

Capstone Project 1: Milestone Report

airbnb

Places to stay Monthly stays Experiences Online Experiences **NEW**

LOCATION Where are you going? CHECK IN / CHECK OUT Add dates GUESTS Add guests **Search**

We may be apart, but we'll get through this together.

Online Experiences
Unique activities we can do together, led by a world of hosts.

Monthly stays
Make Airbnb your home, for stays of a month or longer.

Frontline stays
Find or provide accommodations for COVID-19 responders.

airbnb

Los Angeles, CA Add dates Add guests **Search**

Entire cottage in Topanga
Pirates of the Caribbean Getaway
2 guests · 1 bedroom · 1 bed · 1.5 baths
Wifi · Free parking · Kitchen · Washer
★ 4.90 (617) Superhost \$95 / night

Entire apartment in Malibu
Idyllic, Romantic Beachfront Apartment in Malibu
2 guests · 1 bedroom · 1 bed · 1 bath
Wifi · Free parking · Kitchen · Washer
★ 4.99 (195) Superhost \$399 / night

Entire guesthouse in Mount Washington
Private Pool House with Amazing Views!
3 guests · 1 bedroom · 2 beds · 1.5 baths
Wifi · Free parking · Kitchen · Washer
★ 4.95 (192) Superhost \$185 / night

Entire guest suite in Malibu
Ocean View Malibu Hideaway
<https://www.airbnb.com>

A map of Los Angeles with price markers for various neighborhoods. Prices range from \$69 to \$350 per night. Key locations labeled include Chatsworth, Northridge, Reseda, Canoga Park, Encino, Woodland Hills, Topanga State Park, Bel Air, Westwood, Brentwood, Pacific Palisades, Santa Monica, Culver City, Mid City, Hollywood, Chinatown, Los Feliz, Glendale, Burbank, Pasadena, Altadena, Alhambra, Monterey Park, El Segundo, Hawthorne, Manhattan Beach, Redondo Beach, Torrance, Carson, Lakewood, Bell Gardens, South Gate, Lynwood, Compton, Paramount, and Huntington Park.

Introduction

Airbnb is one of the world's largest marketplaces for unique, authentic places to stay and things to do, offering over 7 million accommodations and 50,000 handcrafted activities, all powered by local hosts. An economic empowerment engine, Airbnb has helped millions of hospitality entrepreneurs monetize their spaces and their passions while keeping the financial benefits of tourism in their own communities. With more than three quarters of a billion guest arrivals to date, and accessible in 62 languages across 220+ countries and regions, Airbnb promotes people-to-people connection, community and trust around the world.

Problem

Although Airbnb and other sites provide some general guidance, there are currently no free and accurate services which help hosts price their properties using a wide range of data points. Also there is no way for travellers to figure out the correct price range based on the features they are looking for.

Paid third party pricing software is available, but generally hosts are required to put in their own expected average nightly price ('base price'), and the algorithm will vary the daily price around that base price on each day depending on the day of the week, seasonality, how far away the date is, and some other factors.

It is very important to get the listings' price correctly, particularly in big cities like New York, San Francisco, Los Angeles, Boston where competition is so high and some small differences in prices can make a big difference.

This project aims to solve this problem, by using machine learning to predict the base price for properties in Los Angeles.

Clients

I have two types of client clients, it can be used for both hosts and airbnb for better pricing optimization & maximize the revenue.

1. Airbnb hosts

If the listing price is high, then chances of listings getting booked will go down. Also if the price is low, then the host will be missing out on a lot of potential income. This project will help Airbnb hosts to price their listings competitively and thereby increase the bookings. The prices predicted by this model can also be used by existing hosts to increase bookings which will result in more earnings.

2. Guests looking to book Airbnb Listings

When guests are on Airbnb they do not have any means of knowing if the price that the hosts are asking for is reasonable for given features or not. This project will help guests to find real value for their money in the Airbnb marketplace.

Dataset

The dataset I am using is available on [Inside Airbnb](#) for [Los Angeles](#) location. "Inside Airbnb" is an independent, non-commercial set of tools and data that allows us to explore how Airbnb is really being used in cities around the world.

As per the information available on [Inside Airbnb](#) the dataset was last scraped on July 8, 2019 and contains information of all 44,504 Los Angeles Airbnb Listings on Airbnb site on July 8, 2019.

```
In [14]: listings.size
```

```
Out[14]: 4729720
```

```
In [16]: listings.shape
```

```
Out[16]: (44620, 106)
```

Data Wrangling

I started this project because as of today Airbnb doesn't have a system where a host can go and find out what is the competitive price they can charge for their listing on Airbnb and at the same time guests don't know if they are getting real value for their money or not.

I took Airbnb LA data and I am working towards predicting Airbnb listing prices based on various conditions. This model will help hosts and guests both. First step after loading the data is Data Wrangling. It involves looking at the data, removing outliers, taking care of missing data and null values, converting data into a format that can be analyzed and studied.

The original dataset contained 44620 Airbnb listings and 106 features but I dropped many of those.

1. Some of the features are free text variables, like the host description of the property and all the written reviews. To include that data, I would have to perform Natural Language processing operations and at present it is out of the scope of this project. So those features were dropped.
2. Several features only contained one category, so they were dropped too.
3. Most listings don't offer the experiences feature, so it was also dropped.
4. Some columns contained mostly 'None' or 'NaN' values and therefore were dropped.
5. There were some features which were duplicates, so those were also dropped.
6. Data in money columns was stored as string, so I fixed those data types and converted those to floats.
7. Categorical data was converted to numeric type.
8. Biggest challenge was the amenities column. It had a lot of useful information but extracting that info was not easy. I used various python

functions to extract the data from this column and then find out the most important items and then create separate columns for these items. Data in these columns was filled as amenities present or absent.

9. I did draw various plots to see the relationships between various features and price of the listing. I also draw graphs to visualize price distribution for these listings.
10. After performing all these operations I saved the clean data into a separate csv file. This file will be loaded in the next phase (EDA phase) and data will be studied further.

Exploratory Data Analysis

During exploratory data analysis, following questions were asked:

1. What are the most important features affecting listing price?
2. Does location/neighborhood affect listing price?
3. If location/neighborhood affects listing price, what made these neighborhoods special? and ask follow-up questions if needed.

Initial Findings

1. As expected, listing price is correlated to accommodates, bedrooms, bathroom, neighborhood, and property type.
 - Interestingly, some neighborhoods (like Hollywood Hills West, Beverly Hills, Bel-Air, Hollywood Hills) seem to have higher listing prices.
 - However there might be some exceptions:
 - There might be extremely high listing prices although there's only one bedroom, but rarely happened.
 - If the property has a lot more bathrooms, ex. 8, but it's a hostel, the listing price would not increase proportionally.

- If the property has a lot more bathrooms, ex. 8, but can only accommodate 2 people, the listing price would not increase proportionally, either.
 - If the property is a villa which can accommodate 14 or more people but its accommodates vs. bedrooms ratio and/or accommodates vs. bathrooms ratio are higher, that means more travelers need to share the bedrooms / bathrooms, the listing price would not increase proportionally, either.
2. Property-type-wise, most of the property type of LA listings are house and apartment.
 3. In terms of bedrooms and bathrooms, 1 bedroom 1 bathroom Airbnbs which can accommodate 1 ~ 4 people are most common in LA.
 - About 52% of LA Airbnb listings are of this type.
 - For the rest of Airbnbs, not many of them can accommodate more than 10 people. In fact, only 2.75% of listings can accommodate more than 10 people.