

# Исследование недвижимости

Контрольная точка №2 по Анализу данных

Серкибаев Аким, Э-2209

# Знакомство с данными

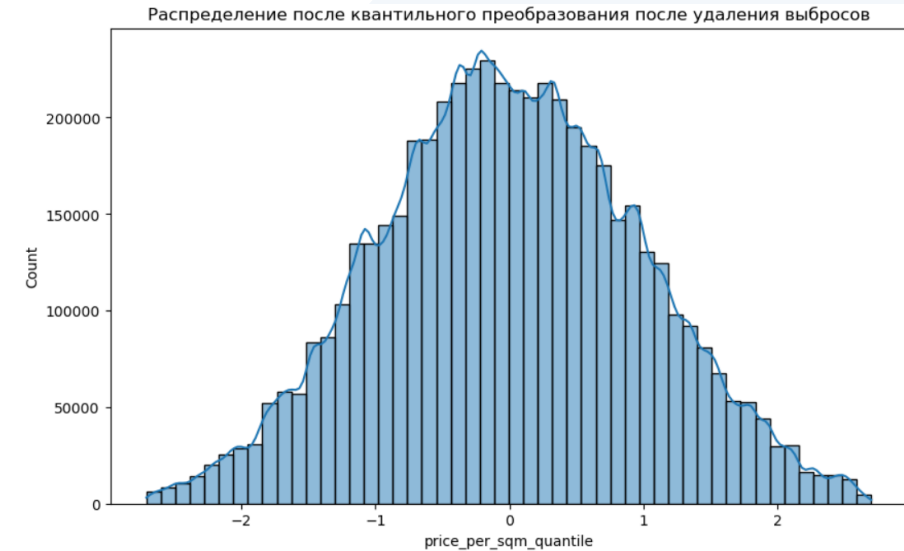
Выполнена предобработка двух датасетов, они были объединены

Проверены логические ошибки данных:

- Площадь кухни > общей площади — строки удалены.
- Номер этажа > общего количества этажей — заменён на максимальный.
- Объекты с 0 или отрицательными этажами отсутствуют.
- Минимальная площадь квартиры — 6.5 кв.м. Строки с меньшими значениями удалены, завышенные площади обработаны с учётом выбросов по цене за кв.м.

Удалены выбросы:

- Использовано квантильное преобразование (лучше всего приближено к нормальному распределению).
- Выбросы по стоимости за кв.м. удалены по 1.5 IQR, составив 0.7% данных.



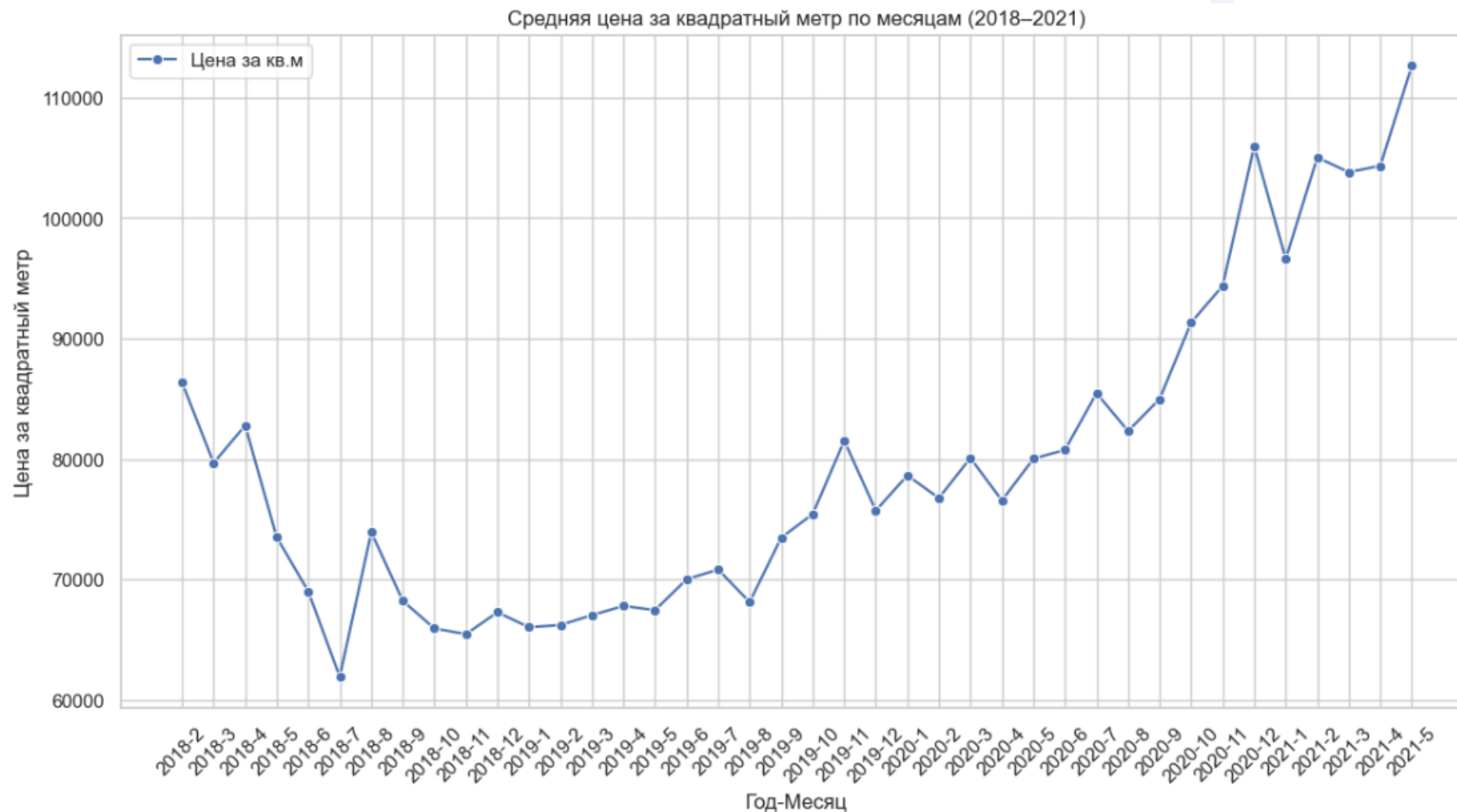
# Восстановление пропусков в столбце с площадью кухни

**LightGBM** – это градиентный бустинг, оптимизированный для скорости и больших объемов данных. Он строит деревья по принципу leaf-wise, что позволяет быстро находить наиболее важные разбиения. Модель устойчива к выбросам, поддерживает многопоточность и категориальные признаки. Однако для максимальной производительности требуется тщательная настройка гиперпараметров

**Тренд модели:** корень из ошибки уменьшался, что свидетельствует об улучшении качества модели

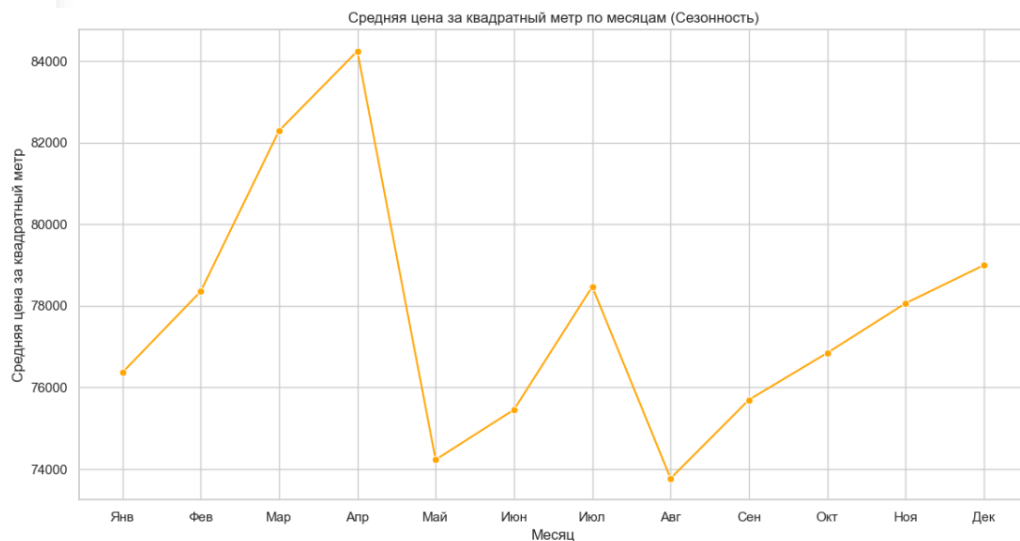
**Итоговый RMSE** на валидационных данных равен **0.238218**

# Когда выгодно покупать / продавать объект недвижимости?



В 2018–2019 годах цены снижались и были стабильны, затем в 2020–2021 наблюдался резкий рост из-за пандемии, изменений в экономике, повышенного спроса на жильё, низких ставок и инфляции

# Когда выгодно покупать / продавать объект недвижимости?



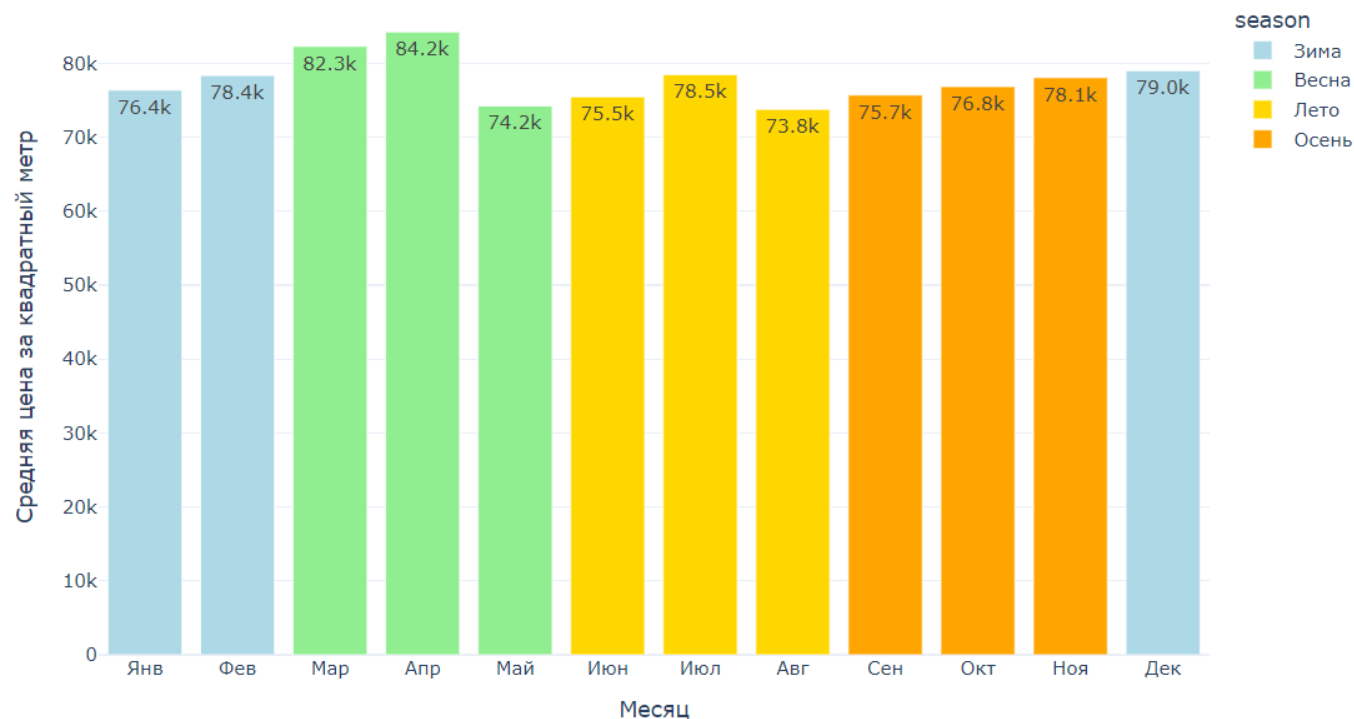
**Продавать** выгодно весной, особенно в марте-апреле, когда цены достигают пика.

**Покупать** лучше в августе, когда цены минимальны из-за низкой рыночной активности.

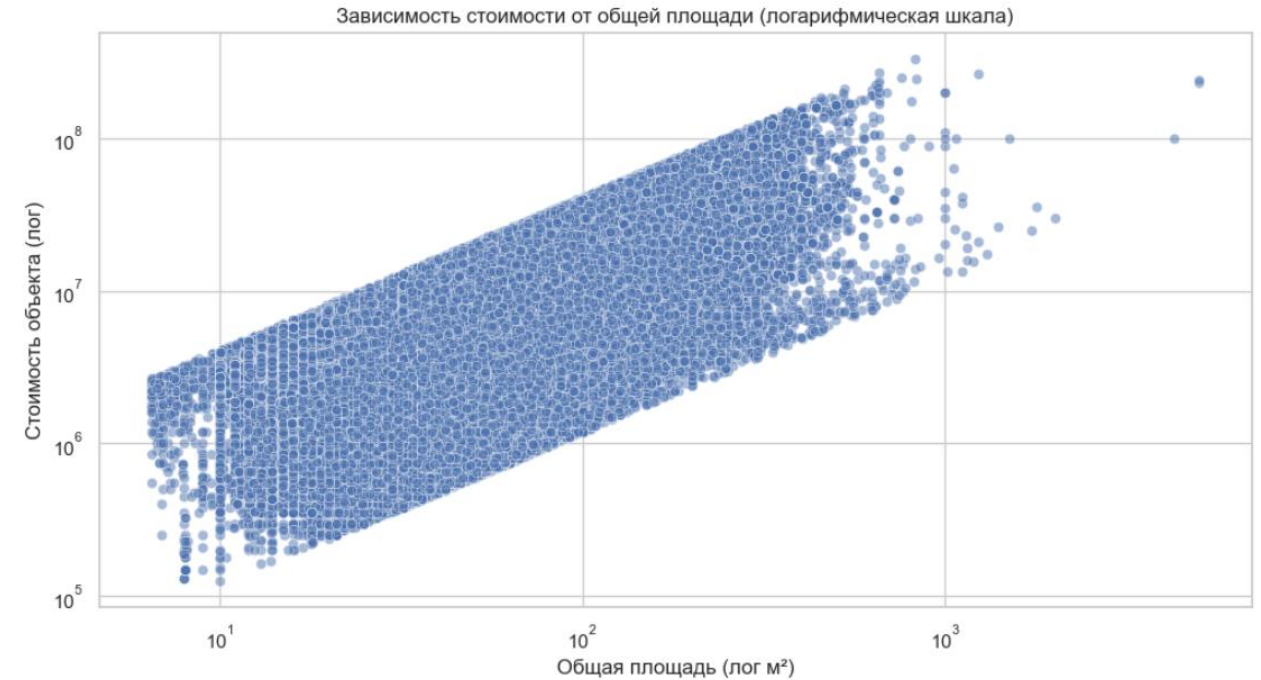
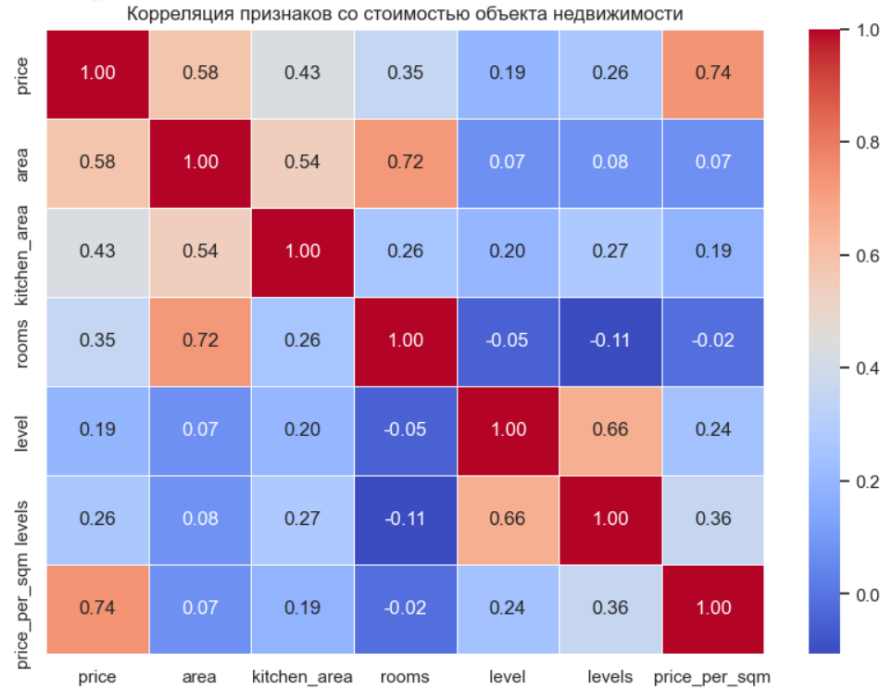
**Весна** — лучшее время для продажи, **конец лета** — для покупки.

Цены на квадратный метр достигают пика весной (март-апрель), снижаются летом, особенно в августе, и остаются на уровне 76–78 тыс. Зимой, осенью и летом, при этом весна выделяется самой высокой средней стоимостью — около 81 тыс.

Средняя цена за квадратный метр по месяцам с разбивкой по сезонам



# От каких признаков зависит стоимость объекта недвижимости?

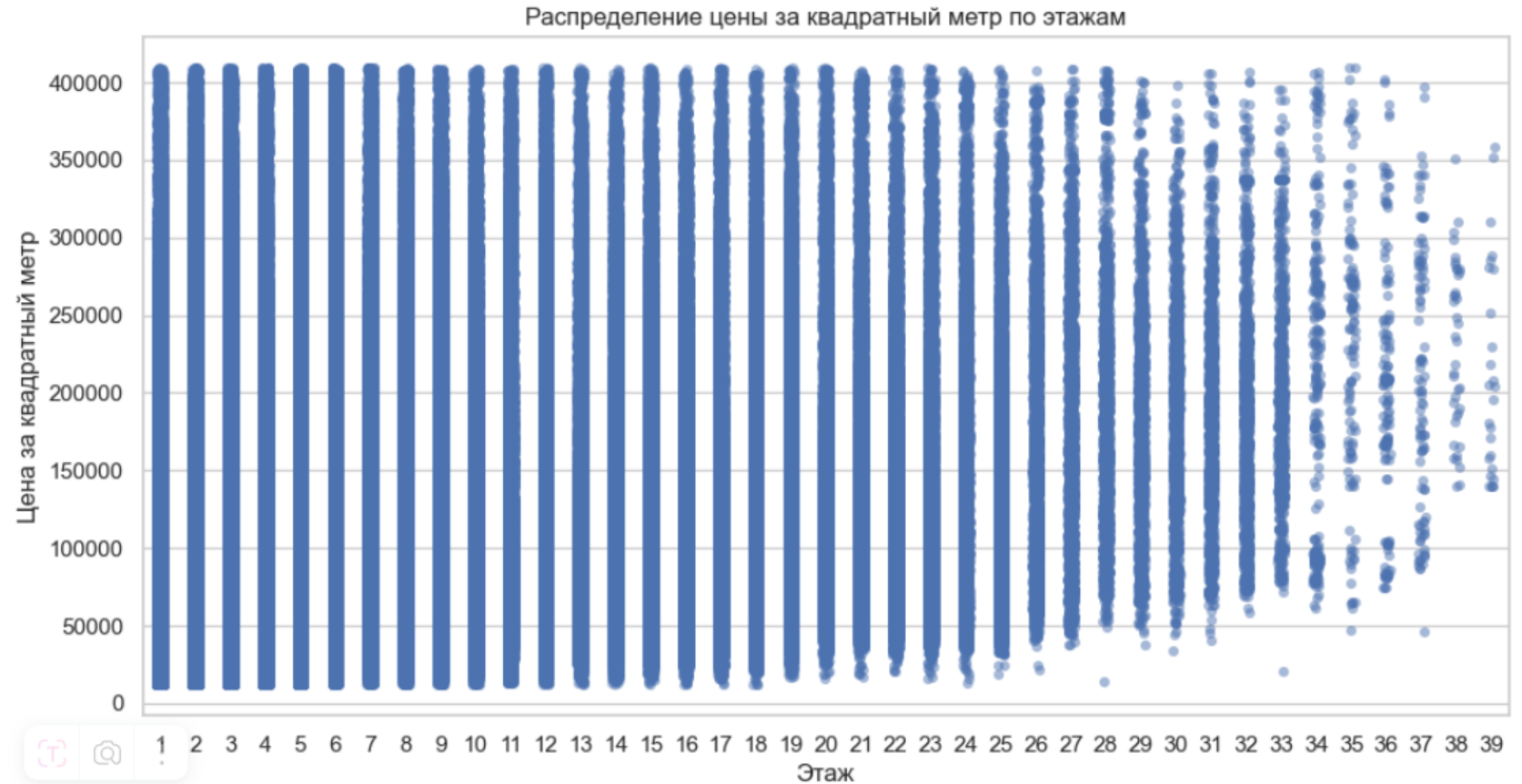


Большинство объектов имеют площадь до 300 м² и низкую стоимость, но редкие выбросы с площадью до 5000 м² и высокой стоимостью связаны с элитной недвижимостью. В логарифмическом масштабе видна сильная положительная зависимость между площадью и стоимостью, с небольшим рассеиванием на больших значениях: элитная недвижимость

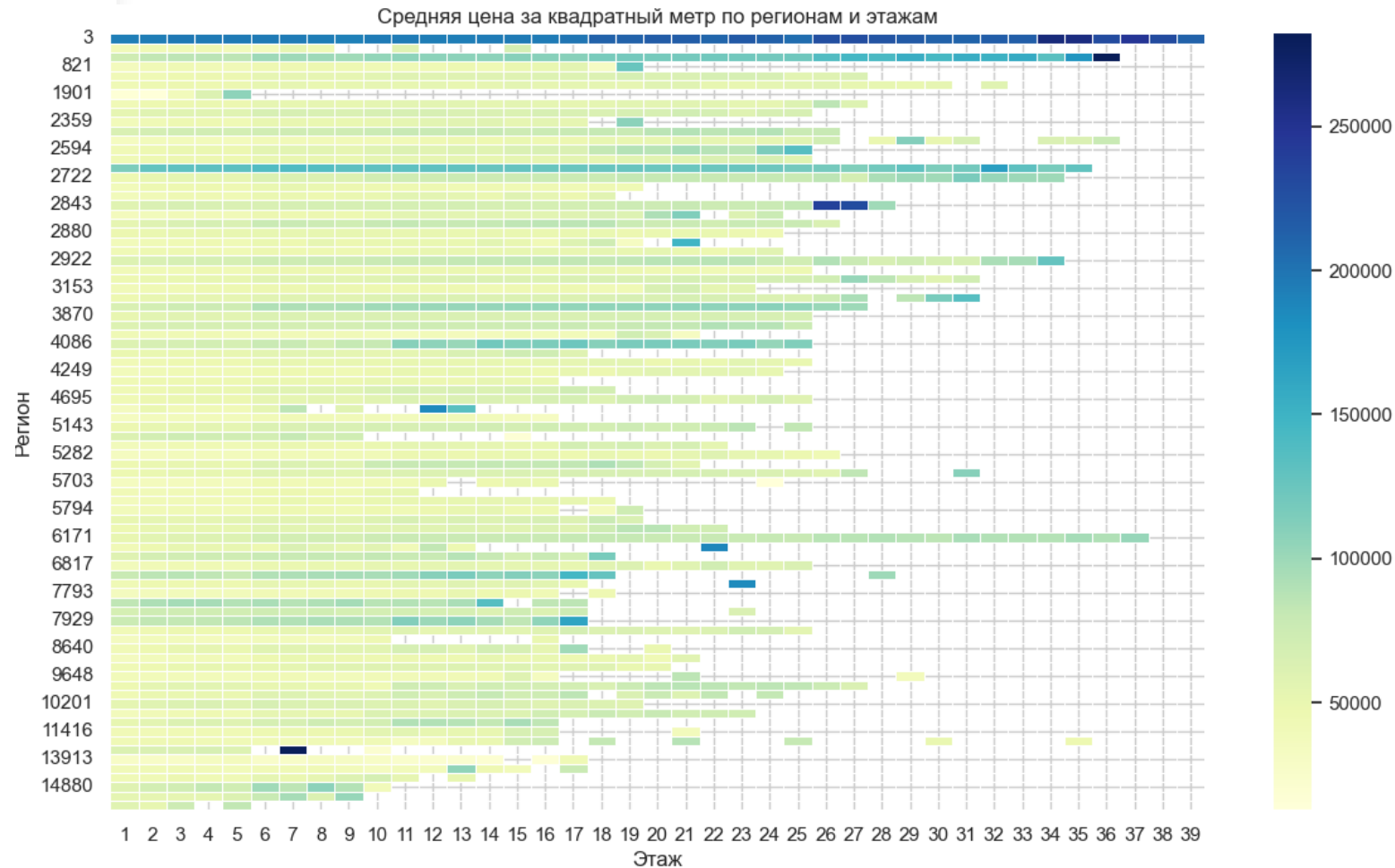
# От каких признаков зависит стоимость объекта недвижимости? +

**Цены за квадратный метр** сильно различаются **между регионами**, аномально высокие значения связаны с элитной или столичной недвижимостью.

**Этажность** менее влияет на цены, но на этажах выше 25 наблюдается разброс из-за элитных объектов и спроса на виды.



# От каких признаков зависит стоимость объекта недвижимости?



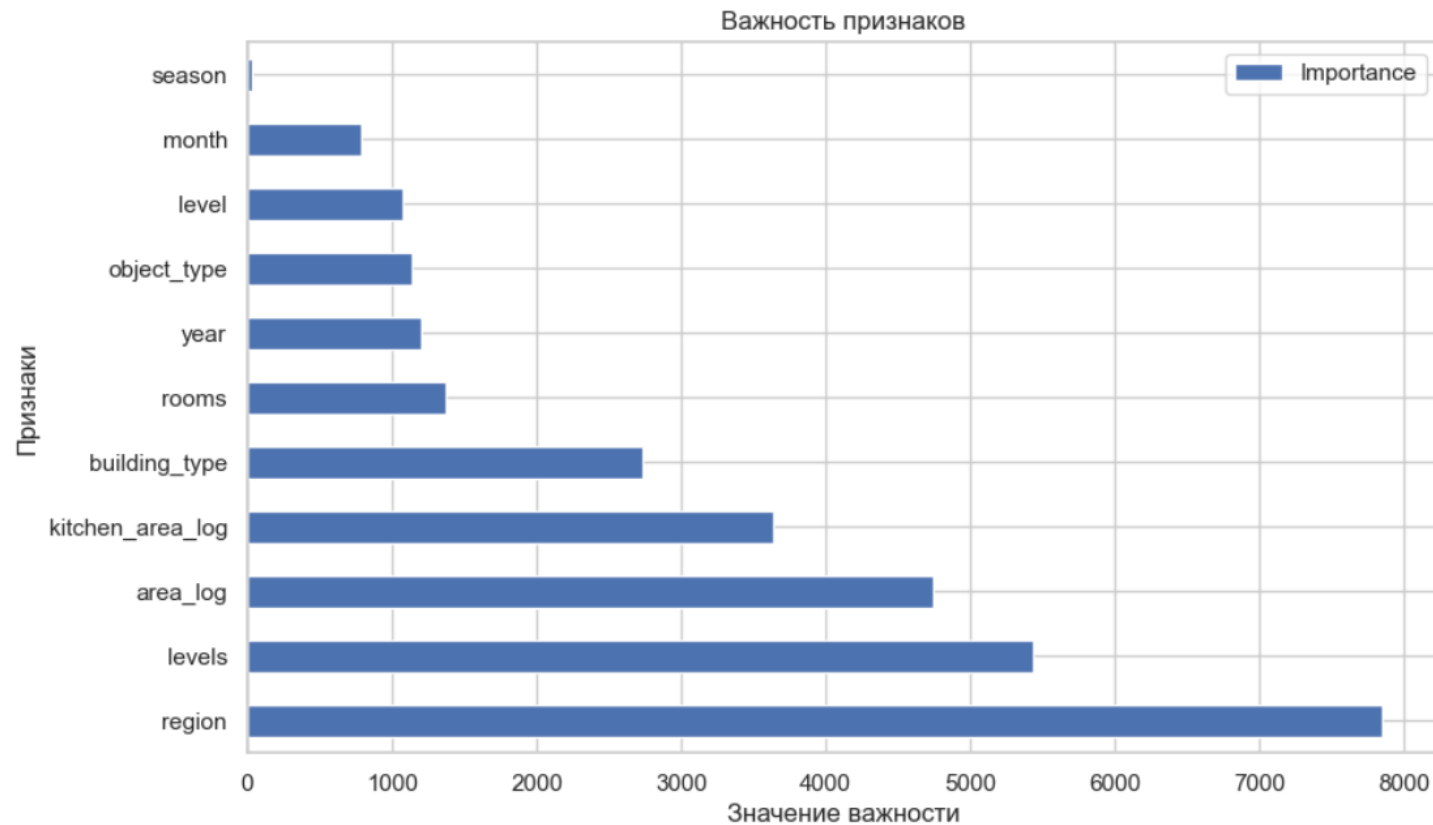
**Высокие цены** за квадратный метр сосредоточены **в крупных регионах**, таких как Москва и Санкт-Петербург, и на отдельных этажах. В престижных регионах, например, Москве, цены остаются высокими даже на низких этажах, тогда как в других, как Башкортостан, пик цен наблюдается на средних и высоких этажах из-за вида или особенностей рынка. В большинстве регионов цены умеренные и равномерно распределены



# От каких признаков зависит стоимость объекта недвижимости?

Для анализа с помощью **LightGBM** данные были логарифмированы, отобраны нужные признаки и закодированы категориальные значения.

**Проверка важности признаков** показала, что ключевое влияние на цену оказывают region, levels и area\_log. Умеренное влияние наблюдается у rooms, year и month. Признак season имеет минимальное значение для предсказания и считается наименее значимым.



# Проверка собственной гипотезы статистическими методами

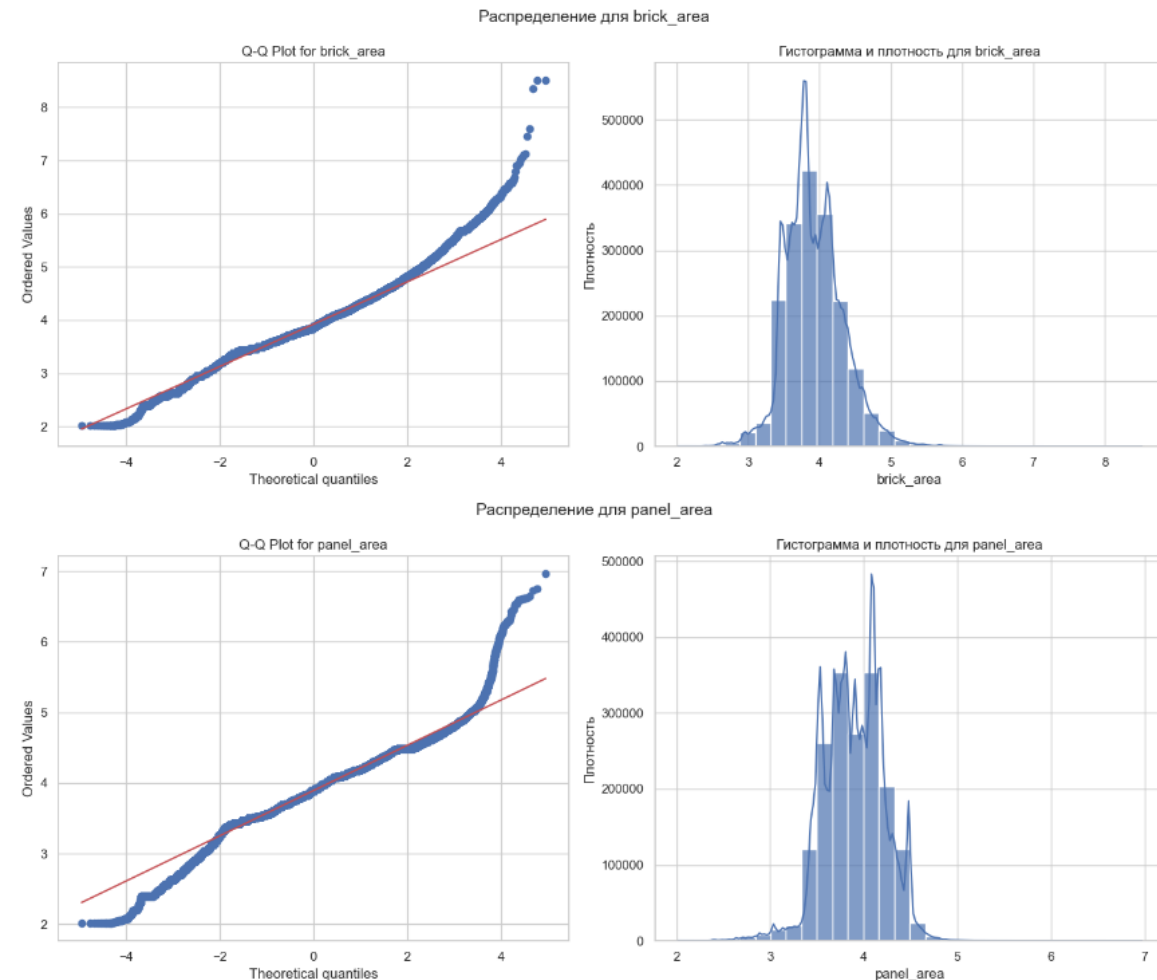
**$H_0$  (нулевая гипотеза):** Средняя площадь объектов в кирпичных зданиях равна средней площади объектов в панельных зданиях

**$H_1$  (альтернативная гипотеза):** Средняя площадь объектов в кирпичных зданиях не равна средней площади объектов в панельных зданиях

Использовался t-тест, Манна-Уитни и бутстреп

Разница средних площадей между кирпичными и панельными зданиями составляет 0.0299 и статистически значима.

**Нулевую гипотезу** о равенстве средних **отвергаем** при уровне значимости  $\alpha=0.05$



# Создание модели прогноза цены объекта недвижимости

**Хороший  
показатель  
качества модели**

Обработка данных, грамотный выбор предикторов и создание модели

**Коэффициент детерминации  $R^2 = 0.8556$**   
показывает, что 85.56% дисперсии логарифмированной цены объясняется моделью, что указывает на её высокое качество. Оставшиеся 14.44% дисперсии могут быть связаны с неучтёнными факторами или шумом в данных

**Благодарю за  
внимание!**



# Исследование недвижимости

Контрольная точка №2 по Анализу данных

Серкибаев Аким, Э-2209