

АНАЛИЗ РЕНТАБЕЛЬНОСТИ ЛОКАЦИЙ ДЛЯ АВТОМОБИЛЬНЫХ ЗАПРАВОЧНЫХ СТАНЦИЙ (АЗС)

Иванов И.Д.

ООО «БСТ Диджитал», Москва, Россия
ivanzivanov@yandex.ru

Топыркин А.Д.

ООО «БСТ Диджитал», Москва, Россия
topyrkinalexei@yandex.ru

Аннотация. Разработан концептуальный подход для решения задачи размещения, позволяющий коммерческим организациям, владеющим автомобильными заправочными станциями, оценивать степень влияния географических характеристик локаций на показатели рентабельности. Проведен сбор, обработка и анализ пространственных данных, на основе которых построена модель прогнозирования финансовых результатов.

Ключевые слова: геоаналитика, рентабельность, Python, Яндекс.API, XGBoost, OLS, непараметрическая статистика.

Введение

Вопрос рентабельности бизнеса для коммерческих организаций возрастает в эпоху перехода к цифровым технологиям при высокой конкуренции и открытости информации. Для выявления наилучших мест размещения в прошлом было достаточно простого подсчета потока людей и/или автомобилей. Сегодняшние реалии не позволяют экстенсивно расширять сеть и увеличивать обороты за счет открытия новых точек, так как в таких локациях возрастает количество потенциальных конкурентов или же целевая аудитория конкретных районов может в целом не попадать под потребности конкретной фирмы. Именно поэтому в современном мире на первое место и выходят исследования, за счет которых в дальнейшем лицо принимающее решение и осуществляет свой выбор по открытию, например, новой заправочной станции, магазина, отделения, офиса и т.п.

Исследования такого рода получили название «геоаналитика». Геоаналитика – это лишь один из многих терминов, используемых для принятия решений по размещению объекта на основе анализа данных о его местоположении. Синонимами слова «геоаналитика» выступают такие термины как: разведка местоположения, геолокация, геопространственный анализ, географический анализ и многие другие. Все эти термины описывают одну и ту же основную идею: применение географии к конкретным бизнес-процессам для получения аналитических сведений.

Применение геоаналитики позволяет решать целый ряд задач, например, таких как: визуализация данных с географической точки зрения; интерактивное картографирование; географический анализ и поиск; кластеризация и построение тепловых карт; тематическое картографирование; проектирование и управление территориями; оптимизация маршрутов; прогнозирование.

Важность геоаналитики заключается в том, что она позволяет компаниям максимизировать показатели в зависимости от размещения новых точек. В частности, когда речь идет о данных системы управления взаимоотношениями с клиентами (CRM), отчетов и таблиц, то этого недостаточно для того, чтобы на основе них принимать решения. Геоаналитика использует данные, которые есть у организации и отображает их на карте, упрощая понимание и анализ полезной информации.

1. Подходы к оценке рентабельности и решению задач размещения

Анализом рентабельности занимались большое количество зарубежных и российских экономистов, но каждый из них трактовал понятие рентабельности по-разному в отдельных аспектах. Многие экономисты считают, что рентабельность и прибыльности понятия идентичные, другие экономисты соотносят рентабельность с понятием доходности.

Юджин Бригхэм и Майкл Эрхардт в своей книге «Финансовый менеджмент» дают следующее определение для понятия «рентабельность». «Рентабельность бизнеса – это конечный результат действия целого ряда управленческих решений и структуры компании. Рентабельность продаж вычисляется с помощью деления чистой прибыли на величину продаж и показывает долю прибыли в каждом долларе продаж» [1].

В своей работе «Методика финансового анализа деятельности коммерческих организаций» Шермет Анатолий Данилович, определил рентабельность, как – основной показатель эффективности

хозяйственной деятельности. Показатели рентабельности рассчитываются как относительные показатели финансовых результатов, полученных предприятием за отчетный период [2].

Щедрая Ю. Л. и Шинин В. В. в статье «Рентабельность в системе хозяйственного механизма», писали следующее об рентабельности. Это норма прибыли, которая проявляется в соотношении прибыли к себестоимости продукции или производственных фондов [3].

Савицкая Глафира Викентьевна, считала, что «рентабельность – это степень доходности, выгодности, прибыльности бизнеса [4].

Вячеслав Никитич Масленников писал, что рентабельности фирмы определяет отношение всей массы полученного чистого дохода к стоимости основных производственных фондов и оборотных средств или к полной себестоимости реализованной продукции [5].

Большая российская энциклопедия содержит в себе следующее определение: «РЕНТАБЕЛЬНОСТЬ (от нем. *rentabel* – доходный, выгодный, прибыльный), показатель деятельности предприятия по степени использования затрачиваемых ресурсов (природных, материальных, трудовых, финансовых и т. д.)» [6].

В экономическом словаре Александра Николаевича Азрилияна для понятия «рентабельность» приводится следующее определение – это один из основных стоимостных качественных показателей эффективности производства на предприятии, объединении, отрасли в целом, характеризующий уровень отдачи затрат и степень использования средств в процессе производства и реализации продукции [7].

Налоговый кодекс Российской Федерации определяет рентабельность, как «отношение прибыли от продаж к выручке от продаж, исчисленной без учета акцизов и налога на добавленную стоимость» [8].

В работе будем подразумевать под рентабельностью ту трактовку, которую сформировали Юджин Бригхэм и Майкл Эрхардт, так как она наиболее точно и полно обобщает все остальные определения.

Измерить рентабельность можно с помощью целого ряда различных показателей: рентабельности собственного капитала, валовой рентабельности, рентабельности продаж, рентабельности активов, рентабельности основных средств, рентабельности персонала, рентабельности инвестиций, рентабельности чистых активов, рентабельности производства, рентабельности наценки и, конечно, рентабельности себестоимости.

Для того, чтобы определиться, какая рентабельность лучше всего подойдет нам обратимся к самой задаче исследования. Целью работы является оценка локаций для выявления мест, которые будут иметь наибольшую рентабельность. В большинстве определений представленных выше, можно было заметить, что большинство ученых сходятся к тому, что основными показателями для расчета рентабельности являются доходы и расходы. Так как в рамках исследования будет решаться задача размещения, которая подразумевает под собой прогнозирование экономических величин на основе различных внешних факторов, описывающих конкретную местность, то автоматически появляется одно ограничение. Расходы организации – это показатель комплексный, который зачастую зависит не только от внешних переменных, но и от огромного количества внутренних факторов, поэтому прогнозирование рентабельности для определенной локации на основе лишь внешних данных невозможно. Однако, можно оценить другие составляющие рентабельности, а именно доход или объем продаж определенной точки в каком-то конкретном месте. Бесспорно выручка АЗС или пролив также зависит и от внутренних факторов (площади заправки, количества колонок и касс, доступности до транспортных артерий). Но их гораздо проще оценить в моменте времени (фактически, все эти величины становятся нам известны уже в тот момент, когда мы только начинаем планировать постройку), чем, например, затраты, которые понесёт конкретная станция от потенциального изменения стоимости нефти или бензина. Именно поэтому в рамках исследования будет прогнозироваться 2 величины – это объем пролива и выручка магазина АЗС в определенной локации, которые в дальнейшем могут быть использованы для расчета показателей рентабельности. Таким образом, наиболее подходящим для этой задачи типом рентабельности является рентабельность продаж. Юджин Бригхэм и Майкл Эрхардт описали рентабельность продаж, как отношение чистой прибыли к выручке:

$$P = \frac{I-C}{I}, \quad (1)$$

где I – выручка фирмы в результате реализации произведенной ею продукции, а C – затраты фирмы, понесенные в процессе реализации продукции [1].

Высокая значимость количественной оценки взаимосвязи между изменениями географического расположения объекта и изменениями показателей рентабельности обусловлена тем, что выявленные закономерности лежат в основе инвестиционного анализа организации.

По мнению ряда ученых использование географических данных для решения задач размещения является, в достаточной степени эффективным методом (Вебер и Чепмен, 2011 г. [9]; Гринева и Топыркин, 2023 [10, 11]; Харрис, Слейт и Уэббер, 2005 [12]). В иностранной литературе даже выработался отдельный термин «Location Intelligence», который можно интерпретировать, как «аналитика, основанная на местоположении». Появился этот тип аналитики в результате развития такого молодого направления в науке, как геоинформатика (geoinformatics).

Патрик Вебер и Дэйв Чепмен [9] из университетского колледжа Лондона в своей работе о инновационном подходе к расположению бизнеса приводят в качестве примера столицу Великобритании, город Лондон, и показывают, как разбиение города на районы может помочь в классификации и выявлении оптимальных мест под размещение различных коммерческих организаций.

Наталья Владимировна Гринева и Топыркин Алексей Дмитриевич из Финансового университета и Российской академии народного хозяйства и государственной службы в своей первой статье описали процесс сбора (обработки и хранения), анализа и отображения географических данных для решения задачи размещения коммерческих организаций [10].

В своей второй работе авторы показали процесс моделирования экономической эффективности локаций в сфере торговли на основе результатов и данных, полученных из первой статьи [11].

Ричард Харрис, Питер Слейт и Ричард Уэббер из Бристольского университета описали полный алгоритм действий, реализуемых в процессе решения задачи размещения. От момента сбора географических данных и отображения их в ГИС до анализа и построения алгоритма классификации районов [12].

2. Гипотеза и методика исследования

Основная идея статьи – разработка моделей оценки рентабельности локаций на основе пространственных и внутренних данных для автомобильных заправочных станций.

Объектом исследования является механизм формирования оценки рентабельности при выборе локации для размещения точек АЗС. Предметом исследования являются эконометрические модели оценки рентабельности локаций при решении задачи размещения.

Целью работы является моделирование оценки влияния пространственных и внутренних факторов на показатели рентабельности различных локаций в рамках деятельности одной организации.

Методика исследования включает в себя следующие этапы:

- сбор данных:
 - обзор источников данных, показателей их глубины, доступа;
 - определение методов сбора и хранения данных.
- анализ и обработка данных для моделирования:
 - очистка и заполнение пропусков в данных;
 - препроцессинг данных для моделирования;
 - анализ зависимостей между объясняемой и объясняющими переменными;
- фичер-инжиниринг:
 - генерация гипотез и выявление потенциальных зависимостей;
 - создание дополнительных факторов.
- определение правил моделирования:
 - выбор метрик оценки точности моделей;
 - выбор методов анализа качества модели;
 - формирование плана построения моделей.
- разработка моделей оценки рентабельности локаций на основе пространственных и внутренних данных для автомобильных заправочных станций:
 - оценка точности моделей;
 - оценка качества моделей;
 - формирование прогнозов;
 - визуализация результатов.

3. Исходные данные

В основу исследования легли данные о 163 российских автомобильных заправочных станциях. Целевые переменные представляют собой объем продаж в литрах и суммарную выручку магазина в рублях каждой из АЗС за год. В работе будут использоваться их нормированные от 0 до 1 версии, которые затем были прологарифмированы. Подробнее на диаграммы рассеяния и гистограммы распределения целевых переменных можно посмотреть на рисунках 1-4. На рисунках 1-2 представлены графики объема продаж в литрах за год для каждой АЗС.

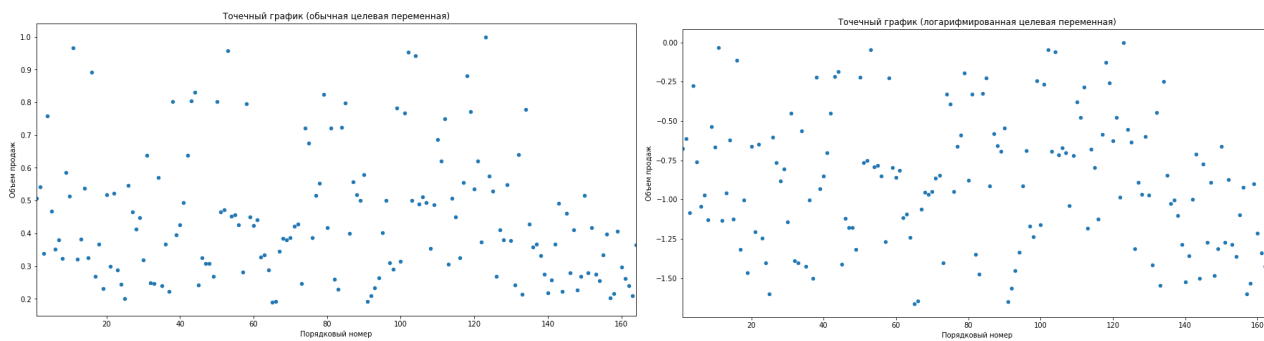


Рис. 1. Точечные графики объема продаж АЗС (обычный и логарифмированный вид)

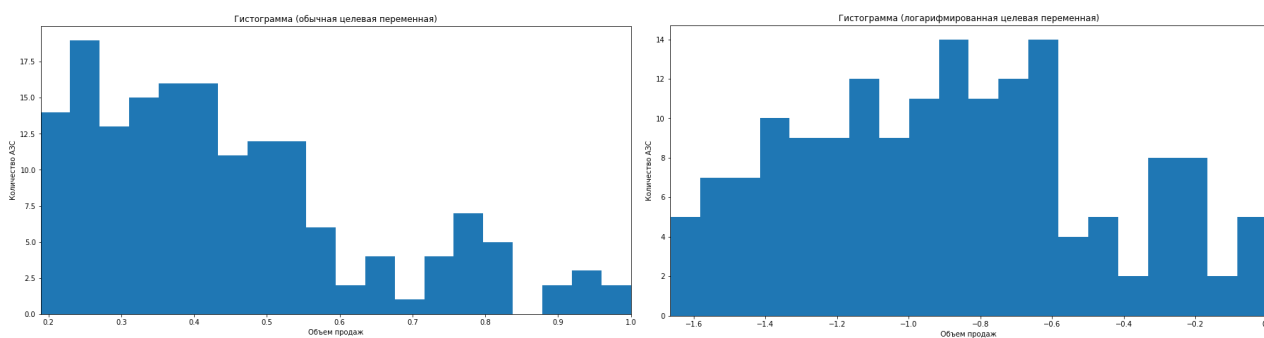


Рис. 2. Гистограммы объема продаж АЗС (обычный и логарифмированный вид)

Рисунки 3-4 демонстрируют те же графики, но для выручки магазина в рублях за год для каждой АЗС.

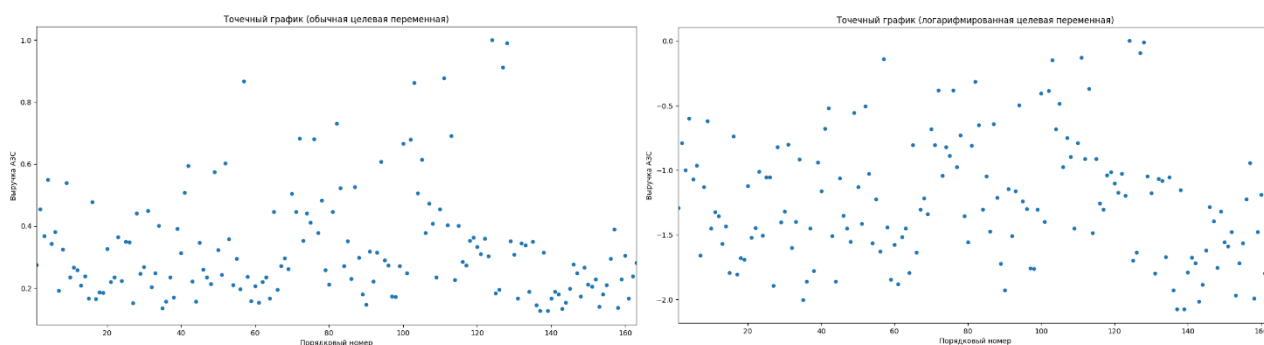


Рис. 3. Точечные графики выручки магазинов АЗС (обычный и логарифмированный вид)

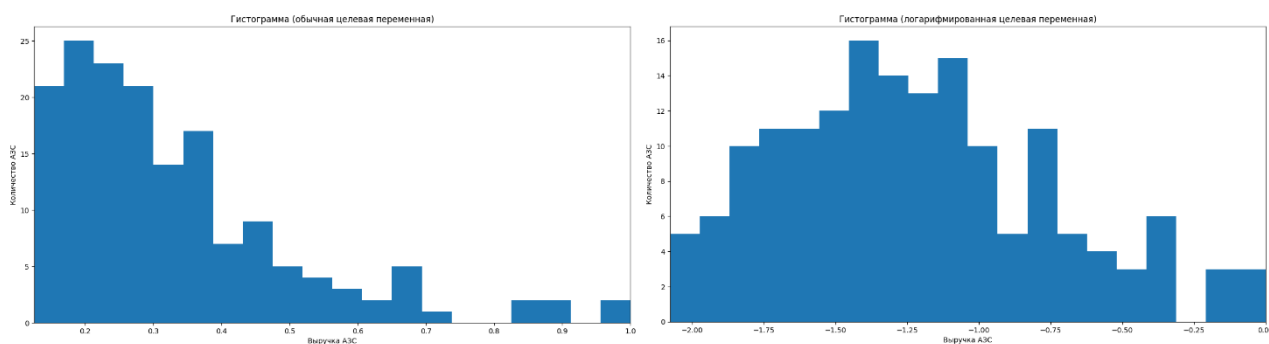


Рис. 4. Гистограммы выручки магазинов АЗС (обычный и логарифмированный вид)

Географические данные, которые используются в процессе моделирования разбиты на 5 основных групп:

- мобильные данные;
- данные по трафику;
- данные о недвижимости;
- данные о бизнес-среде;
- данные о постройках.

Всего суммарно эти пять категорий насчитывают 907 факторов, если отбросить те переменные, которые для каждой из АЗС обладают либо нулевыми значениями, либо пропусками, то в итоге остается 851 переменная.

Таким образом, в итоговом датасете насчитывается 851 пространственный фактор, они делятся на следующие категории: мобильные данные (работающие, жители, транзит. Они в свою очередь детализируются по полу, возрасту и доходу); трафик (пешеходный и автомобильный. Детализируется по работающим, проживающим и транзиту); недвижимость (стоимость 1 квадратного метра недвижимости и количество квартир); данные по бизнес-единицам (количество организаций в определенной сфере); постройки (количество объектов и их площадь). Географические факторы используются для получения сведений о конкурентах, бизнес-структуре районов, о количестве жителей и работающих, об интенсивности движения пешеходов и автомобилистов в различное время суток, о количественных и качественных характеристиках людей на определенной территории, о стоимости недвижимости, что является своего рода маркером «успешности» жизни в локации. Краткая характеристика данных представлена в таблице 1.

Таблица 1. Характеристики пространственных данных

Тип данных	Источник	Исходный формат	Конечный формат	Инструменты и методы
Мобильные данные	Сотовые операторы	Полигон	Полигон	PostgreSQL
Трафик	Агрегаторы информации по мобильным приложениям	Гексагон (10)	Гексагон (10)	PostgreSQL
Недвижимость	Сайт «ЦИАН»	Точка	Гексагон (10)	requests, pandas, geopandas, h3, PostgreSQL
Бизнес	API Яндекс.Карт	Точка	Гексагон (10)	requests, Яндекс.API поиск по орг., JSON, pandas, geopandas, (Selenium, Playwright, Cypress, Puppeteer), h3, PostgreSQL
Постройки	Сайт «OpenStreetMap»	Полигон	Гексагон (10)	OSMnx, pandas, geopandas, h3, PostgreSQL

Источник: составлено авторами

4. Анализ данных

Для определенных в работе двух целевых переменных исходное пространство признаков было разделено на 6 категорий по типу данных: данные по трафику, данные о недвижимости, данные о постройках, данные о бизнесе, другое (внутренняя информация, тестируемые гипотезы), количество касс. Это позволяет увидеть основные зависимости между объемом продаж/выручкой магазинов АЗС

и конкретными факторами, а также в рамках этого подхода становится более простым выявление значимости переменных. Первичный корреляционный анализ для 851 фактора производился посегментно. Ниже будут представлены его результаты в формате графиков и таблиц.

По объему продаж анализ корреляции показал следующее. Итоговые данные сведены в таблицу 2.

Таблица 2. Сегментарный анализ корреляции для объема продаж АЗС за год

Категория	Кол-во факторов	MIN корреляция	MAX корреляция	MEAN корреляция	VAR корреляция	STD корреляция
Количество касс	1	0,666	0,666	0,666	-	-
Другое	12	-0,453	0,182	-0,022	0,037	0,192
Данные по трафику	264	0,121	0,549	0,324	0,007	0,082
Данные о бизнесе	544	-0,315	0,366	0,137	0,005	0,070
Данные о недвижимости	20	0,251	0,512	0,339	0,004	0,060
Данные о постройках	10	-0,228	-0,088	-0,156	0,002	0,043

Источник: составлено авторами

Аналогично таблица 3 с сегментарным корреляционным анализом для выручки магазинов АЗС.

Таблица 3. Сегментарный анализ корреляции для выручки магазина АЗС за год

Категория	Кол-во факторов	MIN корреляция	MAX корреляция	MEAN корреляция	VAR корреляция	STD корреляция
Количество касс	1	0,659	0,659	0,659	-	-
Другое	12	-0,426	0,347	0,017	0,046	0,214
Данные по трафику	264	-0,146	0,662	0,093	0,010	0,101
Данные о бизнесе	544	-0,567	0,405	-0,034	0,006	0,077
Данные о недвижимости	20	0,238	0,454	0,291	0,003	0,052
Данные о постройках	10	-0,208	-0,065	-0,162	0,003	0,053

Источник: составлено авторами

1) Относительно сегмента «Другое» и целевой переменной в формате объема продаж минимальное и максимальное значение коэффициентов корреляции составили соответственно -0,453 и 0,182 для следующих переменных «Расположение в центральной части страны» и «Мистлово». Для выручки магазинов наименьшее и наибольшее значение коэффициентов корреляции были равны -0,426 и 0,347 у переменных «Другое» и «КАД». Всего факторов в этом сегменте: 12. Средняя корреляция по всем переменным составила -0,022 и 0,017 соответственно.

2) Для сегмента «Данные по трафику» и целевой переменной в формате объема продаж минимальное и максимальное значение коэффициентов корреляции составили соответственно 0,121 и 0,549 для следующих переменных «Автомобильный трафик рабочих в радиусе 500 метров» и «Автомобильный трафик в радиусе 150 метров». Для выручки магазинов наименьшее и наибольшее значение коэффициентов корреляции были равны -0,146 и 0,662 у переменных «Пешеходный трафик рабочих в радиусе 150 метров» и «Автомобильный трафик в радиусе 150 метров». Всего факторов в этом сегменте: 264. Средняя корреляция по всем переменным составила 0,324 и 0,093 соответственно.

3) Для сегмента «Данные о бизнесе» и целевой переменной в формате объема продаж минимальное и максимальное значение коэффициентов корреляции составили соответственно -0,315 и 0,366 для следующих переменных «Количество магазинов в радиусе 500 метров» и «Количество банков в радиусе 500 метров». Для выручки магазинов наименьшее и наибольшее значение коэффициентов корреляции были равны -0,567 и 0,405 у переменных «Количество организаций в радиусе 300 метров» и «Интенсивность ритейла в радиусе 150 метров». Всего факторов в этом сегменте: 544. Средняя корреляция по всем переменным составила 0,137 и -0,034 соответственно.

4) Относительно сегмента «Данные о недвижимости» и целевой переменной в формате объема продаж минимальное и максимальное значение коэффициентов корреляции составили соответственно 0,251 и 0,512 для следующих переменных «Стоимость 1 кв. м. недвижимости в радиусе 1000 метров» и «Количество квартир в радиусе 5-ти минутной пешей доступности». Для выручки магазинов наименьшее и наибольшее значение коэффициентов корреляции были равны 0,238 и 0,454 у переменных «Количество квартир в радиусе 150 метров» и «Стоимость 1 кв. м. недвижимости в радиусе 1500 метров». Всего факторов в этом сегменте: 20. Средняя корреляция по всем переменным составила 0,339 и 0,291 соответственно.

5) По сегменту «Данные о постройках» и целевой переменной в формате объема продаж минимальное и максимальное значение коэффициентов корреляции составили соответственно -0,228 и -0,088 для следующих переменных «Количество частных домов в радиусе 2000 метров» и «Количество частных домов в радиусе 75 метров». Для выручки магазинов наименьшее и наибольшее значение коэффициентов корреляции были равны -0,208 и -0,065 у переменных «Количество частных домов в радиусе 2000 метров» и «Количество частных домов в радиусе 75 метров». Всего факторов в этом сегменте: 10. Средняя корреляция по всем переменным составила -0,156 и -0,162 соответственно.

Финальные набор факторов для модели на основе объема продаж можно видеть в таблице 4.

Таблица 4. Набор факторов в модели с целевой переменной объем продаж АЗС за год

Тип данных	Источник	Исходный формат
Количество касс	Количество касс	0,666
Автомобильный трафик в радиусе 150 метров	Данные по трафику	0,549
Общий трафик в радиусе 300 метров	Данные по трафику	0,523
Количество квартир в радиусе 5-ти минутной пешей доступности	Данные о недвижимости	0,512
Пешеходный трафик в радиусе 300 метров	Данные по трафику	0,470
Количество банков в радиусе 500 метров	Данные о бизнесе	0,366
Количество магазинов в радиусе 500 метров	Данные о бизнесе	-0,315
Расположение в центральной части страны	Другое	-0,453

Источник: составлено авторами

Аналогичный набор факторов, но только для выручки магазинов, находится ниже в таблице 5.

Таблица 5. Набор факторов в модели с целевой переменной выручка магазина АЗС за год

Тип данных	Источник	Исходный формат
Количество касс	Количество касс	0,659
Автомобильный трафик в радиусе 150 метров	Данные по трафику	0,662
Стоимость 1 кв. м. недвижимости в радиусе 1500 метров	Данные о недвижимости	0,454
Общий трафик в радиусе 100 метров	Данные по трафику	0,423
Интенсивность ритейла в радиусе 150 метров	Данные о бизнесе	0,405
Количество банков в радиусе 500 метров	Данные о бизнесе	-0,331
Количество организаций в радиусе 150 метров	Данные о бизнесе	-0,391
Количество организаций в радиусе 300 метров	Данные о бизнесе	-0,567

Источник: составлено авторами

5. Моделирование

Установим правила построения моделей: для набора признаков, собранных ранее географических, социально-демографических и внутренних факторов, оставшихся после корреляционного анализа, поэтапно генерировались комбинации различных размеров; от одной переменной до количества переменных, после которого пропадал какой-либо значительный прирост метрик (процедура Forward Selection). Далее каждая из комбинаций оценивалась в режиме кросс-валидации с разбиением на 10 фрагментов. Метрикой оценки точности модели выступило MAPE (средняя относительная ошибка аппроксимации). Из сформированных комбинаций факторов по значению MAPE выбиралась комбинация с наилучшим показателем точности и количеством признаков не более 10. Последующее изменение комбинаций факторов основывалось на исключении мультиколлинеарных признаков (среди совокупности пар факторов удалялись те, которые характеризовались значением коэффициента корреляции больше 0,8). Помимо этого производилась фильтрация признаков, которые

были бы схожи на практике, например, дневной пешеходный трафик в радиусе 300 и 500 метров. Исключение факторов реализовывалось в соответствии с тем, чтобы рассчитанное на кросс-валидации с 10 перемешенными фолдами значение MAPE (средней относительной ошибки аппроксимации), не увеличивалось.

Основным алгоритмом в процессе построения моделей стал градиентный бустинг в реализации LightGBM, так как именно он показал наилучшие результаты относительно точности и скорости обучения.

В заключении для наилучшей по показателю точности комбинации реализовывалась оптимизация гиперпараметров. Целочисленный поиск для следующих аргументов модели: максимальная глубина дерева (max_depth), минимальное количество наблюдений в листе (min_child_samples), количество деревьев (n_estimators), количество узлов в дереве (num_leaves). Непрерывный поиск использовался для параметров: регуляризации lambda (reg_lambda) и alpha (reg_alpha), доли обучающей выборки, используемой для построения деревьев (subsample), шага градиентного бустинга (learning_rate). В качестве функции потерь при оптимизации гиперпараметров использовалась средняя относительная ошибка аппроксимации. Основным оптимизационным алгоритмом при подборе значений параметров являлся Tree-structured Parzen Estimator (TPE) из библиотеки hyperopt.

Для каждой из целевых переменных было отстроено по 7623 модели градиентного бустинга. Процесс изменения метрик точности в рамках добавления факторов в наилучшие комбинации для каждой из регрессант представлены в таблицах 6-7.

Таблица 6. Изменение MAPE для лучшей комбинации в процессе добавления факторов в модель с целевой переменной объем продаж АЗС за год

№	Фактор	MAPE (insample), %	MAPE (cross-validation), %
1	Количество касс	33,527	33,84
2	Общий трафик в радиусе 300 метров	27,671	31,468
3	Количество банков в радиусе 500 метров	25,067	30,959
4	Пешеходный трафик в радиусе 300 метров	20,14	30,854
5	Автомобильный трафик в радиусе 150 метров	16,204	27,349
6	Расположение в центральной части страны	15,129	26,102
7	Количество магазинов в радиусе 500 метров	13,851	25,483
8	Количество квартир в радиусе 5-ти минутной пешей доступности	13,512	24,466

Источник: составлено авторами

Таблица 7. Изменение MAPE для лучшей комбинации в процессе добавления факторов в модель с целевой переменной выручка магазина АЗС за год

№	Фактор	MAPE (insample), %	MAPE (cross-validation), %
1	Количество касс	32,473	33,108
2	Стоимость 1 кв. м. недвижимости в радиусе 1500 метров	21,906	30,427
3	Автомобильный трафик в радиусе 150 метров	10,331	27,605
4	Интенсивность ритейла в радиусе 150 метров	7,577	25,326
5	Количество организаций в радиусе 300 метров	6,706	24,263
6	Общий трафик в радиусе 100 метров	5,716	23,757
7	Количество банков в радиусе 500 метров	5,721	23,096
8	Количество организаций в радиусе 150 метров	5,574	22,878

Источник: составлено авторами

На рисунках 5-6 представлены гистограммы со степенью важности факторов внутри каждой из моделей, чем выше переменная расположена на графике, тем сильнее она оказывает влияние на формирование финального прогноза модели.

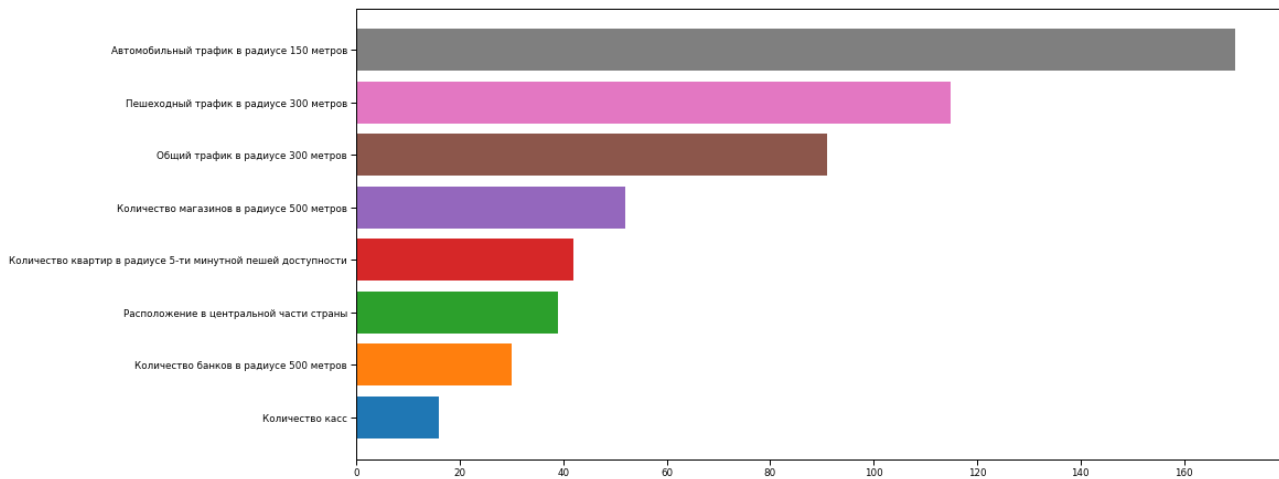


Рис. 5. Feature importance модели с целевой переменной объем продаж АЗС за год

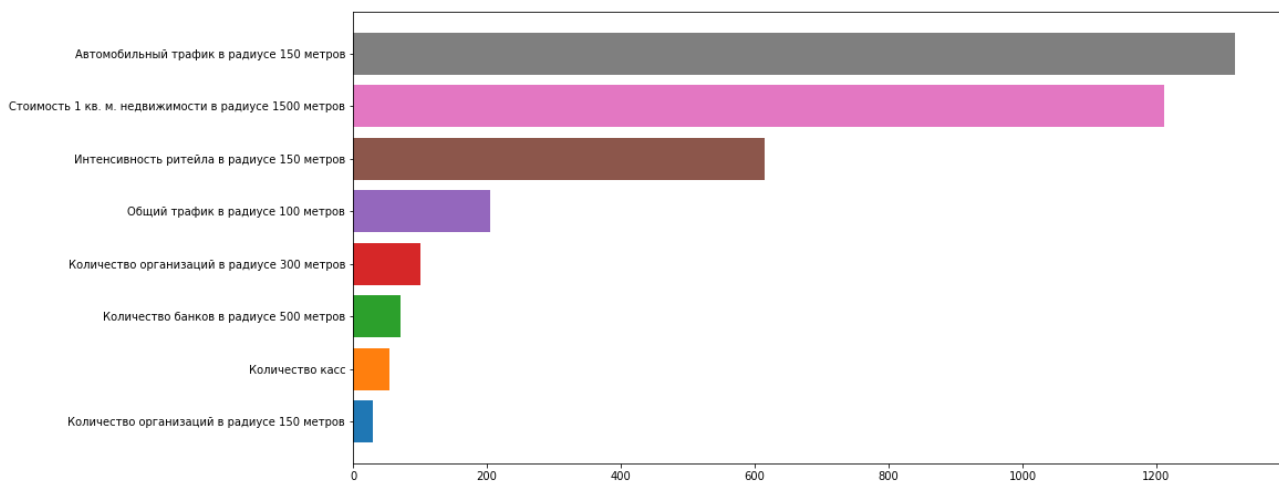


Рис. 6. Feature importance модели с целевой переменной выручка магазина АЗС за год

Для интерпретации результатов градиентного бустинга воспользуемся визуализацией вклада факторов на основе SHAP-values. В отличие от графиков на рисунках 5-6, где показано лишь модульное влияние факторов, SHAP-values позволяет узнать и о характере влияния.

На рисунках 7-8 представлены графики SHAP-values для каждой из моделей. Чем выше переменная расположена на рисунке, тем сильнее она влияет на предсказания. Если красные точки располагаются с правой стороны графика, а синие с левой, то фактор характеризуется прямой связью с объясняемой переменной и наоборот. Ширина разброса характеризует параметры исходной величины.

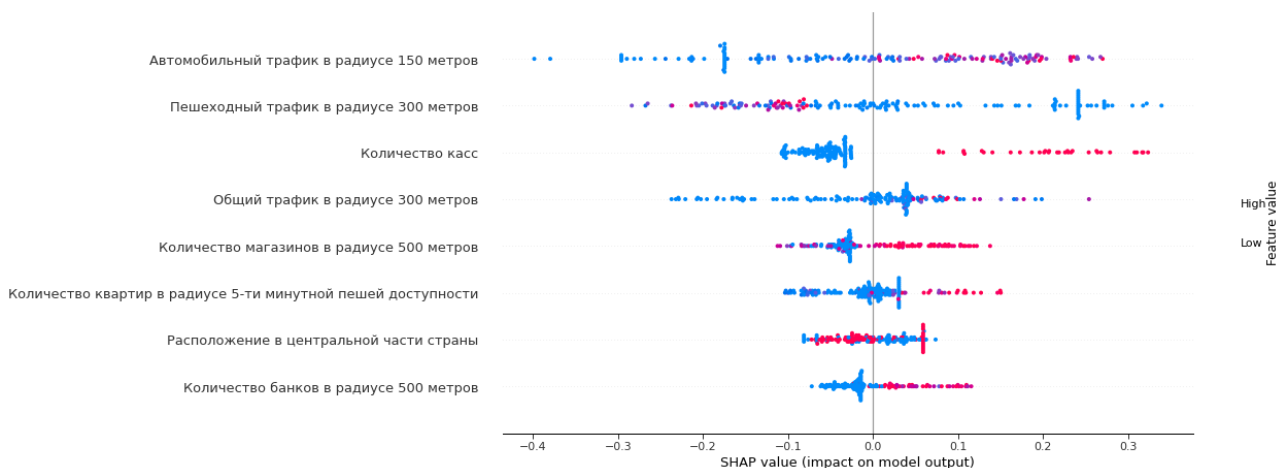


Рис. 7. SHAP-values модели с целевой переменной объем продаж АЗС за год

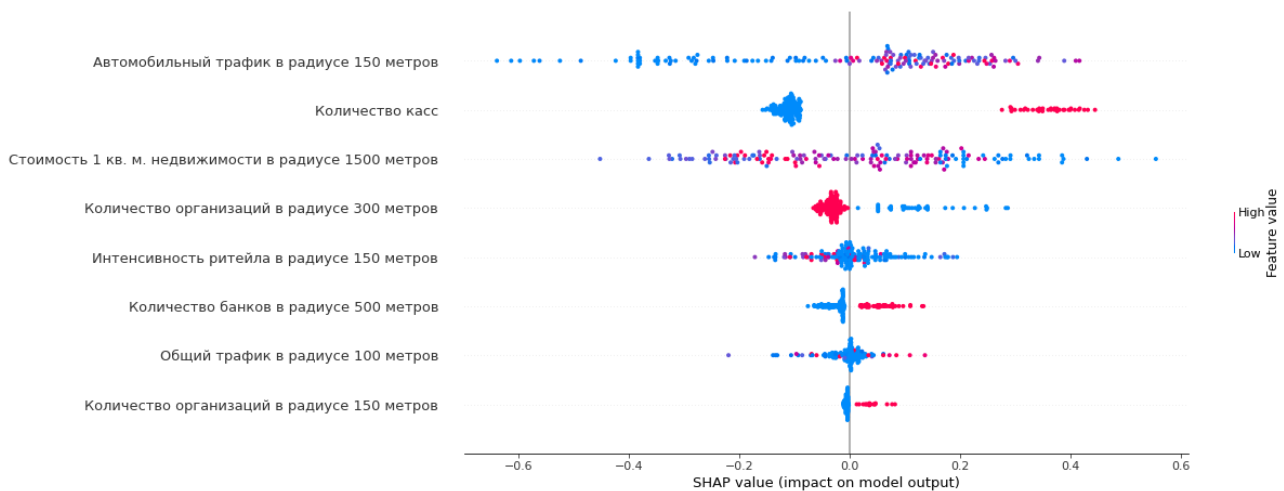


Рис. 8. SHAP-values модели с целевой переменной выручка магазина АЗС за год

Как показывает анализ рисунка 7, наибольший вклад несет фактор, связанный с максимальным автомобильным трафиком в радиусе 150 метров, а зависимость почти от всех факторов приближена к прямой. Более сложными зависимостями характеризуются пешеходный трафик в радиусе 300 метров и расположение в центральной части страны. С точки зрения интерпретации, неоднозначная зависимость от пешеходного трафика вполне логична. При этом от фактора, характеризующего расположение заправочной станции в центральном регионе вполне логичным является то, что объем пролива в этом округе в среднем ниже, чем в остальных, что подтверждается при анализе исходных данных. Анализ рисунка 8 показывает, что наибольшее влияние на прогноз также оказывает автомобильный трафик в радиусе 150 метров. При этом большая часть факторов напрямую влияют на величину прогноза. Исключениями могут служить только лишь стоимость 1 кв. м. недвижимости в радиусе 1500 метров, количество организаций в радиусе 300 метров и интенсивность ритейла в радиусе 150 метров. Нелинейность в характер зависимости от недвижимости привносит тот факт, что в рамках работы исследовались не только заправочные станции расположенные на трассах, но также азс, находящиеся в городской черте. Наивысшими показателями выручек зачастую обладали заправки, находящиеся на удалении от городов, именно поэтому и обратная зависимость от количества организаций и интенсивности ритейла.

6. Заключение

В статье был сформулирован подход к анализу и дальнейшему прогнозированию рентабельности при оптимальном выборе локаций для размещения автомобильных заправочных станций.

На примере двух целевых переменных: объем продаж в литрах и выручка магазина в рублях для каждой заправочной станции был проведен анализ рентабельности локаций. Основной целью которого являлось моделирование показателей объема продаж и доходности для каждой объясняемой переменной в рамках различных мест потенциальной работы.

Таким образом, был рассмотрен ряд понятий и подходов к решению задачи размещения на основе методов машинного обучения и исследования пространственных данных. Проведен анализ и обработка данных для моделирования, выбраны методы моделирования. Проведен корреляционный анализ, который показал перечень факторов наиболее тесно связанных с целевыми переменными, а также позволивший ограничить исходное пространство объясняющих факторов. Построены модели, выбраны метрики точности, а также подобраны гиперпараметры для наилучших алгоритмов. Реализовано построение моделей градиентного бустинга и проведена их содержательная интерпретация.

Анализ рентабельности локаций в рамках решения задачи размещения – это одно из важнейших направлений развития оффлайн бизнеса в современном мире. Именно поэтому большинство крупных организаций в сфере предоставления услуг или ритейла так или иначе занимаются исследования в области геоаналитики, применение которой диктуется возрастающей конъюнктурой рынка.

Литература

1. Бригхэм Ю., Эрхардт М. Финансовый менеджмент. 10-е изд. / Пер. с англ. под ред. к. э. н. Е. А. Дорофеева. – СПб.: Питер, 2009. – 960 с.

2. *Шеремет А. Д.* Методика финансового анализа деятельности коммерческих организаций / А. Д. Шеремет, Е. В. Негашев. – М.: Издательский Дом «Инфра-М», 2013. – 208 с.
3. *Щедрая Л. И., Шинин В. В.* Рентабельность в системе хозяйственного механизма // Вопросы политической экономии. 1988. N 197. – С. 18.
4. *Савицкая Г. В.* Экономический анализ: учебник для студентов вузов, обучающихся по экономическим направлениям и специальностям / Г. В. Савицкая. – 11-е изд., испр. и доп. – М.: Новое знание, 2005. – 651 с.
5. *Масленников В. Н., Афремов В. М.* Экономические методы определения рентабельности в промышленности: Вопросы регулирования рентабельности / В. Н. Масленников, В. М. Афремов. – М.: Финансы, 1975. – 190 с.
6. Большая российская энциклопедия: [в 35 т.] / гл. ред. Ю. С. Осипов. – М.: Большая российская энциклопедия, 2004-2017.
7. *Азрилиян И. Н.* Большой экономический словарь; под ред. А. Н. Азрилияна. – 2-е изд., перераб. и доп. – М.: Институт новой экономики, 1997. – 856 с.
8. Налоговый кодекс Российской Федерации (часть первая) от 31.07.1998 № 146-ФЗ (ред. от 29.05.2023) [Электронный ресурс], URL: https://www.consultant.ru/document/cons_doc_LAW_19671/
9. *Weber P., Chapman D.* Location Intelligence: An Innovative Approach to Business Location Decision-making // Transactions in GIS. – 2011. – Vol. 15 (3). – P. 309–328.
10. *Гринева Н.В., Топыркин А.Д.* Геоаналитика: сбор, анализ и отображение географических данных для решения задач размещения коммерческих организаций // Инновации и инвестиции. 2023. N 5. – С. 296–302.
11. *Гринева Н.В., Топыркин А.Д.* Анализ экономической эффективности локаций в сфере торговли и влияния на нее внешних факторов // Проблемы экономики и юридической практики. 2023. N 2 (19). – С. 184–196.
12. *Harris R.J., Sleight P., Webber R.* Geodemographics, GIS and Neighbourhood Targeting. Chichester, UK: John Wiley & Sons, 2005. – P. 288.