

# **Документация**

**Дипломная работа по теме:  
“Анализ бронирования отелей  
(поиск инсайтов, составление  
рекомендаций стейкхолдерам)”**

Профессия “Аналитик данных”, DAU-84  
Ивандаева Юлия Адорьевна

**г. Москва, 2025**

## Оглавление

Введение.....	3
Блок 1. Описание исходного датасета и типов данных.....	4
Блок 2. Преобразование данных.....	5
Блок 3. Очистка Данных.....	6
3.1 Аномалии в данных.....	8
Блок 4. Анализ данных для стейкхолдеров.....	9
4.1 Анализ общих бронирований.....	9
4.1.1 Динамика бронирований по типу отеля .....	9
4.1.2 Динамика бронирований по месяцам.....	9
4.1.3 Распределение по типу клиента .....	10
4.2 Отмены бронирований.....	11
4.2.1 Факт отмены бронирований.....	11
4.2.2 Доля отмен по типу отеля .....	12
4.2.3 Доля отмен по месяцам.....	12
4.2.4 Доля отмен по типу депозита .....	13
4.2.5 Распределение отмен бронирования по странам.....	14
4.3 Поведенческий анализ отмен.....	15
4.3.1 Доля отмен по каналам дистрибуции .....	15
4.3.2 Доля отмен по сегментам.....	16
4.3.3 Доля отмен по типу клиента.....	17
4.3.4 Отмены по lead_time.....	18
4.3.5 Влияние на отмену бронирования.....	20
Итоги проекта и заключение.....	21

# Введение

## Цели проекта:

В рамках работы отельной индустрии, провести анализ имеющихся у отелей данных с целью определения поведения клиентов и отмены бронирований для составления рекомендаций по их уменьшению и определению основных стейкхолдеров отчетности.

## Бизнес-Задачи:

1. Создание формы единого отчета на основе предварительно подготовленных и очищенных данных, что позволит компании создать единую иерархию метрик;
2. Понять, какие сегменты и каналы дают больше всего отмен, а какие — стабильные заезды, что будет способствовать снижению потерь от отмен бронирований;
3. Определение влияния факторов, таких как: источник бронирования (lead\_time, agent, distribution\_channel), сезонность и пр., чтобы оптимизировать условия бронирования и тарифы;

## Стейкхолдеры:

- руководители отеля/отельная сеть
- отдел продаж
- маркетинговая команда
- финансовый отдел
- партнерские агентства.

## Интересы:

- Увеличить доход.
- Понять причины отмен.
- Оптимизировать загрузку номеров в пиковые и низкие сезоны.
- Повысить удовлетворенность клиентов.

## Блок 1. Описание исходного датасета и типов данных (32 столбца)

Для исследования был взят датасет “[Hotel bookings demands](#)” со статистикой бронирования отелей.

№	Имя Столбца	Описание	Тип данных
1	<i>Hotel</i>	Тип отеля	object
2	<i>Is_canceled</i>	Отмена брони	int64
3	<i>lead_time</i>	Количество дней между датой бронирования и датой заезда	int64
4	<i>arrival_date_year</i>	Год заезда в отель	int64
5	<i>arrival_date_month</i>	Месяц заезда в отель	object
6	<i>arrival_date_week_number</i>	Номер недели заезда	int64
7	<i>arrival_date_day_of_month</i>	День месяца заезда	int64
8	<i>stays_in_weekend_nights</i>	Количество ночей пребывания в выходные	int64
9	<i>stays_in_week_nights</i>	Количество ночей пребывания в будни	int64
10	<i>adults</i>	Количество взрослых	int64
11	<i>children</i>	Количество детей	float64
12	<i>babies</i>	Количество малышей	int64
13	<i>meal</i>	Тип питания	object
14	<i>country</i>	Код страны гостя	object
15	<i>market_segment</i>	Источник бронирования	object
16	<i>distribution_channel</i>	Каналы распределения бронирования	object
17	<i>Is_repeated_guest</i>	Постоянный гость	int64

18	<i>previous_cancellations</i>	Предыдущие отмены	int64
19	<i>previous_bookings_not_canceled</i>	Предыдущие не отмененные бронирования	int64
20	<i>reserved_room_type</i>	Забронированный тип номера	object
21	<i>assigned_room_type</i>	Назначенный тип номера	object
22	<i>booking_changes</i>	Количество изменений в бронировании	int64
23	<i>deposit_type</i>	Тип предоплаты	object
24	<i>agent</i>	ID турагенства	float64
25	<i>company</i>	ID компании, организации	float64
26	<i>days_in_waiting_list</i>	Количество дней в листе ожидания	int64
27	<i>customer_type</i>	Тип гостя	object
28	<i>adr</i>	Цена за ночь	float64
29	<i>required_car_parking_spaces</i>	Необходимые парковочные места	int64
30	<i>total_of_special_requests</i>	Общее количество специальных запросов	int64
31	<i>reservation_status</i>	Статус бронирования	object
32	<i>reservation_status_date</i>	Дата статуса бронирования	object

## Блок 2. Преобразование данных

В ходе исследования качества данных были сделаны следующие изменения:

- В столбце “**children**” – Изменен тип данных на “**int64**”
- В столбце “**agent**” – Изменен тип данных на “**int64**”
- В столбце “**reservation\_status\_date**” – Изменен тип данных на “**datetime64**”

### Блок 3. Очистка данных

Результат итогового вида обработанного датасета, а также выводы по причинам очистки данных приведены в таблице:

№	Новое имя Столбца	Преобразование данных	% NaN	Очистка данных
1	<i>Hotel</i>	object	0.00	Без изменений. Нет пустот
2	<i>is_canceled</i>	int64	0.00	Без изменений. Нет пустот
3	<i>lead_time</i>	int64	0.00	Без изменений. Нет пустот
4	<i>arrival_date_year</i>	int64	0.00	Без изменений. Нет пустот
5	<i>arrival_date_month</i>	object	0.00	Без изменений. Нет пустот
6	<i>arrival_date_week_number</i>	int64	0.00	Без изменений. Нет пустот
7	<i>arrival_date_day_of_month</i>	int64	0.00	Без изменений. Нет пустот
8	<i>stays_in_weekend_nights</i>	int64	0.00	Без изменений. Нет пустот
9	<i>stays_in_week_nights</i>	int64	0.00	Без изменений. Нет пустот
10	<i>adults</i>	int64	0.00	Без изменений. Нет пустот
11	<i>children</i>	int64	0.0034	Пустоты заменены на 0
12	<i>babies</i>	int64	0.00	Без изменений. Нет пустот
13	<i>meal</i>	object	0.00	Значения 'Undefined' были заменены на моду, их было 1169
14	<i>country</i>	object	0.4	Пустоты были заменены на моду, т.к. их количество не существенно (488)
15	<i>market_segment</i>	object	0.00	Значения 'Undefined' были удалены, т.к. их всего 2
16	<i>distribution_channel</i>	object	0.00	Значения 'Undefined' были удалены, т.к. их всего 5
17	<i>is_repeated_guest</i>	int64	0.00	Без изменений. Нет пустот

<b>18</b>	<b>previous_cancellations</b>	int64	0.00	Без изменений. Нет пустот
<b>19</b>	<b>previous_bookings_not_canceled</b>	int64	0.00	Без изменений. Нет пустот
<b>20</b>	<b>reserved_room_type</b>	object	0.00	Без изменений. Нет пустот
<b>21</b>	<b>assigned_room_type</b>	object	0.00	Без изменений. Нет пустот
<b>22</b>	<b>booking_changes</b>	int64	0.00	Без изменений. Нет пустот
<b>23</b>	<b>deposit_type</b>	object	0.00	Без изменений. Нет пустот
<b>24</b>	<b>agent</b>	int64	13.4	Пустоты были заменены 0
<b>25</b>	<b>company</b>	object	94.3	Пустоты были заменены на 'missing', т.к. этот столбец относится только к корпоративным бронированиям.
<b>26</b>	<b>days_in_waiting_list</b>	int64	0.00	Без изменений. Нет пустот
<b>27</b>	<b>customer_type</b>	object	0.00	Без изменений. Нет пустот
<b>28</b>	<b>adr</b>	float64	0.00	Без изменений. Нет пустот
<b>29</b>	<b>required_car_parking_spaces</b>	int64	0.00	Без изменений. Нет пустот
<b>30</b>	<b>total_of_special_requests</b>	int64	0.00	Без изменений. Нет пустот
<b>31</b>	<b>reservation_status</b>	object	0.00	Без изменений. Нет пустот

32	<b>Reservation_status_date</b>	datetime	0.00	Без изменений. Нет пустот
----	--------------------------------	----------	------	---------------------------

### 3.1 Аномалии в данных

Были обнаружены выбросы в столбцах 'children', где насчитывалось свыше 3 и равнялось 10 детям и 'babies', где свыше 2 малышей в одном номере – очень редко, данные строки были удалены. Кроме того, количество дней между бронированием и заселением ('lead\_time') превышало 334 и даже было свыше 600, все строки со значением превышающим 334 дня были удалены, как аномальные. Более того, 'stays\_in\_weekend\_nights' > 10 и 'stays\_in\_week\_nights' > 20 были расценены как очень редкие и маловероятные, было принято решение об удалении строк, превышающих данные значения. Что касается средней ставки за день ('adr'), были найдены как отрицательные значения (что невозможно), так и превышающие 1000, что так же является очень маловероятным, такие строки были удалены. И наконец, был создан дополнительный столбец 'total\_guests', который включал в себя 'adults' + 'children' + 'babies', в результате которого были выявлены бронирования с числом гостей равным 0, такие строки были удалены как бессмысленные.

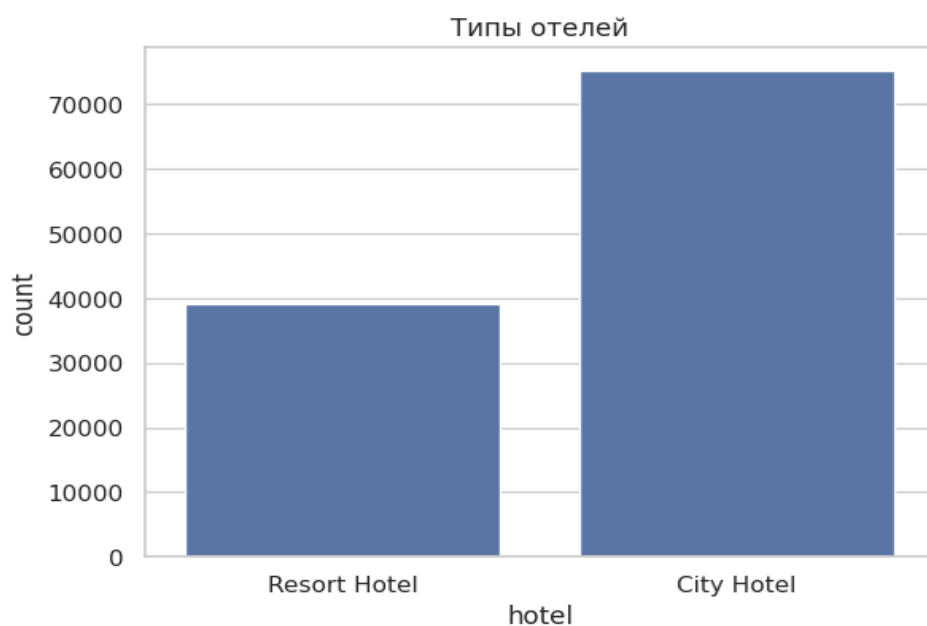


## Блок 4. Анализ данных для стейкхолдеров

Целью блока является поиск тенденций и инсайтов для составления рекомендаций стейкхолдерам.

### 4.1 Анализ общих бронирований

#### 4.1.1 Динамика бронирований по типу отеля



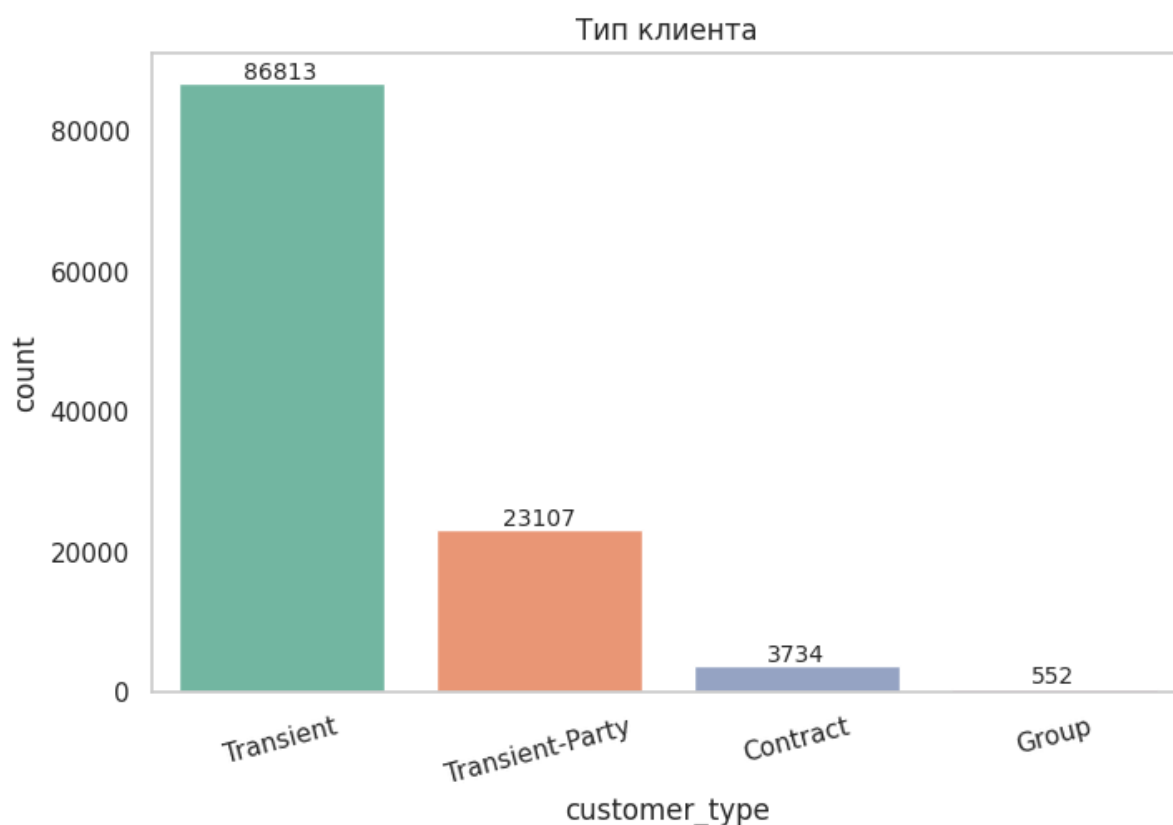
#### 4.1.2 Динамика бронирований по месяцам



### Вывод по пунктам 4.1.1 и 4.1.2:

По графику видно, что брони городских отелей почти в 2 раза выше курортных. Чётко прослеживаются регулярные пики и падения — это говорит о сезонном характере бронирований. Максимумы приходятся на весну/лето (апрель–август) — это логично: больше людей путешествуют летом. Минимумы — на конец осени и зиму (ноябрь–февраль). Это типичное поведение для туристических и курортных отелей. Периодичность повторяется в 2015, 2016 и 2017 годах — примерно одинаковая структура. Особенно заметен рост в 2017 году (пик ~6000 бронирований). Есть подозрительно резкие провалы — например, в ноябре 2015 и январе 2016 (падения ниже 2500): это может быть связано с праздниками, экономическими или техническими факторами.

### 4.1.3 Распределение по типу клиента



### Вывод:

Transient — доминирующий сегмент: это индивидуальные клиенты, бронирующие самостоятельно. С ними связано наибольшее количество отмен. Transient-Party — также важный сегмент: вероятно, частные группы, бронирующие

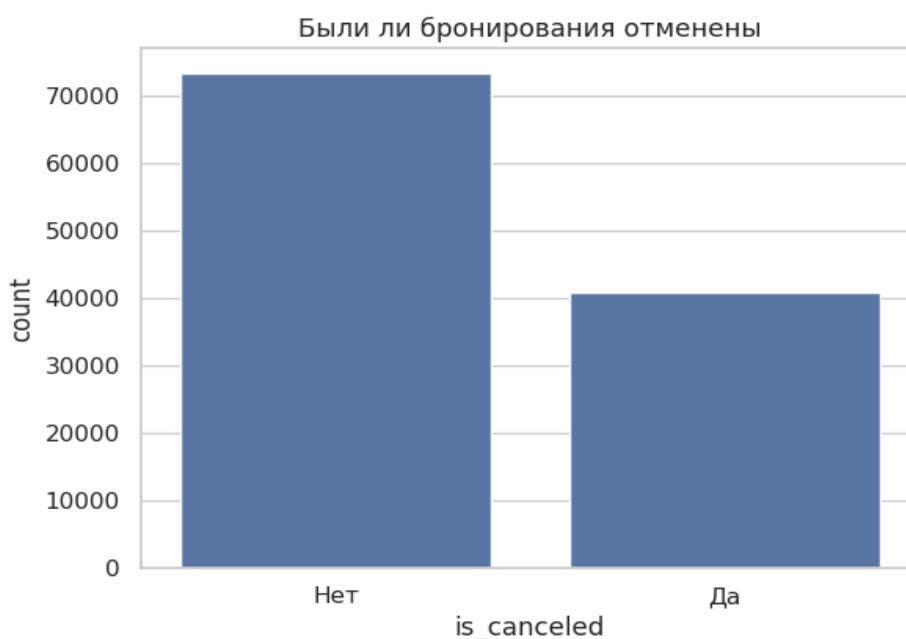
неофициально. Отменяют меньше, чем Transient — это хороший знак. Contract — стабильные клиенты: хоть их мало, у них низкая доля отмен. Group — практически отсутствует: вероятно, отель не ориентирован на крупные организованные заезды. Или группы регистрируются как другие типы клиентов.

### **Рекомендация Стейкхолдерам:**

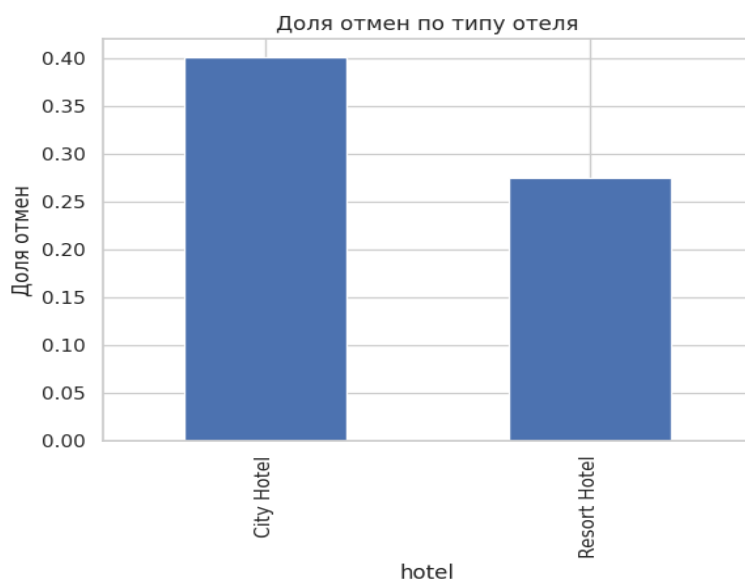
Увеличить маркетинговые инвестиции в сегмент City Hotel, так как он показывает более стабильный и высокий спрос. Курортные отели: сфокусировать продвижение на высокие сезоны (весна–лето) и разработать специальные акции в низкий сезон (осень–зима). Ввести динамическое ценообразование с повышением цен в пиковый период. Планировать штат сотрудников и ресурсы (уборка, персонал, поставки) с учётом сезонных колебаний. Продумать специальные зимние предложения или праздничные акции в периоды спада. Оптимизировать маркетинг и спецпредложения под Transient группу. Предлагать удобные условия отмен/изменения брони. Развитие сегмента Contract: несмотря на малую долю, контрактные клиенты — это стабильный поток. Рекомендуется: инициировать партнёрства с компаниями и бизнесами; предложить скидки при длительном сотрудничестве.

## **4.2 Отмены бронирований**

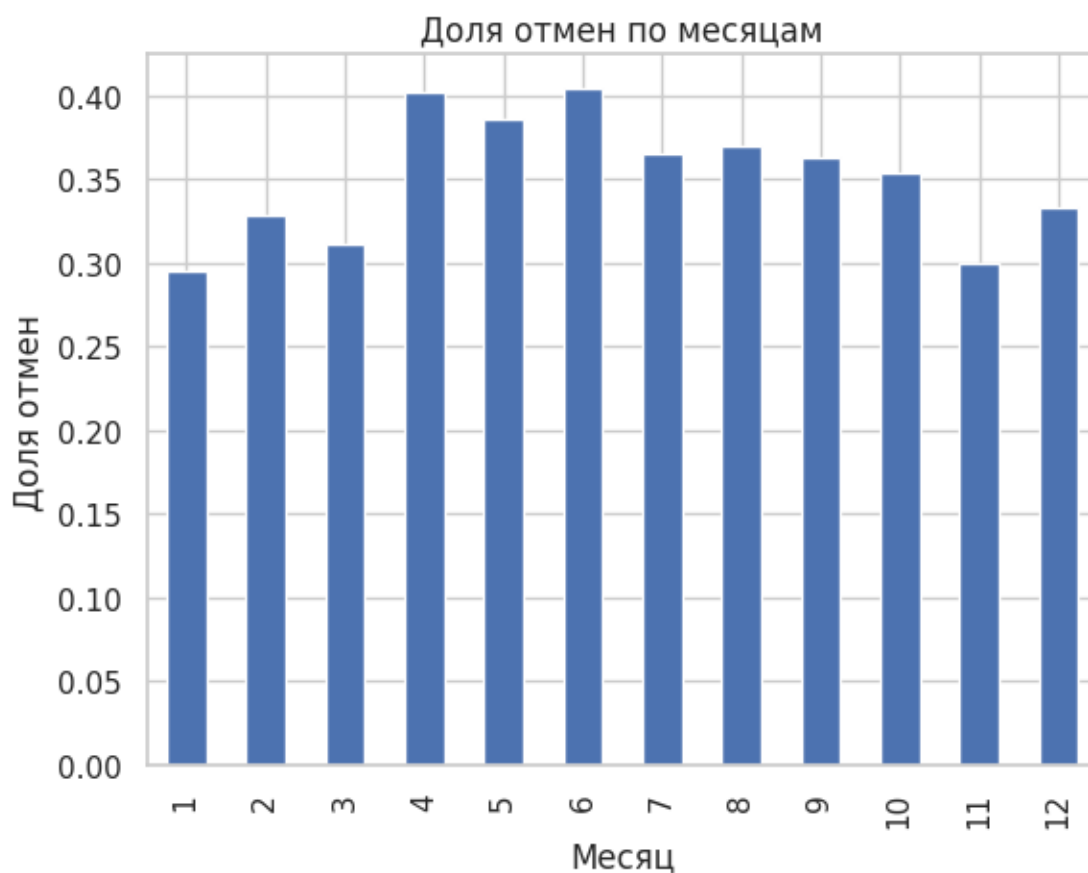
### **4.2.1 Факт отмены бронирований**



#### 4.2.2 Доля отмен по типу отеля



#### 4.2.3 Доля отмен по месяцам

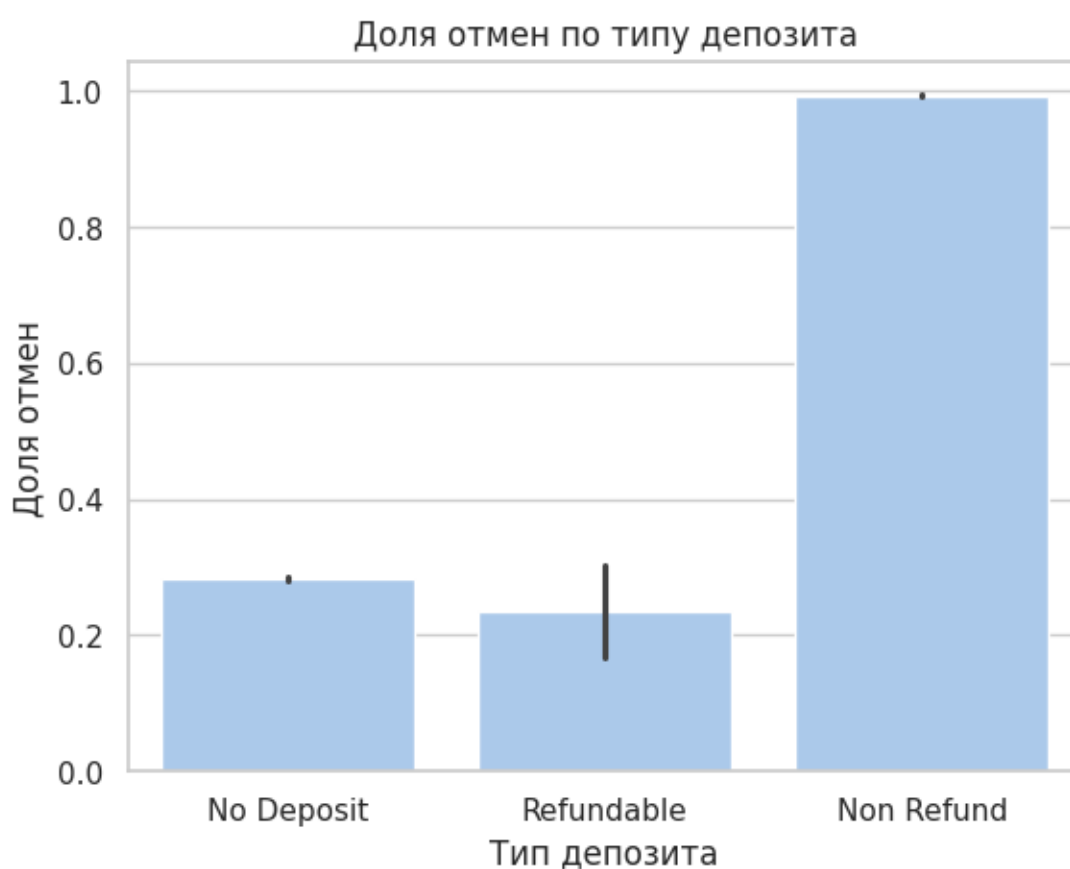


#### Выводы по пунктам 4.2.1, 4.2.2 и 4.2.3:

Больше всего отмены приходятся на городские отели (40%), что не удивительно, ведь там количество бронирований

соответственно больше. Пик отмен приходится на весну и раннее лето (апрель-июнь), это может быть связано с нестабильностью планов, сезонными колебаниями спроса, отпускными корректировками. Минимум отмен – зимние месяцы (январь, ноябрь), возможно, это связано с более уверенными корпоративными или деловыми поездками. Также, хочется отметить летние месяцы (июль, август) – чуть ниже, но всё ещё высокий уровень отмен. Вероятно, клиенты планируют отпуск, но вносят коррективы.

#### 4.2.4 Доля отмен по типу депозита



#### Вывод:

Non Refund (невозвратный тариф): почти 100% бронирований отменяются, несмотря на то, что возврат средств невозможен.

Это аномальное поведение — вероятно, оно связано с ошибками в данных или автоматическими отменами (например, системой ОТА или туроператора).

Refundable и No Deposit: доли отмен составляют около 23–29%.

При этом Refundable (с возвратом депозита) показывает чуть меньшую долю отмен, чем No Deposit, что может показаться неожиданным — брони по нему совершаются более осознанно.

#### 4.2.5 Распределение отмен бронирования по странам

Доля отмен по странам



#### Вывод:

На карте видно, что некоторые страны (США) демонстрируют высокую долю отмен — ~24%. Это выше средней по выборке (в среднем по набору — около 37% по всем данным, но нужно уточнить по группам).

#### Рекомендация Стейкхолдерам:

Ужесточить политику отмен для городских отелей: ввести невозвратные тарифы. Применить дифференцированную политику отмен: мягче — при раннем бронировании, жестче — при бронировании на ближайшие даты. Промо-акции при раннем бронировании с фиксированными датами. Период повышенного риска отмен: ужесточить условия отмен в весенние месяцы (например, штраф за отмену за 14 дней до заезда). Внедрить депозитную модель бронирования: удержание части суммы при отмене. Лето — пик отпусков и спонтанных изменений планов: продвигать тарифы с гибкой отменой по более высокой цене. Ввести опции изменения даты заезда без штрафов — вместо полной отмены. Январь и ноябрь = низкая

отменяемость: продвигать в это время корпоративные тарифы и деловые пакеты. Риск-менеджерам следует ужесточить условия отмен в проблемных странах. Продуктовым стратегам ввести только возвратные тарифы для стран с высокой лояльностью. Перенаправить бюджеты на рынки с высокой долей стабильных заездов. Проверить агентов, каналы и платформы с высоким процентом отмен.

## 4.3 Поведенческий анализ отмен

### 4.3.1 Доля отмен по каналам дистрибуции

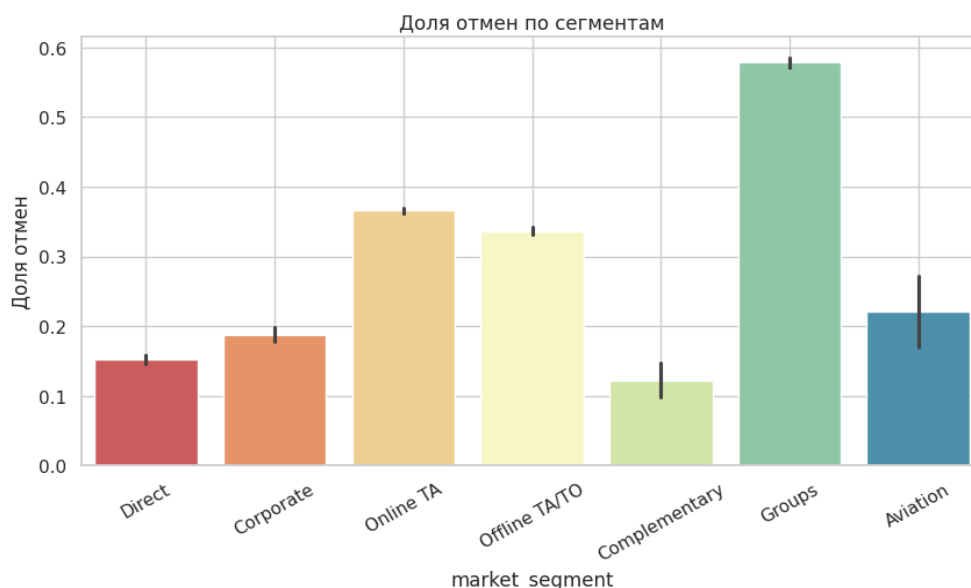


#### Вывод:

Наивысшая доля отмен бронирований наблюдается через канал TA/TO (Travel Agencies/Tour Operators) — почти 40%. Это может говорить о: высокой гибкости или политике отмен у туроператоров, спонтанности или неустойчивости таких бронирований, групповом бронировании с высокой вероятностью отмены. Корпоративный сегмент (Corporate) — вторая по уровню отмен (~23%). Возможные причины: командировки часто отменяются или переносятся по рабочим обстоятельствам. Каналы GDS (Global Distribution Systems) и Direct (прямые бронирования) имеют наименьшие доли отмен — 19% и 17% соответственно. Это говорит о более надёжных и

обоснованных бронированиях, особенно если клиент бронирует напрямую.

### 4.3.2 Доля отмен по сегментам

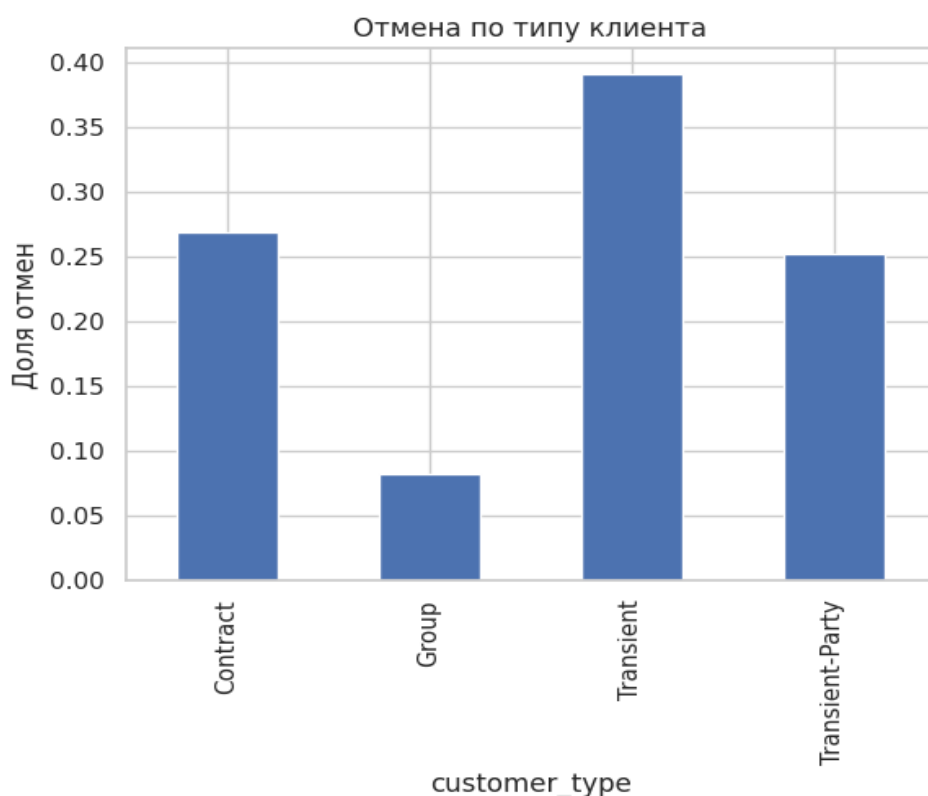


#### Вывод:

Наибольшая доля отмен наблюдается в сегменте Groups — почти 60%, что говорит о высокой нестабильности групповых бронирований. Online TA (Travel Agencies) и Offline TA/TO (Tour Operators) — также показывают высокий уровень отмен (35–37%). Это может быть связано с тем, что клиенты через посредников чаще бронируют с запасом и затем меняют планы. Самая низкая доля отмен — у клиентов из сегмента Complementary (бесплатные брони), а также Direct (самостоятельные брони) — около 15%. Это указывает на большую серьезность намерений при прямом бронировании. Сегмент Corporate демонстрирует средний уровень отмен (~19%), что характерно для деловых поездок: они устойчивее, но тоже подвержены корректировкам. Aviation показывает умеренные значения, но имеет широкий интервал доверия — то есть нестабильные данные, возможно, из-за малого объема.



### 4.3.3 Доля отмен по типу клиента



#### Вывод:

Transient (временные гости): самая высокая доля отмен — почти 40%. Это индивидуальные гости, бронирующие напрямую или через онлайн-агентства. Они чаще меняют планы, возможно из-за гибких условий брони или сезонных колебаний цен.

Contract (контрактные клиенты): доля отмен — около 27%. Такие брони обычно делаются компаниями/организациями по договору. Отмены здесь могут быть связаны с изменением командировок или бизнес-планов.

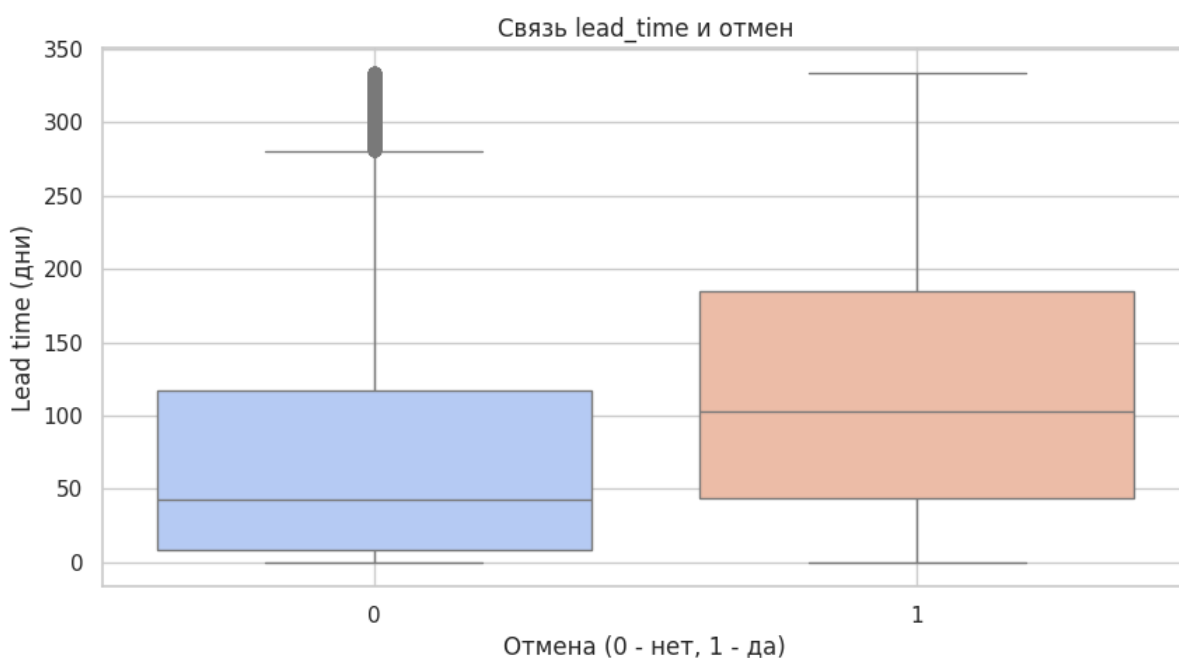
Transient-Party (временные гости в составе группы): доля отмен — около 25%, схожа с Contract. Это индивидуальные клиенты, путешествующие с группой, но оплачивающие отдельно. Могут быть чувствительны к изменениям в составе группы.

Group (групповые бронирования): самая низкая доля отмен — около 8%. Групповые поездки часто организуются заранее и реже отменяются (например, туры, конференции).

## Рекомендации Стейкхолдерам:

Пересмотреть политику отмен и внести дифференциацию по каналам: чем выше историческая доля отмен — тем строже условия. Снизить риски от сегмента Groups: ввести более строгие условия отмены или депозиты для групповых бронирований. Настроить уведомления и подтверждения за несколько дней до заезда. Укрепить канал Direct: продвигать прямые бронирования через сайт или мобильное приложение. Предлагать дополнительные бонусы (например, ранний заезд, бесплатный завтрак) для клиентов, бронирующих напрямую. Пересмотреть работу с Online TA и Offline TA/TO: анализировать агентов с высоким уровнем отмен — возможно, есть смысл ограничить сотрудничество. Ввести более гибкую политику возврата с динамическими условиями — например, ближе к дате заезда делать возврат менее выгодным. Работа с корпоративными клиентами: предлагать гибкие тарифы с гарантией брони, особенно для постоянных клиентов.

### 4.3.4 Отмены по lead\_time



## Вывод:

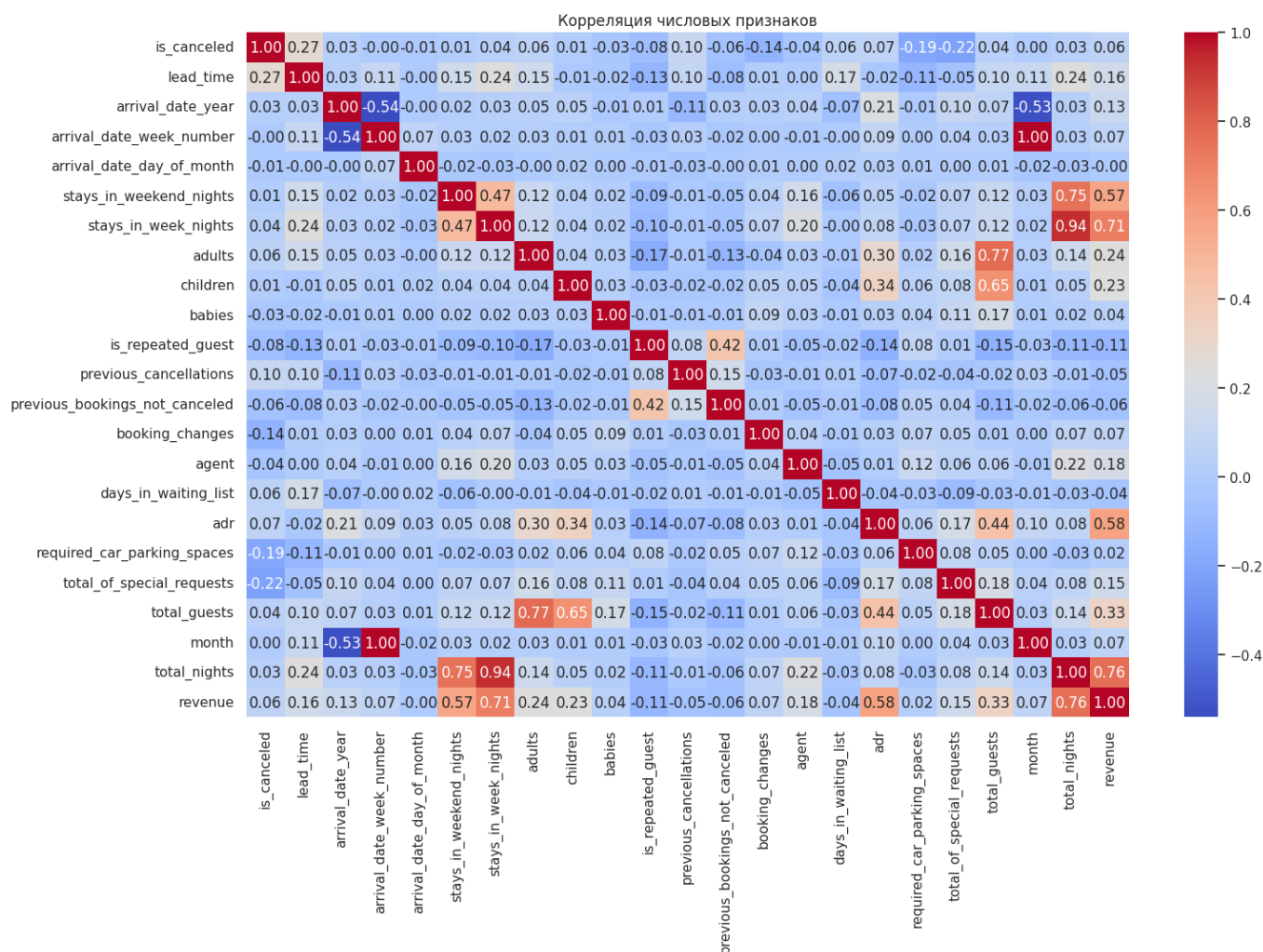
На основании анализа видно, что: отменённые бронирования имеют существенно более высокий lead\_time (медиана ~100 дней), чем неотменённые (медиана ~30 дней). То есть, чем раньше клиент бронирует заезд — тем выше вероятность, что он его отменит. Это логично: за 2–3 месяца до поездки планы у

клиентов могут поменяться (отпуск, визы, билеты, здоровье и т.д.). Отмены с длинным lead\_time — это упущенная потенциальная прибыль, особенно в высокий сезон.

### **Рекомендации Стейкхолдерам:**

1. Ввести дополнительные условия для длинных бронирований. Для бронирований с lead\_time > 60–90 дней: предлагать только невозвратные или частично возвратные тарифы. Требовать депозит или предоплату, чтобы снизить мотивацию к отмене.
2. Оптимизировать тарифную политику. Динамически регулировать цены: чем ближе дата заезда — тем более гибкий тариф (меньше отмен). Чем раньше — тем выше цена или жёстче условия отмены.
3. Внедрить систему напоминаний. Автоматические письма за 1–2 недели до заезда: подтверждение брони. Это поможет сократить «забытые» или «ненужные» брони и дать шанс переиспользовать номер.

### 4.3.5 Влияние на отмену бронирования



#### Вывод:

По матрице корреляций мы видим положительную корреляцию: `lead_time` (+0.27) — чем больше времени между бронированием и заселением, тем выше шанс отмены; `adr` (+0.06) — слабо, но рост цены может немного увеличивать вероятность отмены. `previous_cancellations` (+0.09) — клиенты, уже отменявшие бронирования, склонны делать это снова. Отрицательная корреляция: `is_repeated_guest` (-0.08) — постоянные гости реже отменяют бронирования; `total_of_special_requests` (-0.22) — чем больше пожеланий, тем меньше вероятность отмены: клиент настроен серьезно.

## **Итоги проекта и заключение**

В нынешнее время отельная индустрия набирает популярность всё больше и для владельцев бизнеса были найдены нужные инсайты и рекомендации, которые смогут помочь им в построении стратегии на будущие годы и улучшения своих показателей.

### **По бизнес-задачам:**

1. Для задачи построения единого отчета были преобразованы и очищены, имеющиеся у компании сырые данные. Это позволит выработать единую иерархию метрик и внедрить их в компанию.
2. Анализ поведения клиентов и структуры отмен позволил определить ключевых стейкхолдеров и выделить значимые метрики, влияющие на успех бизнеса. OTA (Online TA, TA/TO) и групповые сегменты показали высокий уровень отмен. Напротив, прямые и корпоративные бронирования — более стабильные. Менеджерам по работе с клиентами заранее учитывать риски отмен, использовать информацию о сегментах и каналах при принятии решений о заселении и удержании клиентов. Менеджеры по маркетингу на основе данных об источниках бронирования и сезонности могут настраивать рекламные кампании, ориентируясь на надёжные сегменты и минимизируя затраты на нестабильные. Финансовые аналитики и топ-менеджмент — получают инструмент для прогнозирования доходов, выявления провалов в загрузке и стратегического планирования по улучшению показателей.
3. Анализ ключевых факторов, влияющих на вероятность отмены (в том числе: количество дней до заселения, источник бронирования, канал дистрибуции, сезонность и депозит), позволил выявить поведенческие паттерны клиентов и определить, какие условия наиболее сильно влияют на нестабильность бронирований. Длительный срок между бронированием и заездом значительно повышает вероятность

отмены. Весна и начало лета — периоды, когда требуется ужесточение условий возврата.

### **Рекомендации:**

В процессе анализа было выявлено, что ряд бизнес-гипотез ограниченно подтверждаются из-за нехватки детализированных данных, что затрудняет построение точных выводов и прогностических моделей. Для повышения информативности и аналитической ценности отчёта рекомендуется расширить исходный датасет следующими колонками:

#### **ID клиента или идентификатор гостя:**

Позволит отслеживать поведение конкретного клиента во времени: определить долю повторных заездов и лояльность гостей; оценить цикл жизни клиента и рассчитать LTV; сегментировать клиентов по их ценности и частоте.

#### **Цена за бронирование на момент подтверждения: (фиксированный тариф)**

Позволит сравнивать динамику цен, оценивать гибкость ценовой политики и находить перекосы: отслеживать влияние сезонности, каналов и сегментов на тариф; выявлять заниженные или завышенные тарифы; сопоставлять с ADR и отклонениями при отменах.

#### **Дата фактического заезда / выезда:**

Позволит точнее рассчитывать сезонные пики, выявлять задержки и недоезды: определить, совпадают ли плановые даты с реальными; анализировать частоту переносов или no-show; построить реальные оссурансу-модели по дням и неделям.

#### **Причина отмены (если доступна):**

Ключевая колонка для построения прогнозов отмен: позволит различать добровольные отмены, технические сбои и форс-мажоры.

#### **Маркетинговые активности или канал привлечения:**

Позволит увязать поведение клиента с источником трафика: определить эффективность рекламных кампаний; выделить каналы

с высокой долей отмен или, наоборот, с высокой конверсией; поможет отделу маркетинга корректно перераспределять бюджет.

### **Заключение:**

Несмотря на отдельные сложности с качеством данных, структурой признаков и интерпретацией, благодаря пошаговой очистке и логическому подходу к анализу удалось сформировать чистый датасет, выявить ключевые зависимости и подготовить наглядные рекомендации для бизнеса.