Цена на жилье растет

В отличие от зарплаты



Манна-Уитни band

ПРЕДАНАЛИЗ

Структура данных

Датасет по проектам
Подробные данные по каждому объекту недвижимости

17 столбцов

21 300 строк

Датасет по сделкам
Информация о покупке квартир с 2019 по 2025 год

28 столбцов

50 955 строк

ПРЕДАНАЛИЗ

Основные переменные



Площадь лота
в квадратных метрах



Сумма бюджета (цена лота)
в рублях



Даты планируемого и фактического ввода в эксплуатацию
год, месяц и день



Застройщик, этажей в здании, общая проектная площадь ЖК
вспомогательные переменные

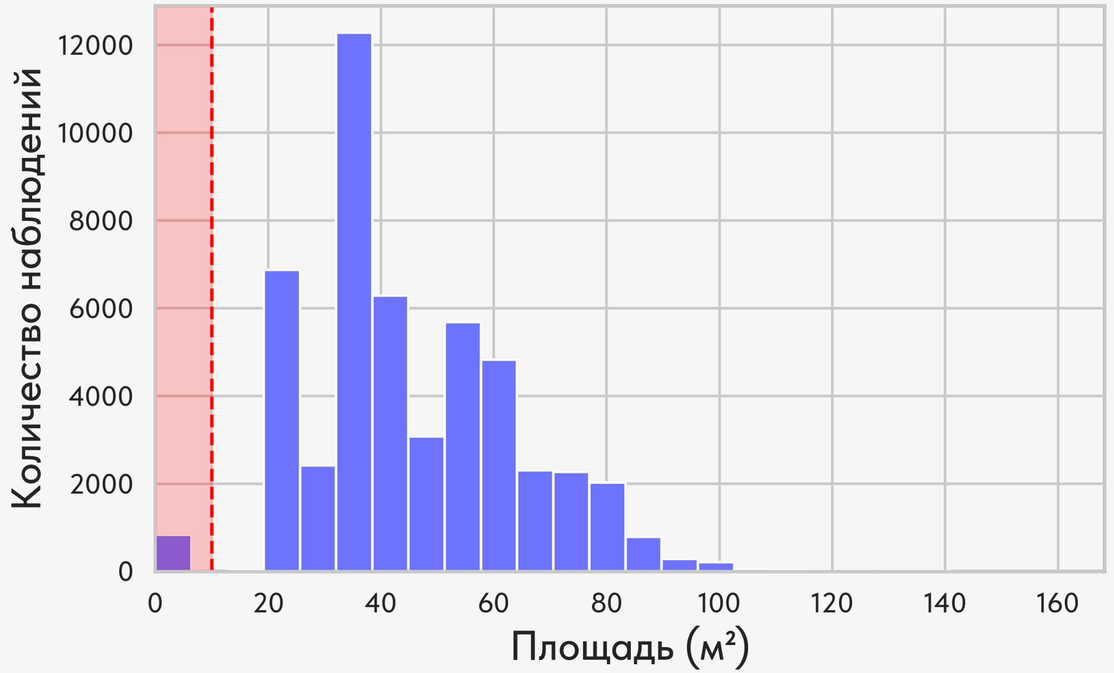
Площадь квартир и нюансы базы

У нас есть столбец “количество сделок”, в котором есть значения больше одного → Это значит, что похожие лоты агрегированы с помощью усреднения → Удаляем такие немногочисленные строки, чтобы повысить точность исследования

Очистим аномальные значения:
нулевая площадь объекта

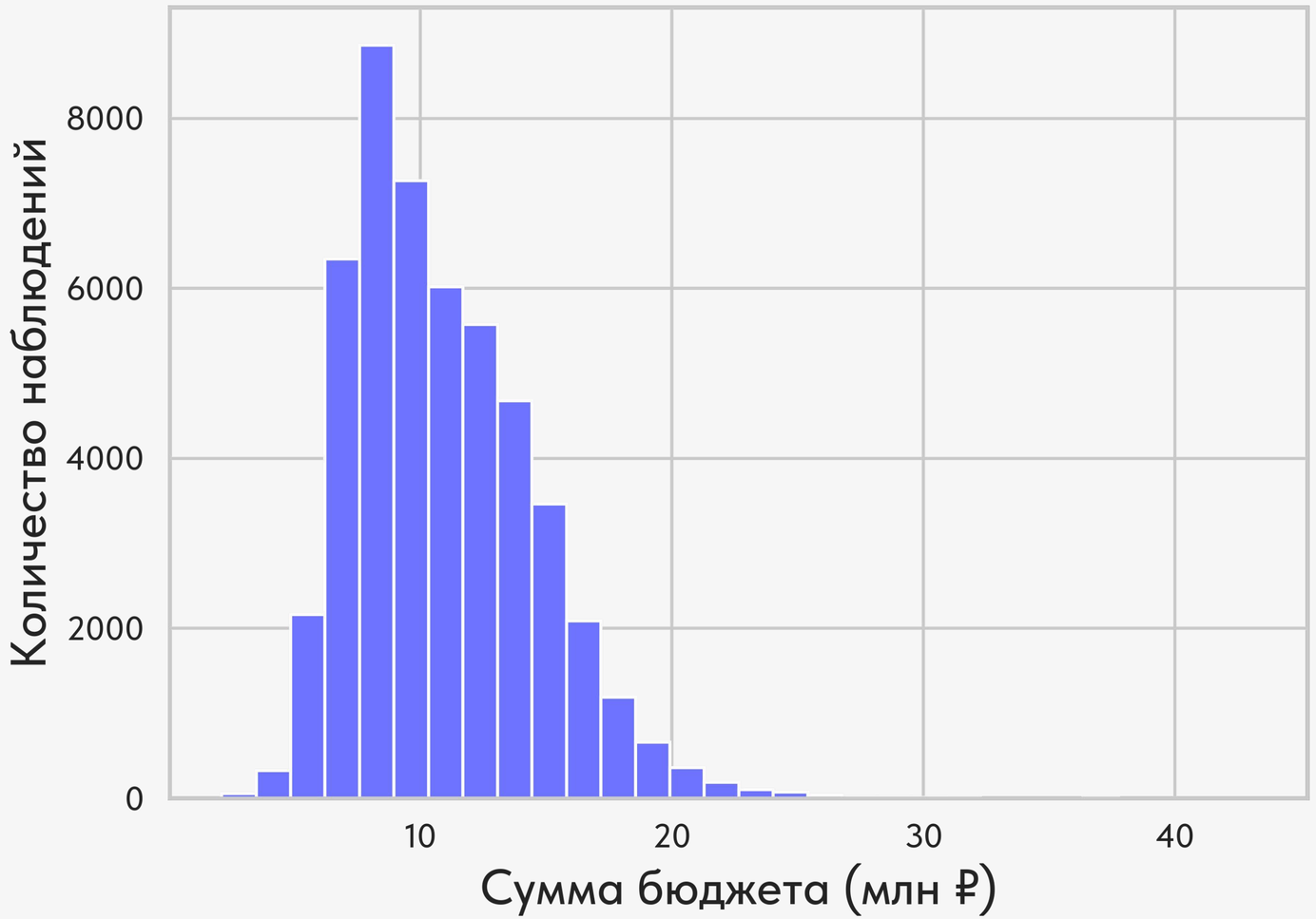
46
среднее

18
стандартное отклонение



ПРЕДАНАЛИЗ

Распределение цен за квартиры



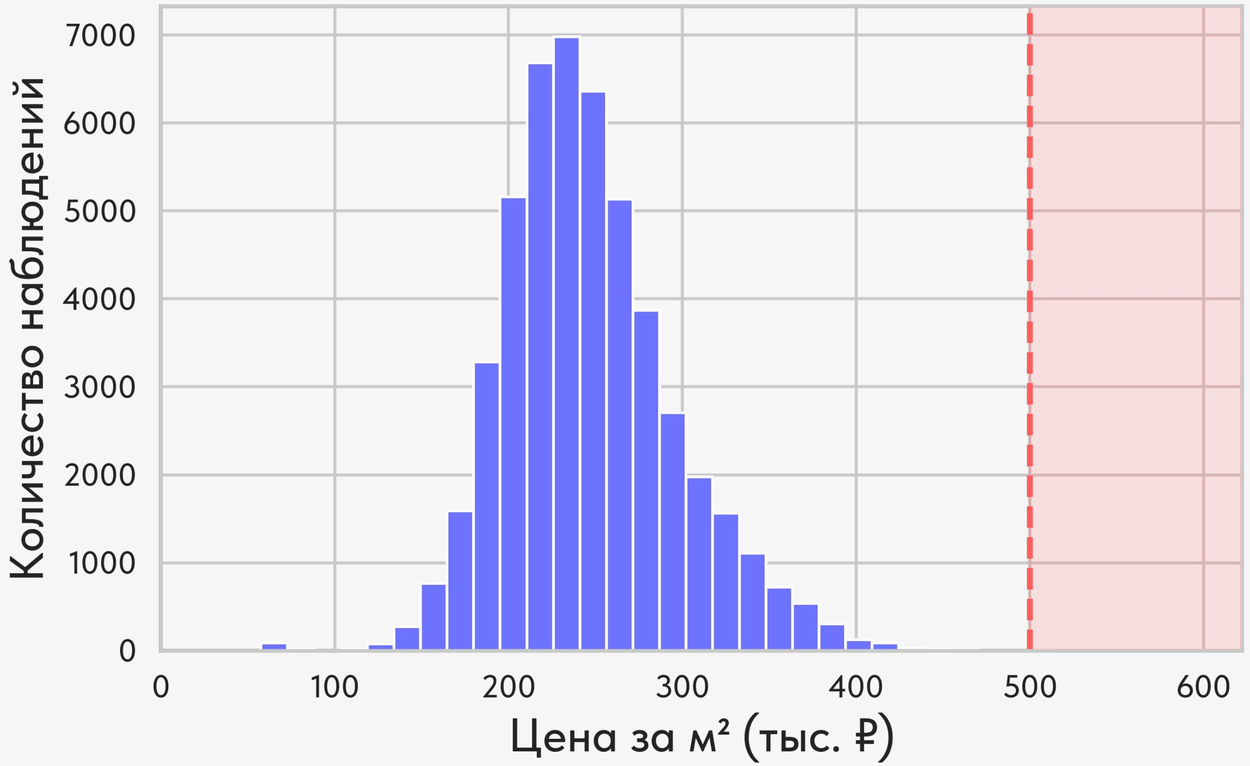
11
среднее

10
медиана

4
стандартное отклонение

Создание новой переменной

Цена за квадратный метр квартиры
рассчитывается путем деления общей цены лота на его площадь



246
среднее

241
медиана

49
стандартное
отклонение

Удаляем значения цены больше 500 тыс.

ПРЕДАНАЛИЗ

Данные по отдельным корпусам

Агрегируем данные по корпусам по следующим причинам:

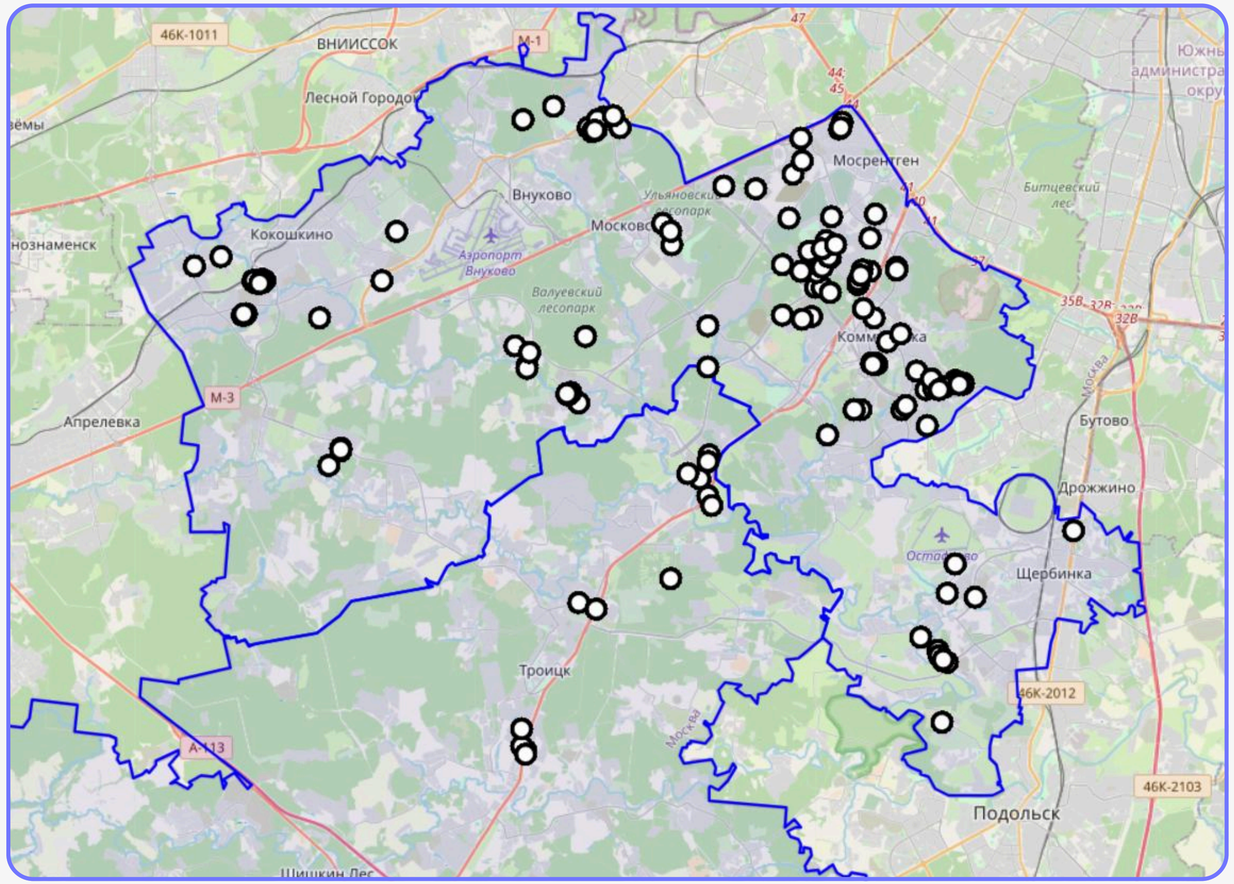
1. **Дом сдается целиком**
   Это помогает нам объединить квартиры, сданные в одно время в одном корпусе

2. **Помогает сделать данные независимыми**
   Позволит нам проводить статистические тесты и строить регрессии

В наших данных 419 корпусов

Географические данные

Во время анализа мы нашли каждый ЖК на карте, исходя из колонки с адресом



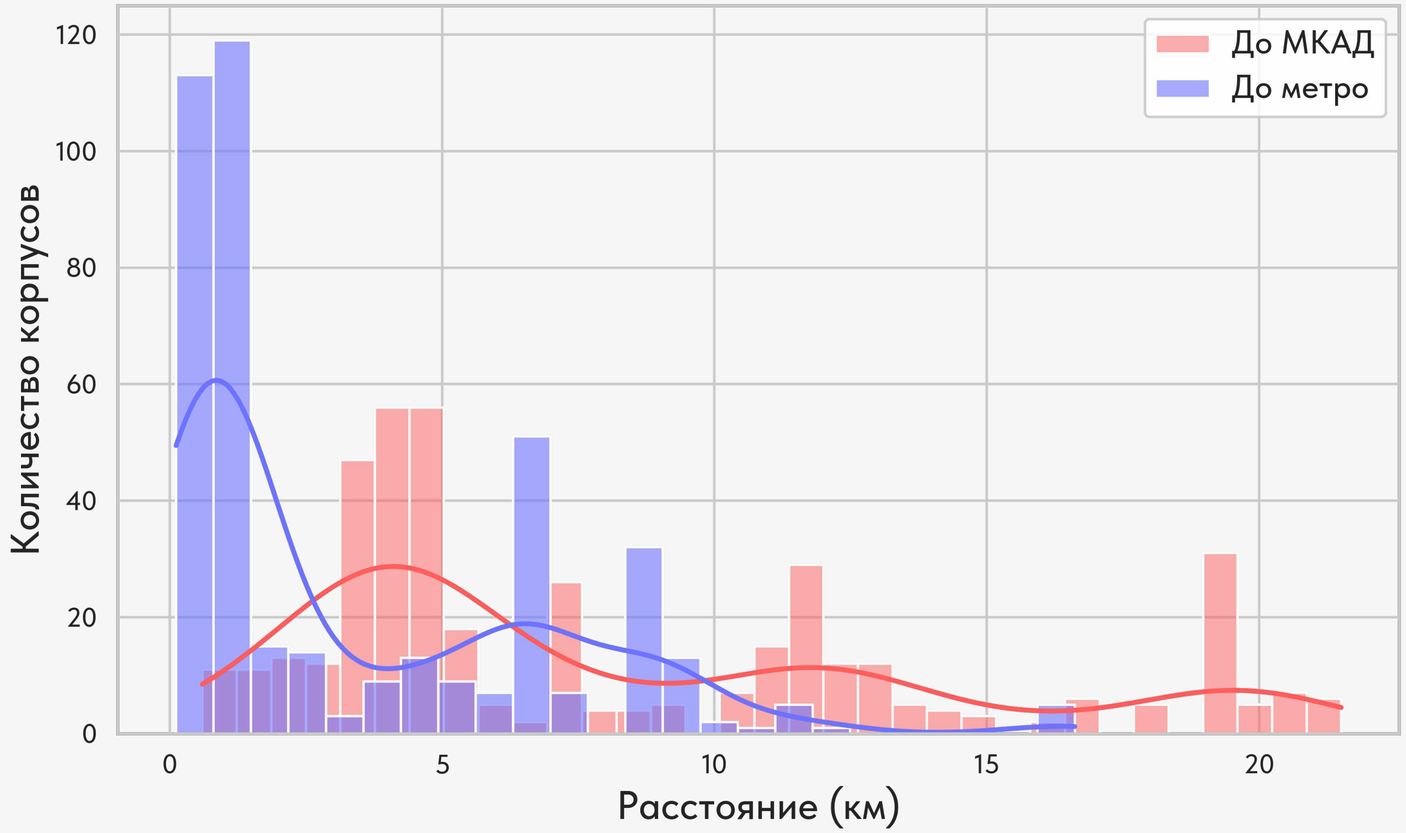
Синим цветом обозначены границы Новой Москвы

Белым обозначены ЖК

Заметим, что вся Новая Москва находится за МКАДом, и ЖК неравномерно располагаются относительно неё

Создание новой переменной

Расстояние ЖК до МКАД и метро измеряется в километрах



МКАД	Метро
8 среднее	3 среднее
5 медиана	1 медиана
6 стандартное отклонение	4 стандартное отклонение

ПРЕДАНАЛИЗ

Обработка датасета

Удалим пропущенные и нерелевантные для нас значения

1. **Ещё не сданные корпуса**
   Удаляем, поскольку одна из наших основных переменных содержит данные о сдаче ЖК

2. **Пропущенные значения в переменной плановая дата РВЭ**
   Удаляем, поскольку эта переменная понадобится нам при анализе

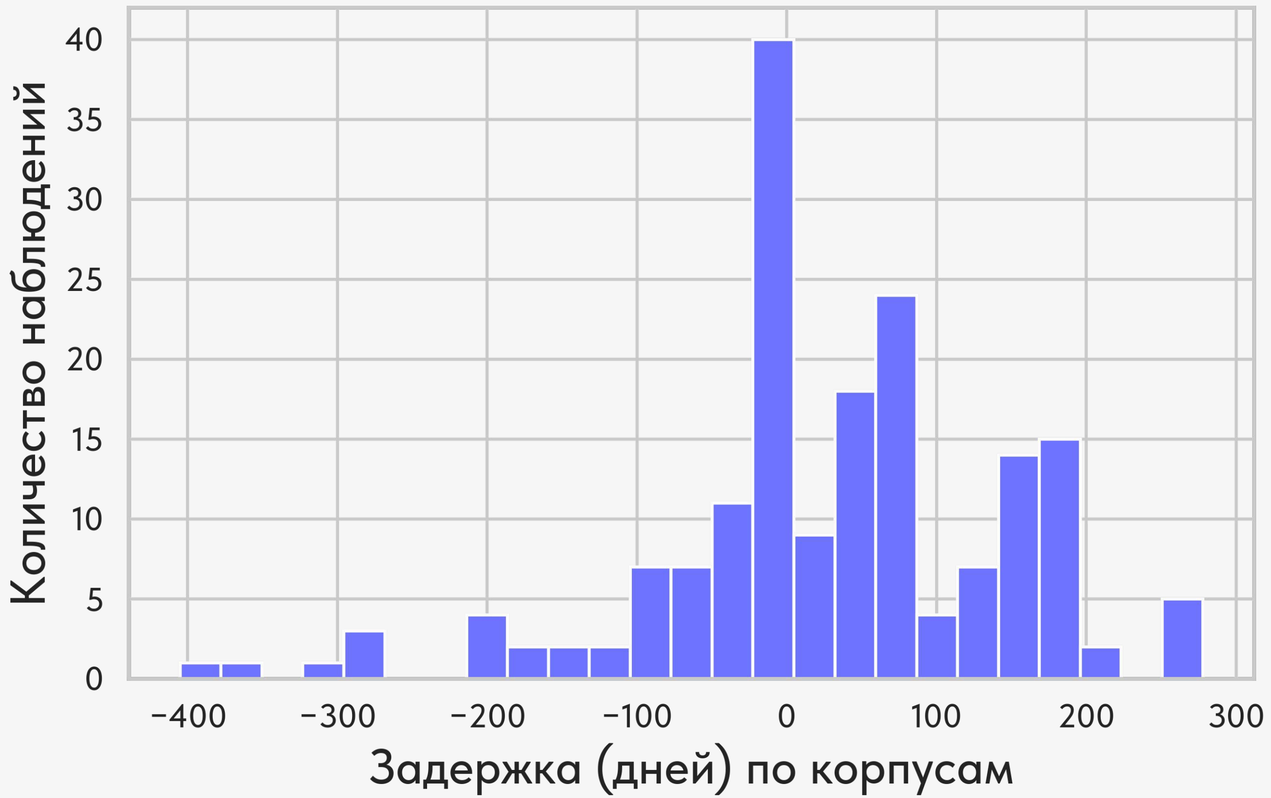
3. **Расстояние до МКАД больше 19 километров**
   Очень далекие районы, которые только начали развиваться[1]

419 корпусов → 179 корпусов

1. информация от составителя нашего датасета

Создание новой переменной

Задержка сдачи корпуса в днях
разница между фактической и планируемой датой ввода в эксплуатацию



31
среднее

32
медиана

116
стандартное отклонение

отрицательная задержка значит, что дом сдали раньше, чем планировалось

54%
Задержанных домов

Наши предположения

Мы думаем, что средняя цена за квадратный метр может влиять на задержку, потому что:

1 **Высокая цена**

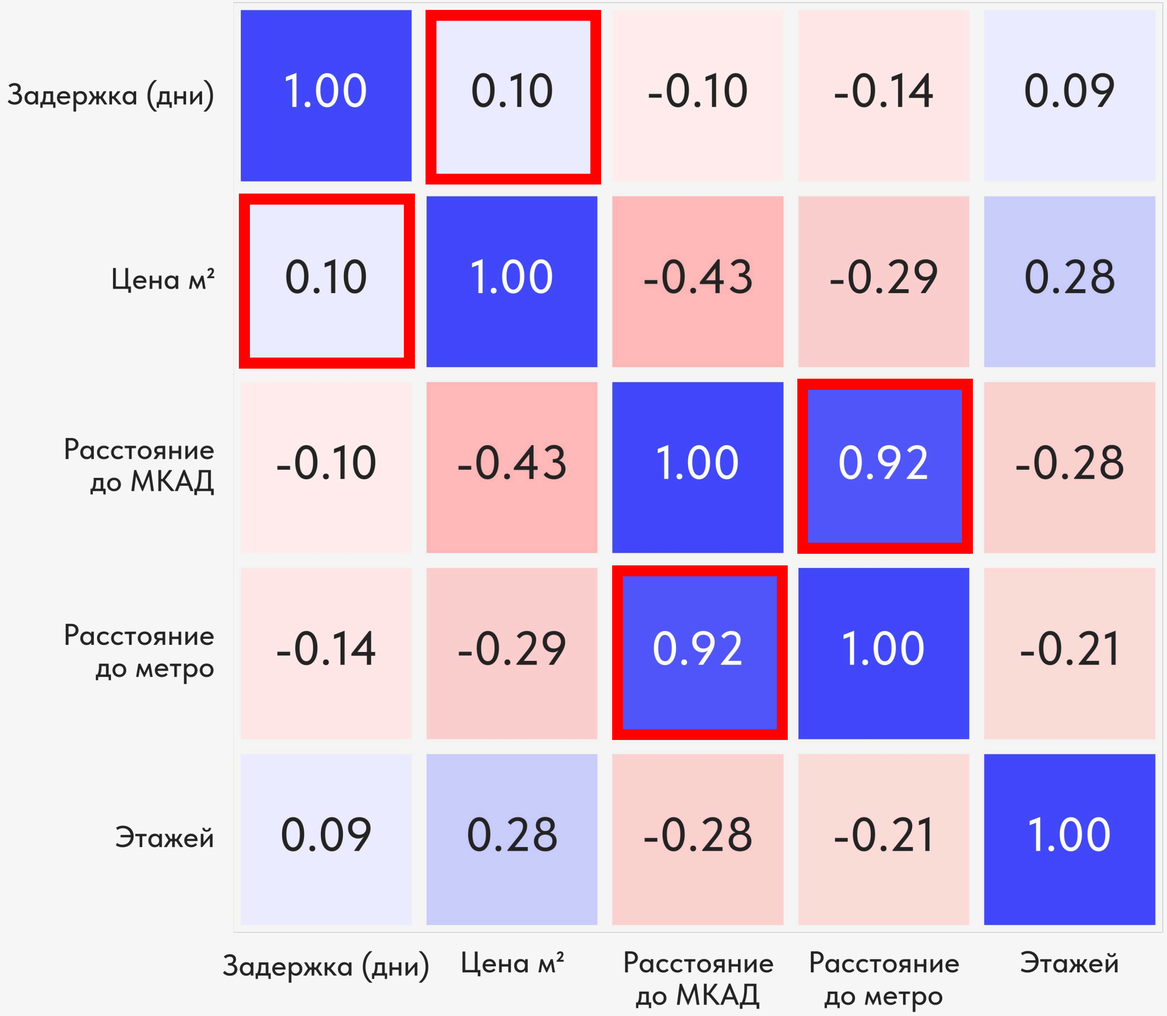
→ признак хорошей репутации застройщика
→ мотивация сдать объект в срок, чтобы получить деньги быстрее

2 **Низкая цена**

→ возможна нехватка бюджета
→ застройщик вынужден экономить, что ухудшает организацию строительства

Но влияет ли она на самом деле?

Взаимосвязь переменных



Слабая корреляция между ценой за квадратный метр и задержкой

Сильная корреляция между расстоянием до МКАД и расстоянием до метро

Средняя корреляция между расстоянием до МКАД и количеством этажей

Выводы из преданализа

В датасете много несданных корпусов ЖК
Их чистка сделала наш датасет сильно меньше по сравнению с изначальным

Корреляция между расстоянием до метро и до МКАД высокая
Это говорит о том, что эти переменные взаимозаменяемы

Корреляция задержки с расстоянием до МКАД -0.1
Знак у корреляции отрицательный, что кажется контринтуитивным

↓

Но почему мы считаем это контринтуитивным?

ОБОСНОВАНИЕ

Логическое обоснование

Удалённость от МКАД коррелирует с недостаточной развитостью коммуникаций[1] [2]

Подключение к каждому из видов ЖКХ требует отдельных технических условий, согласований и работ

Чем дальше объект, тем выше вероятность, что мощности существующих сетей недостаточно

Это вынуждает застройщика финансировать и организовывать строительство новых сетей

Любые задержки по коммуникациям срывают последующие этапы и сдачу объекта

Повышается задержка объекта

1. TRACC: при отдолении от ближайшего шоссе, ухудшаются дороги и падает уровень транспортной доступности
2. International Journal of Geography & Environmental Management: при уменьшении транпортной доступности возрастает сложность проведение коммунальных систем
3. IRN: коммуникации часто становятся причиной задержки

Иследовательский вопрос

Что влияет на задержку корпуса ЖК при его сдаче?

Гипотеза

Чем дальше от МКАДа расположен корпус ЖК, тем больше задержка его ввода в эксплуатацию

МАТМОДЕЛЬ

Теоретическое обоснование выбора модели

Выбрана линейная регрессия, так как она помогает нам отследить влияние сразу нескольких количественных признаков на количественную целевую переменную

Принимаемый уровень значимости:
0.05

Предикторы для регрессии:

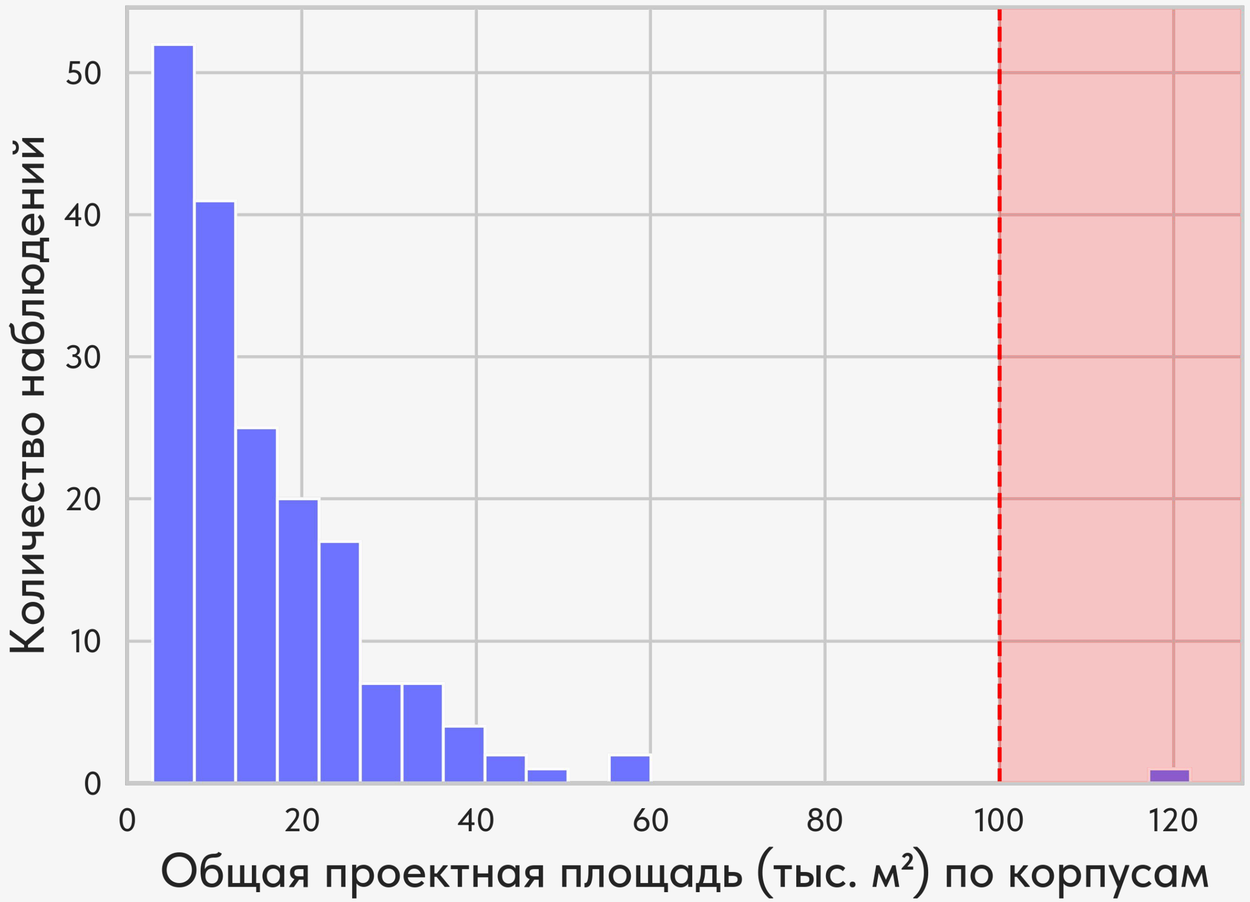
- Расстояние до МКАДа
- Этажи
- Цена за кв. метр
- Общая проектная площадь
- One-hot encoding по крупным застройщикам

контрольные переменные, которые также могут сильно повлиять на задержку

Все предикторы мы сначала стандартизировали

АНАЛИЗ

Распределение проектных площадей



Удаляем значения с общей проектной площадью выше 100 000 кв. метров (по 3IQR)

179 корпусов

↓

178 корпусов

после очистки:

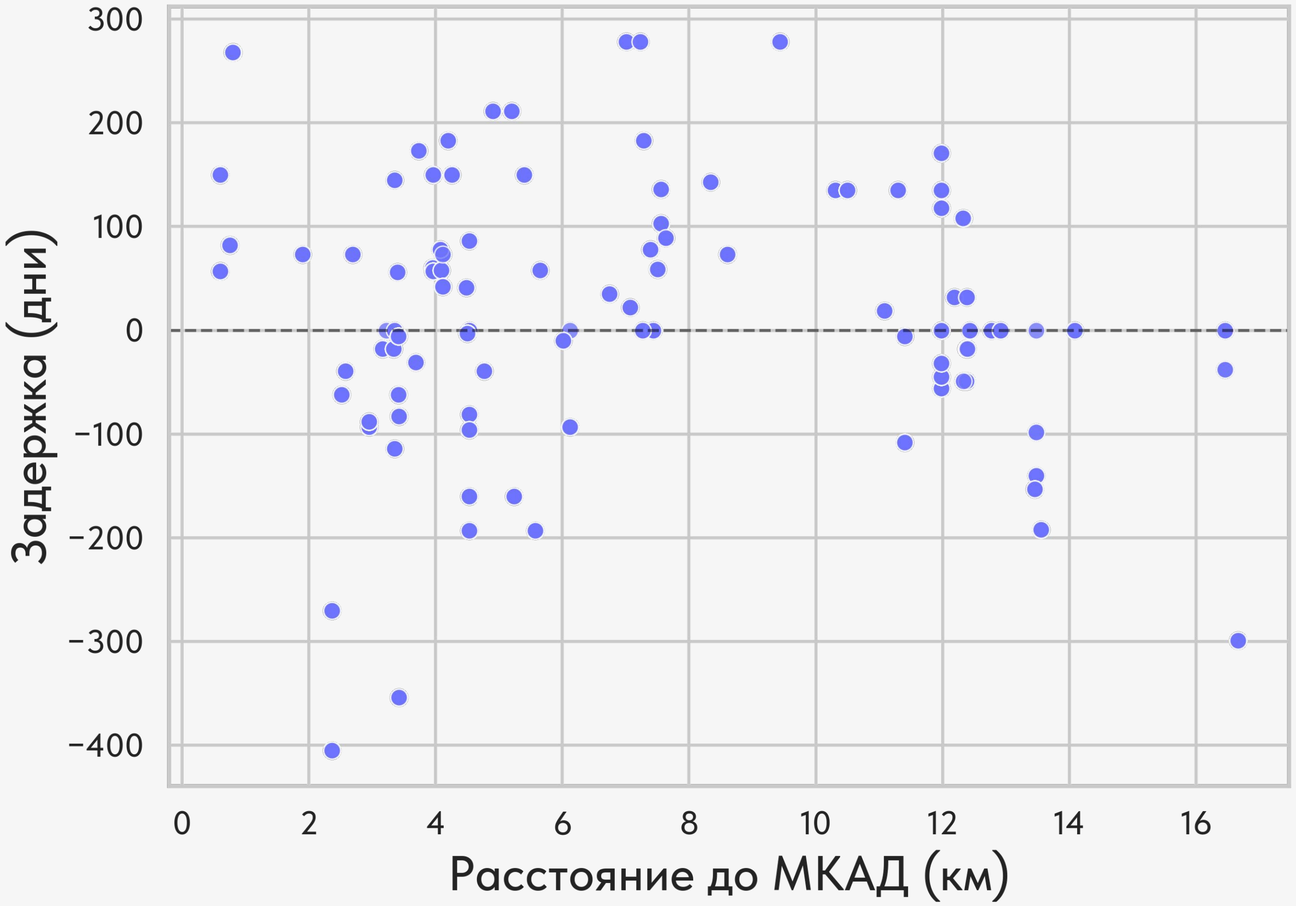
15
среднее

11
медиана

11
стандартное отклонение

АНАЛИЗ

Анализ взаимосвязи задержки и расстояния



Мы построили график зависимости задержки от расстояния до МКАДа

↓

Видим, что зависимость квадратичная

↓

Добавляем в линейную регрессию квадрат расстояния до МКАДа

МАТМОДЕЛЬ

Математическая модель

Признаки	Коэффициент	p-value	Значимость
const	30.97	0.00	✓
Цена за кв. метр	-8.32	0.35	⚡
Расстояние до МКАДа	11.87	0.36	⚡
Квадрат расстояния до МКАДа	-37.74	0.00	✓
Этажей	18.44	0.03	✓
Общая проектная площадь	-21.38	0.02	✓
ПИК	-75.53	0.00	✓
A101	-21.85	0.15	⚡
Самолет	-18.61	0.05	⚡

One-hot по крупным застройщикам

Зависимость задержки от расстояния до МКАДа имеет квадратичную зависимость

Проверка результатов линейной регрессии

VIF по переменным в рамках нормы, что говорит об отсутствии мультиколлинеарности между предикторами

Признаки	VIF
Цена за кв. метр	1.60
Расстояние до МКАДа	3.40
Квадрат расстояния до МКАДа	2.17
Этажей	1.42
Общая проектная площадь	1.59

4.77
Cond. No.

Cond. No. говорит о том, что в модели в целом отсутствует мультиколлениарность

Показатель Durbin-Watson для автокорреляции: ≈2 → В модели отсутствует автокорреляция

Вывод: критерии для построения модели были выполнены

МАТМОДЕЛЬ

Проверка на устойчивость по годам сдачи

Не проверяем устойчивость на годах до 2023, так как там всего около 10 корпусов

Год	Кол-во наблюдений	Коэффициент при квадрате расстояния	p-value	Значимость
2023	47	-6.43	0.65	
2024	65	-129.26	0.00	✓
2025	57	-64.07	0.00	✓

Проверка на устойчивость подтвердилась на всех группах, кроме 2023

Почему на 2023 году не подтвердилось?

В 2023 году мораторий[1] на взыскание неустойки
с застройщиков действовал только первые полгода

↓

Застройщики были гораздо больше заинтересованы[1]
в том, что бы не задерживать какие-либо проекты

1. РБК Недвижимость: отмена моратория повысит дисциплину застройщиков

Проверка на устойчивость по конструкции

Мы разделили на группы по типу конструкции объекта

Группа	Кол-во наблюдений	Коэффициент при квадрате расстояния	p-value	Значимость
Монолит-блоки	131	-17.42	0.047	✓
Другое	48	-106.45	0.000	✓

Проверка на устойчивость подтвердилась на всех группах

Интерпретация

Гипотеза подтвердилась частично

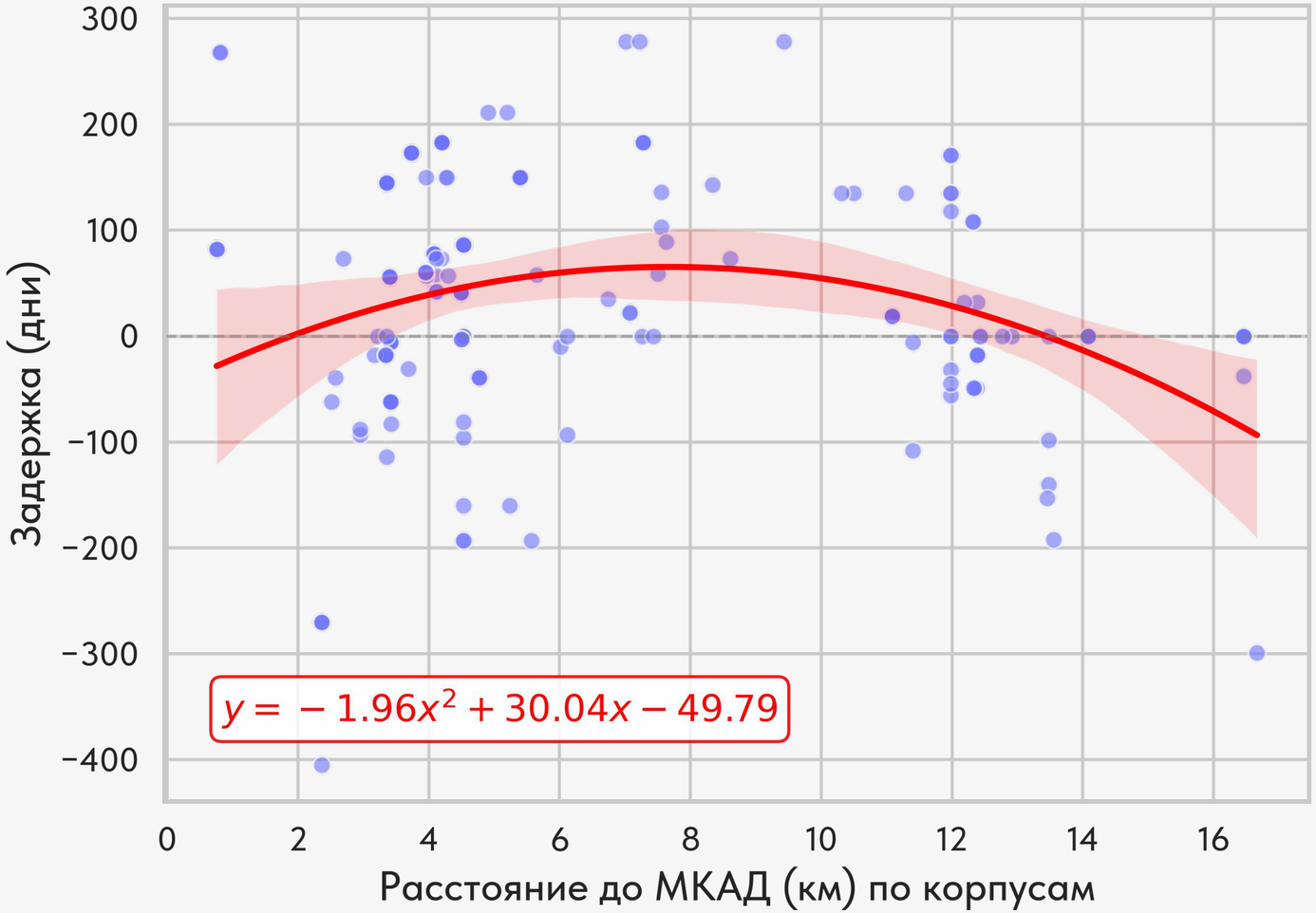
При отдалении от МКАД задержка сдачи объекта возрастает, а после своего максимума снижается
Задержка сдачи зависит от расстояния до МКАД квадратично

Проверка на устойчивость пройдена по всем годам сдачи ЖК, кроме 2023, а также на всех группах по типу конструкции объекта

То, что проверка в 2023 не пройдена, обусловлено отсутствием моратория на задержки

АНАЛИЗ

Уточнение зависимостей



Мы построили линейную регрессию без контрольных переменных и стандартизации

↓

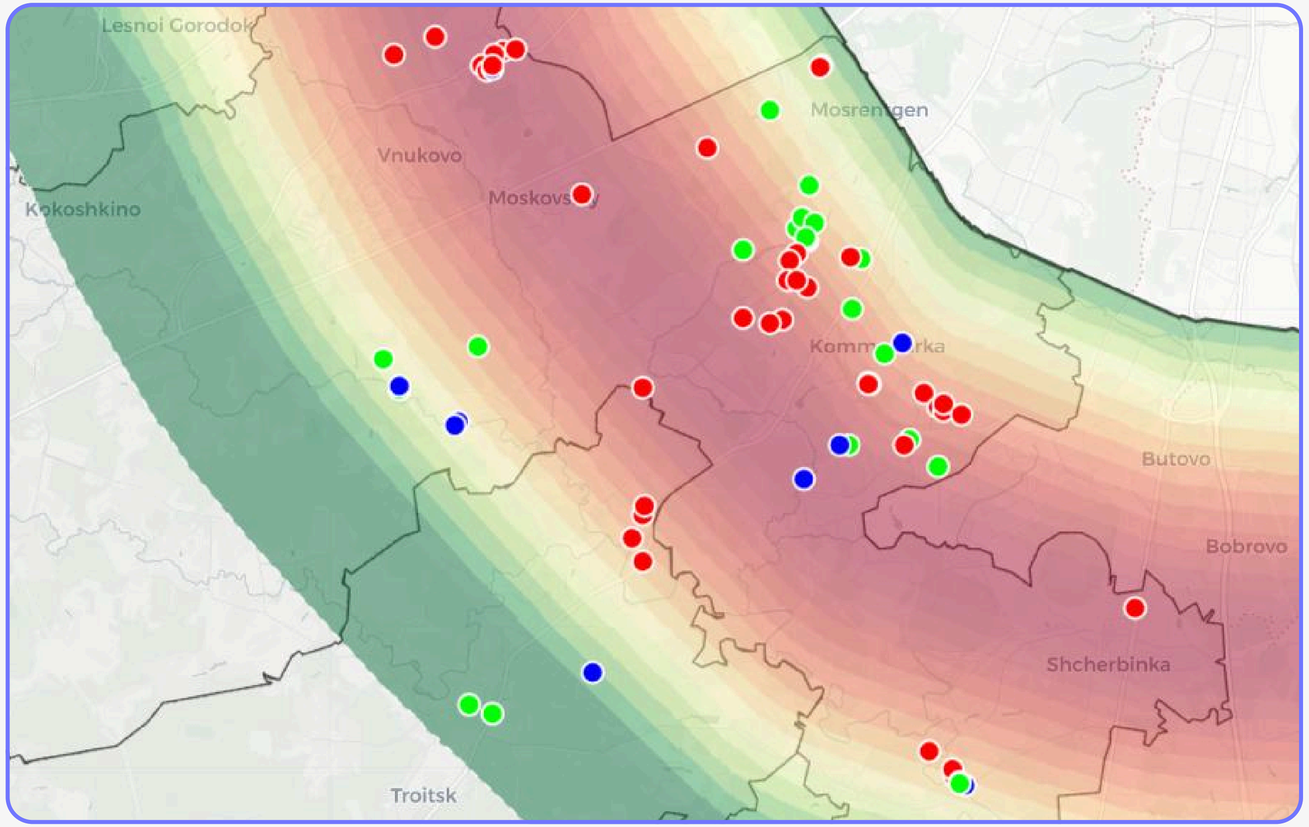
Получили уравнение зависимости задержки от расстояния

Это нужно нам для визуализации зависимости

АНАЛИЗ

Визуализация зависимости

Мы нанесли на карту ЖК и задержку по зонам: красный цвет означает большую задержку, а зеленый – меньшую



Визуально можно выделить три зоны:

- около МКАДа: задержка небольшая
- средняя зона: задержка большая
- ближе к Троицку: задержка небольшая

Альтернативное направление влияния

Зона около МКАДа

Ближе к Старой Москве

↓

Лучше транспортная доступность[1]

↓

Больший спрос от покупателей, для которых важна доступность к Москве

Зона около Троицка

Нет выраженных экологических проблем[2]

↓

Больший спрос от покупателей, которые работают в этом же районе или семей, которым важна экологическая ситуация

↓

Благодаря достаточному финансированию от банков у застройщиков не возникает проблем с тем, чтобы сдавать ЖК в срок

1. irn.ru: Транспортная доступность новых территорий Москвы
2. Новый очаг: плюсы и минусы Троицка

Альтернативное направление влияния

Территория между Троицком и МКАДом

В этой зоне возникает район, с неразвитой городской инфраструктурой и плохой экологической составляющей

↓

В данной зоне ниже спрос от покупаетелей

↓

При низком спросе у застрощика возникают проблемы[1] с наполнением денег с эскроу-счетов, из-за чего появляются ограничения на кредиты

↓

Возникают задержки сдачи ЖК

1. Коммерсант: девелоперы пытаются не допустить затоваривания и снижения цен

Релевантные исследования

IRN: От трети до половины новостроек в Москве сдается с просрочками

1

Расстояние от МКАДа влияет на задержки в сдаче

В нашем исследовании это также подтверждается. В статье, получено положительное влияние расстояния на задержку, однако говорится лишь про расстояния до 5 км и в зоне 30 км от МКАДа. Наше же исследование охватывает расстояние до 19 км

2

Разница между задержкой квартир массового сегментта и премиум сегмента не велика

В нашем исследовании это отражается отсутствием значимости средней стоимости квартир в ЖК на влияние задержки

3

Нехватка поступлений от дольщиков и банков является одними из ключевых в задержке

Мы используем этот тезис для качественной интерпретации получившихся результатов в нашей модели

ВЫВОДЫ И РЕКОМЕНДАЦИИ

Policy implication

Для дольщика

Использование информации о влиянии расстояния объекта до МКАДа на задержку позволит более корректно оценивать сроки сдачи квартиры. Это позволит покупателям планировать свои финансы, ведь им, скорее всего, придётся какое-то время платить и за аренду, и за новую ипотеку

Для банка

При оценке девелоперских проектов использовать дифференцированную ставку дисконтирования, в которой премия за риск зависит по мере удаления объекта от МКАД, чтобы корректно учитывать снижение спроса и дополнительные рыночные риски таких локаций

Для государства

Ужесточение контроля за ресурсоснабжающими организациями, особенно в "зоне сильных задержек" между Троицком и МКАД, ведь зачастую проблема задержки обусловлена незаинтересованностью ресурсоснабжающих компаний делать все в срок [1]

1. Ведомости

Ограничения



Нет данных, почему произошла задержка
Например, на стадии подключения коммуникаций или завозе материалов



Нет данных по ЖК, в которых все квартиры уже проданы
Возможно, тенденция в ЖК, которые уже продали квартиры, отличаются от найденных нами

ЗАКЛЮЧЕНИЕ

Наша команда



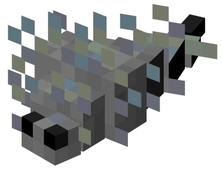
Лука
Дуванов



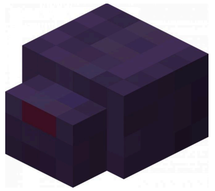
Матвей
Мерзликин



Аня
Тихомирова



Илья
Власов



Руслан
Сафарян

ПРИЛОЖЕНИЕ

а ничо тот факт что

Манна-Уитни band

ПРИЛОЖЕНИЕ

Почему приведено мало релевантных исследований?

Общие исследования по задержкам в России или даже по Московскому региону часто "растворяют" специфику Новой Москвы в более крупных массивах данных

Мы анализируем "микросрез" более общей проблемы. Наша работа является их углублением и конкретизацией многих исследований про проект. Мы берем известные макрофакторы задержек и проверяем, как один из них проявляется в условиях "чистого эксперимента", каким является Новая Москва

ПРИЛОЖЕНИЕ

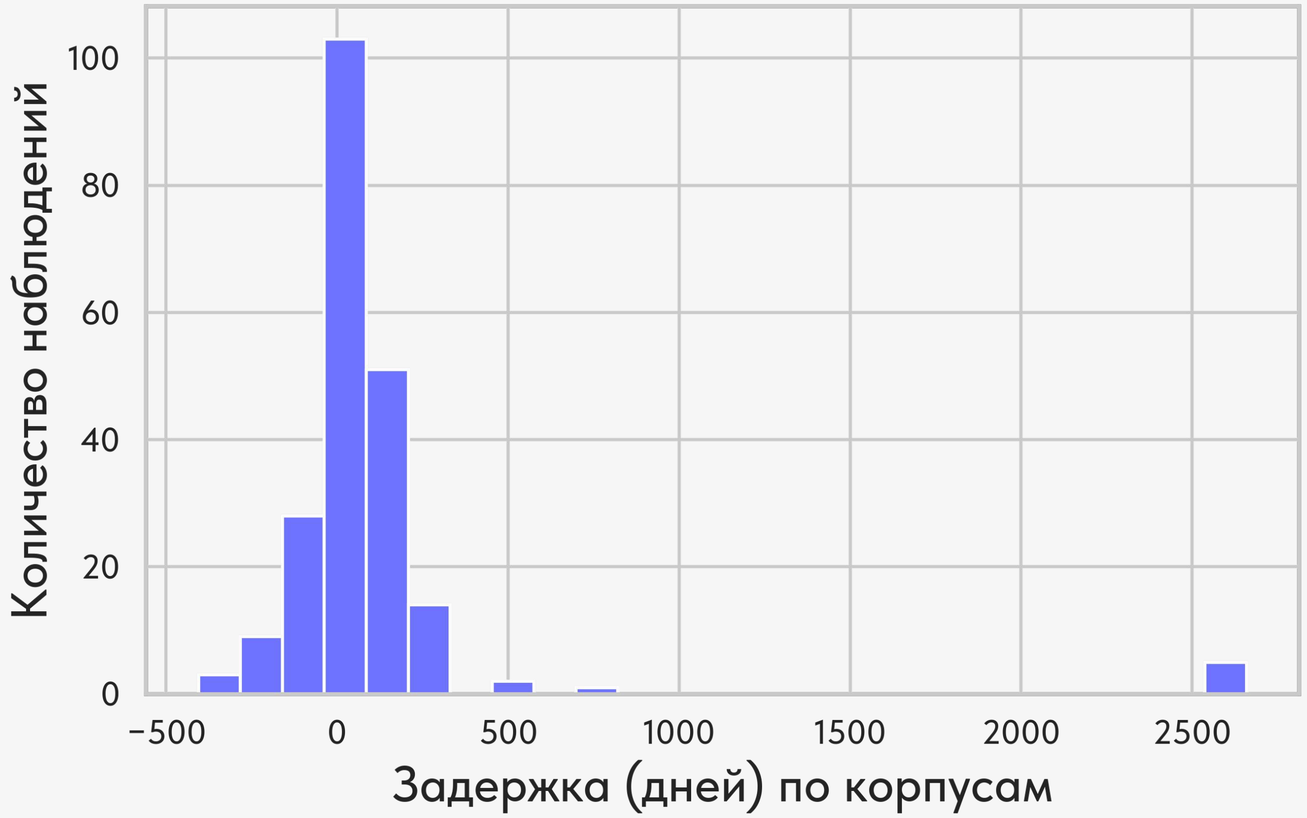
Распределение по датам





ПРИЛОЖЕНИЕ

Задержка сдачи ДО очистки



108
среднее

41
медиана

409
стандартное отклонение

ПРИЛОЖЕНИЕ

Влияние этажности дома

Этажность дома может оказать положительное влияние на задержку домов https://www.ijirr.com/sites/default/files/issues-pdf/3248_0.pdf

ПРИЛОЖЕНИЕ

проверка на устойчивость по годам

2023:

	coef	std err	t	P>|t|	[0.025	0.975]
const	-13.1738	18.205	-0.724	0.474	-50.027	23.679
Расстояние до МКАД	-22.6948	16.331	-1.390	0.173	-55.756	10.366
Общая проектная площадь	-5.8181	11.884	-0.490	0.627	-29.875	18.239
Цена квадратного метра	1.2815	14.156	0.091	0.928	-27.375	29.938
Этажей до	-20.2128	13.686	-1.477	0.148	-47.919	7.493
Самолет	-8.6214	15.567	-0.554	0.583	-40.135	22.892
ПИК	-82.8704	17.105	-4.845	0.000	-117.498	-48.243
A101	-7.5556	19.568	-0.386	0.702	-47.169	32.058
Квадрат расстояния до МКАД	-7.0602	15.445	-0.457	0.650	-38.328	24.207

2024:

	coef	std err	t	P>|t|	[0.025	0.975]
const	176.9834	40.664	4.352	0.000	95.523	258.444
Расстояние до МКАД	176.2204	37.188	4.739	0.000	101.725	250.716
Общая проектная площадь	-21.8555	14.477	-1.510	0.137	-50.856	7.145
Цена квадратного метра	27.1606	18.178	1.494	0.141	-9.254	63.575
Этажей до	16.7627	15.548	1.078	0.286	-14.384	47.910
Самолет	13.9562	21.242	0.657	0.514	-28.597	56.509
ПИК	-8.0306	21.822	-0.368	0.714	-51.745	35.684
A101	55.7335	28.994	1.922	0.060	-2.349	113.816
Квадрат расстояния до МКАД	-152.0142	39.060	-3.892	0.000	-230.260	-73.768

2025:

	coef	std err	t	P>|t|	[0.025	0.975]
const	118.0242	11.659	10.123	0.000	94.582	141.466
Расстояние до МКАД	-17.7787	16.034	-1.109	0.273	-50.016	14.459
Общая проектная площадь	4.1549	14.265	0.291	0.772	-24.527	32.837
Цена квадратного метра	10.3394	11.794	0.877	0.385	-13.373	34.052
Этажей до	-11.5166	12.487	-0.922	0.361	-36.623	13.590
Самолет	23.1031	12.593	1.835	0.073	-2.216	48.422
ПИК	-52.1992	13.984	-3.733	0.001	-80.316	-24.082
A101	-55.8544	23.948	-2.332	0.024	-104.005	-7.704
Квадрат расстояния до МКАД	-32.5680	7.382	-4.412	0.000	-47.411	-17.725

ПРИЛОЖЕНИЕ

проверка на устойчивость по конструкции

Монолит-Блоки:

	coef	std err	t	P>|t|	[0.025	0.975]
const	79.1299	11.768	6.724	0.000	55.839	102.421
Расстояние до МКАД	9.8231	13.271	0.740	0.461	-16.442	36.088
Общая проектная площадь	-21.2272	9.189	-2.310	0.023	-39.413	-3.041
Цена квадратного метра	-16.8463	9.584	-1.758	0.081	-35.815	2.122
Этажей до	28.1610	9.172	3.070	0.003	10.008	46.314
Квадрат расстояния до МКАД	-17.4200	8.691	-2.004	0.047	-34.620	-0.220

Остальное:

	coef	std err	t	P>|t|	[0.025	0.975]
const	-52.9375	13.829	-3.828	0.000	-80.845	-25.030
Расстояние до МКАД	484.3322	82.831	5.847	0.000	317.172	651.493
Общая проектная площадь	-3.6898	15.373	-0.240	0.811	-34.714	27.334
Цена квадратного метра	38.1699	16.313	2.340	0.024	5.249	71.090
Этажей до	23.0092	17.514	1.314	0.196	-12.336	58.355
Квадрат расстояния до МКАД	-432.8372	82.235	-5.263	0.000	-598.794	-266.881

ПРИЛОЖЕНИЕ

проверка на устойчивость по застройщику

Пик:

	coef	std err	t	P>|t|	[0.025	0.975]
const	-108.6071	17.201	-6.314	0.000	-144.190	-73.024
Расстояние до МКАД	400.2463	109.546	3.654	0.001	173.633	626.859
Общая проектная площадь	6.7032	17.523	0.383	0.706	-29.546	42.953
Этажей до	39.4022	24.708	1.595	0.124	-11.710	90.514
Квадрат расстояния до МКАД	-354.9624	114.221	-3.108	0.005	-591.248	-118.677

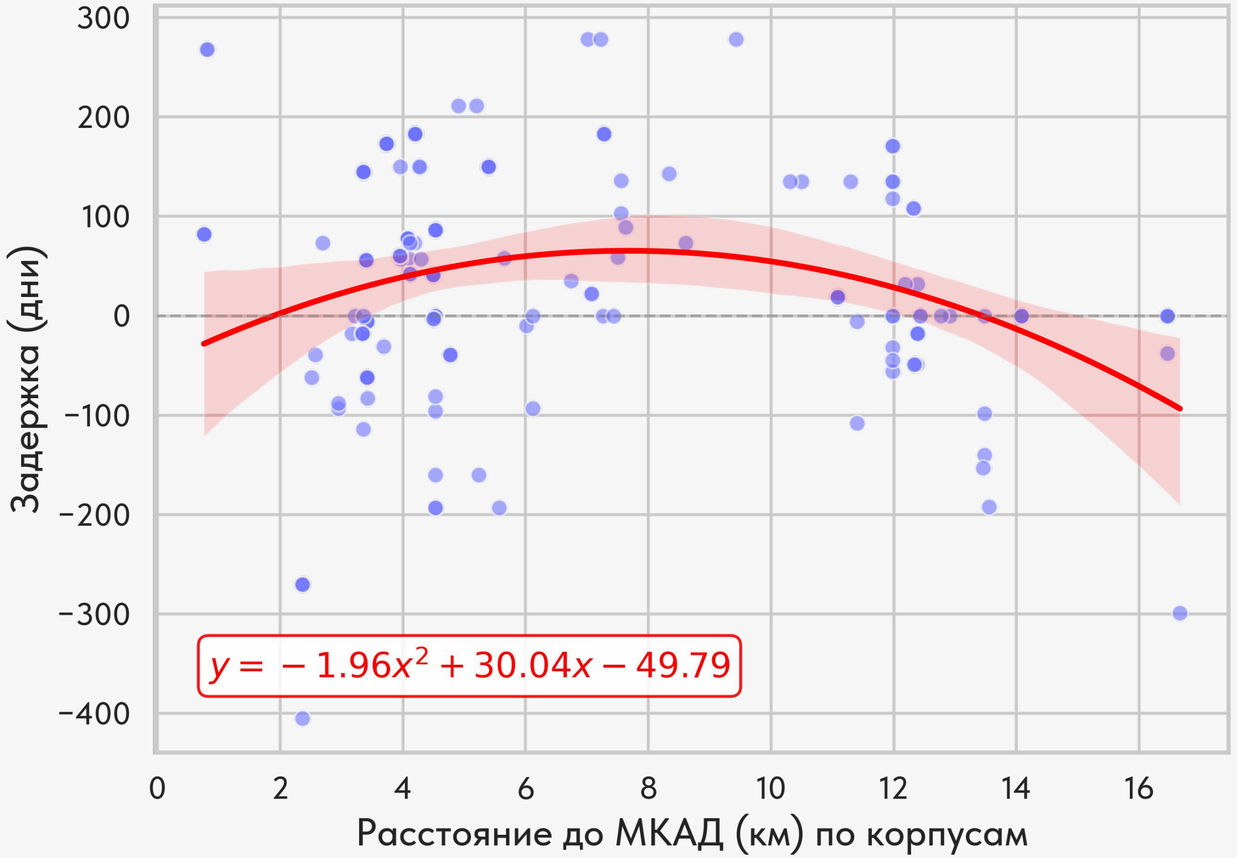
Остальные:

	coef	std err	t	P>|t|	[0.025	0.975]
const	56.8477	7.436	7.645	0.000	42.152	71.543
Расстояние до МКАД	99.0784	35.803	2.767	0.006	28.319	169.838
Общая проектная площадь	-15.7015	8.163	-1.924	0.056	-31.834	0.431
Этажей до	18.3864	8.123	2.263	0.025	2.332	34.440
Квадрат расстояния до МКАД	-114.9915	35.334	-3.254	0.001	-184.824	-45.159

У ПИК распределение похоже на параболу, решили проверить, мало ли только ПИК у нас все это создает

ПРИЛОЖЕНИЕ

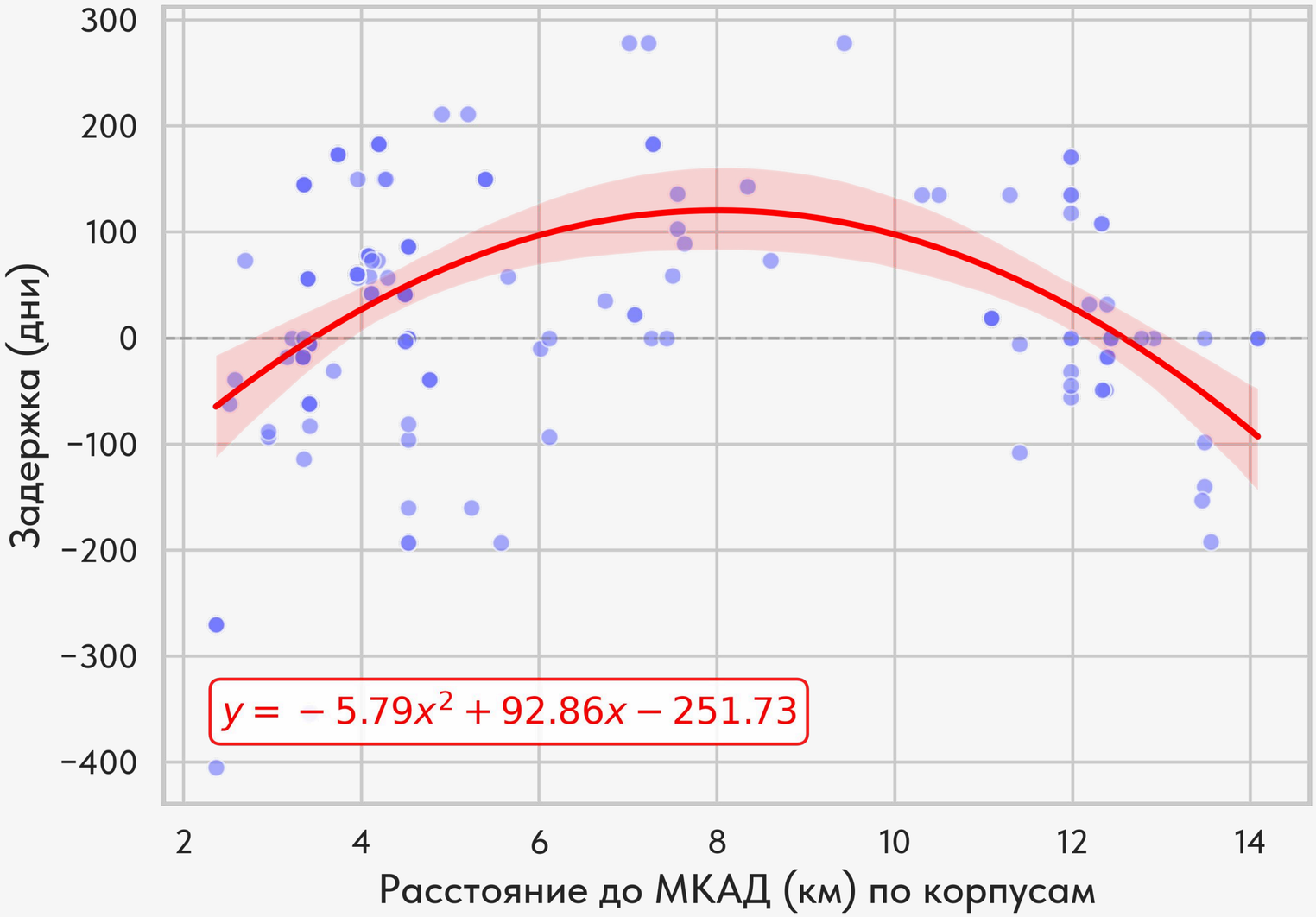
ЛИНРЕГ БЕЗ КОНТРОЛЬНЫХ ПЕРЕМЕННЫХ



	coef	std err	t	P>|t|	[0.025	0.975]
const	-49.7900	34.996	-1.423	0.157	-118.859	19.279
mkad_dist_km	30.0430	10.498	2.862	0.005	9.324	50.762
mkad_dist_km_sq	-1.9600	0.614	-3.194	0.002	-3.171	-0.749

убрали данные с расстоянием
до МКАД <1км
И больше 16 км - **6,2%**

178 → 167



ПРИЛОЖЕНИЕ

Эскроу счет

Счёт эскроу - безопасный способ оплаты для работы с застройиком. Деньги хранятся на счету, а не передаются банку

Однако, ситуация с эскроу счетом влияет на взаимоотношения застройщика и банка

Чем больше денег на эскроу счете, тем меньше процентная ставка по кредиту. Тем самым застройщику выгодно держать там больше денег. Отсутствие некоторого количества денег может задержать стройку

Застройщик не может получить средства до выполнения условий договора

1. Количество денег на эскроу счете влияет на работу с банком:

2. Застройщики часто сдвигают сдачу из-за нехватки средств на эскроу счетах

ПРИЛОЖЕНИЕ

Общая проектная площадь: боксплот



ПРИЛОЖЕНИЕ

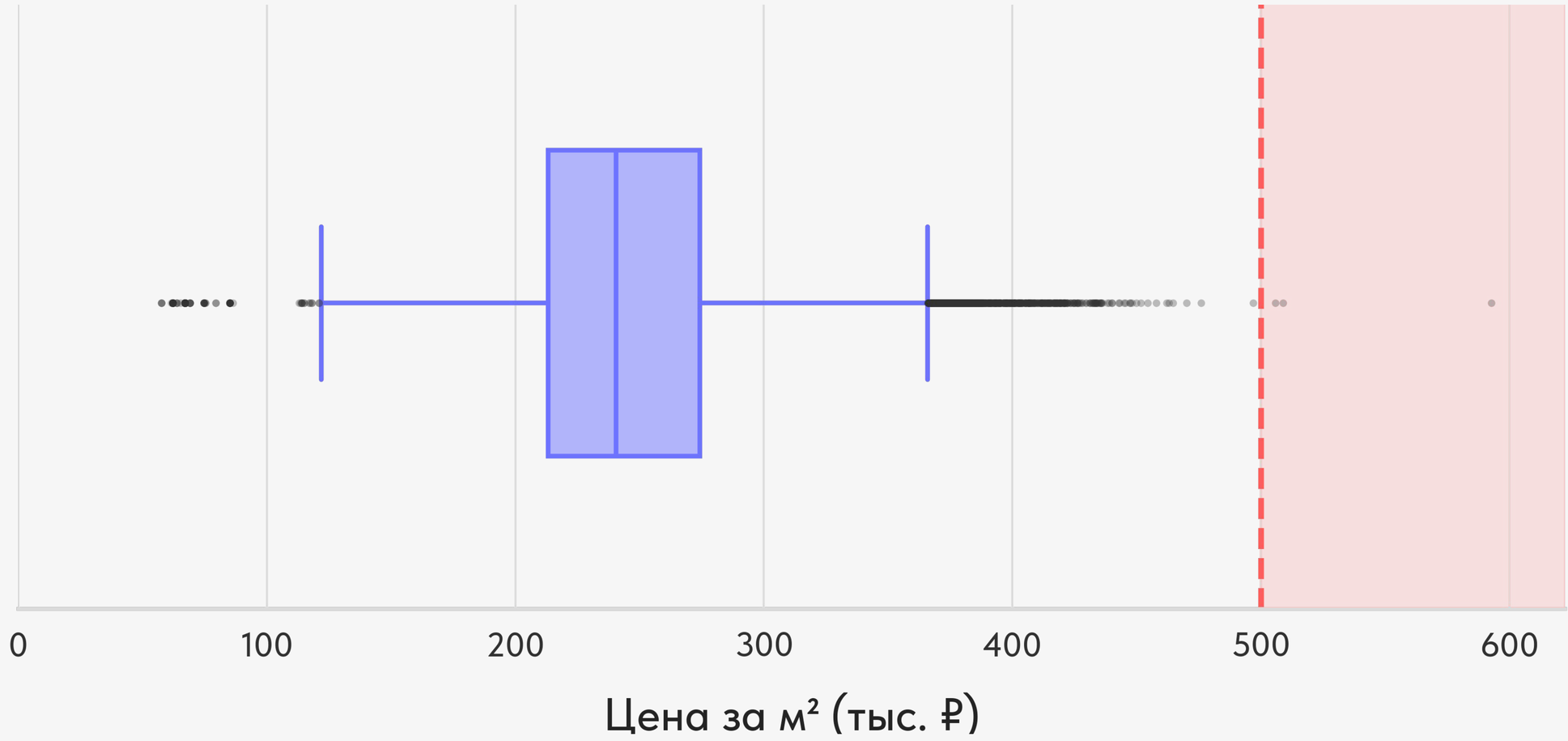
Распределение этажности по корпусам



Медиана: 15.0,
Среднее: 15.03
Ст. отклонение: 3.27

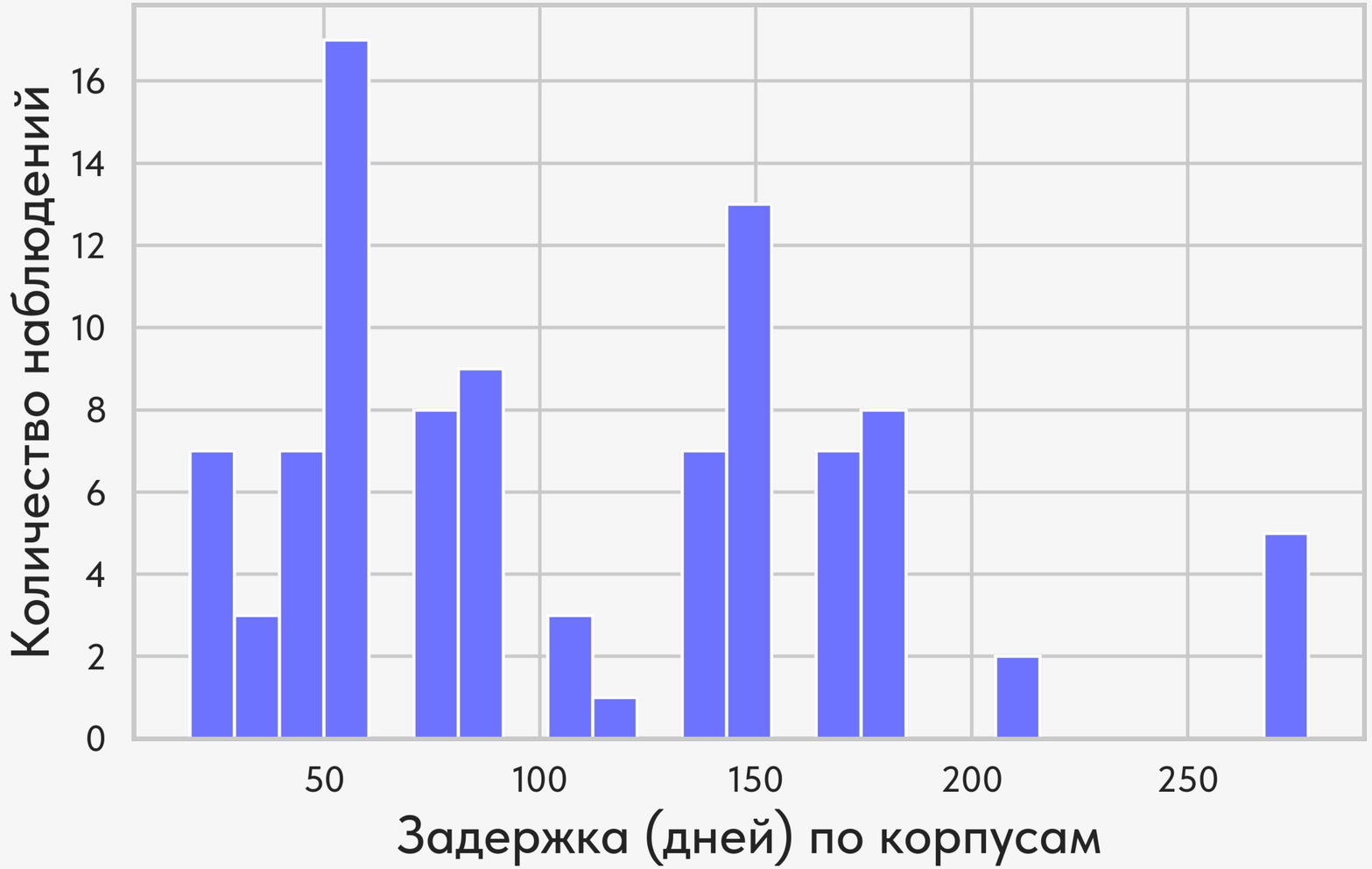
ПРИЛОЖЕНИЕ

Цена за кв. метр: боксплот



ПРИЛОЖЕНИЕ

Задержка сдачи (только положительная)



110
среднее

86
медиана

66
стандартное отклонение

ПРИЛОЖЕНИЕ

Как мы считаем задержку ввода в эксплуатацию?

1. Находим самый ранний план застройщика (“Дата ввода в эксплуатацию” в первой сделке в корпусе)
2. “Плановая дата РВЭ” в домах, введёных в эксплуатацию - фактическая дата ввода в эксплуатацию
3. Задержка = “Плановая дата РВЭ” - Первая “Дата ввода в эксплуатацию”

Durbin-Watson

$$d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

e_t - остатки

Condition number

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)},$$