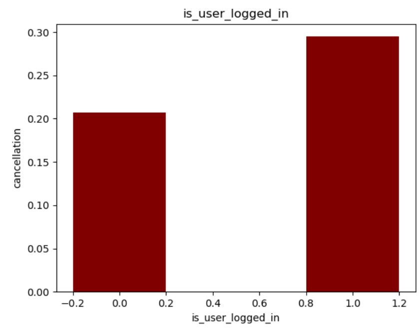
michaelmerzin ,323913277 | sean.cherny,213716608| ofir.rosen, 325538189| yonmiroshnik, 318209939

1.2.3 Churn prediction Model

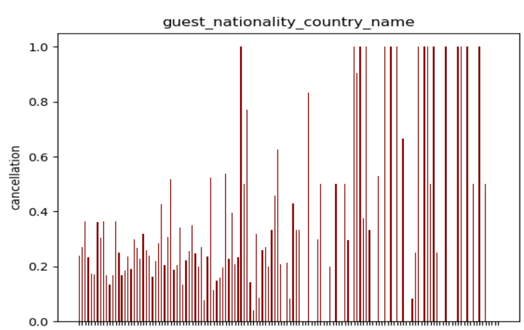
- During our analysis in the "Eda.py" code, we discovered two significant features that have a notable impact on reservation cancellations.

1. The first significant feature is "is_user_logged_in." We observed that the cancellation rate for not logged in users is 50% higher compared to logged in users. This finding suggests that users who are logged into their accounts have a lower tendency to cancel their reservations.

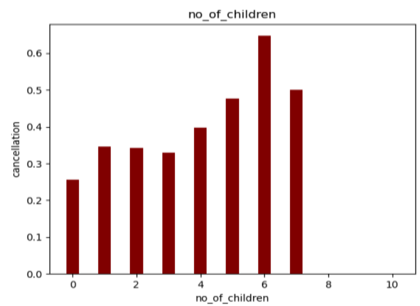


2. The second significant feature we identified is "guest_nationality_country_name" We noticed that certain countries exhibited high cancellation rates. For example, Russia had a cancellation percentage of 0.51, Pakistan had 0.52, and Mauritius had 0.76. On the other hand, we found that Thailand, despite having 4,306 reservations in total, had a cancellation percentage of only 0.17, while Malaysia, with 7,368 reservations, had a cancellation percentage of only 0.16. These observations indicate that the nationality of a guest may influence their likelihood of canceling a reservation.

   However, it's important to note that the cancellation percentages we observed were based on a relatively low number of reservations for each country. Therefore, drawing definitive conclusions about the impact of nationality on cancellation rates requires further investigation and analysis.
   We can see that the correlation between cancellation rates and nationality is very high.



3. The third significant feature we discovered is the number of children (“no_of_children”). Upon analyzing the data, we observed a trend where an increase in the number of children corresponded to a higher cancellation rate. While this correlation may seem intriguing, it's essential to approach it with caution and consider alternative explanations. In conclusion, to take your kids while travel , or don’t make kids .



<u>Important marks</u>

- During the EDA (Exploratory Data Analysis) process and data analysis, several important marks and understandings emerged. One crucial observation was the significance of normalizing the size of bars when creating visualizations for different features. Failing to normalize the bars can lead to misinterpretation and erroneous conclusions.
- Let's consider an example to illustrate this point. Suppose we have a feature representing the number of hotels in different star categories, such as 1-star, 2-star, 3-star, and so on. Initially, when plotting a bar chart without normalizing the sizes, it might appear that the number of hotels in a specific star category is a highly significant feature. However, this can be misleading.
- If think that this is a significant feature , its can ruin your predict by adding a noise