Functional Specifications

Background

Homeowners and potential home buyers often rely on Zillow's Zestimate to estimate the market price of houses. The Zestimate predicts a home's value using publicly available housing data and a proprietary machine learning formula. It serves as a starting point for homeowners who are looking to sell their property and enables home buyers to gain a better understanding of the market.

Such a feature would be similarly useful to predict the market rental price of a short-term rental unit on Airbnb. A guest could compare the price of an Airbnb listing to its predicted price to know if they are getting a good deal. A host could easily determine a good price to list their unit.

User Profile

Guests:

People looking to rent an Airbnb unit. At least 18 years old. Has basic web-browsing and interface navigation experience. Potentially has no experience with short-term rentals.

Hosts:

People looking to list their home on Airbnb. At least 18 years old. Has basic web-browsing and interface navigation experience. Potentially has no experience with short-term rentals.

Data Sources

http://insideairbnb.com/get-the-data.html
Includes multiple csv files for many cities compiled by Airbnb. 'listings.csv' will be primarily used and will be combined amongst several cities. Variables of interest include neighborhood, latitude, longitude, room_type, price, minimum_nights, number_of_reviews, last_review, reviews_per_month, calculated_host_listings_count, and availability_365.

https://info.kingcounty.gov/assessor/DataDownload/default.aspx
Includes multiple csv files on property data in King County. Could potentially be used to predict rental prices for units that have never been listed on Airbnb. Feature would be limited to King County. 'parcel.csv' has data about every aspect of each property in the area.

Use Cases

1. Host wants to get a recommended listing price.
   - User clicks "Recommend a rental price". Form is displayed
   - User fills out form about unit specifications, clicks submit
   - Specifications are inputted into model
   - Page shows an interval with the predicted rental price of the unit

2. Guest wants to determine if a listing is overpriced
   - User clicks "explore map", city buttons are displayed
   - User clicks the city the listing is in, map displays
   - User enters address into search bar, map zooms in to address

- Map displays estimated rental price. A color indicating the value of the actual listing (good, average, poor) is displayed

Component Specification

Data manipulation

CreateAirbnbMatrix

Reads the listing.csv file of Airbnb data for a particular city and converts it into a matrix that can be used in predictive modelling.

Input: listing.csv file from Airbnb
Output: numerical matrix in numpy ndarray form
Pseudocode:

```
Dataframe = Read listing.csv using pandas
Categorical_columns = ['neighborhood_cleansed', 'neighborhood_group_cleansed',
'property_type', 'room_type', 'bed_type', ]
Dict_list = []
For column in Categorical_columns:
        Categories = Dataframe[column].unique()
        Dict_list.append({ i : Categories[i] for i in range(Dataframe[column].nunique())

Host_df = new dataframe with columns 'neighborhood', 'neighborhood_group', 'property_type',
'room_type', 'bed_type', 'latitude', 'longitude', 'accommodates', 'bathrooms', 'bedrooms',
'beds', one column for each amenity, 'availability_365', 'minimum_nights',
'number_of_reviews', 'price'

For column in Host_df:
  If column is numeric:
    Copy same values from Dataframe
  If column is categorical:
    Use Dict_list to convert string into ordinal int value
  If column is an amenity:
    Populate Host_df[each amenity] with 1s or 0s if they are present in Dataframe['amenities']
Host_matrix = Host_df.as_matrix()
Return Host_matrix
```

Visualized Pricing Model

User Data Entry:

User interact with an interface to select listing features. Selections will need to be saved to table (potential to see history) or captured as parameters to run through model.

For Hosts - Predicting Listing Prices:

The model will be trained with Airbnb listing data. We will look at different feature selection options, if necessary, to avoid boosting effects from highly correlated features or to eliminate low importance features. Model coefficients will be stored in a table. Users will input specifications for their listing like the number of rooms available, neighborhood, shared unit, etc. and the visualization will display the predicted price range for that listing.

For Guests - Rating Current Listings:

Using the stored model coefficients from our pricing prediction model, we will run listing data through and save predicted values to a table. We will use the deviations from our predicted price to grade each listing as good, average, or poor. User will interact with listing features like neighborhood and price range to filter which display on map.