

## Description

### Context

A significant number of hotel bookings are called off due to cancellations or no-shows. The typical reasons for cancellations include change of plans, scheduling conflicts, etc. This is often made easier by the option to do so free of charge or preferably at a low cost which is beneficial to hotel guests but it is a less desirable and possibly revenue-diminishing factor for hotels to deal with. Such losses are particularly high on last-minute cancellations.

The new technologies involving online booking channels have dramatically changed customers' booking possibilities and behavior. This adds a further dimension to the challenge of how hotels handle cancellations, which are no longer limited to traditional booking and guest characteristics.

The cancellation of bookings impact a hotel on various fronts:

1. Loss of resources (revenue) when the hotel cannot resell the room.
2. Additional costs of distribution channels by increasing commissions or paying for publicity to help sell these rooms.
3. Lowering prices last minute, so the hotel can resell a room, resulting in reducing the profit margin.
4. Human resources to make arrangements for the guests.

### Objective

The increasing number of cancellations calls for a Machine Learning based solution that can help in predicting which booking is likely to be canceled. INN Hotels Group has a chain of hotels in Portugal, they are facing problems with the high number of booking cancellations and have reached out to your firm for data-driven solutions. You as a data scientist have to analyze the data provided to find which factors have a high influence on booking cancellations, build a predictive model that can predict which booking is going to be canceled in advance, and help in formulating profitable policies for cancellations and refunds.

### Data Description

The data contains the different attributes of customers' booking details. The detailed data dictionary is given below.

### Data Dictionary

- Booking\_ID: the unique identifier of each booking
- no\_of\_adults: Number of adults
- no\_of\_children: Number of Children
- no\_of\_weekend\_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- no\_of\_week\_nights: Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel
- type\_of\_meal\_plan: Type of meal plan booked by the customer:
  - Not Selected – No meal plan selected
  - Meal Plan 1 – Breakfast
  - Meal Plan 2 – Half board (breakfast and one other meal)
  - Meal Plan 3 – Full board (breakfast, lunch, and dinner)

- **required\_car\_parking\_space:** Does the customer require a car parking space? (0 - No, 1- Yes)
- **room\_type\_reserved:** Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels Group
- **lead\_time:** Number of days between the date of booking and the arrival date
- **arrival\_year:** Year of arrival date
- **arrival\_month:** Month of arrival date
- **arrival\_date:** Date of the month
- **market\_segment\_type:** Market segment designation.
- **repeated\_guest:** Is the customer a repeated guest? (0 - No, 1- Yes)
- **no\_of\_previous\_cancellations:** Number of previous bookings that were canceled by the customer prior to the current booking
- **no\_of\_previous\_bookings\_not\_canceled:** Number of previous bookings not canceled by the customer prior to the current booking
- **avg\_price\_per\_room:** Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
- **no\_of\_special\_requests:** Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
- **booking\_status:** Flag indicating if the booking was canceled or not.

## Submission Guidelines

1. There are two ways to work on this project:

**i. Full-code way:** The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

**ii. Low-code way:** The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

Submission type	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	Submission Format
Full-code	Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch	Perform exploratory data analysis to identify insights and recommendations for the problem	Focus on code writing: 10-20% grading on the quality of the final code submitted	Solution notebook from the full-code template submitted in .html format	.html

Low-code	Learners who aspire to be in managerial roles in the future- focussed on solution review, interpretation, recommendations, and communicating with business		Focus on business presentation : 10-20% grading on the quality of the final business presentation submitted	Business presentation in .pdf format with problem definition, insights, and recommendations	.pdf
----------	--	--	---	---	------

Please follow the below steps to complete the assessment. **Kindly note that if you submit a presentation along with the notebook, ONLY the presentation will be evaluated.** Please make sure that all the sections mentioned in the rubric have been covered in your submission.

*i. Full-code version*

- Download the full-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Clearly write down insights and recommendations for the business problems in the comments.
- Submit only the solution notebook prepared from the learner notebook [format: .html]

*ii. Low-code version*

- Download the low-code version of the learner notebook.
- Follow the instructions provided in the notebook to complete the project.
- Prepare a business presentation with insights and recommendations to the business problem.
- Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

- it is submitted post-deadline, or,
- more than 1 file is submitted.

*Best Practices for Full-code submissions*

- The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
- The notebook should be run from start to finish in a sequential manner before submission.

- It is important to remove all warnings and errors before submission.
- The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
- Please refer to the FAQ page for common project-related queries.

#### *Best Practices for Low-code submissions*

- The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
- The key points in the presentation should be the following:
  - Business Overview of the problem and solution approach
  - Key findings and insights which can drive business decisions
  - Business recommendations
  - Focus on explaining the key takeaways in an easy-to-understand manner.
  - The inclusion of the potential benefits of implementing the solution will give you the edge.
- Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
- The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
- Please refer to the FAQ page for common project-related queries.

Happy Learning!

## Scoring guide (Rubric) - INN Hotels Rubric

Criteria	Points
<p>Define the problem and perform an Exploratory Data Analysis</p> <ul style="list-style-type: none"> <li>- Problem definition - Univariate analysis - Bivariate analysis - Use appropriate visualizations to identify the patterns and insights - Key meaningful observations on individual variables and the relationship between variables</li> </ul>	12
<p>Data Pre-processing</p> <p>Prepare the data for modelling: - Missing value Treatment (if needed) - Outlier Detection(treat, if needed) - Feature Engineering (if needed) - Data split</p>	4
<p>Model building - Logistic Regression</p> <ul style="list-style-type: none"> <li>- Build the Logistic Regression model using statsmodels library and comment on the model statistics - Check assumptions of Logistic Regression - Provide interpretations based on coefficients obtained from the logistic regression model - Comment on model performance</li> </ul>	9
<p>Model Performance evaluation and improvement - Logistic Regression</p> <ul style="list-style-type: none"> <li>- Try and improve the model performance by changing the classification threshold - Comment on model performance after changing the threshold</li> </ul>	6
<p>Model building - Decision Tree</p> <ul style="list-style-type: none"> <li>- Build the model and comment on the model performance.</li> </ul>	5
<p>Model Performance evaluation and improvement - Decision Tree</p> <ul style="list-style-type: none"> <li>- Try and improve the model performance by pruning - Comment on model performance - Find the decision rules and check feature importance</li> </ul>	10
<p>Actionable Insights &amp; Recommendations</p> <ul style="list-style-type: none"> <li>- Compare decision tree and Logistic regression models - Conclude with the key takeaways for the business</li> </ul>	6

Criteria	Points
Presentation / Notebook - Overall quality - Structure and flow - Crispness - Visual appeal - Conclusion and Business Recommendations OR - Structure and flow - Well commented code - Conclusion and Business Recommendations	8
	Points 60