Project Description

The customer of this study is the hotel chain "As a guest".

To attract customers, this hotel chain has added the ability to book a room without prepayment to its website. However, if the customer canceled the booking, then the company suffered a loss. For example, hotel employees could buy groceries for the guest's arrival or simply not have time to find another client.

To solve this problem, you need to develop a system that predicts reservation abandonment. If the model shows that the reservation will be canceled, then the client is asked to pay a deposit. The amount of the deposit is 80% of the room rate for one day and the cost of one-time cleaning. The money will be debited from the client's account if he still cancels the reservation.

Business Metrics and Other Data

The main business metric for any hotel chain is its profit. The hotel's profit is the difference between the cost of the room for all nights and the cost of service: both in the preparation of the room and in the stay of the guest.

The hotel has several types of rooms. Depending on the type of room, the rate is for one night. There are also cleaning costs. If the client has rented a room for a long time, then they are cleaned every two days.

Hotel room rates:

- Category A: 1,000 per night, 400 one-time service;
- Category B: 800 per night, 350 per day;
- Category C: 600 per night, 350 per day;
- Category D: 550 per night, 150 one-time service;
- Category E: 500 per night, 150 per night;
- Category F: 450 per night, 150 per day;
- Category G: 350 per night, 150 per day.

Seasonal coefficients are used in the hotel's pricing policy: in spring and autumn, prices increase by 20%, in summer - by 40%.

The hotel's losses in case of cancellation of a room reservation are the cost of one cleaning and one night, taking into account the seasonal coefficient.

A budget of 400,000 rubles has been set for the development of the forecasting system. At the same time, it should be taken into account that the implementation of the model should pay off during the test period. Development costs should be less than the revenue that the system will bring to the company.

Execution Instructions

Step 1. Open the data files

File paths:

`/datasets/hotel_train.csv` - Data to train the model.

`/datasets/hotel_test.csv` - Data for model testing.

Step 2. Pre-processing and exploratory data analysis

- Review the data and correct it if necessary.
- Conduct an exploratory analysis of the data presented. Describe the features found: omissions, outliers.

Step 3. Compute a business metric

Estimate the hotel's profit without implementing deposits.

Step 4. Developing an ML model

- Train different models and evaluate their quality by cross-validation. Choose the best model and test it on a test sample. Describe the findings.
- Select the metric you want to train.
- Estimate the profit that the selected model will bring in a year.

Step 5. Identify the signs of an "untrustworthy" customer

Based on exploratory data analysis, describe a customer who is prone to cancellation.

Step 6. Write a general conclusion

Describe the model that solves the problem with the greatest benefit to the business. Give recommendations to the business based on the results of the work done.

Typography

Complete the job in Jupyter Notebook. Fill in the program code in the cells of the type, and the text explanations in the cells of the type. Use formatting and headings. `code` `markdown`

Description of the data

The tables contain the same columns: `hotel_train` `hotel_test`

- `id` - Record number
- `adults` – number of adult guests;
- `arrival_date_year` – year of arrival;
- `arrival_date_month` – month of arrival;
- `arrival_date_week_number` - Check-in week;
- `arrival_date_day_of_month` – day of arrival;
- `babies` – the number of infants;
- `booking_changes` — the number of changes to the order parameters;
- `children` – number of children from 3 to 14 years old;
- `country` – citizenship of the guest;
- `customer_type` – Type of customer:
  - `Contract` – an agreement with a legal entity;
  - `Group` - group check-in;
  - `Transient` – is not related to a contract or a group check-in;
  - `Transient-party` – Not related to a contract or group check-in, but related to a booking type. `Transient`
- `days_in_waiting_list` — how many days the order has been waiting for confirmation;

- `distribution_channel` — order distribution channel;
- `is_canceled` – cancellation of the order;
- `is_repeated_guest` - A sign that the guest is booking a room for the second time.
- `lead_time` – the number of days between the date of booking and the date of arrival;
- `meal` – Order options:
  - `SC` – there are no additional options;
  - `BB` – breakfast is included;
  - `HB` – breakfast and lunch included;
  - `FB` – Breakfast, lunch and dinner are included.
- `previous_bookings_not_canceled` — the number of confirmed orders from the client;
- `previous_cancellations` — the number of canceled orders from the client;
- `required_car_parking_spaces` – the need for space for the car;
- `reserved_room_type` – the type of room booked;
- `stays_in_weekend_nights` – the number of nights on weekends;
- `stays_in_week_nights` – number of nights on weekdays;
- `total_nights` – total number of nights;
- `total_of_special_requests` is the number of special marks.

How will my project be checked?

Your project will be evaluated according to specific criteria. Before solving a case, study them carefully.

What you pay attention to when checking projects:

- how well you prepare and analyze data;
- what models you build and how do you check their quality;
- whether you follow the instructions and leave comments on them;
- whether you keep track of the project structure;
- Do you keep the code neat and do not duplicate it?

- How you design a notebook in Jupyter Notebook

- what conclusions do you draw?

Everything you need to complete the project is in the topics you've completed.

Success!