# Project proposal

**(Hotel Booking)**

Description:

This project about Hotel Booking dataset to do EDA and predict the booking that were canceled or not using Classification algorithms.

This dataset contains observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and 31st of August 2017, including booking that effectively arrived.

The dataset provided by Kaggle website:

<https://www.kaggle.com/mojtaba142/hotel-booking>

Question/need:

1-Select what year was the most reservation in?

2-What is the cancellation rate for each hotel?

3-What is the month in which there were the most cancellations, and what is the month in which there were the least cancellations?

4-Is the number of changes of the reservation affected by (market\_segment)?

5-Is booking changes affected by deposit type or by days in waiting list?

6-How many hotel reservation in weekend night vs week nights?

Data Description:

This dataset contains 119390 observations and 36 columns

|  |  |  |
| --- | --- | --- |
| Description | Type | Feature name |
| Average Daily Rate. Calculated by dividing the sum of all lodging transactions by the total number of staying nights. | Float | ADR |
| Number of adults | Integer | Adults |
| ID of the travel agency that made the booking. | Float | Agent |
| Day of the month of the arrival date | Integer | ArrivalDateDayOfMonth |
| Month of arrival date with 12 categories: “January” to “December” | Object | ArrivalDateMonth |
| Week number of the arrival date | Integer | ArrivalDateWeekNumber |
| Year of arrival date | Integer | ArrivalDateYear |
| Code for the type of room assigned to the booking. | Object | AssignedRoomType |
| Number of babies. | Integer | Babies |
| Number of changes/amendments made to the booking from the moment the booking was entered on the Property Management System until the moment of check-in or cancellation. | Integer | BookingChanges |
| Number of children. Sum of both payable and non-payable children. | Float | Children |
| ID of the company/entity that made the booking or responsible for paying the booking. | Float | Company |
| Country of origin. Categories are represented in the International Standards Organization (ISO) 3155–3:2013 format. | Object | Country |
| Customer credit card number. Artificially created. | Object | CreditCard |
| Type of booking, assuming one of four categories: Contract ,Group (when the booking is associated to a group), Transient (when the booking is not part of a group or contract, and is not associated to other transient booking), and Transient-party (when the booking is transient, but is associated to at least other transient booking). | Object | CustomerType |
| Number of days the booking was in the waiting list before it was confirmed to the customer. | Integer | DaysInWaitingList |
| Indication on if the customer made a deposit to guarantee the booking. | Object | DepositType |
| Booking distribution channel. The term “TA” means “Travel Agents” and “TO” means “Tour Operators”. | Object | DistributionChannel |
| Customer email. Artificially created. | Object | Email |
| Type of hotel. Categories are presented in City Hotel and Resort Hotel | Object | Hotel |
| Value indicating if the booking was canceled (1) or not (0) | Integer | IsCanceled |
| Value indicating if the booking name was from a repeated guest (1) or not (0). | Integer | IsRepeatedGuest |
| Number of days that elapsed between the entering date of the booking into the Property Management System and the arrival date. | Integer | LeadTime |
| Market segment designation. In categories, the term “TA” means “Travel Agents” and “TO” means “Tour Operators”. | Object | MarketSegment |
| Type of meal booked. Categories are presented in standard hospitality meal packages. | Object | Meal |
| Customer name. Artificially created | Object | Name |
| Customer phone number. Artificially created | Object | PhoneNumber |
| Number of previous bookings not canceled by the customer prior to the current booking. | Integer | PreviousBookingsNotCanceled |
| Number of previous bookings that were canceled by the customer prior to the current booking. | Integer | PreviousCancellations |
| Number of car parking spaces required by the customer | Integer | RequiredCarParkingSpaces |
| Reservation last status | Object | ReservationStatus |
| Date at which the last status was set. | Object | ReservationStatusDate |
| Code of room type reserved. Code is presented instead of designation for anonymity reasons. | Object | ReservedRoomType |
| Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel. Calculated by counting the number of weekend nights from the total number of nights. | Integer | StaysInWeekendNights |
| Number of week nights (Monday to Friday) | Integer | StaysInWeekNights |
| Number of special requests made by the customer (e.g. twin bed or high floor). | Integer | TotalOfSpecialRequests |

Goal:

The goal of this project is:

Predict if someone canceled the reservation or not by using Classification algorithms.

About target column (is canceled) have two classes:

|  |  |
| --- | --- |
| Class 2 | Class 1 |
| 1 (was canceled) | 0 (was not canceled) |
| 44224 | 75166 |

Tools:

-Numpy and Pandas for data manipulation.

-Scikit-learn for modeling.

Matplotlib and Seaborn for plotting.-

-Tableau for interactive visualizations.

-Sklearn for linear regression.

- Xgboost.

-Web Scraping.