

## Описание проекта

Заказчик этого исследования — сеть отелей «Как в гостях».

Чтобы привлечь клиентов, эта сеть отелей добавила на свой сайт возможность забронировать номер без предоплаты. Однако если клиент отменял бронирование, то компания терпела убытки. Сотрудники отеля могли, например, закупить продукты к приезду гостя или просто не успеть найти другого клиента.

Чтобы решить эту проблему, вам нужно разработать систему, которая предсказывает отказ от брони. Если модель покажет, что бронь будет отменена, то клиенту предлагается внести депозит. Размер депозита — 80% от стоимости номера за одни сутки и затрат на разовую уборку. Деньги будут списаны со счёта клиента, если он всё же отменит бронь.

## Бизнес-метрика и другие данные

Основная бизнес-метрика для любой сети отелей — её прибыль. Прибыль отеля — это разница между стоимостью номера за все ночи и затраты на обслуживание: как при подготовке номера, так и при проживании постояльца.

В отеле есть несколько типов номеров. В зависимости от типа номера назначается стоимость за одну ночь. Есть также затраты на уборку. Если клиент снял номер надолго, то убираются каждые два дня.

Стоимость номеров отеля:

- категория А: за ночь — 1 000, разовое обслуживание — 400;
- категория В: за ночь — 800, разовое обслуживание — 350;
- категория С: за ночь — 600, разовое обслуживание — 350;
- категория D: за ночь — 550, разовое обслуживание — 150;
- категория E: за ночь — 500, разовое обслуживание — 150;
- категория F: за ночь — 450, разовое обслуживание — 150;
- категория G: за ночь — 350, разовое обслуживание — 150.

В ценовой политике отеля используются сезонные коэффициенты: весной и осенью цены повышаются на 20%, летом — на 40%.

На разработку системы прогнозирования заложен бюджет — 400 000. При этом необходимо учесть, что внедрение модели должно окупиться за год. Затраты на разработку должны быть меньше той выручки, которую система принесёт компании.

## Инструкция по выполнению

### Шаг 1. Откройте файлы с данными

Пути к файлам:

/datasets/hotel\_train.csv — данные для обучения модели.

[Скачать датасет](#)

/datasets/hotel\_test.csv — данные для тестирования модели.

[Скачать датасет](#)

## Шаг 2. Предобработка и исследовательский анализ данных

- Проверьте данные и откорректируйте их, если потребуется.
- Проведите исследовательский анализ представленных данных. Опишите найденные особенности: пропуски, выбросы.

## Шаг 3. Вычислите бизнес-метрику

Оцените прибыль отеля без внедрения депозитов.

## Шаг 4. Разработка модели ML

- Обучите разные модели и оцените их качество кросс-валидацией. Выберите лучшую модель и проверьте её на тестовой выборке. Опишите выводы.
- Выберите метрику для обучения.
- Оцените прибыль, которую принесёт выбранная модель за год.

## Шаг 5. Выявите признаки «ненадёжного» клиента

На основе исследовательского анализа данных опишите клиента, склонного к отказу от брони.

## Шаг 6. Напишите общий вывод

Опишите модель, которая решает задачу с наибольшей выгодой для бизнеса. Дайте бизнесу рекомендации по результатам проделанной работы.

# Оформление

Выполните задание в Jupyter Notebook. Заполните программный код в ячейках типа `code`, текстовые пояснения — в ячейках типа `markdown`. Используйте форматирование и заголовки.

## Описание данных

В таблицах `hotel_train` и `hotel_test` содержатся одинаковые столбцы:

- `id` — номер записи;
- `adults` — количество взрослых постояльцев;
- `arrival_date_year` — год заезда;
- `arrival_date_month` — месяц заезда;
- `arrival_date_week_number` — неделя заезда;
- `arrival_date_day_of_month` — день заезда;
- `babies` — количество младенцев;
- `booking_changes` — количество изменений параметров заказа;
- `children` — количество детей от 3 до 14 лет;

- `country` — гражданство постояльца;
- `customer_type` — тип заказчика:
  - `Contract` — договор с юридическим лицом;
  - `Group` — групповой заезд;
  - `Transient` — не связано с договором или групповым заездом;
  - `Transient-party` — не связано с договором или групповым заездом, но связано с бронированием типа `Transient`.
- `days_in_waiting_list` — сколько дней заказ ожидал подтверждения;
- `distribution_channel` — канал дистрибуции заказа;
- `is_canceled` — отмена заказа;
- `is_repeated_guest` — признак того, что гость бронирует номер второй раз;
- `lead_time` — количество дней между датой бронирования и датой прибытия;
- `meal` — опции заказа:
  - `SC` — нет дополнительных опций;
  - `BB` — включён завтрак;
  - `HB` — включён завтрак и обед;
  - `FB` — включён завтрак, обед и ужин.
- `previous_bookings_not_canceled` — количество подтверждённых заказов у клиента;
- `previous_cancellations` — количество отменённых заказов у клиента;
- `required_car_parking_spaces` — необходимость места для автомобиля;
- `reserved_room_type` — тип забронированной комнаты;
- `stays_in_weekend_nights` — количество ночей в выходные дни;
- `stays_in_week_nights` — количество ночей в будние дни;
- `total_nights` — общее количество ночей;
- `total_of_special_requests` — количество специальных отметок.

## Как будут проверять мой проект

Ваш проект будут оценивать по конкретным критериям. Прежде чем решать кейс, внимательно изучите их.

На что обращают внимание при проверке проектов:

- насколько хорошо вы подготавливаете и анализируете данные;
- какие модели строите и как проверяете их качество;
- все ли шаги выполняете по инструкции и оставляете ли к ним комментарии;
- следите ли вы за структурой проекта;
- поддерживаете ли аккуратность кода и не дублируете ли его;
- как вы оформляете блокнот в Jupyter Notebook;
- какие выводы делаете.

Всё необходимое для того, чтобы выполнить проект, есть в темах, которые вы прошли.