

# Binary Classification with a Tabular Reservation Cancellation Dataset

Авторы:

Трифонов Степан  
Кривенко Михаил

The Kaggle logo, featuring the word "kaggle" in a white, lowercase, sans-serif font, centered within a solid blue rectangular background.



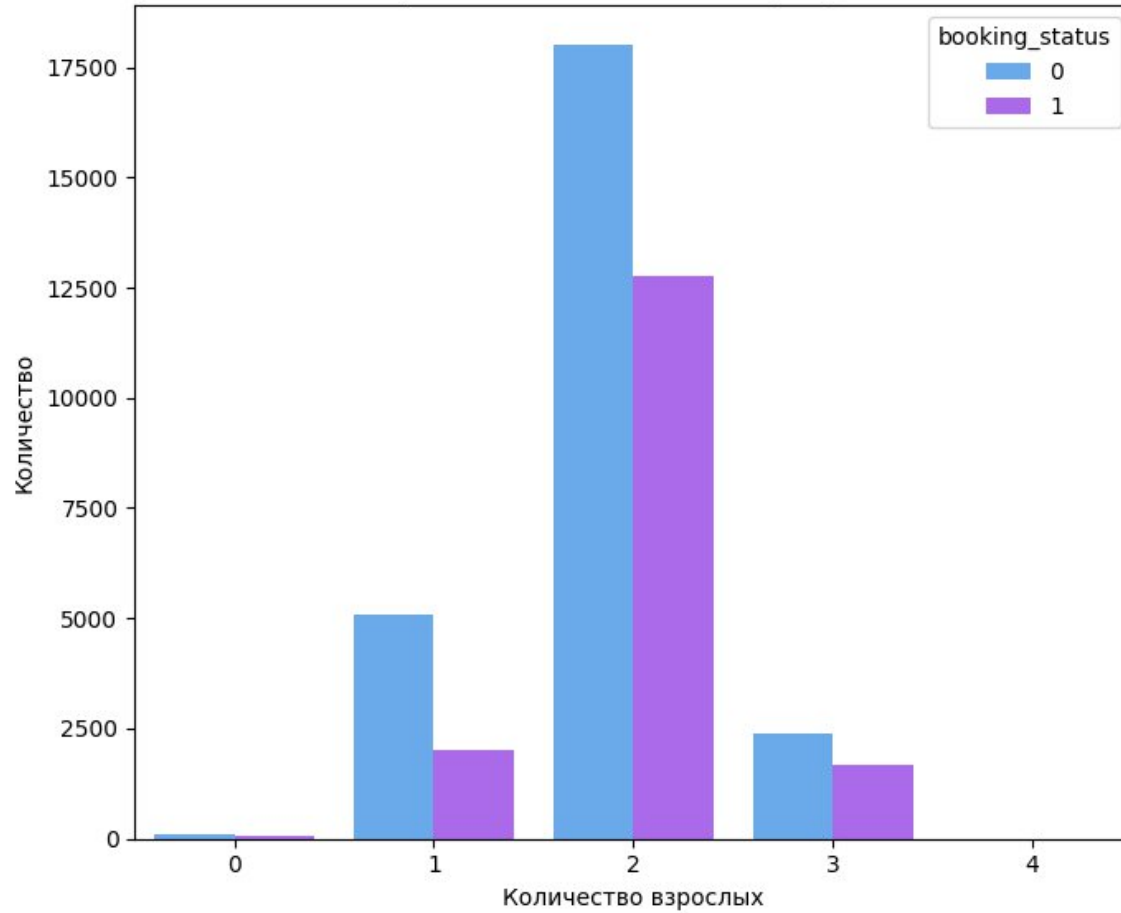
# Description of the data and the task

```
data.info()
# Column Non-Null Count Dtype
0 id 42100 non-null int64
1 no_of_adults 42100 non-null int64
2 no_of_children 42100 non-null int64
3 no_of_weekend_nights 42100 non-null int64
4 no_of_week_nights 42100 non-null int64
5 type_of_meal_plan 42100 non-null int64
6 required_car_parking_space 42100 non-null int64
7 room_type_reserved 42100 non-null int64
8 lead_time 42100 non-null int64
9 arrival_year 42100 non-null int64
10 arrival_month 42100 non-null int64
11 arrival_date 42100 non-null int64
12 market_segment_type 42100 non-null int64
13 repeated_guest 42100 non-null int64
14 no_of_previous_cancellations 42100 non-null int64
15 no_of_previous_bookings_not_canceled 42100 non-null int64
16 avg_price_per_room 42100 non-null float64
17 no_of_special_requests 42100 non-null int64
18 booking_status 42100 non-null int64
dtypes: float64(1), int64(18)
memory usage: 6.1 MB
```

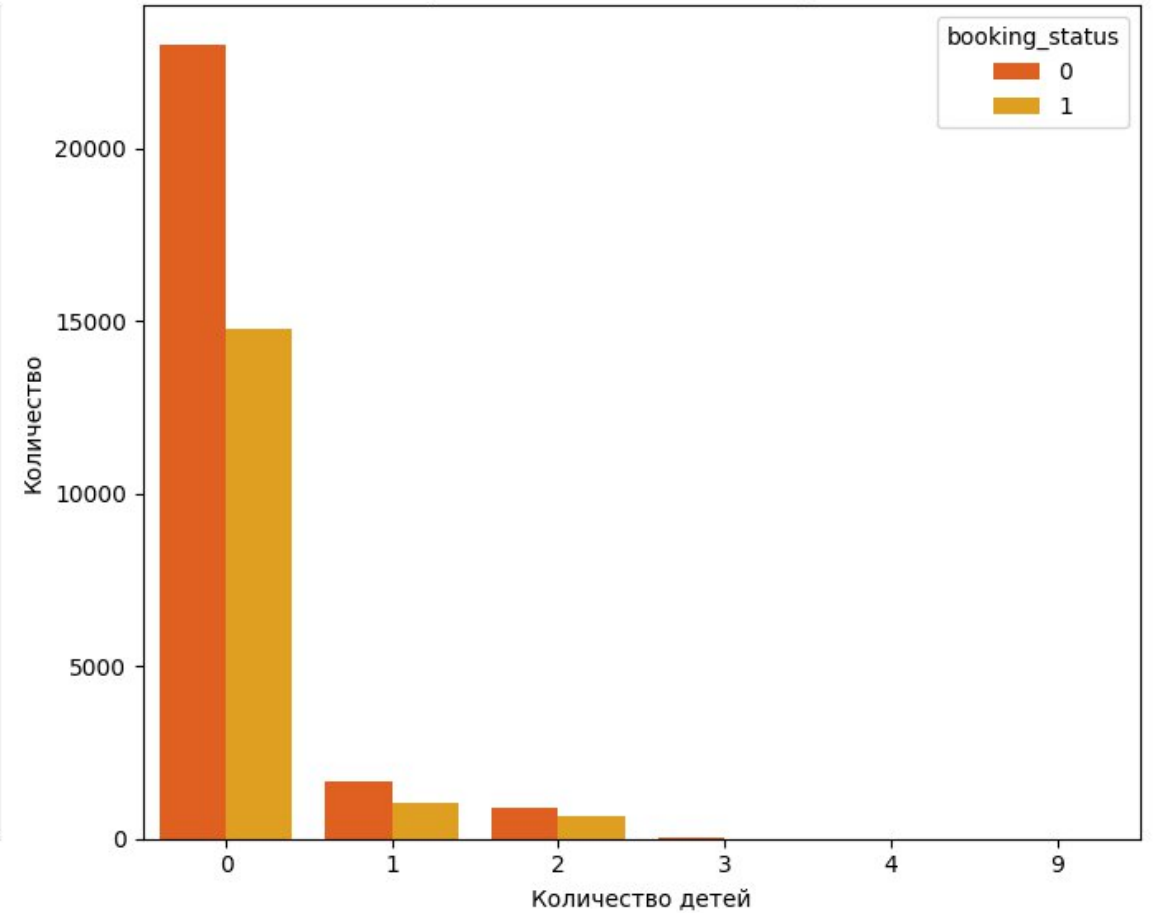
- The dataset for this competition (both train and test) was generated from a deep learning model trained on the **Reservation Cancellation Dataset**
- It is necessary to predict the target variable - whether the reservation was canceled

# Data visualization

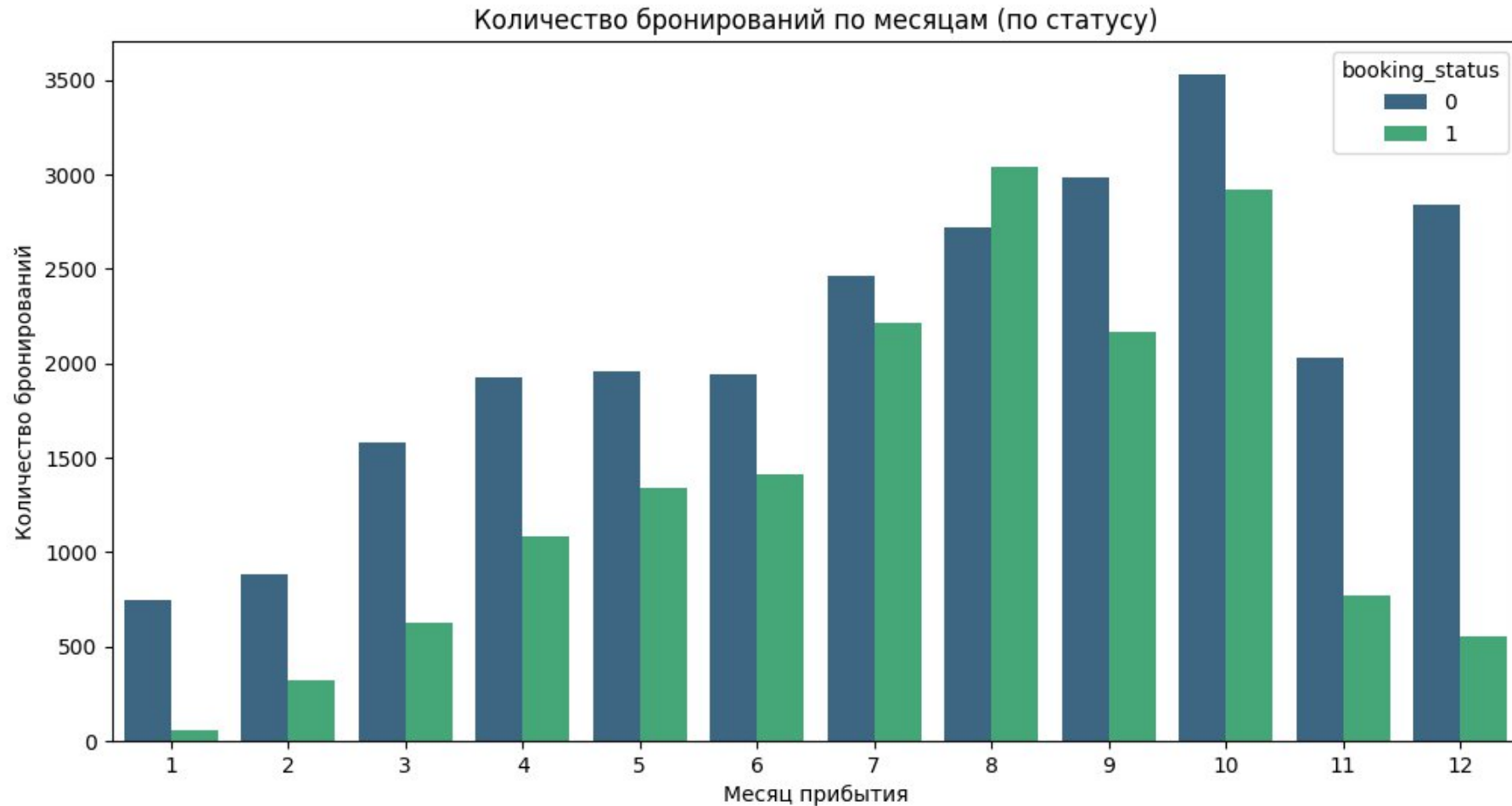
Распределение отмен по числу взрослых



Распределение отмен по числу детей

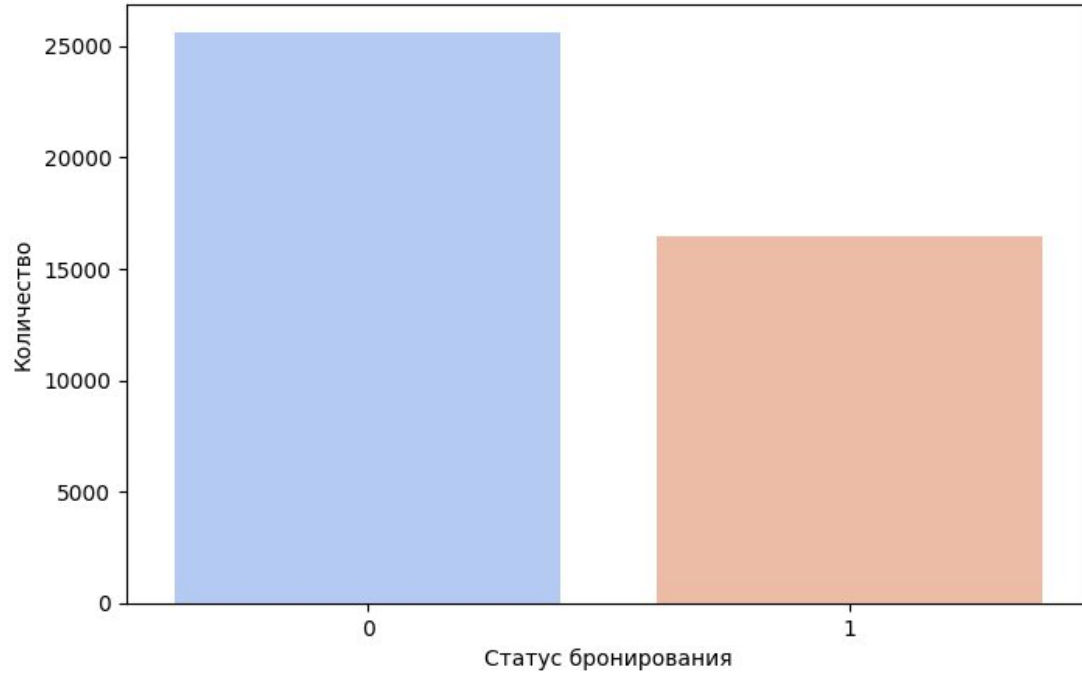


# Data visualization

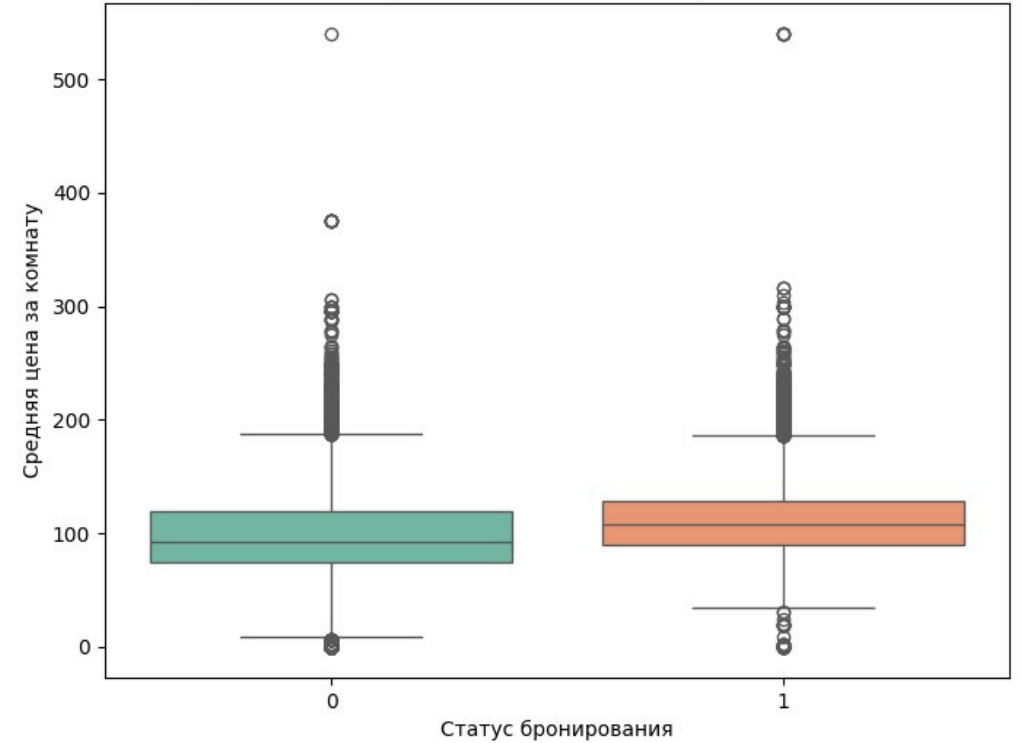


# Data visualization

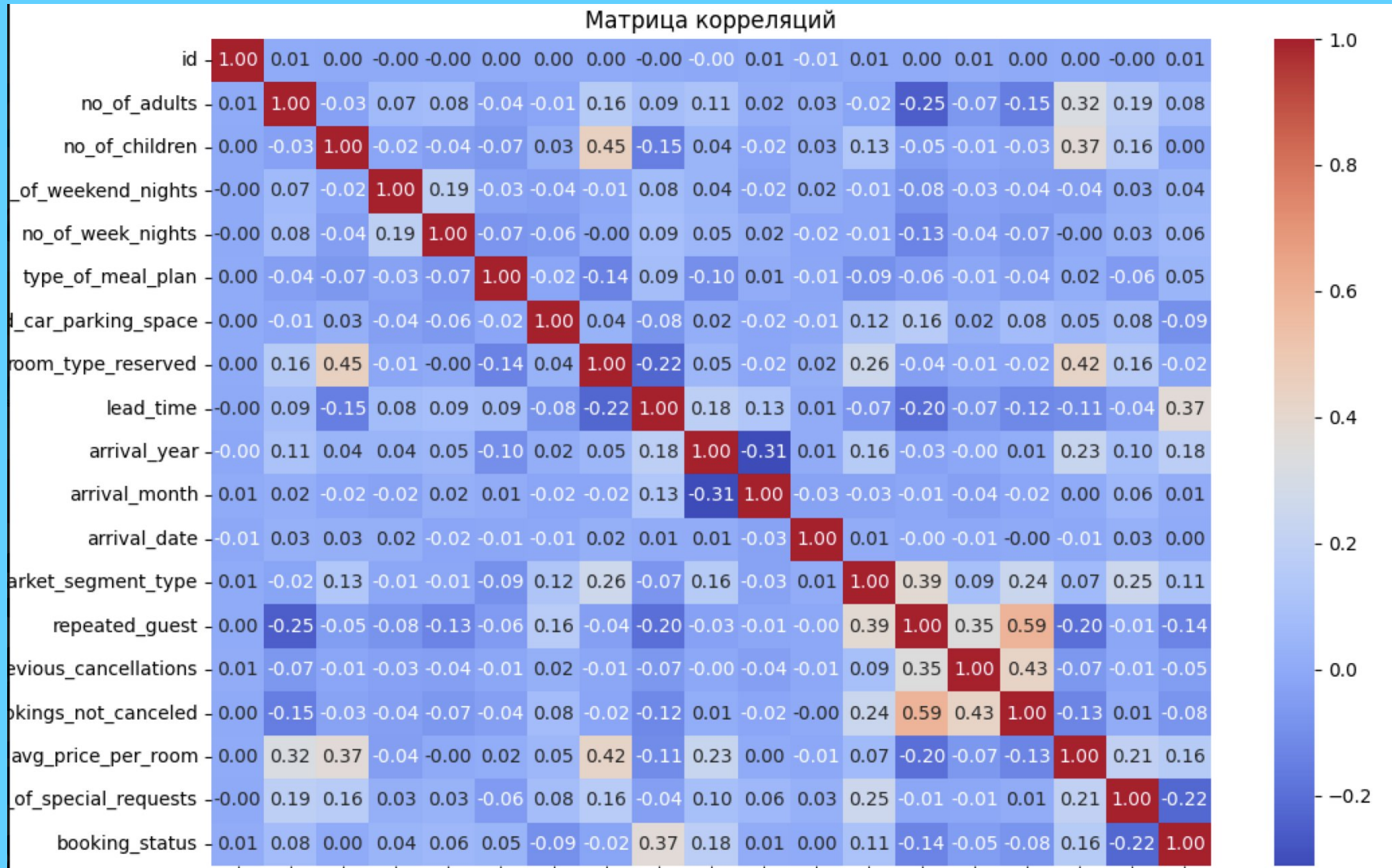
Распределение отмен бронирования



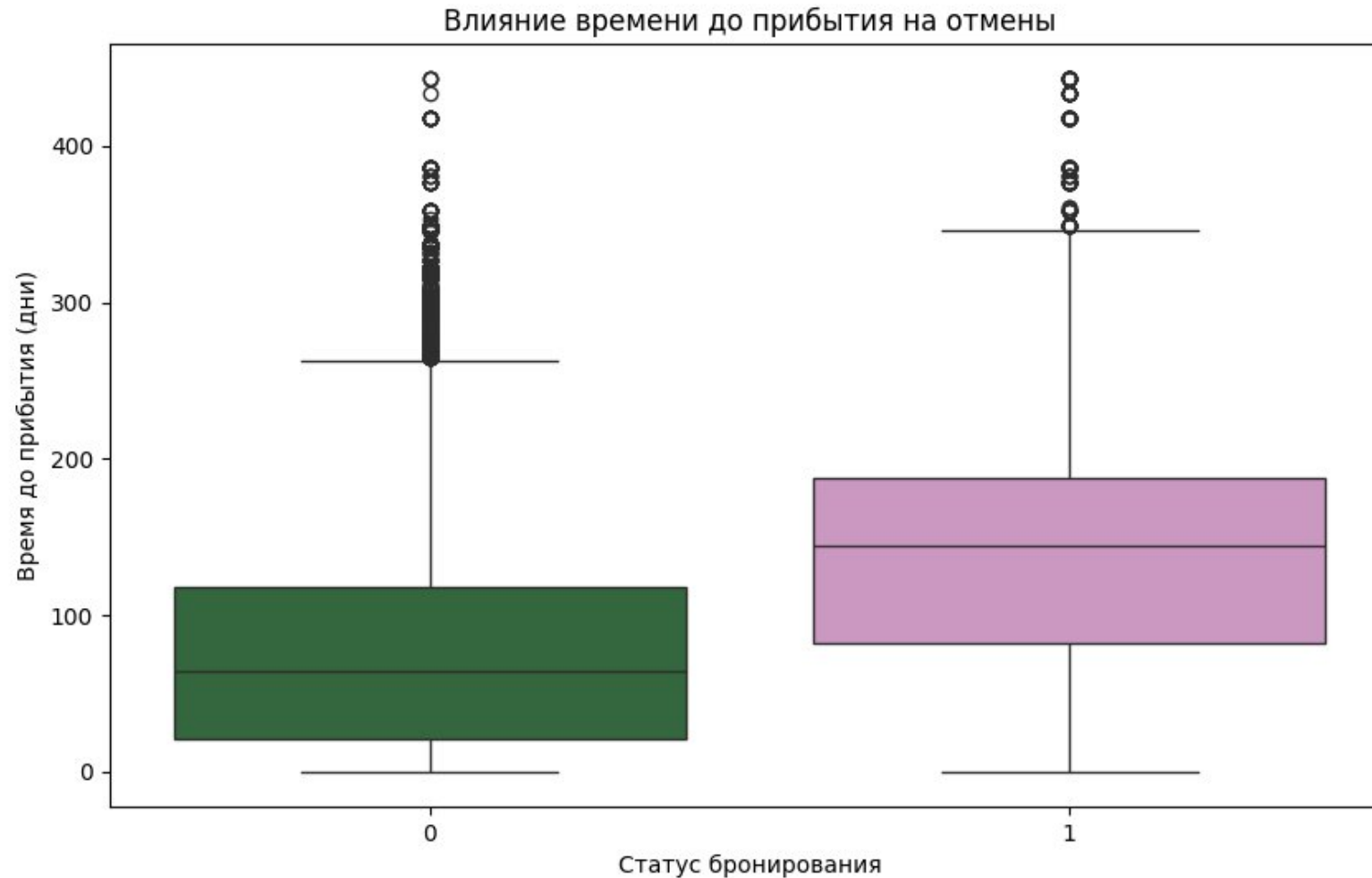
Цена за комнату в зависимости от статуса бронирования



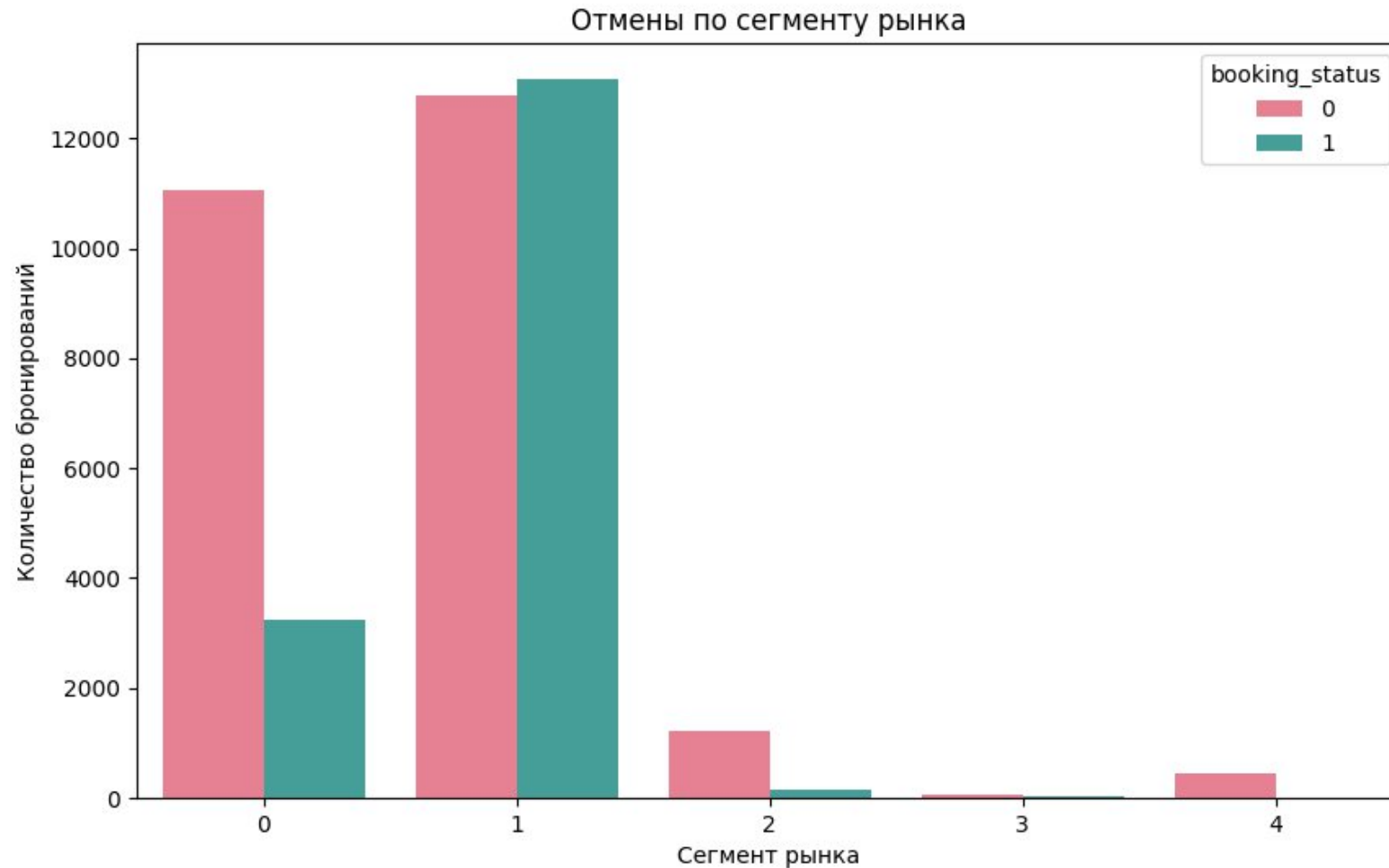
# Data visualization



# Data visualization



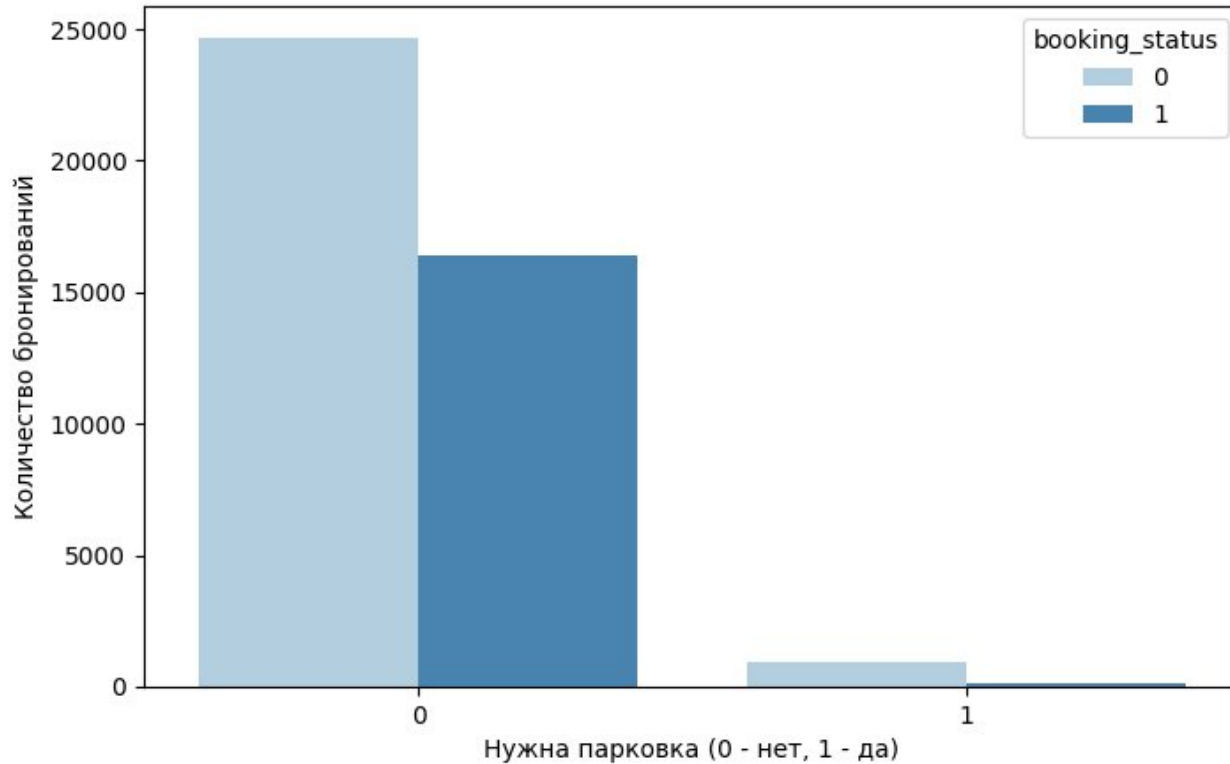
# Data visualization



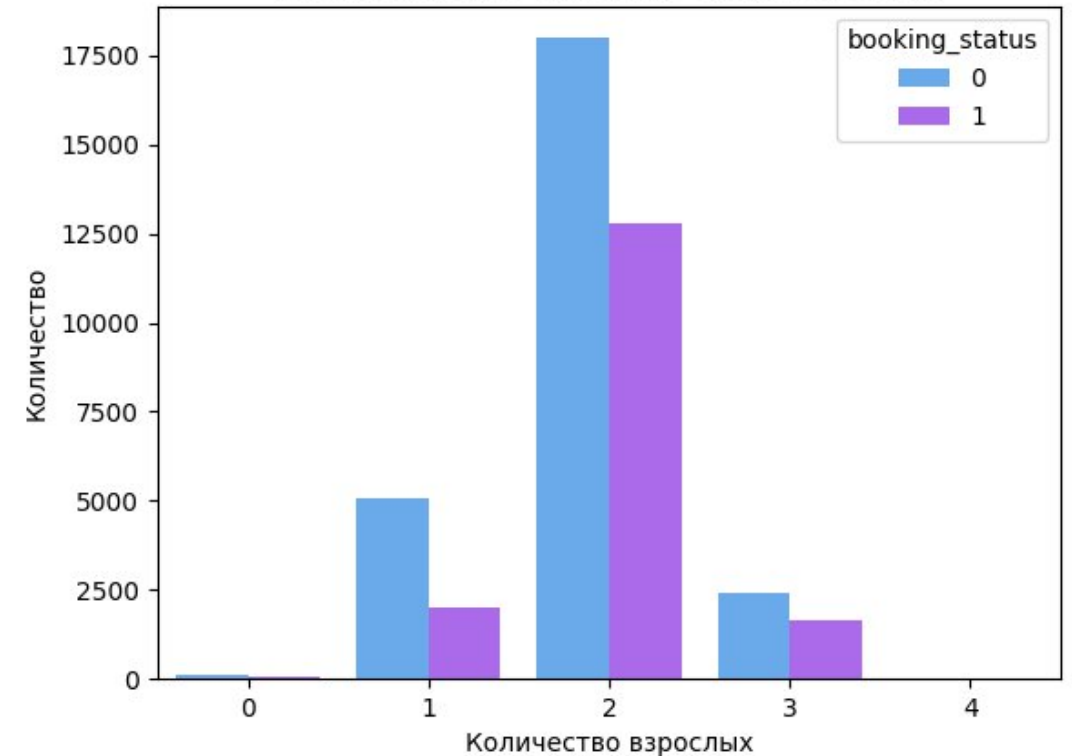


# Data visualization

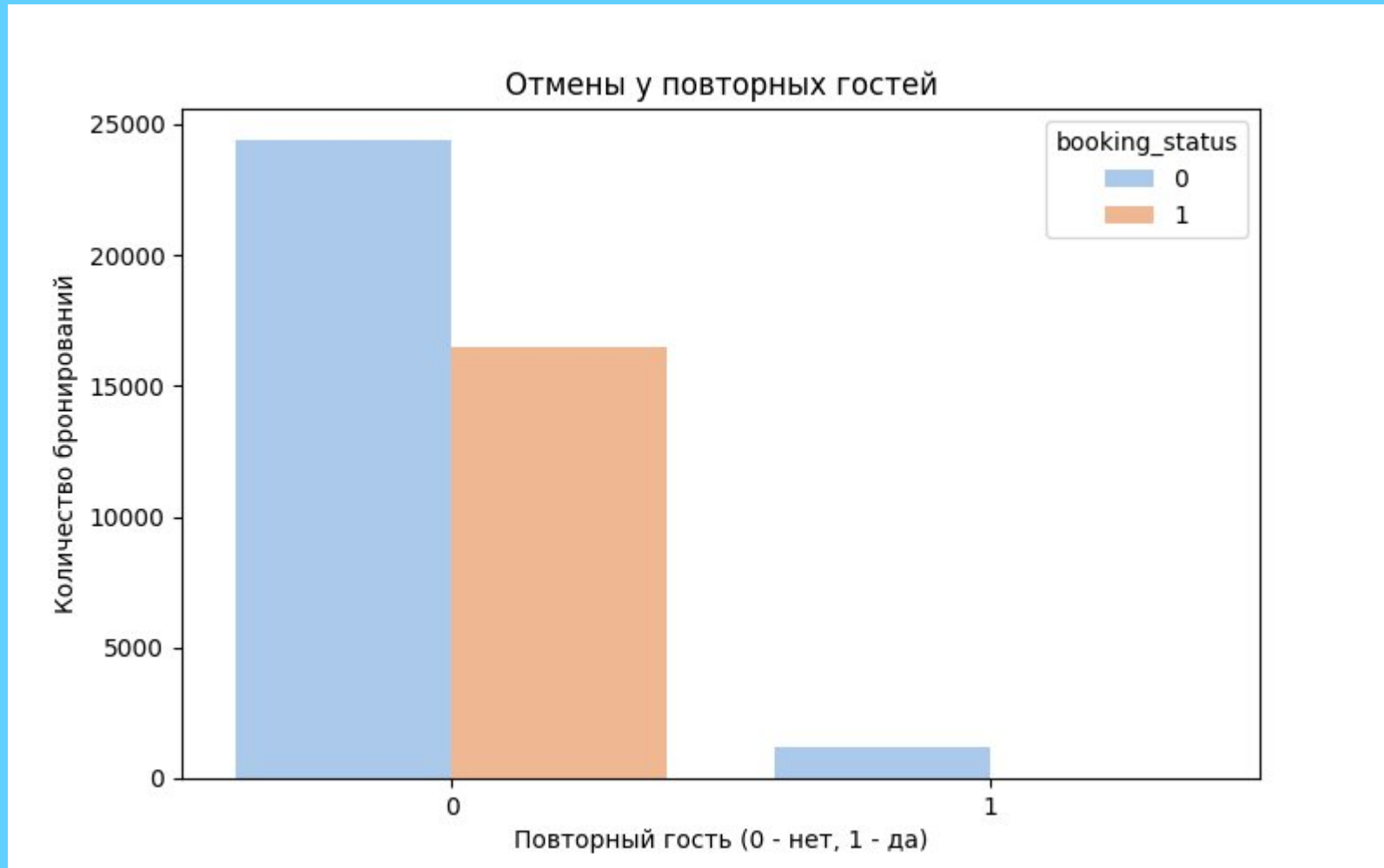
Отмены в зависимости от необходимости парковочного места



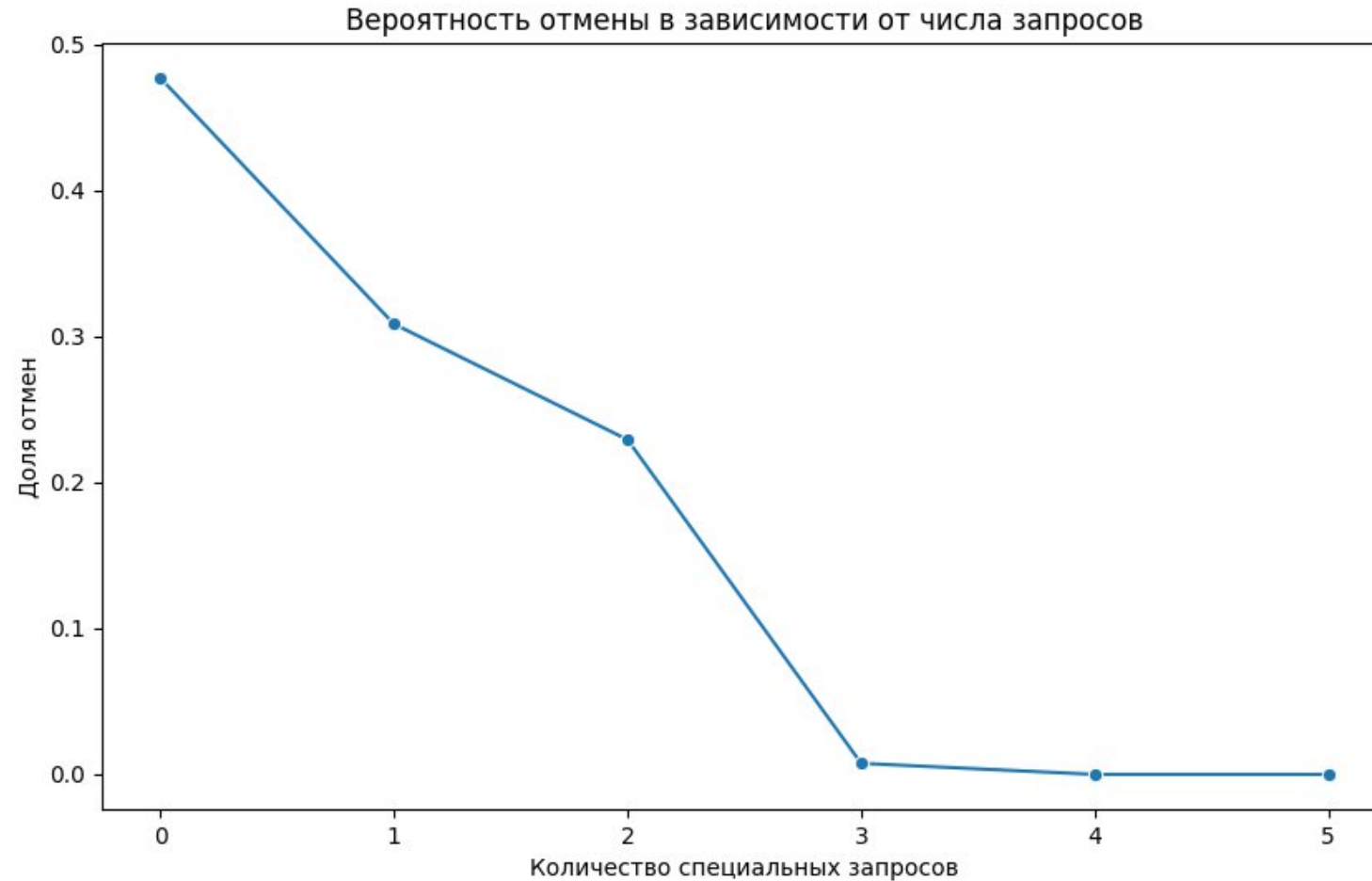
Распределение отмен по числу взрослых



# Data visualization



# Data visualization



# Data preparation and Train

- Adding polynomial features
- Detecting outliers
- Detection of categorical features

## 1. **LightGBM (LGBMClassifier):**

- Gradient boosting on root trees, optimized for speed and performance.

## 2. **XGBoost (XGBClassifier):**

- Another popular implementation of gradient boosting, known for its efficiency and accuracy in classification and regression tasks.

## 3. **CatBoost (CatBoostClassifier):**

- A specialized library for working with categorical invitations; gradient boosting is used.


Stacking/Blending. Meta-training that uses the results of the predictions of three models to improve the final predictive ability. This is achieved by optimizing the weights of the models using the Nelder-Mead method to maximize the ROC AUC estimate.

Use `early_stopping_rounds`



# Result


× Submission Details


**submission\_final.csv**  
Complete (after deadline) · 1m ago

Score: 0.92804

Private score: 0.91960

UPLOADED FILES

 submission\_final.csv (729 KiB)



DESCRIPTION

Enter a description

0 / 500

SELECT FOR FINAL SCORE

☐ Select this submission to be scored for your final leaderboard score

Cancel

Save

Private Score ⓘ	Public Score ⓘ
0.91960	0.92804

75 private perc

0,9036

85 private perc

0,91136

95 private perc

0,91795

kaggle