

What is the problem you want to solve?

For capstone 1 I will be using Airbnb's new user bookings data. The problem I will try to solve using this data is to predicting where a newly registered user will book his/her first trip from 12 possible outcomes of the destination country, thus this make it an **multiclass classification problem**. The goal here would be to come up with an Machine learning model which accurately predicts the country (from the options) where the user is more likely to book his / her first trip. The model will make its predictions by analyzing variety of data such as user informations, country information, device used etc.

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

Client here is the company (Airbnb) by accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community, decrease the average time to first booking, and better forecast demand.

What data are you using? How will you acquire the data?

I will be using data provided by Airbnb for their competition on Kaggle. The data comes in about 5 .csv files, and will be acquired directly from Kaggle.

Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

I will cover the classical stages of a Data Science project, which includes data acquisition and wrangling, exploratory data analysis, application of

inferential statistics and deeper data analysis using machine learning models to be defined.

The initial analysis will involve visual analytics to determine the spread of data, variance, and measures of central tendency. This will also help determine if there are any data preparation tasks to be done. Followed by statistical tests, this statistical analysis will also help determine which variables to focus on for predictive modelling.

What are your deliverables? Typically, this includes code, a paper, or a slide deck.

As required, I will submit all Jupyter notebooks I will develop, a final report, and a presentation slide deck.

The code in jupyter notebook, and A Slide deck about my findings.