Hackathon 2023 - 1.2.4

Amit Benbenishti, Adi Shnaidman, Shoham Mazuz, Omer Ventura

Introduction

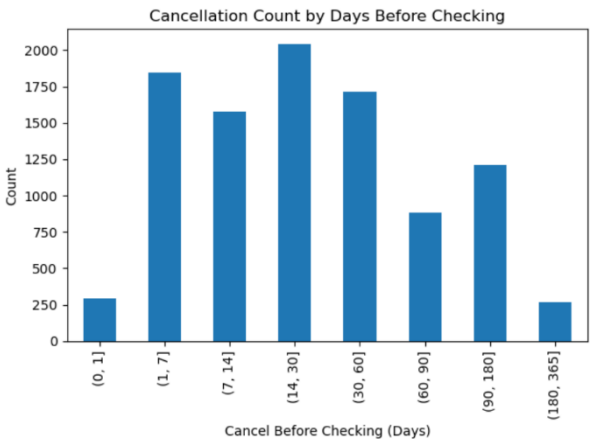
In previous questions, we developed models and conducted data analysis to understand cancellation patterns in the hotel dataset provided by Agoda. In Question 1, we identified relevant features strongly correlated with the risk of cancellation. In Question 2, we estimated the cost of a vacation using metadata such as hotel type, location, and star rating.
Data Analysis on Cancellation: Based on our findings, several factors significantly influenced the cancellation risk. These include the time gap between check-in and reservation dates, penalty policies (after detailed analysis), and other relevant factors.

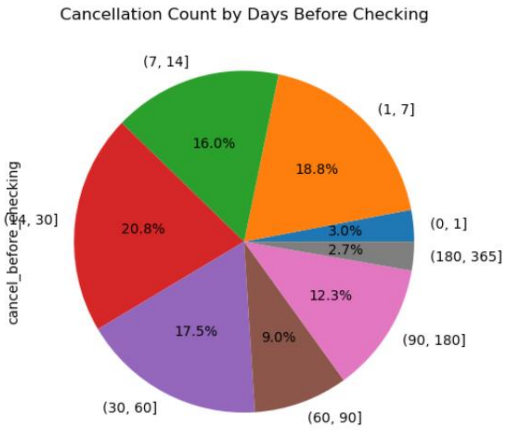
Cancellation Policy Formula

To formulate our cancellation policy, we examined the percentage of cancellations at various time intervals before check-in:

- 1 day before check-in
- 1 week before check-in
- Between 1 week and 2 weeks before check-in
- Between 2 weeks and 1 month before check-in
- Between 1 month and 3 months before check-in
- Between 3 months and half a year before check-in
- Between half a year and a year before check-in

We also analyzed the cumulative percentage of cancellations leading up to each relevant date to understand the overall proportion of cancellations.
The following graphs illustrate our findings:



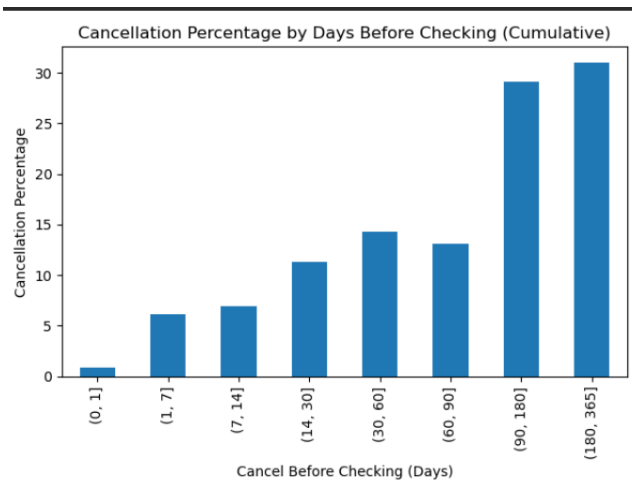


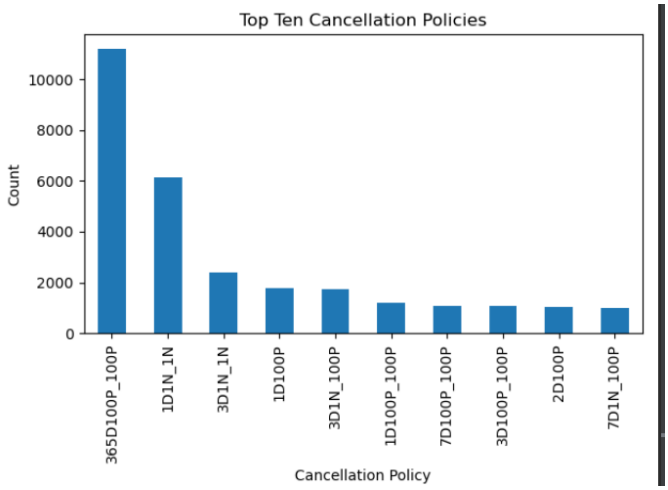
Looking at the graphs, we observe that the majority of cancellations occur between 14 days and 30 days or 1 week before the check-in date. In terms of percentage, the highest cancellation rates are seen for orders that are placed far in advance. This is evident from the graph and further supported by the fact that most penalties are associated with 1-year cancellations.

Furthermore, we notice that the second most common cancellation policy is the 1-day policy. We propose modifying this policy to impose a higher penalty for the time range between 2 weeks and 1 month, as this timeframe exhibits a higher cancellation percentage and is where a larger portion of the population cancels.

Formula

Based on these observations, we propose the following formula for calculating the cancellation penalty:





Cancellation Penalty = (Cancellation Popularity) * (0.7 * Model Suggestion for Cancellation) * (Model Price Suggestion)

- Here's an explanation of the components:
- "Cancellation Popularity" represents the cancellation rate for a specific time period relative to the overall yearly cancellations. It provides a measure of the likelihood of cancellation during that time period.
- 0.7 is the probability for making the right decision, multiplied by the prediction made by our cancellation model. This factor takes into account the accuracy of our model in predicting cancellations.
- Model Price Suggestion is the price recommendation provided by our model. It incorporates various factors such as the number of children, floor level, location, and other relevant metadata to determine an appropriate price for the rooms.

By using this formula, hotels can effectively price their rooms based on metadata alone, utilizing our model's predictions and considering the cancellation rate for different time periods. This approach enables better decision-making and optimization of room prices.