Predicting Hotel Cancellation and Price

1. **Problem Statement**

Developing a model to predict who is going to cancel their booking to help the tourism industry. Develop a model to predict the hotel price considering other multiple factors.

2. **Background of the problem**

Hotel cancellation is one of the biggest challenges to control in tourism industry. A lot of hotels let you cancel within 24 hours notice. Some hotels have a no refund policy however, still many hotels authorize customers to cancel within certain period time notice at a rated penalty or without a penalty. This is a good customer service which can help the customers to plan out their trips more reliably however, there is a risky factor for hotels. Therefore, predicting the actual demand and price can help to reduce the unexpected profit loss. This can help the industry by reducing the unexpected risk, as well as ready with enough facilities.

3. **Success Criteria**

- Determine the key features that lead to hotel cancellation and estimating the price
- Build a model based on the key features to predict who is going to cancel or not with at least 70-80% of accuracy.
- Build a model based on the key features to predict the nightly hotel price and have a reasonable Mean Square Error, R2, Mean Absolute Error.

4. **Data Source**

The dataset is sourced from https://www.kaggle.com/jessemostipak/hotel-booking-demand and consists of a dataset: hotel_bookings.csv file.

- Hotel_bookings.csv file is composed of total of 32 columns of features and 40,060 observations of H1 and 79,330 observations of H2 collected from the article Hotel Booking Demand Dataset written by Nuno Antonio and more for Data in Brief, Volume 22, February 2019.
- Features: hotel, is_canceled, lead_time, arrival_date_year, arrival_date_month, arrival_date_week_number, arrival_date_day_of_month, stay_in_week_nights, adults, children, babies, is_repeated_guest, previous_cancellation, previous_bookings_not_canceled, booking_change, agent, days_in_waiting_list, adr, required_car_parking_spaces, total_of_special requests

5. **Constraints & Scope**

- Dataset might be related to specific region rather than overall of the US or Worldwide

- Only developing a model based on the features available in the dataset; there might be additional important features to predict the models

6. Approach

Multiple steps will be taken to develop a predictive models for this project as well as to analyze the resulting predictions.

1. The hotel_bookings.csv file will be imported and cleaned via python. Missing values will be handled appropriately based on specific factors.
2. Exploratory Data Analysis
3. Classification model for predicting cancellation
4. Regression Model for predicting price
5. Evaluation of models (R2, MSE, MAE)

7. Deliverables

The final draft of the project will be presented in the form of a formal project report and slide deck. Jupyter Notebook will be delivered with detailed steps of codes and analysis of the project.