**Hotel Booking Cancellation**

Reservation cancellation is not uncommon in the hotel industry. Therefore, if we can predict if a guest would cancel a reservation, hotels could contact guests that most likely to cancel to confirm more efficiently and to resell the room to optimize revenues. As a data-scientist we have to help the company to predict whether the customer will cancel or not. We have to do some data analysis and we have to work on Machine Learning Model(s) to select the model that predicts the cancellation by comparing them with the accuracy scores of different ML models.

The data has been provided by Kaggle ( <https://www.kaggle.com>). The dataset contains 119,391 observations with 32 features about booking information for a city hotel and a resort hotel, such as Arrival Time, when the booking was made, length of stay, the number of adults, children, and/or babies etc. There are 32 columns 12 of them are Categorical and 20 are Numerical.

**Exploratory Data Analysis (EDA):**

* From where the most guests are coming?
* How much do guests pay for a room per night?
* Which are the busiest month?
* How long do people stay at the hotels?
* How many bookings were canceled?
* Which month have the highest number of cancelations?

**Tools:**

In this project will use the following libraries: pandas, numpy, matplotlib.pyplot, seaborn and sklearn.

**Data Pre-processing:**

* Dropping columns that are not useful.
* Creating Numerical and Categorical data frames.
* Handling non-numeric data.
* Dealing with Missing Data and Correlation Matrix.
* Reviewing correlations between features.
* Splitting data into training and test set.

**Model Building and Comparison:**

1. Logistic Regression
2. KNN
3. Decision Tree
4. Random Forest

Regards,

Manal Mohammed AlSaleh