

Project Proposal

Question/need:

- What is the framing question of your analysis, or the purpose of the model/system you plan to build?

I want to build a model that predict in advance whether a hotel customer will cancel his booking or not. Also , I want to identify the most important factors that can lead to booking cancellation.

- Who benefits from exploring this question or building this model/system?

Knowing if a customer is more likely to cancel his booking or not helps the booking company and hotels plan for cancellation and refund policies as well as targeting customers with offers and discounts.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data?

I am planning to use “Hotel booking demand” dataset available on Kaggle.
<https://www.kaggle.com/jessemostipak/hotel-booking-demand>

- What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

The individual sample/unit of analysis in this project is a single booking made by a hotel customer. There are 32 features related to the booking, including booking date, number of adults, deposit type and previous cancellations.

- If modeling, what will you predict as your target?

I will predict if the booking was canceled or confirmed.

Tools:

- How do you intend to meet the tools requirement of the project?

I am planning to use scikit-learn for modeling and seaborn and plotly for visualization

Are you planning in advance to need or use additional tools beyond those required? No

MVP Goal:

- What would a [minimum viable product \(MVP\)](#) look like for this project?

A Jupiter notebook that trains and evaluates a regression model that predict booking cancellation and a heatmap of the most important features.