**PROJECT PROPOSAL**

**Title:** Hotel Booking Cancellation Analysis

**Data Source:** Kaggle- <https://www.kaggle.com/duygut/eda-booking-cancelation-prediction/data>

hotel\_bookings.csv is the file in above link

**Data description:**

There are 119390 observation and 32 variables in the dataset.

The dataset has a column named “hotel” which compares various booking information between two hotels: a city hotel and a resort hotel.

The variables contain details about arrival, location, previous records of cancellations, guests, status of reservation etc:-

**Motivation:**

Consumers are always looking for ways to minimize their cost of buying something. When they find out that they can buy the same thing at a lower price than they paid for, they would attempt to cancel and repurchase, and that’s what usually happens with hotel bookings. A huge loss of income occurs because of unsold rooms due to last minute cancellations. Every day in the U.S alone, more than 221,000 hotel bookings are canceled, resulting in a market which sees $8.6 billion wasted annually.

Further, the COVID 19 pandemic has sent shockwaves of disruptions to travel plans worldwide. The global hospitality industry is overwhelmed by the large number of cancellations spurred by the virus. It’s better to have 70% of cancellations and knowing about them in good time than 20% of cancellations that are notified at the last minute. Therefore, predicting reservations that can be canceled and preventing these cancellations will create a surplus value for the institutions.

**Project ideas:**

To identify the factors that contribute to hotel booking cancellations, this helps to understand what factors can be controlled to avoid huge cancellations.

Logistic regression, CART, K-NN(optional) can be used to understand this. Hyperparameter tuning, Gridsearch can also be implemented.