Safa Alsafari 10-7-2021

Project Proposal

Hotels booking cancelation make it harder to accurately forecast and optimize occupancy which in turn results in revenue loss. The goal of this project is to predict in advance weather a hotel customer will cancel his booking or not. Predicting future booking cancellation can help hotels plan for cancellation and refund policies, staffing schedules as well as targeting customers with offers and discounts. It is also important to understand key booking cancellation factors and how those factors relate to booking cancellation.

To achieve this, I am planning to use "Hotel booking demand" dataset available on Kaggle(https://www.kaggle.com/jessemostipak/hotel-booking-demand).

The dataset consists of 119,390 observations with 32 features. 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, lead time, number of adults, children, babes, deposit type and previous cancellations.

Using these features, I will train and test different machine learning models that predict if the booking will be canceled or confirmed. For exploratory data analysis and visualization, I will use pandas, numpy, seaborn and plotly libraries while for model training and evaluation, I will use scikit-learn and keras.

MVP for this project will be a Jupiter notebook that trains and evaluates a classification model that predict booking cancellation and a heatmap of the most important features.