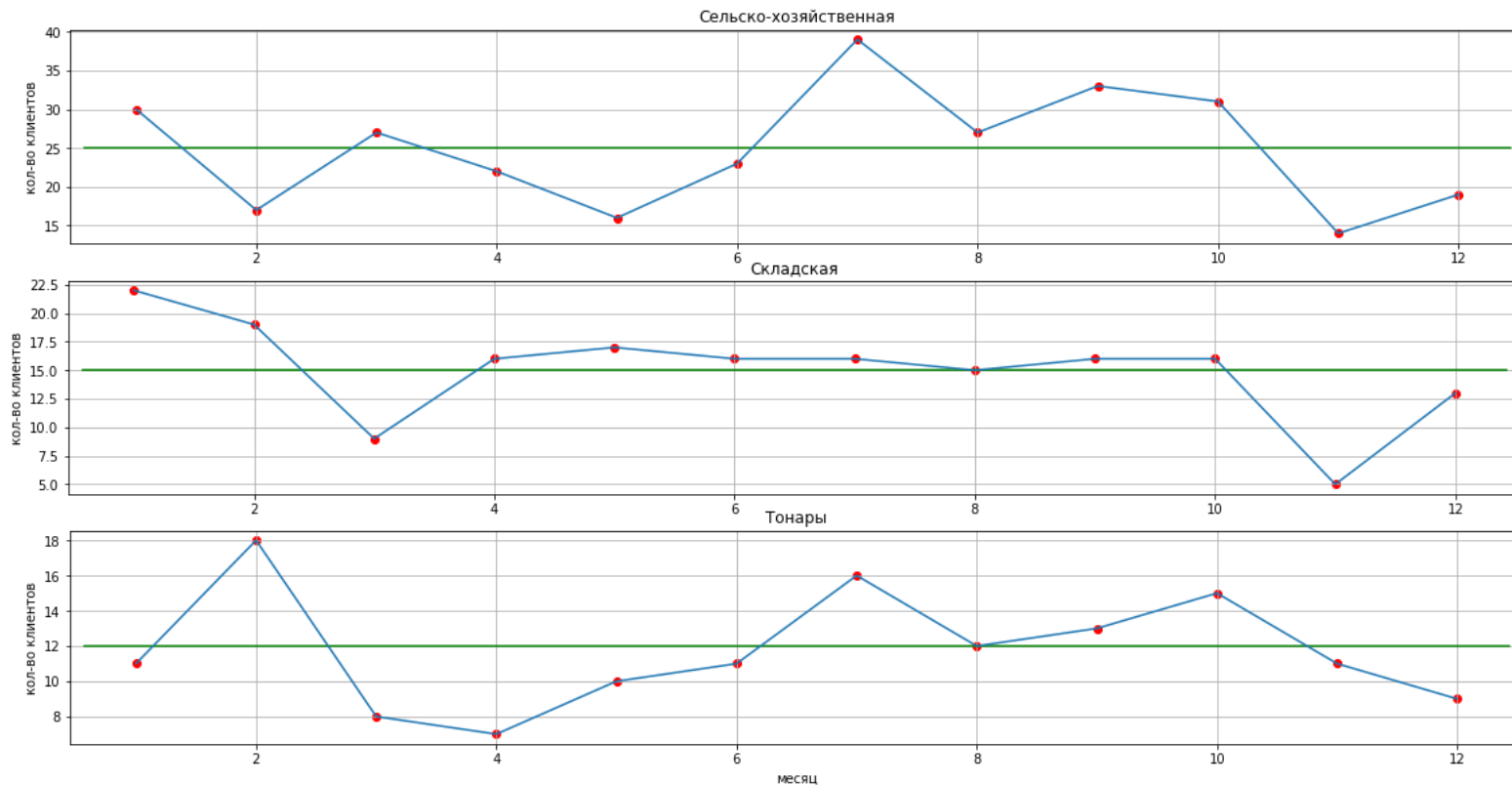
Частота обращений по месяцам

■ ТОП-3 по количеству клиентов



Частота обращений по месяцам

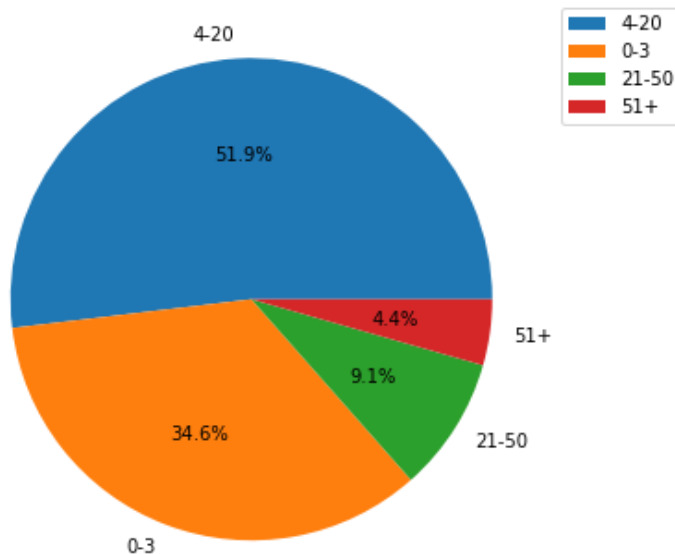
- 3 характерных графика по количеству клиентов



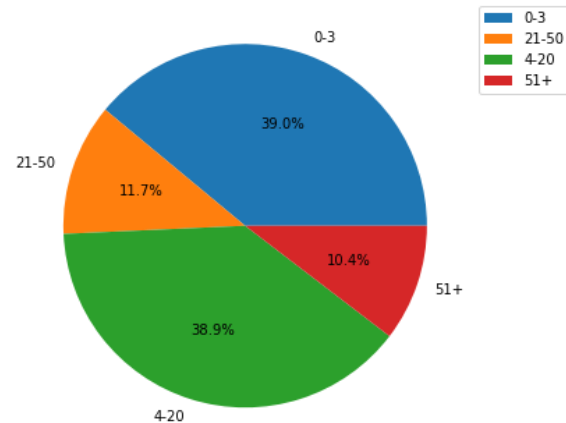
Сегментация по размеру парка машин

- Предложен деление на сегменты '0-3', '4-20', '21-50' и '51+'.
- При общей доле сегмента '4-20' в 52% доля выручки составляет 39%, в то же время при общей доле сегмента '0-3' в ~35% доля выручки составляет также 39%. Тоже самое характерно для сегментов '21-50' и '51+', где сегмент '51+' при меньшей доле дает большую выручку.

Разбивка по сегментам

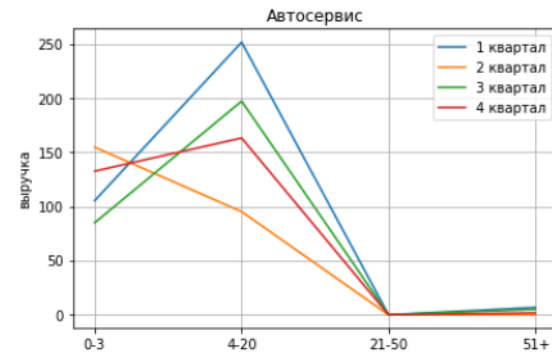
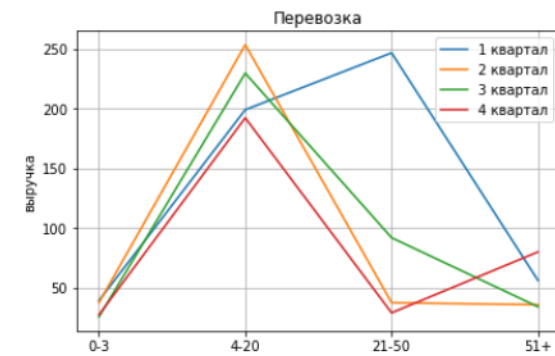
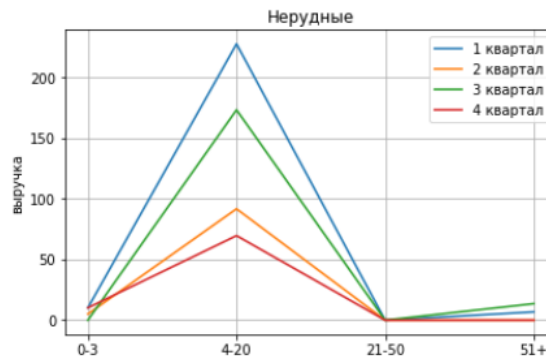
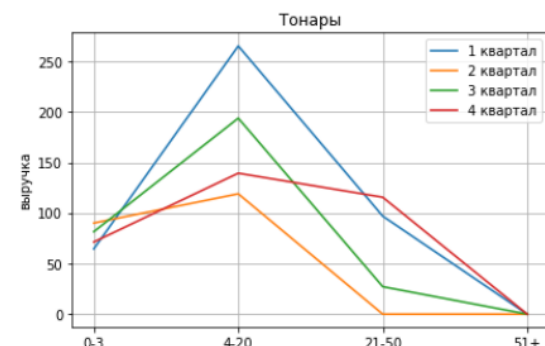
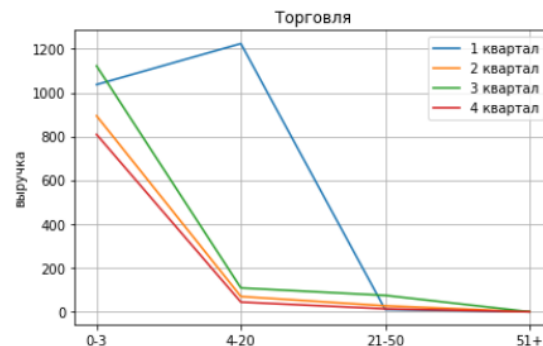
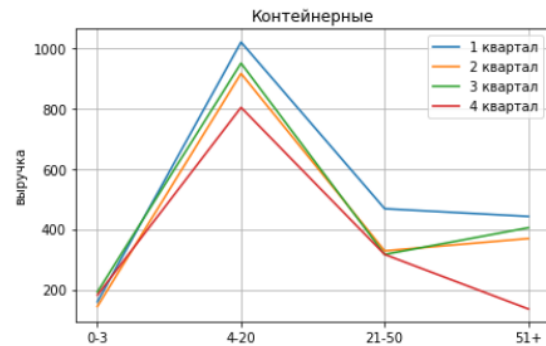


Выручка по сегментам



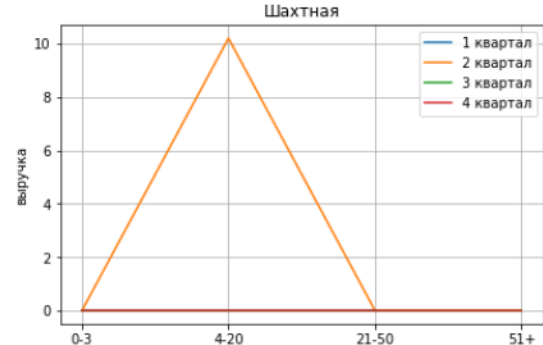
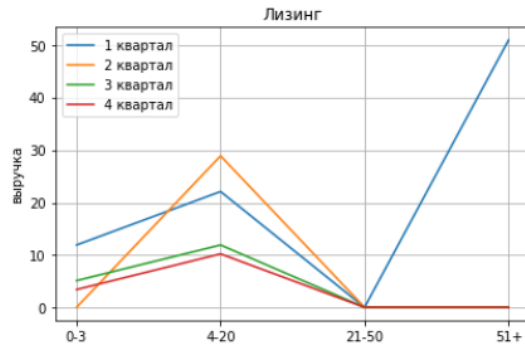
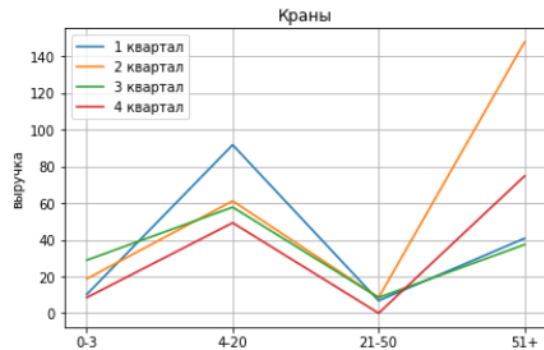
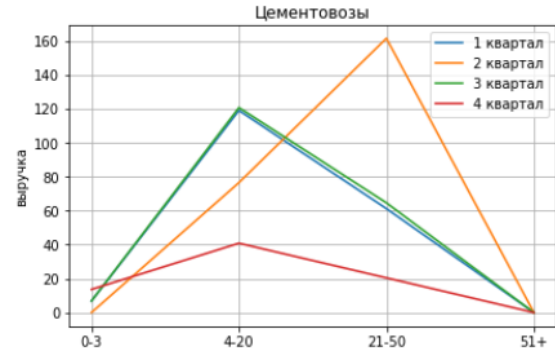
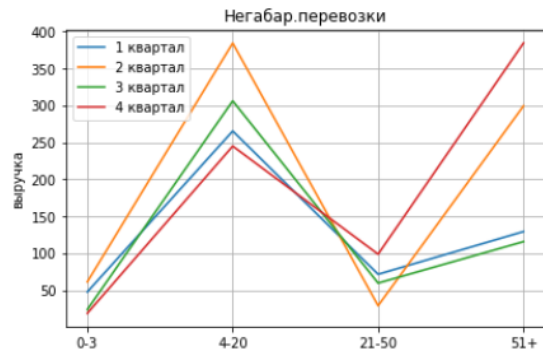
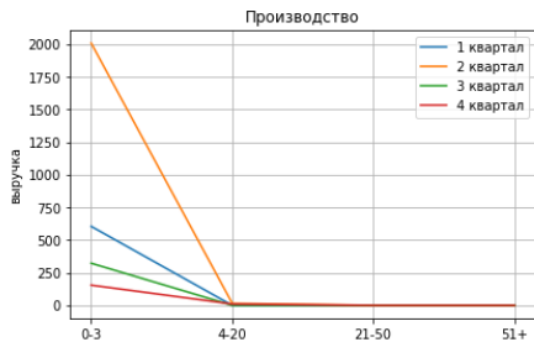
Поквартальная выручка в разрезе сегментов

- 6 отраслей с акцентом выручки на 1-й квартал



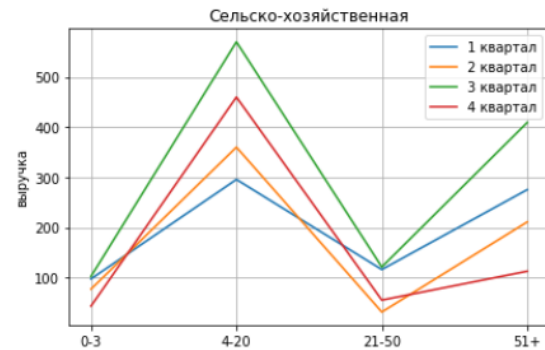
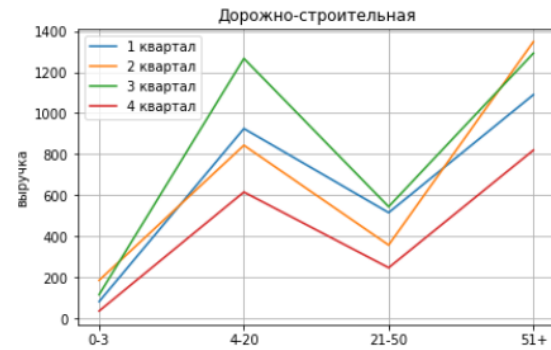
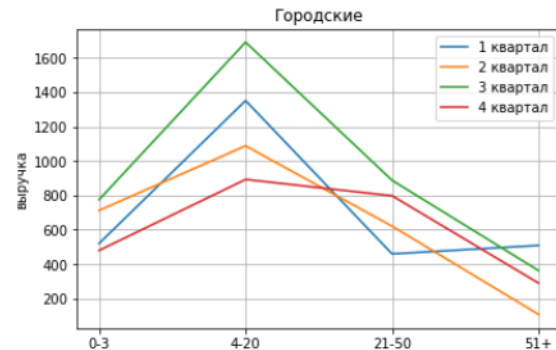
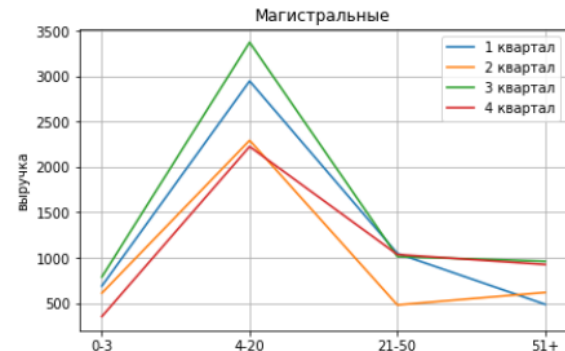
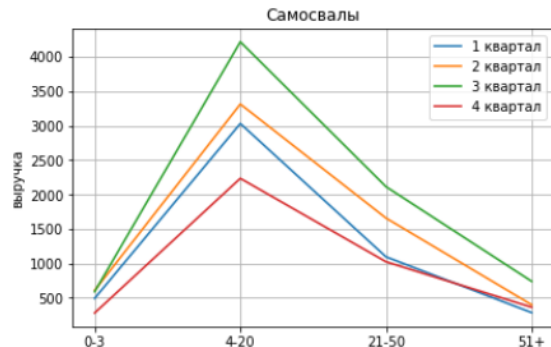
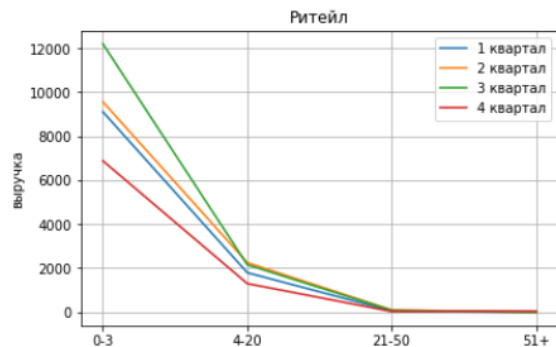
Поквартальная выручка в разрезе сегментов

- 6 отраслей с акцентом выручки на 2-й квартал



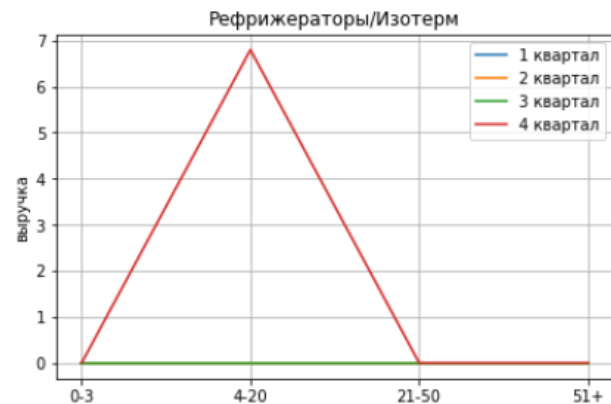
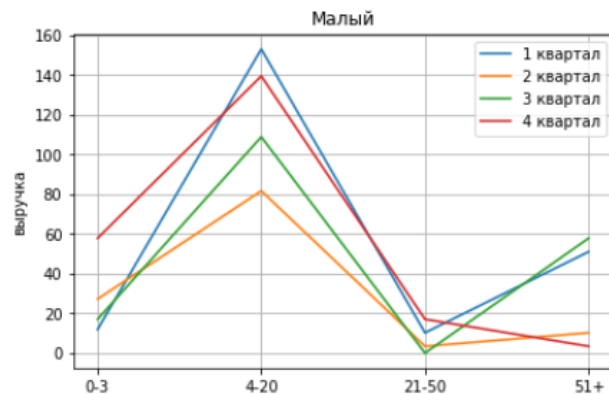
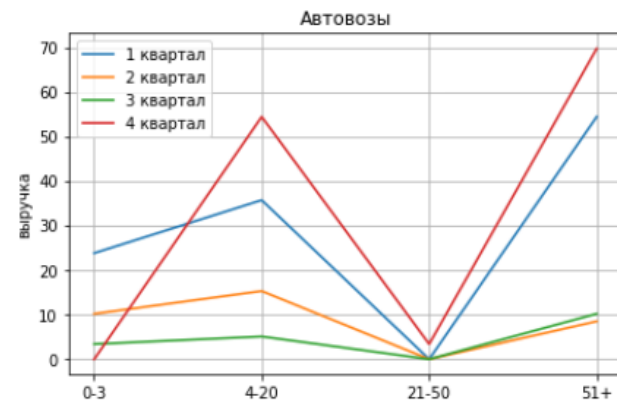
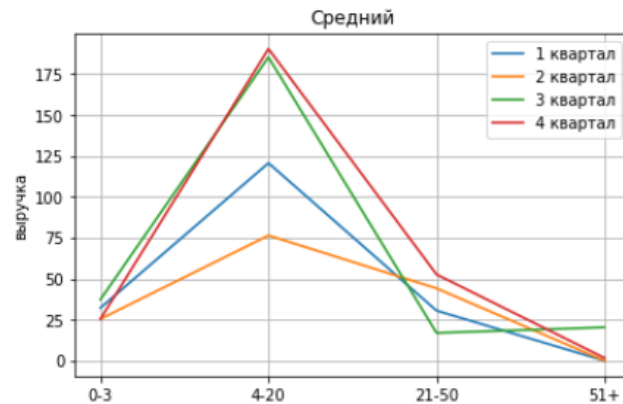
Поквартальная выручка в разрезе сегментов

- 6 отраслей с акцентом выручки на 3-й квартал



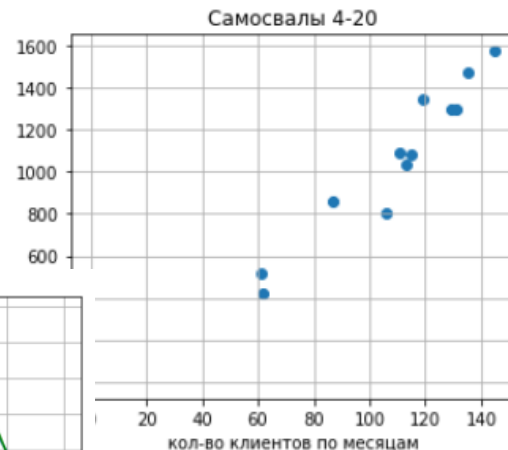
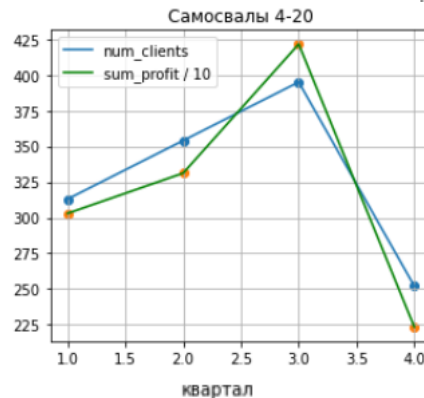
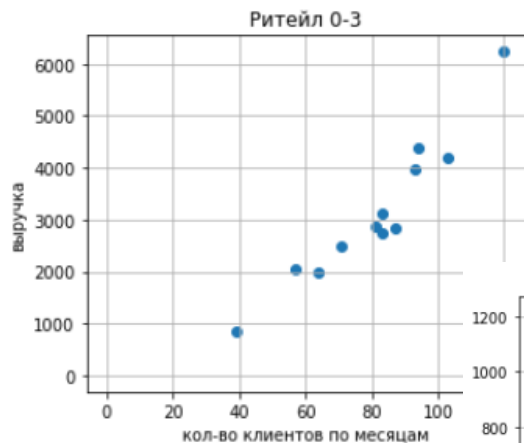
Поквартальная выручка в разрезе сегментов

- 4 отрасли с акцентом выручки на 4-й квартал



Линейная связь между кол-вом клиентов и выручкой

- Показывает линейную зависимость между двумя исследуемыми переменными.
- Например отрасли «Ритейл» и «Самосвалы» являются лидерами по выручке в и имеют в сегментах '0-3' и '4-20' соответственную выраженную лин.связь при том, что эти сегменты наиболее прибыльны среди других сегментов отрасли.



Дополнительное исследование данных с помощью ML

- В данной задаче исследовалась зависимость целевой переменной (итоговая сумма, столбец "Итого") от других переменных, имеющих в предоставленной таблице, которая не дополнялась новыми фичами.

	Код клиента	Размер парка	Месяц	Итого	Отрасль	Сегмент	Номер_сегмента	Квартал
0	00-00049327	4	2	1020.0	Торговля	4-20	2	1
1	00-00005692	24	9	642.6	Самосвалы	21-50	3	3

- Метрика R2 на трейне и на тесте получилась очень низкой и ошибка MSE получилась слишком высокой, это говорит о том, что признаки имеют плохую объясняющую способность для целевой переменной.

LightGBM



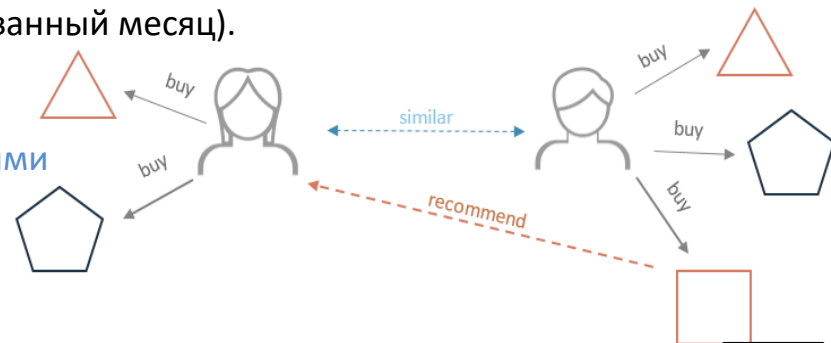
CatBoost

- Для эффективного исследования зависимостей необходим более качественный датасет.

Дополнительное исследование данных для RecSys

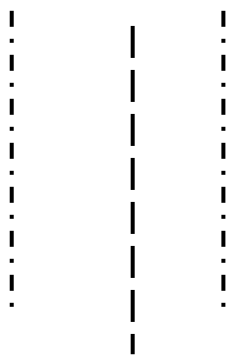
- В данной задаче исследовалась возможность создания рекомендательной системы для определения месяца, в котором клиент снова обратится в компанию.
- Исследовались наиболее популярные у клиентов из разных отраслей и разных сегментов месяцы, в которых клиенты чаще всего обращаются в компанию. Затем эти месяцы рекомендованы клиентам. В исследовании применена **простая одноуровневая рекомендательная модель без дополнительных библиотек**.
- Предложенная метрика precision@2 получилась достаточно высокой (на трейне 0.51, на тесте 0.40), это говорит о том, что предложенная модель может быть полезной в случае разработки рекомендаций для клиента (например, отправки дополнительного уведомления клиенту в мобильном приложении в рекомендованный месяц).

- Исследование данных в целях формирования рекомендаций не ограничивается предложенными и реализованными действиями, а лишь является неким **baseline** в данной постановке задачи.



Для выработки дальнейших выводов предлагается общий порядок действий:

- Изучение клиентов из разных отраслей, формирование образа клиента(его потребностей и т.д.);
- Изучение специфики бизнес-процессов компании, в том числе логистики;
- Изучение ценовой политики для клиентов разных сегментов в отрасли;
- Изучение охвата клиентов, посещаемости ресурсов, конверсии по отраслям;
- Изучения влияния внешних факторов на исследуемые показатели компании;
- Изучение и выявление закономерностей в данных для обоснования целевой переменной в данных;
- Изучение потребительского поведения и возможности внедрения рекомендательных систем, uplift-моделирования.



The end

Для изучения данных использовался Python (Pandas);
Для визуализации данных – Matplotlib;
ML – Catboost, LightGBM;
RecSys – Python (top-popular).

Данные инструменты были выбраны ввиду широкого функционала и скорости работы.

Для построения дашбордов целесообразно
применение средств BI.