

LEVEL Причальный

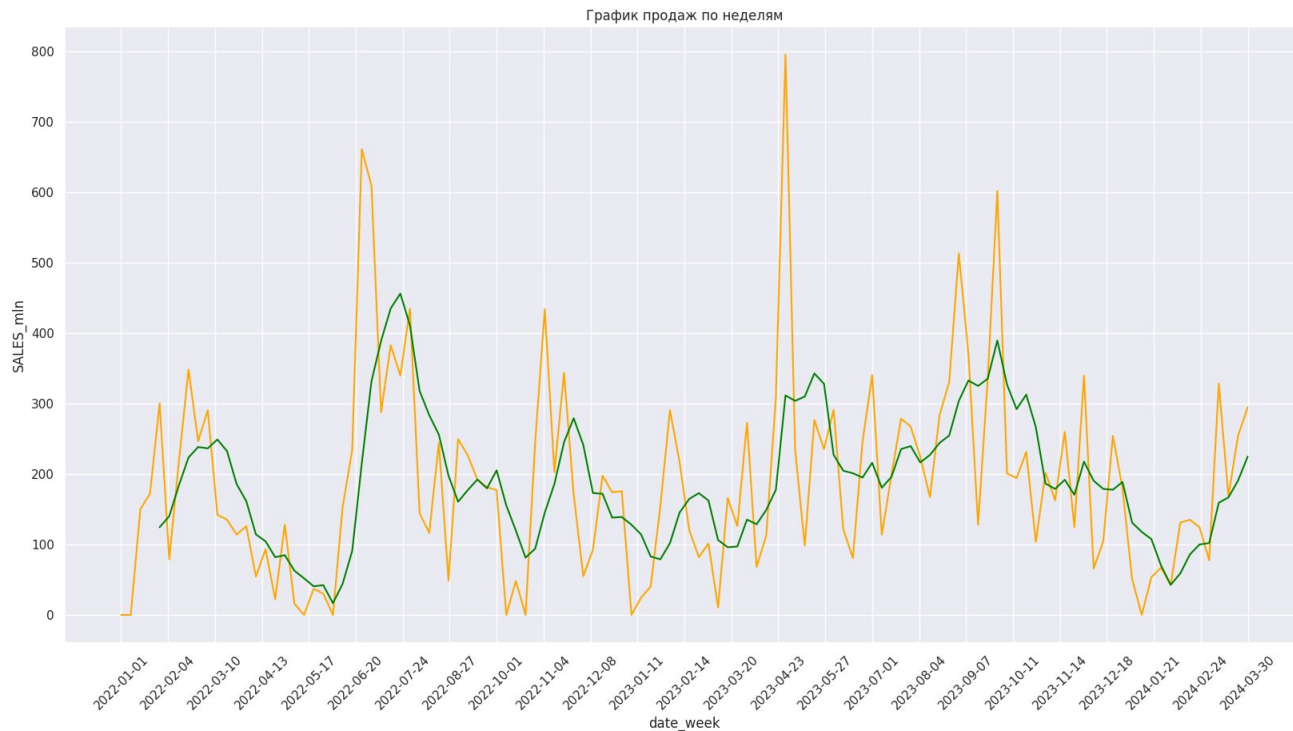
EDA, media contribution analysis

Plan

- EDA
- Media contribution (Marketing mix modeling)
- Suggestions/Insights

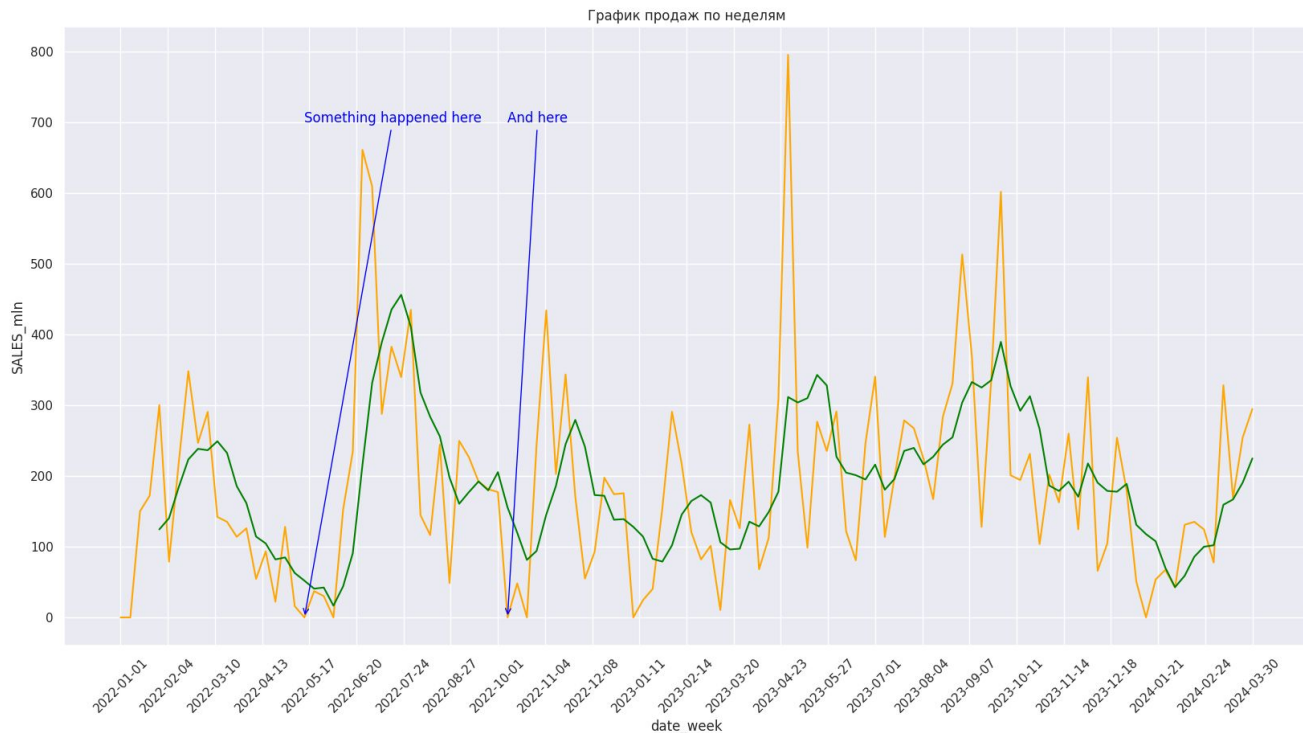
EDA

Начнем с анализа графика целевой переменной



EDA

Начнем с анализа графика целевой переменной



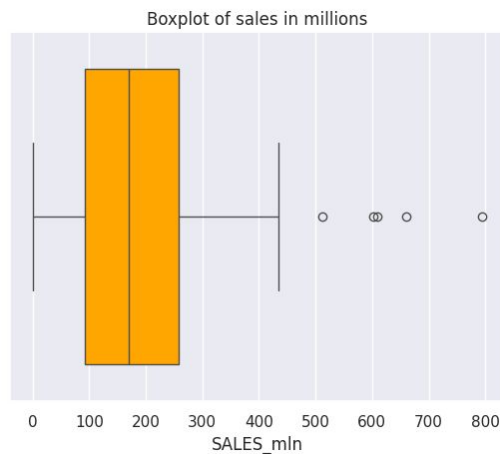
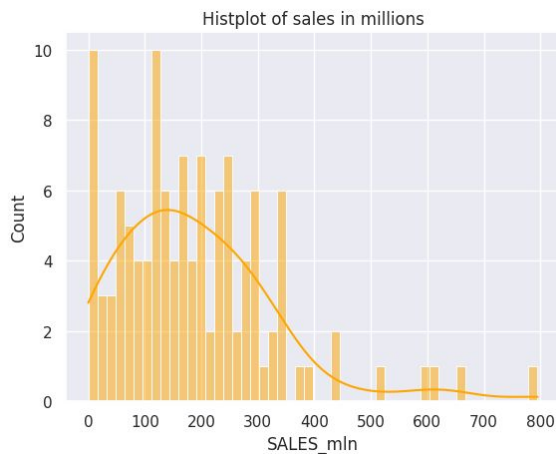
В рамках рассматриваемого периода есть некоторые существенные события, которые очевидно откладывают свои отпечатки.

Они отмечены на графике синими стрелками.

Также сразу бросается в глаза слом тренда - начиная с 2023 года продажи становятся заметно стабильнее и в среднем выше.

EDA

Начнем с анализа графика целевой переменной

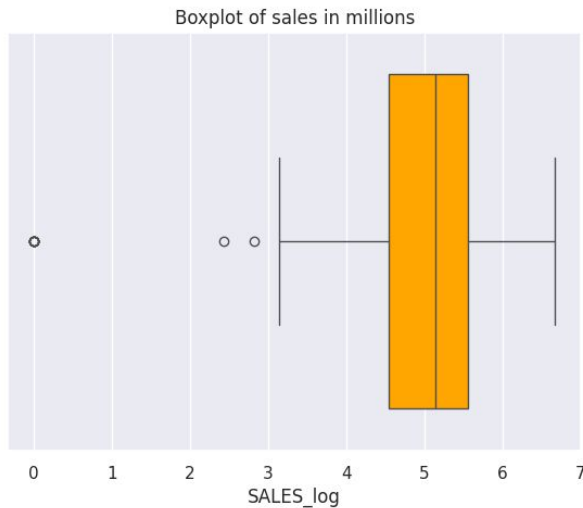
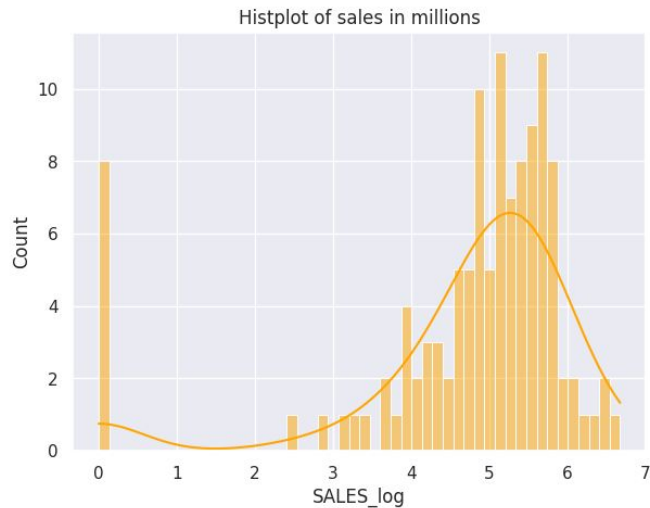


Распределение у продаж смещенное. Видны сильно большие по значениям выбросы.

Возможно хорошей идеей будет масштабировать распределение логарифмированием.

EDA

Начнем с анализа графиков целевой переменной



И действительно это привело к улучшению распределения.

Теперь гистограмма продаж сильно напоминает нормальное распределение.

На boxplot-е также заметно, что большие по значениям выбросы пропали.

Однако не все так идеально - появились нулевые значения

EDA

Начнем с анализа графиков целевой переменной

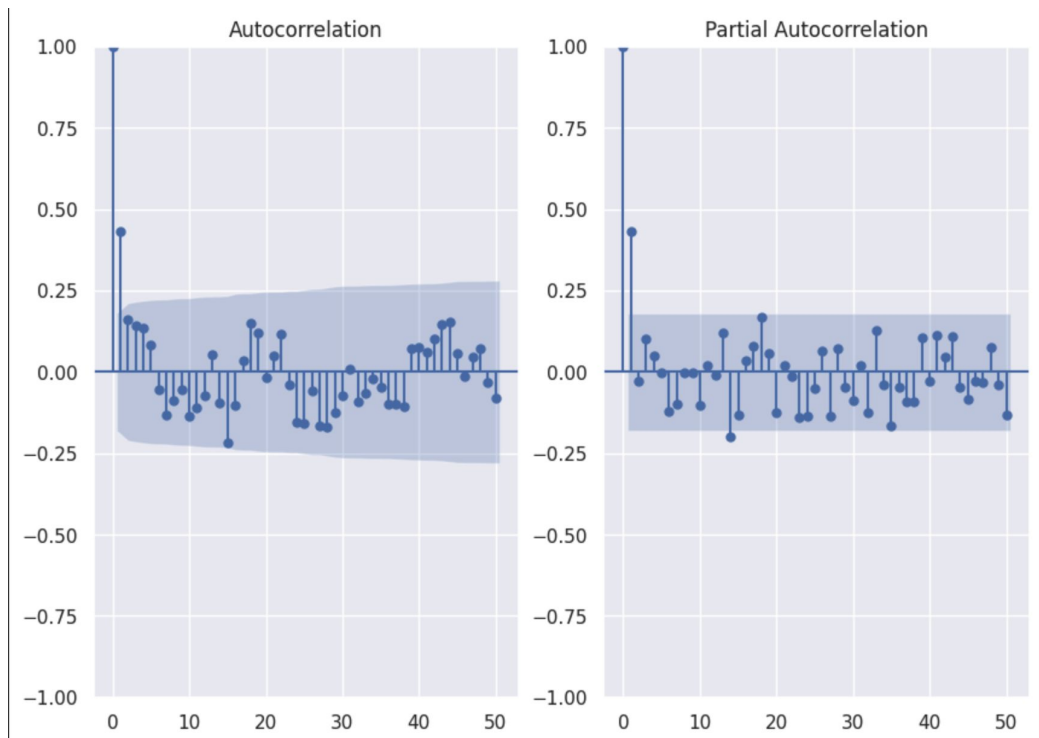


Особенно хорошо стабильность, проявляющаяся в 2023 году становится заметна на графике логарифмированных продаж.

Мы больше не наблюдаем резких просадок в нулевые значения

EDA

Начнем с анализа графиков целевой переменной



Построим графики автокорреляции и частичной автокорреляции.

Возможно мы сможем наблюдать некоторую сезонность в наших данных/тренды.

По графикам acf и $pacf$ видно, что оснований думать о цикличности/сезонности у нас нет.

С каждым сдвигом корреляция убывает и уходит в область малозначимых значений.

EDA

Перейдем к корреляционному анализу



Для начала посчитаем корреляцию с целевой переменной у не медийных признаков.

Таковыми являются различные варианты скидков, агентские и цена за квадратный метр.

Мы видим, что очень сильное влияние на целевую переменную оказывают цена за квадратный метр и агентские.

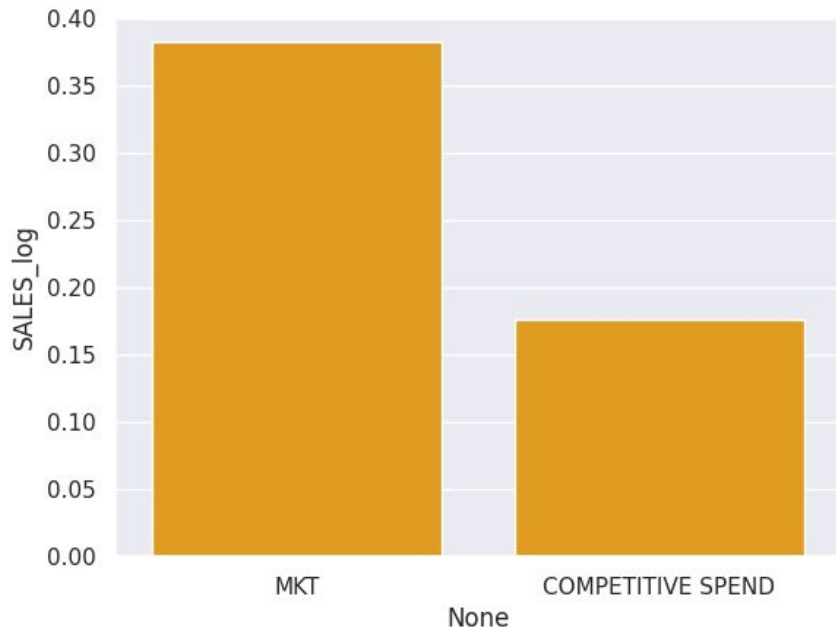
Однако, тут не стоит переоценивать величину этого открытия - агентские очевидным образом растут при увеличении выручки.

С ценой за квадратный метр скорее всего похожая ситуация - повышенный спрос повышает цену за квадратный метр. Но в этом случае еще может играть роль механизм упущенной выгоды - клиенты боятся упустить хорошую цену и начинают покупать еще больше, чем еще больше повышают спрос.

Скидки не особо сильно влияют на целевую переменную. Если и можно что-то выделить - то совсем небольшой положительный вклад скидок 20, 25%. Скидка же 15% имеет отрицательную взаимосвязь с продажами. Кажется, что такие неочевидные результаты стоит проанализировать и изменить подход к скидкам.

EDA

Перейдем к корреляционному анализу



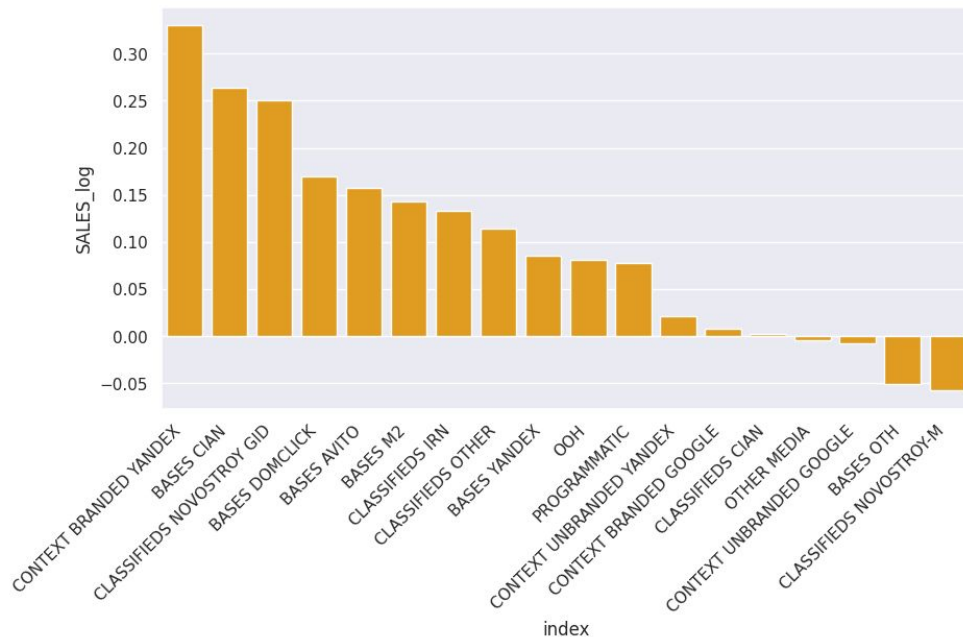
Теперь посмотрим на признаки, связанные с показателями конкурентов.

Тут мы видим некоторую взаимосвязь между выручкой конкурентов и нашей выручкой. Что логично, когда рынок растет - выручка растет у всех.

Также присутствует слабая взаимосвязь с тратами конкурентов. Скорее всего она объясняется тем же, что и связь с выручкой - растущий рынок тянет за собой траты на маркетинг.

EDA

Перейдем к корреляционному анализу



Теперь взглянем на главные признаки - траты по медиа каналам.

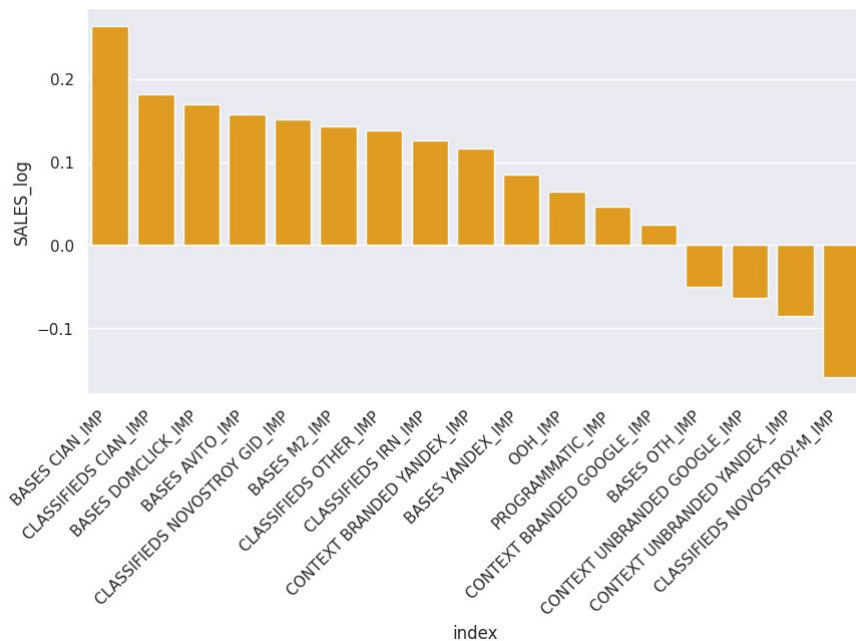
Топ-3 медиа канала по корреляции с целевой переменной:

- Context branded yandex
- Bases cian
- Classifieds novostroy gid

Это подсказывает нам, на какие каналы стоит обратить внимания, но в выводах стоит быть аккуратным. Корреляция мало что говорит об эффективности этих каналов

EDA

Перейдем к корреляционному анализу

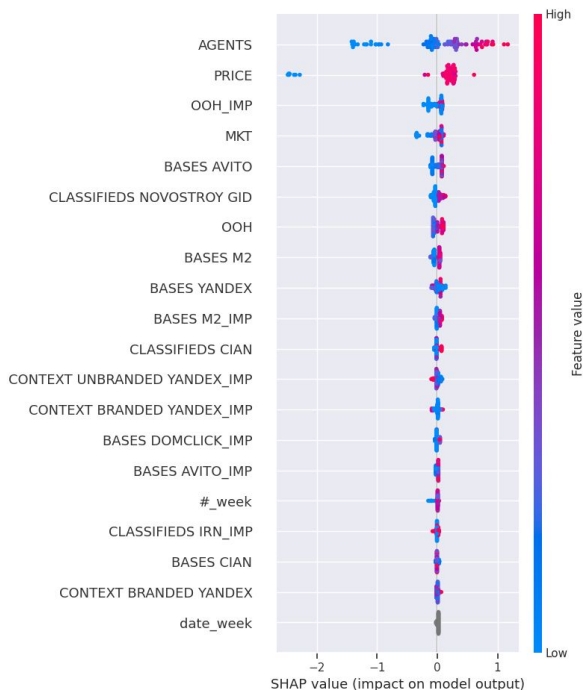


В исследовательских целях можно посмотреть и на корреляцию показов с целевой переменной.

Видно, что показы на специализированных базах и досках объявлений дают больший прирост, по сравнению с каналами контекстной рекламы - это интересное наблюдений.

EDA

SHAP analysis



Еще один интересный подход - SHAP анализ.

Говоря простым языком, он позволяет нам посмотреть на те признаки, на которые больше всего обращала внимания модель предсказывающая целевую переменную - в нашем случае продажи.

Моделью в данном случае была Catboost Regressor с умным подбором параметров через Optuna. Модель специально была выбрана не линейной.

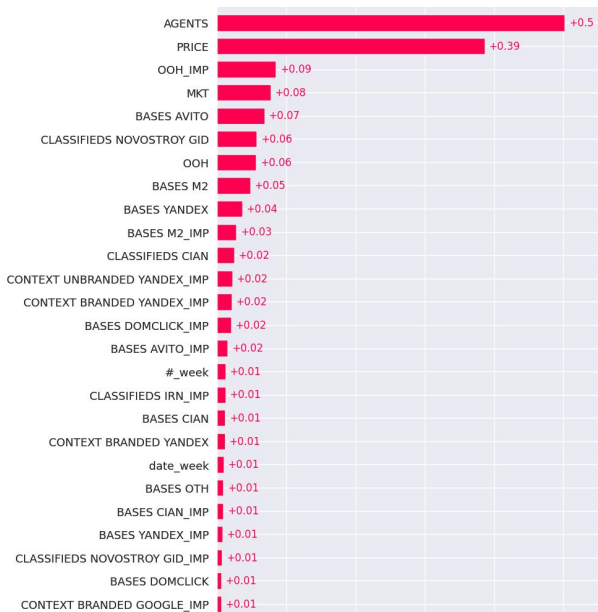
Ранее мы проводили корреляционный анализ признаков, который улавливает линейные закономерности. Теперь хотелось бы проанализировать более сложные варианты зависимостей признаков от целевой переменной.

В целом, результаты дублируют выводы корреляционного анализа. Однако в первых рядах по влиянию у нас появились гости - наружная реклама и ее показы.

Такие примеры показывают, что корреляция не всегда способна уловить сложные связи из-за своей специфики - линейности.

EDA

SHAP analysis



Еще один интересный подход - SHAP анализ.

Говоря простым языком, он позволяет нам посмотреть на те признаки, на которые больше всего обращала внимания модель предсказывающая целевую переменную - в нашем случае продажи.

Моделью в данном случае была Catboost Regressor с умным подбором параметров через Optuna. Модель специально была выбрана не линейной.

Ранее мы проводили корреляционный анализ признаков, который улавливает линейные закономерности. Теперь хотелось бы проанализировать более сложные варианты зависимостей признаков от целевой переменной.

В целом, результаты дублируют выводы корреляционного анализа. Однако в первых рядах по влиянию у нас появились гости - наружная реклама и ее показы.

Такие примеры показывают, что корреляция не всегда способна уловить сложные связи из-за своей специфики - линейности.

EDA

Feature parsing/extraction

Попробуем сгенерировать новые признаки из уже существующих и добавить другие, возможно полезные для анализа признаки, которые мы спарсим из открытых источников.

Из открытых источников была взята информация о:

- Ключевой ставке ЦБ (API цб рф)
- Различным показателям ипотечного рынка России (статистика цб рф):
 - Объем/число предоставленных кредитов
 - Задолженность (отдельно просроченная) по предоставленным кредитам
 - Средний срок кредитования
 - Средняя процентная ставка
- Курсе евро, доллара и юаня (API)

EDA

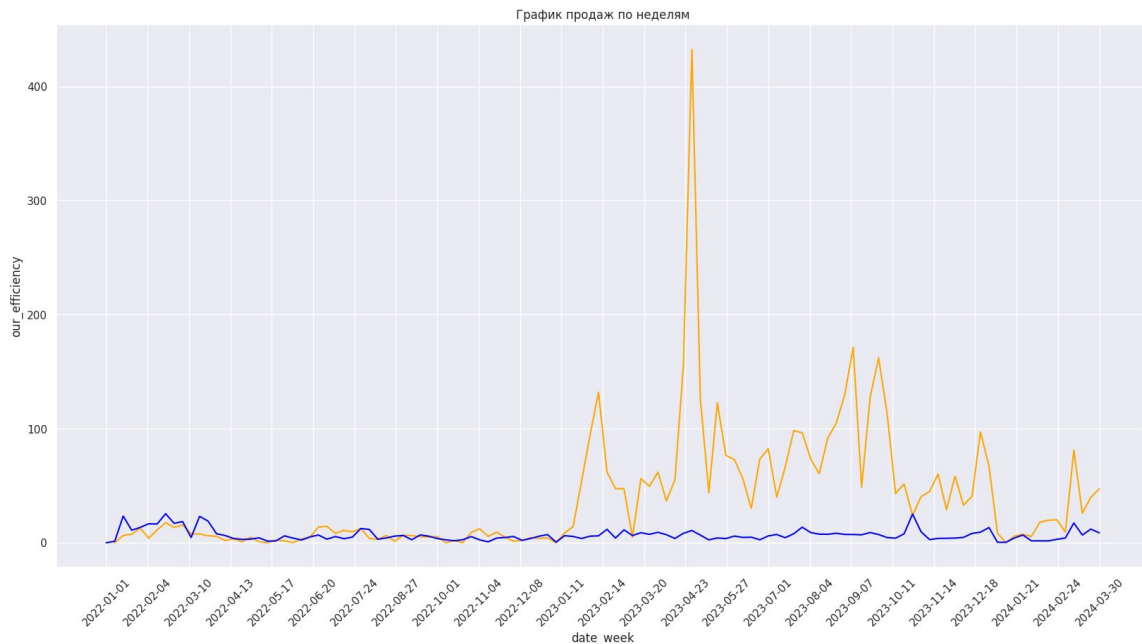
Feature parsing/extraction

Также были сгенерированы следующие признаки:

- CPV (cost per view) для каждого канала
- Media efficiency (income/spends) для каждого канала
- Adstock + saturation для каждого канала

EDA

Feature parsing/extraction



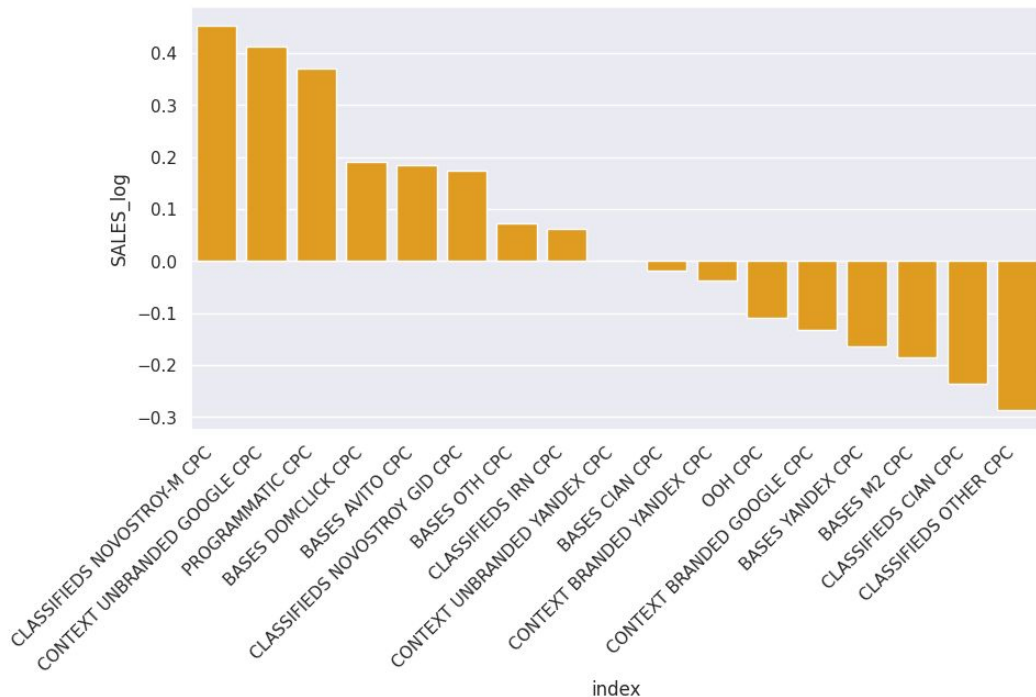
Начнем с признаков сгенерированных по уже имеющимся признакам. Интересные выводы можно получить оценивая эффективность наших медиа расходов в соотношении с продажами.

На графике видно, что в том самом 2023 году происходит резкое изменение, которое не только улучшает исследуемое нами соотношение, но и дает нам серьезный выигрыш по сравнению с конкурентами.

Еще один намек на то, чтобы более подробно изучить этот период.

EDA

Feature parsing/extraction



На этом графике мы видим значения корреляции признаков cost per view по каждому из медиа каналов.

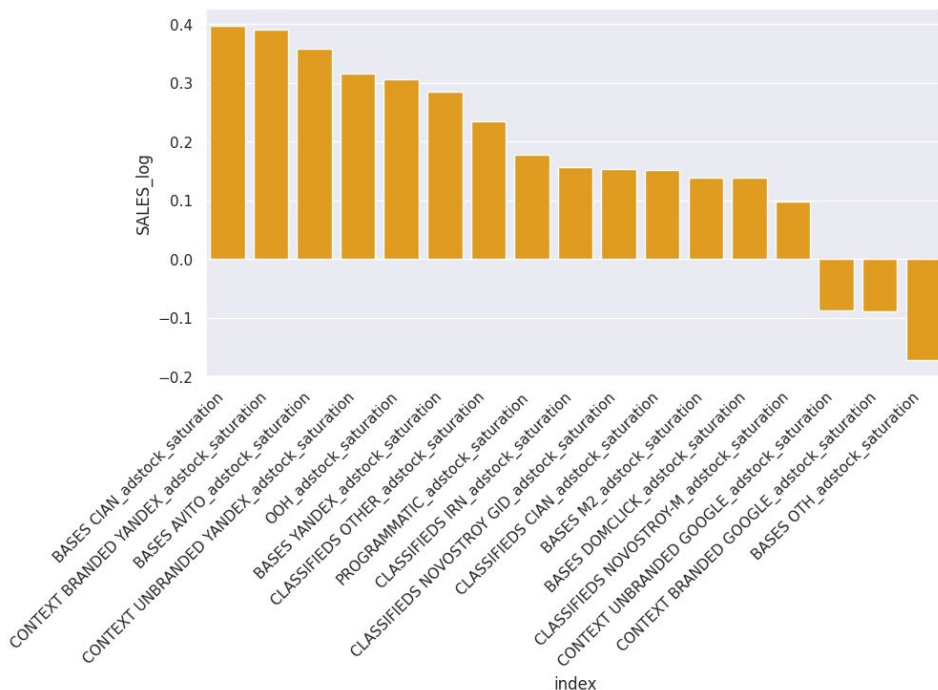
Интерпретация этих данных может быть разной, так как данные признаки состоят из двух других переменных признаков. А значит изменение в любом из них может вести к изменению значений cost per view.

Однако наличие местами положительной, а местами отрицательной корреляции должно нас как минимум наталкивать на мысль о том, что данные показатели нужно исследовать более подробно.

Применение к признакам таких

EDA

Feature parsing/extraction



Кажется, что у рекламы должно быть два свойства: аккумулярование и насыщение.

Для того чтобы учитывать оба этих свойства мы производим с нашими тратами по медиа каналам два преобразования:

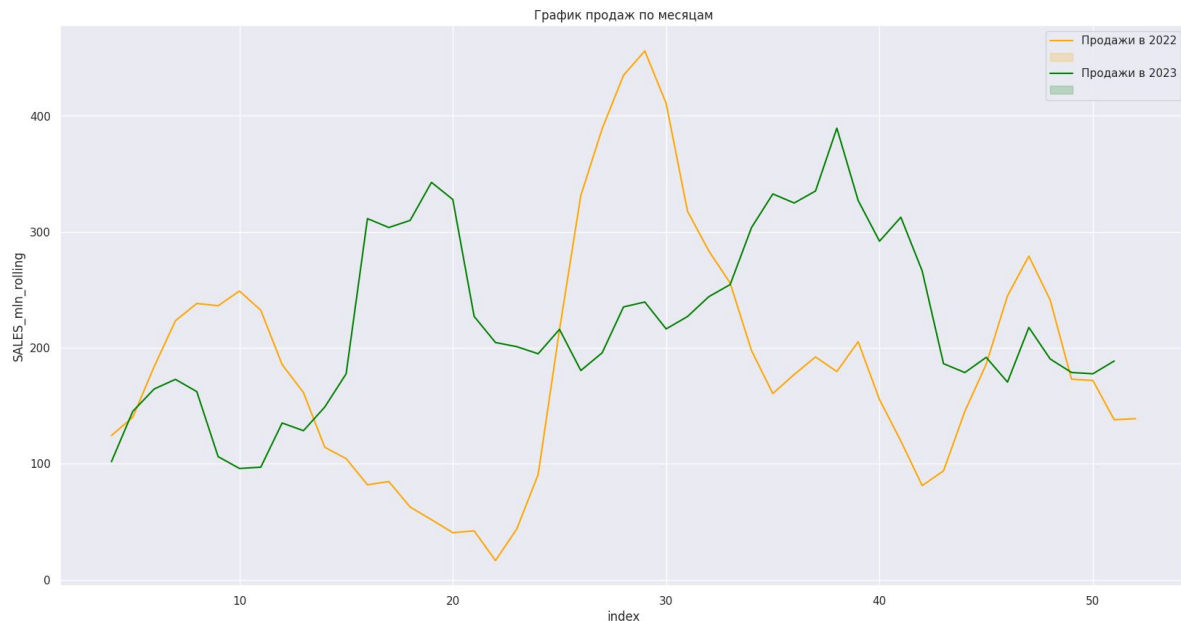
- Adstock
- Saturation (log)

Результат корреляции полученных признаков с целевой переменной мы можем наблюдать на графике - видим что в среднем корреляция по каналам повысилась.

Это подтверждает нашу теорию о том, что рекламе нужно время на действие и взвешивание инвестиций.

EDA

Анализ перехода в 2023



На графике мы видим сравнение по месяцам 2022 и 2023 годов.

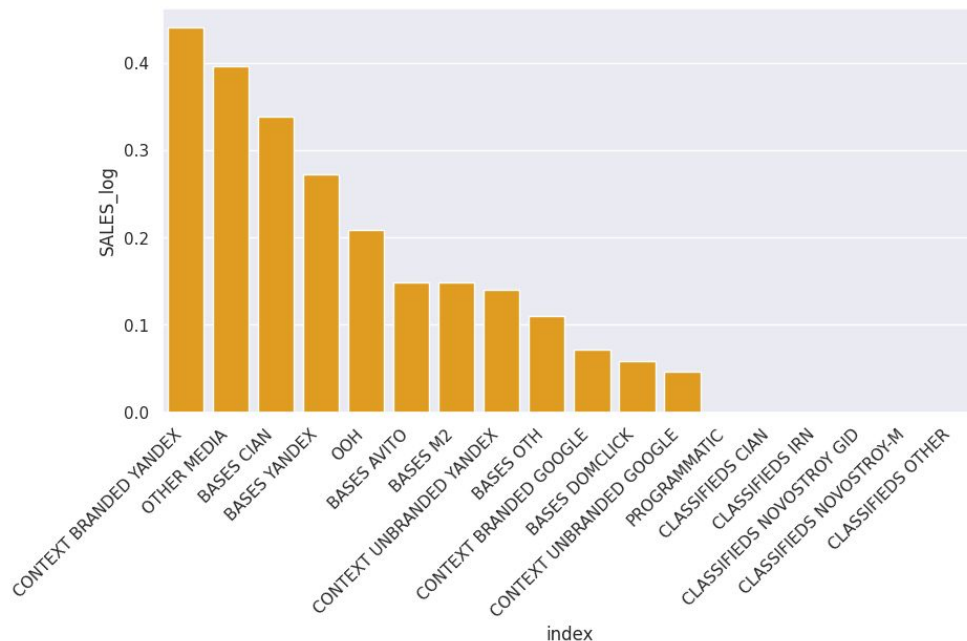
Предварительно было проведено преобразование подсчетом скользящего среднего для сглаживания временных рядов.

Этот график наглядно подтверждает нашу теорию: в 2023 что-то поменялось и поменялось в лучшую сторону.

Попробуем понять что.

EDA

Анализ перехода в 2023



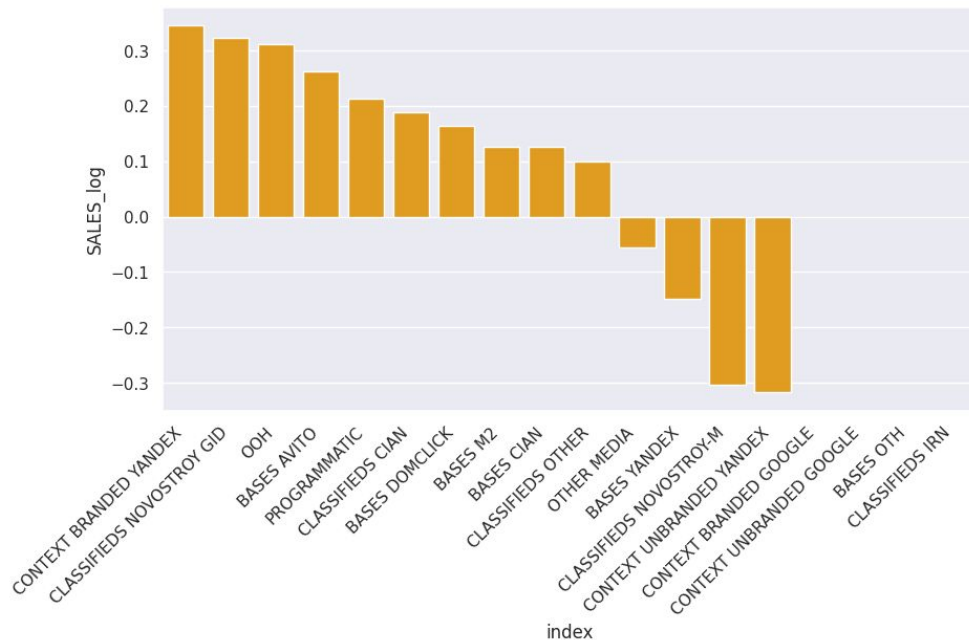
По нашему набору данных видно, что при переходе в 2023 год изменилась стратегия и набор используемых каналов.

До перехода 2023 года не использовались программтики и площадки объявлений.

Также огромную долю от медиа расходов занимали other media. Да, в таблице корреляций мы видим, что other media действительно связано с целевой переменной - но стоят ли затраты того?

EDA

Анализ перехода в 2023



В 2023 году были на порядок сокращены расходы по каналу other media. Также началось использование классифайдов и программатиков.

Расходы на маркетинг упали на порядок, а продажи не то что не упали - а даже выросли!

Отсюда мы можем сделать вывод о том, что новые каналы компенсировали уменьшение расходов в other media при не сопоставимо меньших затратах.

EDA

Анализ перехода в 2023

	index	0
0	D15	0.000000e+00
1	D20	4.150943e-01
2	D25	0.000000e+00
3	D30	1.509434e-01
4	PRICE	3.248581e+05
5	MKT	1.064540e+09
6	COMPETITIVE SPEND	1.619484e+08
7	AGENTS	1.459707e+06
8	key_rate	1.052830e+01
9	usd	6.839206e+01
10	cny	1.027208e+01

	0
D15	0.000000e+00
D20	0.000000e+00
D25	3.461538e-01
D30	9.615385e-02
PRICE	3.783268e+05
MKT	1.612465e+09
COMPETITIVE SPEND	2.417261e+08
AGENTS	1.191236e+06
key_rate	9.961538e+00
usd	8.540565e+01
cny	1.372164e+01

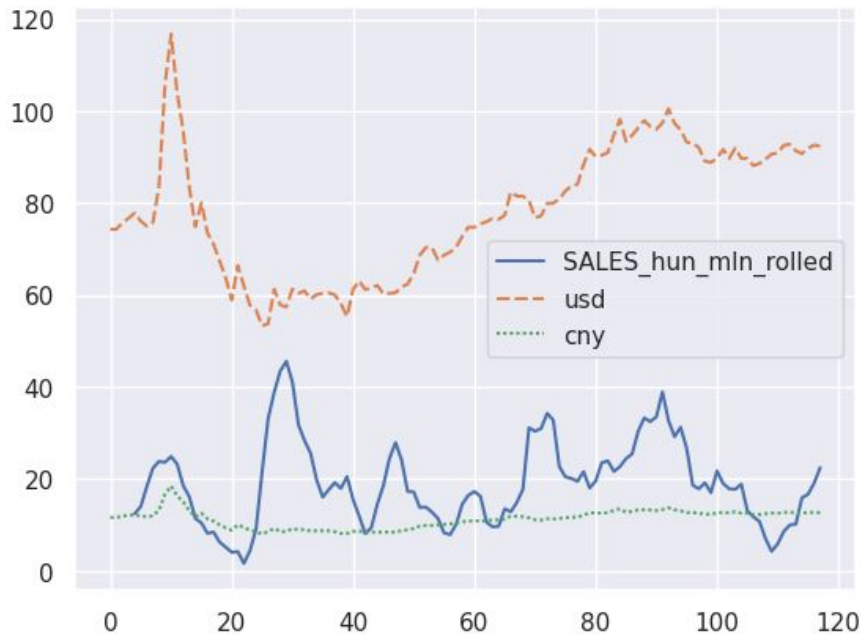
Для сравнения, посмотрим на прочие (не медийные признаки).

Тут мы видим либо не актуальные для прогноза продаж, либо слабо коррелированные с продажами признаки.

Поэтому их изменение вряд ли повлекло тот самый подъем продаж в 2023.

EDA

Связь с курсом валют



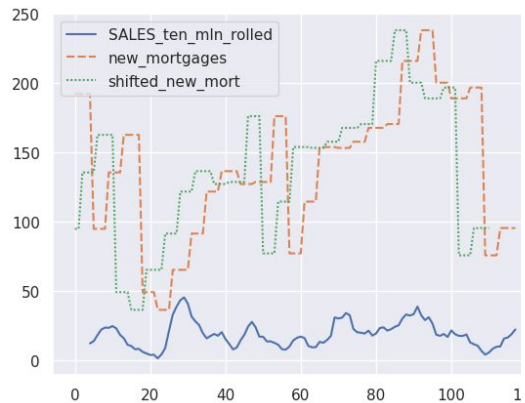
На совместном графике курса валют и продаж мы можем наблюдать совсем слабый совместный рост.

На основе графика можно сказать, что вряд ли курс доллара существенно повлиял на продажи.

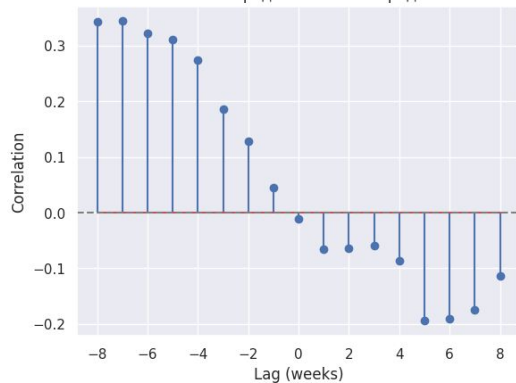
Да, при сильном росте валют (падении рубля) есть тенденция на увеличение покупок недвижимости - но таких серьезных скачков у нас не происходило.

EDA

Связь с ипотечными метриками



Cross-correlation: Количество предоставленных кредитов за месяц vs. sales

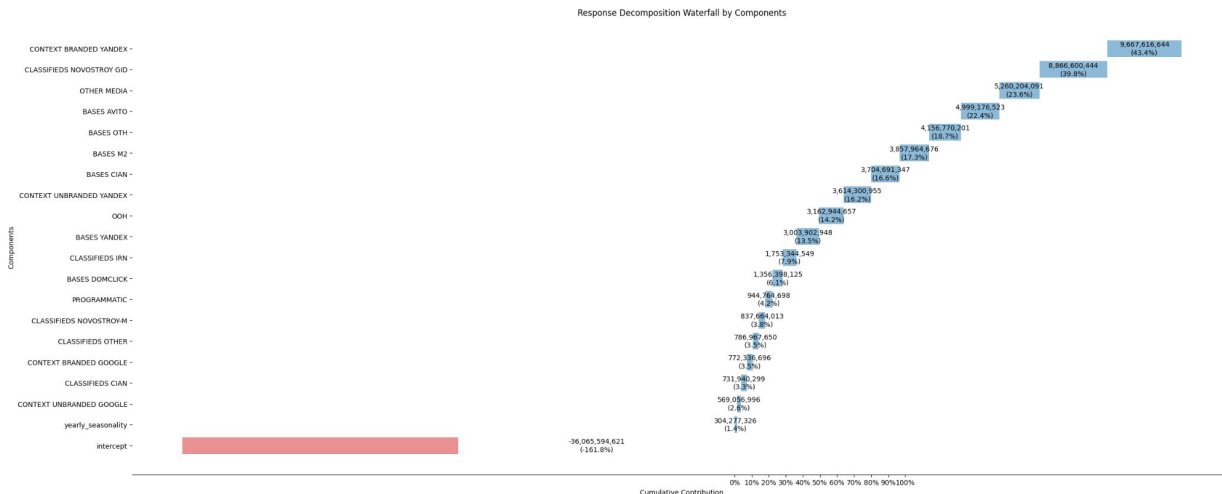


Единственная из ипотечных метрик, которая имела хоть небольшое влияние на продажи оказалась количество выданных кредитов.

Что в принципе логично, и отсюда мы можем сделать лишь вывод, что большая часть наших квартир продается в ипотеку, а не за наличные деньги.

Marketing mix modeling

Media contribution



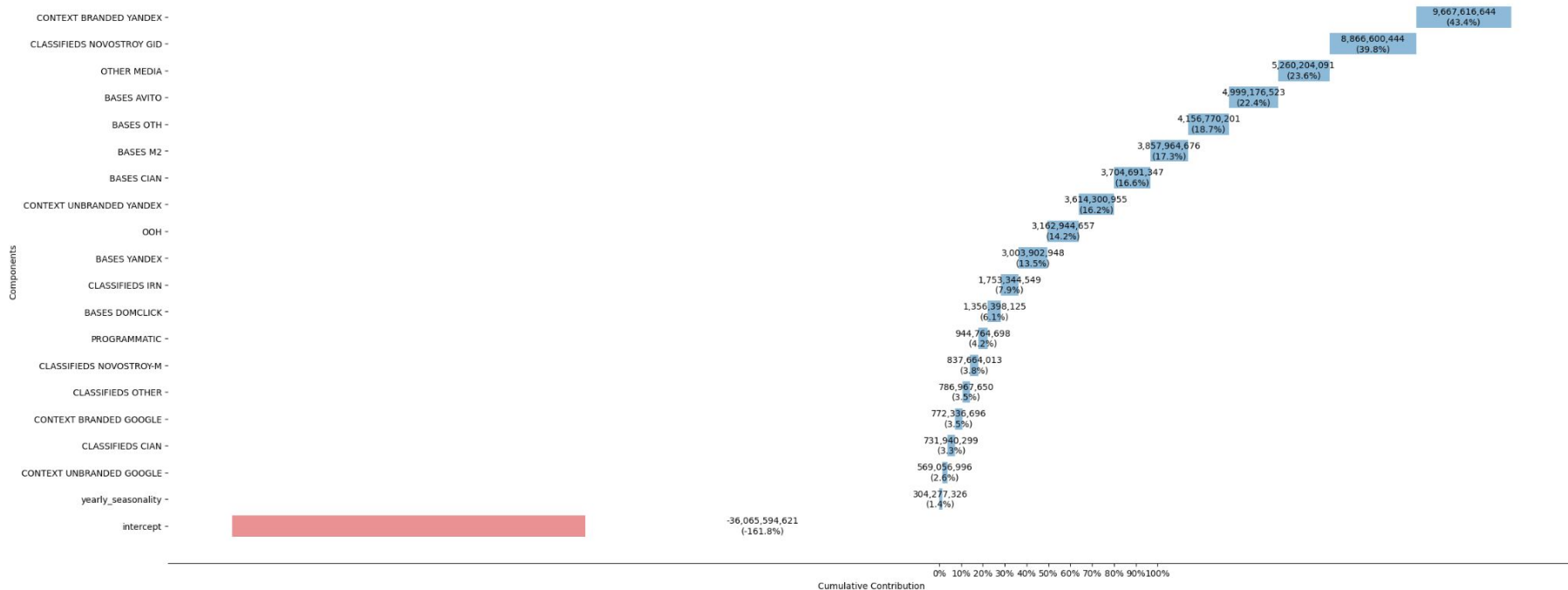
Теперь перейдем к основной части исследования - построению модели media contribution.

Мы будем использовать для построения модели фреймворк руМС. Он позволяет строить модели на основе байесовского подхода, который особенно выгоден в случаях малого числа данных. У нас как раз такой случай.

Плюсом на выходе мы можем получить распределение признаков, чего не бывает с классическими моделями машинного обучения

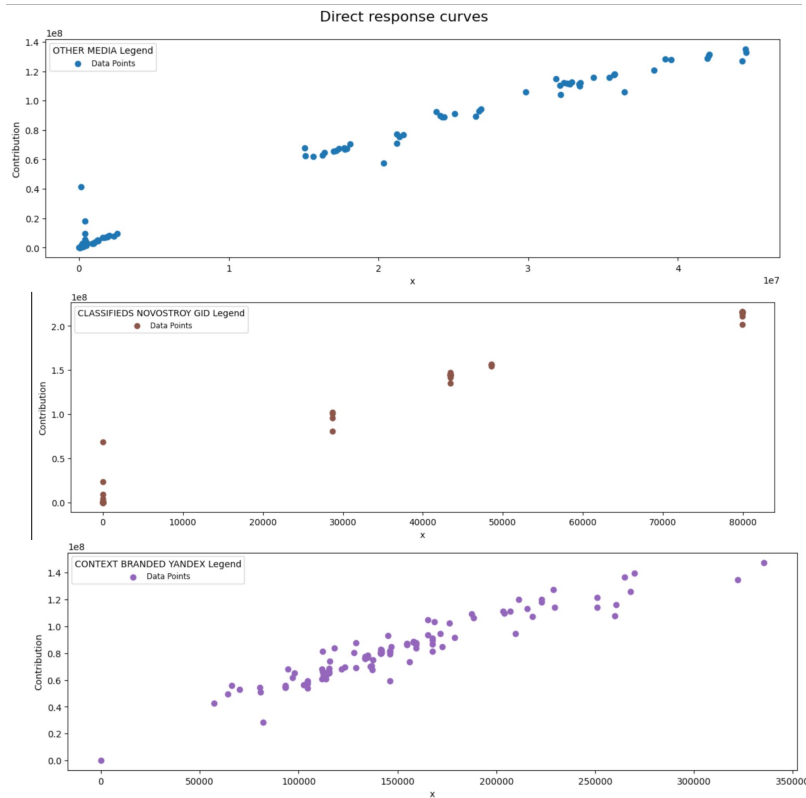
Media contribution

Response Decomposition Waterfall by Components



Marketing mix modeling

Media contribution



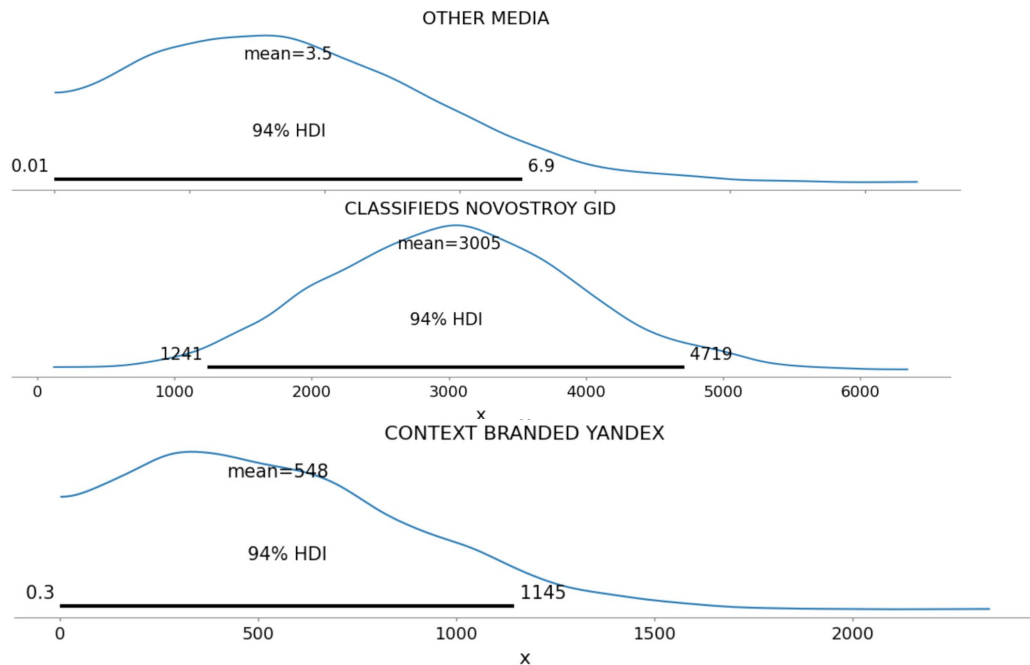
Приведем диаграммы рассеивания для трех наиболее важных по contribution признаков:

- Other media
- Classified Novostroy gid
- Context branded yandex

Выводы модели в какой-то степени подтверждают выводы нашего корреляционного и shap анализом!

Marketing mix modeling

Media contribution



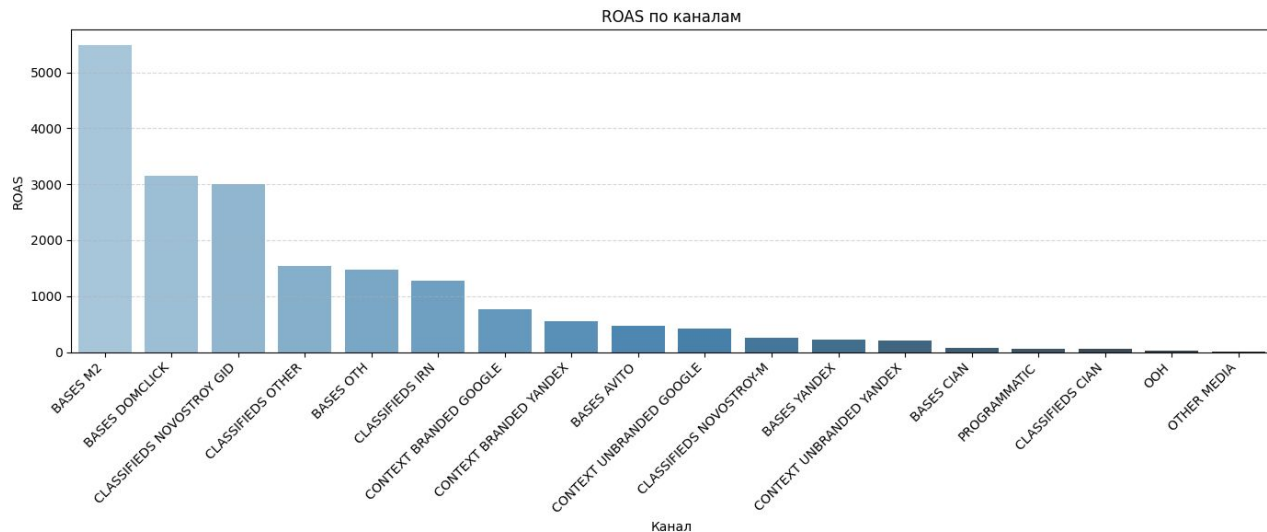
А теперь посмотрим на оценку метрики ROAS нашей моделью и увидим интересные выводы.

Как раз поэтому other media при высокой корреляции с продажами поглощала такие объемы средств - на вложенный рубль мы получали всего лишь 3.5 рублей дохода.

Тогда как контекстная реклама яндекса на тот же вложенный рубль возвращает 548 рублей, а классифайд новострой гит - аж целых 3005!

Marketing mix modeling

Media contribution



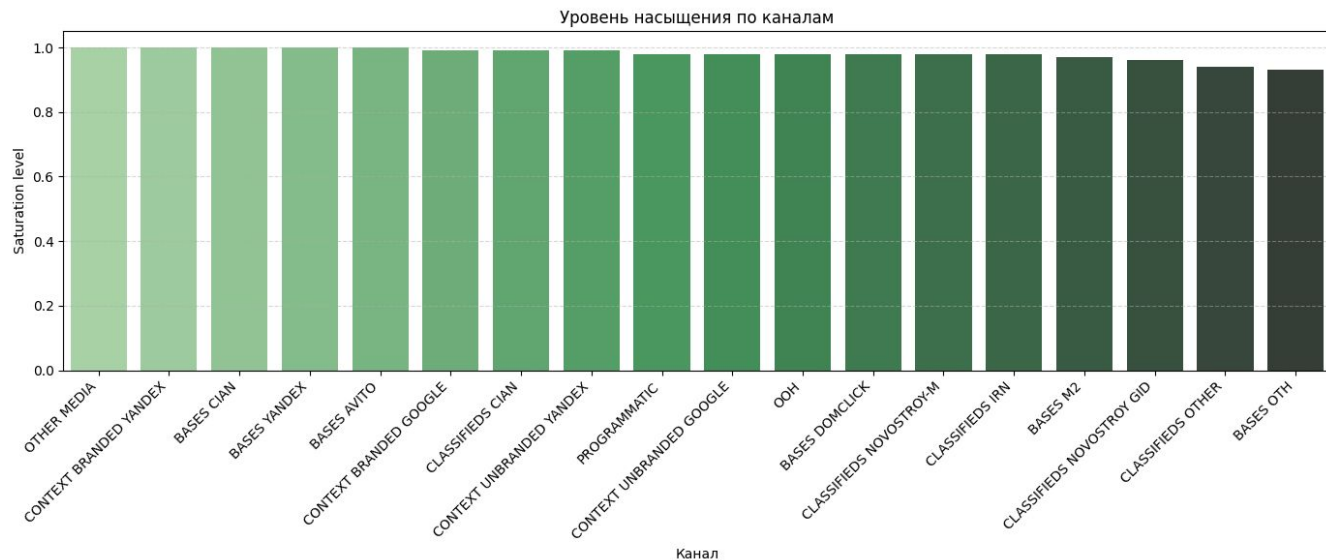
Теперь взглянем на средний ROAS по всем признакам.

Мы видим, что у other media самый низкий показатель из всех задействованных каналов.

При этом задействованные после 2023 каналы классифайд показывают одни из лучших показателей ROAS среди всех используемых каналов.

Marketing mix modeling

Media contribution



Однако существует и другая сторона медали.

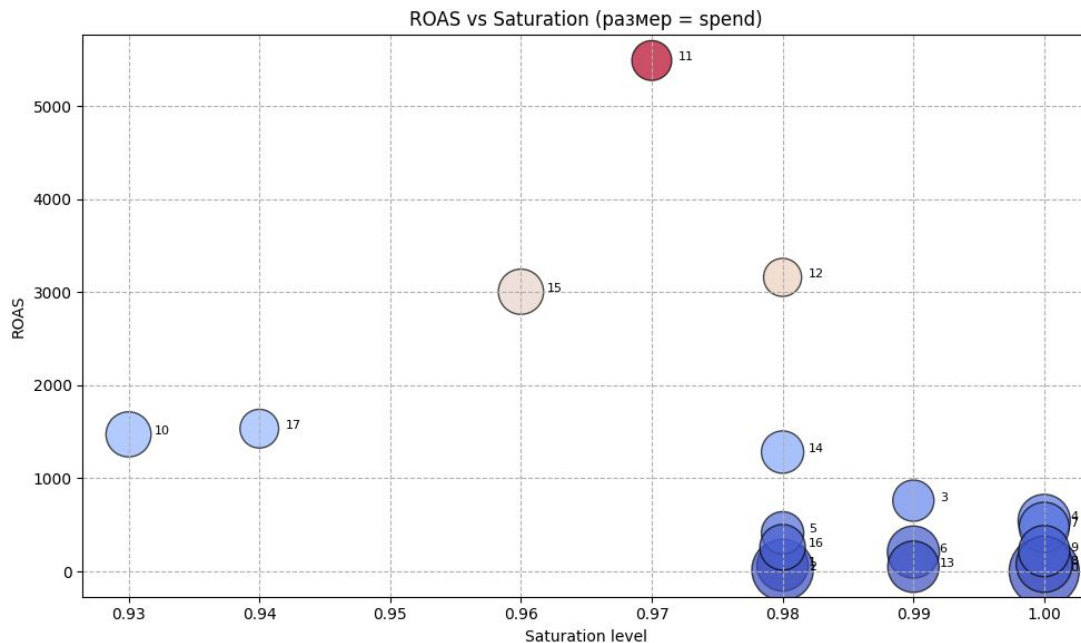
Как упоминалось ранее, медиа каналы имеют свойства насыщения - когда каждая дополнительная инвестиция приносит меньшие результаты.

И мы видим, что модель считает, что все используемые нами каналы находятся на пределе сатурации.

К плюсам можно отнести то, что некоторые из наиболее эффективных каналов, используемых имеют мизерный запас - но ситуация требует альтернативных решений

Marketing mix modeling

Media contribution

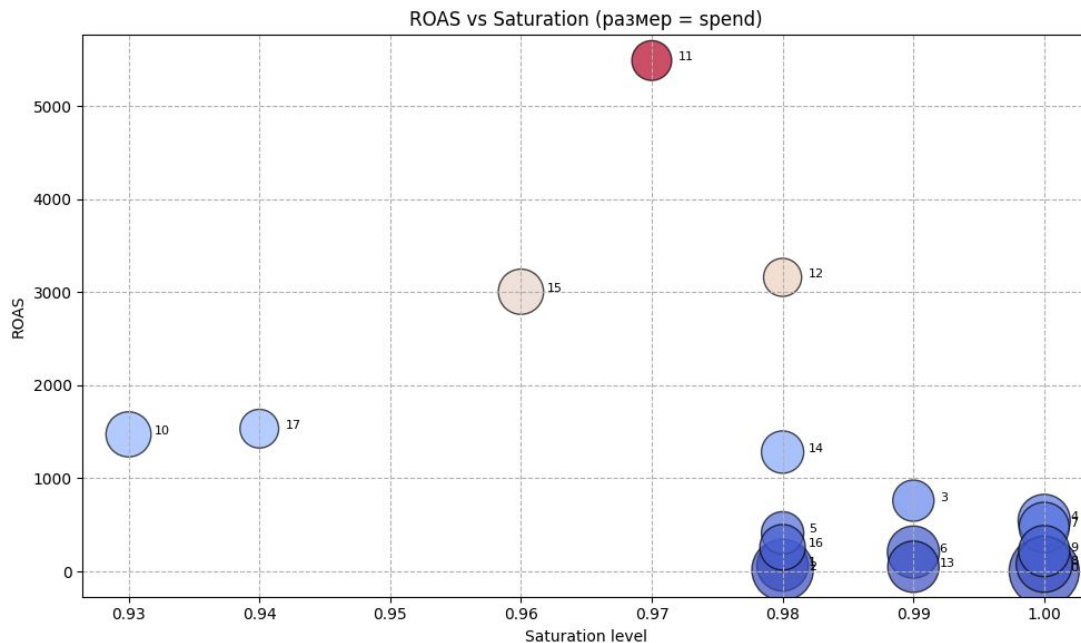


Обратим внимание на еще один график, который сопоставляет уровень сатурации и метрику ROAS.

И на нем видно, что некоторые из наиболее эффективных признаков еще имеют некоторый запас (который стремится к нулю)

Marketing mix modeling

Media contribution



Обратим внимание на еще один график, который сопоставляет уровень сатурации и метрику ROAS.

И на нем видно, что некоторые из наиболее эффективных признаков еще имеют некоторый запас (который стремится к нулю)

Suggestions/Insights

Учитывая весь приведенный выше анализ, мы можем придти к следующим идеям развития маркетинговой стратегии в рамках нашей задачи:

- Other media
 1. Остановить инвестиции в other media было хорошей идеей
 2. Sentiment analysis (next step) - был бы хорошим инструментом для оценки медиа инвестиций и возможно позволит бы повысить ROAS у other media
 3. Не всегда наиболее эффективные каналы являются самыми затратными
- Новые каналы
 - Поиск новых каналов окупается (классифайды)
 - Нужно активнее использовать каналы с высоким ROAS и запасом по сатурации
 - Почти все каналы полностью израсходовали запас сатурации, мы нуждаемся в поиске новых каналов

Suggestions/Insights

Other Media

1. Канал other media потреблял огромное количество ресурсов при низких показателях окупаемости. Прекратить инвестиции в него отдать предпочтение другим, более прогнозируемым вариантам было хорошей идеей.
2. Однако это не значит, что канал не эффективен. Возможно стоит вопрос неэффективного использования бюджета. Можно попробовать автоматизировать сбор обратной связи с помощью анализа соцсетей. К примеру sentiment analysis комментариев у блоггеров, у которых была заказана такая компания. Это позволит более эффективно подбирать каналы внутри other media.
3. Хороший пример того, как большие вливания не гарантируют лучшего результата. Да, вклад other media в продажи был серьезный, однако он не сравним с теми ресурсами, которые канал требовал, тем более в сравнении с альтернативами

Suggestions/Insights

Новые каналы

1. Добавление новых каналов в начале 2023 года с лихвой окупилось и позволило очень сильно увеличить эффективность рекламной стратегии.
2. В ближайшей перспективе есть только одна опция улучшать показатели - давить на каналы с высоким ROAS и хоть каким-то запасом по сатурации.
3. Почти все каналы насыщены до предела. Необходимо проводить поиск новых каналов. Возможно проводить статистические тесты на влияние различных наборов новых каналов на продажи. Текущий набор каналов почти израсходовал свои возможности. Что точно стало ясно - поиск новых каналов возможен (в 2023 получилось) и зачастую приводит к скачку в эффективности рекламной стратегии. Поэтому инвестиции в research с большой вероятностью окупят себя с головой.