**Оглавление**

[1. Сравнительное исследование источников данных 1](#_Toc184565491)

[1.1 Витрина данных ФИАС 1](#_Toc184565492)

[1.2 Витрина данных 2ГИС 1](#_Toc184565493)

[1.3 Парсинг OpenStreetMap 2](#_Toc184565494)

[1.4 Парсинг 2ГИС 2](#_Toc184565495)

[2. Анализ погрешностей и их влияние 2](#_Toc184565496)

[3. Сравнительная точность 3](#_Toc184565497)

[4. Детализация категорий 3](#_Toc184565498)

[5. Итоговые выводы 3](#_Toc184565499)

[6. Рекомендации 4](#_Toc184565500)

# **Сравнительное исследование источников данных**

## **Витрина данных ФИАС**

* Предоставляет наиболее простую и обобщённую классификацию, где основная категория — "Дом" (9704 записей).
* Подходит для базового анализа жилого фонда без необходимости детализированной классификации зданий.

**Итог**: ФИАС хорошо охватывает жилые дома, но игнорирует встроенные помещения, такие как магазины, офисы или другие объекты.

## **Витрина данных 2ГИС**

* Значительное количество записей "Не определено" (8421).
* Классификация объектов ("Жилой дом", "Малоэтажный жилой дом", "Административное здание", "Общежитие") выглядит адекватной и полезной для задач, связанных с анализом типов зданий.
* Присутствует возможность учитывать встроенные помещения, такие как кафе, магазины и торговые центры, что может быть полезно в ряде сценариев.
* Встречаются уникальные типы, такие как "Кафе, бар", "Религиозное сооружение", "Аэропорт", "Торговый центр".

**Итог:** витрина данных 2ГИС сохраняет приемлемый баланс между детализацией и точностью.

## **Парсинг OpenStreetMap**

* Большое количество записей "Не определено" (2709) и ошибок (118).
* Присутствует множество уникальных категорий, таких как "Церковь", "Караульное помещение", "Промышленное здание", "Железнодорожный вокзал".
* Выявлены ошибки - переклассификация жилых домов как магазинов, церквей, гаражей и других встроенных объектов, что значительно снижает качество данных.
* Несмотря на хорошую детализацию, результаты для жилых домов оказываются менее надёжными, так как значительное количество зданий классифицировано неверно.

**Итог**: к использованию не рекомендуется. Многоквартирные дома с магазинами на первых этажах ошибочно определяются как коммерческие здания, что в значительной степени искажает анализ жилого фонда.

## **Парсинг 2ГИС**

* Данные по жилым домам выглядят более надёжными по сравнению с OpenStreetMap. Основные категории, такие как "Жилой дом", "Малоэтажный жилой дом" и "Частный дом", распределены корректно.
* Значительное количество записей "Не определено" (8586) остаётся проблемой.

**Итог**: 2ГИС предоставляет более точные данные о жилых домах обеспечивают лучшую основу для анализа жилого фонда.

# **Анализ погрешностей и их влияние**

* OpenStreetMap: Погрешности классификации делают его менее надёжным для анализа жилых домов. Использовать эти данные без дополнительной проверки нельзя.
* 2ГИС: несмотря на большое количество "Не определено", структура классификации кажется более надёжной.
* ФИАС: Простой подход к классификации делает его устойчивым к ошибкам, но непригодным для задач, требующих высокой детализации.

# **Сравнительная точность**

Доля "Не определено" является ключевым показателем точности:

* ФИАС: минимальная (0%), но за счёт обобщения типов зданий.
* Витрина 2ГИС: высокая (8421).
* Парсинг Open Street Map: средняя (2709).
* Парсинг 2ГИС: высокая (8586).
* Open Street Map выделяется как наиболее точный источник при значительном количестве "Не определено".

# **Детализация категорий**

* Open Street Map предоставляет наибольшую детализацию, включая редкие категории зданий (например, "Казарма", "Терраса", "Ангар"), что делает его ценным для анализа разнообразных типов зданий.
* 2ГИС хорошо классифицирует жилые дома, административные и образовательные здания, но часто не определяет категорию здания.
* ФИАС даёт обобщённые данные, которые могут быть полезны только в базовом анализе.

# **Итоговые выводы**

**Наиболее точный источник для анализа жилых домов: *Витрина данных 2ГИС***. Несмотря на высокую долю "Не определено", классификация объектов выглядит более надёжной.

**Источник для вспомогательных данных:** Парсинг 2ГИС.

**Источник, требующие уточнения:** OpenStreetMap. Для использования необходимо проводить дополнительную ручную проверку или исключать категории, подверженные ошибочной классификации (например, "Магазин", "Церковь", "Гаражи").

# **Рекомендации**

1. **Использовать данные из витрины 2ГИС** для анализа жилых зданий как наиболее точный и надёжный источник.
2. **Исключение OpenStreetMap.** Из-за высокого уровня ошибок использовать данные OpenStreetMap только для специфических категорий (например, промышленных или религиозных объектов) после ручной проверки.
3. **Совмещение данных.** Для улучшения точности возможно объединение данных ФИАС и 2ГИС, учитывая, что ФИАС может служить базовым уровнем для проверки присутствия зданий, а 2ГИС — для уточнения типов.