Capstone 1 – Project Proposal

Predicting Successful Price Points for Prospective Airbnb Host's

Zach Palamara

What is the problem you want to solve?

Founded in 2008, Airbnb is a growing and popular alternative to traditional lodging. It is essentially an open market where as a host, you can list a spare room, entire apartment or home for prospective travelers to book and use as a domicile during a trip. Currently there are no free resources for hosts to use when trying to price their Airbnb listing on the open market. Since the amount of listings for Airbnb rentals has exploded over the last several years, being able to competitively price your rental space is a must in order to maintain a high booking rate and rate of return for your property. My goal is to develop a tool for hosts to use to gain insight on how to appropriately price their Airbnb rental space in the market they reside in.

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?

I'm aiming to target new Airbnb hosts in Austin, TX looking to competitively price their listings as well as existing hosts who may want to revisit their pricing strategy to optimize returns. Based on my project these hosts will be able to appropriately price their Airbnb listings based on insight from all of the other Airbnb listings in their local area.

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

I will be using data from [Inside Airbnb](#), which is an independant, non-commercial set of tools that allows you to explore how Airbnb is being used in various cities around the globe. They scrape publicly available data from Airbnb which includes listing descriptions, reviews, calendar data and much more. Inside Airbnb has pre prepared .csv files as well as geo-spatial data in the form of geo-json files, which can be easily downloaded for free directly from their website. There is also another site, [AirDNA](#), which scrapes higher quality data, but at a price.

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.

The first step in this process will be to conduct exploratory data analysis on the data sources. This will begin by examining the raw data and evaluating how to properly clean and prepare it. One of my goals is to do this in a reproducible format so that data from other cities around the world can easily be analyzed using my methods. Once my data is prepared and cleaned, I will begin the EDA process and try to answer the following questions to narrow down the scope of what to target in the regression analysis to be able to figure out what price point and Airbnb host should set their listing at. Upon first look at the data and after reviewing similar projects on Medium, it looks like a regression model approach will be the best way to solve most of these problems.

1. *Which areas/neighborhoods in Austin, TX are the most expensive and have the most listings?*
2. *What are the most common listing types? (ie: entire home/apartment or shared space?)*
3. *How long have different hosts been listing properties through Airbnb and how many listings do host's have on average?*
4. *How do Airbnb prices correlate with the number of people a listing accommodates?*
5. *How favorable are reviews and what do the distributions look like?*
6. *What are the most common amenities, and which amenities are likely to increase the price of a listing?*
7. *What is the value of being a superhost? Is it worth getting verified?*

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

- 15-20 page report (see outline below)
- Slide deck
- Github repository with python scripts etc.

Outline of Report
1. Table of Contents
2. Executive Summary
3. Problem Statement
   a. Brief description of the data what's useful
4. Data Preparation and EDA
   a. Cleaning the Data

        b.   Process Methodology
        c.   Regression Model etc.
        d.   Flow Chart
5. Results
        a.   Charts/Visualizations
6. Conclusions and Findings
        a.   Actionable Insights