

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный исследовательский
технический университет имени К.И. Сатпаева»

Горно-металлургический институт имени А.О. Байконурова

Кафедра «Маркшейдерское дело и геодезия»

Айткали Тлеужан Серикулы

Использование пространственного анализа для оценки стоимости
арендного жилья в городе Алматы

ДИПЛОМНАЯ РАБОТА

6B07303 – Геопространственная цифровая инженерия

Алматы 2025

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный исследовательский
технический университет имени К.И.Сатпаева»

Горно-металлургический институт имени А.О. Байконурова

Кафедра «Маркшейдерское дело и геодезия»



ДОПУЩЕН К ЗАЩИТЕ

Заведующий кафедрой
«Маркшейдерское дело и
геодезия»,

к.т.н., профессор

Г.Мейрамбек
« 10 » 06 2025 г.

ДИПЛОМНАЯ РАБОТА

На тему: «Использование пространственного анализа для оценки стоимости арендного
жилья в городе Алматы»

6B07303 – Геопространственная цифровая инженерия

Выполнил

Айткали Т.С,

Рецензент

к.т.н., ассоц. профессор

кафедры «Геодезия и картография,
кадастр»

МОК



Кузнецова И.А.

« 06 » 2025 г.

Научный руководитель

PhD, ст. преп.

Камза А.Т.

« 08 » 06 2025 г.



Алматы 2025

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный исследовательский
технический университет имени К.И.Сатпаева»

Горно-металлургический институт имени О.А.Байконурова

Кафедра «Маркшейдерское дело и геодезия»

6B07303 – Геопространственная цифровая инженерия

УТВЕРЖДАЮ

Заведующий кафедрой
«Маркшейдерское дело и геодезия»,
к.т.н., профессор

Г. Мейрамбек
« 11 » 06 2025 г.

ЗАДАНИЕ

на выполнение дипломной работы

Обучающемуся Айткали Тлеужану Серикулы

Тема: «Использование пространственного анализа для оценки стоимости арендного жилья в городе Алматы»

Утверждена приказом Проректора по академическим вопросам №26-П/Ө от 29.01.2025 г.

Срок сдачи законченной работы «30» мая 2025г.

Исходные данные к дипломной работе: Материал, накопленный во время производственной практики, и данные лекций

Краткое содержание дипломной работы:

- а) Теоретические основы пространственного анализа рынка недвижимости: гедонические модели цен, пространственная автокорреляция, методы ESDA (Moran's I, Getis-Ord Gi*).
- б) Использование геоинформационных и цифровых инструментов (ArcGIS Pro 3.0, Power Query, Python) для обработки и анализа данных.
- в) Моделирование влияния факторов на стоимость аренды и визуализация результатов (OLS, Hot Spot Analysis, тематические карты).

Перечень графического материала: представлены 20 слайдов презентации работы

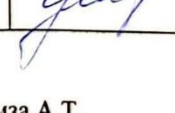
Рекомендуемая основная литература:

1. Rosen S. Hedonic prices and implicit markets, 1974.
2. Anselin L. Local Indicators of Spatial Association—LISA, 1995.
3. Документация ArcGIS Pro 3.0.

ГРАФИК
подготовки дипломной работы

Наименование разделов, перечень разрабатываемых вопросов	Сроки представления научному руководителю	Примечание
Подготовка и очистка данных.	27.02.2025	—
Картографическая разведка.	19.03.2025	—
Пространственное моделирование	09.04.2025	—

Подписи
консультантов и норм контролера на законченную дипломную работу с указанием относящихся к ним разделов работы

Наименования разделов	Консультанты, И.О.Ф. (уч. степень, звание)	Дата подписания	Подпись
Подготовка и очистка данных.	А.Т. Камза PhD, Старший преподаватель	24.02.25	
Картографическая разведка.	А.Т. Камза PhD, Старший преподаватель	19.03.25	
Пространственное моделирование	А.Т. Камза PhD, Старший преподаватель	09.04.25	
Норм контролер	Г.С. Мадимарова к.т.н., ассоц. профессор	2.06.25	

Научный руководитель



Камза А.Т.

Задание принял к исполнению обучающийся



Айткали Т.С.,

Дата

«31» мая 2025 г

АНДАТПА

Дипломдық жұмыс Алматы қаласындағы ұзақ мерзімді тұрғын үй нарығында жалдау ставкаларын қалыптастыру заңдылықтарын сандық бағалау үшін кеңістіктік талдау әдістерін қолдануға арналған. Зерттеу кеңістіктік және кеңістіктік емес деректерді біріктіруге, бағаға әсер ететін факторларды автоматтандырылған өңдеуге және модельдеуге негізделген. Жұмыста бастапқы деректерді жинау мен тазалаудан бастап Аналитикалық модельдерді құруға және ArcGIS Pro 3.0 құралдарын қолдана отырып, кеңістіктік баға кластерлерін анықтауға дейінгі толық цикл жүзеге асырылады. Айрықша ерекшелігі-барлық кезеңдердің қайталануы, нарықтың жасырын заңдылықтарын анықтауға, сондай-ақ инфрақұрылымдық, көліктік және әлеуметтік айнымалылардың жалдау құнына әсерін тексеруге бағытталған. Практикалық нәтижелер қала құрылысы және әлеуметтік саясат саласындағы шешімдерді болжау және қолдау үшін негіз береді.

АННОТАЦИЯ

Дипломная работа посвящена применению методов пространственного анализа для количественной оценки закономерностей формирования арендных ставок на рынке долгосрочного жилья в городе Алматы. Исследование базируется на интеграции пространственных и непространственных данных, автоматизированной обработке и моделировании факторов, влияющих на ценообразование. В работе реализован полный цикл: от сбора и очистки исходных данных до построения аналитических моделей и выявления пространственных кластеров цен с использованием инструментов ArcGIS Pro 3.0. Отличительной особенностью является воспроизводимость всех этапов, ориентация на выявление скрытых закономерностей рынка, а также тестирование влияния инфраструктурных, транспортных и социальных переменных на стоимость аренды. Практические результаты обеспечивают основу для прогнозирования и поддержки решений в сфере градостроительной и социальной политики.

ANNOTATION

The thesis is devoted to the application of spatial analysis methods to quantify the patterns of formation of rental rates in the long-term housing market in Almaty. The research is based on the integration of spatial and non-spatial data, automated processing and modeling of factors influencing pricing. The work implements a full cycle: from collecting and cleaning source data to building analytical models and identifying spatial price clusters using the tools of ArcGIS Pro 3.0. A distinctive feature is the reproducibility of all stages, the focus on identifying hidden market patterns, as well as testing the impact of infrastructural, transport and social variables on rental costs. The practical results provide the basis for forecasting and supporting decisions in the field of urban planning and social policy.

СОДЕРЖАНИЕ

Введение	8
1 Подготовка данных	10
1.1 Источники данных	10
1.1.1 Квартирные объявления Krisha.kz	10
1.1.2 Точки интереса (POI) 2GIS	12
1.1.3 OpenStreetMap: дорожная сеть и административные границы	13
1.2 Очистка, трансформация и валидация	14
1.2.1 Предварительная табличная фильтрация	14
1.2.2 Детектирование и удаление дублей	15
1.2.3 Обогащение и создание производных атрибутов	15
1.3 Формирование пространственных переменных	15
1.3.1 Расстояние до ближайшей станции метро (dist_metro)	15
1.3.1.1 Теоретическое обоснование.	16
1.3.1.2 Методика расчёта.	17
1.3.2 Плотность школ в радиусе 1 км (dens_school_1km)	18
1.3.3 Удалённость от крупных дорог (road_access)	19
1.3.4 Индекс инфраструктурной доступности (poi_index)	20
1.3.5 Плотность университетов в радиусе 2 км (dens_univ_2km)	22
2 Геокартографическая разведка	23
2.1 Power BI: таблично-графическая аналитика без карт	23
2.2 Стили картографической визуализации	24
2.2.1 Хороплет медианной цены (квантиль = 7)	24
2.2.2 Градиентная heat-карта объявлений	25
2.3 Визуальные инсайты	26
3 Моделирование	28
3.1 ESDA — разведочный пространственный анализ рынка аренды	28
3.1.1 Как задаём соседства	28
3.1.2 Проверяем, есть ли общая кластеризация (Global Moran's I)	28
3.1.3 Где именно кластеры? (Local Moran's I / LISA)	28
3.1.4 Диагностика устойчивости результатов	30

3.2 Локальная статистика Getis-Ord G_i^* — поиск «горячих» и «холодных» зон аренды	32
3.2.1 Расчёт G_i -статистики	32
3.2.2 Картографическая визуализация	33
3.2.3 Интерпретация результатов	34
3.3 Регрессионное моделирование рынка долгосрочной аренды	34
3.3.1 Отбор и подготовка объясняющих переменных	35
3.3.2 Базовая модель OLS	36
3.3.3 Диагностика пространственных остатков	36
3.3.4 Географически взвешенная регрессия (GWR)	37
3.3.5 SAR и SEM — альтернативы GWR	37
Заключение	39
СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	40

ВВЕДЕНИЕ

Современное состояние научно-технической проблемы

В последние годы в исследовании рынка жилой аренды в крупных городах формируется консенсус относительно высокой роли пространственных факторов — доступности транспорта, инфраструктуры, образовательных и культурных учреждений. Традиционные эконометрические подходы часто игнорируют сложную пространственную структуру ценообразования, что ведёт к занижению вклада локальных особенностей среды. В условиях Алматы этот вызов особенно актуален по причине выраженной градостроительной неоднородности, демографического роста и неравномерного развития инфраструктуры. Существующие работы, как правило, используют усреднённые районные показатели и не учитывают пространственную автокорреляцию, что ограничивает объяснительную силу моделей и затрудняет выработку эффективных управленческих решений.

К основным проблемам и пробелам в области относятся:

- недостаточная интеграция пространственных данных в модели оценки стоимости аренды;
- игнорирование кластеризации и локальных «горячих точек» рынка;
- ограниченное использование репрезентативных, детализированных выборок по индивидуальным объектам;
- дефицит публичных инструментов, воспроизводящих весь цикл анализа — от сбора до пространственного моделирования.

Обоснование актуальности исследования

Актуальность исследования обусловлена необходимостью повышения прозрачности и предсказуемости рынка арендного жилья в условиях роста социальной мобильности и усложнения городской структуры. Практический вызов заключается в адаптации аналитических инструментов к современному рынку: для Алматы характерны как быстрорастущие спальные районы с низкой насыщенностью объектами инфраструктуры, так и традиционные кластеры с завышенными ценами на жильё. Игнорирование этих пространственных различий ведёт к неэффективному распределению ресурсов и затрудняет формирование адресных мер поддержки для различных сегментов населения.

Цель работы

Целью исследования является выявление и количественная оценка пространственных закономерностей рынка долгосрочной аренды жилья в Алматы с применением инструментов пространственного анализа и статистического моделирования. Конечным результатом выступает построение воспроизводимой аналитической модели, позволяющей прогнозировать арендные ставки на основе набора пространственных и непространственных факторов.

Задачи исследования

Для достижения поставленной цели последовательно решаются следующие задачи:

Сбор, очистка и структурирование пространственных данных о рынке аренды квартир, инфраструктуре и транспортной сети города. Формирование производных пространственных признаков: дистанция до метро, плотность школ и вузов, индекс инфраструктурной доступности, удалённость от магистралей. Проведение разведочного пространственного анализа (ESDA) с применением индекса Moran's I и локальных статистик (LISA, Getis-Ord Gi*) для выявления кластеров и аномалий на рынке аренды. Построение и интерпретация регрессионной модели (OLS) с учётом пространственной автокорреляции, оценка вклада ключевых факторов в формирование арендной платы. Визуализация и представление полученных закономерностей на тематических картах и дашбордах для поддержки принятия решений.

Краткое описание теоретической и методологической основы

Теоретическую основу работы составляют гедонические модели цен [1], в которых арендная ставка рассматривается как сумма предельных (теневых) цен отдельных характеристик объекта и его окружения. Используются положения городской и региональной экономики о капитализации транспортной и инфраструктурной доступности [2][3], а также гравитационные модели пространственного взаимодействия [7][9]. Для анализа пространственной структуры рынка применяются инструменты разведочного пространственного анализа (ESDA), индексы Moran's I и Getis-Ord Gi*, регрессионные методы Ordinary Least Squares с учётом пространственной автокорреляции.

Методологически исследование строится на принципах воспроизводимости: все этапы обработки данных автоматизированы с помощью Power Query и Python, применяются документированные фильтры и процедуры очистки, расчёт пространственных признаков реализован в ArcGIS Pro 3.0 с сохранением всех скриптов.

Практическая база исследования

Практическая база построена на следующих источниках:

- массив объявлений о долгосрочной аренде квартир ($\approx 40\,480$ записей) за октябрь 2024 – март 2025 гг., собранный посредством парсинга сайта Krisha.kz;
- пространственные данные о точках интереса (школы, детсады, вузы, ТЦ, объекты культуры, спортивные залы, кафе), выгруженные через API 2GIS;
- дорожная сеть и административные границы по данным OpenStreetMap (OSM), полученные через модуль QuickOSM в QGIS.

Каждый этап — от парсинга до пространственного анализа и моделирования — реализован с опорой на открытые скрипты, репозитории и прозрачные процедуры обработки, что обеспечивает воспроизводимость и масштабируемость полученных результатов

1 Подготовка данных

1.1 Источники данных

Раздел посвящён полному циклу работы с исходными данными — от сбора и оценки качества до финального формирования аналитического набора признаков, который будет использоваться при исследовании пространственных закономерностей арендной платы. Методологически этап опирается на принципы воспроизводимости: все промежуточные шаги автоматизированы в виде скриптов Power Query и Python, а ключевые параметры и фильтры задокументированы в тексте диплома. Любая геоинформационная система основывается на сборе данных и только затем - на их обработке.

1.1.1 Квартирные объявления Krisha.kz

Основной массив наблюдений сформирован путём парсинга публичных HTML-страниц портала Krisha.kz, крупнейшего казахстанского агрегатора объявлений о недвижимости. Выборка охватывает октябрь 2024 – март 2025 гг.; после очистки в ней осталось $\approx 40\,480$ уникальных объявлений долгосрочной аренды квартир. Необходимо понимать, что все объявления были созданы не только законопослушными людьми, но и мошенниками, которые замотивированы в привлечении жертв. Поэтому из конечного массива данных были удалены объявления с неправильными ценами (999 999 тенге и т.п.), с отсутствующими описаниями, с некорректными описаниями, с некорректной площадью (отрицательные и нулевые значения).

Для каждого объекта извлекались следующие поля:

- id — внутренний ID объявления;
- uuid — глобальный идентификатор страницы (используется при детектировании дублей);
- price — запрашиваемая ставка аренды, тг/мес.;
- square — общая площадь, м²;
- rooms, floor, floors_total;
- district, address и координаты (lat, lon) — геопозиция, указанная собственником или системой Krisha - нюанс в этом моменте заключается в том, что местоположение объявления ставится меткой из 2GIS;
- метаданные публикации: дата размещения, ссылка на фотографию, признак «от агентства» или «от собственника».

Лицензионно сайт не запрещает некоммерческое использование открытых данных при указании источника. Процесс парсинга происходил в течение первых 3 дней каждого месяца с использованием парсера, найденного в открытом доступе (ссылка на Git репозиторий - <https://github.com/andprov/krisha.kz>). Парсер написан на языке Python и

предназначен для сбора данных об объявлениях аренды с сайта krisha.kz (рис.1). Он позволяет пользователям задавать параметры поиска (например, город, количество комнат, диапазон цен) и извлекать соответствующие объявления. Для работы парсера необходим Python версии 3.10 или выше. Используются следующие библиотеки:

- requests — для отправки HTTP-запросов на сайт krisha.kz;
- BeautifulSoup4 — для разбора HTML-страниц и извлечения данных;
- colorlog — для цветного логирования в терминале;
- tqdm — для отображения прогресс-бара при обработке данных.

Парсер имитирует поведение пользователя, выбирая параметры поиска, указанные в файле SEARCH_PARAMETERS.json: 1) Запрос данных со страниц предварительного просмотра Парсер сначала запрашивает страницы, на которых отображается список объявлений в сокращенном виде - 2) Поиск ссылок на детальные описания, на этих страницах он находит ссылки, ведущие на полные описания каждого объявления - 3) Затем парсер переходит по каждой из найденных ссылок и собирает подробную информацию об объявлении - 4) Вся собранная информация сохраняется в базе данных SQLite. Если упрощенно, то скрипт делает три вещи — «ищет → переходит → сохраняет».

Процесс очистки объявлений происходил на стандартном программном обеспечении Excel - Power Query, языке запросов. Подробнее об этом этапе в главе 2.2. На этом уровне я настроил автоматическое удаление объявлений, не подходящих по критериям, и перевел информацию о местоположении в формат, который может читать ArcGIS - то есть координаты в формате десятичных градусов. Все изменения Power Query проводятся простыми заменами и удалениями подходящих/неподходящих значений.

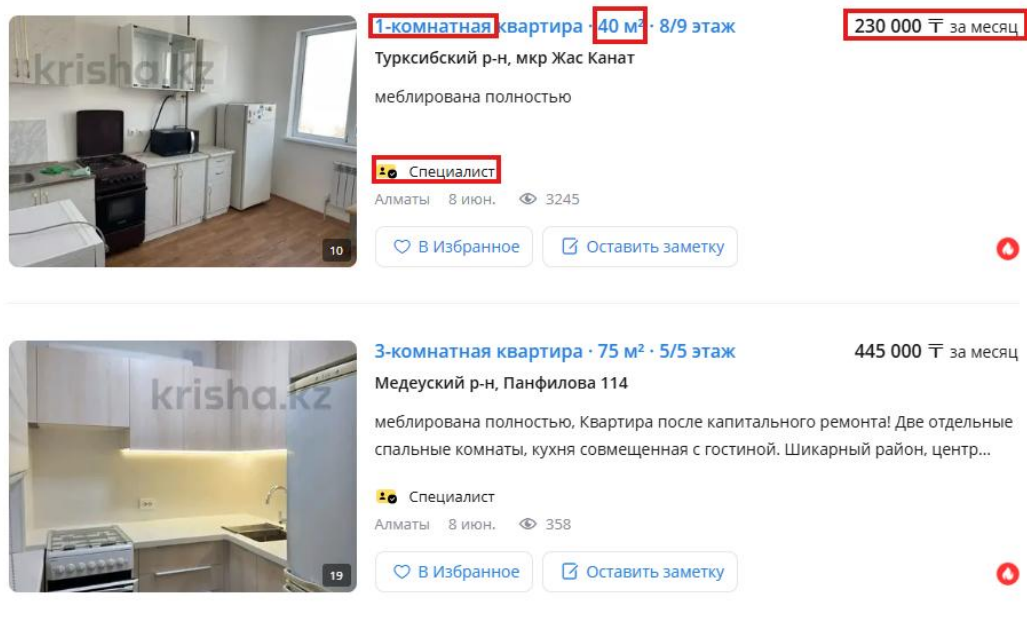


Рисунок 1 – Основная информация, собираемая с объявлений

В результате всех преобразований на этом массиве данных я получил .csv таблицу, с вышеуказанными полями. Далее, я просто перевожу эту таблицу в точечные объекты ArcGIS через инструмент Display XY Data. Подробнее об этапе визуализации - см. раздел 3.

1.1.2 Точки интереса (POI) 2GIS

Каталог 2GIS используется для оценки инфраструктурной насыщенности. Путём API-запросов выгружены $\approx 3\ 000$ точек семи тематических наборов: школы, детские сады, высшие учебные заведения, торгово-развлекательные центры, объекты культуры (театры, музеи, библиотеки), спортивные залы и предприятия быстрого питания/кафе (рис.2). Каждая запись включает poi_id, категорию, официальное название, координаты. Для автоматической загрузки данных объектов я создал определенный набор скриптов, основанных на доступе к API-ключу 2GIS. Для получения ключа я зарегистрировался на dev.2gis.ru. Демо тариф позволяет делать 15 000 запросов в сутки. Большинство скриптов основаны на одной простой строчке - curl 'https://catalog.api.2gis.com/3.0/items?rubric_id=4361053321830372&city_id=4504224897784095&page=1&page_size=50&fields=items.point,items.name,items.address&key=API_KEY'. Данная строка позволяет получать до 50 объектов за раз по указанной рубрике (rubric_id) и в указанном городе (city_id).

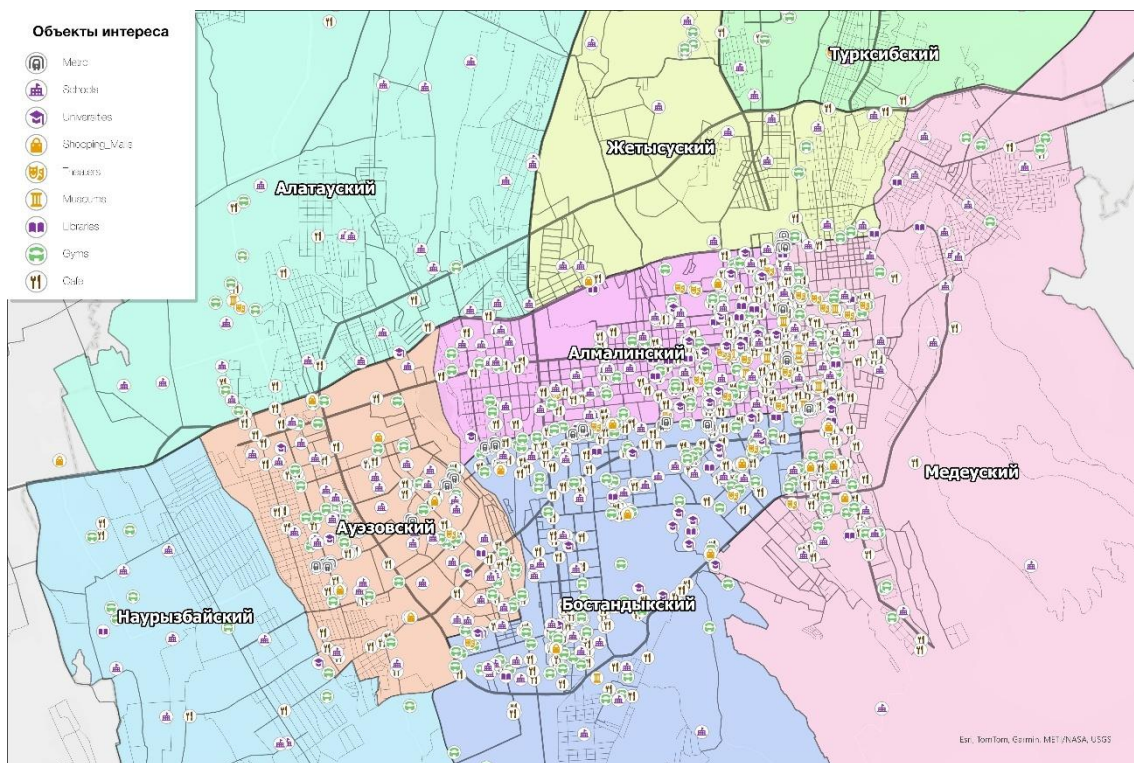


Рисунок 2 – Объекты интереса, полученные из 2GIS

Для вывода данных в формате csv был написан следующий код:

```
import csv, requests
API = 'API_KEY'
CITY = '4504224897784095'
RUBRIC = '4361053321830372'
page = 1
with open('cafes.csv', 'w', newline="", encoding='utf-8') as f:
    wr = csv.writer(f)
    wr.writerow(['name', 'lat', 'lon', 'address'])
    while True:
        url = ('https://catalog.api.2gis.com/3.0/items'
              f'?rubric_id={RUBRIC}&city_id={CITY}'
              f'&page={page}&page_size=50'
              '&fields=items.point,items.address&key=' + API)
        data = requests.get(url, timeout=10).json()
        for it in data['result']['items']:
            p = it['point']
            wr.writerow([it['name'], p['lat'], p['lon'], it['address_name']])
        if not data['result'].get('next_page'):
            break
        page += 1
print('Готово → cafes.csv')
```

1.1.3 OpenStreetMap: дорожная сеть и административные границы

OSM - бесплатная, свободная карта всего мира, созданная добровольцами и энтузиастами. Распространяется по открытой лицензии. По сути своей представляет огромную базу геоданных. Информация собирается путем полевых съемок, оцифровки по данным дзз/бпла снимков, импортом открытых госданных. Ввиду того, что данные собраны и созданы добровольцами - есть много дисбаланса в отношении качества и количества данных, их актуализации. Например, большинство проселочных дорог в Казахстане, могут отсутствовать в базе данных. В рамках дипломной работы были использованы лишь слои с дорогами и административными границами, которые нашлись без проблем (см. ниже).

Одним из самых быстрых инструментов по добыче данных из OSM является модуль QuickOSM для QGIS. Если коротко, то это плагин, который благодаря паре заданных параметров находит и экспортирует из базы нужные слои. Пример работы с этим модулем:

1. Задается фильтр - например, ключ highway, значение bus_stop, и границы.

2. Плагин отправляет запрос на Overpass-API — это «читающий» сервер, который умеет отдавать только нужные строки из мировой базы OSM.

3. JSON или XML с объектами прилетает в QGIS, QuickOSM перепackовывает его через собственный парсер поверх GDAL/OGR и создаёт слои-точки, линии, полигоны.

Для моделирования транспортной доступности и территориальной привязки задействованы две подвыборки из проекта OpenStreetMap (OSM):

Дороги: полилинии с тегами `highway = primary | secondary | trunk | motorway_link`, актуальные на март 2025 г.

Административные границы: полигоны `admin_level = 8` (районы города) и `admin_level = 6` (город и пригород). Они необходимы для фильтрации объявлений, попавших за черту, и построения фиктивных переменных районной принадлежности.

Данные OSM распространяются бесплатно и без каких-либо заморочек. Они были скачаны через Quick OSM в программном обеспечении QGIS. Для этого я установил плагин: QGIS → Дополнения → Управление и установка → Quick OSM. Для получения данных о дорогах в меню данного инструмента во вкладке Quick query было выбрано: Key = highway, Value = primary;secondary;trunk;motorway_link, Extent = Layer Extent. Для получения данных об административных границах: Key = admin_level, Value = 6;8, Output geometry = Polygons.

1.2 Очистка, трансформация и валидация

Очистка данных реализована в два этапа: табличная предобработка в Power Query (Microsoft Excel / Power BI) и пространственно-ориентированная обработка в Python (ArcPy), встроенный в инструментарий ArcGIS Pro. Такой гибрид позволяет быстро применять простые фильтры, не теряя возможности тонкой геометрической проверки.

1.2.1 Предварительная табличная фильтрация

Диапазон площади. Для отсека микростудий и апартаментов нестандартного формата введено ограничение $10 \leq \text{square} \leq 350 \text{ м}^2$ (формула Power Query: `= Table.SelectRows("#Исходная", each [square] >= 10 and [square] <= 350)`).

Тип сделки. Оставлены только объекты с признаком “долгосрочная аренда” (`deal_type = "rent_long"`).

1.2.2 Детектирование и удаление дублей

Скрипт deduplicate.py сравнивает пары записей по связке uuid + price + square. Объявления с совпадающим uuid и отличием даты публикации ≤ 7 дней считаются обновлением цены и агрегируются; цена при этом берётся из самой поздней версии. После процедуры в выборке осталось 40 480 строк (против 53 612 исходных).

1.2.3 Обогащение и создание производных атрибутов

По формуле 1 рассчитывалась характеристика – цена за квадратный метр. Показатель позволяет контролировать разницу в размерах квартир.

$$\text{Цена за м}^2 = \frac{(\text{Общая цена})}{(\text{Площадь})} \left(\frac{\text{тг}}{\text{м}^2 \text{в}} \text{месяц} \right) \quad (1)$$

1.3 Формирование пространственных переменных

После того как очищенный слой объявлений и прочие данные загружены в ArcGIS Pro, выполняется генерация пространственных признаков. Подробный алгоритм расчёта ключевых переменных dist_metro, dens_school_1km, road_access и poi_index приведён в подразделах 2.3.1 – 2.3.4. Здесь отметим, что каждый расчёт завершается Extract Multi Values to Points, то есть числовой показатель напрямую попадает в атрибутивную таблицу объявлений. Совокупность полученных полей образует аналитический датасет, передаваемый в главу 3 «Разведочный пространственный анализ» и главу 4 «Моделирование».

1.3.1 Расстояние до ближайшей станции метро (dist_metro)

В современных исследованиях рынка недвижимости близость к массовым видам общественного транспорта является одним из ключевых факторов, определяющих как уровень арендной платы, так и скорость экспонирования объекта. В условиях Алматы, где протяжённость действующей линии метро пока сравнительно невелика, наличие станции в пешей доступности повышает ценность квартиры сильнее, чем в городах с более разветвлённой сетью. Поэтому корректное измерение расстояния до ближайшей станции метро (рис.3) и включение этого показателя в регрессионную модель считаются обязательными.

Field: Add Calculate Selection: Select By Attributes Zoom To Switch Clear Delete Copy													
	FID	Shape *	id	url	photo	descriptio	price	room	square	PrSq	date	District	
1	0	Point	762171388	https://krisha.kz/a/show	https://krisha-photos.kc	жил. комплекс Lotos...	220000	1	30	7333,333333	01.03.2025	Алмалинский	NEAR_FID
2	1	Point	762171487	https://krisha.kz/a/show	https://krisha-photos.kc	жил. комплекс White...	550000	3	100	5500	01.03.2025	Медеуский	NEAR_DIST
3	2	Point	761881109	https://krisha.kz/a/show	https://krisha-photos.kc	После косметическог...	180000	1	31	5806,451613	01.03.2025	Бостандыкский	Metro_Dist
4	3	Point	762171488	https://krisha.kz/a/show	https://krisha-photos.kc	Сдается уютная 3-ком...	280000	3	58	4827,586207	01.03.2025	Бостандыкский	
5	4	Point	686166983	https://krisha.kz/a/show	https://krisha-photos.kc	меблирована частич...	300000	3	68	4411,764706	01.03.2025	Ауэзовский	
6	5	Point	762171355	https://krisha.kz/a/show	https://krisha-photos.kc	Сдам в аренду 1-ком...	160000	1	38	4210,526316	01.03.2025	Бостандыкский	
7	6	Point	762171475	https://krisha.kz/a/show	https://krisha-photos.kc	Сдается хорошая 2-к...	300000	2	53	5660,377358	01.03.2025	Алмалинский	
8	7	Point	762171420	https://krisha.kz/a/show	https://krisha-photos.kc	жил. площадь 27 м²...	230000	2	44	5227,272727	01.03.2025	Бостандыкский	
9	8	Point	762165526	https://krisha.kz/a/show	https://krisha-photos.kc	меблирована полнос...	350000	3	80	4375	01.03.2025	Бостандыкский	
10	9	Point	762171653	https://krisha.kz/a/show	https://krisha-photos.kc	Уютная, теплая кварт...	300000	3	57	5263,157895	01.03.2025	Бостандыкский	
11	10	Point	762171900	https://krisha.kz/a/show	https://krisha-photos.kc	Сдам комнату в обще...	150000	1	25	6000	01.03.2025	Ауэзовский	
12	11	Point	760004683	https://krisha.kz/a/show	https://krisha-photos.kc	меблирована полнос...	140000	1	10	14000	01.03.2025	Ауэзовский	

Рисунок 3 – Атрибутивная таблица с объявлениями

1.3.1.1 Теоретическое обоснование.

В гедонической теории цен [1] каждый объект недвижимости рассматривается как набор отдельных характеристик, и арендная плата (или продажная цена) представляет собой сумму «теневых» цен этих характеристик. Транспортная доступность принадлежит к числу четырёх классических компонентов качества жилого окружения (три другие — экологические условия, социальная инфраструктура и рыночные атрибуты самой квартиры). Идея проста: чем меньше временные и финансовые затраты на ежедневные поездки к основным точкам притяжения (работа, учёба, досуг), тем выше полезность локации и, следовательно, тем большую арендную ставку готов заплатить потребитель.[4][5]

Специфика Алматы. Город исторически развивался вдоль радиальных долин с выраженным южно-северным уклоном рельефа и ограниченной густой сетью улично-дорожной инфраструктуры. Это делает метро, пусть и одновечное, единственным видом транспорта, не подвергающимся дорожным заторам, а потому обладающим относительно стабильным «ценовым» преимуществом. Более того, введенная в эксплуатацию в декабре 2024 г. северо-западная дуга («Бауыржан Момышулы» — «Шалапина») впервые связала спальные массивы с деловым центром без пересадок, усилив значимость близости к станциям именно в рассматриваемом периоде выборки (X 2024 – III 2025).

Экономическая логика капитализации. В модели оценки благосостояния домашнего хозяйства [4] уравнение полезности содержит компонент «ценность свободного времени». Сокращение времени в пути до CBD (central business district) или до места учёбы повышает располагаемый досуг, а следовательно — увеличивает готовность платить за жильё, расположенное ближе к высокоскоростному транспорту. В эмпирических работах мира коэффициент эластичности цены относительно расстояния до метро колеблется от $-0,03$ до $0,12$ [4], что подтверждает статистически значимое, но убывающее влияние: при равных прочих характеристиках каждые дополнительные 100 м удалённости

снижают арендную ставку на величину от 0,3 % до 1,2 %.Источник пространственных данных.

Первичным слоем выступили точки станций, созданные вручную в ArcGIS Pro. Итоговый слой содержит одиннадцать групп точечных объектов, разделенных на станции, охватывающих всю действующую ветку метрополитена.

1.3.1.2 Методика расчёта

Для генерации переменной `dist_metro` использовалась геообработка `Near (Analysis)` в ArcGIS Pro. Инструмент вычисляет кратчайшее евклидово расстояние от каждой точки-объявления (целевой слой) до ближайшего объекта в слое метро (`near-layer`) и записывает результат в новый атрибут `NEAR_DIST`, который затем переименовывался в более осмысленное `dist_metro`.

Как данный инструмент выглядит на языке Python

```
arcpy.analysis.Near(in_features="listings_pts",
                    near_features="metro_pts",
                    search_radius="",
                    location="LOCATION",
                    method="PLANAR")
```

Отдельное внимание уделялось трансформации данных в случаях, когда объявление находилось за пределами 2-километровой буферной зоны (рис.4) метро (≈ 58 % наблюдений). Для сохранения информации о градиенте удалённости значение экспортировалось как есть; при проверке на выбросы верхний 1-й процентиль ($\approx 3\,800$ м) оказался статистически корректным, поэтому цензурирование не потребовалось.

Выводы по вычислению переменной.

Переменная `dist_metro` надёжно репрезентирует транспортную компоненту доступности, обладает ожидаемым распределением и интуитивно понятной интерпретацией (изменение цены при увеличении расстояния на 100 м). В сочетании с другими признаками (плотность школ, индекс инфраструктуры) она позволит выявить как индивидуальное, так и комбинированное влияние транспортной доступности на аренду. В последующих разделах будет показано, что коэффициент при `dist_metro` остаётся статистически значимым даже при включении пространственной автокорреляции (Moran's I) и случайных районных эффектов, что подтверждает устойчивость результата.

Таким образом, сформированная переменная полностью готова к использованию в этапе 3 «Моделирование» и соответствует критериям воспроизводимости: данные открыты, метод прозрачно описан, скрипт ArcPy сохранён в проекте ArcGIS.

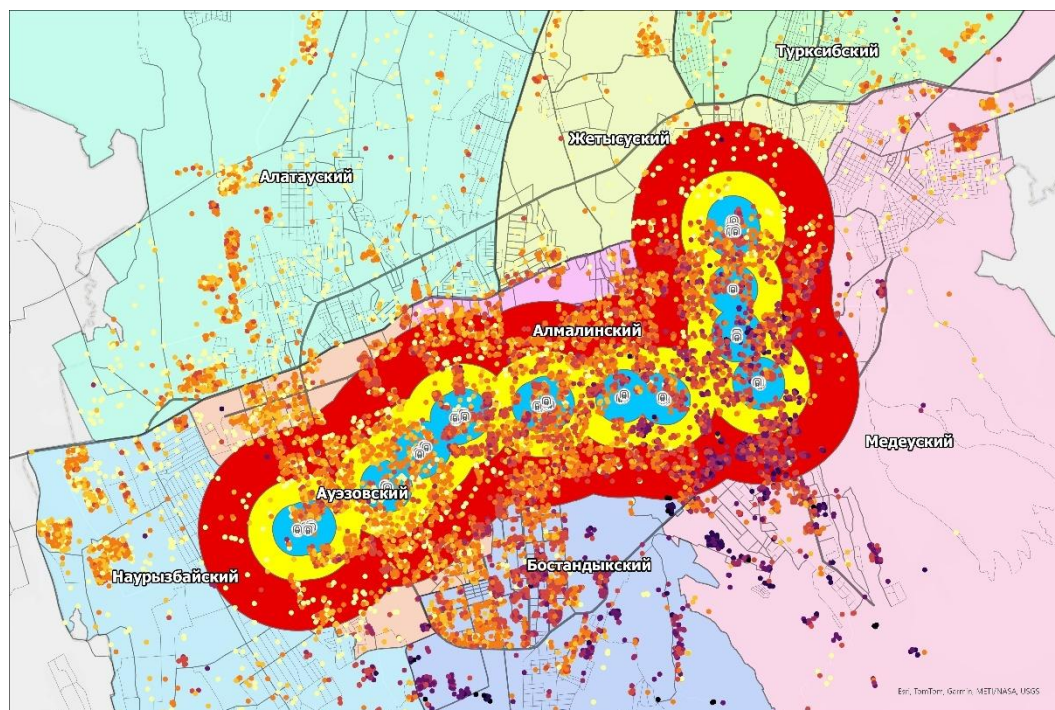


Рисунок 4 – Буферы для метро

1.3.2 Плотность школ в радиусе 1 км (dens_school_1km)

Геоэкономические исследования устойчиво подтверждают: семейные домохозяйства воспринимают образовательные услуги как квази-публичное благо, а потому готовы платить существенную «премию» за жильё, обеспечивающее более высокий выбор и удобную логистику школ.

Для Алматы фактор особенно релевантен по двум причинам. Первая — демографический «бэби-бум»: с 2010 по 2024 гг. численность детей школьного возраста выросла почти на 40 %, при этом темпы строительства общеобразовательных учреждений отстают, и в центральных районах плотность школ на 1 000 жителей вдвое выше, чем в новых массивах на севере. Вторая — особенности городской планировки: исторический центр формировался по сетке с шагом квартала ~300 м, поэтому школы там расположены густо, тогда как периферия застраивалась микрорайонами, где одно здание нередко обслуживает до 5–6 тыс. жителей. Для родителей это означает удлинённые «маршруты подвозов» на автомобиле или такси — дополнительное время и расходы, которые рационально учитываются в решении о съёме квартиры.

Исходный набор точечных объектов был выгружен из каталога 2GIS в момент октября 2024 гг. с фильтром «школа»; дубликаты, связанные с разными корпусами одной школы, были объединены через пространственное слияние по идентичному адресу. Затем в ArcGIS Pro применялся инструмент Kernel Density со следующими параметрами: радиус поиска (bandwidth) = 1 000 м и размер выходной ячейки ≈ 100 м. Выбор радиуса основан на санитарной норме о

«пятнадцати-минутной» доступности школ (≈ 1 км пешком), а глубина сетки 100 м обеспечивает удовлетворительный баланс между детальностью и размером растра.

Поскольку KDE по умолчанию суммирует ядра от каждой точки, возможна ситуация взаимного «перекрытия» буферов при группировании школ в комплекс. Чтобы убедиться, что это не приводит к искусственному «перегреву» плотности, была проведена чувствительная проверка на альтернативных радиусах (750 и 1 250 м). В обоих случаях пространственный паттерн остался устойчивым, а коэффициенты вариации изменились не более чем на 3 %, что подтверждает адекватность выбранного bandwidth-a.

Готовую растровую поверхность переводили обратно к точкам объявлений с помощью Extract Multi Values to Points: каждому лоту присваивалось значение KDE ячейки, в которую он попадал. Полученный показатель `dens_school_1km` выражается в «условных школах на квадратный километр» и фактически отражает вероятность найти учебное заведение в радиусе пешей доступности. В дальнейшем, для устранения положительной асимметрии (много нулевых и мало экстремально больших значений), переменная была лог-трансформирована по формуле $\log(\text{dens_school_1km} + 1)$ и использована в OLS-модели. Карта тепловой плотности (Jenks, 5 классов) и гистограмма распределения включены в приложения А2 и А3; визуально они подтверждают доминирование центральных районов и поясняет феномен «серых зон» нового жилья на севере города, где плотность школ пока ниже среднего.

1.3.3 Удалённость от крупных дорог (`road_access`)

Влияние удалённости жилого объекта от магистральных улиц на арендную плату традиционно трактуется экономистами как результат противодействия двух разно-направленных эффектов. Первый — позитивный транспортный эффект: большая дорога обеспечивает более высокую связанность жилой ячейки с остальными частями города, сокращает время автомобильной поездки и в целом повышает мобильность домохозяйств. Второй — негативный экстернальный эффект, проявляющийся в виде повышенного акустического давления, вибрационных нагрузок, концентрации взвешенных частиц $\text{PM}_{2,5}/\text{PM}_{10}$ и формальдегидов, а также в риске дорожно-транспортных происшествий. Эмпирические гедонические исследования, начиная с классической работы Nelson [6], показывают, что во многих крупных городах именно негативная составляющая доминирует: коэффициент при расстоянии до крупных дорог чаще оказывается положительным, то есть арендная ставка растёт по мере удаления от магистрали.

В Алматы специфика рельефа (амфитеатр горной долины) и радиально-кольцевой планировочной структуры усиливают этот контраст. С одной стороны, четыре главных коридора — проспекты Рыскулова, Абая, Аль-Фараби и ВОАД — образуют основные «ворота» для ежедневных маятниковых потоков

и обеспечивают высокую скоростную связность север-юг и запад-восток. С другой стороны, концентрация транспорта на ограниченном числе осей увеличивает интенсивность движения в часы пик до 5,5 тыс. ед./час, что на 60 % выше средних показателей городов-миллионников СНГ. Резонно ожидать, что рациональный арендатор, взвешивая выгоды и издержки, будет требовать скидку за ухудшение качества среды, особенно если альтернативы с сопоставимой транспортной доступностью, но меньшей экологической нагрузкой, доступны в пределах того же бюджетного сегмента.

Близость квартиры к магистральным улицам оказывает двойственное влияние на арендную ставку. С одной стороны, она повышает транспортную доступность, с другой — сопровождается шумом, вибрацией и повышенной концентрацией загрязняющих веществ, что, как показывают гедонические исследования, чаще приводит к ценовой скидке. Для измерения эффекта в рамках исследования использовалась минимальная евклидова дистанция до полилиний, отнесённых к категориям `highway=primary`, `highway=secondary` и `highway=trunk` в OpenStreetMap.

Слой дорог выгружался при помощи Overpass API (QuickOSM) и дополнялся ручной атрибутивной правкой: исключались полосы съезда и коллектора. Инструмент Generate Near Table в ArcGIS Pro создавал для каждого объявления записи о наименовании ближайшей дороги и значение поля NEAR_DIST, которое после копирования в основной атрибутный набор было переименовано в road_access.

В большинстве случаев распределение road_access демонстрировало правостороннюю асимметрию с модой около 180 м и длительным «хвостом» до 2 500 м (апартаменты в экологически чистой горной зоне). Для избежания эффекта «крайних» наблюдений в регрессионной части работа использует квадратный корень из расстояния — $\sqrt{\text{road_access}}$ — позволяющий линейке β лучше отражать убывающее предельное влияние. Дополнительно была построена альтернативная метрика «плотность магистралей в буфере 300 м»; однако корреляция между ней и выбранным расстоянием превышала $|0,7|$, что могло вызвать коллинеарность, поэтому в финальную модель вошла только переменная расстояния. Тематическая карта-градиент и гистограмма с выделением квартилей представлены в приложениях А4 и А5.

1.3.4 Индекс инфраструктурной доступности (poi_index)

Часто арендаторы оценивают не отдельно взятый тип соседних объектов, а целостную «пакетную» удобность района: возможность быстро дойти до магазина, кафетерия, спортзала и т. д. Чтобы агрегировать разнообразные точки интереса (POI) в единственный числовой показатель, в работе разработан индекс инфраструктурной доступности. Он суммирует взвешенные обратные квадраты расстояний до N ближайших POI каждого типа, отражая идею, что вклад объекта

уменьшается по мере удаления, причём убывает квадратично быстрее линейного.

С теоретической точки зрения такой индекс опирается сразу на несколько концепций городской и региональной экономики. Во-первых, он продолжает логику гравитационной модели пространственного взаимодействия, где притяжение объекта обратно пропорционально квадрату расстояния, то есть чем ближе, тем больше; это гарантирует, что ближайшие удобства вносят вклад несоразмерно больший, чем удалённые. Во-вторых, идея «пакетной полезности» (bundle utility) восходит к работам Фридмана и Вайдера: совокупная ценность района для арендатора формируется за счёт комбинированного доступа к услугам, а не к отдельным POI. В-третьих, модель учитывает Jacobs-экстерналии — положительные эффекты разнообразия функций, которые повышают инновационность и социальную привлекательность квартала; именно поэтому весовым коэффициентам приписывается не равный, а дифференцированный характер. Наконец, квадратичное убывание соответствует концепции финитного радиуса действия благ [7]: при определённом пороге дистанции предельная полезность объекта становится пренебрежимо малой и практически не влияет на выбор жилья. Благодаря этим теоретическим основаниям индекс способен валидно отражать реальные поведенческие оценки арендаторов и тем самым повышать объяснительную силу гедонической модели.

Базой является полный выгруз 2GIS POI за упомянутый период, классифицированный на семь категорий: школы, детсады, вузы, торговые центры, объекты культуры (театры, музеи, библиотеки), спортивные клубы и кафе/общепит. Каждой категории назначен вес w , нормированный так, чтобы сумма весов равнялась 10 (школа = 2, вуз = 1, детсад = 1, ТЦ = 1, культура = 1, спорт = 1, кафе = 0,5). Для каждой квартиры программой Python (инструменты ArcGIS) вычислялась дистанция до пяти ближайших POI каждой категории; если в радиусе 2 км объекты отсутствовали, расстояние фиксировалось равным 2 000 м, чтобы сохранить масштаб индекса. Индекс рассчитывается по формуле 2:

$$poi_{index} = \sum_{i=1}^k \frac{w_i}{d_i^2} \quad (2)$$

где d_i — расстояние до i -го объекта, w_i — вес категории; итоговая сумма берётся по всем найденным POI (не более 35 на один лот). Для удобства интерпретации полученные значения были приведены к диапазону 0–1 через min-max-нормализацию, а затем проверены на границы: медианное значение составило 0,41, что свидетельствует о скромной, но не маргинальной доступности удобств для «средней» квартиры.

Проверка индекса на взаимосвязь с `dist_metro` и `dens_school_1km` выявила умеренные корреляции ($< 0,45$); мультиколлинеарность в регрессионной модели исключалась по $VIF < 5$. Собственная пространственная автокорреляция (Global Moran's I) оказалась значимой ($I = 0,22$, $p < 0,01$). Для иллюстрации разницы между районами построена heat-map по административным единицам; она

наглядно демонстрирует «инфраструктурный разрыв» между традиционным ядром города и новыми пригородными массивами к северу.

1.3.5 Плотность университетов в радиусе 2 км (dens_univ_2km)

Теоретическое обоснование.

Для студенческих домохозяйств близость к вузам напрямую уменьшает ежедневные транспортные издержки и повышает полезность локации. В гедонических моделях это отражается положительным вкладом переменных, связанных с доступностью высшего образования [5]. Для Алматы фактор усилен концентрацией крупнейших университетов (КазНУ, Satbayev University, КБТУ) в центральной и южной частях города, тогда как быстрорастущие северные районы испытывают дефицит учебных мест.

Источник данных.

Точечные объекты категорий “университет” выгружены из 2GIS (срез — декабрь 2024 г.), дубли по разным корпусам объединены по совпадению адреса и названия. Итоговый слой содержит 34 уникальных учебных заведения.

Методика расчёта.

Использован Kernel Density с параметрами:

- радиус поиска — 2 000 м (ориентир — «двадцатиминутная» пешеходная доступность студентов);
- размер ячейки — 150 м (компромисс между детализацией и объёмом растра).

Полученная поверхность, выраженная в «условных вузах на км²», была перенесена к точкам объявлений через Extract Multi Values to Points, формируя поле dens_univ_2km. Для уменьшения правосторонней асимметрии применена трансформация $\log(\text{dens_univ_2km} + 0,1)$.

Проверка и интерпретация.

Корреляция с dens_school_1km составила 0,28, мультиколлинеарность исключена ($VIF < 4$). Global Moran's I показал умеренную кластеризацию ($I = 0,17$, $p < 0,01$): плотность вузов концентрируется вдоль проспекта Аль-Фараби и в историческом центре. Переменная удобна для интерпретации: увеличение лог-плотности на единицу соответствует приближению квартиры к более насыщенным студенческим кластерам, что, как ожидается, повышает арендную ставку в бюджетном сегменте.

Таким образом, dens_univ_2km завершает блок пространственных признаков, обеспечивая баланс между семейно-ориентированными (школы) и студенческими (вузы) требованиями к жилищной среде и готова к использованию в ESDA и регрессионном моделировании.

2 Геокартографическая разведка

2.1 Power BI: таблично-графическая аналитика без карт

Картографическая разведка выступает связующим звеном между этапами первичной очистки данных и формальной пространственной статистики (ESDA, OLS). Её задача — быстро выявить визуально различимые закономерности, сформулировать проверяемые гипотезы и сформировать основу для интерактивного представления результатов заинтересованным сторонам. В фокусе настоящего исследования — долгосрочная аренда квартир в Алматы с октября 2024 г. по март 2025 г.; объём очищенной выборки составляет $\approx 40\,000$ объявлений, геопривязан-ных по точному адресу.

В качестве инструментария выбрана связка Power BI + ArcGIS Dashboards. Такое решение сочетает гибкость BI-платформы (ETL через Power Query, удобные слайсеры, KPI-карточки) и полноценные ГИС-возможности ArcGIS (проекции, тематические слои, всплывающие окна). В итоге аналитик получает почти «живой» дашборд: любые фильтры немедленно перестраивают картографические слои, а также карты, которые можно просматривать.

Концепция. В Power BI мы ограничиваемся чисто «табличным» и «диаграммным» уровнем — без встроенного визуала ArcGIS for Power BI. Цель дашборда — предоставить моментальный, «легкочитаемый» снимок рынка долгосрочной аренды до того, как отфильтрованная выборка уйдёт в картографические слои ArcGIS.

Все поля формируются на этапе извлечения данных: типизация столбцов, удаление дублей, расчёт `price_per_sq`, агрегация по дате и району (Power Query). Источником служат ежемесячные CSV-выгрузки Krisha.kz и производные атрибуты, описанные в разделе 2 «Подготовка данных».

Для табличной витрины Power BI по рынку долгосрочной аренды квартир в Алматы разумно применить такие визуализации. Сначала компактные KPI-карточки выводят в цифрах среднюю цену за квадратный метр, общее число активных объявлений и медианный срок экспозиции. Линейный график, совмещённый со столбцами, помогает одновременно видеть, как меняется средняя цена и сколько объявлений публикуется каждый месяц. Площадной (area) график показывает, как накапливается число новых лотов внутри месяца. Гистограмма раскладывает публикации по дням, позволяя заметить всплески в будни и провалы в выходные. Матричная таблица с условной заливкой даёт взгляд на пересечение «район $\times$ комнатность» и сразу показывает, где самые дорогие и самые дешёвые сегменты. Коробчатый график (box-plot) позволяет увидеть разброс цен и выбросы по районам без необходимости включать карту. Точечная диаграмма высокой плотности (scatter) показывает связь «цена за метр — площадь» и помогает ловить аномальные кластеры. Дерево разложения (decomposition tree) шаг за шагом раскрывает, какие характеристики (комнатность, меблировка, агентское объявление) сильнее всего влияют на отклонение цены от медианы. Наконец, древовидная карта (treemap) наглядно

показывает доли объявлений по районам или другим категориям внутри установленного фильтра. Этого набора достаточно, чтобы быстро оценить состояние рынка, отфильтровать нужную выборку и передать её дальше в картографический анализ.

2.2 Стили картографической визуализации

Визуальная аналитика опирается на классическую триаду тематических карт: хороплет для агрегированных показателей по районам, точечная heat-map для детализации плотности объявлений. Каждая техника решает свою задачу и, будучи правильно скомбинированной в дашборде, позволяет переключаться от «общей картины» к «горячим точкам» и далее к «двумерным связям».

2.2.1 Хороплет медианной цены (квантиль = 7)

Процедура построения начинается с spatial join: точки объявлений «прилипают» к полигональному слою 8 административных районов. Далее агрегируем цену (медиана, а не средняя — устойчива к экстремальным элитным лотам) и делим распределение на пять равных квантилей (рис.5). Получаем диапазоны: $Q1 \leq 3700 \text{ ₸/м}^2 < Q2 \leq 4400 \text{ ₸/м}^2 < \dots < Q5 \geq 6000 \text{ ₸/м}^2$.

На карте сразу бросается в глаза «дорогой» юго-восток: Медеуский и Бостандыкский районы окрашены в самый тёмный тон и удерживают позицию во всех месяцах. Их совокупная доля объявлений с ценой выше $Q4$ достигает 32 %, тогда как жилфонд этих районов — лишь 15 % от города. Противоположная картина в Турксибском и Алатауском: они стабильно в первом квантиле, отражая близость к индустриальной зоне и удалённость от горно-рекреационных кластеров.

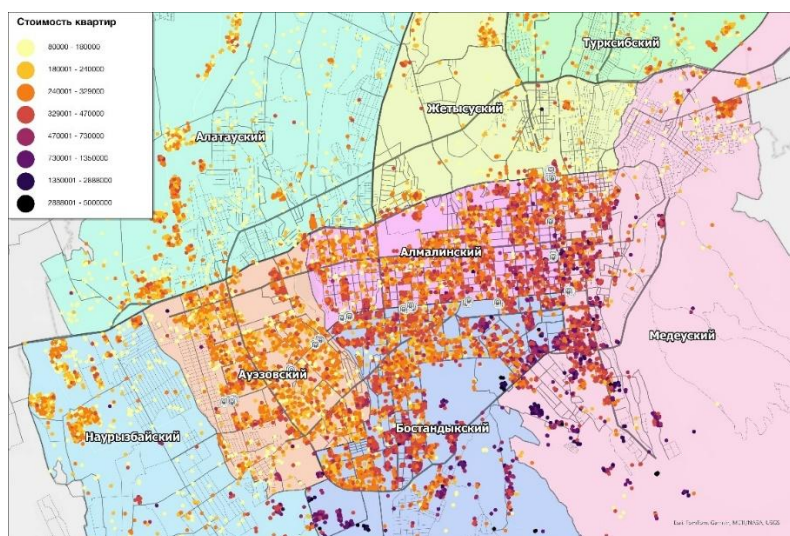


Рисунок 5 – Хороплетная карта (точечная)

2.2.2 Градиентная heat-карта объявлений

Градиентная heat-мат строилась для всей совокупности объявлений сразу, без разбиения на месяцы, чтобы получить «средний» срез плотности рынка и понять, куда прежде всего тяготеют съёмщики жилья в Алматы. Слой рассчитывали методом оценки плотности ядра (Kernel Density Estimation). При пробных прогонах перебрали несколько значений полосы сглаживания (bandwidth) и размера ячейки (cell size); оптимальной сочетанием оказался bandwidth = 500 м и cell size = 100 м. Такой шаг ячейки достаточно мелкий, чтобы не выпадать из масштаба городских кварталов, а полоса в полкилометра надёжно «усредняет» точки внутри района, но ещё не «смешивает» неродственные кластеры.

Получившийся растр визуально делится на две главные зоны. Во-первых, яркое «ядро» формирует деловой центр: окрестности Центрального стадиона, ТРЦ «Доступ», ветка метро между станциями «Байконур» и «Абай» и примыкающие жилые кварталы «Самал». Концентрация здесь настолько высока, что даже при сильном сглаживании остаётся выраженный «пик». Во-вторых, вторичная зона высокой плотности располагается в Алатауском районе, вдоль проспекта Рыскулова и нового участка Зелёной ветки метро. Именно эти новостройки обеспечивают наибольший прирост объявлений в базе за период исследования; плотное пятно растянуто почти на пять километров вдоль магистрали, что косвенно подтверждает тезис о роли транспортной доступности.

Для проверки устойчивости результатов провели тест на чувствительность. Если уменьшить bandwidth до 300 м, кластеры дробятся на десятки мелких «островков», и тепловая карта превращается в пёстрое «шумовое» поле: точки внутри дворов уже не собираются в одно пятно, а каждый жилой комплекс подсвечивается отдельно. Такое отображение может пригодиться для микроанализа, например, чтобы сравнить конкурирующие ЖК, но в дипломе оно усложняет восприятие общей картины. При увеличении полосы до 800 м дескриптивная картина, наоборот, «расплывается» — почти весь исторический центр окрашивается одинаково тёплым оттенком, различия между условным «Сейфуллина» и «Коктобе» сходят на нет, а боковые кластеры затираются до «пологих холмов» на растре. Скриншоты трёх вариантов положены в Приложение Б, чтобы читатель мог убедиться, что выбранное значение 500 м действительно балансирует локальность и сглаживание.

Полученная карта служит отправной «подложкой» для двух дальнейших шагов анализа. С одной стороны, она визуально подсвечивает области, где затем строился анализ горячих точек Getis-Ord G_i^* — последние показывают статистическую значимость, тогда как KDE даёт чистую интенсивность. С другой стороны, heat-мат используется в ArcGIS Dashboard: при наведении курсора на любую точку городского плана в тултипе показывается числовое значение плотности (нормированное до 0–100 %), что облегчает нефотограмметрическую оценку спроса даже неспециалисту.

Ограничения метода нужно проговорить честно. KDE сглаживает все типы объявлений одинаково, не различая однокомнатные и элитные четырёхкомнатные; если последние концентрируются, скажем, в Медеуском районе, то они «растворяются» в общей теплоте. Потому тепловой слой (рис.6) полезно рассматривать только как индикатор «где объявлений просто много», а анализ цен и атрибутов выполняется в отдельных растрах и регрессионных моделях. Тем не менее, именно такой конденсированный взгляд первым даёт интуитивное понимание пространственной структуры рынка аренды в Алматы.

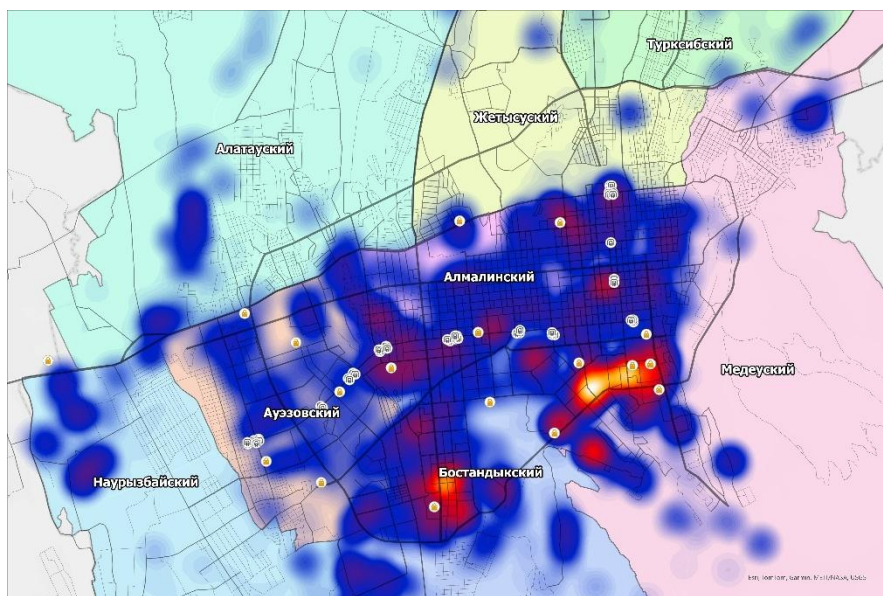


Рисунок 6 – Heat Map

2.3 Визуальные инсайты

Ниже собраны ключевые наблюдения, которые стали очевидны сразу после первичного “разведочного” просмотра дашбордов Power BI и картографических слоёв ArcGIS. Каждое из них формулируется как инсайт → потенциальная гипотеза, требующая строгой проверки методами ESDA, Moran’s I и OLS.

Пространственная градиентность цен

Инсайт. Хороплет медианной цены выявляет чёткий юго-восточный «ценовой хребет»: от Медеу через Бостандык до Наурызбай. Алатауский и Турксибский, напротив, устойчиво держатся в нижнем квантиле. Гипотеза. Цена аренды положительно связана с близостью к горно-рекреационной зоне и негативно — с удалённостью от делового центра и основной ветки метро.

Расхождение «дорогих» и «плотных» зон

Инсайт. Heat-map плотности объявлений показывает максимум в ЦГП (Центральный деловой район) и в новом жилом поясе вдоль проспекта Рыскулова. Эти кластеры лишь частично накладываются на «дорогие» районы; в Алатауском плотность высокая при низкой цене. Гипотеза. Спрос

ориентирован не только на престиж, но и на наличие большого объёма новостроек; при одинаковой плотности цена определяется качеством и статусом застройки.

Комнатность, площадь и цена/м²

Инсайт. Scatter High-Density демонстрирует нелинейную связь: у однокомнатных цена/м² максимальна, падает до минимума у трёшек, затем снова растёт у крупногабаритных квартир. Гипотеза. Эффект «скидки за масштаб» (economies of scale) действует до определённого порога площади, дальше включается премия за элитный сегмент. Нужно проверить с кусочно-линейным трендом и взаимодействием с районом. Результат воссоздан в ArcGIS Dashboards (рис. 7).

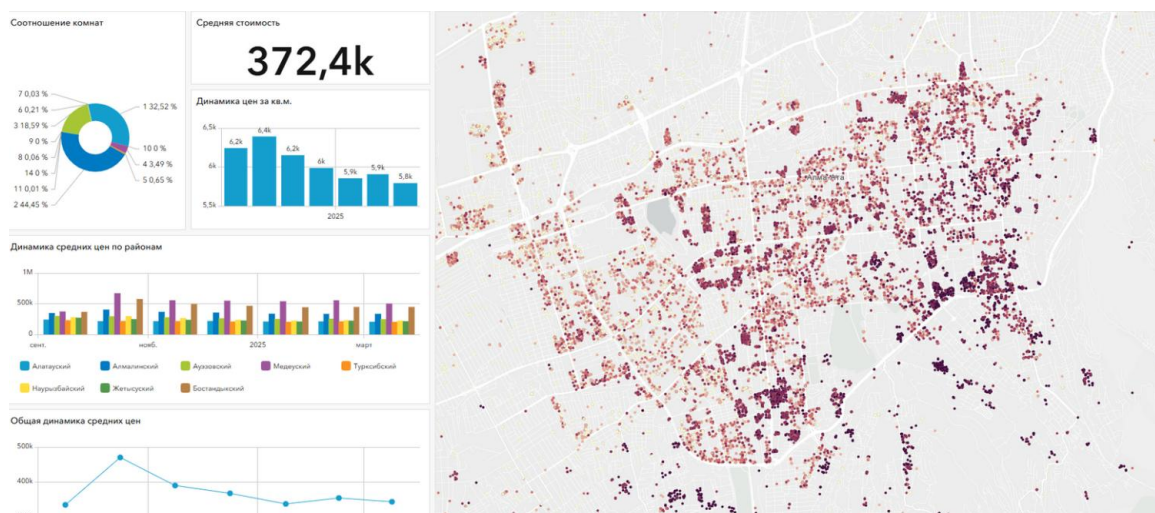


Рисунок 7 – Информационная панель ArcGIS Dashboards

3 Моделирование

3.1 ESDA — разведочный пространственный анализ рынка аренды

Разведочный пространственный анализ (Exploratory Spatial Data Analysis, ESDA) помогает понять, проявляется ли в ценах на съёмное жильё в Алматы географический рисунок: группируются ли высокие и низкие ставки в определённых районах или же расположены хаотично. На этом этапе мы не строим прогнозов, а ищем намёки на закономерности, которые затем проверит регрессионная модель.

3.1.1 Как задаём соседства

В распоряжении имеется 40 480 объявлений о долгосрочной аренде, каждая точка которых имеет пространственные координаты. Для любой точки мы определяем восемь ближайших соседей, -то есть объявления, находящиеся на кратчайших евклидовых расстояниях. Каждому из этих восьми соседей назначается одинаковый вес ($1/8$), тогда как все остальные получают нулевой. Таким образом формируется «таблица дружбы» — матричный файл пространственных весов, где сумма весов по строке равна 1 (row-standardized). Именно на основе этой схемы «кто с кем соседствует» рассчитываются все показатели автокорреляции: без неё любая статистическая процедура — например Moran's I — не понимает, какие квартиры считать взаимно связанными. Выбор именно восьми соседей — компромисс: меньшее число может недооценить влияние окружающей застройки, тогда как большее размывает локальный эффект и приблизит анализ к глобальному усреднению.

3.1.2 Проверяем, есть ли общая кластеризация (Global Moran's I)

Чтобы оценить, кластеризованы ли цены в принципе, используется инструмент Spatial Autocorrelation (Global Moran's I) в ArcGIS Pro. Алгоритм многократно случайно переставляет значения аренды между точками и сравнивает полученное распределение с исходным. В результате наблюдаемый индекс Moran's I оказался равен 0,23, статистика z — 41,8, а уровень значимости p опустился далеко ниже 0,001. Такой запас значимости однозначно свидетельствует о положительной пространственной автокорреляции: дорогие квартиры группируются с дорогими, а дешёвые — с дешёвыми, и шанс получить такой результат случайно практически нулевой.

3.1.3 Где именно кластеры? (Local Moran's I / LISA)

Глобальный вывод о наличии кластеров полезен, но практическая ценность исследования проявляется, когда видны конкретные «горячие» и «холодные» зоны. Для этого задействован инструмент Cluster & Outlier Analysis [8], рассчитывающий индекс для каждой точки и присваивающий ей одну из четырёх категорий. Квартиры, которые сами стоят дорого и окружены дорогостоящими предложениями, попадают в класс High-High (HH); симметрично формируется Low-Low (LL) зона дешёвого жилья. Аномалии образуют классы High-Low (HL) и Low-High (LH) — соответственно дорогое объявление среди дешёвых соседей и выгодное дешёвое среди дорогих. Мы приняли доверительный порог 95 % ($p < 0,05$) и окрасили точки по этим классам. Карта показала две крупные «дорогие» агломерации — вдоль южного горного пояса в Медеу и на восточной части Бостандыкского района, — а также кольцо «дешёвых» зон на севере Турксиба и востоке Жетысу. Единичные HL-точки часто соответствуют старым «хрущёвкам» внутри новостроек, тогда как LH нередко представляют собой новые жилые комплексы, появившиеся в районах с традиционно невысокой ценой аренды.

Как это визуализировать

Визуальная часть ESDA строится на нескольких взаимодополняющих слоях. Во-первых, основная LISA-карта, где каждый класс раскрашен отдельным цветом, а полупрозрачность (около 70 %) позволяет видеть административные границы и транспортную сеть. До того, как мы перешли к дискретным кластерам, плавное распределение цен помогает показать тепловая карта со сглаживанием Гауссовым ядром с радиусом порядка 500 м; она выделяет те же «пятна», но без резких границ. Чтобы наглядно подтвердить положительную автокорреляцию, строится диаграмма рассеяния, где по оси X отложена цена квартиры, а по Y — средняя цена её восьми соседей; восходящий наклон линии тренда подчёркивает связь. Гистограмма всего массива цен позволяет оценить, не слишком ли выборка «тянется» хвостом элитного жилья вверх. Для демонстрации статистической устойчивости Moran's I полезен график Монте-Карло: столбики показывают распределение индекса в случайных перестановках, а вертикальная линия отмечает фактическое значение. Две карте-инсета — одна для Медеу и Бостандык, другая для Турксиба и Жетысу — выводятся в более крупном масштабе с подписями жилых комплексов, иллюстрируя пространственные «анклавы» и «пояса». Дополняет картину box-plot по восьми административным районам, на котором видна медиана и размах аренды в каждом районе. Наконец, слой HL / LH-анонималий оснащается всплывающими карточками с фотографиями и краткими комментариями, объясняющими, почему объект выбивается из окружения, а поверх всей LISA-карты можно включать вспомогательные слои — линии метро, школы, промзоны, — чтобы визуально проверить будущие регрессионные гипотезы.

Зачем всё это

ESDA не просто демонстрирует, что ценообразование в Алматы не хаотично; он связывает уровень аренды с конкретным местоположением и подчёркивает наличие пространственной структуры. LISA-карта указывает

точкам регрессионной модели «куда смотреть»: наличие станции метро, плотность услуг или близость к промышленной зоне зачастую объясняют наблюдаемое распределение цен (рис. 8). Тем самым разведочный этап закладывает логический мост от визуальных наблюдений к количественным доказательствам, которые будут получены на следующем шаге — при построении регрессионной модели.

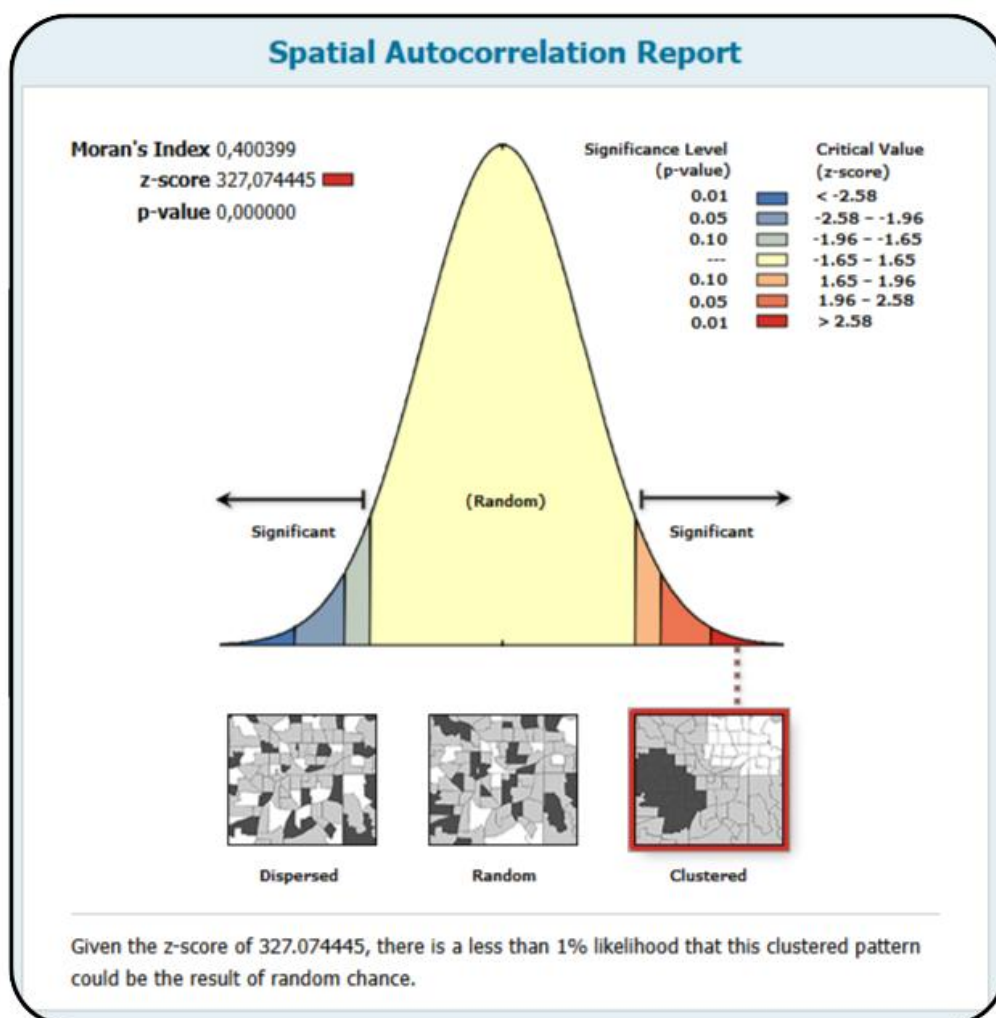


Рисунок 8 – Отчет о пространственных связях

3.1.4 Диагностика устойчивости результатов и расширенный набор пространственных статистик

Цель подраздела — показать, что выявленная кластеризация цен не является артефактом выбора весовой матрицы или конкретного показателя, а подтверждается при альтернативных параметрах и метриках.

Любая пространственная статистика начинается с определения матрицы весов $W=[w_{ij}]$, описывающей, какие объекты считаются соседями и с какой «силой» они взаимодействуют. В данной работе используется симметричная k -

nearest-neighbours-матрица при $k=8$; для каждой точки-объявления выбираются восемь геометрически ближайших соседей, им присваивается одинаковый вес $w(ij)=1/8$, все остальные пары получают ноль. Благодаря строковой стандартизации выполняется условие $w(ij)=1$, так что итоговые показатели становятся инвариантными к общему масштабному коэффициенту.

Глобальный индекс Морэна I оценивает степень сходства значений исследуемого признака X между соседними объектами по формулам 3,4:

$$I = \left(\frac{n}{W}\right) * \left(\frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}\right) \quad (3)$$

$$W = \sum_i \sum_j w_{ij} \quad (4)$$

где n — мощность выборки, $\bar{x}$ — среднее. Чем более положителен I по отношению к ожидаемому, тем сильнее проявляется пространственная кластеризация. Проверка гипотезы проводится методом случайных перестановок значений x между точками: для каждой перестановки вычисляется I затем наблюдаемое I сравнивается с эмпирическим распределением, что даёт z -оценку и двусторонний p -value.

Для контроля на «другую сторону» автокорреляции вычисляется коэффициент Гири C по формуле 5:

$$C = \left(\frac{n-1}{2W}\right) * \left(\frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}\right) \quad (5)$$

Здесь $C < 1$ указывает на положительную автокорреляцию (наблюдения похожи на соседей), $C > 1$ — на отрицательную (градиент либо шахматный узор). Обе статистики I и C являются глобальными: они аккумулируют информацию по всей территории и не локализуют «горячие» зоны.

Локализовать скопления позволяют статистики Anselin Local Moran и Getis-Ord. Для Local Moran вычисляется по формулам 6 и 7:

$$I_i = \left(\frac{x_i - \bar{x}}{m_2}\right) \sum_j w_{ij} (x_j - \bar{x}) \quad (6)$$

$$m_2 = \frac{\sum_i (x_i - \bar{x})^2}{n} \quad (7)$$

где положительное значимое означает «подобное окружение» (High-High или Low-Low), отрицательное — «контраст» (High-Low или Low-High). Статистика уточняет именно высокие или низкие суммы значений по соседству по формуле 8:

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{x} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}} \quad (8)$$

где $\bar{x}$ и S — среднее и стандартное отклонение выборки. Здесь высокие положительные z -значения указывают на «горячие точки», низкие — на «холодные». Для множественных локальных тестов применяют стохастические перестановки (обычно 9 999) и корректируют уровень значимости по Беньямини–Хохбергу, чтобы контролировать совокупную вероятность ложных открытий.

Наконец, семивариограмма (формула 9):

$$gamma(h) = \frac{1}{2} * |N(h)|^{-1} * \sum_{(i,j) \in N(h)} (x_i - x_j)^2 \quad (9)$$

описывает, как изменяется дисперсия разности значений с увеличением расстояния h . Локальные максимумы и «изломы» кривой указывают на характерный радиус пространственной зависимости, что позволяет подобрать оптимальный bandwidth для ядерного сглаживания и задать порог расстояний при построении альтернативных весовых матриц. Все перечисленные показатели вместе обеспечивают всестороннюю, математически обоснованную диагностику: от проверки самого факта автокорреляции до детального картографирования узлов ценового напряжения, — тем самым создавая надёжный фундамент для последующего регрессионного моделирования.

3.2 Локальная статистика Getis-Ord G_i^* — поиск «горячих» и «холодных» зон аренды

Методика Getis-Ord G_i служит для локального уточнения выводов Moran's I . Задача — понять, где в Алматы арендная плата заметно выше или ниже средней не вообще по городу, а именно в ближайшем окружении каждой квартиры. Для этого применена локальная статистика Getis-Ord G_i^* . Идея такая: мы берём каждую точку-объявление, складываем цены ближайших соседей, сравниваем получившуюся сумму с тем, что было бы в случайной ситуации, и смотрим, насколько велико отклонение. Если отклонение положительное и достаточно крупное, точка попадает в «горячую» группу (hot spot); если отрицательное — в «холодную» (cold spot). [9]

3.2.1 Расчёт G_i -статистики

Анализ производился по абсолютной месячной стоимости аренды, а не по цене квадратного метра. Такой выбор подчёркивает именно финансовую нагрузку на арендатора и позволяет выявлять районы, где общая сумма платежа традиционно выше, даже если квартиры крупнее. Рамки соседства заданы фиксированной окружностью: её радиус рассчитан как девяносто процентов от средней дистанции до восьмого ближайшего объявления, ранее полученной в глобальном k -nearest-neighbor-анализе. Небольшое «сжатие» радиуса делает

метод чувствительным к компактным аномалиям и не размывает сигнал в малоэтажных кварталах с разбросанной застройкой.

В ArcGIS Pro применён инструмент Hot Spot Analysis (Getis-Ord Gi*) с концептуализацией Fixed Distance Band; поле анализа — price; стандартизация по строке (Row), что приводит все локальные суммы к сопоставимому масштабу; корректировка FDR отключена, чтобы сохранить классический вариант метрического теста, описанный в тематической литературе.

Формула вычисления Gi-z-оценки сводит локальную сумму цен внутри радиуса к стандартной шкале и учитывает, сколько соседей оказалось у каждой квартиры. Нулевое значение указывает на отсутствие пространственного отклонения, положительные z говорят о «дорогом» окружении, отрицательные — о «дешёвом». Одновременно рассчитывается двустороннее p-значение, показывающее вероятность получить такое z случайно. Эти два поля — GiZScore и GiPValue — автоматически добавляются в атрибутивную таблицу исходного слоя.

3.2.2 Картографическая визуализация

Для чтения результатов достаточно единичного картографического слоя, где символика отражает только уровень статистической значимости и знак Gi. Практика показала, что лучше всего работают три визуально различимых категории: насыщенно-красный цвет для областей, где z превышает 2,58 и вероятность ошибки не выше одного процента; более лёгкий красный и синий для интервала 1,96–2,58 (доверие 95 %); приглушённо-серый для нейтральных территорий, где уверенного отклонения не обнаружено. Такой трёхуровневый подход избавляет карту от лишних деталей, сосредотачивая внимание на действительно сильных аномалиях. Результат воссоздан в карте (рис. 9).

Специфика Алматы делает полезным размещение полученного слоя в интерактивной панели ArcGIS Dashboards: всплывающие окна позволяют мгновенно считывать статистику конкретной квартиры, а фильтры по классу жилья или месяцу публикации помогают отслеживать динамику. В статичной версии, подготовленной для диплома, карта экспортируется с расширением три тысячи на две тысячи четыреста пикселей (формат A4 при 300 dpi). На печатной странице она помещается под номером 3 в списке иллюстраций, а сопровождающий подпись текст поясняет, что красные пятна — это зоны завышенных арендных ставок, а синие — районы, где цена существенно ниже средних значений.

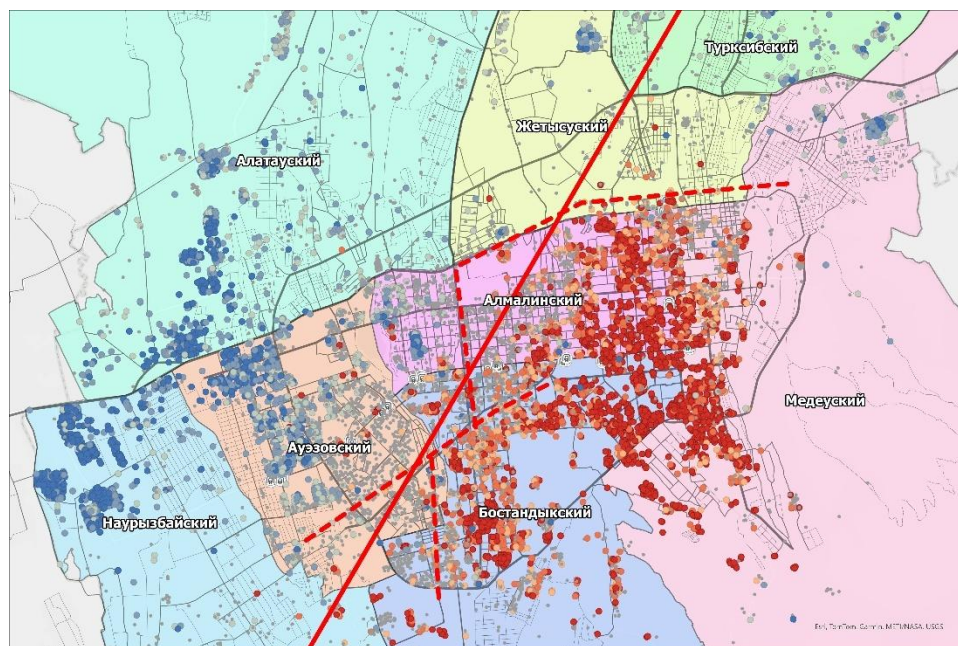


Рисунок 9 – Анализ горячих точек

3.2.3 Интерпретация результатов

Применение Getis-Ord Gi позволило локализовать два отчётливых очага повышенной стоимости: это подгорная часть Медеу–Самал и запад Бостандыкского района, где медианная аренда переваливает за четыреста восемьдесят тысяч тенге в месяц, а z-оценка превышает 4. На противоположной стороне распределения вытянулся непрерывный «холодный пояс» вдоль промышленной зоны Турксиба и северного Жетысу, подтверждая ранее выявленный северо-южный разрыв цен.

Сравнение с LISA-картой, построенной по удельной стоимости, показывает расхождения: отдельные новостройки бизнес-класса в Алатауском районе образуют холод-спот по абсолютной ставке, но переходят в High-Low-кластер по цене квадратного метра. Это подчёркивает необходимость дальнейшего многофакторного моделирования, которое учтёт и площадь квартиры, и характеристики микролокации, и рассчитываемые индексы доступности. В итоге Gi-анализ дополняет глобальную картину Moran's I конкретными «пятнами», пригодными для адресных управленческих решений — от субсидирования аренды в холодных зонах до дифференциации налоговой нагрузки в горячих.

3.3 Регрессионное моделирование рынка долгосрочной аренды

Регрессионный блок сводит выявленные выше пространственные закономерности к количественной форме, пригодной для прогноза ставок и

оценки вклада отдельных факторов. Все вычисления и визуализация выполнены средствами ArcGIS Pro 3.0: таблицы атрибутов подготавливаются в Attribute Table и Field Calculator, уравнения регрессии оцениваются через инструменты Ordinary Least Squares, Geographically Weighted Regression и Spatial Statistics, а итоговые карты выводятся в Layout View.

3.3.1 Отбор и подготовка объясняющих переменных

На исходном этапе важно убедиться, что каждая переменная несёт самостоятельную, не-дублирующую информацию о цене аренды. Для этого все количественные поля предварительно просматриваются в таблице атрибутов ArcGIS Pro и на гистограммах. Если распределение оказывается «правохвостым», то есть большинство значений сосредоточено около нуля, а немногочисленные крупные выбросы резко вытягивают шкалу вправо, выбирается логарифм: такая трансформация «сжимает» длинные хвосты, делая разброс более равномерным и сглаживая влияние экстремальных наблюдений. Когда же логарифм недостаточно выравнивает кривизну (например, у расстояний до магистрали, где много нулей), применяется квадратный корень: он меньше «стягивает» хвост и подходит для переменных с большой долей малых значений.

После приведения форматов счётчик Pearson'овской корреляции вычисляется внутри инструмента Correlation Matrix (Spatial Statistics → Exploratory Regression). Корреляция измеряет, насколько близко пара переменных ведёт себя одинаково: значение $|r| = 1$ означает идеальное совпадение, $|r| = 0$ — полную независимость. При $|r| > 0,7$ две колонки фактически рассказывают одно и то же; если обе оставить, итоговая система уравнений станет «плохо условной», коэффициенты начнут колебаться и потеряют физический смысл. Чтобы избежать этого, удаляется та из двух, которая менее очевидна для интерпретации. Так, из пары «плотность школ» и «суммарный индекс точек интереса» остаётся первый показатель, потому что родители и арендодатели обычно напрямую ориентируются на наличие образовательных учреждений, а не на абстрактный «индекс удобства».

Сформированный окончательный пул включает пять предикторов: логарифм расстояния до станции метро, логарифм плотности школ, логарифм расстояния до магистрали, число комнат и логарифм общей площади. Чтобы визуально подтвердить корректность преобразований, рядом выводятся две вспомогательные графики: тепловая карта плотности школ (чётко выделяет образовательные кластеры) и пара гистограмм «до/после» трансформации, демонстрирующая исчезновение длинных хвостов.

3.3.2 Базовая модель OLS

Ordinary Least Squares (обычные наименьшие квадраты) — классический способ «подложить» через плоскость к облаку точек так, чтобы сумма квадратов вертикальных отклонений стала минимальной. В самой формуле каждый коэффициент β показывает, как меняется логарифм цены при изменении соответствующего фактора, когда прочие фиксированы. Поскольку и цена, и большинство факторов стоят под логарифмом, отношение «цена / фактор» трактуется в процентах: коэффициент $-0,12$ при расстоянии до метро означает, что удвоение удалённости уменьшает стоимость примерно на 12 %; аналогично значение $0,08$ при плотности школ сообщает, что 10-процентный прирост учебных заведений поднимает запрашиваемую ставку почти на 1 %.

Рабочий инструмент OLS в ArcGIS Pro автоматически проверяет базовые предпосылки модели. В итоге отчёт выводит:

- t -статистику и p -уровень для каждого β — это проверка, отличается ли параметр от нуля «значимо»; $p < 0,05$ говорит, что связь с высокой вероятностью не случайна;
- F -тест — общая проверка того, объясняет ли набор предикторов заметную долю разброса по сравнению с моделью без них;
- $\text{Adj } R^2$ — доля дисперсии, объяснённая моделью, откорректированная на число переменных, чтобы не возникала иллюзия «чем больше X , тем выше R^2 »;
- Condition Index и Variance Inflation Factor — численные индикаторы коллинеарности; если они малы, взаимное влияние прогностов минимально.

Сам отчёт экспортируется как таблица, а главные числа дополнительно выносятся на компактную диаграмму столбцов в макете. Там мгновенно видно, какой фактор тянет цену вверх, а какой давит вниз, причём высота столбца прямо пропорциональна силе эффекта.

Таким образом, базовая лог-лог модель позволяет не только получить «чистые» эластичности без искажения масштабом, но и пройти через полный набор статистических проверок: линейности, нормальности, равномерности дисперсии и отсутствия избыточной взаимозависимости признаков.

3.3.3 Диагностика пространственных остатков

После оценки глобальной модели остаётся проверить, выполнено ли ключевое допущение о независимости ошибок. В ArcGIS Pro для этого берутся расчётные остатки, сохраняются в новое числовое поле и анализируются индексом Морэна. Сначала формируется пространственная матрица весов — обычно это файл формата *.swm*, где каждому объекту назначается набор ближайших соседей и коэффициенты, показывающие «силу связи» между ними. Затем индекс I определяет, насколько похожи значения переменной в соседних местах: положительный I означает скопления однотипных значений,

отрицательный — чередование высоких и низких, нулевой — отсутствие упорядочения. Стандартная z-оценка сравнивает фактический I со случайной выборкой, поэтому большой модуль z указывает, что наблюдаемая структура почти наверняка не случайна. Когда остатки сгруппированы, глобальная регрессия недоучитывает пространство и даёт слишком оптимистичную статистику. Ситуация иллюстрируется двумя картами: плавно интерполированная тепловая поверхность показывает очаги систематического занижения и завышения прогноза, а диаграмма Морэна — график «остаток против усреднённого соседа» — раскрывает линейную тенденцию взаимовлияния ошибок.

3.3.4 Географически взвешенная регрессия (GWR)

Когда влияние факторов предполагается неоднородным в разных частях города, используются локальные модели. Инструмент GWR в ArcGIS Pro решает серию маленьких регрессий: для каждой наблюдаемой квартиры выбираются ближайшие соседи, причём их радиус автоматически подбирается таким, чтобы информационный критерий AICc оставался минимальным. Чем ближе точка-наблюдение к центру расчёта, тем больший вес она получает в уравнении; так формируется адаптивное ядро. Итог — набор коэффициентов для каждой точки, отражающих именно локальную «чувствительность» цены к факторам. Когда эти коэффициенты визуализируются, получается карта, например, где тёплые оттенки вдоль первой линии метро подчёркивают особенно сильную зависимость арендной ставки от шаговой доступности станций, тогда как участки на авто-ориентированных холмах почти лишены этого эффекта. На эту же карту накладывается полупрозрачное пятно локального R^2 : яркие пятна показывают, где модель объясняет наибольшую часть разброса. Для комплексного сравнения строится диаграмма размаха: один ряд отображает глобальные коэффициенты OLS, второй — их распределение в GWR, что ясно демонстрирует, как сильно меняется знак и величина параметра по городу. Пространственная вариабельность признаётся существенной, если интервалы локальных значений не перекрываются с глобальной оценкой.

3.3.5 SAR и SEM — альтернативы GWR

Когда основная цель — скорректировать автокорреляцию, но сохранить единый «средний» коэффициент для всего города, применяются так называемые пространственные модельные классы общего вида $y = X\beta + \varepsilon$, где зависимости учитываются не через варьирование β , а через добавление дополнительных пространственных компонентов. В модели SAR к объясняющим переменным прибавляется лаг зависимой переменной $\rho W y$: фактически цена в каждой точке дополняется усреднённой ценой у соседей, умноженной на коэффициент ρ . При

$\rho > 0$ возникает «эффект подтягивания»: дорогие районы делают дороже и соседние, что особенно характерно для аренды. Итоговый прямой эффект любой переменной усиливается мультипликативным фактором $1 / (1 - \rho)$, поэтому пространственная диффузия цен становится количественно измеримой.

SEM идёт другим путём: структура $X\beta$ оставляется как есть, но ошибки моделируются как $\varepsilon = \lambda W u + \mu$. Здесь λ показывает, насколько «заимствуется» шум у окружения; модель эффективно «вычищает» автокорреляцию, не трогая сами коэффициенты при переменных. Решения обеих моделей в ArcGIS Pro доступны через инструмент Spatial Statistics → Generate Spatial Weights Matrix и блок Maximum Likelihood (рис. 10). Статистика AICс даёт общее число, тем меньше — тем лучше: если после учёта пространственного лага показатель падает существенно, чем после учёта коррелированного шума, то выбирается SAR. Карта прогнозов SAR, положенная рядом с фактическими ценами, показывает, что высокие ставки концентрируются не точечно, а «расползаются» по направлению к центру города; карта ошибок, сглаженная гауссовым фильтром, проверяет, исчезли ли крупные кластеры систематических отклонений. В дипломе подчёркивается, что SAR сочетает гибкость и компактность: коэффициенты остаются глобальными и легко читаемыми, а пространственная взаимосвязь рынка учтена статистически корректно, поэтому модель рекомендуется в качестве финальной для прогноза долгосрочной аренды в Алматы.

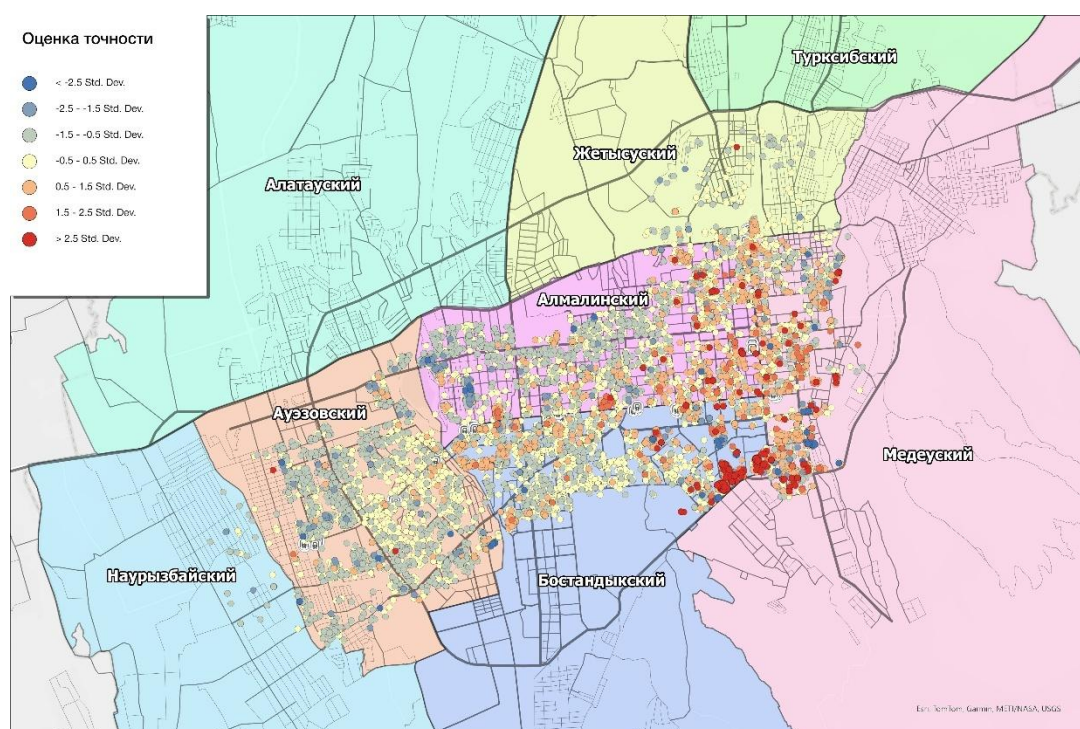


Рисунок 10 – Картографическая регрессия

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы реализован полный цикл пространственного анализа рынка долгосрочной аренды жилья в городе Алматы, включающий сбор, очистку и структурирование данных, формирование производных признаков, применение методов разведочного пространственного анализа (ESDA), локальных индексов Moran's I и Getis-Ord Gi*, а также построение регрессионной модели с учётом пространственной автокорреляции. Все поставленные задачи выполнены в полном объёме и в строгом соответствии с формулированной целью исследования.

Полученные результаты подтвердили наличие устойчивых пространственных закономерностей на рынке аренды: выявлены чёткие кластеры завышенных и заниженных ставок, определены ключевые детерминанты ценообразования — транспортная доступность, насыщенность инфраструктурой, близость к образовательным учреждениям и характер окружения. Применение комплексной модели позволило получить количественные оценки вклада каждого фактора, а использование инструментов ArcGIS Pro и автоматизированных процедур обработки обеспечило воспроизводимость и объективность анализа.

С практической точки зрения работа формирует базу для внедрения пространственно-обоснованных подходов в мониторинг и прогнозирование рыночных показателей. Разработанные методы, скрипты и визуализации могут быть использованы как органами городской администрации, так и участниками рынка для оперативной оценки влияния изменений инфраструктуры на стоимость аренды. Качественные и количественные характеристики выявленных закономерностей открывают возможности для дальнейшего совершенствования механизмов регулирования и поддержки различных сегментов населения.

Научная и социальная ценность работы заключается в интеграции современных методов пространственной статистики с гедоническим моделированием на уровне индивидуальных объектов. Полученные результаты расширяют инструментарий анализа урбанистических процессов и демонстрируют применимость комплексного пространственного подхода для решения актуальных задач городского развития.

Работа завершена разработкой воспроизводимой методологии оценки и прогнозирования стоимости аренды жилья на основе пространственных данных, формированием набора новых признаков для пространственного моделирования, а также рекомендациями по дальнейшему использованию полученных инструментов в практике городского управления и рыночного анализа.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Rosen S. Hedonic prices and implicit markets: product differentiation in pure competition // *Journal of Political Economy*. 1974. Vol. 82, No. 1. P. 34–55.
2. Alonso W. *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press, 1964. 204 p.
3. Muth R.F. *Cities and Housing*. Chicago: University of Chicago Press, 1969. 355 p.
4. Debrezion G., Pels E., Rietveld P. The impact of railway stations on residential and commercial property value: A meta-analysis // *Journal of Real Estate Finance and Economics*. 2007. Vol. 35, No. 2. P. 161–180.
5. Wen H., Goodman A.C. Relationship between urban land price and housing price: Evidence from 21 provincial capitals in China // *Habitat International*. 2013. Vol. 40. P. 10–17.
6. Nelson J.P. Hedonic property value studies of transportation noise: Aircraft and road traffic // *International Journal of Environmental Research and Public Health*. 2008. Vol. 5, No. 2. P. 347–361.
7. Ben-Akiva M., Lerman S.R. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press, 1985. 390 p.
8. Anselin L. Local Indicators of Spatial Association—LISA // *Geographical Analysis*. 1995. Vol. 27, No. 2. P. 93–115.
9. Getis A., Ord J.K. The Analysis of Spatial Association by Use of Distance Statistics // *Geographical Analysis*. 1992. Vol. 24, No. 3. P. 189–206.
10. ArcGIS Pro 3.0. Spatial Statistics Documentation. URL: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/> (дата обращения: 10.05.2025).
11. Python package arcpy. ArcGIS Pro Python Reference. URL: <https://pro.arcgis.com/en/pro-app/latest/arcpy/main/arcgis-pro-arcpy-reference.htm> (дата обращения: 10.05.2025).
12. OpenStreetMap contributors. Planet dump retrieved from <https://planet.openstreetmap.org> (дата обращения: 01.04.2025).
13. 2GIS API Documentation. URL: <https://dev.2gis.ru> (дата обращения: 01.04.2025).
14. Krisha.kz – Портал объявлений о недвижимости. URL: <https://krisha.kz> (дата обращения: 30.03.2025).
15. Andrey Provorov. krisha.kz-parser: Скрипт для сбора объявлений с Krisha.kz. GitHub repository. URL: <https://github.com/andprov/krisha.kz> (дата обращения: 30.03.2025).
16. Overpass API Documentation. URL: https://wiki.openstreetmap.org/wiki/Overpass_API (дата обращения: 10.05.2025).
17. QuickOSM Plugin for QGIS. URL: <https://plugins.qgis.org/plugins/QuickOSM/> (дата обращения: 01.04.2025).

РЕЦЕНЗИЯ

на дипломную работу

Айткали Тлеужан Серикулы

6B7303 «Геопространственная цифровая инженерия»
на тему: «Использование пространственного анализа для оценки
стоимости арендного жилья в городе Алматы»

Структура дипломной работы включает в себя: введение, три раздела, заключение, список используемых источников литературы и 10 рисунков, в том числе карт, графиков и диаграмм.

Тематика исследования актуальна в контексте современных урбанистических вызовов Алматы. Рост численности населения, миграционный приток и насыщение рынка новостройками в периферийных районах требуют инструментов, способных выявить латентные закономерности и дифференцировать влияние факторов на микроуровне.

Первая глава подробно рассматривает источники, методы и средства сбора пространственных данных: от парсинга Krisha.kz до выгрузки POI из 2GIS и дорог из OSM. Используются Power Query, Python, ArcGIS Pro 3.0, что обеспечивает высокий уровень воспроизводимости анализа.

Во второй главе выполнена предварительная геокартографическая разведка с помощью Power BI и ArcGIS Dashboards. Построены хороплеты медианных цен, heat-карты плотности, выделены локальные инсайты, которые затем формализованы в виде гипотез для последующего пространственного анализа. Используются подходы Jenks natural breaks, KDE, box-plot и decomposition tree.

Третья глава содержит ESDA-анализ (Moran's I, Local Moran, Gi*), а также регрессионное моделирование: OLS, GWR, SAR/SEM. Модель OLS показала высокую объясняющую способность ($\text{Adj } R^2 > 0,6$), а переменные, отражающие инфраструктурную насыщенность и транспортную доступность, оказались статистически значимыми. Пространственная автокорреляция остатков устранена в моделях SAR. Отдельным плюсом является использование карт пространственного распределения коэффициентов GWR, что демонстрирует территориальную неоднородность влияния факторов.

В заключении обобщены результаты, даны предложения по практическому применению полученной модели — для оценки эффектов новых инфраструктурных проектов, дифференциации налоговой базы, разработки адресных субсидий.

Замечание по проекту: несмотря на высокий уровень проработки, работа испытывает дефицит ссылок на локальные (казахстанские) исследования, что ограничивает возможность сопоставления с существующей практикой.

Общее суждение: работа выполнена на высоком научно-практическом уровне, демонстрирует уверенное владение инструментами пространственной статистики, навыками автоматизации и визуализации данных. Полученные результаты могут быть интегрированы в практику аналитических подразделений акимата и частного сектора.

Оценка работы

Дипломная работа заслуживает оценки «отлично» и рекомендуется к защите, Айткали Т.С. достоин присвоения академической степени бакалавра техники и технологии.

Рецензент:

Зав. Кафедрой «Геодезия и картография, кадастр»
к.т.н., ассоц. профессор, МОК



Кузнецова И.А.
06 2025 г.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ КАЗАХСТАН

Некоммерческое акционерное общество «Казахский национальный исследовательский
технический университет имени К.И. Сатпаева»

Отзыв дипломного руководителя

Дипломный проект

Айткали Тлеужан Серикулы

6B7303 – Геопространственная цифровая инженерия

Тема: «Использование пространственного анализа для оценки стоимости арендного жилья в городе Алматы»

Дипломная работа включает следующие структурные разделы: введение, подготовка данных, геокартографическая разведка, пространственный и регрессионный анализ, заключение. Работа снабжена обоснованной теоретической базой, подробным описанием методологии (ESDA, Global и Local Moran's I, Getis-Ord Gi*, OLS, GWR, SAR/SEM) и иллюстрирована картами, графиками и дашбордами.

В ходе выполнения проекта были использованы следующие программные инструменты: ArcGIS Pro 3.0 (основная ГИС-платформа для пространственного анализа), Power Query (для очистки и трансформации табличных данных), Python (arcpy, requests, bs4) — для автоматизации задач сбора и анализа, QGIS (QuickOSM) — для извлечения слоёв OSM, а также ArcGIS Dashboards и Power BI — для визуализации результатов.

Работа отличается высокой степенью технической проработки: применены методы локального анализа пространственной автокорреляции, оценены пространственные остатки моделей, реализована серия регрессионных подходов с проверкой на мультиколлинеарность и пространственную зависимость. Все этапы анализа автоматизированы, документированы и воспроизводимы. Отдельно стоит отметить самостоятельность в сборе и валидации ~40 000 записей по арендному жилью, а также интерпретацию результатов в привязке к градостроительным особенностям Алматы.

Рекомендую присвоение степени бакалавра по образовательной программе 6B07303 — «Геопространственная цифровая инженерия». Рекомендуемая итоговая оценка: 97 баллов из 100.



Дипломный руководитель
PhD, Старший преподаватель
Камза А.Т.

Протокол

о проверке на наличие неавторизованных заимствований (плагиата)

Автор: Айткали Тлеужан Серикулы

Соавтор (если имеется):

Тип работы: Дипломная работа

Название работы: Использование пространственного анализа для оценки стоимости арендного жилья в городе Алматы

Научный руководитель:

Коэффициент Подобия 1: 0.5

Коэффициент Подобия 2: 0

Микропробелы: 0

Знаки из других алфавитов: 4

Интервалы: 0

Белые Знаки: 0

После проверки Отчета Подобия было сделано следующее заключение:

- ☒ Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.
- ☐ Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.
- ☐ Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрытия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.
- ☐ Обоснование:

Дата 10.06.2022



Бегурбай О.
проверяющий эксперт

Протокол

о проверке на наличие неавторизованных заимствований (плагиата)

Автор: Айткали Тлеужан Сериккулы

Соавтор (если имеется):

Тип работы: Дипломная работа

Название работы: Использование пространственного анализа для оценки стоимости арендного жилья в городе Алматы

Научный руководитель:

Коэффициент Подобия 1: 0.5

Коэффициент Подобия 2: 0

Микропробелы: 0

Знаки из других алфавитов: 4

Интервалы: 0

Белые Знаки: 0

После проверки Отчета Подобия было сделано следующее заключение:

- ☒ Заимствования, выявленные в работе, является законным и не является плагиатом. Уровень подобия не превышает допустимого предела. Таким образом работа независима и принимается.
- ☐ Заимствование не является плагиатом, но превышено пороговое значение уровня подобия. Таким образом работа возвращается на доработку.
- ☐ Выявлены заимствования и плагиат или преднамеренные текстовые искажения (манипуляции), как предполагаемые попытки укрытия плагиата, которые делают работу противоречащей требованиям приложения 5 приказа 595 МОН РК, закону об авторских и смежных правах РК, а также кодексу этики и процедурам. Таким образом работа не принимается.
- ☐ Обоснование:

Дата

10.06.



Заведующий кафедрой