This data set contains booking information for a city hotel and a resort hotel, 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. Using the booking information about these hotels, the challenge is to build a model to predict booking cancellation based on the following fields:

Data Description:

1 - hotel (str) - Hotel (H1 = Resort Hotel or H2 = City Hotel)

In hotel. The resort hotel is a luxury facility that is intended primarily for vacationers and is usually located near special attractions, such as beaches and seashores, scenic or historic areas, ski parks, or spas.

City-centre/Downtown hotels- Generally located in the heart of city within a. short distance from business and shopping centers. These hotels are more. expensive than the suburban hotels. The clientele to these hotels is business men and high income tourists.

2 - lead\_time (int) - Number of days that elapsed between the entering date of the booking into the PMS and the arrival date.

PMS - software that facilitates a hotel's reservation management and administrative tasks.

3 - arrival\_date\_year (int) - Year of arrival date

4 - arrival\_date\_month (str) - Month of arrival date

5 - arrival\_date\_week\_number (int) - Week number of year for arrival date

6 - arrival\_date\_day\_of\_month (int) - Day of arrival date

7 - stays\_in\_weekend\_nights (int) - Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

8 - stays\_in\_week\_nights (int) - Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

9 - adults (int) - Number of adults

10 - children (float) - Number of children

11 - babies (int) - Number of babies

12 - meal (str) - Type of meal booked. Categories are presented in standard hospitality meal packages: (Undefined/SC = no meal package, BB = Bed & Breakfast, HB = Half board – breakfast and one other meal - usually dinner, FB = Full board – breakfast, lunch and dinner)

13- country (str) - Country of origin. Categories are represented in the ISO 3155–3:2013 format

14- market\_segment (str) - Market segment designation (TA = Travel Agents, TO = Tour Operators)

The difference is that the tour operator creates packages and sells them either to the consumer or the travel agent. Whereas, the travel agent only sells the packages on behalf of the tour operator and does not create packages.

15 - distribution\_channel (str) - Booking distribution channel (TA = Travel Agents, TO = Tour Operators).

hotel distribution channel is anywhere — online or offline.

16 - is\_repeated\_guest (int) - Value indicating if the booking name was from a repeated guest (1) or not (0)

17 - previous\_cancellations (int) - Number of previous bookings that were cancelled by the customer prior to the current booking

18 - previous\_bookings\_not\_canceled (int) - Number of previous bookings not cancelled by the customer prior to the current booking

19 - reserved\_room\_type (str) - Code of room type reserved. Code is presented instead of designation for anonymity reasons

20- assigned\_room\_type (str) - Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons

21 - booking\_changes (int) - Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation

22 - deposit\_type (str) - Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: (No Deposit = no deposit was made, Non Refund = a deposit was made in the value of the total stay cost, Refundable = a deposit was made with a value under the total cost of stay)

23 - agent (float) - ID of the travel agency that made the booking

24 - company (float) - ID of the company/entity that made the booking or responsible for paying the booking. ID is presented instead of designation for anonymity reasons

25 - days\_in\_waiting\_list (int) - Number of days the booking was in the waiting list before it was confirmed to the customer.

Bookings on the Waiting List are unconfirmed bookings. This means that the resource in question doesn’t currently have enough availability to accommodate the booking. So, the booking either needs to move to another date, another resource or it needs to be reduced to a point where it fits within the availability.

26 - customer\_type (str) - Type of booking, assuming one of four categories: (Contract = when the booking has an allotment or other type of contract associated to it, 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, Transient-party = when the booking is transient, but is associated to at least other transient booking.

The main differences between these customer types are the size of the group, the booking duration, and the nature of the booking agreement. Transient and transient-party bookings involve individual guests or small groups for short stays, contract bookings are negotiated agreements with organizations for a specified period, and group bookings are made for larger gatherings or events. Each type has its own implications for hotel operations, revenue management, and guest services.

27 - adr (float) - Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights.

ADR stands for "Average Daily Rate" in the hotel industry. It is a key performance metric used to measure the average revenue earned per available room in a hotel over a specific period of time, usually calculated on a daily basis. ADR is an essential indicator for hoteliers as it provides valuable insights into the hotel's pricing strategy and revenue management.

To calculate the Average Daily Rate (ADR), you use the following formula:

ADR = Total Room Revenue / Number of Rooms Sold

For example, if a hotel's total room revenue for a given month is $50,000, and the number of rooms sold during that month is 500, then the ADR would be:

ADR = $50,000 / 500 = $100

This means that, on average, each room sold during that month generated $100 in revenue for the hotel.

28 - required\_car\_parking\_spaces (int) - Number of car parking spaces required by the customer

29 - total\_of\_special\_requests (int) - Number of special requests made by the customer (e.g. twin bed or high floor)

30 - reservation\_status (str) - Reservation last status, assuming one of three categories: (Canceled = booking was canceled by the customer, Check-Out = customer has checked in but already departed, No-Show = customer did not check-in and did inform the hotel of the reason why)

31 - reservation\_status\_date (str) - Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to understand when was the booking canceled or when did the customer checked-out of the hotel.