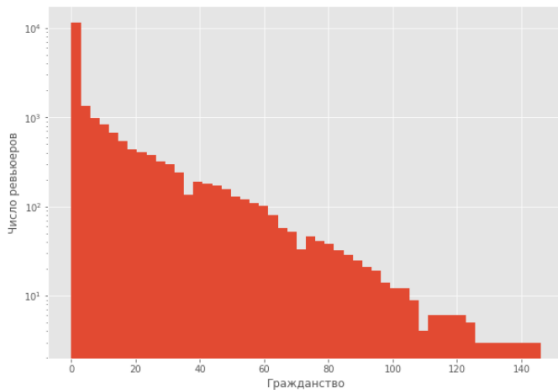


Датасет

- ▶ 500k отзывов на отели;
- ▶ Для каждого отзыва дана его оценка, положительный и отрицательный отзывы;
- ▶ Есть информация об отелях и гражданстве заполняющих;
- ▶ Всего 17 признаков.

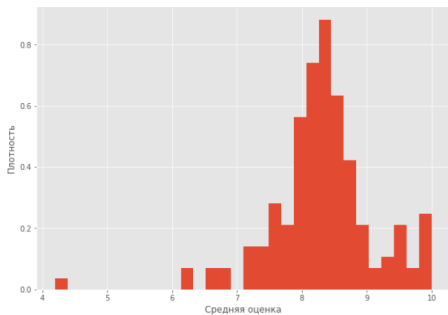
Информация о гражданстве



Number		Country
0	1	Andorra
1	1	Antigua Barbuda
2	1	Benin
3	1	Botswana
4	1	Cambodia

Number		Country
0	413	United Arab Emirates
1	591	Ireland
2	792	Australia
3	1382	United States of America
4	9498	United Kingdom

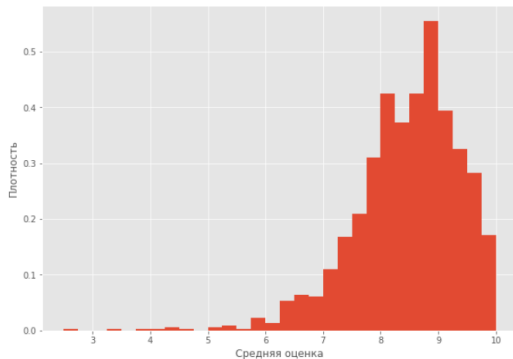
Гражданство и оценка отелей



	country	mean_review	count
0	Zambia	6.300	1
1	Cambodia	6.300	1
2	Syria	6.700	1
3	Algeria	6.875	8
4	Macau	4.200	1
5	St Maarten	6.700	2
6	Armenia	6.800	7

	country	mean_review	count
0	Equatorial Guinea	10.0	1
1	Ethiopia	9.6	2
2	Faroe Islands	9.6	1
3	Andorra	9.6	1
4	Puerto Rico	9.9	4
5	Botswana	10.0	1
6	Paraguay	10.0	1
7	Eritrea	9.6	1
8	Guatemala	9.8	2
9	Costa Rica	9.8	2
10	Venezuela	10.0	2

Название отелей и их оценки



	Hotel name	mean_review	count
0	ABaC Restaurant Hotel Barcelona GL Monumento	2.500000	1
1	Roomz Vienna	5.850000	2
2	Savoy Hotel Amsterdam	5.785714	14
3	HCC St Moritz	5.966667	3

Значимость признаков

- ▶ Берём числовые признаки;
- ▶ Строим XGBoost регрессию на признак оуценки ревьюера;
- ▶ Смотрим на значимость признаков.

	0	1
0	Additional_Number_of_Scoring	0.030939
1	Average_Score	0.218641
2	Review_Total_Negative_Word_Counts	0.449774
3	Total_Number_of_Reviews	0.007655
4	Review_Total_Positive_Word_Counts	0.288506
5	Total_Number_of_Reviews_Reviewer_Has_Given	0.004485
6	days_since_review	0.000000

Корреляция числовых признаков

