

Приложение № 15
к отчету о результатах
контрольного мероприятия

Результаты аналитических исследований

I. Сравнение «эффективности» реализации (продаж / аренды) земельных участков разными субъектами публичной собственности

Исследовательский вопрос

Земельные участки реализуются (продаются, сдаются в аренду) разными субъектами (не берем реализацию частными собственниками): на федеральном уровне (Росимущество, агент Российской Федерации – ДОМ.РФ), Субъектами Российской Федерации, муниципалитетами.

Цель анализа – оценить имеются ли различия в конечной цене реализации земельного участка в зависимости от того реализовывался ли земельный участок Российской Федерацией, Субъектами Российской Федерации или муниципалитетами¹.

Нулевая гипотеза состоит в том реализация земельных участков на федеральном уровне (Росимуществом или агентами Российской Федерации) дает меньшую цену, чем реализация муниципалитетами или Субъектами Российской Федерации.

Данные

Данные о торгах земельными участками, самих земельных участках были получены из открытых данных об извещениях о торгах, результирующих протоколах об итогах торгов земельными участками из открытого информационного ресурса «ГИС Торги» (<https://torgi.gov.ru/new/public/opendata/reg>).

Формирование датасетов

Шаг 1.

С «ГИС Торги» были выгружены данные об извещениях о торгах за период с сентября 2021 по декабрь 2022 года и о результирующих протоколах об итогах торгов с сентября 2021 по апрель 2023 года². Новая версия сайта «ГИС Торги» устроена таким образом, что непосредственно в момент выгрузки невозможно указать тип имущества, по которому нужна выгрузка извещений, протоколов, в этой связи первично потребовалось выгрузить 138 906

¹ Для целей настоящей аналитической записки, если отдельно не уточнено, то под реализацией понимаем покупку или заключение договора аренды земельного участка.

² Необходимо отметить, что в 2022 году была введена в эксплуатацию новая версия сайта torgi.gov.ru. Начало периода выгрузки – сентябрь 2021 года, соответствует началу появления данных на новой версии сайта torgi.gov.ru, окончание периода выгрузки – моментом формирования финального датасета с результатами торгов для анализа. До 17.04.2023 функционировала также старая версия сайта torgi.gov.ru, содержащая данные о торгах земельными участками за более ранние периоды, в процессе исследования получить данные из раздела открытых данных со старой версии сайта torgi.gov.ru не удалось, т.к. с 17.04.2023 данные со старой версии сайта torgi.gov.ru стали недоступными. В ходе мероприятия были сделаны запросы в Федеральное казначейство, Минэкономразвития России, Минцифры России с целью получить данные со старой версии сайта torgi.gov.ru о торгах земельными участками. В силу того, что старая версия сайта torgi.gov.ru выведена из эксплуатации, ограничений на программное обеспечение для возможной работы с архивом старой версии сайта/torgi.gov.ru, данные со старой версии сайта torgi.gov.ru за более ранние периоды получить не удалось.

извещений (каждое извещение – отдельный файл в формате json) и 121 483 протоколов. Общий объем данных в файлах составил примерно 20 Gb. Детально процедуру выгрузки файлов из открытых данных «ГИС Торги» можно посмотреть в прилагаемых скриптах на языке python: load_new_torgi_notifications.ipynb (извещения) и load_new_torgi_protocols.ipynb (протоколы).

Шаг 2.

Вторым шагом для формирования датасета для обработки и анализа данных был отбор файлов (извещений и протоколов), которые относятся к торгам земельными участками. После фильтрации файлов, относящихся к торгам земельными участками, осталось 32 713 файлов - извещения и 13 367 файлов - протоколов, детальнее с процедурой нужного типа имущества можно ознакомиться в скриптах: copy_notices_ZK_ZKPT_701PP_200FZ.ipynb (извещения), copy_protocols_ZK_ZKPT_701PP_200FZ.ipynb (протоколы).

Шаг 3.

Третий шаг объединение отдельных файлов с извещениями и отдельных файлов с протоколами в общие файлы со всеми извещениями и всеми протоколами (детальнее в скрипте concat_files_all_types.ipynb).

Записи об извещениях о торгах и о протоколах содержат иерархические структуры. Это означает, что в одном извещении, например, может быть выставлено несколько земельных участков на реализацию, т.е. извещение содержит составное поле «*lots*» (лоты), в котором в формате словаря описаны характеристики выставленных для реализации земельных участков: один земельный участок – один лот, в одном извещении могло быть более 10 лотов. Аналогично в протоколах в одном поле приведена информация обо всех участниках торгов, число участников могло варьироваться от 0 (никто не участвовал) до нескольких десятков. Чтобы далее можно было работать с датасетом такие составные поля требовалось развернуть в разрезе земельных участков (лотов), и оставить нужные для исследования поля. Такая развертка происходила на четвертом этапе (скрип ZK_parsing.ipynb).

Шаг 4.

На четвертом шаге формирования датасета было сделано:

4.1. Удалены неактуальные версии извещений.

Дубликаты извещений появляются, например, в случае исправления ошибок, опечаток организаторами торгов, в этом случае появляется новая запись об извещении с указанием номера версии извещения (поле «*version*»), для каждого извещения была оставлена только одна – последняя версия.

После исключения таких дубликатов из 32 713 извещений осталось 29 380.

4.2. Из составных извещений с несколькими лотами была выделена информация по каждому лоту (земельному участку).

После развертывания поля с лотами в датасете с извещениями оказалось 52 091 записей о торгах земельными участками.

4.3. Развёрнутое поле с основными характеристиками земельного участка.

Характеристики земельного участка также содержались в одном поле: один участок – одно поле с характеристиками. При этом разные участки могли содержать разный набор полей с характеристиками.

К характеристикам земельного участка относились (открытый список):

- площадь,
- кадастровый номер,
- вид разрешённого использования,
- цель использования,
- числится ли в государственном реестре объектов культурного наследия.

Поле вид разрешенного использования содержало 143 варианта использования, при этом присутствовала многоуровневая классификация видов разрешённого использования, например для разрешённого использования – производственная деятельность были следующие варианты:

- «6.0 Производственная деятельность»,
- «6.1 Недропользование»,
- «6.2.2001 Автомобилестроительная промышленность»,
- «6.2 Тяжелая промышленность»,
- «6.3.2001 Фармацевтическая промышленность»,
- «6.3 Легкая промышленность»,
- «6.4 Пищевая промышленность»,
- «6.5 Нефтехимическая промышленность»,
- «6.6 Строительная промышленность»,
- «6.7 Энергетика»,
- «6.8 Связь»,
- «6.9.2001 Складские площадки»,
- «6.9 Склад»,
- «6.11 Целлюлозно-бумажная промышленность»,
- «6.12 Научно-производственная деятельность».

В силу многообразия вариантов видов разрешённого использования для целей дальнейшего анализа был оставлен только первый уровень градации видов разрешенного использования (до первого знака).

Несколько лотов было без указания вида разрешённого использования, такие лоты были исключены из дальнейшего рассмотрения, после исключения в датасете с извещениями осталось 52 057 записей.

4.4. Развёрнутое поле с дополнительной информацией о лоте.

В извещении к дополнительной информации о лоте в частности относилось:

- вид договора (продажа или аренда),
- срок аренды,
- ограничение прав на земельный участок,
- участок только для граждан,
- цель предоставления земельного участка,
- участок только для крестьянских (фермерских) хозяйств,
- участок только для субъектов малого и среднего предпринимательства,
- право лица, с которым заключен договор аренды.

4.5. Удалены неактуальные версии протоколов.

Дубликаты протоколов, как и извещений, появляются, например, в случае исправления ошибок, опечаток организаторами торгов, в этом случае появляется новая запись о протоколе с указанием номера версии протокола (поле «version»), для каждого протокола была оставлена только одна – последняя версия.

Кроме того, в ряде случаев организаторы торгов при формировании новой версии протокола присваивают ей новый номер протокола, т.е. для одного лота может быть два протокола с различными номерами. Дублирующие протоколы были исключены из датасета, оставлен только последний экземпляр протокола.

После исключения таких дубликатов из 13 367 протоколов осталось 12 601.

4.6. Развернутое поле с результатами торгов.

Результаты торгов также содержаться в одном поле. В поле в частности содержится информация об участниках торгов, номер участника, допущен или нет к торгу, цена предложенная участником, место участника по результатам торгов.

4.7. Для каждого лота подсчитано число участников, допущенных до торгов.

4.8. Датасет с протоколами объединен с датасетом с извещениями.

Далее, до процедуры поиска похожих земельных участков, реализуемых Российской Федерации, Субъектами Российской Федерации, муниципалитетами, обработка датасетов осуществлялась в PolyAnalyst:

<https://polyanalyst.ach.achrf.ru/polyanalyst/static/paclient/pabclient.html?locale=rus&prjUUID=253ab669-2e1f-435c-944e-4b91c7b38fb&objId=224>

Основная информация о земельных участках, содержится в извещениях, поэтому протоколы необходимо было объединить с извещениями. Объединения датасетов осуществлялось по составному ключу: номер извещения – номер лота. Для небольшой части протоколов не нашлось информации в датасете с извещениями.

После объединения датасетов в объединенном датасете осталось 12 523 записи о протоколах с информацией из извещений.

4.9. Фильтрация записей с заполненной информацией об организаторе торгов.

В извещениях присутствует информация о том, кто реализует участок. Данное поле ключевое для данного исследования, т.к. именно оно содержит название субъекта реализации. Из интересующих нас вариантов заполнения данного поля: РосИмущество, ДОМ.РФ, Субъект Российской Федерации, Муниципалитет.

Не для всех записей данное поле было заполнено, после исключения записей с пустым значением по данному полю в объединенном датасете осталось 12 103 записи.

4.10. Разделение датасета на аренду и продажу.

Сравнительная эффективность реализации анализировалась по типам реализации, отдельно для аренды, отдельно для продажи.

После разделения датасетов в датасете с арендой стало 8 926 записей, в датасете с продажами 2 427 записи.

4.11. Исключены записи с некорректными начальными – итоговыми ценами.

В датасетах присутствовали записи с явно некорректными ценами, когда начальная цена лота была выше цены реализации. Такие записи были исключены из анализа.

В датасете с арендой осталось 8 899 записей, в датасете с продажами – 2 419.

4.12. Исключены малочисленные записи по категориям.

Из датасетов были исключены записи с редкими категориями разрешенного использования (обеспечение обороны и безопасности, деятельность по особой охране и

изучению природы), земельные участки особо охраняемых территорий и специального назначения, земельные участки с пометкой «не образованы».

После данных процедур исключения малочисленных категорий в датасете с арендой осталось 8 447 записей, в датасете с продажами – 2 225 записей.

Преобразование и создание признаков

Ряд числовых признаков (площадь земельного участка, начальная цена участка на торги, итоговая цена реализации земельного участка) в датасете имели очень большой разброс значений и распределение данных признаков совсем не походило на нормальное (см. матрицы диаграмм рассеивания ниже).

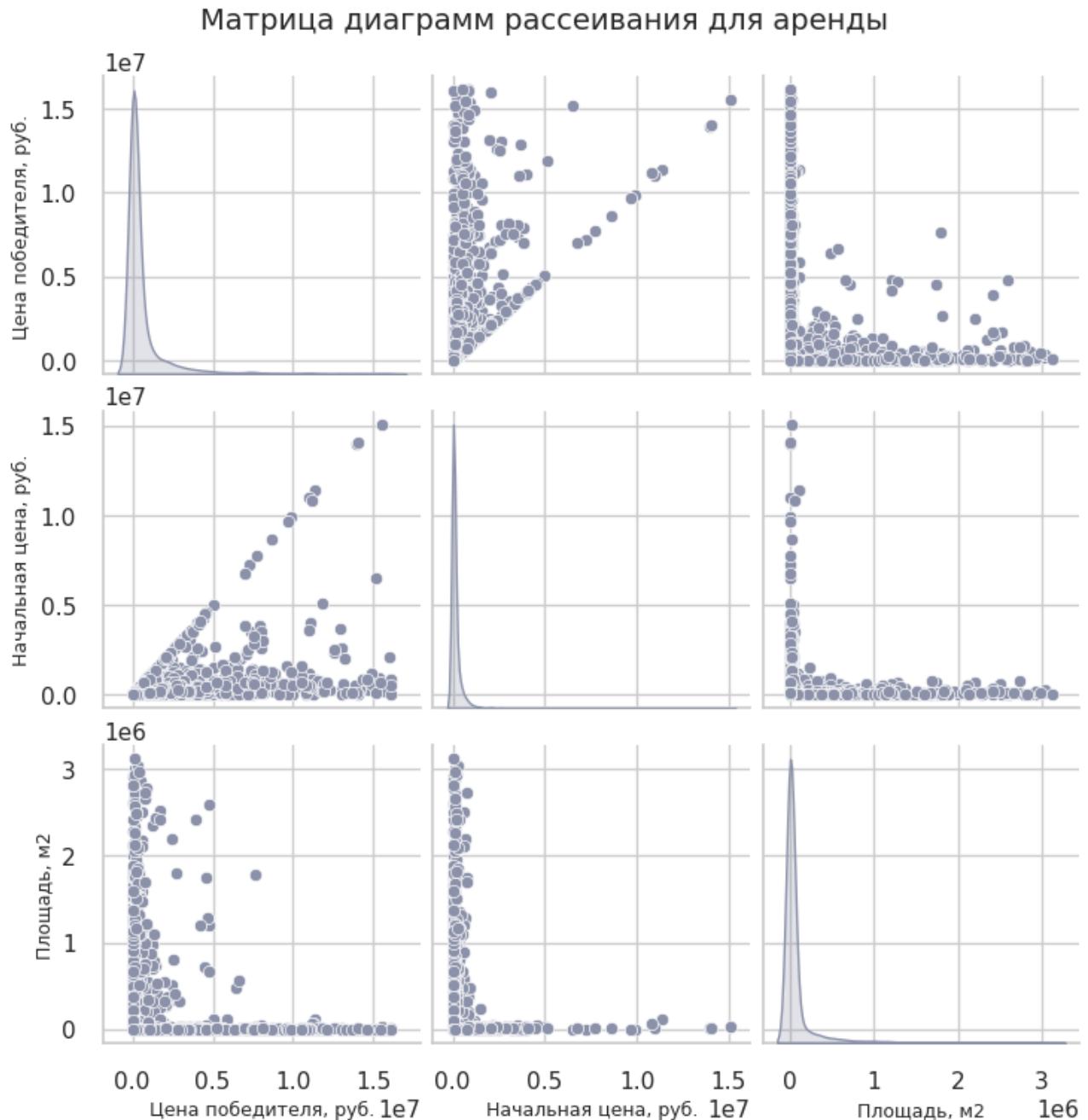


Рис. 1. Матрица диаграмм рассеивания итоговой цены, начальной цены и площади для датасета с арендой земельных участков.

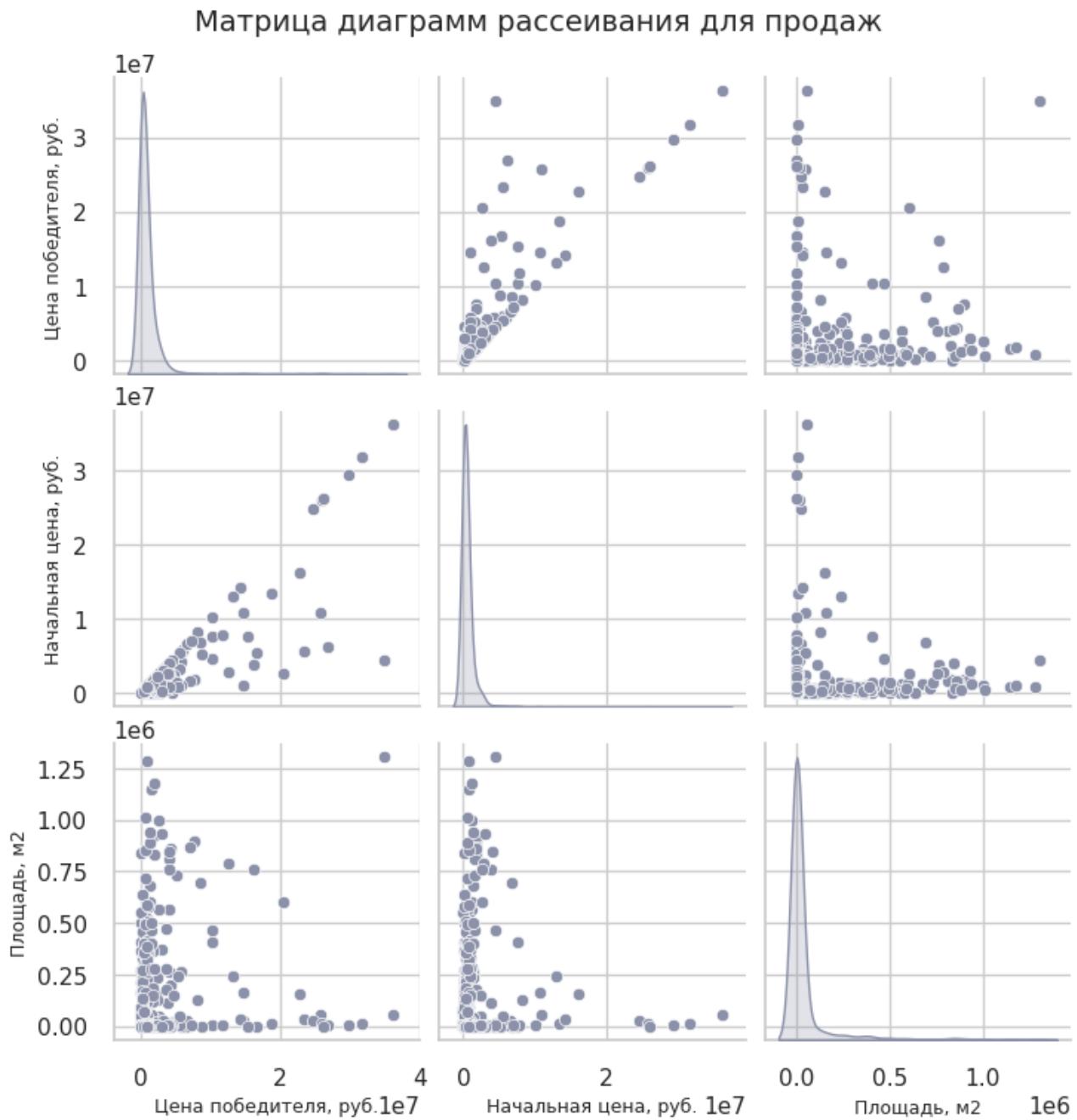


Рис. 2. Матрица диаграмм рассеивания итоговой цены, начальной цены и площади для датасета с продажами земельных участков.

Для целей дальнейшего использования в исследовании показатели были преобразованы, а именно был взят десятичный логарифм от показателей.

В результате приведенные выше матрицы диаграмм рассеивания стали следующими:

Матрица диаграмм рассеивания для аренды

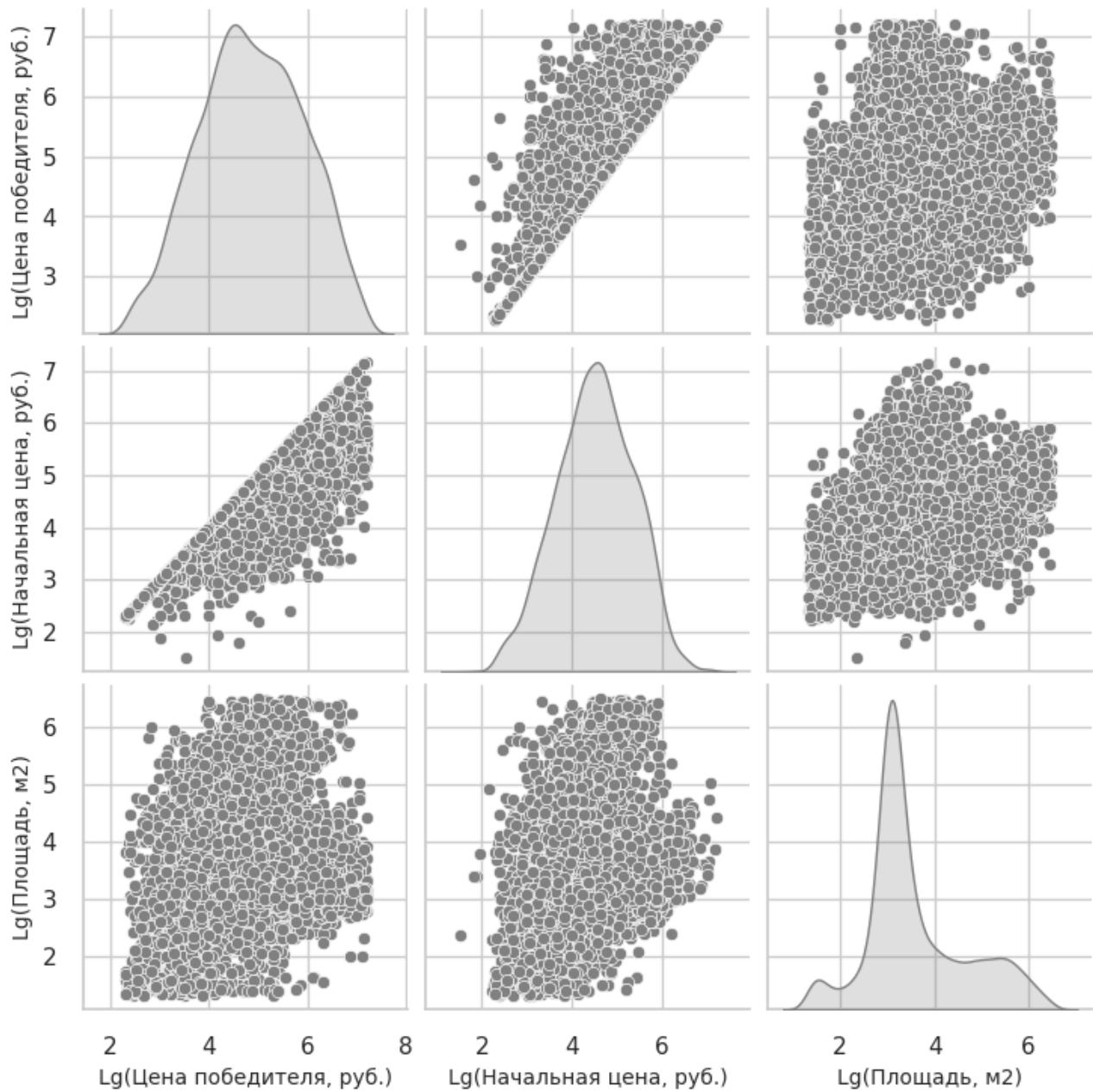


Рис. 3. Матрица диаграмм рассеивания десятичных логарифмов от итоговой цены, начальной цены и площади для датасета с арендой земельных участков.

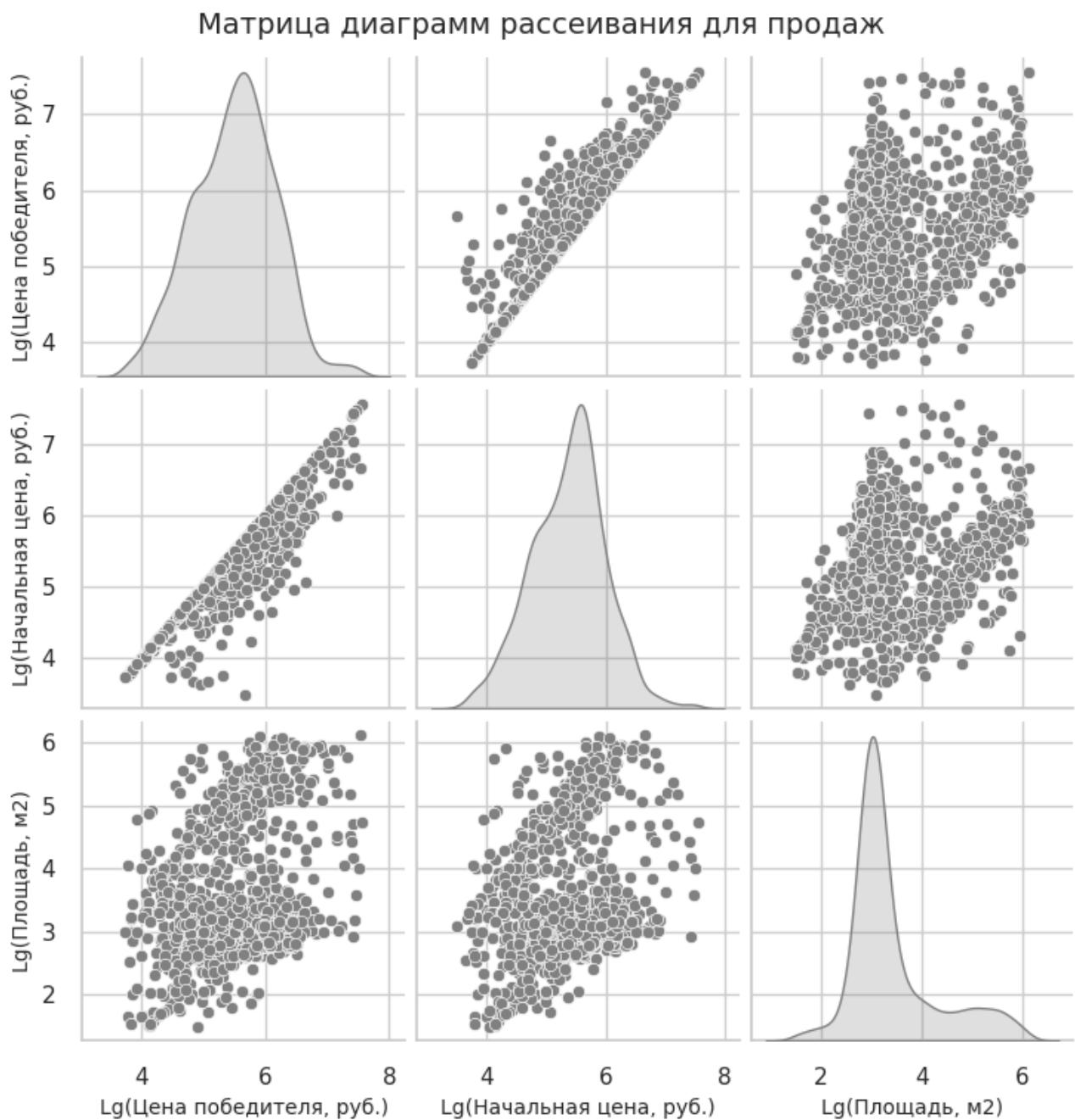


Рис. 4. Матрица диаграмм рассеивания десятичных логарифмов от итоговой цены, начальной цены и площади для датасета с продажами земельных участков.

На основе категориальных признаков: вида разрешенного использования, назначения земельного участка, ограничений на земельный участок были созданы дамми-переменные (1 – если данный признак присутствует у земельного участка, 0 если признак не присутствует у данного участка).

Были созданы дамми-переменные по числу участников торгов:

- были - 1 или нет - 0 участники в торгах,
- был ровно один участник - 1, иначе - 0,
- было ровно два участника - 1, иначе - 0,
- было более двух участников -1, иначе – 0.

Создана дамми-переменная проводились ли торги на электронных торговых площадках – 1, или торги не через электронные площадки.

Создан дополнительный признак – доля депозита для участия от начальной цены лота.

Формирование целевой и контрольной групп для исследований

К данному этапу исследования в датасетах осталось (число записей):

Кто реализует	Датасет с арендой	Датасет с продажами
Российская Федерация (через Росимущество или ДОМ.РФ)	85	50
Субъект Российской Федерации	2 437	284
Муниципалитеты	5 925	1 891
Всего записей	8 447	2 225

К целевой группе относятся участки, которые реализует Российской Федерации.

Контрольные группы нужно составить из участков, которые реализуют Субъекты Российской Федерации и муниципалитеты.

Число земельных участков, которые реализует Российской Федерации, значительно меньше числа земельных участков, которые реализуют Субъекты Российской Федерации, муниципалитеты.

Формирование контрольных групп осуществлялось на основании имеющихся качественных характеристик земельных участков:

- вида разрешенного использования:
 - сельскохозяйственное использование,
 - жилая застройка,
 - общественное использование объектов капитального строительства,
 - предпринимательство,
 - производственная деятельность,
 - транспорт,
 - земельные участки (территории) общего пользования,
 - земельные участки общего назначения,
- земли населенных пунктов, земли сельскохозяйственные,
- имеются или нет ограничения для земельного участка,
- экономического района нахождения земельного участка³,

а также количественной характеристики площади (логарифма площади земельных участков).

Отбор контрольной группы шел из принципа, чтобы были наименьшие различия в качественных признаках и как можно меньшие различия в площади.

Для датасета с арендой были отобраны муниципальные участки без различий по качественным характеристикам и с не более 50% различием по площади⁴ и участки Субъектов Российской Федерации с не более чем одним различием в качественных характеристиках и с не более 50% различием по площади, то есть для каждого участка, реализованного Российской

³ В соответствии с Общероссийским классификатором экономических регионов, раздел 2. Экономические районы (утвержден постановлением Госстандарта Российской Федерации от 27.12.1995 № 640 с последующими изменениями).

⁴ Здесь и в датасете с продажами параметры подобраны экспериментально, варьируя разные ограничения и оценивая сколько земельных участков попадает в контрольную группу, для скольких земельных участков, реализованных Российской Федерацией не находится ни одного участка – «соседа» из субъектов Российской Федерации, муниципалитетов с такими ограничениями.

Федерации, искались участки, реализованные Субъектами Российской Федерации и муниципалитетами, с не более чем одним (для Субъектов Российской Федерации) и точно такими же (для муниципалитетов) качественными характеристиками как у участка Российской Федерации и площадь которых не более чем на 50% меньше площади участка Российской Федерации и не более чем на 50% больше площади участка Российской Федерации.

В итоге для датасета с арендой получилось из 85 участков Российской Федерации для 29 земельных участков не удалось найти «соседа» из участков Субъектов Российской Федерации, для 25 участков не удалось найти «соседа» из участков муниципалитетов.

Земельные участки, реализованные Российской Федерацией, для которых не был найден ни один «сосед» (ни среди Субъектов Российской Федерации, ни среди муниципалитетов) были исключены из дальнейшего анализа.

Всего осталось 65 земельных участка, реализованных Российской Федерации, а к ним 340 участков Субъектов Российской Федерации и 352 участка муниципалитетов. Необходимо отметить, что с учетом размах ограничений для одних земельных участков Российской Федерации был найден всего 1-2 «сосед», для каких-то «соседей» было более 20. Для того чтобы сократить разброс соседей по количеству и оставить наиболее «близкие» по характеристикам участки к участкам Российской Федерации были составлены урезанные датасеты:

- с не более пятью ближайшими «соседями» для каждого земельного участка Российской Федерации (осталось 132 участков Субъектов Российской Федерации и 185 участка муниципалитетов),

- с не более тремя ближайшими «соседями» для каждого земельного участка Российской Федерации (осталось 98 участков Субъектов Российской Федерации и 133 участка муниципалитетов),

- только с одним ближайшим «соседом» для каждого земельного участка Российской Федерации (осталось 42 участка Субъектов Российской Федерации и 55 участка муниципалитетов).

Для датасета с продажами отбор контрольной группы осуществлялся по аналогичному принципу, но для продаж характеристики участков, которые реализует Российская Федерация, более существенно (чем в случае аренды) отличаются от характеристик участков, которые реализуют Субъекты Российской Федерации и муниципалитеты. В результате для датасета с продажей были отобраны муниципальные участки с не более двумя различиями по качественным характеристикам и с не более 50% различия по площади и участки Субъектов Российской Федерации с не более тремя различиями по качественным характеристикам и с не более 50% различия по площади.

Из 50 исходных участков Российской Федерации для 9 не нашлось ни одного «соседа» ни среди участков Субъектов Российской Федерации, ни для участков муниципалитетов.

Всего для датасета с продажами осталось 41 земельных участка, реализованных Российской Федерации, а к ним 259 участков Субъектов Российской Федерации и 1 102 участков муниципалитетов:

- с не более пятью ближайшими «соседями» для каждого земельного участка Российской Федерации осталось 80 участок Субъектов Российской Федерации и 126 участков муниципалитетов,

- с не более тремя ближайшими «соседями» для каждого земельного участка Российской Федерации осталось 51 участка Субъектов Российской Федерации и 81 участка муниципалитетов,

- только с одним ближайшим «соседом» для каждого земельного участка Российской Федерации осталось 17 участков Субъектов Российской Федерации и 29 участка муниципалитетов.

Детальнее процедуру отбора схожих участков можно посмотреть в прилагаемых скриптах (ZK_rent_knn_find_ver4.ipynb, ZK_sale_knn_find_ver3.ipynb).

Оценка доли торгов без участников

Часть торгов оканчиваются совсем без участников, посмотрим насколько отличаются доли торгов без участников.

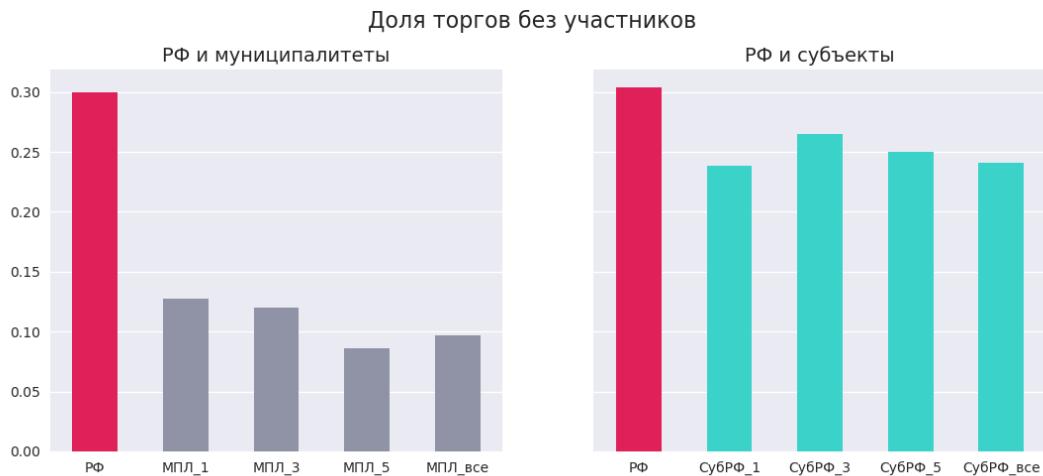


Рис. 5. Доли торгов по аренде земельных участков без участников.

На диаграмме обозначения MPL_1, MPL_3, MPL_5, MPL_all соответствуют датасетам муниципальных земельных участков с самым близким «соседом», с 3-мя ближайшими «соседями», с 5-ю ближайшими соседями, по всей контрольной группе муниципальных участков. Обозначения SubRF_1, SubRF_3, SubRF_5, SubRF_all, соответствуют датасетам земельных участков Субъектов Российской Федерации с самым близким «соседом», с 3-мя ближайшими «соседями», с 5-ю ближайшими соседями, по всей контрольной группе земельных участков Российской Федерации.

Визуально видно, что доля торгов без участников для Российской Федерации доля торгов без участников выше, чем для остальных. Был проведен статистический тест t-критерий Уэлча о равенстве средних, имеющих необязательно равные дисперсии. Результаты приведены в таблице:

	a	b	a_mean	b_mean	statistic	pvalue
0	MPL_1	RF	0.127	0.300	-2.305	0.023
1	MPL_3	RF	0.120	0.300	-2.721	0.008
2	MPL_5	RF	0.086	0.300	-3.381	0.001
3	MPL_all	RF	0.097	0.300	-3.296	0.002
4	SubRF_1	RF	0.238	0.304	-0.720	0.473

5	SubRF_3	RF	0.265	0.304	-0.500	0.618
6	SubRF_5	RF	0.250	0.304	-0.738	0.463
7	SubRF_all	RF	0.241	0.304	-0.942	0.349

Из таблицы видно, что для всех датасетов с муниципальными участками есть статистически значимое на уровне 95% вероятности различие в доле торгов без участников.

Ситуация с продажами по данной части следующая:

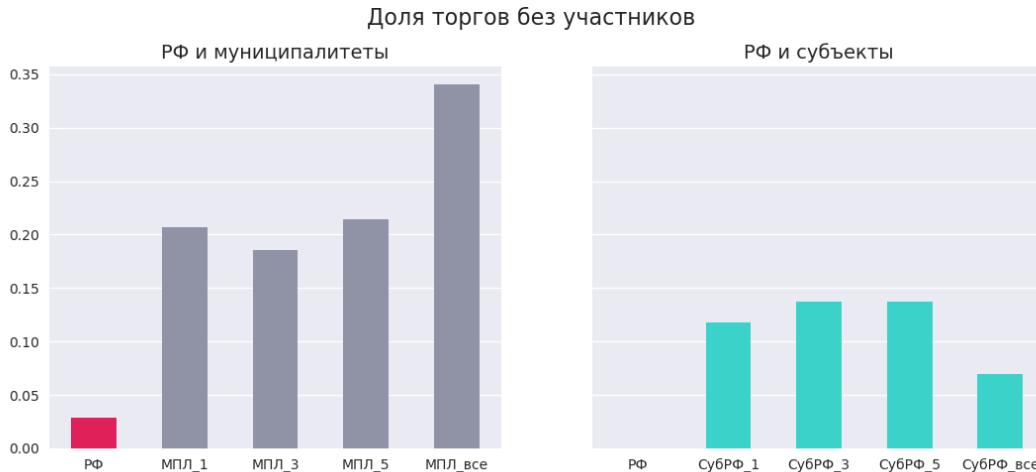


Рис. 6. Доли торгов по продаже земельных участков без участников.

	a	b	a_mean	b_mean	statistic	pvalue
0	MPL_1	RF	0.207	0.029	2.182	0.036
1	MPL_3	RF	0.185	0.029	3.013	0.003
2	MPL_5	RF	0.214	0.029	3.993	0.000
3	MPL_all	RF	0.340	0.029	9.759	0.000

Из таблицы также видно, что для всех датасетов с муниципальными участками есть статистически значимое на уровне 95% вероятности различие в доле торгов без участников. При этом здесь для продаж иная тенденция, чем для аренды. Доля торгов по продажам земельных участков для Российской Федерации, где не приходят участники, существенно ниже, чем аналогичная доля для продаж муниципалитетами.

Исследование различий в определении начальной цены лотов

Количественные признаки (независимые переменные и целевая переменная) были проанализированы на взаимную корреляцию, результаты представлены на тепловых картах корреляций ниже.

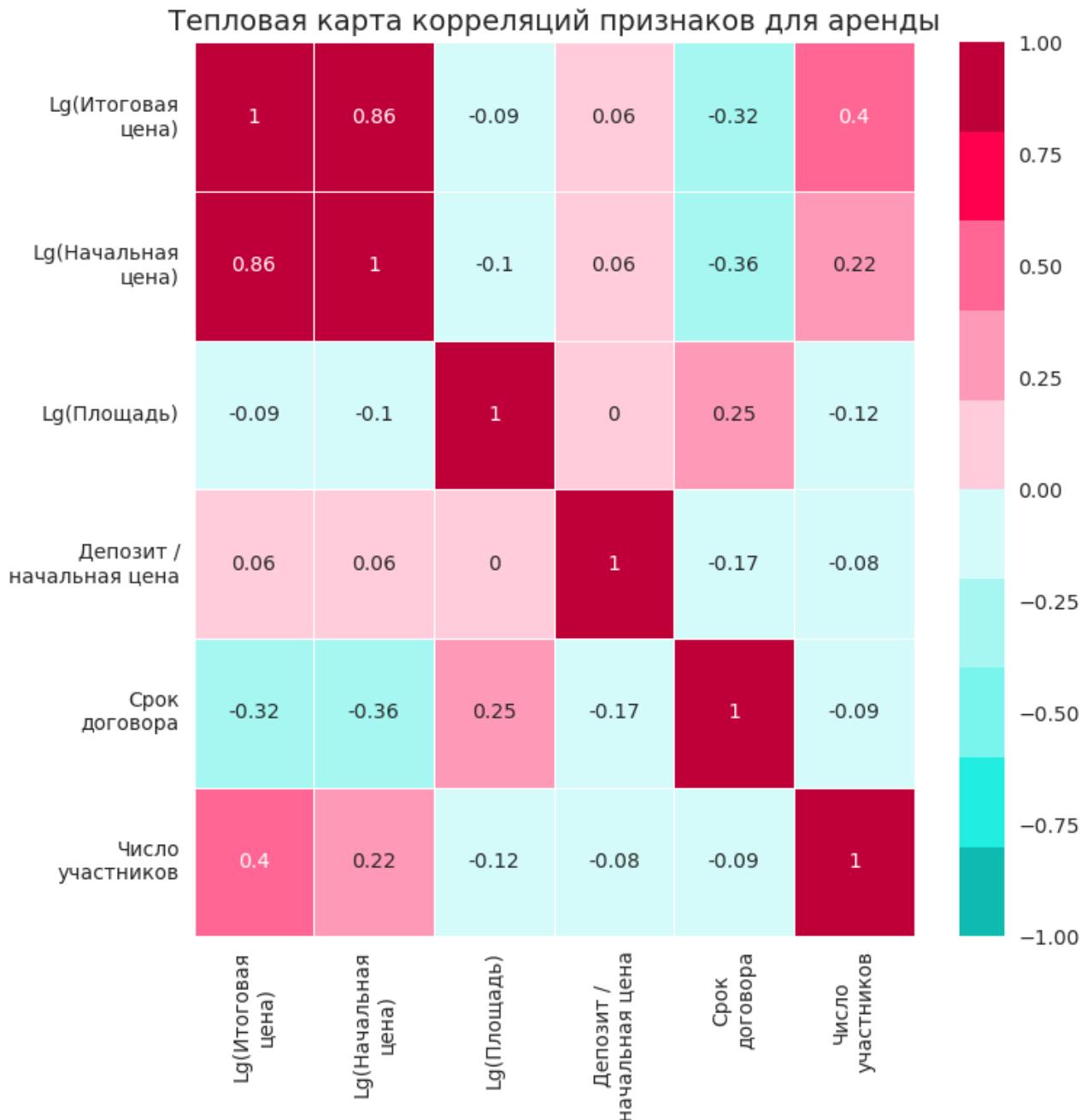


Рис. 7. Термометрическая карта матрицы корреляций числовых признаков датасета с арендой земельных участков



Рис. 8. Термометрическая карта матрицы корреляций числовых признаков датасета с продажей земельных участков

Видно, что как для аренды, так и для продажи ключевым для итоговой цены является начальная цена лота.

Так как начальная цена наиболее сильно положительно коррелирует с итоговой ценой, то в дальнейшем исследовании, чтобы понять влияние интересующих нас факторов, а в частности, кто реализует земельный участок, мы в качестве целевой переменной будем брать не итоговую цену (логарифм), а разность (логарифмов) между итоговой ценой и начальной ценой, таким образом, мы сможем получить оценку того приводит ли и если да, то во сколько

раз изменение независимой переменной к изменению отношения итоговой цены к начальной цене.

Так как начальная цена наиболее важна, то сначала посмотрим, есть ли различия при установке начальной цены.

Оценка осуществлялась линейной регрессией описываемой следующей формулой, для датасета с арендой:

$\text{lg_lotPriceMin} \sim \text{Saler_SubjectRF} + \text{Saler_Municipal} + \text{Lg_SquareZU} +$
 'Ограничение_прав_на_ЗУ_Имеются' + '02. Жилая застройка' + '04. Предпринимательство' +
 '06. Производственная деятельность' + 'ЗУ_Населенных пунктов' + 'Срок договора лет' +
 'Economic_Region_Северо_Кавказский_район' + 'Economic_Region_Уральский_район' +
 'Economic_Region_Центральный_район' + 'Economic_Region_Поволжский_район' +
 'Economic_Region_Западно_Сибирский_район' + 'Economic_Region_Дальневосточный_район'
 + 'Economic_Region_Восточно_Сибирский_район' + 'Economic_Region_Северный_район'.

Для датасета с продажами:

$\text{lg_lotPriceMin} \sim \text{Saler_SubjectRF} + \text{Saler_Municipal} + \text{Lg_SquareZU} +$
 'Ограничение_прав_на_ЗУ_Имеются' + '02. Жилая застройка' + '04. Предпринимательство' +
 '06. Производственная деятельность' + 'Economic_Region_Московская_область' +
 'Economic_Region_Северо_Кавказский_район' + 'Economic_Region_Уральский_район' +
 'Economic_Region_Центральный_район' + 'Economic_Region_Поволжский_район' +
 'Economic_Region_Западно_Сибирский_район' +
 'Economic_Region_Центрально_Черноземный_район' +
 'Economic_Region_Дальневосточный_район' +
 'Economic_Region_Восточно_Сибирский_район' + 'Economic_Region_Северный_район'.

где:

lg_lotPriceMin – десятичный логарифм начальной цены,

Saler_SubjectRF – дамми-переменная, равная 1, если реализует Субъект Российской Федерации и 0 иначе,

Saler_Municipal – дамми-переменная, равная 1, если реализует муниципалитет и 0 иначе,

'Ограничение_прав_на_ЗУ_Имеются' – дамми-переменная, равная 1, если есть информация об ограничениях на земельный участок, 0 – иначе,

'02. Жилая застройка', '04. Предпринимательство', '06. Производственная деятельность' – дамми-переменные, равные 1, если у участка есть соответствующий вид разрешенного использования, 0 - иначе⁵,

'ЗУ_Населенных пунктов' – дамми-переменная, равная 1, если земельный участок населенного пункта, 0 – если сельское хозяйство⁶,

⁵ Одним из распространённых видов разрешенного использования является '01. Сельскохозяйственное использование', но вместе с указанными выше видами разрешенного использования, почти во всех случаях получалось, что один из этих признаков равен 1, остальные 0. Чтобы избежать линейной зависимости между данными 4 признаками признак '01. Сельскохозяйственное использование' исключался из регрессии, по сути все 0 в признаках '02. Жилая застройка', '04. Предпринимательство', '06. Производственная деятельность' для датасетов в исследовании означало, что вид разрешенного использования '01. Сельскохозяйственное использование'.

⁶ В регрессии для продаж не использовался, так как в данном датасете почти все участки были с признаком земель населенных пунктов.

'Срок договора лет' – число лет, на сколько заключается договор аренды⁷,

'Economic_Region...' – дамми-переменная, равная 1, если земельный участок расположен в данном экономическом районе, 0 – иначе.

Результаты регрессий следующие:

OLS_rent_start_price_res_all							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.527				
Model:	OLS	Adj. R-squared:	0.516				
Method:	Least Squares	F-statistic:	47.91				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	6.63e-107				
Time:	12:00:32	Log-Likelihood:	-645.66				
No. Observations:	757	AIC:	1327.				
Df Residuals:	739	BIC:	1411.				
Df Model:	17						
Covariance Type:	HC3						
	coef	std err	z	P> z	[0.025	0.975]	
Intercept	1.3864	0.402	3.447	0.001	0.598	2.175	
Saler_SubjectRF	-0.3423	0.085	-4.021	0.000	-0.509	-0.175	
Saler_Municipal	-0.2140	0.081	-2.634	0.008	-0.373	-0.055	
Lg_SquareZU	0.7325	0.049	15.091	0.000	0.637	0.828	
Q('Ограничение_прав_на_ЗУ_Имеются')	-0.0567	0.072	-0.783	0.433	-0.198	0.085	
Q('02. Жилая застройка')	0.2920	0.272	1.072	0.284	-0.242	0.826	
Q('04. Предпринимательство')	0.6785	0.265	2.559	0.011	0.159	1.198	
Q('06. Производственная деятельность')	0.1937	0.260	0.746	0.456	-0.315	0.703	
Q('ЗУ_Населенных_пунктов')	1.4894	0.271	5.502	0.000	0.959	2.020	
Q('Срок договора лет')	-0.0098	0.001	-7.383	0.000	-0.012	-0.007	
Q('Economic_Region_Северо_Кавказский_район')	-0.4339	0.319	-1.359	0.174	-1.059	0.192	
Q('Economic_Region_Уральский_район')	-0.5882	0.319	-1.845	0.065	-1.213	0.037	
Q('Economic_Region_Центральный_район')	-0.6058	0.319	-1.899	0.058	-1.231	0.019	
Q('Economic_Region_Поволжский_район')	-0.5694	0.328	-1.738	0.082	-1.211	0.073	
Q('Economic_Region_Западно_Сибирский_район')	-0.3987	0.323	-1.233	0.218	-1.033	0.235	
Q('Economic_Region_Дальневосточный_район')	-0.5024	0.323	-1.556	0.120	-1.135	0.131	
Q('Economic_Region_Восточно_Сибирский_район')	-0.5070	0.342	-1.480	0.139	-1.178	0.164	
Q('Economic_Region_Северный_район')	-0.6557	0.298	-2.198	0.028	-1.240	-0.071	
Omnibus:	5.952	Durbin-Watson:	2.021				
Prob(Omnibus):	0.051	Jarque-Bera (JB):	5.658				
Skew:	0.170	Prob(JB):	0.0591				
Kurtosis:	2.748	Cond. No.	727.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 9. Аренда на датасете со всеми «соседями»

⁷ В регрессии с продажами не использовался, так как продажи без срока.

OLS_rent_start_price_res_5							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.453				
Model:	OLS	Adj. R-squared:	0.427				
Method:	Least Squares	F-statistic:	15.48				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	1.17e-33				
Time:	12:02:21	Log-Likelihood:	-364.24				
No. Observations:	382	AIC:	764.5				
Df Residuals:	364	BIC:	835.5				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	1.5070	0.655	2.301	0.021	0.224	2.790	
Saler_SubjectRF	-0.2334	0.099	-2.347	0.019	-0.428	-0.038	
Saler_Municipal	-0.2550	0.089	-2.875	0.004	-0.429	-0.081	
Lg_SquareZU	0.7030	0.081	8.696	0.000	0.545	0.861	
Q('Ограничение_прав_на_ЗУ_Имеются')	0.0316	0.104	0.304	0.761	-0.172	0.235	
Q('02. Жилая застройка')	0.3131	0.271	1.154	0.249	-0.219	0.845	
Q('04. Предпринимательство')	0.7077	0.272	2.599	0.009	0.174	1.242	
Q('06. Производственная деятельность')	0.2735	0.271	1.009	0.313	-0.258	0.805	
Q('ЗУ_Населенных_пунктов')	1.3850	0.313	4.430	0.000	0.772	1.998	
Q('Срок_договора_лет')	-0.0089	0.003	-3.218	0.001	-0.014	-0.003	
Q('Economic_Region_Северо_Кавказский_район')	-0.4362	0.526	-0.829	0.407	-1.467	0.595	
Q('Economic_Region_Уральский_район')	-0.6390	0.520	-1.229	0.219	-1.658	0.380	
Q('Economic_Region_Центральный_район')	-0.5187	0.524	-0.989	0.322	-1.546	0.509	
Q('Economic_Region_Поволжский_район')	-0.6367	0.534	-1.191	0.234	-1.684	0.411	
Q('Economic_Region_Западно_Сибирский_район')	-0.4126	0.521	-0.792	0.428	-1.434	0.609	
Q('Economic_Region_Дальневосточный_район')	-0.3506	0.515	-0.681	0.496	-1.360	0.658	
Q('Economic_Region_Восточно_Сибирский_район')	-0.4997	0.535	-0.935	0.350	-1.548	0.548	
Q('Economic_Region_Северный_район')	-0.6418	0.506	-1.267	0.205	-1.634	0.351	
Omnibus:	7.311	Durbin-Watson:	1.865				
Prob(Omnibus):	0.026	Jarque-Bera (JB):	6.259				
Skew:	0.237	Prob(JB):	0.0437				
Kurtosis:	2.589	Cond. No.	566.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 10. Аренда на датасете с 5-ю ближайшими «соседями»

OLS_rent_start_price_res_3							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.469				
Model:	OLS	Adj. R-squared:	0.437				
Method:	Least Squares	F-statistic:	11.60				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	6.71e-24				
Time:	12:02:59	Log-Likelihood:	-273.74				
No. Observations:	296	AIC:	583.5				
Df Residuals:	278	BIC:	649.9				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	2.0047	0.846	2.370	0.018	0.347	3.663	
Saler_SubjectRF	-0.2308	0.107	-2.154	0.031	-0.441	-0.021	
Saler_Municipal	-0.2324	0.094	-2.485	0.013	-0.416	-0.049	
Lg_SquareZU	0.6487	0.085	7.655	0.000	0.483	0.815	
Q('Ограничение_прав_на_ЗУ_Имеются')	0.0964	0.104	0.925	0.355	-0.108	0.301	
Q('02. Жилая застройка')	0.4168	0.280	1.487	0.137	-0.133	0.966	
Q('04. Предпринимательство')	0.8663	0.261	3.317	0.001	0.354	1.378	
Q('06. Производственная деятельность')	0.4631	0.249	1.860	0.063	-0.025	0.951	
Q('ЗУ_Населенных_пунктов')	1.1141	0.287	3.880	0.000	0.551	1.677	
Q('Срок_договора_лет')	-0.0129	0.004	-3.468	0.001	-0.020	-0.006	
Q('Economic_Region_Северо_Кавказский_район')	-0.5668	0.749	-0.757	0.449	-2.035	0.901	
Q('Economic_Region_Уральский_район')	-0.8345	0.737	-1.132	0.258	-2.279	0.610	
Q('Economic_Region_Центральный_район')	-0.6926	0.744	-0.931	0.352	-2.150	0.765	
Q('Economic_Region_Поволжский_район')	-0.8872	0.750	-1.183	0.237	-2.357	0.583	
Q('Economic_Region_Западно_Сибирский_район')	-0.5782	0.738	-0.784	0.433	-2.024	0.868	
Q('Economic_Region_Дальневосточный_район')	-0.2915	0.711	-0.410	0.682	-1.686	1.103	
Q('Economic_Region_Восточно_Сибирский_район')	-0.8103	0.748	-1.083	0.279	-2.277	0.657	
Q('Economic_Region_Северный_район')	-0.7950	0.723	-1.099	0.272	-2.212	0.622	
Omnibus:	7.677	Durbin-Watson:	1.984				
Prob(Omnibus):	0.022	Jarque-Bera (JB):	7.308				
Skew:	0.336	Prob(JB):	0.0259				
Kurtosis:	2.625	Cond. No.	561.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 11. Аренда на датасете с 3-мя ближайшими соседями»

OLS_rent_start_price_res_1							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.471				
Model:	OLS	Adj. R-squared:	0.409				
Method:	Least Squares	F-statistic:	6.111				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	1.65e-10				
Time:	12:04:21	Log-Likelihood:	-146.59				
No. Observations:	162	AIC:	329.2				
Df Residuals:	144	BIC:	384.8				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	2.5676	1.771	1.449	0.147	-0.905	6.040	
Saler_SubjectRF	-0.2666	0.149	-1.788	0.074	-0.559	0.026	
Saler_Municipal	-0.2853	0.125	-2.279	0.023	-0.531	-0.040	
Lg_SquareZU	0.6449	0.118	5.485	0.000	0.414	0.875	
Q('Ограничение_прав_на_ЗУ_Имеются')	0.0828	0.148	0.561	0.575	-0.206	0.372	
Q('02. Жилая застройка')	0.0697	0.410	0.170	0.865	-0.734	0.873	
Q('04. Предпринимательство')	0.6370	0.405	1.574	0.115	-0.156	1.430	
Q('06. Производственная деятельность')	0.0054	0.362	0.015	0.988	-0.704	0.715	
Q('ЗУ_Населенных_пунктов')	1.4251	0.420	3.391	0.001	0.601	2.249	
Q('Срок_договора_лет')	-0.0135	0.005	-2.659	0.008	-0.024	-0.004	
Q('Economic_Region_Северо_Кавказский_район')	-0.9969	1.676	-0.595	0.552	-4.281	2.288	
Q('Economic_Region_Уральский_район')	-1.3366	1.661	-0.805	0.421	-4.592	1.919	
Q('Economic_Region_Центральный_район')	-1.3778	1.664	-0.828	0.408	-4.638	1.883	
Q('Economic_Region_Поволжский_район')	-1.3318	1.676	-0.795	0.427	-4.616	1.953	
Q('Economic_Region_Западно_Сибирский_район')	-1.1894	1.661	-0.716	0.474	-4.446	2.067	
Q('Economic_Region_Дальневосточный_район')	-0.7647	1.642	-0.466	0.641	-3.983	2.454	
Q('Economic_Region_Восточно_Сибирский_район')	-1.2420	1.668	-0.745	0.456	-4.511	2.027	
Q('Economic_Region_Северный_район')	-1.3204	1.651	-0.800	0.424	-4.557	1.916	
Omnibus:	1.171	Durbin-Watson:	1.868				
Prob(Omnibus):	0.557	Jarque-Bera (JB):	1.158				
Skew:	0.089	Prob(JB):	0.561				
Kurtosis:	2.626	Cond. No.	573.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 12. Аренда на датасете с самым ближайшим соседом»

OLS_sale_start_price_res_all							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.336				
Model:	OLS	Adj. R-squared:	0.327				
Method:	Least Squares	F-statistic:	28.01				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	7.90e-77				
Time:	13:12:40	Log-Likelihood:	-953.32				
No. Observations:	1402	AIC:	1943.				
Df Residuals:	1384	BIC:	2037.				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	4.7957	0.318	15.103	0.000	4.173	5.418	
Saler_SubjectRF	-0.9916	0.145	-6.837	0.000	-1.276	-0.707	
Saler_Municipal	-1.1196	0.137	-8.164	0.000	-1.388	-0.851	
Lg_SquareZU	0.4410	0.063	7.018	0.000	0.318	0.564	
Q('Ограничение_прав_на_ЗУ_Имеются')	-0.1214	0.060	-2.012	0.044	-0.240	-0.003	
Q('02. Жилая застройка')	0.6366	0.110	5.796	0.000	0.421	0.852	
Q('04. Предпринимательство')	0.5226	0.168	3.116	0.002	0.194	0.851	
Q('06. Производственная деятельность')	0.3474	0.167	2.084	0.037	0.021	0.674	
Q('Economic_Region_Московская_область')	-0.0330	0.080	-0.412	0.680	-0.190	0.124	
Q('Economic_Region_Северо_Кавказский_район')	-0.3700	0.070	-5.314	0.000	-0.506	-0.234	
Q('Economic_Region_Уральский_район')	-0.2869	0.058	-4.945	0.000	-0.401	-0.173	
Q('Economic_Region_Центральный_район')	-0.0492	0.048	-1.031	0.303	-0.143	0.044	
Q('Economic_Region_Поволжский_район')	-0.1728	0.066	-2.600	0.009	-0.303	-0.043	
Q('Economic_Region_Западно_Сибирский_район')	0.0821	0.084	0.979	0.327	-0.082	0.246	
Q('Economic_Region_Центрально_Черноземный_район')	-0.2230	0.073	-3.037	0.002	-0.367	-0.079	
Q('Economic_Region_Дальневосточный_район')	-0.1847	0.094	-1.970	0.049	-0.368	-0.001	
Q('Economic_Region_Восточно_Сибирский_район')	-1.1483	0.118	-9.754	0.000	-1.379	-0.918	
Q('Economic_Region_Северный_район')	-0.0274	0.093	-0.294	0.769	-0.210	0.155	
Omnibus:	41.558	Durbin-Watson:	1.823				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	72.098				
Skew:	-0.232	Prob(JB):	2.21e-16				
Kurtosis:	4.009	Cond. No.	72.2				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 13. Продажи на датасете со всеми «соседями»

OLS_sale_start_price_res_5							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.562				
Model:	OLS	Adj. R-squared:	0.529				
Method:	Least Squares	F-statistic:	18.45				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	5.61e-34				
Time:	13:13:03	Log-Likelihood:	-208.53				
No. Observations:	247	AIC:	453.1				
Df Residuals:	229	BIC:	516.2				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	5.6427	0.602	9.378	0.000	4.463	6.822	
Saler_SubjectRF	-0.9286	0.160	-5.809	0.000	-1.242	-0.615	
Saler_Municipal	-1.2764	0.140	-9.143	0.000	-1.550	-1.003	
Lg_SquareZU	0.3867	0.093	4.161	0.000	0.205	0.569	
Q('Ограничение_прав_на_ЗУ_Имеются')	0.6345	0.600	1.057	0.290	-0.542	1.811	
Q('02. Жилая застройка')	0.4618	0.164	2.809	0.005	0.140	0.784	
Q('04. Предпринимательство')	0.4619	0.184	2.516	0.012	0.102	0.822	
Q('06. Производственная деятельность')	0.0215	0.257	0.084	0.933	-0.482	0.525	
Q('Economic_Region_Московская_область')	-0.5507	0.341	-1.614	0.107	-1.220	0.118	
Q('Economic_Region_Северо_Кавказский_район')	-0.7610	0.451	-1.686	0.092	-1.646	0.124	
Q('Economic_Region_Уральский_район')	-0.7721	0.352	-2.192	0.028	-1.462	-0.082	
Q('Economic_Region_Центральный_район')	-0.6339	0.368	-1.722	0.085	-1.355	0.087	
Q('Economic_Region_Поволжский_район')	-0.7703	0.372	-2.070	0.038	-1.500	-0.041	
Q('Economic_Region_Западно_Сибирский_район')	-0.4278	0.385	-1.110	0.267	-1.183	0.328	
Q('Economic_Region_Центрально_Черноземный_район')	-0.8786	0.359	-2.450	0.014	-1.581	-0.176	
Q('Economic_Region_Дальневосточный_район')	-0.6117	0.348	-1.757	0.079	-1.294	0.071	
Q('Economic_Region_Восточно_Сибирский_район')	-1.4616	0.383	-3.816	0.000	-2.212	-0.711	
Q('Economic_Region_Северный_район')	-0.9573	0.376	-2.549	0.011	-1.693	-0.221	
Omnibus:	6.741	Durbin-Watson:	2.018				
Prob(Omnibus):	0.034	Jarque-Bera (JB):	7.793				
Skew:	-0.249	Prob(JB):	0.0203				
Kurtosis:	3.714	Cond. No.	69.5				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 14. Продажи на датасете с 5-ю ближайшими «соседями»

OLS_sale_start_price_res_3							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.641				
Model:	OLS	Adj. R-squared:	0.602				
Method:	Least Squares	F-statistic:	29.69				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	3.87e-40				
Time:	13:13:37	Log-Likelihood:	-136.79				
No. Observations:	173	AIC:	309.6				
Df Residuals:	155	BIC:	366.3				
Df Model:	17						
Covariance Type:	HC3						
coef	std err	z	P> z	[0.025	0.975]		
Intercept	6.1574	0.488	12.620	0.000	5.201	7.114	
Saler_SubjectRF	-0.9305	0.158	-5.901	0.000	-1.240	-0.621	
Saler_Municipal	-1.1774	0.141	-8.377	0.000	-1.453	-0.902	
Lg_SquareZU	0.4177	0.097	4.312	0.000	0.228	0.608	
Q('Ограничение_прав_на_ЗУ_Имеются')	0.3250	1.617	0.201	0.841	-2.844	3.494	
Q('02. Жилая застройка')	0.3002	0.152	1.977	0.048	0.003	0.598	
Q('04. Предпринимательство')	0.4057	0.182	2.224	0.026	0.048	0.763	
Q('06. Производственная деятельность')	-0.0625	0.222	-0.281	0.779	-0.498	0.373	
Q('Economic_Region_Московская_область')	-1.0070	0.116	-8.655	0.000	-1.235	-0.779	
Q('Economic_Region_Северо_Кавказский_район')	-1.1814	0.405	-2.920	0.004	-1.974	-0.388	
Q('Economic_Region_Уральский_район')	-1.3296	0.145	-9.153	0.000	-1.614	-1.045	
Q('Economic_Region_Центральный_район')	-1.1825	0.190	-6.217	0.000	-1.555	-0.810	
Q('Economic_Region_Поволжский_район')	-1.3215	0.190	-6.956	0.000	-1.694	-0.949	
Q('Economic_Region_Западно_Сибирский_район')	-0.9879	0.240	-4.119	0.000	-1.458	-0.518	
Q('Economic_Region_Центрально_Черноземный_район')	-1.4041	0.155	-9.074	0.000	-1.707	-1.101	
Q('Economic_Region_Дальневосточный_район')	-1.1192	0.160	-7.006	0.000	-1.432	-0.806	
Q('Economic_Region_Восточно_Сибирский_район')	-1.8572	0.258	-7.186	0.000	-2.364	-1.351	
Q('Economic_Region_Северный_район')	-1.5313	0.180	-8.522	0.000	-1.883	-1.179	
Omnibus:	0.909	Durbin-Watson:	1.971				
Prob(Omnibus):	0.635	Jarque-Bera (JB):	0.675				
Skew:	-0.147	Prob(JB):	0.713				
Kurtosis:	3.088	Cond. No.	74.6				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 15. Продажи на датасете с 3-мя ближайшими «соседями»

OLS_sale_start_price_res_1							
Dep. Variable:	lg_lotPriceMin	R-squared:	0.776				
Model:	OLS	Adj. R-squared:	0.725 <th></th> <th></th> <th></th> <th></th>				
Method:	Least Squares	F-statistic:	13.23				
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	2.66e-15				
Time:	13:16:20	Log-Likelihood:	-52.016				
No. Observations:	87	AIC:	138.0				
Df Residuals:	70	BIC:	180.0				
Df Model:	16						
Covariance Type:	HC3						
		coef	std err	z	P> z	[0.025	0.975]
Intercept		5.5635	0.683	8.151	0.000	4.226	6.901
Saler_SubjectRF		-1.0023	0.188	-5.320	0.000	-1.372	-0.633
Saler_Municipal		-0.9261	0.158	-5.867	0.000	-1.235	-0.617
Lg_SquareZU		0.5880	0.106	5.568	0.000	0.381	0.795
Q('02. Жилая застройка')		0.0153	0.163	0.094	0.925	-0.303	0.334
Q('04. Предпринимательство')		0.2390	0.167	1.430	0.153	-0.088	0.566
Q('06. Производственная деятельность')		-0.2910	0.307	-0.947	0.343	-0.893	0.311
Q('Economic_Region_Московская_область')		-0.6242	0.452	-1.380	0.168	-1.511	0.262
Q('Economic_Region_Северо_Кавказский_район')		-0.6021	0.544	-1.108	0.268	-1.667	0.463
Q('Economic_Region_Уральский_район')		-1.1226	0.440	-2.550	0.011	-1.985	-0.260
Q('Economic_Region_Центральный_район')		-1.4316	0.482	-2.970	0.003	-2.376	-0.487
Q('Economic_Region_Поволжский_район')		-1.1452	0.510	-2.244	0.025	-2.145	-0.145
Q('Economic_Region_Западно_Сибирский_район')		-1.1014	0.487	-2.263	0.024	-2.056	-0.147
Q('Economic_Region_Центрально_Черноземный_район')		-1.2853	0.514	-2.501	0.012	-2.293	-0.278
Q('Economic_Region_Дальневосточный_район')		-0.7871	0.440	-1.789	0.074	-1.649	0.075
Q('Economic_Region_Восточно_Сибирский_район')		-1.4574	0.488	-2.987	0.003	-2.414	-0.501
Q('Economic_Region_Северный_район')		-1.3410	0.545	-2.463	0.014	-2.408	-0.274
Omnibus:	2.838	Durbin-Watson:	1.922				
Prob(Omnibus):	0.242	Jarque-Bera (JB):	1.900				
Skew:	0.147	Prob(JB):	0.387				
Kurtosis:	2.338	Cond. No.	93.1				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 16. Продажи на датасете с самым ближайшим соседом»

Результаты регрессий показывают:

Начальные цены лотов для земельных участков, реализуемых Российской Федерацией статистически значимо отличаются (дороже) начальных цен участков, которые реализуют Субъекты Российской Федерации, муниципалитеты.

В целом результаты исследования различий в начальной цене согласуется с результатами формирования контрольных групп – Российская Федерация реализует земельные участки, которые сильно выделяются по сравнению с участками, которые реализуют Субъекты Российской Федерации и муниципалитеты. Делать какие-либо сравнения в эффективности продаж земельных участков Российской Федерацией нужно будет с учетом указанной оговорки.

Исследование эффективности реализации земельных участков

Оценка осуществлялась линейной регрессией описываемой следующей формулой, для датасета с арендой:

diff_lg_winner_price_lg_start_price ~ Saler_SubjectRF + Saler_Municipal + Lg_SquareZU + depozit_to_priceMin_ratio + Is_ETP + Two_participants + More_then_two_participants + 'Ограничение_прав_на_ЗУ_Имеются' + '02. Жилая застройка' + '04. Предпринимательство' + '06. Производственная деятельность' + 'ЗУ_Населенных_пунктов' + 'Срок_договора_лет' + 'Economic_Region_Северо_Кавказский_район' + 'Economic_Region_Уральский_район' + 'Economic_Region_Центральный_район' + 'Economic_Region_Поволжский_район' + 'Economic_Region_Западно_Сибирский_район' + 'Economic_Region_Дальневосточный_район' + 'Economic_Region_Восточно_Сибирский_район' + 'Economic_Region_Северный_район'.

Для датасета с продажами:

diff_lg_winner_price_lg_start_price ~ Saler_SubjectRF + Saler_Municipal + Lg_SquareZU
 + depozit_to_priceMin_ratio + Is_ETP + Two_participants + More_then_two_participants+
 'Ограничение_прав_на_ЗУ_Имеются' + '02. Жилая застройка' + '04. Предпринимательство' +
 '06. Производственная деятельность' + 'Economic_Region_Московская_область' +
 'Economic_Region_Северо_Кавказский_район' + 'Economic_Region_Уральский_район' +
 'Economic_Region_Центральный_район' + 'Economic_Region_Поволжский_район' +
 'Economic_Region_Западно_Сибирский_район' +
 'Economic_Region_Центрально_Черноземный_район'
 'Economic_Region_Дальневосточный_район'
 'Economic_Region_Восточно_Сибирский_район' + 'Economic_Region_Северный_район'.

где:

diff_lg_winner_price_lg_start_price – разность десятичных логарифмов итоговой цены победителя и начальной цены лота,

depozit_to_priceMin_ratio – отношение размера депозита для участия в торгах к начальной цене лота,

Is_ETP – дамми-переменная, проводились ли торги на электронной площадке – 1, или нет 0,

Two_participants – дамми-переменная, в торгах участвовало ровно 2 участник – значение 1, 0 – иначе,

More_then_two_participants – дамми-переменная, в торгах участвовало более двух участников – 1, 0 – иначе⁸,

Остальные переменные соответствуют описанию выше в разделе с анализом начальной цены лотов.

⁸ Случай, когда в торгах участвовал только один участник соответствует тому, что обе переменные Two_participants и More_then_two_participants равны 0. Отдельная дамми-переменная на одного участника не вводится, чтобы избежать линейной зависимости между данной переменной и переменными Two_participants и More_then_two_participants.

Результаты регрессий следующие:

OLS_rent_diff_winner_price_participants_all						
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.314			
Model:	OLS	Adj. R-squared:	0.290			
Method:	Least Squares	F-statistic:	12.74			
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	4.79e-36			
Time:	12:06:12	Log-Likelihood:	-337.34			
No. Observations:	623	AIC:	718.7			
Df Residuals:	601	BIC:	816.2			
Df Model:	21					
Covariance Type:	HC3					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	0.1263	0.337	0.375	0.708	-0.534	0.786
Saler_SubjectRF	0.1241	0.077	1.617	0.106	-0.026	0.275
Saler_Municipal	0.0409	0.072	0.565	0.572	-0.101	0.183
Lg_SquareZU	-0.0910	0.053	-1.719	0.086	-0.195	0.013
depozit_to_priceMin_ratio	0.0549	0.045	1.212	0.225	-0.034	0.144
Is_ETP	-0.1525	0.040	-3.809	0.000	-0.231	-0.074
Two_participants	0.4704	0.048	9.803	0.000	0.376	0.565
More_then_two_participants	0.8650	0.110	7.896	0.000	0.650	1.080
Q('Ограничение_прав_на_ЗУ_Имеются')	0.0357	0.054	0.667	0.505	-0.069	0.141
Q('02. Жилая застройка')	-0.3826	0.258	-1.481	0.139	-0.889	0.124
Q('04. Предпринимательство')	-0.5995	0.218	-2.754	0.006	-1.026	-0.173
Q('06. Производственная деятельность')	-0.5840	0.219	-2.669	0.008	-1.013	-0.155
Q('ЗУ_Населенных_пунктов')	0.3929	0.231	1.704	0.088	-0.059	0.845
Q('Срок_договора_лет')	0.0030	0.001	2.235	0.025	0.000	0.006
Q('Economic_Region_Северо_Кавказский_район')	0.3057	0.189	1.614	0.107	-0.066	0.677
Q('Economic_Region_Уральский_район')	0.3232	0.188	1.715	0.086	-0.046	0.693
Q('Economic_Region_Центральный_район')	0.4385	0.195	2.245	0.025	0.056	0.821
Q('Economic_Region_Поволжский_район')	0.3500	0.196	1.789	0.074	-0.034	0.734
Q('Economic_Region_Западно_Сибирский_район')	0.4436	0.191	2.328	0.020	0.070	0.817
Q('Economic_Region_Дальневосточный_район')	0.5959	0.178	3.354	0.001	0.248	0.944
Q('Economic_Region_Восточно_Сибирский_район')	0.2334	0.197	1.188	0.235	-0.152	0.619
Q('Economic_Region_Северный_район')	0.2064	0.168	1.231	0.218	-0.122	0.535
Omnibus:	224.991	Durbin-Watson:	2.028			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	868.084			
Skew:	1.652	Prob(JB):	3.15e-189			
Kurtosis:	7.746	Cond. No.	788.			

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 17. Аренда, итоговый эффект на датасете со всеми «соседями»

OLS_rent_diff_winner_price_participants_5							
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.407	Model:	OLS	Adj. R-squared:	0.364
Method:	Least Squares	F-statistic:	12.90	Date:	Wed, 08 Nov 2023	Prob (F-statistic):	1.04e-30
Time:	12:07:36	Log-Likelihood:	-160.87	No. Observations:	315	AIC:	365.7
Df Residuals:	293	BIC:	448.3	Df Model:	21	Covariance Type:	HC3
	coef	std err	z	P> z	[0.025	0.975]	
Intercept	0.3928	0.471	0.833	0.405	-0.531	1.317	
Saler_SubjectRF	0.0680	0.089	0.763	0.446	-0.107	0.243	
Saler_Municipal	0.0384	0.080	0.480	0.631	-0.118	0.195	
Lg_SquareZU	-0.0921	0.063	-1.468	0.142	-0.215	0.031	
depozit_to_priceMin_ratio	0.0082	0.053	0.155	0.877	-0.096	0.112	
Is_ETP	-0.1839	0.056	-3.307	0.001	-0.293	-0.075	
Two_participants	0.5510	0.066	8.361	0.000	0.422	0.680	
More_then_two_participants	0.7387	0.100	7.381	0.000	0.543	0.935	
Q('Ограничение прав на ЗУ Имеются')	0.0814	0.065	1.249	0.212	-0.046	0.209	
Q('02. Жилая застройка')	-0.3123	0.274	-1.139	0.255	-0.850	0.225	
Q('04. Предпринимательство')	-0.6249	0.224	-2.787	0.005	-1.064	-0.185	
Q('06. Производственная деятельность')	-0.5225	0.227	-2.299	0.022	-0.968	-0.077	
Q('ЗУ Населенных пунктов')	0.2264	0.248	0.912	0.362	-0.260	0.713	
Q('Срок договора лет')	-0.0012	0.002	-0.579	0.563	-0.005	0.003	
Q('Economic_Region_Северо_Кавказский_район')	0.3396	0.298	1.139	0.255	-0.245	0.924	
Q('Economic_Region_Уральский_район')	0.2510	0.295	0.851	0.395	-0.327	0.829	
Q('Economic_Region_Центральный_район')	0.3555	0.304	1.168	0.243	-0.241	0.952	
Q('Economic_Region_Поволжский_район')	0.1475	0.300	0.491	0.623	-0.441	0.736	
Q('Economic_Region_Западно_Сибирский_район')	0.3497	0.296	1.182	0.237	-0.230	0.929	
Q('Economic_Region_Дальневосточный_район')	0.5911	0.284	2.081	0.037	0.034	1.148	
Q('Economic_Region_Восточно_Сибирский_район')	0.1176	0.305	0.386	0.700	-0.480	0.715	
Q('Economic_Region_Северный_район')	0.1412	0.283	0.499	0.617	-0.413	0.695	
Omnibus:	106.904	Durbin-Watson:	2.140				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	330.146				
Skew:	1.523	Prob(JB):	2.04e-72				
Kurtosis:	6.985	Cond. No.	588.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 18. Аренда, итоговый эффект на датасете с 5-ю ближайшими «соседями»

OLS_rent_diff_winner_price_participants_3							
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.416	Model:	OLS	Adj. R-squared:	0.359
Method:	Least Squares	F-statistic:	9.169	Date:	Wed, 08 Nov 2023	Prob (F-statistic):	3.65e-20
Time:	12:09:19	Log-Likelihood:	-131.84	No. Observations:	236	AIC:	307.7
Df Residuals:	214	BIC:	383.9	Df Model:	21	Covariance Type:	HC3
	coef	std err	z	P> z	[0.025	0.975]	
Intercept	0.1064	0.573	0.186	0.853	-1.018	1.230	
Saler_SubjectRF	0.0913	0.101	0.906	0.365	-0.106	0.289	
Saler_Municipal	0.0346	0.082	0.421	0.673	-0.126	0.196	
Lg_SquareZU	-0.0869	0.074	-1.182	0.237	-0.231	0.057	
depozit_to_priceMin_ratio	0.0062	0.058	0.108	0.914	-0.107	0.119	
Is_ETP	-0.2027	0.071	-2.837	0.005	-0.343	-0.063	
Two_participants	0.5532	0.081	6.863	0.000	0.395	0.711	
More_then_two_participants	0.6914	0.108	6.401	0.000	0.480	0.903	
Q('Ограничение прав на ЗУ Имеются')	0.0595	0.072	0.826	0.409	-0.082	0.201	
Q('02. Жилая застройка')	-0.2010	0.356	-0.565	0.572	-0.898	0.496	
Q('04. Предпринимательство')	-0.6965	0.232	-2.998	0.003	-1.152	-0.241	
Q('06. Производственная деятельность')	-0.6293	0.232	-2.708	0.007	-1.085	-0.174	
Q('ЗУ Населенных пунктов')	0.3149	0.263	1.196	0.232	-0.201	0.831	
Q('Срок договора лет')	-0.0008	0.003	-0.243	0.808	-0.007	0.006	
Q('Economic_Region_Северо_Кавказский_район')	0.5620	0.355	1.582	0.114	-0.134	1.258	
Q('Economic_Region_Уральский_район')	0.5632	0.346	1.626	0.104	-0.116	1.242	
Q('Economic_Region_Центральный_район')	0.7576	0.366	2.071	0.038	0.041	1.475	
Q('Economic_Region_Поволжский_район')	0.4175	0.351	1.189	0.234	-0.271	1.106	
Q('Economic_Region_Западно_Сибирский_район')	0.6197	0.343	1.806	0.071	-0.053	1.292	
Q('Economic_Region_Дальневосточный_район')	0.7937	0.278	2.856	0.004	0.249	1.338	
Q('Economic_Region_Восточно_Сибирский_район')	0.3679	0.350	1.052	0.293	-0.318	1.054	
Q('Economic_Region_Северный_район')	0.2715	0.292	0.931	0.352	-0.300	0.843	
Omnibus:	69.197	Durbin-Watson:	2.199				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	170.013				
Skew:	1.340	Prob(JB):	1.21e-37				
Kurtosis:	6.179	Cond. No.	583.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 19. Аренда, итоговый эффект на датасете с 3-мя ближайшими «соседями»

OLS_rent_diff_winner_price_participants_1							
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.436	Model:	OLS	Adj. R-squared:	0.323
Method:	Least Squares	F-statistic:	5.469	Date:	Wed, 08 Nov 2023	Prob (F-statistic):	1.67e-09
Time:	12:09:55	Log-Likelihood:	-52.752	No. Observations:	127	AIC:	149.5
Df Residuals:	105	BIC:	212.1	Df Model:	21	Covariance Type:	HC3
	coef	std err	z	P> z	[0.025	0.975]	
Intercept	-0.2001	0.716	-0.280	0.780	-1.603	1.203	
Saler_SubjectRF	-0.0126	0.125	-0.101	0.920	-0.258	0.233	
Saler_Municipal	-0.0107	0.108	-0.099	0.921	-0.223	0.201	
Lg_SquareZU	-0.0594	0.084	-0.704	0.482	-0.225	0.106	
depozit_to_priceMin_ratio	0.0398	0.071	0.562	0.574	-0.099	0.179	
Is_ETP	-0.2474	0.098	-2.512	0.012	-0.440	-0.054	
Two_participants	0.4058	0.103	3.951	0.000	0.204	0.607	
More_then_two_participants	0.6290	0.120	5.240	0.000	0.394	0.864	
Q('Ограничение прав на ЗУ Имеются')	-0.0260	0.085	-0.307	0.759	-0.192	0.140	
Q('02. Жилая застройка')	-0.1383	0.482	-0.287	0.774	-1.083	0.807	
Q('04. Предпринимательство')	-0.5188	0.411	-1.261	0.207	-1.325	0.288	
Q('06. Производственная деятельность')	-0.5050	0.397	-1.270	0.204	-1.284	0.274	
Q('ЗУ Населенных пунктов')	0.3346	0.430	0.779	0.436	-0.508	1.177	
Q('Срок договора лет')	-0.0015	0.003	-0.508	0.611	-0.007	0.004	
Q('Economic_Region_Северо_Кавказский_район')	0.8004	0.535	1.497	0.134	-0.247	1.848	
Q('Economic_Region_Уральский_район')	0.6882	0.517	1.331	0.183	-0.325	1.701	
Q('Economic_Region_Центральный_район')	0.9912	0.555	1.786	0.074	-0.096	2.079	
Q('Economic_Region_Поволжский_район')	0.6948	0.533	1.304	0.192	-0.349	1.739	
Q('Economic_Region_Западно_Сибирский_район')	0.9483	0.536	1.769	0.077	-0.102	1.999	
Q('Economic_Region_Дальневосточный_район')	0.9260	0.429	2.160	0.031	0.086	1.766	
Q('Economic_Region_Восточно_Сибирский_район')	0.5098	0.518	0.984	0.325	-0.505	1.525	
Q('Economic_Region_Северный_район')	0.4805	0.537	0.895	0.371	-0.572	1.533	
Omnibus:	54.066	Durbin-Watson:	2.354				
Prob(Omnibus):	0.000	Jarque-Bera (JB):	188.506				
Skew:	1.533	Prob(JB):	1.17e-41				
Kurtosis:	8.121	Cond. No.	588.				

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 20.Аренда, итоговый эффект на датасете с 1-м ближайшим «соседом»

OLS_sale_diff_winner_price_participants_all						
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.281			
Model:	OLS	Adj. R-squared:	0.266			
Method:	Least Squares	F-statistic:	21.61			
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	5.82e-67			
Time:	13:17:37	Log-Likelihood:	267.84			
No. Observations:	1008	AIC:	-491.7			
Df Residuals:	986	BIC:	-383.5			
Df Model:	21					
Covariance Type:	HC3					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	0.2500	0.096	2.606	0.009	0.062	0.438
Saler_SubjectRF	-0.0016	0.039	-0.041	0.967	-0.078	0.075
Saler_Municipal	-0.0149	0.031	-0.478	0.632	-0.076	0.046
Lg_SquareZU	-0.0409	0.020	-2.077	0.038	-0.080	-0.002
depozit_to_priceMin_ratio	-0.0004	0.017	-0.020	0.984	-0.034	0.033
Is_ETP	-0.0516	0.015	-3.468	0.001	-0.081	-0.022
Two_participants	0.2136	0.018	11.935	0.000	0.179	0.249
More_then_two_participants	0.3639	0.050	7.239	0.000	0.265	0.462
Q('Ограничение_прав_на_ЗУ_Имеются')	-0.0278	0.024	-1.134	0.257	-0.076	0.020
Q('02. Жилая застройка')	-0.0404	0.038	-1.061	0.289	-0.115	0.034
Q('04. Предпринимательство')	-0.0550	0.042	-1.303	0.193	-0.138	0.028
Q('06. Производственная деятельность')	-0.0492	0.042	-1.162	0.245	-0.132	0.034
Q('Economic_Region_Московская_область')	-0.0046	0.035	-0.130	0.896	-0.074	0.065
Q('Economic_Region_Северо_Кавказский_район')	0.0541	0.032	1.683	0.092	-0.009	0.117
Q('Economic_Region_Уральский_район')	0.0140	0.023	0.604	0.546	-0.031	0.060
Q('Economic_Region_Центральный_район')	-0.0421	0.022	-1.878	0.060	-0.086	0.002
Q('Economic_Region_Поволжский_район')	0.0040	0.027	0.148	0.882	-0.049	0.057
Q('Economic_Region_Западно_Сибирский_район')	-0.0228	0.030	-0.749	0.454	-0.082	0.037
Q('Economic_Region_Центрально_Черноземный_район')	-0.0464	0.032	-1.444	0.149	-0.109	0.017
Q('Economic_Region_Дальневосточный_район')	0.1508	0.040	3.771	0.000	0.072	0.229
Q('Economic_Region_Восточно_Сибирский_район')	0.0153	0.027	0.576	0.565	-0.037	0.067
Q('Economic_Region_Северный_район')	-0.0600	0.024	-2.550	0.011	-0.106	-0.014
<hr/>						
Omnibus:	713.460	Durbin-Watson:	1.972			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13597.790			
Skew:	3.021	Prob(JB):	0.00			
Kurtosis:	19.949	Cond. No.	120.			

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 21. Продажа, итоговый эффект на датасете со всеми «соседями»

OLS_sale_diff_winner_price_participants_5						
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.414			
Model:	OLS	Adj. R-squared:	0.348			
Method:	Least Squares	F-statistic:	10.62			
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	3.73e-22			
Time:	13:18:14	Log-Likelihood:	57.010			
No. Observations:	208	AIC:	-70.02			
Df Residuals:	186	BIC:	3.405			
Df Model:	21					
Covariance Type:	HC3					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	0.1376	0.158	0.872	0.383	-0.172	0.447
Saler_SubjectRF	-0.0112	0.070	-0.159	0.874	-0.149	0.127
Saler_Municipal	0.0322	0.035	0.930	0.352	-0.036	0.100
Lg_SquareZU	-0.0335	0.028	-1.175	0.240	-0.089	0.022
depozit_to_priceMin_ratio	0.0618	0.073	0.845	0.398	-0.082	0.205
Is_ETP	0.0027	0.043	0.063	0.950	-0.081	0.087
Two_participants	0.2441	0.036	6.816	0.000	0.174	0.314
More_then_two_participants	0.4396	0.097	4.536	0.000	0.250	0.630
Q('Ограничение_прав_на_ЗУ_Имеются')	0.0604	0.116	0.523	0.601	-0.166	0.287
Q('02. Жилая застройка')	-0.0322	0.061	-0.529	0.597	-0.151	0.087
Q('04. Предпринимательство')	-0.1080	0.060	-1.792	0.073	-0.226	0.010
Q('06. Производственная деятельность')	-0.1130	0.072	-1.569	0.117	-0.254	0.028
Q('Economic_Region_Московская_область')	0.0188	0.100	0.188	0.851	-0.177	0.214
Q('Economic_Region_Северо_Кавказский_район')	0.0365	0.114	0.321	0.748	-0.186	0.259
Q('Economic_Region_Уральский_район')	0.0372	0.081	0.460	0.646	-0.121	0.196
Q('Economic_Region_Центральный_район')	-0.0638	0.089	-0.718	0.473	-0.238	0.110
Q('Economic_Region_Поволжский_район')	0.0509	0.090	0.567	0.571	-0.125	0.227
Q('Economic_Region_Западно_Сибирский_район')	-0.0795	0.105	-0.755	0.450	-0.286	0.127
Q('Economic_Region_Центрально_Черноземный_район')	0.0315	0.115	0.273	0.785	-0.195	0.258
Q('Economic_Region_Дальневосточный_район')	0.1764	0.099	1.779	0.075	-0.018	0.371
Q('Economic_Region_Восточно_Сибирский_район')	-0.0054	0.095	-0.057	0.955	-0.191	0.180
Q('Economic_Region_Северный_район')	-0.0025	0.090	-0.027	0.978	-0.179	0.174
Omnibus:	164.408	Durbin-Watson:	2.047			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2283.751			
Skew:	2.953	Prob(JB):	0.00			
Kurtosis:	18.121	Cond. No.	90.1			

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 22. Продажа, итоговый эффект на датасете с 5-ю ближайшими «соседями»

OLS_sale_diff_winner_price_participants_3						
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.517			
Model:	OLS	Adj. R-squared:	0.437			
Method:	Least Squares	F-statistic:	17.19			
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	8.71e-28			
Time:	13:18:51	Log-Likelihood:	66.806			
No. Observations:	150	AIC:	-89.61			
Df Residuals:	128	BIC:	-23.38			
Df Model:	21					
Covariance Type:	HC3					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	0.2505	0.140	1.783	0.075	-0.025	0.526
Saler_SubjectRF	0.0242	0.081	0.300	0.764	-0.134	0.182
Saler_Municipal	0.0244	0.040	0.609	0.543	-0.054	0.103
Lg_SquareZU	-0.0360	0.026	-1.383	0.167	-0.087	0.015
depozit_to_priceMin_ratio	-0.0208	0.071	-0.293	0.770	-0.160	0.118
Is_ETP	-0.0278	0.052	-0.535	0.592	-0.130	0.074
Two_participants	0.2592	0.050	5.201	0.000	0.161	0.357
More_then_two_participants	0.4332	0.101	4.303	0.000	0.236	0.631
Q('Ограничение_прав_на_ЗУ_Имеются')	0.1451	0.206	0.705	0.481	-0.258	0.548
Q('02. Жилая застройка')	-0.0162	0.050	-0.324	0.746	-0.114	0.081
Q('04. Предпринимательство')	-0.0836	0.063	-1.336	0.182	-0.206	0.039
Q('06. Производственная деятельность')	-0.0890	0.070	-1.263	0.206	-0.227	0.049
Q('Economic_Region_Московская_область')	-0.1106	0.060	-1.855	0.064	-0.227	0.006
Q('Economic_Region_Северо_Кавказский_район')	-0.0005	0.101	-0.005	0.996	-0.199	0.198
Q('Economic_Region_Уральский_район')	-0.0313	0.046	-0.684	0.494	-0.121	0.058
Q('Economic_Region_Центральный_район')	-0.1602	0.065	-2.479	0.013	-0.287	-0.034
Q('Economic_Region_Поволжский_район')	-0.0282	0.061	-0.462	0.644	-0.148	0.091
Q('Economic_Region_Западно_Сибирский_район')	-0.1502	0.098	-1.530	0.126	-0.343	0.042
Q('Economic_Region_Центрально_Черноземный_район')	-0.0775	0.130	-0.595	0.552	-0.333	0.178
Q('Economic_Region_Дальневосточный_район')	0.0926	0.072	1.282	0.200	-0.049	0.234
Q('Economic_Region_Восточно_Сибирский_район')	-0.0615	0.068	-0.909	0.363	-0.194	0.071
Q('Economic_Region_Северный_район')	-0.0833	0.055	-1.501	0.133	-0.192	0.025
Omnibus:	63.254	Durbin-Watson:	1.939			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	222.164			
Skew:	1.596	Prob(JB):	5.72e-49			
Kurtosis:	8.036	Cond. No.	120.			

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 23. Продажа, итоговый эффект на датасете с 3-мя ближайшими «соседями»

OLS_sale_diff_winner_price_participants_1						
Dep. Variable:	diff_lg_winner_price_lg_start_price	R-squared:	0.595			
Model:	OLS	Adj. R-squared:	0.453			
Method:	Least Squares	F-statistic:	6.012			
Date:	Wed, 08 Nov 2023	Prob (F-statistic):	4.59e-08			
Time:	13:20:20	Log-Likelihood:	37.867			
No. Observations:	78	AIC:	-33.73			
Df Residuals:	57	BIC:	15.76			
Df Model:	20					
Covariance Type:	HC3					
	coef	std err	z	P> z	[0.025	0.975]
Intercept	0.1942	0.215	0.901	0.367	-0.228	0.616
Saler_SubjectRF	0.0346	0.108	0.320	0.749	-0.178	0.247
Saler_Municipal	0.0117	0.070	0.167	0.867	-0.126	0.150
Lg_SquareZU	-0.0245	0.044	-0.552	0.581	-0.111	0.062
depozit_to_priceMin_ratio	0.0149	0.117	0.127	0.899	-0.215	0.244
Is_ETP	-0.0330	0.082	-0.404	0.686	-0.193	0.127
Two_participants	0.2999	0.098	3.057	0.002	0.108	0.492
More_then_two_participants	0.4245	0.129	3.282	0.001	0.171	0.678
Q('02. Жилая застройка')	-0.0288	0.055	-0.520	0.603	-0.137	0.080
Q('04. Предпринимательство')	-0.0325	0.086	-0.379	0.705	-0.200	0.136
Q('06. Производственная деятельность')	-0.0966	0.097	-0.997	0.319	-0.287	0.093
Q('Economic_Region_Московская_область')	-0.1260	0.107	-1.182	0.237	-0.335	0.083
Q('Economic_Region_Северо_Кавказский_район')	-0.0513	0.113	-0.452	0.651	-0.274	0.171
Q('Economic_Region_Уральский_район')	-0.0261	0.080	-0.325	0.745	-0.184	0.132
Q('Economic_Region_Центральный_район')	-0.1072	0.095	-1.132	0.258	-0.293	0.078
Q('Economic_Region_Поволжский_район')	-0.0016	0.116	-0.014	0.989	-0.228	0.225
Q('Economic_Region_Западно_Сибирский_район')	-0.1149	0.183	-0.627	0.531	-0.474	0.245
Q('Economic_Region_Центрально_Черноземный_район')	-0.2434	0.260	-0.937	0.349	-0.752	0.266
Q('Economic_Region_Дальневосточный_район')	0.0674	0.135	0.499	0.618	-0.197	0.332
Q('Economic_Region_Восточно_Сибирский_район')	-0.0369	0.115	-0.322	0.747	-0.261	0.188
Q('Economic_Region_Северный_район')	-0.0929	0.104	-0.895	0.371	-0.296	0.111
Omnibus:	65.627	Durbin-Watson:	2.206			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	487.266			
Skew:	2.443	Prob(JB):	1.55e-106			
Kurtosis:	14.228	Cond. No.	131.			

Notes:
[1] Standard Errors are heteroscedasticity robust (HC3)

Рис. 24. Продажа, итоговый эффект на датасете с самым ближайшим «соседом».

Результаты регрессий показывают:

Отсутствие статистически значимых различий в итогах торгов влияния на итоговую цену реализации земельного участка в зависимости от того реализуется участок Российской Федерации или Субъектом Российской Федерации, реализуется участок Российской Федерации или муниципалитетом.

На всех датасетах оценка коэффициентов при объясняющих переменных приняли участие два участника, приняли участие более двух участников является статистически значимым.

Диапазон оценок для коэффициента при переменной «два участника» для аренды варьируется от 0,4056 до 0,5532. С учетом, что в качестве зависимой переменной выступала разность десятичных логарифмов от итоговой цены победителя и начальной цены, это означает, что участие в торгах двух участников приводит к росту итоговой цены в 2,5 – 3,6 раза ($10^{0,4058} - 10^{0,5532}$) по сравнению с тем, когда в торгах участвует только один участник.

Диапазон оценок для коэффициента при переменной «более двух участников» для аренды варьируется от 0,6290 до 0,8650. Это означает, что участие в торгах более двух участников приводит к росту итоговой цены в 4,3 – 7,3 раза ($10^{0,6290} - 10^{0,8650}$) по сравнению с тем, когда в торгах участвует только один участник.

Диапазон оценок для коэффициента при переменной «два участника» для продаж варьируется от 0,2136 до 0,2999. Это означает, что участие в торгах более двух участников

приводит к росту итоговой цены в 1,6 – 2,0 раза ($10^{0,2136} - 10^{0,2999}$) по сравнению с тем, когда в торгах участвует только один участник.

Диапазон оценок для коэффициента при переменной «более двух участников» для продаж варьируется от 0,3639 до 0,4396. Это означает, что участие в торгах более двух участников приводит к росту итоговой цены в 2,3 – 2,8 раза ($10^{0,3639} - 10^{0,4396}$) по сравнению с тем, когда в торгах участвует только один участник.

Влияние на итоговые цены торгов числа участников торгов хорошо иллюстрируют следующие диаграммы рассеивания:

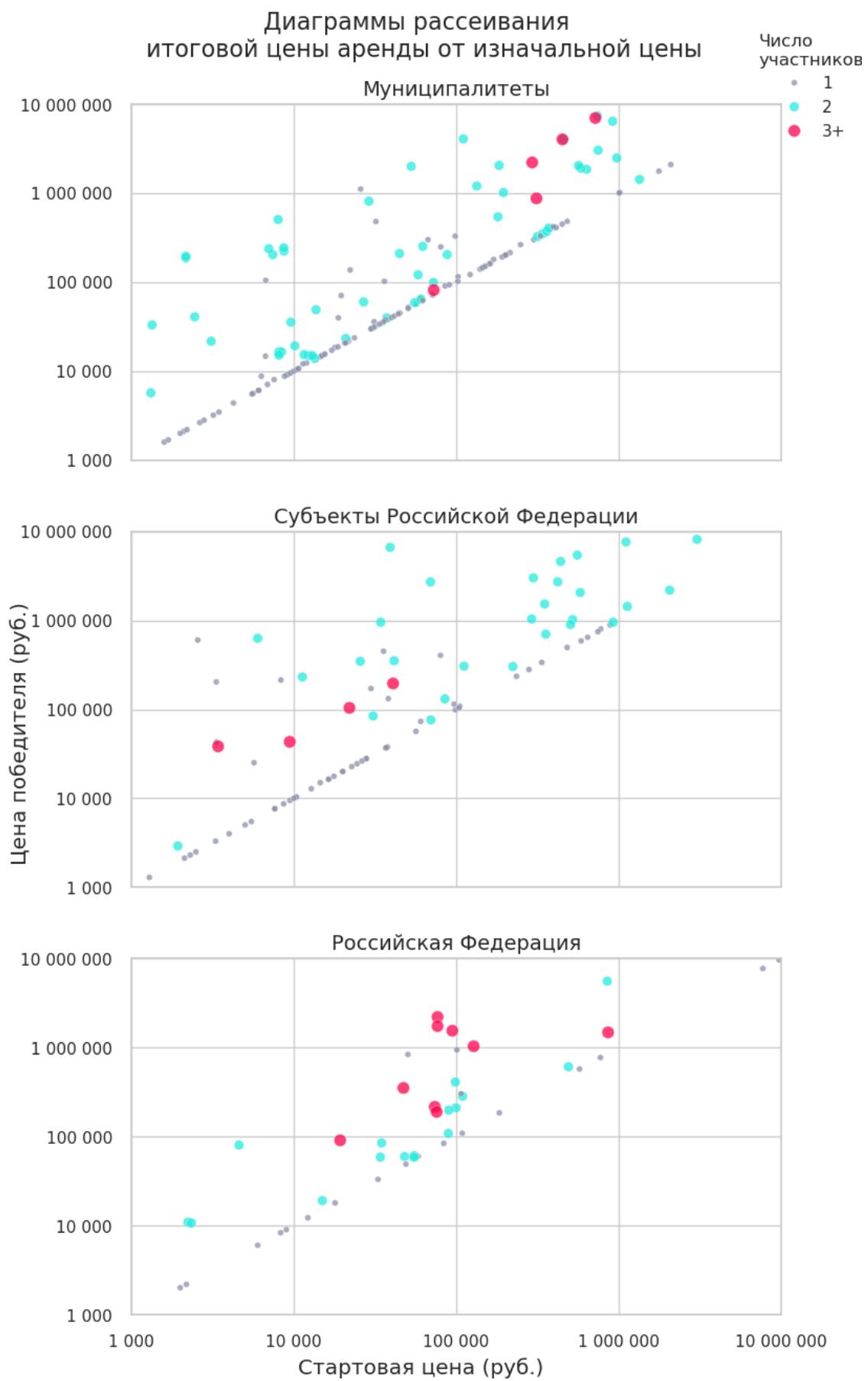


Рис. 25. Диаграммы рассеивания итоговой цены от стартовой цены для аренды

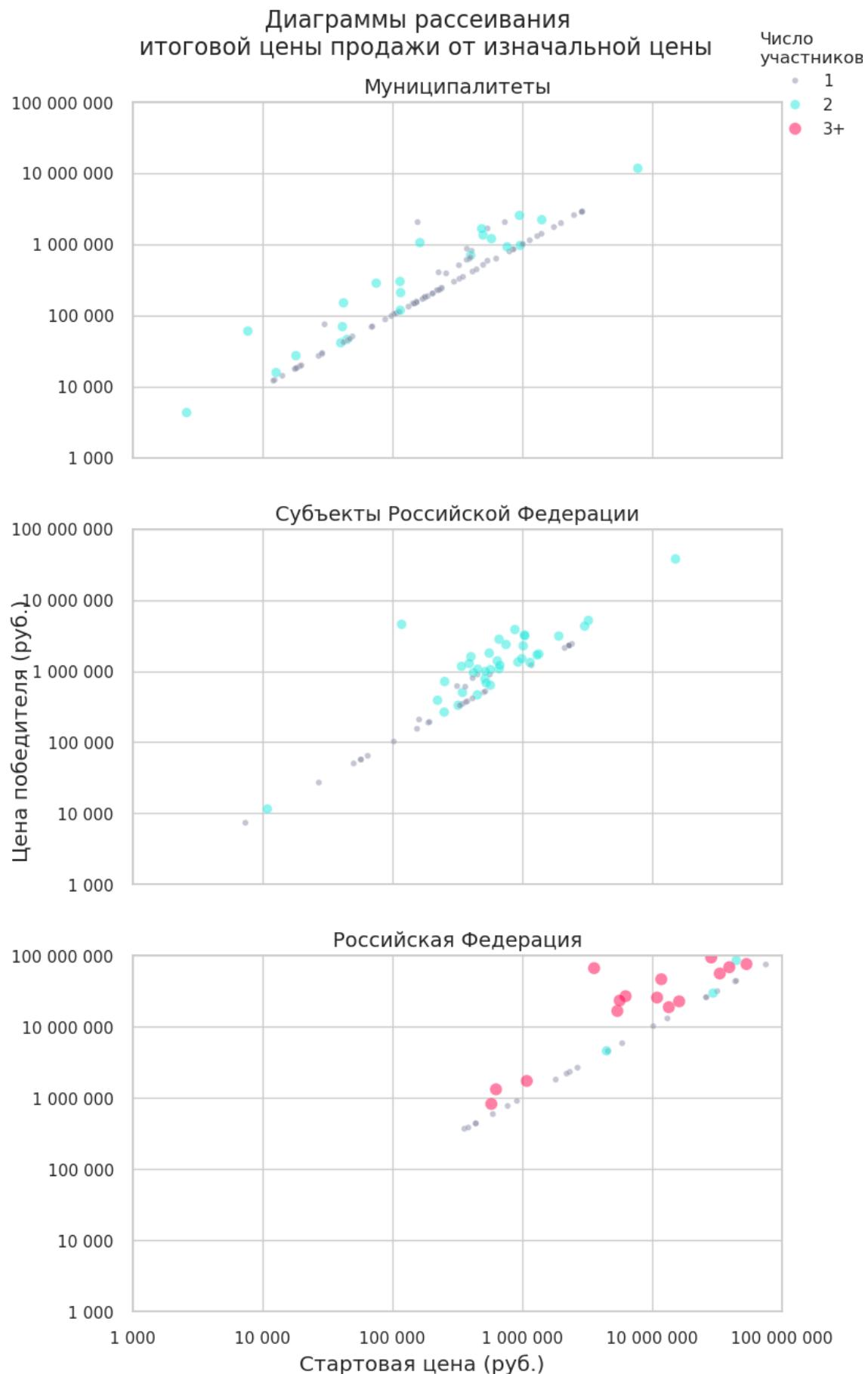


Рис. 26. Диаграммы рассеивания итоговой цены продажи для каждого изучаемого уровня

Приведенные выше диаграммы рассеивания для продаж также хорошо иллюстрируют насколько в частности по стартовой цене отличаются участки, продаваемые Российской Федерацией от участков, продаваемых Субъектами Российской Федерации и муниципалитетами.

Выводы:

1. Во всех случаях не выявлены статистически значимые отличия между Российской Федерацией и Субъектами Российской Федерации, между Российской Федерацией и муниципалитетами при формировании итоговой цены реализации земельных участков, как для аренды земельных участков, так и для продажи земельных участков.
2. Российская Федерация реализует земельные участки, которые существенно отличаются по своим характеристикам в частности по начальной цене от земельных участков, которые реализуются Субъектами Российской Федерации и муниципалитетами.
3. Для Российской Федерации и муниципалитетов на уровне 5% значимости имеется статистически значимое различие в доле торгов, в которых в торгах не принял участие ни один участник.
Вывод о статистически значимых различиях между долей торгов, в которых не принял участие ни один участник, для Российской Федерации и Субъектов Российской Федерации сделать нельзя.
4. Во всех случаях (для различных вариантов расчетов) на 5% уровне значимости выявлен положительный эффект влияния числа допущенных участников, принявших участие в торгах, на итоговую цену реализации участка. Значение оценки варьируется в зависимости от набора данных, на котором производилась оценка, при этом получен следующий диапазон оценок:
 - a. участие в торгах за право заключения договора аренды земельного участка 2-х участников приводит к росту итоговой цены в 2,5 – 3,6 раза по сравнению с тем, когда в торгах участвовал только один участник;
 - b. участие в торгах за право заключения договора аренды земельного участка более 2-х участников приводит к росту итоговой цены в 4,3 – 7,3 раза по сравнению с тем, когда в торгах участвовал только один участник;
 - c. участие в торгах на продажу земельного участка 2-х участников приводит к росту итоговой цены в 1,6 – 2,0 раза по сравнению с тем, когда в торгах участвовал только один участник;
 - d. участие в торгах на продажу земельного участка более 2-х участников приводит к росту итоговой цены в 2,3 – 2,8 раза по сравнению с тем, когда в торгах участвовал только один участник.

Скрипты Python:

1. `load_new_torgi_notifications.ipynb` и `load_new_torgi_protocols.ipynb` – скачивание открытых данных с сайта torgi.gov.ru.
2. `copy_notices_ZK_ZKPT_701PP_200FZ.ipynb`
`copy_protocols_ZK_ZKPT_701PP_200FZ.ipynb` – отбор извещений и протоколов, относящихся к земельным участкам.
3. `concat_files_all_types.ipynb` – объединение разрозненных файлов извещений и протоколов в один датасет.

4. ZK_parsing.ipynb – развертывание внутренней структуры извещений и протоколов в отдельные признаки.
5. ZK_rent_knn_find_ver4.ipynb, ZK_sale_knn_find_ver3.ipynb – отбор земельных участков, реализуемых Субъектами Российской Федерации и муниципалитетами с характеристиками, схожими с характеристиками земельных участков, реализуемых Российской Федерацией.
6. ZK_regression_v4.ipynb, ZK_sale_regression2.ipynb – реализация регрессионного анализа.

**II. Аналитическая записка
по оценке влияния приобретения (покупка, аренда) земельных участков у
Российской Федерации, субъектов Российской Федерации,
муниципалитетов на показатели деятельности компаний – приобретателей
земельных участков**

Исследовательский вопрос

Цель анализа – оценить, влияет ли приобретение земельного участка на торгах у Российской Федерации, Субъектов Российской Федерации, муниципалитетов на показатели хозяйственной деятельности компании⁹.

Нулевая гипотеза состоит в том, что показатели деятельности компании существенно не меняются после приобретения земельного участка. Альтернативная гипотеза состоит в том, что данные показатели изменяются.

Данные

Список компаний – участников торгов земельных участков был составлен на основе открытых данных о результирующих протоколах об итогах торгов земельными участками из данных открытого информационного ресурса «ГИС Торги» (<https://torgi.gov.ru/new/public/opendata/reg>) (далее – компании с участием) в 2021-2022 годах.

Финансовые показатели компаний были выгружены из информационной системы СПАРК.

Для исследования были выбраны следующие данные о компаниях: статус организации (действующая, реорганизуется, ликвидируется), регион регистрации, ОКАТО, ОКТМО, ОКВЭД, ОКОПФ, ОКФС, размер компании, налоговый режим, среднесписочная численность, доходы, расходы, уплачиваемые налоги, а также данные балансов за 2018-2022 годы:

- Активы (чистые, оборотные),
- Капитал и резервы,
- Долгосрочные обязательства,
- Краткосрочные обязательства,
- Выручка,
- Себестоимость продаж,
- Коммерческие расходы,
- Управленческие расходы,
- Доходы от участия в других организациях,
- Прочие доходы,
- Проценты к получению,

⁹ Для целей настоящей аналитической записки под приобретением понимаем покупку или заключение договора аренды земельного участка.

- Прибыль (убыток) до налогообложения,
- Чистая прибыль (убыток)
- Коэффициент соотношения заемных и собственных средств;
- Рентабельность продаж;
- Рентабельность затрат.

Выгрузка из информационной системы СПАРК осуществлялась как для организаций, которые участвовали в торгах земельными участками, так и организаций, которые участие в торгах не принимали (далее – компании без участия).

Всего изначально данные содержали информацию о 129 показателях. В процессе проведения исследования показатели подверглись предобработке, преобразованию для возможности использования в алгоритмах машинного обучения, категориальные переменные были переведены в числовые – дамми-переменные (принимающие значение 0 или 1 в зависимости от характеристик признака), также были созданы дополнительные производные переменные (подробнее будет изложено ниже).

Формирование исходного датасета компаний с участием

Из «ГИС Торги» были выгружены данные о компаниях, которые в 2021-2022 годах участвовали в торгах на право заключения договора аренды земельного участка у государства.

Изначально всего было выгружено 5059 записей об участниках торгов, но эта информация включает физических лиц, которые не входят в предмет исследования.

Данные в частности содержали информацию об основном государственном регистрационном номере (ОГРН), идентификационном номере налогоплательщика (ИНН), номере извещения, номере протокола, номере лота, порядковом номере лота.

На основе данных был сформирован список ИНН для выгрузки финансовых показателей из информационной системы СПАРК (всего 1 файл).

После исключения физических лиц, дубликатов по ИНН (один участник мог участвовать в нескольких торгах) осталось 919 записей.

Формирование исходного датасета компаний без участия

Для сравнительного анализа показателей компаний, которые приобретали земельные участки, и которые не приобретали земельные участки, требовалось создать объемный датасет из компаний, которые не участвовали в аукционе.

В силу ограничений информационной системы СПАРК на размер выгрузки (не более 10 000 записей за выгрузку) формирование датасета

компаний, которые не участвовали в аукционе осуществлялось следующим образом:

- Для каждого Субъекта Российской Федерации делалась либо одна выгрузка (если число компаний в субъекте до 10000), либо 2 выгрузки (если число компаний в субъекте более 10000), в этом случае первый список формировался по названию компаний в алфавитном порядке по возрастанию, а второй по названию компаний в алфавитном порядке по убыванию.
- С сайта Банка России (https://cbr.ru/registries/rcb/inn_ogrн/) был взят список ИНН компаний, эмитентов ценных бумаг, векселедателей, по состоянию на начало 2021 - 2022 года
- Были взяты компании, которые не находились в статусе «недействующие» по состоянию на момент выгрузки – май 2023 года.¹⁰

Всего из информационной системы СПАРК было сделано 184 выгрузки. Далее файлы с показателями были объединены в один датафрейм

После исключения дубликатов по ИНН размер исходного датасета компаний без участия в аукционе, составил 1 038 422 записи. Всего в информационной системе СПАРК имеется информация о 3,3 млн. действующих компаний, т.е. наш сформированный датасет составляет 31% от всех компаний, имеющихся в информационной системе СПАРК. С учетом порядка формирования выборки (для каждого субъекта Российской Федерации, отсутствии избирательности при отборе компаний в выборку) можно считать, что сформированный датасет является репрезентативным для всех компаний.

Полнота данных

Данные датасетов содержали пропуски (см. график). Белый цвет на графиках ниже соответствует пропускам данных. Видно, что заполняемость данных в датасете компаний без участия заметно ниже, чем в датасете компаний с участием.

² По компаниям в статусе «недействующие» в информационной системе СПАРК показатели заполнены с очень большим числом пропусков, с учетом этого для анализа брались данные только по действующим компаниям.

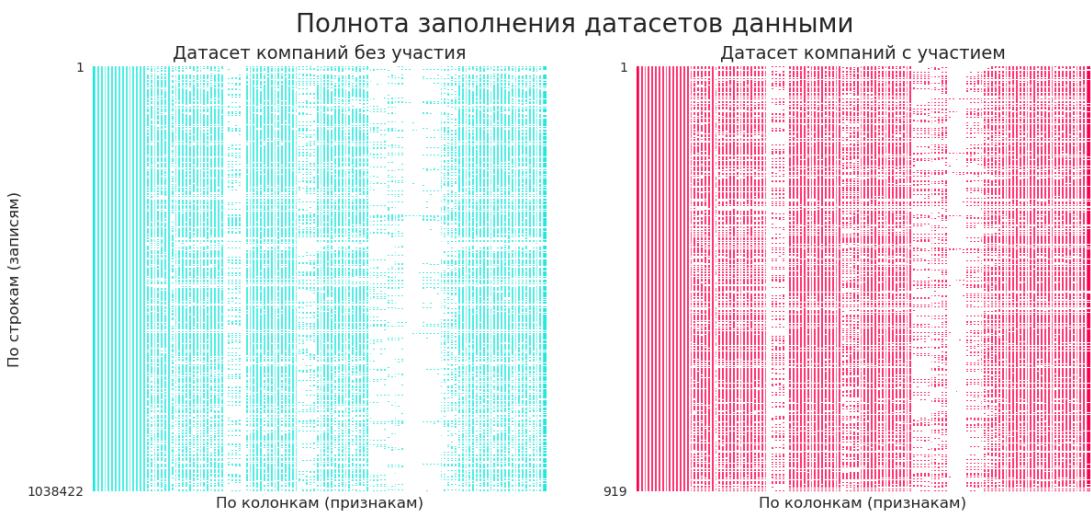


Рисунок 1. График заполнения данных в датасетах

Добавление новых признаков

По первым двум цифрам кода ОКВЭД можно определить отрасль и класс вида деятельности. В датасеты были добавлены переменные с первыми двумя цифрами кода ОКВЭД и класс ОКВЭД.

В датасете компаний с участием поле налоговый режим содержит различные комбинации из следующих режимов: УСН, ЕНВД, ЕСХН. И для датасета компаний с участием и для датасета компаний без участия создадим три дамми-переменные, характеризующие наличие (1) или отсутствие (0) соответствующего налогового режима.

Первичное приведение в соответствие показателей датасета компаний с участием к показателям датасета компаний без участия

Для целей проведения анализа на сопоставимых данных в датасете компаний без участия были оставлены только те записи по классу ОКВЭД, которые есть в датасете компаний с участием.

Аналогично для кода организационно-правовой деятельности ОКОПФ, в датасете компаний, без участия, были оставлены записи с кодами ОКОПФ, которые есть в датасете компаний с участием.

Кроме того, в датасете компаний с участием 97,5% компаний – это коммерческие организации, а среди коммерческих организаций 92,3% это: общества с ограниченной ответственностью, непубличные акционерные общества, публичные акционерные общества и акционерные общества. Для дальнейшего анализа и в датасете компаний с участием и в датасете компаний без участия оставим записи только для коммерческих организаций в форме обществ с ограниченной ответственностью, непубличных акционерных обществ, публичных акционерных обществ и сельскохозяйственной деятельности.

Созданы соответствующие дамми-переменные для этих категорий организационно-правовых форм.

Для поля форма собственности – «Код ОКФС» выделены формы собственности с иностранным участием и с государственным или муниципальным участием и созданы отдельные две дамми-переменные для данных признаков.

Очистка данных

Для приведения данных в вид, позволяющий их исследовать, были сделаны следующие шаги.

Шаг 1:

Удалены все записи, для которых в датасетах пустые все финансовые показатели.

Шаг 2.

Для показателей Доходы 2018-2022, Расходы 2018-2022, Налоги 2018-2022, Выручка 2018-2022, Чистые активы 2018-2022, Оборотные активы 2018-2022, Капитал и резервы 2018-2022, Себестоимость продаж 2018-2022, удалены все записи, для которых 2 и более пропуска в 5-и показателях. Если из 5-и показателей пропуск в 1-м окаймлен другими показателями, то он заполнялся линейной интерполяцией. Если из 5-и показателей пропуск в 1-м, и он не окаймлен другими показателями, то такие записи исключались.

Для показателей Краткосрочные обязательства 2018-2022, Коммерческие расходы 2018-2022, Управленческие расходы 2018-2022, Доходы от участия в других организациях 2018-2022, Проценты к получению 2018-2022, Прибыль (убыток) до налогообложения 2018-2022, Коэффициент соотношения заемных и собственных средств 2018-2022, Рентабельность продаж 2018-2022, Рентабельность затрат 2018-2022 пропуски заполнены 0.

Показатели Налог на прибыль 2018-2022, Прочие доходы 2018-2022 в связи с большим количеством пропусков в данных исключены из дальнейшего анализа.

Шаг 3.

После исключения записей с пропусками в датасетах осталось:

Датасет для компаний с участием – 475 записей.

Датасет для компаний без участия – 226 832 записей.

Метод

Для оценки использовался квазиэкспериментальный метод мэтчинга, основной идеей которого является сравнение экспериментальной (подвергается воздействию) группы и контрольной (не подвергается воздействию) группы.

Экспериментальная группа была сформирована из компаний победителей, которые участвовали в торгах (далее – компании с приобретением ЗУ):

Контрольная группа была сформирована из организаций, которые не арендуют землю у государства (далее – компании без приобретения ЗУ)

Датасет для компаний с приобретением ЗУ – 154 записей.

Датасет для компаний без приобретения ЗУ – 226 832 записей.

Преобразование целевой переменной

Из графиков (рисунок 2) видно, что в выборках присутствуют выбросы (хвосты), которые сильно выбиваются на фоне остальных значений переменных.

Перед преобразованием исключим сильно выбивающиеся значения по доходам за 2021 год, которые могут привести к искажению результатов.

Считаем, что значение является сильно выбивающимся, если оно лежит ниже 0.01-квантили и выше 0.99-квантили значений переменных.

Исключение по 2021 году осуществилось в связи с тем, что данные этого года наиболее близки к данным 2022 года.

Мы не исключаем выбросы по 2022 году, за исключением случаев некорректно заполненных данных, чтобы не вмешиваться в ход самого эксперимента.

Из значения переменных доходов 2022 года была исключена организация, превышающая показатель в 10^{14} , что не соответствует данными финансовой отчетности из ГИР БО¹¹.

После исключения данных получили:

датасет для компаний с приобретением ЗУ – 150 записей,

датасет для компаний без приобретения ЗУ – 226 831 записей.

Масштабируем данные натуральным логарифмом от показателей (модуля показателей) доходов за 2022 год.

¹¹ Государственный информационный ресурс бухгалтерской (финансовой) отчетности.

Гистограммы распределения доходов за 2022 гг.
До и после преобразования

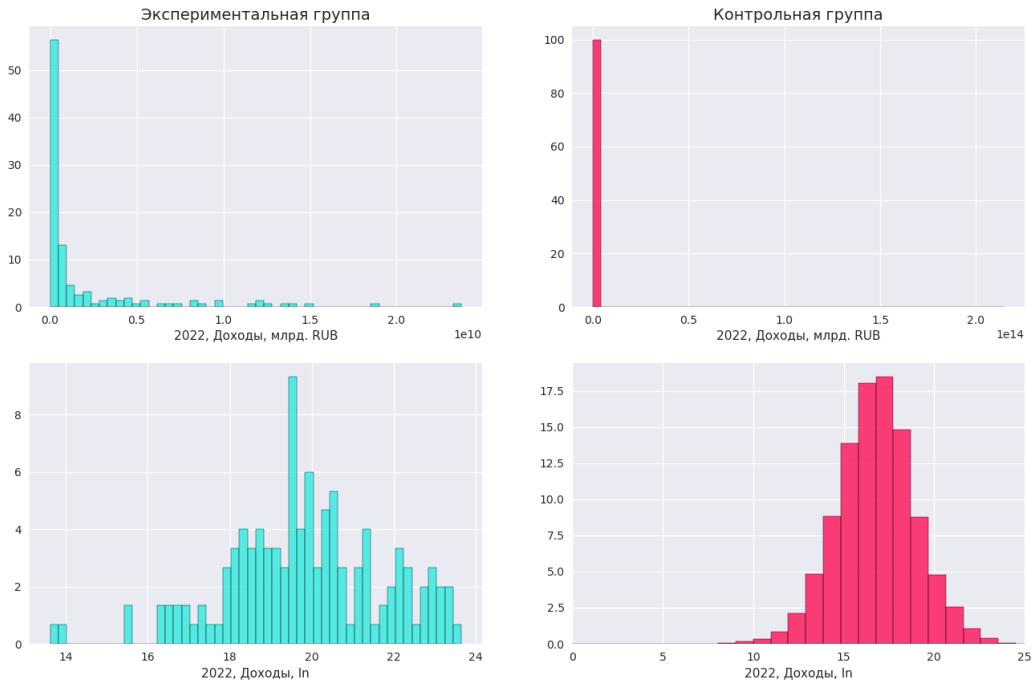


Рисунок 2. Гистограмма распределения доходов в руб.

Сопоставление компаний с приобретением ЗУ и без приобретения ЗУ

Для того чтобы оценить влияет ли аренда участка земли у государства на доходы 2022 года компаний из экспериментальной группы, необходимо найти такие компании из контрольной группы, чьи результаты финансово-хозяйственной деятельности за период 2018-2021 год были максимально похожими на финансовые показатели организаций за тот же период в экспериментальной группе. После нахождения максимально похожих организаций необходимо вычислить некую «дельту», возникающую в результате приобретения ЗУ.

Под «дельтой» будем понимать эффект воздействия τ_i следующего вида:

$$\tau_i = Y_i(1) - Y_i(0)$$

Где

$i = 1, \dots, N_1$ – множество организаций экспериментальной группы

$Y_i(1)$ – доход 2022 года i -ой организации из экспериментальной группы

$Y_i(0)$ – доход 2022 года организации из контрольной группы, которая соотносится с i -ой организацией из экспериментальной группы

Следовательно, необходимо использовать такой алгоритм подбора организаций, который для целевого значения (доходы 2022 года) каждой компании из экспериментальной группы сопоставит значение дохода (среднее

значение доходов 2022 года, если будет найдено несколько и более похожих организаций) из контрольной группы по схожим значениям ковариат.

Исходный список признаков для сопоставления компаний:

- Расходы 2018-2021;
- Налоги 2018-2021;
- Чистые активы 2018-2021;
- Оборотные активы 2018-2021;
- Капитал и резервы 2018-2021;
- Краткосрочные обязательства 2018-2021;
- Выручка 2018-2021;
- Себестоимость продаж 2018-2021;
- Коммерческие расходы 2018-2021;
- Управленческие расходы 2018-2021;
- Доходы от участия в других организациях 2018-2021;
- Проценты к получению 2018-2021;
- Прибыль (убыток) до налогообложения 2018-2021;
- Чистая прибыль (убыток) 2018-2021.

Отбор соответствующей компании осуществим с помощью метода **к-ближайших соседей** – метрический алгоритм для автоматической классификации объектов или регрессии, который группирует объекты, сопоставляя векторы ковариат.

В качестве метрики «близости» показателей из датасетов с приобретением ЗУ и без приобретения ЗУ было выбрано Евклидово расстояние.

Формально, каждой организации i из экспериментальной группы сопоставляется организация $m(i)$ из контрольной группы с наименьшим расстоянием ковариат.

$$m(i) = \underset{j:j \in D_0}{\operatorname{argmin}} \|X_j - X_i\|$$

Где:

$\|X_j - X_i\|$ – Евклидово расстояние между ковариатами векторов X_j и X_i

D_0 – исходное множество организаций, которые не приобретали земельные участки.

Будем использовать сопоставление: одна организация из контрольной группы сопоставляется одной организации из экспериментальной группы, то есть каждому целевому показателю их экспериментальной группы будет сопоставляться не среднее значение доходов за 2022 год нескольких похожих организаций из контрольной группы, а доход 2022 года первой взятой nearestнейшей организацией.

Запишем индексы сопоставленных организаций в вектор J_m

$$J_m = \{m(1), m(2), \dots, m(N_1)\}$$

Где:

N_1 – количество участников экспериментальной группы.

Выбранная норма не инвариантна к масштабу объектов (диапазону значений для признаков). Это означает, что одни признаки могут быть восприняты менее значимы только по тому, что диапазон из значений меньше, чем диапазон значений других признаков. По этой причине, прежде чем применять алгоритм классификации объектов, требуется масштабировать объекты так, чтобы они были примерно в одном масштабе.

Признаки были масштабированы при помощи формулы стандартизации:

$$Z = \frac{x - \mu}{\sigma}$$

где:

x – случайная величина;

μ – значение математического ожидания;

σ – стандартная ошибка случайной величины.

На полученных масштабированных признаках (ковариатах) был рассчитан нормализованный коэффициент для проверки предположения о том, что ковариаты экспериментальных и контрольных групп должны иметь одинаковое распределение (покрытие). Отсутствие покрытия означает, что группы уже по своей сути будут разными, и любые выводы, которые мы можем сделать будут спорными (ошибочными).

Нормализованный коэффициент для признаков экспериментальной группы и контрольной группы был рассчитан по формуле:

$$\text{Normalized score} = \frac{\overline{X}_{treat} - \overline{X}_{control}}{\sqrt{(S^2_{treat} + S^2_{control})/2}}$$

где:

$\overline{X}_{treat}, \overline{X}_{control}$ – выборочные средние значения сравниваемых признаков в экспериментальной группе и контрольной группе

$S^2_{treat}, S^2_{control}$ – выборочные дисперсии сравниваемых признаков в экспериментальной группе и контрольной группе

Из двух групп были исключены признаки с нормализованным коэффициентом вне интервала $(-0.5, 0.5)$.

Дополнительно из группы были исключены признаки с высокой корреляцией. Такое решение было продиктовано с проблемой оценки (неустойчивости) и анализа общего результата при мультиколлинеарность.

Отбор осуществлялся при помощи коэффициентов VIF (variance inflation factor), которые показывают насколько сильно связаны друг с другом регрессоры модели. Чтобы определить коэффициенты VIF, соответствующий регрессору x^j , нужно оценить вспомогательную регрессию в которой слева стоит x^j , а справа все остальные объясняющие переменные исходной модели. После этого вычисляется коэффициент VIF по формуле:

$$VIF = \frac{1}{1 - R^2}$$

где:

R^2 – это коэффициент детерминации из оцененной вспомогательной регрессии.

Исключили все значения с коэффициентами $VIF > 10$.

В результате были отобраны следующие признаки для сопоставления компаний:

- Расходы 2018-2019;
- Налоги 2018, 2020;
- Капитал и резервы 2018;
- Себестоимость продаж 2018, 2019, 2021;
- Чистая прибыль (убыток) 2018, 2019, 2020;
- Краткосрочные обязательства 2021
- Коммерческие расходы 2018, 2021;
- Управленческие расходы 2018-2021;

Получили следующие распределения целевого показателя, сопоставленных по финансовым показателям за период 2018-2021:

Гистограмма распределения натурального логарифма доходов за 2022 год до сопоставления компаний

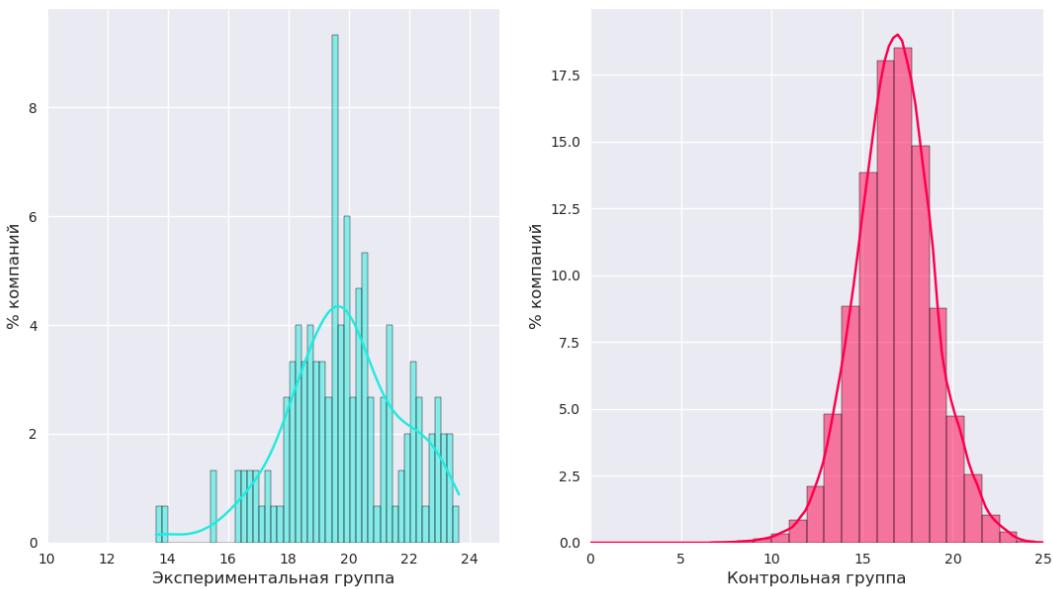


Рисунок 3. Гистограмма распределения натурального логарифма доходов за 2022 год до сопоставления компаний

Гистограмма распределения натурального логарифма доходов за 2022 год после сопоставления компаний

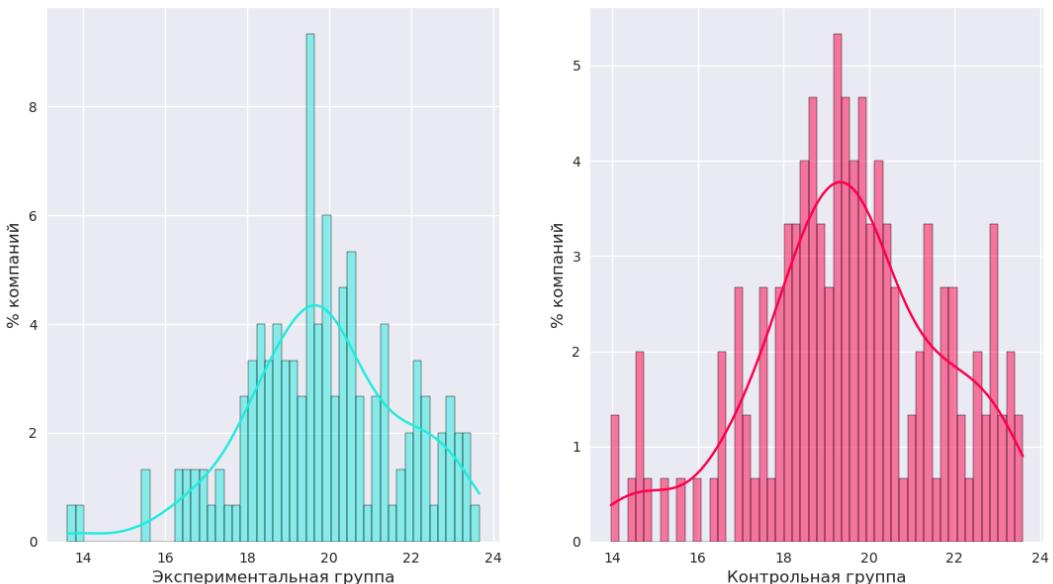


Рисунок 4. Гистограмма распределения натурального логарифма доходов за 2022 год после сопоставления компаний

Графики выше показывают, что удалось сопоставить схожие по признакам компании в датасетах с арендой и без аренды.

Оценка эффектов

После сопоставления экспериментальной группы и контрольной группы получили, что каждому значению дохода за 2022 года из экспериментальной группы подбиралось значение дохода за 2022 год из контрольной группы¹².

Формально получаем:

$$\hat{Y}_{i(W_i=0)} = Y_{m(i), m(i) \in J_m}$$

где:

$i = 1, \dots, N_1$ – множество организаций экспериментальной группы

$\hat{Y}_{i(0)}$ – доход 2022 года $m(i)$ организации из контрольной группы сопоставленный доходу 2022 года i организации из экспериментальной группы

W_i – индикатор принадлежности к контрольной ($W_i = 0$) или экспериментальной группе ($W_i = 1$)

Получение состоятельной оценки доходов экспериментальной группы по доходам контрольной группы требует (см. указанную выше статью Abadie & Imbens (2011)) дополнительного поправочного члена, который получается из оценки регрессии:

$$\hat{\mu}_{W_i=0}(x) = \hat{\beta}_{00} + \hat{\beta}'_{01}x$$

Коэффициенты регрессии оцениваются методом наименьших квадратов

$$(\hat{\beta}_{00}, \hat{\beta}_{01}) = \min_{\beta_{00}, \beta_{01}} \sum_{i: W_i=0} (Y_{m(i)} - \beta_{00} - \beta'_{01} X_{m(i)})^2$$

Где:

$W_i = 0$ – индикатор принадлежности организаций к контрольной группе

$i = 1, \dots, N_1$ – множество организаций экспериментальной группы

Получаем сопоставление доходов 2022 года экспериментальной и контрольной группы:

$$\hat{Y}_{i(w_i=0)} = Y_{m(i), m(i) \in J_m} + \hat{\mu}_0(X_i) - \hat{\mu}_0(X_{m(i), m(i) \in J_m})$$

Где:

$\hat{\mu}_0(X_i)$ – оценка линейной регрессией целевого показателя участников экспериментальной группы, если бы они были в контрольной группе;

¹² Alberto Abadie & Guido W. Imbens (2011) Bias-Corrected Matching Estimators for Average Treatment Effects, Journal of Business & Economic Statistics, 29:1, 1-11, DOI: 10.1198/jbes.2009.07333)

$\hat{\mu}_0(X_{m(i), m(i) \in J_m})$ – оценка линейной регрессии целевого показателя на факторах $m(i)$ организации из контрольной группы для которых существует максимально похожие факторы организации i из экспериментальной группы.

Используя подобранные значения доходов 2022 года контрольной группы к доходам 2022 года экспериментальной можно оценить эффект воздействия на участников, которые арендуют землю у государства с помощью метода среднего эффекта воздействия на подвергшихся воздействию (average effect of the treatment on treated)

$$\hat{t}_M^{bcm,t} = \frac{1}{N_1} \sum_{i: W_i=1} \{Y_i - \hat{Y}_{i(0)}\}$$

Где:

N_1 – количество участников экспериментальной группы;

$Y_i - \hat{Y}_{i(0)}$ – разница между доходами 2022 года экспериментальной группы и подобранными значениями доходов 2022 года из контрольной группы.

Проверим гипотезу о статистической значимости показателя $\hat{t}_M^{bcm,t}$, его статистическое отличие от нуля.

$$\begin{cases} H_0: \hat{t}_M^{bcm,t} = 0 \\ H_a: \hat{t}_M^{bcm,t} \neq 0 \end{cases}$$

Проверку гипотезы будем осуществлять при помощи t – теста (t – критерий Стьюдента)

Применялась следующая методика расчета:

1. Предварительно была рассчитана t – статистика по формуле:

$$t = \frac{\bar{X} - H_0}{\frac{\sigma}{\sqrt{n}}}$$

Где:

\bar{X} – случайная величина выборочного среднего;

H_0 – значение математического ожидания;

$\frac{\sigma}{\sqrt{n}}$ – стандартная ошибка случайной величины.

2. Определили уровень значимости $\alpha = 5\%$ при котором H_0 отвергается, если t-статистика превышает своё критическое значение ($t > |t_{kp}^{\frac{\alpha}{2}}|_{n-1}$)
3. Рассчитали диапазон значений, который может содержать параметр генеральной совокупности (построили доверительный интервал) по формуле:

$$\bar{X} \pm t * \left(\frac{\sigma}{\sqrt{n}} \right)$$

Поскольку р-значение (0,607438) больше, чем альфа (0.05), мы не можем отвергнуть нулевую гипотезу теста. У нас нет достаточных доказательств, чтобы сказать, что средний эффект воздействия на участников экспериментальной группы отличен от нуля (Таблица 1).

Таблица 1. Результаты расчета среднего эффекта воздействия на подвергшихся воздействию (экспериментальную группу).

Treatment Effect Estimates: Matching

statistic	s.e.	t	P> t	confidence interval	95%
0.514826	1.505616	0.042035	0.607438	[-0.179628, 0.306206]	

В результате получили отсутствие статистической значимости влияния аренды земельного участка государства на участников экспериментальной группы – эффект воздействия статистически не отличается от нуля.

Вывод:

Приобретение земельного участка не оказывает статистически значимого влияние на доходы компаний в первый год после приобретения участка в сравнении с доходами схожих компаний, которые не приобретали земельные участки.

Литература

Alberto Abadie & Guido W. Imbens (2011) Bias-Corrected Matching Estimators for Average Treatment Effects, Journal of Business & Economic Statistics, 29:1, 1-11, DOI: 10.1198/jbes.2009.07333)

Скрипты Python

7. land_part2_v15.ipynb – консолидация и обработка данных, сопоставление, оценка эффектов.