## Capstone Project: Hotel Booking Cancellation Prediction

### Overview

The goal of this project is to build a machine learning model capable of predicting hotel booking cancellations using customer and booking data. This helps hoteliers in managing revenue, inventory, and overbooking strategies efficiently. The project follows the supervised learning paradigm using structured tabular data.

### **Install**

This project requires ****Python 2.7**** and the following Python libraries installed:

* [NumPy](http://www.numpy.org/)
* [Pandas](http://pandas.pydata.org/)
* [matplotlib](http://matplotlib.org/)
* [scikit-learn](http://scikit-learn.org/stable/)
* [time](https://docs.python.org/2/library/time.html)
* Seaborn
* XGBoost

### Dataset

The dataset used is a cleaned version of the **Hotel Booking Demand Dataset**, available from:

[UCI Machine Learning Repository](https://www.sciencedirect.com/science/article/pii/S2352340918315191" \t "_new)

Files used: hotel\_bookings-1.csv: Original raw dataset.

hotel\_booking\_Model\_data.csv: Cleaned and preprocessed dataset for model training.

### Setup

There is no specific environment or deployment setup required. The notebooks can be run in any Jupyter-compatible environment with the above libraries installed.

### Project Workflow

### ****Exploratory Data Analysis**** Notebook: Exploratory\_Data\_Analysis\_1.ipynb Key tasks: Handling nulls, feature relationships, outliers, distributions

**Data Cleaning & Preprocessing**  
Notebook: Data\_cleaning1.ipynb  
Tasks: Encoding categorical variables, removing irrelevant features, handling class imbalance.

**Model Building & Prediction**  
Notebook: Machine\_Learning\_Prediction\_.ipynb  
Models used: Logistic Regression, Random Forest, XGBoost.  
Evaluation metrics: Accuracy, Precision, Recall, Confusion Matrix

### **Submission Files**

The following files are included as part of this submission:

1. README.md: Project overview and installation.
2. hotel\_bookings-1.csv: Raw dataset.
3. hotel\_booking\_Model\_data.csv: Processed dataset
4. Exploratory\_Data\_Analysis\_1.ipynb: EDA notebook.
5. Data\_cleaning1.ipynb: Data cleaning and transformation.
6. Machine\_Learning\_Prediction\_.ipynb:
7. Model building and evaluation.