Capstone Project Proposal Template

**Notes:**

* This should take no more than one hour to complete – the clearer you are about the business problem you’re working to solve with your ML-driven solution, the easier your proposal will be to complete
* This will be uploaded to your repo, which will be a part of your final submission
* Due date for submission is 1/16

**Instructions:**

1. Download this document as a Word Doc
2. Answer each question using a few sentences, at most
3. Save your completed proposal as a PDF
4. [Create a project GitHub repo](https://github.com/new) (if you have yet to do so)
5. [Add your instructor as a collaborator](https://docs.github.com/en/account-and-profile/setting-up-and-managing-your-personal-account-on-github/managing-access-to-your-personal-repositories/inviting-collaborators-to-a-personal-repository) (username dodgy719) to your project repo
6. Add your mentor as a collaborator
7. Push your proposal PDF (created in Step 3) up to your repo
8. Copy the URL corresponding to the location of the PDF in your repo
9. Submit the copied URL using [this link](https://my.learn.co/courses/586/quizzes/6353?module_item_id=79223)

**[project name]**

**Business Understanding**

* What problem are you trying to solve, or what question are you trying to answer?
  + Which customers are most likely to cancel their hotel reservation? Or Is it possible to know what percentage of hotel guests for any given date will cancel their hotel reservations based on a few key traits.
* What industry/realm/domain does this apply to?
  + Hotel Industry
* What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
  + It would be interesting to learn what makes someone most likely to cancel their reservation

**Data Understanding**

* What data will you collect?
  + A dataset on hotel booking
* Is there a plan for how to get the data (API request, direct download, etc.)?
  + Direct download datasets from Kaggle
* What are the features you’ll be using in your model?
  + I’ll be using every column from the dataset.
  + Booking\_ID: unique identifier of each booking
  + no\_of\_adults: Number of adults
  + no\_of\_children: Number of Children
  + no\_of\_weekend\_nights: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel
  + no\_of\_week\_nights: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel
  + type\_of\_meal\_plan: Type of meal plan booked by the customer:
  + required\_car\_parking\_space: Does the customer require a car parking space? (0 - No, 1- Yes)
  + room\_type\_reserved: Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels.
  + lead\_time: Number of days between the date of booking and the arrival date
  + arrival\_year: Year of arrival date
  + arrival\_month: Month of arrival date
  + arrival\_date: Date of the month
  + market\_segment\_type: Market segment designation.
  + repeated\_guest: Is the customer a repeated guest? (0 - No, 1- Yes)
  + no\_of\_previous\_cancellations: Number of previous bookings that were canceled by the customer prior to the current booking
  + no\_of\_previous\_bookings\_not\_canceled: Number of previous bookings not canceled by the customer prior to the current booking
  + avg\_price\_per\_room: Average price per day of the reservation; prices of the rooms are dynamic. (in euros)
  + no\_of\_special\_requests: Total number of special requests made by the customer (e.g. high floor, view from the room, etc)
  + booking\_status: Flag indicating if the booking was canceled or not.

**Data Preparation**

* What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
* What are some of the cleaning/pre-processing challenges for this data?
  + There is no cleaning to be done because there are no null values in this dataset

**Modeling**

* What modeling techniques are most appropriate for your problem?
  + Because the outcome is binary, if someone cancels or doesn’t cancel Logistic Regression will work best.
* What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
  + The target variable is Booking Status (indicates if the booking was cancelled or not)
* Is this a regression or classification problem?
  + Regression problem

**Evaluation**

* What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?
  + I will be using a confusion matrix to see how my model performs.

**Tools/Methodologies**

* What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
  + Neural Networks