

MMMG team

PROBLEM Tech



Задача – предсказание **отмены** бронирования номера в отеле.

Business



Задача – повышение доходов отелей.

Базовое решение и анализ данных



Наша стратегия заключалась в глубоком анализе и использовании множества инструментов для достижения высокой метрики. Это позволило выявить оптимальные подходы к предобработке данных.

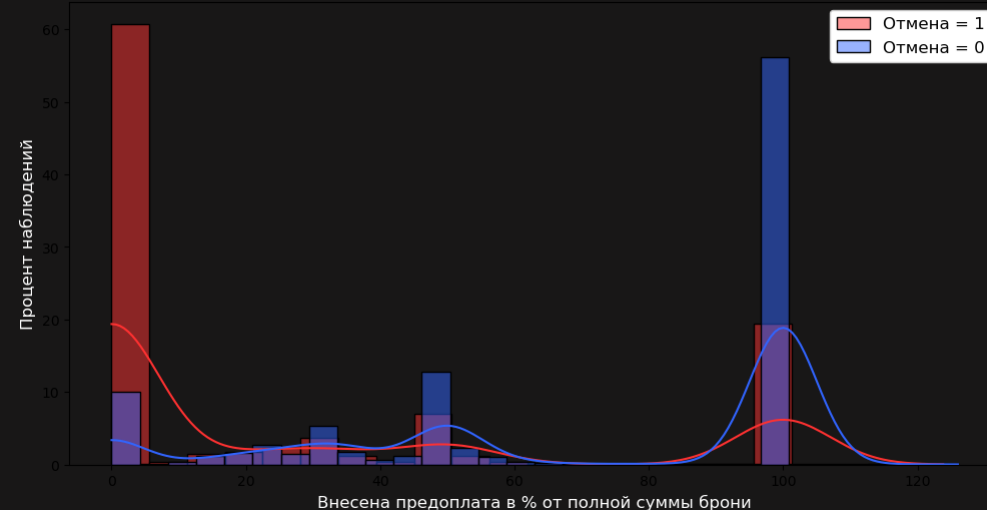
Ознакомившись с данными и увидев, что присутствует много категориальных признаков, мы решили применить алгоритм CatBoost для первоначального исследования ситуации. Модель и её предсказательная способность показали себя хорошо.

EDA

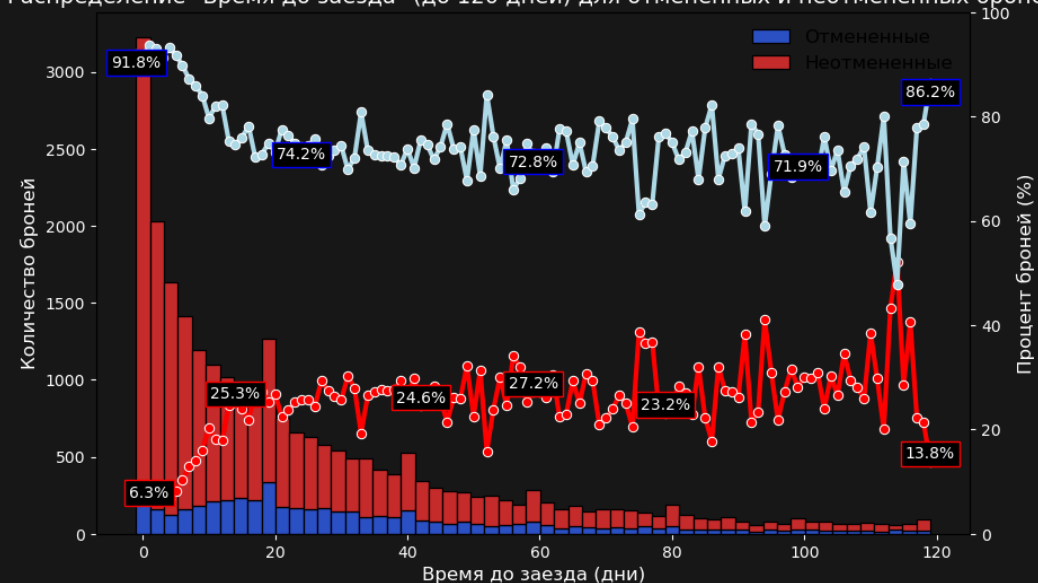
Способы оплаты без **отмены**



Распределение предоплаты в зависимости от отмены



Распределение "Время до заезда" (до 120 дней) для отмененных и неотмененных броней



Обработка данных

1. Проанализировав **уникальные значения** источников записей, мы соединили все похожие.
2. Были удалены или преобразованы признаки с **высокой корреляцией**, чтобы избежать **мультиколлинеарности** и **улучшить обучаемость модели**.
3. Для колонок, где было несколько категорий номера, мы разделили их и оставили только самые популярные варианты.

Корреляционная матрица числовых признаков

	Номеров	Гостей	Отмена	День недели заезда	Время до заезда	День недели бронирования	Месяц заезда
Номеров	1	0.52	0.029	-0.0035	0.1	-0.0041	0.012
Гостей	0.52	1	0.024	0.029	0.11	0.0027	0.0056
Отмена	0.029	0.024	1	-0.013	0.16	-0.012	0.001
День недели заезда	-0.0035	0.029	-0.013	1	-0.07	0.026	0.034
Время до заезда	0.1	0.11	0.16	-0.07	1	0.015	0.024
День недели бронирования	-0.0041	0.0027	-0.012	0.026	0.015	1	0.00048
Месяц заезда	0.012	0.0056	0.001	0.034	0.024	0.00048	1

Создание новых признаков

Мы создали несколько новых признаков, которые оказались полезными для повышения качества модели:

Внесена предоплата (%)

День недели заезда

День недели бронирования

Длительность пребывания

Время до заезда (дни)

Месяц заезда

* Часть **признаков** была **отсеяна** с помощью анализа важности признаков в **CatBoost**.

КАК найти гостиницы?

1

1. Исследуем

“Способ оплаты”

2. Исследуем

“Категория номера”

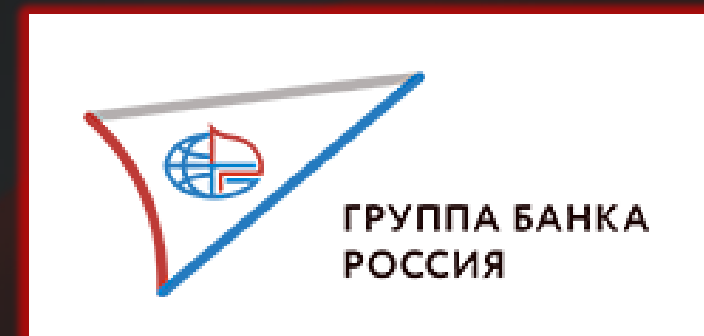
2

3. Находим

Гостиницу **“Игора”**
и
“Точка на карте”



+



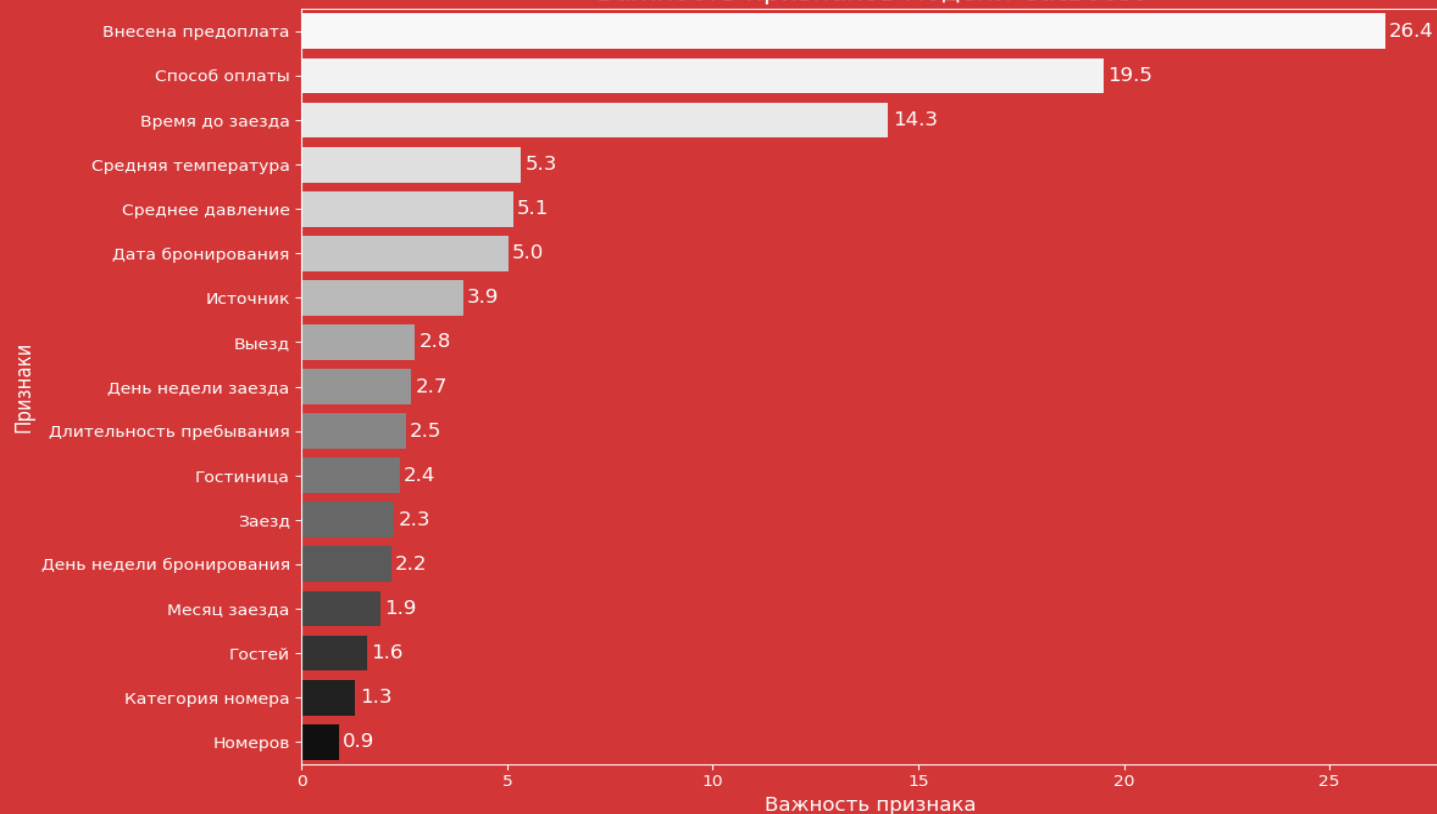
Добавление данных о погоде

Мы добавили данные о погодных условиях (температура и атмосферное давление) в местах нахождения отелей.

Эти данные были усреднены по времени между моментом бронирования и датой заезда, чтобы учесть возможное влияние погоды на поведение клиентов.



Важность признаков модели CatBoost

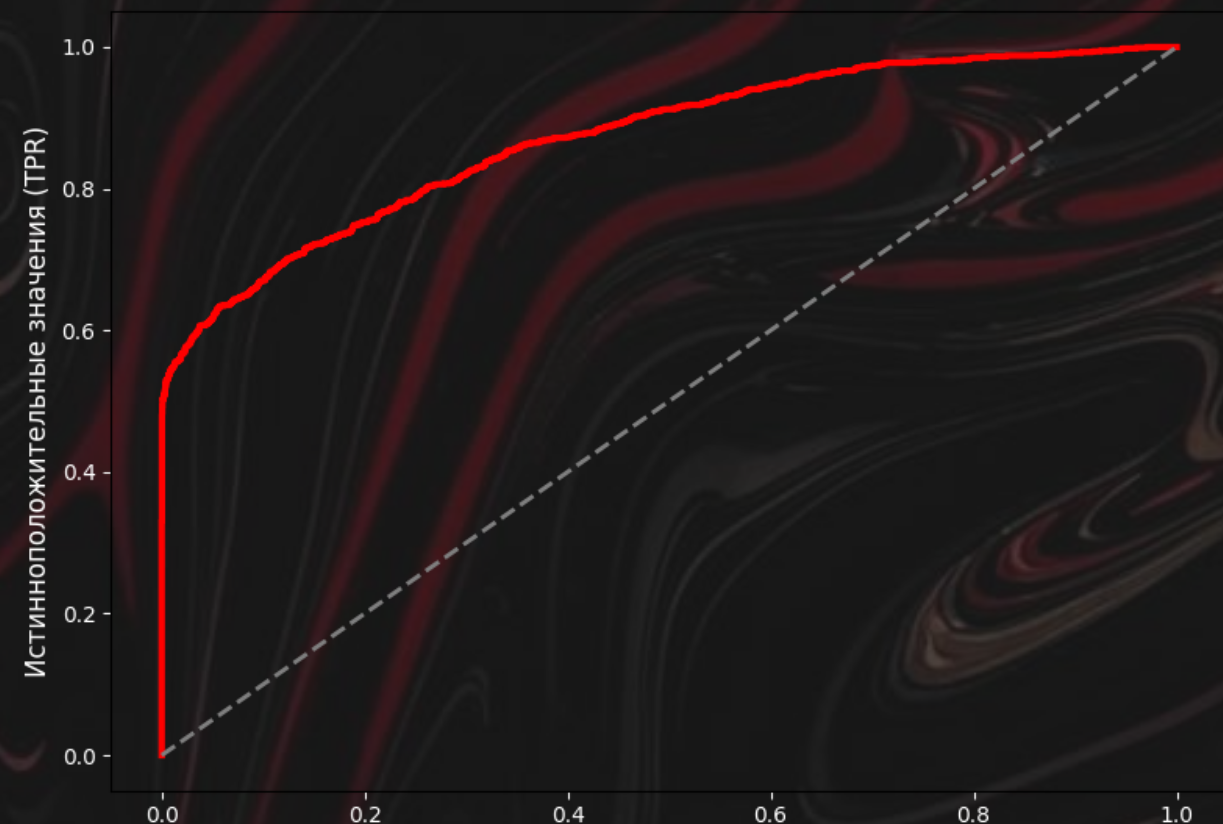


Ход работы

- **CatBoost** на 5-фолдовой кросс-валидации показала хорошие результаты.
- На графике, построенном с помощью встроенного анализа важности признаков в CatBoost, видно, что добавленные данные о погоде имеют существенное значение

МОДЕЛИ

ROC-AUC



OPTUNA



0.8624

МЕТРИКИ

Развитие модели

Пробовали:

- В процессе работы мы также тестировали другие модели, такие как XGBoost и LightGBM, однако их результаты были хуже по сравнению с CatBoost.
- Попытки очистки данных от выбросов и использование бинарного признака "Выходной день" также не дали положительного эффекта.

Предположение:

- Моделирование временных зависимостей.
- Применить oversampling для балансировки классов.
- Использовать другие виды ансамблей моделей.
- Сгенерировать более сложные зависимости между данными.
- Более детально исследовать влияние погоды и других региональных данных на целевую переменную.

Расходы

Переменные затраты	Сумма	Источник
Разработка сервиса под ключ	500 000	Workspace

Постоянные затраты	Сумма	Источник
Системный администратор	110 000 в месяц	hh.ru
Служба клиентской поддержки	76 000 * 2 в месяц	hh.ru
Аренда помещения	88 000 в месяц	Циан
Маркетинг b2b	30 000 000 в год	Рекламное агентство
Маркетинг b2c	50 000 000 в год	Smm –агенства
IT обслуживание сервиса	210 000 * 2	hh.ru

USP-B2C

Отели Авиабилеты Ж/д билеты Автобусы Туры

Санкт-Петербург

Заезд 7 авг

Выезд 14 авг

2 взрослых

Найти

Отели и гостиницы в Санкт-Петербурге

Все фильтры

Популярные

Цена

РЕКЛАМА

Лучшие каникулы с нами

ДО 2000 БАЛЛОВ ПО ПРОГРАММЕ ЛЮБИМЫХ ИТ-ПАРТНЕРОВ

КАШБЭК ДО 10% при оплате ПСБ

Вдохновиться

Приозерск

6 000 ₽

цена за 1 ночь за 1 номер

Выбрать

Рекомендуем

10% на экскурсии

5.0 10393 отзыва · Гостиница

На эти даты отличная погода

2000 бонусов на услуги отеля

Можно разделить на 6 месяцев

Страховка при поздней отмене

Port Comfort by Sennaya Square 4★

4 600 ₽

цена за 1 ночь за 1 номер

Выбрать

Рекомендуем

Отели в Санкт-Петербурге на карте

Бронируйте отели с кэшбеком до 20%

7871 ₽

2000 ₽

10500 ₽

3480 ₽

9500 ₽

4800 ₽

7700 ₽

3839 ₽

11800 ₽

6000 ₽

4200 ₽

5130 ₽

5300 ₽

3240 ₽

Скидка 10% на гастрономии при бронировании любого отеля

КАШБЕК ДО 15% ПРИ БРОНИРОВАНИИ 3 И БОЛЕЕ ЧЕЛОВЕК

2% тебе 2% другу

2 УРОВЕНЬ ПУТЕШЕСТВЕННИКА

CatBoost-B2B

НАЗВАНИЕ	Описание	ПЛЮСЫ ДЛЯ ОТЕЛЕЙ
Меняющийся % овербукинга	На основе результатов полученных от нашей модели мы можем предсказать количество отмен в каждый период времени и выставить на прожладу броней больше равное этому количеству	Это поможет отелям уменьшить количество пустующих из-за отмен номеров
Полная предоплата	На основе статистических данных мы выяснили что люди делающие полную предоплату реже отменяют бронирование	Это поможет отелям уменьшить количество отмен
Начисление бонусов	Мы предлагаем начислять бонусы в размере 10% от стоимости услуги в отеле при бронировании отеля	Это спровоцирует покупателей пользоваться дополнительными услугами отеля
Оценка погоды	Наш сервис будет на этапе выбора показывать предложения во время которых будет хорошая погода	Это поможет покупателям лучше планировать отдых и реже делать отмены

ИТОГИ:

Время окупаемости
менее года

Расходы в год	Доходы за год	Итог за год
500 000 – создание сервиса	115 000 000	2 500 000
9 240 000 – зарплаты сотрудникам и аренда помещения		
80 000 00 – маркетинг		
115 000 000 * 20 % – налоги		

- **Стоимость проекта 500 000 р**
- **Средняя выручка 4000 р**
- **Целевая аудитория студенты**

Уменьшение отмен: овербукинг, полная предоплата, оценка погоды

NPV = 10.598.041

MMG состав

Капитан – Денис Маликов

Денис Маликов – DS/DA

Артём Таратин – DS

Даниил Аль-Натор – DS

Илья Обухов – DE

Максим Чудасов – PM



**Спасибо за
внимание!**