# Project Proposal

**Topics Covered:**

1. **Introduction to the problem**
2. **Client details and decision based on analysis**
3. **Details of Data**
4. **Proposed Outline of the approach**
5. **Deliverables**

**Introduction:**

**AirBnB new user Bookings**

The problem I am going to solve is concerning a travel bookings company which provides primarily the accommodation bookings. It started by covering typical bread and breakfast chain initially, but now covers a large group of hotels as well as individual rooms and houses listed over its website.

I am going to predict the destination country a user might book as her first booking through AirBnB. The company likes to customize their users’ experience as per their tastes and requests and hence this might help them with their goal.

**Client Details and decision based on the analysis:**

The client for this problem is AirBnB as well as Kaggle, as I would be submitting the code to them as well. AirBnB would be interested to know what destination their customer might book in advance so that it can help them in preparing well to provide their customers with a customized experience which ultimately leads to great user experience and provide them with a competitive edge. It will also help in maintaining demand and supply as prior information of the booking will lead to better managed supply for the accommodation.

**Details of Data:**

Data used here has been taken from Kaggle website. It contains six different files namely:

* Countries.csv
* Age\_gender\_bkts.csv
* Train and test data has been separately provided as train\_users\_2.csv and test\_users.csv
* Sessions.csv
* Sample\_submission\_NDF.csv

The first file contains the details of the possible destination countries along with the longitude, latitude and the area of the country in kms.

Age gender buckets data contains the details of the destination country visited by the users by their gender, age bucket and the population of the city visited in the country.

Training and Test set contains total of 16 different variables including the user IDs, destination country, age, date of account creation, date of first booking, which affiliate channels and providers were used etc. This is a very specific information per user and shall be used ultimately to model for a solution.

Sessions data set is a huge data file which contains the actions taken by the users during different sessions spent by them over the website.

Sample submission file is the format for submission of the output got after running the model over the test data. This is for the submission over Kaggle website.

**Proposed Outline of the approach:**

1. Data Wrangling & Cleaning
2. Exploratory Data analysis
3. Feature Engineering & Construction
4. Modeling using different techniques learnt during the workshop
5. Testing the model on the test data and going back to step three as per requirement

**Deliverables:**

The deliverables for the capstone project will be the code files, technical documentation or the paper, explaining the problem, the approach and further research and three concrete recommendations to the client based on the findings, and a slide deck to be presented to the client.