



Концепция ранжирования районов города

**по уровню самодостаточности
на основе больших данных**

Артем Ерохин

Lead DS

 **@Gofat**



Как определить самодостаточность?

Самодостаточность – способность человека или группы к самостоятельному удовлетворению своих собственных потребностей.

Для городского планирования определение самодостаточности можно почерпнуть в работе [1]:

Самодостаточный город может быть определен, как определенный периметр, внутри которого популяция самодостаточна, то есть:

- экономика внутри города полностью обеспечивает занятость населения
- инфраструктура поддерживает достаточный уровень обеспечения потребностей населения в области услуг, культуры и безопасности

Концепция 20-минутного города

В наше время, популярной градостроительной концепцией является идея т.н. «**20-минутного города**».

Эта концепция широко применяется в мировой практике городского планирования, например в:

- Портленде (США)
- Мельбурне (Австралия)
- Париже (Франция)

*В рамках концепции предполагается, что все (или большинство) повседневных нужд жителей города могут быть удовлетворены в радиусе **15-20 минут пешей ходьбы от дома**.*

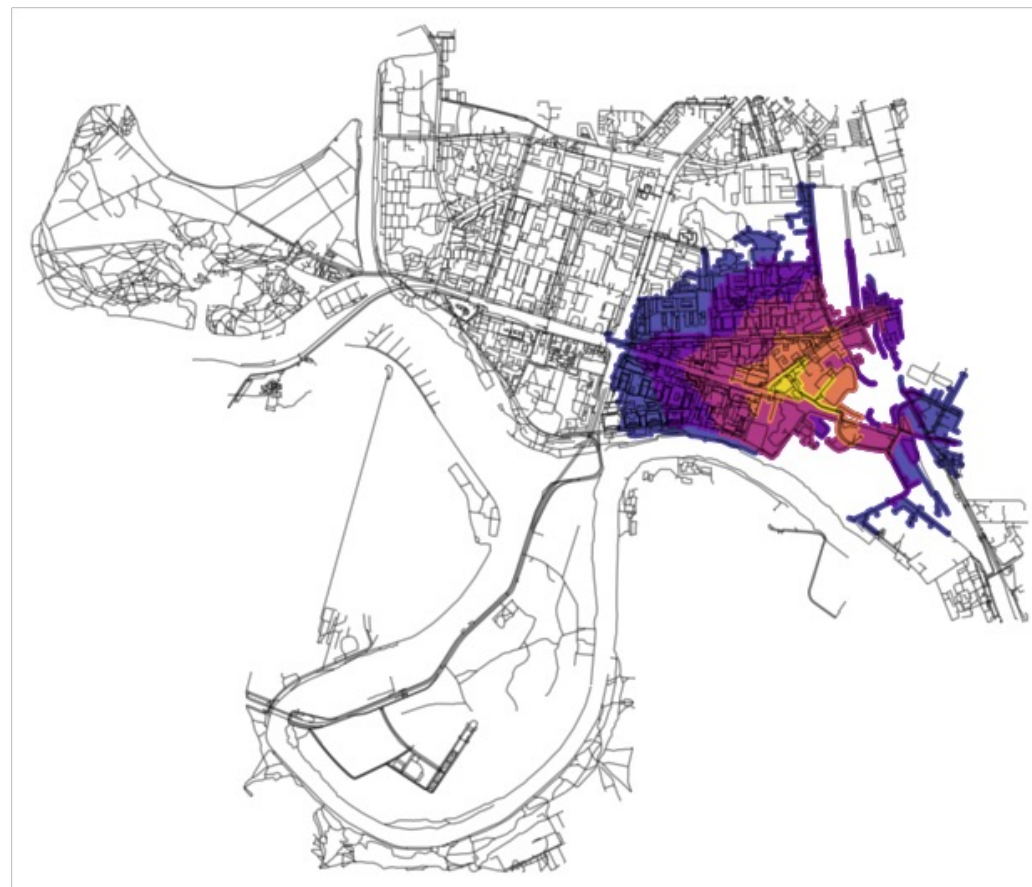


Но как найти зону доступности?



Для анализа зон пешей (или транспортной) доступности часто используется метод построения *изохрон*.

Изохрона или зона транспортной доступности по времени — это линия/полигональный объект, очерчивающий район, в который может добраться житель, двигаясь из определенной точки, в течение заданного отрезка времени, с учетом его скорости перемещения.



Выбор источников данных



В рамках анализа сконцентрируемся на следующих областях:

1. **Экономика** (наличие рабочих мест)
2. **Медицина** (доступность поликлиник для взрослых и детей)
3. **Транспорт** (доступность остановок НГПТ, станций метро)
4. **Образование** (наличие детских садов и школ)
5. **Быт** (наличие супермаркетов, аптек, банковских отделений)
6. **Отдых** (доступность парков и зон отдыха, кинотеатров, библиотек, ресторанов)

Использовать будем следующие наборы данных:

1. **Исходный источник данных** о количестве жителей и рабочих мест в зоне 500*500 м;
2. **Данные data.mos.ru** по объектам социальной инфраструктуры;
3. **Данные Open Street Maps (OSM)** для получения графа пешеходных дорог, точек притяжения (points of interest), парков и рекреационных зон, дополнительные данные об учебных заведениях
4. **Данные портала API HeadHunter** для получения информации о станциях метро, МЦК, МЦД



Работа с данными

- Для каждой зоны 500×500 м, получим ее центр и найдем ближайшую вершину пешеходного графа дорог;
- От вершины графа отложим *20-минутную изохрону* (полигон, в рамках которого в любую точку можно дойти менее чем за 20 минут);
- Для каждой зоны внутри района получим *пересечение между изохроной и важными точками (полигонами)* для каждого направления (экономика, медицина и т.д.);
- Будем считать наличие в пешей доступности объектов какого-либо направления за количество ее жителей, а отсутствие за 0, после чего, просуммируем по всем зонам. Тогда можно получить коэффициент *число жителей с пешей доступностью направления / все жители района**;
- Для оценки достаточности числа рабочих мест, используем коэффициент *суммарное число работающих человек / суммарное число жителей* во всех зонах района (на основании данных исходного набора данных)**;
- Полученные показатели приведем к значениям в промежутке $[0, 1]$. Посчитаем *среднее и разброс* ряда показателей.

* В случае, когда в одном направлении сочетаются несколько вариантов объектов (напр. детские и взрослые поликлиники), будем усреднять коэффициенты каждого варианта для получения общего коэффициента направления

** Формально, тут не учтена пешая доступность, но при ее учете возможен множественный подсчет одних и тех же рабочих мест. Поэтому более корректным на начальном этапе будет использовать такое приближение



Интегральный критерий оценки

Каждая из сфер городской жизни может оцениваться и восприниматься жителями и специалистами по-разному. Поэтому, возможно расширить методологию, используя опрос жителей и специалистов.

Например, работе [2] используется метод взвешивания Гилфорда, который приводит показатели к общим числовым оценкам. Для первого приближения будем использовать *взвешенное среднее по каждому из показателей* (где веса получены опросом граждан и специалистов).

В данной работе, *показателям выдан равный вес*, т.к. оценка одного человека не является репрезентативной жителям города и экспертному сообществу.

Итоговым интегральным показателем будет *средняя доступность благ* из каждой области (медицина, транспорт и т.д.) для выбранного района.

Пример: Хорошево-Мневники



Можно заметить, что большинство областей на среднем или высоком уровне. С некоторым снижением в области Экономики.

Такие результаты можно объяснить тем, что район имеет достаточно большую площадь – поэтому доступ к некоторым благам может быть затруднен.

При более детальном анализе зон доступности (пример см. в приложении 1), можно заметить, что проблемы с доступностью возникают на границах района – по улице Берзарина и около ст. м. Терехово.



Пример: Митино

По сравнению с предыдущим районом, заметно, что практически все показатели находятся на высоком уровне, за исключением Экономики.

Это может объясняться тем, что Митино – спальный район за пределами МКАД с активным ростом числа новостроек и ростом количества жителей. Но с куда более скромным ростом числа локальных рабочих мест.

Более детальный анализ зон доступности (см. приложение 1) показывает следующие результаты – ядро района имеет отличную пешую доступность благ. Чего не скажешь об окраинах района.



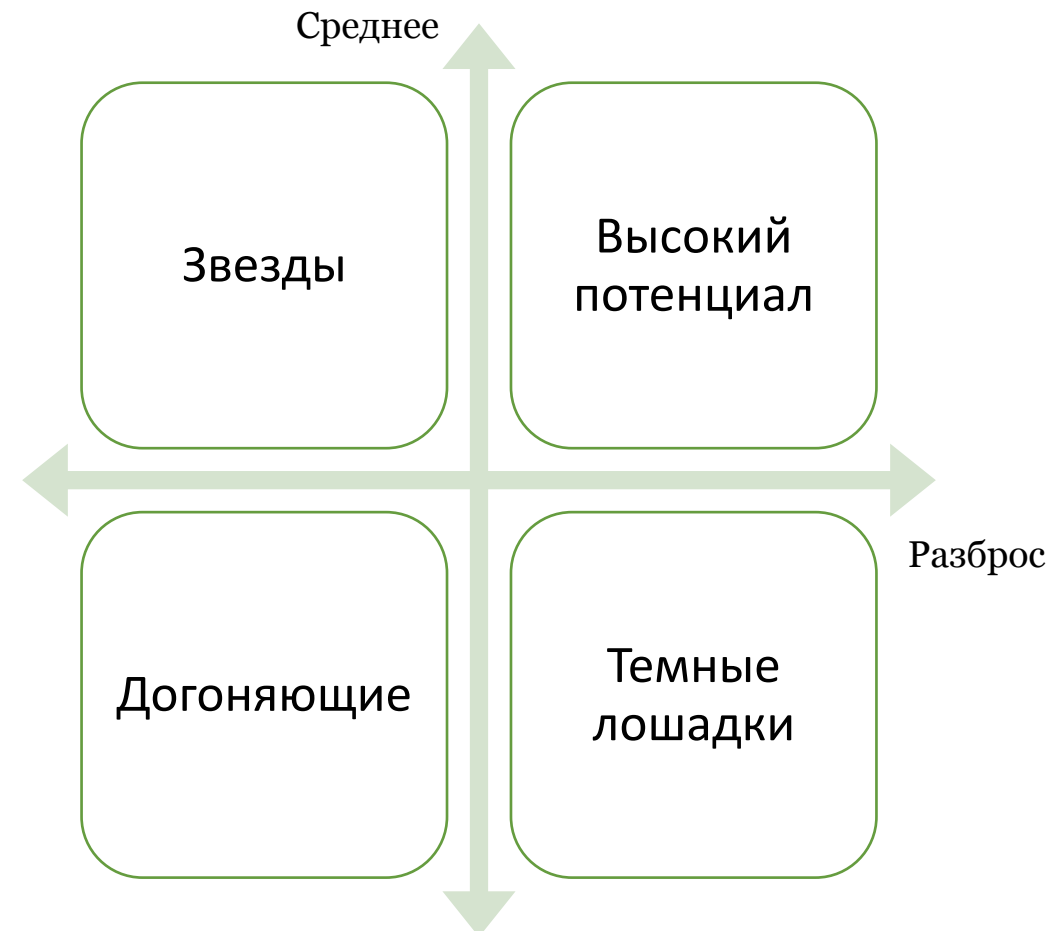
Оценка районов внутри округа



Для ранжирования районов внутри округа будем использовать аналог **матрицы BCG** [3].

Разделим районы на матрицу 2*2 по показателям:

- Средняя доля жителей с доступом к сфере
- Разброс доли жителей с доступом к сфере

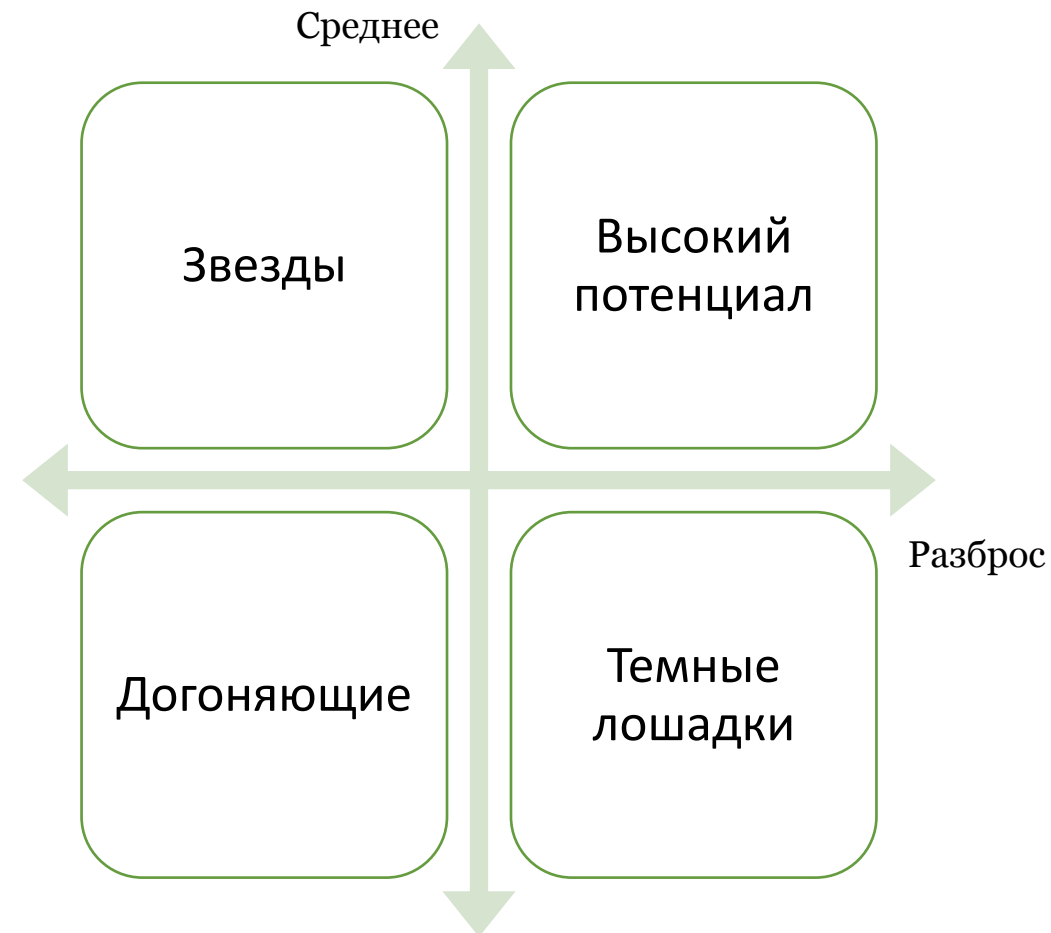


Оценка районов внутри округа



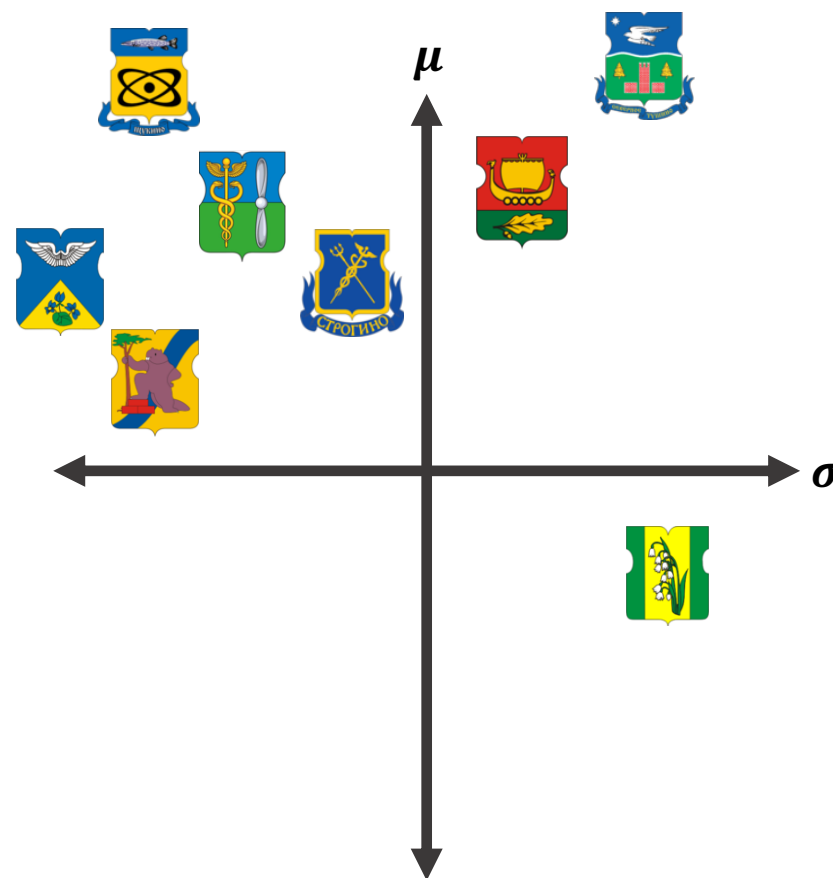
Каждый район относится к своей категории:

1. **Звезды** – районы с наилучшей самодостаточностью;
2. **Высокий потенциал** – районы с высоким уровнем многих показателей, но с небольшими «просадками» по некоторым из них;
3. **Темные лошадки** – районы с невысокими показателями, имеющие некоторые положительные стороны;
4. **Догоняющие** – районы, имеющие стабильно низкие показатели по всем сферам



Пример: СЗАО

	Эк-ка	Мед-на	Транс-т	Обр-е	Быт	Отдых	μ	σ
Северное Тушино	0.36	0.91	0.93	1	1	0.82	0.84	0.24
Щукино	0.53	0.66	0.93	1	0.97	0.91	0.83	0.19
Митино	0.34	0.76	0.93	0.97	0.98	0.82	0.8	0.24
Южное Тушино	0.5	0.6	0.86	1	1	0.78	0.79	0.21
Покровское -Стрешнево	0.74	0.64	0.77	0.96	0.95	0.59	0.78	0.15
Строгино	0.4	0.77	0.87	0.97	0.95	0.69	0.78	0.21
Хорошево-Мневники	0.48	0.62	0.63	0.92	0.91	0.65	0.7	0.18
Куркино	0.4	0.47	0.46	0.92	0.93	0.46	0.61	0.25



Использование методики в процессе принятия управленческих решений



На примере СЗАО г. Москвы:

- Большинство районов можно отнести к «Звездам». Есть некоторые недочеты у районов **Митино** и **Северное Тушино** (в основном, связанные с низким числом локальных рабочих мест в районах, а также меньшей доступностью медицины в Митино). И достаточно серьезные недочеты у **Куркино**. Дополнительно можно заметить, что у района **Хорошево-Мневники** есть риск перехода в худшую категорию;
- Отсюда можно сделать следующие выводы развития районов – *увеличение количества локальных рабочих мест* (например, за счет организации Технопарков, коворкингов, льгот на размещение офисов малого и среднего бизнеса внутри районов);
- В случае **Хорошево-Мневники** – контроль существующих рисков за счет размещения дополнительных поликлиник на территории района, дополнительного открытия кинозалов и библиотек;
- В случае **Куркино** – повышение транспортной доступности района (находится на окраине Москвы) за счет увеличения числа маршрутов/остановок НГПТ и возможного внесения в планы развития метрополитена. А так же контроль рисков за счет введения в эксплуатацию дополнительных поликлиник на территории района, открытия библиотек.

Пути дальнейшего развития методологии

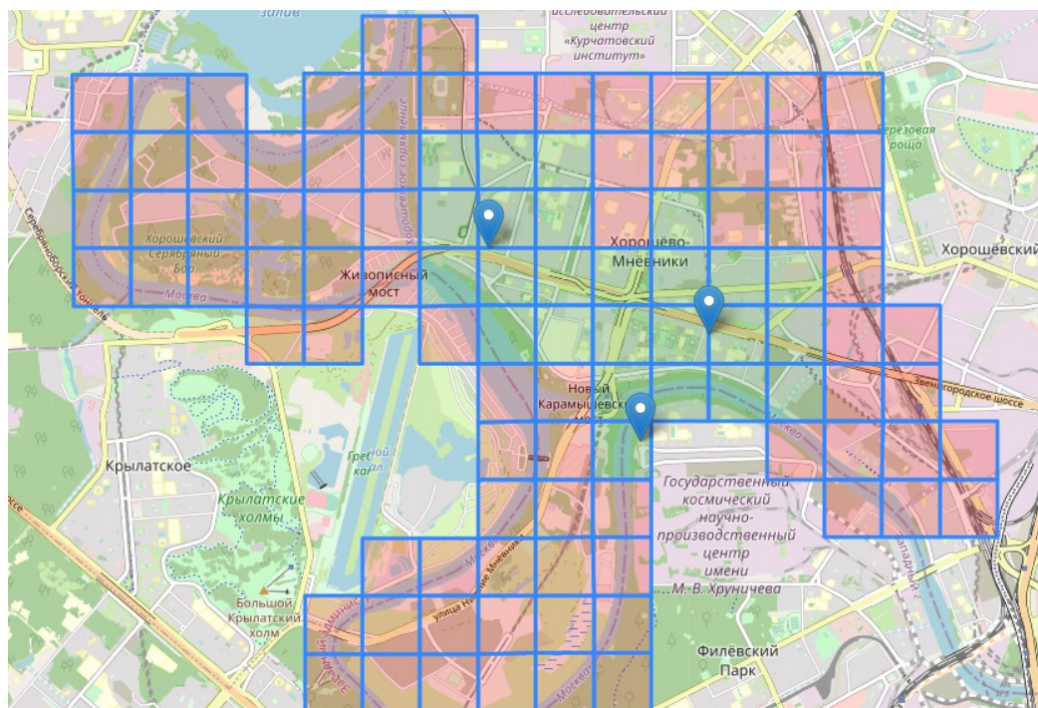


В порядке приоритета:

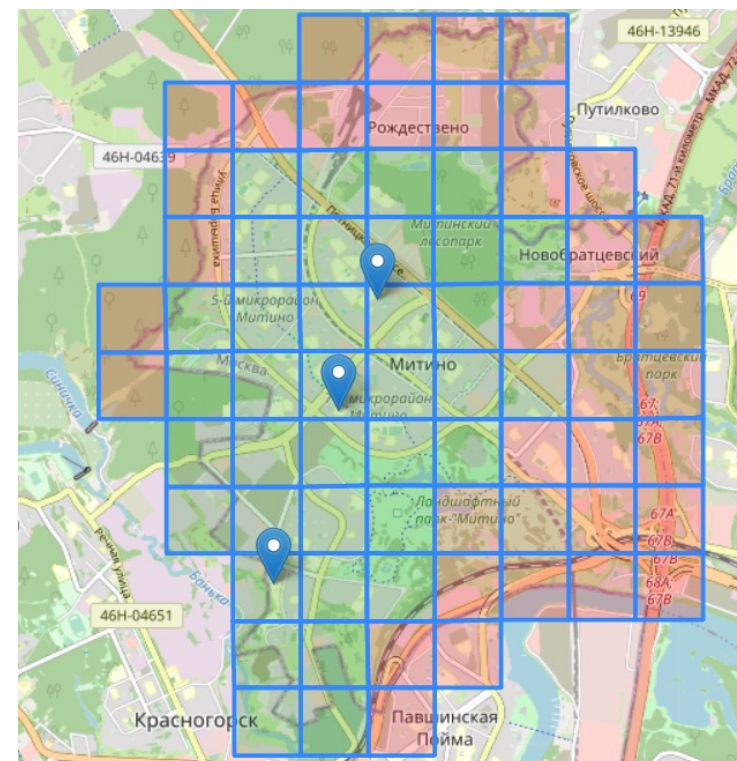
1. Уточнение *наиболее важных областей интересов жителей* для оценки самодостаточности и способов реализации потребностей в этих областях;
2. *Внедрение процессов оценки качества данных (data quality)*, оценка существующих наборов данных и исправление проблем качества данных для этих наборов;
3. Обновление методологии *с учетом связи блага и района для пограничных областей* (напр. жителям может быть не важно, куда идти в кино, но важна территориальная принадлежность школы/детского сада);
4. Выявление «важности» каждой области для восприятия уровня качества жизни – *получение весовых коэффициентов для каждой области интересов жителей*;
5. Внедрение в методологию *оценки загруженности инфраструктуры* (даже при территориальной близости к месту жительства, школа/больница могут быть перегружены, что снижает качество получаемых услуг);
6. Обновление методологии – *внедрение оценок дополнительных количественных показателей* (напр. уровня загрязнения воздуха, уровня преступности, цены недвижимости и т.п.);
7. Внедрение процесса *периодической переоценки результатов и отслеживания динамики показателей по районам*. Адекватной периодичностью кажется переоценка раз в квартал.

Приложение 1

Пример визуализации зон доступности взрослых поликлиник в районе Хорошево-Мневники *



Пример визуализации зон доступности взрослых поликлиник в районе Митино *



* Можно заметить, что зоны частично задевают другие районы города, что несколько снижает качество получаемых выводов и результатов. Правильный учет географической принадлежности жителей может быть одним из направлений развития методологии.