Факультет географии и геоинформационных технологий

НУГ «Пространственный анализ и моделирование городских процессов»

Москва
2026

Семинар НУГ

Моделирование дневной активности населения на основе механистических моделей и моделей машинного обучения на примере Хельсинки

Кокурина Екатерина Константиновна
студентка 3 курса Факультета географии и геоинформационных технологий НИУ ВШЭ

Актуальность

Дневная плотность населения - важнейшая метрика для анализа городских систем. Ее анализ полезен для планирования городских территорий и оптимального размещения объектов социальной, экономической и транспортной инфраструктуры.

Проблема

Статистические данные о дневной плотности населения труднодоступны и не всегда имеют высокое пространственное разрешение, что значительно усложняет прогнозирование распределения населения на территории городов в дневное время суток

Цель

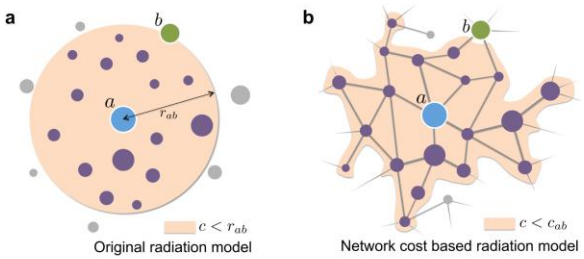
разработка оптимального подхода к моделированию дневной плотности населения на основе механистических моделей и алгоритмов машинного обучения

Критерии оптимальности: минимальный набор входных данных и высокое качество модели

Источник: Global daytime population density, 2019. https://www.mbi-geodata.com/en/2019/07/26/global-daytime-population/

Механистические модели

- основаны на изученных законах и закономерностях
- менее требовательны к качеству и объему данных по сравнению с моделями машинного обучения
- высокая интерпретируемость



$$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

T_i - общее число людей, совершающих поездки из округа i

m_i и n_j - численность населения в округах i и j

s_{ij} - общее население внутри круга с центром в i и радиусом до j

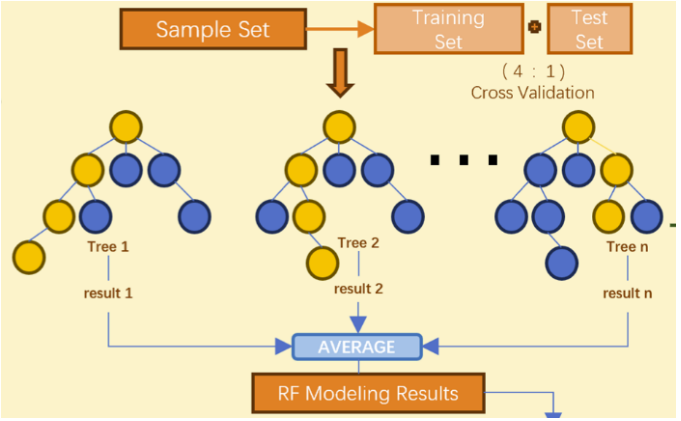
Barbosa H. et al. Human mobility: Models and applications //Physics Reports. – 2018. – T. 734. – C. 1-74.

Возможно ли интегрировать эти подходы для обеспечения более точного прогнозирования дневной плотности населения при минимальном наборе исходных данных?

Гибридные модели

ML-подходы

- автономный поиск зависимостей в данных
- повышенная чувствительность к данным



Chen Y. et al. Modeling the spatial distribution of population based on random forest and parameter optimization methods: a case study of Sichuan, China //Applied Sciences. – 2024. – T. 14. – №. 1. – C. 446.

Данные

- Агрегированные сеточные данные мобильных операторов с разрешением 250 на 250 метров о 24-часовом распределении населения в будние дни (с понедельника по четверг)

- OpenStreetMap: транспортная сеть, точки интереса (POI), данные о землепользовании

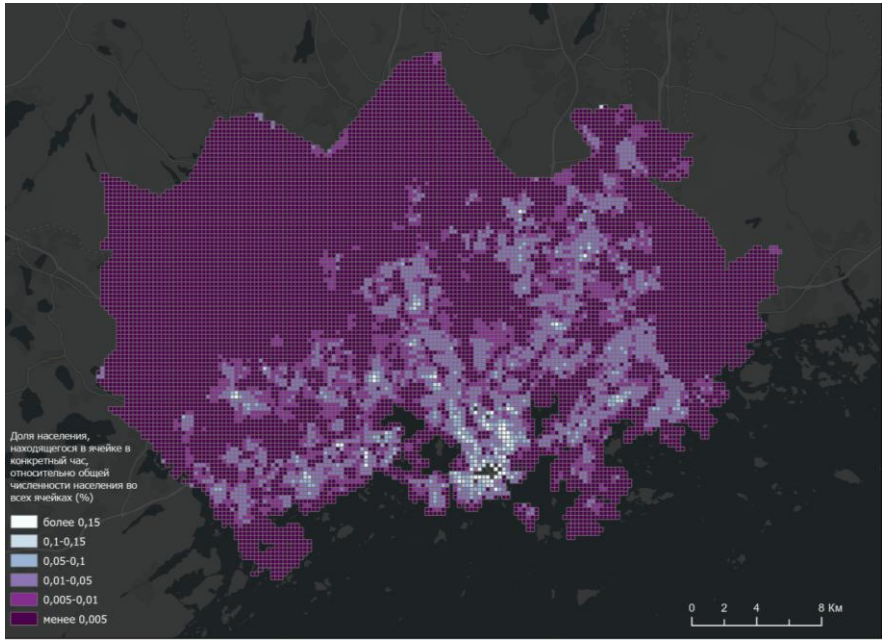
- для валидации результатов моделей

- в качестве источника данных о ночном населении

- ячейки сетки - для агрегации вспомогательных данных

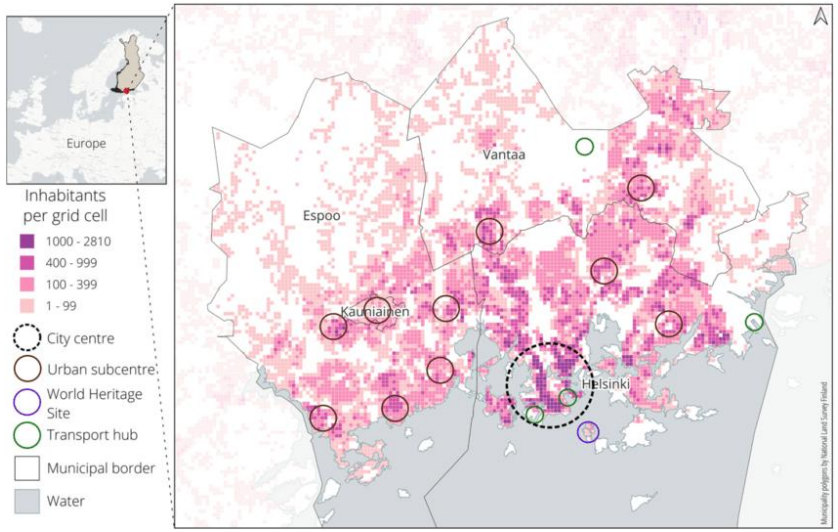
- для построения механистических моделей: точки интереса

- в качестве предикторов в моделях машинного обучения

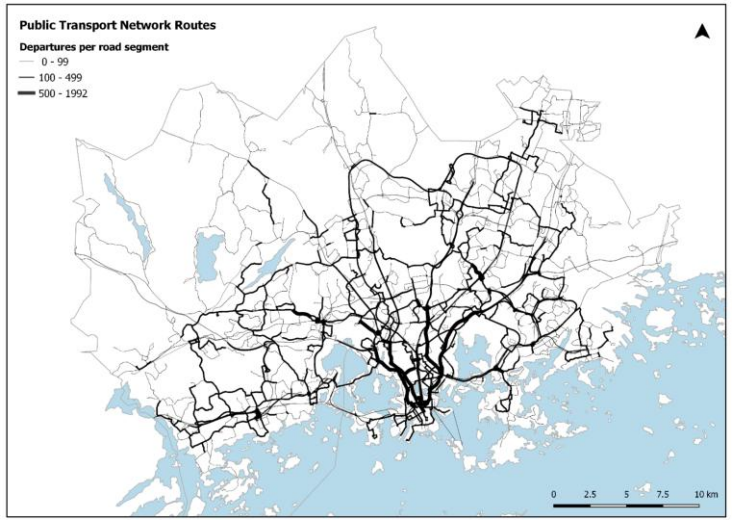


Пример визуализации исходных сеточных данных (14:00 будний день)

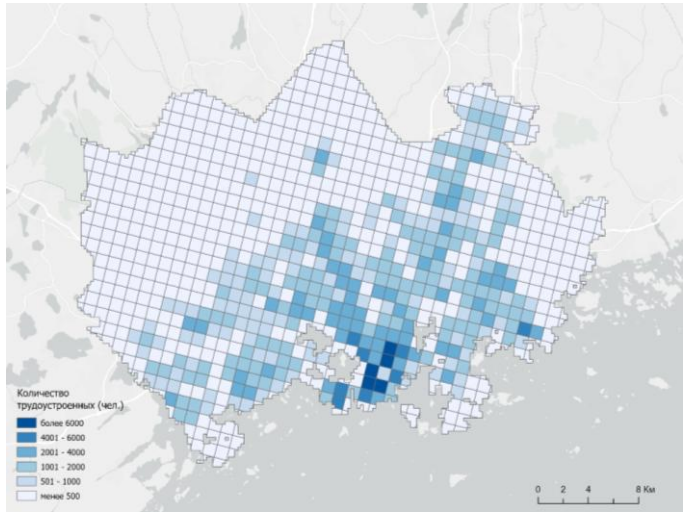
Пространственная структура и особенности размещения населения в Столичном регионе Хельсинки



Плотность населения Столичного региона Хельсинки с основными локациями (Väisänen et al., 2022)

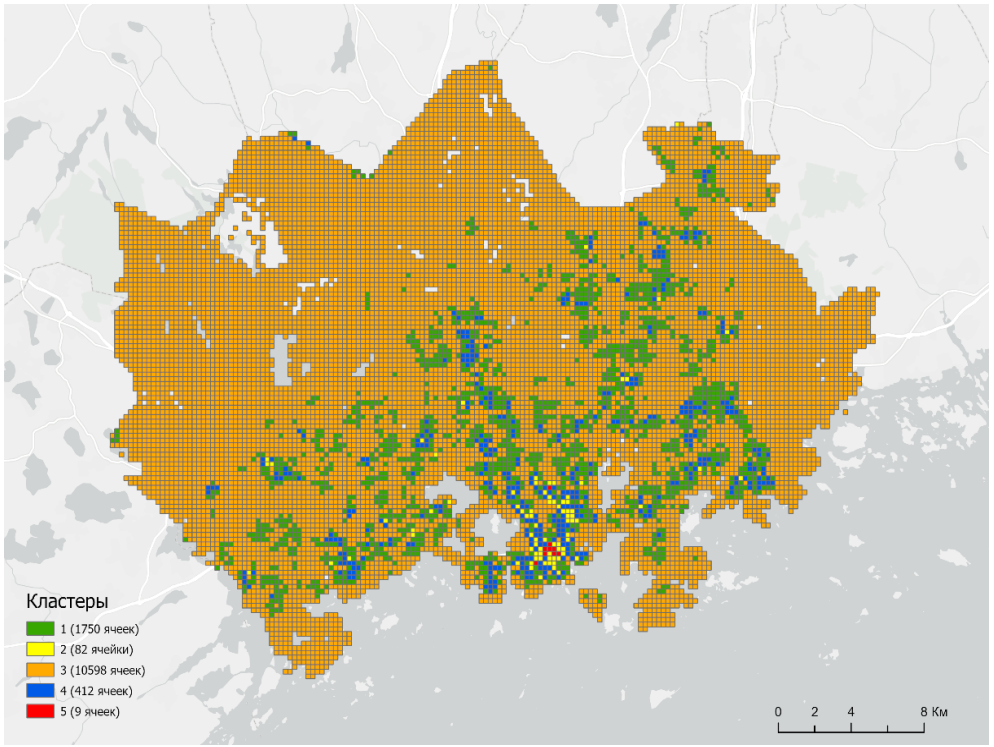


Маршруты общественного транспорта и количество отправлений в сутки на каждом участке улично-дорожной сети в Столичном регионе Финляндии в 2016 году (Bergroth, 2019)

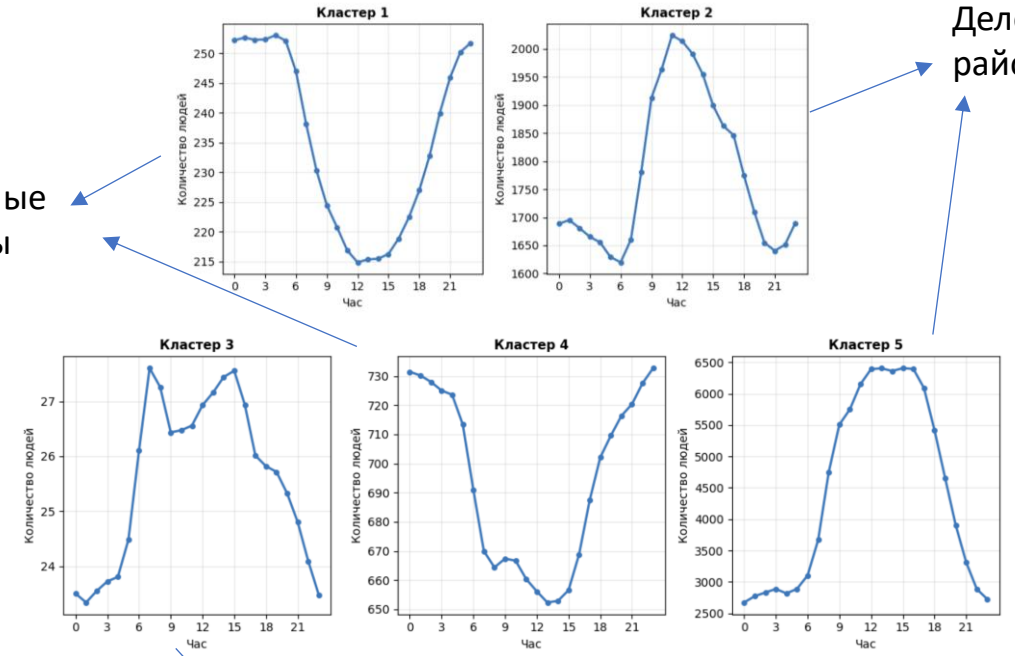


Число трудоустроенных на 2021 год по ячейкам регулярной сетки 1 км по данным GEOSTAT, 2021

Пространственно-временная динамика плотности населения в Столичном регионе Хельсинки и ее факторы



Кластеризация методом k-means (5 кластеров)



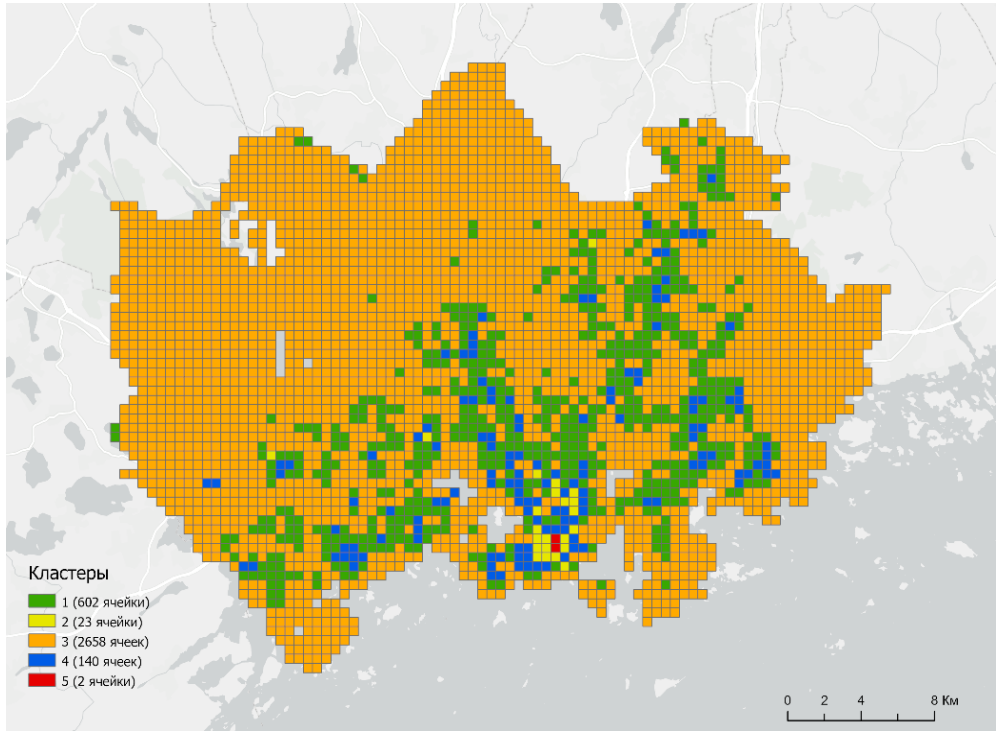
Деловые районы

Спальные районы

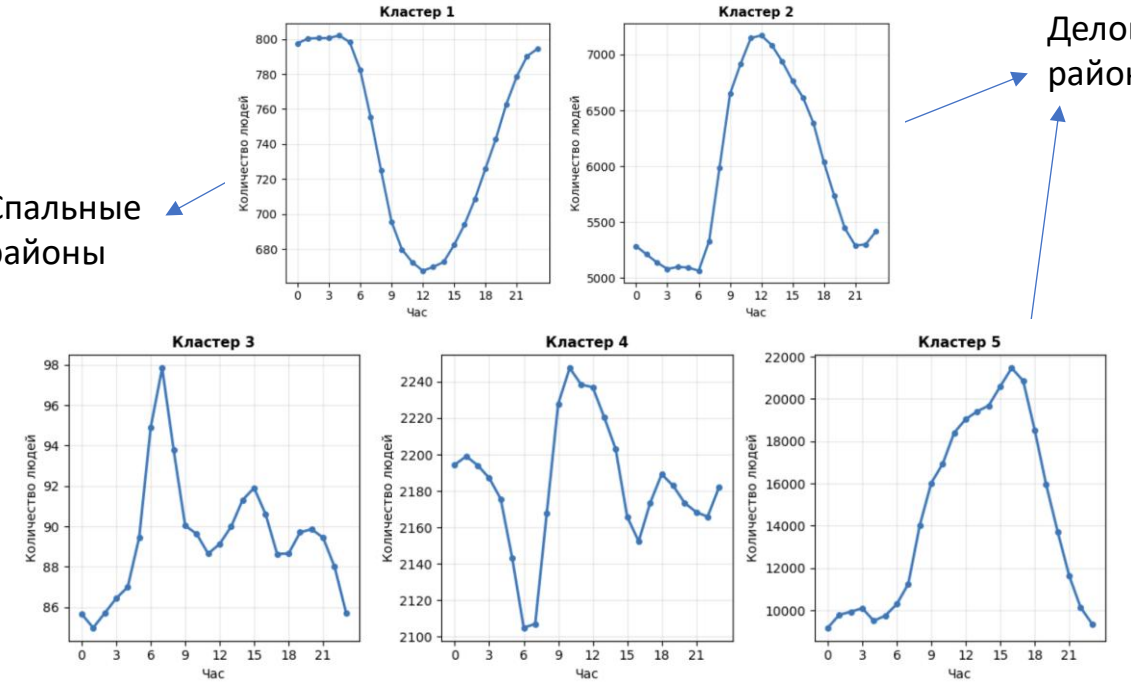
Транзитные, малонаселенные районы

- Средняя дневная плотность: 11-13 часов
- Средняя ночная плотность: 22-6 часов

Пространственно-временная динамика плотности населения в Столичном регионе Хельсинки и ее факторы



Кластеризация методом k-means (5 кластеров)



Деловые районы

Спальные районы

Транзитные, малонаселенные районы

- Средняя дневная плотность: 11-13 часов
- Средняя ночная плотность: 22-6 часов

Механистические модели

✗ Гравитационная

$$T_{ij} = \frac{m_i^a n_j^b}{f(r_{ij})}$$

a и b - настраиваемые параметры модели

Промежуточных возможностей

✗

$$T_{ij} = O_i \frac{e^{-LV_{ij-1}} - e^{-LV_{ij}}}{1 - e^{-LV_{in}}}$$

L – это калибровочный параметр модели, отвечающий за вероятность остановки в промежуточном пункте

✓ Модель выбора приоритетных возможностей

$$T_{ij} = O_i \cdot P_{ij} = O_i \cdot \frac{m_j/S_{ij}}{\sum_j m_j/S_{ij}}$$

(Liu et al., 2019)

✓ Радиационная

$$T_{ij} = T_i \cdot \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

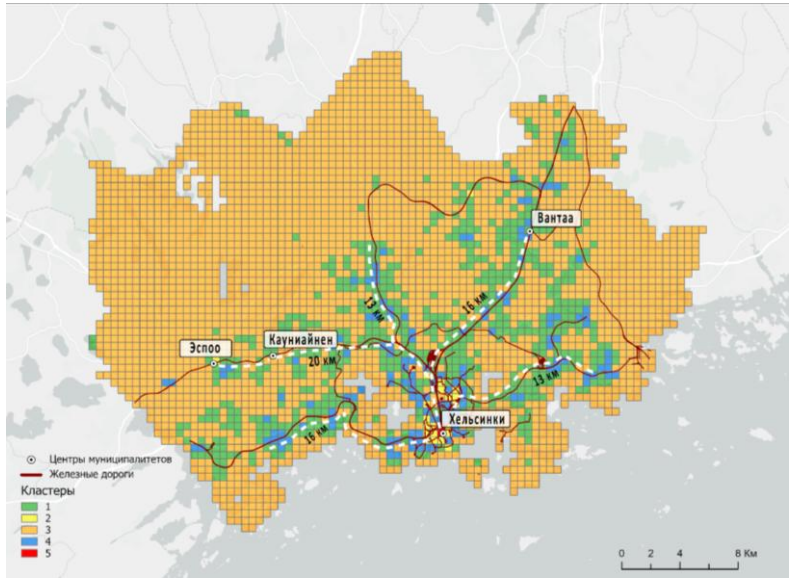
(Simini et al., 2012)

✓ Модель возможностей, взвешенных по численности населения

$$T_{ij} = T_i \frac{m_j(1/S_{ji} - 1/M)}{\sum_{k \neq i} m_k(1/S_{ki} - 1/M)}$$

(Yan et al., 2014)

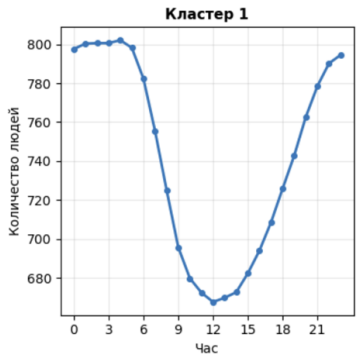
Радиационная модель (RAD)



r = 15 км

$$T_{ij} = T_i \cdot \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

- **m_i**: численность населения в *i* (ночное население)
- **n_j**: количество POI в *j*
- **s_ij**: сумма POI во всех ячейках, находящихся в радиусе от i до j
- **T_i**: общий отток из *i* (**0.6** * m_i)



56,7% процентное изменение ночной плотности относительно дневной в кластере 1 (спальный)

Модель выбора приоритетных возможностей (OPS)

$$T_{ij} = O_i \cdot P_{ij} = O_i \cdot \frac{m_j/S_{ij}}{\sum_j m_j/S_{ij}}$$

- **O_i * 0,6 -** общий отток из *i*
- **m_j -** кол-во POI в точке назначения
- **S_ij** - накопленное число POI от i до j включительно

Модель возможностей, взвешенных по численности населения (PWO)

$$T_{ij} = T_i \frac{m_j(1/S_{ji} - 1/M)}{\sum_{k \neq i} m_k(1/S_{ki} - 1/M)}$$

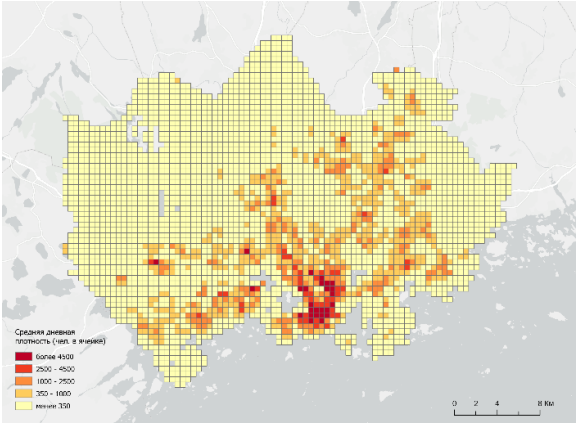
- **m_j -** кол-во POI в точке назначения
- **S_ji** - накопленное население в радиусе вокруг точки назначения j
- **M** - сумма населения во всех ячейках
- **T_i**: общий отток из *i* (**0.6** * m_i)

GitHub

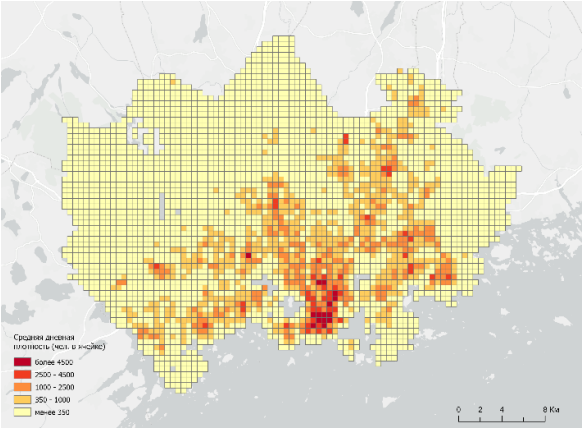
Сравнение точности механистических моделей

Метрика	Радиационная модель	**Модель выбора приоритетных возможностей**	Модель возможностей, взвешенных по численности населения
R^2	0.71	**0.75**	0.67
MAE	158.59	**148.42**	173.00
RMSE	505.32	**470.17**	537.86
Корреляция Пирсона	0.86	**0.88**	0.85

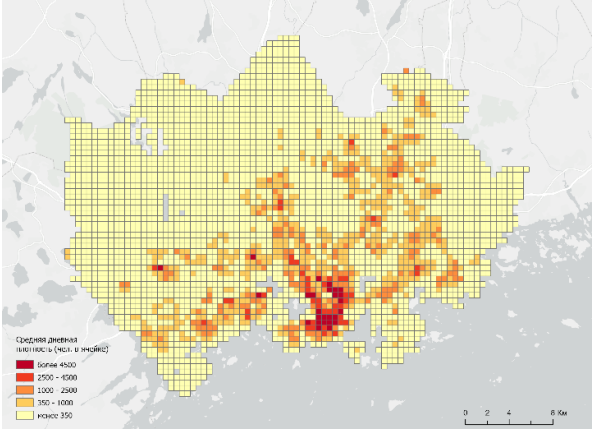
Фактическая средняя дневная плотность населения



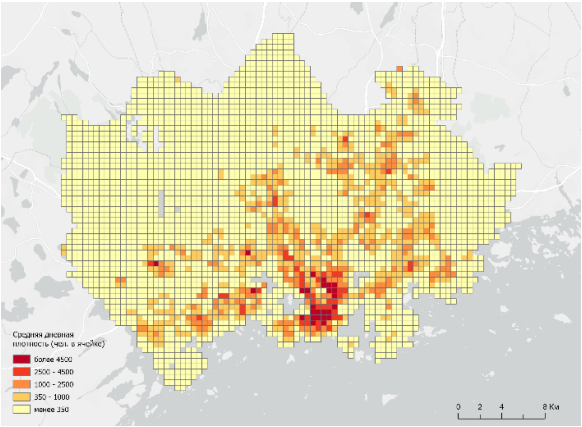
Модель выбора приоритетных возможностей



Радиационная модель

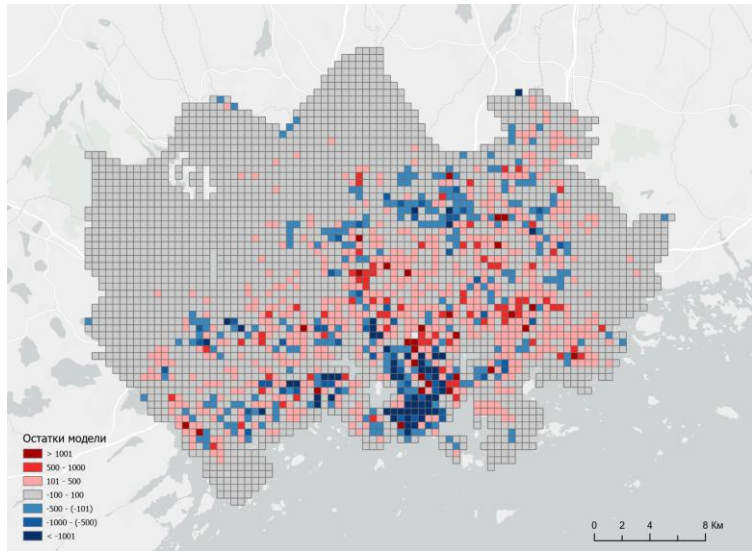


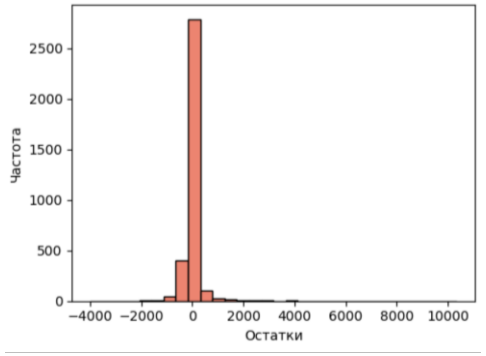
Модель возможностей, взвешенных по численности населения



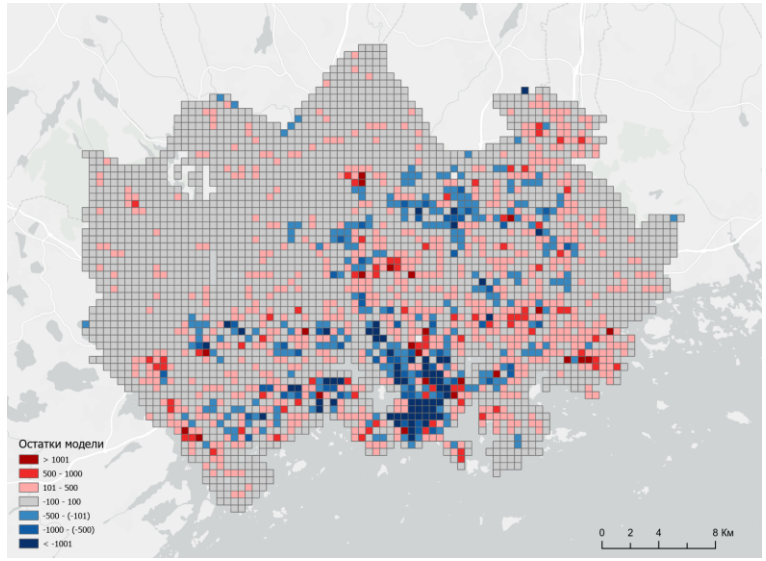
Остатки моделей

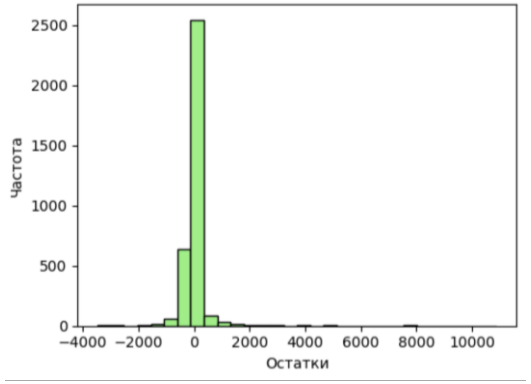
RAD



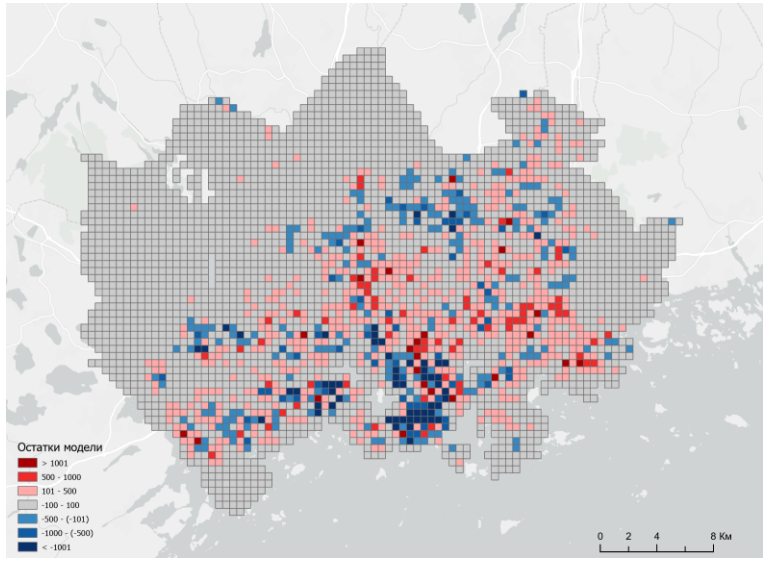


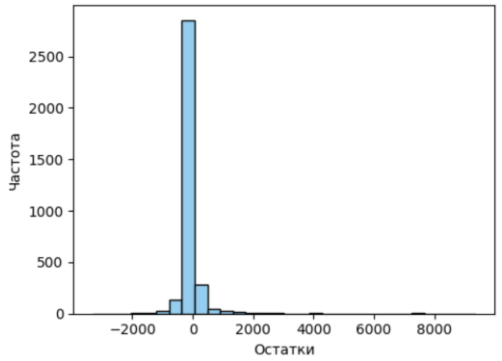
PWO





OPS

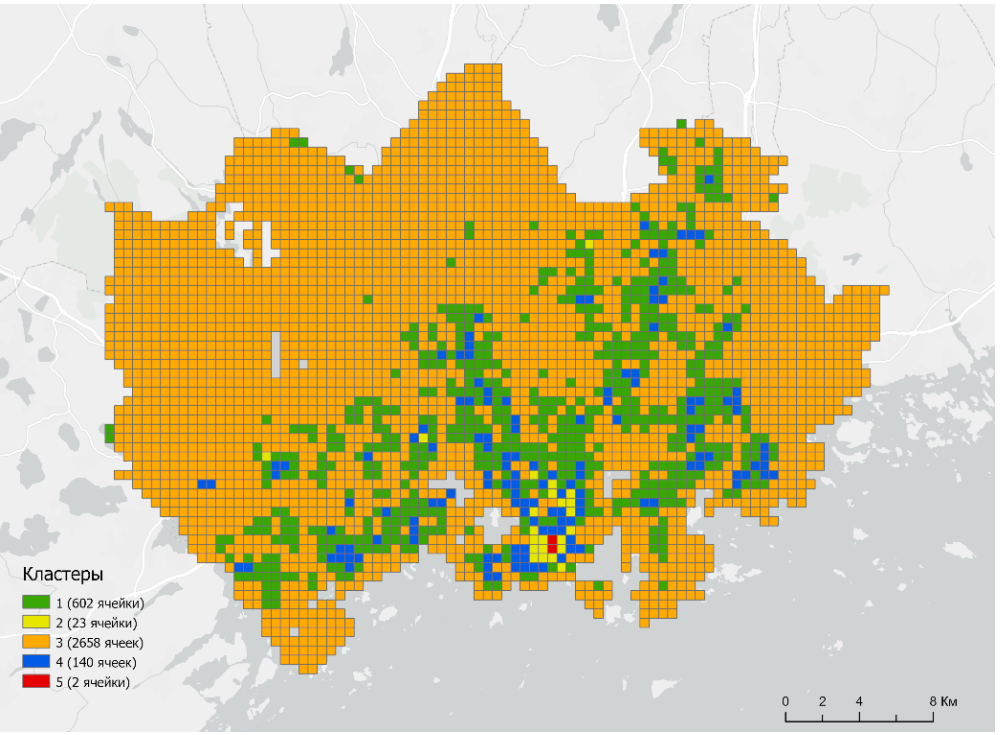




Сравнение точности механистических моделей

MAE/среднюю дневную плотность в кластере

Кластеры	Радиационная модель	Модель выбора приоритетных возможностей	Модель возможностей, взвешенных по численности населения
1 (спальные районы)	0.45	0.44	0.43
2 (деловые районы)	0.39	0.33	0.44
3 (транзитные районы)	0.76	0.67	0.94
4 (транзитные районы)	0.31	0.32	0.33
5 (деловые районы)	0.52	0.46	0.56



Модели машинного обучения (Random Forest)

n = 500
Train:test 70:30

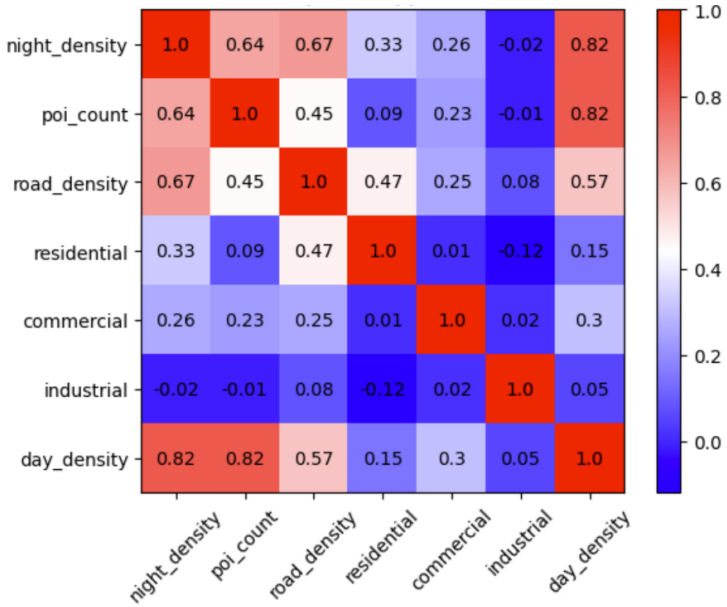
- Случайное разделение

R^2 = 0,82 MAE = 118,08 RMSE = 342,5 Корреляция Пирсона = 0,91

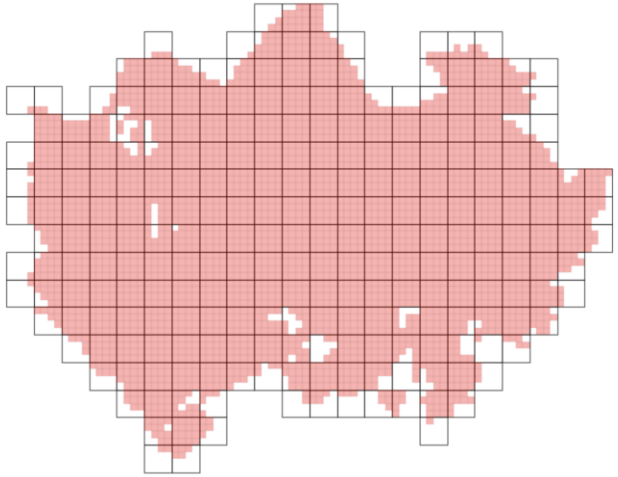
- С применением пространственной кросс-валидации

R^2 = 0,79 MAE = 135,38 RMSE = 408,81 Корреляция Пирсона = 0,89

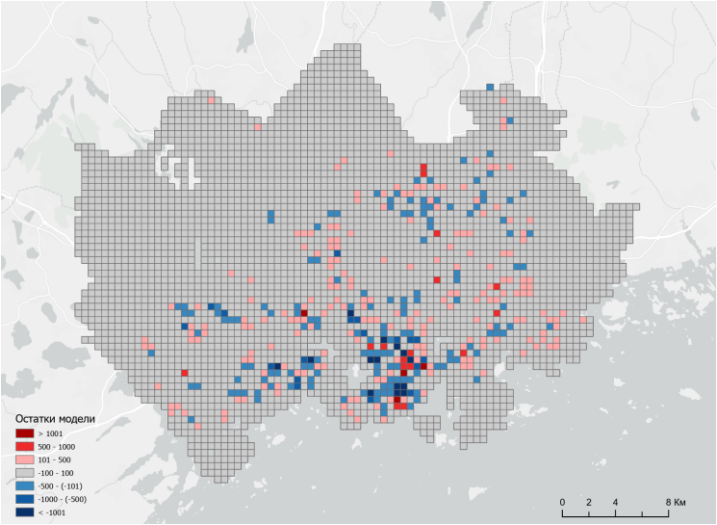
Матрица корреляции



Пространственные блоки 2000 на 2000 м



Остатки модели



Гибридные модели

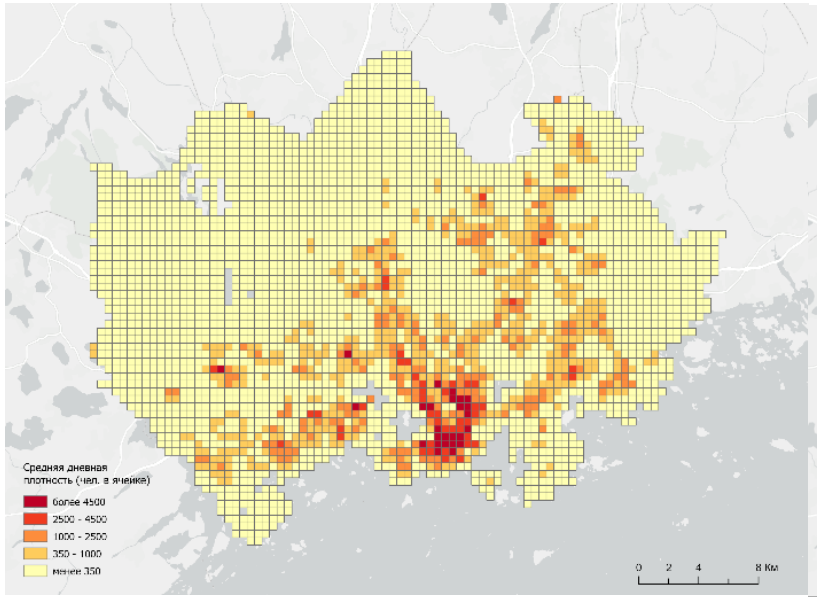
- предсказанное моделью выбора приоритетных возможностей (OPS) как признак в RF

  R^2 = 0,73 (80:20) MAE = 149,91 RMSE = 483,54 Корреляция Пирсона = 0,86

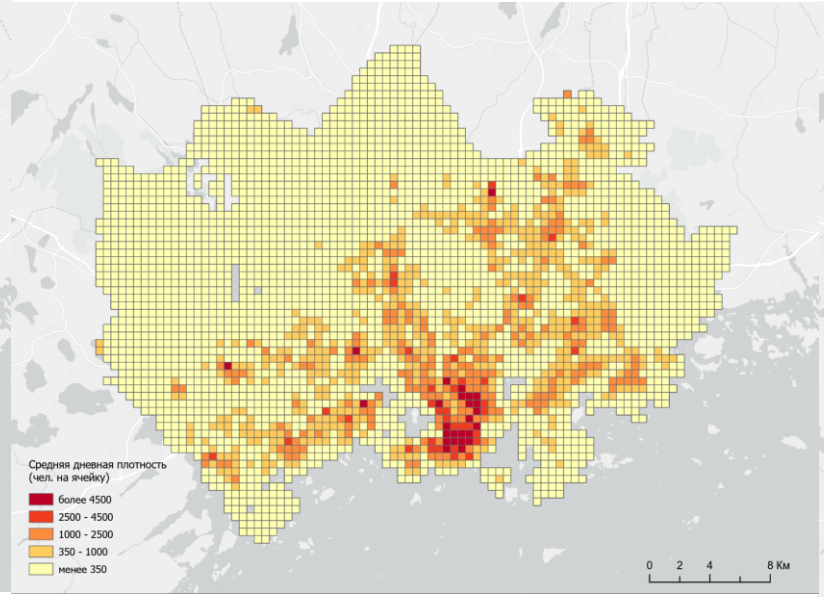
- RF предсказывает ошибку OPS используя признаки

  R^2 = 0,83 (70:30) MAE = 130,46 RMSE = 334,58 Корреляция Пирсона = 0,92

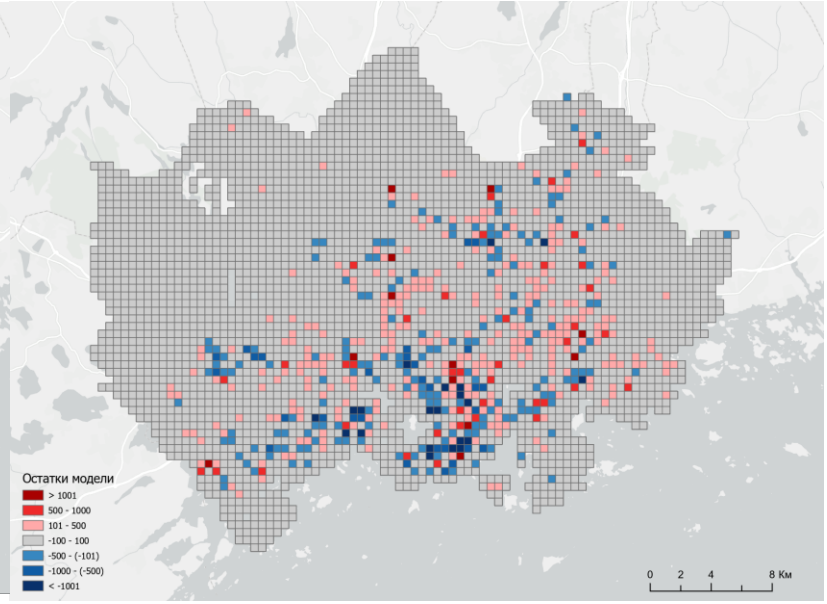
Фактическое



Предсказанное



Остатки



Выводы

1) Механистические модели, построенные с использованием открытых данных о точках интереса (POI) из OpenStreetMap, могут использоваться для предсказания мобильности на уровне агломераций, однако имеют ряд ограничений, связанных с занижением дневной плотности в центральных районах и переоценкой плотности на периферии
2) Лучший результат из механистических моделей показала **модель выбора приоритетных возможностей**
3) **Модели машинного обучения**, построенные с использованием дополнительных наборов открытых данных (категорий землепользования, плотности дорожной сети), показывают более высокую точность, но также занижают плотность в центральных районах
4) **Гибридная модель**, разработанная на основе модели выбора приоритетных возможностей и алгоритма случайного леса, является оптимальной для прогнозирования дневной плотности. Переоценка дневной плотности в наиболее «активных» районах может быть полезна для качественного планирования территорий

Список источников

1) Bergroth C. et al. A 24-hour population distribution dataset based on mobile phone data from Helsinki Metropolitan Area, Finland //Scientific data. – 2022. – Т. 9. – №. 1. – С. 39.
2) Chen Y. et al. Modeling the spatial distribution of population based on random forest and parameter optimization methods: a case study of Sichuan, China //Applied Sciences. – 2024. – Т. 14. – №. 1. – С. 446.
3) Liu E., Yan X. New parameter-free mobility model: Opportunity priority selection model //Physica A: Statistical Mechanics and its Applications. – 2019. – Т. 526. – С. 121023.
4) Liu L. et al. Population spatialization in Zhengzhou city based on multi-source data and random forest model //Frontiers in Earth Science. – 2023. – Т. 11. – С. 1092664.
5) Ravenstein E. G. The laws of migration. – Royal Statistical Society, 1885.
6) Ren Y. et al. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges //Nature communications. – 2014. – Т. 5. – №. 1. – С. 5347.
7) Rojradtanasiri P., Tamura J., Kobayashi M. Estimating ambient population density using physical features from GIS and machine learning: a study based on Japanese neighborhood //Journal of Asian Architecture and Building Engineering. – 2025. – Т. 24. – №. 6. – С. 5787-5802.
8) Yan X. Y. et al. Universal predictability of mobility patterns in cities //Journal of The Royal Society Interface. – 2014. – Т. 11. – №. 100. – С. 20140834.
9) Wang J. et al. Urban human mobility: Data-driven modeling and prediction //ACM SIGKDD explorations newsletter. – 2019. – Т. 21. – №. 1. – С. 1-19.
10) Bergroth C. A 24-hour dynamic population distribution dataset based on mobile phone data from Helsinki Metropolitan Area, Finland [Электронный ресурс]/ URL: https://zenodo.org/records/6106064
11) Geofabrik. Download OpenStreetMap for Finland [Электронный ресурс] // Geofabrik Download Server. Karlsruhe, 2026. URL: https://download.geofabrik.de/europe/finland.html
12) GEOSTAT [Электронный ресурс] // GISCO. Geodata. Population distribution: Eurostat census grid 2021. URL: https://ec.europa.eu/eurostat/web/gisco/geodata/population-distribution#Eurostat

ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •