

РЕКОМЕНДАЦИИ ПО УЛУЧШЕНИЮ СЕРВИСОВ ТАКСИ НА ОСНОВЕ АНАЛИЗА ДАННЫХ О ПОЕЗДКАХ

Аннотация. В статье описана актуальность применения программных методов и средств для сервиса такси. Проведен анализ данных о поездках такси с целью выявления неэффективных практик и формирования рекомендаций по улучшению обслуживания. Исследование включает в себя предварительную обработку данных, анализ распределения скоростей и зависимостей средней продолжительности поездки и количества заказов от времени суток. С помощью метода кластеризации данных предложено выделить значимых мест для точки подачи машины и высадки пассажиров.

Ключевые слова: такси, анализ данных, визуализация, кластеризация, улучшение сервиса, рекомендации.

Введение. Сфера такси является одной из самых динамично развивающихся в современном мире. В условиях постоянно растущего числа заказов и жесткой конкуренции на рынке таксомоторных перевозок, когда успешными остаются компании, которые умеют правильно организовать обслуживание клиентов, необходимо постоянно улучшать эффективность сервиса и повышать качество обслуживания. Неоптимальное распределение и использование автомобилей такси часто приводит к неудовлетворенности клиентов и снижению доходов водителей. Исходя из результатов опроса ВЦИОМ [1], 21% пассажиров остаются недовольными сроками ожидания такси, а пожелание расширить таксопарки выразили 14% опрошенных, что говорит о существовании потребности в улучшении обслуживания клиентов.

Опубликовано достаточно большое количество работ, посвященных применению анализа данных и машинного обучения в сфере деятельности такси. Однако существующие исследования в основном уделяют внимание поведенческим шаблонам пассажиров и социальным атрибутам, обуславливающим их [2-4], особенностям передвижения водителей [5] или прогнозированию стоимости поездок [6], но не рассматривают конкретные меры по увеличению эффективности и прибыльности работы такси, которые можно вывести из имеющихся данных о передвижении автомобилей. Благодаря анализу данных о поездках, компании такси могут определить периоды пикового спроса и соответственно оптимизировать расписание работы водителей, выявить наиболее популярные маршруты и увеличить количество машин на них, чтобы уменьшить время ожидания клиентов. Проведение анализа данных о поездках может помочь выявить проблемные аспекты в работе службы такси, а также оно имеет высокий потенциал для повышения эффективности работы и улучшения качества обслуживания.

Проблема исследования. Основная проблема заключается в неоптимальном распределении и использовании автомобилей такси, что приводит к увеличению времени ожидания для клиентов и снижению доходов водителей. Цель исследования — составление рекомендаций по улучшению сервисов для вызова такси с использованием различных методов и программных средств анализа данных.

Методы и средства. Исследование основано на данных за период с 28 декабря 2023 года по 11 января 2024 года, включающих информацию о времени, скорости и местоположении

автомобилей, отправленную в случайный момент в ходе выполнения ими заказов. Для обработки, анализа и визуализации данных использовались библиотеки `pandas`, `numpy`, `scipy`, `seaborn`, `matplotlib`, `folium`, для проведения кластеризации (метод К-средних) применена библиотека `scikit-learn` языка Python.

Результаты. При изучении функционирования сервисов такси была выдвинута гипотеза о том, что на их работу могут негативно сказываться следующие тенденции:

1. Простаивание такси в отсутствие заказов.
2. Долгое ожидание пассажирами.
3. Отказ водителей ехать на дальние расстояния или в удаленные места.

Для того, чтобы разработать рекомендации по решению вышеперечисленных проблем, было проведено исследование данных о работе такси за период с 28 декабря 2023 года по 11 января 2024 года.

Первым этапом работы с данными стала их предварительная обработка. В ходе исследования проведена очистка и подготовка данных о поездках такси, включающая замену пропущенных значений, преобразование типов и удаление выбросов, таких как, например, аномально большие скорости.

Затем проведена оценка качества данных и их структуры. Были изучены основные характеристики данных, такие как средняя продолжительность поездок, среднее расстояние, средняя скорость автомобилей такси, популярные места назначения и т. д. Также проведен анализ распределений признаков и корреляций между ними.

Сделаны основные выводы об особенностях работы такси. Сезонную компоненту не стоит учитывать, так как обрабатываются данные одного сезона средней полосы России.

Средняя скорость передвижения такси в течение недели имеет тенденцию к возрастанию в начале и к концу недели, наиболее низкая средняя скорость наблюдается в пятницу. Скорость автомобилей такси достигает своего максимума в ночные часы (с 22:00 до 01:00) и наименьшего значения днем (с 11:00 до 15:00), что связано с увеличением трафика в утреннее и дневное время.

Средняя продолжительность поездки достигает своего максимума в пятницу. Длительность выполнения заказов увеличивается в период с 09:00 до 12:00, достигает минимума в период с 18:00 до 20:00. Кроме того, проведено сравнение средней продолжительности поездки в каждый из дней рассматриваемого периода (см. рис. 1). Выявлены дни с наиболее продолжительным средним временем в пути — 28, 29, 30 декабря. Это объясняется увеличением трафика на дорогах, вызванным изменениями в потребительском поведении людей перед периодом праздников.

Аналогично длительности выполнения заказов, количество заказов увеличивается в утренний и дневной период (с 09:00 до 13:00) и уменьшается в вечернее и ночное время (с 18:00 до 01:00). Распределение заказов по дням показывает, что к концу недели количество заказов увеличивается, достигает пика в четверг. На рис. 2 представлено распределение количества заказов по дням рассматриваемого периода. Можно заметить, что в первые дни января количество заказов снижается по сравнению с периодом перед Новым годом. Эта тенденция также подтверждает вывод о влиянии потребительского поведения на увеличение трафика.

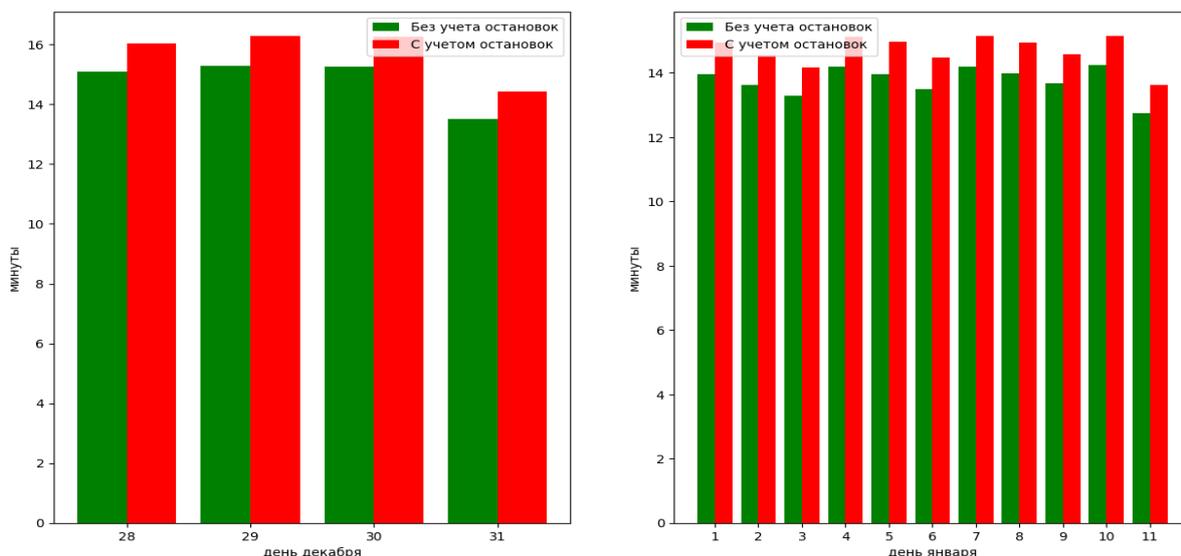


Рис. 1. Среднее время исполнения заказа по дням рассматриваемого периода

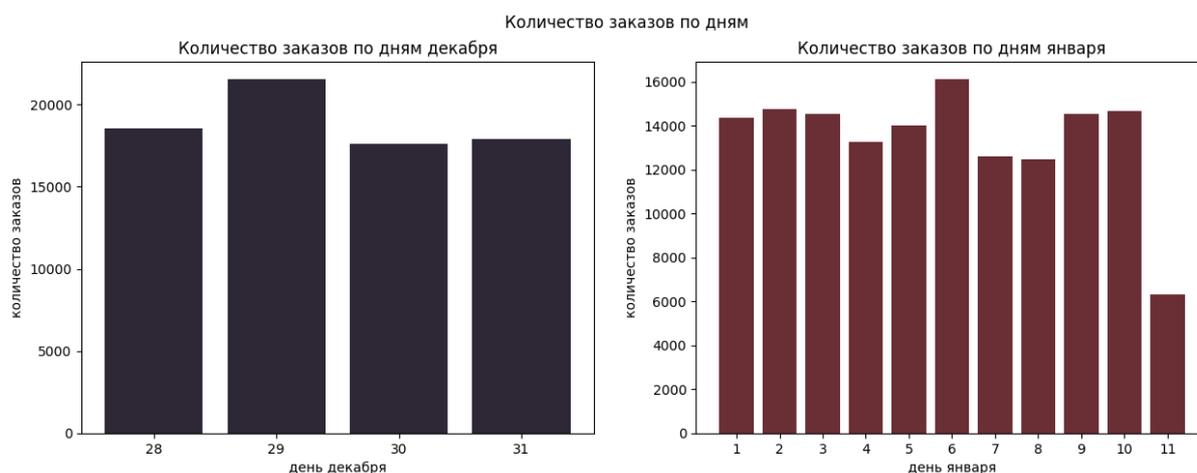


Рис. 2. Количество заказов по дням рассматриваемого периода

На основе данных о скорости автомобилей такси, продолжительности поездок и количестве заказов авторами предложены следующие рекомендации:

1. В период с 28 по 30 декабря нужно увеличить количество автомобилей для подачи.
2. Внедрить систему акций на длительные поездки (например, на поездки больше n км. распространяется скидка 10%).
3. Увеличить количество автомобилей для подачи в четверг и пятницу, а также в период с 09:00 до 12:00.

При анализе географического распределения отправленных автомобилями данных выявлено, что большинство посадок и высадок происходит в центральных районах населенных пунктов, что говорит о том, что именно в этих областях следует увеличить количество свободных автомобилей.

Авторами проанализированы записи, связанные с посадками и высадками в аэропортах и железнодорожных вокзалах. Можно заметить, что большое количество высадок в аэропортах произошло 9 января, что связано с возвращением людей домой после новогодних

праздников. На рис. 3 изображено распределение посадок и высадок в зоне аэропортов по часам дня. Большинство посадок происходит в 09:00, 23:00 и 01:00. Также в 00:00 и 08:00 высадок больше, чем посадок, что может быть связано с тем, что люди приезжают в аэропорт в среднем за час до вылета.



Рис. 3. Количество посадок и высадок в аэропортах по часам

Большое количество посадок и высадок в районах железнодорожных вокзалов произошло 29 декабря, что, возможно, связано с началом новогодних праздников. Наибольшее количество высадок характерно для 01:00 и 22:00, в эти временные точки количество высадок значительно превышает количество посадок. Максимум числа посадок наблюдается в 09:00 и 23:00.

Исходя из вышеперечисленных тенденций, связанных с работой такси в зонах аэропортов и железнодорожных вокзалов, можно сделать вывод о том, что важно адаптировать распределение автомобилей такси и их подачу в этих районах в соответствии с пиковыми временами посадок и высадок. Например, за час до наступления периода наибольшего количества посадок следует увеличить количество доступных автомобилей, чтобы клиенты могли своевременно добраться до аэропорта или вокзала.

Проведенный корреляционный анализ между признаками, характеризующими перемещение автомобилей такси (средняя скорость автомобиля и средняя продолжительность пути) и показателями погодных условий (температура, атмосферное давление, влажность воздуха, уровень облачности, дальность видимости, высота снежного покрова, наличие осадков и других погодных явлений) выявил, что связь между этими факторами не значительна.

Далее в рамках исследования был проведен кластерный анализ с применением алгоритма k-means для выявления основных групп поездок. Определено оптимальное количество кластеров с помощью метода локтя, произведена кластеризация всех точек, представленных в наборе данных, а также за отдельные временные периоды, и произведена интерпретация результатов. В полученных кластерах выявлены центры, которые можно считать точками, в которых водителю следует находиться, чтобы быстрее доехать до места подачи потенциального заказа (см. рис. 4).



✗ центр кластера

Рис 4. Кластеризация точек в период с 8 до 11 часов

Основываясь на результатах кластеризации, нами выдвинута рекомендация о том, что во время «простаивания» автомобилей следует направлять $x\%$ (где x — процент поездок, совершаемых в пределах одного района) водителей в центры кластеров, чтобы они могли быстрее доехать до места подачи потенциального заказа, $y\%$ (где y — процент посадок, происходящих в центральном районе) водителей отправлять в центральный район населенного пункта, а для оставшегося процента водителей внедрить систему поощрения на поездки в отдаленные районы.

Далее проведен анализ скоростей автомобилей во время поездки и сравнение их с допустимыми на законодательном уровне. Выявлено, что превышение допустимой скорости в черте города (60 км/ч) составляет около 7% случаев, в то время как для автомобилей такси,двигающихся за пределами города, превышение ограничения в 90 км/ч наблюдается уже в 21% случаев. Превышение допустимой скорости в черте города на 5 км/ч, а вне черты города на 10 км/ч увеличивает вероятность возникновения дорожно-транспортного происшествия в 2 раза [7], а также создает дискомфорт для пассажиров, поэтому для обеспечения безопасности как водителей, так и клиентов необходимо выявлять случаи превышения рекомендуемой скорости и принимать меры по их устранению.

Заключение. В результате исследования выявлены основные закономерности в данных о поездках такси, которые могут быть использованы для улучшения сервиса. На основе полученных данных можно выделить следующие рекомендации для улучшения сервисов такси:

1. Увеличить количество автомобилей для подачи в праздники.
2. Увеличить количество автомобилей для подачи в период, когда количество заказов достигает своего максимума.
3. Адаптировать распределение автомобилей такси и их подачу в аэропортах и железнодорожных вокзалах в соответствии с пиковыми временами посадок и высадок.
4. Определить точки, в которых автомобилям будет наиболее выгодно простаивать, и направлять к ним $x\%$ (где x — процент поездок, совершаемых в пределах одного района) водителей, чтобы они могли быстрее доехать до места подачи потенциального заказа, $y\%$

(где u — процент посадок, происходящих в центральном районе) водителей отправлять в центральный район населенного пункта, а для оставшегося процента водителей внедрить систему поощрения на поездки в отдаленные районы.

5. Выявить случаи превышения рекомендуемой скорости и внедрить системы мониторинга скорости движения, чтобы предотвратить превышение допустимых скоростей и уменьшить риск ДТП.

СПИСОК ЛИТЕРАТУРЫ

1. Такси в России: мнение пользователей. — Текст: электронный // ВЦИОМ. Новости: [сайт]. — URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/taksi-v-rossii-mnenie-polzovatelei> (дата обращения: 23.04.2024).
2. Du J. Analysis of Urban Residents' Travelling Characteristics and Hotspots Based on Taxi Trajectory Data / J. Du, C. Meng, X. Liu. *Applied Sciences*. — 2024. — № 14. — URL: <https://www.mdpi.com/2076-3417/14/3/1279> (date of the application: 23.04.2024). — Text: electronic
3. Yu L. Understanding Intra-Urban Trip Patterns from Taxi Trajectory Data / L. Yu, C. Kang, S. Gao, Y. Xiao [et al.] // *Journal of Geographical Systems*. — 2012. — № 14. — P. 464-483. — URL: https://www.researchgate.net/publication/235349383_Understanding_Intra-Uban_Trip_Patterns_from_Taxi_Trajectory_Data (date of the application: 23.04.2024). — Text: electronic.
4. Yang Q. Taxi Operation Optimization Based on Big Traffic Data / Q. Yang, Z. Gao, X. Kong, R. Azizur, J. Wang [et al.]. — Direct text // 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops. — Beijing, 2015. — P. 127-134.
5. Wang Y. Detecting Anomalous Trajectories and Behavior Patterns Using Hierarchical Clustering from Taxi GPS Data / Y. Wang, K. Qin, Y. Chen, P. Zhao // *ISPRS International Journal of Geo-Information*. — 2018. — № 7. — URL: <https://www.mdpi.com/2220-9964/7/1/25> (date of the application: 23.04.2024). — Text: electronic.
6. Андриянов Н.А. Применение методов факторного анализа и нейронных сетей для прогнозирования цены заказа службы такси / Н.А. Андриянов. — Текст: непосредственный // Управление развитием крупномасштабных систем: Труды Четырнадцатой международной конференции, Москва, 27-29 сентября 2021 года. — Москва, 2021. — С. 955-962.
7. Добромиров В.Н. Скорость как фактор влияния на безопасность дорожного движения / В.Н. Добромиров, С.С. Евтюков. — Текст: электронный // *Современные проблемы науки и образования*. — 2013. — № 5. — URL: <https://science-education.ru/ru/article/view?id=10248> (дата обращения: 23.04.2024).