

The goal of any data mining project is to extract knowledge and patterns from data using a wide range of methods and techniques. Within big datasets, important relationships can be uncovered that give insights and can be used to make predictions that impact the business.

This project aims to develop models that are capable of predicting whether or not clients will cancel their hotel bookings. For this purpose, data was collected from several hotel clients, containing samples of clients who cancelled their bookings and clients who didn't.

When customers cancel their bookings, hotels incur losses, resulting in annual losses. In this project, we will analyse various factors, including the booking lead time, customer type, market segment, and more, to forecast whether an individual is likely to cancel their booking.

Using the algorithms you've learned, the aim is to develop predictive models based on the data provided. The project must follow the CRISP-DM methodology and include code for its phases in the Python language, annotated in Markdown.

- data exploration and preparation;
- data pre-processing;
- creation of models using data mining algorithms;
- evaluation of the models created.

The project must be submitted with a report describing, in as much detail as possible, the process you followed to obtain your solutions. The report must include the data mining goals, the most relevant data graphical figures and their interpretation, an explanation of the cleaning and pre-processing of data performed, an interpretation or evaluation of the models created, and commitments assumed in their development.

The work will be mainly assessed by the quality of the data analysis process, followed by the conclusions reached, policies or actions proposed, and the accuracy of the models. But more important than the accuracy of the models is the description of the analysis process and conclusions extracted from the data.

### Deadline and submission instructions

- The project should be submitted to Moodle in the discipline area by **24:00 on November 3, 2024**. After this date, we will penalize the notes by 10%, and we won't accept any projects after November 5th.
- Place the code and report in a ZIP file named MINDD-NumberX-NumberY.zip, where NumberX represents the number of group elements.
- The project will be presented and evaluated individually with the PL teacher during the 8th and 9th weeks, with previous schedule.