

Hotel booking Project Proposal

This repository to meet requirements for data science bootcamp with SADIA.

Predict the hotel booking's cancellation.?

A study conducted by property management systems (PMS) provider Guestline has revealed insights on how hoteliers can avoid booking cancellation. The survey amongst 2,000 participants delved into how people book their hotels, and more importantly, their main priorities and desires when choosing one.

Dataset

We have data set **hotel-booking-demand** that represents booking data for a city hotel and a resort hotel with 1193900 records and 32 features. Includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things. I use this data set for extract useful information to hotels companies or booking apps.

This dataset can be found at [Kaggle](https://www.kaggle.com/jessemostipak/hotel-booking-demand/code):

<https://www.kaggle.com/jessemostipak/hotel-booking-demand/code>

This dataset contains hotel for the following type:

- City hotel
- Resort hotel

The dataset is available as the .csv file. a sample of data is shown in the following table:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights	
0	Resort Hotel	0	342	2015	July	27	1	0	

adults	...	deposit_type	agent	company	days_in_waiting_list	customer_type	adr	required_car_parking_spaces	total_of_special_requests	reservation_status
2	...	No Deposit	NaN	NaN	0	Transient	0.0	0	0	Check-Out

The most important features for this study:

Hotel, Hotel Resort Hotel or City Hotel

is_canceled, Value indicating if the booking was canceled (1) or not (0) (label variable, the target)

lead_time, Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

arrival_date_year, Year of arrival date

arrival_date_month, Month of arrival date

arrival_date_week_number, Week number of year for arrival date

arrival_date_day_of_month, Day of arrival date

stays_in_weekend_nights, Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

stays_in_week_nights, Number of weeknights (Monday to Friday) the guest stayed or booked to stay at the hotel

adults, Number of adults

The logistic Regression Analysis is the process of predicting a Label based on the features at hand so, here we use regression to know relationship between cancellation and other features.

Also, in this particular study, the KNN algorithm is used to classify hotel bookings in terms of cancellation risk (1 = model predicts that the customer will cancel their booking, 0 = customer is not predicted to cancel their booking).

Tools

I used Jupyter notebook to write codes and import libraries to achieve the goal of this dataset, such as:

numby, matplotlib, pandas used to EDA and prepare data to train a model a. and for regression use sklearn, logisticRegression (function) will be used to analyzes the relationship between two or more features it is type of supervised learning to train the model. And KNeighborsClassifier to KNN.

TO DO:

- I will do wrangling data and exploratory data analysis before used the model.
- Visualize data and Fit models
- **Note:** some features might be increased or changed and the model as well.