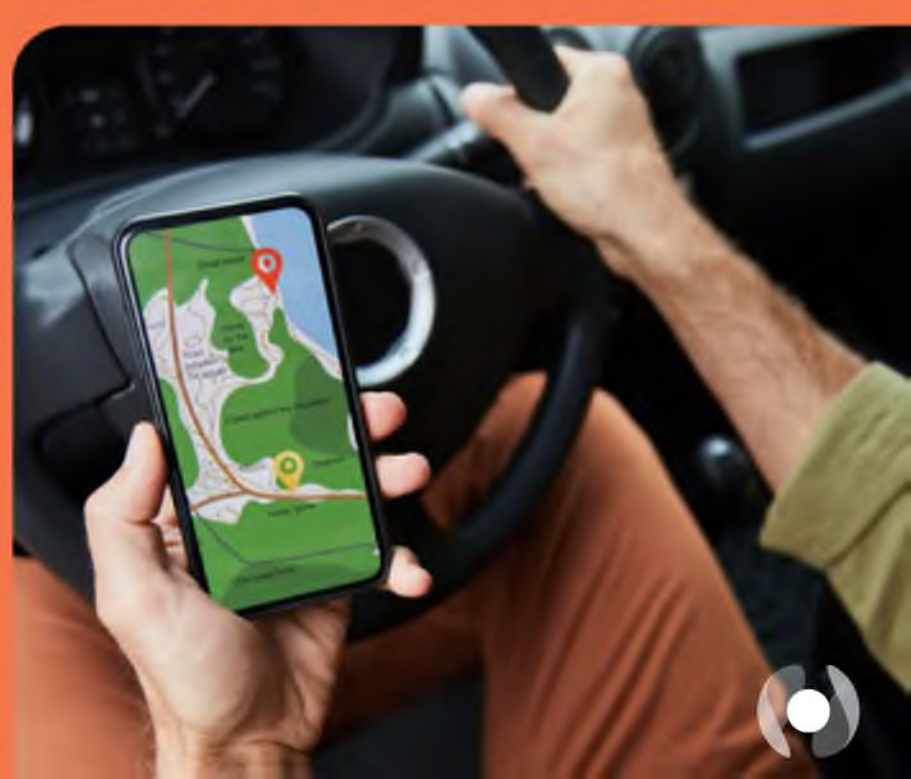
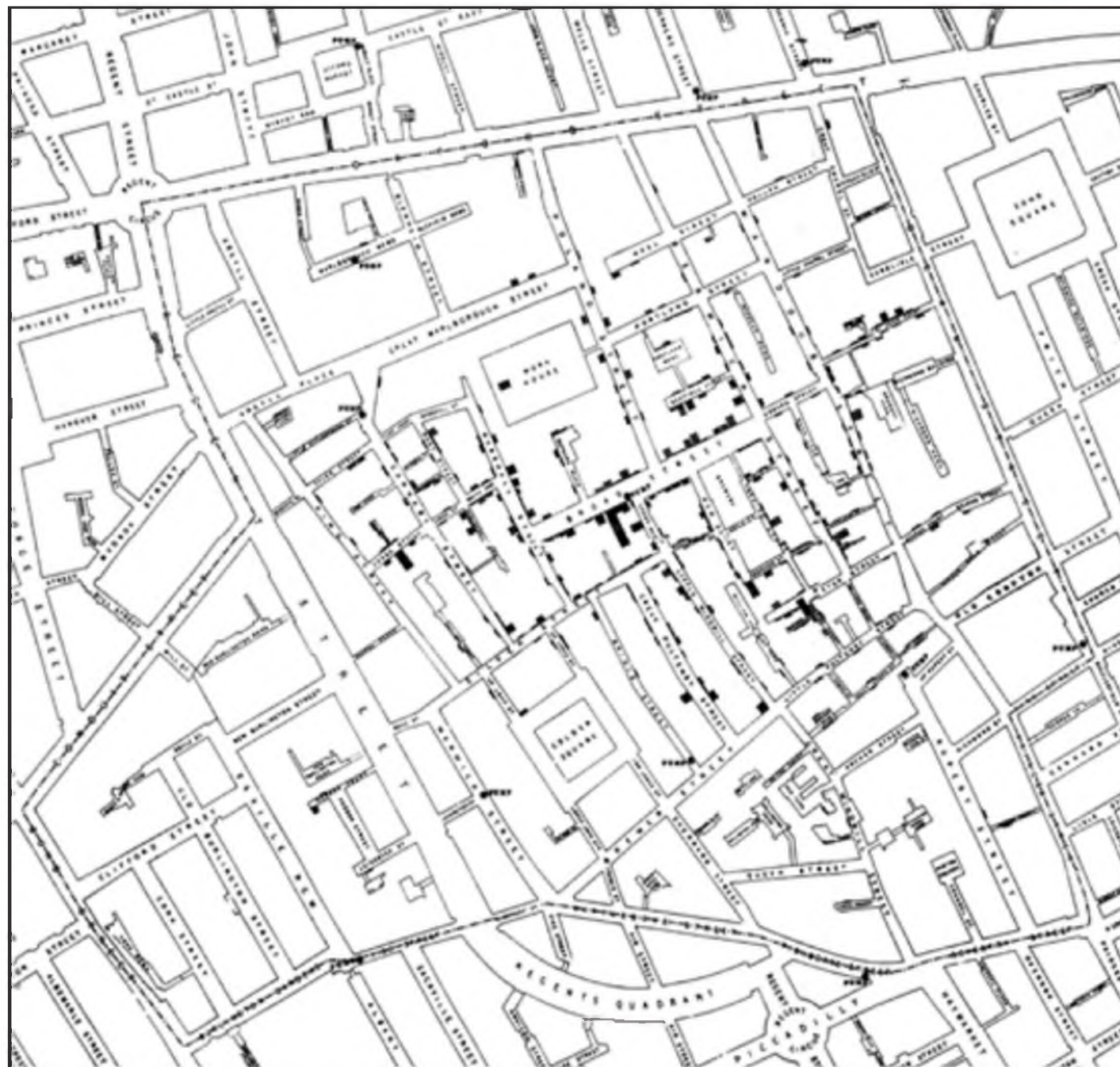


# Облачные решения в пространственном анализе больших данных

Ольга Ведерникова, ООО «Эпсилон Метрикс»



# 1854, Лондон



# Анализ пространственных данных

Скорее всего, у вас уже есть пространственные данные

Но как можно сделать их полезными в масштабе?

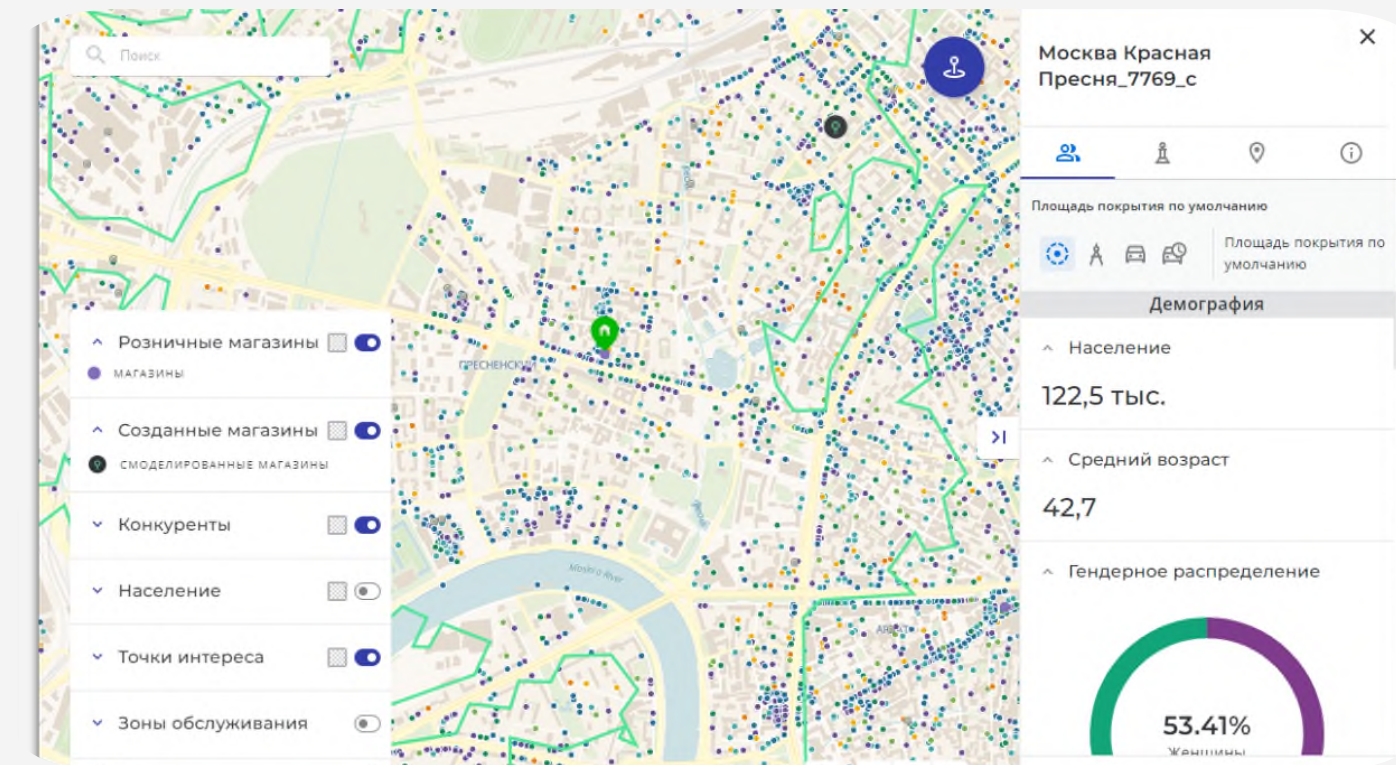
И как использовать для решения отраслевых задач, например, →

55.745361, 37.580293  
Смоленская улица,  
дом 5,  
Москва,  
119121  
Золотое Кольцо

```
SELECT * FROM `bigquery-public-data.new_york_taxi_trips.tlc_yellow_trips_20
```

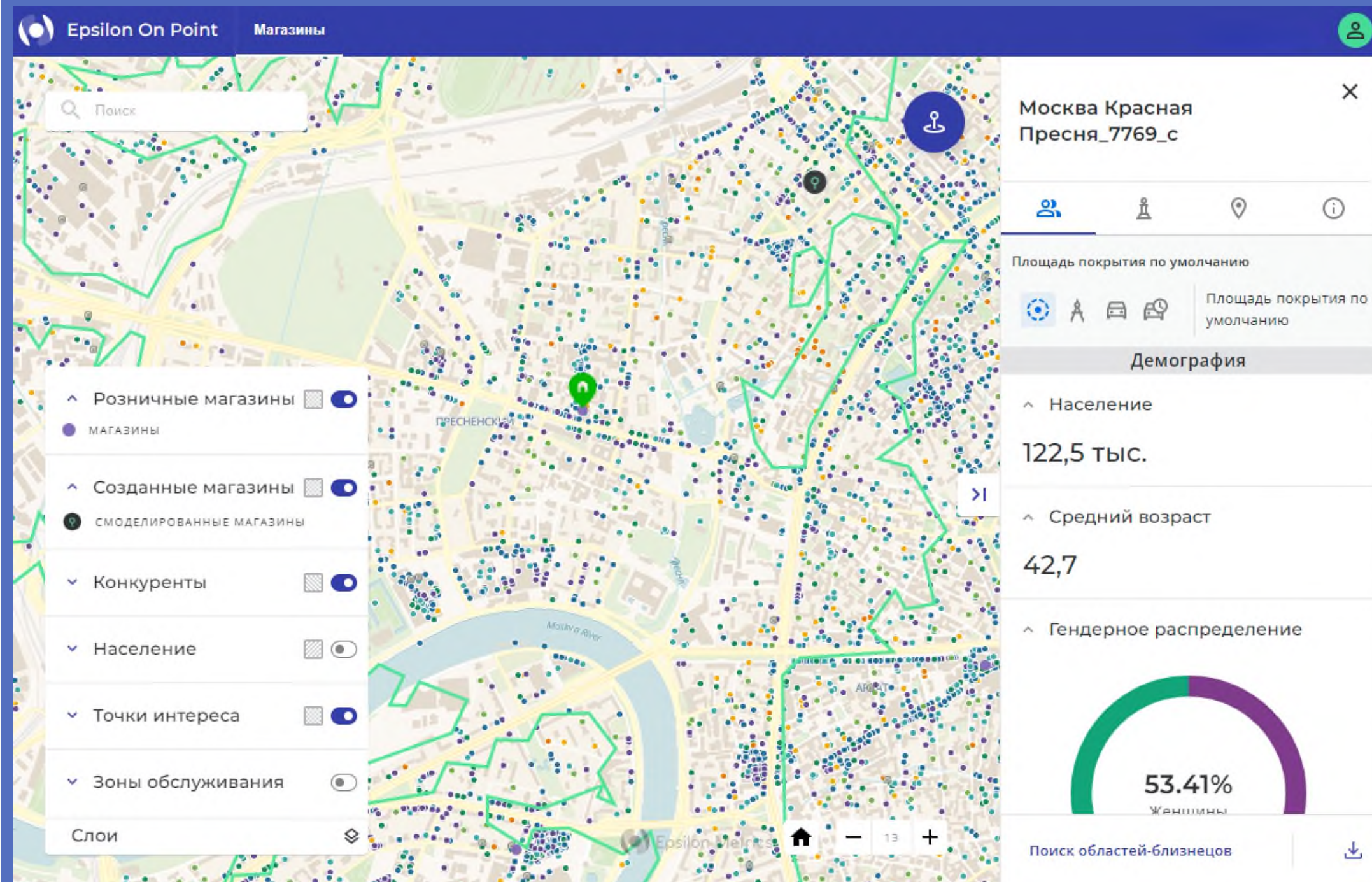
pickup_datetime	dropoff_datetime	passenger_count
2015-10-19T14:42:59	2015-10-19T14:47:06	
2015-11-24T14:28:25	2015-11-24T12:37:33	

Rows per page: 100 1 - 100 of 140205128



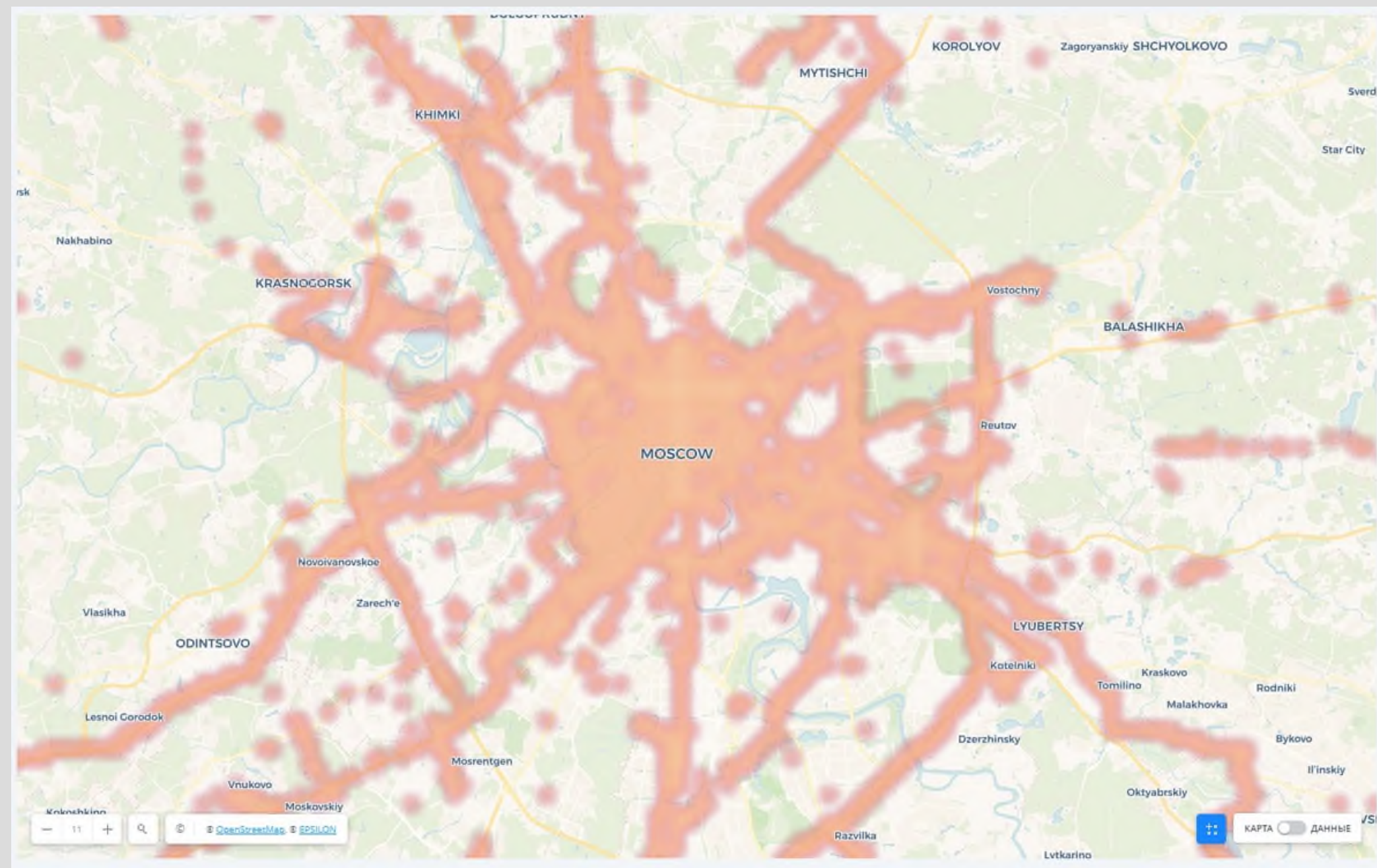
# Поиск лучших мест для новых магазинов и прогноз выручки

Использовать данные о продажах для оценки локаций и выбора оптимальных



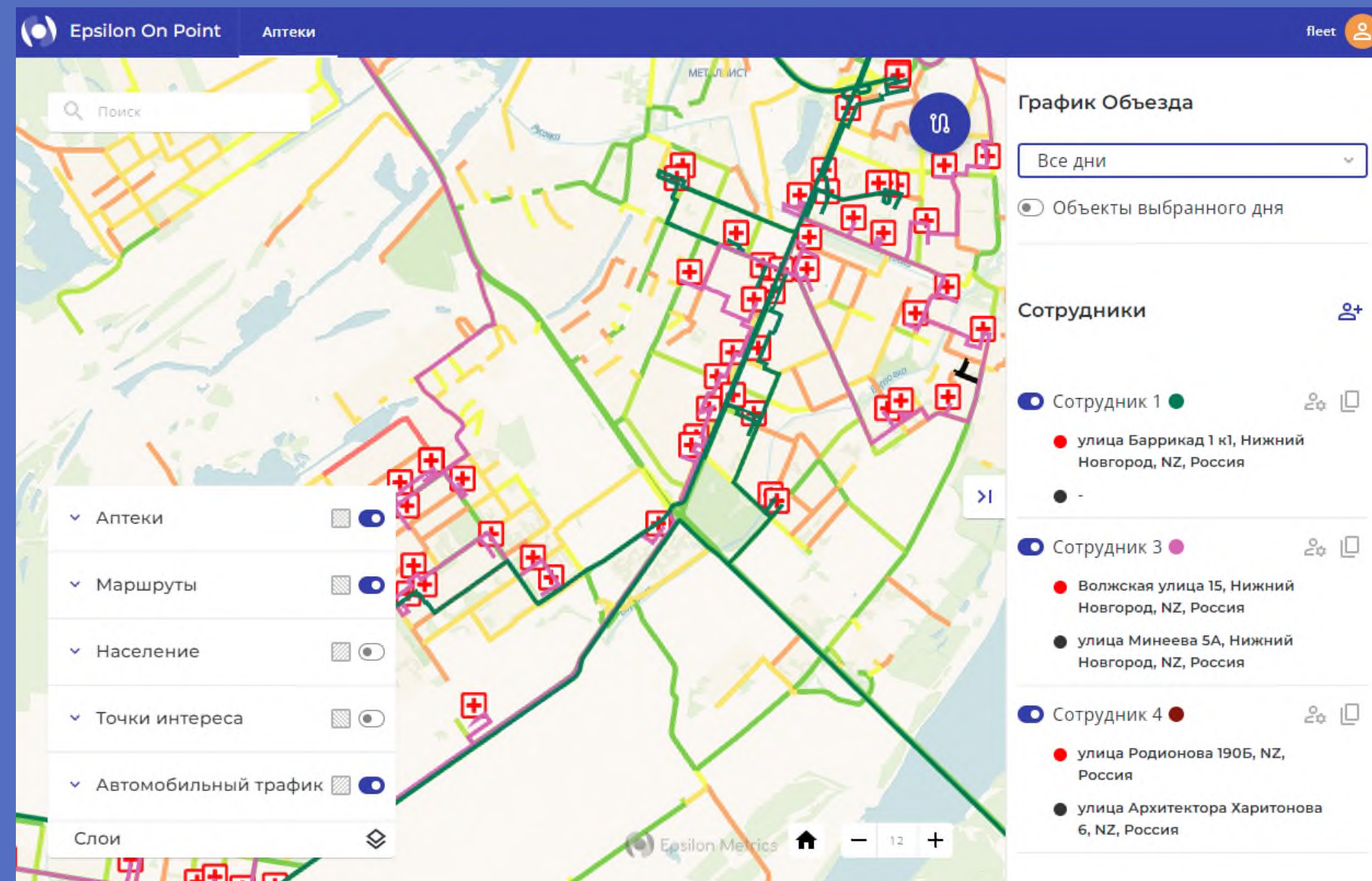
# Оптимизация покрытия сотовой связи

Использовать данные о текущем  
покрытии и башнях сотовой  
связи для оптимизации покрытия



# Оптимизация маршрутов

Использовать данные о заказах, продажах, трафике, пробках, графике курьеров, стоимости поездок для приоритизации точек для посещения и построения планов поездок и оптимальных маршрутов



# О чём мы поговорим





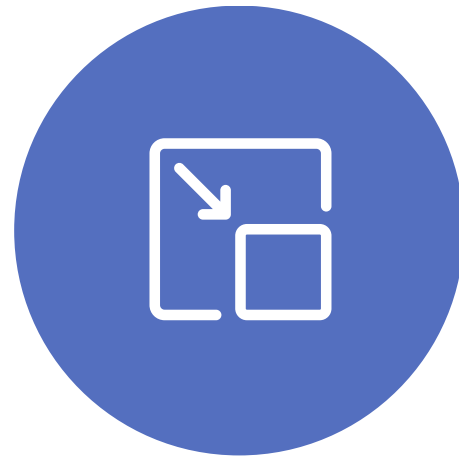
1 ГДЕ



# Облако данных

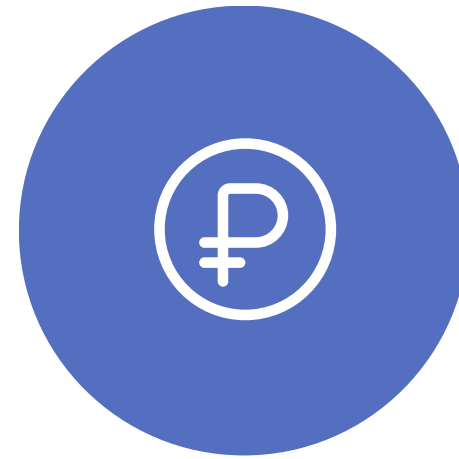


# Доступ к облачным данным



## Масштабируемость

С помощью возможностей облачных вычислений



## Экономическая эффективность

Благодаря разделению ресурсов для вычислений и для хранения



## Доступность

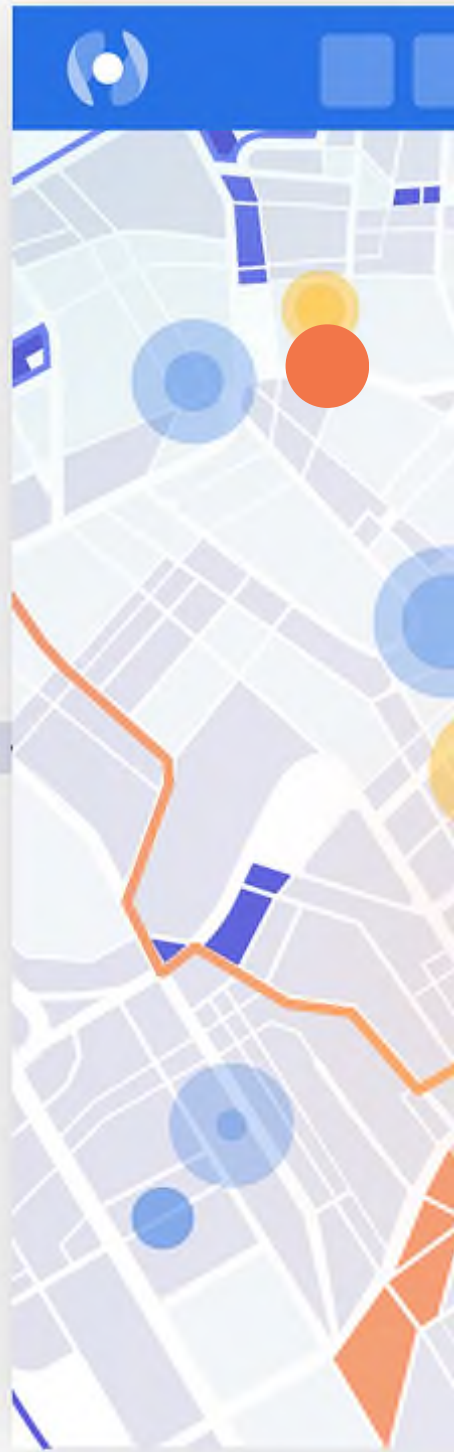
За счёт использования SQL

Удобный способ обмена данными и их использования





Пространственное  
расширение



Мы  
разработали  
облачную  
платформу



# Разделение вычислений и хранения

Подключаетесь к облачному хранилищу и больше не перемещаете данные между разными файлами, базами и хранилищами. Сохраняете один раз, и затем используете в разных сервисах и комбинируете с другими данными.

**ВЫЧИСЛИТЕЛЬНЫЕ  
МОЩНОСТИ**



```
SELECT data FROM myTable JOIN providerTable
```

**ОБЛАЧНЫЕ ХРАНИЛИЩЕ**

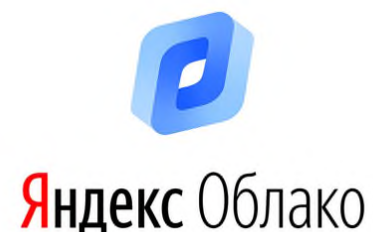
PROVIDER 1

PROVIDER 2

PROVIDER 3

...

data.parquet  
ENTERPRISE DATA



# GeoParquet: на пути к геопространственной совместимости между облаками данных

Простой способ добавления **гео-составляющей** к Parquet для возможности работы **облака геоданных**



Инициатива сообщества и Open Geospatial Consortium



GitHub - [opengeospatial/geoparquet](https://github.com/opengeospatial/geoparquet)

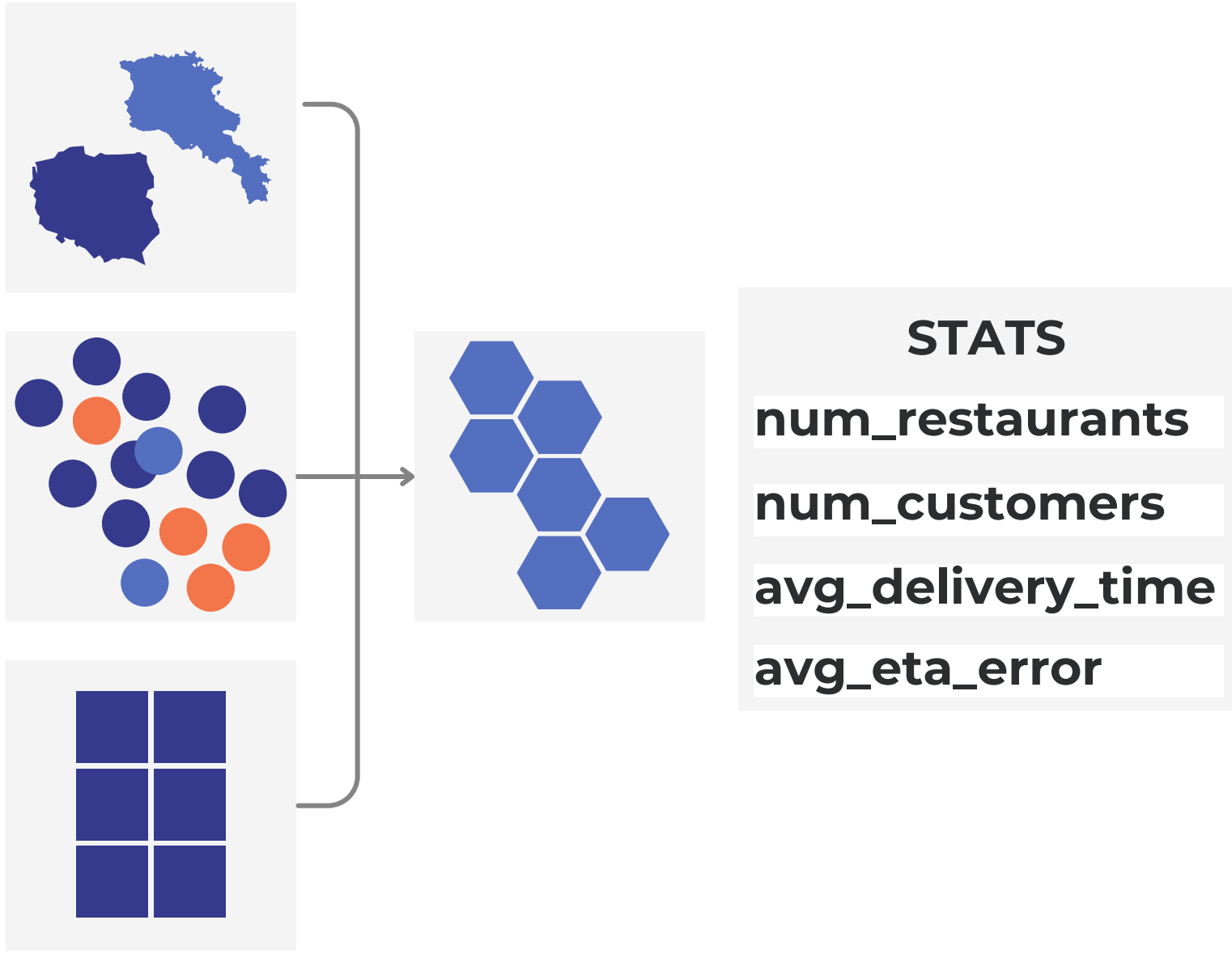
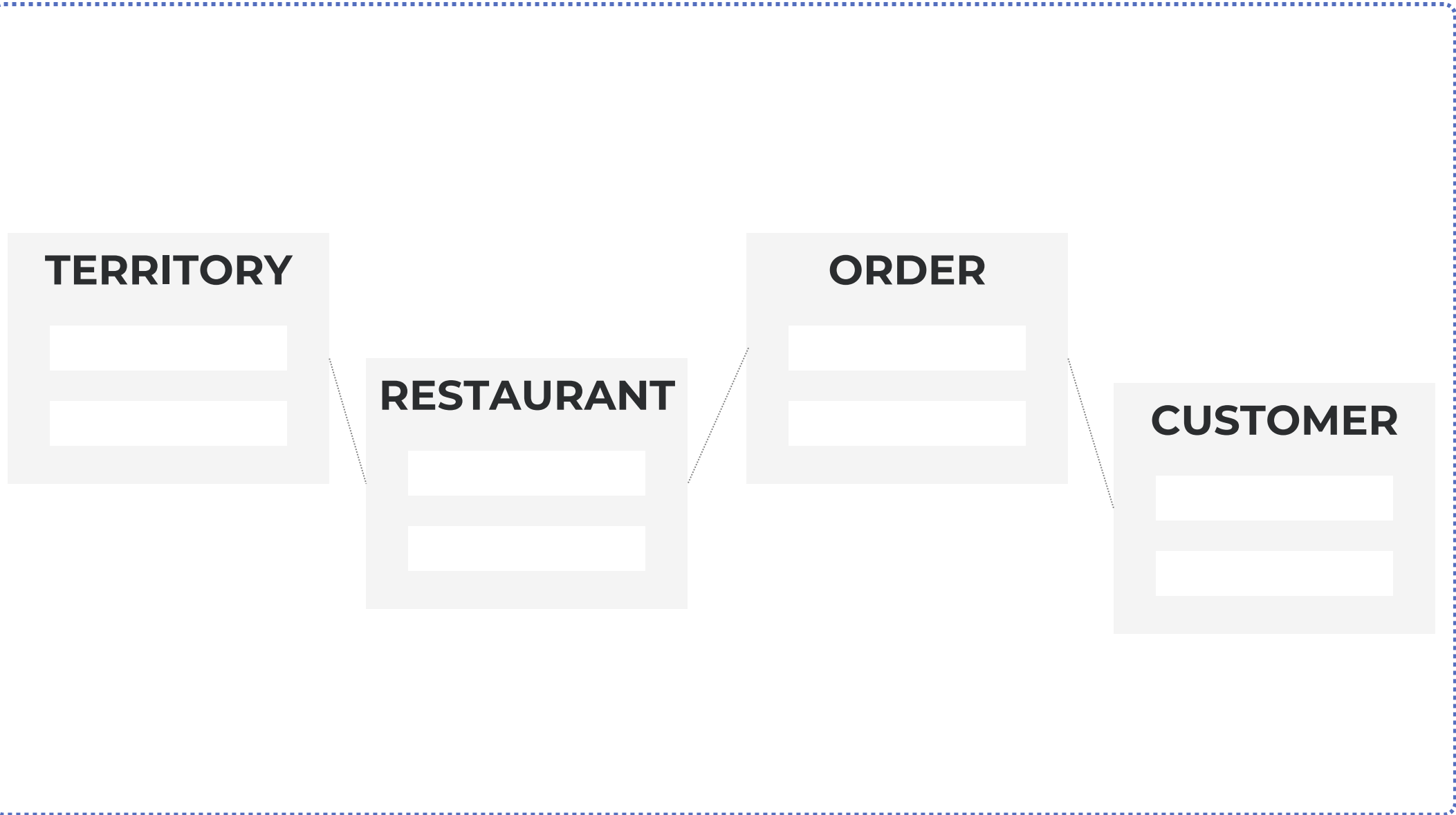
2

КАК



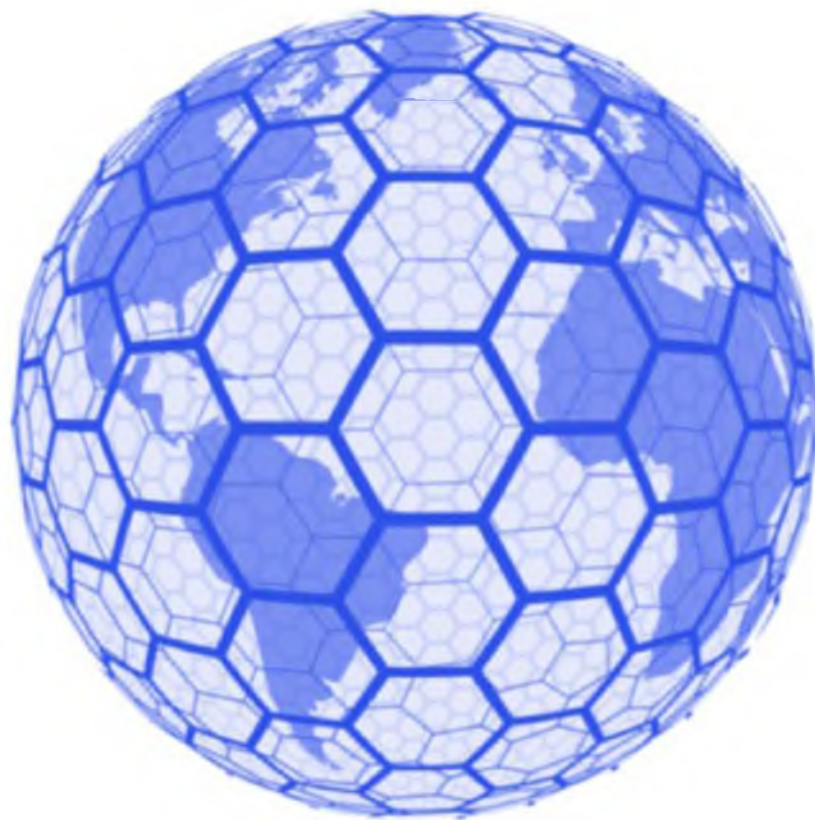
# Пространственные индексы

# Как вы моделируете данные?



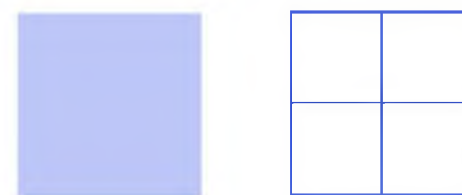
# Пространственные иерархические индексы

Различные стратегии разделения пространства на дискретные сетки

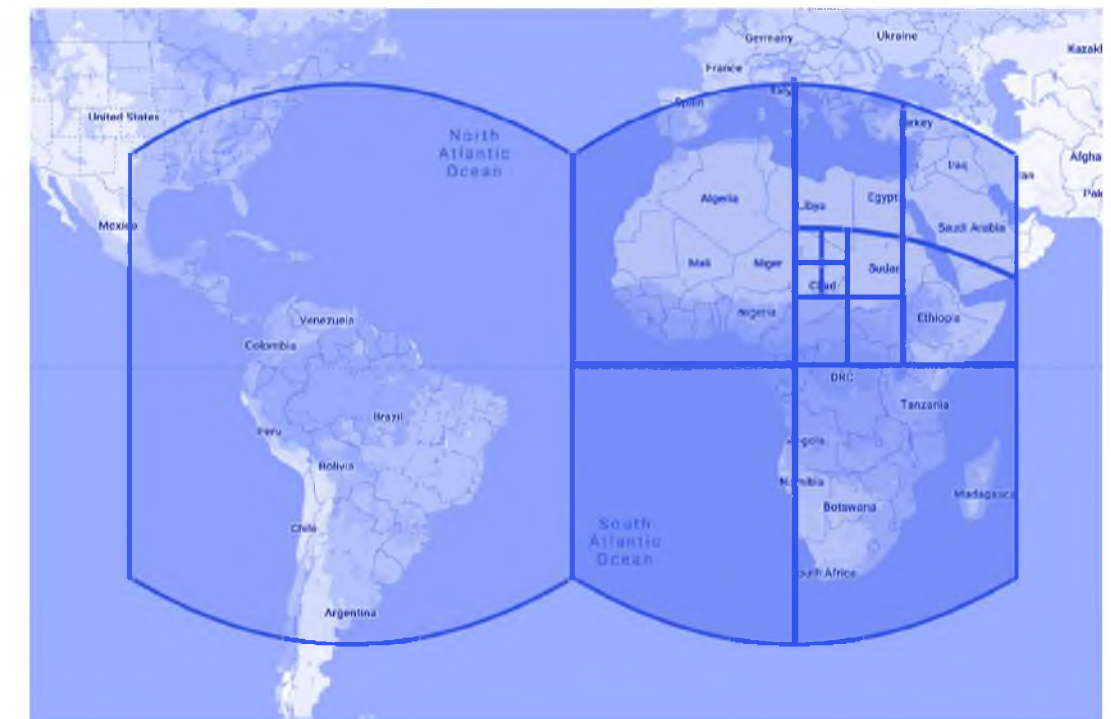


**H3**

(0,0)	(1,0)	(2,0)	(3,0)	(4,0)	(5,0)	(6,0)	(7,0)
(0,1)	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)	(7,1)
(0,2)	(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)	(7,2)
(0,3)	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)	(7,3)
(0,4)	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)	(7,4)
(0,5)	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)	(7,5)



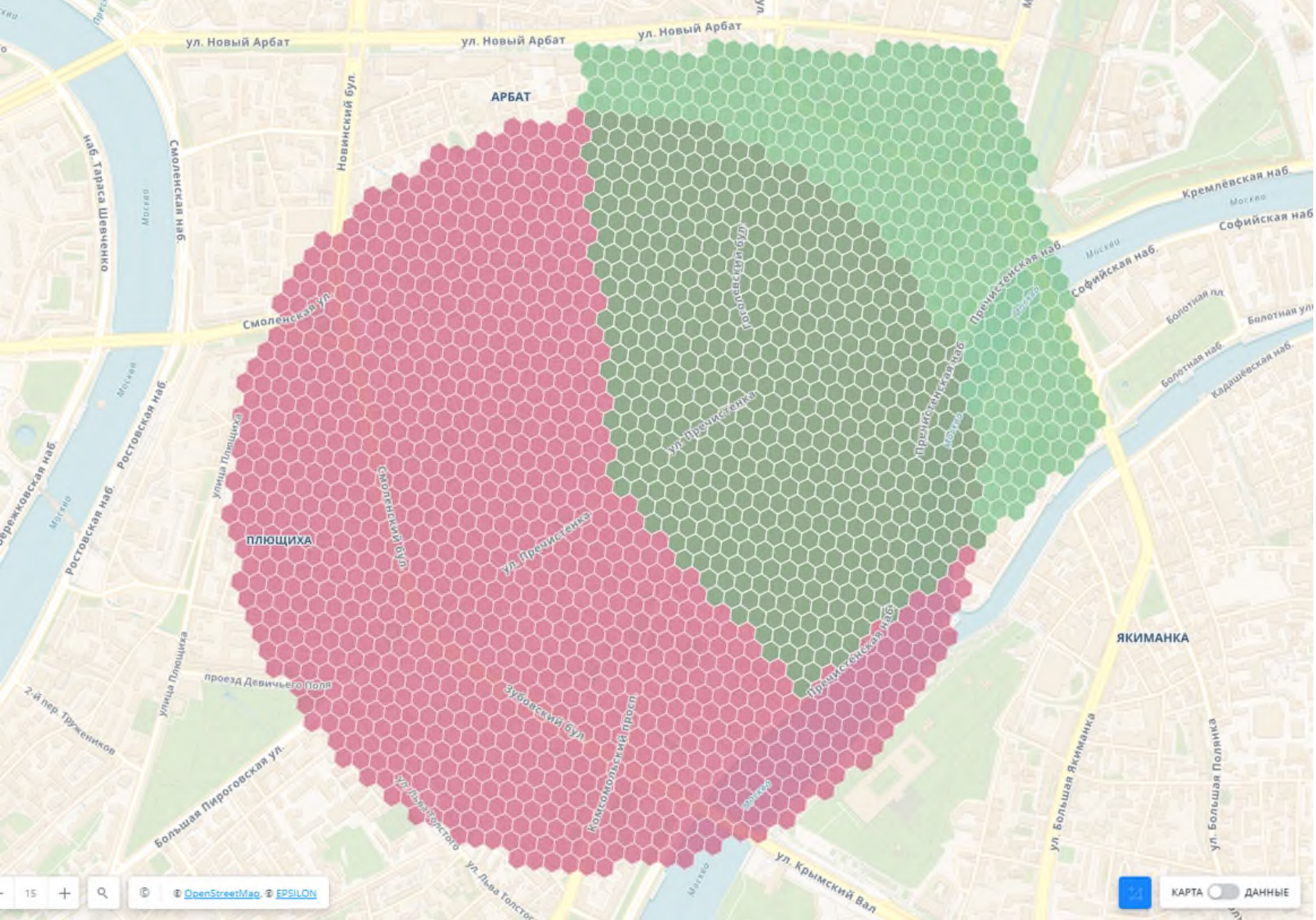
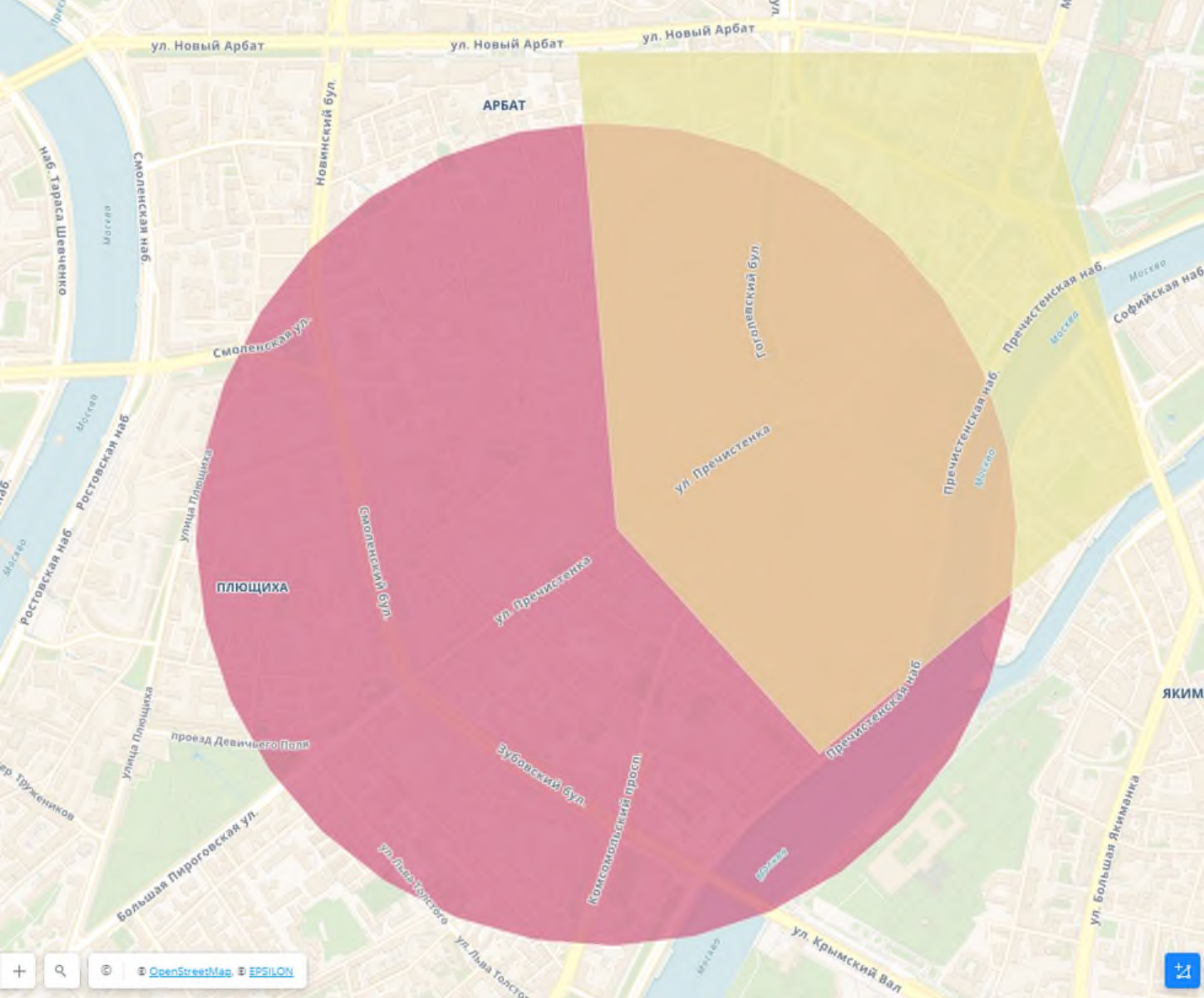
**S2**



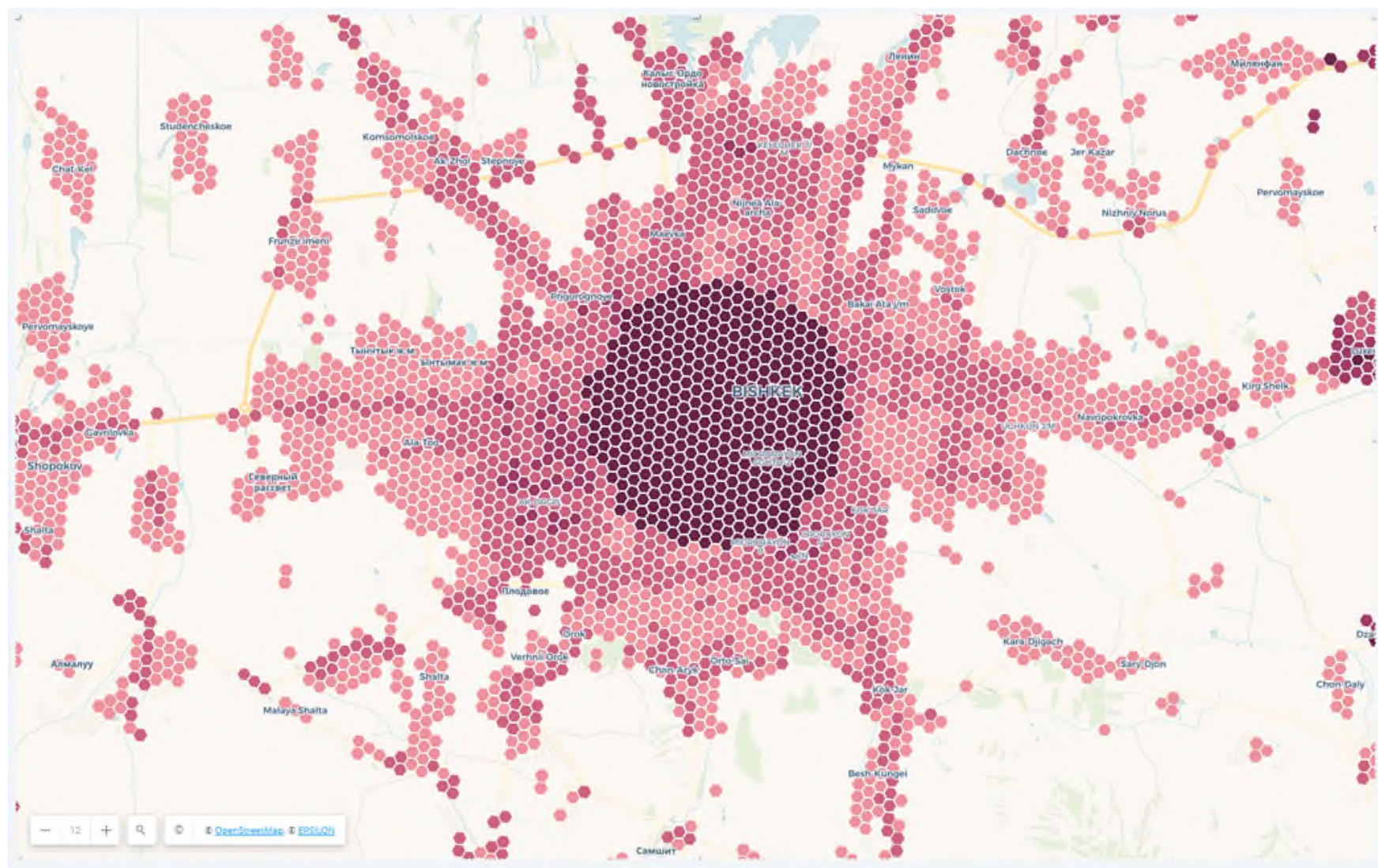
**QuadBin**



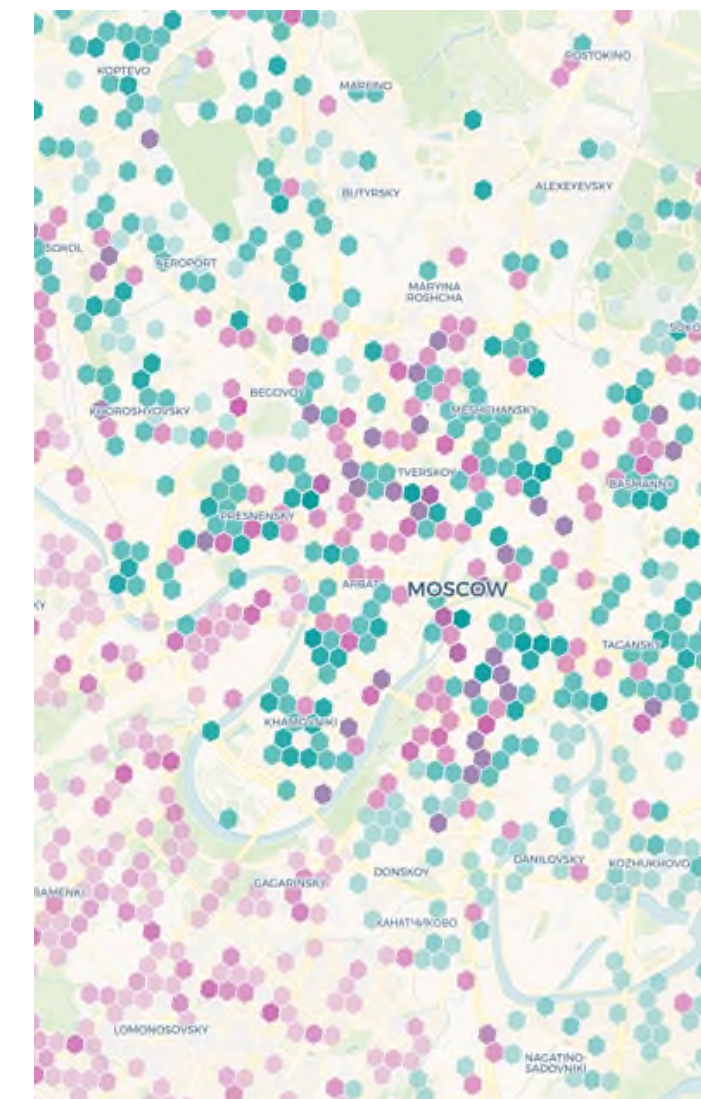
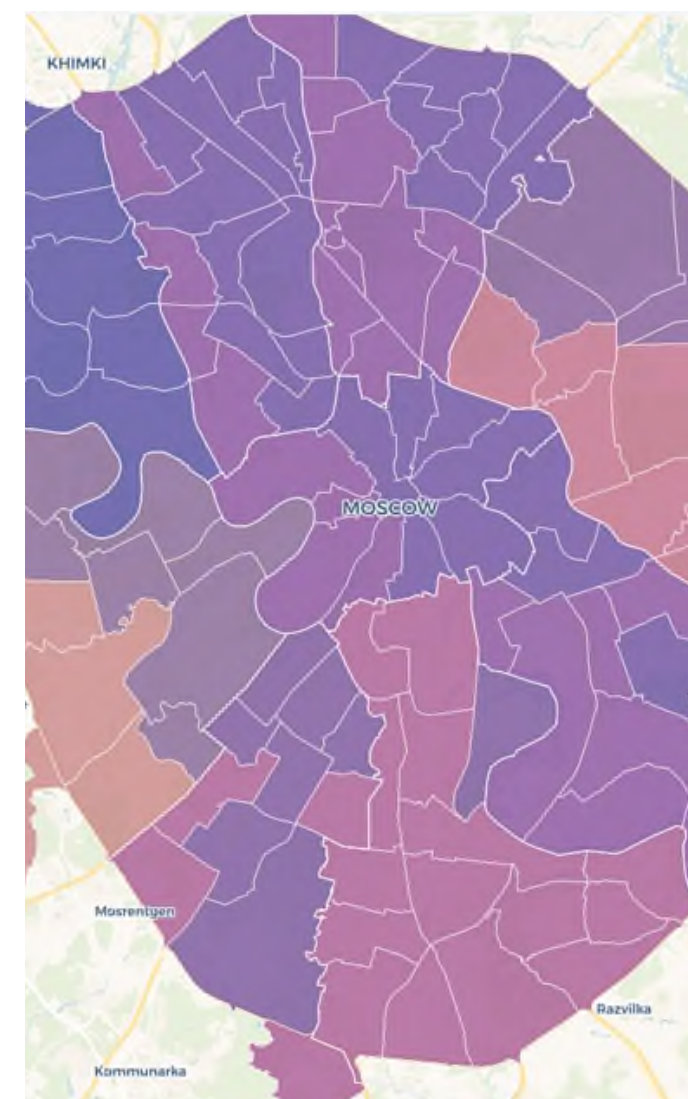
# Вместо геометрий объектов - индексы



# Что ещё дают пространственные индексы



**Объединение данных  
из разных источников**






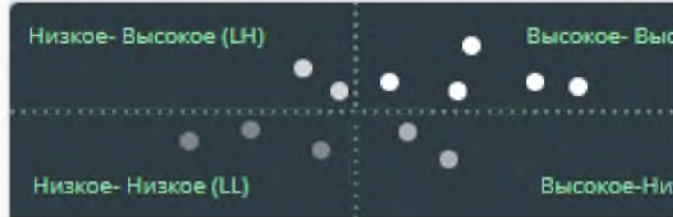




**Быстрый анализ и  
визуализация**

## Добавить новый анализ

Выбрать анализ, который вы хотите добавить

ВСЕ [СОЗДАТЬ И ОЧИСТИТЬ](#) [АНАЛИЗИРОВАТЬ И ПРОГНОЗИРОВАТЬ](#) [ПРЕОБРАЗОВАТЬ](#)

 <p><input type="radio"/> <b>Создание многоугольников из точек</b> Создать многоугольники из точек, используя выпуклые оболочки, вогнутые оболочки, ограничивающие круги или ограничивающие рамки. <a href="#">ПОДРОБНЕЕ</a></p>	 <p><input type="radio"/> <b>Фильтр по значению столбца</b> Сохранить или удалить строки с выбранным значением столбца. <a href="#">ПОДРОБНЕЕ</a></p>	 <p><input type="radio"/> <b>Создать линии из точек</b> Создать линии из точек, используя одну точку, значения столбца или второй слой. <a href="#">ПОДРОБНЕЕ</a></p>
 <p><input type="radio"/> <b>Процент строк подвыборки</b> Подвыборка строк в наборе данных на основе указанного процента. <a href="#">ПОДРОБНЕЕ</a></p>	 <p><input checked="" type="radio"/> <b>Вычислить кластеры точек</b> Дополнить столбец cluster_no, чтобы пространственно разделить точки на определенное количество групп. <a href="#">ПОДРОБНЕЕ</a></p>	 <p><input type="radio"/> <b>Обнаружение выпадающих значений и кластеров</b> Использовать Морана I для выбора геометрии, в которой значения пространственно сгруппированы, а выпадающие значения находятся рядом. <a href="#">ПОДРОБНЕЕ</a></p>
		

# Полный набор геоаналитических инструментов и сервисов, работающих в хранилище данных



# Каталог наборов данных



## Финансы

Данные о транзакциях продавцов и банкоматов от ведущих банков и компаний, выпускающих кредитные карты.



## Мобильность

Мобильные устройства и данные GPS дают представление о моделях движения человека.



## Демография

Самые последние данные переписи, включая: возраст, доход, типы домохозяйств и многое другое.



## Жилье

Статистика, цены и история недвижимости для принятия решений и инвестиций.



## Дорожное движение

Данные из приложений маршрутизации и GPS для анализа моделей трафика и пассажиров.



## Точки интереса

Данные о местонахождении коммерческих предприятий, ресторанов, школ, достопримечательностей и

3

КТО



# Low-code и No-code в пространственном анализе



# Конструктор workflow: демократизация пространственного SQL

The screenshot displays the Epsilon Workflows interface. At the top, the title bar shows "Epsilon Workflows" and the user "admin4example". The main workspace is divided into three vertical panels:

- Left Panel (Components):** Contains a search bar and several categories of components: "Агрегация (5)" with options like "Группировка по", "Итого", "Подсчет", "Разгруппировка"; "Статистика точек в полигонах"; and "Импорт/Экспорт (1)" with "Получить таблицу по имени".
- Center Panel (Workflow):** A visual flowchart with three main stages:
  - Загрузка данных магази В:** A data source box for "moscow\_stores" with 6,530 records.
  - Загрузка данных о покрытии:** A data source box for "towers\_st" with 20,609 records.
  - Разбиение данных на НЗ:** A blue-shaded area containing a sequence of operations: "Выбор" (Selection), "Буфер" (Buffer), "Заполнение НЗ" (Fill), "Перекрестное соединение" (Join), and "Группировка по" (Group by).
  - Расчет выруч:** A final stage containing a "Вызов процедуры" (Call procedure) component.
- Right Panel (Details):** Provides metadata for the "moscow\_stores" source, including the source name, user, and database. It also shows the table type, number of rows (6,530), and options to "Показать таблицу" or "Показать карту".

At the bottom of the interface, there are tabs for "Результаты", "Сообщения", "Данные", "Карта", and "SQL". The "Карта" tab is active, showing a map of Moscow with red dots representing the data points. A legend in the bottom right of the map indicates "Предварительный просмотр на основе" with two options: "the\_geom" (selected) and "the\_geom\_webmercator".

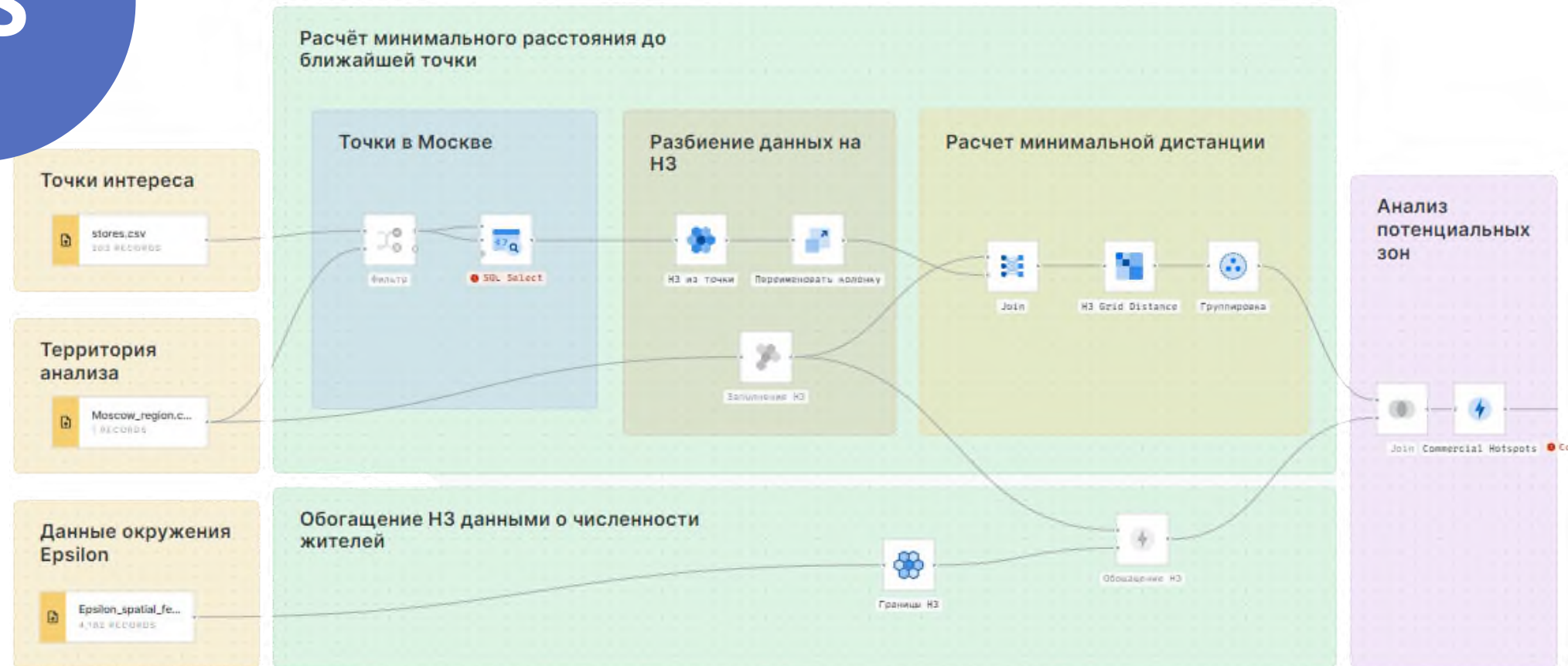
# Low-code в действии

```

BEGIN
/*===== native.buffer (v1) [8385c367-75e1-4dd0-9e67-fa393f2c577c] =====*/
BEGIN
DROP TABLE IF EXISTS "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".ttt.workflow_df925e633c45d708_368c94057b53218b_result;
CREATE TABLE "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".ttt.workflow_df925e633c45d708_368c94057b53218b_result AS
SELECT
ST_TRANSFORM(
ST_BUFFER(
ST_TRANSFORM(
the_geom,
4326
)::geography,
1000
)::geometry,
ST_SRID(the_geom)
) AS the_geom_buffer,
*
FROM "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".admin4example.moscow_stores;
DROP INDEX IF EXISTS idx_b6025a7155fe1e90;
CREATE INDEX idx_b6025a7155fe1e90
ON "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".ttt.workflow_df925e633c45d708_368c94057b53218b_result
USING GIST (the_geom_buffer);
END;

/*===== native.h3polyfill (v1) [87c5e892-4e6f-4435-ba22-7251406289d8] =====*/
BEGIN
IF EXISTS(SELECT 1 FROM pg_extension WHERE extname = 'h3' /*plv8*/) THEN
DROP TABLE IF EXISTS "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".ttt.workflow_df925e633c45d708_e3332d02;
CREATE TABLE "epsilon_db_user_67be497f-8954-4894-9bbb-f352458d15ef_db".ttt.workflow_df925e633c45d708_e3332d02b071afad
WITH __h3 AS
(
/*SELECT
epsilon-un.epsilon.H3_POLYFILL(
the_geom_buffer,
9,
'центр'
) h3s
*/

```





# Что мы обсудили



# Начните прямо сейчас

Посетите наш сайт

<https://epsilonmetrics.ru/request-live-demo/>



ПОГОВОРИТЕ С НАМИ

## Запросить демонстрацию

Запланируйте 20-минутную встречу с нашими экспертами, чтобы понять, как вы можете использовать пространственный анализ в своей организации.

Ваше имя*	Ваше Фамилия*
Корпоративная почта*	
Организация*	
Ваша роль в организации?*	
Численность персонала*	

Да, я хотел бы получать маркетинговые сообщения о продуктах, услугах и мероприятиях Epsilon Metrics. Я могу отписаться в любой момент.

Регистрируясь, я подтверждаю, что мои личные данные будут обрабатываться в соответствии с [Соглашением о конфиденциальности](#).

Сохранить



# Спасибо!



[epsilonmetrics.ru](https://epsilonmetrics.ru)



[ceo@epsilonmetrics.ru](mailto:ceo@epsilonmetrics.ru)



@OLGAVVEDERNIKOVA

**Sk**

**СКОЛКОВО**