Priceline Take-Home Assessment Case Study
Role: Machine Learning Engineer

Submission Requirements:

- Complete the case study below, ensuring that all deliverables are provided.
- Submit your final deliverables as a zip file or through a shared repository link.
- The case study should take approximately **3-5 hours** to complete. Focus on clarity, structure, and scalability over complexity.

Deliverables:

1. A runnable Python script(s) containing:
   - Exploratory Data Analysis (EDA).
   - Preprocessing and feature engineering steps.
   - Machine learning pipeline implementation.
   - Model training, evaluation, and comparison.
   - Visualization of results (if applicable).
2. Clean, well-documented, and modular production-level code.
3. A README file containing:
   - A summary of your approach and findings.
   - Recommendations for improving the system.
   - Instructions to set up and run the script, including any dependencies or environments needed.

Guidelines:

- You are free to use any open-source libraries.
- Clearly state any assumptions you make.
- Focus on providing a structured and logical approach.

Good luck, and we look forward to reviewing your submission!

Case Study: Hotel Booking Probability Prediction

Background: When a customer visits our website and provides search parameters such as check-in date, check-out date, destination, and number of rooms, they are redirected to our hotel listing page. Here, a list of hotels is displayed, and as the customer browses, clicks and bookings may result from their interactions.

The company wants to improve its ability to predict which hotel among those listed has the highest probability of being booked. This will optimize the ordering of hotel listings and improve customer satisfaction.

Task:

1. Data Understanding and Exploration
2. Feature Engineering
3. Pipeline Development
4. Model Development
5. Model Evaluation
6. Scalability and Deployment

Dataset:

The dataset provided includes hotel search records, with labels indicating whether the hotels were clicked or booked. While this dataset should be your primary focus, you are welcome to use any external data sources that could enhance your analysis. A data dictionary is provided to clarify any questions you may have.

- searchId: Unique ID for the search
- checkInDate: Check-in date for the search
- checkOutDate: Check-out date for the search
- searchDate: Date when the search occurred
- destinationName: Name of the destination for the search
- numRooms: Number of rooms requested in the search
- userId: Unique ID of the customer
- vipTier: Customer's VIP program membership level
- deviceCode: Device type code used for the search
- signedInFlag: Whether the customer was signed in during the search
- rank: Rank of the listed hotel in the search results
- hotelId: Unique hotel ID of the listed hotel
- brandId: Unique brand ID of the listed hotel's brand
- starLevel: Star level of the listed hotel
- customerReviewScore: Average review score of the listed hotel from post-stay surveys
- reviewCount: Number of reviews the listed hotel received from surveys
- minPrice: Current price of the lowest-price room in the listed hotel
- minStrikePrice: Original price of the lowest-price room before discount
- freeBreakfastFlag: Whether the listed hotel offers free breakfast
- freeInternetFlag: Whether the listed hotel offers free internet
- bookingLabel: Whether the listed hotel led to a booking through the search
- clickLabel: Whether the listed hotel led to a click through the search