Final Project

Project Goal:

The overall goal of the case is to provide actionable insight, based on the data available.

The dataset contains real-life hotel stay data, with each row representing a hotel booking.

Your team's goal is to understand some key drivers for why people cancel hotel reservations and/or better predict who will cancel. To do this, you will use all of the skills you have developed in the labs/homework to make sense of a novel dataset, to perform some essential analyses on the dataset, and to explain/document what you have done, and your insights generated.

The Data:

Here are the variables you will find in your data file:

- **IsCanceled**: Categorical Value indicating if the booking was canceled (1) or not (0)
- **LeadTime**: Integer, Number of days that elapsed between the entering date of the booking into and the arrival date
- **StaysInWeekendNights**: Integer, Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
- **StaysInWeekNights**: Integer, Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
- **Adults**: Integer, Number of adults
- **Children**: Integer, Number of children
- **Babies**: Integer, Number of babies
- **Meal**: Categorical, Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal package; BB – Bed & Breakfast; HB – Half board (breakfast and one other meal – usually dinner); FB – Full board (breakfast, lunch and dinner)
- **Country**: Categorical, Country of origin. Categories are represented in the ISO 3155–3:2013 format
- **MarketSegment**: Categorical, Market segment designation. In categories, the term "TA" means "Travel Agents" and "TO" means "Tour Operators"
- **IsRepeatedGuest**: Categorical, Value indicating if the booking name was from a repeated guest (1) or not (0)
- **PreviousCancellations**: Integer, Number of previous bookings that were cancelled by the customer prior to the current booking

- **PreviousBookingsNotCanceled**: Integer, Number of previous bookings not cancelled by the customer prior to the current booking
- **ReservedRoomType**: Categorical, Code of room type reserved. Code is presented instead of designation for anonymity reasons
- **AssignedRoomType**: Categorical, Code for the type of room assigned to the booking. Sometimes the assigned room type differs from the reserved room type due to hotel operation reasons (e.g. overbooking) or by customer request. Code is presented instead of designation for anonymity reasons
- **BookingChanges**: Integer, Number of changes/amendments made to the booking from the moment the booking was entered on the PMS until the moment of check-in or cancellation
- **DepositType**: Categorical, Indication on if the customer made a deposit to guarantee the booking. This variable can assume three categories: No Deposit – no deposit was made. Non Refund – a deposit was made in the value of the total stay cost. Refundable – a deposit was made with a value under the total cost of stay.
- **CustomerType**: Categorical, Type of booking, assuming one of four categories: Contract - when the booking has an allotment or other type of contract associated to it; Group – when the booking is associated to a group; Transient – when the booking is not part of a group or contract, and is not associated to other transient booking; Transient-party – when the booking is transient, but is associated to at least other transient booking
- **RequiredCardParkingSpaces**: Integer, Number of car parking spaces required by the customer
- **TotalOfSpecialRequests**: Integer, Number of special requests made by the customer (e.g. twin bed or high floor)

Project Deliverables:

There are three deliverables for this project:

1. A report (e.g., word document) that describes the work done. This is a technical document and you should feel free to use technical terms. You must explain any visualization in your document.
2. The R code for your analysis (the team needs to figure out how to share code / work together)
3. A presentation of the actionable insight achieved. This actionable insight should be submitted in the form of a presentation, where the insights generated are explained. You must record your presentation, and it should be 10 minutes long (e.g., a powerpoint presentation with your voice explaining the slides).

The analysis should include exploratory analysis (e.g., histograms, scatter plots), mapping visualizations and several machine learning techniques.

Hints:

For exploratory work:

   a. Histograms and boxplots of numeric variables are typically useful.
   b. Producing tables of categorical response variables is often helpful.
   c. Dividing the data in cancelled and not-cancelled subsets and doing boxplots, using each of other grouping variables is often useful.
   d. Barplots of cancelations across different categories is often useful

2. Since there is geographic info, you should make some sort of map:
   a. To create a color gradient world map the rworldmap package and the joinCountryData2Map() and mapCountryData() functions might be helpful
3. It might be helpful to create a new column that is Detractor (it could be Boolean) and try to understand rules for predicting a Detractor.
4. For your recommendations (in the presentation):
   a. Provide a summary covering all of your results in language that is suitable for a manager to understand. Most managers do not know too much about statistics, so you probably should not quote terms like “R-squared” or “p-value” but rather describe your results in plain language.

b. Your presentation should conclude with at least one substantive recommendation to the managers.

c. Important: Your recommendation(s) MUST be connected with one or more of your data science results; it MUST NOT be based on your own personal experience with hotels, etc.

Data:

The data file is located at:
https://intro-datascience.s3.us-east-2.amazonaws.com/Resort01.csv