

**T5 Boot camp**

**Data science**

Project proposal

(Hotel Booking )

By:

Najd Almutairi

Instructor:

Ahmed Abu Ghida

Design :-

Hotel wants to get the maximum profit and minimum possible loss but in the realty . losses can not be avoided ,one that caused it was the party canceling the hotel booking .As the potential benefits that should be obtained are losses due to customers canceling orders,especially those who order in large quantities in groups.

### **Data**

This data set contains booking information for a city hotel and a resort hotel. This dataset have **119390** **data points**and **32 features**, and 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.

## Algorithms

This dataset contain information of hotel booking, We will perform exploratory data analysis to get insight from the data. We will try to answer the following Questions:

* How Many Booking Were Cancelled?
* What is the booking ratio between Resort Hotel and City Hotel?
* From which country most guest come?
* Which was the most booked accommodation type (Single, Couple, Family)?

After that we will make predictive model to predict whether the booking will be cancelled or not using :

- RandomForestClassifier

- k-means clustering

**Tools**

* **Pandas**: a library offers data structures and operations for manipulating numerical tables and time series.
* **Numpy**: a library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.
* **Matplotlib**: a plotting library for the Python programming language and its numerical mathematics extension NumPy.
* **Seaborn**: a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.
* **Sklearn** : a library used for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering.