Basic Info

Project Title: DataBNB

Repository: https://github.com/electricaldata/databnb.git

Authors:

Stanislav Kuzmenko, u1201109@utah.edu (u1201109)

Elli Hudacek, u1369395@utah.edu (u1369395)

Randy Martinez, u1406143@utah.edu (u1406143)

Background and Motivation

Our team chose to explore and visualize Airbnb data because of the vast amount of publicly available information the platform generates. As young adults who enjoy traveling and often use Airbnb ourselves, we are especially interested in understanding what drives pricing, which neighborhoods and properties offer the best value, and how much reviews really matter. These questions are not only relevant to travelers seeking vacation rentals or short-term housing, but also to individuals who may be considering renting out their own place. By analyzing and visualizing patterns in price, location, reviews, and more, we aim to uncover insights that make the Airbnb marketplace more transparent and useful for both consumers and hosts.

Project Objectives

The main question for our project is: what factors contribute most to the price and quality of Airbnb listings, and how do location and reviews shape the experience for both travelers and hosts?

Some potential, more specific questions we will visualize and analyze to answer our main question are:

- How do listings and prices vary across neighborhoods?

- Which areas offer the best value for travelers?

- What types of listings dominate the market?

- Do higher review scores or a greater number of reviews correspond to higher prices?

- How far outside the city center should a traveler stay to maximize value?

- Which neighborhoods or types of rentals receive the best reviews?

There are many benefits to exploring these questions, both for travelers and hosts. This analysis and visualization will help create transparency in the Airbnb marketplace, giving travelers confidence that they are booking a quality place, especially those seeking affordable options. By examining the relationships between price, location, and reviews, travelers can identify the best balance of cost and experience. For hosts, this project offers insights into what makes a listing high-quality and competitively priced, which can help increase occupancy rates and overall success.

Data

The data for this project will come from Inside Airbnb a website that provides different datasets by Country/City as a csv download. https://insideairbnb.com/get-the-data/

Data Processing

No, we do not expect to need substantial data cleanup. After a brief review of the dataset shows that it is already well-organized and complete. The main task will be handling missing values where certain data points are not applicable. These missing values will need to be appropriately represented as nulls, zeros, or another suitable way to ensure accurate reporting and analysis. For our project objective and questions, we plan to derive quantities such as average price per neighborhood, price distribution (min, max, quartiles), overall ratings, neighborhood rating (using ratings of all homes in the area), distance of each listing to city center (possibly use Haversine formula). Potentially mess around with other simple calculations that represent quality, price, review score, and location. We will implement data processing using Python (pandas) in either Pycharm or VSCode to get the data ready for our visualizations.

Visualization Design

How will you display your data? Provide some general ideas that you have for the visualization design. Develop **three alternative prototype designs for your visualization**. Create **one final design that incorporates the best of your three designs**. Describe your designs and justify your choices of visual encodings. We recommend you use the Five Design Sheet Methodology.

Since this data is inherently location-based, a map would be the most obvious starting point. The data already corresponds to locations across the US and other countries. This would make it easy to begin with a heat map where the most popular Airbnb destinations would stand out.
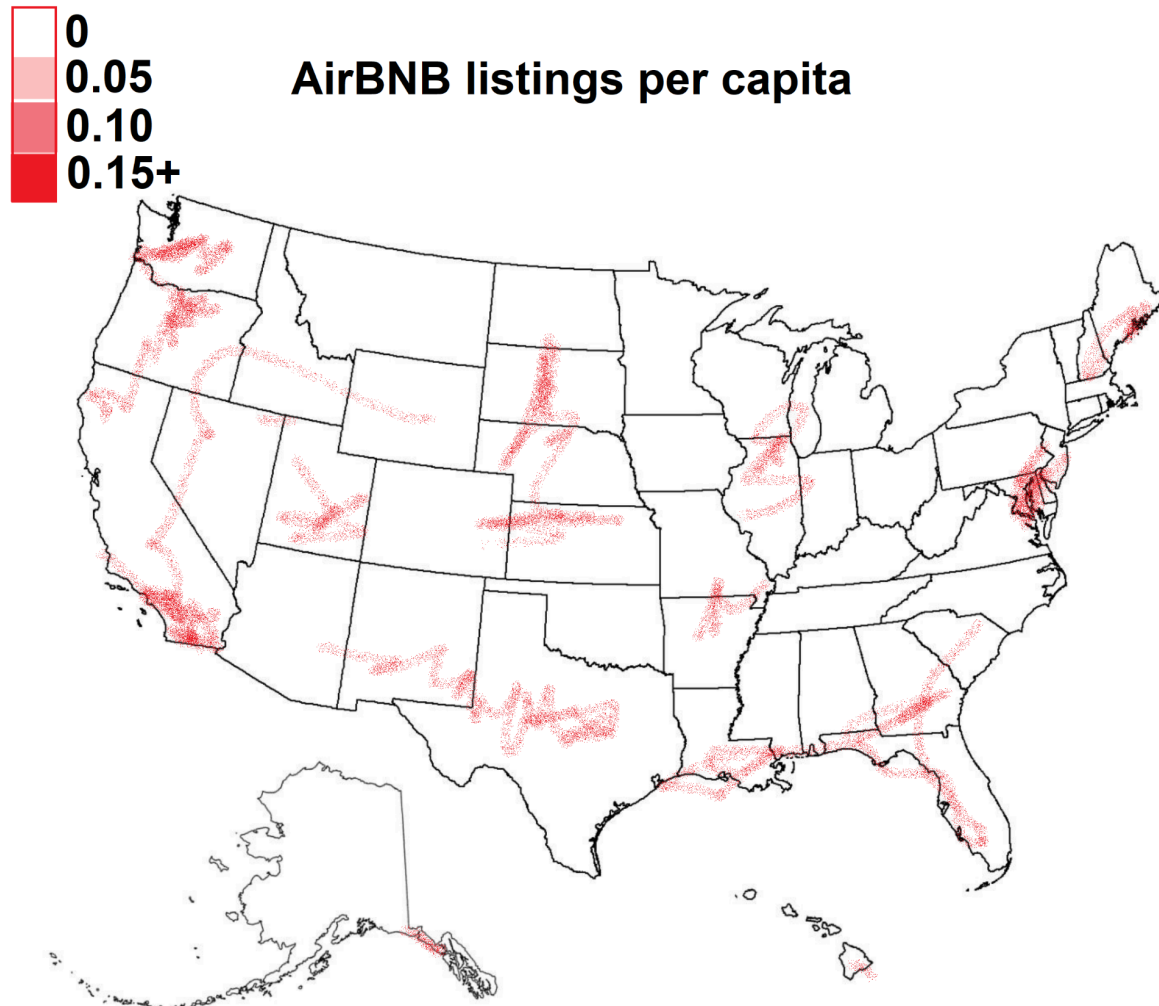
This map may show some overlap with popular tourist destinations, states with more relaxed zoning laws, or other trends. It would allow zooming into states or popular cities for which we have data. However, it would only answer a few of our targeted questions.

Another possible map would be that of user reviews. We could employ a two-color system to describe this, using blue for positive reviews (4-5 stars) and red for negative ones (1-2 stars). A benefit of this is that we could reserve the value channel for review count. Showing locations with many reviews in stronger shades than those with few or no reviews.

Average AirBNB Review by Region

Legend:
- low score (1-2 stars)
- medium score 3stars
- high score (4-5 stars

many reviews <> few reviews
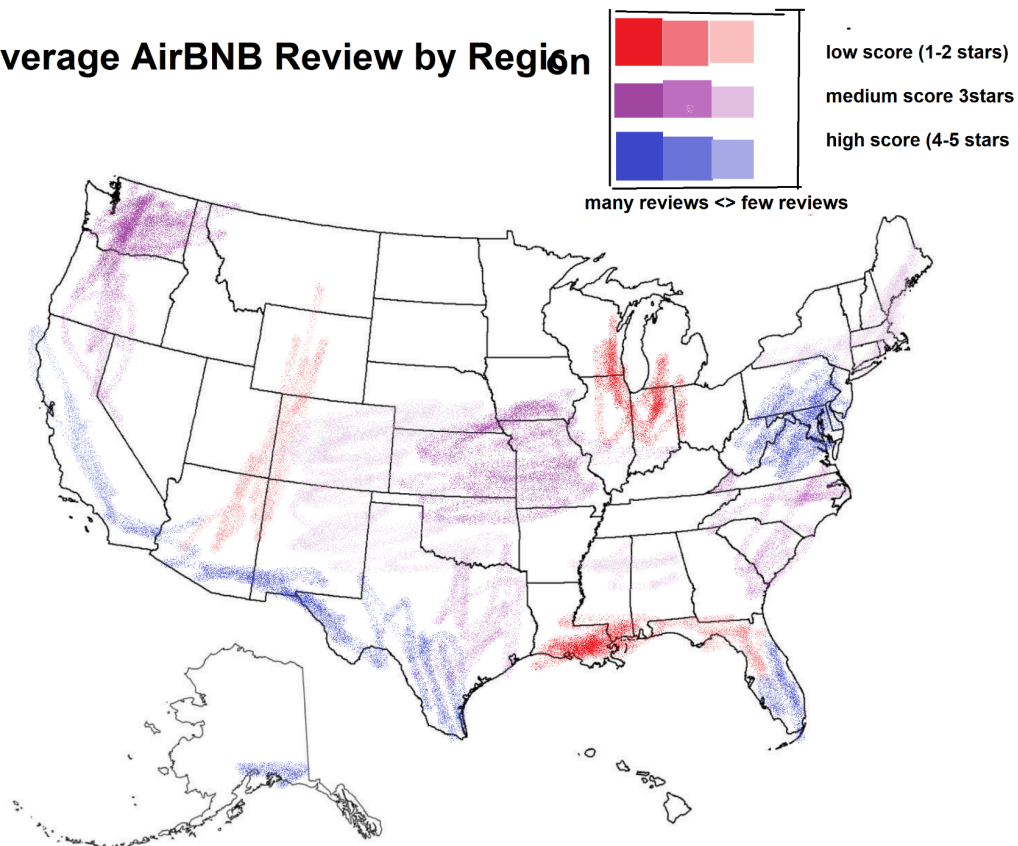
This would lead to a slightly more complicated map (information-wise) but would be more useful to people who want to visit regions with strong Airbnb listings. This map, too, could have zoom functionality for specific areas, cities, or states.

Average AirBNB Review by Region
in Washington (state)

low score (1-2 stars)

medium score 3stars

high score (4-5 stars
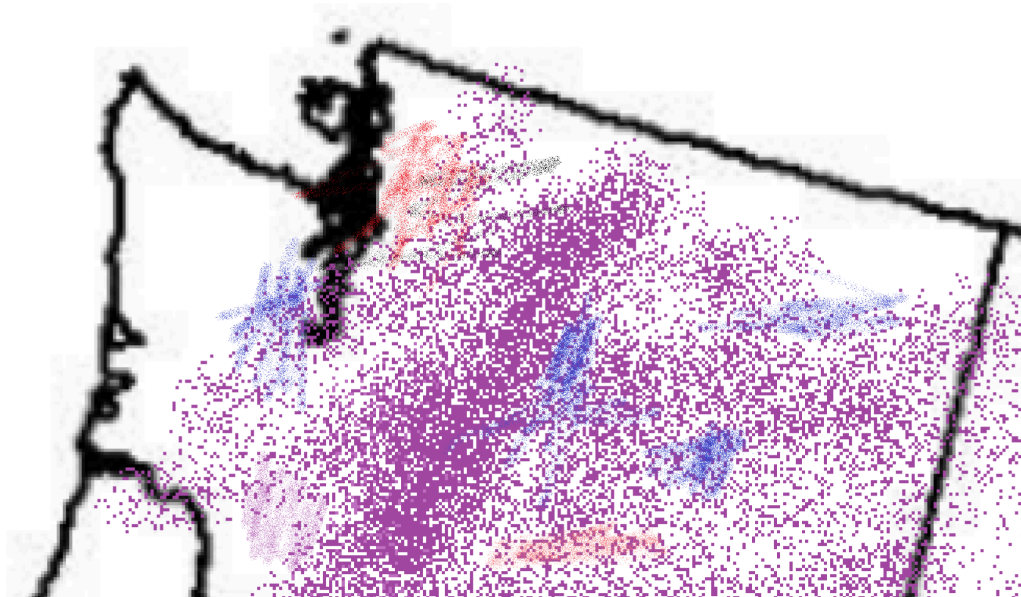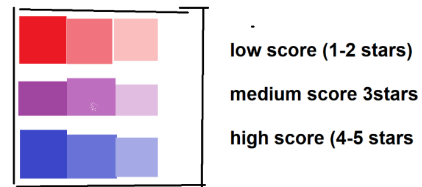
Another important question that the previous maps haven't answered is how price correlates to review score, review count, listing type, and possibly even time of the year. To account for these, we could display how each attribute matches up to price using bar charts, histograms, or even correlation plots.

AirBNB prices in the US depending on their review score
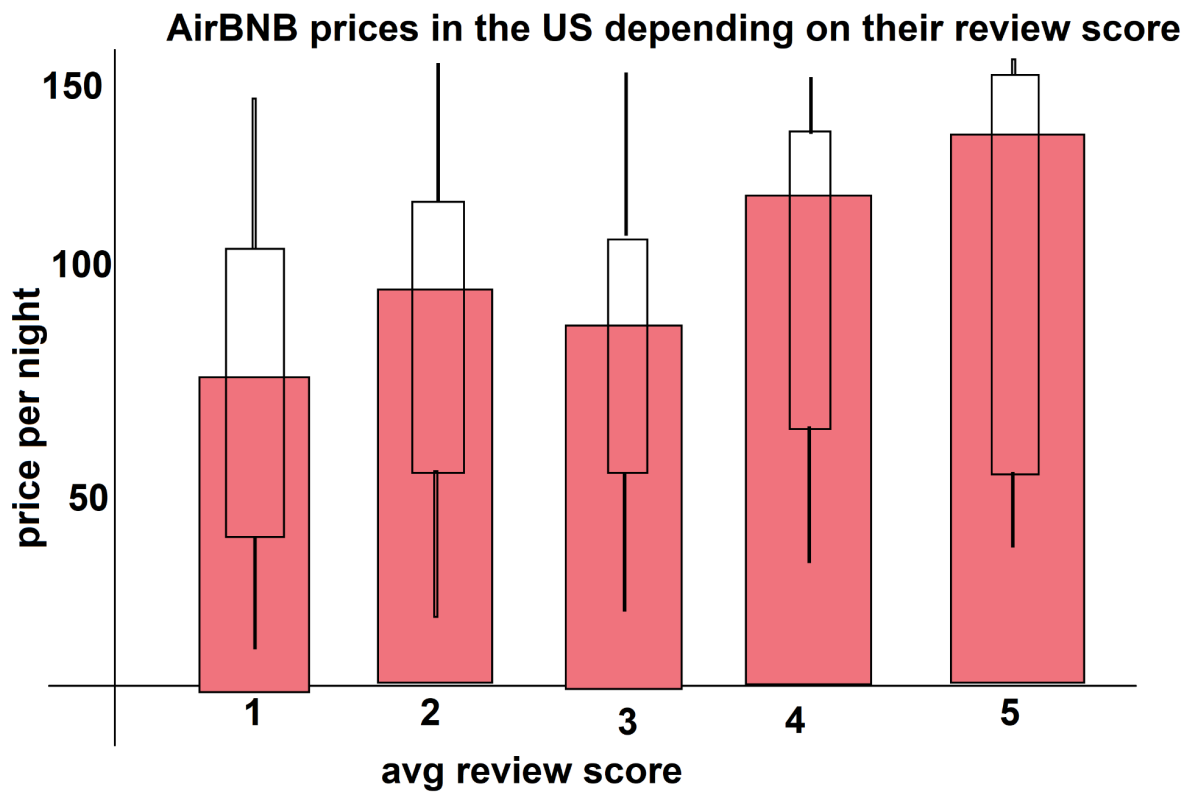
We could even include or overlap boxplots to show how much each category or group varied in price.

Nightly AirBNB price trend over a year

price per night

mar   jun   sep   dec

time of year
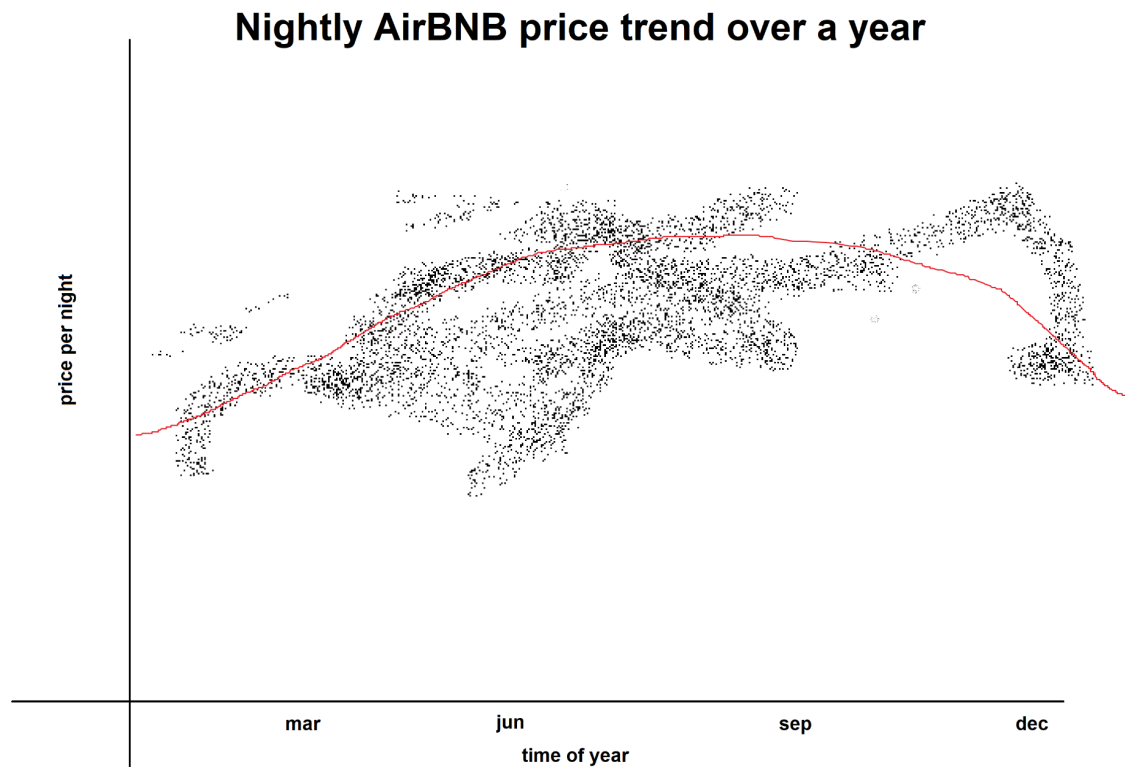
This type of graph would show more data points but have an overarching trendline estimated by an algorithm.

Overall, the final model may be a mix of all of these - we may have a tab with a map, which you can click on to see the review scores in a particular region, and check a sidebar to see what the minimum and maximum prices are in a particular year.

Average AirBNB Review by Region
low score (1-2 stars)
medium score 3stars
high score (4-5 stars)
many reviews <> few reviews

Average AirBNB Review by Region
in Washington (state)
low score (1-2 stars)
medium score 3stars
high score (4-5 stars)

AirBNB prices in the US depending on their review score

Nightly AirBNB price trend over a year

AirBNB prices in WA depending on their review score

Nightly AirBNB price trend in WA over a year

Must-Have Features

- Map of listing with pricing and review information:
  Each Airbnb listing will represent a point on a map. We will encode features such as pricing and review info into each point. We will explore several avenues to accomplish this, such as using a color gradient, varying size, or changing shape. Making this a must-have feature allows users to see the effects of pricing and review how their geographical behavior changes. This feature will also bring insight into knowledge about quality and affordability.

- A choropleth map for different neighborhoods:
  Having a color-coded choropleth map that shows some summary statistics. As with our first feature, we will manipulate color, size, and shape to accomplish this feature. A chronopleth map will allow users to identify the neighborhoods or different zip codes that offer the best prices and reviews. Users can also analyze the summary statistics to make a better decision about the Airbnb they choose to stay at.

- Panel that allows filtering:
  Having a panel for filtering is a must-have filter because it allows users to navigate throughout graphics to get their different questions answered. If we don't have a panel, we would need to have all the graphics on one page, including all our insight and findings, which would overwhelm the user.

- Zooming in at a specific location:

This feature will allow users to narrow down their search to specific neighborhoods or areas. Having this feature will improve clarity and interpretation of the graphics. Having this feature will also allow users to compare nearby neighborhoods to their search.

Optional Features

- Seasonal comparison logic:
  We could have a time slider and compare how the data changes as you slide the time slider. This would require time series analysis for the data set. We could color-code this as well, but the time series analysis will help us provide a better insight into the color coding. This type of feature is complex but would bring more seasonal insight.
- Best value:
  We could make a recommendation and determine which are the best values. In order for us to determine the best values, we must model and feature engineer the required features to come to the conclusion on which are the best values. Our best value recommendations may be incorrect, which could cause issues for users.
- Text review sentiment analysis:
  Having this feature can help users make a decision based on sentiment from other users rather than a data-based approach. Some users prefer hearing from other humans rather than the actual numbers of the data.

Project Schedule.

Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

When	What	Who
Week 4	**Project Proposal**	All
Week 5/6	**Project Review**	All
Week 6/7/8	Data Collection Data Cleaning Data Analysis Site Creation	Elli Elli/Randy All Stas/Randy
Week 9	Visualization Prototype	All
Week 10	**Project Milestone**	All

Week 11/12/13	Final Data Check Site Tweaking Visualization Enhancement	All Stas All
Week 14	**Screencast Presentation**	All
Week 15 (Thanksgiving)	Final Touches	All
Week 16	**Final Submission** **Feedback**	All All