

БОЛЬШАЯ МАТЕМАТИЧЕСКАЯ МАСТЕРСКАЯ 2022

г. Тюмень

проект

«Применение автоматизированной системы сбора и обработки данных в урбанистике и образовании»

Команда проекта:

*Анна Авдюшина, Анна Алексеева, Алина Аралова, Елизавета Егорова,
Кристина Шелемеха*

Современные подходы к образованию состоят из компетентностной модели подготовки, но внедрение такого метода требует кардинальных изменений всех компонентов. Из-за переходного периода образуется квалификационный разрыв между обще-профессиональными компетенциями программного инженера, полученными в ходе обучения, и квалификационными требованиями по решению нестандартных задач и разработке программных средств, недостаточный уровень практико-ориентированности образовательных программ. Это в совокупности приводит к длительной адаптации выпускников на рабочем месте.

Практико-ориентированный поход — ключевой элемент в реализации «технологической» магистратуры для опережающей подготовки кадров и обеспечения потребностей организаций и предприятий в оперативном и точном реагировании на внутренние и внешние изменения. К нам обратился Курганский государственный университет для актуализации современной методологической основы преподавания дисциплины «Интеллектуальный анализ данных» в рамках реализации образовательного процесса по направлению подготовки «09.04.04 Программная инженерия (Методы и алгоритмы интеллектуальной обработки данных в информационных вычислительных системах)». И целью работы на Большой математической мастерской стало формирование метапредметных компетенций студентов с помощью комплекта интерактивных кейсов, представляющих собой платформу анализа данных, которая позволяет создавать модели машинного обучения на примере анализа городских данных.

Мы выделили несколько компетенций, которые можно будет приобрести с помощью такого подхода.

- Учащийся способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте.

- Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач.
- Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
- Способен применять на практике новые научные принципы и методы исследований.
- Способен самостоятельно приобретать с помощью информационных технологий и использовать в практической деятельности новые знания и умения, в том числе в новых областях знаний, непосредственно не связанных со сферой деятельности.

Первый кейс, который мы разработали для включения в методическое пособие — это модель геттоизации района. Выбор обусловлен тем, что существует большое количество социологических и урбанистических исследований относительно процесса появления городских гетто, но при этом нет формализации выделенных критериев. В то же время модель можно верифицировать за счет многолетней истории исследований районов-гетто.

Геттоизация

Это процесс пространственный концепт социального неравенства, пространственная дезорганизация и деградация определенных городских территорий.

Из статьи Чернышевой Любови Алексеевны «Российское гетто: воображаемая маргинальность новых жилых районов»

В академической литературе понятие «гетто» критикуют за то, что оно является скорее категорией моральной оценки, чем аналитически значимой категорией. Оно не отражает процессы городской сегрегации, скрывая многообразие процессов, в которых конструируются границы и артикулируется неравенство. Подобную критику можно адресовать и экспертам, которые обращаются к понятию «гетто» для анализа новых жилых районов в России. Они рассматривают его как факт, как универсальную объективную характеристику, схватывающую социально-демографические и морфологические характеристики городских территорий, не уделяя внимание детальному исследованию мест, к которым они применяют эту категорию.

Альтернативный подход — обратить внимание на процессы, разворачивающиеся вокруг новостроек, учесть в анализе этих процессов то, каким значением наполняется понятие гетто на практике (в конкретном кейсе) и какую роль в этих процессах играют дискурсивные практики экспертов, связанные с названием района. Иными словами, что бы ни стояло за понятием «гетто» в академическом языке социологов города или в публичном

дискурсе европейских стран, в современной России категория гетто формирует свою собственную реальность и устанавливает особый феномен.

В этой статье рассматривается как гетто Мурино (новый жилой район в Петербурге), поэтому мы выбрали этот район как отправную точку нашего исследования и как район, на котором будем верифицировать разрабатываемую модель.

Математическая модель геттоизации

Из научной литературы мы выделили следующие критерии, характеризующие гетто:

- массовая застройка
- плохие, ухудшающиеся жилищные условия
- высокий уровень преступности
- депопуляция
- социальная изоляция жителей
- недобровольная сегрегация
- исключение из сфер труда, предполагающих долгосрочный трудовые контракты
- высотные жилые здания до 30 этажей и пустоты между ними, не образующие качественных общественных пространств.
- низкая транспортная доступность жилья
- отсутствие необходимых социальных услуг
- отсутствие границ дворов (нет понимания, где заканчивается один двор и начинается второй)
- разочарование собственников(много квартир выставлены на продажу или в аренду)
- превращение газонов в парковки
- падение стоимости жилья
- пивные дешевые магазины
- место проживания мигрантов
- инфраструктурная ненасыщенность района
- аварийность в районе
- объем объявлений об аренде

Однако, чтобы привести все эти критерии в программный вид, нужно большое количество закрытых данных, что не подходит для студенческой работы. Поэтому мы выделили 3 критерия, на которых будем проверять гипотезу, что Мурино является гетто:

- плотность алкомаркетов на квадратный километр;
- объем пустот, неблагоустроенных мест в районе;
- объем и тип объявлений об аренде в районе.

Рабочий процесс

Какие программные средства мы были использованы:

- интерактивные среды разработки: Google Colaboratory, Jupyter Notebook;
- пакеты языка Python: osmnx, folium, pandas, seaborn, geojson, numpy;
- Overpass turbo — инструмент анализа и извлечения данных для OpenStreetMap;
- cURL — программа командной строки, позволяющая взаимодействовать с множеством различных серверов.

Так как построить классическую математическую модель в случае гетто не получится, мы решили провести сравнительно-сопоставительный анализ. Сравнимые районы — Мурино (гетто-район) и Озерки (территория, не описываемая жителями и исследователями как гетто, но являющаяся спальным районом с многоэтажной застройкой).

Сколько пивных магазинов на районе

Почему мы считаем преимущественно пивные магазины?

Алкоголь продается во многих продуктовых магазинах и наличие продуктовых — не признак маргинализации среды. При этом продажа алкоголя в формате магазина разливных напитков позволяет обходить ограничения по реализации алкоголя в ночное время. Поэтому мы выделили объекты с тегом #beer и #alcohol на Open street map и категорию «Пивной магазин» на Яндекс.Картах. Объединили данные, убрали дублирующиеся объекты.

Весь процесс состоял из следующих этапов.

Этап 1. Создание запроса с использованием Overpass API.

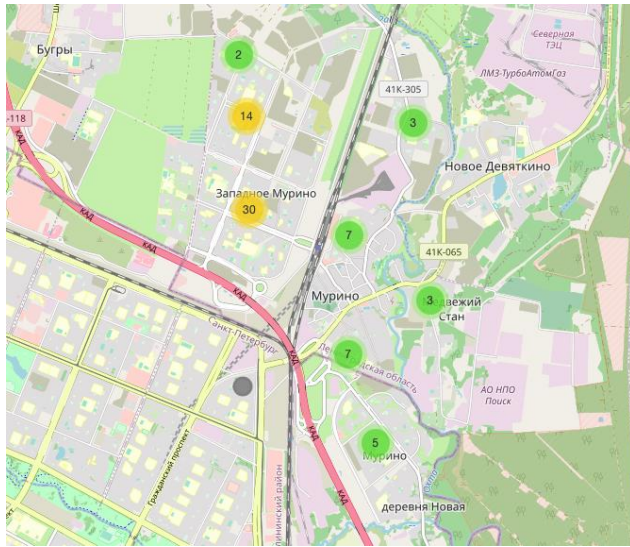
Этап 2. Сохранение данных в специализированном формате GeoJSON.

Этап 3. Определение данные, у которых есть тег названия алкогольного магазина.

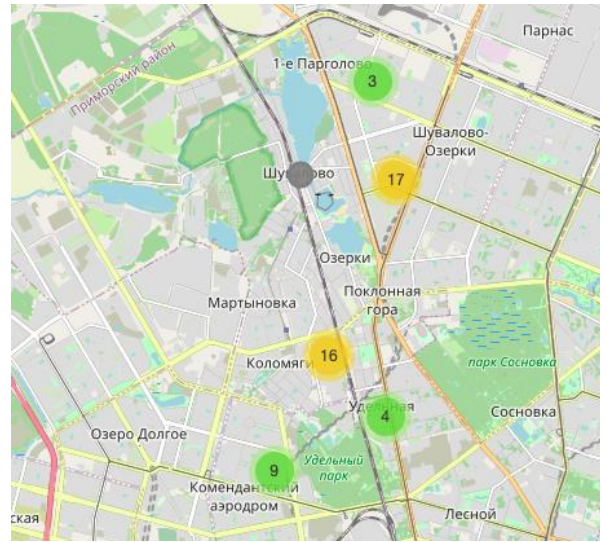
Этап 4. Сохранение данные в датафрейм.

Этап 5. Объединение данные с сервиса OpenStreerMap и Yandex Maps, исключение повторяющихся точек.

Этап 6. Использование библиотеки folium. Нанесение координат алкогольных магазинов на карту. Использование библиотеки для кластеризации геокоординат.



Мурино



Озерки

Кластера с пивными магазинами

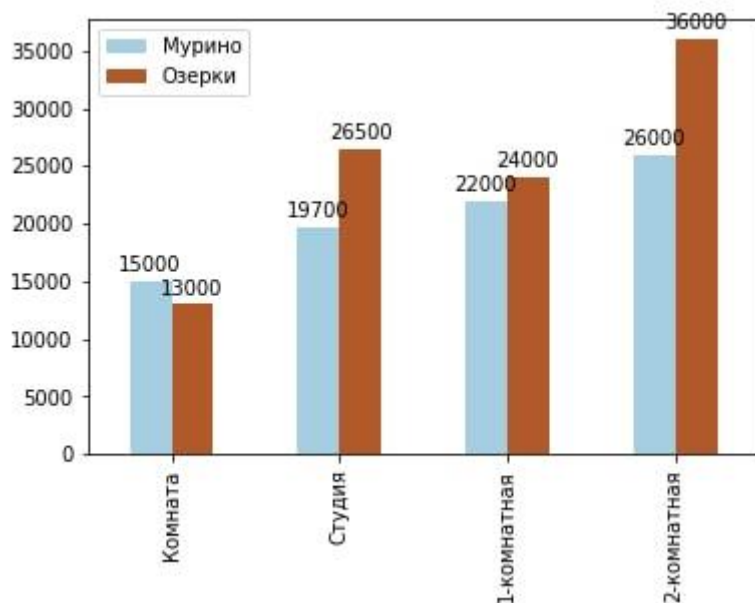
Плотность пивных магазинов:

- Мурино — 5,3892 пивнушки/кв.км
- Озерки — 2,008 пивнушки/кв.км

Гипотеза подтвердилась, в Мурино — районе как гетто — плотность магазинов разливных напитков в 2,68 раза выше чем в Озерках — типичном спальном районе.

Какое количество квартир сдается

Собрали данные о стоимости аренды и объеме объявлений в районах с Циан и Яндекс, сравнили цены и количество объявлений. В анализ взяли комнаты и квартиры, в которых 1, 2, 3 комнаты. Самая большая сложность заключается в получении общего количества квартир в муниципальных округах, так как эти данные являются кадастровыми и их предоставляет комитет имущественных отношений. Мы сравнили долю квартир, которые сдают в Выборгском районе Санкт-Петербурга и в Мурино, и выяснили, что в этих районах сдается схожий объем жилья: 0,3% и 0,4%, соответственно. Что не подтверждает гипотезу о том, что люди используют потенциальные гетто для любительских инвестиций или покидают их со временем. Район имеет признаки гетто в пространственном и социокультурном отношении, но депопуляции не наблюдается.



Распределение стоимости квартир

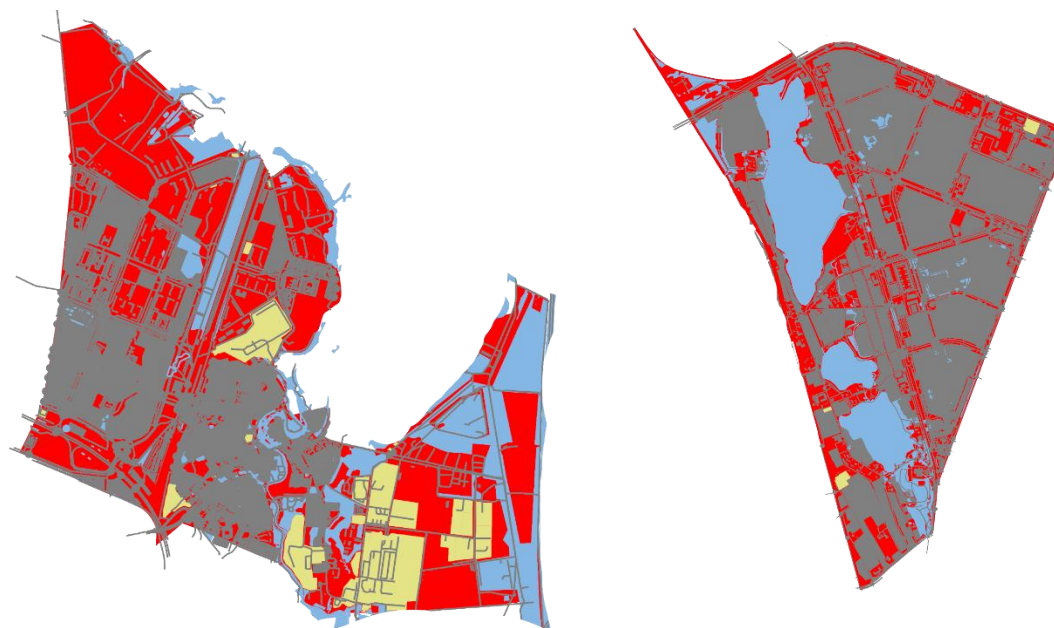
В Мурино:

- комнаты дороже на 15%
- студии дешевле на 35%
- 1-комнатные дешевле на 9%
- 2-комнатные дешевле на 38%

Улучшение: В дальнейшем развитии кейса планируется оценивать благосостояние население и включить эти данные в оценку количества квартир определенного типа.

Сколько пустот в районе

Один из признаков геттоизации — большой объем «ничьих» мест (это пустыри, заброшки и пр.). Границы «своего» сужаются до порога квартиры, люди не стремятся и не требуют благоустройства прилегающих к дому территорий.



Мурино

Озерки

Карта с цветовой разметкой

Весь процесс состоял из следующих этапов.

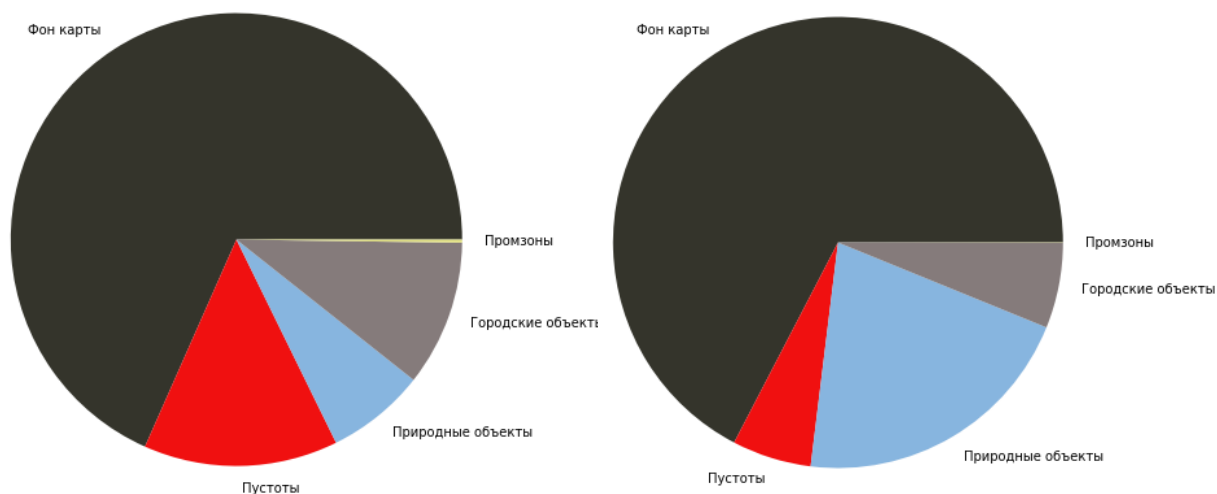
Этап 1. Создание запроса с использованием Overpass API для получения полигонов и объектов с тегами #parking, #building, #leisure.

Этап 2. Сохранение данных в специализированном формате GeoJSON.

Этап 3. Визуализация территорий района с помощью библиотеки osmnx, где пустоты обозначены серым цветом

Этап 4. Подсчет количества пикселей цвета пустот с использованием библиотеки OpenCV

Этап 5. Определение процентного соотношения пустот относительно всего района



Мурино — 44,29%

Озерки — 7,46%

Процент пустот в районах

Итак, в Мурино практически половина территории — это пустоты. Да, район только застраивается, но на карте видно, что даже в центре, где плотная застройка, между домами — пустота, а в Озерках большой процент территории — это зоны рекреации и зеленые насаждения.

Результаты

Измерение геттоизации района оказалась хорошим примером, чтобы показать потенциал кейс-метода. В рамках проекта провели исследование, состоящее из этапов:

- уточнение термина и выделение признаков геттоизации в научной литературе;
- сбор географических данных (точки, полигоны, линии) определенного типа из двух источников, удаление дублирующихся сведений:
 - более 40 типов запросов к данным;
 - 2 728 174 географических объектов;
- анализ плотности расположения объектов на квадратный километр;
- визуализация данных: нанесение данных на карту, кластеризация данных;
- сбор географических данных (объемные объекты), исключенные данных с площади района, анализ неблагоустроенных территорий;
- сбор числовых данных (аренда жилья и формат квартир), построение сравнительных гистограмм, иллюстрирующих объем арендного жилья в разных районах.

Полученный алгоритм может применяться не только при создании методического пособия для студентов, но и при:

- разработке проектов благоустройства;
- приведении научных исследований в урбанистике;
- верификации результатов социологических исследований.

Сами автоматизированные решения не новы. Но трудности вызывает не поиск метода как таковой, а его применение на конкретной базе данных или к конкретному запросу. К тому же часто для получения результата методы приходится сочетать. Полученный алгоритм демонстрирует различные подходы к работе с данными и показывает пример того, как можно преобразовать социологическое исследование в четкие количественные характеристики на открытых данных. Дальнейшее развитие модели геттоизации послужит интерактивным примером для создания математической модели сравнительного типа.