Round 01:- Kaggle Competition

Case Study

Data Strom 2.0

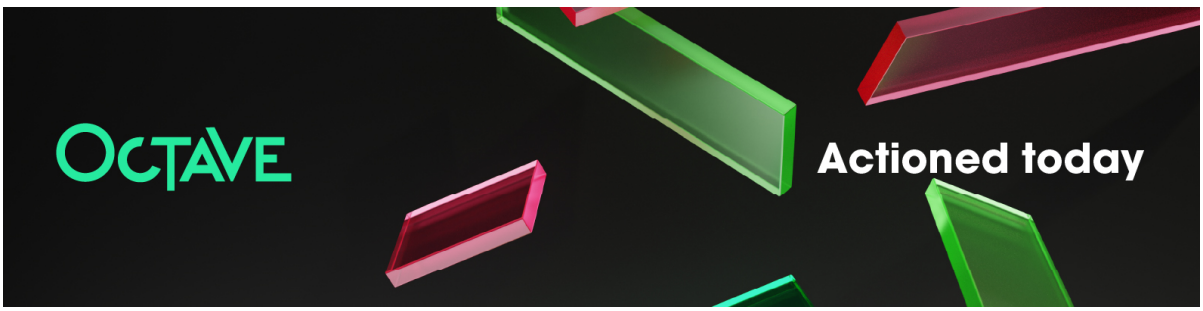
OCTAVE,

Rotaract Club of University of Moratuwa,

&

Rotaract Club of Faculty of Science, University of Colombo

March, 2021



1 Hotel Reservations

1.1 Business Problem:

The hotel industry is one of the fastest growing industries in the world, catering to many customers who require overnight accommodation. Due to various reasons, a traveler's plan may change and result in a booking cancellation or a no-show. In recent years, the cancellation rate for bookings has become quite high and poses a challenge for many hotels to take necessary precautions.

To deal with the potential revenue loss, it's important to understand whether a given traveler will cancel or not show for their booking reservation and develop interventions to minimize these circumstances.

1.2 Use Case definition

Hotel Chain **A** decided to use data analytics techniques to derive the possibility of a given booking reservation get cancelled. They have collected over 15 months of booking reservation data at their three hotel types: City Hotels, Airport Hotels and Resorts. They are seeking to understand the potential cancellations, no-show and check-in reservations through data analytics, which can be useful to identify key parameters for booking reservations.

In this round of Data Storm 2.0, you are required to provide an analytics solution to Hotel Chain **A** in order to solve their booking reservation cancellation problem.

1.3 Data Sources

You are provided with the following data source files to develop a data analytics solution.

1. **Historical training dataset** – Historical booking reservation dataset consist of 27500 booking reservations with more than 20 attributes that hotel staff have collected (*Hotel-A-train.csv*). Please refer *Hotel-A-data-dictionary.csv* for more details about the attributions.

2. **Historical validation dataset** – Validation dataset (*Hotel-A-validation.csv*)

3. **Testing dataset** – You are required to test your solution using *Hotel-A-test.csv*. Final outcome for each reservation-id in the test data should be submitted in a csv format which contains *reservation-ID, likelihood class* format.

4. **Data Dictionary** – Data dictionary file contains the information about data attributes (*Hotel-A-data-dictionary.csv*).

1.4 Deliverables & Evolution metrics

In this competition, you are required to submit the following.

1. **Analytical Solution** – Analytical solution to predict the possibility of a given booking reservation get cancelled, be a no show or a check-In. You are required to submit a csv file, which contains the reservation-ID and the predicted Booking status using the given test set. You can use F_1 metric for evaluation – **50 Points**

   F_1 Score is given by;

$$F_1 = \frac{2 \times \text{ Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

   where;

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

2. **Technical Report** – Report of your solution with clearly defined steps, features, feature engineering steps, modelling approaches, evaluation metrices, all the necessary plots/figures and interesting business findings that you can derive from this analysis – **50 Points**.

   (a) What is the expected revenue loss due to booking cancellations?

(b) Discuss any additonal attributes that hotel management should collect

(c) List down several interventions that the management team can take to reduce the revenue loss.