

Федеральное государственное автономное образовательное
учреждение высшего образования

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет экономических наук

КУРСОВАЯ РАБОТА

Анализ детерминант привлекательности районов города на основе
агентного моделирования и эконометрических методов

по направлению подготовки Экономика
образовательная программа «Экономика и статистика»

Работу выполнили:

Егорова Анастасия Тимофеевна, БСТ192

Золотарев Антон Олегович, БСТ182

Попова Диана Олеговна, БСТ196

Руководитель:

доцент, кандидат экономических наук,

Пильник Николай Петрович

Москва 2021

Содержание

1 Мотивация исследования	3
2 Теоретический обзор используемых методов	5
2.1 Обзор постановок задачи геоаналитики и лучшие практики их решения	5
2.2 Учёт расположения статических объектов для оптимального расположения точек продажи	7
2.3 Методы моделирования пассажиропотоков	8
2.4 Учёт имеющихся конкурентов по отрасли для оптимального расположения точек продажи	11
2.5 Подходы к пониманию оптимальности расположения. Формирование теории размещения производства	17
3 Решение задачи оптимального размещения объекта на реальных данных	22
3.1 Постановка задачи	22
3.2 Обзор имеющихся данных	22
3.3 Анализ возможностей применения существующих моделей оптимального расположения объекта	27
3.4 Ядерная оценка плотности	29
3.5 Создание переменных для анализа	31
3.6 Применение моделей для предсказания успешности заведения	35
4 Решение задачи с учётом недостатков предыдущих методов	39
4.1 Функционал качества модели	39
4.2 Оптимизация заданного функционала	41
5 Заключение	44

1 Мотивация исследования

Оценка привлекательности коммерческой недвижимости является сложным и важным экономическим процессом, зависящим от огромного количества факторов, влияющих на итоговый успех предприятия. Важность оценки привлекательности коммерческой недвижимости в районе с точки зрения получения прибыли обуславливается необходимостью эффективно расположить объекты, минимизируя при этом необходимые затраты на размещение и максимизируя выгоду от оптимального расположения.

Поиск оптимального расположения объекта коммерческой недвижимости в районе зависит от таких факторов, как, например, спрос на услугу/товар в исследуемом районе, имеющиеся конкуренты в окрестности (возможно наличие монополий), которые могут перетягивать спрос, выбор оптимальной торговой площади из доступных для аренды, через которую проходит наибольший поток людей. Непосредственно задача оптимального размещения объектов не сформулирована в общем виде ввиду всеобъемлющего характера данной темы, поэтому для начала сформулируем задачу.

Необходимость развития методологических аспектов оценки привлекательности объектов коммерческой недвижимости обусловлена, во-первых, социальными и экономическими функциями сферы коммерческой недвижимости; а во-вторых, ее возрастающей ролью в экономике в последние годы.

По итогам данной работы были рассмотрены и применены разные постановки задачи оптимального расположения объектов на данных о торговых предприятиях г. Москва, была предпринята попытка получить некоторую модель, которая при анализе заданных факторов с весами (в рассматриваемом случае для упрощения были использованы такие факторы, как доступные для аренды торговые площади, конкурентные предприятия в районе, пешеходные потоки и население в районе) выдаёт оптимальное расположение. В начале работы с задачей имелись следую-

щие данные: геолокация жилых домов района с количеством заселенных квартир в качестве весов, геолокация конкурентных заведений с количеством положительных отзывов в качестве весов, геолокация, цена и площадь доступных для аренды торговых площадей внутри района. В результате обработки этих данных модель выдаёт оптимальную точку для расположения того или иного коммерческого объекта в рамках рассматриваемого района г. Москва.

В качестве данных для модели по трем районам были выбраны следующие коммерческие направления объектов недвижимости: отрасль общественного питания (рестораны, кафе и пр.), бытовых услуг, а также продовольственные торговые предприятия. В процессе разработки модели было принято решение сузить спектр исследуемых объектов и остановиться на предприятиях, оказывающих услуги общественного питания. Данное решение мотивировано насыщенностью и высоким потенциалом рынка «B2C» в России, а именно сегмента общественного питания, который показывает тенденцию роста в разрезе последних лет. После кризиса данной области, связанного с эпидемиологической ситуацией в России и мире, многие предприятия общественного питания закрылись из-за правовых ограничений и сильного снижения трафика, что еще раз указывает на потенциал к росту после отмены ограничительных мер. Соответственно, в данной работе проявляется интерес к существующему на данный момент трафику на данный момент и оптимальному размещению коммерческих объектов именно общественного питания ввиду изменившейся ситуации на рынке после частичной отмены ограничительных мер. Оптимальное расположение объекта коммерческой недвижимости в разработанной модели зависит от положения статичных объектов в зоне доступности, а именно метро, остановок наземного общественного транспорта и жилых домов. Выбор данных факторов обусловлен, во-первых, тем, что при поиске оптимального расположения предприятия стоит учитывать наличие транспортных сетей в зоне доступности, так как это является ключевым фактором, влияющим на доходность предприятия. Это объясняется высоким трафиком потен-

циальных потребителей, который позволяет генерировать тем больше прибыли, чем ближе предприятие находится к станциям метро или же остановкам наземного транспорта. Данный фактор особенно важен для предприятий питания, так как на данный момент рынок кафе и ресторанов является одним из самых перспективных и молодых в России, а также, по мнению экспертов, происходит увеличение среднего класса в РФ, и современный житель Москвы все чаще и чаще пользуется услугами ресторанов и кафе. Для жителя мегаполиса с увеличивающимся темпом жизни важен комфорт и удобство и поэтому быстро оказываемые услуги питания будут пользоваться большим спросом возле метро и остановок транспорта.[7] Во-вторых, на проходимость предприятия также влияет и расположение жилых домов в торговой зоне объекта, то есть предприятие общественного питания, находящееся на регулярном маршруте “остановка транспорта - дом” потенциально может генерировать большую прибыль, нежели объект, в зону обслуживания которого данный маршрут не выходит. Стоит отметить, что пандемия COVID-19 сильно изменила потребительские привычки населения: в России наблюдается рост рынка доставки еды, и близость ресторана или кафе к жилым домам выгодна как с точки зрения возможности для очного посещения, так и скорости доставки.

2 Теоретический обзор используемых методов

2.1 Обзор постановок задачи геоаналитики и лучшие практики их решения

Поставленная нами урбанистическо-экономическая задача происходит из проблемы оптимального размещения объектов, тщательно изученной в англоязычной академической среде во второй половине XX века. Ключевыми двигателями к её дальнейшему исследованию тогда стали задачи о р-медиане, р-центре, простом размещении завода и квад-

ратичном присвоении [15, с. 36]. Приведём формальную постановку задачи оптимального размещения объекта(по аналогии с [14, с. 5110]):

$i = 1, 2, \dots, n$ - индексы существующих конкурентных объектов, аналогичных по функционалу

$j = 1, 2, \dots, m$ - индексы потенциальных потребителей, предъявляющих спрос на соответствующий товар или услугу.

$a_j = (a_{j1}, a_{j2})$ - координаты статичного расположения потребителя j

d_j - спрос со стороны потребителя j

s_i - вес на рынке конкурента i

$q_i = (x_i, y_i)$ - координаты расположения предприятия конкурента i

$d(q_i, a_j)$ - расстояние между конкурентом i и потребителем j

w_{ij} - объём потребления блага или услуги потребителем j в предприятии i

В результате постановки задачи будет решаться проблема минимизации издержек для нового предприятия через поиск его оптимального расположения:

$$\begin{cases} (x^*, y^*) = \operatorname{argmin} C(x, y | \theta) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} d(q_i, a_j) \\ \sum_{i=1}^n w_{ij} = d_j, j = 1, 2, \dots, m \\ \sum_{j=1}^m w_{ij} \leq s_i, i = 1, 2, \dots, n \\ w_{ij} \geq 0, g_i(x, y) \geq 0 \end{cases}$$

Решение данной задачи может быть осуществлено через метод релаксаций Лагранжа или с помощью приближенных методов оценки как NP-полная задача [13, с. 60—72].

Очевидно, у такой постановки задачи есть свои недостатки, главным из которых кажется предположение о линейной зависимости увеличения издержек на обслуживание клиента при увеличении расстояния. В итоговом функционале нашей модели для решения практической задачи

по поставленной теме данный недостаток будет учтён.

Также следует добавить, что для задачи оптимального размещения объекта есть множество модификаций, учитывающих случайность расстояния ввиду факта постоянного движения клиентов, формирующихся в полноценные пассажиропотоки [14, с. 5111—5112], другие модификации будут рассмотрены в последующих главах теоретического обзора.

2.2 Учёт расположения статических объектов для оптимального расположения точек продажи

Приведённая в предыдущем разделе постановка задачи рассматривает потенциальных потребителей товара или услуги как статические объекты, то есть их расположение относительно предприятий-конкурентов и нового предприятия, оптимум для которого ищется, должно быть зафиксировано.

Вообще, важность учёта расположения статических объектов обуславливается самой постановкой задачи и подчёркивается как в практически-ориентированных [3], так и в теоретических статьях [11, с. 101—102], посвящённых данной теме. Этот фактор является системообразующим и позволяет моделировать непрерывную оценку плотности через ядерную оценку плотности [12, с. 312] для изучаемой зоны, что способствует качественной визуализации и, соответственно, более обоснованному выбору оптимальной точки расположения нового предприятия.

Для задач геомаркетинга статическими объектами принято считать такие точки, из которых к нам может прийти потенциальный покупатель, то есть жилые дома, место работы, остановки общественного транспорта и далее.

В практической части нашей работы в качестве статических объектов будут приняты жилые комплексы, учитывающие примерное количество проживающих в каждом из них, а также станции метро и автобусные остановки, генерирующие массивный пласт ежедневных пассажиропотоков в городе Москва.

2.3 Методы моделирования пассажиропотоков

Модели оценки привлекательности объекта коммерческой недвижимости основаны на обратной зависимости пространственных и временных издержек потребителя на то, чтобы добраться до этого объекта. Однако если человек проходит через какое-то место «по дороге», и оно не является конечной точкой маршрута, издержки на посещение объекта складываются только из временных затрат. Поэтому для нахождения оптимального размещения точки необходимо учитывать пешеходный и транспортный трафик.

В ходе работы введены два допущения, связанные с подходом к моделированию пассажиропотоков в целом. Во-первых, жители города рассмотрены на макроскопическом уровне, то есть через анализ потоков людей без учета действий отдельно взятого жителя. Во-вторых, действия и маршрут человека являются независимыми: в расчет не берутся движения людей группами или толпами.

Приведём основополагающие факторы, влияющие на формирование трафика:

- ”Потокообразующие факторы, то есть размещение объектов, порождающих передвижения: места проживания, приложения труда, культурно-бытового обслуживания и т.д.
- Характеристики транспортной сети (количество и качество улиц, дорог, судоходных путей, маршруты и т.д.)
- Поведенческие факторы: мобильность населения, предпочтения при выборе маршрутов и т.д.”[4, с. 75]

Моделирование пассажиропотоков принято разделять на 4 этапа:

1. Оценка численности населения, объемов ежедневных передвижений в целом
2. Разделение по способам передвижения - пешим или на транспорте

3. Построение матриц корреспонденций, считающих объем трафика между парами объектов

4. Распределение корреспонденций по всем возможным маршрутам

Пассажир оценивает пути передвижения на основе «обобщенной цены пути»: чем она выше, тем ниже привлекательность маршрута. Цена может составляться из факторов денежных затрат и личных предпочтений пассажира, однако наиболее значимым показателем является время.

Матрица корреспонденций служит количественной характеристикой структуры передвижений. Наиболее распространенными моделями расчета матрицы являются гравитационная и энтропийная модели и модель промежуточных возможностей Стaufфера. Гравитационный подход основан следующим образом: «корреспонденция из района i в район j пропорциональна общему объему отправления из центра i общему объему прибытия в центр j и некоторой функции $c(t_{ij})$, зависящей от транспортного расстояния t_{ij} между центрами i и j » [4, с. 76]. Энтропийная модель более интерпретируема и строится на описании поведения пассажиров, которые «случайным образом распределяются по некоторому набору возможных состояний. Независимый и случайный выбор всеми пассажирами своих состояний приводит к тем или иным макроскопическим состояниям системы» [4, с. 76].

Второй тип подходов представляет модель промежуточных возможностей Стaufфера. В ходе решения задачи подсчета объемов передвижения Стaufфер предполагает, что на трафик влияет не столько расстояние между начальной и конечной точками маршрута, сколько число альтернативных пунктов прибытия, встречающихся на пути, то есть число альтернативных возможностей посещения.

Оба подхода являются диаметрально противоположными: модели гравитационного типа учитывают расстояние между объектами, в то время как модели промежуточных возможностей фокусируются на альтернативных точках прибытия, поэтому зачастую используются комбинации двух концепций.

В рамках упрощения модели вместо построения и расчета корреспонденций по парам объекта взят только один базовый маршрут пассажира: от дома до метро. Во-первых, этот маршрут является одним из самых популярных: ежедневно метро использует около 6 млн человек, что составляет чуть меньше половины населения города. Во-вторых, его достаточно просто построить, что облегчает задачу подсчета проходимости.

Стоит отметить, что в модифицированной модели Д. Хаффа трафик учитывается только исходя из населения дома: чем больше число квартир, и, соответственно, число жителей в доме, тем больше будет проходимость объекта коммерческой недвижимости, расположенного в этом доме. Однако данное упрощение совершенно не учитывает потокообразующий фактор, поэтому не будет использовано в дальнейшей модели.

Подсчет проходимости предполагает присваивание определенных весов объектам, обладающим потокообразующим характером. Как уже было оговорено в предыдущем подразделе, в данной работе предполагается считать три места концентрации людей: ближайшие станции метро, остановки общественного транспорта и дома. Важность отдельно взятого жилого дома будет определяться пропорционально числу жителей в нем. Остановки общественного транспорта будут обладать весом, равным среднему числу жителей домов в некотором радиусе. Станции метро будет присвоено наибольшее значение, равное ее среднесуточной загруженности. Коэффициент умножается на евклидово расстояние между точкой, в которой предполагается открытие объекта коммерческой недвижимостью, и потокообразующими факторами. Чем ближе в рамках доступных для аренды помещений к такому месту находится объект, тем выше вероятность высокой проходимости в нем. Так, например, объекты, расположенные возле станций метро, будут считаться наиболее проходимыми, так как около них концентрируется большая часть населения района.

2.4 Учёт имеющихся конкурентов по отрасли для оптимального расположения точек продажи

Одним из ключевых и важнейших факторов при расчете оптимального расположения объекта коммерческой недвижимости также является конкуренция, а именно конкурирующие предприятия, которые могут располагаться в районе. При анализе конкурентной среды следует определиться, стоит ли вообще выходить на какой-либо рынок (в зависимости от направленности объекта) в некотором районе, будет ли это выгодно с точки зрения прибыльности коммерческого предприятия в перспективе, или же последний не выдержит конкуренции и окажется убыточным.

Перед анализом и учетом конкурентов в районе, в котором непосредственно требуется найти оптимальное расположение объекта коммерческой недвижимости, во-первых, следует определить спрос на товары или услуги, предоставляемые предприятием, ведь возможно то, что последнее предлагает потребителям, окажется невостребованным на конкретном рынке в районе и предприятие окажется убыточным. Также стоит учитывать целесообразность размещения какого-либо предприятия в целом. Здесь стоит упомянуть о таком виде анализа, как геомаркетинг. Геомаркетинговый анализ является сравнительно новым методом исследования объектов коммерческой недвижимости. Он представляет собой разновидность маркетингового анализа, включающего в себя исследование инфраструктуры отрасли, в которой собирается разместиться компания, также рассматривая конкурентную среду конкретной территории, то есть используя помимо маркетинговых инструментов еще и географические. Можно заключить, что сам анализ – это исследование географически локализованной маркетинговой информации. В рамках разработки модели, представленной в данной работе, методы геомаркетинга могут быть использованы для учета конкурентов в районе, то есть для подготовки функционала непосредственно самой модели. Функционал должен учитывать расстояние до конкурентных предприятий (чем более

конкурент популярен у потребителей, тем дальше от него стоит расположиться).

Однако для начала стоит разобраться в некоторых понятиях, которые будут решающими при выборе используемых методов для модели: зона обслуживания, зона доступности и зона охвата. Данные определения являются ключевыми в геомаркетинге и при разработке модели и проведении самого анализа стоит определиться с каким именно понятием нужно будет работать. Зона обслуживания коммерческого объекта представляет собой торговую зону, то есть непосредственно ту территорию, на которой потенциально будут продаваться товары или предоставляться услуги. Торговая зона может расширяться при, например, успешности предприятия. Также есть несколько видов областей торгового обслуживания, а именно основная, второстепенная и окраинная (периферийная). Основная торговая зона охватывает около 50–80% потенциальных потребителей товаров и/или услуг, предоставляемых предприятием, и находится в пешей доступности, то есть клиент может преодолеть расстояние до объекта коммерческой недвижимости за 10–15 минут (радиус зоны обхвата при расположении предприятия в центре окружности 800–1000 м). То есть это ближайшая к предприятию территория и генерирующая основную прибыль за счет высокой плотности потребителей. К этой зоне можно отнести, например, продуктовые магазины, магазины бытовой химии. Второстепенная зона обслуживания объекта включает в себя 15–25% потребителей, распределенных с меньшей плотностью, нежели в основной зоне обслуживания. В данной зоне радиус обслуживания составляет около 2–2,5 км и сюда входят более крупные магазины, например супермаркеты, в которые потребители приезжают примерно раз в неделю. Периферийная торговая зона непосредственно охватывает «оставшихся» потребителей, которые не вошли ни в основную, ни во второстепенную зоны обслуживания. Радиус данной зоны обслуживания составляет 10–12 км и сюда входят, например загородные торговые центры.[8]

Зона доступности позволяет оценить целесообразность размещения

объекта коммерческой недвижимости непосредственно на какой-либо территории. Данная область намечается на карте и показывает участок, внутри которого потенциальный потребитель может добраться до объекта за конкретное фиксированное время. Однако стоит отметить, что то, насколько большая зона доступности, не играет решающей роли, так как в первую очередь при ее анализе обращают внимание на уровень конкуренции внутри нее, а также на качественные и количественные характеристики потенциальных потребителей на выделенной территории (например, количество людей, проживающих на ней, их средний уровень заработной платы, количество конкурентов в отрасли на данном участке и т. д.). Также стоит отметить некоторую ограниченность использования данной метрики, так как она не позволяет результативно оценить расположение конкурентов в зоне доступности. Для более эффективной оценки оптимального расположения чаще используют зону охвата.

Зона охвата отражает фактическую ограниченную территорию, на которой проживают потенциальные потребители и с которой они будут приезжать или приходить на будущий объект коммерческой недвижимости. Зона охвата строится с помощью моделирования поведения потребителей, однако нельзя путать ее с областью транспортной доступности. Разница состоит в том, что потребители из области транспортной доступности возможно придут или приедут на объект, а в случае зоны охвата они точно захотят приобрести товары или воспользоваться услугами с учетом конкурентов, находящихся рядом. Для построения зон охвата используется модель Хаффа, которая будет описана в следующем подразделе. Также стоит отметить, что зона охвата не является статичной, так как могут меняться население окружающей территории, конкурентная среда (открытие новых предприятий или закрытие некогда существующих), транспортное сообщение, а следовательно, и сама зона.

Теперь обратимся к наиболее распространенным методам геомаркетингового анализа. Среди них наблюдается некоторая иерархия, а именно деление на базовые, расширенные и продвинутые и в зависимости от

объекта коммерческой недвижимости, который планируется построить, стоит использовать свои методы.

Базовые методы геомаркетингового анализа могут быть использованы для объектов локального уровня, а именно для небольших предприятий, зона обслуживания которых подразумевает пешую доступность, составляющую около 10 минут. К базовым методам относят простейшие способы географического анализа, а именно построение буферных зон, геокодирование (преобразование географических координат в адрес и наоборот и нанесение на карту для дальнейшего анализа), маршрутизацию, кластеризацию объектов в районе и др. Построение буферной зоны подразумевает выделение области некоторого радиуса вокруг объекта коммерческой недвижимости. Буферная зона может быть построена вокруг точечных, линейных и полигональных объектов и иметь разное буферное расстояние, которое задается экзогенно. Также у объекта может быть несколько буферных зон, которые могут накладываться друг на друга. Большинство ГИС-приложений обладает функционалом для построения буферных зон в качестве аналитического инструмента, который может быть использован при анализе оптимального расположения объекта коммерческой недвижимости. Маршрутизация представляет собой проложение оптимального маршрута по некоторым экзогенно заданным параметрам, в основу которого может ложиться теория графов, манхэттенское расстояние и прочие способы, и, соответственно, позволяет положить начало дальнейшему исследованию спроса на товар или услугу, так как строит поведение потенциальных потребителей на географически локализованной территории.

Расширенные методы рекомендуется использовать для объектов районного уровня, которые находятся в 10 минутах транспортной доступности. К расширенным способам геомаркетингового анализа относят сегментирование потенциальных потребителей, оценку конкуренции на основе модели пространственной дифференциации Хотеллинга и прочих аналогов, построение модели Рейли. Модель Рейли позволяет найти некоторую границу между двумя розничными торговыми зонами, то есть

«точку безразличия», в которой потенциальный потребитель с равной вероятностью захочет приобрести товар или воспользоваться услугами одного из двух предприятий. Граница зависит от расположения центров двух объектов коммерческой недвижимости и соответственно от их привлекательности для самого потребителя. Данная граница рассчитывается по следующей формуле:

$$[\text{рейли}] \quad R_A = \frac{R_{AB}}{1 + \sqrt{\frac{N_b}{N_a}}}$$

где R_A - граница площади охвата центра A ; R_{AB} — расстояние между центрами ; N_A — население вокруг центра A ; N_B — население вокруг центра .

Продвинутые инструменты геомаркетингового анализа используются для предприятий городского уровня, то есть для объектов, которые располагаются в 60–90 минутах транспортной доступности. Для методов данного уровня необходимо глубокое понимание направленности объекта коммерческой недвижимости, использование практически полного функционала ГИС-систем с применением методов математического анализа. К таким инструментам можно отнести построение гравитационной модели притяжения (модели Хаффа), графов маршрутов и прочее. [6] Модель Хаффа предоставляет возможность расчета оптимального расположения объекта коммерческой недвижимости, в котором предприятие может максимизировать свою полезность, прямо зависящую от размера предприятия и обратно от расстояния до него. Ключевая идея модели заключается в том, чтобы минимизировать расстояние от потребителя до объекта. Привлекательность непосредственно самого объекта коммерческой недвижимости рассчитывается по следующей формуле:

$$A_{ij} = \frac{S_j}{T_{ij}^\lambda}$$

где A_{ij} - привлекательность объекта j для покупателя i , S_j - размер

объекта j , T_{ij} - время, потраченное покупателем i на путь до объекта j , λ - параметр, отражающий субъективный эффект влияния разных типов объектов на трату времени (рассчитывается эмпирически). [9, с. 27]

Определив привлекательность объекта для потребителей, можно вычислить вероятность, с которой потенциальные покупатели воспользуются предлагаемыми услугами или приобретут товары. Это можно сделать по формуле:

$$[9, 30] \quad P_{ij} = \frac{A_{ij}}{\sum_{j=1}^n A_{ij}}$$

Соответственно данная модель позволяет рассчитать будущий товарооборот и посещаемость объекта, а также потенциальную долю рынка, которую сможет занять объект.[10] При анализе оптимального расположения объекта коммерческой недвижимости нельзя ограничиваться методами какого-либо определенного уровня; стоит проводить разновневое исследование с использованием различных инструментов. На первом этапе анализа проводится сбор информации как таковой, то есть геокодирование объектов в районе, маршрутизация, визуализация с использованием функционала различных карт. В результате будет получена некоторая база адресов, площадей доступных для аренды, среди которых будет производиться поиск оптимального расположения, конкурентные предприятия, маршруты потенциальных потребителей. После создания базы для исследования уже используются модели в зависимости от географической направленности предприятия и его размеров.

Вторым не менее важным фактором при поиске оптимального расположения в разрезе исследования конкуренции на рынке является устройство рынка как таковое. Структура рынка требует особого внимания, так как рынок может быть монополизирован, также возможно наличие сетевых магазинов, уже завоевавших доверие потребителей. Всему вышеперечисленному следует уделять должное внимание при учете конкурентов и поиске оптимального расположения, ведь в последствии данные

факторы сильно повлияют на итоговое оптимальное положение предприятия. Для определения структуры рынка также могут быть использованы инструменты геомаркетингового анализа, о которых написано ранее.

2.5 Подходы к пониманию оптимальности расположения. Формирование теории размещения производства

Ключевым вопросом в определении наилучшего расположения объекта является критерий «оптимальности»: относительно чего необходимо максимизировать функцию, чтобы говорить о хорошем расположении объекта с точки зрения как потребителя, так и производителя? В ходе развития теории размещения производства ответ на этот вопрос подвергался изменениям.

Можно выделить три этапа формирования теоретического обоснования оптимального размещения объекта. Первый этап представлял собой построение базовых концепций, основанных на классическом подходе к экономике и абстрагированном от конкретностей моделировании. Примером такого теоретического подхода может служить локационный треугольник Лаунхардта и основанная на данной концепции теория изодапан Вебера. Задача Вебера обобщает поиск точки Ферма («необходимо найти для трёх точек такую четвёртую, чтобы суммарное расстояние от исходных точек до неё было минимальным») [5, с. 12] и задачу поиска геометрической медианы: необходимо найти такую точку на плоскости, которая минимизирует издержки доставки из данной точки в n пунктов потребления. Отличительной чертой задачи Вебера от локационного треугольника Лаунхардта является дополнение о том, что издержки не одинаковы в любой точке, и затраты на транспорт, труд, сырье и материалы зависят от местоположения. Однако существуют линии, в которых совпадают издержки отклонения, и изменения производства происходят

одинаковым образом. Эти линии Вебер назвал изодапанами. Задача Вебера предполагала решение через минимизацию издержек. Формально она выглядит следующим образом [5, с. 12]:

”Пусть имеется пространство $L = x : x \in R^2$, p -число предприятий, которые необходимо разместить на L , m - число клиентов, зафиксированных точками на L , w_j - стоимость обслуживания клиента j , $X_i = (x_i, y_i)$, $X_i \in L$ - координаты размещения предприятия i , $i = 1, 2, \dots, p$, $A_j = (a_j, b_j)$, $A_j \in L$ - координаты размещения клиента j , $j = 1, 2, \dots, m$.

Пусть $d(X_i, z_{ij})$ - расстояние между предприятием i и клиентом j , а $z_{ij} \in \{0, 1\}$ - логическая переменная, свидетельствующая об обслуживании клиента i предприятием j .“

Тогда целевая функция будет сформулирована следующим образом:

$$F(X_i, z_{ij}) = \sum_{i=1}^p \sum_{j=1}^m z_{ij} w_j d(X_i, A_j) \rightarrow \min_{X, Z}$$

$$s.t. \begin{cases} \sum_{i=1}^p z_{ij} = 1, j = 1, 2, \dots, m \\ X_i \in L, i = 1, 2, \dots, p \end{cases}$$

Другим примером базового подхода к теории размещения производства является модель линейного города Х. Хотеллинга, основанная на некооперативной игре двух производителей. По условию модели существуют две фирмы А и В, которые продают однородную продукцию по ценам p_A и p_B соответственно вдоль единственной существующей улицы. Задача состоит в поиске оптимальной цены или оптимального расположения в зависимости от условия и предпосылок. Отличие этой концепции заключается в выбранном параметре целевой функции: модель Х. Хотеллинга предполагает максимизацию прибыли по цене при существующих фиксированных издержках фирмы.

Второй этап формирования теории размещения производства представляет собой создание концепций, которые обобщают подходы основоположников этапа «чистых» теорий. Сетка Кристаллера, или теория центральных мест, разработанная В. Кристаллером в 1933 г., определяла оптимальность расположения с геометрической точки зрения, косвенно говоря о задаче минимизации транспортных издержек. В соответствии с концепцией существует сетевая структура регионов, которая обеспечивает равномерный доступ к объектам коммерческой недвижимости. Данная система иерархична, и чем больше социально-экономическое развитие региона, тем больше уровней в ней находится, и тем больше услуг предоставляет населенный пункт для нижестоящих поселений. Сетка Кристаллера имеет форму шестиугольных ячеек. Изначально задача предполагала большое количество упрощений и решала проблему оптимизации транспортных затрат, однако потом была серьезно доработана А. Лёшем.

Третий этап, начавшийся в 1950-х, ознаменован созданием первой полной теории пространственного экономического равновесия. А. Лёш дополнит модель В. Кристаллера и совершил переход к новому пониманию оптимальности: размещение фирм должно происходить в условиях максимизации прибыли, а не минимизации издержек. Более того, необходимо учитывать перманентный рост числа фирм, заполняющих рыночное пространство. Лёш продолжил развитие теории центральных

мест и математически объяснил рыночное взаимодействие производителей и потребителей, где каждая экономическая переменная имеет определенную точку в пространстве.

Все модели, созданные в рамках теории размещения производства, первостепенно учитывают интересы производителей. Эффективность размещения строится на основе экономической выгоды фирмы, в то время как задача учета привлекательности объекта для потребителей остается нерешенной. Модель Д. Л. Хаффа, сформулированная в 1963 году, фокусируется на оценке привлекательности торговой недвижимости с точки зрения потребителя. Хафф выдвинул гипотезу о прямой зависимости привлекательности объекта коммерческой недвижимости от размера объекта и обратно пропорциональной издержкам покупателя на путь до объекта. Тем не менее, модель Хаффа учитывает исключительно площадь объекта и временные затраты потребителей, хотя эти показатели могут по-разному влиять на них, а также не приводят к качественным выводам, так как площадь не всегда может демонстрировать качество магазина. В 2017 году отечественными исследователями была предложена модификация модели Д. Л. Хаффа. Критерием оптимальности расположения стал максимум ожидаемого числа посетителей в предположении, что он обеспечит максимальную прибыль для фирмы. Данный критерий оптимальности требует предпосылки о том, что существует некоторый средний чек, и происходит регрессия трат потребителей к среднему значению. Именно с точки зрения максимизации ожидаемого числа посетителей предполагается решение задачи оптимального расположения объектов в данной работе. Этот подход позволит учесть двустороннюю структуру рынка и выстроить понятие оптимальности с точки зрения как привлекательности для клиента, так и экономической выгоды для фирмы.

Непосредственно методами определения оптимального местоположения торгового объекта являются аналоговый подход, метод контрольного списка и регрессионный анализ. Аналоговый подход предполагает прогнозирование прибыли и оборота фирмы на основе характеристик

точек конкурентов. Для данного метода важно сравнить параметры, которые, в среднем, одинаковы для фирм одного типа, чтобы построить достаточно точный прогноз. Недостаток подхода заключается в отсутствии алгоритмичного решения и сложности в получении информации об экономическом состоянии конкурента. Для магазинов шаговой доступности, среди которых большинство точек не являются сетевыми и не публикуют собственную финансовую отчетность в открытый доступ, это ограничение является существенным.

Вторым методом определения оптимального расположения объекта коммерческой недвижимости является метод контрольного списка. Метод предполагает комплексное сравнение вариантов размещения по фиксированному набору параметров (для каждого типа объекта он является разным), значения которых оцениваются экспертизой. Обычно среди параметров встречаются такие данные, как транспортный или пешеходный трафик, потребительский потенциал территории, расположение относительно других конкурентов, социально-демографические особенности района и др. Главной проблемой данного подхода является необходимость очень высокой осведомленности эксперта. Проведение подобного сравнения неквалифицированным специалистом приведет к несосто ятельной оценке доступных вариантов.

Наконец, третий метод – регрессионный анализ. Он основан на сбере информации о схожих по типу и формату торговых точках (не таких близких, как в аналоговом подходе) и построении регрессии оборота на выбранных данных. В данном методе также используются заданные экспертом параметры, однако их проще предсказать, так как возможна оценка факторов с помощью статистического анализа и выявление наиболее значимых. Подобный метод достаточно популярен для определения местоположения продовольственных магазинов, отелей и отделений банка.

Описанные выше методы используются для конкретного местоположения объекта, оставляя без внимания потребительские предпочтения. Проблему учета потребностей клиента решает модель выявленных предпочтений, которая, во-первых, позволяет получить наиболее точ-

ный количественный результат, а во-вторых, улучшить интерпретируемость результатов, что очень хорошо для решения практических задач при выборе расположения более чем для одного объекта. Данный подход реализован в модифицированной модели оценивания привлекательности торговой недвижимости Д. Л. Хаффа и будет использован в дальнейшем при анализе существующих конкурентов.

3 Решение задачи оптимального размещения объекта на реальных данных

3.1 Постановка задачи

Главной частью работы станет исследование оптимального расположения предприятий общественного питания в районах города Москва, основываясь на данных о конкурентных предприятиях похожего функционала в окрестности, а также данных об источниках потенциального спроса в виде близлежащих жилых комплексов, станций метро и автобусных остановок.

3.2 Обзор имеющихся данных

Для использования в работе с Портала открытых данных Москвы и Портала Реформа ЖКХ были взяты данные по адресам и численности населения жилых домов, местоположения станций Московского метрополитена и среднесуточной проходимости для них, местам размещения остановок общественного транспорта и количеству маршрутов, проходящих через конкретную остановку, а также информация о различных типах объектов коммерческой недвижимости – продовольственных торговых предприятиях, отрасли общественного питания и бытовых услугах. Впоследствии задача была сужена до работы с данными по заведениям общественного питания. Ниже приведена визуализация найденных данных.

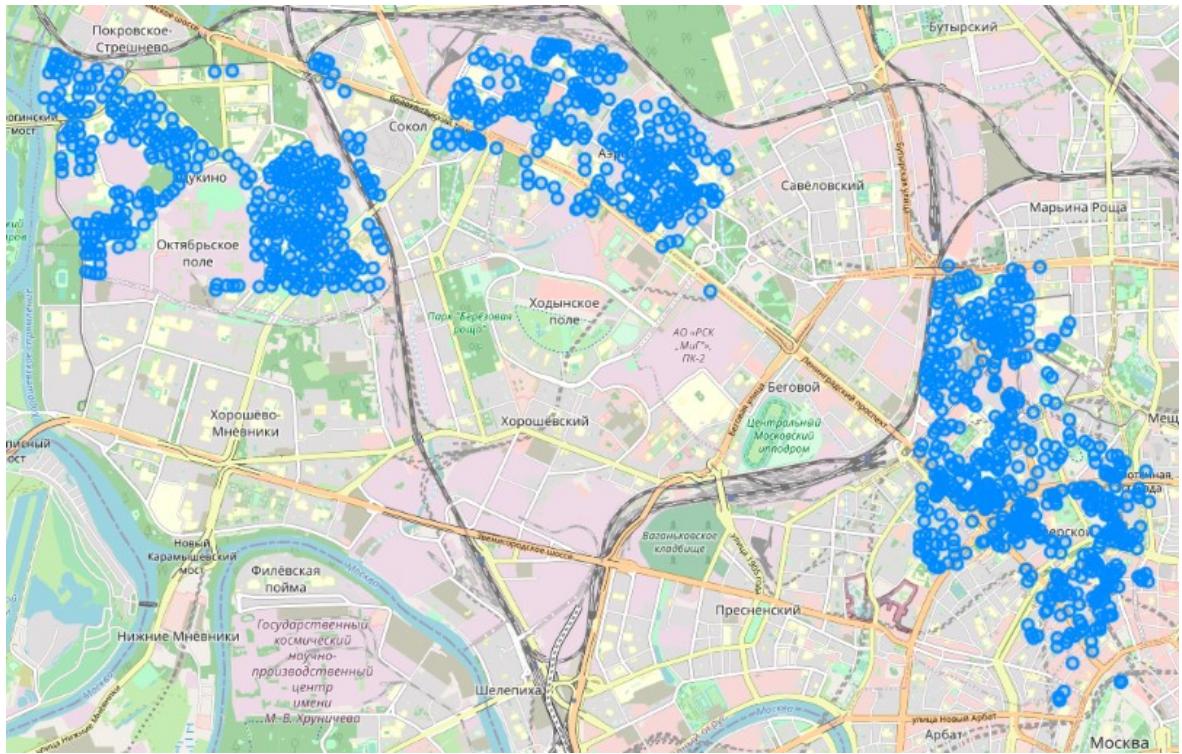


Рис. 1: Распределение расположения жилых домов города Москва

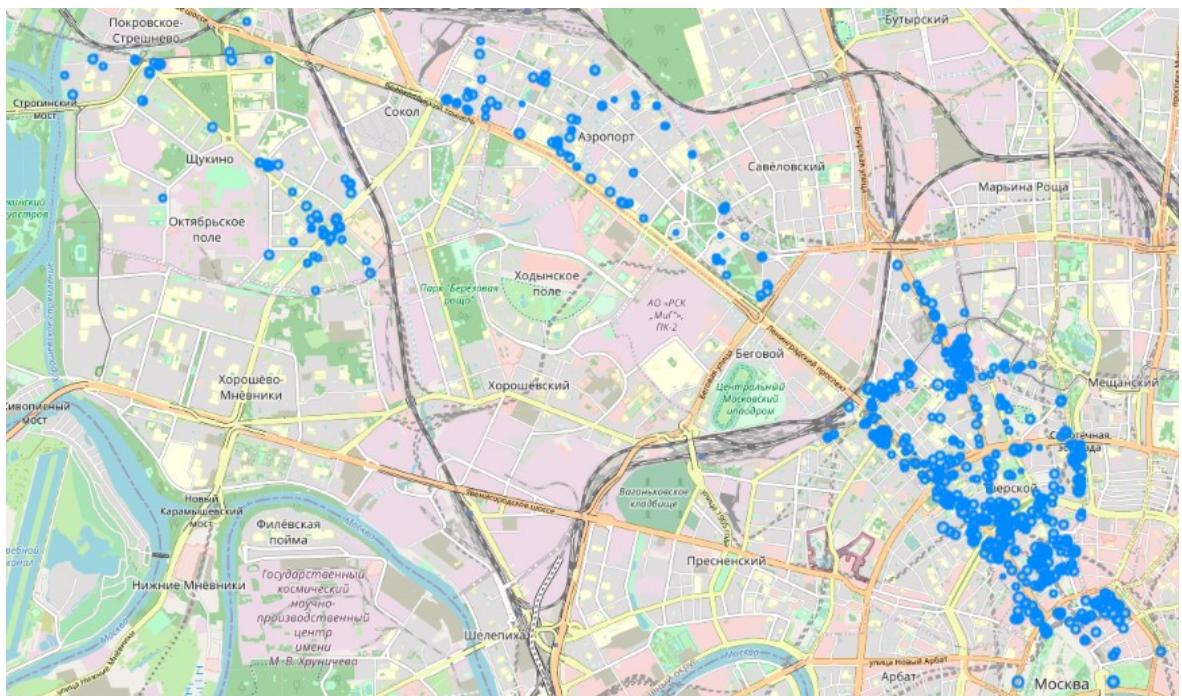


Рис. 2: Распределение расположения предприятий общественного питания города Москва

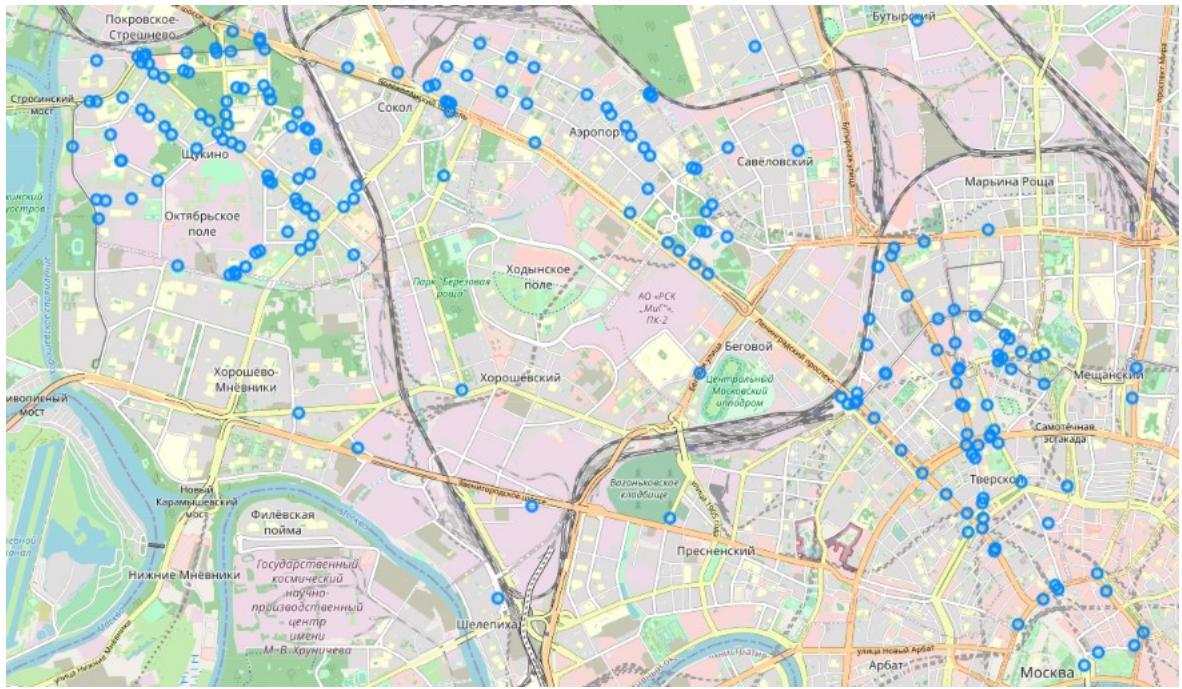


Рис. 3: Распределение расположения автобусных остановок города Москва

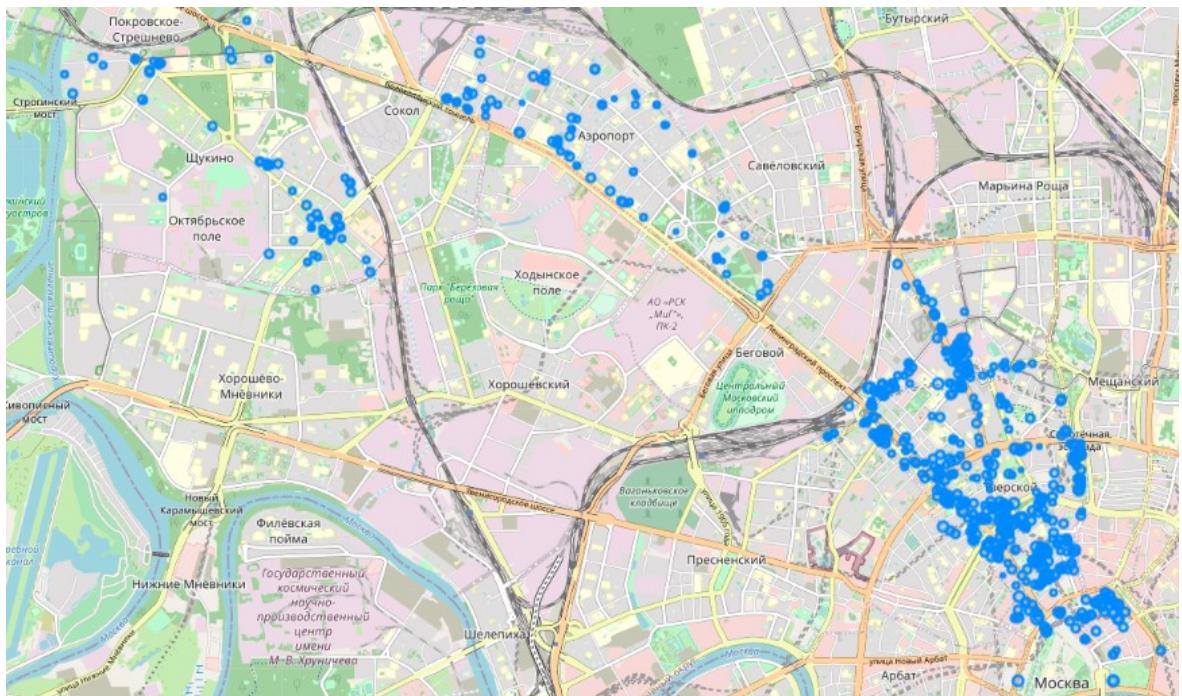


Рис. 4: Распределение расположения выходов из станций метро города Москва

Информация о жилых домах, особенно численности населения в каждом доме, позволит оценить потенциальное число клиентов предпrij-

тия в заданном радиусе. С помощью информации о расположении станций метро и остановок общественного транспорта можно отследить основные места концентрации людей и оценить объект с точки зрения транспортной доступности, что является одним из ключевых факторов привлекательности предприятия для клиента. Наконец, данные о работающих в настоящее время заведениях позволяют провести анализ конкурентов, проверить гипотезу о зависимости рейтинга предприятия от его местоположения, оценить, какую часть спроса может получить объект в данной точке, и насколько важно расстояние до ближайшего конкурента.

Основываясь на выбранных данных, необходимо понять, возможно ли, используя по большей мере географические показатели, предсказать спрос на услугу и назвать расположение оптимальным с точки зрения привлекательности для потребителя.

Поскольку решение задачи поиска оптимального расположения достаточно затруднительно для всех районов Москвы, а исследование только одного района было бы не универсальным, необходимо было выбрать несколько типичных частей города. Деление происходило с помощью транспортных колец Москвы: внутри Третьего транспортного кольца, внутри Московского центрального кольца и в пределах Московской Кольцевой Автомобильной дороги. Так, случайным образом были выбраны районы Тверской, Аэропорт и Щукино. Основные различия заключались в том, как формируется проходимость, насколько для района важна развитая транспортная сеть и какое число объектов коммерческой недвижимости расположено в каждой части города.

В связи с выбранной дифференциацией районов были выдвинуты гипотезы о наиболее важных факторах для каждой группы: для районов внутри Третьего транспортного кольца основную проходимость составляют люди, приезжающие в центр города на работу или с целью провести досуг, но в большинстве своём не его жильцы. Это связано с особенностями инфраструктуры, а также высокой ценой на недвижимость. Тогда, при поиске оптимального расположения объекта необхо-

димо учесть, что, скорее всего, маршрут каждого жителя будет содержать в себе точку станции метро или остановки общественного транспорта, и трафик в несколько раз больше населения района. Поэтому, для Тверского района важность близости к станции метро предположительно будет достаточно большой. Для района в пределах МЦК проходимость составляет некоторая доля жителей данной местности и доля тех, кто приезжает туда ежедневно на работу, учебу или за проведением досуга. Внутри подобных районов чаще всего сосредоточены крупные торговые центры, привлекающие большие потоки людей. В данном случае распределение весов, предположительно, будет более равномерным, и метро будет менее важным фактором в сравнении с центральным районом. Третий выбранный район - Щукино - расположен за МЦК, но в пределах МКАД, и его можно назвать "спальным", то есть большая часть его жителей с утра или днем выезжает за пределы района. Поскольку в районах такого типа сосредоточены только дома и базовая инфраструктура, обслуживающая их, проходимость внутри самого района формируют только его жильцы. Здесь, скорее, вес станции метро или остановок общественного транспорта все еще будет иметь весомое влияние, однако более слабое, чем для двух предыдущих случаев.

Методы агентного моделирования не будут использованы в работе напрямую, но, тем не менее, предполагается моделирование потребительского поведения. Допускаем, что у всех горожан существуют регулярные маршруты, включающие общественный транспорт, что они посещают более популярные сетевые заведения и реже малоизвестные, а также, что время, затраченное на дорогу до объекта, и пройденное расстояние, являются наиболее значимыми факторами привлекательности заведения для потребителя.

Итак, с помощью доступной информации о заселенности, транспортной доступности и конкурентной среде в выбранных районах предполагается создание новых переменных, объясняющих оптимальность расположения объекта коммерческой недвижимости.

3.3 Анализ возможностей применения существующих моделей оптимального расположения объекта

В ходе теоретического анализа проблемы были выявлены три возможных решения задачи оптимального расположения объектов коммерческой недвижимости: модель А. Вебера, модифицированная модель Д. Хаффа и модель, самостоятельно разработанная авторами на основе существующих.

Напомним, что задача Вебера позволяет найти одну точку, в которой сумма стоимости перевозки до трех точек потребления минимальна: именно в этом месте предполагается разместить производство. Несмотря на простоту, она вполне успешно может быть применена в современных условиях. Например, с ее помощью можно найти оптимальное размещение дополнительного склада крупного супермаркета. Тем не менее, данный подход нерелевантен для решения поставленной задачи. Во-первых, модель трактует оптимальное расположение через минимизацию издержек, а не максимизацию прибыли, что не соответствует выбранному авторами ходу решения. Во-вторых, задача Вебера предполагает сосредоточение спроса в одной точке, что не может считаться справедливым для объектов коммерческой недвижимости. Наконец, данная модель игнорирует существование конкуренции. В условиях рыночной экономики такое упрощение модели вызовет проблемы с ее предсказательной мощностью. Таким образом, задача Вебера не может считаться лучшим решением задачи поиска оптимального расположения объекта в силу ее несоответствия необходимым базовым предпосылкам.

С помощью модифицированной модели Хаффа можно оценить торговую привлекательность объекта, функционал в ней учитывает качество предприятия и время, которое потребитель тратит на дорогу до него:

$$A_{ij} = \frac{Q_j}{T_{ij}^{\lambda_k}},$$

где A_{ij} - привлекательность объекта j для потребителя i ;

Q_j - качество j-го объекта коммерческой недвижимости;

T_{ij} - время, затраченное покупателем i на путь до объекта j;

λ_k - параметр, который описывает влияние времени корреспонденции из k-го сектора до объекта j на временные затраты потребителя [1, с. 59]

Параметр Q является функцией многих переменных:

$$[1, 60] \quad [Q = \sum_{l=1}^L x_l, x_l = \{0,1\}, l = [1, L]]$$

Переменные, которые оценивают качество предприятия, выбираются самостоятельно в зависимости от применения модели (например, чистота помещения предприятия, размер торговой площади и проч.). Модель Хаффа применялась при поиске оптимального расположения магазина шаговой доступности в микрорайоне Октябрьский города Пермь. Поскольку исследуемая площадь была достаточно маленькой, исследователям удалось качественно оценить каждое предприятие вручную, однако в условиях районов Москвы подсчет Q автоматически невозможен из-за нехватки такого типа статистической информации, а эмпирический сбор данных достаточно затруднителен. Также в данной модели не учитывается расположение относительно конкурентов, из-за чего нельзя оценить распределение предложения и, соответственно, решить поставленную задачу поиска оптимального расположения. Таким образом, модифицированная модель Хаффа в современном мире широко применяется для поиска оптимального расположения торговых центров, государственных предприятий (больницы, школы), для оценки зеленых насаждений, однако не может быть выбрана в качестве основной при поиске оптимального расположения объекта общественного питания. Следовательно, было принято решение разработать собственную модель на основе изученного материала.

3.4 Ядерная оценка плотности

Непараметрический гистограммный метод ядерной оценки плотности, или KDE (здесь и далее KDE - Kernel Density Estimation), позволяет построить функцию плотности распределения случайной величины, но не может помочь справиться с поставленной задачей поиска оптимального расположения объекта общественного питания, так как нанесение всех домов и ресторанов на карту не подойдет для оценки плотности района в целом и потенциала размещения предприятия. Тем не менее, зная плотность распределения домов и заведений общественного питания в районах, при нанесении данных плотностей на карту эмпирически можно оценить оптимальное расположение объекта коммерческой недвижимости в определенном районе.[2]

С помощью метода KDE в Python были получены следующие функции, которые далее были нанесены на карты районов:



Рис. 5: Плотность распределения ЖК и предприятий общественного питания в районе Тверской

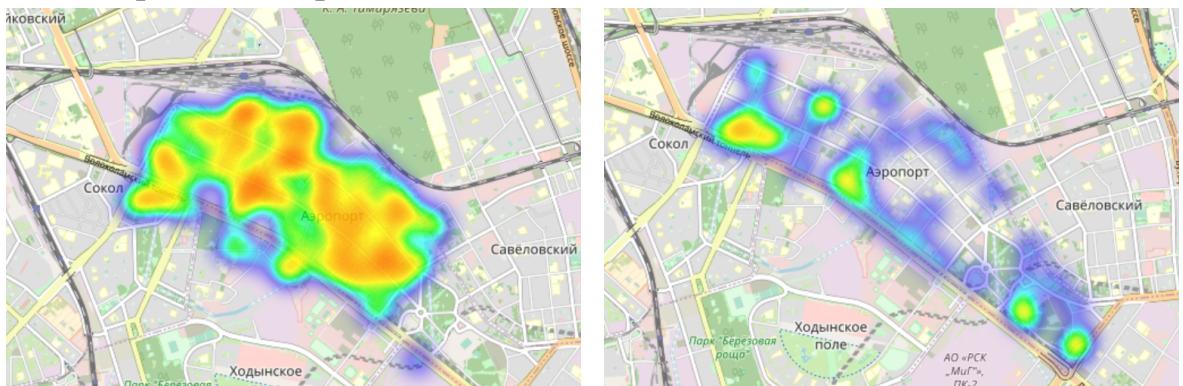


Рис. 6: Плотность распределения ЖК и предприятий общественного питания в районе Аэропорт

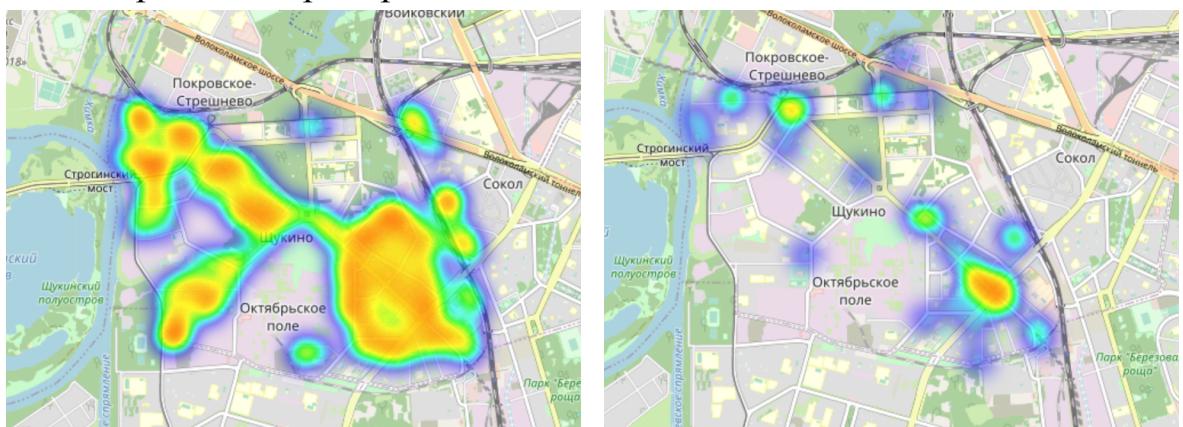


Рис. 7: Плотность распределения ЖК и предприятий общественного питания в районе Щукино

Делая выводы из полученных графиков, можно заключить, что наиболее перспективным рынком с точки зрения открытия нового предприятия общественного питания является район Аэропорт, так как он наиболее густо заселен и не сильно застроен потенциальными конкурентами. Менее перспективным, однако подходящим для поставленной цели, является рынок района Щукино, так как здесь также находится небольшое количество конкурентов, но плотность населения меньше, чем в районе Аэропорт. Из трех районов самым неподходящим для реализации нового объекта коммерческой недвижимости в сегменте общественного питания является рынок Тверского района, так как он сильно конкурентен и из-за особенностей инфраструктуры новое предприятие общественного питания может оказаться убыточным.

Однако минусом данного подхода является отсутствие чёткой карты секторов в каждом из районов, которые наиболее сильно нуждаются в открытии предприятия общественного питания в своей окрестности. Сложностью реализации вычитания одной тепловой карты из другой является разная размерность каждого из секторов, что помешало получить точный вывод для приведённого метода. Однако приведённая выше визуализация имеет место на существование при выборе оптимальной точки расположения нового предприятия в силу предоставления возможности визуально оценить в каком из секторов района может быть выгодно открыть новое предприятие по типу "кофейня у дома".

3.5 Создание переменных для анализа

Информация о жилых комплексах, остановках общественного транспорта, станциях метро и конкурентах в сегменте позволила создать переменные для построения различных моделей. Многие категориальные переменные, такие как район, тип объекта и др. были даны в изначальных наборах данных. Однако наибольший интерес представляли числовые переменные, основанные на подсчете географических расстояний, которые предстояло создать самостоятельно.

Ключевым элементом сравнения объектов коммерческой недвижимости стал подсчет евклидова расстояния от каждой из точек общественного питания до ближайших объектов, генерирующих спрос в виде одной станции метро, трех остановок общественного транспорта и десяти жилых домов. Числа, используемые при подсчете, выбраны исходя из следующих предположений:

1. Полезность от нахождения точки общественного питания около станции метро убывает с ростом числа близко расположенных станций. Иными словами, наибольшую важность для подсчета оптимальности расположения имеет факт наличия хотя бы одной станции метро по близости, так как это возможность легко добраться до точки. Однако если рядом с объектом коммерческой недвижимости находится две и более станции, это скорее добавляет вариативности для горожанина, и вносит в привлекательность объекта не такой большой вклад.
2. Посещение заведений общественного питания в спальных районах (Аэропорт и Щукино) не является повседневной потребностью горожан, и ее регулярность достаточно трудно отследить. Для обеспечения достаточного числа посетителей кафе, баров и ресторанов необходимо взять большой радиус, поэтому расстояние считалось до 10 ближайших домов.

В Тверском районе, расположенном в центре и имеющем около 110 заведений общепита на 1 км², а поэтому привлекательном для москвичей и туристов, посещение кафе и ресторанов скорее является повседневной потребностью, и регулярность сильно выше, чем в спальных районах. Однако и для Тверского района нужно выбирать больший радиус, чтобы набрать необходимое число потенциальных посетителей уже не в силу отсутствия потребности, а в силу низкой плотности населения. Вероятно, что при построении регрессии на условную меру привлекательности заведения именно в Тверском районе регрессоры расстояний до ближайших жилых домов будут иметь

наименьшую значимость.

3. Число остановок общественного транспорта, взятое в модели, зависит от количества рассматриваемых ближайших домов. Исходя из подсчета доли остановок общественного транспорта от общего числа домов, был сделан вывод, что, в среднем, на одну остановку в Москве приходится три жилых дома. Именно в этом соотношении и было выбрано 3 остановки общественного транспорта на 10 жилых домов.

После подсчета расстояния от заведения общественного питания до 5 ближайших объектов в модель было добавлено медианное значение для каждого из типов данных.

Для оценки пассажирского трафика рядом с объектом было посчитано расстояние до ближайших домов и станции метро в радиусе 1 км и остановки общественного транспорта в радиусе 400 метров. Суммарное число жителей домов, находящихся в заданном радиусе, было принято за переменную потенциальных клиентов заведения среди жителей района. В предположении, что чем больше маршрутов проходит через остановку общественного транспорта, тем выше в ней трафик, была создана переменная с информацией о числе автобусных маршрутов, проходящих через ближайшую остановку. С ее помощью предполагается оценить наземный трафик. Среднесуточная проходимость ближайшей станции метро была добавлена как третья переменная.

Важными для добавления в модель показались две бинарные переменные: является ли заведение сетевым и находится ли оно на специально отведенном месте для предприятий общественного питания – фудкорте. Если объект находился на фудкорте или по соседству с конкурентами, что справедливо в большей мере для заведений Тверского района, то, скорее, ключевым фактором в его привлекательности для клиента стал тип, предлагаемый ассортимент и то, является ли он популярным. Переменная с информацией о том, сетевое это предприятие, или нет, позволила объяснить, почему находящиеся по соседству объекты могут сильно

различаться с точки зрения привлекательности для клиента.

Таким образом, в итоговый набор данных попали следующие переменные: район; название заведения и его тип; находится заведение на фудкорте или нет; является сетевым или нет; медианное расстояние до ближайшего дома из 10 минимальных; медианное расстояние до ближайшей станции метро; медианное расстояние до ближайшей остановки из 3 минимальных; число потенциальных клиентов в радиусе километра среди жителей близлежащих домов; число маршрутов, проходящих через остановки; среднесуточная загруженность ближайшей станции метро.

В качестве целевой переменной был выбрана мера привлекательности объекта. Поскольку в открытом доступе нет информации о рентабельности заведений, их среднесуточной проходимости и других данных, которые позволили бы достаточно объективно оценить успешность предприятия, для оценки его привлекательности из Яндекс-карт было взято число отзывов о заведении и его оценка по 5-балльной шкале. Гипотеза о мультиколлинеарности этих двух переменных была отклонена: коэффициент парной корреляции составил 0.7, что приемлемо для дальнейшего анализа. Мера рассчитывалась по следующей формуле:

$$y_i = \ln(100 \times x_1)^{x_2},$$

где x_1 - число отзывов, x_2 - оценка заведения по 5-балльной шкале

Используя при подсчете логарифм и оценку заведения в степени, можно добиться адекватного отражения ключевой разницы между заведениями и разделить их на популярные и малоизвестные с помощью числовой меры. В то же время показатель привлекательности не сильно варьировался для предприятий со схожими оценками, но разным числом отзывов.

3.6 Применение моделей для предсказания успешности заведения

Итоговый набор данных состоит из трех блоков: категориальных переменных, данных по расстояниям и данных по загруженности в заданном радиусе. Для прогнозирования оптимального расположения объекта с помощью этой информации были применены классическая линейная регрессия, построение решающих деревьев и типологическая регрессия.

Первым способом решения задачи стала попытка построения линейной регрессии на признаках, генерирующих пассажиропотоки, и признаках, обуславливающих расстояние до объектов, генерирующих спрос, без дифференциации по районам. Фактор мультиколлинеарности переменных (VIF) оказался не больше 5, что свидетельствует об отклонении гипотезы о мультиколлинеарности переменных. Несмотря на то, что скорректированное значение коэффициента детерминации получилось очень хорошим для обеих регрессий ($R^2 = 0.857$ и 0.940 для регрессии по расстояниям и загруженности соответственно), при добавлении константы в модели качество сильно снижается, что говорит о проблемах в их работе и недостоверности результатов. Более того, построение регрессии без разделения по районам сильно сказывается на учете уникальности и специфики каждого типа частей города.

В качестве второго способа был выбран алгоритм решающего дерева. Несмотря на высокие показатели качества ($R^2 = 0.82$), модель оказалась сильно переобученной: дерево дробило наблюдения до 1 наблюдения в листе. Ограничения на минимальное число объектов в листе снизило R^2 до 0.36, что свидетельствует о низкой предсказательной способности.

Итоговым решением было построение типологической регрессии для трех типов районов с разделением по показателям расстояния и загруженности. Качество линейной типологической регрессии оказалось очень низким: для Тверского района R^2 составил 0.02. Поэтому в модель были добавлены категориальные переменные и применен алгоритм catboost –

один из типов градиентного бустинга, что позволило сильно улучшить качество модели. При этом качество регрессий по признакам, обуславливающим расстояние до объектов, генерирующих спрос, оказалось наилучшим для всех трех групп.

Тип регрессии	Тверской район	Район Аэропорт	Район Щукино
Регрессия по признакам, генерирующими пассажиропотоки, R^2	0.777	0.689	0.731
Регрессия по признакам, обуславливающим расстояние до объектов, генерирующих спрос, R^2	0.877	0.749	0.861

Таблица 1: Доля объяснённой дисперсии в соответствии с разными моделями категориального градиентного бустинга для каждого из районов

Можно сделать вывод, что нельзя называть оптимальным расположение заведения, где только хорошая проходимость, и необходимы другие объясняющие переменные. Таким образом, для всех трех районов наиболее важным признаком оказался тип заведения, что достаточно логично, учитывая специфику выбранной отрасли. Для Тверского района значительный эффект на целевую переменную также оказывают медианное расстояние до дома и расстояние до ближайшей станции метро. Гипотеза о том, что для районов этого типа важность близости к метро будет наибольшей по сравнению с другими, подтвердилась.

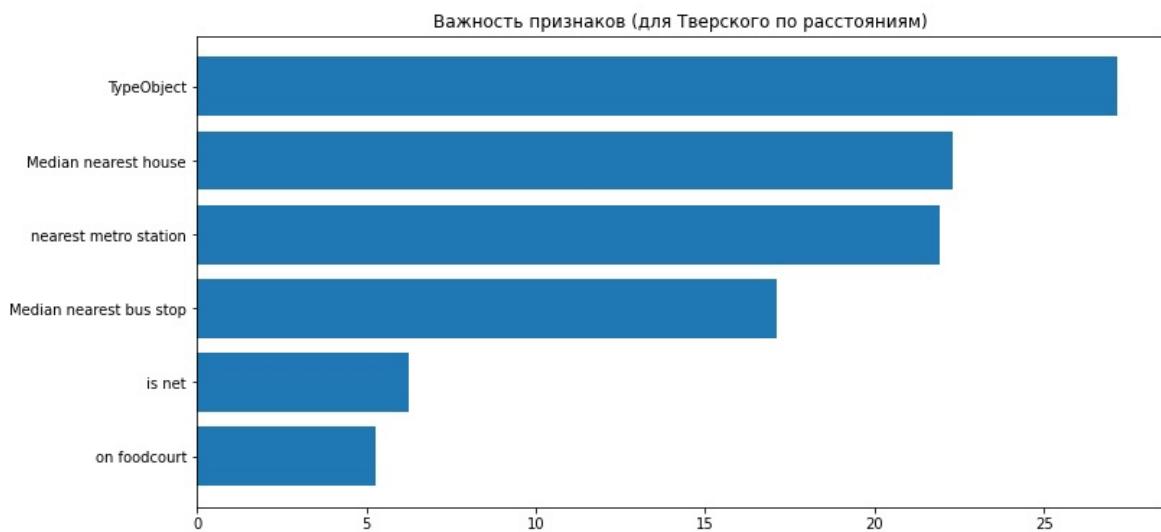


Рис. 8: Значимость объясняющих переменных для района Тверской

В Щукино второй по важности является переменная медианного расстояния до остановки общественного транспорта, что может объясняться хорошо развитой сетью наземного транспорта. Стоит отметить, что для данного района важность всех переменных, кроме типа заведения, достаточно слабая. Это может быть связано с тем, что в Щукино предприятия общественного питания расположены очень неравномерно.

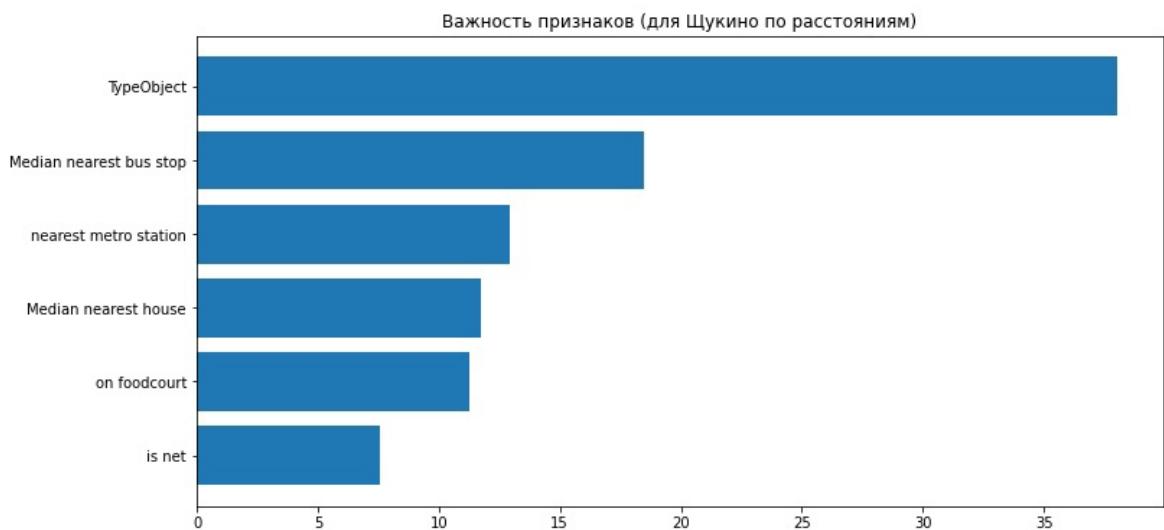


Рис. 9: Значимость объясняющих переменных для района Щукино

Наконец, для района Аэропорт ситуация противоположная: многие факторы достаточно сильно важны для описания целевой переменной. Наиболее значимыми являются расстояния до потокообразующих объектов: ближайшей остановки, дома, станции метро. Данный вывод подтверждает поставленную ранее гипотезу о том, что в районах этого типа спрос на услугу составляют как жители района, так и те, кто приехал туда на работу, учебу или провести досуг, и поэтому важно расположение предприятия вблизи этих объектов.

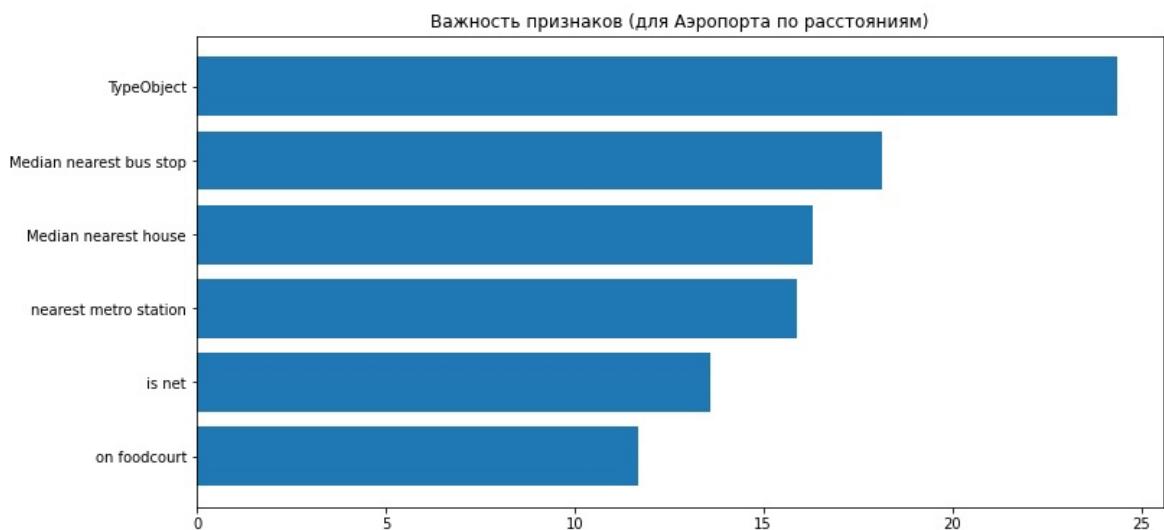


Рис. 10: Значимость объясняющих переменных для района Аэропорт

Очевиден минус приведённого подхода. Возможно оценить наилучшие заведения общественного питания, но нет никакой гарантии, что фактором успеха для них стала именно выбранная геолокация. Более того, если принимать решение о расположении объекта в ближайшей окрестности наилучшего (по значению разработанной зависимой переменной) объекта, то вполне вероятно, что новое предприятие с гораздо менее выдающейся репутацией будет работать в убыток в долгосрочной перспективе.

4 Решение задачи с учётом недостатков предыдущих методов

4.1 Функционал качества модели

Обосновав минусы всех приведённых в предыдущей главе результатов, стала очевидной необходимость разработки функционала, способного адекватно представить особенности размещения нового предприятия, учитывая необходимость отталкиваться от конкурирующих заведений и притягиваться к точкам, генерирующими максимальное количество спроса.

Приведём постановку задачи и опишем реализованный функционал.

$[x, y]$ - геолокация оптимального расположения объекта

$S = (s_1, \dots, s_m)$ - вектор геолокаций конкурентов.

$H = (h_1, \dots, h_n)$ - вектор геолокаций жилых домов.

$d(a, b) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$ - евклидово расстояние.

$W_h = (W_{h_1}, \dots, w_{h_n})$ - вектор количества заселённых квартир по жилым домам.

$W_s = (W_{s_1}, \dots, w_{s_m})$ - вектор количества положительных оценок по конкурентным предприятиям.

c - мера условной важности фактора жилых домов, позволяющая варьировать значение важность конкурирующего предложения и потенци-

ального спроса в зависимости от характеристики района.

$$F([x,y], D, S) = \left(\sum_{i=1}^m \frac{W_{s_i}}{W_s} \cdot e^{d([x,y], s_i)} \right)^{-1} + c \cdot \sum_{j=1}^n \frac{W_{h_j}}{W_h} \cdot e^{d([x,y], h_j)} \xrightarrow{x,y} \min$$

Данная функция удовлетворяет следующим весьма логичным принципам геоаналитики:

- Увеличение расстояния до конкурентных предприятий положительно влияет на потенциальную прибыль заведения, при этом влияние меры увеличения расстояния возрастает всё с большими темпами
- Чем более популярен конкурент у потребителей, тем дальше от него необходимо открывать своё предприятие
- Уменьшение расстояния до жилых домов положительно влияет на потенциальную прибыль заведения, при этом влияние меры уменьшения расстояния возрастает всё с большими темпами
- Чем больше людей в находящемся рядом доме, тем больший спрос может удовлетворить потенциально открывающееся предприятие

В результате поиска минимального значения для заданной функции двух переменных при достаточно большом количестве возможных точек расположения внутри района возможно получить тепловую карту, являющуюся аналогом градиента для заданной функции потерь.

Рассмотрим синтетический пример для малого количества точек потенциального расположения объекта:

$$H = ([0; 0], [1; 0], [0; 1], [1; 1])$$

$$W_h = (20, 50, 60, 100)$$

$$S = ([0,5; 0,5], [0,2; 0,8])$$

$$W_s = (50, 60)$$

$$F([x,y], D, S) = \frac{\frac{20}{230} \cdot e^{\sqrt{(x-0)^2 + (y-0)^2}} + \frac{50}{230} \cdot e^{\sqrt{(x-1)^2 + (y-0)^2}}}{\left(\frac{50}{110}\right) \cdot e^{\sqrt{(x-0.5)^2 + (y-0.5)^2}} + \left(\frac{60}{110}\right) \cdot e^{\sqrt{(x-0.2)^2 + (y-0.8)^2}}} + \\ + \frac{\frac{60}{230} \cdot e^{\sqrt{(x-0)^2 + (y-1)^2}} + \frac{100}{230} \cdot e^{\sqrt{(x-1)^2 + (y-1)^2}}}{\left(\frac{50}{110}\right) \cdot e^{\sqrt{(x-0.5)^2 + (y-0.5)^2}} + \left(\frac{60}{110}\right) \cdot e^{\sqrt{(x-0.2)^2 + (y-0.8)^2}}}$$

$$F([1,1], D, S) = \frac{\frac{2}{23} \cdot e^{\sqrt{2}} + \frac{11}{23} \cdot e + \frac{10}{23}}{\frac{5}{11}e^{\sqrt{0.5}} + \frac{6}{11}e^{\sqrt{0.68}}} = 0,966$$

$$F([0.5,0.5], D, S) = \frac{e^{\sqrt{0.5}}}{\frac{5}{11} + \frac{6}{11}e^{\sqrt{0.18}}} = 1,574$$

$$F([0.9,0.9], D, S) = 0,960$$

Внимательно рассмотрев веса каждого из домов и конкурентов, можно убедиться в том, что данная функция адекватно учитывает окружающую среду и старается найти такое расположение, при котором "потенциальная сфера влияния" нового предприятия будет максимальна.

Поскольку данная функция не может быть оптимизирована через известные методы оптимизации по типу градиентного спуска, было принято решение искать оптимум через симплекс-метод неполного перебора возможных точек расположения нового предприятия общественного питания.

4.2 Оптимизация заданного функционала

В результате реализации функции в Python были найдены оптимальные значения функционала в каждом из районов, полученные через симплекс-метод, реализованный на основе случайного выбора геолокаций к конкурентов по району через мультистарт с добавлением случайного шума к геолокации. Данное решение было принято на основе озвученных ранее предпосылок и на практике предоставляет устойчивую оценку опти-

мальной точки размещения объекта в соответствии со структурой описанной функции потерь.

Важно отметить, что на представленных ниже графиках в качестве весов для случайно сгенерированных точек были взяты обратные значения от функции потерь.

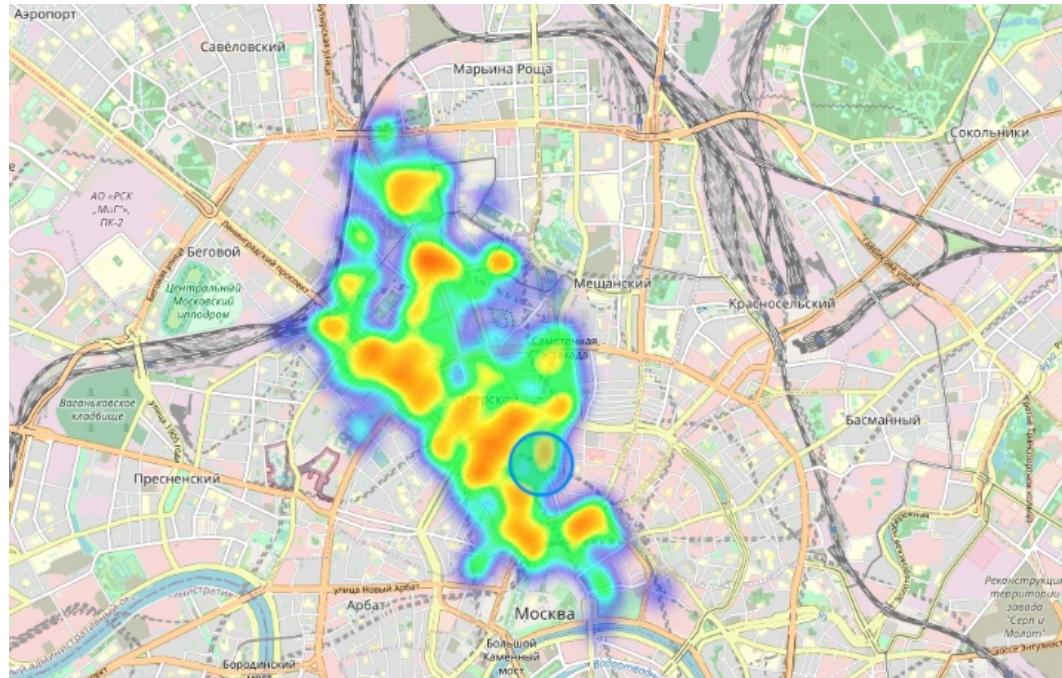


Рис. 11: Аналог градиента для заданной функции потерь в районе Тверской

Так, в районе Тверской оптимальное георасположение нового предприятия общественного питания должно располагаться в окрестности станции метро "Чеховская"; в районе Щукино оптимальная геолокация нового препдриятия общественного питания находится на юго-восточном краю района, в окрестности станции метро "Октябрьское поле"; в районе Аэропорт - в спальной части района поблизости с рижским направлением железной дороги.

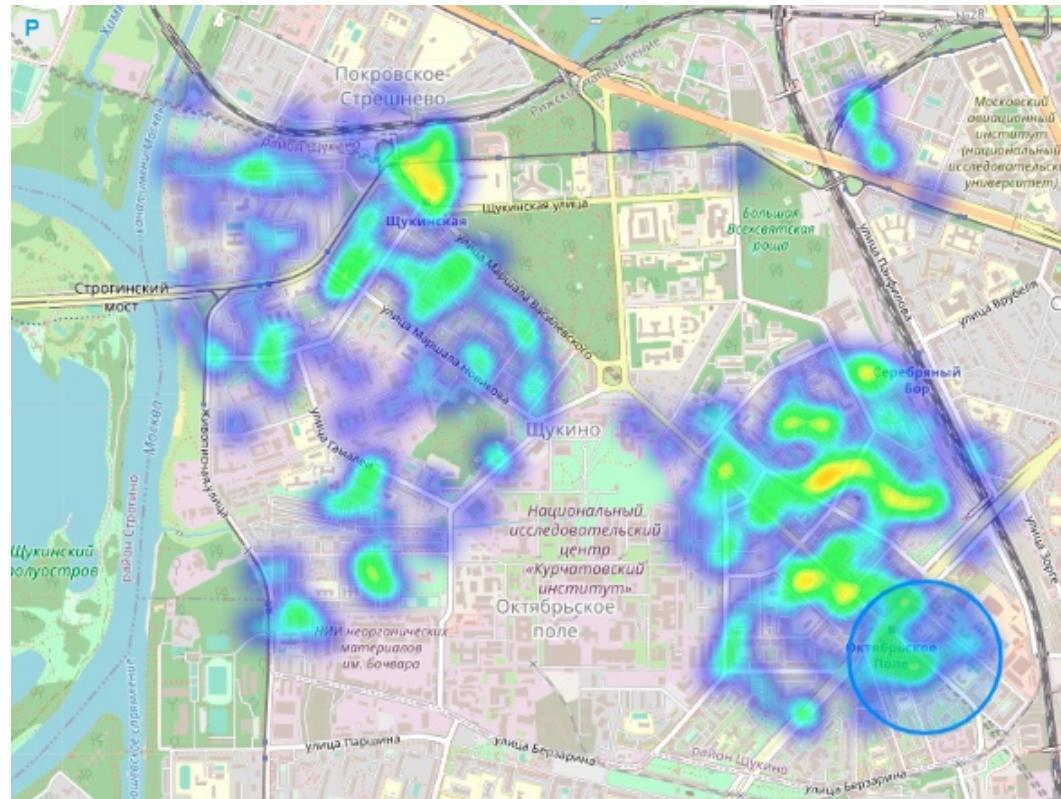


Рис. 12: Аналог градиента для заданной функции потерь в районе Щукино



Рис. 13: Аналог градиента для заданной функции потерь в районе Аэропорт

5 Заключение

В ходе работы были рассмотрены различные методы решения задачи поиска оптимального расположения объекта. В силу объема и специфики данных использование существующих моделей оказалось нереlevantным. С помощью метода ядерной оценки плотности удалось оценить распределение жилых домов и заведений общественного питания в выбранных районах и эмпирически выбрать потенциальный оптимум. Построение типологической регрессии с использованием градиентного бустинга оказалось наилучшей моделью для предсказания привлекательности предприятия и выявления наиболее значимых факторов. Альтернативным методом стало создание функционала качества, основанного на симплекс-методе, и с помощью него удалось найти оптимум расположения не только эмпирически, но и математически.

Таким образом, оптимальными для расположения оказались районы типа Аэропорт, расположенные за Третьим транспортным кольцом, но внутри Московского центрального кольца: в них достаточно высокая плотность населения и хорошо развитая транспортная система, при этом сравнительно небольшое число конкурентов, благодаря чему у предприятия есть шанс свободно войти на рынок и получить некоторую «сферу влияния».

Полученные в ходе выполнения работы результаты могут быть применены в качестве методических рекомендаций для предприятий, которые планируют открытие новых точек в городе Москва, а также полезны для учета при городском планировании размещения объектов.

Перспектива дальнейшей работы над данной задачей состоит, во-первых, в расширении числа районов: нетронутыми оказались части города, где метро находится достаточно далеко, и формируется один магистральный маршрут до этой станции, поэтому заведениям общественного питания выгодно располагаться вдоль этого маршрута. Во-вторых, хотя выбранная в работе отрасль является быстрорастущей и актуальной для анализа, было бы интересно расширить рассмотрение задачи

для других отраслей: бытовых услуг или культурно-досуговых объектов. Предположительно, распределение значимых факторов для этих предприятий будет иначе зависеть от расположения относительно жилых домов и транспортной системы. В-третьих, модель может быть модифицирована и расширена с помощью добавления новых переменных – расстояния до ближайшего торгово-развлекательного центра, например. Наконец, развить имеющиеся результаты позволит моделирование направления пассажиропотоков с помощью теории случайных процессов, а затем изучения паттернов поведения людей через методы обучения с подкреплением.

Список литературы

1. Алексеев А. О., Клейменова А. А., Спиринова В. С. Обоснование места для открытия магазина шаговой доступности (на примере микрорайона Октябрьский города Перми // Вестник ПНИПУ. Прикладная экология. Урбанистика. №1. — 2017. — С. 59—61.
2. Антонов А. В., Зюляева Н. Г., Чепурко В. А. Исследование метода ядерной оценки плотности распределения // Надежность. — 2007. — № 1. — С. 3.
3. Бакаева В. В. Как правильно выбрать место для новых магазинов торговой сети? // Российское предпринимательство №6(2). — 2011. — Т. 12. — С. 160—166.
4. Железнов С. В., Князева Л. С. Обзор методов моделирования пассажиропотоков // Конгресс Международного форума «Великие реки». — 2012. — С. 75—77.
5. Золотарев А. О. Анализ различных постановок и методов решения задачи размещения объектов // Курсовая работа, НИУ ВШЭ. — 2020. — Т. 1. — С. 3—30.
6. Имангалин А. Ф. Геомаркетинг, от простого к сложному // URL: <http://smartloc.ru/list/blog/articles/geomarketin>. — 2015.
7. Маныч И. С. Рынок кафе и ресторанов: поведение потребителей // Бизнес. Общество. Власть. — 2014. — № 18. — С. 157—168.
8. Сазанович А. А. Виды зон торгового обслуживания. — 2015.
9. Спиринова В. С. Методы управления торгово-развлекательными комплексами в условиях неопределенности на основе субъектно-ориентированного моделирования // Диссертация на соискание ученой степени кандидата экономических наук. — 2017. — С. 27—38.
10. Суцков С. Расчет доли рынка торгового объекта с помощью модели Хаффа (Huff Model)[электронный ресурс] // Site: stasgeomarketing.wordpress.com. — 2009.
11. Church R., ReVelle C. The maximal covering location problem // Papers of the regional science association. T. 32. — Springer-Verlag. 1974. — С. 101—118.
12. Hart T., Zandbergen P. Kernel density estimation and hotspot mapping // Policing: An International Journal of Police Strategies & Management. — 2014.
13. Jakob K., Pruzan P. M. The simple plant location problem: Survey and synthesis // European journal of operational research. — 1983. — Т. 12. — С. 36—81.
14. Mousavi S. M., Niaki S. T. A. Capacitated location allocation problem with stochastic location and fuzzy demand: A hybrid algorithm // Applied Mathematical Modelling. — 2013. — Т. 37, № 7. — С. 5109—5119.

15. Solving the simple plant location problem by genetic algorithm / J. Kratica [и др.] // RAIRO-Operations Research-Recherche Opérationnelle. — 2001. — Т. 35, № 1. — С. 127—142.