

AirBnB Spatial Data Visualizer

Project Report

Ameya Padole
apado003@ucr.edu

Karthik Harpanahalli
ksharp009@ucr.edu

Naman Mittal
nmitt005@ucr.edu

Shreyas Ghuge
sghug001@ucr.edu

Simran Saha
ssaha031@ucr.edu

Varun Sapre
vsapr002@ucr.edu

Abstract:

Analyze open data from AirBnB to perform exploratory data analysis and overlay results on interactive maps based on user-selected filters like host and offerings. Using our custom algorithm, we rank these offerings and recommend the best Airbnb based on an analysis of reviews. The output will be an interactive map with different overlays such as the number of listings, prices, hotspots, and availability based on the user input filters.

Introduction:

Airbnb is frequently utilized by travelers looking for less expensive lodging options than those offered by hotels as well as by individuals looking for lodging for one or two nights. These users frequently receive insufficient listing information. For instance, if a user was aware of the average price per neighborhood, they might only consider homes in those areas. Visualizations are a powerful tool for explaining complex concepts to non-technical users. In order to help users choose wisely when looking at rental listings, we created user-friendly interactive maps that convey information.

In addition to visualizations, we provide customers a variety of filter options, including price, kind of apartment, number of bedrooms, and more. Users can use these filters to find residences that closely match their needs. Additionally, each of these filters returns a number of results that correspond to the consumers' preferences. Geospatial data can be visualized using a variety of technologies, including programs like ArcGIS. However, to accomplish the same, we

used Python packages like folium because we wanted to develop our own tool and hence the corresponding user interface.

Folium is a Python library that uses the leaflet.js component of the Python ecosystem to plot data on Open Street Maps. Folium offers features that enable the rendering of various plots and graphics on Open Street Maps, enabling the customization of maps.

Python-based Flask is a microweb framework. Due to the fact that it doesn't require any specific tools or libraries, it is categorized as a microframework. It lacks any components where pre-existing third-party libraries already provide common functions, such as a database abstraction layer, form validation, or other components. It was utilized in our project to create the web application.

Related Work and Taxonomy:

Papers are distinguished based on the factors considered by the authors for the Airbnb geo-spatial data in their papers and cities they have chosen. The following was obtained:

Sr. No.	Features Mentioned in Related Papers	Locations
1	Geographic, Social, and Economic indices	Austin, Los Angeles, Manhattan (New York City), New Orleans, Oakland, San Diego, San Francisco, and Seattle
2	Listing, Reviews, Calendar	New York City
3	Airbnb, peer-to-peer, spatial correlation, spatial econometrics, tourism supply	Budapest
4	Airbnb, GIS, Peer-to-peer accommodation	Los Angeles
5	Airbnb, GIS, spatial analysis	New York City, Chicago, and Los Angeles
6	Airbnb; sharing economy; price; GWR; factors	Metro Nashville, Tennessee
7	Sharing Economy, Airbnb, Socioeconomic, Spatiotemporal, Los Angeles, Spatial Autocorrelation, Regression	Spanish Mediterranean Arc cities

8	Airbnb; price determinants; touristic cities; accommodation prices; ordinary least squares; quantile regression method; sustainability; urban environment; collaborative consumption	Canary Islands
Team Project	GIS, Spatial, Airbnb, GLM, Prices, Time-series	Los Angeles, CA San Francisco, CA

Our Contribution:

Dataset:

To build the base of our airbnb-like website, we need procedures to read the raw-data from airbnb and then clean the data, process it and finally build interactive maps with key information clusters marked on the map. The data provided by Airbnb is in the form of csv files for each city. We read this and build pandas dataframes for easy manipulation and filtering of data.

Each row of the dataset contains 83 columns describing a listing which provides complete information about the property, the host, location, neighborhood, availability, reviews etc. We filtered out the most important factors for most of the users and considered them as part of the project. Complete description of the dataset used can be found here - [+ Inside Airbnb Data Dictionary](#).

Normalization and Ranking Algorithm:

We chose to design a custom ranking algorithm to show more relevant hosts in a selected geographical area. We considered all available filters and data that can give us a better insight into what a user needs. For the weights, we first normalized the reviews using min-max normalization, max normalization and sum normalization. Then we use a simple formula where we inverse the price of the property and multiply with the total review score along with the computed normalization.

- **Number of ratings** : Ratings left by patrons for a host. This field needs to be normalized.
- **Price** : Price of the property.
- **Review Score** : Review score of the property on a scale of 5.

The calculated rankings are added to the new columns in the dataframe and based on the dropdown menu selected by the user, we sort the data by the selected column. And fetch the top 10 to be displayed on the map.

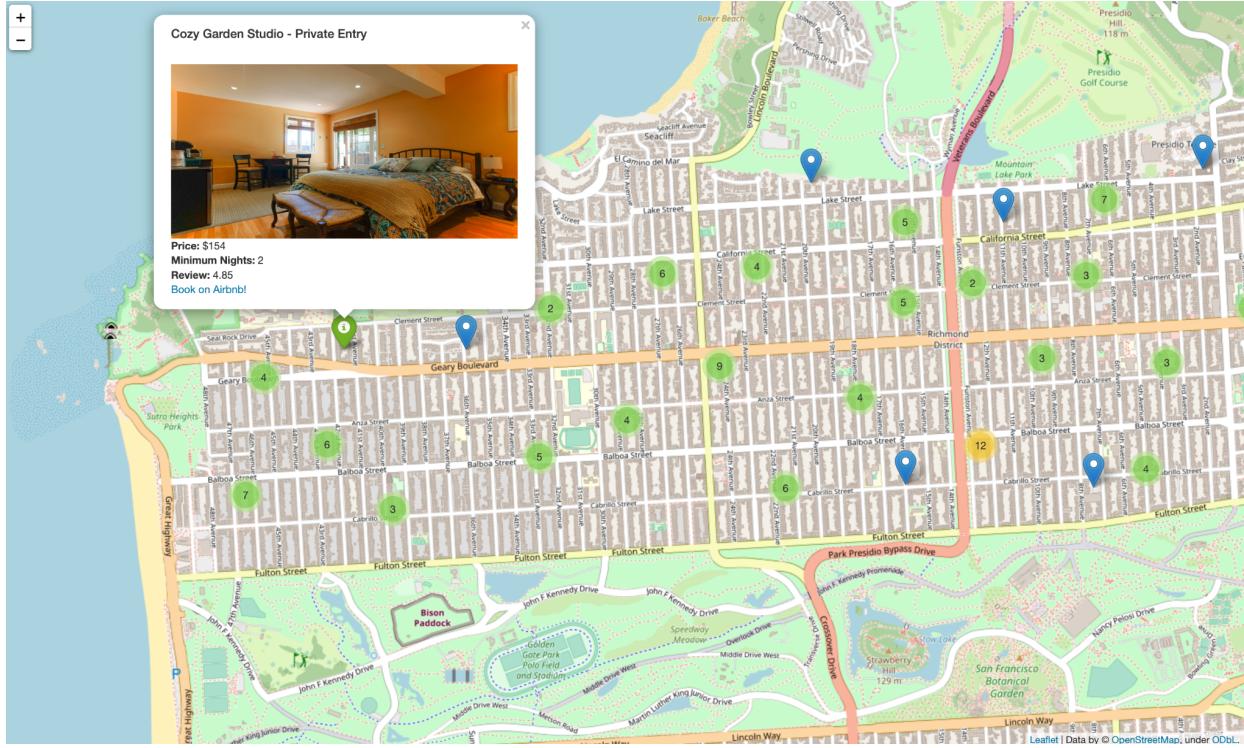
Comparing / Choosing user filters:

Our project is focused on user preferences. Any user should have a variety of options. Airbnb hosts and provides a variety of user filters to its customers. After comparing various user filters listed in our midterm report, we decided to include filters like:

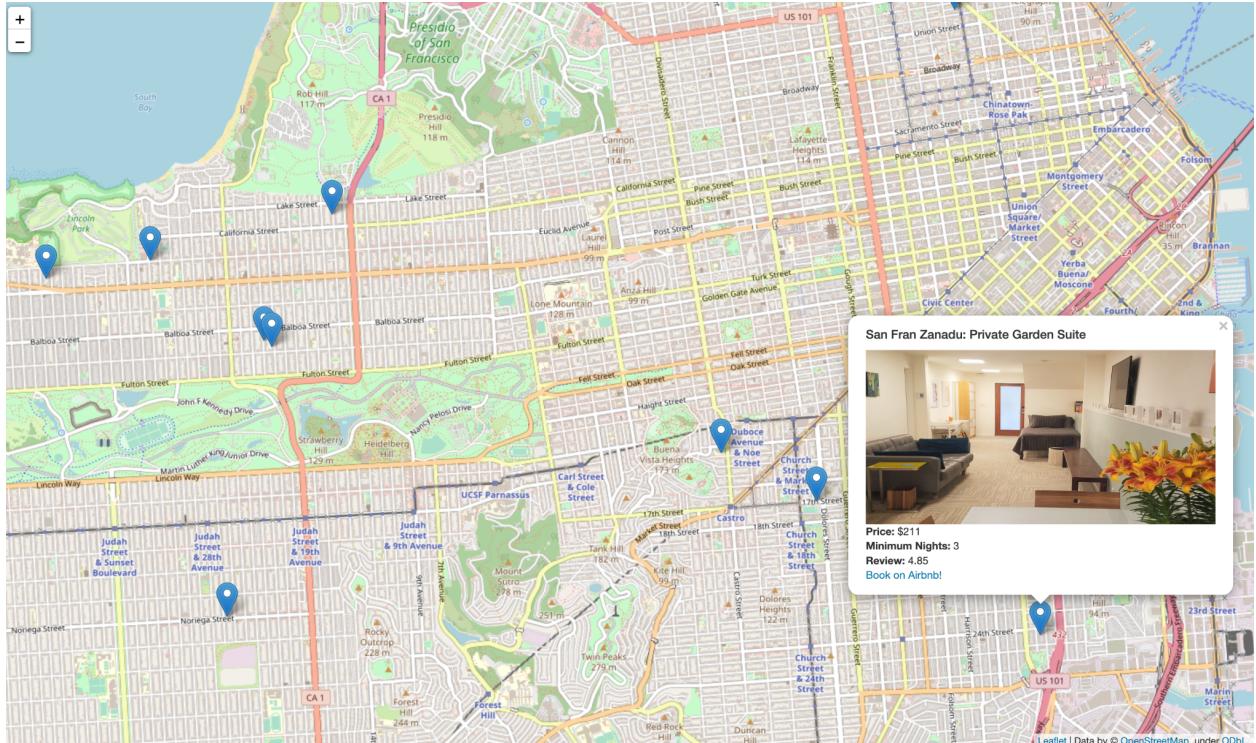
- **Room Type:** for most online hotel forums, room type is one of the most used user filters by customers. People use this filter to get the desirable type of room. For this particular project we have added multiple room types. Following are the one's we're using -
 - Private room
 - Shared hotel room
 - Entire home/apt
- Ref: <https://www.airbnb.com/help/article/479>
- **Number of nights:** The monetary model of hotels such as Airbnb is governed by how many nights a user might occupy the place. To imitate the same behavior, we also added a filter where as an end user you can choose the number by yourself.
- **Max price (per night):** While users select the number of nights they'd spend, the max price determines the total cost of their stay. They can sort and select what maximum price they'd like to spend per night. This user filter allows you to do so by selecting a range of values.
- **Min rating:** The user filter gives the user much more clarity about the place they're going to stay by choosing the minimum ratings. Users can select the minimum number of aggregated stars / rating out of 5 they'd like to see while the filter for places.

Data visualization:

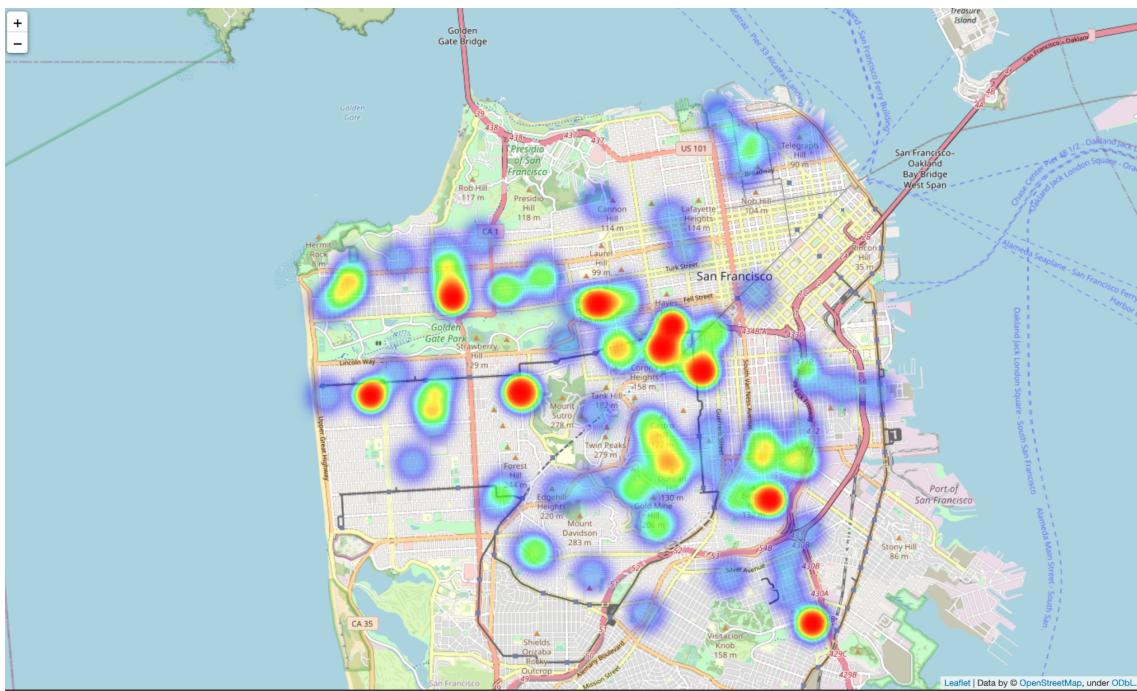
Rendering Map with Markers for all Listings of the City -



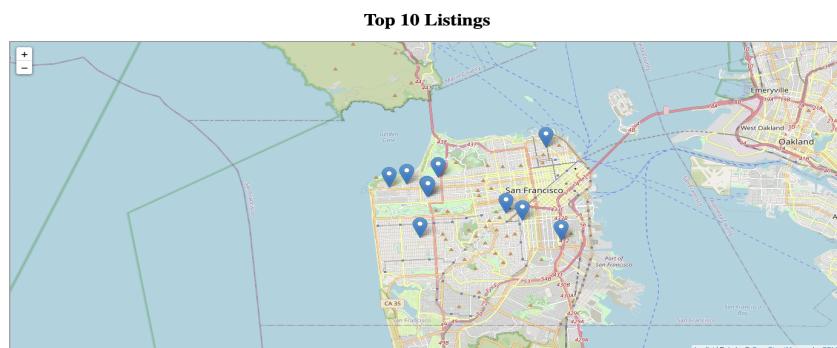
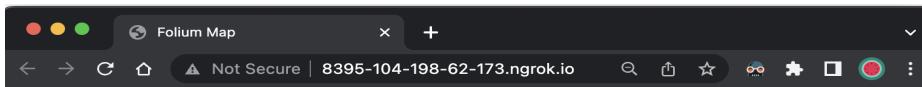
Rendering Top 10 Ranked Listing -



Rendering Heat Map for the areas in San Francisco -



Rendering Web-UI with Maps -



Heatmap



Flask application from user filters:

For the filters mentioned in the later section, we use ipython widgets to enable selection of filters and changing values on-the-go. As mentioned in the midterm report, we've put the selection of filters in the UI thus giving it a more realistic feel. After this, we fetch the values from the filters, and therefore use them to create views of the pandas dataframe. Using these views we display the output in the maps using the folium library in python.

Further, each filter change triggers a map update if the map was previously built already. Folium map gives functionality to aggregate clusters based on the zoom distance. Thus the more you zoom-in into the map, the clusters break into smaller distinct clusters.

Complete Code for our project can be found - [here](#)

Future Work:

1. The same model can be applied to various other services/ products that have a geospatial component associated with them.
2. We can fine-grain the existing model by incorporating various ranking factors such as and not limited to:
 - a. **Wishlisting pace:** We can factor in the pace at which an AirBnB property is being wishlistied.
 - b. **Social Connections:** A social network aspect can help narrow down the choices of users significantly. Say if a users' mutual connections are hosts of an AirBnB, we can list them as a close match due the mutual connections between the host and the end user.
 - c. **Safety equipment:** This filter will allow the user to list AirBnB based on the factors such as if there are fire-alarms, smoke detectors and carbon monoxide sensors present at the facility.
 - d. Other noteworthy filters that can be implemented are **Hosts' verification status, commitment rate**, etc.
3. The model can be supplemented with a rich UI to allow non-technical users to toggle between various overlays and filters.

Conclusion:

In conclusion, we have created a functional geospatial model which provides interactive map overlays using custom in-house ranking algorithms based on ratings, availability, price and review score. With comparison filters such as room type, duration of stay, price and rating.

References

1. Analyzing and predicting the spatial penetration of Airbnb in U.S. cities
<https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0156-6>
2. Exploratory Data Analysis and Visualization of Airbnb Dataset (Non-Academic - Article)
 - a. http://www.columbia.edu/~sg3637/airbnb_final_analysis.html
3. On the determinants of Airbnb location and its spatial distribution
<https://journals.sagepub.com/doi/abs/10.1177/1354816618825415>
4. The visualization of the spatiality of Airbnb in Budapest using 3-band raster representation:
https://www.researchgate.net/publication/315714846_The_visualization_of_the_spatiality_of_Airbnb_in_Budapest_using_3-band_raster_representation
5. Cities reshaped by Airbnb: A case study in New York City, Chicago, and Los Angeles
<https://journals.sagepub.com/doi/full/10.1177/0308518X19853275>
6. Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach: <https://www.mdpi.com/2071-1050/9/9/1635/pdf>
7. <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-4-W2/107/2017/isprs-annals-IV-4-W2-107-2017.pdf>
8. The What, Where, and Why of Airbnb Price Determinants:
<https://www.mdpi.com/2071-1050/10/12/4596>
9. Zervas G, Proserpio D, Byers J (2016) The rise of the sharing economy: estimating the impact of airbnb on the hotel industry. Research Paper (2013-16)